Jasmine Young
IDS 702 Final Project Report
November 22nd, 2020

Understanding North Carolina Evictions

Summary

Are county demographics associated with trends in a county's eviction rate? In this analysis we seek to understand if racial, economic, or other factors appear to correlate with trends in eviction rate. We also seek to understand if these trends vary by year or county. In order to answer these questions we use a hierarchical linear model with eviction rate as the response variable and individual intercepts for each county. We found that a higher African-American population, a higher proportion of renters, and our two middle property value groups are associated with increases in the eviction rate. We also found that year was a significant predictor of eviction rate and counties had a substantial amount of variance and deviation from one another.

Introduction

What is an eviction? An eviction is the removal of a tenant from a rental property by the landlord. An eviction filing begins the eviction process, but not all eviction filings go to court and become full-fledged evictions. Evictions can happen for many reasons - including unpaid rent, criminal activity, or breach of lease. Evictions are also often indicative of underlying problems in a community. Evictions can be a symptom of wealth inequality, gentrification, job insecurity, residential segregation, or a lack of economic opportunity.

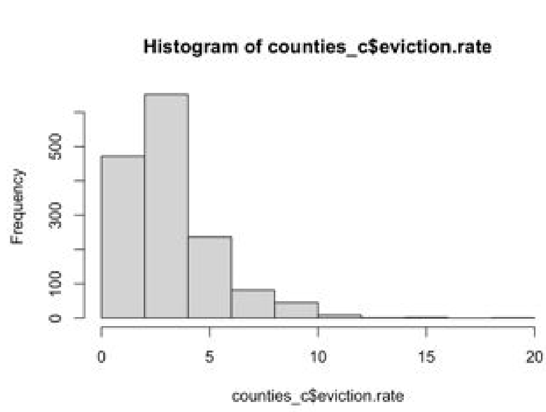
In this analysis we will seek to understand how certain county demographics correlate with or influence eviction rate. We are specifically interested in four inference questions: Does the racial makeup of a county appear to influence the eviction rate? Does the proportion of renters in a county appear to influence the eviction rate? Do property values or other economic factors appear to influence the eviction rate? Do these trends vary by year or county?

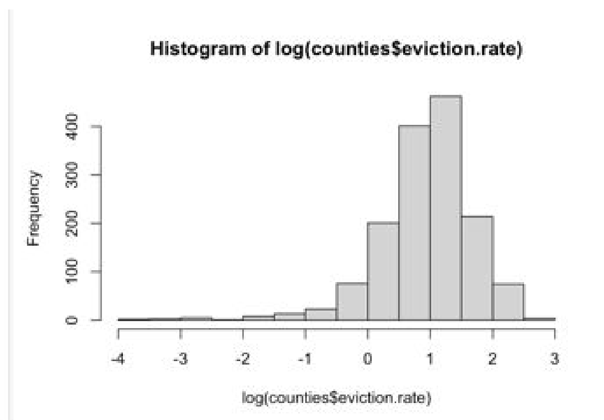
These inference questions are important because they may give us insight into some of the underlying causes of eviction. Racial demographic information could be an indicator of residential segregation in a county. The proportion of renters gives us insight to the culture of the county - and whether it is more common to own or rent your home. Property values and other economic metrics can be an indicator of wealth inequality in a county. And lastly, looking at how counties and years differ from each other allows us to dig more deeply into eviction trends across the state and over time. In this report we will discuss the data used for this analysis, the model selection process and final model, and finally our overall conclusions.

Data

For this analysis we will use data from The Eviction Lab at Princeton University. Our dataset includes eviction and demographic data for 100 North Carolina counties from 2000 - 2016. The data includes information on evictions, eviction filings, population, poverty rate, renter occupancy, median rent, median household income, median property value, and race for each county. In order to deal with missing data we dropped the few observations (less than 20) that did not have a measured eviction rate.

We made important discoveries during our exploratory data analysis. We found that our response variable, eviction rate, was not following a normal distribution. Using the logarithm of eviction rate gave us a more normally distributed response variable.





We also found that some of our other variables were heavily skewed. We decided to create our own categories within Population, Median Household Income, and Median Property Value. Creating categories for these variables improved the clarity of our model and helped produce more normal distributions. We were unable to identify any key interactions in our dataset because of the lack of categorical variables.

Model

Final Model

This analysis employed a hierarchical linear regression model to estimate the impact of certain variables on eviction rate. We decided to use the logarithm of eviction rate as our response variable. Our model included predictors for county population, poverty rate, the percent of renter occupied homes in the county, the median gross rent for the county, the county's median household income, the county's median property value, the year, the calculated rent burden, the percent African-American, the percent Hispanic, and the percent Asian. Our model also includes an individual intercept for each county.

$$\text{Log}(EvictionRate_{ij}) \; = \; (\beta_0 + \gamma_{0j}) + \; \beta_1 Population_{ij} + \beta_2 PovertyRate_{ij} + \beta_3 \% RenterOccupied_{ij} + \beta_4 MedianGrossRent_{ij} + \; \beta_5 MedianHouseholdIncome_{ij} + \beta_6 MedianPropertyValue_{ij} + \; \beta_7 Year_{ij} \; \beta_8 RentBurden_{ij} + \; \beta_9 \% AfricanAmerican_{ij} \; \beta_{10} \% Hispanic_{ij} + \; \beta_{11} \% Asian_{ij}$$

$$i = 1, \ldots n_j \, ; \quad j = 1, \ldots J$$

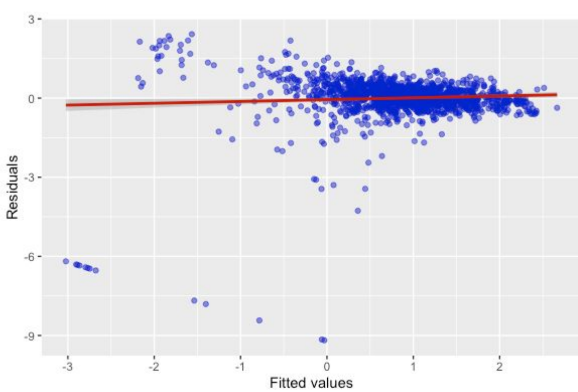
$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

$$\gamma_{0j} \sim N(0, \tau_0^2)$$

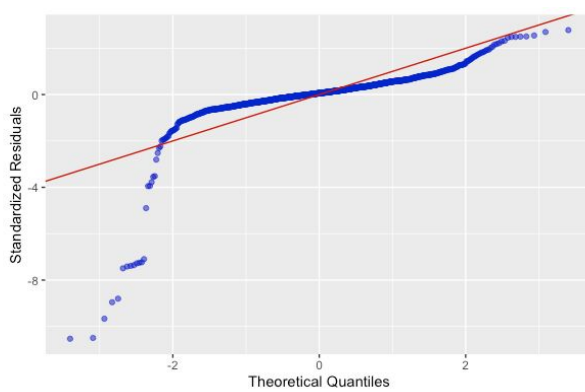
Model Selection/Validation

For this analysis we looked into traditional models and hierarchical models. Because the data lent itself to natural groupings by county, we decided to include individual intercepts for each county and test year as a predictor. We also used model assumption satisfaction to choose a model structure, and found that our hierarchical model with individual intercepts for each county best satisfied our regression assumptions. Within the hierarchical model we tested certain predictors using anova tests. Our anova test showed that year and race significantly improved our model, so we included them in our final model.

We also noticed an issue of multicollinearity with “% White” and “% African-American”. We decided to remove “% White” as a predictor because our anova test determined the “% White” predictor did not significantly improve our model. Removing “% White” solved our issue of multicollinearity. We confirmed again at the end of our Model Selection process that our model assumptions were reasonably met. Our Independence/Equal Variance and Normality assumptions are shown in the graphs below. Our assumptions are reasonably met for counties with higher eviction rates, but assumptions suffer for lower eviction rates.



Residuals vs Fitted values



Normal Q-Q

Model Results

Our model results are shown in the table below.

	log_er		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.14	-0.82 – 1.11	0.770
pop_cat [1]	0.15	-0.09 – 0.40	0.219
pop_cat [2]	0.08	-0.42 – 0.58	0.766
pop_cat [3]	0.46	-0.46 – 1.39	0.326
poverty.rate	-0.03	-0.06 – 0.00	0.054
pct.renter.occupied	0.04	0.02 – 0.06	**<0.001**
median.gross.rent	-0.00	-0.00 – 0.00	0.589
mhi_cat [1]	0.18	-0.02 – 0.38	0.071
mhi_cat [2]	0.17	-0.23 – 0.56	0.404
mpv_cat [1]	0.49	0.29 – 0.69	**<0.001**
mpv_cat [2]	0.40	0.10 – 0.70	**0.009**
mpv_cat [3]	0.48	-0.28 – 1.23	0.214
year-7	-0.70	-1.00 – -0.39	**<0.001**
year-6	-0.56	-0.86 – -0.27	**<0.001**
year-5	-0.54	-0.83 – -0.24	**<0.001**
year-4	-0.47	-0.77 – -0.17	**0.002**
year-3	-0.91	-1.27 – -0.55	**<0.001**
year-2	-0.74	-1.10 – -0.38	**<0.001**
year-1	-0.44	-0.80 – -0.08	**0.017**
year [0]	-0.71	-1.07 – -0.35	**<0.001**
year [1]	-0.46	-0.82 – -0.10	**0.012**
year [2]	-0.50	-0.91 – -0.09	**0.017**
year [3]	-0.55	-0.98 – -0.11	**0.013**
year [4]	-0.77	-1.20 – -0.34	**<0.001**
year [5]	-0.79	-1.22 – -0.36	**<0.001**
year [6]	-0.44	-0.87 – -0.00	**0.048**
year [7]	-0.46	-0.89 – -0.03	**0.036**
year [8]	-0.42	-0.85 – 0.01	0.058
rent.burden	0.01	-0.02 – 0.03	0.489
pct.af.am	0.03	0.02 – 0.04	**<0.001**
pct.hispanic	0.03	-0.00 – 0.06	0.091
pct.asian	-0.09	-0.22 – 0.04	0.192
Random Effects			
σ^2	0.82		
$\tau_{00\ name}$	0.34		
ICC	0.29		
N_{name}	99		
Observations	1499		
Marginal R^2 / Conditional R^2	0.248 / 0.470		

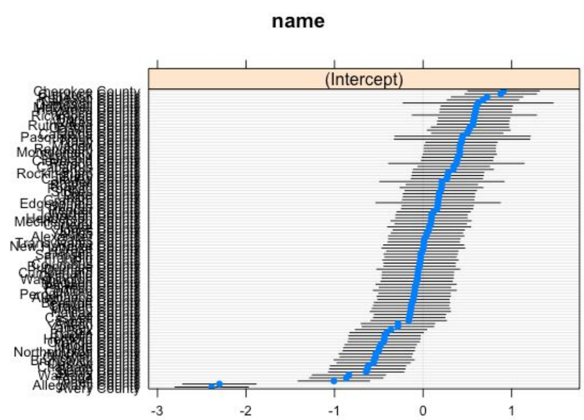
We begin with our first inference question - Does the racial makeup of a county appear to influence the eviction rate? We found that the percentage of African Americans in a county significantly influences the eviction rate with over 99% significance. We notice a coefficient of 0.03, which once exponentiated results in a 3.05% multiplicative increase associated with eviction rate for each additional percent of the population that is African-American. The percentage of hispanic and asian populations did not significantly impact eviction rate.

Our second inference question asks - Does the proportion of renters appear to influence the eviction rate? We found that the percentage of renter occupied homes significantly influences the eviction rate with over 99% significance. We notice a coefficient of 0.04, which once exponentiated results in a 4.08% multiplicative increase associated with eviction rate for each additional percent of homes that are renter occupied.

Our third inference question asks - Do property values or other economic measures appear to influence the eviction rate? We found that the first Median Property Value group (100k - 150k) and the second Median Property Value group (150k - 250k) appear to significantly influence eviction rate compared to counties with a Median Property Value below 100k with over 99% significance. Counties with a MPV of 100k-150k have a coefficient of 0.49, which when exponentiated results in a 63.23% multiplicative increase associated with eviction rate compared to counties with a MPV below 100k. Counties with a MPV of 150k-250k have a coefficient of

0.40, which when exponentiated results in a 49.18% multiplicative increase associated with eviction rate compared to counties with a MPV below 100k.

Our last inference question asks - Do these trends vary by Year or County? We found that all years are associated with significant downward trends in eviction rate compared to the year 2000, except for 2016, at the 95% level or greater. All years show significant multiplicative decreases in eviction rate. They range from a decrease of 59.75% - 35.60% in the eviction rate compared to the year 2000. In the chart below we see that there is a decent amount of variation by county. We note two extreme outliers on the left-hand side of our chart below beyond -2. Our counties vary with a standard deviation of 0.56 within county and 0.91 between counties.



Conclusions

In conclusion, certain county demographics are associated with increasing or decreasing trends in eviction rate. We found that counties with larger African-American populations and more renters are more likely to have higher eviction rates. We also found that counties with median property values from 100k - 150k and 150k - 250k were associated with significant increases in eviction rate. This indicates that our poorest and richest counties experience less evictions than our mid-range counties. Counties in the middle may be experiencing combinations of higher and lower income individuals that result in supply/demand issues when it comes to housing. Lastly, we found that year and county played significant roles in understanding eviction rate and there was variance between years and counties.

This analysis also contained some limitations and potential for future analysis. Our regression assumptions, though improved, were not ideal. We were limited by the scarcity of data in certain categories, like high property values, and the lack of normality within certain variables. In the future it may be beneficial to study eviction filings in addition to eviction rates, as well as expand this study to more states. Other groupings, like zip code and neighborhood, also might allow for a more granular analysis of eviction trends and community demographics.