

PHPM 672/677 Assignment #2: Variables & Conditionals
Due date: Submit in E-Campus by 11:59pm Wed 1/29
Mid point check due date: Email by 11:59pm Wed 1/22 (next week)

Submission. Submit on E-Campus by 11:59pm the day before the class they are due.

REMEMBER no datasets please.

1. Midpoint email (1 point)
 - What dataset will you use for assignment 2 ?
2. Final Submission: zip up all files and upload
 - Commented code (SAS: lnameN.sas, where N indicates the assignment number and lname is your last name)
 - Outputs from your code (2 files. SAS: lnameN.log AND lnameN.lst or lnameN.html)
 - Readme.txt file answers to questions
 - Lab 2 (program, log, lst/html)

Late Assignments. No late assignment is accepted for this assignment because this assignment feeds into the next assignment.

Elegance: There is always more than one way to say something, but some ways will be more “elegant” than others. You will learn to recognize elegant expressions as you become more familiar with a language and use the elegant idioms as you become more skilled.

Guideline for assignment grading (Total of 8)

- Assignment (Total 4)
 - 1: Submitted code that does not run.
 - 2: Mostly running but incorrect.
 - 3: Correct and meets requirements (i.e uses programming constructs required)
 - 4: Correct & Elegant. Comments.
- Answers to questions on the assignment (Total 1)
- Midpoint Check email (Total 1)
- Lab (Total 2) – DON’T forget to submit this with this assignment

Required & recommended readings for this assignment

- <https://stats.idre.ucla.edu/sas/modules/creating-and-recoding-variables-in-sas/>
- <https://stats.idre.ucla.edu/sas/modules/using-sas-functions-for-making-and-recoding-variables/>
- <https://stats.idre.ucla.edu/sas/modules/ubsetting-data-in-sas/>
- <https://stats.idre.ucla.edu/sas/modules/labeling/>

Assignment 2: Variables & Conditionals

In this assignment, you will learn the basics of working with one dataset. By the end of this assignment, you should be able to

- To write conditional logic codes
- Subset columns (variables) & rows (observations) from a given table
- Recode, rename variables and calculate new variables
- Label variables and values

Setting Up

1. Create a working directory, often referred to as pwd (present working directory), where you will work on this assignment. For example, “assign2”. You will be writing code in assign2/

2. Create a sub directory data/. When you are writing programs, it is often a good idea to put your code in a different directory from your datasets, because if you are writing a lot of code, or working with a lot of datasets, you will stay more organized.

Pick a problem (See P3.2): Pick a content area you are interested to explore, with datasets you know you can access. Talk to your advisors if you have Research Assistantship about potential datasets in your content area. You will be working with this dataset for many of the assignments so that you can use the class to explore your interest. OR you may continue to use the google flue dataset from Assignment 1.

Getting Data: Pick a dataset. It can evolve over the semester. Requirements for the dataset

- At least 20 variables, both numerical and categorical variables
- At least 1000 observations

Midpoint check: Email Dr. Kum a short note by next week on what dataset you will be using to do assignment 2.

Describing the Data (See P3.3)

- Describe the dataset, including the number of obs and variables that are in your subset of data
- Check your data. Examine each variable for its range (min to max) or number of categories. Discuss any anomalies you see in this data (such as an abnormally high or low values)

P1. Cleaning and Manipulating the Data

Your program should do the following. Please indicate BEFORE the code in comments where each of these items occurs in your code.

- P1.1 subset vars (You need at least one of each continuous, categorical, binary, id var. If you are missing a type, create a new variable of that type.)
- P1.2 subset obs (be creative if you do not need this for your analysis)
- P1.3 rename at least 3 variables
- P1.4 label at least 3 variables
- P1.5 label values for at least 3 variables (at least one must be permanently labeled, and one must be temporarily label)
- P1.6 Recode at least 3 variables (use your imagination, if not essential to your analysis)
- P1.7 Construct at least 3 new variables (use your imagination, if not essential to your analysis)
- P1.8 Save out your new data permanently

P2. Learning Your Data (Descriptive Analysis)

Your program should also do the following. Please indicate BEFORE the code in comments where each of these items occurs in your code.

- P2.1 List each type of variable (continuous, categorical, binary, id). (see P3.3)
- P2.2 Create summary statistics for all your continuous & binary variables
- P2.3 Create tabulations for each categorical variables
- P2.4 Answer one interesting question using at least 3 variables (see P3.4)

P3. Readme file

Create a text file called readme.txt in your working directory (assign2/) Answer the following questions in the readme file send as attachment with code and output.

- P3.1 State what you were not able to complete, if any.
- P3.2 Briefly describe the problem you are interested to pursue
- P3.3 Briefly describe what data you will be using. Include number of obs, vars, and total file size.
- P3.4 List each type of variable (continuous, categorical, binary, id). (P3.4)
- P3.5 Answer one interesting question using at least 3 variables. State your questions and answer clearly. Then explain your answer. (P2.4)