One Thing About...the Rhetoric of Statistics

Stephen T. Ziliak

One thing about statistics we might consider is the rhetoric of a commonplace phrase, "Let the data speak for themselves."

Wait. What? Come to think of it, from business and politics and journalism to academia and our own professional statistical societies, we'd do well to study the *rhetoric of statistics*, period—from the rhetoric of data and models and graphs to the reasons people give for making real-world decisions based on *p*-values (Wasserstein, et al. 2019). (Crazy, but true—many still do.)

The absurd implication of letting the data speak for themselves is that a data-producing author or speaker (or perhaps you, the data-receiving audience) is getting in the way of data. Too loud, too intrusive. "You're giving too much of—or the wrong kind of—*interpretation* of the data; let the data speak!" exhorts the Data Speaks for Themselves camp.

Alternatively, the commonplace phrase implies that both the speaker *and* the audience are crowding out data. Data are sitting quietly, after all, right there at the end of table; printed on sheets of paper or projected onto a screen, and yet can't get a word in edgewise to silence the voluminous rhetoric and finally speak forthemselves.

If only.

Data are figures, and figures are figures of speech. In truth, our basic ways of speaking and writing about statistical figures are literally bursting with figures of speech.

A model is a metaphor. Take the simple linear regression model, for instance, $Y = \beta X + \alpha Z + \varepsilon$. The character of variable Y is seen by perspective change, that is, by metaphorical extension (Burke. 1945. Appendix D, "Four Master Tropes). The perspective change—the metaphor—is a redirection of character Y by way of characters X and Z, expressed quantitatively by slope coefficients α and β, and notably by ε (the assumed-to-be random, independent, and identically distributed error term).

Some models, such as the Keynesian expenditure model of macroeconomic gross domestic product, $Y = C + I + G + Net\ Exports$, are better seen as *synecdoches*.

A synecdoche is a multivariate metonymy, a representation of a complex system, built on metonymic reduction and metaphorical extension. In the United States, and in economics generally, gross domestic product (GDP) estimation is said to "represent" the full chaos making up the final value of goods and services in the entire national economy, from consuming (C) a delicious dinner at Gene and Georgetti's restaurant in Chicago to new investment (I) nationwide in infrastructure (one hopes) for charging all those new electric vehicles; the total value of economic activity is said to be "represented" by this one simple reduction: synecdoche.

Rhetoricians discover synecdoche, metonymy, and metaphor in a split second. Most others—professional statisticians included—do not. Yet it's those same figures of speech that do most of the persuading.

Data, to repeat, are a rhetorical construct, too, from start to finish. "Data" in Latin is the plural form of *datum*, meaning in English something like "given," when in fact it's better to think of numerical selections as *capta*; that is, "taken."

Start with naming, classification, and group-wise categorization (Linnaeus, Galton, and Pearson come to mind). Continue to the choice of the units of measurement, sample selection strategy, normal or non-normal distributional assumptions, time period, context, geographic span, target population, mode of presentation, parameter estimation, *p*-values, and more; types of data are selections—figures of speech taken and shaped and presented for the purpose of persuasion.

When I was studying the history of public assistance in America, I was alarmed to find that as recently as the 1920s, the U.S. Census Bureau, influenced at the time by eugenics,

went county by county, collecting reams of data on what they called "insane Negro paupers living in almshouses," and displayed such data between 1880 and 1923 in a series of government-produced books they called *Paupers in Almshouses*.

The mean and median are, for example, metonymic reductions, using a "part" of the evidence—a selection of the whole distribution—to stand in for the "whole." The naive statistician and scientist, like the eager politician or journalist writing on deadline, do their best to persuade us they're doing purely "objective" research; they're "unbiased" or "revealing the truth" or "letting the data speak for themselves," but we know better. Or should. If statistical studies are purely objective, impartial, impersonal, and unbiased, free from judgment, why do replications and attempted reproductions fail so miserably in fields from economics to psychology?

Admitting that statisticians use rhetoric, and must, is no defect; it's an advantage. Take the official unemployment rate as measured by the United States government. The unemployment rate claims to shine a light on the percentage of jobless people, for example, but it does not include those (many) people who grow discouraged or are plain disbarred from looking for a job in the above-ground market. Excluding discouraged workers and former convicts from the jobless estimate is a rhetorical choice—one that Martin Luther King, Jr. and, more recently, the economist Mat Forstater have strongly argued against on grounds of pure reason and social justice.

The word "statistics" itself is a metaphor with nationalist ambitions (and not so innocent; *https://www.ucl.ac.uk/news/2021/jan/ucl-makes-formal-public-apology-its-history-and-legacy-eugenics*); at origin, it was a neologism deriving from the German *statistika*; that is, using numbers for "statecraft." Nowadays, we literally apply the word statistics to all areas of inquiry: government, business, personal, and other. Statistics the word, in other words, is itself a dead metaphor.

Discovering and probing the rhetoric of statistics is not a shameful act. We are no less scientific for understanding that the data doesn't speak for itself. On the contrary, we are *more* scientific; looking at its own language is what a mature science does, or ought to do. For instance, we are better off, not worse, for Kahneman and Tversky's discoveries of "framing" effects in behavioral economics.

Statisticians will sometimes say "let the data speak for itself" but most—and certainly most Bayesians—have set aside such childish positivism for a more mature rhetoric and understanding about "warrants-" or "degrees of belief," in Savage's and Jeffreys's sense. We consider the total body of evidence, our values and judgments, showing why we believe what we believe.

In other words, being rhetorically aware of statistics is prudent, and justice-promoting, too. Years ago, in a beautiful book called *A Rhetoric of Motives*, one of America's most influential rhetoricians—Kenneth Burke (1950)—had statistics, economics, and other human sciences in mind when he observed, "But in cases where a decision is still to be reached, a yielding to the form is preparation for assent to the matter identified with it." C

Further Reading

Burke, Kenneth. 1945. *Grammar of Motives. https://archive.org/details/grammarofmotives1945burk/page/328/mode/2up*.

Burke, Kenneth. 1950. *A Rhetoric of Motives*. Berkeley, CA: University of California Press.

Forstater, Mat. *https://realprogressives.org/podcast_episode/episode-55-the-economics-of-martin-luther-king-jr-with-mathew-forstater/*.

U.S. Census Bureau. *Paupers in Almshouses in the United States in 1923*. 1924. *https://hsus.cambridge.org/HSUSWeb/toc/showChapter.do?id=Bf*.

U.S. Census Bureau. *Public Health Reports (1896–1970). JSTOR* 39(41): 2572–74. *http://www.jstor.org/stable/4577326*. Accessed September 26, 2023.

Wasserstein, Ronald L., Schirm, Allen L., and Lazar, Nicole A. 2019. Moving to a World Beyond "$p < 0.05$." *The American Statistician* 73:sup1, 1–19. DOI: *10.1080/00031305.2019.1583913*.

About the Author

Stephen T. Ziliak is probably best known for his book (with Deirdre N. McCloskey) *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*. In 1996, he was awarded a PhD Certificate in the rhetoric of the human sciences at the same time he completed a PhD in economics, both at the University of Iowa. He has taught courses about the rhetoric of economics and statistics to postdocs and PhD, JD, MA, and BA students at universities across Europe and the United States.