**Getting Started with SAS**

How To Get Started

The amount of statistics you need to run analysis on a basic research project is not overly complicated. As all med students are required to understand and interpret statistics for step 1, 2, and 3 and all orthopedic residents are required to interpret research papers for OITE, pretty much everyone that is reading this document understands the basics of statistical theory. For that reason, it should not be too difficult to apply your statistics knowledge to real-life analysis of a dataset.

If you dedicate yourself for 1-2 days on statistics and learning how to use these simple tools, I promise it will make research much easier and make the research rotation much easier. I use SAS because it is available for free for students and there is a ton of support for it online.

Do the following to get into the SAS ecosystem:

1. Go here: <https://welcome.oda.sas.com/login>
2. Click on “*Don’t have a SAS Profile?*” and create an account. Make sure to use your .edu email address (so your @eid.utexas.edu email address)
3. Go to your email and verify your email
4. Go back to <https://welcome.oda.sas.com/login> and login with your credentials
5. You will get to a screen like this:

A screenshot of a computer

Description automatically generated

Click on SAS Studio to go to the virtual environment

1. You are then in a virtual SAS environment. Essentially think of this is a place where you can upload files, create SAS programs and see the output of their programs. The good thing about this software is it is all online, your files are saved in the cloud and you don’t need to download anything on your own computer. Once you get the hang of where everything is, it is relatively easy to use

Graphical user interface, text, application

Description automatically generated

* **Green:** This box is essentially the folders that you have in your personal cloud, essentially think of it as your dropbox folder or google drive where you can save your programs and upload your datasets. You can create folders here and move programs/datasets around in these folders
* **Red:** click this to create a new file, such as a program or folder
* **Blue:** click this to upload a file from your computer to the cloud. You won’t be able to click this unless you are clicked on a folder in your green box. So if you want to add a file in your Files (Home) folder, click the folder in the green box and then press the blue button and go through the steps to upload a file
* **Black:** This is the window where you can edit and run programs. There are 3 windows you can see in your program – Code (which it is on now), Log (in orange, which will show you errors in your program when you run it), and Results (in grey, which will show you the results of your program when you run it)

SAS Basics

So now comes the programming part. Programming is essentially a series of steps where you tell a calculator to perform a task. Each of these “tasks” require a command. At a high level, you can break your program into a series of commands/tasks in sequence. So what do we need to do and what are the associated commands?

1. We need to import our excel file so that the program can analyze it 🡪 *proc import*
2. We then need to see the distribution of data to see if there are any missing variables and patterns in your variables 🡪 *proc univariate*
3. We need to edit the data. Let’s say we want to analyze the effect of age on risk of cancer and we want to see if those of age >45 years old have a higher risk than those <45 years old. We need to create a new variable for each patient saying are they older than 45 or are they not. We need to do that by editing the data which is relatively easy in SAS. We just use the *data* procedure.
4. Conduct bivariate analysis 🡪 variety of procedures based on the test you want to run, *proc ttest* for t-tests, *proc ANOVA* for ANOVA, *proc freq*for chi-square, and *proc corr* for correlation
5. Conduct linear regression 🡪 *proc reg*
6. Conduct logistic regression 🡪 *proc logistic*

This is the essentials of what we are going to be doing. Taking a dataset, editing it, and then running tests on it. Once you figure out how to speak the SAS language, it is easy to replicate.

SAS Syntax

* If you ever want to make a comment in SAS that will not be read by the program but is just there in case you want to annotate what you are doing, start with a slash button and an asterix so “/\*” and make your comment and then end with an asterix and another slash button “\*/”
* Every line in SAS should be ended with a semicolon. If you don’t end a line with a semicolon it will assume the next line is part of that same line until there is a semicolon

Essentially there are two types of commands: A data command OR a procedure command.

* A data command allows you to edit a dataset that you have uploaded. This is the syntax:

data NEWDATASET;

set OLDDATASET;

[insert actions which edit the dataset]

run;

* So here you are making a newdataset from an old dataset. The actions that you can do include creating new variables based off of old variables, abstracting data from variables (such as year from a date variable), editing variables (lets say you only want the first 3 letters from a data point like only the first 3 characters of a license plate)
* You need to end this data step with a line that says “run;”
* A procedure command allows you to upload or analyze a dataset. This is the syntax:

Proc PROCEDURENAME data = DATASET-TO-ANALYZE;

[insert which variables you want to analyze and what you want to do]

run;

* If you notice, the first line of a procedure command is different than the data command. Instead of “setting” it as an old dataset in the next line, you have to assign it to a dataset within the same line.
* The meat of the procedure is after this line but it varies depending on what you are doing.
* It ends with the same “run;” line

If you want to read more about SAS syntax or have a cheatsheet, this is a good PDF link: http://ivyproschool.com/blog/wp-content/uploads/2015/08/SAS-CHEAT-SHEET.pdf