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 healthyR data. Now we want a more detailed picture of the relationship between LOS and total health costs.

Problem 1: Multivariate Regression.

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. As you do this, consider whether any of the dependent or independent variables should have a transformation. Also report robust standard errors of your choice. Defend your choices. Update the table (or add a column to the initial table). Interpret your coefficient on LOS 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 LOS changed once you included these dummy variables. How does this impact the way we talk about the causal relationship between LOS and costs?

d. Now include an interaction term between LOS and being enrolled in any type of Medicare. 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 health costs might cause increased LOS, rather than vice versa. Why is this test imperfect? (If you don't believe reverse causality is an issue, defend this assertion, with data if possible.)