Class Exercise 1

This exercise is based upon Chapter 1 of the SAS Certification Prep Guide. We will learn how to perform simple queries using the SQL procedure. Be sure to include portions of the output to support your answers.

Let’s read in some admissions data (the Fall 09 follow-up to the ten percent subsample of Fall 08 admissions). The enrollment files are password-protected; I will include the password in the assignment posting. I recommend that you save a non-password protected version by pasting the columns in a new workbook and saving the workbook (a tab-delimited text file works too). Upload the file to your course folder in SAS Studio, then import it (name it **sample**) and run the following commands for an initial look at the data.

Note the order of the columns in the output will match the order in which the columns appeared in the Excel file.

proc sql outobs=10;

select \*

from sample;

quit;

We could also list all the columns in the **select** clause. Note the order of the columns in the output will match the order specified in the **select** clause.

proc sql outobs=10;

select class, degree, cltotgpa, instcd, gender, country, stcd, cntyid, satv, satm, regstat, enroll, ipeds, housing

from sample;

quit;

Let’s limit the columns to d**egree**, **stcd**, **housing**, and **cltotgpa**. Let’s also subset the data to students whose **class** is equal to ‘Junior’.

proc sql;

select degree, stcd, housing, cltotgpa

from sample

where class= 'Junior ';

quit;

We can also create a new column within the **select** clause. Let’s create a new column named **satt**, which will be the sum of the **satv** and **satm** columns. SAS may import the satm and satv columns as character variables. For this clause to work correctly in that case, you may need to change the missing values in **satm** and **satv** prior to import, or you can use a **data** step to convert **satm** and **satv** to numeric variables (e.g., **satm\_new=input(satm,3.); drop satm; rename satm\_new=satm;**)—no changes are needed provided **satm** and **satv** are imported as numeric variables.

proc sql outobs=10;

select degree, stcd, housing, cltotgpa, sum(satm, satv) as satt

from sample

where class= 'Junior ';

quit;

The new column will appear in the output, but will not be kept unless a table is created; we suppress outobs=10 since we are creating a new data set.

proc sql;

create table sample2 as

select degree, stcd, housing, cltotgpa, sum(satm, satv) as satt

from sample

where class= 'Junior ';

quit;

*Graduate students: Recreate sample2 (including the construction of satt, and the restriction to juniors) using regular data step statements. Print the data set (do not use a* ***where*** *statement in proc print).*

We can use the **order by** statement to sort the data in the table **sample2** by the new variable we created. The code below sorts the output by **satt** from highest to lowest. Explain why the **proc sql** view is sorted, but the **proc print** output is not sorted. Remove the word **desc** to sort the data lowest to highest.

proc sql outobs=10;

select degree, stcd, housing, cltotgpa, satt

from sample2

order by satt desc;

quit;

proc print data=sample2 (obs=10);

run;