Food Insecurity in Alemeda County: Modeling Food Need and Access

Biyonka Liang [1] **Sidney Le** [* 1] **Hubert Luo** [* 1] **Ash Mohan** [* 1] **Suhas Rao** [* 1] **Reese Williams** [* 1] **Rick Yang** [* 1] **Katherine Yen** [* 1] **Alex Yi** [* 1]

Abstract

Analysis and predictive modelling of pounds of food unmet by local supply in Alameda County.

1. Introduction

In partnership with the Alameda County Community Food Bank (ACCFB), we worked to address issues of food insecurity and food pantry access in Alameda County. Defined by the United States Department of Agriculture Economic Research Service, food insecurity is a state of hunger and "reduced quality, variety, or desirability of diet" resulting from poor access to food, either from structural issues (low availability) or from economic issues (low income).

In 2010, it is estimated that some 49 million people in the US live in food insecure households, which face serious health and developmental risks as a result of their food insecurity. In Alameda County alone, Feeding America estimated that 14.3% of people, roughly 1 in 7, were food insecure in 2015.

A key part of the solution is creating a cheap, or even free, and accessible supply of healthy food, which food bank networks such as the ACCFB attempt to do. However, no matter how well their network creates access, their supply of food and funding is limited.

The core of this issue becomes misallocation; besides vague ideas about which regions might have more people experiencing food insecurity, there has been no systematic way to estimate which regions need what quantity of food. It is generally clear that food insecurity is spatially defined, but how local factors come into play in estimating food need thus became our goal.

2. Background Research & Literature Review

Luckily, the issue of food insecurity is not one which researchers are unfamiliar with. Feeding America is the largest hunger-relief organization in the United States, and

---

*Equal contribution [1]Department of Statistics, University of California, Berkeley.

they produced an invaluable tool for mapping hunger across the country based on research done by Craig Gundersen & Ziliak (Gundersen & Ziliak, 2018) The paper lays out a fixed effect model, which can be thought of as a linear regression model that takes in data that is split based on two grouping variables, usually location and year, to estimate food insecurity rates on the state level. Using the coefficients for the state-level fixed effect model, the researchers then go down to the county level and use country data to estimate food insecurity rates. This fixed effect model developed here becomes especially important for us in the modelling part of the project.

In the course of our research, although this paper (Gundersen & Ziliak, 2018) became a core part of our work, a large number of other papers were found that were also useful in developing our intuitions for working with food insecurity. One that was particularly helpful was the Yen et al. paper, published in 2008. (Yen et al., 2008) They performed econometric analyses of the Food Stamp Program in relation to food insecurity, and actually found that a dynamic, two-way causal relationship between the two actually led to lower food insecurity being correlated to food stamp usage. While this makes sense intuitively, it reveals that the data we are examining cannot be analyzed without context.

3. Exploratory Data Analysis

With a strong footing in existing research, we set off to better understand the Alameda County from the perspective of the data. As food insecurity is determined by an innumerable number of environmental and economic variables, the hope was that by examining the geographical distribution of some specific variables we identified as important, we could get a good grasp of the whole landscape of need and hunger across the county.

The first variable we looked at to understand need was the percentage of burdened households by census tract, which can be more specifically defined as households which spend more than 30% of their income on rent or mortgage. It appears as though household burden, in this narrow definition, is distributed somewhat randomly across the county, although there are some small census tracts which contain a disproportionately high percentage of burdened households.
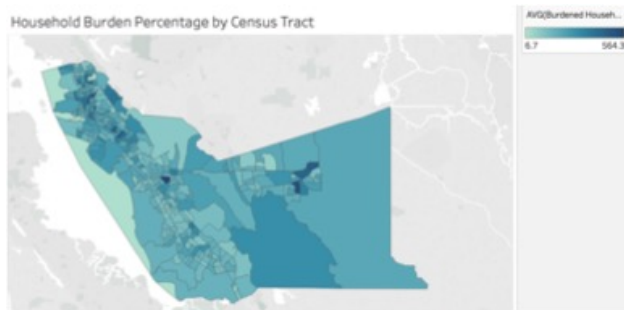
Figure 1. Household Burden Percentage by Census Tract in Alameda County.
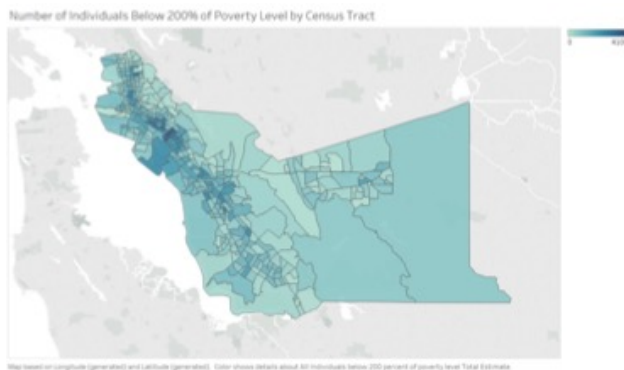

Figure 2. Number of Individuals Below 200% of Poverty Level by Census Tract in Alameda County.

Approaching this issue from the opposite direction, we sought to look at the distribution of low-income people. Here, we found that most census tracts have low counts of very low-income people, and a small number of census tracts clustered around Oakland have high numbers of low-income people.

We also examined some other similar variables, such as unemployment rate and enrollment in social services and welfare programs, but they largely mirrored the data already discussed. To more directly approach the issue of food access, we looked at an index called the food affordability ratio.

Essentially, it maps the relative affordability of food on a 0 to 1 scale, where 0 is affordability parity (all people being able to afford food) and 1 is lack of affordability. The most interesting result of this exploration was the seasonality—food appears to get marginally more expensive as the year goes on. On the whole, affordability has stayed within a small range over the time period analyzed.
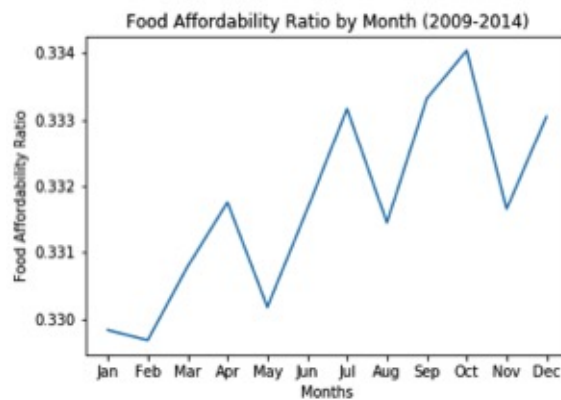

Figure 3. Food Affordability Ratio by Month in Alameda County from 2009 to 2014.

4. Predictive Modelling

While we were very lucky that the Gundersen & Ziliak paper specified a well-researched model for estimating food insecurity that we could use, (Gundersen & Ziliak, 2018) what we actually want to estimate is pounds of food need that is unmet by a local food bank (which will be referred to simply as "pounds unmet").

The data that was provided by ACCFB and Feeding America for the actual observed pounds unmet for the past year would be used to train and validate our models, so it was important that we could properly specify what we were trying to estimate in the first place. Thus, our model would be attempting to unearth an empirical relationship between food insecurity/local variables and pounds unmet, one that had not yet been researched.

What we ended up taking from the Gundersen & Ziliak model was their selected variables and their approach for taking rates for a large geographic unit and projecting downwards to a more granular geographic unit. In our case, it made the most sense to work with census tracts, since that was the most granular, local level we could work on, and we used county-level data to project downwards onto census tracts. Both the identified variables and this projection approach were vetted by the Gundersen & Ziliak research, so we could feel confident in using both. (Gundersen & Ziliak, 2018)

The variables identified by the paper were the following: 1) unemployment, 2) poverty rates, 3) median income, 4) percentage of population that is Hispanic, and 5) percentage of population that is Black. In addition to these five variables, we also included census tract population to be able to factor in food insecurity rates in the model.
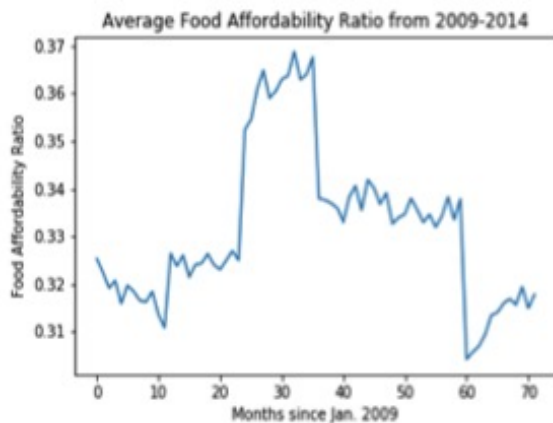
Figure 4. Average Food Affordability Ratio in Alameda County from 2009 to 2014.

The multiple regression model did relatively well, but the residuals were clearly patterned—telltale signs of poor model performance. We theorized that this was the result of the difference between the fixed effects model and the linear regression model. We attempted to transform variables in a number of ways to account for the residual patterns, and eventually we found a strong fit. It appeared that our predictors were strong predictors for the cube root of pounds unmet, with r-squared of 0.8969. The model held up remarkably well; the residuals were well-distributed and the model made it through 10-fold cross-validation with little problem.

5. Future Work

Although we were able to successfully develop and test a predictive model, the knowledge we gained along the way in both exploratory data analysis and discussions with the ACCFB left us with many tasks left to do.

The data and the experience of the experts both support the idea of seasonality—food need changes in cyclical ways as time progresses. Seasonality of food need makes sense intuitively, as well. One might expect that, as the month progresses and paychecks are spent, more and more people become food insecure. The approach to seasonal data requires time series models, which we attempted to implement but lacked the data to do so.

Time series models require data across a long period of time with data for the intervals. In the case of food need rising at the end of the month, one might need data for every day or every week of the month, and that level of granularity is almost impossible to find. Almost all existing data is produced on the order of years, and sometimes on the order of months.

Over the course of the project, we were in contact with the ACCFB to try to produce some of this data through customer surveys, but the funds for such a large scale survey project were not available.

References

Gundersen, C. and Ziliak, J. P. Food Insecurity Research in the United States: Where We Have Been and Where We Need to Go. *Applied Economic Perspectives and Policy*, 40(1):119–135, 02 2018. ISSN 2040-5790. doi: 10.1093/aepp/ppx058. URL https://doi.org/10.1093/aepp/ppx058.

Yen, S. T., Andrews, M., Chen, Z., and Eastwood, D. B. Food stamp program participation and food insecurity: An instrumental variables approach. *American Journal of Agricultural Economics*, 90(1):117–132, 2008. ISSN 00029092, 14678276. URL http://www.jstor.org/stable/30139495.