Stat 657
Assignment 07 - SAS

Scope:

This assignment will primarily utilize, but is not limited to, techniques covered in lectures 5 through 10. You will practicing the use of the SQL set operators but may need to use joins, subqueries, and/or inline views in conjunction with the set operators in each of your SQL procedures.

Specific Instructions for this Assignment:

This assignment uses the three datasets; ncaam03, ncaam04 and ncaam06 that were downloaded earlier in the semester. These three data sets contain information about men's college basketball teams who were in the NCAA national championship tournament. The name of the data set corresponds to the year of the tournament. The LIBNAME statement used to access these data sets must be set to readonly to preserve permanent data. REMINDER: There are some pesky little differences in the layout of these three tables that can cause you unnecessary delays if you are not paying attention to detail. We know that some of the team names were entered inconsistently but we are going to disregard that for this assignment. You are **not** to clean up the data.

You are allowed to use hard-coded values to create columns that are not in the original data. This assignment is to be done using only SQL procedures. All output is written to PDF. Note that the date does not appear on the output until the second part of the assignment. The PDF output fiel contains no bookmarks. Use a filename statement to specify the name and location of the output file. No data sets are created during this assignment.

1. Create a report that combines the 2003 and 2004 statistics of only those players who played in both the 2003 and 2004 NCAA Championship tournaments. Match the tables on both the player name and team so we do not inadvertently combine data from different teams. The output must contain the year, team, the team's seed that year, player and player's points per game (PPG) that year. Sort the output by player and then by descending PPG. <u>The successful solution will use at least one SQL set operator.</u> HINT: Do not try to absorb the entire problem at once. Work on your code in segments. Start with code that will identify players who were in both tournaments. (TIP: Sometimes an inner join with its join criteria is sufficient to identify the records that are common members of two or more tables.) Then add that to code that will extract the data you need from a single year. Finally, add the code that combines the two years together. There are a number of acceptable methods to arrive at the desired output. In addition to the set operator, successful solutions MAY include one or more joins, subqueries, and/or in-line views. (You are not required to use all of these in your solution.)
2. Create a report of teams who played in all three of the tournaments for which data were provided. Include the team name, the team's seed each year, the team's average PPG of all the

players on their roster each year, and the year. Sort the output by team and year. The approach to solving this problem is similar to Assignment 04 except you are using SQL tools this time. Use the INTERSECT operator to determine which schools were in all three tournaments.

3. Three PDF files must be uploaded to WebAssign. Convert your program and the SAS log to PDF files. The third file will contain all the requested output from ODS PDF. The program must contain a completed header block and comment blocks for each step. Except for style, options must be set so the output looks like that in the Assignment output posted on eCampus.