STAT 440 – Homework 2

Students are encouraged to work together on homework. However, sharing or copying any part of the homework is an infraction of the University's rules on Academic Integrity.

Final submissions must be uploaded to our Compass 2g site on the Homework page. No email, hardcopy, or late submissions will be accepted.

The HW Report should include the output generated from the following exercises:
1-bd, 2-bd, 3-bcef

Getting the program file ready

a. Create a folder on the hard drive with the following pathname – C:\440\hw2. Save all data files accompanying this assignment in that folder. If you cannot create the folder because you are working on a university computer and don't have permission, create the …\440\hw2 folder elsewhere.
b. Assign the library reference **hw2** to the folder 'C:\440\hw2'. Use this library as your permanent library for this assignment. If you could not create the folder, assign the library reference **hw2** to your …\440\hw2 folder.
   Note: If you are using a folder other than 'C:\440\hw2', you must change any pathname references in your program file to 'C:\440\hw2' before submitting your homework.

Submitting your work to Compass 2g

You are to submit two (and only two) files for your homework submission.

1. Your SAS program file which should be saved as **HW*n*_*YourNetID*.sas**. For example, my file for the HW2 assignment would be HW2_dunger.sas. All program statements and code should be included in one program file.

2. Your Report including all relevant output to address the exercises. For this homework, use ODS to send your results to a Portable Document Format (PDF) file called **HW*n*_*YourNetID*.pdf**. For example, my file for HW2 would be HW2_dunger.pdf. Only include your final set of output. Do not include output for every execution of your SAS program.

You have an unlimited number of submissions, but only the last one will be viewed and graded. Homework submissions must always come as a pair of files, as described above.

1. **Australian AIDS Survival Data**

   The raw data file **AUaids.dat** contains data on patients diagnosed with AIDS in Australia before July 1, 1991. In order of occurrence, it contains the following variables: observation number, state of origin, sex, date of diagnosis, date of death or end of observation, status at end of observation, reported transmission category, and age at diagnosis.

   a. Write a DATA step so that the **AUaids.dat** raw data file is used to create a permanent SAS data set called **AUaids_*NetID***. Choose appropriate names, labels, formats, and other attributes as needed for the variables.
   b. Print the first 10 observations from the data portion of your new SAS data file, excluding observation numbers. (Include results in the HW Report.)

   c. Repeat part (a) by writing a DATA step to read in **AUaids.dat**. But this time the output data set should only contain those patients who are males under the age of 30 and who contracted AIDS through the receipt of `"blood"`. The output data set should be permanent and titled **under30_*NetID***. Choose appropriate names, labels, and other attributes as needed for the variables.
   d. Print the entire data portion of your new SAS data file, excluding observation numbers. (Include results in the HW Report.)

2. Patient allergy records

Medical data is stored in the raw data file **allergy.dat**. The first six fields are always as follows:

Field	Description	Notes
1	ID Number	5-character code
2	Last Name	Longest value is 9 characters
3	First Name	Longest value is 11 characters
4	Plan Type	1-character code
5	Blood Type	Longest value is 3 characters
6	Allergy Code	1-character code; Y=Yes, N=No

If the patient has an allergy, then the rest of the record is as follows:

7	Allergy Type	2-character code indicating type of allergy
8	Number of Dependents	

If the patient does not have an allergy, then the rest of the record is as follows:

7	Number of Dependents	

a. Write a DATA step to read the values of **allergy.dat** into SAS. The output data set is to be a temporary SAS data file called **allpatients_*YourNetID***. Choose appropriate names, labels, and other attributes as needed for the variables.

b. Print the data portion of your new SAS data file. Include all variables except patient ID and plan type. (Include results in the HW Report.)

c. Write a new DATA step to read the values of **allergy.dat** into SAS similar to the one in part (a), but this time include only patients with allergies. The output data set is to be a temporary SAS data file called **allergies_*YourNetID***. Choose appropriate names, labels, and other attributes as needed for the variables.

d. Print the data portion of your new SAS data file. Include all variables except patient ID and plan type. (Include results in the HW Report.)

3. Hockey data

The raw data set **calories.txt** contains the results of a survey of 40 food items claiming to be "lite," "reduced-calorie," "low-calorie," "diet," "low-fat," "no-fat," or "health" foods. Each food is classified based on its distribution as nationally advertised, regionally distributed, or locally prepared.

Field	Name	Description
1	Food	Kind of food
2	Per_gram	Percentage difference between measured calories and labeled calories per gram (100% x (measured-labeled)/labeled)
3	Per_item	Percentage difference between measured calories and labeled calories per item
4	Classification	N if nationally advertised, R if regionally distributed, L if locally prepared

a. Write a DATA step to read the values of **calories1.txt** into SAS using formatted input. The output data set is to be a temporary SAS data file called **calories1_*YourNetID***. Choose appropriate names, labels, and other attributes as needed for the variables.
b. Print the descriptor portion of your new SAS data file. (Include results in the HW Report.)
c. Using **calories1**, print a list of all foods who measured fewer calories than the label actually displays in both metrics. That is, both Per_gram and Per_item value are less than 0. (Include results in the HW Report.)

d. Write a DATA step to read the values of **calories2.txt** into SAS using list input. The output data set is to be a temporary SAS data file called **calories2_*YourNetID***. Choose appropriate names, labels, and other attributes as needed for the variables.
e. Print the descriptor portion of your new SAS data file. (Include results in the HW Report.)
f. Using **calories2**, print a list of all locally prepared foods. (Include results in the HW Report.)