HAD 5744: Quantitative Methods I
Dr. Hoagland

Assignment #2

Late assignments will be penalized at the rate of 10% per day

Please submit a knitted R-Markdown file with text and code for this assignment. Don't forget to acknowledge your group members (although each of you must submit your own assignment) and to include your AI statement.

For the problem indicated with an asterisk (*), please submit a video of no more than 3 minutes explaining both how you got to your solution and what it means. Refer to the syllabus for complete instructions here.

Note: This problem continues directly from Assignment 1 – you can use your code from before to load your cleaned dataset. Recall that we are working with the NHANES and the following variables: PhysActiveDays, SleepHrsNight, Gender, Age, Education, HHIncome, Work, BMI, Diabetes, Depressed, and SmokeNow.

Problem 1: Multivariate Regression. We are interested in understanding the drivers of physical activity. That is, our dependent variable will be the count of days in a week that a respondent is physically active days, PhysActiveDays. Our primary question is: how does a person's sleep impact changes in PhysActiveDays?

a. Start with the regression you ran in part (e) of the last assignment. Update this regression with the full structure you placed in your DAG. When you do this, use robust standard errors of your choice (with a sentence defending that choice). Update the table (or add a column to the initial table). Interpret your coefficient on SleepHrsNight again. How has it changed relative to the last assignment?

b. How do we interpret the values of the other regressors (in their contexts)?

c. * Which of the covariates that you included changed the story between last week's regression and part (a), and why? Focus on the levels of one of your categorical variables. Provide some supplemental regressions or figures that show why the coefficient on SleepHrsNight changed once you included these dummy variables. Plot the relationship between that variable and both SleepHrsNight and PhysActivityDays. How does this impact the way we talk about the causal relationship between SleepHrsNight and PhysActivityDays?

d. Now include an interaction term between SleepHrsNight and a dummy variable for high income, defining high income as the top 4 income brackets in your data. Update your regression table to *appropriately* include this interaction term. Interpret the results of this coefficient and its implications for your research question.

e. Are you concerned about reverse causality in this assessment at all? If so, provide some evidence "testing" whether PhysActivityDays causes another variable in your data. Why is this test imperfect? (If you don't believe reverse causality is an issue, defend this assertion, with data if possible.)