Journal of Business and Economics, ISSN 2155-7950, USA
July 2019, Volume 10, No. 7, pp. 606-611
DOI: 10.15341/jbe(2155-7950)/07.10.2019/002

http://www.academicstar.us

Veracity Versus Illusion in the Contents of an Initial Course in Statistics

Juan Carlos Abril, María de las Mercedes Abril
(Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad Nacional de Tucumán, Argentina)

Abstract: Without Statistics it would be impossible to give a satisfactory explanation to a large number of human events, and without it, it would be absolutely impossible to foresee them in all their scope, to guide the criteria with which they are to be judged, to take in time the most conducive measures to avoid or accelerate them, to keep them unchanged or to modify them.

It is our main goal to analyze some of the current contents of many initial courses of Statistics as they should be changed and be updated immediately.

A very important point to take into account is that this type of course is usually in some cases the first and sometimes the only course of statistics that students see during their degree training. We have the responsibility to offer the best and most up to date material, thus fulfilling the premise that universities should always be at the forefront of knowledge.

Key words: statistics; courses; contents; knowledge

JEL codes: A22, C10

1. Introduction

The fundamental notion in statistical theory is that of a group or aggregate, a concept for which statisticians use a special word: "*population*". This term will be used to designate any collection of objects under consideration, whether animate or inanimate; for example we can consider populations of human beings, plants, errors in the measurement of a scale, barometric pressure on different days, and even more populations of ideas, such as the possible ways in which a hand of cards can be distributed. The common notion to all these things is that of the aggregate.

Statistics is the branch of the scientific method that deals with the data obtained by counting or measuring the properties of populations of natural phenomena. In this definition "*natural phenomena*" includes all the facts of the outside world, whether human or not.

It may be good to point out that "*Statistics*", the name of the scientific method, is a collective noun, has a capital "*S*" and is singular. The word "*statistics*" applies to the numerical material the method uses, and in that case it is not capitalized and plural. We also have the singular word "*statistic*" that is defined as a function of the observations in samples of some population. "*Statistic*" in this sense has plural "*statistics*".

Juan Carlos Abril, Ph.D. in Statistics, Professor, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), National University of Tucumán; research areas/interests: time series analysis, volatility, econometrics. E-mail: jabril@herrera.unt.edu.ar.

2. The Importance of Statistics

The enormous importance of Statistical studies is well known when it comes to examining and comparing the evolutionary process of any manifestation of human activity through time and space.

Without Statistics it would be impossible to give a satisfactory explanation to a large number of human events, and without it, it would be absolutely impossible to foresee them in all their scope, to guide the criteria with which they are to be judged, to take in time the most conducive measures to avoid or accelerate them, to keep them unchanged or to modify them and to discover with swiftness and certainty the factor that prevails in this or that direction on the final result in each case.

Many examples show us that people of high intellectual formation considered Statistics, on the one hand, as an essential part of the civic and democratic life of every society. This leads to the need for all of us, as members of civil society, to have a proper "*statistical culture*" to conveniently benefit from the information provided. On the other hand, it is also considered an area whose technical content should be developed by those who have achieved appropriate training so that the information that reaches civil society accessible in its interpretation, accurate in its appraisals and true in its numbers.

Herbert George Wells said "*Statistical thinking will become as necessary to be a responsible citizen as it is to know how to read and write*".

On the other hand, Francis Galton once said "*I have to write about a great subject (Statistics), but I strongly feel my literary inability to do it intelligibly without sacrificing its accuracy and its veracity*".

3. Some Historical Questions

Alexandre Moreau de Jonnés, a French statistician, points this out from the earliest times, inscribed in the Pentateuch with the name "arithmi", which means numbers. The first empires frequently collected population censuses or recorded commercial transactions of different products. The Roman Empire was one of the first states to collect extensively data on the population of the empire, geographical area and wealth. The counting was repeated several times by different soldiers (that is, sampling techniques were applied). The most frequent value (what in modern terms means the mode) thus determined was taken as the most probable value of the number of bricks (today we would say that this is a first approximation to the method of maximum likelihood). Multiplying this value by the height of each brick used in the wall allowed the Athenians to estimate the height of the stairs needed to climb the walls.

In general and throughout the history of mankind, statistical studies were very simple, and consisted of a description using tables, graphs, percentages, averages, etc. of a case, situation or problem based on the data obtained, without making any kind of inference about the population from which they came. The main reason for this form of work was that probabilities had not been introduced as a tool within the Statistics; therefore empirically it happened that implicitly it was considered that the elements of a population all had the same probability of occurrence or of being observed. This approach was maintained until almost the end of the 19th century.

Although the origins of the statistical theory can be placed in the 18th century when there were major advances in the field of probability, that is the area of knowledge that begins to be used as a basis for the scientific development of the discipline, the concept of modern statistics emerges at the end of the 19th and early 20th

century and is produced in three stages.

The first stage, which begins at the early 20th century, was led by the works of Francis Galton and Karl Pearson, who transformed Statistics into a mathematically rigorous discipline by introducing probabilities as its basic tool. This was achieved not only to be used in scientific analysis, but also in industry and politics.

Another great precursor of the application of mathematics in statistical investigations was Adolphe Quetelet (1796-1874). Quetelet taught that the more individuals are subjected to observation, their individual peculiarities disappear, being that physical, intellectual or moral in any given set of social phenomena. In effect, the scientific basis of modern statistics was found by Adolphe Quetelet, Director of the Brussels Astronomical Observatory with his Law of Large Numbers.

From these advances, it is clear that each population under study has a probabilistic structure on which subsequent work should be based. At this time the following fundamental concepts of Statistical Science are defined:

- **Parameter:** Characteristic, usually numerical, of a population. Parameters are fixed, non-random and usually unknown magnitudes.
- **Estimator:** It is a rule that expresses how to calculate estimation as a function of the information obtained from the sample. It is stated in general a formula. Estimators, being functions of the respective sample, are random; therefore they have a probabilistic structure. They give us sample information about the parameters. When the function of the sample does not depend on the parameters, it is called a ***statistic***.
- **Estimation:** The value that an estimator or statistic takes in a particular case, that is, for a given sample.

On the other hand, statistical societies begin their work in the 19^{th} century. Indeed, the oldest and one of the most important is The Royal Statistical Society of Great Britain founded in 1834 and Florence Nightingale was its first female member. As we said, mathematical statistics owes a lot to the works on probabilities performed by Jacob Bernoulli, De Moivre, Daniel Bernoulli, Laplace, Leibnitz, Maclaurin, d'Alambert, Condorcet, Fourier, Gauss and Quetelet. From the illustrious Quetelet we have the idea of periodically gathering in congresses and meetings, both official and private. Statistical congresses began in 1853 in Brussels, and followed in Paris, Vienna, London, Berlin, Florence, The Hague, St. Petersburg, Budapest, etc. All these congresses have been fecund in teachings concerning statistical research, and if we haven't seen in all countries equal efficiency in their statistical services was due to an uneven development of public administrations, from which such services emanate, because, as Quetelet said: from such sources, such statistics.

The second stage takes place between 1910 and 1929, it was initiated by William Gosset (whose pseudonym was Student) and reached its peak with the works of Ronald Fisher. This involved the development of better models for the design of experiments, hypothesis tests and techniques used with data from small samples.

One of the great advances of this period was made by Fisher (1921) when he introduced the principle of maximum likelihood, the current basis of a great variety of estimation procedures and tests.

The third stage, which mainly saw the refinement and expansion of the first developments, emerged in the 1930s from the collaborative work between Eagon Pearson and Jersy Neyman.

This process of developing Statistics as a science with applications in practically all branches of knowledge continues since then at a sustained pace, producing new advances at increasing rates, which come to light and come to us, the statisticians, through a large number of high-level publications, both theoretical and applied. It would be impossible to try to make in these few pages even a simple list of the great precursors of Statistics.

Therefore it is recommended to read the bibliography given at the end of this work. Nowadays, statistical methods are applied in all fields that involve decision making, to make an accurate inference from a set of data and to make decisions in the face of uncertainty.

4. Illusion in the Contents

On this section we will analyze some of the current contents of many initial courses of Statistics that must be changed and be updated immediately. Because these issues are totally outdated and inadequate, and because there is a resistance to modernizing them, we call them contents of dreams and not truthful as it would be those that should be taught.

Many courses begin with an introduction and definition of Statistics, sources of statistical information and data banks. This may be acceptable as long as it does not take too long. Then, they continue with information summary techniques, population and sample concepts, position measures, variability and asymmetry. Here the problem begins: students do not receive a minimum training in probabilities. They are not taught to distinguish between parameter, estimator and estimate; they do not know the basic concept of random variable and its distribution.

Sampling concepts cannot be understood in their real dimension because there is no probabilistic basis. This implies that the very essence of modern statistics, inference is not understood and cannot be developed in its full dimension. When the subject of estimation is introduced, it is treated as a calculation problem and there is no distinction between estimator as a random variable and estimation as the value of that estimator achieved for the only sample available. The properties of the estimators are, at best, enunciated without providing arguments or proofs that justify their characteristics. This is repeated in the regression and correlation topics where the estimation of parameters is taken as an algebraic question without taking into account the properties of those estimators and the conditions that these properties have to be satisfied.

A separate paragraph deserves the principle of maximum likelihood. It tries to be taught it but in a totally unintelligible way. Of course, the properties of the estimators achieved by this method are not studied. It should be noted that given the increasing computing capacity available today, this is the method, or one of its variants, that is most used in practical situations.

The subject of chronological or time series is approached through the classic decomposition, that is: trend, seasonality and irregular; to which you could add cycles but it is not done. It is taught that the trend can be estimated in two ways:

1) through a regression in time using least squares, and
2) through moving averages (centered).

In case a) the model is easy to estimate using simple least squares, but suffers from the disadvantage that the trend is deterministic. In general, this is very restrictive. Indeed, in economics, for example, if a variable is considered to have a deterministic tendency, it would mean that any economic impulse of any intensity will have no long term effects, since everything will return to its given tendency (April, 1997).

In case b) the estimation of the trend suffers from multiple contraindications, but the most important thing is that the statistical properties of these estimators are unknown (April, 2011). Since in this approach, seasonality estimation and any other component are interrelated with the estimate of the trend in cases a) and b), these estimates will be unsatisfactory. Nowadays the theoretical subject has evolved enormously and there are many

computer packages that make estimates in the most efficient way possible.

An important issue we observed is that, in general, these initial statistics subjects are not taught by statistics professionals. On the contrary, they tend to be people with adequate training in other areas but with little advanced training in statistical science.

Finally it is observed that modern computer packages specialized in statistics are rarely used. But here one must be careful, because using packages automatically without adequate knowledge of the theory can do a lot of damage to students.

5. Veracity in the Contents

An initial course in Statistics should contain, at least, the following topics, in the following order

1) Data visualization and descriptive statistics
2) Probability Theory
3) Random variables
4) Current distributions of random variables
5) Sampling distributions of statistics
6) Point and interval estimation
7) Simple linear regression and correlation
8) Other topics considered useful and important

The topic "*descriptive statistics*" should be offered in complete harmony with the rest of the contents. We must distinguish between the calculations of population parameters that could be achieved here and the estimations to be presented in later topics.

Good bibliographical references for this course are Larsen R. J. and M. L. Marx (2011) and Newbold P., W. L. Carlson and B. M. Thorne (2012).

A very important point to take into account is that this type of course is usually in some cases the first and sometimes the only course of statistics that students see during their degree training. We have the responsibility to offer the best and most up to date material, thus fulfilling the premise that universities should always be at the forefront of knowledge.

It should be emphasized that it is necessary to introduce the use of the computer and specific packages to analyze data, and that teachers and instructors must have advanced training in Statistics.

It is imperative that we follow this path of updating and modernization in teaching our subject.

References

Abril Juan Carlos (1997). "Series de tiempo irregulares: Un enfoque unificado", in: *Conference during the XXV Coloquio Argentino de Estadística*, Sociedad Argentina de Estadística, November 1997.

Abril Juan Carlos (2008). "La estadística y la ciencia estadística", *XIII. Reunión Científica del Grupo Argentino de Biometría*, Tucumán.

Abril Juan Carlos (2011). *Análisis de la Evolución de las Técnicas de Series de Tiempo. Un Enfoque Unificado*, Estadistica, Vol. 63, pp. 5-56.

Fisher Ronald A. (1921). "On the mathematical foundations of theoretical statistics", *Phil. Trans*., A, pp. 222, 309.

Galton Francis (1909). *Memories of My Life*, London: E. P. Dutton and Company.

Larsen R. J. and M. L. Marx (2011). *An Introduction to Mathematical Statistics and Its Applications* (5th ed.), Prentice-Hall: New Jersey.

Newbold P., Carlson W. L. and Thorne B. M. (2012). *Statistics for Business and Economics* (8th ed.), Pearson. London.

Stigler S. M. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*, Harvard University Press, Cambridge, Massachusetts and London.

Stuart A. and Ord J. K. (1987). *Kendall's Advance Theory of Statistics* (6th ed.), Vol. 1, Charles Griffin and Company Limited: London.

Stuart A. and Ord J. K. (1991). *Kendall's Advance Theory of Statistics* (5th ed.), Vol. 2, Edward Arnold: London.

Tucídides (Thucydides) (1985). *History of the Peloponnesian War*, Penguin Books: New York.