Analysis of Political Party's Spending Behavior

Nathaniel Marcuson, Bella Lin, Yanni Zhou, Jeremy Lu

June 1, 2021

1 Introduction

According to the nonpartisan Center for Responsive Politics, in the 2020 election cycle, the amount spent on both the presidential and congressional campaigns hit nearly $14 billion, and Democrats have nearly doubled their spending compare to that of the Republicans. In the final week of the election, Democrats devoted $6.9 billion while Republicans put in $3.8 billion into the 2020 fight. With these figures on money spending, we build our project to examine the spending behaviors of the two political parties.

One previous study conducted by David Breaux titled as *"It's Money that Matters"* in 1991 finds that by increasing their own level of campaign spending, candidates are able to negatively affect the opponent's share of the primary vote. Another study conducted in 2003 by Robin Kolodny and David Dulio titled *"Political Party Adaptation in US Congressional Campaigns"* looks at how national political party committees spend the money for individual candidates in congressional races in 1998 and 2000 election cycles, and conclude that politicians spend their money mostly on hiring political consultants. Our goal is to compare the spending trends between the two political parties, not only on their overall distribution, but also on specific expenditure categories.

We focus on the 2022 election cycle and are interested in whether the Democrats and Republicans have similar spending behavior. We will adopt a permutation test and bootstrap resampling method to test our hypothesis. Our data comes from the Federal Election Commission's spending category, this includes all the types of disbursements candidates and committees make.

Our project can be broken into two parts. First, we test whether the contributions of Democrats and Republicans's spending come from the same distribution, and whether the two party's spending have a close average. In the second part, we take a closer look at specific types of expenditure category, especially TV advertising and canvassing, to test whether the Democrats and Republicans spend money on similar things.

2 Data Description

Our data set contains a total of 1334 rows and 23 columns. There are 1281 afflicted total expenditures, with 704 being Democratic and 577 being Republican. Among the total of 79 unique candidates, 36 of them identified as Democrats, and the remaining 43 as Republicans. Within this data set, we identified 3 important variables that are relevant to our analysis: candidate party affiliation, expenditure amount, and spending purpose including digital advertising, full page ads, staff salary and benefits, etc, which are described as `cand_pty_aff`, `exp_amo`, and `pur` in the data set, respectively. Candidate party affiliation is a categorical variable, which is either Republican or Democratic, corresponding to the part of the candidate associated with the spending.

Expenditure amount is a quantitative variable ranging from a minimum of 25 cents to a maximum of around $525,000, and describes how much of an individual expenditure was. Comparing the distributions for each party roughly in Figure 1, we see that both are skewed right, and share a similar shape, with most of the spending being concentrated around 0 - $20,000.

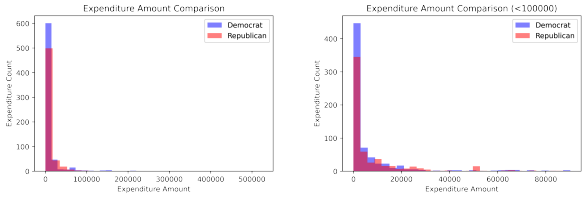


Figure 1: Histograms Comparing Expenditure By Party

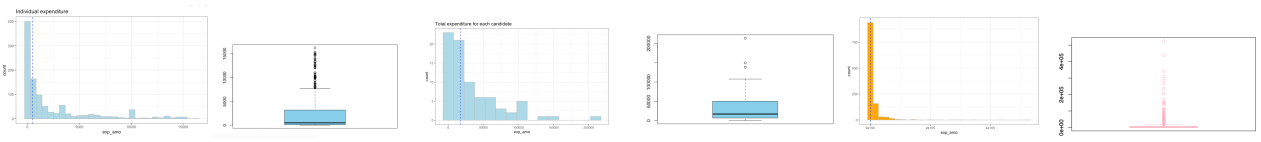
And finally, purpose is a categorical variable referring to what the spending went towards, such as *"Digital Advertising, Full Page Ad, Staff Salary and Benefits, etc..."* There are 289 unique purposes, but only 28 saw 10 or more occurrences. "Field Canvassing Expenses" was the most frequent purpose, with 150 expenditures, followed by "Door to Door Canvassing" and "Digital Advertising", with 66 and 37 each. "TV advertisements", which we also analyze in this paper, saw 21 total expenditures.

3 Analysis

3.1 Overall Spending

In the first part of analysis, there are two questions to be addressed statistically for exploring the expenditures in election cycle based on party affiliations. Do the contributions to Democrats and Republicans come from the same distribution? Do Democrats and Republicans spend money in a similar manner (have close average, median and IQR)?

To start with, we emphasize the issues related to expenditures through variables of expenditure amounts, names of candidates and the related party affiliations. In order to analyze the dataset effectively and explore the potential impacts that may be caused by the outliers, we will apply the same analytic procedures for three datasets with different data transformations based on the original dataset. The first dataset named **individual dataset** filters out outliers from the original dataset according to the quantile calculations. After calculating the IQR of the orginal dataset, any observation with expenditure amounts over 16441.68 will be filtered as they are below 0.25 **quantile** $- 1.5 \cdot$ **IQR** $= -9625.005$ and over 0.75 **quantile** $+ 1.5 \cdot$ **IQR** $= 16441.68$. The second dataset named **candidate dataset** uses a similar procedure except for an extra step to sum up expenditure amounts for observations with the same candidate id since there are multiple expenditures made by the same person. After calculating the ranges for the outliers, the observations with expenditure amounts above -127690.8 and below 238324.6 are left while others are filtered out. Then, the last dataset that will be used is the **original dataset**, which does not filter out any observation for comparing the effects of outliers on the analysis.



(a) individual (b) individual (c) candidate (d) candidate (e) original (f) original

Figure 2: Density histograms and Box plots for different datasets

Now we are ready to start analyzing the difference between distributions of expenditures amounts for Democrats and Republicans as the first question of interest described, and we will use permutation tests in this case since we do not have suitable assumptions for the distributions.

Let $X_1, \cdots, X_n$ be i.i.d. expenditures of cdf F from Republican party, and $Y_1, \cdots, Y_m$ be i.i.d. expen-

ditures of cdf G from Democratic party. The null hypothesis is

$$H_0 : F = G, H_1 : F \neq G$$

Permutation P-value

	median	mean	IQR
individual	0.8228177	0.01469853	0.3167683
candidate	0.3516648	0.829917	0.19988
with outliers	0.3024698	0.2932707	0.00539946

Figure 3: P-values

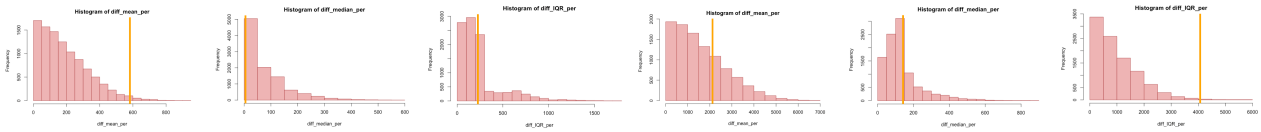
Applying the test statistics of differences in median, mean and IQR for expenditures made for different parties, we want to test the null hypothesis to see if the distributions for expenditures from Republican party and Democratic party are different. Given the preview of the sample test statistics and p-values from permutations, we can tell that most of the p-values are greater than 0.05 and the corresponding test statistics are the largest among the three test statistics used.

Therefore, based on the given previews, we cannot reject the null hypothesis in most cases at 5% level of significance except for two special cases with p-values of 0.0147 and 0.000540, which may indicate the possible issues from the individual dataset and the original dataset.

Permutation Sample Test Stats

	median	mean	IQR
individual	5	582.2118	226.455
candidate	6110.295	1431.206	7723.025
with outliers	141.75	2123.605	4069.968

Figure 4: Test Stats



(a) ind mean (b) ind median (c) ind IQR (d) origin mean (e) origin median (f) origin IQR

Figure 5: Density histograms for p-values using individual and original dataset

Firstly, for **individual dataset**, the p-values calculated using differences in median and IQR are 0.8228 and 0.3168 accordingly, which tend to be high in this case. And only the test statistics of difference in mean give a small p-value of 0.0147 which could reject the null hypothesis at 5% level of significance while all other p-values are greater than 0.05.

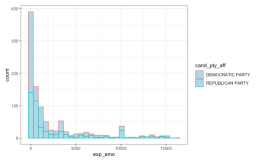


Figure 6: individual dataset possible issue

However, we can tell that the expenditures from Democratic Party tend to concentrate on the smaller scale, that is, observations with Republican party affiliation have larger expenditures amount in general than those with Democratic party affiliation. So when we filter out the outliers, we tend to filter out more observations with large amount of expenditures from Republican party which results in the large difference between means. This is a possible reason why there is a small p-value of

0.017 from the permutation using difference in mean values as test statistic.

Next, if we use the **original dataset**, we get the p-values of 0.3025 and 0.2933 calculated using differences in median and mean. And only the test statistics of difference in IQR give a very small p-value of 0.005340 which let us reject the null hypothesis at 5% level of significance.

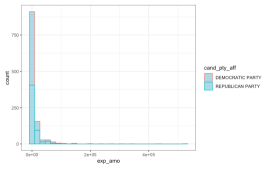


Figure 7: original dataset possible issue

However, we can tell that the possible issue with the original dataset could be caused by high variation of the expenditure amounts. Since the values of expenditures for observations vary between \$10 to \$400,000, which results in the large difference between IQRs. And the sample IQR test statistic will be larger accordingly, which leads to the small p-value.

Lastly, we use the **candidate dataset**. Based on the permutation test, we get a p-value of 0.3517, 0.8299 and 0.1999 from test statistic of median, mean and IQR accordingly. In this case, all the p-values calculated using all three test stats tend to be higher. And we could not reject the null hypothesis at 5% level of significance using either test statistic.

Also from the histogram for the distribution of observations with different colors indicating different parties, we can tell that the values of expenditure amounts from both parties are more similar to each other compared to the first dataset. So this dataset may be a better choice as there are more reasonable observations left after filtering and we get a more reasonable result as well.

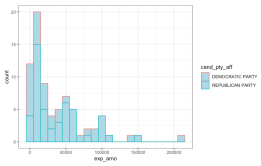


Figure 8: more reasonable candidate dataset

In order to analyze the second question which focus on the detailed differences of expenditures instead of differences in distributions for distinct party affiliation, we will use empirical bootstrap. Let the estimate be the ratio of test statistics θ. The null hypothesis is:

$$H_0 : \theta = 1, H_1 : \theta \neq 1$$

We will use the same three datasets as we analyze the first question. As we can see from the table, again we can reject the null hypothesis only when we use ratio between means using individual dataset and ratio between IQR using original dataset since the value under null hypothesis is not covered in 95%

Ind: Bootstrap Variance, MSE and CI

	median	mean	IQR
Variance	0.1542335	0.01609513	0.06137463
MSE	0.1944949	0.01614068	0.07516454
95% CI	[0.607, 2.135]	[1.056, 1.557]	[0.873, 1.755]

(a) individual

With outlier: Bootstrap Variance, MSE and CI

	median	mean	IQR
Variance	0.06760745	0.02360258	0.1342669
MSE	0.06793464	0.02380914	0.1392368
95% CI	[0.731, 1.743]	[0.568, 1.157]	[1.074, 2.522]

(b) original

Cand: Bootstrap Variance, MSE and CI

	median	mean	IQR
Variance	0.9809205	0.1233769	0.4709655
MSE	1.092718	0.1293955	0.4773927
95% CI	[0.537, 4.080]	[0.610, 1.981]	[0.517, 3.177]

(c) candidate

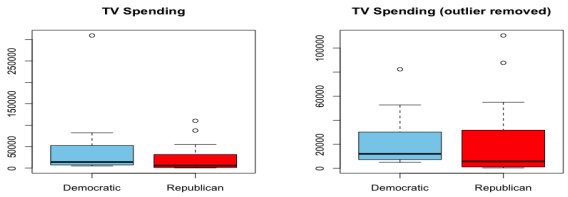
Figure 9: Tables for p-values using three dataset

confidence intervals. So there are similar issues with individual and original datasets, and the reasons are also similar to the first question.

In summary, we cannot tell there is a difference between either the distributions of expenditures from Democratic and Republican party, or the differences in median, mean and IQR of expenditures from these two parties. Also, we notice that the most reasonable dataset among individual, candidate and original datasets is the candidate dataset since there are less possible issues as mentioned.

3.2 TV Expenditures

There are many types of spending that campaigns engage in, however, none might be as far reaching as TV advertisements. In this section, we will analyze the TV spending of the Democrats and Republicans.



(a) (b)

Figure 10: Boxplots of TV spending with and without the outlier

Democrats have 10 expenditures, while Republicans have 11. We can see that without the Democrats' outlier of over $200,000, that the distribution appears to be quite similar. However, for this analysis, we will include the outlier.

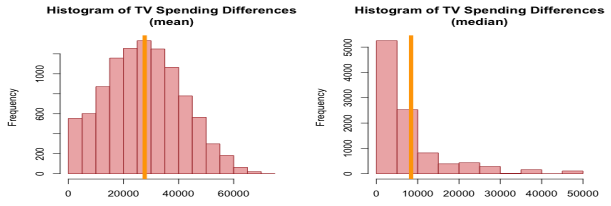
In order to challenge the null hypothesis that Democrats and Republicans spend their TV contributions similarly, we perform permutation tests on the absolute difference in the mean and median between the two samples. After performing the permutation tests (B=10000), we get the following histograms

TV Expenditures

	median	mean
Democrat	14120.92	53205.62
Republican	5727.64	25501.35
Abs. Diff.	8393.28	27704.27

Figure 11: Test Stats

of our test statistics, with the original statistics as the vertical lines.



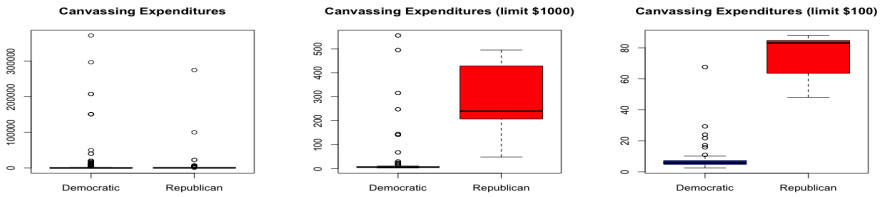
(a) (b)

Figure 12: Density histograms for permutation test statistics

Our p-value for the mean is 0.4823, and for the median it is 0.2670. We cannot reject the null hypothesis that the means and medians of TV spending for Democrats and Republicans come from the same distribution. TV advertisements are both necessary for campaigns and very costly - regardless of the candidate's party. With this in mind, our results make sense.

3.3 Canvassing Expenditures

This section will take a similar approach to the previous section on TV spending, however we will be looking at the expenditures for canvassing. As opposed to TV spending, which had a few number of large expenditures, canvas expenditures were far more numerous and tended to be less costly.



(a) (b) (c)

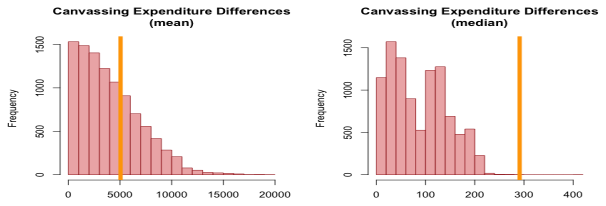
Figure 13: Density histograms for permutation test statistics

We can immediately see that the spending habits of Democrats and Republicans differ vastly, concerning canvassing, in the boxplots below. Though it may appear that Republicans spend a lot more on canvassing than Democrats, that is not the case. Democrats spent $2,188,502 while the Republicans only spent $513,585.40. The Democrats had more outliers than the Republicans, and had a much larger sample size with 207 expenditures, as opposed to the Republicans' 93.

We perform permutation tests on the absolute difference of mean and median with B=10000, and get the following histograms, with the original statistics as vertical lines.

Canvassing Expenditures		
	median	mean
Democrat	6.99	10572.48
Republican	298.2	5522.423
Abs. Diff.	291.21	5050.053

Figure 14: Test Stats



(a) (b)

Figure 15: Density histograms for permutation test statistics

The p-value for the difference in means is 0.3238; for the medians it is very small at 0.0001. We can clearly see that the difference in medians is significant from the boxplots above. However, when we look at the mean contributions, there does not appear to be a significant difference between the Republicans and Democrats. This shows a differing style of canvassing expenditures between the Democrats and Republicans, but we do not have enough evidence to believe that their mean canvassing expenditure comes from a different distribution.

4 Conclusion

We aimed to investigate the spending patterns between parties, since these behaviors can strengthen candidate results while weakening their opposition during voting. Using empirical bootstrap and permutation tests on several metrics, we found that there was no statistically significant difference between Democratic and Republican overall party candidate spending in general if we apply a more reasonable dataset instead of datasets with possible issues like impacts from outliers. Looking more closely at specific purposes, we see that there is no difference in spending for TV advertisements while there is for canvasing expenditures. A key observation was that overall, Republicans tend to spend higher amounts on single expenditures, with this happening at a significant rate for canvassing. For future analysis, we can consider looking at additional purposes such as Full Page Ads, or simply performing the same analysis on more data, as the election cycle will be continuing for several more months.

Bibliography

Breaux, David A., and Anthony Gierzynski. *"It's Money That Matters": Campaign Expenditures and State Legislative Primaries."* Legislative Studies Quarterly 16, no. 3 (1991): 429-43. Accessed May 13, 2021. http://offcampus.lib.washington.edu/login?url=https://www.jstor.org/stable/440106

Kolodny, Robin, and David A. Dulio. *"Political Party Adaptation in US Congressional Campaigns: Why Political Parties Use Coordinated Expenditures to Hire Political Consultants."* Party Politics 9, no. 6 (November 2003): 729–46. https://doi.org/10.1177/13540688030096004.

"Expenditures." OpenSecrets, www.opensecrets.org/campaign-expenditures.

"Spending: by the Numbers." FEC.gov, www.fec.gov/data/spending-bythenumbers/.