

PH7525: Statistical Computing (SAS programming)

Assignment 4

Due date: 07/12/2021 @ 11:59pm

For full credit, upload: SAS Program, the log and output file. PDF or RTF/DOCX files preferred.

Part A: NHANES 2013-14 total Cholesterol data (50 points)

The [National Health and Nutrition Examination Survey \(NHANES\)](#) is a program of studies designed to assess the health and nutritional status of adults and children in the United States. NHANES is a major program of the National Center for Health Statistics (NCHS). We use the NHANES 2013-2014 total cholesterol (Laboratory Data) and demographic data for this assignment.

1. Download two SAS dataset TCHOL_H and DEMO_H from iCollege.
2. Read and merge the two datasets (make sure to use the appropriate merge method)
3. Select participants aged ≥ 20 years
4. Calculate the mean total cholesterol concentration (in millimoles per litre (mmol/L)) as follows: (i) the unweighted mean; (ii) use simple weight (WTMEC2YR); and (iii) use complex survey weight method (i.e., specify the cluster and strata variables)
5. Calculate the mean total cholesterol concentration (in millimoles per litre (mmol/L)) by 10-year age group (i.e., generate a variable for 10-year age group and use it in the CLASS statement). Calculate: (i) the unweighted mean; and (iii) use complex survey weight method (i.e., specify the cluster and strata variables)
6. Repeat #5, separately for men and women (i.e., the mean total cholesterol for men and women by age group).
7. Is the peak mean total cholesterol concentrations by age similar to men and women?

Part B: Atlanta Area Historical Daily Temperatures for the month of January (50 points)

Official data for DAILY DATA FOR A MONTH - daily maximum, minimum and average temperature (degrees F) for the Atlanta area culled from the NOAA Online Weather Data (<https://w2.weather.gov/climate/xmacis.php?wfo=ffc>).

An Excel file with three tabs (worksheets) containing daily temperature data for Atlanta area recorded for the month of January in 1980, 2000, and 2020 is on iCollege, Assignment 4 submodule (folder). There are 3 tabs (worksheets) in the Excel file: Y2020, Y2000, and Y1980.

1. Read the 3 worksheets onto your local drive
2. Check the contents of each data. Check the variable names and data type
3. RENAME, KEEP, DROP as appropriate to merge the three datasets

- Note that you can only merge two datasets at a time
 - The date variable will not work as key variable as the datasets are from different years
 - Generate a new key variable (unique id): `id = _N_ (or idnum = day(date))`
 - Use variable 'id' as a key variable
4. Find the monthly average temperature for the month of January for 1980, 2000, and 2020
 5. What do you observe from the monthly averages?
 6. Generate weekly average temperature (group variable 'id' as follows: id 1 to 7 = 1; 8 to 15=2; 16 to 23 =3; 24 to 31=4, use the IN option)
 7. Use PROC Univariate with PLOT option to generate histograms and box plots for each year