Stat 604
Assignment 14 – SAS

This assignment reinforces the concepts covered in lessons SAS01 through SAS14. You will create and apply a user defined format. You will create a data set by reading in a raw data file.

Download the OKSchools.csv file from the **Assignment Data Files** section on eCampus to a folder on your computer for use in this assignment. Perform each of the exercises listed below. To the extent you have been taught to control it, your output should match that in the PDF file posted on eCampus.

The OKSchools.csv file contains selected enrollment information about schools in the state of Oklahoma. Prior to writing your code, be sure to open the file in a text editor and become familiar with the contents. The first line of the OKSchools.csv file contains the names of the variables. Use notepad, Text Wrangler, or some other text editing program to open this file and examine the data and the file layout. You will need to make certain deductions about this data (such as column widths) based on your observations.

1. Begin your program with the required header and filename statements. Use a fileref to access the CSV file. As always, your program must include comments in the appropriate places.
2. For this assignment the output file must be created with the pages in a landscape layout. The date is to only be displayed on the final section of the output. The SAS output PDF file should start on page number 2.
3. Create a user defined format that can be used to display the school Division based on the number of students in the HSTotal variable. The size divisions are as follows:
   6A – 1251 and above
   5A – 721 to 1250
   4A – 375 to 720
   3A – 181 to 374
   2A – 107 to 180
   A – 70 to 106
   B – 0 to 69
   Non-HS – missing or other values
4. Within the same procedure, create a second format that can be applied to the STRatio variable to classify the value into Class Size categories. The cutoff values for these categories are shown below. The format must be created so that every value fits within one of these categories.
   Very Small – any value less than 10
   Small – 10 up to but not including 14
   Medium – 14 up to but not including 18
   Large – 18 up to but not including 22
   Very Large – 22 and up
   Unknown – Missing values
5. Write a data step similar to the one shown in the lectures that converts the csv file to a SAS dataset in the work library. This data step will include infile, input, and length statements. You will need to make certain deductions about this data, such as column widths, based on your observations. You will need to use the firstobs option for the infile statement to tell SAS to skip the column titles that are in the raw data. Unlike the data step option by the same name, this

option does not use parenthesis. Be sure to check the log very carefully when you read in the raw data. You should not get invalid data errors when reading this data. If you do, then you most likely do not have your data step set up correctly.

6. Print the first 30 observations of the new data set. Suppress the printing of observations numbers. Make the title and footnote match the sample output posted on eCampus.
7. Use the FREQ procedure to produce a distribution of class sizes based on the Student-Teacher Ratio (STRatio). Apply a temporary label to the variable and use the format you created in step 4 to create the groupings as shown. The statistics shown on your report must match those posted on eCampus. Use a single title statement to get the titles to match the sample output.
8. Use the SUMMARY procedure to create an output data set that contains the average Student-Teacher Ratio (STRatio) grouped by Division. Even though HSTotal is numeric, we are going to use it as a classification variable and apply our user-defined format to determine the Divisions. You will need to use the appropriate option on the proc summary statement to ensure that non-high school values are captured properly. Research Help on the SUMMARY procedure as needed to determine which option to use.
9. Print out the data portion of the summary data set. Make sure the date is printed at the top of the page for this output step. Use other enhancements to supply titles as shown. Note that there is a blank line between the two titles on this page. Suppress the printing of observation numbers. Apply other enhancements and options including a subsetting clause as necessary to get your output to match the output in eCampus.
10. At the end of your program include “housekeeping” statements to ensure that titles and footnotes do not get carried over to any subsequent output generated during this SAS session.
11. Convert the program and log to PDF files and submit them to WebAssign along with your SAS output.