**Relationship of class by income as related to socioeconomic variables regarding the environment (using the *Measure of America* Dataset)**

**Caitlin Jinmae Falk**

*Lab Lead: Daiki Tagami*

*Professor: Anthony Donoghue*

*Course: Statistical Thinking with Python*

Link to code: https://github.com/caitlinjinmaefalk/Final-project

The richest one percent of the world’s population is responsible for more than twice as much carbon pollution (Gore). According to a new report with the Stockholm Environment Institute, between 1990 and 2015, the richest one percent was responsible for 15 percent of emissions during that time period (Gore). There is decades-long research about the inextricable link between racism, economic class, our economic system and climate change. The environmental justice movement has inspired examination of these issues. In 2012, the senior editor of *Scientific American* wrote that “if we are ever to cope with climate change in any fundamental way, radical solutions on the social side are where we must focus…the relative efficiency of the next generation of solar cells is trivial by comparison” (Girgenti).

My research question is examining the effect of economic class and environmental variables. The purpose of this study is to investigate the effect of income inequality on states’ indicators of pollution as well as other socioeconomic variables which might pertain to the environment and adverse effects upon it. Using Census data in the *Measure of America* dataset (which is categorized by state), we will examine that relationship using statistical hypothesis testing. One of the greatest difficulties of this research project is data limitations and the fact that it examines only one dataset at a very high level of geography. Ideally, we would want to explore the relationships at a lower level of geography like county or census tract. While many studies point to a correlation between income, income inequality, capitalism and environmental injustice, the tests I ran do not necessarily indicate that with overwhelming evidence. I make note of why those findings may or may not be reliable throughout this study. It would be ideal to have observations in samples at the level of individual people to examine the relationship of racism, class, and income as related to environmental variables.

Due to the evidence extracted from datasets outside of this study, and due to confounding variables mentioned throughout, we still think it is important to conclude that wealthier states (and more specifically, wealthier individuals) distribute wealth more equitably in order to avoid the most catastrophic effects of what is most certainly, the demise of our species on this planet.

A picture containing chart

Description automatically generated

**Figure 1.** First, I examined the relationship between GDP and carcinogen distribution across states. The average carcinogen level of low GDP states in the United States was 17,442,512.653 lbs (95% CI = [(5,917,498.14, 28,967,527.17)]), and the average carcinogen level of high GDP states in the United States was 12,822,930.56 lbs (95% CI = [-2,892,394.64, 28,538,255.76)]). A two-sample t-test was conducted to compare the average carcinogen level of low GDP and high GDP states. There was not a significant difference in the carcinogen level of these states (p>0.05). The average difference was 4,619,582.09 lbs (95% CI = [(-14,868,809.47, 24,107,973.66)]). With this large difference I would’ve expected to reject the null hypothesis. However, the high variance may be due in part to the outliers and is probably the reason the null hypothesis could not be rejected.

Chart, scatter chart

Description automatically generated

**Figure 2.** Simple linear regression analysis was conducted to examine the relationship between median earnings and water consumption (per capita). A significant regression equation was found at p<0.05 but not significant at p<.001. The equation was:

Water Consumption (gallons per day per capita) = 9644.98 + median earnings\*(-0.26)

Thus, states’ water consumption (per capita) decreased 0.26 gallons for every dollar increase in median earnings (95% CI = [-0.455, -.060]). The correlation coefficient between the two variables was -0.35 (p<0.05, p>0.01). This implies that, there is only a weak negative correlation between water consumption and median earnings.

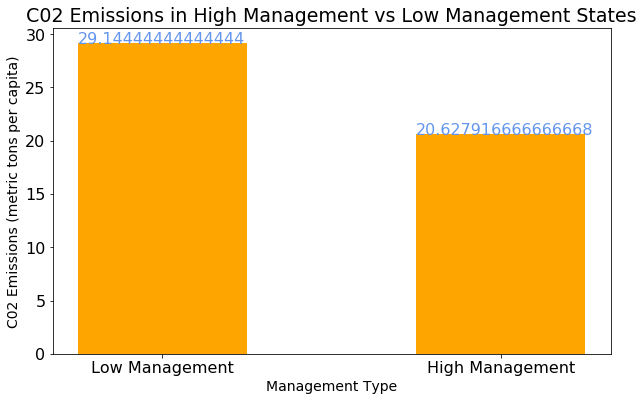
Chart, bar chart

Description automatically generated

**Figure 3.** A chi-square test was conducted to examine any association between urban versus rural states and “high voter turnout”. It was hypothesized that urban states might have a higher voter turnout. However, since this data was binned at the median, the results just show that there was almost an equal number of rural states with high voter turnout as there were urban states with high voter turnout. While this result is trivial, it does demonstrate the utility of a chi-square test. I checked the modelling assumption, and the expected value of each cell was greater than five.

The chi2 test statistic was 0.0204 which is less than 3.84 and the p-value calculated was 0.999 (p>0.05) meaning that I cannot reject the null hypothesis in favor of the alternative.

The 95% confidence interval of proportion difference is: (-0.294, 0.254). Therefore, we concluded that the data provides evidence that the proportion of states with high voter turnout is not different than the proportion of rural states with high voter turnout (chi2(1, N=51). The observed proportion difference was -0.020 (95% CI = [-0.249,0.254]). Future tests should explore whether rural states really have lower turnout than urban states.



**Figure 4.** A mean and confidence interval were calculated in order to examine the difference between C02 emissions in high vs low management sector states. The null hypothesis is that the C02 in low management states is equal to the C02 level in high management states. The alternative hypothesis is that the C02 levels are not equal in high management vs low management states. I think the alternative hypothesis could go one of two ways. The first theory I have is that the C02 level in high management states could be higher because I would expect that it would be in line with trends where richer countries with service economies (more people in management) tend to produce more carbon emissions. However, it could also be that low management states produce more C02 emissions and that could be explained by environmental racism and the lack of infrastructure or jobs that might produce more carbon dioxide in lower management states vs jobs that produce less C02 emissions in higher management states.

The average C02 level of low management states in the United States was 29.144 metric tons per capita (95% CI = [21.320, 36.969]) and the average C02 level of high management states was 20.628 metric tons per capita (95% CI = [14.035, 27.220]). A two-sample t-test was conducted to compare the average C02 levels of low management vs high management states. There was not a significant difference in the C02 level of these states (p>0.05). I am perplexed because the two means lie outside each other’s confidence intervals, so I am surprised p>0.05. The average difference was 8.516 metric tons per capita (95% CI is: [-1.7146631963070025, 18.747718751862543]). With this large difference I would’ve expected to reject the null hypothesis. However, there may be a large variance as indicated by the wide confidence interval and this is probably the reason the null hypothesis could not be rejected.

Chart, scatter chart

Description automatically generated

**Figure 5.** Simple linear regression analysis was conducted to examine the relationship between poverty rate and energy consumption (BTUs per capita). My null hypothesis is that energy consumption is equal regardless of poverty rate. The alternative hypothesis is that energy consumption decreases with poverty rate. There was not a significant regression found at the p<0.05 level. The equation was:

Energy Consumption (BTUs per capita) = 3.734e+05 + poverty rate\*(856.8705)

Thus, the states’ energy consumption (per capita) increased by 856.870 BTUs per capita for every one percent increase in poverty rate (95% CI = -1.5e+04, 1.67e+04]). The correlation coefficient between the two variables was 0.015 which is very low (p>.05). This implies that, there is no correlation or relationship between state level poverty rate and energy consumption. The null hypothesis cannot be rejected. While my findings may be insignificant, this may be due to the tremendous outliers in a few of the states, particularly in the low poverty region. If outliers were removed, there might be a slight positive relationship.

Chart, histogram

Description automatically generated

**Figure 6.** I only included this extra visualization to demonstrate the use of a grouped histogram and it’s value. Rural states seem to be slightly skewed towards the low end of GDP because some of the rural states are lower than any of the urban states whereas, high GDP states may be skewed towards higher GDP as shown by the peak in urban states around 48000 and the one outlier urban state (District of Columbia). This relationship was important to examine for as an indicator of economic inequality between urban and rural states which effects the economic resources available for infrastructure spending.

In conclusion of this study, our statistical analysis of this very limited dataset provided some evidence that there is a correlation between income as an indicator of class and environmental and socioeconomic variables. In the future, I hope to use the skills developed in this course to analyze datasets which might explore this hypothesis in more depth and further the line of research regarding this topic.

Works Cited

Girgenti, Guido. *Winning the Green New Deal*. Simon & Schuster, 2020.

Gore, Tim. Stockholm Environment Institute (2020). *Confronting Carbon Inequality.* United Kingdom: Oxfam International. Accessed 23 Dec 2020.

United States Bureau of the Census. *Measure of America.* Provided in class Sep 2020.