Mini Review

Open Access

Interpreting the cumulative frequency distribution of Socio-Economic data

Abstract

The ogive-the accumulated frequencies of quantitative variables of socio-economic statistical data - is treated, if at all, as a curiosity, without trying to interpret its contribution. This essay shows how to extract revealing features of the underlying local-historical situation of society from the ogive that otherwise remain unnoticed.

Keywords: Socio Economic Statistical Data, Quantitative Characteristics, Frequency Distributions, Ogive

Volume 8 Issue 1 - 2019

Othmar W Winkler
Georgetown University, USA

Correspondence: Othmar W Winkler, Professor Emeritus, Georgetown University, Washington, DC 20057, USA, Email winklero@georgetown.edu

Received: October 22, 2018 | **Published:** January 23, 2019

Mini Review

1. The purpose of statistics in the social sciences is to describe and analyze situations of society. Unfortunately, statisticians, authors of textbooks and editors of statistical journals show little interest in the final, most important stage of the statistical process. The analysis of the statistical output, the numbers, is left to economists and sociologists, not done by statisticians properly speaking.

Phenomena in medicine and the life sciences - such as "tuberculosis", "HIV" or "cancer" - are of a different nature, than those of society, like "Inflation," "Foreign Trade," "Employment" or "Crime and Policing." Their work requires different statistical approaches and methodologies. The tools for bio-and medical statistics, differ from those for social science statistics like the tools of a watchmaker from those of a carpenter.

2. Numeric characteristics, ideally from precise measurements, arranged in Frequency Distributions and prominent in bio-medical statistics, are less frequent and less important in the social sciences. Nonetheless, textbooks for statistics in business, economics and the social sciences[1], following the lead of bio-and medical statistics and their ease to invite calculations[2], give them preferred attention. When economists speak of "measuring," e.g.' Price Levels,' they imply innocently but deceptively the use of precision measuring instruments like in the bio-medical statistics. But this is not the case.

3. Data of numeric characteristics, arranged as a Frequency Distribution only reveal a narrow, usually minor aspect of the studied Phenomenon. In biology and medicine these assembled measurements tend to form symmetrical shapes. Their ideal contour, a symmetrical bell-shape, formulated mathematically by Gauss, is variously known as the "Normal Distribution," as the "Error Curve, Gaussian Distribution or Gauss Curve." That became the center of a mathematical foundation of statistics leading to statistical inference, random sampling, model building, Decision Theory, and more.

To adjust this ideal of symmetry to the reality of asymmetries in distributions of socio-economic data, deviations from symmetry were reluctantly accepted. The following examples of socio-economic data, which may surprise workers in the medical and bio sciences, show how unrealistic the assumption of symmetry actually is (Figures 1–7).

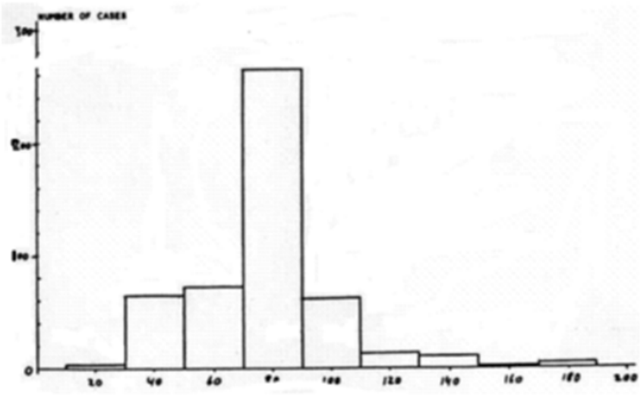

Figure 1 482 wages of cotton mill workers, Caracas Venezuela.

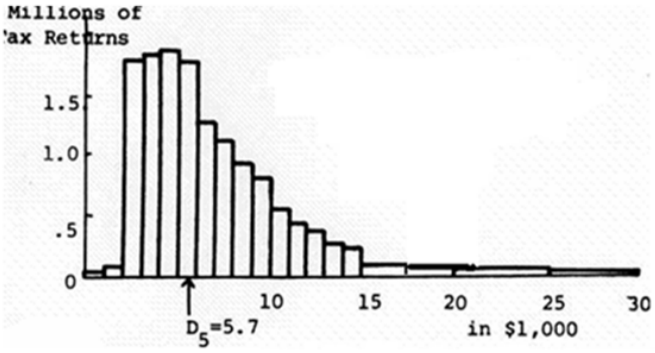

Figure 2 Individual Income Tax returns.

[1]David R Anderson, Dennis J Sweeney, Thomas A Williams. Essentials of Modern Business Statistics with Microsoft Office Excell. South-Western, Mason, OH, 2012 devotes only 15 lines, on p.56/7 to 'Ogive' Yfj. Also in Basic Business Statistic s-Concepts and Applications, 14th edn, by Mark L Berenson, David M Levine, Kathryn A Szabat, David F Stephan, published by Pearson, 2019 presents two examples of two polygon lines in each picture but no comment of ay kind.

[2]Note that Frequency Distributions are discussed toward the end of my book in chapter 8, to account for their diminished role in the social sciences, not as usual, in the beginning chapters of textbooks on statistics. See: Interpreting Economic and Social Data–A Foundation of Descriptive Statistics. Springer. 2009. E.g. in the website of *statlit.org*, look for 'Academic books'.

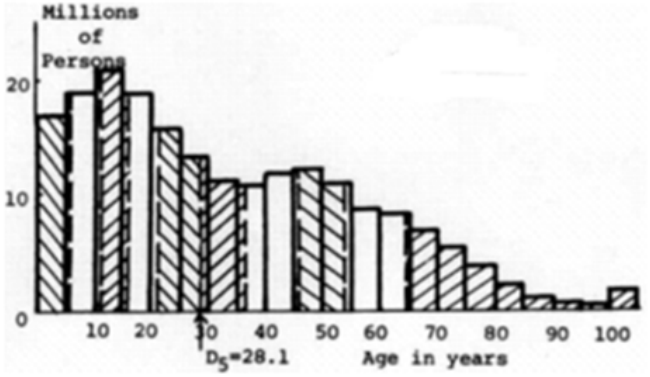

Figure 3 US population, by Age.

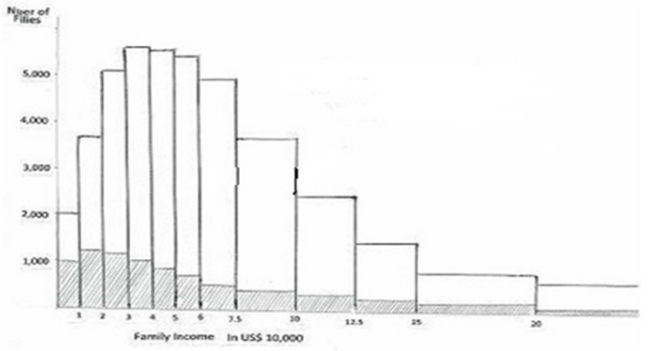

Figure 4 Family income in the past 12 months – white and black (shaded) families in USA.

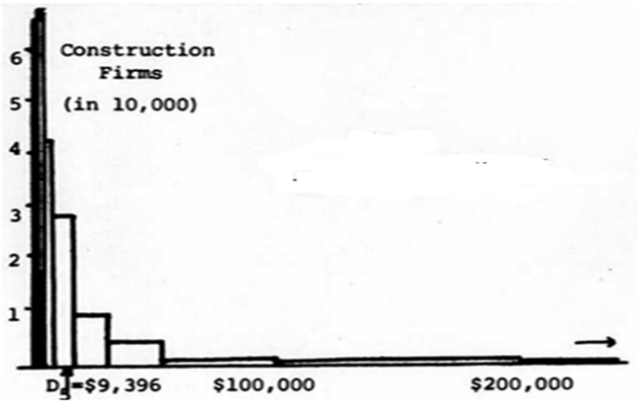

Figure 5 US construction firms by gross Income.

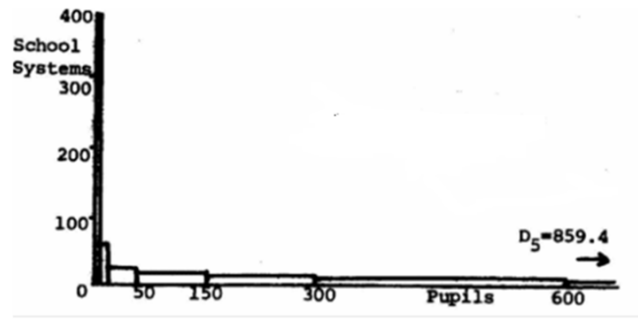

Figure 6 US school systems by number of pupils.

The standard deviation, the measures of centrality - the mean, median, mode, and measures of asymmetry - are of scant practical value for actual Frequency Distributions in the social sciences. Bell shape or symmetry in frequency distribution of socio-economic data just does not happen.

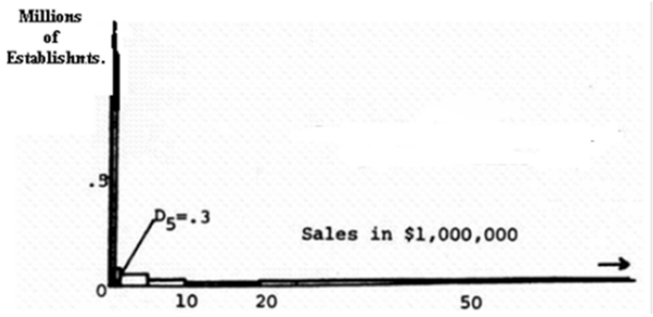

Figure 7 Retail trade establishments by size of Sales.

The measures of Frequency Distributions only make sense for data distributions that re essentially symmetric, or "normal." Asymmetries are considered deviant forms. However, asymmetries are typical in the Frequency Distributions in the social sciences. Most of these Frequency Distributions have their highest frequency at the beginning of the horizontal, X-scale. The right tail end, in many distributions, extends far into the distant high X-scale values.

4. The Ogive. It is Ironic that after carefully explaining how to proceed to aggregate the data of a Frequency Distribution in successive steps into the ascending cumulative S-shaped profile - named an OGIVE because of the similarity to that architectural design - the meaning and use of the rearrangement of the data have never been explored or explained. The slope of the Ogive shows the degree of resistance an additional counting unit, arriving at the scene, will encounter (Figure 8&9).

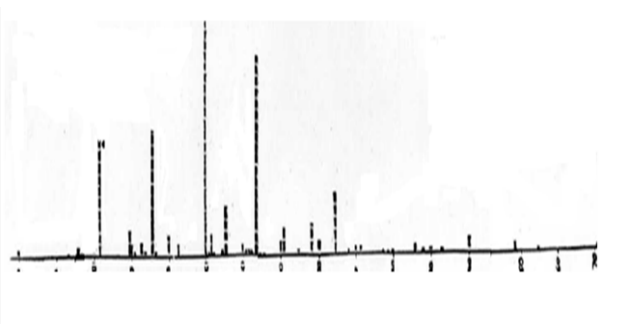

Figure 8 Plot of the 482 individual wages before grouping, shown in Figure 1.

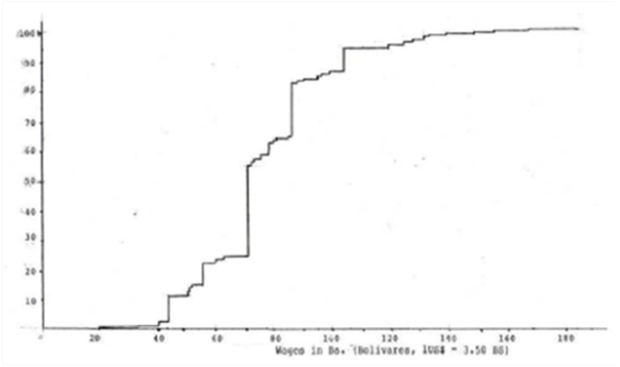

Figure 9 Ogive of the ungrouped values of the 482 data in Figure 8.

To construct an OGIVE the class frequencies of a given class and those of all previous classes are addend and plotted at the upper end of a given class, beginning with the first, lowest class interval to the last class at the end of the Frequency Distribution. In this accumulated

Citation: Winkler OW. Interpreting the cumulative frequency distribution of Socio-Economic data. *Biom Biostat Int J.* 2019;8(1):13–15.
DOI: 10.15406/bbij.2019.08.00263

form, the data represents the same information as in the shape of the Frequency Distribution from which it derives. But in addition, the Ogive highlights features of the structure of the society that are not directly perceptible in the data of the Frequency Distribution.

5. To explain what the Ogive can accomplish, the following intuitive example appears well suited. Imagine the S-shaped ogive of the 482 workers – Figure 9, also Figures 8 and 1 - as the geological profile of a beach: its profile begins emerging from the ocean, gently rising. Then the profile of that beach rising in small steps of the rocky backdrop, abruptly becomes a steep cliff like a wall of rocks. Toward the end of the beach the firm, flat surface of the higher terrain of the hinterland begins. Depending on the intensity and strength of an arriving wave, it will easily roll over that flat, then rising part at the beginning of that beach, and then run against that steeper cliff, further up. Most of those new incoming waves will break somewhere at that vertically rising portion of the beach profile. Only few, rare waves will arrive with sufficient force and momentum to overrun the flat part, the steep rising cliff of the beach and wash over the high, flat part at the end. The resistance of the given contour of the land to the further advancement of that strong wave becomes less and less as the wave reached into the high part of that imaginary beach of an ogive.

6. Now apply this comparison of an ocean wave moving against the beach to the situation of a new arrival at the place of employment. Its salary structure is symbolized by the graph of the ogive, Figure 9. Like such an ocean wave, a newcomer most likely will not be able to go beyond that 'cliff', at the corresponding X-value of 70 Bs[3]. He may have no difficulty getting a modestly paying menial job, like a wave easily rolling over that first flat part of the imagined beach profile of the ogive. If such an arrival is intelligent, has physical strength or useful skills, he or she can reach the steep part of the ogive, like the medium strong ocean wave reaching but not passing beyond that steep vertical portion of the beach profile. It will become harder for him/her to further advance and he will get stuck where the majority of new arrivals likely will end up like those many ocean waves being stopped at that steep portion of the beach. Now if the new arrival is smart, has employable skills and knowledge, such a newcomer - like the strong Ocean wave will roll over the steep portions of that "beach profile"- the new arrival will move up to the flat part where he finds decreasing and minimal resistance to even easier and further advancement economically and socially. Although this information is also contained in the data of the original Frequency Distribution, (Figure 8) it becomes visible in the Ogive.

Following is a series of Ogives (data not shown here) from different social and economic situations. In analogy to the discussion about the ogive in Figure 9, you can figure out how an imaginary new arrival would fare in those economic, technical or societal structures underlying these ogives (Figure 10–13).

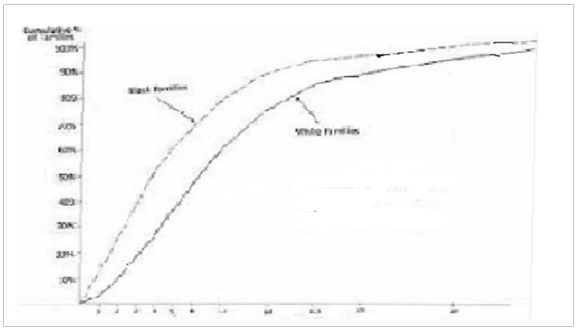

Figure 10 Ogive of the family Income Data in Figure 4.

3Bs. = Bolivars, Venezuelan currency. At that time US $1= 3.30Bs

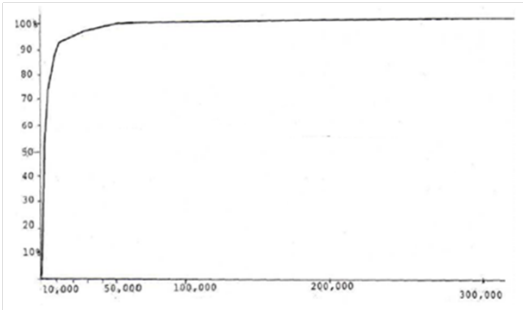

Figure 11 Ogive of the number of township Governments by Size of population.

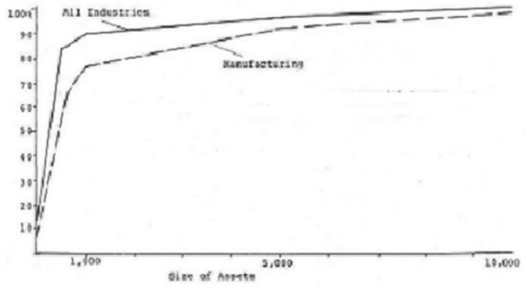

Figure 12 US manufacturing corporations and all corporations with assets up to $10,000 - Income Tax 2007.

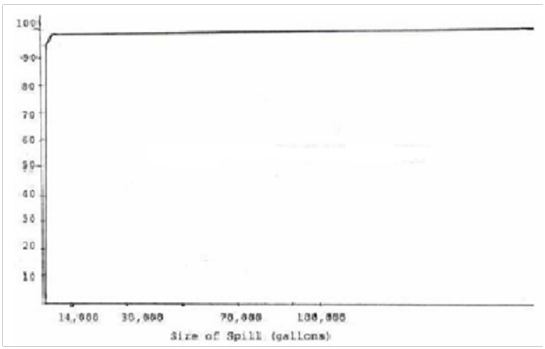

Figure 13 Recorded Oil spills in US waters.

Acknowledgments

None.

Conflicts of interest

Author declares that there is no conflict of interest.

Citation: Winkler OW. Interpreting the cumulative frequency distribution of Socio-Economic data. *Biom Biostat Int J.* 2019;8(1):13–15. DOI: 10.15406/bbij.2019.08.00263