PROJECT: Lagged Cross-Correlation Analysis of Weekly COVID-19 Cases and Deaths in LA

DATE: 07-16-2023

TO: Dr. Blanchard, Supervisor

FROM: Sushovan Adhikari, Data Analyst/Scientist

RE: Report on Cross-Correlation Analysis of COVID-19 Data

Introduction:

This report presents the results of an analysis conducted on the first three waves of COVID-19 data, focusing on weekly-averaged cases and deaths from the disease. The primary objective of this analysis was to examine the lagged cross-correlations between cases and deaths, providing insights into the relationships between these variables over time. The analysis involved investigating the lag of peak correlation for each wave. The findings indicate that for the first wave, deaths lagged behind cases by approximately seven weeks (-7). In the second wave, the lag was found to be -15 weeks, while in the third wave, it was -9 weeks.

Furthermore, this analysis allowed for the quantification of the deaths/case ratio during each wave. The results revealed slope values of 0.038 (with a standard error of 0.004) for the first wave, 0.0146 (with a standard error of 0.00048) for the second wave, and 0.0128 (with a standard error of 0.00055) for the third wave.

Experimental:

Data for this analysis was obtained from the State of Louisiana's public data repository, consisting of three CSV files. These files contained weekly average cases and death cases for the first, second, and third waves of COVID-19.

To explore the time-based relationship between weekly average cases and deaths, a cross-correlation analysis was conducted which is represented by the following equation:

$$R_{xy}[k] = \sum_{n=-\infty}^{\infty} x[n]y[n-k] \quad \text{………..} \quad (1)$$

Where k is the time lag applied to signal y, and $R_{xy}[k]$ is the cross-correlation between x and y at lag k.

The correlate() function from the SciPy.Signal module was utilized to calculate the cross-correlation values $R_{xy}[k]$ between the two data vectors: weekly average cases (x) and weekly average deaths (y). The lag of peak correlation was determined by identifying the index of the maximum value in the resulting correlation vector using the argmax() function from NumPy.

In addition to the cross-correlation analysis, linear regression was performed to investigate the association between the shifted average weekly cases (based on the peak lag) and average weekly death cases. The lingress module from the SciPy library was employed for this purpose. The slope value obtained from linear regression provided a quantification of the deaths/case ratio.

Results:

The results can be summarized as follows: Figures 1, 2, and 3 display the plots of weekly average cases and average weekly deaths for the first, second, and third waves, respectively. The weekly average cases have been shifted according to the determined lag value. Linear regression analyses were performed to examine the relationship between the shifted weekly average cases and average weekly deaths for each wave. Figures 4, 5, and 6 show the corresponding linear regression plots for the first, second, and third waves.

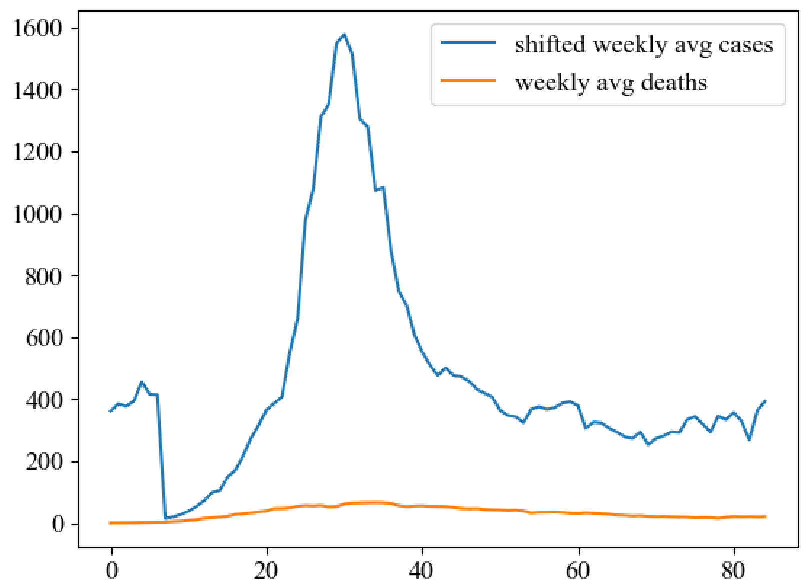

Figure 1: Plot of Weekly Avg Cases (shifted based on lag) and Avg Weekly Death Cases for First Wave.

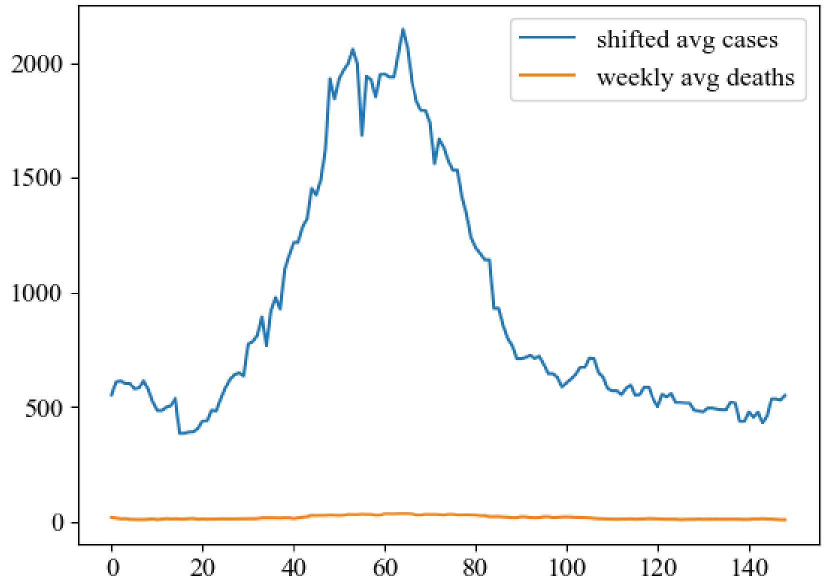

Figure 2:Plot of Weekly Avg Cases (shifted based on lag) and Avg Weekly Death Cases for Second Wave.

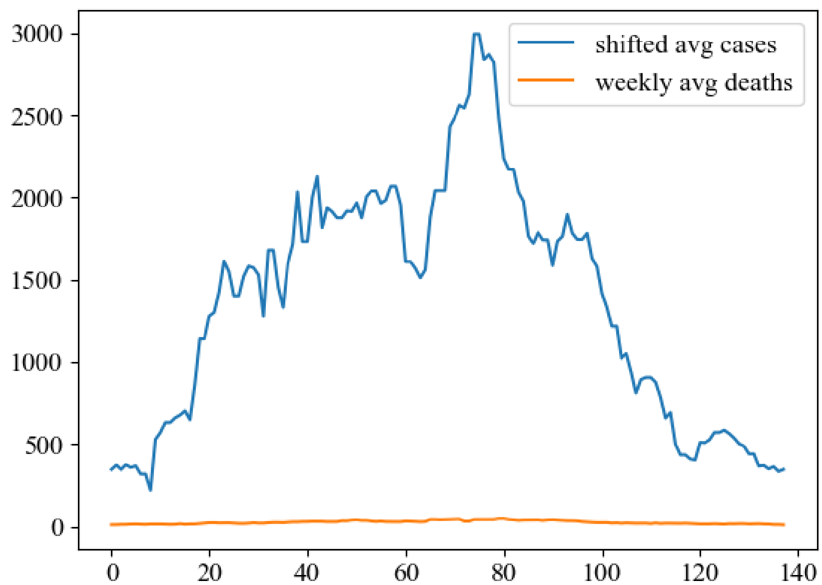

Figure 3:Plot of Weekly Avg Cases (shifted based on lag) and Avg Weekly Death Cases for Third Wave.

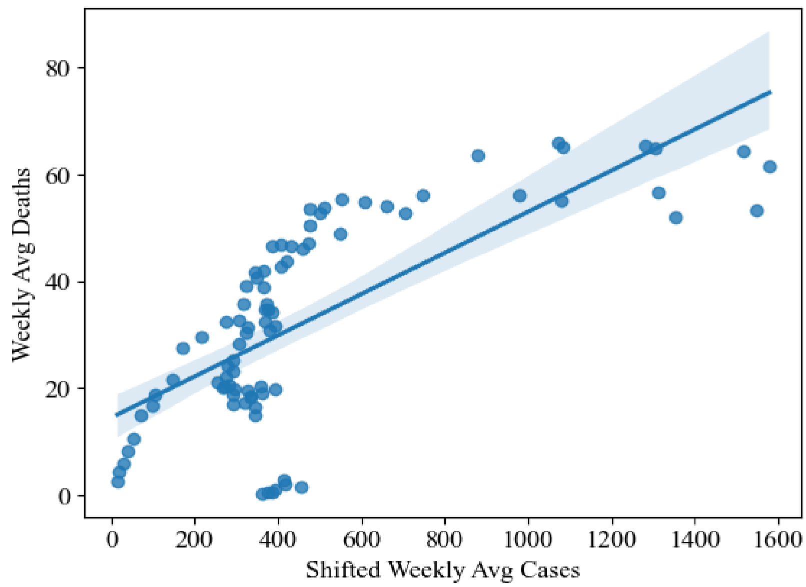

Figure 4:Linear Regression of Shifted Weekly Avg Cases vs Deaths for First Wave.

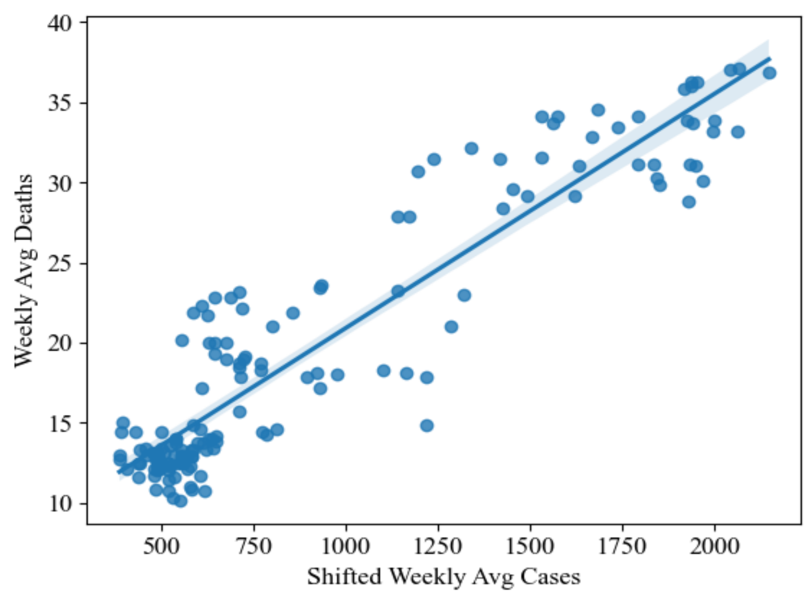

Figure 5:Linear Regression of Shifted Weekly Avg Cases vs Deaths for Second Wave.

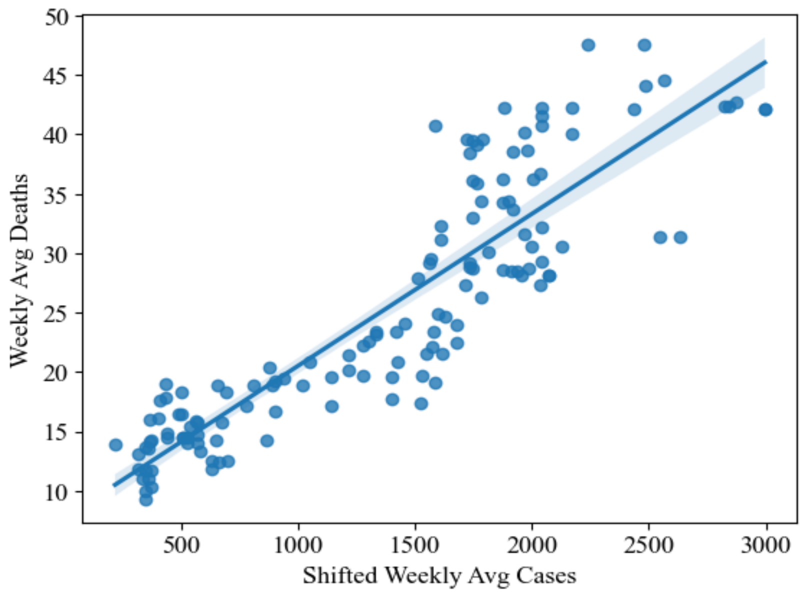

Figure 6:Linear Regression of Shifted Weekly Avg Cases vs Deaths for Third Wave.

Discussion:

The analysis of the first three waves of COVID-19 data revealed significant insights into the relationship between average weekly cases and deaths, as well as the timing and deaths/case ratio for each wave. Figure 1, 2, and 3 plots represent the adjusted data, accounting for the lag present in the time series. These plots provide visual representations of the relationship between average weekly cases and deaths for the first, second, and third waves, respectively. By examining the plots, we observed variations in the lag between increases in cases and subsequent rises in deaths for each wave. Specifically, the first wave showed a lag of approximately seven weeks (-7), the second wave had a longer lag of around fifteen weeks (-15), and the third wave exhibited a shorter lag of about nine weeks (-9).

To further understand the impact of the disease, linear regression analysis was conducted. Figure 4, 5, and 6 present the linear regression plots, illustrating the relationship between the shifted average weekly cases and average weekly deaths for the first, second, and third waves, respectively. The results of the linear regression analysis yielded slope values that quantify the deaths/case ratio for each wave. In the first wave, the slope was found to be 0.038 with a standard error of 0.004. For the second wave, the slope was 0.0146 with a standard error of 0.00048. Lastly, the third wave exhibited a slope of 0.0128 with a standard error of 0.00055. The first wave had the highest slope of 0.038, indicating a relatively higher mortality rate per case. The second wave had a lower slope of 0.0146, and the third wave had an even lower slope of 0.0128. This decreasing trend suggests a potential improvement in the management of cases and healthcare response, resulting in a lower mortality rate in the subsequent waves.

Conclusions:

In conclusion, the analysis of the first three waves of COVID-19 data has provided valuable insights into the relationship between average weekly cases and deaths, as well as the timing and deaths/case ratio for each wave. The adjusted plots and examination of the lag between increases in cases and subsequent rises in deaths revealed variations in the durations between these events across the waves. The linear regression analysis quantified the deaths/case ratio, demonstrating a decreasing trend in mortality rates per case from the first wave to the subsequent waves.

The insights gained from this analysis contribute to our understanding of the dynamics of COVID-19 during the examined waves and provide valuable information for guiding public health interventions and strategies. They underscore the significance of timely responses, effective healthcare management, and continued vigilance in reducing mortality rates and limiting the impact of the disease.