Predicting Food Insecurity in the United States by FIPS Code County

Daniel Kim, Advised by Professor Jonathan David Reuning-Scherer

Yale University, Department of Statistics & Data Science

Introduction

- Food Insecurity is a prevalent issue in the United States; over 13 million households experienced food insecurity even before COVID-19
- Understanding the landscape of food insecurity can have benefits in creating preemptive programs to target areas of high need
- Can we predict food insecurity by FIPS Code County? (Federal Information Processing Standards code)

PCA

- To explain 65.3% of the variance in the data, the first two principal components would be needed.
  - 1st principal component is more important than the second; makes up 47.8% of the variance to the second component's 17.5%.
- 1st principal component seems indicative of characteristics of poverty
- 2nd principal component seems to be indicative of joblessness

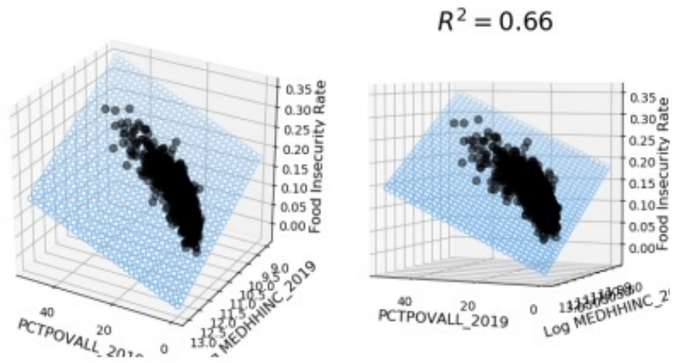
Loading Coefficients for PCA

	Comp 1	Comp 2	Comp 3	Comp 4	Comp 5	Comp 6
Unemployment Rate	0.37	0.40	0.43	0.68	0.23	0.08
Proportion of Disabled Workers	-0.01	-0.82	0.56	0.12	0.05	-0.02
Cost Per Meal	-0.30	0.40	0.69	-0.50	-0.08	0.15
% of All in Poverty	0.54	0.06	0.15	-0.24	-0.30	-0.73
Log Median Household Income	-0.52	0.08	0.01	0.12	0.55	-0.64
% of Adults with less than High School Diploma	0.46	-0.06	-0.04	-0.45	0.74	0.17

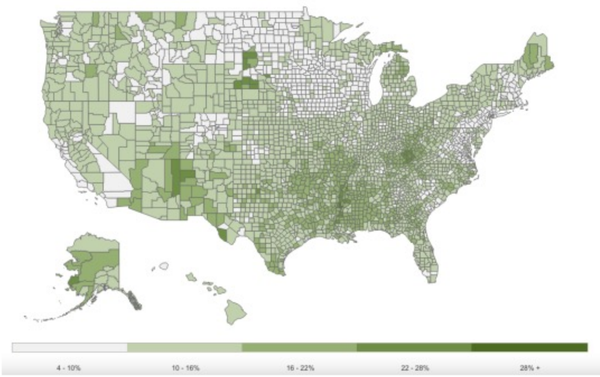
Cluster Analysis

- Are geographically close counties similar to one another in terms of food insecurity predictors?
- Euclidean Distance and Complete Agglomeration of 10 randomly selected FIPS Counties and their 9 respective counties pointed to yes; closer counties tend to be clustered together.
- This is confirmed by the literature stating that "low–food insecurity peer groups are often right beside each other" (Lou).

Linear Regression Visualization based on % of All in Poverty and Log Median Household Income

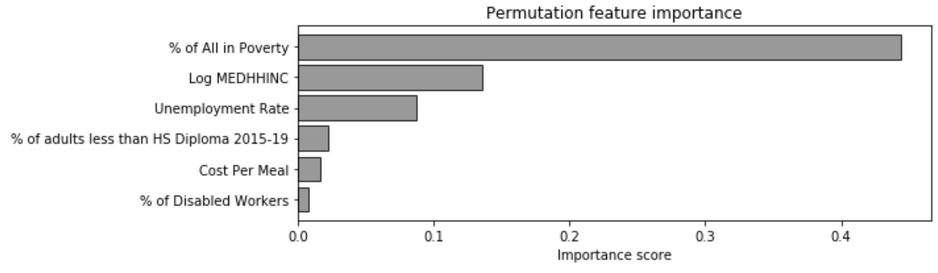


U.S. Map of Food Insecurity Rates



Predictive Modeling

- Linear Regression had lowest RMSE (0.020072183893) and highest R^2 (0.7361560701424712) values out of all the models.
- Backwards Stepwise Regression and Significance of Predictors showed that Percent of All in Poverty, Unemployment Rate, % of Disabled Workers, Cost Per Meal, Log Median Household Income, % of People w/ Less than High School Diploma
- Percent of All in Poverty and Log of Median Household Income were most crucial predictors (89.6% of the proportion of the variance in Food Insecurity Rate explained.)
- Outperformed Ridge Regression and Random Forest Models; also offered most interpretability



Conclusion

- Representative of metrics of financial instability, these characteristics point to the notion that economic distress is correlated with food insecurity.
- This seems to agree with recent policy as federal income support in the form of stimulus checks and supplemental UI benefits helped boost food insecurity rates
- The linear regression model is the most favorable, giving clear indications on the importance of the predictors and also providing the lowest errors.