STAT 475 FINAL EXAMINATION

*Due on Thursday, December 12, by 10PM, sent electronically*

*in a single pdf file to stat475f24@mail.com*

**PERSONALIZE YOUR CODES:** In the problems below, all outputs must contain titles with your first name, last name, and date and time. You will need to use macro variables to achieve that.

**Problem 1. (Use PROC SQL wherever possible).** Use the SAS dataset *snowstationsdata.sas7bdat*, to answer the following questions:

1. How many lines does the file contain?

2. What are the variable names in the imported file? (Hint: use proc contents).

3. How many different snow stations are the data given for?

4. How many years have been the data recorded for each station?

5. Make an ordered list of years and give the number of stations that had measurements that

year.

6. Compute the highest value of max\_SWE for each station. (Note: max\_SWE=maximum

annual Snow Water Equivalent). Order by station name.

7. You might have noticed by now that a station name ‘ADIN MOUNTAIN’ had been

misspelled as ‘ADIN MOUTAIN’. Fix this typo.

8. How many different regions are involved in snow data collection?

9. How many stations are there in each region?

10. Create a data set that contains variables STATION\_ID, STATION\_NAME,

TOTAL\_MAX\_SWE, REGION, ELEVATION, LATITUDE, and LONGITUDE, and

which entries are the snow stations that had been in operation for the most recent 50

years (from 1963 to 2012). That is, you have to take the stations that have records from

1963 to 2012, and compute the highest max\_SWE values (call the TOTAL\_MAX\_SWE)

over the duration of these 50 years.

**Problem 2. (Don’t use PROC SQL anywhere).** Investigators surveyed 347 schools in several school districts. Data set surveyresults.sas7bdat contains the results of this survey. Questions 1, 2, 4, 5, 6, and 7 were answered on a five-point Likert scale (0 through 4). Questions 3a-3h and 8a-8e were yes/no (1/0) questions. Note that Question 2 had three missing responses. The investigators were interested in how a school’s response to this survey relates to its Academic Performance Index (API). They procured publically available API information for all schools in that region (file api.sas7bdat). Answer the following questions:

1. Add the column of API scores to the surveyresults.sas7bdat file. Hint: merge the two files on SchoolName and remove all entries for which q1 values are missing. Double-check that your resulting file contains the 347 schools.
2. Impute the three missing values for q2 by the mean of the entire column. Hint: you should compute that mean and make it a global macro variable.
3. Compute the total of the survey results (variable “Total”). The total, in this case, is defined as the sum of:

*q1,*

*q2,*

*the mean of q3a through q3h,*

*reverse coded q4,*

*q5,*

*q6,*

*reverse coded q7,*

*and the mean of q8a through q8e.*

A reverse-coded variable is defined by recoding 0 into 4, 1 into 3, 2 into 2, 3 into 1, and 4 into 0. Hint: reverse coding can be accomplished with one simple arithmetic operation.

1. Plot Total versus API. Present the graph.
2. Print the school name, API, and Total for the top ten schools with the highest APIs. The variable Total should be formatted to show three decimal places.

**Problem 3.** (**Use PROC SQL everywhere**). The data set “NBA\_roster.csv” contains the player’s name, team abbreviation, age, height, weight, and season. Import the data into SAS and answer the following questions:

1. For each season, identify the youngest player(s) and output season, name, team abbreviation, and age.
2. For each season, identify the oldest player(s) and output season, name, team abbreviation, and age.
3. For each season identify the shortest player(s) and output season, name, team abbreviation, and height.
4. For each season identify the tallest player(s) and output season, name, team abbreviation, and height.
5. For each season identify the lightest player(s) and output season, name, team abbreviation, and weight.
6. For each season identify the heaviest player(s) and output season, name, team abbreviation, and weight.