

STATS 201 Lab Class 4

CAFE Voting Analysis

This question is concerned with a vote on a bill in the US senate on the Corporate Average Fuel Economy (CAFE) standard in 2002. The bill was widely held to be beneficial to automotive manufacturers, as a vote of NO would have forced them to increase fuel economy across their fleets. It is of interest to determine whether a senator's voting decision was related on their political ideology, and/or the amount of money contributed to them by individuals or political action committees associated with the automotive industry. The variables are as follows:

- **party**: indicating the political party of a senator; either **D** for Democrat or **R** for Republican.
- **contributions**: the total lifetime dollar amount contributed to a senator by the automotive industry, rounded to the nearest ten thousand dollars, and given in tens of thousands of dollars (e.g., a value of 2 indicates \$20,000 worth of contributions).
- **yes**: the number of senators who voted YES on the bill for a particular combination of **party** and **contributions**.
- **no**: the number of senators who voted NO on the bill for a particular combination of **party** and **contributions**.

The data are as follows:

party	contributions	yes	no
D	0	8	21
D	1	2	7
D	2	7	2
D	3	2	1
R	0	3	3
R	1	17	1
R	2	13	1
R	3	10	1

We are interested in determining if the amount of contributions from the automotive industry is related to the probability a senator votes YES on the bill. We are also interested to determine if senators from one party are more likely to vote YES on the bill than the other. Also, does the effect of contributions depend on the political party? Conduct a full analysis, and include **Methods and Assumption Checks** along with an **Executive Summary**.

In your report, add to the plot code that is provided (to plot contributions vs proportion voting YES) to also show the fit of your chosen model.

Hints

General comments

- Your assignment should be written using R Markdown in RStudio. It should include the code you used and its output, including plots.

- There are many examples of **Methods and Assumption Checks** and **Executive Summaries** in your course book. It is a good idea to find a case study with a similar analysis, and use this to guide you, but remember to write things in your own words. Additionally, each data set may have its own specific questions of interest, so do not only base your answers on examples from other case studies.
- There are some more specific guidelines for the **Methods and Assumption Checks** and **Executive Summaries** below.

Methods and assumption checks

Unless told otherwise, this should include the following:

- A brief comment on any plots of the data.
- A brief description of how you arrived at your final model. For example, if you fitted multiple models, you may comment on what changes you made, and why.
- A brief description of any concerns with the model assumptions. If you have no concerns then it is enough to say “All model assumptions appear to be satisfied.”
- The model equation, making sure to define all model terms.
- Reporting the R-square of the model.

Executive Summary

Unless told otherwise, this should include the following:

- A brief statement summarising what your analysis was investigating, and the model that was used.
- Interpretations of the important findings. For example, was there evidence to suggest that your explanatory variables were related to the response variable? Where appropriate, use confidence intervals to express the magnitude of the relationship.
- Answers to any specify research questions.
- This should be written without too much technical detail. An intelligent person who is *not* a statistician should be able to understand what your findings were. Imagine that you are trying to explain the results of your analysis to a friend who is not taking this course.

For the plot

- Try to predict the probability of voting YES for contributions ranging from 0 to 3. You can create such a vector using `seq(0, 3, length.out = 1000)`.
- Use the `lines()` function to superimpose your lines onto the plot.