Proposal: Expenditures in NY Municipalities

Richard A. Levine*
San Diego State University
STAT 794: Statistical Communication in Data Science

September 9, 2020

Background

Two New York municipalities (Warwick and Monroe) are interested in predicting changes in expenditures resulting from proposed construction of new housing projects. The issue is that if the expenditures increase, the town may have to generate more funds through say property tax.

Specific Aims

- Study the relationship between a set of NY municipality demographics and expenditure per person.
- Predict expenditure per person for the years 2005 and 2025 in each of the two municipalities, Warwick and Monroe.

Data

The data set was obtained from the New York municipality governments. The data set contains three identifiers for the municipality (ID, State, County codes) and six predictors, as shown in the following table. The variable "percent intergovernmental" represents the percentage of revenue coming from state and federal grants and subsidies. All other variables are self-explanatory.

Variable	Defintion
ID	Identity number
ST	State code
CO	County code
EXPEN	Expenditure per person
WEALTH	Wealth per person
POP	Population
PINT	Percent intergovernmental
DENS	Density
INCOME	Mean income per person
GROWR	Growth rate

*Email: rlevine@sdsu.edu

The data is from 1992 and includes 916 municipalities. The municipalities wish to build a model that can predict expenditures in the years 2005 and 2025 for which the six demographic and income-related predictors are provided by expenditure per person is not available. We note that though this is a future prediction from 1992, we are viewing the years 2005 and 2025 as new observation labels. More importantly is predicting expenditures over expected changes in the predictors resulting from the new housing projects. Since these changes fall within the range of the data set, we are not extrapolating per se.

Methods

The specific aims will be addressed through a regression model of expenditures on the six potential predictors. The analysis plan will include an exploratory data analysis to assess each variable individually and the relationship between expenditures and the six predictors. The exploratory data analysis will inform the model building: potential transformations required, non-linear relationships, and challenges to regression modeling assumptions including influential points and missing data. The regression models considered will be assessed through a regression diagnostic work-up and an analysis of predictive performance via ROC curves and corresponding evaluation statistics.

Broader Impacts

The regression model will allow us to draw inferences on the relationship between each demographic variable in the model and municipality expenditures. We will also present inferences on municipality expenditures in the desired years 2005 and 2025. These results will inform the municipalities on the impact of the new housing projects on future expenditures. We will also make recommendations on further data collection and modeling as the municipalities assess the impact of and required resources for new planned projects. We will share our findings in a data analysis report to the municipalities, including documented code for carrying out the regression analysis.