**BIOS810/CPH651: Assignment #1**

**Points: 16**

This assignment will give you practice reading into SAS data from several types of formats. There are 3 parts to the assignment. You may create one program by appending the code for each part or create separate programs for each part. You may find it helpful to write the program on a piece of paper before you work on the computer. It is very important to look at the log after running your program each time to make sure there are no errors.

What to hand in?

Include the SAS program for each part answering any questions. Copy your SAS program and output into the homework, Word document. The answers to the questions can be put in the same document. *Remember to answer any questions that are in italics*.

All data files are posted on Canvas.

If you are having issues getting the data into Canvas, review the video, **Uploading data from Canvas into SAS Studio [Excel or raw text files].**

1. (3 points) The data file organization.txt contains made-up data on 4 employees in an organization. The file is a text file where colons were used as delimiters. To indicate missing values, two colons were entered. The contents of the file is as follows:

**organization.txt**

Roberts:M234:45000

Chien:M74777::

Walters::75000

Rogers:F7272:78131

Write a SAS program that reads the data from **organization.txt** and creates a temporary SAS dataset called **organization**. Name the variables LastName (character), EmpNo (character) and Salary (numeric). Use either an INFILE or FILENAME statement to bring in the data (Don’t copy the data into your program as if it were instream data). For example if using INFILE, the first two statements of your program should be the following:

**DATA organization;**

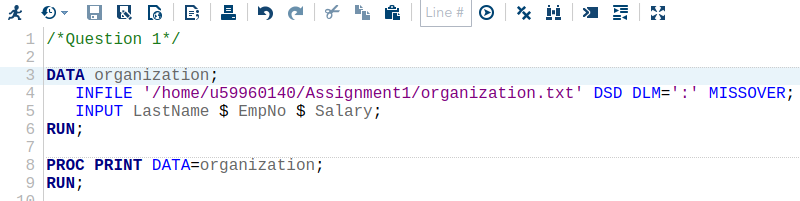
**INFILE '/home/u49671821/SAS\_Class/organization.txt' ;**

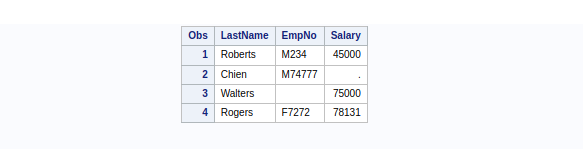
(Note: You will need to upload the file into SAS Studio. In SAS Studio, right click on the data file and select Properties to determine the location of the file for the INFILE statement.)

Display the values of the variables using PROC PRINT and include the output. Run the program, making sure the data is read in correctly.

[Hint: Recall we discussed data that was comma delimited in class]

| DATA organization;  INFILE '/home/u59960140/Assignment1/organization.txt' DSD DLM=':' MISSOVER;  INPUT LastName $ EmpNo $ Salary;  RUN;  PROC PRINT DATA=organization;  RUN; |
| --- |





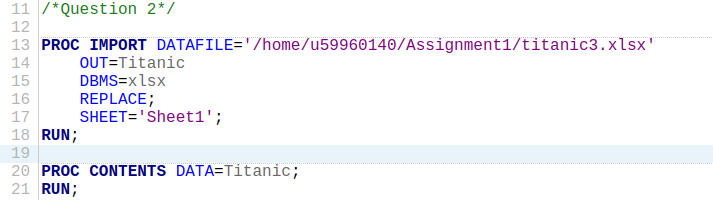
1. This problem will give you practice reading data into SAS from an Excel spreadsheet. If you are having problems importing the Excel spreadsheet into SAS Studio, refer to the SAS Studio Information Module that has a video explaining how to read an Excel spreadsheet into SAS Studio. Importing Excel files are also covered in the Importing Data into SAS Studio video under Week 2 Videos.

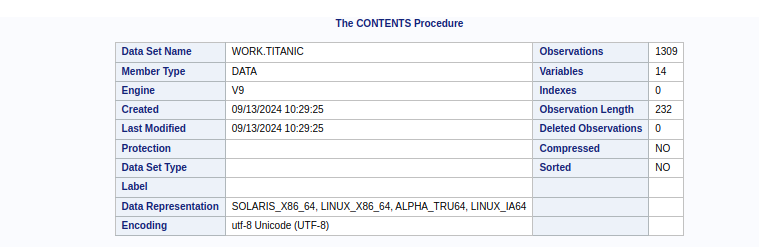
The data file **titanic3.xlsx** contains data on individual passengers on the Titanic. The variables in creation order are pclass, survived, name, sex, age, sibsp, parch, ticket, fare, cabin, embarked, boat, body, and home.dest. .

[Note: You will not need a data step for this problem]

* 1. (3 points) Write a program to read in the **Titanic3** dataset, creating a SAS dataset called **Titanic**. Include the code that is generated by the Import Wizard or code from SAS Studio that read in the Excel sheet. Run a PROC CONTENTS on the dataset which shows the variable names of the dataset and number of observations. *How many observations are in the dataset?* Include the output from PROC CONTENTS in your homework.

| PROC IMPORT DATAFILE='/home/u59960140/Assignment1/titanic3.xlsx'  OUT=Titanic  DBMS=xlsx  REPLACE;  SHEET='Sheet1';  RUN;  PROC CONTENTS DATA=Titanic;  RUN; |
| --- |

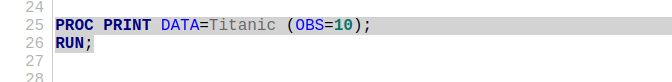


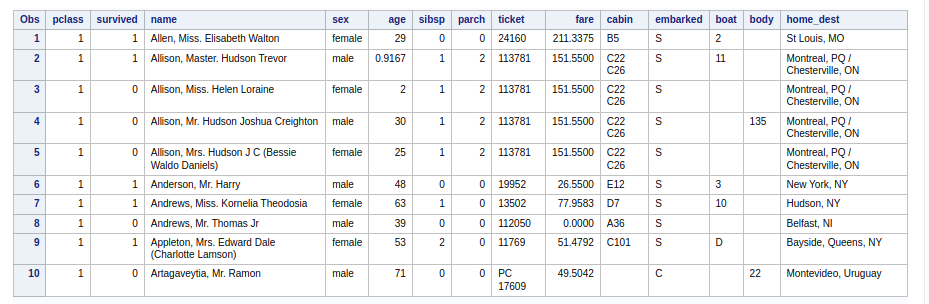


Number of observations is 1309.

* 1. (1 point) Display all variables using PROC PRINT. Use the OBS option available with PROC PRINT to limit the number of observations displayed to 10 and include the output in your homework.

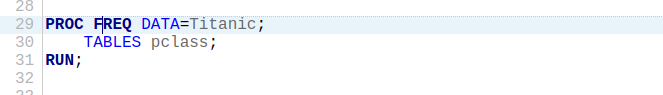
| PROC PRINT DATA=Titanic (OBS=10);  RUN; |
| --- |

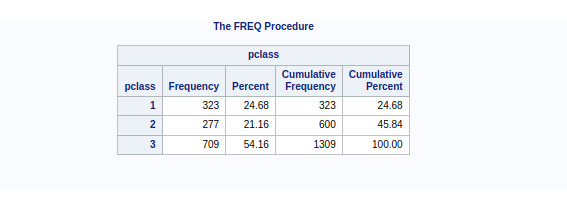




* 1. (1 point) **pclass** refers to passenger class (1st, 2nd, 3rd), and is a proxy for socio-economic class. Using PROC FREQ show the frequency distribution of **pclass** for the passengers.

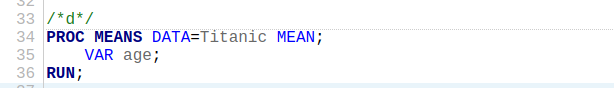
| PROC FREQ DATA=Titanic;  TABLES pclass;  RUN; |
| --- |

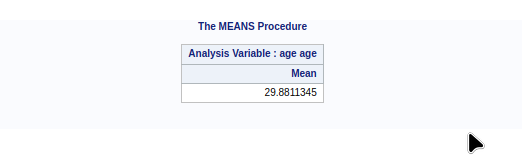




* 1. (2 points) Using PROC MEANS display the average age of the passengers. *What is the average age*? Include the output from PROC MEANS and the answer to the question.

| PROC MEANS DATA=Titanic MEAN;  VAR age;  RUN; |
| --- |





The average age is 29.8811345.

1. The text file **geocaching.txt** posted on Canvas has data values arranged as follows:

| **Variable** | **Description** | **Starting Column** | **Ending Column** | **Data Type** |
| --- | --- | --- | --- | --- |
| Name | Cache name | 1 | 20 | Char |
| LongDeg | Longitude degrees | 21 | 22 | Num |
| LongMin | Longitude minutes | 23 | 28 | Num |
| LatDeg | Latitiude degrees | 29 | 30 | Num |
| LatMin | Latitude minutes | 31 | 36 | Num |

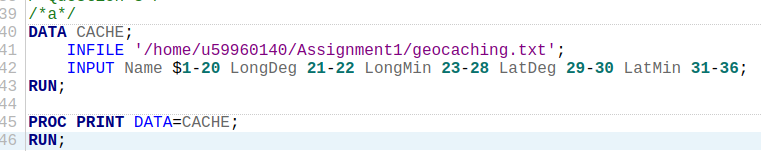
Again, you will need to data file into SAS Studio and bring it into SAS with code such as what follows:

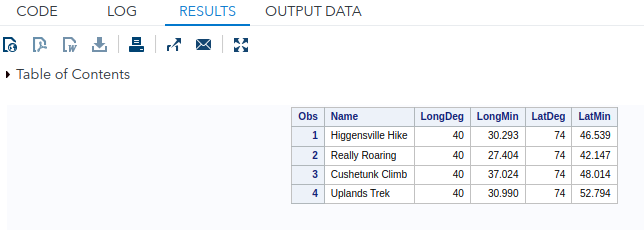
**DATA CACHE;**

**INFILE '/home/u49671821/SAS\_Class/geocaching.txt' ;**

1. (3 points) Create a temporary SAS data set called CACHE using the data file. Use column input to read the data values. Include the code of your SAS program and output from PROC PRINT in your homework to show the program ran correctly.

| DATA CACHE;  INFILE '/home/u59960140/Assignment1/geocaching.txt';  INPUT Name $1-20 LongDeg 21-22 LongMin 23-28 LatDeg 29-30 LatMin 31-36;  RUN;  PROC PRINT DATA=CACHE;  RUN; |
| --- |





1. (3 points) Use formatted input (i.e. pointers and informats) to read the data values instead of column input. Call the second temporary SAS dataset CACHE2. Include the code of your SAS program output from PROC PRINT in your homework to show the program ran correctly.

| DATA CACHE2;  INFILE '/home/u59960140/Assignment1/geocaching.txt';  INPUT @1 Name $CHAR20. @21 LongDeg 2. @23 LongMin 6. @29 LatDeg 2. @31 LatMin 6.;  RUN;  PROC PRINT DATA=CACHE2;  RUN; |
| --- |

