

Analysis of the relationship between city size and municipal spending efficiency in California cities*

Subtitle

Terrell Enoru

September 22, 2025

California's rising government debt over the past decade necessitates careful examination of municipal spending efficiency. To analyze spending efficiency I use data on expenditure per capita for all California cities over the past 20 years. I use regression analysis to model the relationship between population size(the predictor) and per capita spending(the response variable). Then hypothesis testing is used on the slope to determine whether large cities benefit from economies of scale. The results were FILL THIS IN which is significant because FILL THIS IN.

Introduction

Data

Each observation in this dataset represents the expenditure per capita of a particular city in California at the end of a fiscal year. The key variables are entity name which is the name of the city,fiscal year(July 1 of the current year to June 30 of the next year) which is the year for which each expenditure per capita is calculated ,total expenditure which is the gross expenditure of a particular city over the fiscal year, estimated population which is the population of the city based on the Housing Unit Method which extrapolates population by taking housing units adjusting for occupation and then using that to estimate household population then combining that with estimates for "group quarters population"(essentially everyone in nontraditional households like dorms and prisons) to get the overall population estimate(more details in reference),and expenditures per capita which is the total expenditures

*Project repository available at: <https://github.com/tycebot/MATH261A-project>.

divided by estimated population. The data is grouped by city, year, as well as binned into three groups for large, medium and small sized cities. The limitations for the dataset are as follows: the estimated population does a bad job of measuring homelessness because of its reliance on housing units, the years are limited to a twenty year period from 2003 to 2023, the estimated population can double count those that move until they register their new address, and the expenditures measures are both skewed over the covid period (2020-2023). None of the records are missing. By analyzing how much money is spent by a city over an extended period of time for cities with a variety of different populations we can reasonably ascertain the relationship between expenditure and the population. However the key assumptions are that the population measurement is accurate relative to other cities in the dataset (meaning if it underestimates or overestimates the population then it does it consistently) and that the cities are correctly reporting their expenditures to the state.

per capita expenditure visualizations

total expenditure visualizations

estimated population visualizations

#Methods We adopt a simple linear regression model with expenditure per capita as the response and population as the predictor. Let Y_i denote expenditure per capita and X_i denote population for the Y_i year in our dataset. The model can be written as $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ where β_0 represents the expected population when the expenditure is 0, β_1 represents the average increase in expenditure per capita for every individual added to the population, and ϵ_i represents the change in expenditure per capita not captured by the population. This model is implemented using the `lm()` function from the R programming language. The `lm` function is the a function used to create linear models. **ADD SECTION ON HYPOTHESIS TESTING**

Results

References