The Effect of COVID-19 Lock Downs
on Air Quality Index
MAT 494

Daniel REES

November 15, 2021

Date Performed:	November 2021
Instructor:	Professor Wang

1 Introduction

The novel coronavirus spread throughout the United States during early 2020 and has now reshaped the living structure of many Americans. One of the most prevalent techniques used to prevent the spread of the virus are the international lock downs. The idea of a "lockdown" in the United States consisted of a federal mandate that required public and private sectors to reduce person-to-person contact in order to decrease ICU demand that overcrowded hospitals. Part of the lockdown mandate included an international travel ban as well as promoted at-home working. Inevitably, this created a decrease in transportation to and from work as well as the need for work-related trips. Ultimately, the reduction in fuel emissions during the lockdown was inevitable during these safety precautions. This could have a significant effect on the air quality index in cities with metropolitan life.

The integration of a stay at home ban has significantly decreased the amount of cars on the road which release nitrogen oxides, carbon monoxide as well as ozone. Eliminating these chemicals from the atmosphere would be beneficial

in preserving clean air for our community. The only question is how much? Climate activists and environmentalists emphasize the importance of living in a sustainable environment by recycling, driving electric vehicles, and avoid littering. However, the impact that these techniques have on the environment are faulty. In this experiment, Air Quality Index will be evaluated before and after lockdown to determine the stay at home order's effect on how pollution is altered by carbon emissions. It is anticipated that a reduction in emissions would conclude a decrease in Air Quality Index. In order to retrieve the most accurate data possible, COVID and air quality statistics were taken from Los Angeles and Phoenix due to their high traveling capacity.

2 Linear Regression

Regression analysis is a linearly classified evaluation of two variables (typically numerical). The objective of regression is to determine how two variables are correlated and to what extent. Linear Regression is a perfect example of outlining the cause and effect of an experiment. Similar to a linear model, regression analysis uses a dependent and independent variable in the form of an the equation:

$$y = mx + b$$

This equation demonstrates how the dependent variable(s) alters the independent one. A more detailed example of how the slope and intercept can be calculated from a data set is:

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

Although two variable may have a relationship, defining the strength of the relationship is important in determining how relevant the data is. The correlation coefficient finds how strong the independent variable is on the dependent variable. A negative coefficient represents negative correlation, and a positive coefficient represents a positive correlation. The closer the coefficient is to -1.00 or 1.00 the stronger it is. The equation to find the correlation coefficient is represented as:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}$$

Finding the correlation coefficient gives an immediate interpretation of how the linear relation will move in the future, regardless if the data is relevant or not. When testing for variable relevance in a model, the p-value will explain which variables are the most efficient in a model. Typically a p-value ¡ 0.05 proves significance in a model with a 95 percent confidence interval. This type of evaluation is useful when evaluating a multi-variable model in order to see which variables must be included to determine future values of the dependent variable.

In order to test the relevance of COVID cases to AQI, a combination of regression equations, correlation coefficients and scatter plots will help determine the relevance of the data set. Since a decrease in AQI is anticipated with an increase in COVID cases, a negative correlation coefficient should be expected.

3 Data

The data for Phoenix and Los Angeles AQI levels are reported from a series of state county and independent data collection sites. There are two types of AQI levels reported from different types of pollution. The first, comes from smaller particles is classified as particle matter 2.5 ($PM_{2.5}$). The 2.5 represents the size

of the air molecules diameter measured by microns. The smaller microns are more harmful to the ozone layer and are contributed by most human activity from fuel emissions to burning fire. The other particles are about four times as big in the form of particle matter 10 (PM_{10}). These particles measure 10 microns in diameter are produced through dust, pollen and mold. For the relevance of this experiment, particle matters with a 2.5 diameter will be recorded due to it's ability to account for transportation.

For COVID cases, the data recorded for both cities are reported through a series of state county and independent data collection sites. The numbers that are being evaluated are the new cases per day. This makes more sense as opposed to cumulative COVID data because the days with the highest new cases represents the peak of the spread. In order to produce the most accurate relationship between COVID cases and AQI levels, data from the beginning of March to the end of the year will be evaluated. The relevance of using only the first peak of COVID cases as opposed to the all peaks is because the extent of lock downs were more serious during the first outbreak as the next outbreaks occurred during a more lenient lockdown due to high vaccination status. Each day of new COVID cases reported is correlated to the average AQI of that same day.

4   Results

Refer to last two pages for all graphs and figures.

5   Conclusions

In figure 1, the scatter plot shows a downtrend that may not be heavily supported by a strong correlation. However, it is shown that after about 5000 cases

a day, the AQI does not seem to exceed a 75 AQI. Immediately, this demonstrates the downward trend that will later be classified through the regression report.

In figure 2, the scatter plot shows a pretty discrete trend that does not show any noticeable correlation at first glance. Opposite to the previous scatter plot, higher AQI levels are reported after 2000 cases which is contradictory to the original hypothesis.

Figure 3, is the regression analysis report of Los Angeles AQI levels. The report provides every number in determining the significance of the correlation. From the report, the equation can be written as:

$$y = -0.006\,x + 54.740$$

Where 'y' represents the AQI level and 'x' represents the number of new COVID cases that day. The equation states that for every 1 unit increase in COVID cases (1 new case), the AQI index will decrease by 0.006. The correlation coefficient reads 0.010 which states that 1 percent of the graphs movement can be explained by the number of COVID cases. Although this number is not very critical in determining the reason for AQI drop, it can be noted that there was a substantial decrease in the all time AQI highs which was most likely attributed to the lack of emissions during these lock downs. The report also provide a p-value of 0.081 which is greater than the 95 percent confidence interval of 0.05. However, with a 90 percent confidence interval, the dependent variable proves it's significance.

Figure 4 is the regression analysis report of Phoenix AQI levels. The equation for the AQI levels of Phoenix can be written as:

$$y = 0.004\,x + 30.785$$

Where 'y' represents the AQI level and 'x' represents the number of new COVID

cases that day. Contrary to the hypothesis, the equations states that for every unit increase in COVID cases, the AQI level will increase by 0.004. This completely rejects our hypothesis that AQI levels will decrease as COVID cases begin to rise. The correlation coefficient reads 0.081 which states that 8.1 percent of the variation of the graph is due to the number of COVID cases. The coefficient of the regression report shows that the COVID cases have more of an impact on Phoenix' positive correlation that Los Angeles' negative correlation. It is important to note the presence of out outliers in both of these graph. Los Angeles would test all time AQI highs before lockdown as Phoenix reported all time high before, during and after lockdown. Evaluating the extent of how these cities differ in correlation is important in determining why the experiment contradicts the initial hypothesis.

6   Discussion

The results of the experiment showed partial support of the original hypothesis. Los Angeles demonstrated a decrease in AQI levels while Phoenix showed the opposite. By examining potentials factors that could alter the outcomes of this experiment, the extent of lockdown mandates comes into question. Los Angeles has a higher population that produced higher positive test counts. The probability that population congestion created a more vivid response to COVID lockdown could be a potential factor. It is also notable to add that towards the end of the year 2020, Arizona held the highest infected rate per capita in the US for about a 7 day period. Lock downs were still in place however many of them had been modified to promote personal contact in some essential business. These could all have an impact on the overall potential of these graphs.

References

[1] County of Los Angeles Public Health, 2021, Cumulative and Daily Cases and Deaths by Date [Data File], Retrieved from $http$: $//dashboard.publichealth.lacounty.gov/covid19_surveillance_dashboard/$"

[2] Enigma Forensics, 2021, Arizona Coronavirus COVID-19 Statistics [Data File], Retrieved from $https : //enigmaforensics.com/coronavirus-covid-19-data-tracking/$

[3] United States Environmental Protection Agency, 2021, Daily AQI by County [Data File], Retrieved from $https$: $//aqs.epa.gov/aqsweb/airdata/download_files.htmlAQI$
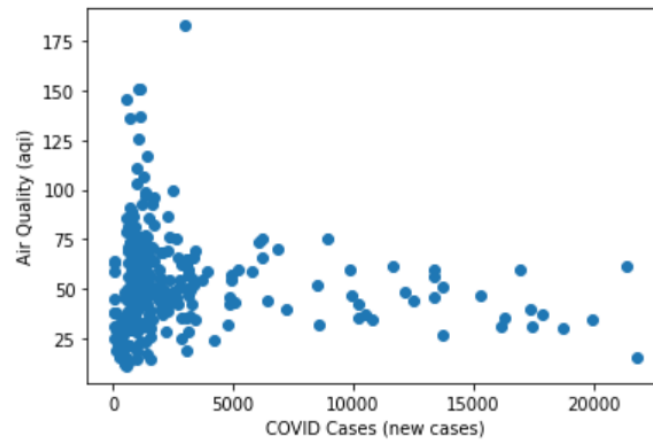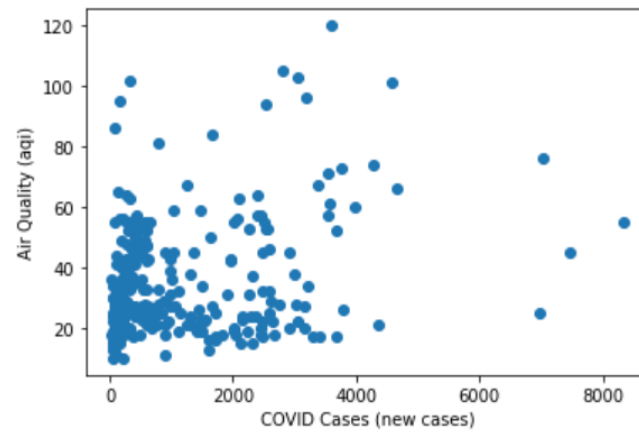
Figure 1: Los Angeles Scatterplot

Figure 2: Phoenix Scatterplot

```
                        OLS Regression Results
==============================================================================
Dep. Variable:                     aqi   R-squared:                       0.010
Model:                             OLS   Adj. R-squared:                  0.007
Method:                  Least Squares   F-statistic:                     3.069
Date:                 Sun, 14 Nov 2021   Prob (F-statistic):             0.0808
Time:                         20:58:40   Log-Likelihood:                -1408.9
No. Observations:                  306   AIC:                             2822.
Df Residuals:                      304   BIC:                             2829.
Df Model:                            1
Covariance Type:             nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          54.7398      1.677     32.645      0.000      51.440      58.039
cases          -0.0006      0.000     -1.752      0.081      -0.001    7.58e-05
==============================================================================
Omnibus:                       103.058   Durbin-Watson:                   0.412
Prob(Omnibus):                   0.000   Jarque-Bera (JB):              366.005
Skew:                            1.440   Prob(JB):                      3.33e-80
Kurtosis:                        7.517   Cond. No.                      5.78e+03
==============================================================================
```

Figure 3: Los Angeles Regression Report

```
                        OLS Regression Results
==============================================================================
Dep. Variable:                     aqi   R-squared:                       0.081
Model:                             OLS   Adj. R-squared:                  0.078
Method:                  Least Squares   F-statistic:                     24.81
Date:                 Sun, 14 Nov 2021   Prob (F-statistic):           1.11e-06
Time:                         21:18:32   Log-Likelihood:                -1223.1
No. Observations:                  283   AIC:                             2450.
Df Residuals:                      281   BIC:                             2458.
Df Model:                            1
Covariance Type:             nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          30.7851      1.419     21.696      0.000      27.992      33.578
cases           0.0040      0.001      4.980      0.000       0.002       0.006
==============================================================================
Omnibus:                        71.512   Durbin-Watson:                   0.624
Prob(Omnibus):                   0.000   Jarque-Bera (JB):              141.993
Skew:                            1.295   Prob(JB):                      1.47e-31
Kurtosis:                        5.309   Cond. No.                      2.29e+03
==============================================================================
```

Figure 4: Phoenix Regression Report