Research Design

Katelyn Diaz, Sarah Fleming, Shevaughn Holness

4/1/2022

Abstract

The purpose of our analysis is to investigate the relationship between a county's uninsured rates and our two primary demographic determinants: median household income and non-US citizenship percentage. Using data from the 2019 American Community Survey collected by the U.S. Census Bureau, we derive six models with a county's uninsured rates as our dependent variable. Our analysis reveals that household median income and non-US citizenship are statistically significant predictors of a county's uninsured rates, but only the non-US citizenship percentage holds a substantive effect on the uninsured population. Our analysis supports evidence regarding the correlation between demographic determinants and health insurance rates.

Background

Despite the United States spending 19% of its GDP on healthcare in 2020 – the most of any other advanced country—health insurance rates have been declining unevenly since 2017, especially as it pertains to rates among marginalized communities (Medicare and Medicaid Services n.d.). Health insurance not only serves an important role in reducing medical debt but it has also been linked to a significant decrease in mortality (Card, Dobkin, and Maestas 2007). Despite the overall benefits to health, nearly 8.6% of the US population did not have health insurance in 2020 (US Census Bureau n.d.d).

While health insurance has undergone a multitude of reforms over the past decade, not all demographics have felt the reforms evenly, with certain marginalized communities facing an unsteady downward trajectory due to frequent policy changes. With the passing of the Affordable Care Act in 2010, there were significant expansions in terms of healthcare coverage, with racial minorities seeing a significant increase in the insured population. However, later policy changes made by the Trump administration reversed these trends, with most minorities having an increase in the uninsured rates (Damico 2021). In light of these shifts occurring in uninsured rates, we aim to investigate the demographic determinants of county-level uninsured rates.

The two county-level demographic determinants of interest for our analysis are county median household income and the percentage of non-US citizen residents. Household median income is a significant factor regardless of employment status because higher-income households can afford private insurance. Yet, with over half of the population getting employer-provided health insurance, marginalized communities who face obstacles to employment – such as racial minorities, the disabled, and undocumented immigrants – are all historically shown to have higher rates of being uninsured (Damico 2021). Jennifer Tolbert and Kendal Orgera also found that 73.7% of uninsured adults attributed their lack of health insurance due to the high cost of coverage (Tolbert, Nov 06, and 2020 2020).

Literature also seems to suggest that citizenship status is a crucial factor in determining uninsured rates, as undocumented people also often face barriers to employment, meaning that they may be less likely to get a job with benefits, and thus, get health insurance from their employers. Michael S. Cohen's and William L. Schpero (2018) also found that mixed-citizenship status households were less likely to have health insurance due to the fear that it would warrant further inspection into the legality of household members (Cohen and Schpero 2018).

As such, we are going to investigate the relationship between county US. citizenship percentage and county median household income on uninsured rates for a county. The null hypothesis states that there is no significant relationship between either variable and uninsured rates whereas the alternative hypothesis states there is a significant relationship between the variables and uninsured rates. We suspect that we will reject the null hypothesis. In other words, we hypothesize that there is a significant relationship between both US. citizenship percentage and county median household income on uninsured rates for a county.

Methods

Data

The dataset we will be using comes from the 2019 American Community Survey (ACS) collected by the United States Census Bureau; we chose 2019 as the year of interest as it is the most up-to-date and validated ACS (US Census Bureau n.d.a), (US Census Bureau n.d.c). The ACS collects demographic, housing, and economic information on the county level for the United States, Puerto Rico, and Guam. The U.S Census Bureau collected this information through the use of mailed questionnaires, telephone interviews, and visits from Census Bureau field representatives. This data is collected from 3.5 million household addresses annually and only in counties with at least 65,000 people (US Census Bureau n.d.b).

The unit of analysis for this dataset is county, and there are eight hundred fifty-one total observations. From this dataset, we chose to focus on counties that are in the United States which leads to our dataset having eight hundred twenty-one observations. As we can only include a portion of the counties in the U.S., our dataset is a sample as the population of interest is all counties in the U.S.

Variables

The analysis conducted includes the following variables: median household income, disability status, marriage status, employment status, US citizenship status, uninsured, and majority racial group.The response variable was the percentage of people who are uninsured, which is a continuous variable. The key variables analyzed were the continuous variables median income per county in dollars, and percentage of people who are citizens of the US.

The covariates included the continuous variables percentage of people with a disability, percentage of people married, and percentage of people employed, along with the reference binary variable of if the county is majority white. For our analysis, race was mutated into a categorical variable consisting that stated which racial group held the majority in that county. The reference group is `majority_nonwhite` (which indicates the county is not majority white). The other group in the model are `majority_white` (which indicates the county is majority white).

Analysis

Null Hypothesis 1: The percentage of Non-US citizens in a county does not have a significant relationship with county uninsured rates.

Alternative Hypothesis 1 : The percentage of Non-US citizens in a county has a significant relationship with county uninsured rates.

Null Hypothesis 2: The median household income in a county does not have a significant relationship with county uninsured rates.

Alternative Hypothesis 2: The median household income in a county has a significant relationship with county uninsured rates.

Our primary hypothesis is that the percentage of Non-US Citizens inside a county as well as the median household income of a county have a significant relationship with the percentage of uninsured people in that county. As such, we want to start with simple linear regression to show the relationship between each of those variables independently before moving to multiple regression to view their relationship with uninsured rates when controlling for other variables. Two of our models will be simple linear regression models that have either percentage of non US-citizens inside the county or median household income of the county as an explanatory variable for percentage of people who are uninsured. This would show if there was a relationship between the explanatory and the response variable when not controlling for other factors.

We would then create numerous multiple regression models. The third model we create will have only the percentage of Non US-Citizens inside a county and the median household income as explanatory variables. We will include a full model that has all the covariates as well as the independent variables as a focus. The last few models will be the model produced after accounting for any potential multicollinearity.

All models will be shown; the model that we will choose to discuss is the model that is indicated to be best fitted model by its R-Squared Adjusted, AIC, and BIC. We will check regression diagnostics for each model before running them.

Results

Table 1: Predicting County Uninsured Rates

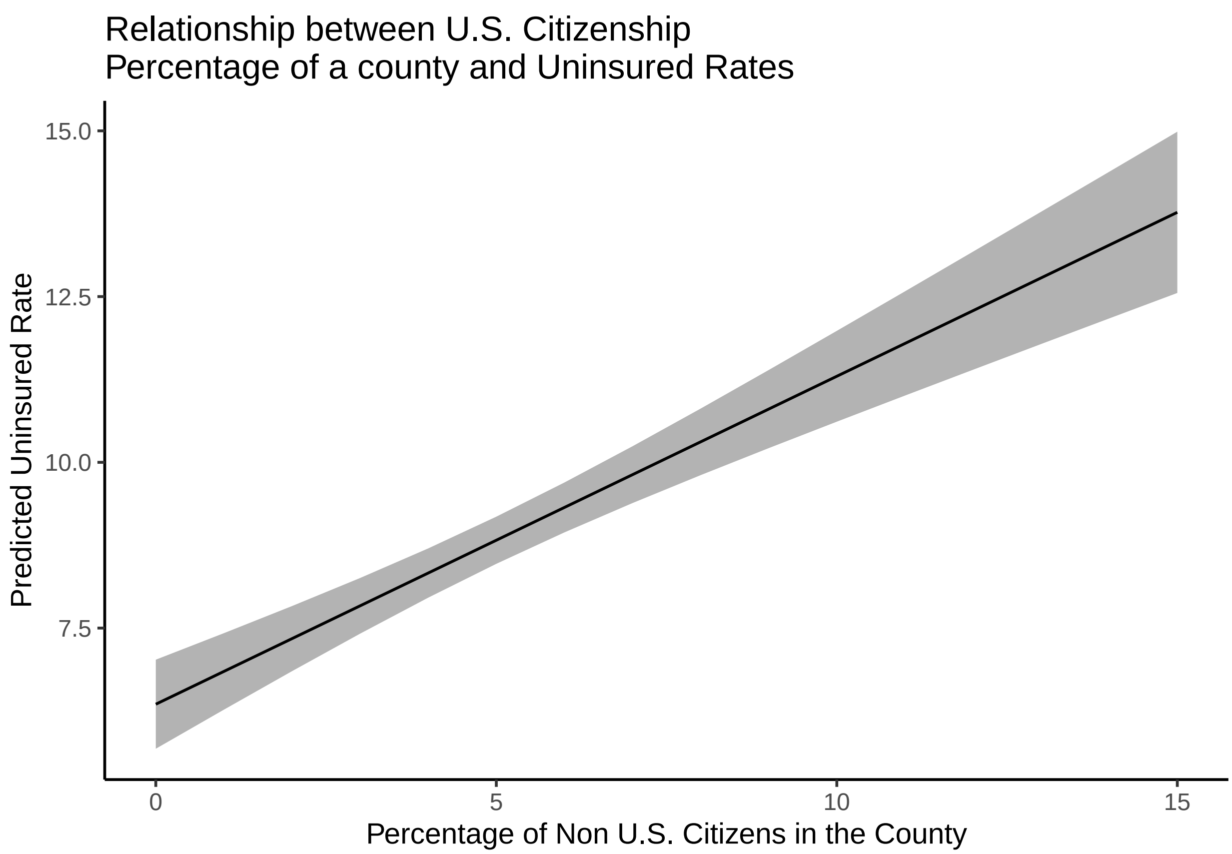
	Dependent variable:					
	uninsured_pct					
	(1)	(2)	(3)	(4)	(5)	(6)
Non US Citizenship Percentage	0.191***		0.423***		0.346***	0.495***
	(0.048)		(0.056)		(0.056)	(0.059)
Household income (log)		−0.711***	−1.752***	−0.627***		−1.429***
		(0.209)	(0.240)	(0.228)		(0.233)
Employment Percentage				−0.266***	−0.307***	−0.272***
				(0.078)	(0.076)	(0.073)
Disabled Percentage				−0.012	0.270***	0.197**
				(0.082)	(0.083)	(0.081)
Marriage Percentage				−0.049	−0.038	−0.049*
				(0.030)	(0.029)	(0.028)
Racial Majority: White				−3.125***	−1.716***	−1.748***
				(0.623)	(0.627)	(0.604)
Constant	8.147***	18.119***	29.015***	38.653***	25.804***	42.564***
	(0.316)	(2.637)	(2.875)	(7.085)	(6.270)	(6.626)
Adjusted R^2	0.030	0.022	0.127	0.112	0.166	0.227
Akaike Inf. Crit.	2,676.759	2,680.671	2,627.998	2,639.074	2,609.245	2,574.523
Bayesian Inf. Crit.	2,689.230	2,693.141	2,644.626	2,668.173	2,638.344	2,607.779
Note:						*p<0.1; **p<0.05; ***p<0.01

The non-applicable observations (due to missing data) were filtered out, then an analysis was conducted on a sample size of 472 counties. Using the adjusted R-Squared, AIC, and BIC, it was determined that model 6 - the full model - held the most explanatory power. A VIF test also showed that there was no multicollinearity in our model. The analysis was conducted at a .05 significance level. Our model shows that there is a statistically significant relationship between both county median household income and county US citizenship percentage on county uninsured rates. Thus, we reject both of the null hypotheses which stated that neither variable had a significant relationship with uninsured rates.

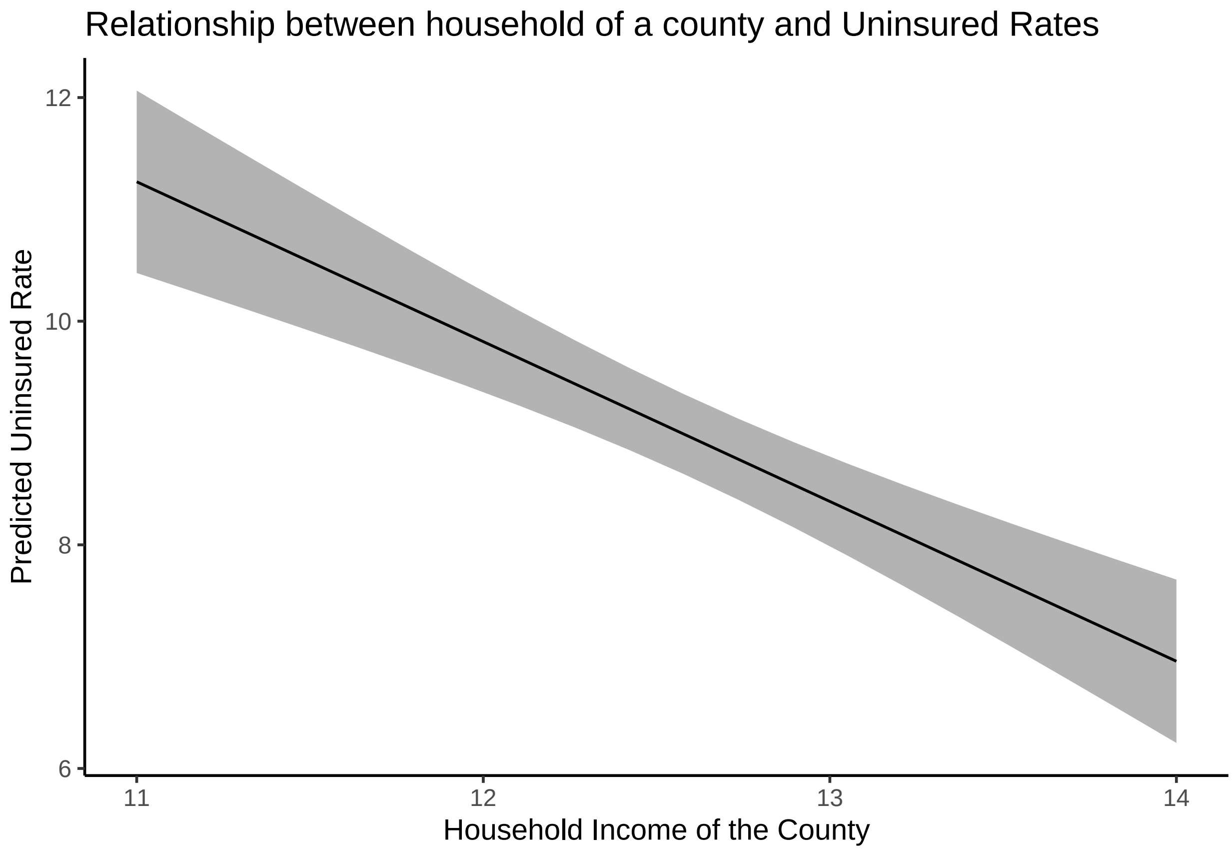
Most of our covariates, with the exception of marriage percentage, also had a significant relationship with county uninsured percentages. Employment percentage of a county (coefficient = -0.272) as well as being a majority white county (coefficient = -1.5748) had a negative relationship with uninsured rates which was expected. Percentage of people with disabilities (coefficient = 0.192) had a positive relationship which was also consistent with the literature. Because marriage percentage is not significant, the data supports it has no relationship with uninsured rates.

In order for our model to have a linear relationship, we had to log transform household median income. The log transformation of household income had a significant negative relationship with uninsured percentage. After transforming the coefficient for interpretability, we found that (on average) a one percent increase in county household median income is associated with a 0.01429 increase in uninsured rates. This means that there is a significant negative relationship between uninsured rates and household income. However, a 0.01429 increase in percentage of uninsured rate is a very small number. Thus, we do not consider this result substantively significant.

We see the opposite relationship for citizenship percentage, which has a significant positive relationship with uninsured percentage.We found that (on average) a one percent increase in a county's citizenship percentage is associated with a .495 increase in uninsured rates. We believe that that coefficient is a large enough value to be considered substantively significant.



The graph above shows the relationship between the percentage of non US citizens in a county and uninsured rates holding all other variables in the model constant. It shows that there is a clear positive linear relationship between the percent of non-US citizens in a county's population and the uninsured rate in that county.



The graph above shows the relationship between the county median household income and uninsured rates holding all other variables in the model constant. It shows that there is a clear negative linear relationship between the percent of non-US citizens in a county's population and the uninsured rate in that county.

Conclusion

The focus of this research was to investigate the demographic determinants of county-level healthcare uninsured rates with specific focus on US citizenship status and median household income. More specifically, we aimed to determine if there was a significant relationship between percentage of people in a county and county uninsured rates as well as county median household income and county uninsured rates.

When creating a simple linear regression model that examined each variable on its own, without covariates, we saw that each had significant explanatory power for uninsured rates. When we combined both variables and covariates into a single model, which happened to be our best fitting model, both variables were shown to have a statistically significant relationship with county uninsured rates. County uninsured rates have a negative relationship with county median household income whereas it had a positive relationship with county US citizenship percentage. Although both variables were statistically significant, we believe that only US citizenship percentage is substantively significant because the coefficients for were very small.

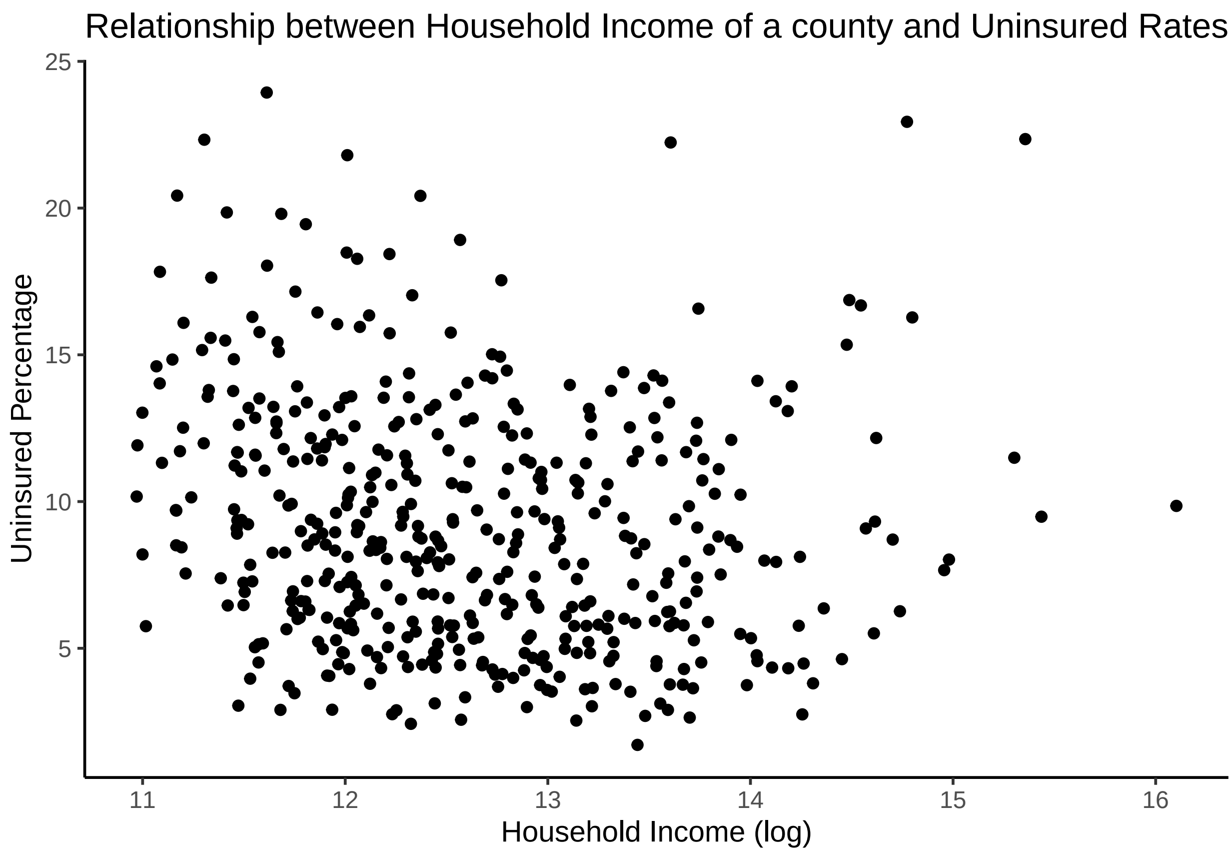
One big limitation of our study was that the ACS only surveys places with over 65,000 people; this means that many smaller US counties were left out of the survey. This could mean that the findings of our analysis may only be representative of large counties rather than all counties of any size. Were we to re-do this analysis, we would utilize the most recent US census which would have information on all counties.

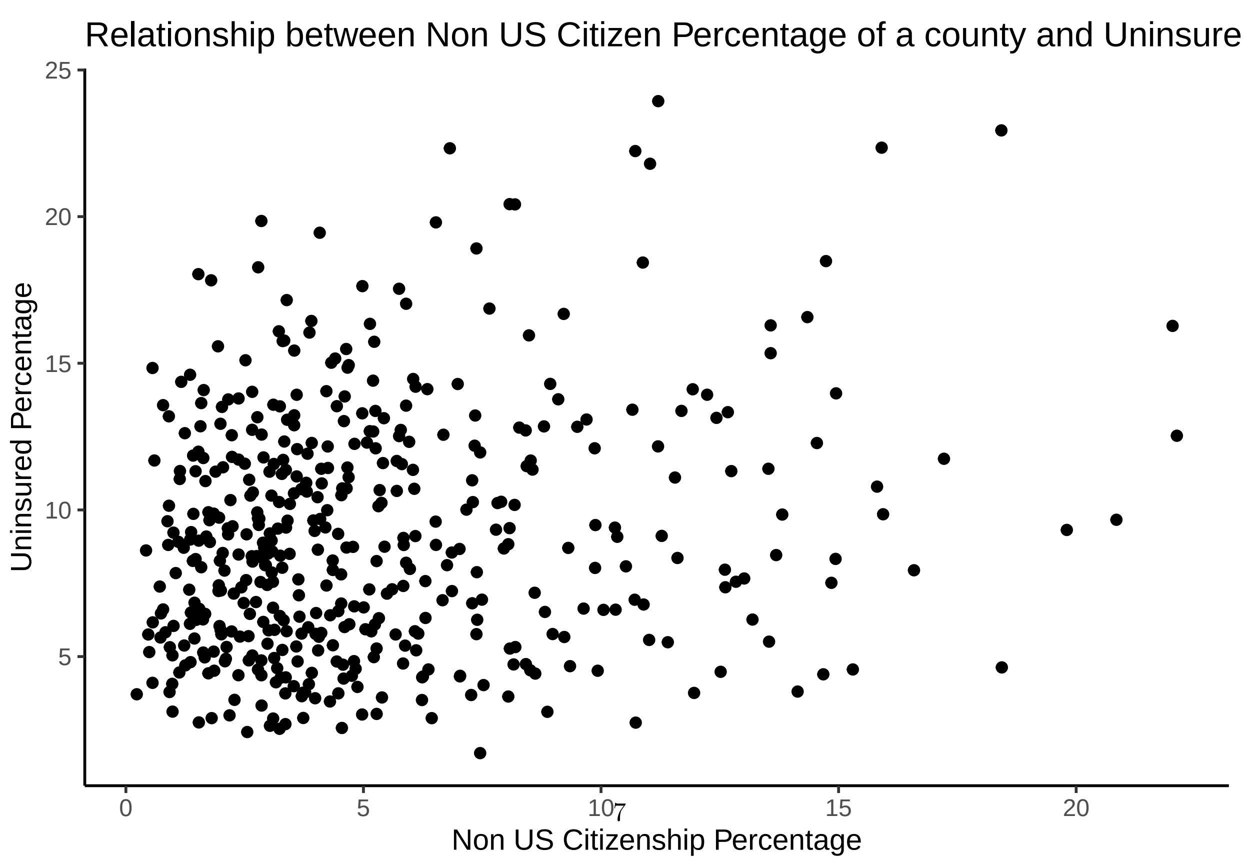
Another limitation is that we did not control for the geographic component of healthcare insurance. In 2019, 39 states elected to expand Medicaid as part of the Affordable Care Act to nearly all adults 137% above the poverty line, significantly increasing the insured rates of lower-income households (Kaiser Family Foundation). As these reforms were not spread evenly across all regions, controlling for the change could have had a significant effect on certain coefficients including our two variables of interest.

Lastly, we cannot draw any causal conclusion about the relationship between county median household income or county US citizenship percentage and uninsured rates. We can only point to there being a significant relationship between the variables and uninsured rates based on the data we have. More nuanced studies that account for a plethora of factors would have to be done to draw a causal relationship.

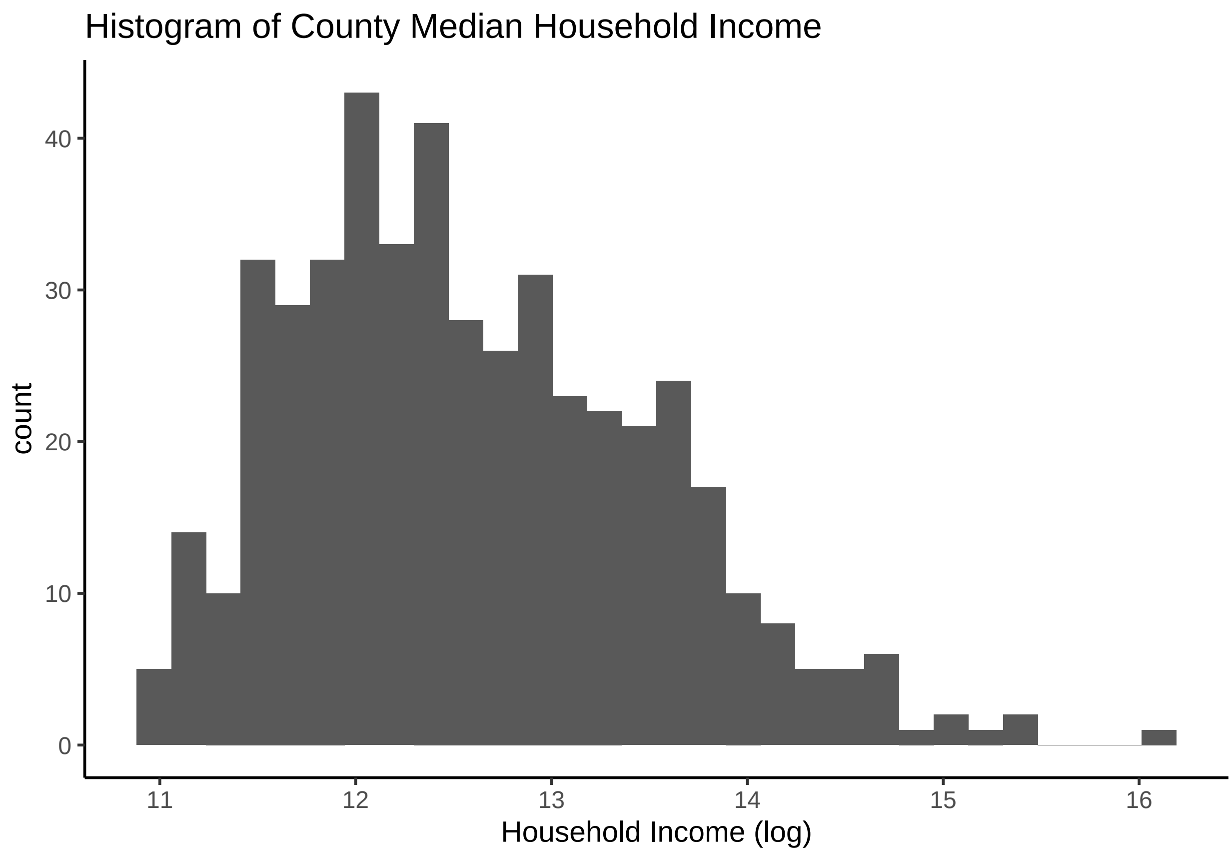
Appendix

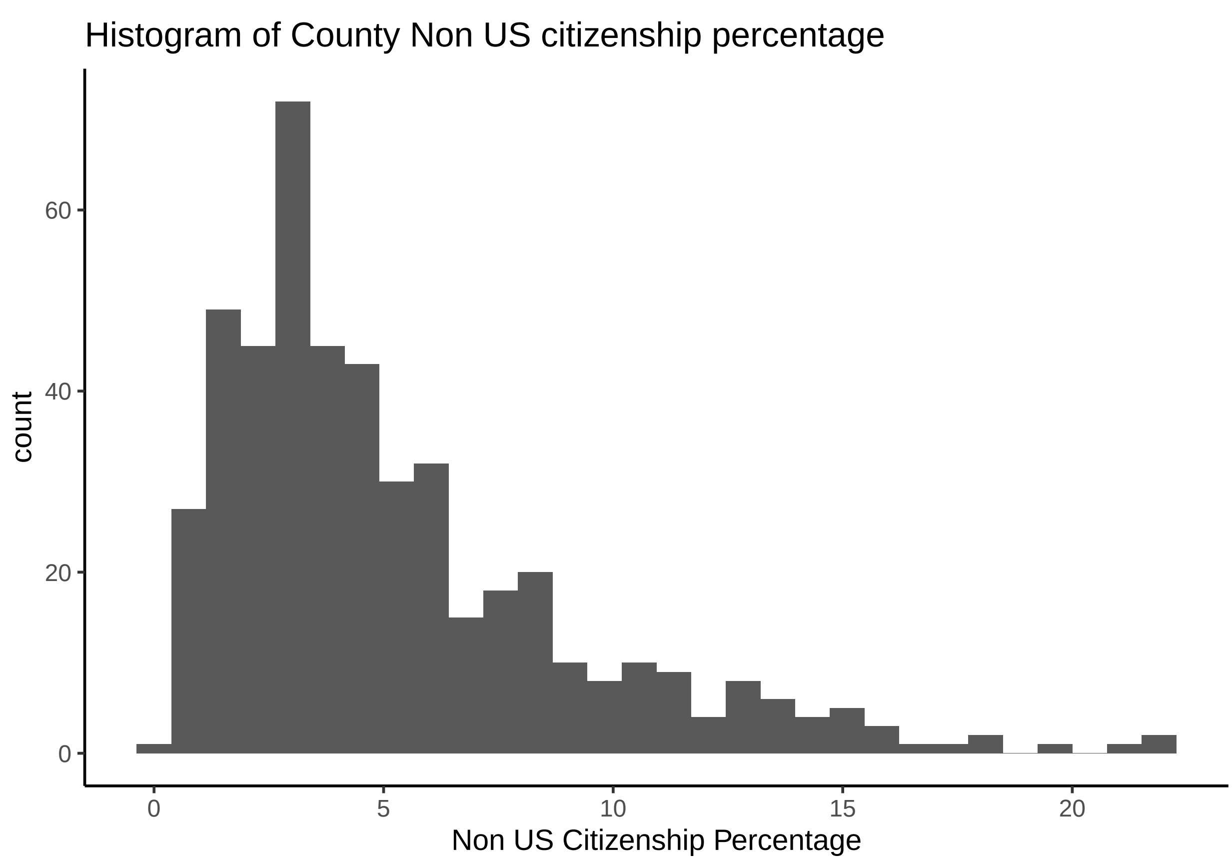
The scatterplots for our variables of interest





Histograms of variables of interest

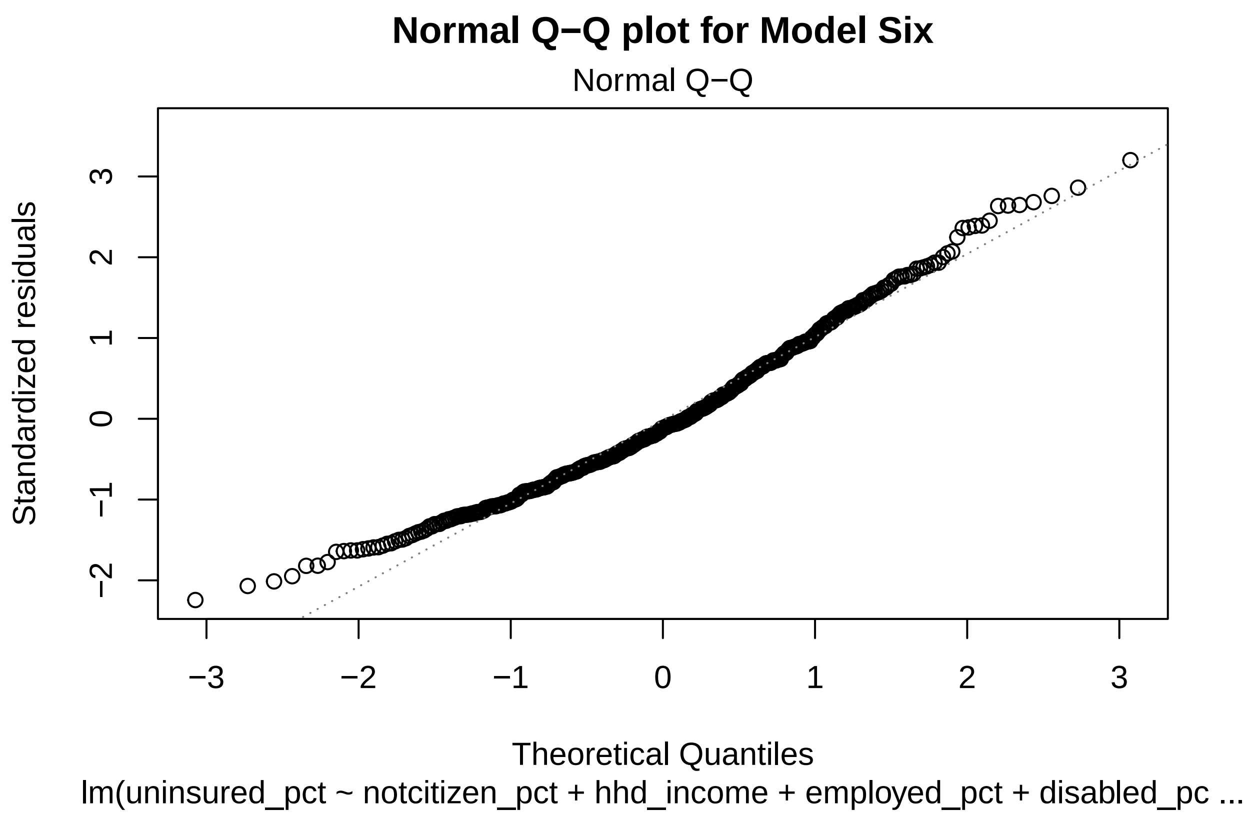


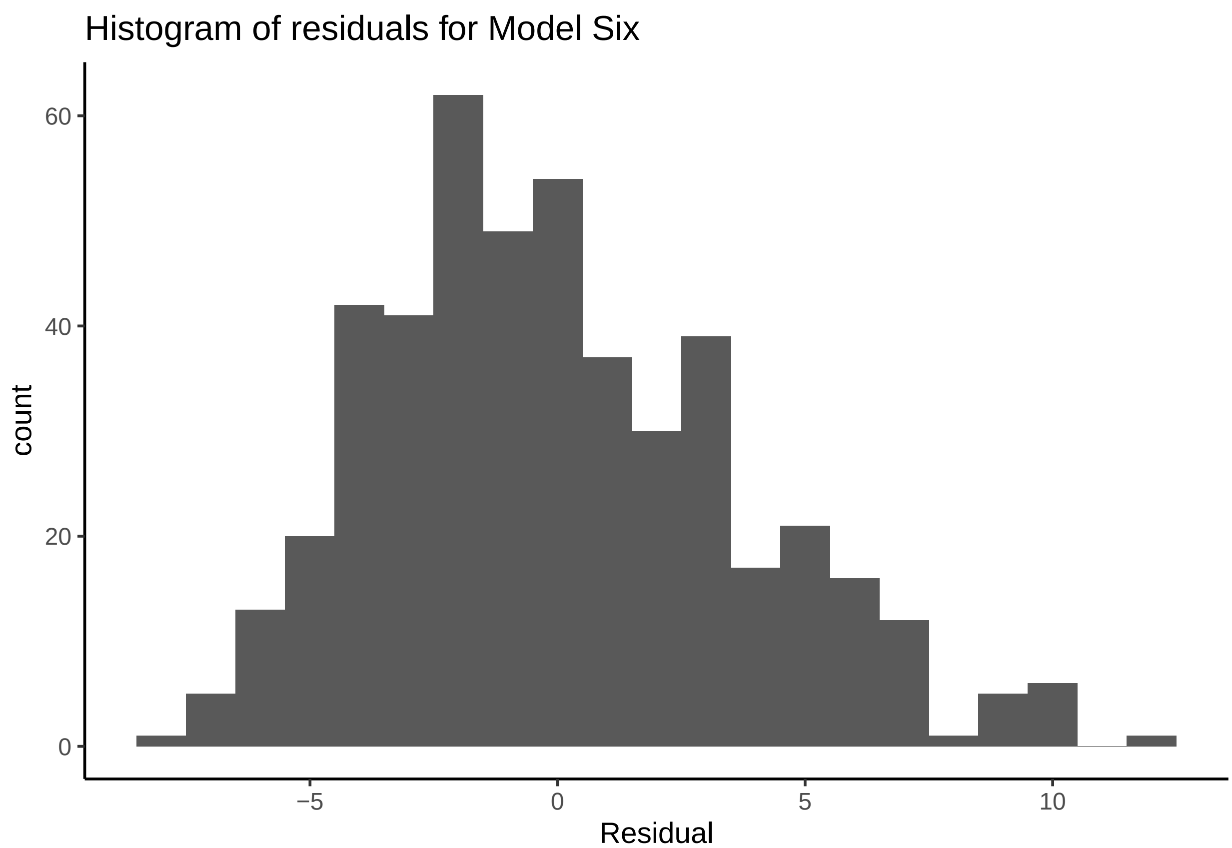


Variance Inflation Factor tests for model in the order of model 3, 6, 4, 5.

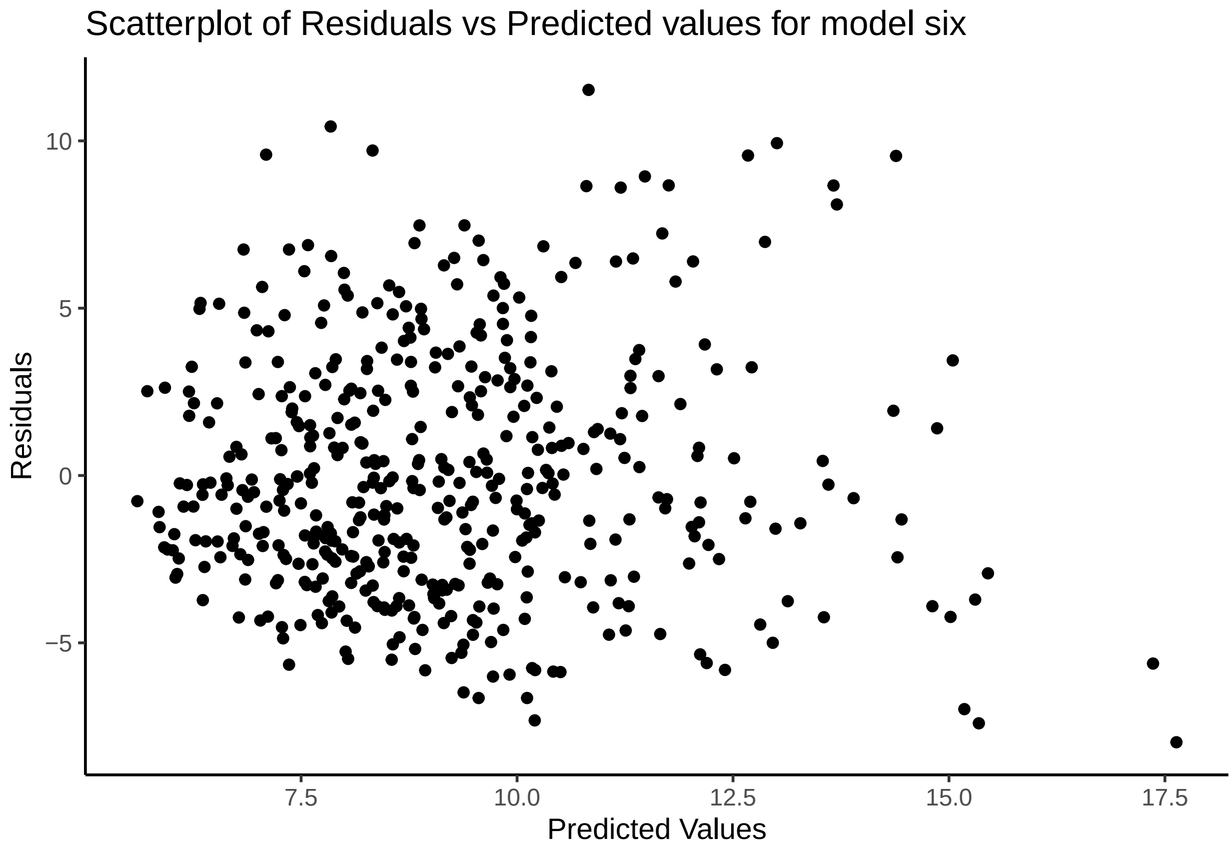
Testing models for assumptions

Tests for model six. This model did pass assumptions.



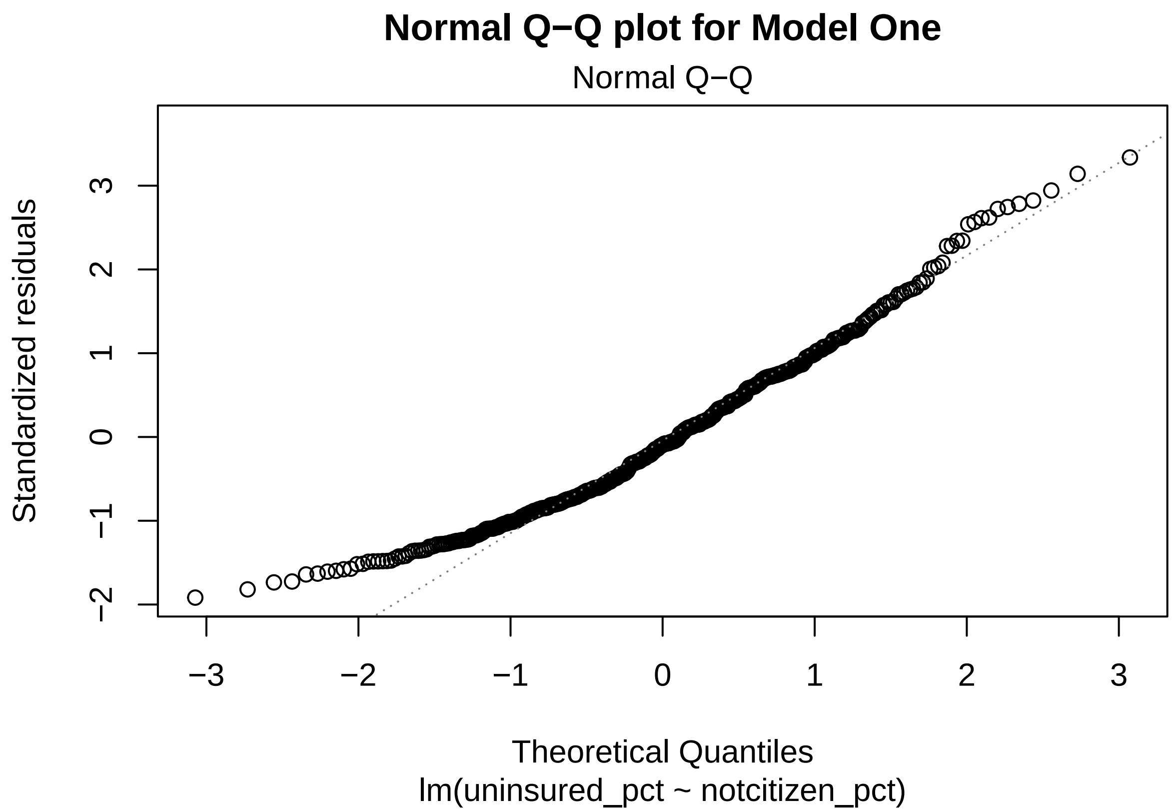


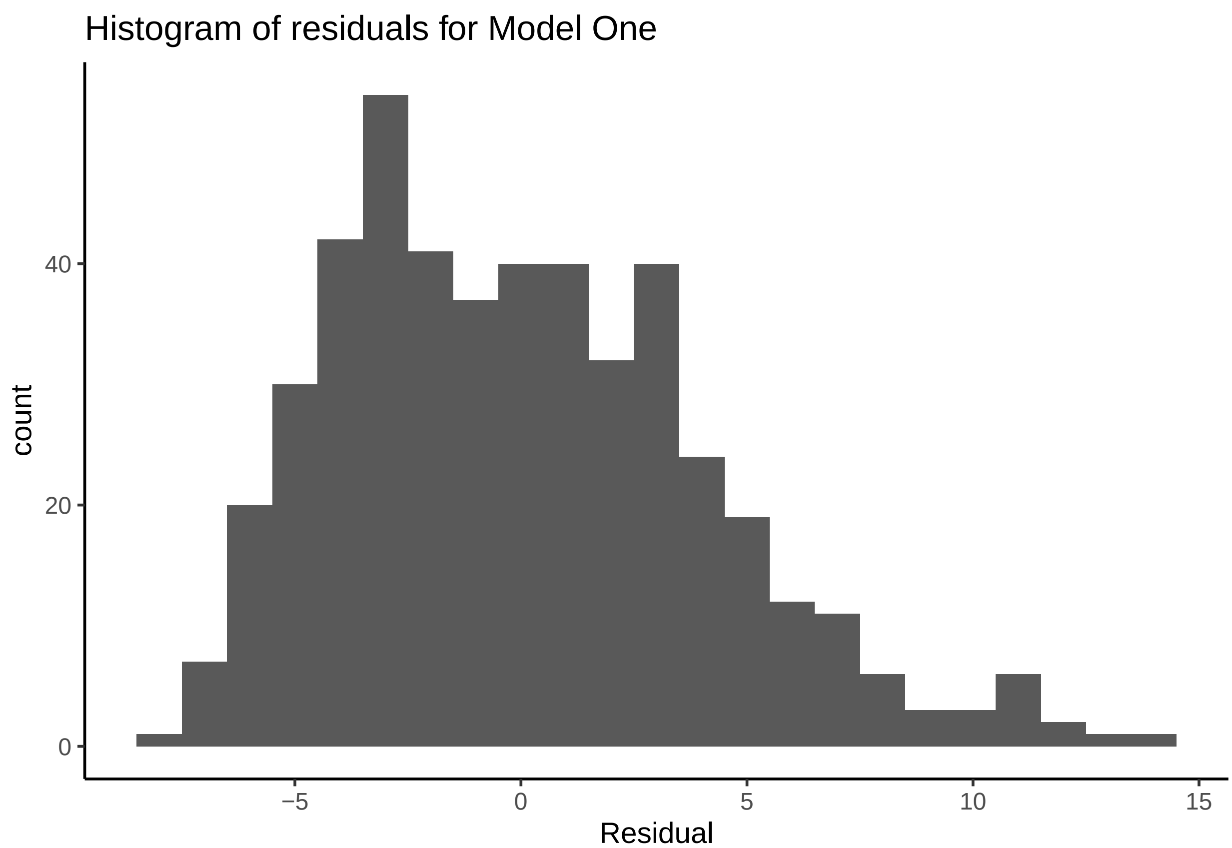
Histogram of residuals for Model Six
count
60
40
20
0
−5
0
5
10
Residual

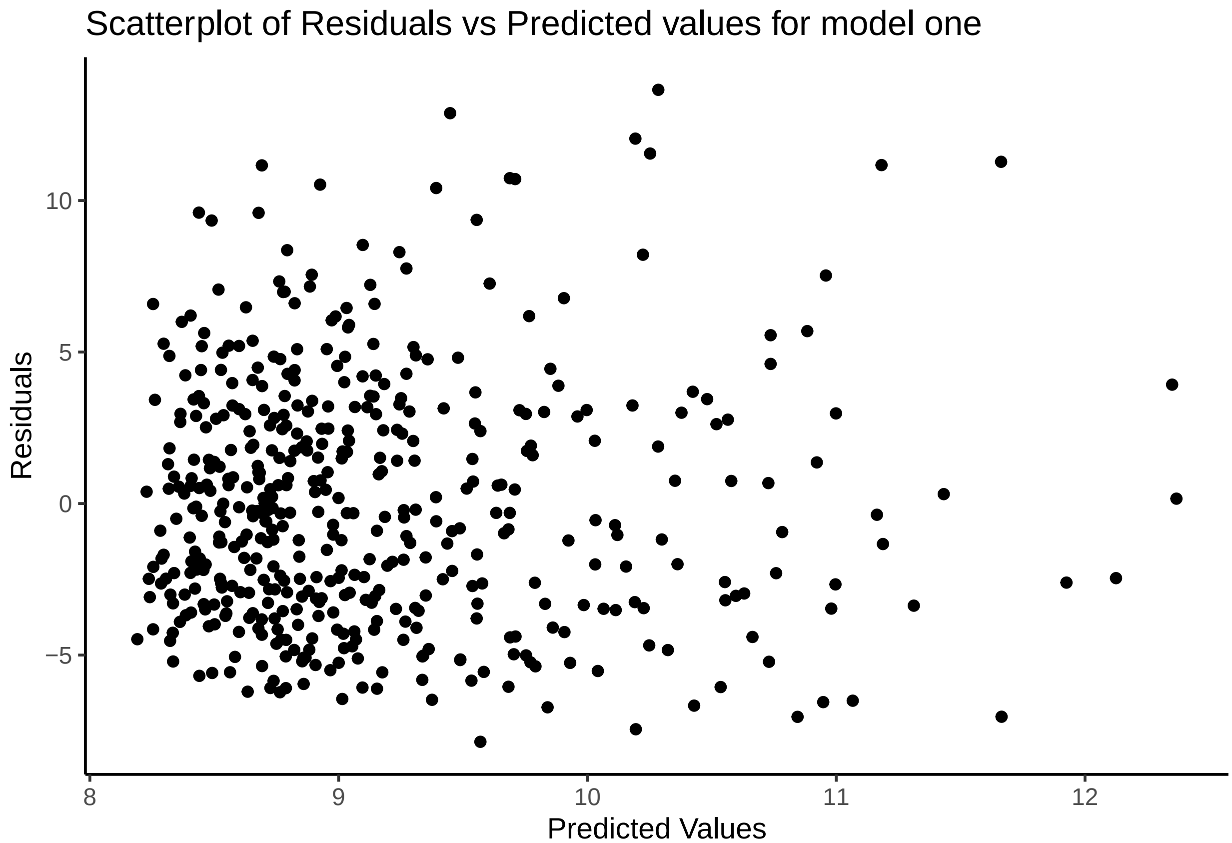



```
## NULL
```

Tests for model one. This model did pass assumptions.

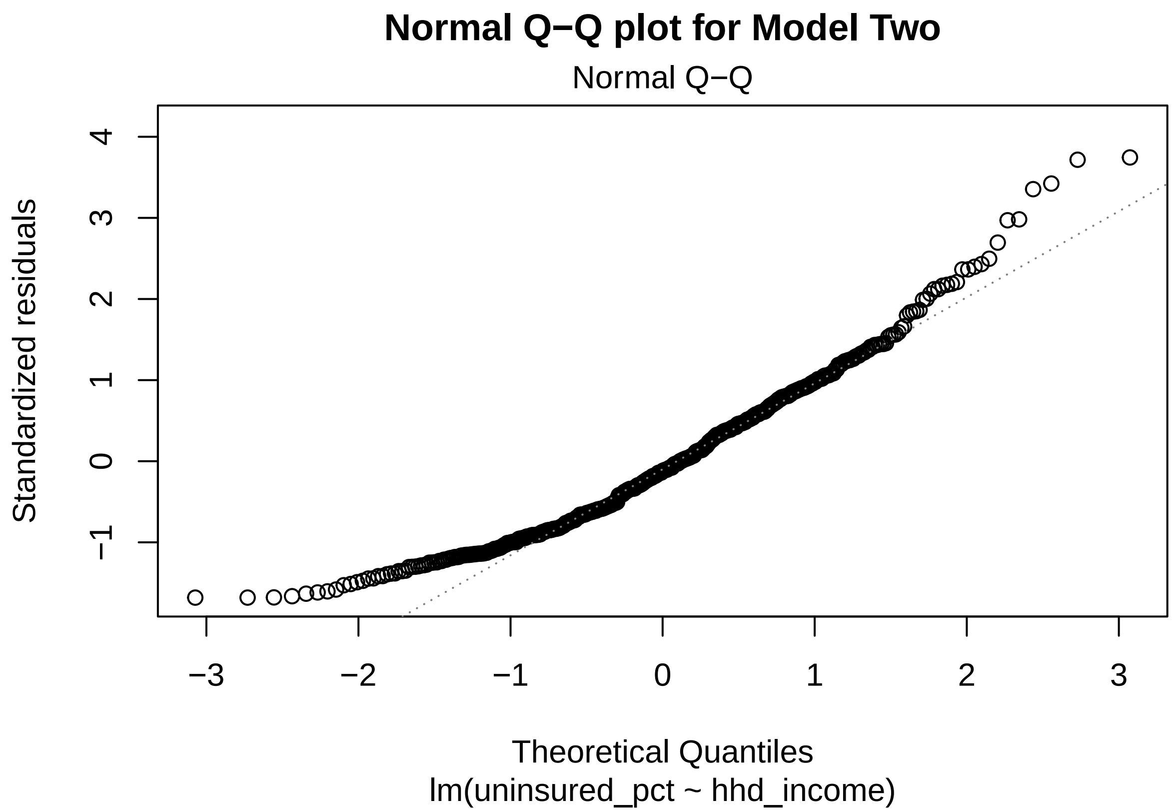

Normal Q−Q plot for Model One
Normal Q−Q
Standardized residuals
−2
−1
0
1
2
3
−3
−2
−1
0
1
2
3
Theoretical Quantiles
lm(uninsured_pct ~ notcitizen_pct)

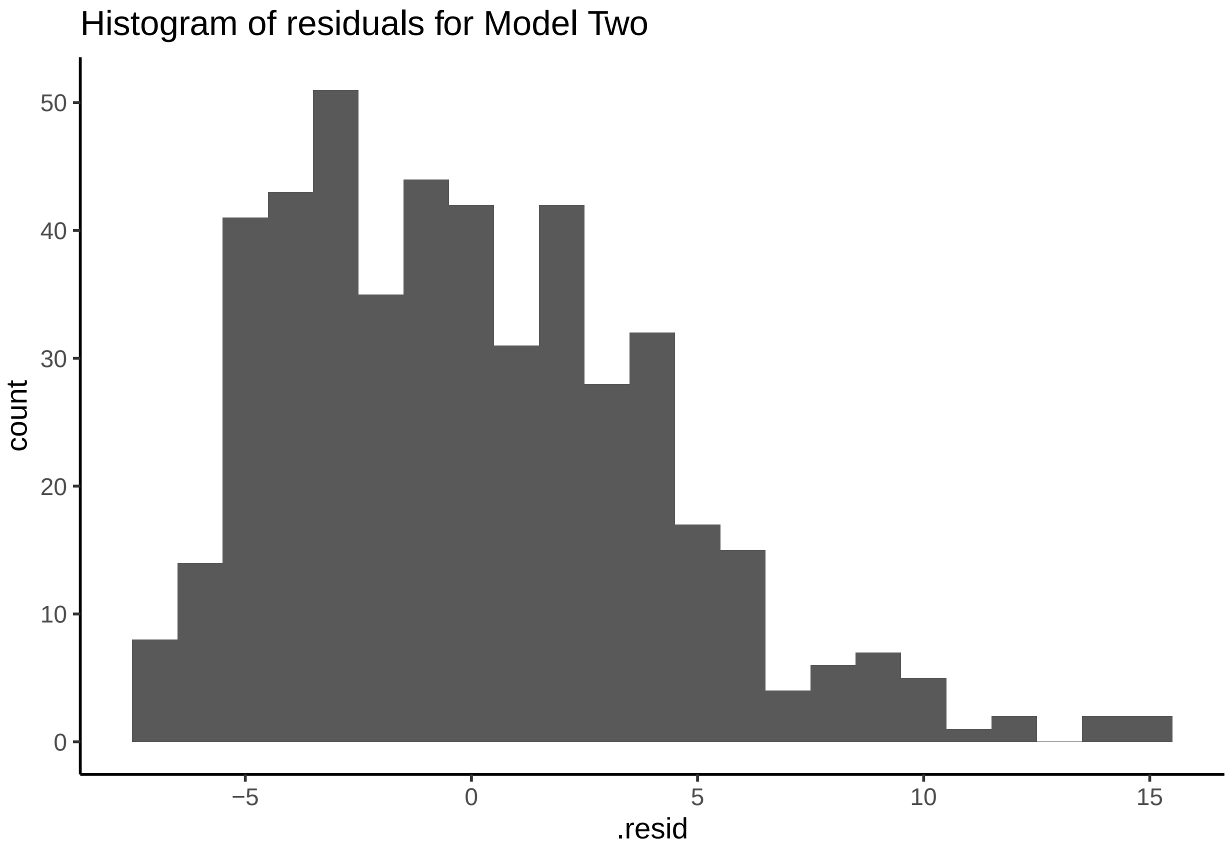

Histogram of residuals for Model One
count
40
20
0
−5
0
5
10
15
Residual

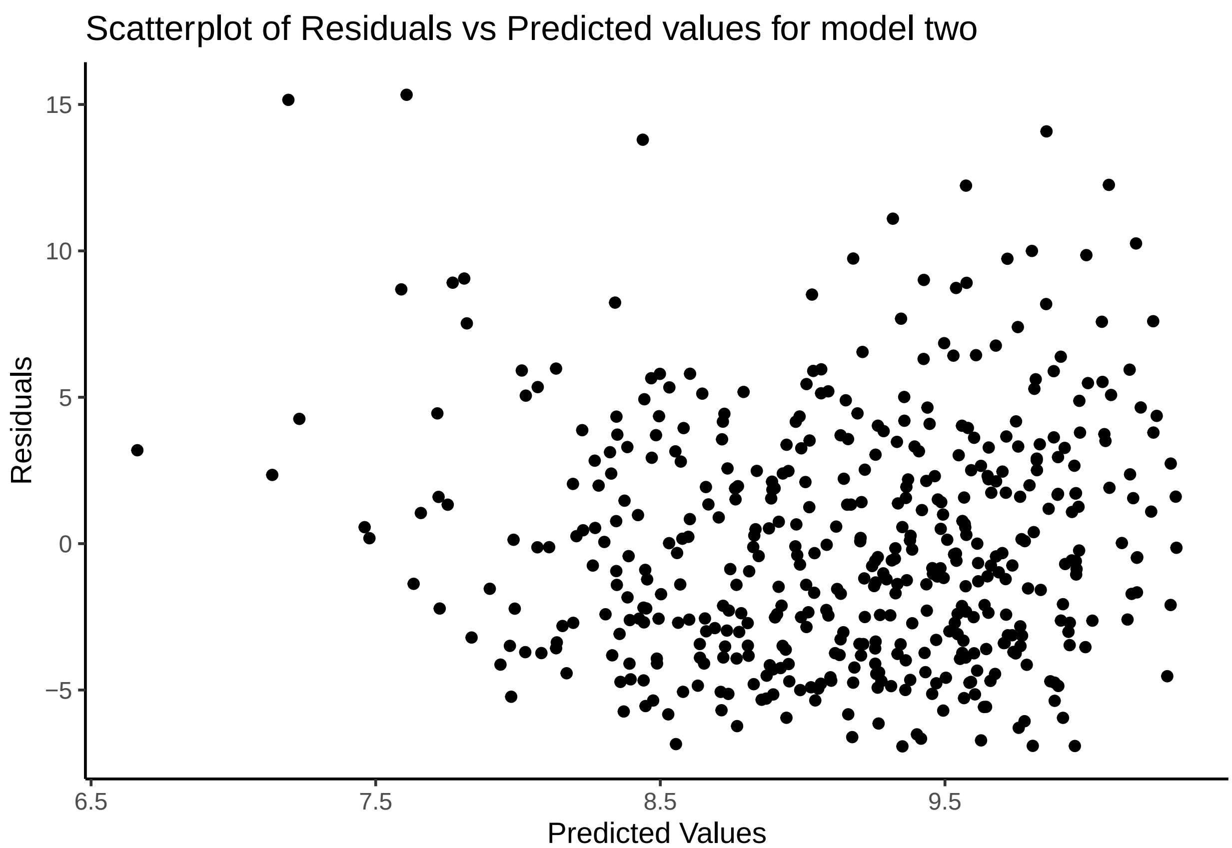



```
## NULL
```

Tests for model two. This model did pass assumptions

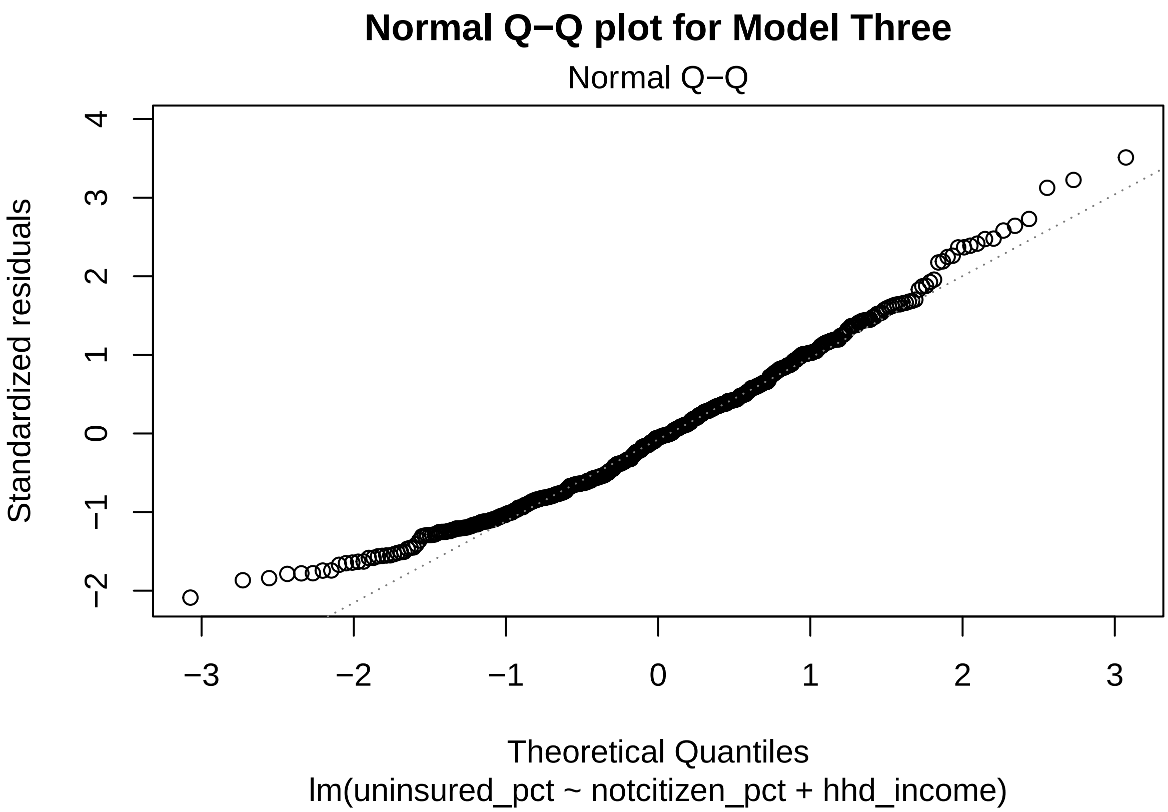

Normal Q−Q plot for Model Two
Normal Q−Q
Standardized residuals
−1
0
1
2
3
4
−3
−2
−1
0
1
2
3
Theoretical Quantiles
lm(uninsured_pct ~ hhd_income)

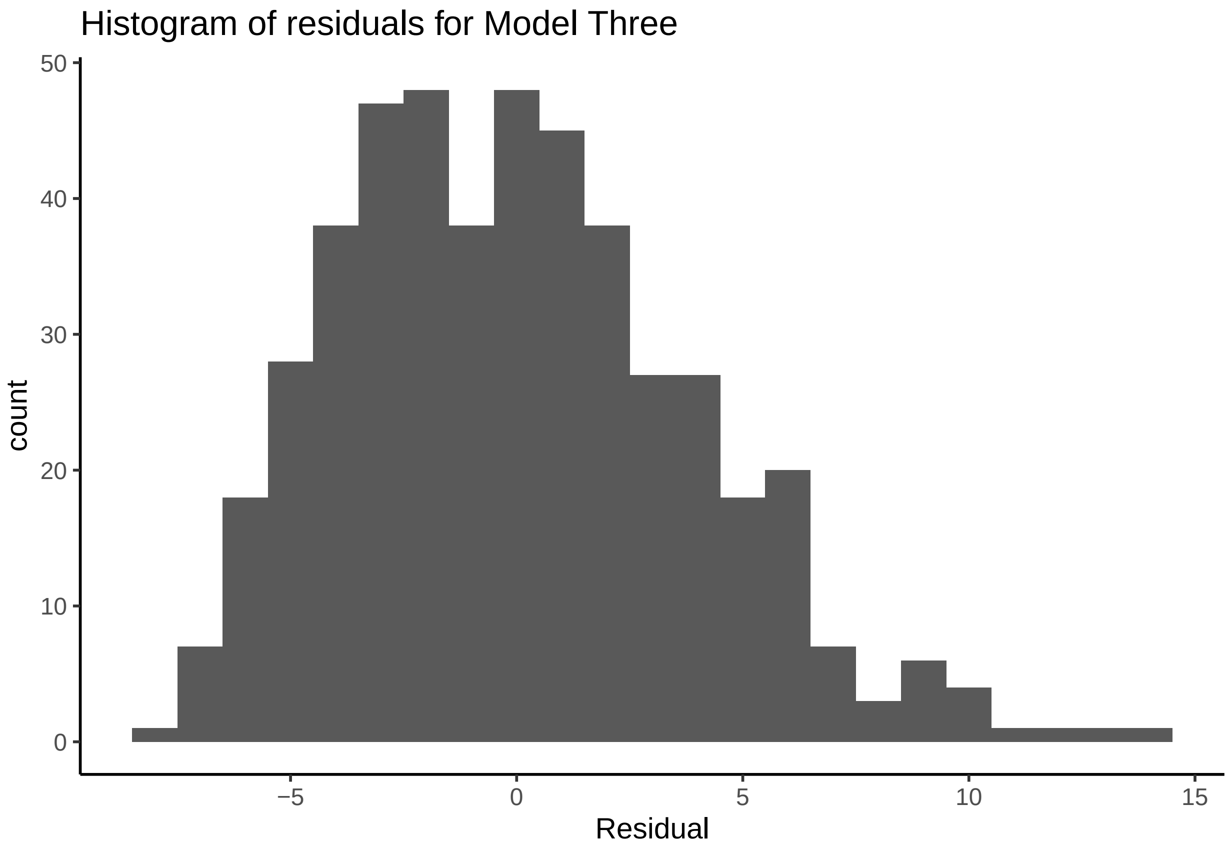

Histogram of residuals for Model Two
count
50
40
30
20
10
0
−5
0
5
10
15
.resid

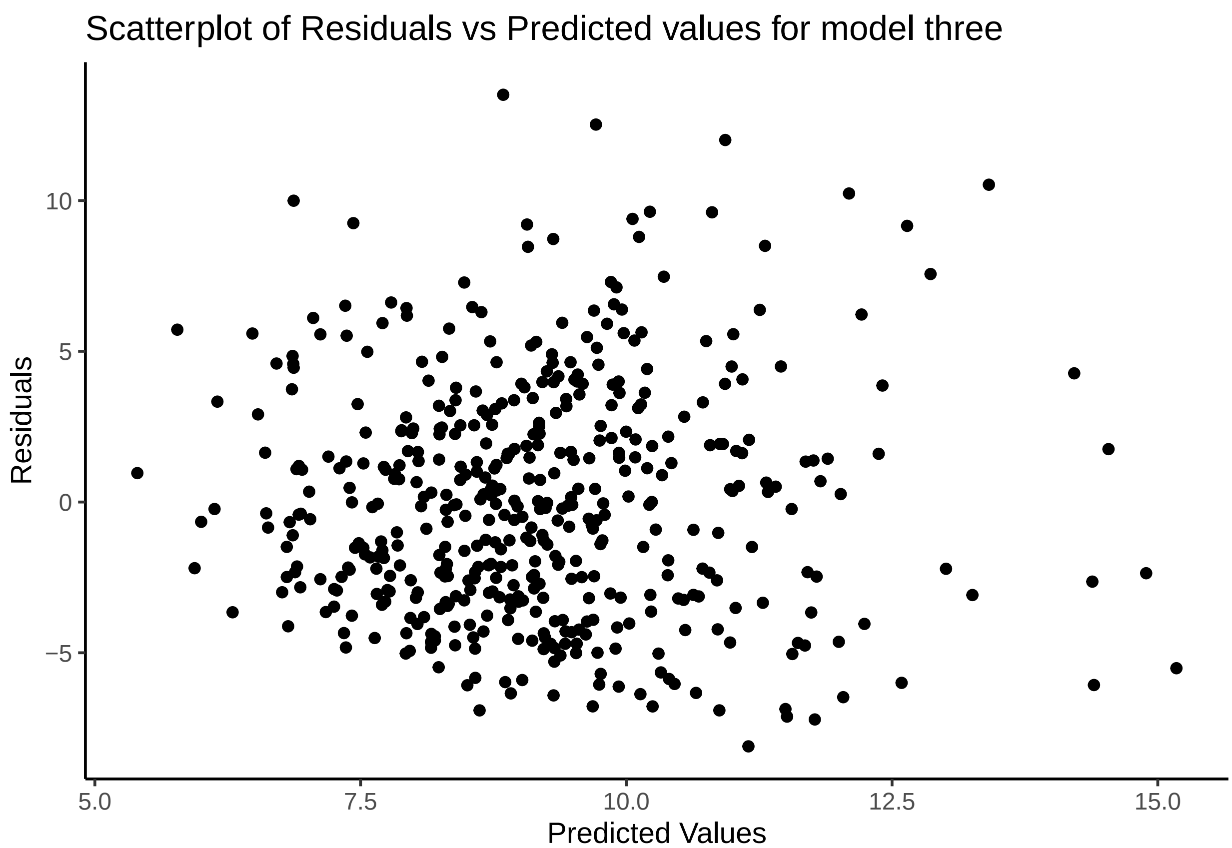



```
## NULL
```

Tests for model three. This model did pass assumptions.

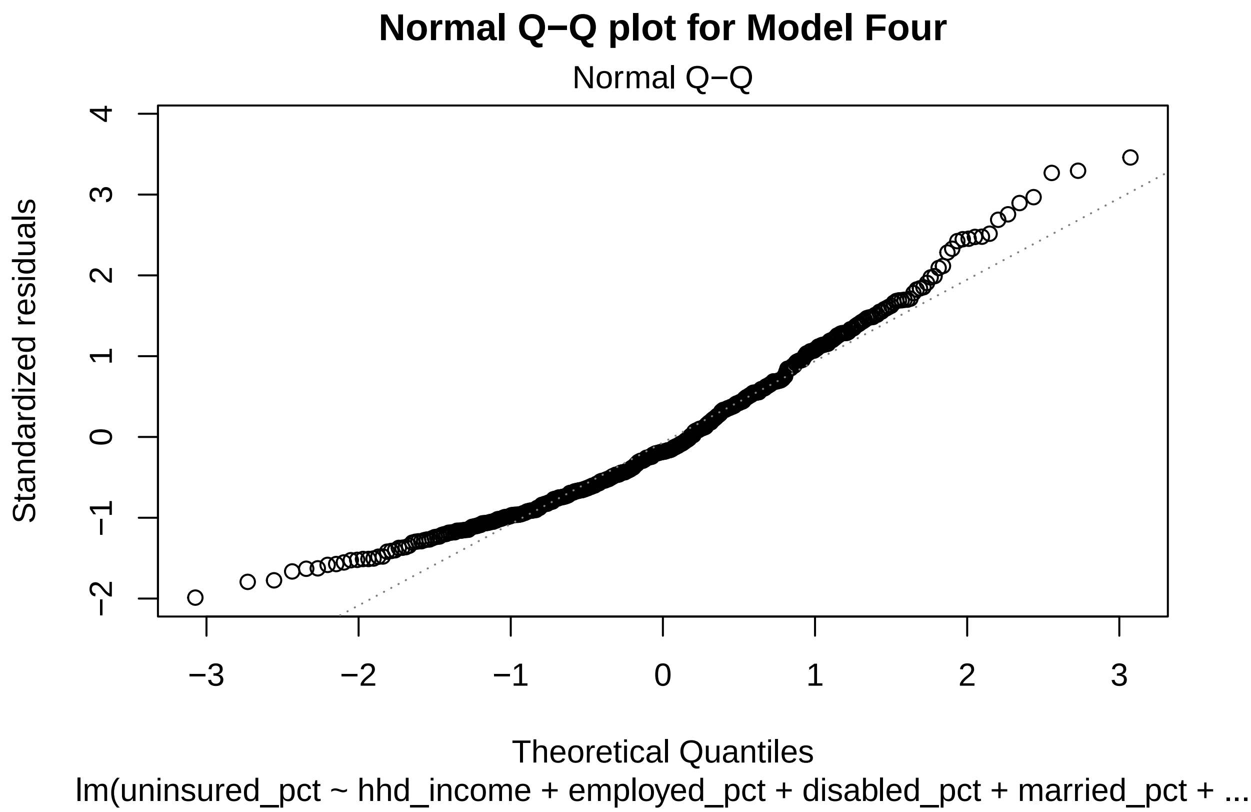

Normal Q−Q plot for Model Three
Normal Q−Q
Standardized residuals
4
3
2
1
0
−1
−2
−3
−2
−1
0
1
2
3
Theoretical Quantiles
lm(uninsured_pct ~ notcitizen_pct + hhd_income)

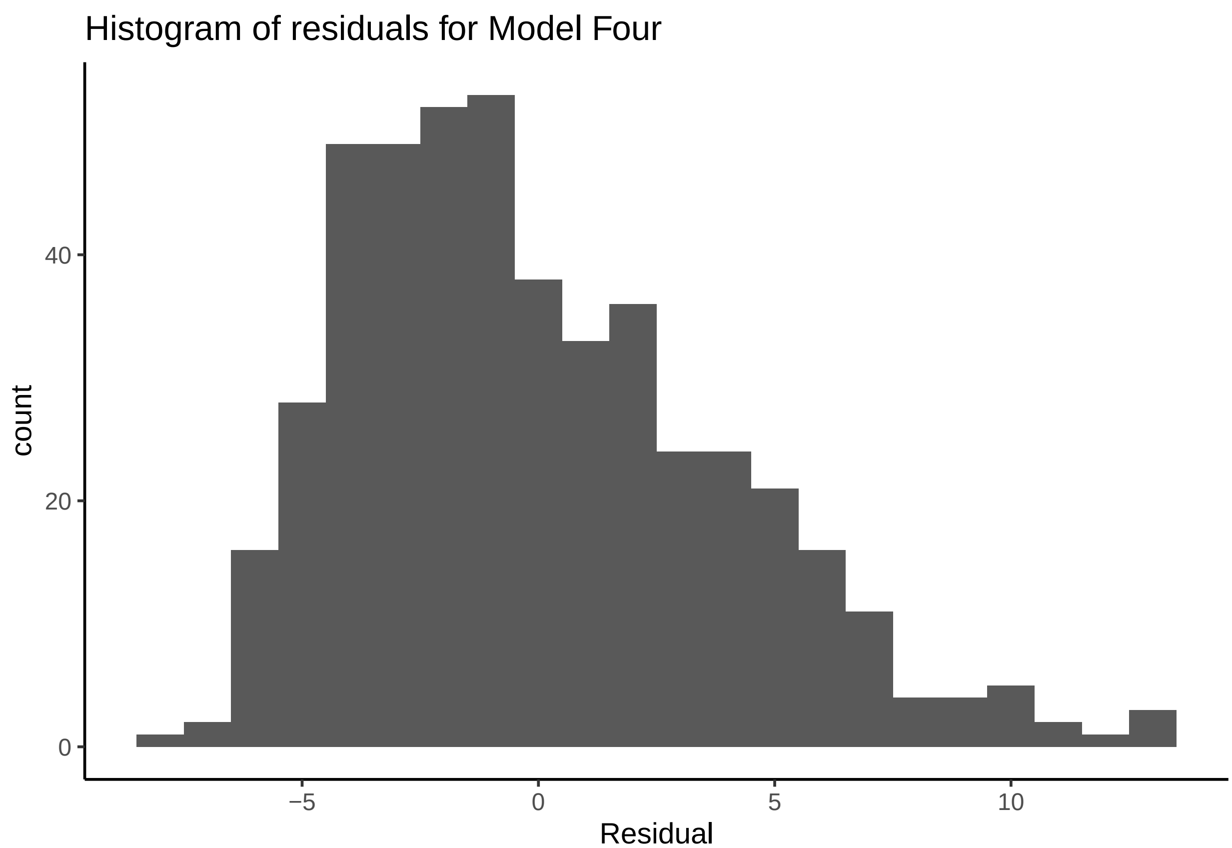

Histogram of residuals for Model Three
count
50
40
30
20
10
0
−5
0
5
10
15
Residual

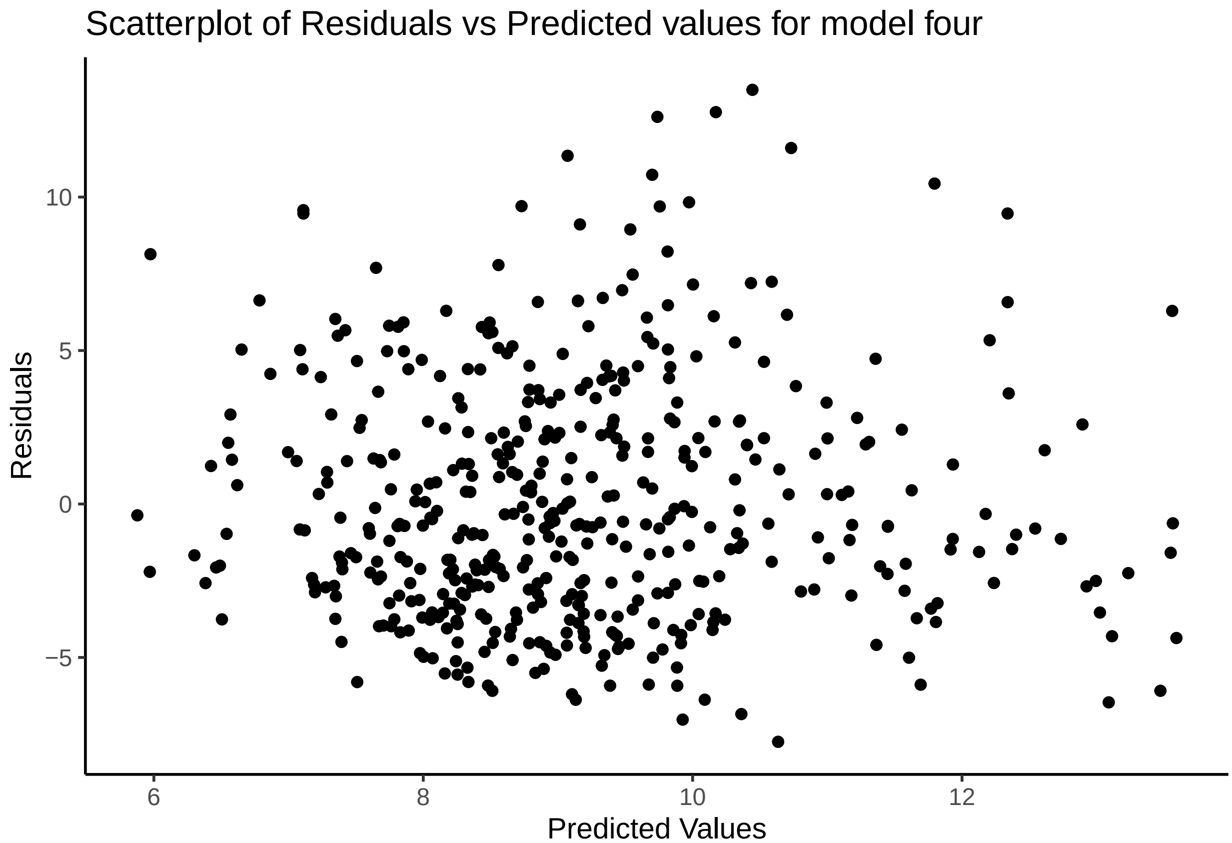



```
## NULL
```

Tests for model four. This model did pass assumptions.

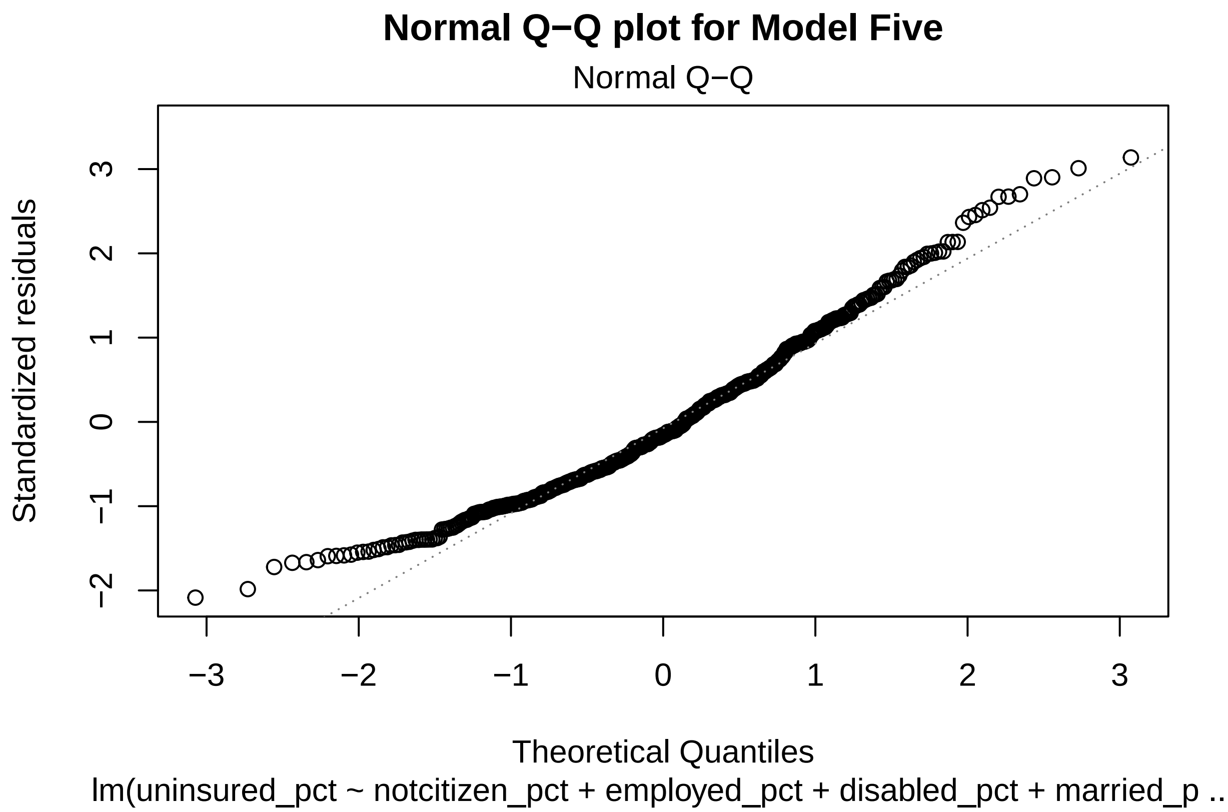

Normal Q−Q plot for Model Four
Normal Q−Q
Standardized residuals
−2
−1
0
1
2
3
4
−3
−2
−1
0
1
2
3
Theoretical Quantiles
lm(uninsured_pct ~ hhd_income + employed_pct + disabled_pct + married_pct + ...

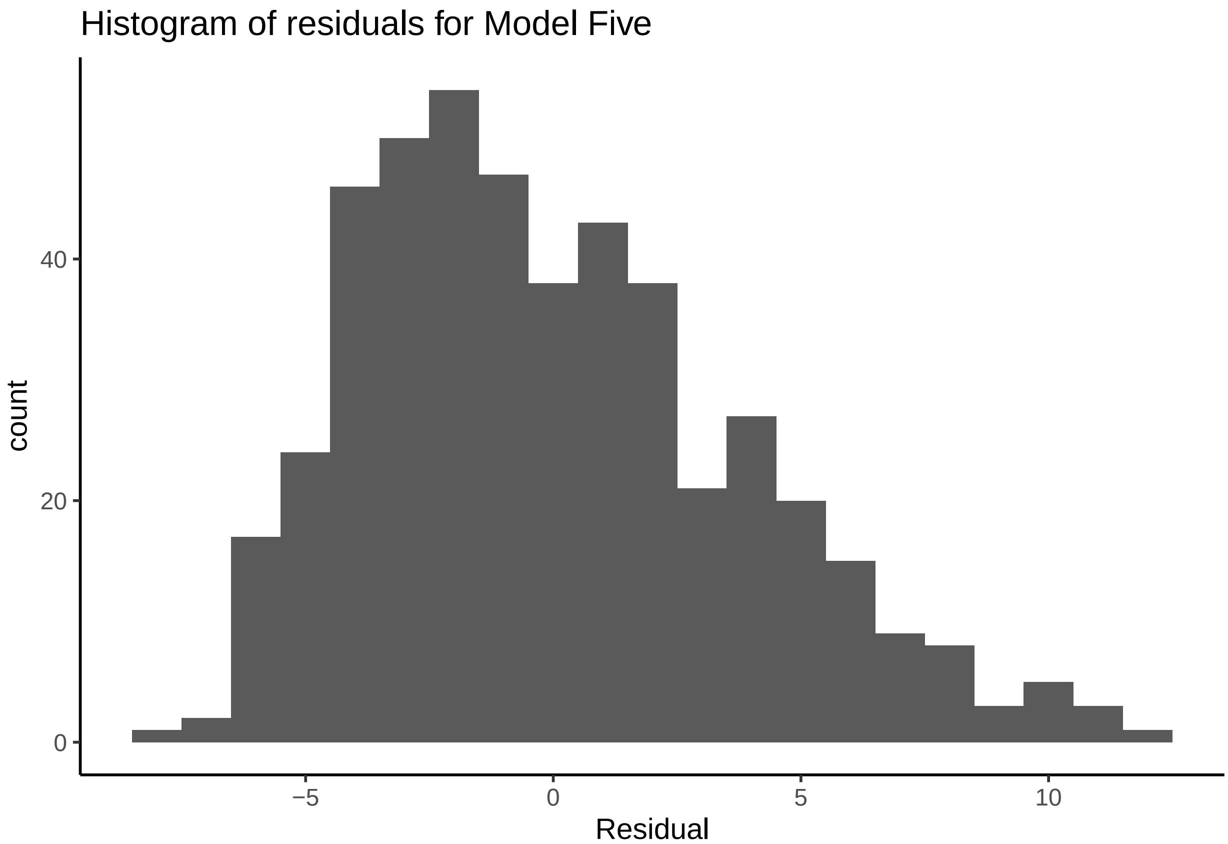

Histogram of residuals for Model Four
count
40
20
0
−5
0
5
10
Residual



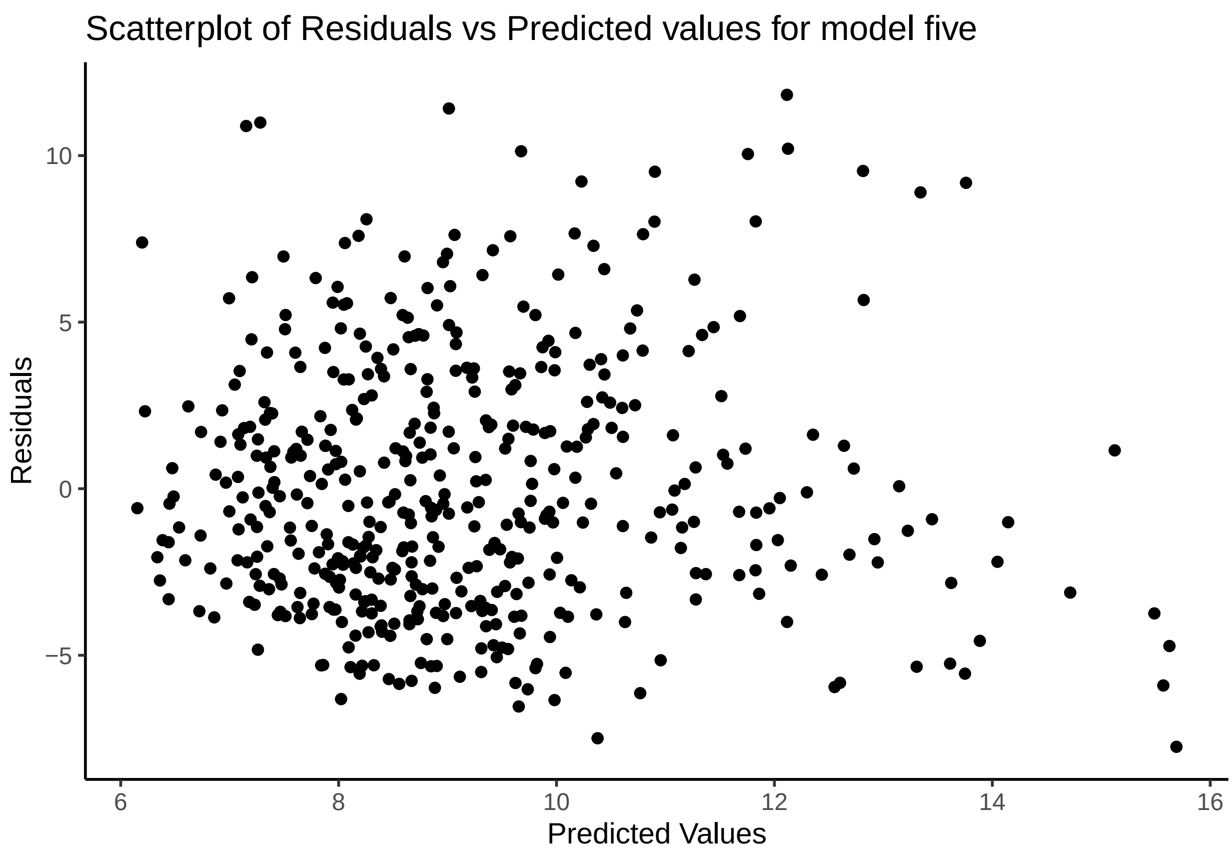

```
## NULL
```

Tests for model five. This model did pass assumptions.


Normal Q−Q plot for Model Five
Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(uninsured_pct ~ notcitizen_pct + employed_pct + disabled_pct + married_p ...
−3
−2
−1
0
1
2
3



Histogram of residuals for Model Five
count
40
20
0
−5
0
5
10
Residual




```
## NULL
```

Bibliography

Card, David, Carlos Dobkin, and Nicole Maestas. 2007. "Does Medicare Save Lives?" Working {Paper} 13668. National Bureau of Economic Research. https://doi.org/10.3386/w13668.

Cohen, Michael S., and William L. Schpero. 2018. "Household Immigration Status Had Differential Impact On Medicaid Enrollment In Expansion And Nonexpansion States." *Health Affairs* 37 (3): 394–402. https://doi.org/10.1377/hlthaff.2017.0978.

Damico, et. al, Anthony. 2021. "Health Coverage by Race and Ethnicity, 2010-2019." *KFF*. https://www.kff.org/racial-equity-and-health-policy/issue-brief/health-coverage-by-race-and-ethnicity/.

Medicare, Center for, and Medicaid Services. n.d. "NHE Fact Sheet CMS." Accessed April 1, 2022. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NHE-Fact-Sheet.

Tolbert, Jennifer, Anthony Damico Published: Nov 06, and 2020. 2020. "Key Facts about the Uninsured Population." *KFF*. https://www.kff.org/uninsured/issue-brief/key-facts-about-the-uninsured-population/.

US Census Bureau. n.d.a. "2020 Data Release Schedule." *Census.gov*. Accessed April 1, 2022. https://www.census.gov/programs-surveys/acs/news/data-releases/2020/release-schedule.html.

———. n.d.b. “American Community Survey (ACS) Questions and Answers.” Accessed April 1, 2022. https://www.bls.gov/lau/acsqa.htm.

———. n.d.c. “Explore Census Data.” Accessed April 1, 2022. https://data.census.gov/cedsci/table?q=United%20States&t=Health%20Insurance&g=0100000US%240500000&tid=ACSST5Y2019.S2701.

———. n.d.d. “Health Insurance Coverage in the United States: 2020.” *Census.gov.* Accessed April 1, 2022. https://www.census.gov/library/publications/2021/demo/p60-274.html.