Final

Biost 540

General Instructions:

- The assignment is to be completed **individually**. You should not talk to other students about the project.
- As with the midterm, any general questions you have can be posted to the discussion board. For any questions that specifically pertain to your analysis, please email Katie and Yiqun.

Background of the data:

The Americans' Changing Lives (ACL) survey series is an ongoing nationally representative longitudinal study focusing especially on differences between Black and White Americans in middle and late life. These data constitute the first, second, third, fourth, and fifth waves in a panel survey covering a wide range of sociological, psychological, mental, and physical health items. Wave I of the study began in 1986 with a national face-to-face survey of 3,617 adults ages 25 and up. Subjects were re-interviewed in 1989, 1994, 2001/2 and 2011. **We will consider a (lightly edited) subset of the data with individuals who are 50 years or younger in 1986.**

ACL was designed and sought out to investigate the following: (1) the ways in which a wide range of activities and social relationships that people engage in are broadly "productive," (2) how individuals adapt to acute life events and chronic stresses that threaten the maintenance of health, effective functioning, and productive activity, and (3) sociocultural variations in the nature, meaning, determinants, and consequences of productive activity and relationships. Among the topics covered are interpersonal relationships (spouse/partner, children, parents, friends), sources and levels of satisfaction, social interactions and leisure activities, traumatic life events (physical assault, serious illness, divorce, death of a loved one, financial or legal problems), perceptions of retirement, health behaviors (smoking, alcohol consumption, overweight, rest), and utilization of health care services (doctor visits, hospitalization, nursing home institutionalization, bed days). Also included are measures of physical health, psychological well-being, and indices referring to cognitive functioning. **We will be focusing on a measure of functional impairment.**

Demographic information provided for individuals includes household composition, number of children and grandchildren, employment status, occupation and work history, income, family financial situation, religious beliefs and practices, ethnicity, race, education, sex, and region of residence.

Citation:

House, James S. Americans' Changing Lives: Waves I, II, III, IV, and V, 1986, 1989, 1994, 2002, and 2011. Inter-university Consortium for Political and Social Research [distributor], 2018-08-22. https://doi.org/10.3886/ICPSR04690.v9

Note: for the purposes of this analysis we will ignore any considerations surrounding the survey sampling design as that is beyond the scope of this class.

Variables:

Note: the data is provided in *wide* format:

1. ID: Subject ID
2. age: age of subject at baseline
3. sex: sex of subject (Male or Female)
4. race: race of subject (W = White American or AA = African American)

5. ses: socioeconomic status (Low, Middle, Upper)
6. wave (columns):
    - w1: wave 1 (1986)
    - w2: wave 2 (1989)
    - w3: wave 3 (1994)
    - w4: wave 4 (2002)
    - w5: wave 5 (2011)
7. functional impairment indicator at each wave
    - 0: no functional impairment
    - 1: at least some functional impairment

Scientific Goal:

To assess trends in the odds of functional impairment over time and determine how this effect varies across racial groups.

Specific Questions:

1. Perform an appropriate exploratory analysis of the data.
2. Propose a model that would allow you to assess whether changes in odds of functional impairment over time vary by racial group, adjusting for baseline age, sex, socioeconomic status, and **past functional impairment status** using **all available data**. Is past functional impairment status an important predictor of current functional impairment status?

In answering the remaining question, carefully consider the strategies for handling missing data in this study for your **primary analysis** and for your **sensitivity analysis**.

3. Propose a model that would allow you to examine differences in trends in functional impairment over time by racial group, adjusting for baseline age, sex, and socioeconomic status and answer the following questions:
    - Characterize functional impairment at baseline (wave 1). Is there a difference in functional impairment at baseline between African Americans and White Americans?
    - Characterize functional impairment over time. Is there a difference in the rate of change in functional impairment over time comparing the two racial groups?

Format:

In order to answer the scientific questions organize your short report as follows. Format: maximum of **3** pages, font size 12, line spacing 1.5, with the following sections:

1. **Introduction**: provide a brief background for the problem and the questions you will address
2. **Methods**: describe the statistical methods that you utilized (for both descriptive and inferential analyses). Justify any key decisions that you made.
3. **Results**: present the results of your analyses appropriate for a general scientific audience (point estimates, 95% CIs, etc.). Provide careful interpretation for any key regression coefficients that address the question(s) of interest.
4. **Discussion**: provide a brief discussion on the key results and highlight any limitations to the data and analysis
5. **Tables and Figures**: Include these at the end of your report. Provide appropriate captions and be sure to refer to them in the main text. Please note that unedited output from any statistical software is not acceptable. **You are allowed a maximum of 5 tables and figures (combined).**

6. **Code**: provide a copy of the code that you utilized for the analysis. This code should have adequate documentation so that others could replicate the analysis.

Note: pages for tables, figures and code are not counted towards your maximum of 3 pages for the extended abstract.

You may find this code to be useful for wrangling the data:

```
dat.long <- melt(dat, id=c("id", "sex", "race", "ses", "age"))
dat.long$year <- 1986
dat.long$year[dat.long$variable=="w2"] <- 1989
dat.long$year[dat.long$variable=="w3"] <- 1994
dat.long$year[dat.long$variable=="w4"] <- 2002
dat.long$year[dat.long$variable=="w5"] <- 2011
```