Analysis of Poverty Rate

Calvin He, January 2018

Executive Summary

There are many features that correlates with the poverty rate in a US county, but the ones that stand out are:

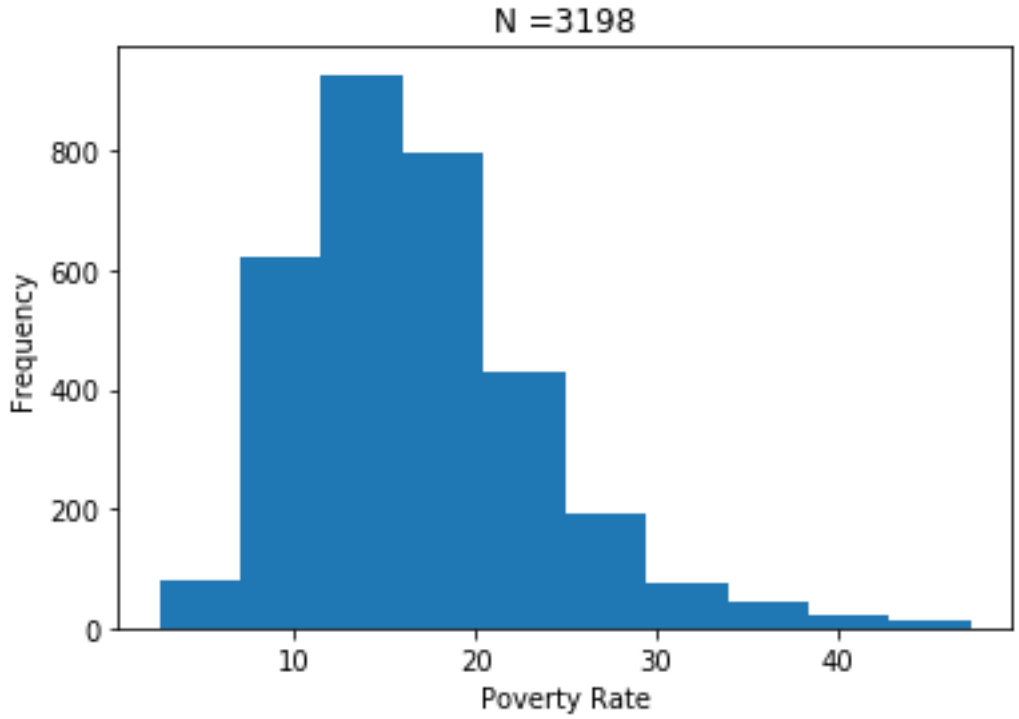
- **Percent Civilian Labor** - Civilian labor force, annual average, as percent of population. Poverty rate is negatively correlated with civilian labor percentage.
- **Percent Unemployment** - Unemployment, annual average, as percent of population. Poverty rate is positively correlated with unemployment.
- **Percent Uninsured Adults** - Percent of adults without health insurance. Poverty rate is positively correlated with percentage of uninsured adults.
- **Percent Uninsured Children** - Percent of children without health insurance. Poverty rate is negatively correlated with percentage of uninsured children.
- **Percent Below 18 years of age** - Percent of population that is below 18 years of age. Poverty rate is negatively correlated with percentage of population below 18 years of age.
- **Percent nonhispanic African American** - Percent of population that identifies as African American. Poverty rate is positively correlated with percentage of nonhispanic African Americans.
- **Percent American Indian or Alaskan Native** - Percent of population that identifies as Native American. Poverty rate is positively correlated with percentage of American Indian or Alaskan Native.
- **Percent Adults less than a High School Diploma** - Percent of adult population that does not have a high school diploma. Poverty rate is positively correlated with percentage of adults with less than a High School Diploma.
- **Percent Diabetes** - Percent of population with diabetes. Poverty rate is negatively correlated with percentage of population with diabetes.

Data Exploration

The metric we are interested in analyzing and predicting is the poverty rate of a US county. The poverty rate of a population is defined as the rate in which families/individuals have incomes that fall below the poverty line, which in 2016 was (for various categories, not all inclusive):

- $12,486 for an individual under 65 years of age
- $11,511 for those 65 years or older
- $19,318 for a family of three (two adults and one child)
- $28,643 for a family of five (two adults and three children)

Our data contains 3198 samples, each from a US county. The poverty rate has a median of 15.8 and a mean of 16.8 with a standard deviation of 6.70. A histogram of the data is shown below:



Economic Typology

We looked at how economic topology affected the poverty rate. The following is a table of median poverty rate given the economic topology:

Economic Typology	Poverty Rate	Count
Farm Dependent	13.6	482
Manufacturing Dependent	15.5	494
Government Dependent	19.35	390
Mining Dependent	16.1	254
Recreation	13.9	312

Economic Typology is defined as such: County Typology Codes "classify all U.S. counties according to six mutually exclusive categories of economic dependence and six overlapping categories of policy-relevant themes. The economic dependence types include farming, mining, manufacturing, Federal/State government, recreation, and nonspecialized counties. The policy-relevant types include low education, low employment, persistent poverty, persistent child poverty, population loss, and retirement destination."

It can be seen that the government dependent economic topology has the highest median poverty rate. This is consistent with the finding that percentage civilian labor is negatively correlated with poverty rate according to the linear regression model. This suggests that poverty rate is smaller for counties when more people work in the civilian sector as opposed to the public sector.

Urban Influence Category

We also looked at how the urban influence category is related to the poverty rate.

Urban influence category is defined as: Urban Influence Codes "form a classification scheme that distinguishes metropolitan counties by population size of their metro area, and nonmetropolitan counties by size of the largest city or town and proximity to metro and micropolitan areas."

A table of all urban influence categories and their median poverty rates are shown below:

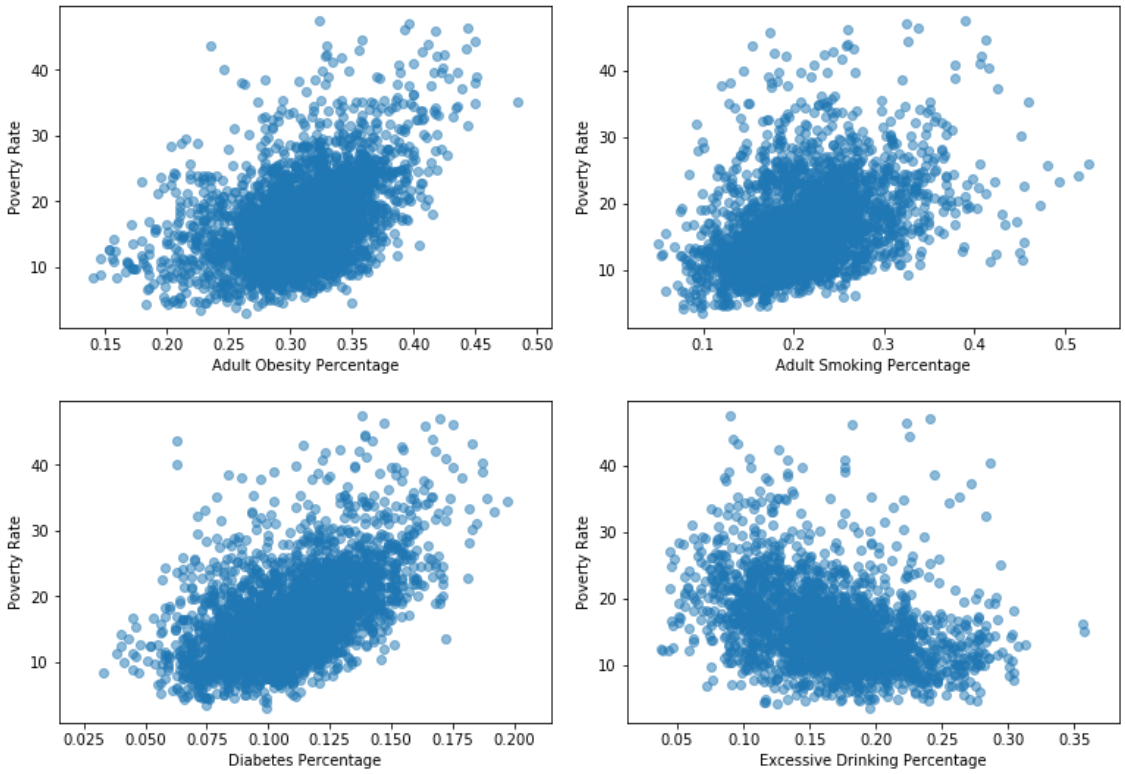
Urban Influence Category	Median PR
Noncore adjacent to a large metro area	15.8
Micropolitan adjacent to a large metro area	16.8
Noncore adjacent to micro area and contains a town of 2,500-19,999 residents	18.25
Large-in a metro area with at least 1 million residents or more	11.9
Micropolitan not adjacent to a metro area	16.6
Noncore not adjacent to a metro/micro area and does not contain a town of at least 2,500 residents	14.9
Noncore adjacent to a small metro with town of at least 2,500 residents	18.9
Small-in a metro area with fewer than 1 million residents	15.5
Noncore adjacent to micro area and does not contain a town of at least 2,500 residents	14.55
Noncore not adjacent to a metro/micro area and contains a town of 2,500 or more residents	13.5
Noncore adjacent to a small metro and does not contain a town of at least 2,500 residents	16.1
Micropolitan adjacent to a small metro area	17.7

We also compared metro and nonmetro classifications for the Rural-Urban continuum codes, defined as: "form a classification scheme that distinguishes metropolitan counties by the population size of their metro area, and nonmetropolitan counties by degree of urbanization and adjacency to a metro area. The official Office of Management and Budget (OMB) metro and nonmetro categories have been subdivided into three metro and six nonmetro categories. Each county in the U.S. is assigned one of the 9 codes."

It was found that the median poverty rate for metro classification, which consisted of 1128 counties, was 14.3, while that for nonmetro classification, with 2070 counties, was 16.8.

Health Indicators

We looked at how certain health indicators (adult obesity percentage, adult smoking percentage, diabetes percentage, and excessive drinking percentage) affected the poverty rate. The following are scatter plots for the various indicators looked at:



It can be seen that there are positive correlations between poverty rate and adult obesity, adult smoking, and diabetes, while there is a negative correlation between poverty rate and excessive drinking.

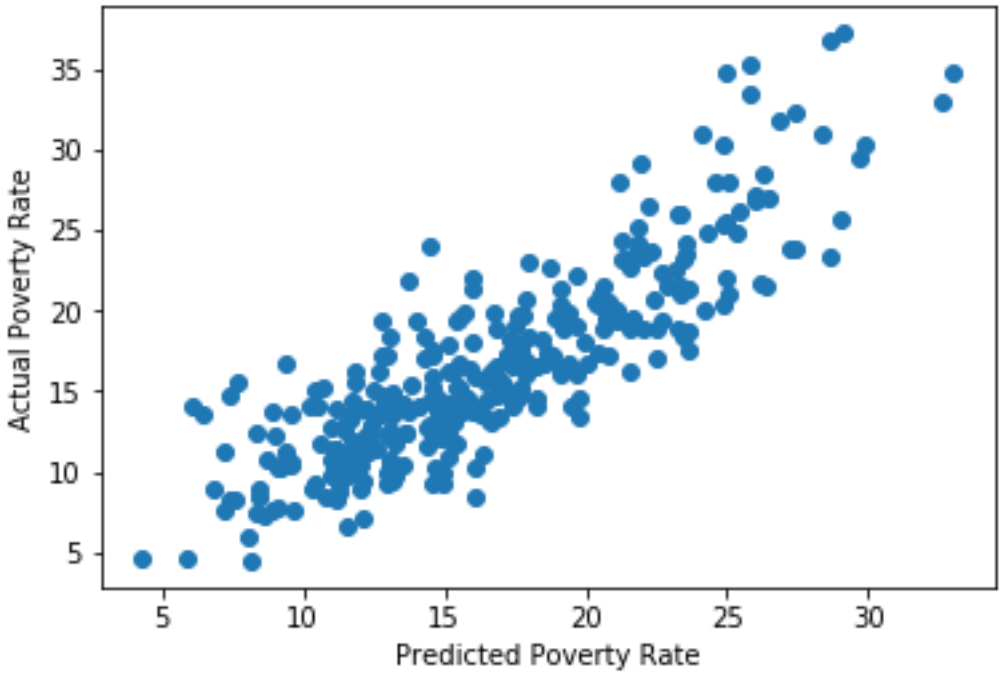
Age and Rural-Urban Classification

We also looked at how the combination of age and rural-urban classification codes (RUCC) related to the poverty rate. The following is a table of median poverty rates according to their age & RUCC combination:

Classification	Median Poverty Rate
Elderly and Metro	14.7
Elderly and Nonmetro	15.6
Nonelderly and Metro	14.3
Nonelderly and Nonmetro	18.95

Data Modeling

The regression model used to model the data was Linear Regression. The data was split so that 90% was training data and 10% was testing data.



The sum of squared errors for this model is 3178.54.

The complete table of coefficients for the linear regression model is given below:

Feature	Coefficient
row_id	-1.77E-05
area__rucc	-0.074368292
area__urban_influence	0.12752852
econ__economic_typology	-0.10019815
econ__pct_civilian_labor	-19.05922449
econ__pct_unemployment	29.23650083
econ__pct_uninsured_adults	23.648731
econ__pct_uninsured_children	-24.56468971
demo__pct_female	2.314815927
demo__pct_below_18_years_of_age	-42.64136749
demo__pct_aged_65_years_and_older	-10.87374923
demo__pct_hispanic	10.7453501
demo__pct_non_hispanic_african_american	18.84573377
demo__pct_non_hispanic_white	8.882711051
demo__pct_american_indian_or_alaskan_native	25.81583644
demo__pct_asian	-0.125719005
demo__pct_adults_less_than_a_high_school_diploma	21.46143751
demo__pct_adults_with_high_school_diploma	-6.605299983
demo__pct_adults_with_some_college	-6.144205046
demo__pct_adults_bachelors_or_higher	-8.711932478

demo__birth_rate_per_1k	0.074663745
demo__death_rate_per_1k	0.136674021
health__pct_adult_obesity	2.940286921
health__pct_adult_smoking	1.833823949
health__pct_diabetes	-19.28509544
health__pct_low_birthweight	6.520794491
health__pct_excessive_drinking	-2.686946135
health__pct_physical_inacticity	6.257773306
health__air_pollution_particulate_matter	0.111361847
health__homicides_per_100k	0.038367587
health__motor_vehicle_crash_deaths_per_100k	0.014461622
health__pop_per_dentist	-5.18E-05
health__pop_per_primary_care_physician	-0.000205876
yr	-0.325524553