<u>Assignment 2</u> – Trends in NHIS
Analysis of Large-Scale Data
2025

Assignment: The goal is to analyze trends in health policy topics using NHIS. Choose one topic or health condition of interest to you that is reported for at least 10 years in NHIS (you can choose the 10+ year window from 1963 to present). Try to choose a topic that you think has been trending in some interesting way over time – this will make the exercise more engaging – and don't choose one that is sex-specific.

First, recreate your own new extract from the IPUMS site. From the site, select the health topic you want to study, as well as age, sex, and general health status (1-5 scale). Once the extract is complete, read the data into Stata and save the file.

Next, process and clean the data you read in, using the principles we have discussed in class. As discussed, treat any refused/don't know/unknown/not applicable category as missing. Make sure all the variables are comparable over time.

Once you are confident your data is clean and ready to analyze, complete the following tasks. Be sure to use comments in your code so we understand any assumptions or decisions you're making.

1. Plot the trend in your chosen health topic over the 10 year period.

2. Test if the trend significantly varies over time in two ways: regress the health condition against a <u>linear</u> time trend (the 'year' variable), and regress the health condition against <u>dummy variables</u> for each year. Briefly explain what you determined from these two tests.

3. Test if the yearly trend is significant using the linear time trend only, but also control for sex, age, and general health status. To add general health status, create a binary dummy variable for whether health is excellent or very good as opposed to good, fair, or poor. Briefly explain why we might want to control for these variables.

4. Building on the regression from Q3, test if the trend <u>differs by gender</u>. To do this, add to the regression interactions between the gender dummy and linear time trend. Provide a brief interpretation of your results.

5. Repeat Q4, but instead of using the linear time trend, assess if the trend <u>differs by gender by using the year dummies</u>. To do this, add to the regression interactions between the gender dummy and each of the year dummies (excluding the linear time trend and its interaction with gender). Provide a brief interpretation of your results.

6. Repeat Question 4, but now weight the data using the appropriate NHIS weight (perweight or sampweight). Briefly explain the reason why we might want to use weights when using data like NHIS.

7. Use the IPUMS website to learn more about your chosen outcome variable (e.g., exact wording over time, universe of respondents who the question was asked of, etc.). Thinking about the variable you've chosen and NHIS in general, <u>list at least two</u> biases, limitations, or concerns you might have about your analysis.