The 2016 United States (U.S.) Presidential Election:

An Econometric analysis of U.S. Census Bureau data and Federal Election Commission results

Maxim Zibitsker

Professor McKinney

Econometrics Capstone Spring 2018

Table of Contents

Introduction

The United States Presidential election occurs every four years and most recently there has been a lot of discussion regarding the results of the 2016 election in which Donald Trump became the 45th president of the United States. Pertaining to the paper at hand we will primarily assess the relationships between state voter outcome and data retrieved from the United States Census Bureau to better understand which political party each state voted for in 2016. Determining whether factors such as education, income, and race predict voting behavior in the 2016 election will be accomplished by reviewing literature associated with general election theory as well as econometric regression conducted on previous elections.

The general overarching question to be answered is whether factors of race, education, and income affected the results of the 2016 United States Presidential Election. To further our understanding of the voting patterns of citizens across the United States of America (U.S.A.) it will be essential to create a regression function where the dependent variable will be the percentage of Republican votes in each state. This will allow us to determine whether various explanatory variables accounting for state-level demographic data can accurately model the results of the 2016 United States Presidential Election.

Review of Literature

Presidential elections offer economists with data that establishes an understanding of economic indicators, monetary and fiscal policy, and patterns identifying voter behavior. Current literature addresses whether the 2016 Election results were influenced by variables such as race and gender, as well as the influx of fake news published on social network outlets.

The paper, *Voting Patterns in 2016: Exploration using multilevel regression and postratification (MRP) on pre-election polls*, assesses state and demographic data obtained from YouGov pre-election polls. Their proposed model is useful since it uses categorical variables to identify whether they played a critical role in their analysis on the 2016 election results. The authors were able to create a function that measured an electorate by addressing the following six categorical variables:

• State residency • Ethnicity • Gender • Marital status • Age • Education. The variables are represented where "[e]ach variable v has L_V levels. State residency has fifty levels. Ethnicity has four levels: Black, Hispanic, Other, and White. Gender has two levels. Marital status has three levels: Never married, Married, Not married. Age has four levels, corresponding to the left-closed intervals of age: [18, 30), [30, 45), [45, 65), [65, 99). Education has five levels: No High School, High School, Some College, College, Post Graduate" (Trangucci 4). They use the categorical variables to address both Election results graphs and Poststratification graphs, which are essential to understanding the patterns of voting behavior and will be useful to build upon in this paper. Overall, the paper concludes that white voters with lower educational attainment uniformly voted for Trump and that income was not a significant factor in determining whether someone had voted for Trump. Therefore, since the paper at hand will address both income and gender as variables our analysis will not need to test the proven hypothesis, but further build

upon the conclusion that education played a role in voter outcome.

Even though the paper at hand is not necessarily focused on the performance of the economy during a Presidents tenure it is important to understand how the economy is affected by the outcome of the election. The paper, *Presidents and the U.S. Economy: An Econometric Exploration*, tests whether the U.S. economy has performed better when the President of the United States is a Democrat as opposed to a Republican candidate. The authors accomplish this by comparing forecasted GDP growth rates with Presidential attributes, Congress control, and Presidential terms (Blinder, Watson 12-15). This econometric modelling could potentially affect the voter behavior as a variable, which could be added to the model presented in the paper at hand. By focusing on measuring real GDP growth and questioning why the performance gap is both large and significant; Blinder and Watson, determine that "both fiscal and monetary policy actions seem to be a bit more pro-growth when a Republican is president—even though GDP grows significantly faster under Federal Reserve chairmen appointed by Democrats than by Republicans" (Blinder, Watson 36). This provides us with a foundation in testing the hypothesis regarding the argument that President Trump promoting the creation of jobs had statistically benefited his successful voter given the pro-growth under a Republican candidate.

The paper, *The Impact of the supply of fake news on consumer behavior during the 2016 election*, questions which users responded to the fake news that they were exposed to. The authors propose an alternative hypothesis attempting to answer "if the supply of fake news directly affected browsing behavior, then it would be valuable to understand how effective the producers of that fake news were in generating clicks" (Doshi, Raghavan, Weiss, Petit 2). They test their hypothesis empirically by counting visits to fake news websites, the number of articles created in that day, how many social media outlets the individual visited and whether candidates

were mentioned in the articles. In the results, they mention how they were able to "find a link between the production and consumption of fake news during the 2016 US presidential election. Not only did consumers increase their consumption, but they also, increased the probability of visiting a fake news site at all….imply[ing] that the production of fake news would, on average, yield traffic rates similar to those of traditional display advertising" (Doshi, Raghavan, Weiss, Petit 26). As such, it is important to consider this as an important variable in predicting voter outcome in the 2016 Election because fake news generated myriads of clicks. Thus an individual's overall exposure to fake news may act as a critical factor in modelling voter outcome leaving the paper at hand to also discuss and question the accuracy of forecasts conducted for the election.

In the paper, *How surprising was Trump's victory? Notes on predictions in the 2016 U.S. presidential election,* Fred and Alec Wright address whether the prediction models used by various media outlets were wrong in the 2016 United States Presidential Election. The prediction model they created not only outperformed major prediction site estimates, but also assumes that there is a prediction shrinkage effect. They "suggest that an increased emphasis on fundamental statistical tradeoffs of bias and variance, prior to focusing on poll adjustment or demographic behavior, may be the key to improved prediction" (Wright 1). As opposed to pollsters creating the poll aggregations, media outlets had been responsible for the development of poll aggregation, which made the authors question its efficacy. The authors evaluated their hypothesis by "formulating a fit with a smoothing mixed effects model that is sensitive to both national and state-specific trends. State-level standard errors and inter-state correlations were estimated using a parametric bootstrap that resamples outcomes from each polling organization. The results indicate a higher probability for a Trump Victory than was generally recognized"

(Wright 3). The authors were able to create the most accurate predictive model using demographic data and is useful for future election research; however some forecasts that were created had actually been correct.

The authors of the paper, *Assessing the 2016 U.S. Presidential Election Popular Vote Forecasts*, discuss how the PollyVote had been accurately predicted where Clinton would win the election. Overall, the paper tests whether a prediction is accurate or not. The PollyVote model is a model that combines forecasts from the combination of indexes and econometric models, which are based on theory. Their results indicate that "[only] one component, the econometric models, underestimated the Clinton vote. As a result the PollyVote did not perform as well as in previous elections and was only slightly more accurate than the typical forecast" (Armstrong, Graefe, Jones, & Cuzán 15). The paper attests to the fact that in the long run, using combined forecasts will provide researchers with the most accurate forecast.

Methodology

As previously mentioned, the data included in this paper was obtained from the U.S. Census Bureau, as well as from the Federal Election Commission (i.e. government website that publishes state-level election results). The regression equation used is a cross-sectional analysis of state-level data. Even though the regression equation is very similarly represented by using variables that were included in the model presented by Trangucci, limitations exist because data necessary to discuss all the variables in their model would only be possible by accessing YouGov pre-election polls. However, the regression equation does consider the work of Wright, which incorporated demographic data into the model. Since the model that was presented by his work created an accurate predictive model it was necessary to follow the same methodology.

Again, since the data available was limited to state-level results identifying whether an individual's overall exposure to fake news acted as a critical factor in modelling voter outcome, and testing the hypothesis regarding the argument that President Trump promoting the creation of jobs had statistically benefited his success on the state-level was not accomplished.

Also, it was not possible to test the hypothesis regarding the argument that President Trump promoting the creation of jobs had statistically benefited his success on the state-level since the data set would have needed to be restructured into a time-series model accounting for previous election outcomes as well as economic indicators such as GDP growth, accompanied with some form of results from either pre-election or exit polls. Therefore, the various interesting variables and hypotheses tests introduced in the models in the review of literature cannot be accomplished without purchasing data.

With regards to data manipulation, all the categorical variables listed below except for 'Income' (i.e. represented as USD) were originally represented as a proportion of the individuals with these demographics in each state, which made it necessary to multiply the variables by 100 to represent them as a percentage.

The dependent variable (i.e. pct) was computed by assessing the popular vote for Donald Trump, Hillary Clinton, as well as the Other category; then by dividing the number of votes for Donald Trump by the summation of the total vote in each state, and multiplying by 100.

The following categorical variables obtained in the early data collection stage describe factors of Race, Education, as well as Income & Poverty and include:

- White • Black or African American • American • Asian • Native Hawaiian • Hispanic
- High school graduate or higher, percent of persons age 25 years+, 2012-2016
- Bachelor's degree or higher, percent of persons age 25 years+, 2012-2016

- Median household income (in 2016 dollars), 2012-2016

After conducting PROC CORR (see page 29-31), the following variables were omitted from the model:

- White • American • Asian • Native Hawaiian

The variables listed above were omitted from the model because it is more interesting to learn about the relationships between minority races in the U.S. (i.e., Black or African American and Hispanic citizens).

The regression equation used in the model REGRESSION EQUATION

$$\underset{+}{} \quad \underset{-}{} \quad \underset{-}{} \quad \underset{-}{} \quad \underset{-}{}$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon$$

Where X_1 = 'High School', X_2 = 'Bachelor's', X_3 = 'Income', X_4 = 'Black or African American', and X_5 = 'Hispanic'.

Overall from the CORR procedure, we expect a negative relationship between the dependent variable (i.e. pct) and the 'Bachelor's', 'Income', 'Black or African American', and 'Hispanic' variables. With regards to the relationship between the percentage of the popular vote for Donald Trump (i.e. dependent variable) and the other explanatory variables we assess the results of PROC CORR.

The results indicate that the relationship between the percentage of the popular vote for Donald Trump and the 'High School' variable is -0.14323 and has a p-value of (0.3160). The relationship between the percentage of the popular vote for Donald Trump and the 'Bachelor's' variable is -0.82164 and has a p-value of (<0.0001). The relationship between the percentage of the popular vote for Donald Trump and the 'Income' variable is -0.65231 and has a p-value of

(<0.0001). The relationship between the percentage of the popular vote for Donald Trump and the 'Black or African American' variable is -0.19044 and has a p-value 0 of (0.1807). The relationship between the percentage of the popular vote for Donald Trump and the 'Hispanic' variable is -0.31383 and has a p-value of (0.0249). Conclusively, after analyzing each explanatory variable's relationship with the dependent variable and the 'Bachelor's' variable had the strongest correlation in the model since we obtained -0.82164.

The 'High School', 'Bachelor's', 'Income', 'Black or African American', and 'Hispanic' variables all had a negative relationship with the dependent variable.

As indicated on page 27-28 in the appendix with regards to simple statistics we obtained a mean of 88.51569 for the 'High School' variable and a standard deviation of 2.98666, which indicates that 88.52% of persons were a High School graduate or higher in the sample of all 51 states. Additionally, the minimum occurred in the state of California (CA) and was 82.1. The maximum occurred in the state of Montana (MT) and was 92.9.

Regarding the Bachelor's' variable, we obtained a mean of 29.88627 for the 'and a standard deviation of 6.01498, which indicates that 29.89% of persons were a College graduate or higher in the sample of all 51 states. Also, the minimum occurred in the state of West Virginia (WV) and was 19.6. While the maximum was 55.4 and occurred in the District of Columbia (DC).

Additionally, we obtained a mean of $55,730 for the 'Income' variable. This indicates that the average median household income in America is approximately $56,000. The highest income reported was in the state of Maryland (MD) with $76,067. The lowest income reported was in the state of Mississippi (MS) with $40,528.

With regards to the 'Black or African American' variable, we obtained a mean of 11.71373 and a standard deviation of 10.63979; which indicates that 11.71% of persons were either Black

or African American in the sample of all 51 states. The minimum obtained was 0.6 and occurred in (MT). While the maximum was 44.7 in (DC).

In addition, we obtained a mean of 11.73922 for the 'Hispanic' variable and a standard deviation of 10.24350, which indicates that 11.74% of persons were Hispanic in the sample of all 51 states. The minimum occurred in (WV) and was 1.5. The maximum was 48.5 and occurred in the state of New Mexico (NM).

Lastly, we obtained a mean of 48.37742 and a standard deviation of 11.95375, which indicates that the percentage vote for Donald Trump in each state was an average of 48.38%. Additionally, the minimum percentage vote for Donald Trump occurred in (DC) and was 4.08747. The maximum percentage vote for Donald Trump occurred in the state of Wyoming (WY) and was 68.49877.

Pearson Correlation Matrix (see page 29-31)

Multicollinearity occurs when the top values in the Pearson Correlation Matrix are either greater than 0.70 or -0.70, which affect the standard errors and the t-values in the regression.

As such, the relationship between the High School and the 'Bachelor's' variable is 0.40111 and has a p-value of (0.0035). The relationship between the High School and the 'Income' variable is 0.47445 and has a p-value of (0.0004). The relationship between the High School and the 'Black or African American' variable is -0.47461 and has a p-value of (0.0004). The relationship between the High School and the 'Hispanic' variable is -0.43865 and has a p-value of (0.0013). Thus, all the variables had no issues with regards to multicollinearity.

The relationship between the Bachelor's and the 'Income' variable is 0.77780 and has a p-value of (<0.0001). The relationship between the Bachelor's and the 'Black or African American' variable is 0.16278 and has a p-value of (0.2538). The relationship between the Bachelor's and the

'Hispanic' variable is 0.10569 and has a p-value of (0.4604). Thus, the Income variable was the only problem with regards to multicollinearity. The relationship between the Income and the 'Black or African American' variable is -0.04764 and has a p-value of (0.7399). The relationship between the Income and the 'Hispanic' variable is 0.10941 and has a p-value of (0.4447). Thus, all the variables were not a problem with regards to multicollinearity. The relationship between the Black or African American and the 'Hispanic' variable is -0.10702 and has a p-value of (0.4548). Therefore, all the variables were a problem with regards to multicollinearity. Overall, the only issue with regards to multicollinearity occurred between the 'Bachelor's' and the 'Income' variable. We obtained a value of 0.77780, which is relatively severe multicollinearity.

Regarding the Variance Inflation Factor (VIF), (see page 34) we obtained a VIF of 4.63175 for the 'High School' variable. While the VIF for the 'Bachelor's' variable was 3.61198, and the VIF for the 'Income' variable was 2.91351. Lastly we obtained a VIF of 2.76054 for the 'Black or African American', and a VIF of 2.51256 for the 'Hispanic' variable. Therefore, since none of the VIF(s) were greater than 5, then there were no issues with multicollinearity.

PROC REG

As such, when running PROC REG (see page 34), the regression equation in the current model tests the relationships between the percentage vote for Donald Trump (i.e., dependent variable) and the following categorical explanatory variables:

- High School
- Bachelor's
- Income
- Black or African American
- Hispanic

Since the current model has data for every variable for all 51 states in the U.S.A we will be using the following Critical T (50, α = 0.05), which equals 2.009 (Studenmund 519); for each T test below:

T test on the High School Variable:

Ho: $\beta_1 \geq 0$

Ha: $\beta_1 < 0$

T test on the Bachelor's Variable:

Ho: $\beta_2 \geq 0$

Ha: $\beta_2 < 0$

T test on the Income Variable:

Ho: $\beta_3 \geq 0$

Ha: $\beta_3 < 0$

T test on the Black or African American Variable:

Ho: $\beta_4 \geq 0$

Ha: $\beta_4 < 0$

T test on the Hispanic Variable:

Ho: $\beta_5 \geq 0$

Ha: $\beta_5 < 0$

The parameter estimate for 'High School' was 0.49499 and had a t-value of 0.76, with a p-value of 0.4541, which is insignificant. This indicates that for every 1%-point increase in high school graduates the vote for Trump will increase by 0.49499% points. However, the parameter estimate for 'Bachelor's' was -1.55677 and had a t-value of -5.42, with a p-value of (<0.0001), which is significant. This indicates that for every 1%-point increase in bachelors degrees the vote for Trump will fall by -1.55677% points. The 'Income' variable had a parameter estimate of -0.00000111 and had a t-value of -0.65, with a p-value 0.5195, which is insignificant.

This indicates that for every 1 dollar increase in income the vote for Trump will fall by -0.00000111% points. The 'Black or African American' had a parameter estimate of -0.02975 and a t-value of -0.21, with a p-value of 0.8350, which is insignificant. This indicates that for every 1%-point increase in Black or African American citizens in each state the vote for Trump will decrease by 0.02975% points. The 'Hispanic' variable had a parameter estimate of -0.19885 and had a t-value of -1.41, with a p-value of 0.1646, which is insignificant. This indicates that for every 1%-point increase in Hispanic citizens in each state the vote for Trump will decrease by 0.19885% points.

Therefore, we reject the Null Hypothesis for the T test on the Bachelor's variable and fail to reject the Null Hypotheses including the 'High School', 'Income', 'Black or African American', and 'Hispanic' variables.

F test on the model:

Critical F-Value (v_2 = 50,v_1=5, α = 0.05) equals 2.45 (Studenmund 521).

H_O: H_O: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

H_A: H_O is not true.

We obtained an F-value of 25.54 and a p-value of (<0.0001), which is significant. Additionally, the Adjusted R-squared value we obtained was 0.7105. This high Adjusted R-squared value indicates that the model was a good fit for the data set. Since the F-value is greater than our Critical F-Value we reject the Null Hypothesis.

Lastly, the R-squared indicates that 73.94% of the variation in the dependent variable is explained by the model.

Lagrange multiplier test for heteroscedasticity:

Ho: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$

H_A: H_O is not true.

We obtained an R-squared value of 0.7394, and had 50 observations. Thus, the test statistic for the Lagrange multiplier test is equal to 0.7394 *50, which equals 37.47. Our critical value for the chi-square statistic, (df=5, α = 0.05), equals 11.07 (Studenmund 529). Since our test statistic is greater than the critical value we reject the Null Hypothesis and determine that heteroscedasticity is present.

As such, we will not make any changes to the specification criteria of the regression equation and just use the heteroscedasticity consistent Standard Error, t-values, and p-values (see page 38). The hypotheses tests for the heteroscedasticity consistent values are as follows:

T test on the High School Variable:

Ho: $\beta_1 \geq 0$

HA: $\beta_1 < 0$

T test on the Bachelor's Variable:

Ho: $\beta_2 \geq 0$

HA: $\beta_2 < 0$

T test on the Income Variable:

Ho: $\beta_3 \geq 0$

HA: $\beta_3 < 0$

T test on the Black or African American Variable:

Ho: $\beta_4 \geq 0$

HA: $\beta_4 < 0$

T test on the Hispanic Variable:

Ho: $\beta_5 \geq 0$

Ha: $\beta_5 < 0$

Additionally, the new t-Values are computed using the heteroscedasticity consistent values and are the following for each variable:

'High School': $t_k = \frac{0.49499}{0.62520} = 0.79$

'Bachelors': $t_k = \frac{-1.55677}{0.38471} = -4.05$

'Income': $t_k = \frac{-0.00011058}{0.00023230} = -0.48$

'Black or African American': $t_k = \frac{-0.02975}{0.13334} = -0.22$

'Hispanic': $t_k = \frac{-0.19885}{0.15444} = -1.29$

The parameter estimate for 'High School' was 0.49499 and the new t-value we obtain is 0.79, and the p-value is 0.4327, which is insignificant. This parameter estimate indicates that for every 1%-point increase in high school graduates the vote for Trump will increase by 0.49499% points. However, the parameter estimate for 'Bachelor's' was -1.55677 and the new t-value we obtain for the 'Bachelor's' variable is -4.05. The p-value is 0.0002, which is significant. This parameter estimate indicates that for every 1%-point increase in bachelors degrees the vote for Trump will fall by -1.55677% points. The 'Income' variable had a parameter estimate of -0.00000111 and the new t-value we obtain for the 'Income' variable is -0.48. The p-value is 0.6364, which is insignificant. This parameter estimate indicates that for every 1 dollar increase in income the vote for Trump will fall by -0.00000111%-points. The 'Black or African American' had a parameter estimate of -0.02975 and the new t-value we obtain for the 'Black' variable is -0.22. The p-value is 0.8244, which is insignificant. This parameter estimate indicates that for every

1%-point increase in Black or African American citizens in each state the vote for Trump will decrease by 0.02975% points. The 'Hispanic' variable had a parameter estimate of -0.19885 and the new t-value we obtain for the 'Hispanic' variable is -1.29. The p-value is 0.2045, which is insignificant. This parameter indicates that for every 1%-point increase in Hispanic citizens in each state the vote for Trump will decrease by 0.19885% points.

Therefore, we only reject the Null Hypothesis for the t-test on the 'Bachelor's' variable and fail to reject the Null Hypothesis for the t-tests on the 'High School', 'Income', 'Black', and 'Hispanic' explanatory variables.

F test on the model:

With regards to the F-Test, we obtained a Critical F-Value of (v_2= 50,v_1 = 5, α = 0.05), which equals 2.45 (Studenmund 521).

H_O: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$ =0

H_A: H_O is not true.

We obtained an F-value of 25.54 and a p-value of (<0.0001), which is significant (see page 37). Additionally, the Adjusted R-squared value we obtained was 0.7105. This high Adjusted R-squared value indicates that the model was a good fit for the data set. Since the F-value is greater than our Critical F-Value we reject the Null Hypothesis. Lastly, the R-squared indicates that 73.94% of the variation in the dependent variable is explained by the model.

Conclusion

Overall, the results obtained from the CORR, and two REG procedures (i.e. 1st was the regular linear regression equation, and the 2nd was the Heteroscedasticity Consistent linear regression equation) both determined that there was significance between the dependent variable (i.e. pct) and the 'Bachelor's' variable. Additionally, there was relatively severe multicollinearity

between the ‘Bachelor’s’ and the ‘Income’ variable, as well as heteroscedasticity present in the model. With the multicollinearity present in the model OLS estimates are still BLUE, but they are not with heteroscedasticity, which is why it was necessary to use the 2nd model with the heteroscedasticity consistent values. Even though Donald Trump did not win the Popular Vote with 62,984,828 votes (46.09%) against Hillary Clinton’s 65,853,514 votes (48.18%) (see page 27), Trump did win the electoral college vote with 304 votes opposed to Hillary Clinton’s 227 (Federal Elections Committee). Thus, further research would need to assess the county-level results of the 2016 U.S. Presidential Election because it will be able to more accurately describe the election results. Such data can be purchased from *Dave Leip’s Atlas of U.S. Presidential Elections*, and would provide an in-depth glance at the States where voters voted Republican. The data would also enable one to test the hypothesis and discuss whether citizens in large cities voted republican or democratic in 2016.

Bibliography

Armstrong, J. S., Graefe, A., Jones, R. J., & Cuzán, A. G. (2017). Assessing the 2016 U.S. Presidential Election Popular Vote Forecasts, in The 2016 Presidential Election: The causes and consequences of an Electoral Earthquake. Lexington Books, Lanham, MD.

Blinder Alan S., and Mark W. Watson (2015). Presidents and the U.S. Economy: An Econometric Exploration. *Princeton University*, Retrieved from: http://www.princeton.edu/~mwatson/papers/Presidents_AER_2016.pdf.

Doshi, A. R., Petitt, E., Raghavan, S., & Weiss, R. (2018). The Impact of the Supply of Fake News on Consumer Behavior During the 2016 US Election. SSRN Electronic Journal. 10.2139/ssrn.3093397.

Federal Elections Commission (2018). *Federal Elections 2016: Election Results for the U.S. President, the U.S. Senate, and the U.S. House of Representatives.* Retrieved from: https://transition.fec.gov/general/FederalElections2016.shtml.

Trangucci, R. (2018). Voting Patterns in 2016: Exploration Using Multilevel Regression and Poststratification (MRP) on Pre-Election Polls, *University of Michigan,* Retrieved from: arxiv.org/abs/1802.00842.

Studenmund, A. H. Using Econometrics: A Practical Guide (Seventh Edition). Boston, Pearson, 2016.

United States Census Bureau. (2018). *Quick Facts.* Retrieved from https://www.census.gov/quickfacts/fact/table/WY/RHI225216.

Wright, A. and Wright, F. (2017). How surprising was Trump's victory? Notes on predictions in the 2016 U.S. presidential election, *North Carolina State University* Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2900394.

Appendix

Obs	trump	clinton	other	hs	c	y	white	black	am	asian	native	hispanic
1	1318255	729547	75570	0.848	0.240	44758	0.693	0.268	0.007	0.014	0.0010	0.042
2	163387	116454	38767	0.923	0.288	74444	0.661	0.038	0.152	0.063	0.0130	0.070
3	1252401	1161167	15959 7	0.862	0.280	51340	0.833	0.049	0.054	0.034	0.0030	0.309
4	684872	380494	65310	0.852	0.215	42336	0.794	0.157	0.010	0.016	0.0030	0.073
5	4483814	8753792	94399 8	0.821	0.320	63873	0.727	0.065	0.017	0.148	0.0050	0.389
6	1202484	1338870	23889 3	0.910	0.387	62520	0.875	0.045	0.016	0.033	0.0020	0.213
7	673215	897572	74133	0.870	0.303	55322	0.769	0.133	0.013	0.057	0.0020	0.178
8	185127	235603	23084	0.888	0.305	61017	0.701	0.226	0.006	0.040	0.0010	0.092
9	12723	282830	15715	0.900	0.554	72935	0.446	0.447	0.006	0.041	0.0020	0.109
10	4617886	4504975	29717 8	0.872	0.279	48900	0.776	0.168	0.005	0.029	0.0010	0.249
11	2089104	1877963	14766 5	0.858	0.294	51037	0.612	0.320	0.005	0.041	0.0010	0.094
12	128847	266891	33199	0.914	0.314	71977	0.258	0.022	0.004	0.377	0.1020	0.104
13	409055	189765	91435	0.900	0.262	49174	0.933	0.008	0.018	0.015	0.0020	0.123
14	2146015	3090729	29968 0	0.883	0.329	59196	0.772	0.147	0.006	0.055	0.0010	0.170
15	1557286	1033126	14454 6	0.881	0.246	50433	0.856	0.097	0.004	0.022	0.0010	0.068

Obs	trump	clinton	other	hs	c	y	white	black	am	asian	native	hispanic
16	800983	653669	111379	0.917	0.272	54570	0.914	0.037	0.005	0.025	0.0010	0.058
17	671018	427005	86379	0.903	0.316	53571	0.866	0.062	0.012	0.030	0.0010	0.116
18	120297 1	628854	92324	0.846	0.227	44811	0.880	0.083	0.003	0.015	0.0010	0.035
19	117863 8	780154	70240	0.838	0.230	46652	0.632	0.326	0.008	0.018	0.0010	0.050
20	335593	357735	54599	0.919	0.293	50826	0.948	0.015	0.007	0.012	0.0005	0.016
21	943169	167792 8	160349	0.896	0.384	76067	0.593	0.307	0.006	0.066	0.0010	0.098
22	109089 3	199519 6	238957	0.901	0.412	70954	0.818	0.086	0.005	0.067	0.0010	0.115
23	227954 3	226883 9	250902	0.899	0.274	50803	0.796	0.142	0.007	0.031	0.0005	0.050
24	132295 1	136771 6	254146	0.926	0.342	63217	0.850	0.062	0.013	0.049	0.0010	0.052
25	700714	485131	23512	0.830	0.210	40528	0.593	0.377	0.006	0.011	0.0010	0.031
26	159451 1	107106 8	143026	0.888	0.276	49593	0.832	0.118	0.006	0.020	0.0010	0.041
27	279240	177709	40198	0.929	0.299	48380	0.892	0.006	0.066	0.008	0.0010	0.036
28	495961	284494	63772	0.907	0.300	54384	0.889	0.050	0.014	0.025	0.0010	0.107
29	512058	539260	74067	0.854	0.232	53095	0.751	0.096	0.016	0.087	0.0080	0.285
30	345790	348526	49980	0.926	0.355	68485	0.938	0.015	0.003	0.027	0.0010	0.035
31	160193 3	214827 8	123835	0.889	0.375	73702	0.724	0.150	0.006	0.098	0.0010	0.200

Obs	trump	clinton	other	hs	c	y	white	black	am	asian	native	hispanic
32	319667	385234	93418	0.846	0.267	45674	0.825	0.025	0.106	0.017	0.0020	0.485
33	2819533	4556118	345791	0.859	0.347	60741	0.699	0.177	0.010	0.089	0.0010	0.190
34	2362631	2189316	189617	0.863	0.290	48256	0.710	0.222	0.016	0.029	0.0010	0.092
35	216794	93758	33808	0.920	0.282	59114	0.879	0.029	0.055	0.015	0.0010	0.036
36	2841005	2394164	261318	0.895	0.267	50674	0.825	0.128	0.003	0.022	0.0010	0.037
37	949136	420375	83481	0.873	0.245	48038	0.745	0.078	0.092	0.022	0.0020	0.103
38	782403	1002106	216827	0.900	0.314	53270	0.874	0.021	0.018	0.045	0.0000	0.128
39	2970733	2926441	268304	0.895	0.293	54895	0.824	0.118	0.004	0.035	0.0010	0.070
40	180543	252525	31076	0.870	0.325	58387	0.844	0.081	0.010	0.036	0.0020	0.149
41	1155389	855373	92265	0.860	0.265	46898	0.685	0.275	0.005	0.016	0.0010	0.055
42	227721	117458	24914	0.912	0.275	52078	0.852	0.020	0.090	0.015	0.0010	0.037
43	1522925	870695	114407	0.860	0.254	46574	0.787	0.171	0.400	0.018	0.0010	0.052
44	4685047	3877868	406311	0.823	0.281	54727	0.794	0.126	0.010	0.048	0.0010	0.391
45	515231	310676	305523	0.915	0.317	62518	0.911	0.014	0.016	0.025	0.0100	0.138
46	95369	178573	41125	0.919	0.362	56104	0.946	0.013	0.004	0.018	0.0005	0.019
47	1769443	1981473	233715	0.886	0.369	66149	0.700	0.198	0.005	0.066	0.0010	0.091

Obs	trump	clinton	other	hs	c	y	white	black	am	asian	native	hispanic
48	1221747	1742718	352554	0.906	0.336	62848	0.800	0.041	0.019	0.086	0.0080	0.124
49	489371	188794	36258	0.853	0.196	42644	0.936	0.036	0.002	0.008	0.0005	0.015
50	1405284	1382536	188330	0.914	0.284	54610	0.875	0.066	0.011	0.028	0.0010	0.067
51	174419	55973	25457	0.924	0.260	59143	0.928	0.013	0.027	0.010	0.0010	0.100

Obs	high_school_	bachelors_	income_	black_	hispanic_	pct
1	84.8	24.0	44758	26.8	4.2	62.0831
2	92.3	28.8	74444	3.8	7.0	51.2815
3	86.2	28.0	51340	4.9	30.9	48.6716
4	85.2	21.5	42336	15.7	7.3	60.5719
5	82.1	32.0	63873	6.5	38.9	31.6171
6	91.0	38.7	62520	4.5	21.3	43.2510
7	87.0	30.3	55322	13.3	17.8	40.9269
8	88.8	30.5	61017	22.6	9.2	41.7127
9	90.0	55.4	72935	44.7	10.9	4.0875
10	87.2	27.9	48900	16.8	24.9	49.0219

Obs	high_school_	bachelors_	income_	black_	hispanic_	pct
11	85.8	29.4	51037	32.0	9.4	50.7713
12	91.4	31.4	71977	2.2	10.4	30.0387
13	90.0	26.2	49174	0.8	12.3	59.2614
14	88.3	32.9	59196	14.7	17.0	38.7618
15	88.1	24.6	50433	9.7	6.8	56.9400
16	91.7	27.2	54570	3.7	5.8	51.1473
17	90.3	31.6	53571	6.2	11.6	56.6546
18	84.6	22.7	44811	8.3	3.5	62.5196
19	83.8	23.0	46652	32.6	5.0	58.0887
20	91.9	29.3	50826	1.5	1.6	44.8698
21	89.6	38.4	76067	30.7	9.8	33.9093
22	90.1	41.2	70954	8.6	11.5	32.8084
23	89.9	27.4	50803	14.2	5.0	47.4976
24	92.6	34.2	63217	6.2	5.2	44.9248
25	83.0	21.0	40528	37.7	3.1	57.9410
26	88.8	27.6	49593	11.8	4.1	56.7723

Obs	high_school_	bachelors_	income_	black_	hispanic_	pct
27	92.9	29.9	48380	0.6	3.6	56.1685
28	90.7	30.0	54384	5.0	10.7	58.7474
29	85.4	23.2	53095	9.6	28.5	45.5007
30	92.6	35.5	68485	1.5	3.5	46.4587
31	88.9	37.5	73702	15.0	20.0	41.3504
32	84.6	26.7	45674	2.5	48.5	40.0425
33	85.9	34.7	60741	17.7	19.0	36.5156
34	86.3	29.0	48256	22.2	9.2	49.8281
35	92.0	28.2	59114	2.9	3.6	62.9556
36	89.5	26.7	50674	12.8	3.7	51.6877
37	87.3	24.5	48038	7.8	10.3	65.3229
38	90.0	31.4	53270	2.1	12.8	39.0940
39	89.5	29.3	54895	11.8	7.0	48.1833
40	87.0	32.5	58387	8.1	14.9	38.8981
41	86.0	26.5	46898	27.5	5.5	54.9393
42	91.2	27.5	52078	2.0	3.7	61.5308

Obs	high_school_	bachelors_	income_	black_	hispanic_	pct
43	86.0	25.4	46574	17.1	5.2	60.7220
44	82.3	28.1	54727	12.6	39.1	52.2347
45	91.5	31.7	62518	1.4	13.8	45.5380
46	91.9	36.2	56104	1.3	1.9	30.2694
47	88.6	36.9	66149	19.8	9.1	44.4067
48	90.6	33.6	62848	4.1	12.4	36.8327
49	85.3	19.6	42644	3.6	1.5	68.4988
50	91.4	28.4	54610	6.6	6.7	47.2182
51	92.4	26.0	59143	1.3	10.0	68.1726

18 Variables:	trump clinton other hs c y white black am asian native hispanic high_school_ bachelors_ income_ black_ hispanic_ pct

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
trump	51	1234997	1142383	62984828	12723	4685047
clinton	51	1291245	1548085	65853514	55973	8753792
other	51	153548	153315	7830934	15715	943998
hs	51	0.88516	0.02987	45.14300	0.82100	0.92900
c	51	0.29886	0.06015	15.24200	0.19600	0.55400
y	51	55730	9115	2842242	40528	76067
white	51	0.78551	0.13143	40.06100	0.25800	0.94800
black	51	0.11714	0.10640	5.97400	0.00600	0.44700
am	51	0.02763	0.06119	1.40900	0.00200	0.40000
asian	51	0.04357	0.05509	2.22200	0.00800	0.37700
native	51	0.00390	0.01423	0.19900	0	0.10200
hispanic	51	0.11739	0.10243	5.98700	0.01500	0.48500
high_school_	51	88.51569	2.98666	4514	82.10000	92.90000
bachelors_	51	29.88627	6.01498	1524	19.60000	55.40000
income_	51	55730	9115	2842242	40528	76067
black_	51	11.71373	10.63979	597.40000	0.60000	44.70000

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
hispanic_	51	11.73922	10.24350	598.70000	1.50000	48.50000
pct	51	48.37742	11.95375	2467	4.08747	68.49877

Pearson Correlation Coefficients, N = 51 Prob > |r| under H0: Rho=0										
	trump	clinton	other	hs	c	y	white	black	am	asian
trump	1.00000	0.88928 <.0001	0.77668 <.0001	-0.45717 0.0007	-0.05324 0.7106	-0.08823 0.5381	-0.07218 0.6147	0.20151 0.1562	-0.10197 0.4765	0.07834 0.5848
clinton	0.88928 <.0001	1.00000	0.91514 <.0001	-0.39408 0.0042	0.14802 0.2999	0.15009 0.2931	-0.12998 0.3633	0.13186 0.3564	-0.14635 0.3055	0.25694 0.0687
other	0.77668 <.0001	0.91514 <.0001	1.00000	-0.27475 0.0510	0.15882 0.2656	0.18312 0.1984	0.01484 0.9177	-0.05811 0.6854	-0.11995 0.4018	0.23783 0.0928
hs	-0.45717 0.0007	-0.39408 0.0042	-0.27475 0.0510	1.00000	0.40111 0.0035	0.47445 0.0004	0.27577 0.0501	-0.47461 0.0004	-0.03465 0.8093	0.03347 0.8156
c	-0.05324 0.7106	0.14802 0.2999	0.15882 0.2656	0.40111 0.0035	1.00000	0.77780 <.0001	-0.24169 0.0875	0.16278 0.2538	-0.16527 0.2465	0.24199 0.0871
y	-0.08823 0.5381	0.15009 0.2931	0.18312 0.1984	0.47445 0.0004	0.77780 <.0001	1.00000	-0.31274 0.0255	-0.04764 0.7399	-0.11344 0.4280	0.51915 <.0001
white	-0.07218 0.6147	-0.12998 0.3633	0.01484 0.9177	0.27577 0.0501	-0.24169 0.0875	-0.31274 0.0255	1.00000	-0.64849 <.0001	0.02291 0.8732	-0.64046 <.0001
black	0.20151 0.1562	0.13186 0.3564	-0.05811 0.6854	-0.47461 0.0004	0.16278 0.2538	-0.04764 0.7399	-0.64849 <.0001	1.00000	-0.09928 0.4882	-0.08062 0.5738

Pearson Correlation Coefficients, N = 51 Prob > |r| under H0: Rho=0										
	trump	clinton	other	hs	c	y	white	black	am	asian
am	-0.10197 0.4765	-0.14635 0.3055	-0.11995 0.4018	-0.03465 0.8093	-0.16527 0.2465	-0.11344 0.4280	0.02291 0.8732	-0.09928 0.4882	1.00000	-0.10626 0.4580
asian	0.07834 0.5848	0.25694 0.0687	0.23783 0.0928	0.03347 0.8156	0.24199 0.0871	0.51915 <.0001	-0.64046 <.0001	-0.08062 0.5738	-0.10626 0.4580	1.00000
native	-0.15728 0.2704	-0.09496 0.5074	-0.07997 0.5769	0.14717 0.3027	0.03513 0.8066	0.30382 0.0302	-0.58058 <.0001	-0.16124 0.2583	-0.02395 0.8675	0.88395 <.0001
hispanic	0.39827 0.0038	0.48972 0.0003	0.49444 0.0002	-0.43865 0.0013	0.10569 0.4604	0.10941 0.4447	-0.07048 0.6231	-0.10702 0.4548	0.01302 0.9277	0.23246 0.1007
high_school_	-0.45717 0.0007	-0.39408 0.0042	-0.27475 0.0510	1.00000 <.0001	0.40111 0.0035	0.47445 0.0004	0.27577 0.0501	-0.47461 0.0004	-0.03465 0.8093	0.03347 0.8156
bachelors_	-0.05324 0.7106	0.14802 0.2999	0.15882 0.2656	0.40111 0.0035	1.00000 <.0001	0.77780 <.0001	-0.24169 0.0875	0.16278 0.2538	-0.16527 0.2465	0.24199 0.0871
income_	-0.08823 0.5381	0.15009 0.2931	0.18312 0.1984	0.47445 0.0004	0.77780 <.0001	1.00000 <.0001	-0.31274 0.0255	-0.04764 0.7399	-0.11344 0.4280	0.51915 <.0001
black_	0.20151 0.1562	0.13186 0.3564	-0.05811 0.6854	-0.47461 0.0004	0.16278 0.2538	-0.04764 0.7399	-0.64849 <.0001	1.00000 <.0001	-0.09928 0.4882	-0.08062 0.5738

Pearson Correlation Coefficients, N = 51 Prob > |r| under H0: Rho=0										
	trump	clinton	other	hs	c	y	white	black	am	asian
hispanic_	0.39827 0.0038	0.48972 0.0003	0.49444 0.0002	-0.43865 0.0013	0.10569 0.4604	0.10911 0.4447	-0.07048 0.6231	-0.10702 0.4548	0.01302 0.9277	0.23246 0.1007
pct	-0.03665 0.7985	-0.28322 0.0440	-0.28529 0.0424	-0.14323 0.3160	-0.82164 <.0001	-0.65231 <.0001	0.41110 0.0027	-0.19044 0.1807	0.22456 0.1131	-0.46622 0.0006

Pearson Correlation Coefficients, N = 51 Prob > |r| under H0: Rho=0								
	native	hispanic	high_school_	bachelors_	income_	black_	hispanic_	pct
trump	-0.15728 0.2704	0.39827 0.0038	-0.45717 0.0007	-0.05324 0.7106	-0.08823 0.5381	0.20151 0.1562	0.39827 0.0038	-0.03665 0.7985
clinton	-0.09496 0.5074	0.48972 0.0003	-0.39408 0.0042	0.14802 0.2999	0.15009 0.2931	0.13186 0.3564	0.48972 0.0003	-0.28322 0.0440
other	-0.07997 0.5769	0.49444 0.0002	-0.27475 0.0510	0.15882 0.2656	0.18312 0.1984	-0.05811 0.6854	0.49444 0.0002	-0.28529 0.0424
hs	0.14717 0.3027	-0.43865 0.0013	1.00000 <.0001	0.40111 0.0035	0.47445 0.0004	-0.47461 0.0004	-0.43865 0.0013	-0.14323 0.3160
c	0.03513 0.8066	0.10569 0.4604	0.40111 0.0035	1.00000 <.0001	0.77780 <.0001	0.16278 0.2538	0.10569 0.4604	-0.82164 <.0001

Pearson Correlation Coefficients, N = 51 Prob > |r| under H0: Rho=0								
	native	hispanic	high_school_	bachelors_	income_	black_	hispanic_	pct
y	0.30382 0.0302	0.10941 0.4447	0.47445 0.0004	0.77780 <.0001	1.00000 <.0001	-0.04764 0.7399	0.10941 0.4447	-0.65231 <.0001
white	-0.58058 <.0001	-0.07048 0.6231	0.27577 0.0501	-0.24169 0.0875	-0.31274 0.0255	-0.64849 <.0001	-0.07048 0.6231	0.41110 0.0027
black	-0.16124 0.2583	-0.10702 0.4548	-0.47461 0.0004	0.16278 0.2538	-0.04764 0.7399	1.00000 <.0001	-0.10702 0.4548	-0.19044 0.1807
am	-0.02395 0.8675	0.01302 0.9277	-0.03465 0.8093	-0.16527 0.2465	-0.11344 0.4280	-0.09928 0.4882	0.01302 0.9277	0.22456 0.1131
asian	0.88395 <.0001	0.23246 0.1007	0.03347 0.8156	0.24199 0.0871	0.51915 <.0001	-0.08062 0.5738	0.23246 0.1007	-0.46622 0.0006
native	1.00000	0.02163 0.8802	0.14717 0.3027	0.03513 0.8066	0.30382 0.0302	-0.16124 0.2583	0.02163 0.8802	-0.23774 0.0930
hispanic	0.02163 0.8802	1.00000	-0.43865 0.0013	0.10569 0.4604	0.10941 0.4447	-0.10702 0.4548	1.00000 <.0001	-0.31383 0.0249
high_school_	0.14717 0.3027	-0.43865 0.0013	1.00000	0.40111 0.0035	0.47445 0.0004	-0.47461 0.0004	-0.43865 0.0013	-0.14323 0.3160

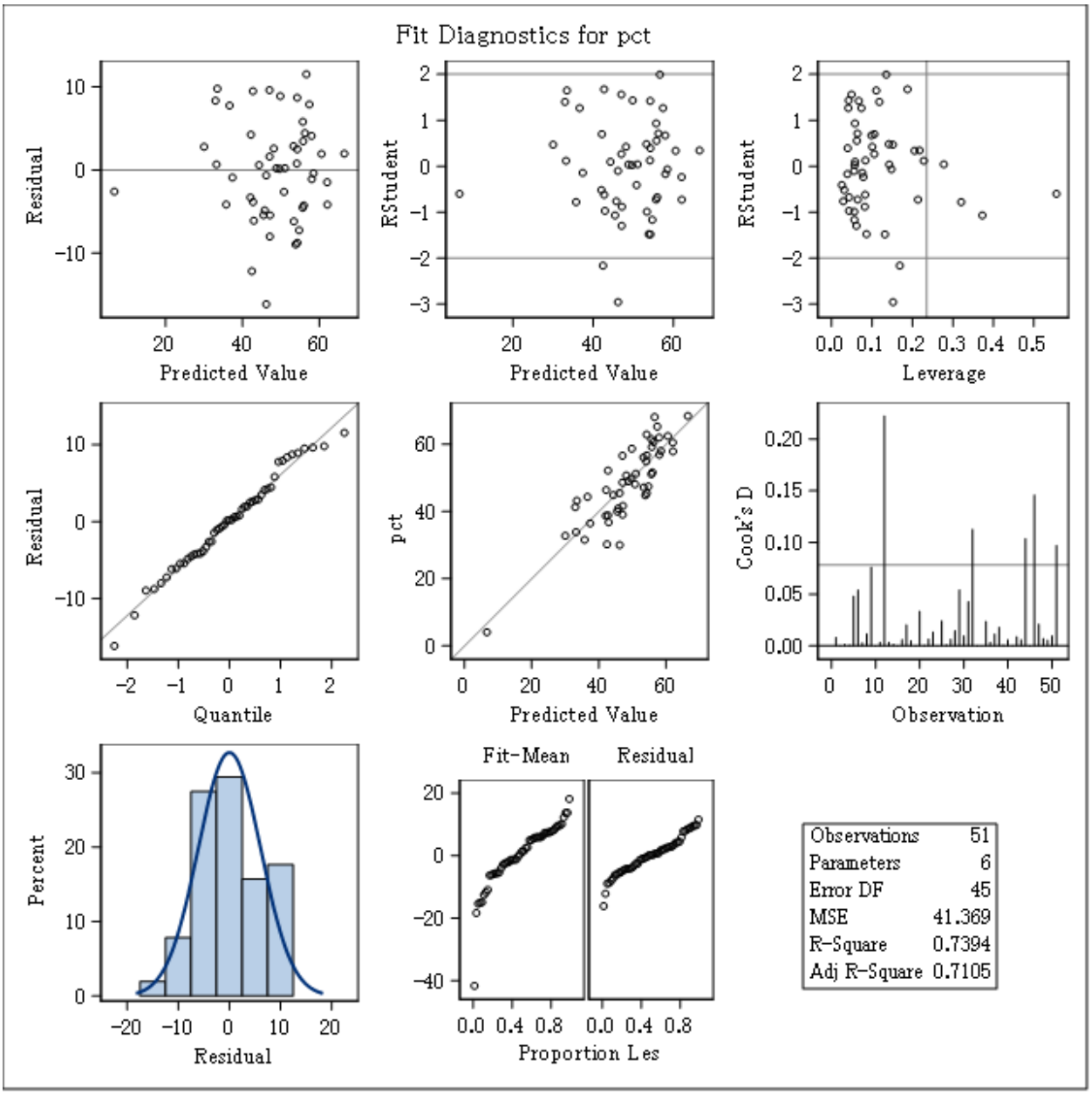
Pearson Correlation Coefficients, N = 51 Prob > |r| under H0: Rho=0								
	native	hispanic	high_school_	bachelors_	income_	black_	hispanic_	pct
bachelors_	0.03513 0.8066	0.10569 0.4604	0.40111 0.0035	1.00000	0.77780 <.0001	0.16278 0.2538	0.10569 0.4604	-0.82164 <.0001
income_	0.30382 0.0302	0.10941 0.4447	0.47445 0.0004	0.77780 <.0001	1.00000	-0.04764 0.7399	0.10941 0.4447	-0.65231 <.0001
black_	-0.16124 0.2583	-0.10702 0.4548	-0.47461 0.0004	0.16278 0.2538	-0.04764 0.7399	1.00000	-0.10702 0.4548	-0.19044 0.1807
hispanic_	0.02163 0.8802	1.00000 <.0001	-0.43865 0.0013	0.10569 0.4604	0.10941 0.4447	-0.10702 0.4548	1.00000	-0.31383 0.0249
pct	-0.23774 0.0930	-0.31383 0.0249	-0.14323 0.3160	-0.82164 <.0001	-0.65231 <.0001	-0.19044 0.1807	-0.31383 0.0249	1.00000

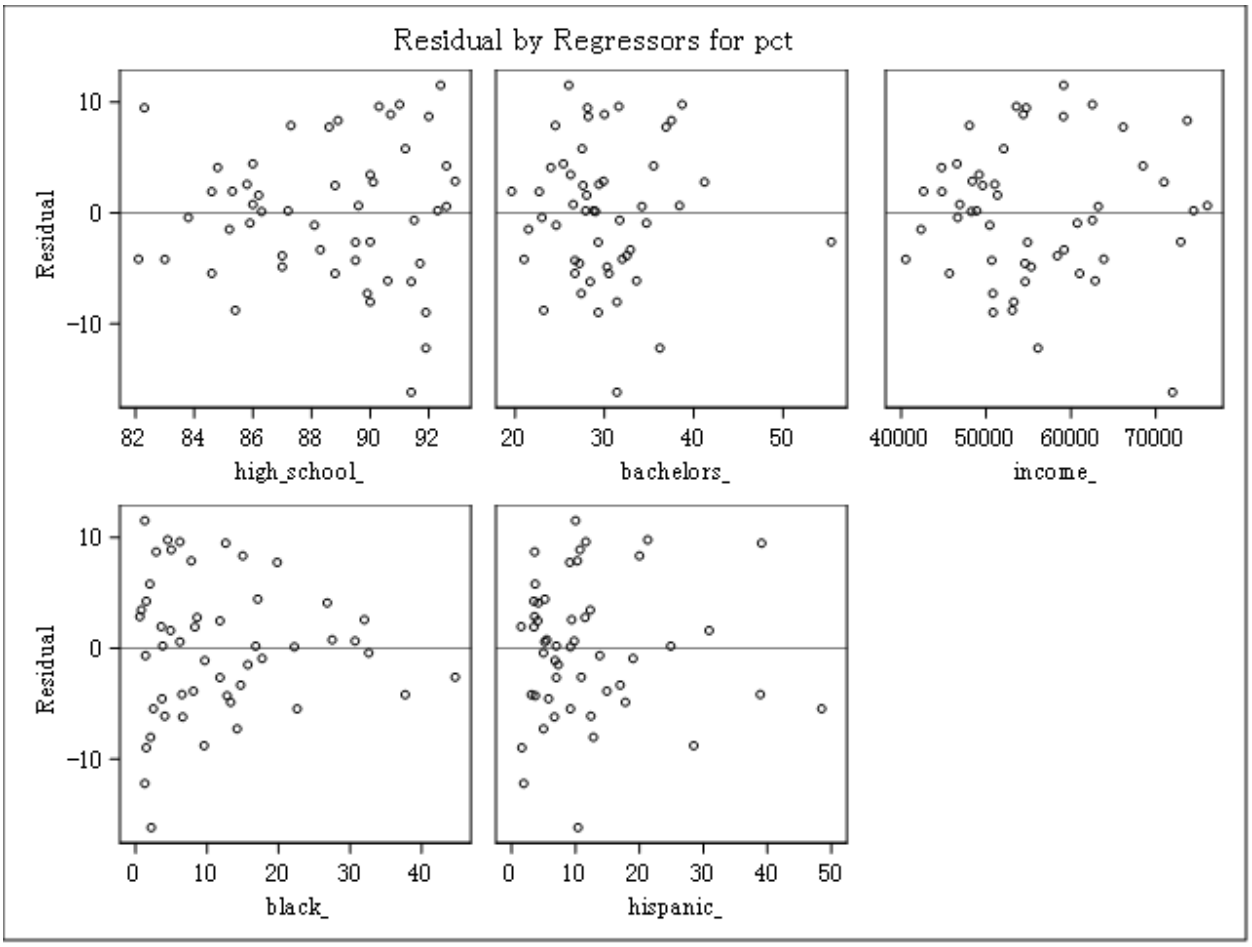
Number of Observations Read	51
Number of Observations Used	51

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	5283.01457	1056.60291	25.54	<.0001
Error	45	1861.58783	41.36862		
Corrected Total	50	7144.60240			

Root MSE	6.43184	R-Square	0.7394
Dependent Mean	48.37742	Adj R-Sq	0.7105
Coeff Var	13.29514		

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	59.93502	54.62165	1.10	0.2784	0
high_school_	1	0.49499	0.65545	0.76	0.4541	4.63175
bachelors_	1	-1.55677	0.28740	-5.42	<.0001	3.61198
income_	1	-0.00011058	0.00017034	-0.65	0.5195	2.91351
black_	1	-0.02975	0.14204	-0.21	0.8350	2.76054
hispanic_	1	-0.19885	0.14075	-1.41	0.1646	2.51256


Fit Diagnostics for pct
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
pct
Predicted Value
Cook's D
Observation
Percent
Residual
Fit-Mean
Residual
Proportion Les
Observations 51
Parameters 6
Error DF 45
MSE 41.369
R-Square 0.7394
Adj R-Square 0.7105


Residual by Regressors for pct
Residual
10
0
−10
82
84
86
88
90
92
high_school_
20
30
40
50
bachelors_
40000
50000
60000
70000
income_
Residual
10
0
−10
0
10
20
30
40
black_
0
10
20
30
40
50
hispanic_

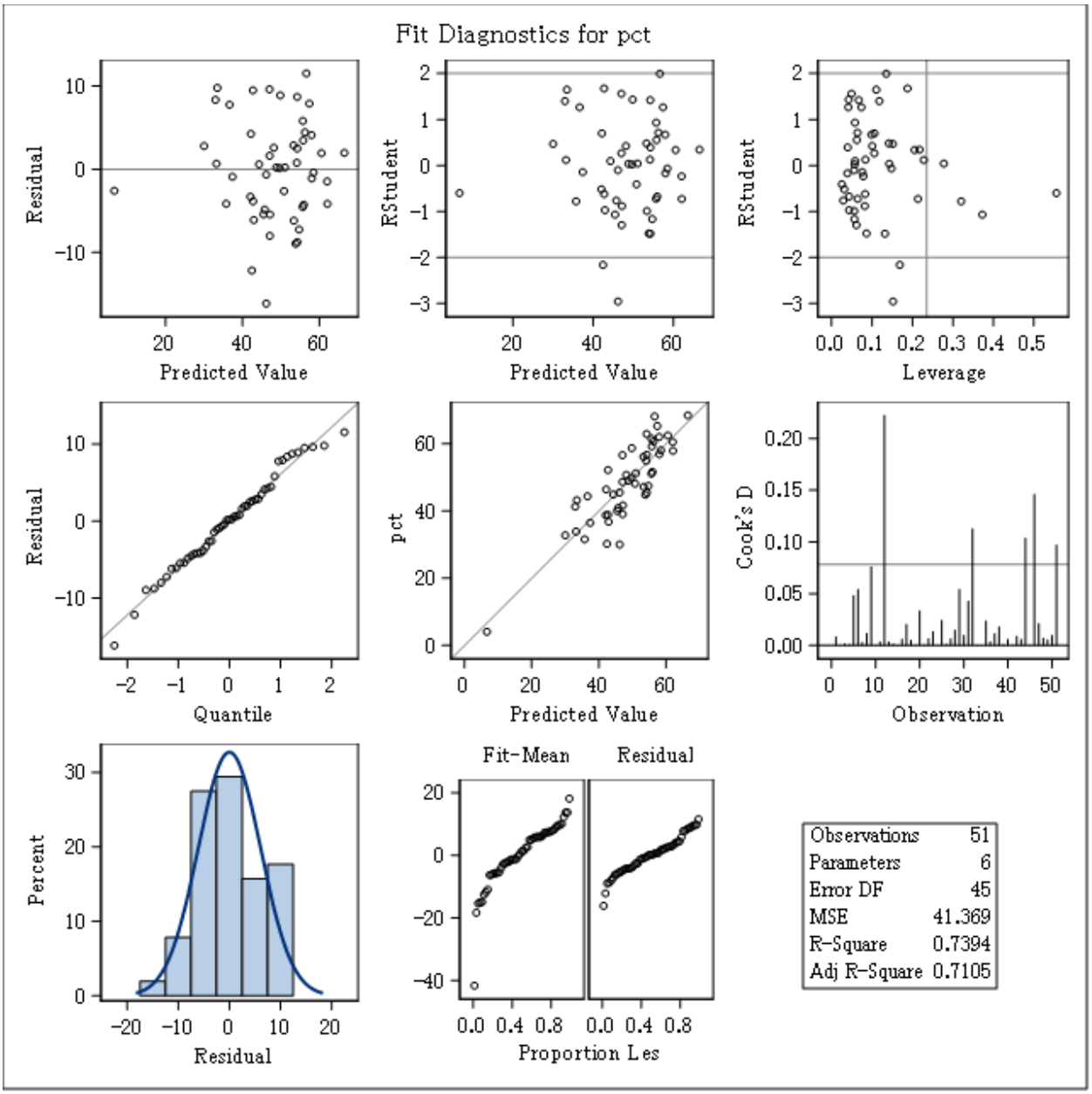
Number of Observations Read	51
Number of Observations Used	51

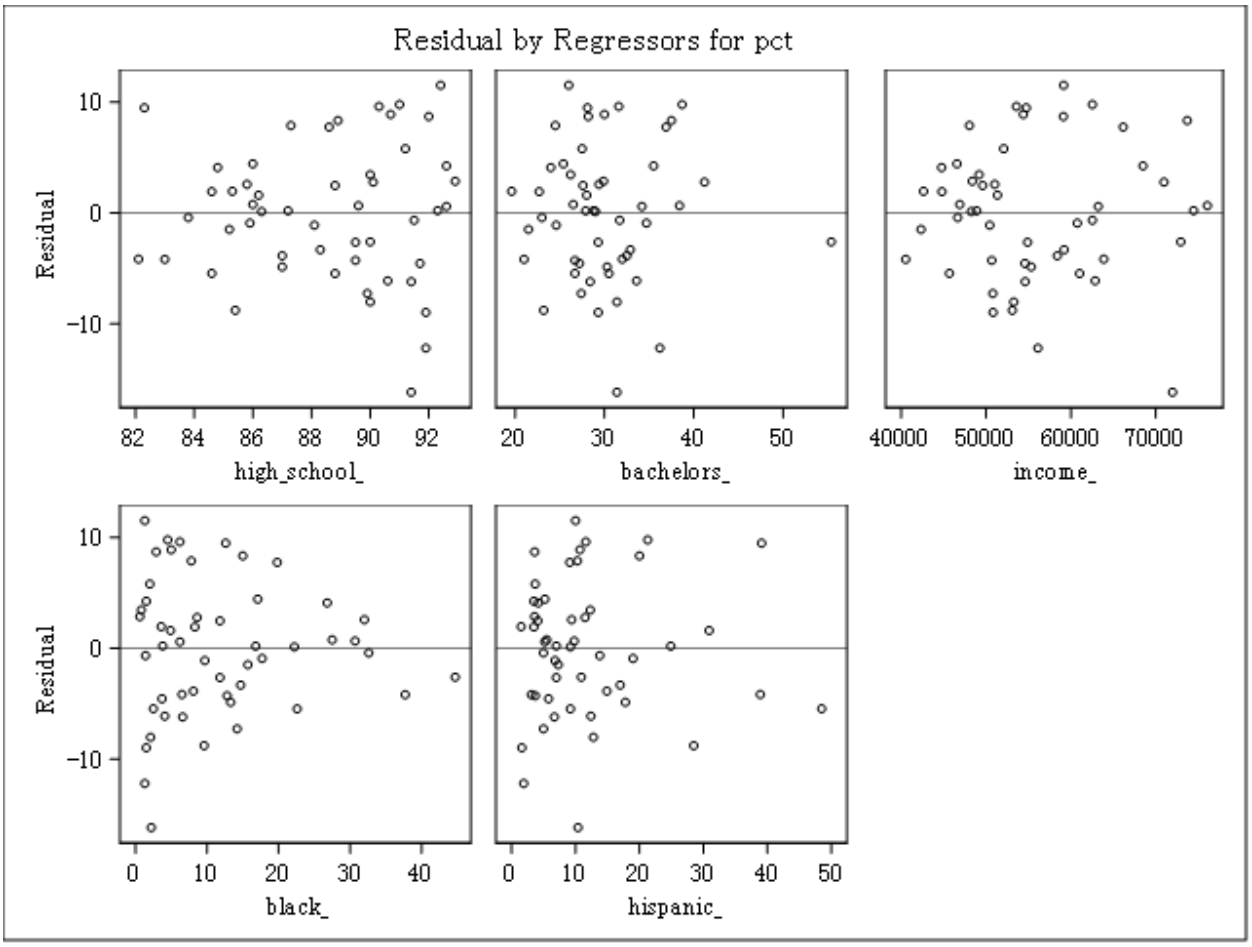
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	5283.01457	1056.60291	25.54	<.0001
Error	45	1861.58783	41.36862		
Corrected Total	50	7144.60240			

Root MSE	6.43184	R-Square	0.7394
Dependent Mean	48.37742	Adj R-Sq	0.7105
Coeff Var	13.29514		

Parameter Estimates								
						Heteroscedasticity Consistent		
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standard Error	t Value	Pr > |t|
Intercept	1	59.93502	54.62165	1.10	0.2784	51.02276	1.17	0.2463
high_school_	1	0.49499	0.65545	0.76	0.4541	0.62520	0.79	0.4327
bachelors_	1	-1.55677	0.28740	-5.42	<.0001	0.38471	-4.05	0.0002
income_	1	-0.00011058	0.00017034	-0.65	0.5195	0.00023230	-0.48	0.6364
black_	1	-0.02975	0.14204	-0.21	0.8350	0.13334	-0.22	0.8244
hispanic_	1	-0.19885	0.14075	-1.41	0.1646	0.15444	-1.29	0.2045

Test of First and Second Moment Specification		
DF	Chi-Square	Pr > ChiSq
20	15.84	0.7264


Fit Diagnostics for pct
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
pct
Predicted Value
Cook's D
Observation
Percent
Residual
Fit-Mean
Residual
Proportion Les
Observations 51
Parameters 6
Error DF 45
MSE 41.369
R-Square 0.7394
Adj R-Square 0.7105


Residual by Regressors for pct
Residual
10
0
−10
82
84
86
88
90
92
high_school_
20
30
40
50
bachelors_
40000
50000
60000
70000
income_
Residual
10
0
−10
0
10
20
30
40
black_
0
10
20
30
40
50
hispanic_

Code for the 1st model

```
ODS RTF FILE='M:\econ304\model.rtf';

data one;
infile 'M:\econ304\maxim.csv' dlm = ',';
input trump clinton other hs c y white black am asian native hispanic;
high_school_ = hs *100;
bachelors_ = c*100;
income_ = y;
black_ = black*100;
hispanic_ = hispanic*100;

pct = 100*trump/(trump + clinton + other);
run;
proc print;
run;
proc corr;
run;
proc reg;
model pct = high_school_ bachelors_ income_ black_ hispanic_/VIF;
run;

ODS RTF CLOSE;
```

Code for the 2nd model

```
ODS RTF FILE ='M:\econ304\hetero.rtf';

data ht;
infile 'M:\econ304\maxim.csv' dlm = ',';
input trump clinton other hs c y white black am asian native hispanic;
high_school_ = hs *100;
bachelors_ = c*100;
income_ = y;
black_ = black*100;
hispanic_ = hispanic*100;
pct = 100*trump/(trump + clinton + other);
run;

proc reg;
model  pct = high_school_ bachelors_  black_ hispanic_/spec hccmethod=3 white;
output out=two residual = res;
run;
data hetero;
set two;
resq = res*res;
run;
proc reg data=hetero;
model resq = high_school bachelors_  black_ hispanic_;
run;
ODS RTF CLOSE;
```