

Review of *Interpreting Economic and Social Data: A Foundation of Descriptive Statistics* by Othmar W. Winkler

Natural scientists apply the fundamental assumptions of statistics to experimental data to draw conclusions about natural phenomena. Social scientists use the same methodology with socio-economic data to create dynamic models of human behavior. Othmar Winkler's *Interpreting Economic and Social Data* calls into question the tendency of social scientists to treat quantitative summary data as objective measurements as in the natural sciences. Winkler's observations on the subject are both thought-provoking and insightful.

Measurement in natural science is performed using uniform building blocks. By contrast, socioeconomic data emerges from 'real-life-objects' which are the projecting agents of socio-economic phenomena: households, firms, contracts, and sales, to name just a few examples. Relying primarily on the statistical survey, these objects are reduced to 'statistical-counting-units' or "still-pictures...somewhat like a photographic snapshot – except that less detail is retained" (Page 29). Since these counting units are typically self-reported and the samples are usually subjectively chosen, it follows they are hardly truly random. Furthermore, socio-economic data can be influenced by a number of factors that are of little concern in natural science, such as the place and time period they data are collected. As Winkler says, "the assumption that [socio-economic] data are only random deviations from some 'true value' is a carry-over from the thinking developed in the natural sciences." (Page 11) and "to analyze them with statistical methods based on inference and on the concept of random sampling, is pseudoscience." (Page 185)

This indictment serves as the basis for much of the book. Winkler, a professor of business and economic statistics, is devoted to rehabilitating the proper treatment and interpretation of socio-economic data. He laments that introductory statistical texts increasingly emphasize inferential statistics to the exclusion of descriptive statistics, the traditional domain of social scientists. This book aims to reverse the trend, and would serve nicely as a complement to the typical formula driven undergraduate or early graduate level text. Although the author strives to be straightforward, the book requires an understanding of concepts such as times series, frequency distributions, probability, and linear regression.

The early chapters focus on the structure and nature of socio-economic data. Through aggregation, statistical-counting-units can be organized into feasible units of analysis; ratios allow these aggregates to be put in to context with each other. After some conceptual framework, Winkler cautions of the "loss of meaning in aggregation" (page 45), meaning that it is easy to lose sight of the phenomenon of interest as the level of aggregation grows. He also warns specification is important; there is a temptation to produce ratios with entirely unrelated aggregates. An example of this would be producing a measure of accidents per hours worked; however, a better ratio would be accidents per hours worked in a given industry, since many industries are not especially dangerous.

The middle and largest portion of the text focuses on longitudinal analysis. All too often, time series data seldom reveal any universal economic laws so much as they are a product of the historical condition and landscape where the data were recorded. Because of this fact, socio-economic time series data utilized in forecasting models will eventually become obsolete as broader societal changes take place. He cautions forecasters against the temptation to treat socio-

economic time series data as "random samples from some hypothetical timeless populations" (Page 94) which can lead to a sense of complacency as the number of observations increases. This word of warning seems pertinent, especially in the aftermath of the financial crisis of 2008, which was caused by a series of rare but interconnected events.

Winkler's discussion of longitudinal analysis drifts into the realm of price statistics. He asserts that transactions of money for goods should be thought of as the real-life-object, and the price paid the statistical-count-unit, which can vary widely depending on the place and time of purchase. Winkler suggests utilizing scanner data to produce a measure of average currency paid per transaction. This would negate the need to alter the basket of goods used in a price index as new products enter the market or are phased out. Winkler also expresses concern that current measures of labor productivity focused on total hours of labor and output may fail to account for capital improvements over time.

Only late in the book does Winkler approach cross-sectional analysis, beginning with a chapter on the interpretation of frequency distributions, central tendency, and dispersion. Turning to regression analysis, he notes linear regression models originated in natural science, and are poorly suited to socio-economic data, which is typically plagued by problems such as heteroskedasticity and low R-values. As a result, the line of best fit produced in regression equations is often fraught with misspecification issues and is likely to miss more complex relationships underlying the data. To remedy this, Winkler suggests using disaggregated data to the extent possible, and cautions against interpreting slope coefficients literally. The following chapter on the intersection of socio-economic statistics and probability carries on in this vein. The most common mistake is the application of statistical inference to populations or deliberately selected samples deemed 'representative'. Modern social scientists tend to "view every situation as a random process or a random experiment, regardless of whether randomization was involved" (Page 187). Winkler describes these misuses of inference as a trend, reinforced by statistics textbooks and journal editors. However, it seems likely the expanded use of mathematical statistics in social science is here to stay.

The book contains numerous diagrams to visually illustrate and reinforce the concepts described verbally. The endnotes of each chapter contain detailed asides and citations should the reader be interested in pursuing any particular subject at length. Several chapters also contain appendices that cover topics which may require a refresher or need additional space to explore concepts mathematically. There are also two short chapters focused on the use of statistics in accounting and geography.

The writing in this book is easy to digest, although in the later portions, it can seem repetitive, likely because it is intended for use as a reference. Overall, this very practical book would serve an aspiring social scientist or an experienced practitioner well by working through its lessons.

Thomas Luke Spreen, economist, Division of Labor Force Statistics, Office of Employment and Unemployment, U.S. Bureau of Labor Statistics, June 9, 2011 To be published in the Book Review Section of the August 2011 issue of the Monthly Labor Review.