STATISTICAL ILLITERACY IN LATIN AMERICA: A CONSEQUENCE OF THE DIFFERENT VISIONS ABOUT THE MEANING OF STATISTICS

Carlos Araújo
Pontificia Universidad Católica de Chile, Chile
araujo@infodevelop.org

The lack of interest of statisticians to reach a consensual characterisation of Statistics is the main reason for the inefficacy in the efforts made until now to install or to improve statistical education in Latin American countries. It is therefore necessary to attain consensus within the international statistical community to produce a characterisation of Statistics that makes possible the establishment of a general common frame within which all the activities related to statistical development should be incorporated. Under this characterisation, some actions are suggested to encourage national and international Statistics associations to support the improvement of the teaching of Statistics and so set some sustainable bases for the establishment of a suitable statistical literacy.

INTRODUCTION

The major problems facing the development of Statistics in Latin America and in part of the Caribbean countries (an area that hereon we will call "the region") arise from the poor development of statistical education that has been unable to make known to professionals (whether statisticians or not), to reputed public users, and to the general public, what Statistics is and what it does.

As a consequence, it is easy to verify that some professionals with little or no knowledge of Statistics become in charge (most of them in good faith) of the technical conduction of actions dealing with statistical development in the region, in national and international agencies, public and private alike.

As for the general public (including professionals and public authorities), there exists the generalised perception that Statistics "*lies," "it is cold," or "it is unreliable."*

Can statistical activities be carried out suitably if those who are technically responsible for this work have a wrong concept of what Statistics means? By being a service discipline, can Statistics develop adequately if its users consider that it "lies," "it is cold," or "it is unreliable"?

It is for this reason that the improvement of statistical education through a proper teaching of Statistics has top priority as a policy to encourage statistical development in the region in general, and to improve statistical literacy in particular.

As it will be seen next, the lack of interest of statisticians in relation to a consensual characterisation of Statistics is the main reason for the inefficacy in the efforts made until now to install or to improve statistical education in the region.

COEXISTENCE OF MULTIPLE, VAGUE, INCOMPLETE AND CONTRADICTORY CHARACTERISATIONS OF STATISTICS

Many professionals and researchers can apply or develop Statistics without fully knowing the purposes of the discipline, but *it turns out impossible to teach Statistics without transmitting, explicitly or implicitly, an awareness of the purposes or objectives of Statistics.*

The attempts to characterise or define Statistics are varied, different and incomplete. This contributes to maintaining confusion among those who teach it, use it or practise it (see, for example, the web site http://www.google.cl/search?hl=es&lr=&oi=defmore&q=define:Statistic). In general , these attempts are characterised by:

1. *Vagueness.* In this category fall, for instance, those who say Statistics is: "*the technology of scientific thinking"; "a bridge between mathematical models and real problems"; "interface between Mathematics, computing and data analysis";* and so on. All these "definitions," and many others of similar kind, bear the inconvenience that they do not characterise Statistics (other disciplines also fit them) and, besides, they are not translated into specific or concrete norms with respect to its nature and applications. It may be for this latter reason that it does not have real influence upon the teaching, the research, or the practise of Statistics.

2. *Limitations by compression.* The main group of such characterisations is made up of those that begin with the phrase "*Statistics is a branch of Mathematics that …"*. Any definition that considers Statistics as a branch of Mathematics implies that the objectives of Statistics are contained within the objectives of Mathematics, which is far from true. The contribution of Mathematics to the development of Statistics should always be adapted to the objectives of the latter, as it happens with any other scientific discipline.
3. *Limitations by extension.* Another group of limiting "definitions" is the one that "defines" Statistics by means of a description of the activities it carries out. They are in the fashion of: *"Statistics collects, organises and analyses data… ."* Besides the fact that other disciplines or techniques can perform the above mentioned activities (for instance, Accountancy), these characterisations always are incomplete.

SOME PROBLEMS ARISING FROM THE CONFUSION ABOUT THE MEANING OF THE TERM "STATISTICS"

The big confusion brought about by the different and sometimes contradictory purposes attributed to Statistics gives rise, among others, to some behaviours described below that directly affect the teaching of Statistics in the region and, consequently, impair the establishment of statistical literacy.

1. *Those originated by considering Statistics as "a branch of Mathematics."* We shall mention two types of educational behaviours about which the examples abound:
   a. *Using Statistics to exercise mathematical applications.* This practice leads to placing a greater emphasis on calculus than on interpretation. A remarkable example can be easily tested when any student of Statistics is asked to interpret the median or the average of a set of numbers.
   b. *Conditioning the teaching of statistical methods to the knowledge of the mathematical concepts that support it or found it.* As will be seen later, this trend limits the formation of 'intelligent' users of Statistics who make up the basis for the development of the research and the practice of Statistics.
2. *Those originated by considering that Statistics "picks up, organises and analyses data..".* Since the operations described have as a final aim the drawing of conclusions, these are frequently presented, explicitly or implicitly, as part of the statistical results in the teaching of Statistics. To assume that statistical hypotheses can be formulated in relation to the various fields from which the data may arise leads to the absurd pretence that Statistics can draw conclusions about the most diverse areas without the scientific knowledge of such areas. This may generate damaging consequences and practices for the perception of Statistics, such as:
   a. *To give origin to professional misgivings and legitimate distrusts with respect to statistical work.* Statistics being a discipline of services, this jeopardises its development directly.
   b. *To encourage or 'justify' specialists from the areas of application to present conclusions based on their particular knowledge and experiences as if they were 'statistical' inferences.* This is one of the main reasons for the discrediting of Statistics among the general public and the principal cause to allow that professionals from other areas manage technically 'Statistic' Units. As an example, it is enough to observe the perception that ordinary people have with respect to the so-called 'Statistic' on poverty or labour in Latin America.
   c. *To introduce the concept of 'representative sample' as a statistical concept.* In the perspective that Statistics picks up and organises data to draw conclusions next, it can be inferred, by oversimplification or ignorance, that statistical work consists in selecting a *representative sample* whose processing might lead to valid conclusions. From the statistical point of view, to hypothesise about the existence of such samples is absurd and perverse since it makes possible the inclusion of subjective aspects in the selection of the sample, and it encourages specialists from the areas to which the data are related, to consider themselves in a better position to define the 'statistical' design of sample selection. After all, who can be better qualified than such specialist to know how representative a sample is?

d. To consider errors or limitations in the registration of the data as 'statistical errors.' For instance, when the data are gathered from administrative registries, their bad quality or their lack of coverage can cause inconsistencies in statistical results obtained from such sources. These inconsistencies are presented and sometimes quantified as 'Statistical Error.' The use of this term provides Statistics with an unfair character of insecurity that may become an additional element of discredit in the eyes of public opinion with respect to this discipline.

THE NEED FOR A COMMON CHARACTERISATION OF STATISTICS

The inefficiencies in the teaching and application of Statistics, some of which are mentioned in the above sections, have direct consequences on the poor perception that potential users and the public in general have with respect to this discipline.

On the other hand, the lack of definition of the term Statistics delays the efforts to identify the adequate conceptual frame for the transmission of the educational message in Statistics. In this way, it is possible today to count on modern, complete and effective methods for teaching Statistics, but totally inefficient to achieve a correct statistical education because the methods are applied in the transmission of inappropriate or simply wrong knowledge.

It is then necessary to reach a common basic characterisation of Statistics that counts on the consensus of the international statistical community, and one that reflects appropriately the nature and purposes of statistical work. This characterisation will provide basic orientations and recommendations in the matter of the teaching and application of Statistics, and consequently start satisfactory policies in statistical education in the region.

The following proposal is presented with the purpose of contributing to initiate work that may lead to the referred consensual characterisation of Statistics on the part of national and international statistical communities, Based on that characterization it may be possible to design and keep efficient policies concerning statistical education in the region, as well as to provide follow-up and evaluation procedures for them.

A PROPOSAL TO CHARACTERISE STATISTICS

Statistics is a scientific discipline that has as a purpose to facilitate or to favour the realisation of inductive inferences based on observations or data by means of: i) the summary of the information contained in the observations, or ii) the measurement and control of the error inherent to all inductive inference. In order to achieve this purpose, the methods, techniques and procedures of Statistics can be classified in:

a) Pre-inferential Statistics, that corresponds to the methods, techniques and general procedures of Descriptive Statistics or Data Analysis, and which has as its objective to summarise the information contained in a set of data without losing relevant aspects for inductive inferences that can be made on the basis of such summary.

b) Inferential Statistics or Statistical Inference, which comprises the methods, techniques and procedures that have as a purpose to quantify and, if possible, to control the inevitable error inherent to every inductive inference based on observations or data (inferential error).

KEY ELEMENTS OF THE PROPOSED CHARACTERISATION

1. *The data or the observations.* The existence of data or observations is a necessary condition for a characterisation of Statistics, and those may arise from the most diverse areas. Statistics must know the way these data or observations were obtained and, on specific occasions, some methods, techniques and statistical procedures are also dedicated to capture them so that they meet certain requirements. Examples of these statistical methods to collect data or observations are grouped under the name of Organisation and Planning of Surveys.
2. *The inductive inference.* The real or potential presence of inductive inference is another necessary condition for statistical work. It must be noted that Statistical Inference renders a service to Inductive Inferences based on observations or data by means of control of inferential error, although Statistical Inference *is not* the only way to formulate inductive inferences. The responsibility for final inductive inferences belongs to the users of statistical services.
3. *Information summary.* The word 'summary' must be understood in a wide sense to include all the operations leading to such summary, for instance, to organise, analyse, interpret, process or

disseminate the information contained in a set of data or observations, etc. It is also convenient to observe that the summary has as a purpose to serve the users to the effect that they can make inductive inferences based on that summary. Therefore, the construction of the summary must be carried out with an awareness of the users' needs. On the other hand, if the user does not understand the summary, he will not be able to make inductive inferences and so Descriptive Statistics will have failed. It is important that the presentation of the summary is simple, clear, and that it includes all the required antecedents for such purposes.

4. *Quantification or control of the error inherent to every inductive inference.* Statistical Inference is distinguished from any other procedure related with the realisation of inductive inferences because its purpose is to quantify or to control the inevitable possibility of error associated to every inductive inference (inferential error). In this way, an application where there is no control of inferential error is not an application of Statistical Inference.

   So, for example, the estimation of parameters without reliability indicators (variance, coefficient of variation, confidence interval, error margin) is not a statistical work, in the same way that the selection of one out of two hypotheses is not if the probabilities associated to committing an error related to that particular selection are unknown. The important and exclusive thing about Statistical Inference methods is not the conclusion reached with that particular inductive inference based on a sample, but the 'measurement' and control of the error associated to such inductive inference.

ADVANTAGES OF THE PROPOSED CHARACTERISATION

1. *It covers and complements the essential purposes of the most usual characterisations of Statistics.* A revision of the purposes associated to other characterisations of Statistics shows that these are directly or indirectly incorporated in the proposed characterisation. And besides, the proposed characterisation corrects the vagueness and limitations of the others.
2. *It is the one the public adopt when they refer to 'statistical information.'* The press and the public in general recognise as an essential objective of Statistics to facilitate the achievement of general conclusions on the basis of the processing of particular observations (inductive inference).
3. *It clearly identifies the exclusivity of its service to other sciences and to decision making.* By situating Statistical Inference as the (only) discipline that allows the scientific treatment of inferring general properties from partial observations or data, it is clear that it distinguishes its aims from those of any other discipline.
4. *It explains the reason why Statistics is present in the most diverse areas of knowledge and in decision making.* There cannot be a characterisation of Statistics without mentioning examples of its applications on different fields, even if the cause for this multiple service capacity is not identified. In the proposed characterisation, this cause is found clearly identified in the support to inductive inferences based on observations or data in the respective fields.

SOME CONSEQUENCES OF THE PROPOSED CHARACTERISATION UPON THE TEACHING OF BASIC STATISTICS

1. *Who should be taught Statistics.* There is no human activity that can make without the inductive inferences based on observations, so Statistics should be part of the bulk of basic knowledge that all individuals need to count with a suitable capacity to apply his intelligence to the surrounding world (cultural wealth).
2. *When the teaching of Statistics should be started.* The teaching of Statistics should be started as soon as the students begin with their first inductive inferences. This happens at a very early age in daily life (for example, when deciding upon waking time), as well as during elementary education through exercises or applied experiments in natural sciences (measurements, associations, etc.) and in the social sciences (comparative charts, graphs and other forms to summarise information).
3. *Who should teach Statistics.* The question itself about who should teach Statistics explains the confusion existing about its aims. For instance, if Statistics is taken as a branch of Mathematics, it will be concluded that the teaching of Statistics, whose aim is to facilitate the making of inductive inferences, will be in charge of teachers with little or no experience in the

scientific practice of this kind of inferences. The only valid answer is that *Statistics should be taught by statisticians.*

Notwithstanding the former, it must be acknowledged that it substantially contributes to the teaching of Statistics the fact that teachers of other disciplines, who deal with inductive inferences in their respective fields, apply Statistics (Pre-Inferential or Inferential Statistic) to reach their respective conclusions. Also, the demonstrations and exercises that the teachers of Mathematics may develop *within their discipline*, and that are directly linked to statistical methods, facilitate the teaching of Statistics too.

4. *How to teach Statistics.* There is no one single way to transmit the educational message. Among other determining factors, teaching depends on the objectives of the teaching activity and on the individual aptitudes and attitudes of each learner. However, it is possible to indicate certain recommendations related to the teaching of Statistics. A few of these are presented below:
   a. *A general recommendation.* In the case of Pre-Inferential Statistics, all teaching should incorporate explicitly or implicitly the data or observations as raw material of statistical work. On the other hand, the presentation of methods and techniques should always take into account that the final aim is to summarise information for the users to achieve conclusions (to make inductive inferences) more easily. However, in the case of Statistical Inference, the basic concepts to be transmitted are those of population, a set of possible samples and associated probabilities, and inductive inference. As for the presentation of methods and techniques, it is necessary bear in mind that their aim is not only to make inferences (they can be made without Statistics), but to measure or to control inferential error.
   b. *About the data.* It is convenient that these data arise from situations of interest for the learners to the effect of obtaining greater enthusiasm and cooperation to assimilate the teaching of the Statistics topic being dealt with. This aspect generally implies the use of computing facilities (at least, electronic sheet and word processor).
   c. *To promote the use of computing facilities as soon as possible.* Easy access to computing services existing today makes it possible to attain this goal.
   d. *The emphasis must be placed on the interpretation of the hypotheses of the methodological model and its results.* The teaching of Statistics must give special emphasis to the hypotheses that support a specific methodology as well as to the results obtained from its application. In particular, it is not wholly necessary to have the mathematical knowledge that supports a specific statistical methodology to teach: i) the objectives of this methodology and to recognise its possible applications; ii) the meanings and limitations of the hypotheses that validate their application (including the use of software); and iii) the interpretation of their results. This is the way to educate an intelligent user of statistical services, whose presence is necessary condition to encourage the development of Statistics and to discourage the activity of those who claim to practise it but know just a little or nothing about the subject.

ACTIONS TO IMPROVE THE TEACHING OF BASIC STATISTICS

Some of the recommendations pointed out in the above section would be implemented more easily if there were coordinated actions to support the improvement of the teaching of Basic Statistics on the part of national and international institutions that gather statistical professionals of the region. Among others, the following supporting actions are suggested:

1. *Distance training in the teaching and application of Statistics to teachers of other areas, in the primary, secondary school and college levels.* The process of improvement in statistical education requires that the teachers of the different levels of formal education (primary school, high school and university) in the different disciplines that deal with inductive inferences, can have enough statistical knowledge to exemplify basic methodological applications of Statistics to the topics they present in their respective subjects.

   The remote training modality in Statistics is efficient for the improvement of these teachers because: i) a large number of teachers need to be trained in the shortest time possible; ii) an agreed conceptual frame for the teaching of Basic Statistics can be ensured; iii) the target

audience for this training is (or should be) people accustomed to studying; and iv) the irregularity of their work schedules makes difficult or impossible any kind of teaching activity with personal attendance sessions.

To carry out the remote training teaching of Statistics it is necessary to prepare written and interactive (for web site) teaching material especially designed for this modality, in separate forms depending on the teaching level (primary, high school and university), and the subject or area of application in different languages. It is suggested that a group of international statisticians, with the support of other national and international associations of similar kind, draw up a project to prepare this material for remote training which, with the support of the interested governments, would be presented to an international agency for funding. Once the material is prepared, the upgrading and improvements might be self-financed by the users.

On the other hand, at the university level it is recommended to establish some coordination between the Statistics Unit of the respective University or Research Institute and other academic departments in order to assist them technically in the degree theses or dissertations that the students prepare in the various careers given. This way it is possible to provide them with statistical methodology for the inductive inferences based on data or observations that are often made in these theses.

2. *'Real' Data Banks Directory for the teaching of Statistics in the most diverse areas of application of this discipline.* There are many Data Banks at the disposal of teachers via the Internet. It would be convenient that a national or international group of professional statisticians take the responsibility for the preparation of a directory of these Data Banks, counting on a web address and a classification system, depending on the area of application from which the data originate.
3. *Files of Study Cases.* The Data Bank Directory mentioned above can be complemented with another, a *Files of Study Cases* of interest for teaching, some of the cases obtained from the Data Bank and others taken from the statistical assistance provided for the Theses of the students of the different professional and technical careers that the university system offers in the countries of the region. The same as the Data Bank Directory, the *Files of Study Cases* should be administered by an association of statistical professionals and catalogued according to the study area.
4. *Statistical education for the public in general.* As stated before, people should have some kind of statistical literacy in terms of the minimal knowledge of Statistics that allows them to understand if a determined judgement or proposition can be based on the statistical results that it indicates. In order to do this, it will be necessary to disseminate the scope and limitations of statistical methods of frequent use as well as the technical quality and the interpretation of statistical results that these methods give rise to. Examples of these methods and their respective results are, among others, those related to the construction of indexes of poverty, labour, prices, quality of life, environmental issues, economic growth, the quality of opinion polls, etc.

   The national associations of professional statisticians, with the support of the international ones, could promote statistical literacy participating in media events (oral, written, television) and in other forms of cultural encounter (theatre), to divulge among the citizens of countries in the region, some guidelines that enable them to recognise the quality of statistical work of general interest, and to have the correct interpretation of their results.