Instructions

You will analyze neighborhood racial composition data with Microsoft Excel and Stata software programs. The data have been cleaned and prepared for you in advance. They are included in the "hw01" folder on Canvas. (In future weeks, you may need to download and upload several zipfiles).

Follow the steps below and answer each question in a separate document that you should print and bring with you to class on the assignment due date. Retain an electronic copy of your assignment and any Stata do-files or modified Excel files you use. Bring them all to class with you on a flash drive, laptop hard drive, and/or CISER account.

Some of the instructions below require you to figure out the steps as you go. If you have any questions (substantive, methodological, computational), you should post them on the course Piazza page. Fellow students may have answers for you, or it might alert me to issues I need to revisit in our class time. Ultimately, what you get from this assignment and future assignments depends on how much you effort you put into it and how actively you engage on Piazza.

This assignment is worth 4 points but you will receive 0 points if it is late or incomplete. It is due by 10:10am on February 3rd.

Question 1

In Microsoft Excel, open the file "hw01_cbsa2010.xlsx" from the HW01 file package. Each tab in the worksheet includes data from the 2010 decennial Census for a specific Core Based Statistical Area (CBSA). Within each sheet, the rows represent census tracts and the columns are variables.

Create a table that summarizes the following information for each CBSA:
- CBSA code, CBSA name
- Number of persons
- Proportion Black, White, Asian, and Hispanic

[Hint: You can look up the CBSA code online to identify its name (example: CBSA code 37980 is the Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metropolitan Statistical Area).]

[Note: the data are coded so that they can be compared across many census years, so there are some inconsistencies to take note of. There are no counts provided for Native American or Alaskan Native persons, persons identifying as "other" race, and persons identifying as multiracial. The Asian category includes Asian, Hawaiian, and Pacific Islander groups but make no distinction by Hispanic identity. The variable "pop" is a sum of the included counts, so it is not accurate of the true 2010 tract population count.]

Question 2

Urban demographers typically use the census tract as an approximation for a neighborhood. What is a census tract?

a) Using the same file in Excel:, create a summary table with the following information:
- For each of the five CBSAs, separately:
    o CBSA code
    o Number of neighborhoods
    o Neighborhood population mean and standard deviation (i.e. average number of persons per census tract)
    o Neighborhood population minimum, median, and maximum
- For all five CBSAs
    o Neighborhood population grand mean and standard deviation

b) How do your descriptive statistics about census tracts compare to the official definition posted by the U.S. Census Bureau? https://www.census.gov/geo/reference/gtc/gtc_ct.html

c) Do any of the CBSAs tend to have larger or smaller census tracts? Does the distribution of census tract size differ across CBSAs?

d) Explain what procedure you took to calculate the grand mean of the five CBSAs (there is more than one way to do it).

Question 3

Use formulas in Excel to calculate the White-Black dissimilarity index for any one of the provided CBSAs (ignoring the presence of other racial groups).

a) What is your answer?

b) Interpret the dissimilarity index you have calculated and interpret it in 1-3 sentences.

c) Does your chosen CBSA have a level of Black-White dissimilarity in 2010 that is higher than the national average reported in 2010 by Logan and Stults (2011)?

d) Provide a screenshot of your Excel sheets so that it is obvious to me you went through the trouble of doing it. Embed that screen shot in your homework submission.

Question 4

Still in Excel, now calculate the White-Black segregation exposure index. Make sure the total population within each tract includes all other groups (not just White and Black).

a) Report and interpret the exposure index you calculated in 1-3 sentences.

b) How would the interpretation of the index change if you calculate without consideration for the other racial groups?

c) Provide another screenshot of one of your sheets to show your calculations, and embed that screen shot in your homework submission.

Question 5

Now, let's move this enterprise into Stata. This first requires that you install an ".ado" user-written program. The lecture slides note how to do this, and remember that there is more information available on CISER's website, along with some even fancier tips on this: https://ciser.cornell.edu/how-to-install-stata-user-written-ado-files-on-the-research-servers/.

Your next task is to import the Excel sheet of raw data into Stata. Each sheet needs to imported separately and saved as a temporary file. Then, you should append the sheets together to create a single dataset with a variable for each CBSA. You can delete your temporary files. The lecture slides note how to do this as well. If you get stuck, you should also ask for coding support on Campuswire.

You should end up with a single dataset that has a row for each tract from all five CBSAs.

Now that you have installed "seg", run the following two commands:

. seg white black, d by(cbsa)

Take a screen shot of the output reported after running these two commands and embed it in your homework answer page.

   a) Compare your findings to the table you generated in Question 3. Do the results for that particular CBSA match your hand calculations in Excel? (Note: if neither match, something is wrong and you should check your work)
   b) Use the command "help seg" and the notes from lecture to figure out the code you need to reproduce the exposure indices you calculated. Write out the required lines of code here, and then check the results against your work from Excel.

Question 6

We are now going to compare segregation indices that are weighted and unweighted by CBSA population.

Use "seg" to estimate White-Asian exposure indices for each CBSA (hint: use the "by(cbsa)" option again). Our goal is to succinctly summarize White-Asian exposure with a single measure summarizing the five CBSAs. Consider the following two equations, where X is White-Asian exposure, j indexes each CBSA, k is the number of CBSAs, and N is the population count:

<div style="display:flex; justify-content:space-between;">

Mean:

$$\bar{X}_{unweighted} = \frac{\sum_{j=1}^{k} X_j}{k}$$

Population-weighted mean:

$$\bar{X}_{weighted} = \sum_{j=1}^{k} X_j * \frac{N_j}{N}$$

</div>

a) Report overall mean and population-weighted mean White-Asian exposure.
b) Which approach better characterizes White-Asian segregation across the five CBSAs. Why?
c) Create a table summarizing dissimilarity and exposure indices for White-Black, White-Hispanic, and White-Asian racial pairings. Report both the mean and the population-weighted mean.
d) Based on the population-weighted mean dissimilarity index, what racial pairing has the highest level of segregation across all five CBSAs generally? Do you reach the same conclusion using the population-weighted mean exposure index? Is this what you expected? Why or why not?
e) Is Figure 2 in Logan and Stults (2011) a presentation of weighted or unweighted segregation indices over time? Do you agree with the decision the authors made? Why or why not?