Stat 657
Assignment 04 - SAS

Scope:

This assignment will primarily utilize, but is not limited to, techniques covered in the first 5 lectures. You may need to refer to SAS Help for more information on the commands required to complete this assignment.

Specific Instructions for this Assignment:

Download the three data sets, ncaam03, ncaam04, and ncaam06 from the Assignment Data Files section of eCampus to the folder on your computer that you are using as the homework library for this class. These data sets contain selected information on some of the top teams who played in the NCAA Men's Basketball Tournament in a specific year. Become familiar with the structure and contents of these datasets. It is up to you to decide if and when these data sets need to be sorted to accomplish the tasks requested below. Whenever sorting is required, you must create a new temporary data set that is a sorted copy, leaving the original data set intact. In the instructions below, any reference to a data set such as ncaam03 may refer to the original data set or the sorted copy. Again, it is up to you, the programmer, to determine which is appropriate. You may discover some dirty data where teams like Boston College are listed differently in the data sets. Do **not** bother to clean this up for this assignment

1. Assign a libref that points to the folder that contains the data being used for homework assignments in this class. Set the readonly option on this libref to protect it from accidentally overwriting data. Assign a fileref for the PDF file that will contain the output for this assignment. The name of the actual file must follow the naming conventions established in the previous assignment.
2. Open a PDF destination to capture the output from the procedures that follow. Create the PDF bookmarks on the PDF file but do not show them by default. Choose one of the available styles that incorporates color and apply it to your pdf output. As a result, the pagination of your output may not match the sample. This is acceptable as long as the titles, headings, etc. either match or conform to the instructions.
3. Concatenate ncaam03 and ncaam04.
4. Use SQL to print the data portion of the new data set. Title and footnotes must match the output on eCampus including skipped lines between titles. (NOTE: Titles and footnotes are created using the same rules as with other procedures. Define titles and footnotes before or within the SQL procedures.)
5. Interleave ncaam03 and ncaam04 based on player name and team. Interleaving was not covered in the lectures. It is similar to concatenation except it also utilizes a by statement. It is possible that two players with a common name may be in the tournament representing two different teams. It is also possible that a player may have transferred to a different school that also made the tournament in 2006. Use of both the player name and team throughout the remainder of this assignment will help ensure that we do not mistakenly treat these scenarios as a single player. Eliminate any variable that does not appear in both data sets.
6. Use the Print procedure and a data set option to print the first 30 records of interleaved data set.
7. Use the Match Merge process to create a data set of only those who played in both 2003 and 2004 tournaments. Again exclude any variables that are not in both data sets.
8. Use SQL to print Player, Team, and PPG from the merged data. The SQL statement must sort the list by descending PPG. The title and footnote shown in the sample output are only place holders for the actual title and footnote that you will use in your solution. Since PPG is in both data sets, your title must specify from which year the PPG value was taken. The footnote must provide a very brief explanation of why this year was used for PPG. (You may need to do some research to get the answer to this. To save you some time, I have created a link in eCampus to a paper that provides the answer.)

9. Match Merge ncaam03, ncaam04, and ncaam06 into a single data set. (NOTE: When I merge data sets, I start with the oldest and go to the newest when I list the data sets on my merge statement.) Use data set options as needed so that the resulting data set has the variables player, team, ppg2003, ppg2004, and ppg2006. As the names imply, the yearly ppg variables are taken from the respective yearly data sets. Use a length statement to set the length of the player variable to match the longest size of the player variables from the three data sets. Use what you know about how SAS creates the PDV to decide where in the data step to place the length statement.
10. Create a SQL procedure with multiple statements. The first statement must write to the log a list of the columns and their attributes from the data set created in the previous step. The second statement will print the data portion of the data set with the columns in the order shown. Use this statement to assign the column labels shown in the sample output.
11. Use a single PROC step to print the descriptor portion of all data sets in your work library.
12. End the program with the appropriate housekeeping steps to ensure titles and footnotes do not carry over to the next program that is executed. As a reminder, this and all future assignments must contain the appropriate header and documentation within the program.
13. Convert your program and log to PDF using the naming convention established in the previous procedure. Submit these files and the ODS output to WebAssign.