Sophia Chebli

Final Project Report

My research question was: Did Income Influence Voter Turnout in Pennsylvania During the 2024 Presidential Election? This project compared median income in Pennsylvania counties with their voter turnout to determine whether there is a positive relationship.

The data was gathered from two sources. The median incomes of Pennsylvania counties were retrieved from the National Institute on Minority Health and Health Disparities. For voter turnout and registration by county, the data were obtained from an existing dataset available on Observable. This data was sourced directly from the Commonwealth of Pennsylvania website, which the state government runs. The data was cleaned from this pre-crafted dataset by deleting many irrelevant columns. Given total voter turnout and total registered voters per county, I divided the former by the latter to obtain the percentage of total voter turnout. Then, adding these numbers to a data set with income for each corresponding county, the final data set was completed.

I uploaded the data to a Google Notebook and then ran an ordinary least squares (OLS) regression as my GLM. The null hypothesis was that there is no linear relationship between income and voter turnout. 'Voter Turnout' was used as the dependent variable (y-axis) and 'Income' as the independent variable (x-axis). There are limitations to this regression, such as the assumption that errors are normally distributed, have constant variance, and that the relationship is linear.

The results indicate a statistically significant positive relationship between income and voter turnout. The R-squared is 0.265, meaning approximately 26.5% of the variance in voter turnout can be explained by income. The income coefficient (1.45e-06) indicates a very slight

increase in voter turnout with rising income. The p-value was 0.000, confirming that this relationship is statistically significant. This allows us to reject the null hypothesis of no linear association, as we found that there is.

In more general terms, wealthier people had a higher voter turnout rate in Pennsylvania in 2024. This could be attributed to many different reasons, including but not limited to job security and overall resources. Higher incomes are typically associated with more secure jobs, which provide greater mobility to take time out of a person's workday to reach polling locations. This is also linked to available resources, as higher incomes typically have more modes of transportation and access to voter education. Having these resources easily accessible is commonly known to increase civic engagement. As mentioned, these are only potential reasons why income could influence voter turnout; our model does not explicitly indicate this. Our conclusion is only that there is a higher turnout as income increases.

Through this project, I was able to take two different data sets, clean and combine them, run an ordinary least squares regression, and transform the results into a scatterplot. The regression indicated a statistically significant positive relationship between the two variables. A future project to further answer and aid this question would be to examine the average distance between polling locations and the majority of the population in these counties to further understand how voter income differs per county .

References

- **Gemini** used to code the statistical regression as well as the scatterplot

- **Income Data:** https://hdpulse.nimhd.nih.gov/data-portal/social/table?age=001&age_options=ageall_1&demo=00011&demo_options=income_3&race=00&race_options=race_7&sex=0&sex_options=sexboth_1&socialtopic=030&socialtopic_options=social_6&statefips=42&statefips_options=area_states

- **Voter Turnout and Registration Data:** https://observablehq.com/@jgusta/pennsylvania-general-turnout-2024#

  - **Source from Data:** https://www.pa.gov/agencies/dos/resources/voting-and-elections-resources/voting-and-election-statistics/election-data#accordion-9f6b1feab5-item-8ad90604b2

- Replication folder