Project 2: Statistical Models

1. Introduction

Our group decided to research the fluctuating stock prices of three big tech companies: Amazon, Tesla, and Google. PBS News reports that Tesla, along with the other big tech companies known as the "Magnificent 7", is growing to be the most valued stock in the market today. As a result, these three companies' stock price fluctuations are essential to study because they affect all of Wall Street. Investopedia says that "the performance of the Magnificent Seven stocks is driven by technological innovation, market dominance, financial performance, brand equity, research and development, and global economic conditions." We wanted to compare the stock prices on different dates to study how big tech stock prices can change. Specifically, our research question is: how do presidential elections affect the stock prices of Amazon, Tesla, and Google? Therefore, we want to run tests to see if the stock prices during the years of U.S. presidential elections vary from the prices during regular years. Many investors become concerned around the time of presidential elections that their shares will diminish in short-term value due to the political scene (USBank.com). Therefore, we wanted to test this to see if stock prices do differ during election years.

2. Population of Interest/Sample Data

We used the data science platform, Kaggle, to find data recording the stock prices of the "magnificent 7 big tech" companies over the last 10 years. One possible limitation of the data set is the limited time period of the data. Because the data only spans the last 10 years, our conclusions might be swayed by trends in tech stocks from the last 10 years. Another limitation is that the data does not look into any other conditions aside from the date that may be affecting prices. For example, we will look into the year 2020, as it is an election year. However, the COVID-19 pandemic heavily affected the stock market that year. Additionally, the US band points out that elections fall around the same time as third-quarter corporate earnings, which also heavily affect the market.

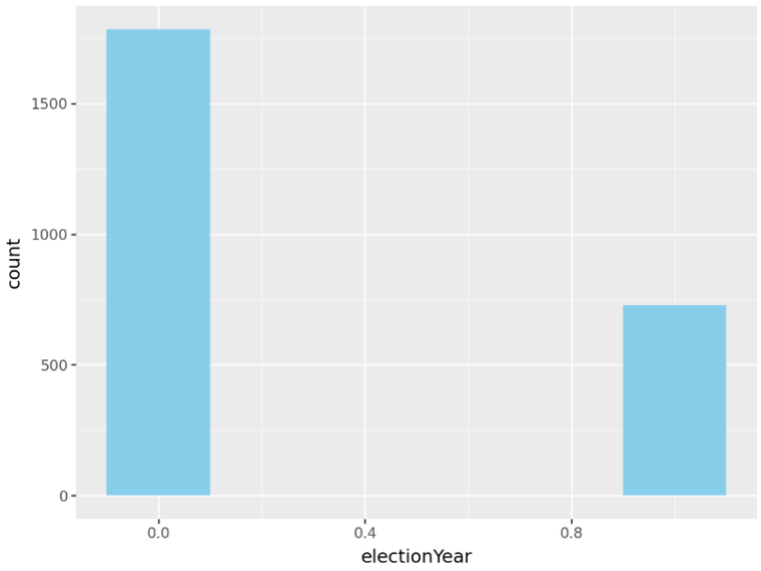
3. Descriptive Statistics + Visualization:

<u>ElectionYear</u>

ElectionYear count :

0 (non-election years): 1785

1 (election years): 728



<u>Amazon</u>

(election + non-election years):

Count = 2513

Std = 53.380589

Mean = 98.996105

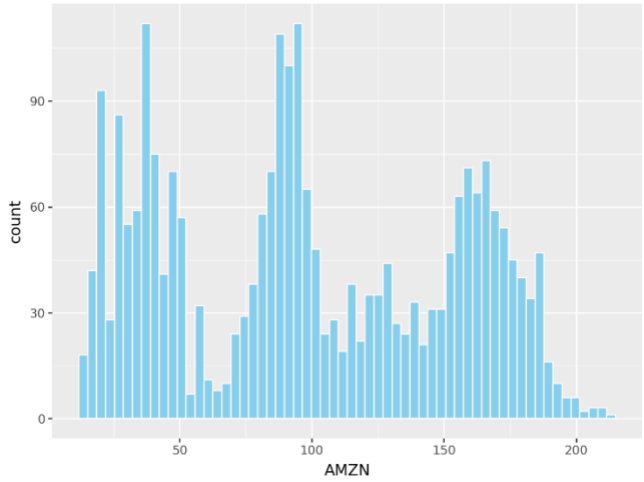
Min = 14.347500

25% = 47.682999

50% = 93.760002

75% = 152.052505

Max = 214.100006



Tesla

(election + non-election years):

Count = 2513

Std = 110.556155

Mean = 111.406961

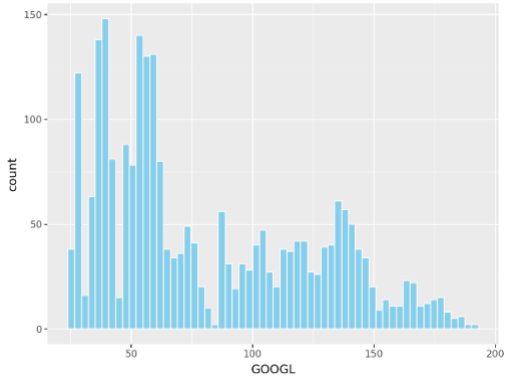
Min = 9.578000

25% = 16.967333

50% = 24.722000

75% = 217.610001

Max = 409.970001



Google

(election + non-election years):

Count = 2513

Std = 42.767798

Mean = 80.986322

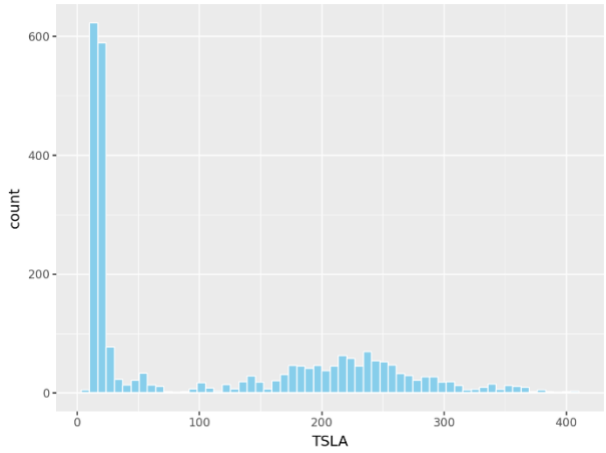
Min = 24.853001

25% = 46.844501

50% = 62.939999

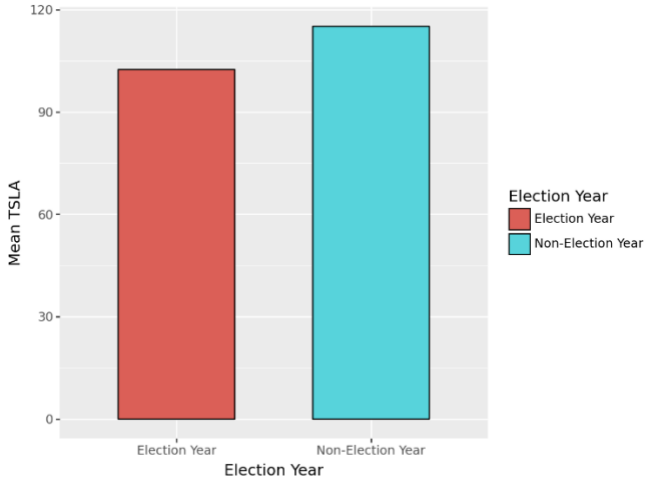
75% = 117.675003

Max = 191.179993

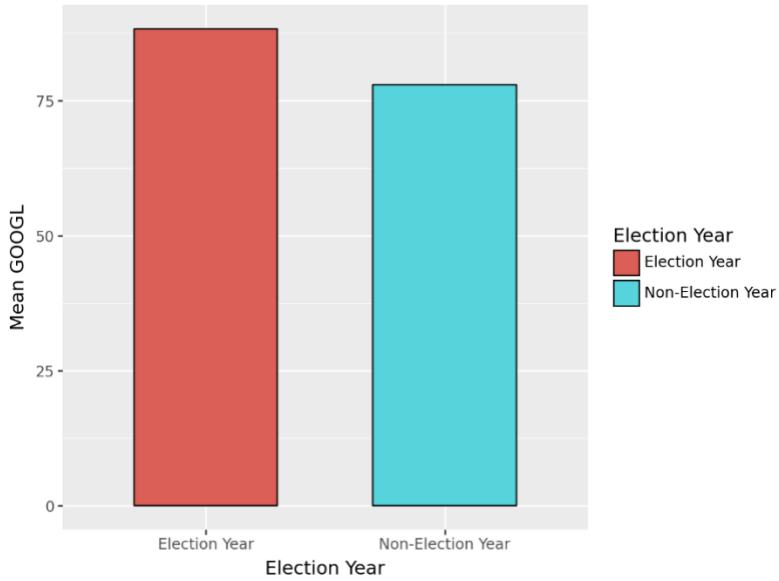


4. Relationships Among Variables:

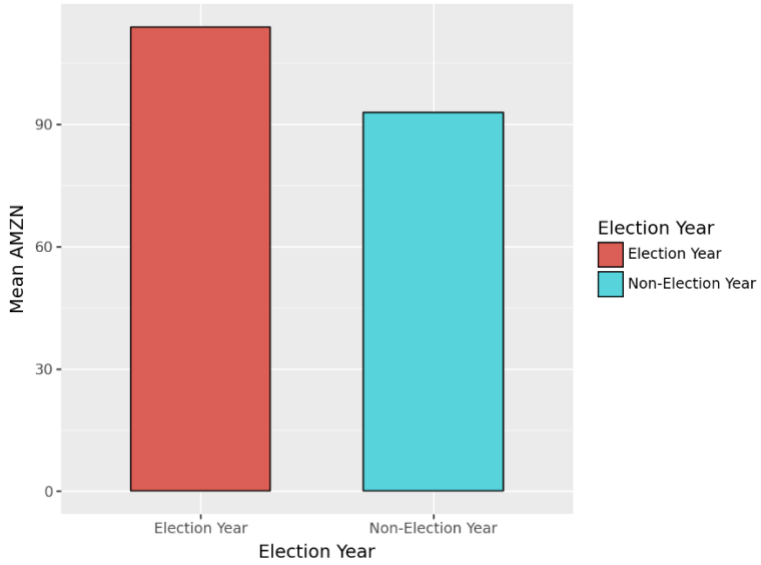
Comparison chart for TSLA group means



Comparison chart for GOOGL group means



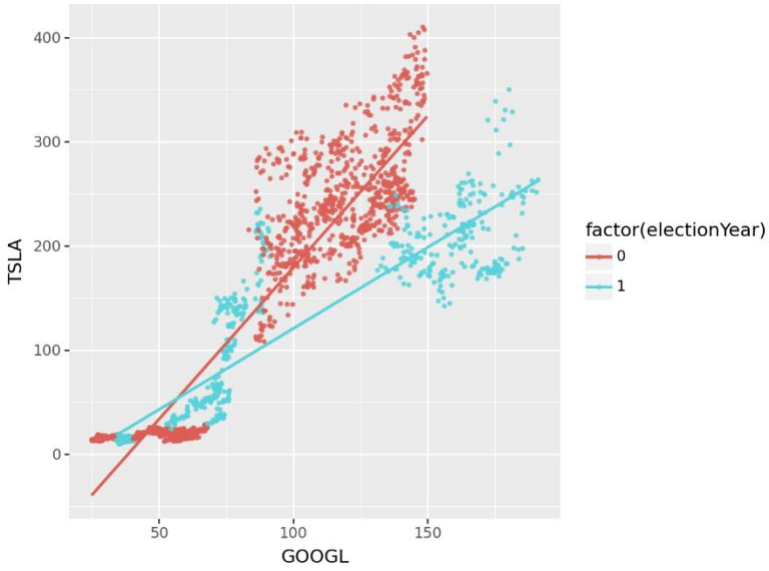
Comparison chart for AMZN group means



5. Multiple Regression + Two Sample T-test:

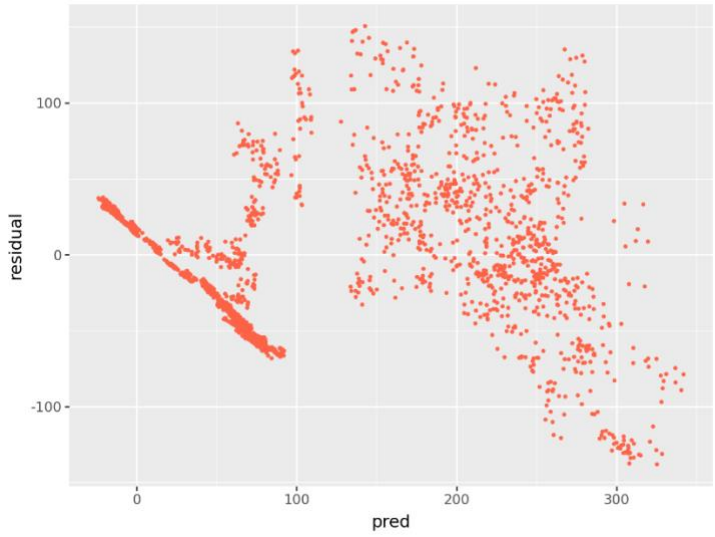
Multiple Regression assumptions and conditions (pred : TSLA , using GOOGL & Election Year)

1. Scatterplot shows a slope (Linear Assumption)

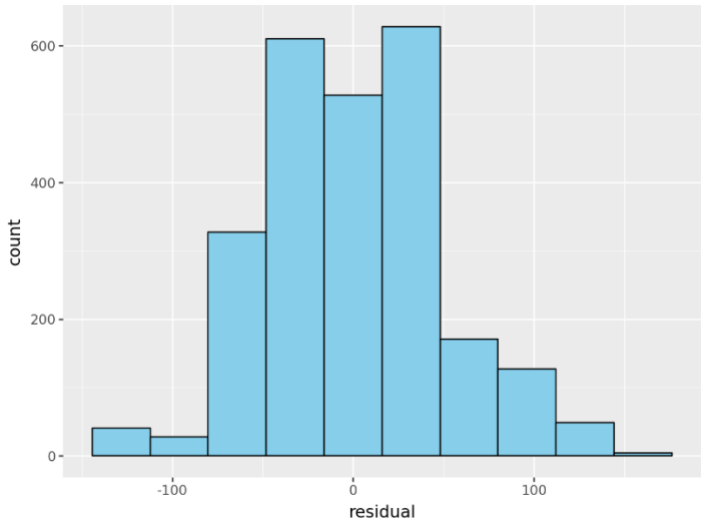


2. Random Samples (Independence Assumption)

3. The scatterplot of residuals is not thicken (Equal Variance Assumption)



4. Distribution of residuals should be nearly normal (Normality Assumption)



The multiple regression meet the conditions

Two sample T-test assumptions and conditions

1. Observations within each sample are independent from each other
2. Individuals are independent
3. Data collected follows a normal distribution

6. Hypotheses & Data Interpretation:

<u>For the Two-Sample T-Test</u>

Null Hypothesis: There is no difference in the mean stock price between election and non-election years.

Alternative Hypothesis: There is a significant difference in the mean stock price between election and non-election years.

TSLA Election vs. Non-Election Year ; *p-value = 0.003618*

AMZN Election vs. Non-Election Year ; *p-value = 1.506149e-15*

GOOGL Election vs. Non-Election Year; p-value = 0.000001

Statistical Interpretation: Given that the p-values for all three stocks are less than 0.05, we reject the null hypothesis, meaning that there is a difference in the mean stock price between election and non-election years

Layperson Interpretation: There is a difference in stock price between election and non-election years

For the Multiple regression

Null Hypothesis: There is no significant relationship between google stock price during the election year and TSLA stock prices which means that the coefficients for Google and election year are 0

Alternative Hypothesis: Google or election year has a significant relation with TSLA prices, meaning that the coefficients for Google or election year are not 0.

R-squared: 0.805

Statistical Interpretation: In our example, r squared = 0.805, suggesting that our model explained 80.5% in TSLA stock prices. Rejecting the null hypothesis.

Layperson: Based on data from our survey, there is a significant relationship between google and election year and TSLA stock prices.

7. 95% Confidence Interval

Regression Equation:

*TSLA = -66.6584 + 2.3300 * GOOGL - 36.6971 * electionYear*

2 predictions

1) Google = 200 , electionyear = 0 (non election year)

TSLA = -66.6584 + 2.3300(200) -36.6971(0)

TSLA = 399.34

4. Google = 1500, election year = 1 (election year)

TSLA = -66.6584 + 2.3300(1500) -36.6971(1)

TSLA = 3391.65

CI for prediction 1 : [393.4052, 405.268683]

- Interpretation: We are 95% confident that the average stock price for Google in non-election years is between 393.4052 and 405.268683.

CI for prediction 2: [3328.087956, 3455.130859]

- Interpretation: We are 95% confident that the average stock price for Google in election years is between 3328.087956 and 3455.130859

8. Conclusions

Based on our multiple regression model and our two sample t-test, we know that there is a difference in stock prices between election years and non-election years. In our multiple regression model, we concluded an r square of 0.805, and consequently rejected our null hypothesis (There is no significant relationship between google and election year and TSLA stock prices which means that the coefficients for Google and election year are 0). When we ran our two sample t-test, we saw the same results, with the p-value turning out to be less than 0.05, rejecting our null hypothesis again (There is no difference in the mean stock price between election and non-election years). However, looking back at our limitations, we must take into consideration how elections fall around the same time as third-quarter corporate earnings, as well as Covid-19 in 2020. With both of these events, it makes more sense that we came to this conclusion but we must also show limitations in our results because of it.

Work Cited:

Shaban, Hazma. “Yahoo Finance - Stock Market Live, Quotes, Business & Finance News.” *Yahoo! Finance*, Yahoo!, 8 Aug. 2024, finance.yahoo.com/.

“How Presidential Elections Affect the Stock Market: U.S. Bank.” *How Presidential Elections Affect the Stock Market | U.S. Bank*, 7 Nov. 2024, www.usbank.com/investing/financial-perspectives/market-news/how-presidential-elections-affect-the-stock-market.html.

Thompson, Cedric. “Magnificent 7 Stocks: What You Need to Know.” *Investopedia*, Investopedia, 29 Oct. 2024, www.investopedia.com/magnificent-seven-stocks-8402262.

Data:

https://www.kaggle.com/datasets/unmoved/magnificent-7-past-10-years-prices-updated-daily

Python Code:

https://colab.research.google.com/drive/17JlJmPyg_qfXuSGe7uhiYp5FRR2QqJm-?usp=sharing