Utilizing US Census Data to Determine What Influences Income

The Outliers: Damary Gutierrez,[redacted]

11/22/2020

Introduction and Data

Income inequality is a serious issue in the United States. A study by Akee, Jones, and Porter examined income inequality and mobility on the basis of race and ethnicity over the period 2000-2014. They found that Whites and Asians accrued higher income shares at all points in the time period[1]. Blacks, Hispanics, and Native Americans were clustered at the low end of the income distribution in each year with Pacific Islanders being slightly higher. They also found that Whites and Asians have the highest levels of within-group inequality and the lowest levels of intra-group mobility. Blacks, Hispanics, and Native Americans have lower intra-group inequality and immobility. There are also many factors besides race that influence income but we included the Akee, Jones, and Porter study because the long history of oppression on Black and Brown individuals in the US has created a very evident gap in wealth and other social indicators.

Thanks to data collected by the US Census Bureau—which was expanded upon in the 2015 American Community Survey 5-year estimates—we can explore the effect of various metrics on income through the use of multiple linear regression. With this data, we are interested in answering the following question: what factors have the most impact on an individual's income within the US (including Puerto Rico)? We can then hypothesize that variables such as the percent of a population within a specific race or ethnicity, the percent of a population under the poverty level, the percent of a population employed in service jobs, and the unemployment rate are useful predictors in the explanation of the variation in median household income per capita within the United States and Puerto Rico.

The data was acquired from tables of the 2015 American Community Survey 5-year estimates and posted on Kaggle.com by a member named MuonNeutrino[2]. The US Census Bureau collects data about the economy and the people living in the United States directly from respondents through censuses every decade and general surveys. In addition, they collect data from additional sources (federal, state, and local governments) and this is called administrative data. The US Census Bureau then combines the administrative data with the survey and census data. The cases of this data set are each county (or county equivalent) within the United States and Puerto Rico.

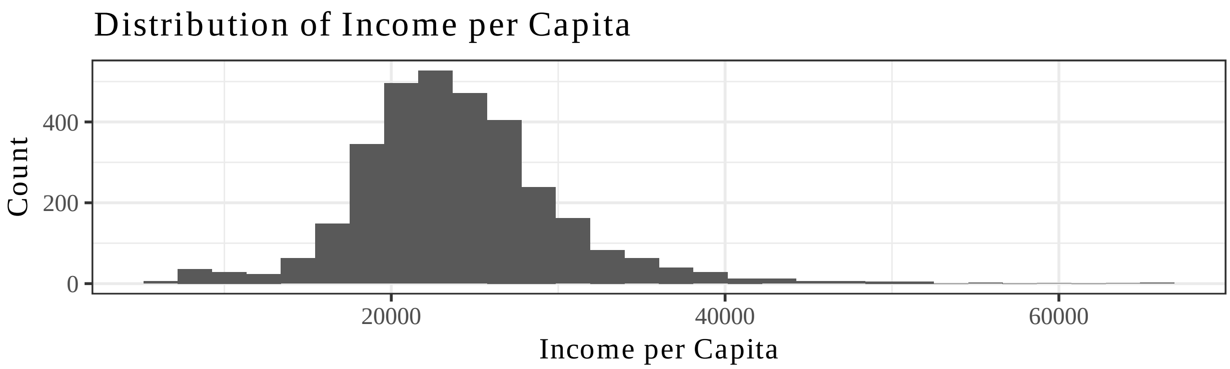
Here is a general description of the relevant variables:

- `IncomePerCap` - average income per capita
- `White` - % of population that is White
- `Asian` - % of population that is Asian
- `Poverty` - % of population living under the poverty line
- `ChildPoverty` - % of children living under the poverty line
- `Professional` - % of employed people working in management, business, science, and arts
- `Service` - % of employed people working in service jobs
- `Office` - % of employed people who work an office job
- `Construction` - % of employed people who work in natural resources, construction, and maintenance
- `Drive` - % of employed people who drive to work
- `Carpool` - % of employed people who carpool to work
- `Walk` - % of employed people who walk to work
- `OtherTransp` - % of employed people who use another form of transportation
- `WorkAtHome` - % of employed people who work from home

- `MeanCommute` - average commute time (in minutes)

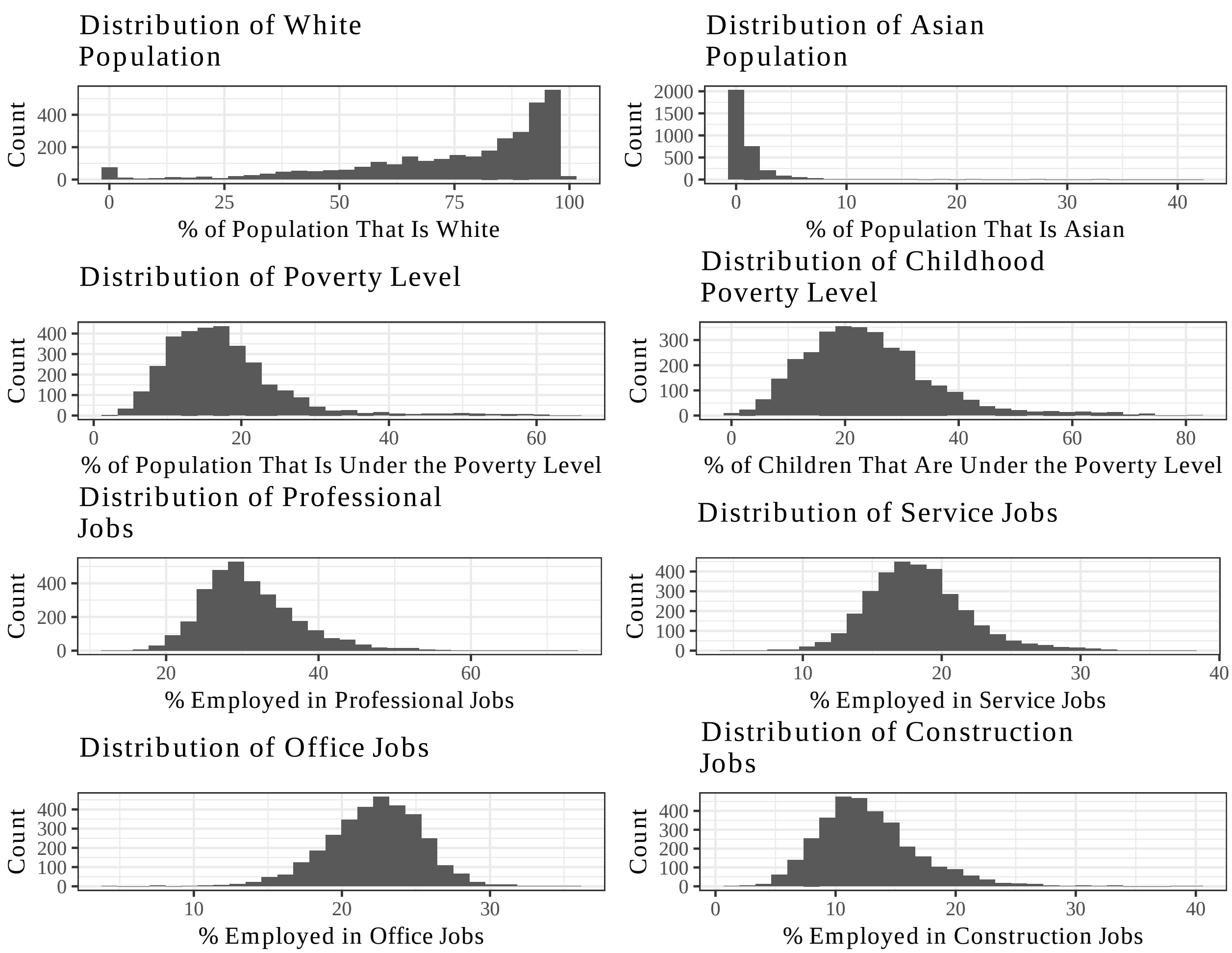
Methodology

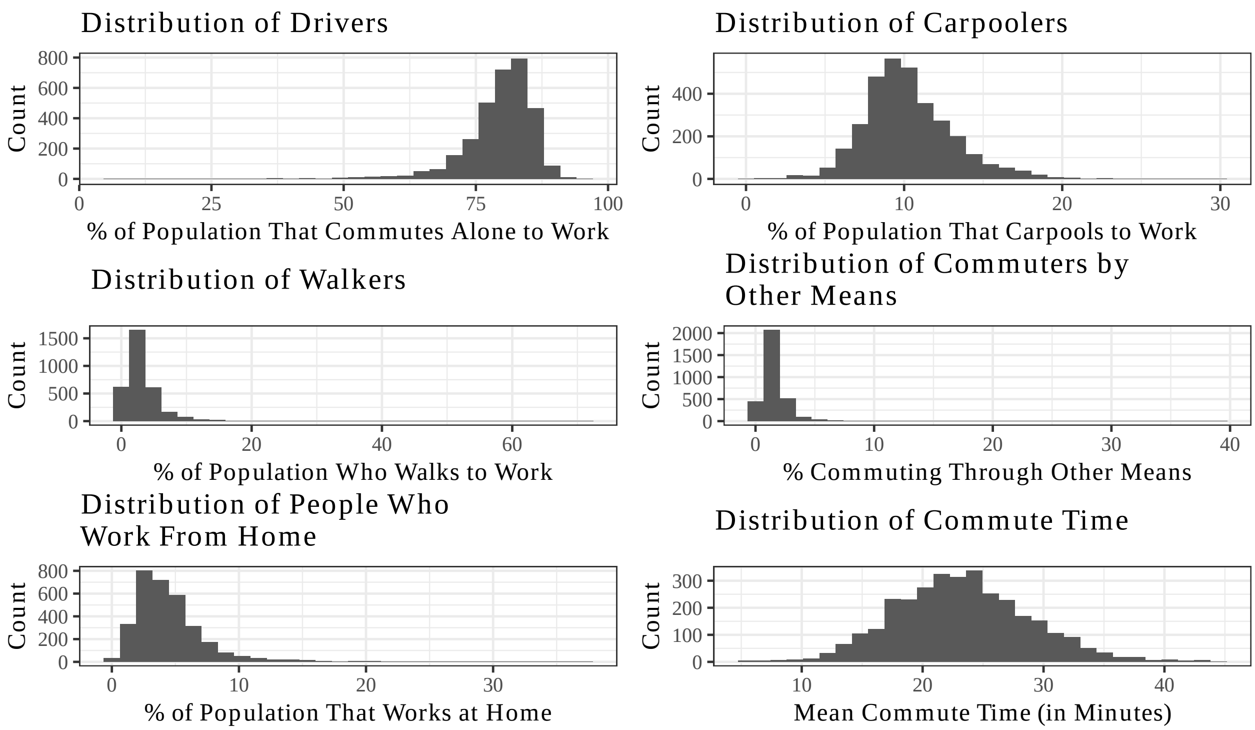
For this analysis, we want to understand which factors have the most impact on an individual's income per capita in the US. Since `IncomePerCapita` is a continuous numerical variable, we must use multiple linear regression to create a model with statistically significant predictor variables.



The visualization above shows the distribution of income per capita of the participants. As the right-skewed graph shows, the majority of people have an income per capita of around $21,000.

Here are the distributions of some relevant predictor variables:





While most of the variables are either right-skewed (`Asian`, `Poverty`, `ChildPoverty`, `Professional`, `Service`, `Construction`, `Carpool`, `Walk`, `OtherTransp`, and `WorkAtHome`) or left-skewed (`White`, `Office`, and `Drive`), `MeanCommute` is surprisingly very normally distributed.

Results

In order to arrive at our final model, we will start off with a model that includes all of the possible predictor variables. We will then check model diagnostics and transform variables if necessary. Once no other changes can be made, we will arrive at our final model.

First, we will begin with a linear model that includes every variable within the data set except for `State` and `County`. Not only would our model become much too long with the addition of `State` and `County`, we also want our model to be generalizable to a non-specific individual in the United States. We do have to recognize that in not using `State` and ‘County, we have to be very careful in how we interpret our results.

As such, our full model has the following intercept and coefficients:

term	estimate
(Intercept)	145598.844
Men	-0.001
Women	-0.057
Hispanic	-16.560
White	-33.952
Black	-19.128
Native	-11.261
Asian	168.249
Pacific	-220.643
Citizen	0.011
Poverty	-509.981
ChildPoverty	63.820
Professional	987.338

term	estimate
Service	654.183
Office	659.471
Construction	667.216
Production	581.459
Drive	-527.040
Carpool	-566.129
Transit	-301.939
Walk	-462.712
OtherTransp	-471.317
WorkAtHome	-484.028
MeanCommute	38.462
Employed	0.047
PrivateWork	-1319.168
PublicWork	-1458.013
SelfEmployed	-1358.741
FamilyWork	-1660.284
Unemployment	-17.784

Next, we will run the step function in order to backwards select the variables that are deemed important within the model with AIC as our selection criteria. The lower the AIC, the better the model fit. Such a process can be found in the **Appendix**.

We then arrive at the following model with significant predictors only:

term	estimate
(Intercept)	168988.737
Women	-0.058
White	-16.394
Asian	168.298
Citizen	0.010
Poverty	-513.346
ChildPoverty	63.297
Professional	409.794
Service	71.063
Office	77.702
Construction	87.561
Drive	-228.204
Carpool	-266.780
Walk	-161.009
OtherTransp	-171.769
WorkAtHome	-185.515
MeanCommute	34.911
Employed	0.048
PrivateWork	-1288.800
PublicWork	-1427.052
SelfEmployed	-1327.877
FamilyWork	-1632.786

We are not done yet, however. Our current model does not contain any interaction variables. Even though we chose to not include `State` in the full model, it could be the case that a state's economic policy greatly impacts variables such as poverty (`Poverty` and `ChildPoverty`), the types of jobs that are most represented

(`Professional`, `Service`, `Office`, and `Construction`), the employment rate (`Employed`), and mode through which work is done (`PrivateWork`, `PublicWork`, `SelfEmployed`, and `FamilyWork`). Due to the fact that `State` is a variable with 51 categories, we will omit this step however. Otherwise, we would arrive at a model with 50+ terms. When drawing conclusions, we must keep the lack of interaction variables in mind.

We can then then check the following conditions for multiple linear regression:

Correlation
We can check that the predictor variables are not too correlated with each other using VIF.

```
##        Women        White        Asian      Citizen      Poverty ChildPoverty
##   390.601380     2.311564     1.823923   189.098566     9.381550     9.060212
## Professional      Service       Office Construction        Drive      Carpool
##     2.957385     1.624325     1.534691     2.263075     9.354550     2.446876
##         Walk  OtherTransp   WorkAtHome  MeanCommute     Employed  PrivateWork
##     4.756640     1.834338     3.748991     1.404654   248.860598 19137.482693
##   PublicWork SelfEmployed   FamilyWork
## 13059.890280  4790.719244    65.328058
```

Using 10 as our threshold, we will remove all variables that have a VIF value that is too large. Variables with a VIF value larger than 10 are variables that are strongly correlated with one or more variables already in the model. By removing these variables, we can minimize instability in our estimation.

As such, our model then becomes the following:

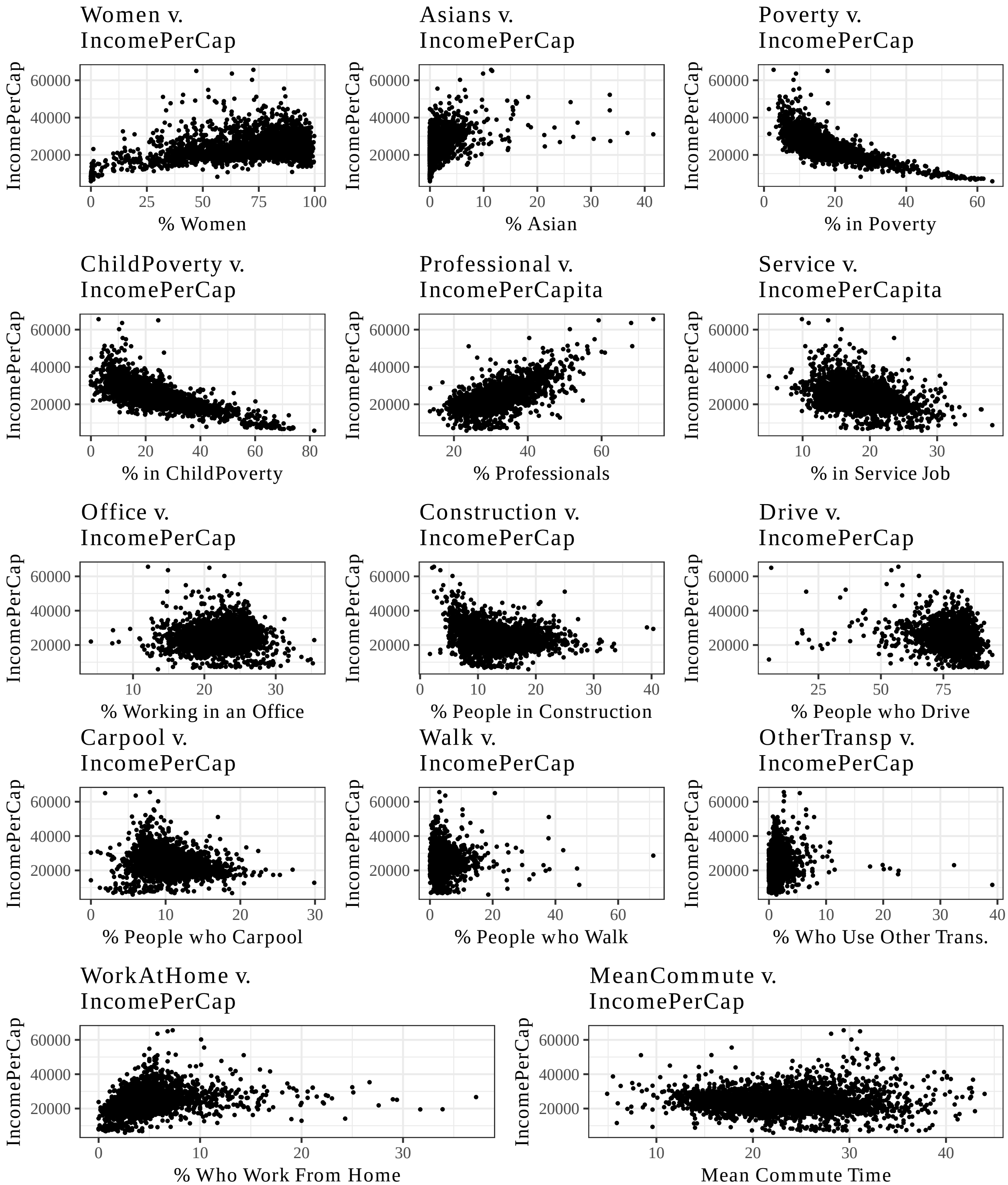
term	estimate
(Intercept)	43201.872
White	-5.386
Asian	252.009
Poverty	-554.622
ChildPoverty	70.909
Professional	367.238
Service	-14.056
Office	79.628
Construction	22.306
Drive	-249.297
Carpool	-308.463
Walk	-228.833
OtherTransp	-210.785
WorkAtHome	-251.301
MeanCommute	36.697

Let’s check correlation again in order to confirm that the new model contains less multicollinearity.

```
##        White        Asian      Poverty ChildPoverty Professional      Service
##     2.143160     1.603518     9.084105     8.931792     2.306344     1.409033
##       Office Construction        Drive      Carpool         Walk  OtherTransp
##     1.477465     1.998665     8.772854     2.329731     4.397720     1.777847
##   WorkAtHome  MeanCommute
##     2.938015     1.364114
```

Since all variables have a VIF value of less than 10, we can proceed to the next conditions with the most recent model as the current model.

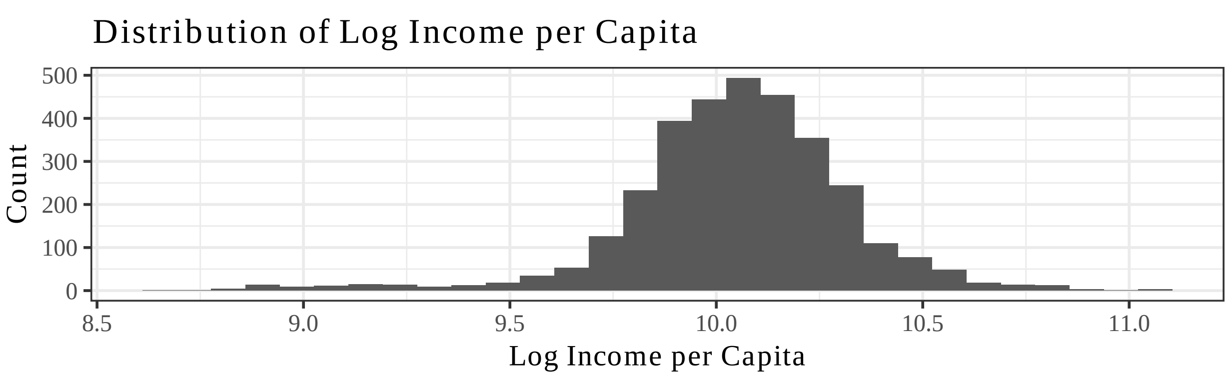
Linearity
Next, we can check the linearity condition. In order for our model to meet the condition, the relationship between the response variable and the predictor variables has to be linear.



According to the graphs, some variables have a more linear relationship than others. For example, poverty and child poverty most definitely show a negative correctional relationship to income per capita, while the women and professional variables show a positive correctional relationship to income per capita. As the percent of people in poverty increase, the income per capita decreases, whereas when the percent of women increase, the income per capita increases. Although it may appear that the other variables such as the percent of people working in an office, the percent of people in service jobs, and the variables pertaining to forms of commute have no linear relationship to income per capita, they actually do. It is just that their linear relationships are much weaker than that of the stronger variables mentioned earlier in the paragraph, but it is still apparent. For example, as the percent of people who carpool increases, so does the income per capita even though the linear relationship is not very strong.

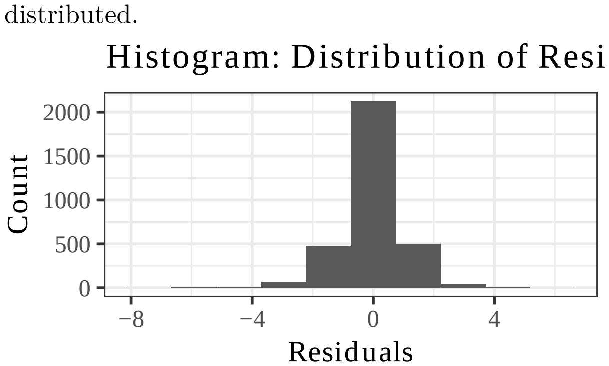
Since certain predictors are very obviously not linear, let's try some transformations on the response variable! As seen in our **Methodology**, `IncomePerCapita` has a right skew. As such, we can log-transform the response variable. This will leave us with the following model with `IncomePerCap` being log transformed:

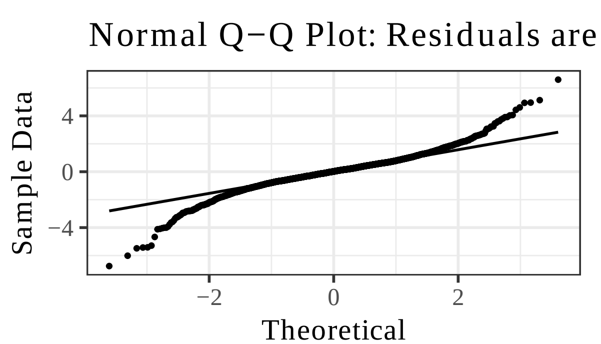
term	estimate	std.error	statistic	p.value
(Intercept)	10.716	0.083	128.400	0.000
White	0.001	0.000	4.767	0.000
Asian	0.010	0.001	11.250	0.000
Poverty	-0.030	0.001	-45.793	0.000
ChildPoverty	0.006	0.000	11.904	0.000
Professional	0.012	0.000	28.555	0.000
Service	0.000	0.001	0.306	0.760
Office	0.004	0.001	6.159	0.000
Construction	0.001	0.001	1.374	0.170
Drive	-0.008	0.001	-11.476	0.000
Carpool	-0.008	0.001	-8.551	0.000
Walk	-0.009	0.001	-8.304	0.000
OtherTransp	-0.004	0.001	-2.806	0.005
WorkAtHome	-0.008	0.001	-8.593	0.000
MeanCommute	-0.001	0.000	-1.756	0.079



Here, the distribution is a bit more normal, so we can carefully proceed and keep in mind that the linearity condition is not met.

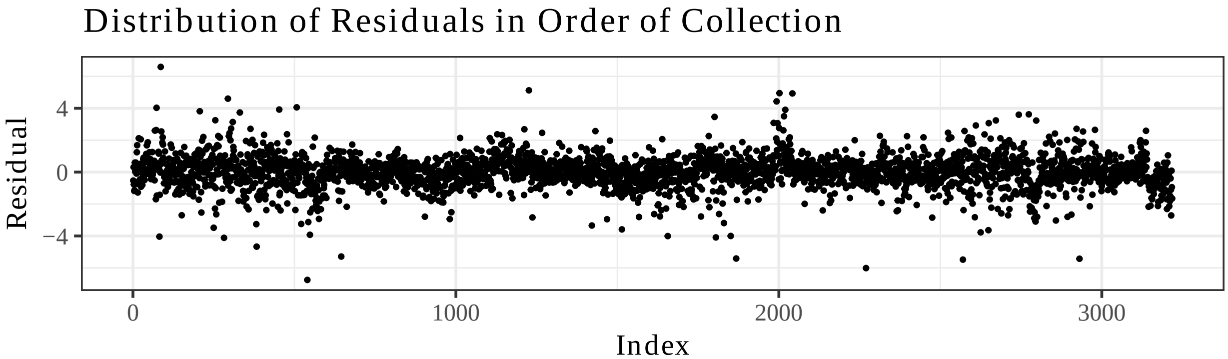
Normality
Next, we can check the normality condition. Here, we are looking for residuals that are nearly normally distributed.





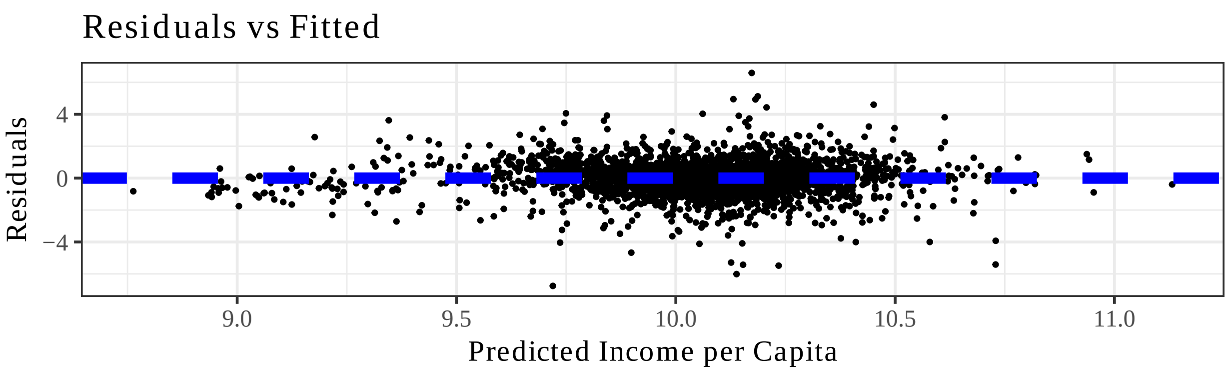
Both the histogram and the Q-Q plot visualization show that the distribution of residuals is nearly normal – normality is almost met. Furthermore, since n > 30, we can relax the normality condition and proceed with the log-transformed model.

Independence
We can then check the independence condition. Here, we are looking to see if the residuals are independent.



There are definitely some trends in this graph; however, it is still fairly scattered. It can be concluded that there is little lack of independence. To fully check for independence, we want to look at how data was collected. Independence condition is met if the sample is random and that the sample size of 1,000 children makes up a small portion of the population of all children.

Equal Variance
Finally, we can check the equal variance condition. Here, we want residuals that have a constant variance.



The plotted points are scattered around the dotted line at y = 0 with pretty constant variance, yet with some patterns that may indicate a lack of equal variance. There is some structure in the residuals in that many of the points are segregated in one specific area; however, they are still fairly scattered. We tentatively conclude that the equal variance condition is not met, and keep this in mind when drawing conclusions.

Discussion

In this project, we investigated the factors that are most statistically significant in predicting an individual's income within the US (including Puerto Rico). In order to create a model, we decided that a multiple linear regression model would produce a model of best fit since our response variable is numerically continuous. While we hypothesized that variables such as the percent of a population within a specific race or ethnicity, the percent of a population under the poverty level, the percent of a population employed in service jobs, and the unemployment rate would be included within the final model, our actual final model consisted of the following predictor variables: `White`, `Asian`, `Poverty`, `ChildPoverty`, `Professional`, `Service`, `Office`, `Construction`, `Drive`, `Carpool`, `Walk`, `OtherTransp`, `WorkAtHome`, and `MeanCommunte`. We also log-transformed our response variable `IncomePerCap`, which changes the interpretation of our model. Such a final model was identified through a backwards model selection process using AIC as the criteria and then checking the conditions. It is interesting to note that while our original hypothesis was correct in identifying that race, poverty level, and job type is important in determining a person's income, it is a bit surprising that the unemployment rate was not significant enough of a variable to include within the final model. Perhaps some of the predictive power is found within the other variables that did make the cut. Another surprising finding is the significance in the variables pertaining to mode of transportation. Since the coefficients for `Drive`, `Carpool`, `Walk`, and `OtherTransp` are negative, this means that there is an inverse relationship between income per capita and the percentage of people who take a certain mode of transportation. While not that surprising, it seems as if the poverty rate is the most effective among the variables in determining income (as determined by its coefficient of -0.03 whose absolute value is larger than the other coefficients).

As seen in our above **Results** section, it is important to note that our model does not fulfill the following model conditions: linearity and equal variance. As such, we would hesitate to use such a model for any sort of real-world decision-making; when drawing conclusions, we should take our results with a grain of salt. Regarding the linearity condition, it is very clear that certain variables (such as `Professional`, `Service`, and `Office`—among others) are not linear. Due to the fact that such variables are "clumped" around certain values, transforming said variables using log transformations and quadratic transformations was not effective. More information regarding other possible transformations will be necessary in order to create a truly effective model. Finding the necessary transformations might also help in checking the equal variance condition. It is also important to note that because our data set is based on counties and not individuals, our interpretations are a bit more difficult to communicate clearly; such a difficultly is further compounded by the fact that many variables are based on percentages and not counts. Additionally, since we log-transformed our response variable, we are predicting the more general median income per capita instead of an exact value. Since some conditions are not met, it could also be the case that a linear regression model is not the best model for this data. Since logistic regression is not applicable because our response variable is continuous and not ordinal, some other form of regression model might be a better fit. Such models are way beyond the scope of this class, however.

Were we to start over, it would be more advantageous to start with a data set that is a bit easier to interpret. Such a data set would consist of individual cases instead of county cases so that the variables would not be based in percentages. Furthermore, it would allow for more categorical variables other than state and county which have too many categories to effectively interpret the results.

Sources

1 : https://link-springer-com.proxy.lib.duke.edu/article/10.1007/s13524-019-00773-7
2 : https://www.kaggle.com/muonneutrino/us-census-demographic-data?select=acs2015_county_data.csv

---

Appendix

Here is the full print out of the backwards selection process:

```
## Start:  AIC=50813.87
## IncomePerCap ~ Men + Women + Hispanic + White + Black + Native +
##     Asian + Pacific + Citizen + Poverty + ChildPoverty + Professional +
##     Service + Office + Construction + Production + Drive + Carpool +
##     Transit + Walk + OtherTransp + WorkAtHome + MeanCommute +
##     Employed + PrivateWork + PublicWork + SelfEmployed + FamilyWork +
##     Unemployment
##
##                Df  Sum of Sq        RSS   AIC
## - Men           1      32682 2.2763e+10 50812
## - Native        1     687788 2.2763e+10 50812
```

```
## - Transit       1    1390659 2.2764e+10 50812
## - Hispanic      1    1790727 2.2764e+10 50812
## - Black         1    2351891 2.2765e+10 50812
## - Walk          1    3261787 2.2766e+10 50812
## - OtherTransp   1    3387171 2.2766e+10 50812
## - WorkAtHome    1    3567561 2.2766e+10 50812
## - Drive         1    4232719 2.2767e+10 50812
## - Production    1    4382344 2.2767e+10 50812
## - Carpool       1    4881685 2.2767e+10 50813
## - Service       1    5549747 2.2768e+10 50813
## - Office        1    5636338 2.2768e+10 50813
## - Construction  1    5775057 2.2768e+10 50813
## - Unemployment  1    6235340 2.2769e+10 50813
## - White         1    7335039 2.2770e+10 50813
## - Professional  1   12647619 2.2775e+10 50814
## - Pacific       1   13082263 2.2776e+10 50814
## <none>                       2.2763e+10 50814
## - PrivateWork   1   17838788 2.2780e+10 50814
## - SelfEmployed  1   18950788 2.2782e+10 50815
## - PublicWork    1   21795924 2.2784e+10 50815
## - FamilyWork    1   28082684 2.2791e+10 50816
## - Citizen       1   80093200 2.2843e+10 50823
## - Asian         1   90872407 2.2853e+10 50825
## - MeanCommute   1   95502868 2.2858e+10 50825
## - Women         1  143862017 2.2906e+10 50832
## - ChildPoverty  1  183256286 2.2946e+10 50838
## - Employed      1  594906032 2.3357e+10 50895
## - Poverty       1 5246367451 2.8009e+10 51479
##
## Step:  AIC=50811.88
## IncomePerCap ~ Women + Hispanic + White + Black + Native + Asian +
##     Pacific + Citizen + Poverty + ChildPoverty + Professional +
##     Service + Office + Construction + Production + Drive + Carpool +
##     Transit + Walk + OtherTransp + WorkAtHome + MeanCommute +
##     Employed + PrivateWork + PublicWork + SelfEmployed + FamilyWork +
##     Unemployment
##
##                Df  Sum of Sq        RSS   AIC
## - Native        1     678349 2.2763e+10 50810
## - Transit       1    1395473 2.2764e+10 50810
## - Hispanic      1    1781326 2.2764e+10 50810
## - Black         1    2334040 2.2765e+10 50810
## - Walk          1    3279223 2.2766e+10 50810
## - OtherTransp   1    3409877 2.2766e+10 50810
## - WorkAtHome    1    3589049 2.2766e+10 50810
## - Drive         1    4254766 2.2767e+10 50810
## - Production    1    4367383 2.2767e+10 50810
## - Carpool       1    4906407 2.2768e+10 50811
## - Service       1    5532468 2.2768e+10 50811
## - Office        1    5619922 2.2768e+10 50811
## - Construction  1    5757702 2.2768e+10 50811
## - Unemployment  1    6275960 2.2769e+10 50811
## - White         1    7317264 2.2770e+10 50811
## - Professional  1   12625491 2.2775e+10 50812
```

```
## - Pacific       1   13077766 2.2776e+10 50812
## <none>                       2.2763e+10 50812
## - PrivateWork   1   17835622 2.2780e+10 50812
## - SelfEmployed  1   18945932 2.2782e+10 50813
## - PublicWork    1   21794193 2.2784e+10 50813
## - FamilyWork    1   28076687 2.2791e+10 50814
## - Citizen       1   83727169 2.2846e+10 50822
## - Asian         1   91576108 2.2854e+10 50823
## - MeanCommute   1   95473971 2.2858e+10 50823
## - ChildPoverty  1  183432105 2.2946e+10 50836
## - Employed      1  634488175 2.3397e+10 50898
## - Women         1  719926339 2.3483e+10 50910
## - Poverty       1 5246348479 2.8009e+10 51477
##
## Step:  AIC=50809.98
## IncomePerCap ~ Women + Hispanic + White + Black + Asian + Pacific +
##     Citizen + Poverty + ChildPoverty + Professional + Service +
##     Office + Construction + Production + Drive + Carpool + Transit +
##     Walk + OtherTransp + WorkAtHome + MeanCommute + Employed +
##     PrivateWork + PublicWork + SelfEmployed + FamilyWork + Unemployment
##
##                Df  Sum of Sq        RSS   AIC
## - Transit       1    1401631 2.2765e+10 50808
## - Walk          1    3295088 2.2767e+10 50808
## - OtherTransp   1    3424461 2.2767e+10 50808
## - WorkAtHome    1    3609825 2.2767e+10 50808
## - Drive         1    4267093 2.2768e+10 50809
## - Production    1    4410444 2.2768e+10 50809
## - Carpool       1    4914792 2.2768e+10 50809
## - Hispanic      1    5348470 2.2769e+10 50809
## - Service       1    5588644 2.2769e+10 50809
## - Office        1    5675782 2.2769e+10 50809
## - Construction  1    5810596 2.2769e+10 50809
## - Unemployment  1    6466123 2.2770e+10 50809
## - Black         1    9683055 2.2773e+10 50809
## - Professional  1   12705652 2.2776e+10 50810
## - Pacific       1   13176970 2.2776e+10 50810
## <none>                       2.2763e+10 50810
## - PrivateWork   1   17805385 2.2781e+10 50810
## - SelfEmployed  1   18909829 2.2782e+10 50811
## - PublicWork    1   21761394 2.2785e+10 50811
## - FamilyWork    1   28032470 2.2791e+10 50812
## - White         1   65915762 2.2829e+10 50817
## - Citizen       1   85150732 2.2848e+10 50820
## - MeanCommute   1   95485117 2.2859e+10 50821
## - ChildPoverty  1  183756354 2.2947e+10 50834
## - Asian         1  271720057 2.3035e+10 50846
## - Employed      1  633815365 2.3397e+10 50896
## - Women         1  721011212 2.3484e+10 50908
## - Poverty       1 5255910465 2.8019e+10 51476
##
## Step:  AIC=50808.17
## IncomePerCap ~ Women + Hispanic + White + Black + Asian + Pacific +
##     Citizen + Poverty + ChildPoverty + Professional + Service +
```

```
##     Office + Construction + Production + Drive + Carpool + Walk +
##     OtherTransp + WorkAtHome + MeanCommute + Employed + PrivateWork +
##     PublicWork + SelfEmployed + FamilyWork + Unemployment
##
##                Df  Sum of Sq        RSS   AIC
## - Production    1    4364685 2.2769e+10 50807
## - Hispanic      1    5462491 2.2770e+10 50807
## - Service       1    5538996 2.2770e+10 50807
## - Office        1    5622374 2.2770e+10 50807
## - Construction  1    5761178 2.2770e+10 50807
## - Unemployment  1    6405436 2.2771e+10 50807
## - Black         1    9761759 2.2774e+10 50808
## - Professional  1   12633760 2.2777e+10 50808
## - Pacific       1   13057241 2.2778e+10 50808
## <none>                       2.2765e+10 50808
## - PrivateWork   1   17950607 2.2783e+10 50809
## - SelfEmployed  1   19062011 2.2784e+10 50809
## - PublicWork    1   21926877 2.2787e+10 50809
## - FamilyWork    1   28177191 2.2793e+10 50810
## - White         1   66226780 2.2831e+10 50816
## - Citizen       1   85107480 2.2850e+10 50818
## - MeanCommute   1   95679289 2.2860e+10 50820
## - OtherTransp   1  134667195 2.2899e+10 50825
## - ChildPoverty  1  183871292 2.2949e+10 50832
## - Walk          1  239028268 2.3004e+10 50840
## - Asian         1  270883899 2.3036e+10 50844
## - WorkAtHome    1  282485698 2.3047e+10 50846
## - Employed      1  632851423 2.3398e+10 50894
## - Women         1  720056611 2.3485e+10 50906
## - Carpool       1  765108731 2.3530e+10 50913
## - Drive         1 1000730717 2.3765e+10 50945
## - Poverty       1 5256230800 2.8021e+10 51475
##
## Step:  AIC=50806.79
## IncomePerCap ~ Women + Hispanic + White + Black + Asian + Pacific +
##     Citizen + Poverty + ChildPoverty + Professional + Service +
##     Office + Construction + Drive + Carpool + Walk + OtherTransp +
##     WorkAtHome + MeanCommute + Employed + PrivateWork + PublicWork +
##     SelfEmployed + FamilyWork + Unemployment
##
##                Df  Sum of Sq        RSS   AIC
## - Hispanic      1    5698938 2.2775e+10 50806
## - Unemployment  1    6344588 2.2775e+10 50806
## - Black         1   10015432 2.2779e+10 50806
## - Pacific       1   13033091 2.2782e+10 50807
## <none>                       2.2769e+10 50807
## - PrivateWork   1   18288743 2.2787e+10 50807
## - SelfEmployed  1   19423613 2.2788e+10 50808
## - PublicWork    1   22306345 2.2791e+10 50808
## - FamilyWork    1   28523491 2.2798e+10 50809
## - White         1   66723148 2.2836e+10 50814
## - Citizen       1   84598993 2.2854e+10 50817
## - MeanCommute   1   96287123 2.2865e+10 50818
## - Office        1  128646066 2.2898e+10 50823
```

```
## - OtherTransp   1  134316257 2.2903e+10 50824
## - Service       1  137690444 2.2907e+10 50824
## - Construction  1  176158907 2.2945e+10 50830
## - ChildPoverty  1  184214744 2.2953e+10 50831
## - Walk          1  237575131 2.3007e+10 50838
## - Asian         1  270433607 2.3040e+10 50843
## - WorkAtHome    1  282312666 2.3051e+10 50844
## - Employed      1  634371462 2.3403e+10 50893
## - Women         1  720098601 2.3489e+10 50905
## - Carpool       1  761613735 2.3531e+10 50911
## - Drive         1  997781808 2.3767e+10 50943
## - Poverty       1 5253332533 2.8022e+10 51473
## - Professional  1 7057624863 2.9827e+10 51674
##
## Step:  AIC=50805.6
## IncomePerCap ~ Women + White + Black + Asian + Pacific + Citizen +
##     Poverty + ChildPoverty + Professional + Service + Office +
##     Construction + Drive + Carpool + Walk + OtherTransp + WorkAtHome +
##     MeanCommute + Employed + PrivateWork + PublicWork + SelfEmployed +
##     FamilyWork + Unemployment
##
##                Df  Sum of Sq        RSS   AIC
## - Black         1    4617075 2.2779e+10 50804
## - Unemployment  1    4890686 2.2780e+10 50804
## - Pacific       1   10776436 2.2786e+10 50805
## <none>                       2.2775e+10 50806
## - PrivateWork   1   18419747 2.2793e+10 50806
## - SelfEmployed  1   19610065 2.2794e+10 50806
## - PublicWork    1   22355383 2.2797e+10 50807
## - FamilyWork    1   28649952 2.2803e+10 50808
## - Citizen       1   85368100 2.2860e+10 50816
## - MeanCommute   1   92472473 2.2867e+10 50817
## - Office        1  125549804 2.2900e+10 50821
## - OtherTransp   1  129006275 2.2904e+10 50822
## - Service       1  132771339 2.2908e+10 50822
## - Construction  1  170992585 2.2946e+10 50828
## - White         1  192503536 2.2967e+10 50831
## - ChildPoverty  1  197941760 2.2973e+10 50831
## - Walk          1  232885499 2.3008e+10 50836
## - WorkAtHome    1  276847704 2.3052e+10 50842
## - Asian         1  297494256 2.3072e+10 50845
## - Employed      1  638761510 2.3414e+10 50893
## - Women         1  725459973 2.3500e+10 50905
## - Carpool       1  757278254 2.3532e+10 50909
## - Drive         1  996084978 2.3771e+10 50941
## - Poverty       1 5499683244 2.8274e+10 51500
## - Professional  1 7055139704 2.9830e+10 51672
##
## Step:  AIC=50804.25
## IncomePerCap ~ Women + White + Asian + Pacific + Citizen + Poverty +
##     ChildPoverty + Professional + Service + Office + Construction +
##     Drive + Carpool + Walk + OtherTransp + WorkAtHome + MeanCommute +
##     Employed + PrivateWork + PublicWork + SelfEmployed + FamilyWork +
##     Unemployment
```

```
##
##                Df  Sum of Sq        RSS   AIC
## - Unemployment  1    5385164 2.2785e+10 50803
## - Pacific       1   10258366 2.2790e+10 50804
## <none>                       2.2779e+10 50804
## - PrivateWork   1   18220667 2.2798e+10 50805
## - SelfEmployed  1   19363240 2.2799e+10 50805
## - PublicWork    1   22142549 2.2802e+10 50805
## - FamilyWork    1   28515599 2.2808e+10 50806
## - Citizen       1   84059138 2.2863e+10 50814
## - MeanCommute   1   90067050 2.2869e+10 50815
## - OtherTransp   1  127536339 2.2907e+10 50820
## - Office        1  133976544 2.2913e+10 50821
## - Service       1  137388897 2.2917e+10 50822
## - Construction  1  187243457 2.2967e+10 50829
## - ChildPoverty  1  194541825 2.2974e+10 50830
## - White         1  199513207 2.2979e+10 50830
## - Walk          1  230552813 2.3010e+10 50835
## - WorkAtHome    1  275195638 2.3055e+10 50841
## - Asian         1  303992514 2.3083e+10 50845
## - Employed      1  638373580 2.3418e+10 50891
## - Women         1  722648571 2.3502e+10 50903
## - Carpool       1  754891702 2.3534e+10 50907
## - Drive         1  993955510 2.3773e+10 50940
## - Poverty       1 5739477944 2.8519e+10 51525
## - Professional  1 7096756759 2.9876e+10 51675
##
## Step:  AIC=50803.01
## IncomePerCap ~ Women + White + Asian + Pacific + Citizen + Poverty +
##     ChildPoverty + Professional + Service + Office + Construction +
##     Drive + Carpool + Walk + OtherTransp + WorkAtHome + MeanCommute +
##     Employed + PrivateWork + PublicWork + SelfEmployed + FamilyWork
##
##                Df  Sum of Sq        RSS   AIC
## - Pacific       1   10095454 2.2795e+10 50802
## <none>                       2.2785e+10 50803
## - PrivateWork   1   17979711 2.2803e+10 50804
## - SelfEmployed  1   19071839 2.2804e+10 50804
## - PublicWork    1   21908286 2.2807e+10 50804
## - FamilyWork    1   28299852 2.2813e+10 50805
## - Citizen       1   80924628 2.2866e+10 50812
## - MeanCommute   1   84882009 2.2870e+10 50813
## - Office        1  132421100 2.2917e+10 50820
## - OtherTransp   1  135434960 2.2920e+10 50820
## - Service       1  136087748 2.2921e+10 50820
## - ChildPoverty  1  194320491 2.2979e+10 50828
## - White         1  194543432 2.2979e+10 50828
## - Construction  1  198559278 2.2983e+10 50829
## - Walk          1  232436764 2.3017e+10 50834
## - WorkAtHome    1  278916434 2.3064e+10 50840
## - Asian         1  304848977 2.3090e+10 50844
## - Employed      1  651170503 2.3436e+10 50892
## - Women         1  722676272 2.3507e+10 50901
## - Carpool       1  760438981 2.3545e+10 50907
```

```
## - Drive         1  998096425 2.3783e+10 50939
## - Poverty       1 6258419163 2.9043e+10 51582
## - Professional  1 7262562861 3.0047e+10 51691
##
## Step:  AIC=50802.43
## IncomePerCap ~ Women + White + Asian + Citizen + Poverty + ChildPoverty +
##     Professional + Service + Office + Construction + Drive +
##     Carpool + Walk + OtherTransp + WorkAtHome + MeanCommute +
##     Employed + PrivateWork + PublicWork + SelfEmployed + FamilyWork
##
##                Df  Sum of Sq        RSS   AIC
## <none>                       2.2795e+10 50802
## - PrivateWork   1   17063236 2.2812e+10 50803
## - SelfEmployed  1   18141084 2.2813e+10 50803
## - PublicWork    1   20922041 2.2816e+10 50803
## - FamilyWork    1   27205418 2.2822e+10 50804
## - Citizen       1   78149796 2.2873e+10 50811
## - MeanCommute   1   87428611 2.2882e+10 50813
## - Office        1  129250890 2.2924e+10 50819
## - Service       1  132155612 2.2927e+10 50819
## - OtherTransp   1  141686941 2.2937e+10 50820
## - Construction  1  193483500 2.2988e+10 50828
## - ChildPoverty  1  194482254 2.2989e+10 50828
## - White         1  196542215 2.2991e+10 50828
## - Walk          1  239932041 2.3035e+10 50834
## - WorkAtHome    1  297364208 2.3092e+10 50842
## - Asian         1  340360181 2.3135e+10 50848
## - Employed      1  672552580 2.3467e+10 50894
## - Women         1  728875143 2.3524e+10 50902
## - Carpool       1  791470834 2.3586e+10 50910
## - Drive         1 1038504924 2.3833e+10 50944
## - Poverty       1 6252135111 2.9047e+10 51580
## - Professional  1 7409127409 3.0204e+10 51706

##
## Call:
## lm(formula = IncomePerCap ~ Women + White + Asian + Citizen +
##     Poverty + ChildPoverty + Professional + Service + Office +
##     Construction + Drive + Carpool + Walk + OtherTransp + WorkAtHome +
##     MeanCommute + Employed + PrivateWork + PublicWork + SelfEmployed +
##     FamilyWork, data = census)
##
## Coefficients:
##  (Intercept)         Women         White         Asian       Citizen
##    1.690e+05    -5.782e-02    -1.639e+01     1.683e+02     1.045e-02
##      Poverty  ChildPoverty  Professional       Service        Office
##   -5.133e+02     6.330e+01     4.098e+02     7.106e+01     7.770e+01
## Construction         Drive       Carpool          Walk   OtherTransp
##    8.756e+01    -2.282e+02    -2.668e+02    -1.610e+02    -1.718e+02
##   WorkAtHome   MeanCommute      Employed   PrivateWork    PublicWork
##   -1.855e+02     3.491e+01     4.817e-02    -1.289e+03    -1.427e+03
## SelfEmployed    FamilyWork
##   -1.328e+03    -1.633e+03
```