

Introduction

This analysis aims to make inferences about California local government spending on law enforcement through bootstrapping and a hypothesis test. The analysis focuses on how much their total budgets local governments (i.e., cities and counties) spend on their law enforcement agencies (i.e., city police and county sheriff departments). The analysis uses a bootstrap to estimate this percentage for all California cities and counties, then uses an A/B test to decide whether this percentage differs by jurisdictions' racial makeup.

The analysis uses a dataset on 157 California cities and counties for 2016 to 2018, which includes data on their law enforcement agency budgets and community demographics. The data were retrieved from the INFO 290 course website, and originally produced by Campaign Zero and their project 8 Can't Wait (Campaign Zero, n.d.).

Prior to beginning the analysis, the data were reviewed to determine whether any data cleaning was needed. In reviewing the data, it was determined that the police arrest csv file provided incorrect labels for three of the population by race columns, based on a comparison against U.S. Census Bureau data (U.S. Census Bureau, 2020). Specifically, the column containing the Hispanic population count was labeled as 'White Population', the column containing the white population count was labeled as 'Black Population', and the column containing the Black population count was labeled as 'Hispanic Population'. These labels were corrected within Jupyter Notebook. No other data cleaning was performed.

Part I: The Bootstrap

A bootstrap was applied to the data to estimate how much California local governments spend on their law enforcement agencies as a fraction of their total budget, on average. The histogram below depicts police¹ budgets as a percentage of the associated local government total budget for the 157 sample jurisdictions within the dataset. As shown in Figure 1, in most local governments in the dataset, police budgets make up approximately 10% to 30% of total budgets, with a median of 16%. The data are right-tailed, with police budgets composing nearly 40% of total budgets in some jurisdictions.

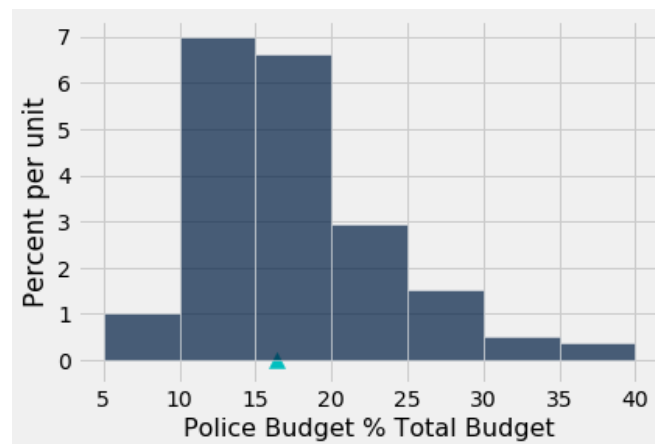


Figure 1. Histogram of police budget as a percentage of total local government budget for 157 California jurisdictions. The median is depicted by the cyan triangle.

¹ From here on, unless otherwise specified, the term 'police' encompasses both city police departments and county sheriff departments

A bootstrap was undertaken by generating 5000 157-sized resamples with replacement from the 157 samples in the data, and for each of these resamples, calculating the median police budget as a percentage of total budget. The 2.5th and 97.5th percentiles of these resampled medians were then calculated to generate a 95% confidence interval. Figure 2 depicts the distribution of the resampled medians and the confidence interval. The findings indicate that we can say with 95% confidence that the true population parameter—the police budget as a percentage of total budget median for California local governments—is between 15.0% and 17.4%. Looking at Figure 1, we can see that the median of the distribution of sample values also lies toward the center of this range.

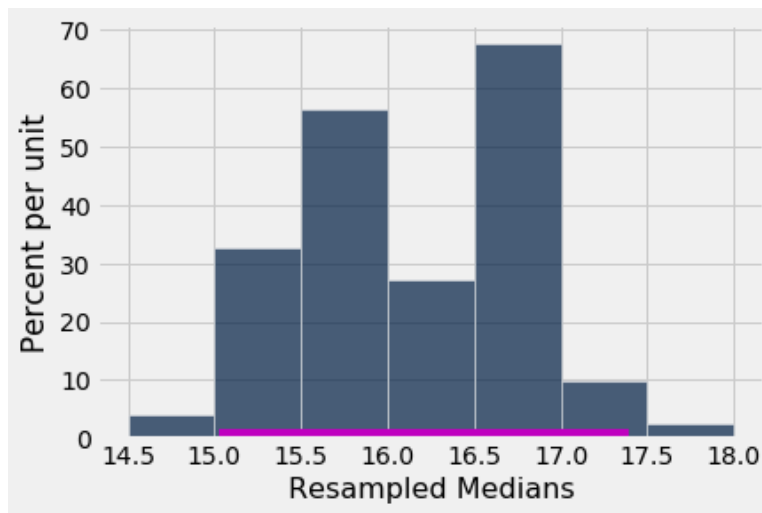


Figure 2. Histogram of resampled median police budget as a percentage of total local government budget from bootstrap process. Histogram includes medians from 5000 resamples of size 157. The 95% confidence interval is depicted in magenta.

Part II: The Hypothesis Test

A hypothesis test was then run to explore the relationship between jurisdiction demographics and the portion of local government budgets allocated to police. To explore this relationship and determine the appropriate hypothesis test type, scatterplots were generated comparing police budget as a percentage of total budget against the percentage of the population that was white, Black, Hispanic, and Asian, with point size varying by the jurisdiction median income (Figure 3). The scatterplots roughly depicted a positive relationship between police budget size and the percentage of population that was Black or Hispanic, a negative relationship between police budget size and the percentage of the population that was white, and no clear relationship between police budget size and the percentage of the population that was Asian.

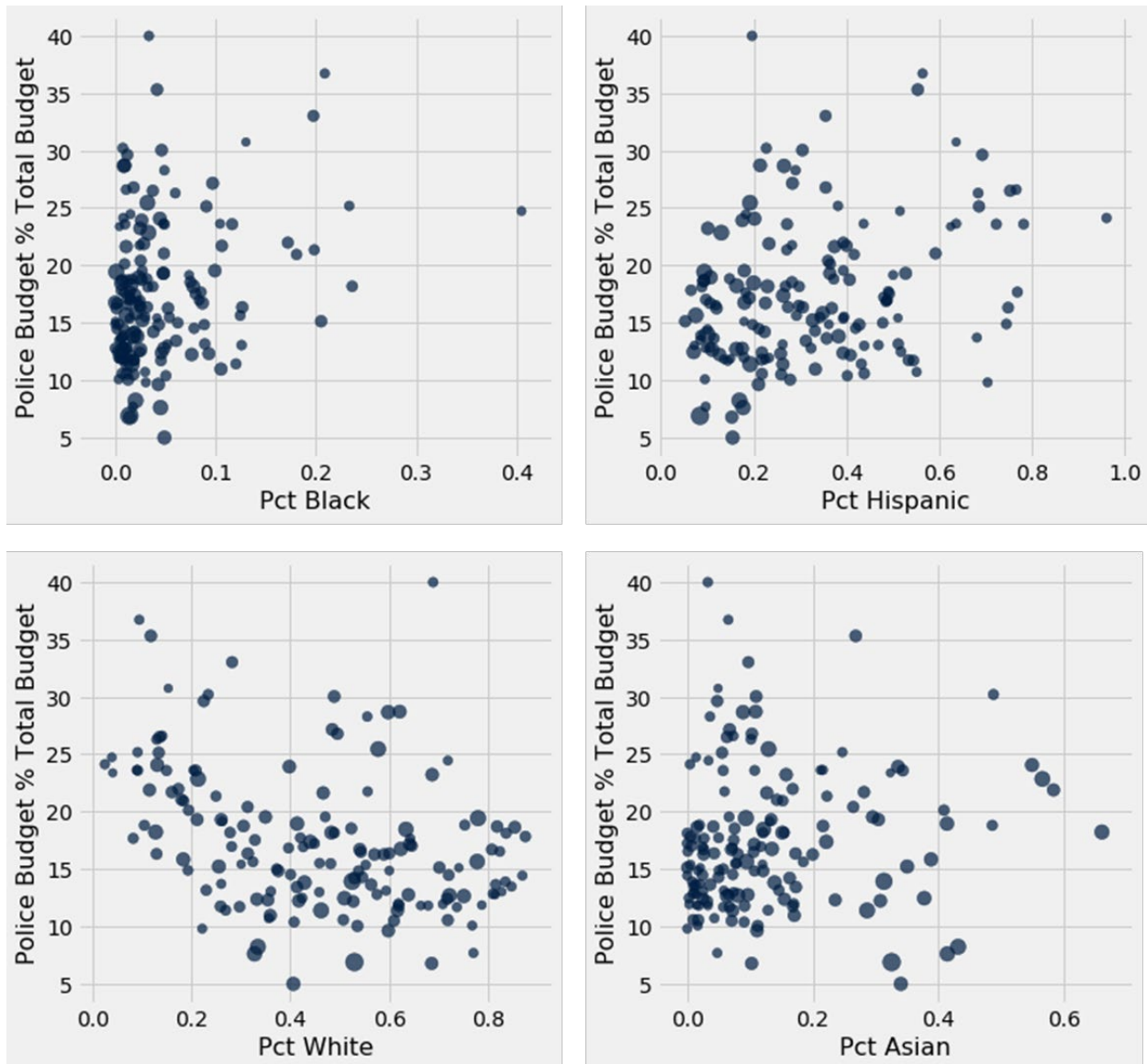


Figure 3. Scatterplots of police budget as a percentage of total local government budget vs. percentage of population that is Black, Hispanic, white, or Asian. Point sizes are proportional to jurisdiction median income.

Based on this, groups were created in the dataset for use in an A/B hypothesis test. A new column within the dataset was created listing each jurisdiction as 'Majority Black & Hispanic' if the combined Black and Hispanic population exceeded 50% of the jurisdiction's total population ($n = 43$), 'Majority White' if the White population exceeded 50% of the jurisdiction's total population ($n = 70$), or 'Neither' if neither of these conditions were satisfied ($n = 44$). To visualize variation in police budget across these three groups and confirm that the groups seemed appropriate for use in a hypothesis test, the histograms in Figure 4 were generated. The left-hand graph provides a histogram for all three groups, while the right-hand graph provides a histogram exclusively for the groups used in the A/B test, to more clearly visualize the differences between these two groups. As seen in the figure, on average, police budget as a percentage of total budget appears to be larger in majority Black and Hispanic communities relative to majority white communities, while the average percentage for communities that are neither majority Black and Hispanic nor majority white appears to lie somewhere between the other two groups' averages.

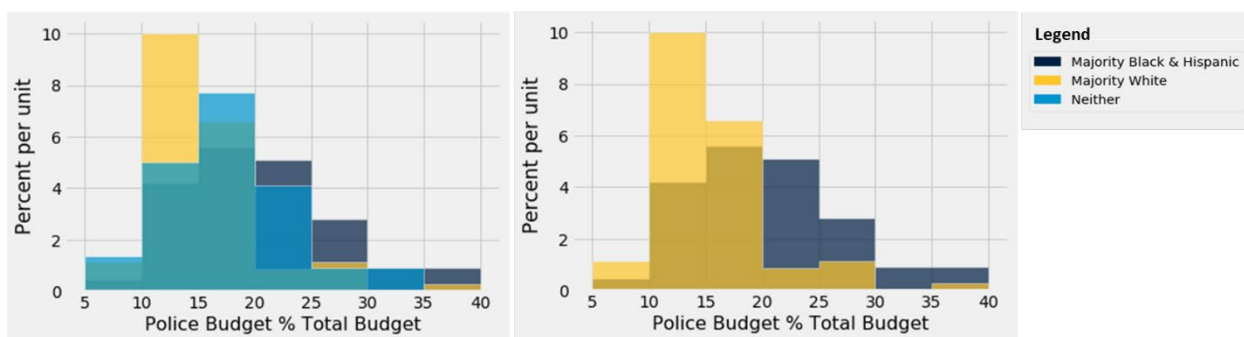


Figure 4. Histograms of police budget as a percentage of total local government budget for jurisdictions that are majority Black and Hispanic (dark blue), majority white (yellow), or neither (cyan). The figure on the left includes histograms for all three groups, while the histogram on the right includes histograms only for the two groups included in the hypothesis test.

Based on this initial exploration, the following null and alternative hypotheses were developed:

Null hypothesis: The distribution of the percentage of jurisdiction budgets allocated to law enforcement agencies is the same for jurisdictions that are majority Black and Hispanic as for jurisdictions that are majority white.

Alternative hypothesis: In jurisdictions that are majority Black and Hispanic, law enforcement budgets make up a larger portion of total jurisdiction budgets, on average, than in jurisdictions that are majority white.

The test statistic was chosen to be the difference between the mean percentage of total jurisdiction budget allocated to law enforcement in majority Black and Hispanic jurisdictions, and the mean percentage allocated to law enforcement in majority white communities.

The A/B test was run by paring down the full dataset to include just the columns listing the A/B test categories (i.e., Majority Black & Hispanic and Majority White) and police budget as a percentage of total jurisdiction budget. The test statistic was then simulated under the null hypothesis 5000 times by shuffling the list of police budget values, adding it to the table with unshuffled categories and values, then calculating the mean of the shuffled values for the two groups, and calculating the difference between the two. These simulated differences were then plotted in a histogram (Figure 5). The observed difference (i.e., the difference between the mean for Majority Black & Hispanic jurisdictions and Majority White jurisdictions using the original unshuffled values) was calculated and added to the histogram. The p-value was calculated by counting the number of simulated

differences that exceeded the observed difference; the p-value was calculated to be 0.00. Using a significance level of 0.01, we reject the null hypothesis that the distributions for the two groups are the same.

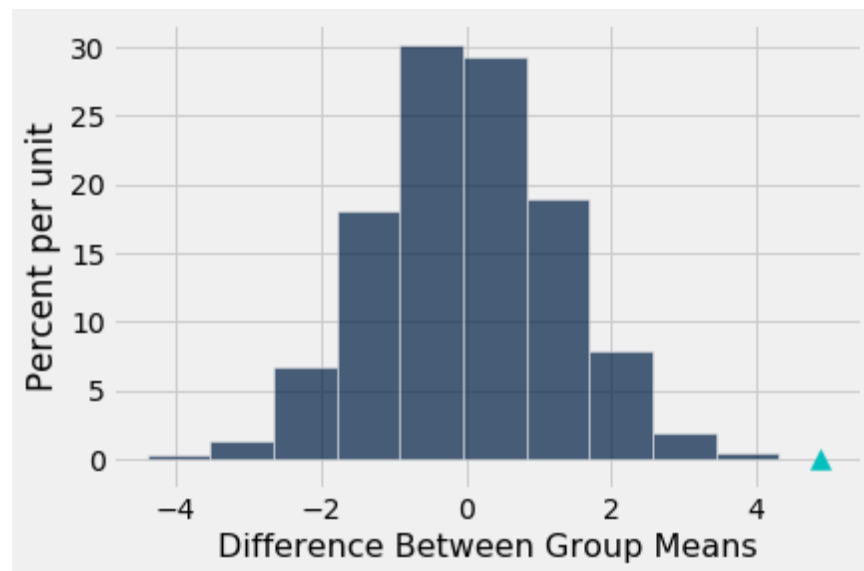


Figure 5. Histogram of differences between group means simulated 5000 times under the null (dark blue) and observed difference between the two group means (cyan triangle).

The test indicates that within the sample of 157 California cities and counties, in jurisdictions where more than 50% of the population is either Black or Hispanic, police budgets make up a larger portion of the total jurisdiction budget than in jurisdictions where more than 50% of the population is white. Given the size of the dataset and that the data come from all 58 California counties, it seems reasonable to generalize these findings across California.

It is important to note that while the data indicate that police budgets differ by community demographics, they do not explain the reason(s) driving this difference. Further research is needed to better understand causation, though this would likely be difficult, as it does not seem feasible to run a randomized controlled experiment to examine causality in this case. However, future research could investigate factors that might influence police budgets, the relationship between police budget sizes and the type of budget allocation processes used, the demographics of the politicians and others who determine police budgets, the strength of police unions, whether the community is engaged in police oversight, race-related tensions and power dynamics within the community, among other factors.

References

- U.S. Census Bureau. (2020). QuickFacts. <https://www.census.gov/quickfacts/>
- Campaign Zero. n.d. Join Campaign Zero. <https://www.joincampaignzero.org/#vision>