Orientation: SAS Studio Workshop

# Reading for this week:

* (Cody) Chapter 1 – Descriptive and Inferential Statistics
* (Cody) Chapter 2 – Study Designs
* (Cody) Chapter 5: Importing Data into SAS
* (OpenIntro) Chapter 1 – Introduction to Data

# Discussion for this week:

*Watch:* Calling Bullshit 4.2: Means & Medians[*https://youtu.be/mc-6-v2c4WM*](https://youtu.be/mc-6-v2c4WM)

*Prompt:* Search your media stream for the use and misuse of averages. Question the presentation of central values and identify whether or not to trust the representation of a dataset. Share one (or more) news articles that presents some form of an “average” (mean, median, or mode) and discuss whether or not that is a good representation of the underlying data. IF you cannot tell, then propose multiple possible interpretations.

# SAS Studio Workshop:

*Goal:* Start working in SAS Studio to put together data analysis workflows for statistics. Import, visualize, and describe data.

*Turn in*: SAS code and answers to questions below.

*Methods:*

1. Open SAS Studio
2. On your computer create a folder for your SAS data and code (try to keep everything under one umbrella for DAN602 here instead of spread out in different project folders)
3. Import data package from the textbook
   1. Go to: <https://support.sas.com/cody> (Log in with SAS credentials)
   2. Create a folder on your computer where you will want to store the files (create a new folder in the folder from #2)
   3. Click on the .zip file on the web page
   4. Read and accept the license agreement.
   5. From your browser, you can either save or open the downloaded zip file.
   6. Save the file to your computer if you think you may want to unzip it again later. (This is recommended because it will save you the trouble of downloading it again from support.sas.com.)
   7. Go to the location where you saved the zip file.
   8. Either double click or right click the zip file.
   9. In the interface that appears, look for a button or a selection to unzip or extract the files.
   10. You will be prompted to choose a location on your computer to put the files. Choose the folder you created for this purpose and confirm your selection.

\*\*\*In SAS Studio\*\*\*

* 1. Open the file in the folder you just set up called ‘*Create\_Datasets.sas*’
  2. Edit line 8 where this code creates a library ‘Stats’ to point to the folder of data
  3. Run the program

\* This code creates a new library called “STATS” that is available in you SAS environment and will be used throughout the textbook exercises.

1. Test out the STATS library with some descriptive statistics

Do TWO (2) of the following:

* 1. For the dataset STATS.EXERCISE calculate the mean of the pushups variable. Then modify this to calculate the mean for males and females.
  2. For the dataset STATS.GRADUATE calculate the mean of the Hours\_Studied variable. Then modify this to calculate the mean for those who graduated and those who did not.
  3. For the dataset STATS.SALARY calculate the mean Weekly\_Salary classified by Gender, Education, or Age\_Group.
  4. For the dataset STATS.RISK calculate the mean Heart\_Attack risk classified by Gender and Age\_Group.

**Put all of the code to do one of these problems into a single program file (if not already written that way) and save this file and turn it in with your answers to the questions in this document.**

1. Set up a program that imports the data file: <https://data.wprdc.org/dataset/c5ae9638-7572-4ed9-a193-384cb2ff4d03/resource/4b97f9fd-2916-45c8-acd8-22c5159ff194/download/vacantpropusps2016q2.csv>

These data are from Pittsburg area housing surveys collected by the USPS to track vacant properties.

* 1. Set up a new library for imported data (new folder + import data -> create library)
  2. Import the CSV file you downloaded
  3. Generate summary statistics (mean/stdev/etc.) for the ‘percent of vacant properties’ column. You can do this with code OR in SAS Studio: Tasks -> Statistics -> Summary Statistics -> select imported data file and “percent of vacant properties” as analysis variable

*What is the average percent of vacant properties in Pittsburg?*

*What is the range?*

*What is the “N” represent? [HINT: you may need to look at the outside metadata for this data file --* [*https://tools.wprdc.org/guides/dash-data-guide/vacantproperty/*](https://tools.wprdc.org/guides/dash-data-guide/vacantproperty/) *]*

* 1. Now, create a graph to visualize the distribution of the percent vacant variable. You can do this in SAS Studio by: Tasks -> Graph -> Histogram
     1. Modify the histogram to have 20 bins (under Appearance tab) and so the bars are blue. E.g.,



*Describe the distribution of % vacant properites in Pittsburg.*

*Is the mean of 0.04% representative of the city?*

*If you lived in a district with 0.08% vacant properties would that be “normal” for the city? Why or why not?*

* 1. **Put all of the code (import, mean, and histogram) into a single program file (if not already written that way) and save this file and turn it in with your answers to the questions in this document.**