STATISTICA, anno LXXI, n. 2, 2011

STATISTICS AND SCIENTIFIC RESEARCH (*)

Paolo Fortunati

When undertaking scientific research, one poses and reproposes a problem which is not always either explicit or conscious, nor does it actually correspond to the obtained results. The problem lies with the meaning of the results, both specific to the research and to the general cognitive process.

In the most profound Galilean sense of the application, the historical test of the matter is that all research is connected to a higher more complex combination of knowledge, especially when the connection is postulated as irrelevant to science.

Within the boundaries of my knowledge and ability, I hesitated between making a statement and putting a problem about the link between statistics and scientific research. It is for the honest need for truth that I chose the latter. Over time it has been witnessed that all fields of human activity resort to statistical standards and techniques, which in turn become so extensive and elaborate, to the point whereby, listing them becomes superfluous. By doing so, varying interpretations are only to be expected.

One becomes aware of the situation by reflecting on the development of both human knowledge and activity and the actual debate about the cognitive impact and the identification of the statistical standards and techniques. Is this debate technical because it only implicates the technical aspect of statistical models predisposed for measurements and research? Or, does this debate influence the scientific knowledge process too? Is it only a matter of convenience that more and more frequently one turns to statistical results or techniques? Or, although contradictorily, does it reveal both historical sociability of all kinds of human knowledge and at the same time a sort of verification which is more-or-less conscious?

When it comes to statistics, can the question be neutral towards the qualification of knowledge?

In specifying the connection between statistics and scientific research, can one not imply the real process through which, in every society, researchers' professional training is developed and research prospects are outlined?

(*) The text reports the Bologna University's inauguration speech made on December 03rd for the academic year of 1957. This article was subsequently published in "*Statistica*" num. 3 in 1958.

If statistical models are tools for definite knowledge, then Gini's assumption of the need for a systematic and methodical verification of the said models would seem valid. However, once the instrumental nature of statistical modals is posed, is verification only one aspect of statistical techniques? There are certainly gaps in my knowledge, just as there could be mistakes in my language and interpretations. However I believe that the "*Concordia Discors*" (Discordant harmony) which constitutes the inalienable heritage of the university in its historical genesis, is the propulsion and the stimulus for research, provided that one can overcome ones need to make ones beliefs known either because the facts haven't given material results yet, or because the results are not credible or perhaps because they have no scientific relevance.

Salvioni and Gini, Bresciani-Turroni and Vinci all taught or studied statistics, Murri and Viola both taught and researched statistics at Bologna University. Each on of these men, either directly or indirectly, felt and understood with genius application, statistical techniques in sectors which are still bristling with objective difficulties and subjective obstacles. I'll restrict the point to these references to show that, beyond myself, there is a past and a present, in and from Bologna, where this question finds formulations and solutions.

Many benevolent scholars from Bologna or elsewhere, statisticians, economists, mathematicians, historians and physicists/naturalists have communicated necessary and valid indications, judgements and arguments, clarifying the limits and prospects of the pre-chosen theme, and concordant and discordant points, but above all clarifying the interferences between technical solutions and their real cognitive meaning. We're already dealing with a result and a type of collaboration which should be a working method to eliminate ambiguous interpretations and unilateral opinions on more complex themes about which, I will scrupulously comment in the final part of this essay.

If one reads the works of Gini and Boldrini carefully in order to define the function of statistics, differences, which don't seem to be just technical, can be found. Boldrini states that working on specific phenomena is not «statistics». Therefore, applied statistics is no longer to be considered as statistics but should be defined according to the field of interest e.g physics, zoology, demography, experimental psychology, inferential statistics and so on. On the other hand however, Statistics represents «the empirical history of natural sciences». By this, I mean «the exclusive method for investigating natural facts whatever their species may be». In fact, the scientific position of statistics has been resolved since Boldrini, firstly by starting from a configured induction as «the systematic collection of facts and their ideal reduction into homogeneous groups» which «always comes from the use of a logical standard system, and therefore appears as a pure and formal moment». Secondly by accepting a distribution between natural science and noumenal science with the introduction of a group of sciences in composite structure. The latter explicitly outlines a given logical-cognitive configuration, in that the development of the pure and formal moment of the exclusive method regarding natural facts, whatever their species, is independent of any classification. By independent, I mean independent of both the context and the con-

tent of the specific scientific investigation. In fact, on this basis, Boldrini stated that one cannot specify towards research and data formation because «while induction is the methodological moment of science and therefore remains outside the specific technical nature of the individual branches of knowledge, the deduction (namely on a research level) gets to the heart of each one and profits from all their logical and technical resources».

In actual fact, in research plans, surveys are predisposed in view of the formulations which should represent the pure logic moment and there are formulations which arise from the analysis of survey results and fuel other research plans. There are also results from the sciences in composite structure, according to the distribution accepted by Boldrini, and from noumenal science, which are also due to statistical survey results. Thus, the method as pure moment seems forced to me, when considered as a logical standard system taken out of the naturalistic research context, which on the other hand, cannot work without it. Since it wouldn't deal with logical standards, if I understand correctly, but valid logical standards for natural sciences.

On the eve of World War I, while teaching at Padova university, Gini specified statistics as a technique, concluding an idea which he had matured whilst as a student then as an assistant at Bologna university.

Since then, Gini never stopped specifying statistics, even when he analysed in his most recent contributions the differential and complementary characteristics of the statistical and experimental methods, as he always pointed out that the substance of statistics was a technique or a set of techniques. From Gini's statement, the first impression is one of an empirical characterisation.

However, it seems to me that the everlastingly valid Galilean sense of application is present in Gini's specific reference to statistics as a technique.

We shall soon see how Gini's polemic about «statistical models» and in particular designed models for the inversion of probability, isn't reduced to a theoretically irreproachable criticism of the models in themselves. Instead it invests the meaning of the models in the context of substantial scientific research and consequently the characteristic of scientific research.

Recently, sciences influenced by logical neopositivism, and in my opinion arbitrary extrapolations in the theory of knowledge of results from quantum physics and relativity, sciences are configured as techniques that have or can have a distinct and incommunicable language, in that satisfying their internal and consistent grammar represents the necessary and sufficient condition for their own validity, which in turn justifies the creation of hypotheses. This consideration has freed science from problems that are not scientific but metaphysical.

If the intention of not addressing given issues on a metaphysical level was enough to free onself from metaphysics, the issue of how to free science from any metaphysics could be sorted out in no time at all!

However, I think that the «creation» *a priori* of a hypothesis is a metaphysical position. As I believe that in the historical development of the scientific knowledge, the theorized criteria regarding the conformity from the facts to the hypotheses is both formalistic and empirical. Firstly, facts are recorded by an op-

erator who cannot help but record, by definition feelings and/or reflections. Secondly, preliminary creations are mental, one cannot say and should not say anything scientific, beyond what the internal language allows. Whoever creates a hypothesis and matches the facts to the said hypothesis within the definite sense as a human being again, can and must go beyond the internal language. At this point, science becomes non-scientific knowledge. Since it appears unlikely that one cannot, or does not, go beyond the said internal language, the choices, the qualifications and the fundamental prospects actually lie outside a scientific commitment. One cannot ignore that schematized guidelines go hand in hand with a reworking of logical, mathematical, probabilistic and statistical instruments.

One should not confuse results obtained through cognitive instruments with a cognitive definition. It is one thing to talk about scientific knowledge that uses cognitive techniques but it is something else to talk about scientific knowledge as a pure and simple cognitive technique.

In my opinion, the profound Galilean sense of the application is not in the technical experiment, nor in the connection between hypotheses and nor is it in the conformity between facts and hypotheses. The hypothesis is already «a definite abstraction» from the tangible which is a functional mediation between reason-nature and reason-fact. In addition, the conformity is also a functional arrangement (not mechanical) not of raw facts but of typical facts (of an experiment, therefore guided by reason) within the hypothesis.

Galilean research is a cycle: tangible-abstract-tangible and not an overlaying or a mechanical combination of independent «plans» or «moments». Not only that; a given unity of knowledge appears implicit in the Galilean application at the exact moment when an historical rift in the unity of knowledge occurs.

If one accepts the definite abstraction as a reason-nature, reason-fact and abstract-tangible functionality, then the debate reflects the presence of reason: the debate moves out of its incompatible position as it is precisely at this point that the transcendent is, or is not, allowed. One cannot assert that the choice is irrelevant and non exacting in the field of science. The choice cannot help but be reflected in the general interpretation of the results of the cognitive activity, whereas the effects of contributions, even decisive, in individual sectors of the activity may not be influenced by the choice.

Having said that and recognizing the incompatibility between the initial and final positions, can it be excluded that one looks for or carries out, from clearly opposing positions and within the limits of the said position, a unified definition of knowledge when following in the footsteps of Galileo?

It seems to me that every schematic answer is not valid, but one can talk about a historical trend towards an unified approach.

The definite abstraction is neither an operation subjected (once and for all) to a set of standards nor an exclusive operation of natural sciences. The consideration of the hypothesis which closes the Galilean cycle isn't a given technique for a given experiment but a functional need for the truth to which expressions always vary according to the application of each definite abstraction; given that each

starting point always corresponds to a historical and progressive integration of induction-deduction and truth comes from reality.

Critical reflection on knowledge, logic, theory of knowledge, all of which evolve and intertwine during the acquisition of knowledge, can be scientifically founded and analysed in that all types of knowledge can be scientific, provided that firstly a connection between all knowledge exists and secondly all knowledge is based on comprehension and acquisition of truth.

The hypothesis, as a temporary definite abstraction, is already a heritage of knowledge. The latter, already acquired by scientists, comes from observation of facts and a combination of knowledge (more or less conscious). Thereby, even in natural sciences, the hypothesis is born from the connection between the human being and nature; in other words, between history and nature.

Therefore, fact and reasoning cannot be considered separately. Thus all scientific knowledge becomes a heritage for every form of knowledge and can therefore be considered as a cognitive technique in other fields. Mathematics, for example, proves this point perfectly.

However, universally valid cognitive techniques cannot be prefigured since certain types of knowledge cannot be identified according to their cognitive techniques.

Statistics cannot be considered as a unique cognitive technique even for limited knowledge, as it cannot be excluded that statistics is, or becomes valid, where definite abstraction and truth exist.

The historical development of statistics can be neither ignored nor completely rejected. The real history of the acquisition of knowledge reveals that, because of the need for a connection between theory and practice, statistics has developed and is still developing as a research technique and a substantial research, in other words as a statistical technique and survey.

In such cases, it appears legitimate to clarify when the term «scientific knowledge» can be used.

However, one needs to characterize statistics as a technique or research since accuracy cannot be reduced to a simple formalism.

The reference to collective phenomena, in order to consider the theoretical study of statistics as a cognitive knowledge, is certainly relevant and should be critically developed and should take into account knowledge and specific interpretations of substantial scientific research.

I have not expressed any reservations about the scientific opportunity and validity of such a theoretical study. I think that the definition and the global comprehension of knowledge is enough to explain that every form of scientific knowledge both integrates and influences each other.

However, an intrinsic instrumental characterization exists in the statistical technique. The knowledge of statistical technique does not «become» cognitive technique for others forms of knowledge; the knowledge of statistical technique arises, develops and has its own scientific validity inasmuch as it is an immediate and possible cognitive technique.

If we ignore the immediate instrumentality of statistical technique, we are likely to fall into formal models which are not applicable or that, by application, twist the cognitive results of the substantial research: and it is difficult to prevent the mistake of thinking that statistical models have a universal and aprioristic validity.

We do not start, at least, from the body of knowledge already acquired towards given fields of reality, to set up and develop statistical models which are used for a further close examination and an extension of knowledge: and we do not worry about the link between the hypothesis on which models devise and the hypothesis on which substantial scientific research is taken.

So we come to set up models in order to measure an unknown reality, and to consider that substantial scientific research should be conducted or even interpreted on the only basis of hypothesis which influence the models.

It is not about – I say it again – bringing into question the need of a theoretical study on Statistics as a research technique: it is about bringing into question concrete criteria for the accomplishment of such a need. And this to restate the borders of the statistical technique's cognitive validity and of the statistical research: and to point the risks incidental to moods and mental outlooks out, according to which everything is given, everything is regulated and adjustable by statistical models and by given statistical models.

Risks are even more considerable, precisely because the progressive awareness of human activity's sociality, of such activity link with scientific knowledge, of the mutual integration of knowledge in all shapes, of Science as a productive social strength, nourishes the need of a mass phenomena knowledge which characterize social life. Each distortion of statistical models leads to distorted images of such phenomena; and each uncritical view of models leads to identify the whole reality and the whole understanding and the whole explanation of reality in the statistical measures influenced by models.

In some fields of research the multiplication of tests constitutes an inescapable and unfortunately often silent elimination of distorted models; in some fields where the substantial research is observational, the choice of models is, on the contrary, difficult, if not even, in given conditions, impossible; so in every sector one must scientifically question that the whole knowledge is achieved and achievable simply through statistical models; and furthermore, through given statistical models.

The real needs of economic and experimentals activity have expanded, for example, the sample survey technique and the statistical inference in fields activity, which do not have similar needs, and in fields where the distortions that in such techniques are or may be implicit, or the hypotheses, on which basis the techniques themselves are in abstract terms valid, may lead to considerable cognitive deformations.

It is a wild-goose chase to trace from a «sample» to a «mass» without an explicit system of hypotheses, which have to be sifted in their whole extent on specific research fields because, for example, it is a wild-goose chase, without a specific system of hypotheses, to equal the probability that a result depends on an acci-

dental combination and the probability that result is generated by an accidental combination.

Gini had the merit of criticizing with deep harshness the statistical models of mass parameters assessment and the choice of hypotheses, which, as everybody knows, constitute the predominant tendency adopted by the Anglo-Saxon Statistics school.

Even after twenty years have passed since the first explicit Gini's stance, which remained isolated at that time because it was considered as detrimental in a quibbling way to the only protection of theoretical Statistics, it does seem to me that a first assessment may be drawn. Not only in Italy, statisticians and probability theorists are persuaded of Gini's criticism logical harshness.

We tend, therefore, to remove interpretation and statement errors in statistical reversal, by regaining the way traced by Laplace and Gauss for probabilistic applications – respectively for intensive and broad dimensions – and by studying, according to Gini's example, other theoretical models (in addition to those made by Laplace and Gauss) based on a legitimate reversal of Bernoulli's theorem and on the shift from the variability index of the «sample» to the one of the «mass».

Becoming aware of the actual logical structure of theoretical models without deforming their own theoretical meaning, certainly is a step forward and it represents a lesson not to be forgotten.

This lesson, in my opinion, has to be expanded beyond the internal and external inspection of all the hypotheses about «models», which Gini has developed and develops theoretically and applicatively.

It doesn't seem to me enough to correct the interpretation of given statistical models, by making all the hypotheses explicit. Statistical models cannot be constructions that justify themselves, but they must also be a thought-out abstraction from the conditions on which they are applied and from the hypotheses of substantial scientific researches.

The application of an examination must, therefore, be present since the statistical model configuration and must be pursued with a deep coherence in the concrete applications of the model itself.

In this respect we must be exact. The deep meaning of bringing back on research's classical track statistical and probabilistic models does not only consist of a logical adjustment and of an examination of all the suggested conditions, as the comparison between verified data and expected data is not, in itself, crucial about the validity of models.

In fact, in trends, which also adapt to such demand, models intended as a knowledge tool tend to turn into knowledge. The hypotheses, i.e., which support the model tend to adopt the role of interpretational abstractions in the substantial scientific research scene.

Using probabilistic patterns, which can grant useful and in need of prediction results, as well as using statistical models, which allow, on an un-probabilistic basis, to describe, within given temporal and spatial borders, with a sufficient approximation, a given fact, does not allow us to give a scientific explanation that mechanically repeats the hypotheses which are the basis of patterns and models.

It could seem new to contrast chance to cause, or to set up the «statistical decision» as a game between «statistician» and nature; but contrast and game, as they are set, have been inquired and criticized as formal in knowledge theory for a long time.

Apart from this one thing seems to be certain: we cannot launch, or endorse, a general idea of knowledge only by statistical models. In the same way we cannot reduce with statistical models the whole scientific comprehension to the pure and simple formulation of this question: «chance or not chance?». Specially if the question finds a solution by ignoring the knowledge, the experiences, the researches and the history of all the researches and of all the researchers, so that each test, each study is substantially a «new sample», towards whom the question is always one: «chance or not chance?».

So we understand how someone could say that all the hypotheses are false because they never coincide with reality, and so the test of hypotheses has no scientific sense. But in this way we'll lose the sense of the mass survey itself, the sense of the hard substantial research, the watchful and troubled sense of scientific research. It must be said, with strength, that we may lose the way of effective collaboration among all researchers, just when we materially work as a team, when a cooperation is not possible, if not in the comprehension of scientific research, in the acquisition of the past knowledge heritage, which is the only one warranty of the scientific nature of innovation and knowledge development.

The way is open, so, to all theoretical expansions of technical statistics and to all statistical models; but statistical models have to be as such and probabilistic patterns have to remain probabilistic patterns.

The way is open, so, to «samples», when to «samples» we must return: but we mustn't argue, even in an illegitimate way, that from samples we always, and in any case, can get what is achievable from the mass, and especially we have to understand that «samples», data-gathering and mass surveys must coordinate and integrate, and not exclude each other.

An unargued transposition of natural science examination technique is not possible in a social field.

Here, the link and the influence of collective phenomenon, of society taken as a whole, is such that it is ingenuous to study a city in a neighbourhood, a country in a city, a rural society in a rural district, a whole society in some expressions of social life. Even from a technical point of view, a «sample», which is representative for data and qualities, it is not necessarily representative for other ones.

A «sample», which is accidental compared to families considered as such, it is no more accidental when it is used to estimate the unemployment rate in a country, because, compared to unemployed people in a family, it is arbitrary to consider that the existing families are equally possible cases.

The way is open, so, to the theoretical research in the field of statistical techniques, which must not lose the technical instrumentality; which must not drain in probabilistic tools, because the variability of concrete collective phenomena can be non-accidental; which must keep the deep meaning of Galilean research. But this is an open invitation to every statistician to face also the difficult way to

substantial research in his own specific field of knowledge. This is the only way to feel and understand what a «statistical model» may and must be; the only way to feel and understand that substantial research can never exhaust in technical statistics nor in statistical survey, even when substantial research may and must be mainly anchored to such survey.

In this sense, there is a possibility and a demand for statistics as an independent substantial scientific research.

In the ferment of scientific development, in order that statistics as a research technique and as a substantial research may give all its essential contribution, it seems to me that some innovations are necessary in the University system and its organisation.

In the University system several divisions have different needs, so that we cannot impose a standard teaching of statistics, if we wish to change the paradoxical existing situation, so, while everyone recurs to statistical techniques and to statistical surveys results, statistics appears only in a few departments with an undifferentiated role as a teaching subject.

In the different schools statistics must be taught by people who know the specific needs of the schools themselves, so in these departments we must have research technicians and substantial researchers, following a direction which I have learnt in Padova, at Pietra's school.

So we will be able to teach statistics at different levels of research technique – and we will not ignore that in such a range of levels there are specific needs for the technical-productive activity -; both for the research technique and, separately, for the applied research; lastly, both mainly as a concrete application, when in given schools – such as the law ones – what is really urgent is to give a characterisation of the essential aspects of life in a country and of the basic elements we need to provide the possibility of one assessment of public statistical surveys and of one mass knowledge of how the public administration works.

In the area of statistics departments, which are not linked to a specific school yet, and which are equipped with all the most modern tools, all statistics teachers must grow together and they must cooperate and organize experience exchanges with all experts.

In the framework of these horizons we must structure preparation courses and specialization courses in statistics departments, for small and well selected groups of young researchers, coming from different departments in several universities, so they can make a substantial research with a critical assimilation to statistical technique, at all levels considered necessary by their own departments. Programmes must not be schematically arranged: they must be discussed and agreed each time with the University deans and with young researchers, who more and more need to be able to estimate knowledge techniques.

We must operate and toil without any regulations, any course and any official programme, as an original, independent and endless development of university forces. For this purpose we do not only have to consider the need of «design of the experiments», but also the need of an organic gathering of observations and experiences, together with the need of a systematic study on existing «documenta-

tion» and on those to organize according to a plan, which has to consider the concrete conditions in which research is made and the hypotheses which direct the research itself. A similar kind of activity must be realized for small and selected teams of graduated people with specific duties in the public administration and in the economic-productive activity.

Finally, first level degrees in statistics (whose enrolled students have a certain level of specific and general knowledge) must provide the intermediate frameworks which constitute the structure necessary for the execution of operations, calculations, data-gathering, through which the research led with statistical technique is realized.

It seems to me that the good managing criterion is to include statistical technique into each level, always in a well-defined, scientific and practical body of knowledge, avoiding the unscientific position which sees in work and research the technical statistician on a position of hierarchical supremacy, imposing rules and models from above.

In this way we could find concrete specific needs and only by this way models, patterns, statistical technique plans will articulate with the effective instrumentality and the application of examination and theoretical need will develop at the same time.

But it is still in University forces and abilities that we have to found a more efficient organization of scientific research.

Nowadays, in our country there is a need for new tools, for increased financial means, for an increased number of technical and assistant staff, as a prerequisite for a possible efficient development of scientific research on every field. Just for this reason, it is necessary that Italian universities have the task of documenting, with an intense and systematic effort of all the researchers, the link between economic-social development and scientific research development, so as to find again in the historical experience the examination that in modern society Science is a social productive force, and as such it needs an organization which grants the highest performance, in order to spread a crucial orientation to public conscience.

According to this effort statisticians must give their contribution, but University directors and college staff in its functions must, in front of the current conditions and the outlook of scientific research, take the initiative for a deep and methodical knowledge even in the collective life at University, from the functioning of Departments to student population. Through this kind of knowledge – which needs surveys and statistical systematic processing – through this modern functioning of «University governance» we must grant that the needs of a modern «research» organization can be satisfied by an independent University direction, avoiding the certain danger that an increasingly public contribution will end up with a bureaucratic coordination from above and a separation from the University lively body of researchers and researches linked to public organizations but not to the University itself.

In 1857-58 the first Italian statistics yearbook was published by Pietro Maestri and Cesare Correnti.

This two authors wrote then: «Ascetic politics, fateful statistics, which have divined and evoked the new Italy would not be sufficient neither to nourish her nor to describe her, nor to educate her. Figures follow sentences and the ideas themselves wait for a reconsecration by numbers».

Today we may smile at statistics in the first Risorgimento atmosphere.

But since one hundred years have passed, I really see in that atmosphere the historical confirmation of statistics as a social function, which is not only applied as a necessary means of communication from science to science about elements and cognitive techniques, but which is realized also by calling us back to keep the needs of human society, for which and in which we work, in mind.

In 1961, in the first centennial of the unification of Italy, Italian Universities must offer to the world a series of documentation, analysis and researches which have to be the expression of their old and new ability to the endless competition of knowledge.