Problem one.
Given the stem.csv data set, answer the following questions:

Transform the numerical variable of stress to above and below the median.
Run a logistic model to predict the odds of high stress from sense of belonging, being prepared for math in high school (Question 3.2 – make into three categories), and the interaction being first generation (Yes/No) and SES.

- Create the plot of odds and summarize the results within context.
- Calculate the accuracy of your model by doing five-fold cross validation.
- Design a different study that would allow you to draw causal conclusions of what affects stress.
- Suppose you were the TA for this class, how would you explain the concepts of
    - Interpretation of zero intercept in logistic model
    - Meaning of Residual deviance
    - Meaning of Null deviance

Problem two.
Suppose we run multiple regression to predict UCLA GPA from high school GPA, parental income, and hours of study per week in freshman year. Suppose you create two regression models one with regular coefficients and one with standardized coefficients.

- What are the major similarities and differences between these two outputs?
- If you were to describe the results to a non-statistical audience and emphasize the most important predictor of UCLA GPA, which method would you use?

Problem three.
Read the report on the sleep data with your group and answer the following questions.

- What did you learn from this report?
- If you were working on this report what other methodology would you have used?
- Do you think what you learned from this report has the potential of application to your final project?
- Do you see any unwarranted causal conclusions in this study?

Problem four
In the prior week, we discussed multivariate regression with multiple numerical outcomes and how it is different from regular multivariate regression. In this problem, we

want you to do the relevant research on ChatGpt or google and answer the following question:

- What is the difference between regular logistic regression and multivariate logistic regression?
- Provide a real world example of multivariate regression in a field of your interest (law, medicine, entrepreneurship, business, psychology, economics, neuroscience, cognitive science) etc.

Friday – short overview of multivariate logistic

Problem five - Friday

If you are interested in learning how to actually run and interpret multivariate logistic regression, have ChatGpt create a fictitious data set in the context of diversity for you with multiple binary outcomes, numerical and categorical predictors and run a complete analysis and interpret the findings for you. Summarize what ChatGpt does as the following:

1. Research question.
2. Predictors and how they were measured.
3. Relevant outputs and plots
4. Interpretation of findings within context.