

STAT 540: Homework 3

You will need to read in 3 different data sets for this assignment; the data sets should be input *as is*; do not alter them. A couple of the data sets seem simple, but can be tricky to develop the correct **INPUT** statement. Unless otherwise specified, you are free to use **PROC IMPORT** or an **INFILE** statement, or to copy-and-paste the text file into your SAS program and use **datalines**. Be sure to document efforts, both unsuccessful and successful, to receive credit. Variables (e.g., dates and social security numbers) should be printed in an appropriate format. *If you are unsuccessful importing your data for the first two problems, enter the data by hand in an Excel spreadsheet and import it so that you can answer the follow-up questions.*

1. Write a SAS program to read (and print) the data file **HW3Q1.txt**; either use **datalines** or a simple **INFILE** statement. The variables for each observation are: Month, Day, Year, LastName, FirstName, Pay. Combine Month, Day, and Year to create the variable Date. Use a subsetting IF to extract records with dates after January 1, 1960 into a new data set.
2. Write a SAS program to read in and print the data file **HW3Q2.txt**; use a simple **INFILE** statement here (you can try **datalines**, but it's even harder to get it to work successfully). The variables for each observation, in order, are: SSN, Gender, Q1, Q2, Q3, Q4, Q5, Q6, Q7.
 - Compute the minimum Q1 score.
 - Now compute the minimum score across all seven quizzes.
 - Create a new data set containing only those records for which the average quiz score of Q1-Q3 is greater than the average quiz score of Q5-Q7.
3. The text file **CongareeGaugeMod.txt** contains information on the flow and stage of the Congaree River collected in 15 minute intervals from October 8-9, 2007. This file was downloaded from USGS website <http://waterdata.usgs.gov/nwis/uv?02169500> using the tab-delimited file option. Variables are: Agency, Gauge, DayTime, Stage, and Discharge. Be sure to inspect the file before proceeding—we are not interested in importing the extensive header. *Try* to import then print the text file **Congaree-GaugeMod.txt** using **PROC IMPORT** or **INFILE**; report any shortcomings in the SAS worksheet you obtain. Students using **PROC IMPORT** should refer to the material in section 2.6 (including optional statements **DATAROWS** and **GUESSING ROWS**) for help; students using **INFILE** can refer to section 2.17 for hints. Grad students: in addition to the above, import the data (from **CongareeGaugeGrad.txt**). Variables are: Agency, Gauge, DayTime, Stage, Stage2, Discharge, Discharge2. Print the data set.