You may consult with classmates or colleagues regarding programming code. Groups of up to **four** students who are in agreement may submit a collective response, with all group members identified. The response of each group must be the independent work of that group. Each group member should have made a substantial contribution to the project.

Answers must be typed and **no response should be longer than 4 single-spaced pages (a minimum 0.5’’ in all margins) including tables, of 11 point font excluding the appendix.** Please submit well-labeled computer output for one of your most important models as an appendix.

The dataset you will use is an imputed dataset from the NHANES II mortality follow-up study that began in 1976. The dataset is located on the “Assignments” tab under “Final Project 2022”.

The goal is to:

Examine the association between alcohol intake (**BOOZE**) and death from cancer (ICD9 140-208) during follow-up.

Requirements:

-There are three common modeling approaches for examining these associations.

You can use 1) Logistic regression, 2) Poisson survival analysis, and 3) the Cox model.

Choose **one** of these as your **primary** analysis and use the other two as sensitivity analyses.

Justify your choice, discussing the advantages and disadvantages.

-Please examine confounding in the association between alcohol consumption and cancer mortality.

[NOTE: If there is missingness in the variables that your group chose to include in the model, you can use complete case analysis. You do not have to do any imputation.]

-Please examine effect modification by **sex** using these two methods for your **primary** model: 1) stratifying by potential effect modifier, and 2) including the product term, and discuss the differences in the results.

-Be sure to check for proportionality if you are fitting a proportional hazards model, and to check for nonlinearity for your **primary** model only. Describe how you are checking for non-proportionality and nonlinearity. If either condition is present, make an appropriate decision as to how to deal with it, and justify that decision.

-Describe the extent to which the results of your sensitivity analyses (the other two modeling approaches) differ from the main analysis.

In reporting your results, please:

1. Write a Methods section which concisely describes and justifies the methods of analysis used, the variable codings used in final models, methods for selection of confounders, methods for handling non-proportionality and non-linearity, and any other issues of methodologic importance that arose while conducting the analysis. (recommended length: 0.5-1 page)
2. Produce tables that summarize the results of the analysis (recommended length: 1-1.5 pages):

* Table 1, describing relevant basic characteristics of study participants according to exposure categories
* Table 2, presenting the age-adjusted and multivariable-adjusted results of the association generated from each of the 3 models
* Table 3, giving the multivariable-adjusted results in each category of the effect modifier (sex), and p-values for interaction (using the primary model of your choice only).

3. Write a Results section that concisely summarizes the data presented in the tables (recommended length: 0.5-1 page).

4. Write a Discussion section that gives bottom-line conclusions from the analysis, evaluates the results with respect to the presence of residual confounding, unmeasured confounding, and selection bias, discusses the sensitivity analyses, discusses any other limitations of the study that are likely to have materially compromised the internal or external validity of the findings, and describes further studies that could be conducted to address each limitation noted. In the limitations, you may state biases that, in your judgment, are not likely to have materially compromised the internal or external validity of the findings, but do not discuss these at length. The entire discussion section should be limited to 0.5 page.