Exercise Lab 1

Mina Moeini
Mina_moeini@sfu.ca

September 4, 2024

1 Some tips for using SAS

- Case Sensitivity: SAS does not differentiate between uppercase and lowercase letters in its code. However, for clarity and consistency, it's a good practice to stick to a consistent case format (e.g., using lowercase for all keywords).
- Semicolons: Always use a semicolon (;) to separate commands. Starting a new line won't terminate a statement. Missing semicolons are a common source of errors in SAS.
- Step Keywords: Commands are grouped into "steps" that begin with specific keywords. The most commonly used are:
  - data: To create, read, or modify datasets.
  - proc: For procedures used to analyze or process data.
- Running Code: Use the run; statement to execute a block of code. This is necessary to tell SAS when a group of instructions should be executed. You can also use quit; after certain procedures (like proc sql).
- Submitting Code: To submit your code, click the submit icon or highlight the section of the code you want to run, right-click, and choose 'Submit Selection.' You can also use keyboard shortcuts, such as F3 or Ctrl + Enter (depending on your SAS environment).
- Color Coding: SAS uses different colors to highlight the components of your code, such as: Keywords (like data, proc) are often colored to stand out. Variables and values might appear in different shades to distinguish their roles. Comments are usually displayed in green to indicate ignored sections.
- Adding Comments: Include comments to document your code. SAS ignores anything between:
  - Block comments: /* This is a comment */

  - Line comments: * This is also a comment;

- Debugging: Always check the log tab after running your code. Look for errors (in red) or warnings (in blue) that might indicate problems in your code. Correct any issues and rerun the affected section.

- Libnames and Filenames: Use libname to assign a shortcut for a library (a folder where datasets are stored). For example, libname mylib 'C:'; will let you reference datasets in that folder as mylib.datasetname. Similarly, filename is used to point to external files.

- Permanent vs. Temporary Datasets: If you don't specify a library, SAS stores datasets in the work library, which is temporary and will be cleared when the session ends. To store datasets permanently, use a library you define with libname.

- Importing Data: You can import data from various sources like CSV, Excel, or databases using proc import for structured files or infile for raw data. The proc import method allows for quick imports, while infile gives more flexibility.

- Exporting Data: Use proc export to save your SAS datasets as external files (e.g., CSV, Excel). It's useful for sharing your processed data with others.

1.1 Exercise

Exercise 1. Read in the following data using three different methods and check contents:

Name	Height	Weight
Jim	1.8	90
Johnson	1.78	80
Jane	1.65	55
Jason	1.90	90
Janis	1.6	60

Exercise 2.

1. Import data from `boneden.xls` from the Lab 1 folder.
2. How many observations are there in this dataset?
3. Create a library named `MYSFUFILES`.
4. Save the imported data to your designated library in a SAS file.

Exercise 3.

1. Review the descriptions of the dataset described below using `Proc contents`.

2. Make 4 subsets:

   (a) The first 10 observations (name it `samhsa_ten`).

   (b) Random sampling of 20 observations (name it `samhsa_sam`).

   (c) Only include people aged 21-24 (code 05) (name it `samhsa_age`).

   (d) A subset that only includes information about `CASEID`, `AGE`, `SEX`, `EDUC` (name it `samhsa_info`).

3. Save the subsets into SAS files.

SAMHSA TEDS-D dataset: TEDS discharges (TEDS-D) datasets can be found here: https://www.icpsr.umich.edu/web/ICPSR/series/238

SAMHSA: The Substance Abuse and Mental Health Services Administration (SAMHSA) is the agency within the U.S. Department of Health and Human Services. The goal of SAMHSA is to reduce the impact of substance abuse and mental illness on America's communities. SAMHDA is an initiative funded under contract with the "Center for Behavioral Health Statistics and Quality" (CBHSQ), "Substance Abuse and Mental Health Services Administration" (SAMHSA), U.S. Department of Health and Human Services (HHS). CBHSQ has primary responsibility for the collection, analysis, and dissemination of SAMHSA's behavioral health data.

CBHSQ promotes the access and use of the nation's substance abuse and mental health data through the Substance Abuse and Mental Health Data Archive (SAMHDA). SAMHDA provides public-use data files, file documentation, and access to restricted-use data files to support a better understanding of this critical area of public health.

The Treatment Episode Data Set – Discharges (TEDS-D) is a national census data system of annual discharges from substance abuse treatment facilities. State laws require certain substance abuse treatment programs to report all of their admissions and discharges to the State. In all States, treatment programs receiving any public funds are required to provide the data on both publicly and privately funded clients. In some States, programs that do not receive public funds are required to provide data as well. TEDS collects this data from the States on all admissions and discharges aged 12 or older.

Data Dictionary: This is a codebook for the "Treatment Episode Data Set – Discharges" (TEDS-D). The dataset includes information on treatment completion, length of stay in treatment, and demographic and substance abuse characteristics of 1.7 million individuals discharged from alcohol/drug treatment in facilities that report to individual state administrative data systems.