**SAS Exercise – Homework 2**

[1](https://www.safaribooksonline.com/library/view/Exercises+and+Projects+for+The+Little+SAS+Book,+Fifth+Edition/9781629598055/xhtml/SelectedSolutions.xhtml#ch02s50).   The World Health Organization (WHO) collected data in countries across the world regarding the outbreak of swine flu cases and deaths in 2009. The data in the file SwineFlu2009.dat include counts per country by month during the epidemic. There are many variables in the raw data file with the following descriptions:

* By date: ID for sorting by first case date
* By continent: ID (X.YY) for sorting by first case date within a continent where X represents continent X, and YY represents the YYth country with the next first case
* Country
* Date of first case reported
* Number of cumulative cases reported on the first day of the month for April, May, June, July, and August (across the columns, respectively)
* Last reported cumulative number of cases reported to WHO as of August 9, 2009
* By date: ID for sorting by first death date
* By continent: ID (X.YY) for sorting by first death date within a continent where X represents continent X, and YY represents the YYth country with the next first death
* Date of first death
* Number of cumulative deaths reported on the first day of the month for May, June, July, August, September, October, November, and December (across the columns, respectively)

Your task is to create a corresponding SAS dataset by reading in this data file.

1. First, examine the raw data file SwineFlu2009.dat using Notepad.

To open in notepad Method 1: Right-click on file, choose Open with 🡪 Other programs 🡪 Notepad.

To open in notepad Method 2: Open Notepad, From the menu: File 🡪 Open (change from .txt to all files to display all files).

1. Write a DATA step to read the file into SAS. Make sure that each variable is assigned a unique (and descriptive) name and is of the correct type – character or numeric.
2. After the dataset is created, locate the file in the work library through the explorer window in SAS. Double click on the dataset to view the data.
3. In SAS, dates can be stored as a special type of numeric data. Modify the DATA step to make sure that the dates are read in the correct SAS date format (not as character). (HINT: Use the correct “informat” in the input statement – e.g., input firstdate ddmmyy8.;)
4. Create a permanent label for each variable based on the preceding descriptions. (HINT: Use the “LABEL” statement).
5. Print a report that describes the contents of the data set including the labels that you’ve created and other attributes of the variables. (HINT: Use PROC CONTENTS).

[2](https://www.safaribooksonline.com/library/view/Exercises+and+Projects+for+The+Little+SAS+Book,+Fifth+Edition/9781629598055/xhtml/SelectedSolutions.xhtml#ch02s54).   A gourmet pizza restaurant is considering adding new toppings to its menu. Each month they survey 10 customers about their preferences for three different toppings. They want data on several different toppings, so they don’t always ask about the same three toppings. Customers rate each topping on a scale of 1(would never order) to 5 (would order often). The restaurant wants to compute average ratings for all toppings, so the ratings variables need to be numeric. The raw data file Pizza.csv has variables for the respondent’s ID, and the ratings for five different toppings: arugula, pine nuts, roasted butternut squash, shrimp, and grilled eggplant. The first two digits in the ID correspond to the month of the survey.

1. Examine the raw data file Pizza.csv and read it into SAS using the IMPORT procedure.
2. Print the data set (on the results screen). Print a report that describes the contents of the data set to make sure all the variables are the correct type.
3. Open the raw data file in a simple editor like WordPad and compare the data values to the output from part b) to make sure that they were read correctly into SAS. In a comment in your program, identify any problems with the SAS data set that cannot be resolved using the IMPORT procedure.
4. Read the same raw data file, Pizza.csv, this time using a DATA step (instead of the IMPORT procedure). Be sure to resolve any issues identified above.
5. Create a new dataset with the average ratings for each topping.

[3](https://www.safaribooksonline.com/library/view/Exercises+and+Projects+for+The+Little+SAS+Book,+Fifth+Edition/9781629598055/xhtml/SelectedSolutions.xhtml#ch03s30).   The new management of a local hotel decided to update their recently acquired (and very outdated) property by installing wireless Internet service for their guests. They are also considering updating their billing system because the method used by the previous owner seems faulty. In order to conduct a billing analysis, they would like some calculations about the guests who stayed with them during the first part of February (this was the first month after the change of ownership). The raw data file Hotel.dat contains variables with information on room number, number of guests, check-in month, day, year, check-out month, day, year, use of wireless Internet service, number of days of Internet use, room type, and room rate.

1. Examine the raw data file Hotel.dat and read it into SAS.
2. Create date variables for the check-in and check-out dates, and format them to display as readable dates.
3. Create a variable that calculates the subtotal as the room rate times the number of days in the stay, plus a per person rate ($10 per day for each person beyond one guest), plus an Internet service fee ($9.95 for a one-time activation and $4.95 per day of use). (HINT: You can subtract dates if they have been stored as SAS dates).
4. Create a variable that calculates the grand total as the subtotal plus sales tax at 7.75%. The result should be rounded to two decimal places.
5. View the resulting data set. In a comment in your program, state the value for the grand total for room 211.

[4](https://www.safaribooksonline.com/library/view/Exercises+and+Projects+for+The+Little+SAS+Book,+Fifth+Edition/9781629598055/xhtml/SelectedSolutions.xhtml#ch03s36).   A study on weight loss in males was conducted over a five-month period. Participants came to a clinic at the beginning of each month, and their weight as well as their responses to six survey questions regarding weight loss were obtained. The raw data file WLSurveys.dat contains variables for subject ID, height (in), weight (lb) at each of five visits, and responses to questions 1-6 at each of five visits. Note that the first six survey responses are from visit one, the next six survey responses are from visit two, and so on for a total of 30 survey responses.

1. Examine the raw data file data WLSurveys.dat and read it into SAS.
2. The survey questions were measured on a scale of 0 (least) to 3 (most). However, the data were recorded inconsistently by data entry clerks; and questions 2, 3, and 5 (for every visit) were accidentally recorded in the reverse order (0 = most and 3 = least). Convert questions 2, 3, and 5 for every visit so that they are measured in the same direction as the other questions.
3. Missing data were recorded as -99. Reclassify these values as a period (.) to signify missing numeric data in SAS.
4. Calculate the subject's body mass index at each of the five visits as weight (lb) divided by height (in) squared, and then multiplied by 703.
5. Researchers would like to study the subjects that are classified as overweight or obese at the final visit. Limit the resulting data set to those patients with a final visit body mass index that is 25.0 or more.
6. View the resulting data set. In a comment in your program, find the tenth observation and state the subject ID, number of question variables with missing values, and the BMI at the last visit.

[5](https://www.safaribooksonline.com/library/view/Exercises+and+Projects+for+The+Little+SAS+Book,+Fifth+Edition/9781629598055/xhtml/SelectedSolutions.xhtml#ch06s31).   The United States Bureau of Labor Statistics publishes various indexes that measure average prices of consumer goods in urban areas. The SAS data set AVEPRICES contains data on the average price of unleaded regular gasoline (per gallon), whole large eggs (per dozen), and whole milk (per gallon) for a 10-year period by month in the United States. The variables in this file are year, month, price, and type of commodity.

1. Examine this SAS data set including the variable labels and attributes. Use one or more procedures to create a data set that contains the average price per year for each commodity.
2. Using one or more procedures, create a new version of this data set that has 12 variables corresponding to each month, that contain the price data for each combination of commodity and year.
3. Combine the averages calculated in part a) with the data set created in part b).
4. In one DATA step, divide the data set from part c) into three data sets that correspond to the three commodities.
5. Add code to the DATA step in part d) that will give the variables for months better names. The final data sets should not include any automatic variables that may have been created along the way.
6. Add a comment to your program that states the number of observations and variables in each of the final data sets created in part e).

[6](https://www.safaribooksonline.com/library/view/Exercises+and+Projects+for+The+Little+SAS+Book,+Fifth+Edition/9781629598055/xhtml/SelectedSolutions.xhtml#ch06s35).   The local school district wants to survey all sixth grade students and their school-aged siblings. There are three different types of surveys: one for the sixth graders, one for their younger siblings, and one for their older siblings. The SAS data set called SCHOOLSURVEY contains data for all sixth graders in the three middle schools in the district (Rachael Carson, Green Valley, and Redwood Grove). The data set also includes data for all their siblings attending schools in the district, which can be linked back to the sixth grader by Family\_ID.

1. Examine this SAS data set including the variable labels and attributes. Add a comment to your program that notes the sort order of the variables in this data set.
2. Create a data set that has one observation for each sixth grader.
3. Combine the data set from part b) with the original data, and create a variable that is the age difference in years between the sixth grader and the sibling.
4. Count the number of older siblings and the number of younger siblings of the sixth grader, and add these variables to the data set of just the sixth graders. Be sure to label the variables appropriately.
5. So the district knows how many surveys to print and distribute to each school, compute the number of sixth graders, and the total number of younger and older siblings by school. Be sure to show the complete names for the schools.
6. For each sixth grader, add variables to your data set for the age difference between the sixth grader and their youngest sibling, and the age difference between the sixth grader and their oldest sibling.
7. For each school, compute the average, minimum, and maximum age difference between the sixth graders and their youngest and oldest siblings. Be sure to show the complete names for the schools.