Problem: Focusing on an average doesn't tell the whole story

Researchers often use averages to describe data. The average (or the mean) of a data set can be used to identify the central value of the group, or what is typical. While valuable, it’s also important to understand the range of data—the highs and lows. What might we miss by focusing on the average?

When considering averages, important questions to ask are: Are the data distributed normally creating a bell-shaped curve? Are the data skewed to one side leaving a tail at either end? When the data are skewed, the average is pulled to one side and is no longer located in the center; thus, the average would not be an appropriate representation of the typical value in the group. Even with normal distribution, considering the range of data values is imperative.

This short brief uses three examples to demonstrate how focusing on averages without also considering the data range can conceal important information: 1) average rural hospital profitability, 2) distance from closed rural hospitals to the next closest hospital, and 3) HIV prevalence by county. The examples look at conclusions that might be drawn from commonly used indicators.

Examples of information that might be missed and how

Example #1: “The average rural hospital was profitable in 2015.” A common measure of profitability is total margin, defined as net income divided by total revenue and reflected as a percentage. If total margin is positive, this indicates that revenue is greater than expenses, and the organization is profitable. In contrast, if total margin is negative, the organization is unprofitable. Profitability is of obvious importance as hospitals need to maintain the necessary infrastructure (building, equipment, staff, technology) to provide services to the population.

![Figure 1: 2015 total margin for all 2,258 rural hospitals](image)
Figure 1 shows the 2015 total margin for every rural hospital in the country (2,258 hospitals). In 2015, the average total margin for rural hospitals was 1.8% indicating that the average rural hospital was profitable in 2015. However, 849 rural hospitals (38%) were unprofitable in 2015 (they had negative total margins as shown by the yellow bars). The average total margin is positive 1.8% because the number of profitable hospitals (1,409) is greater than the number of unprofitable hospitals, but there is still a large number of unprofitable hospitals. Although 1.8% is a relatively small number and may suggest a narrow range of profitability, in reality, some rural hospitals were very unprofitable (-60%) and some were very profitable (+60%).

While it is true that the average rural hospital was profitable in 2015, it would be wrong to conclude that profitability is not a problem for rural hospitals considering a substantial proportion of rural hospitals were unprofitable. It would also be erroneous to conclude that all rural hospitals are barely profitable—a considerable proportion of rural hospitals had substantial total margins (both negative and positive).

Example #2: “When a rural hospital closes, patients must travel an average of 12.5 miles to the next closest hospital.”

Between 2005-2017, 125 rural hospitals closed. When a rural hospital closes in a community, residents must travel to a hospital in another community for inpatient and other types of health care. The additional travel burden is of concern because residents of rural communities are typically older and poorer, more dependent on public insurance programs, and in worse health than urban residents. Geographic distances between rural patients and hospitals may limit access to health care; thus, when policy makers consider access to health care, distance to a hospital is an important policy consideration.

Figure 2 shows the distance from each of the closed rural hospitals to the next closest hospitals. This is a rough estimate of the average additional travel distance to the next hospital for people who live in each community where a rural hospital closed. The graph shows that the average distance to the next hospital is 12.5 miles. A closer look at the data shows that 43% of the closed hospitals are more than 15 miles to the nearest hospital and 15% are more than 20 miles.

Therefore, while it is true that when a rural hospital closes, patients must travel an average of 12.5 miles to the next closest hospital, people in 76 of the 125 communities with closed hospitals are traveling farther than the average. Even if distance to the next nearest hospital was the only downside for a community losing its hospital, it would be wrong to conclude that closure of a rural hospital is not an access problem for many rural residents.
Example #3: “The HIV prevalence rate for Virginia is lower than South Carolina and the national average." The HIV prevalence rate includes newly reported HIV infected individuals (including both HIV and AIDS diagnoses) and represents the number of individuals living with HIV per 100,000 population.

Figure 3 shows the distribution of state HIV prevalence rates for Virginia and South Carolina (a dot for each county rate) and the U.S. prevalence rate as a horizontal line. The Virginia prevalence rate (the average) is lower than the prevalence rates for both South Carolina and the U.S. However, four counties in Virginia have higher HIV prevalence rates than any county in South Carolina.

![Figure 3: 2013 HIV Prevalence by County in V.A. and S.C. and U.S. Average](image)

V.A.'s HIV prevalence rate is lower than S.C. and the nation, but these four V.A. counties have higher rates than the S.C. county with the highest rate.

Therefore, while the average HIV prevalence rate for Virginia is lower than the average rates for both South Carolina and the U.S., it would be wrong to conclude that HIV infection is not a health problem for Virginia. A state may have many counties where there are significant health problems even though the state average may be relatively low.

**Understanding averages requires the context — the full data range**

These examples highlight the importance of considering the full data range (the distribution), not just the average. Depending on how the data are clustered, the average can be skewed in one direction or another and this might lead us to believe a problem is bigger or smaller than it really is. Similarly, focusing exclusively on averages may cause us to overlook sub-problems (e.g., a state average may conceal a problem in many of its counties, a positive profit margin may conceal financial hardship of many facilities, the average distance to a hospital may conceal health care access problems for many communities, etc.). This is particularly important when making rural-urban comparisons due to the wide range of geographic variation across the U.S.

**REFERENCES AND NOTES**