## **‘Exploratory data analysis (EDA) using statistical tools**

By: Ananya Chetia

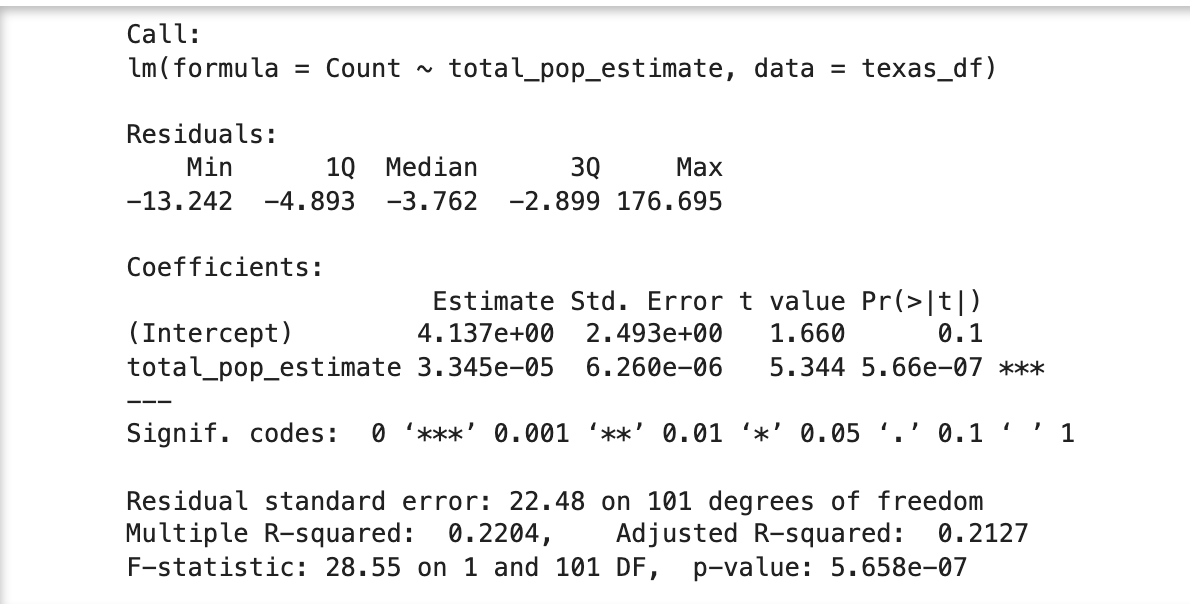
## **Background Information / Dataset Description**

* **Dataset Name:** [Private School Universe Survey (PSS)](https://nces.ed.gov/surveys/pss/pssdata.asp)
* **Link to GitHub Repo:**
* **Description:** The data was extracted from the National Center for Education Statistics for the 2021-22 academic year.It lists all the private schools listed throughout the U.S.
* **Key Objectives for the Analysis:**I looked into why there are over X private schools located in counties where the median income is less than $45,000. I am trying to understand if racial makeup and population count per county impact these decisions.

### Linear Regression

**Question:** Does a higher population impact the number of private schools in the county.

#### 



* **Interpret the results**

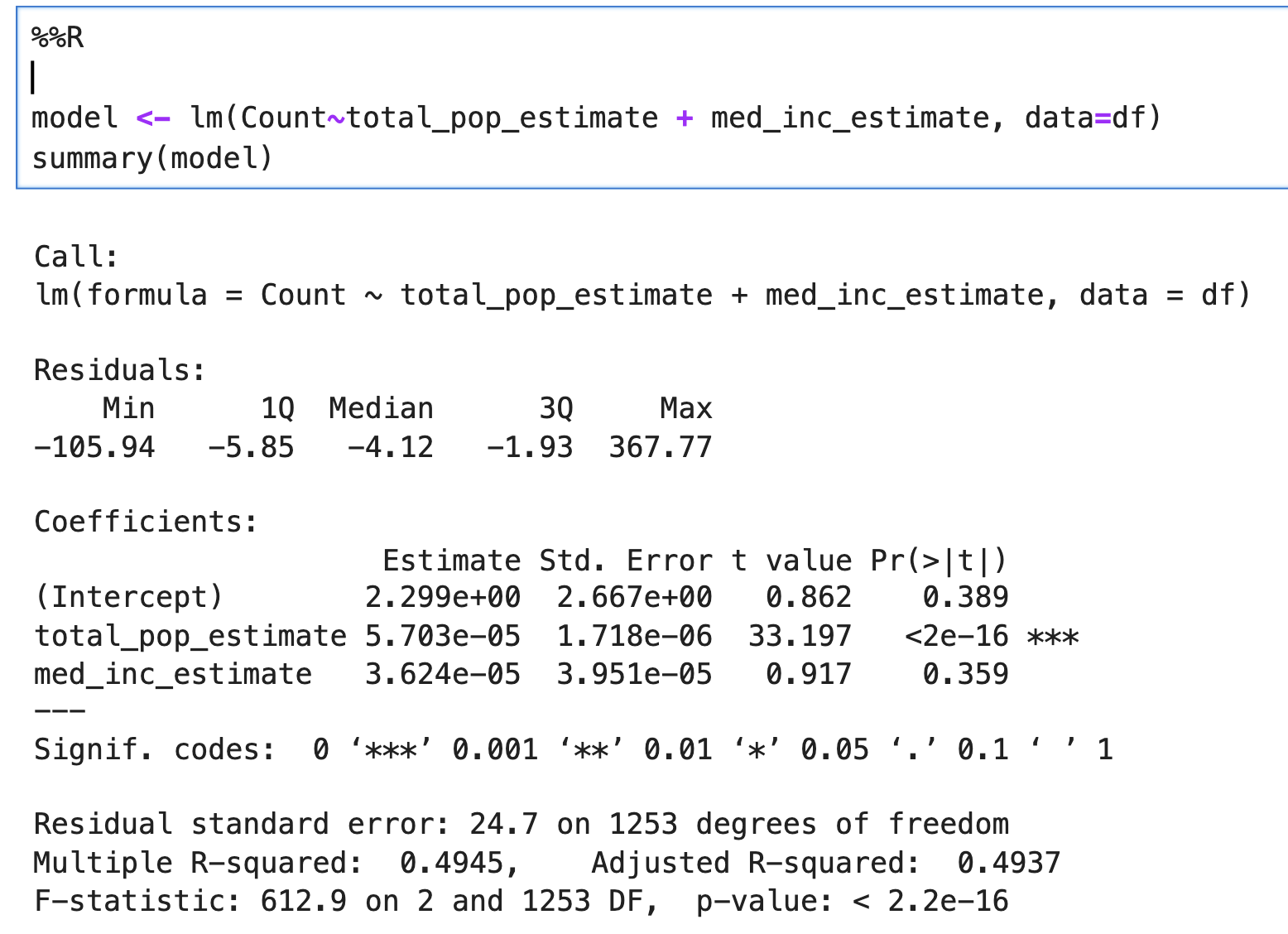
This model summary shows that there's a significant negative relationship between a car's weight and its miles per gallon, implying that as the weight of a car increases, its fuel efficiency decreases.

The coefficient for weight is -5.34, which means for each additional 1,000 pounds of weight, the car's fuel efficiency is expected to decrease by about 5.34 mpg. The model's intercept is at 37.28 mpg, indicating the expected fuel efficiency of a car that weighs 0 pounds (hypothetically).

With an R-squared value of approximately 0.753, the model explains about 75% of the variability in mpg, which is a strong fit for data like this. The extremely low p-value indicates that the results are statistically significant and that the relationship between weight and mpg is not due to random chance.

### 

### Multivariable Linear Regression



* My multivariable linear regression does not indicate a strong predictive relationship. The R square suggests that almost 50% variance can be explained between total population, median income and the number of white people in a county. The P value is low, which suggests it is highly unlikely it was not by chance.

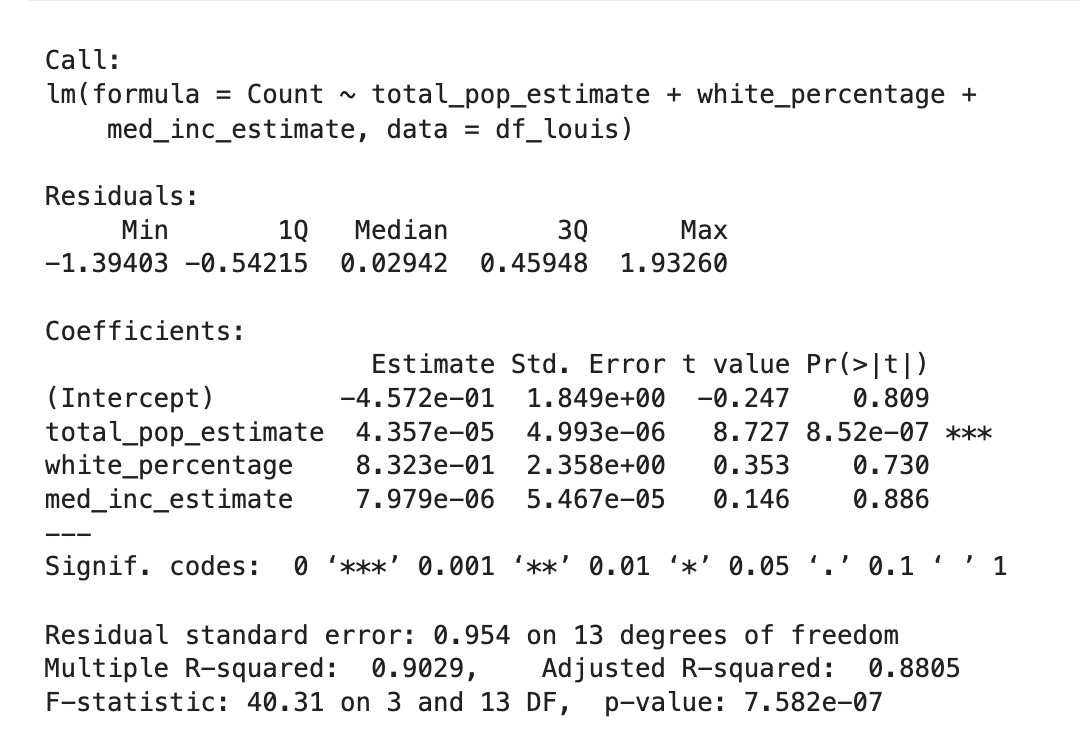
## **Conclusion**

* While I did not find a good multivariable linear regression that can lead to a story, certain schools were outliers and did stand out for me. For example, Claiborne Academy in Louisiana is a private school but it is located in a county where the median income is less than $33,000. Louisiana had 10 private schools where the median income is less than $35,000 and had 45 private schools where the median income is less than $50,000. I was confused why schools in the South like Louisiana, Mississippi and Texas have private schools in low-income communities. My multivariable linear regression could not answer this bit about these outliers. But, a source did. Read below on what they said.

**Sources**

* I spoke to Bobby Harrsion, a reporter at [Mississippi Today](https://mississippitoday.org/author/bharrison/). Harrison is a Mississpi native but reports on schools under the Midsouth Association of Independent Schools, which includes private schools in Louisiana and Mississippi. When I asked Harrison why there are private schools located in low-income counties, Harrison said this is likely because in these counties, white families prefer sending their children off to school in predominantly white schools, which are private schools. Whereas Black families in these southern states and rural counties will send their children off to public schools. That is how, as Harrison put it, segregation still exists here.

**After speaking with a source chart!!!!**

****

So Harrison said that private schools are in low-income counties in Louisiana and Mississippi because these private schools are catered towards white families wanting to send their children to predominantly or all white schools (which is legally not allowed but private schools are the way to go to an all white school in the South).

I decided to use the variables he mentioned to do a new multivariable regression analysis. The R^2 here is around 90%, which means that around 90% of the variance in the number of private schools can be explained by total population, median income and white percentage. Please note, this dataset only shows private schools where the median income is less than $50,000.