Food Deserts and Socioeconomic Factors

Michael Labadie & Samantha Roberts

DATS 6450 - Bayesian Methods for Data Science

**INTRODUCTION**

Ease of access and healthy quality of food sources among various socioeconomic statuses have been noted as significant indicators of disparity. A neighborhood of residence and its geographic distance to healthy food options, such as those provided by supermarkets, has shown significant relationships to a variety of wellbeing factors like income level or quality of educational institutions (Luan, Minaker, & Law, 2016). The Population of Poverty USA (n.d.) website cited USDA survey findings that estimated 12.3% of US households are food insecure. Over half of those reporting in this category receive assistance from one of three Federal food and nutrition assistance programs (The Population of Poverty USA, n.d.). Additionally, poverty-related statistics are available for kids, seniors, and members of racial categories including the following: Native American, Black, Hispanic, White, and Asian.

This project will attempt to find probabilistic distributions between factors including urban/suburban location, poverty rate, age count by category (kids versus seniors), and race by category and accessibility to food options. Discovering the most significant factors could assist in identifying neighborhoods and populations at risk. Additionally, the resulting analytics could provide valuable insight for city officials for future gentrification or new development projects.

**LITERATURE REVIEW**

The importance of a healthy diet has long been noted by researchers and medical professionals. It has become common knowledge that quality foods, like fruits and vegetables, are crucial to wellbeing. A deteriorated diet often consists of high sources of sugar and fats, such as can be found in most fast food options. These sources can lead to serious, chronic medical diseases and conditions that can have a wide effect on the quality of life of these populations (Lamichhane, et al., 2013). At a higher societal level, other living conditions have been found to correlate with food resources that can also have an effect on the quality of life a person experiences. Socioeconomic factors such as race, poverty level, and population density in a designated geographic area have all been found as significantly related to proximity of food sources such as supermarkets and fast food stores (Lamichhane, et al., 2013). Analysis of these complex relationships can be accomplished across multiple geographic locations and various time periods when data is available. Luan, Law, and Quick exemplified the use of a multi-model approach with hyperprior tuning for three different Canadian locations and spanning years 2011 through 2014 (2015). Significant results from this approach could be useful in identifying currently at-risk populations and those who may become at-risk in the future.

These discoveries encourage policymakers to consider a wider perspective for informed economic planning and development of neighborhood areas. Luan, Minaker and Law suggested that there are several interventions to address highly-populated and diverse areas that suffer from lack of quality food resources (2016). The traveling distance was noted as a serious obstacle for many citizens to reach a supermarket when dependent on public transportation, bicycles or walking so they suggested promoting corner store businesses to carry healthy food options. Additionally, encouraging existing businesses to increase fresh produce stocked would be a more cost-effective, less intrusive manner of localized intervention (Luan, Minaker & Law, 2016).

**RESEARCH METHODOLOGY**

Several models with varied feature combinations were evaluated to determine significant factors associated with geographical area and food desert categorization. The core model structure selected to compare the socioeconomic features included bernoulli and beta distributions with Markov Chain Monte Carlo (MCMC) simulation. This resulted in estimating the posterior distributions as affected by the unique data elements for each round. The main parameter assessed was theta as determined by the data being approximated with beta distributions. Just Another Gibbs Sampler (JAGS) was utilized to improve run-time efficiency with a burn-in rate of 500 and 4 chains to ensure optimal steady state evaluation.

**DATA**

The data for this project was accessed from the USDA’s ERS – Food Access Research Atlas (“Food Access Research Atlas.” , n.d.). The original dataset contained 72,864 samples and included 147 columns. The comprehensive set reported on all 50 states plus Washington, D.C. on factors such as population density (e.g. urban), poverty rate, median income, age categories such as kids and seniors, and access to a vehicle. More granular data included population counts and percentages for seven racial categories at various distances from half a mile up to 20 miles. No data samples were removed; however, there were several columns that were excluded during the analysis phase.

**KEY FINDINGS**

There were 33 feature combinations evaluated utilizing the data and model structure delineated in previous sections. Models were separated based on location, whether all locations were included or restricted to Washington, D.C. Consistent features reviewed across both location groups included group quarters, vehicle flag, low income, poverty rate, SNAP participation, and prominent proportion of African Americans. These factors were tested against knowledge of existing low-access food areas at one mile (urban) and ten miles (rural) and half mile, indicating neighborhood proximity to a supermarket.

The multi-feature approach unveiled the rate of vehicle access as one of the most significant factors in low access areas. The flag within the data set designated whether there were greater than or equal to 100 households without vehicle access that were at least a half mile from the nearest supermarket. The ability to reach a supermarket without a vehicle requires alternative means such as public transportation, cycling or walking. This may impede the ability to reach the market as well as limit the amount of food a person could reasonably transport on return. Significant findings were reflected across all locations and within Washington, D.C. The vehicle feature was tested across each of the mileage distances as well. The only non-significant result was low access at one mile (urban) and ten miles (rural) within Washington, D.C. The strongest significant result was in D.C. at a half mile distance, with a mode of -.747 and all data contained well below zero. Both the distance and vehicle features had roughly 50,000 estimated significant samples (ESS) out of the 72,000 original samples with overlapping parameter values and a shrink factor reaching near one within the first 2,000 iterations.

The second significant factor to interact with low access areas at various distances and locations was the low income tract flag. These findings support an interpretation that low income areas are more likely to also suffer from low access to adequate food sources as well. This also supports reviewed literature indicating that socioeconomic factors can be used to delineate current and predict future food deserts. These significant results were found across all locations and within Washington, D.C. The D.C. results at a half mile were slightly stronger with a mode of -.201 than that of nationwide at low access one mile (urban) and 10 miles (rural). The HDI of the difference was entirely below zero, with 0.2% of data possible including or above zero. Each of the individual feature reviews were similar to the vehicle-based trials. Each accomplished roughly 50,000 ESS, with significant overlapping parameter values, and a low shrink factor around one after 2,000 iterations.

The last consistently significant factor across location groups was the poverty rate. The poverty rate was a continuous variable in the data set between 0 and 100 and indicated the share of the population living with income at or below the Federal poverty thresholds for family size. In order to properly assess this feature, it was discretized into four equally divided bins and assessed based on samples falling into each range. Due to the arrangement of this feature and practical interpretations, it was unsurprising that not all bins had probabilistic relationships with low access areas and intra-category interactions. At the national level, all bins were significant at the one mile/10 mile assessment, although there were no significant differences between categories two and four. This suggests there may be no difference between relatively low and high poverty areas across the country. Poverty rate categories in D.C. were similar as each bucket was significantly related at distances one mile/10 miles with no difference between any of the categories themselves. At a half mile distance in D.C., there were similar results as well. The one difference was a significant relationship between category one, indicating zero to very low poverty rate, and category two, the next highest bin. It’s possible this could be due to a severe income difference between the wealthiest neighborhoods and all others within the district limits. Individual theta analysis was very comparable to the previous factor results, with similar ESS, parameter value comparison and shrink factor trends.

**CONCLUSION**

This study assessed the probabilistic relationship using Bayesian Inference between low access food areas with a number of socioeconomic factors through the use of the USDA’s ERS survey results. The results supported the existing literature suggesting a significant relationship between a variety of socioeconomic factors and food deserts. This study’s findings determined vehicle access, low income areas and poverty rates are strong indicators of a food desert. Two of the three features evaluated were related to income levels in neighborhoods. One observation made during the trials was the differing results based on locations. D.C. provided exacerbated results due to the wide range of income levels present across the wards within the district. These delineations were not as clear at the national level and therefore, recommendations for future research would emphasize the importance of proper scaling to identify optimal significant relationships.

Overall, the concern of food deserts and low access to healthy options can have an enduring impact on the greater society. This information could be used by researchers and city officials to maximize policy and neighborhood marketing strategies. Properly informed officials could identify current and potential future food deserts and create change that supports a healthy balance of resources and public services.

**REFERENCES**

“Food Access Research Atlas.” USDA ERS - Food Access Research Atlas, www.ers.usda.gov/data-products/food-access-research-atlas/.

Lamichhane, A. P., Warren, J., Puett, R., Porter, D. E., Bottai, M., Mayer-Davis, E. J., & Liese, A. D. (2013). Spatial patterning of supermarkets and fast food outlets with respect to neighborhood characteristics. Health & Place, 23, 157-164. doi:10.1016/j.healthplace.2013.07.002

Luan, H., Law, J., & Quick, M. (2015). Identifying food deserts and swamps based on relative healthy food access: A spatio-temporal Bayesian approach. International Journal of Health Geographics,14(1). doi:10.1186/s12942-015-0030-8

Luan, H., Minaker, L. M., & Law, J. (2016). Do marginalized neighbourhoods have less healthy retail food environments? An analysis using Bayesian spatial latent factor and hurdle models. International Journal of Health Geographics, 15(1). doi:10.1186/s12942-016-0060-x

“The Population of Poverty USA.” Poverty USA, www.povertyusa.org/facts.