Homework 3 Solutions:

Reading Between the Lines:

Predicting Supreme Court Votes Based on Oral Argument Emotions

Quantitative Methods II - Winter 2025

March 12, 2025

Acknowledgments

Insert your acknowledgments here.

3.1 Preliminaries

Q1: Write a concise, 3-sentence description of the dataset as if you were explaining it to a colleague. What kind of information does each row represent?

A1:

Q2: What is the unit of observation in this dataset? In other words, what does each observation (or row) correspond to?

A2:

Q3: What does it mean when a Supreme Court justice votes in favor of the petitioner? Why is accurately predicting this decision important for policy and legal outcomes? (Hint: Refer to the authors' article for insights.)

A3:

Q4: How many missing values are there in the dataset? Identify which variables have missing data and explain why this might matter for your analysis.

A4:

3.2 Getting to Know the Data and Descriptive Statistics

Q5: Describe the general distribution of the variables `petitioner_vote`, `pitch_diff`, and `petitioner_harvard_pos`. Are there any patterns in how these values are spread out? Do you notice any missing data or extreme values that might be important for analysis?

A5:

Q6: The dataset includes variables based on justices' speech patterns and word choice during oral arguments. How was `pitch_diff` measured, and why was it standardized? How was `petitioner_harvard_pos` created, and what are some potential challenges of using a predefined list of positive words? Could this approach misinterpret tone or context, such as when a justice uses sarcasm?

A6:

Q7: The `petitioner_harvard_pos` variable is based on a predefined dictionary of positive words. What are some ways this method might introduce errors? How could these errors impact the interpretation of the results?

A7:

Q8: The bar chart compares the proportion of votes in favor of the petitioner across three chief justices, depending on whether the Solicitor General submitted an amicus brief. What trends do you observe in how each justice responded to amicus support? What factors might explain differences in their voting patterns?

A8:

[Include HW3_Fig1.png]

Q9: The second bar chart examines how pitch differences in justices' speech relate to their votes across different court periods. What does the chart suggest about the relationship between pitch and voting behavior during the Burger, Rehnquist, and Roberts courts? Do you see any shifts in this pattern over time?

A9:

[Include HW3_Fig2.png]

3.3 Regression Analyses

Q10: What do you learn from the first regression (predicting `petitioner_vote` using `pitch_diff` and `pr_petitioner_pos`)? How do these predictors relate to justices' voting behavior?

A10:

[Include HW3_Table1.txt]

Q11: Why is it important to control for justice-specific effects when running the regression? How do the results change when these controls are added?

A11:

Q12: How does adding court term indicators change the model's explanatory power? What does the adjusted R^2 tell us?

A12:

Q13: In the interaction model for `pitch_diff` and `court_period`, what patterns do you observe? Does the relationship between vocal pitch and voting behavior change across different Supreme Court periods?

A13:

[Include HW3_Fig4.png]

Q14: Compare the six progressively built regression models. What do you learn from the addition of each new predictor or interaction term? How does controlling for these factors refine our understanding of justices' voting behavior?

A14:

[Include HW3_Table2.txt]

Q15: Based on the interaction plots, how does the Solicitor General's amicus support influence justices' voting behavior across different court periods? Similarly, how does the relationship between vocal pitch and sentiment differ based on the use of positive words?

A15: IGNORE. Not answerable.

3.3 Outlier Analysis and Threats to Validity

Q16: Based on the outlier diagnostics (Studentized residuals, leverage, Cook's distance, and DFFITS), which observations stand out as extreme? Do any specific cases appear as outliers across multiple measures?

[Include HW3_Fig6.png]

A16:

Q17: How does removing outliers affect the regression results? Compare the coefficients and standard errors from the full dataset and the outlier-excluded model. Does removing outliers meaningfully change the conclusions, or are the results largely robust?

[Include HW3_Table3.txt]

A17:

Q18 (BONUS): Suppose that the measured pitch differential is related to the true difference in pitch as follows:

$$\text{pitch\_diff}_i = \delta_0 + \delta_1 \text{True\_pitch\_diff}_i + u_i$$

where u_i is an error term with mean zero and is independent of the true pitch differential. How does this type of measurement error affect your estimates? What would happen if u_i had a nonzero mean? What might this indicate about the accuracy of the sound-detection algorithm?

A18:

Q19 (BONUS): Identify two additional sources of measurement error in this study. For each one:

- Name the type of measurement error (e.g., random error, systematic bias).
- Explain how it might arise in the data collection or processing.
- Describe its likely effect on the regression results and the interpretation of findings.

A19:

Q20 (BONUS): Beyond measurement error, what other limitations could affect the validity of this analysis? Consider potential omitted variables, issues with the study design, or alternative explanations for the observed relationships. How would you refine the research to better address the initial question of whether voice pitch predicts justices' votes?

A20: