

Project for STAT 704

You will analyze a data set of your choosing, using a multiple regression method or methods we have discussed in this class.

The data set could be one that is of personal or academic interest to you and that fits one of the multiple regression analyses studied in STAT 704. It should be a real data set which you have not analyzed before and which has not been analyzed in a textbook (using the methods of STAT 704, anyway). It should include a response variable (outcome variable) and SEVERAL potential independent variables (predictors). I recommend picking a data set that has *at least* 30 observations. Try to pick a data set from an area that you have an interest in, whether finance, health, politics, sports, music, economics, etc., so that you can write passionately about your analysis and your conclusions from your study.

Project Part I– Data Set Proposal/Description:

You should write a one- to two-page typed description of the data set you propose to study. You should include details about the response variable, about (potential) predictor variables, and about the number of observations. If there are issues such as obvious potential model violations, multicollinearity, expected interaction effects, etc., comment on these. Discuss the source of the data set, and whether the data come from an experimental or observational study.

You should also include a general proposal for what sort of analysis you plan to do with this data set. If there are any hypotheses/research questions that are of interest from the beginning, you should mention those.

In addition, please include (the plots and data listing do not count against the page limit):

- A printout of the data set (or if it is quite large, a selected part of the data set).
- Separate scatterplots of the response Y against *each* potential predictor variable
- Pairwise scatterplots of *each pair* of potential predictor variables

Briefly discuss any initial conclusions about the analysis that can be gleaned from these plots.

This part is due on or before Friday, October 16, 2026.

Project Part II – Written Report:

You should write a concise report summarizing your analysis. The report should be no longer than six (typed) pages, not counting any SAS/R output, graphs, etc., which you may wish to include as support or illustration for your analysis.

The style of the report is up to you, but the best reports will address many of the questions and details studied in class when we discussed the relevant type of analysis.

Some things to include (depending on the data set and choice of model) might be:

- An introduction and discussion of the data set itself
- A statement of the fitted model, with any relevant interpretations
- Summary results for any relevant hypothesis tests or confidence intervals
- Discussion of variable selection or model building (if appropriate)
- A summary of model assumptions and any violations you may have seen
- Any remedial action you took to fix such violations

- Your overall conclusions about the data, based on your analysis

Part II of the project (the final written project report) will be due ***on or before Monday, November 23, 2026.***

This part is due on or before Friday, November 11, 2026.

Project Part III – Oral “Defense”:

You will arrange a time for a brief oral presentation and defense of your project report with me in my office. This will be after the written report is turned in, and I will have already read your written report. The oral defense **should be scheduled between Nov. 30 and Dec. 7** and the date and time can be arranged with me via email.

For the brief presentation, you should explain the purpose, methods, and results of your regression analysis. This should take *no longer* than five minutes and should be done using only orally and possibly on the chalkboard (no slides). Following this, I will ask you oral questions about your analysis, modeling choices, findings, etc., and you will answer these orally (again, using only the chalkboard if necessary, but no computer or other aids). The entire oral presentation and defense should not take longer than 30 minutes.

The course project will count for 10% of your final grade (1% for Part I, 5% for Part II, and 4% for Part III).