Stat 657
Assignment 05 - SAS

Scope:

This assignment will primarily utilize, but is not limited to, techniques covered in lectures 5 through 7.

Specific Instructions for this Assignment:

1. Download the **scholarship03** data set from the Assignment Data Files section on eCampus to the homework folder that you are using as a permanent library on your computer. This data set contains demographic, scoring, and scholarship data on many of the players who were in the 2003 NCAA Men's Basketball Championship Tournament.

   Assign a libref to your homework data folder. Include the readonly option on the libname statement. Assign a fileref for your PDF output file.

   Turn off page numbering for any output produced by this program.

2. Most of the SQL examples in the lectures only send output to the open ODS destinations. However, we briefly covered a method of writing output to a table using a CREATE statement ahead of the usual SELECT statement. Write a complete PROC SQL step that will create a table as described below:

   a. This completed step will create a data set called **scoring03** in the work library.
   b. The new data set will include the following columns: player, team, region, and PPG from the **scholarship03** dataset that you downloaded.
   c. You will need to compute the overall average PPG of players in the tournament. Remove those whose team has a seed of 15 or 16. You will need to pay close attention to the name and label of this column when writing your code. Construct your query so that this "overall" average PPG value is remerged into the original data as a fifth column.

3. Use a SINGLE proc sql step to create the two reports shown in the output PDF posted on eCampus. Use the Minimal style for this assignment. You will be doing a significant amount of PDF control in this assignment. There is a link under Links on eCampus for a paper entitled "Funny ^Stuff~ in My Code: Using ODS ESCAPECHAR" that will help you with this. You may also find it beneficial to search SAS Help for ODS ESCAPECHAR Statement. Suppress page breaks on the PDF output so that the second report starts on the first page. Since the Title area is only at the top of a page, in addition to the title statement for page 2 you will need to use another ODS option to insert the title text when the second report begins on page 1. (Options listed on the ODS Tip Sheet included in the Course Materials section on eCampus will be used). One of these

options will get the title on the page but it will not be centered. There is information in SAS Help under SAS Products –Base SAS – SAS 9.4 Output Delivery System User's Guide – ODS Statements – Dictionary of ODS Language Statements – ODS ESCAPECHAR Statement. You will need to use an escapechar in this assignment. Pay close attention to the section about the STYLE function. In that Help section there is a link to "Detailed Information for All Style Attributes". You will need to use a style element to align the text of your second "title" line. You can also use this same type of statement with the escapechar and a newline command to add blank lines between the top report and the second "title". It is recommended that you complete the rest of the assignment and then come back to the PDF formatting after everything else is working as desired.

The first report, entitled "Average Scholarships for State Schools", is based on the **scholarship03** dataset downloaded from eCampus.

a. We could use a "contains" operator to subset our data but we want to practice using the find function. Use a single find function in your where clause to subset the data so the report includes only those students from teams with State or the abbreviation St in the team name. St.Josephs is not one of those schools. Construct the find function so that it does not pick up unwanted schools like St.Josephs.
b. The columns in the report are Player, Team, the total scholarship for each player based on the values in amt1 through amt10, the maximum single scholarship per player based on those same 10 values, and the number of scholarships received by each player.
c. Eliminate players who only received a single scholarship.
d. Order the report by team and then by the largest total scholarship to the least within the team.

4. The second report, entitled "2003 NCAA Team Scoring Analysis", is based on the **scoring03** table created above.
    a. There must be only one row for each team included in the report. This must be done WITHOUT remerging summary statistics back in with the original data and WITHOUT using the DISTINCT (or UNIQUE) keyword.
    b. In addition to the team name, the report shows the number of players on the team roster in this data set, the average PPG (points per game) of the players on the team, and a column that represents the team average PPG as a percentage of the overall average (avg_ppg/avg_PPG_all).
    c. Create a final column, labeled PPG Level, with a value of "Above Avg." when the team average PPG is greater than the overall average. Otherwise, the column has a value of "Avg. or Below".
    d. Since some teams in the data set did not include a complete roster, we only want to include teams that had 5 or more players in the data set.
    e. Arrange the output by average team PPG from highest to lowest.
5. Your output must match the output posted on eCampus including page headers, numbers, titles, labels and data formats. Conclude your program with good housekeeping steps. Submit your 3 files in PDF format to WebAssign.