# STAT 440 – Homework 1

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-acd, 2-ace, 3-abce**

## Getting the program file ready

a. Create a folder on the hard drive with the following pathname – C:\440\hw1. 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\hw1 folder elsewhere.
b. Assign the library reference **hw1** to the folder 'C:\440\hw1'. Use this library as your permanent library for this assignment. If you could not create the folder, assign the library reference **hw1** to your …\440\hw1 folder.
   Note: If you are using a folder other than 'C:\440\hw1', you must change any pathname references in your program file to 'C:\440\hw1' 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 HW1 assignment would be HW1_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 HW1 would be HW1_dunger.pdf. Only include your final set of output. Do not include output for every execution of your SAS program. Use the template file **hw1 template.sas** as your guide.

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. Price Data

You will be working with the SAS data file **pricedata** which contains simulated monthly sales data. It is located in the **sashelp** library.

a. Print the descriptor portion of this SAS data file. (Include results in the HW Report.) Read the results to become familiar with the contents of this data set.

b. Read in the SAS data file and create a temporary SAS data file called **pricing_*NetID***.
   - Change the format of Date so that the full date is represented in a ddMMMyyyy structure. For example, JAN90 would show as 01JAN1990.
   - Change the format of the variables Price and Cost so that their values appear monetarily and are rounded to the nearest hundredth. For example, 41.9 would show as $41.90.
   - Change the format of the variable Discount so that the displayed value shows as a percentage rather than a decimal. For example, 0.13 would show as 13%.
   - Change the label of the variable Sale to "Units Sold".
   - Subset the observations so that only those months are included in which more than 500 units were sold and the difference between the Unit Price and Unit Cost was less than $20.
   - Only the variables Date, Sale, Price, Discount, Cost, and ProductName should be in the final data set.

c. Print the descriptor portion of **pricing_*NetID***. (Include results in the HW Report.)

d. Print the data portion of **pricing_*NetID***. (Include results in the HW Report.)

## 2. Birth Weight Data

You will be working with the SAS data file **birthweight** which contains a subset of the 1997 birth weight data from National Center for Health Statistics. The data record live, singleton (not multiples) births to mothers between the ages of 18 and 45 in the United States who were classified as black or white.

| Variable | Description |
|---|---|
| weight | Infant's Birth Weight |
| black | Indicator of Black Mother |
| married | Indicator of Married Mother |
| boy | Indicator of Boy |
| visit | Prenatal Visit: 0 - no visit, 1 - visit in second trimester, 2 - visit in last trimester, 3 - visit in first trimester |
| ed | Mother's Education Level: 0 - high school, 1 - some college, 2 - college, 3 - less than high school |
| smoke | Indicator of Smoking Mother |
| cigsper | Number of Cigarettes Smoked per Day |
| mom_age | Mother's Age |
| m_wtgain | Mother's Weight Gain during Pregnancy |

a. Print the descriptor portion of the SAS data file. (Include results in the HW Report.) Read the results to become familiar with the contents of this data set.

b. Read in the SAS data file and create a permanent SAS data file called **smokers*_NetID*** that contains all observations for mothers who smoked during pregnancy.
   - Include only the following variables in the output data set: weight, boy, visit, smoke, cigsper, and m_wtgain.
   - Create a label for each variable in the output data set.

c. Print the descriptor portion of **smokers*_NetID***. (Include results in the HW Report.)

d. If the mother listed herself as a smoker in the smoke variable, then she should have also listed a positive value for cigsper. Further, if the mother listed herself as a nonsmoker in the smoke variable, then she should have also listed a 0 value for cigsper. Create a temporary SAS data file called **check*_NetID*** that double-checks these conditions. The data set **check*_NetID*** should contain all observations that violate these conditions.

e. Look at the Log notes pertaining to the creation of **check*_NetID***.
   - If the data set contains 0 observations, all observations checked out correctly and have been validated. Print the descriptor portion of **check*_NetID***.
   - But if instead the data set contains 1 or more observations, there is invalid data. Print the data portion of **check*_NetID*** including the weight, smoke, cigsper, and m_wtgain variables.
   - (Only one of these two results will be included in the HW Report.)

## 3. Heart Data

You will be work with the SAS data file **heart** which contains the results of the original 5,209 adults who participated in the Framingham Heart Study which began in 1948 and lasted 20 years. It is located in the **sashelp** library.

a. Print the descriptor portion of this SAS data file. (Include results in the HW Report.) Read the results to become familiar with the contents of this data set.

Without creating a new SAS data set (for parts b and c), use the **heart** data set to …

b. Print a table of all subjects who fall in the Overweight category based on BMI (that is, those males where BMI is between 25 and 30 and BMI = Weight / Height$^2$ * 703) and who died before age 60. Include only the subject's gender, weight, height, and age at death in that order. (Include results in the HW Report.)
c. Print a table of all subjects who died of cancer during the study. Only females who lived to at least 95 and males who lived to at least 85 should be included in the results. Include only the subjects' cause of death, gender, age at the beginning of the study, and age at death should be included. (Include results in the HW Report.)

d. Read in the SAS data file and create a temporary SAS data file called **highrisk*_NetID*** that contains all subjects who were diagnosed with Coronary Heart Disease (CHD) during the study and possessed at least two of the risk factors. The risk factors are as follows:
    - High Cholesterol
    - High Blood Pressure
    - Overweight
    - Heavy or Very Heavy Smoker
e. Print the descriptor portion of **highrisk*_NetID***. (Include results in the HW Report.)