Questions to Raise When Analyzing Data

Whether reading a quantitative argument in the newspaper or constructing your own 60 second summary, the fundamental question to keep in mind is "Does it all make sense?" Any conclusion that claims to be drawn from the analysis of data should be approached with reasoned skepticism. We need constantly to ask questions about how much we trust the data and the conclusions drawn from them.

How Good Are the Data?

- **What definition is being used?**

  During the Great Depression five reputable agencies published five very different unemployment estimates for the same month (See Table below). The variance was due primarily to differences in the definitions of unemployment. Some estimates included people leaving school and seeking employment for the first time and some did not; some measured only blue-collar unemployment, whereas others included unemployed professionals.

Estimates of Unemployment for the Month of November 1935 According to Five Reporting Agencies

Agency Preparing Estimates	Estimate of Number Unemployed
The National Industrial Conference Board	9,177,000
Government Committee on Economic Security	10,913,000
The American Federation of Labor	10,077,000
National Research League	14,173,000
Labor Research Association	17,029,000

Source: Jerome B. Cohen, "The Misuse of Statistics," *Journal of the American Statistical Association,* XXXIII, No. 204, (1938), p. 657. Reprinted with permission from *Journal of the American Statistical Association.* Copyright ©, 1938 by the American Statistical Association.

The different definitions used were likely tied to each group's concerns and objectives. Clearly, very different arguments about the seriousness of the unemployment problem can be made depending on which estimates you choose to use for the number of unemployed people.

- **What is being measured?**

  In the nineteenth century, the "science" of craniology equated intelligence with brain size, and hence skull size. It is possible to measure skull size with extreme accuracy, but is it reasonable to believe that head size is related to intelligence? Today, we attempt to measure intelligence using the Intelligence Quotient and Scholastic Aptitude Tests. Many questions have been raised about what these tests really measure. Some people claim that they are twentieth century versions of craniology.

- **Are the measurements accurate and reliable?**

  In May 1993, *The New York Times* published an article entitled "Job Loss in Recession: Scratch Those Figures" explaining measurement errors in employment figures.

  > *Having declared last June that the recent recession had eliminated many more jobs than anyone had realized, the Labor Department is now canceling that view. The department says it overstated by 540,000 the number of jobs that were created in the late 1980's. And then it overstated how many jobs had disappeared in the recession.*[1]

  The measurements used to count jobs were faulty. Among other things, the U.S. Department of Labor counted paychecks instead of the people receiving paychecks. Any person receiving both a regular check and a separate check for overtime was counted twice.

- **How big is the sample? How was the sample chosen? Is the sample representative of the total population?**

  In his book *Flaws and Fallacies in Statistical Thinking*, Stephen Campbell cites the classic case of a poorly chosen sample. *The Literary Digest* during the 1936 Presidential election

  > *. . . mailed out 10,000,000 ballots and had 2,300,000 returned. On the basis of this unusually large sample, it confidently predicted that Alfred M. Landon would win by a comfortable margin. As it turned out, however, Franklin D. Roosevelt received 60 percent of the votes cast, a proportion representing one of the largest majorities in American presidential history. . . . [It turns out that the] ballots were sent primarily to upper-income types . . . as a result of the magazine's having selected names from lists of its own subscribers, and telephone and automobile owners. . . . In this election, the presence of bias rendered the sample, despite its enormous size, inadequate as a representative subset of the population.*[2]

How Good Are the Conclusions?

Edward R. Tufte in *Data Analysis for Politics and Policy* points out that

> *Almost all efforts at data analysis seek, at some point, to generalize the results and extend the reach of the conclusions beyond a particular set of data. The inferential leap may be from past experiences to future ones, from a sample of a population to the whole population, or from a narrow range of a variable to a wider range. The real difficulty is in deciding when the extrapolation beyond the range of the variables is warranted and when it is merely naive. As usual, it is largely a matter of substantive judgment. . . .*

[1] "Job Loss in Recession: Scratch Those Figures." *The New York Times*, Friday, May 7, 1993, p. D1.

[2] Stephen Campbell, *Flaws and Fallacies in Statistical Thinking*, p. 148, copyright © 1974, Reprinted by permission of Prentice Hall, Inc., Upper Saddle River, New Jersey.

When we read conclusions drawn from data that are often a matter of judgment, we should ask:

- **Is the analysis accurate?**

  [Oct. 6, 1993] Boston researchers disclosed today that they had made a mathematical error in a major study published in July that seemed to provide reassuring news about the risk of breast cancer in women with a family history of the disease. . . . The study of nearly 130,000 women published in the July 21st issue of the Journal of the American Medical Association *found that only 2.5 percent of all breast cancers occurred in women with a family history of the disease. When corrected, the percentage turned out to be 6 percent, much closer to previous estimates.*[3]

  Because the erroneous percentage was so much lower than previous estimates, the report was front page news when it was published.

- **Would alternative conclusions better fit the data?**
  When alternative plausible conclusions are derived from the same data, both conclusions may be wrong or they might both possess some truth. Stephen Campbell argues in his book, *Flaws and Fallacies in Statistical Thinking,* that:

  When equally plausible alternative conclusions can be reached from exactly the same statistical evidence, the logical link between evidence and conclusion offered is probably rather weak.
  Fact: "Fifteen individual nations have improved their infant mortality rates more than the United States since 1950."
  Conclusion offered: The state of health care in the United States is inadequate. Equally plausible alternative conclusion: The state of health care in several other countries—probably underdeveloped countries—has improved markedly since 1950, at least with respect to infant mortality.[4]

- **Are the data selectively reported?**
  In his book *We're Number ONE: Where America Stands—and Falls—In the New World Order,* Andrew Shapiro challenges the assumption that America is "Number One" by showing how "statistics can be found to tell any story." One section is devoted to debunking America's Number One status on spending on education. He claims:

[3] "Cancer study error found. Key finding unchanged on breast disease risk," *The Boston Globe,* Wednesday, October 6, 1993, p. 1.

[4] Stephen Campbell, *Flaws and Fallacies in Statistical Thinking,* p. 111, copyright © 1974, Reprinted by permission of Prentice Hall, Inc., Upper Saddle River, New Jersey.

We're Number One in private spending on education.
But . . . We're Number 17 in public spending on education.

> *Including public and private spending, the United States spends about 7 percent of its gross domestic product (GDP) on education annually, or around $330 billion. By this measure, we do fairly well; only a few developed nations spend more overall relative to the size of their economies. But as with health care expenditures, the spending comparisons change dramatically when we distinguish between public and private sources. The United States is first in private spending on education among the nineteen major industrial nations, but only seventeenth in public spending as a percentage of GDP.*[5]

In most cases you will not have access to the original data used, or have the time and energy to reconstruct the analysis. But you should always ask the kind of questions that we have raised here and critically examine the arguments presented. Keep asking yourself: *Does it all make sense?*

[5] Andrew Shapiro, *We're Number ONE: Where America Stands—and Falls—In the New World Order* (New York: Random House, 1992).