Reading Between the Lines:
Predicting Supreme Court Votes Based on Oral Argument Emotions

GPCO 454 - Quantitative Methods II - Homework 3

You've recently joined a prestigious think tank focused on judicial decision-making. Your latest task involves analyzing whether the emotional tone expressed by U.S. Supreme Court justices during oral arguments can predict how they will ultimately vote. Court observers have long speculated that vocal inflection, particularly pitch variations, reflects emotional arousal, often triggered by disagreement with counsel or colleagues.

Your project builds on cutting-edge research by Dietrich, Enos, and Sen (2019), who used audio data from oral arguments to measure vocal pitch as a proxy for emotional arousal.[1] Their findings suggest that justices are more likely to vote against the side that elicited emotional responses during oral arguments. But the original study left some questions open: How robust are these findings across different types of cases? Do emotional reactions vary by justice ideology or seniority? Can we detect systematic biases in judicial decision-making based on emotion?

To answer these questions, you will complete the following tasks:

1. **Data Exploration:** Familiarize yourself with the dataset, including key variables and descriptive statistics.
2. **Regression Analysis:** Estimate models to test whether emotional arousal (measured by pitch) predicts justices' votes, exploring interactions and controlling for potential confounders.
3. **Model Comparison:** Evaluate the performance of different models to assess the robustness of findings.
4. **Validity Checks:** Identify potential outliers, discuss measurement error, and consider threats to causal inference.

[1] Dietrich BJ, Enos RD, Sen M. Emotional Arousal Predicts Voting on the U.S. Supreme Court. Political Analysis. 2019;27(2):237-243. doi:10.1017/pan.2018.47

1 General Instructions

This homework is designed to help you apply concepts and tools to a real-world research question. Your goal is to use data and multiple regression to analyze whether the emotional tone expressed by U.S. Supreme Court justices during oral arguments can predict their eventual vote.

Here are a few key points to keep in mind:

1. **Focus on Clarity and Interpretation**: The primary objective is to interpret the results of your analysis in the context of the research question. Explain your findings clearly and concisely.

2. **Follow Best Practices in R**: Use the coding techniques from lab sessions to manage data, run regressions, and produce clear outputs.

3. **Submission Format**: Submit two files electronically through Canvas:

- A PDF document (`LastName_PID_HW3.pdf`) containing your answers to all questions, including any written explanations, tables, and figures.
- A clearly labeled R script (`LastName_PID_HW3.R`) that contains all the code used to load data, run analyses, and generate figures and tables.

Both files must be well-organized, and your R script should run without errors. Assume the grader will download the script into the same directory as the data file. Your R script needs to run as a standalone file. Once complete, test your script by running it all the way through to ensure it works from start to finish without any pauses or interruptions.

4. **Collaboration Policy**: You may collaborate with others but must acknowledge all collaborators and tools. If you used AI tools like ChatGPT, specify how they were used.

Example Acknowledgment (to be included at the beginning of your PDF submission):

> **ACKNOWLEDGMENTS:** I, Alexander J. Norling, acknowledge working with John Doe and Jane Smith. I also used ChatGPT for help with data manipulation and improving plot aesthetics. All analyses and conclusions are my own.

5. **Late Submissions**: Late submissions will not be graded. Ensure that your files are submitted on time; the timestamp on the Canvas server will be used to determine timeliness.

6. **Ask for Help**: If you encounter any difficulties, don't hesitate to reach out during office hours or review sessions.

2 About the Data

For this assignment, you will work with a single dataset derived from the original analysis by Dietrich, Enos, and Sen (2019). This dataset has been prepared to facilitate your analysis of the relationship between vocal pitch during Supreme Court oral arguments and justices' votes.

- **Supreme Court Oral Argument Dataset:** This dataset contains observations for each case, including key variables such as vocal pitch (`pitch_diff`), vote outcome (`petitioner_vote`), word counts directed at petitioners and respondents (`petitioner_wc` and `respondent_wc`), emotional tone indicators (`petitioner_harvard_pos` and `respondent_harvard_pos`), case characteristics (`sgpetac`, `term`), and justice demographics (ideology, seniority), among others.

Data Sources and Documentation:

- The dataset is sourced from the replication files provided by Dietrich, Enos, and Sen (2018), available on the Harvard Dataverse.[2]
- A codebook describing variables used for Homework 3 is provided in the file `Codebook_HW3.docx`.

Now is a good time to check the dataset and accompanying codebook provided for this assignment. The dataset contains multiple variables, and the codebook explains key definitions. In practice, data in large projects is often presented in similarly complex and detailed ways. While this may seem daunting at first, these documents are very helpful for ensuring your analysis is accurate. It's crucial to spend time understanding them to avoid errors. Open the dataset to get a sense of what's inside—variable names, data types, and any missing or special values. Remember, you cannot blindly proceed with your regression analysis without first comprehending the structure and content of your data.

[2] Dietrich, Bryce J.; Enos, Ryan D.; Sen, Maya, 2018, "Emotional Arousal Predicts Voting on the U.S. Supreme Court", https://doi.org/10.7910/DVN/JFU71R, Harvard Dataverse, V1

3 Your Tasks

1. R Script. Your R script should thoroughly address **each and every numbered list** outlined in the tasks below. Document the steps you take in your script with meaningful comments that explain your code and provide context for each step. The goal is to ensure that someone with a reasonable understanding of R can follow your logic, understand how you implemented each task, and verify your results. Pay special attention to covering all aspects of the task instructions to demonstrate mastery of the concepts. Save your script as `LastName_PID_HW3.R` (e.g., `Ravanilla_A12345678_HW3.R`). Ensure there are no spaces in the filename.

2. PDF Submission. Your PDF submission should clearly and concisely answer all questions (e.g., **Q1**, **Q2**, etc.) listed under each section. Organize your answers logically and ensure that any figures or tables generated as part of your analysis are properly labeled, referenced, and explained. Save your PDF as `LastName_PID_HW3.pdf` (e.g., `Ravanilla_A12345678_HW3.pdf`). Ensure there are no spaces in the filename.

3.1 Preliminaries

1. Set your working directory and load all necessary packages in your R script. **IMPORTANT:** Comment out the `setwd()` command with a # before submitting your R script to save time when we run and check your script.

2. **Load the Dataset.** Load the dataset `justice_results.tab` into R and assign it to a variable called `justice_data`. Since the dataset is tab-delimited, use the `read.table()` function with the appropriate parameters to ensure correct formatting. Specifically:

   Use the following code in your R script:

   ```
   # Load the dataset
   justice_data <- read.table("justice_results.tab",
                              header = TRUE,
                              sep = "\t",
                              encoding = "ISO-8859-1")
   ```

   This will load the dataset into R and display its structure to ensure that all variables are imported correctly.

3. **Check Structure:** Display the structure of the dataset to view the number of observations, variable names, and data types.

4. **Summary Statistics:** Generate summary statistics for all key variables to understand their distribution, including the range, mean.

5. **Count Unique Cases:** Identify the number of unique cases in the dataset based on the `docket` variable, which represents individual Supreme Court cases.

6. **Check Missing Values:** Assess missing values across all variables to identify potential gaps in the dataset that might affect your analysis.

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?

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

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.)

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.

3.2 Getting to Know the Data and Descriptive Statistics

In this section, you will explore the dataset further, generate key descriptive statistics, and visualize patterns across important variables.

1. **Examine Key Variables:** Generate summary statistics for the following key variables: `petitioner_vote`, `pitch_diff`, and `petitioner_harvard_pos`. This will help you understand their distributions, including means, ranges, and missing values.

2. **Create New Variables:**

   - `high_pitch_diff`: A binary variable equal to 1 if the variable `pitch_diff` is above its mean value and 0 otherwise.
   - `court_period`: A categorical variable identifying the chief justice period (Burger, Rehnquist, or Roberts), based on the term of each case. Use the following periods:
     - **Burger Court:** 1969–1985 terms
     - **Rehnquist Court:** 1986–2004 terms
     - **Roberts Court:** 2005 term onward
   - Verify that the newly created variables are categorical by using the `table()` function to display their frequency distributions.
     Note that Supreme Court terms begin in October and last until September of the following year.

3. **Visualize Amicus Support and Voting Patterns:**

   (a) Filter the dataset to include only cases involving the three Chief Justices: Warren E. Burger, William Rehnquist, and John Roberts.

   (b) Select only the relevant variables: `justiceName`, `petitioner_vote`, and `sgpetac`, and remove any missing values.

   (c) Calculate the proportion of votes in favor of the petitioner for each justice, separately for cases where the Solicitor General submitted an amicus brief (`sgpetac = 1`) and where they did not (`sgpetac = 0`).

   (d) Create a bar plot:

   - The x-axis should indicate whether the Solicitor General submitted an amicus brief (`sgpetac`).

- The y-axis should represent the proportion of votes in favor of the petitioner.
- Use separate panels (facets) for each Chief Justice.
- Assign distinct colors to the two categories of `sgpetac`: blue for "No Amicus" and red for "Amicus."
- Ensure the y-axis ranges from 0 to 1 for proper scaling.

(e) Save the plot as `HW3_Fig1.png` with dimensions 6x4 inches and a resolution of 300 DPI.

4. **Visualize Pitch Differential by Court Period:**

   (a) Filter the dataset to include only cases involving the three Chief Justices: Warren E. Burger, William Rehnquist, and John Roberts.

   (b) Select only the relevant variables: `justiceName`, `petitioner_vote`, `pitch_diff`, and `term`, and remove any missing values.

   (c) Create a new binary variable, `high_pitch_diff`, which equals:

   - "Above Avg. Pitch Differential" if `pitch_diff` is greater than or equal to its mean value.
   - "Below Avg. Pitch Differential" otherwise.

   (d) Create a categorical variable, `court_period`, based on the term of each case:

   - **Burger Court:** 1969–1985 terms
   - **Rehnquist Court:** 1986–2004 terms
   - **Roberts Court:** 2005 term onward

   (e) Calculate the proportion of votes in favor of the petitioner for each combination of `court_period` and `high_pitch_diff`.

   (f) Create a bar plot:

   - The x-axis should indicate whether the pitch differential was above or below the mean (`high_pitch_diff`).
   - The y-axis should represent the proportion of votes in favor of the petitioner.
   - Use separate panels (facets) for each `court_period` (Burger, Rehnquist, Roberts).

- Assign distinct colors to the two categories of `high_pitch_diff`: blue for "Below Avg. Pitch Differential" and red for "Above Avg. Pitch Differential."
- Ensure the y-axis ranges from 0 to 1 for proper scaling.

(g) Save the plot as `HW3_Fig2.png` with dimensions 6x4 inches and a resolution of 300 DPI.

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?

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?

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?

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?

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?

3.3 Regression Analyses

In this section, you will conduct regression analyses to explore how justices' speech patterns during oral arguments relate to their voting behavior.

1. **Create a New Variable:** Construct a variable, `pr_petitioner_pos`, that measures the difference in the proportion of positive words used when addressing the petitioner versus the respondent. This is calculated as the proportion of positive words in petitioner-directed speech minus the proportion of positive words in respondent-directed speech.

2. **Estimate a Basic Regression Model:** Run a linear regression where `petitioner_vote` is the dependent variable and `pitch_diff` and `pr_petitioner_pos` are the independent variables. Call this model m_3_1.

3. **Incorporate Justice-Specific Effects:** Convert `justiceName` into a categorical variable (factor) and include it in the regression as a set of indicator variables to control for justice-specific tendencies. Call this model m_3_2.

4. **Control for Court Term Effects:** Extend the previous regression by adding term-specific indicators to account for time-period effects. Call this model m_3_3. Save the regression results from models m_3_1 to m_3_3 in a formatted table (`HW3_Table1.txt`).

5. **Examine Court Period Interactions:** Create a categorical variable `court_period` that classifies cases into one of three Supreme Court periods:

   - **Burger Court:** 1969–1985
   - **Rehnquist Court:** 1986–2004
   - **Roberts Court:** 2005–present

   Run two regressions:

   - A baseline model using `pitch_diff` and `pr_petitioner_pos`.
   - An interaction model where `pitch_diff` is interacted with `court_period`.

   Visualize the interaction effect using a plot and save the figure as `HW3_Fig3.png`.

6. **Progressive Model Building:** Estimate six regressions with increasing complexity:

   - **Model 1**: A simple regression with `pitch_diff`.
   - **Model 2**: Add `pr_petitioner_pos`.
   - **Model 3**: Add an indicator for whether the Solicitor General submitted an amicus curiae brief supporting the petitioner (`sgpetac`).

- **Model 4**: Includes court period indicators.
- **Model 5**: Builds on Model 4 by adding an interaction between `pitch_diff` and `court_period`.
- **Model 6**: Builds on Model 4 by adding an interaction between `pitch_diff` and `pr_petitioner_pos` (without including the interaction from Model 5).

Save the regression results in a formatted table (`HW3_Table2.txt`).

7. **Interaction Effect Visualization:**
   - Create a plot showing how the probability of a vote in favor of the petitioner changes depending on the interaction between `pitch_diff` and `court_period`. Save this plot as `HW3_Fig4.png`. (Note: This should look similar though not exactly the same as `HW3_Fig3.png`).
   - Create another plot showing how the probability of a vote in favor of the petitioner changes depending on the interaction between `pitch_diff` and `pr_petitioner_pos`. Save this plot as `HW3_Fig5.png`.

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?

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?

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

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?

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?

Q15: IGNORE/DO NOT ANSWER: 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?

3.4 Outlier Analysis and Threats to Validity

In this section, you will conduct an outlier analysis on the final regression model and evaluate potential threats to the validity of your estimates.

1. **Outlier Diagnostics:** Conduct an outlier analysis on the final regression model. Specifically, calculate the following statistics for each observation:

    - **Studentized Residuals:** Measures how much each observation's error deviates from the expected error, standardized by its own variance.
    - **Leverage:** Assesses how much an observation influences the model by examining its distance from the average predictor values.
    - **Cook's Distance:** Identifies observations that, if removed, would significantly change the regression coefficients.
    - **DFFITS:** Measures how much an observation affects its own predicted value in the model.

    Use established statistical thresholds to identify potential and extreme outliers:

    - Observations with **Studentized Residuals** greater than 2 (in absolute value) may be outliers in terms of prediction errors.
    - Observations with **Leverage** greater than $\frac{2k+2}{n}$ (where k is the number of predictors and n is the sample size) are influential in determining regression coefficients.
    - Observations with **Cook's Distance** greater than $\frac{4}{n}$ are highly influential.
    - Observations with **DFFITS** greater than $2\sqrt{\frac{k}{n}}$ indicate significant influence on the model's predictions.

    Finally, classify observations as **outliers** if they exceed any of the above thresholds and as **egregious outliers** if they exceed all of them.

2. **Visualizing Potential Outliers:** Create a scatter plot to visualize potential outliers using leverage and DFFITS values. Each point in the plot represents an observation, with the following elements:

    - The **x-axis** represents the absolute value of DFFITS, indicating how much an observation affects its own predicted value.
    - The **y-axis** represents leverage, measuring how much an observation influences the regression model.

- Points identified as **outliers** based on any diagnostic measure are highlighted in red, while non-outliers remain in black.
- Observations classified as **egregious outliers** (exceeding all threshold criteria) are labeled with their respective observation numbers for further inspection.
- Dashed reference lines mark the statistical thresholds for leverage (blue) and DFFITS (red), helping to identify extreme points.

Save the plot as `HW3_Fig6.png`.

3. **Comparing Regression Models with and without Outliers:** Estimate two regression models to assess the impact of outliers on the results:

   - **Full Dataset Model:** Using the specification in Model 6 under Section 3.3, run a regression using the full dataset without filtering out any observations.
   - **Outlier-Excluded Model:** First, create a new dataset `clean_data` that excludes all observations flagged as outliers in `outlier_df`. Then, rerun the regression on this filtered dataset.

   Generate a regression table comparing the results from both models and save it as `HW3_Table3.txt`.

- **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?
- **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?
- **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?
- **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.

- **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?