The death toll of COVID-19 in the United States has exceeded 107,000 as of early June 2020. While the total number is a grim milestone, it is disparately spread across the United States. Some regions, like San Luis Obispo, California, with a population of 283,111, experienced only one death as of June 3—a death rate of 0.0035 per 1,000 persons—while others, like New York, New York, experienced a death rate of 13.31 per 1,000 persons. This essay quantifies the dispersion in death rates across the United States using county-level data and the Theil index. The Theil index is an “inequality” measure similar to the Gini coefficient, but it is more useful for decompositions into subcategories. For instance, we can measure the overall dispersion in death rates in the United States and then compute whether specific subgroups of counties contributed substantially to the overall dispersion.

In the following equation, let \( N \) denote the number of counties and \( x_i \) denote the death rate—number of deaths per 1,000 persons—in county \( i \). Then, the Theil index \( T \) that captures the dispersion in the death rate across \( N \) counties is

\[
T = \frac{1}{N} \sum_{i=1}^{N} \frac{x_i}{\mu} \ln \left( \frac{x_i}{\mu} \right),
\]

where \( \mu \) is the average death rate across \( N \) counties. If all counties had the same death rate—that is, if there were no dispersion in the death rate across counties—the Theil index would be zero. In our sample, there are 1,753 counties with a total population of 294 million.\(^1\) The average death rate \( \mu \) is 0.24 per 1,000 persons, and the Theil index \( T \) is 0.754.

To decompose the Theil index, suppose we divide the population into deciles so that there are 10 equal-sized groups, each with 10 percent of the population. We rank the counties by population from the least populated (decile 1) to the most populated (decile 10). Of course, the number of counties in each decile will be different since the population in some counties is larger than that in others. (See Geographic Disparity in the U.S. Population.) Using the same formula as above, the Theil index for decile \( j \) would be

\[
T^j = \frac{1}{N^j} \sum_{i=1}^{N^j} \frac{x^j_i}{\mu^j} \ln \left( \frac{x^j_i}{\mu^j} \right),
\]

where \( \mu^j \) is the average death rate in decile \( j \), \( x^j_i \) is the death rate in county \( i \) in decile \( j \), and \( N^j \) is the number of counties in decile \( j \). Figure 1 illustrates the average death rate and the number of counties in each decile. Decile 10 has only six counties, while decile 1 has 998 counties. The average death rate ranges from 0.15 for decile 2, which is below that of the United States, to 1.24 for decile 9, more than five times that of the United States.
The specific decomposition we compute is how much of the overall dispersion is due to dispersion across the 10 deciles and how much of it is due to dispersion within each of the 10 deciles. To this end, we need to calculate how much weight to place on each decile. For instance, some deciles with many counties might account for a larger proportion of the overall dispersion than others that have few counties. The weight on decile \( j \) in our calculation is

\[
s^j = \left( \frac{N^j}{N} \right) \left( \frac{\mu^j}{\mu} \right).
\]

The weight on decile \( j \) would be low if it has few counties and its mean death rate is well below the average of the United States; and, hence, its contribution to the overall dispersion would be low. The overall dispersion can then be decomposed as

\[
T = \sum_{j=1}^{10} s^j T^j + \sum_{j=1}^{10} s^j \ln \left( \frac{\mu^j}{\mu} \right).
\]

Our calculations suggest that most of the overall dispersion is not from across the 10 deciles, but from within the deciles. The within-group dispersion is 0.683, more than 90 percent of the overall dispersion. Furthermore, the within-group dispersion in decile 1 is 0.38, accounting for 50 percent of the overall dispersion.

What accounts for the dispersion in decile 1? We consider a few variables based on susceptible members of the population. In Figure 2, we illustrate three variables: share of the population 60 years of age and older, share of males, and share of African Americans in each of the 998 counties in decile 1. We plot the median, the 90th percentile, and the 10th percentile values for each of these three variables. As the figure illustrates, the key variable is the share of African Americans across these 998 counties: The dispersion in that variable is substantially larger than that in the share of persons 60 years of age and older or in the share of males. The median share of African Americans in decile 1 is almost 4 percent, but the bottom 10 percent of the counties have a share of only 0.6 percent, while the top 10 percent of the counties have a share of more than 44 percent.

The death toll of COVID-19 is disparately spread across the United States, and demographic factors have likely affected this. Specifically, geographic dispersion in the share of African Americans likely accounts for the uneven distribution of COVID-19-related deaths across the United States.

Note

1 The total number of counties in the United States is more than 3,000, but we restrict our sample to counties that had at least one COVID-19 death as of June 3.