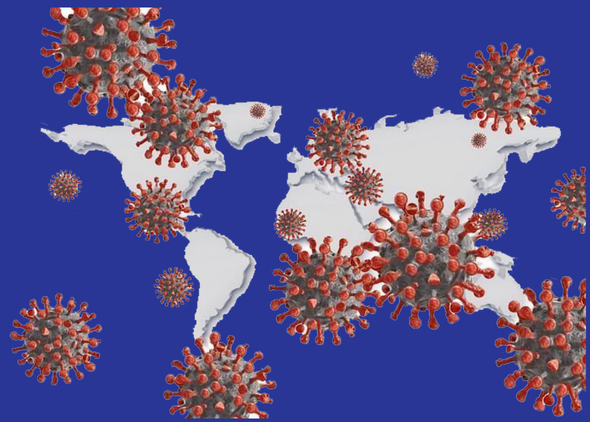
Wilson Consulting Services, LLC

COVID-19: The Correlation Between Positive Tested Cases and Deaths



July 19, 2020

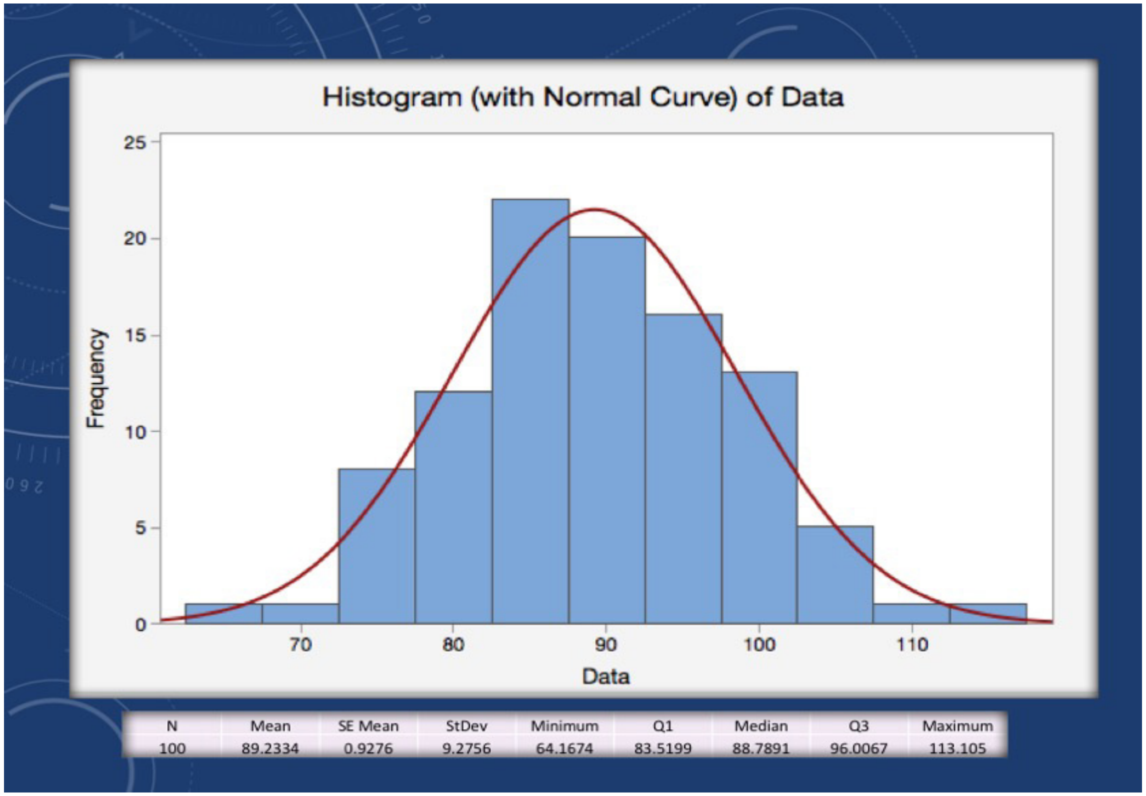
David C. Wilson
Founder / CEO

BBB ACCREDITED BUSINESS

wilsonconsultingservices.net

It must be demonstrated . . .

We are proponents and advocates of literacy in STEM and statistics in a technological and data-driven world.



STEM =∫(Science, Technology, Engineering, Mathematics)

Table of Contents

Note: A copy of this report can be open and downloaded from the link show below.
https://www.wilsonconsultingservices.net/wcs_covid-19_analysis_20.pdf

Part I. Description

Introduction

There has been much discussion regarding whether the increase in positive COVID-19 cases is due to an increase in overall testing or whether the pandemic is truly getting worse, as evidenced by the rising number of positive cases and deaths. The graphs in this paper do not show snapshot pictures of a given day or week of the pandemic; they show a long-term, cumulative view of the virus's impact on countries, states, and some counties. Therefore, this paper does not detail the day-to-day or week-to-week counts of positive cases and deaths from the virus. Instead, the paper provides a long-term view of the pattern of positive cases relative to deaths and shows the virus's impact on various locales. To that end, I have examined thoroughly the cumulative number of positive test results relative to deaths.

Process

In the process of examining the trend in COVID-19-positive cases versus deaths, I determined the best statistical model fitting a long-term examination of the data was the regression analysis model. Regression analysis is a set of statistical processes used to estimate the relationships between a dependent variable (often called the outcome variable) and one or more independent variables (often called predictors, covariates, or features). In this paper, the number of positive COVID-19 cases is the independent variable (horizontal: *x*-axis), and the number of COVID-19 deaths is the dependent variable (vertical: *y*-axis). Choosing the regression model for analysis accomplishes several things: (1) it explains a phenomenon, (2) it predicts factors about the future, or (3) it enables decision-making. To that end, the use of the regression model in this report should be interpreted only as a way of examining the current phenomenon relative to the virus.

I used regression analysis (simple linear regression) to analyze the cumulative positive test cases for 158 countries, the 50 states in the U.S. and the District of Columbia, and all South Carolina counties.

There are large variations in COVID-19-positive cases and deaths across the globe, within the United States, and in South Carolina counties. Therefore, I converted the data for the two sets of variables from base-10 data to logarithmic data to reduce the wide range of quantities to smaller intervals, which allowed me to better discern changes in the graphs showing positive cases and deaths. Going to a logarithmic scale also helped with the nonlinearity of data. (Nonlinearity means the data points do not line up in an approximate straight line fitted through the dots shown on the graph.)

I constructed the graphs in Figures 1, 2, and 3 with the intent of showing how cumulative COVID-19-positive cases and deaths are related, regardless of the number of recovered, active, or inactive cases. Consequently, the cumulative cases and deaths demonstrate the overall long-term effect coronavirus is having across the world and in individual countries, states, and counties. In this report, the counties in South Carolina are shown in Figure 3. All graphs in this report provide a so-called "bird-eye view" of the coronavirus's effects.

Analysis

The red line is the best-fit line, and the blue and red dots each represent points for the number of intersecting cases (*x*-axis and *y*-axis). The name of the country, state, or county is listed near the red dots. The blue dots do not show a name because of lack of

Part I. Description, continued

(Analysis, continued)

space. The name of data points is not necessarily labeled because the purpose of a regression analysis model graph is to show the pattern or proportionality between two variables. For example, 158 countries are represented in the graph in Figure 1, but not all are listed by name because it would be impossible to list all 158 countries corresponding to each dot shown on the graph. A linear equation representing the best fit line is shown on the graph, which is the result of applying regression analysis to the data. The equation with a positive slope shows a correlation between the number of COVID-19-positive cases and deaths. As observed in all three graphs, the logarithmic scale shows a pronounced depiction of the pandemic's behavior without having to use large numbers (such as comparing 10 with 100,000 on a base-10 scale). However, going through the conversion back to base-10 computation is complex; therefore, explaining the mathematical manipulation used in the process is beyond the scope of this paper.* For example, most people remember the equation for a straight line as $y = b + mx$. The graph in Figure 2 shows the equation for the red line (regression line), which is the output from the data: $y = -2.104 + 1.123x$. This equation was constructed from regression modeling of COVID-19 data from the 50 states plus the District of Columbia. The greater the slope ($m = 1.123$) is as shown in the equation, the stronger the relationship is between positive cases and deaths. A zero slope—a flat regression line—means there is no relationship (correlation) between cases and deaths. If the slope becomes a negative number (negative correlation), this means as positive cases increase, COVID-19 deaths will plummet toward zero relative to positive cases. Furthermore, when the information in this equation (Figure 2) is converted back to base-10, the results show if one were to randomly select 30 people across the U.S. who tested positive, there is a probability that one or two will have died or will die from COVID-19. Using the numbers 1 or 2 signifies the mean death rate is 1.000 from a sample of 30 with a 95% confidence interval (CI) with small upper and lower control limits: hence, 95% CI 1.000 [0.796, 1.265]. Because many deaths were clustered among older people in the U.S., including nursing homes, the probability might deviate a bit from the projection provided by the equation. The reader should keep in mind that the equations shown on all three graphs reflect averages, and they likewise suffer from the same limitations inherent in all mathematical modeling.

The COVID-19-positive cases (x-axis) are read horizontally from left to right beginning at the origin, and the deaths (y-axis) are read vertically (up from zero at the origin). Red and blue dots are shown at the intersection of the number of positive cases and deaths. The dots show a positive slope as the number of positive cases increases proportionally, which demonstrates a strong relationship between these two COVID-19-associated entities.

The R-squared is another component of the regression analysis model (see Figures 1, 2, and 3). Therefore, analyzing the relationship between COVID-19 positive-tested cases and death used the R-squared values shown on each graph from the regression analysis testing. The p-value, not shown here, but was included in the computer printout.* Consequently, an analysis of the R-squared value allowed an investigation of the strength between positive-tested cases and deaths without having to view the graphs. Hence, the results are discussed graphically and analytically in this report. The coefficient of determination known as R-squared (R^2), which measures the strength

*If you are interested in how the conversion works for the equation or any other concerns, please contact the author at dave@wilsonconsultingservices.net

Part I. Description, continued

(Analysis, continued)

in percentage between the variables, is a statistical measure of how close the data are to the fitted regression line. If $R^2 = 0$, it suggests there is no relationship between the COVID-19 positive tested cases and the deaths. If $R^2 \neq 0$, it suggests there is a relationship between the positive tested cases and deaths. A regression analysis evaluates the R-squared value to determine its statistical significance. The graphs and their accompanying description depict the following: (1) a positive sloped line, (2) $R^2 \neq 0$ (strength of relationship), and (3) the p-value (p). The parameters that helped determine the significance of the results in proving a strong relationship between the COVID-19 positive tests and COVID-19 deaths are the following: (1) graph: straight line sloping upward; (2) R-squared: $0 \leq R^2 \leq 1$; and (3) p-value: $0 \leq p \leq 1$. A small $p < 0.05$ indicates strong evidence against the null hypothesis of no slope; therefore, one would reject the null hypothesis of no relationship between COVID-19 positive tested cases and deaths. However, Figures 1, 2, and 3 show R-squared values and *p*-values met the criteria outlined above with values significantly greater than zero and *p* significantly less than 1. The findings support scientifically that there is correlation between COVID positive tested cases and COVID deaths in the long term.

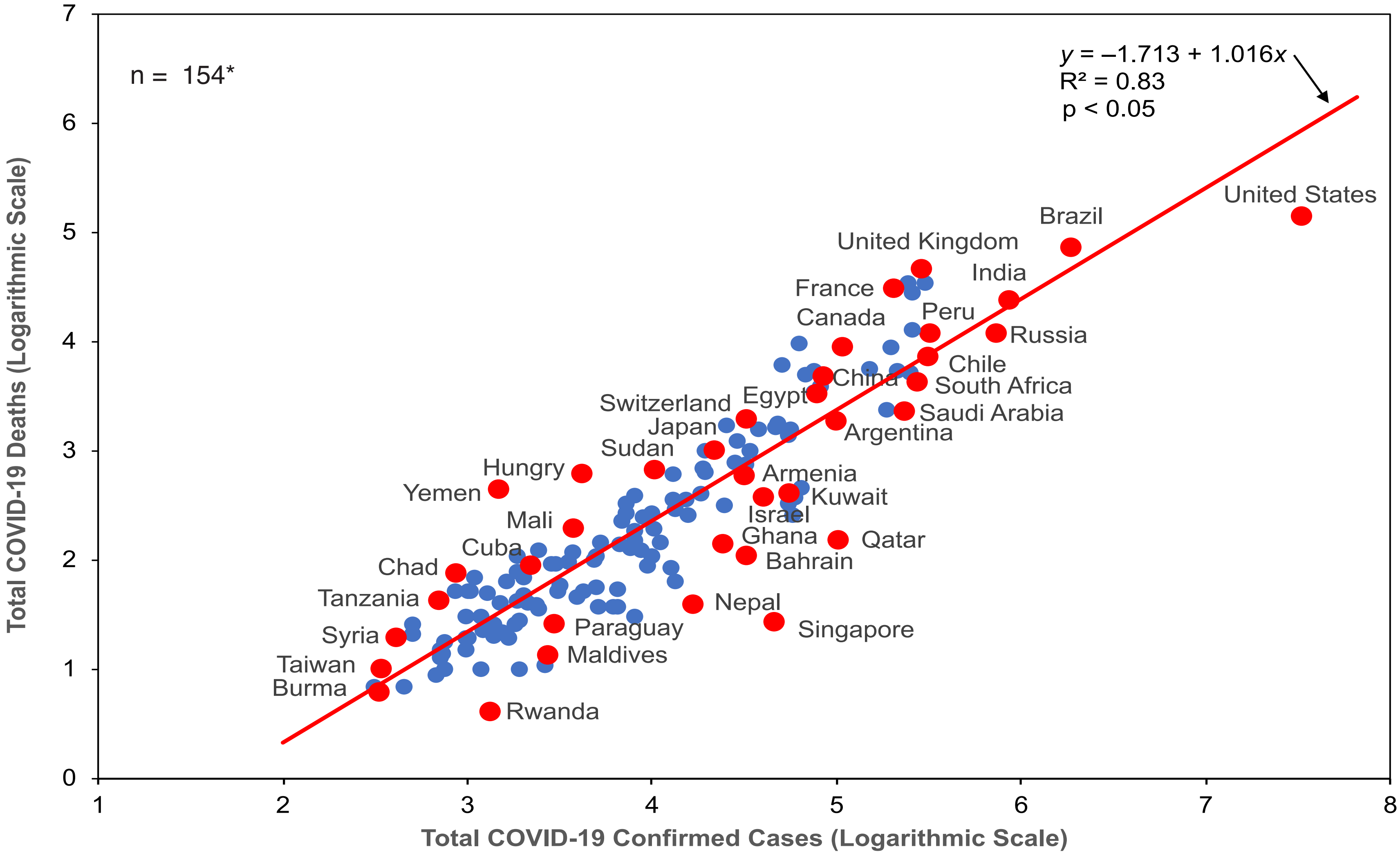
Conclusion

I am a statistical professional, not a medical professional—therefore, my analysis is limited to using data from the medical community to show statistical patterns of the pandemic. More than likely, there are some locales where the number of COVID-19 deaths is decreasing, or remaining unchanged. The model used for this analysis does not respond to a few changes in various locales but shows collectively the impact positive cases and deaths are having globally, within the U.S. and the District of Columbia, and in South Carolina counties. An ideal situation will occur when the regression (straight line) shown on the graph becomes flat and ultimately tips downward (negative slope). We can speculate that a dramatic decrease in deaths means people are not becoming fatally ill from the coronavirus. ■

Part II. Graphs

Worldwide by Country—Regression Analysis Plot for COVID-19

Figure 1: Total COVID-19 Confirmed Cases versus Total COVID-19 Deaths



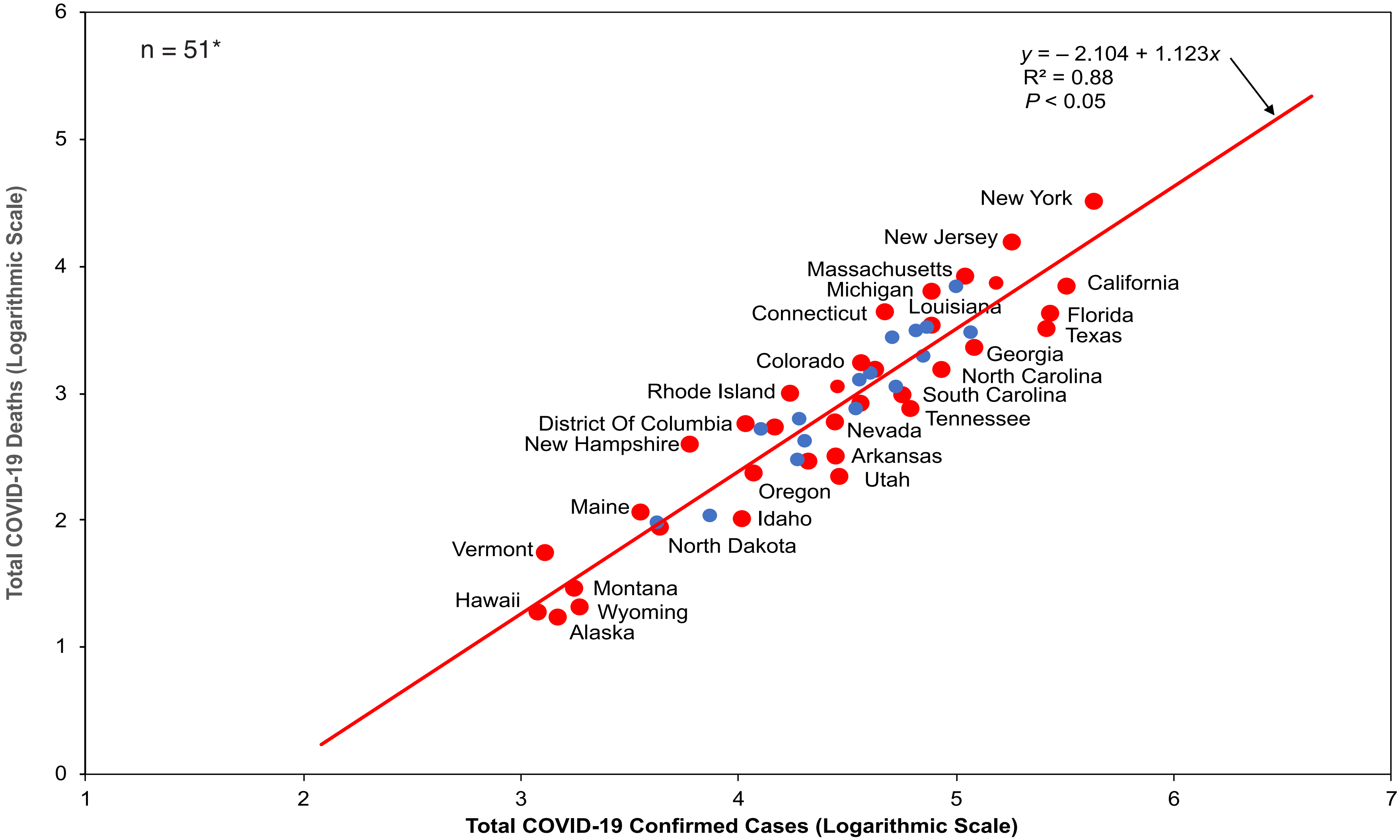
Source: https://www.worldometers.info

*Countries with less than four deaths are not shown in the analysis. The omission does not make any difference because the regression model is an Inferential statistics component, which is based on sampling distribution; therefore, the entire population is not needed nor necessary.

Part II. Graphs

United States by State—Regression Analysis Plot for COVID-19

Figure 2: Total COVID-19 Confirmed Cases versus Total COVID-19 Deaths

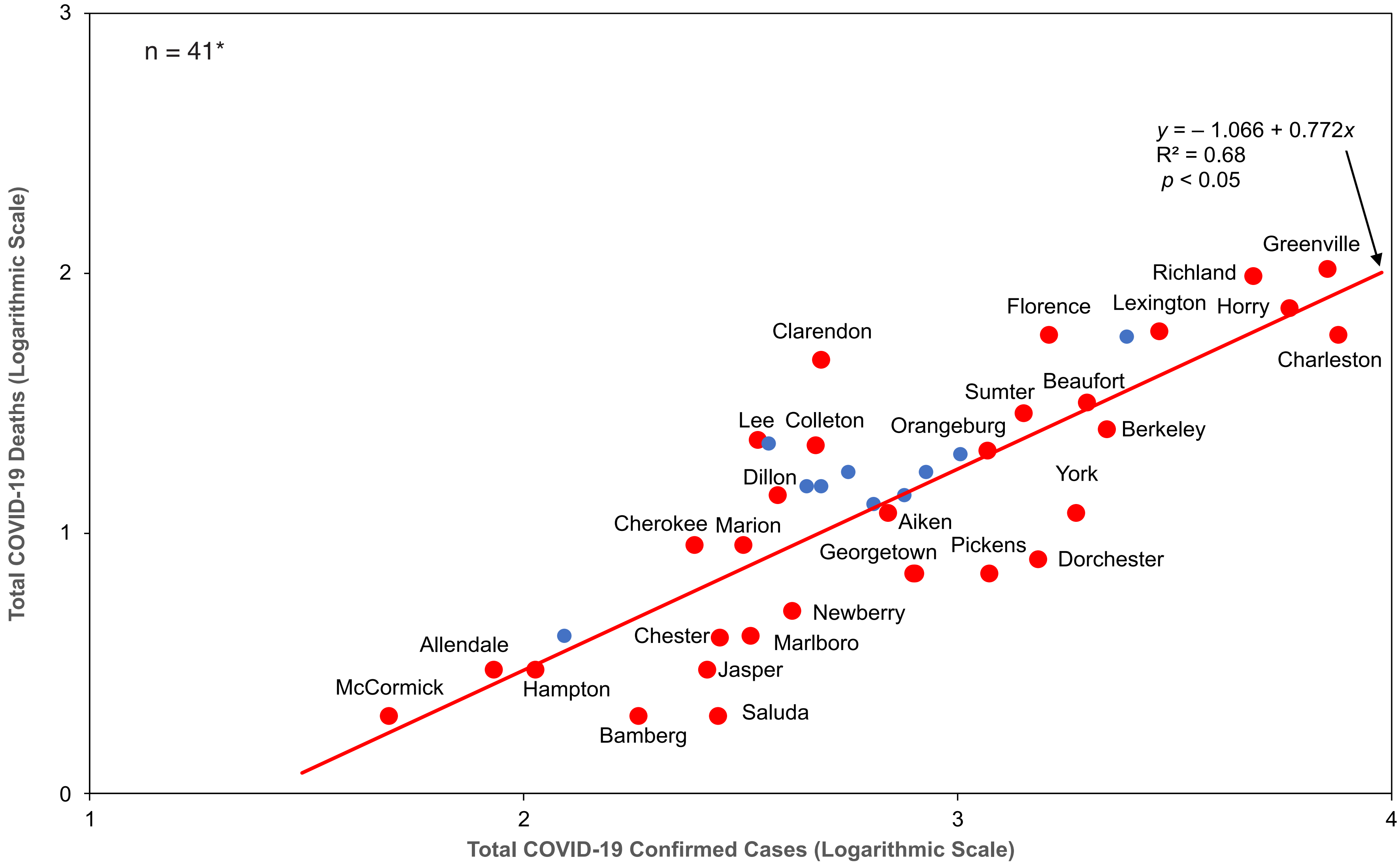


Source: https://www.worldometers.info

*Fifty states, plus the District of Columbia

Part II. Graphs
South Carolina by County—Regression Analysis Plot for COVID-19

Figure 3: Total COVID-19 Confirmed Cases versus Total COVID-19 Deaths



Source: South Carolina Department of Health and Environmental Control

*Counties with less than two deaths are not shown in the analysis. The omission does not make any difference because the regression model is an Inferential statistics component, which is based on sampling distribution; therefore, the entire population is not needed nor necessary.

References

Worldometer
https://www.worldometers.info

South Carolina Department of Health and Environmental Control
https://www.scdhec.gov

About the Author

David C. Wilson is a STEM professional as well as a mathematics and statistics adjunct professor—now retired. He is a statistical consultant, local historian, author, and self-publisher.

Dave attended the public schools of Horry County and is a graduate of the former Chestnut Consolidated High School. After serving three years in the United States Army, he earned his undergraduate and graduate degrees in electrical engineering from the City College of New York and Manhattan College, respectively.

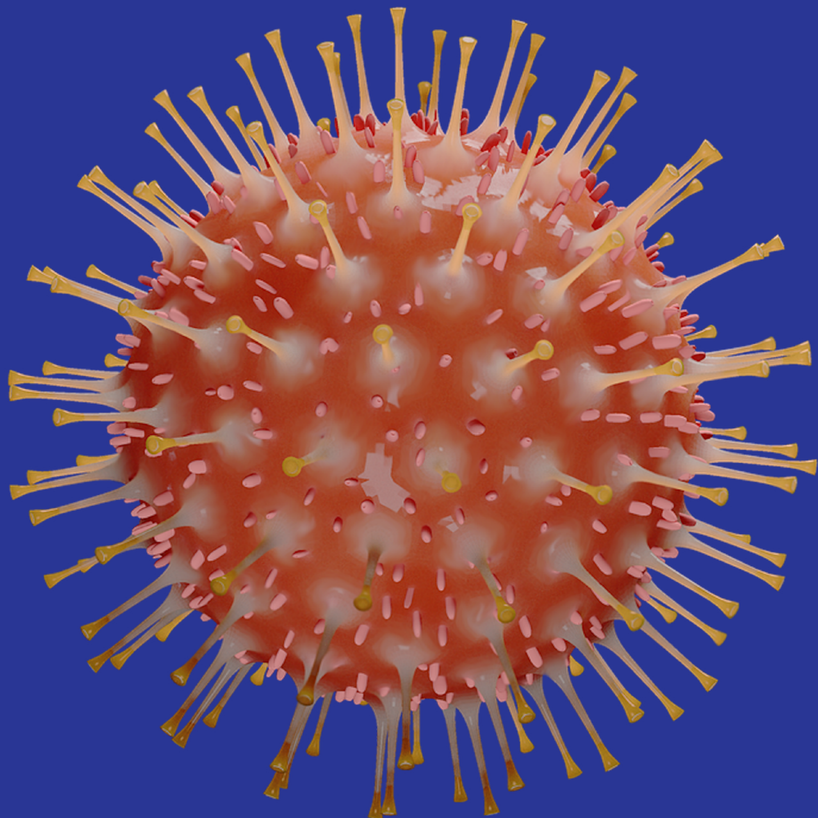
Wilson has worked for IBM, General Electric, and Honeywell. He also taught as an adjunct professor at Dutchess Community College, Quinnipiac University, and Horry Georgetown Technical College. Additionally, he served one year with the prestigious IBM Faculty Loan Program.

His experience spans 30 plus years in engineering and 26 years as an adjunct professor. Additionally, Wilson is a former reader for the College Board/AP Statistics Program and program evaluator for the Accreditation Board for Engineering and Technology/Technology Accreditation Commission (ABET/TAC). Wilson is currently a senior member of the American Statistical Association (ASA), the American Society for Quality (ASQ), and the Institute of Electrical and Electronics Engineers (IEEE). He has received numerous awards for excellence in his professional work such as IBM Division Award and civic work such as The Horry County Board of Architectural Review & Historic Preservation Award.

David C. Wilson
Founder/CEO

He and his wife, Beverly, have two adult sons and six grandchildren. They reside in Conway, South Carolina.

We are in this pandemic together!



wilsonconsultingservices.net

It must be demonstrated . . .