# Problem Set 1

## Sources of Empathy in the Circuit Courts

[YOUR NAME]

Due Date: 2025-02-14 by 11:59PM ET

**If you utilized any assistance to complete this problem set (e.g., classmate, tutor, AI), you should cite it here:** Type your answer here.

---

In this problem set, we will analyze the relationship between the gender composition among a judge's children and voting behavior among circuit court judges. In a recent paper, Adam N. Glynn and Maya Sen argue that having a female child causes circuit court judges to make more pro-feminist decisions. The paper can be found at:

> Glynn, Adam N., and Maya Sen. (2015). "Identifying Judicial Empathy: Does Having Daughters Cause Judges to Rule for Women's Issues?." *American Journal of Political Science* Vol. 59, No. 1, pp. 37–54.

The dataset `dbj.csv` contains the following variables about individual judges:

| Name | Description |
|---|---|
| name | The judge's name |
| child | The number of children each judge has. |
| circuit.1 | Which federal circuit the judge serves in. |
| girls | The number of female children the judge has. |
| progressive.vote | The proportion of the judge's votes on women's issues which were decided in a pro-feminist direction. |
| race | The judge's race (1 = white, 2 = African-American, 3 = Hispanic, 4 = Asian-American). |
| religion | The judge's religion (1 = Unitarian, 2 = Episcopalian, 3 = Baptist, 4 = Catholic, 5 = Jewish, 7 = Presbyterian, 8 = Protestant, 9 = Congregationalist, 10 = Methodist, 11 = Church of Christ, 16 = Baha'i, 17 = Mormon, 21 = Anglican, 24 = Lutheran, 99 = unknown). |
| republican | Takes a value of 1 if the judge was appointed by a Republican president, 0 otherwise. Used as a proxy for the judge's party. |
| sons | The number of male children the judge has. |
| woman | Takes a value of 1 if the judge is a woman, 0 otherwise. |
| X | Indicator for the observation number. |
| yearb | The year the judge was born. |

## Question 1 (5 points)

Load the `judges.csv` file into a data frame called `judges`. Create a cross-tab (of proportions, not counts) of judge gender on the rows and whether the appointing president was Republican on the columns. Save this table with the name `gender_rep_table`. In your write-up, answer the following questions:

- How many judges are in this data set?
- What proportion of the judges are men?
- Is the party composition different for male and female judges?

## Answer 1

Write answer here.

## Question 2 (4 points)

Our outcome in this exercise will be the proportion of feminist rulings on issues related to gender, `progressive_vote`. Create a nicely formatted histogram of this variable and provide a written summary of this graph. Roughly speaking, where is the region of highest density of this variable?

NOTE: nicely formatted means axis labels and a main title that don't contain random R syntax and informative labels.

## Answer 2

Write answer here.

## Question 3 (6 points)

Next, we consider differences between some groups. Create a new factor variable called `gender_party` that takes on separate values for each of the four groups:

- `"Dem. Woman"` for women appointed by Democratic presidents
- `"Rep. Woman"` for women appointed by Republican presidents
- `"Dem. Man"` for men appointed by Democratic presidents
- `"Rep. Man"` for men appointed by Republican presidents

Calculate the mean of `progressive_vote` in each of these groups and save this vector as `gender_party_means`. Plot these means using a barplot.

Briefly interpret the results of the analysis. For example, do any of the results surprise you? Does it appear that partisanship, gender, or both contribute to progressive voting patterns? Should we interpret any of these effects causally? Why or why not?

## Answer 3

Write answer here.

## Question 4 (4 points)

What is the difference in the proportion of pro-feminist decisions between judges who have at least one daughter and those who do not have any? To compute this difference, first create a variable called `any_girls` that is 1 when the judge has at least 1 girl and 0 otherwise. Then, create a subset of the data called `parents` that contains judges that have at least one child. Create an object called `ate` that is the difference in means of `progressive_vote` between judges that have at least one girl versus those that have no girls among those judges with any children.

Why might we worry about interpreting this estimate causally, considering number of children as a possible confounder?

## Answer 4

Write answer here.

## Question 5 (6 points)

Given that the number of children might be a confounder for the relationship between number of girls and voting, let's estimate the effects using statistical control for the number of children using the following steps:

- Create one subset of the data called `girls_123` that restricts to judges with one, two or three children and have at least one girl.
- Create another subset of the data called `nogirls_123` that restricts to judges with one, two or three children and have no girls.
- Calculate the mean of `progressive_vote` within levels of the numbers of kids (`num_kids`) for each of these subsets and save these vectors as `girls_vote_by_nkids` and `nogirls_vote_by_nkids`.
- Use these two vectors to estimate the average treatment effect within levels, saving this vector as `ate_nkids`.

Are these estimated effects largely similar or largely different than what you found using all of the data? What assumption do you need to make to interpret these effects causally? Do you think it is plausible in this case?

## Answer 5

Write answer here.

## Question 6 (EXTRA CREDIT, 5 points)

**This problem is optional. Any points earned on this problem can be applied to lost points on other parts of the problem set. You cannot earn more than the maximum score on the problem set.**

Let's consider the design of this study. The original authors assume that, conditional on the number of children a judge has, the number of daughters is random (as we did in the previous question). If this is true, half of a judge's children should be female, on average. A deviation from this proportion could indicate that a gender preference among judges due a stopping rule such as "have children until we get one girl," which would violate the randomization assumption.

To check this assumption, use the subset of judges with at least one child and create and save a vector that contains the total number of girls (across all judges) for each level of the number of children. Create and

save another vector that is the total number of children (across all judges) for each level of the number of children. (HINT: you might check out the `sum` function to help carry out these steps.) Next, divide the total number of girls vector by the total number of children vector to create a new vector that is the proportion girls for each family type. Create a barplot that plots these proportions on the y-axis with the number of children on the x-axis. This barplot should have (a) informative labels on each axis, (b) a y-axis range that runs from 0 to 1, and (c) a horizontal line at 0.5 to compare against. Does it appear that there is strong gender preference/selection happening among judges?

## Answer 6

Write answer here.