

THE INTERPRETATION OF STATISTICS IN SOCIOLOGY

Bulletin de Méthodologie Sociologique n°114, p. 50-58.

<http://bms.sagepub.com/content/114/1/50>

Philippe Cibois

(University of Versailles - St Quentin en Yvelines)

Résumé

L'interprétation des statistiques en sociologie, et en particulier de l'analyse des données est confrontée au problème de l'interprétation réaliste ou non des types-ideaux que le chercheur peut repérer. A propos d'une enquête sur la bonne volonté scolaire des enfants issus de l'immigration, on montrera comment on peut hésiter entre une interprétation réaliste et une interprétation constructiviste des résultats. On élargira la question par un dialogue fictif entre des partisans des deux camps.

Abstract

The interpretation of statistics in sociology, in particular in data analysis, faces the problem of interpretation realistic or not the ideal-types that the researcher can identify. About an investigation of the willingness of school children from immigrant families, it will show how one can choose from a realistic interpretation and constructivist interpretation of results. It will broaden the issue by a fictitious dialogue between supporters of both camps.

Mots-clés

Analyse des données, réalisme, constructivisme, philosophie spontanée des chercheurs

Keywords

Data analysis, realism, constructivism, spontaneous philosophy of scientists

INTRODUCTION

There was fairly active talk about mathematics in sociology in the 1970's when, for example, Raymond Boudon, who had trained with Paul Lazarsfeld at Columbia, titled his thesis "L'analyse mathématique des faits sociaux" ("Mathematical Analysis of Social Facts", in English; Boudon, 1967). During the same period, Jacques Maitre (1972) was able to relate "religious sociology and mathematical methods". But there had been pioneers such as Georges Guilbaud (1952) who had rediscovered the Condorcet effect and was behind the creation of the "Centre de mathématiques sociales" (Center for Social Mathematics) or CAMS, within what would become the "Ecole des Hautes Etudes en Sciences Sociales" (School for Advanced Studies in the Social Sciences) which replaced the Sixth Section of the "Ecole Pratique des Hautes Etudes (Applied School of Advanced Studies) or EPHE. The CAMS is still active and still publishes its journal, *Mathématiques et Sciences Humaines* (Mathematics and Humanities, <http://msh.revues.org/>).

At the time, we were dreaming of mathematizing sociology; for example, Raymond Boudon thought the word of clarification done by using mathematical language would allow sociology access to scientific maturity (Boudon, 1971:7). But if you look at what was done, we see that, except for graph theory, the problems discussed revolved around correlation and causality, and were in the end statistical problems.

Since that time, textbooks have abandoned the use of the word "mathematics" which has been replaced by that of "statistics". One studies the analysis of contingency tables, and since Jean-Paul Benzécri, in a factorial form (Benzécri, 1973), and in a log-linear form. Upon all this is grafted, in the case of survey data analysis, the use of significance tests.

A REAL EXAMPLE OF USE

To the left of the graphic we find a number of terms which taken together show a respect for academic standards: being late is bad, you should never forget a book or a notebook, you should do homework, particularly on Wednesday or Saturday, never go to bed late, you prepare school stuff before dinner, you listen to the teacher's advice on what to do, you do not do homework while listening to the radio, you do homework a few days before they are due.

If answer terms for various questions are near each other in the graphic, it means they have been selected in a particular way by a subset of respondents.

From this simple evidence,¹ we proceed to the interpretation: this subset of the population that desires to meet the academic standards can be described by generalizing to an attribute we call here "educational goodwill." The status of this attribute, what is it? Is it the discovery of a reality that exists independently of the researcher in the same way that a geological fault is detected on the surface by differences in vegetation? What is not in doubt here is the fact that individuals have given answers to a questionnaire that described behavior which they had perhaps not thought of before. Did they answer respecting their actual practices, probably not all, but even if their actual practices are different, their responses remain marked by what may be called, without exaggeration, "goodwill" toward the educational system, which is not the case for all respondents.

Indeed, on the right side of the graphic, we have the opposite attitude, not "bad faith", but a "casual" attitude toward school (they often forget their school stuff, they do homework with the TV or the radio on, they do homework just before its due, they prepare their school stuff in the morning, they go to bed later). This attitude can be modulated between those at the top of the graphic, who basically assume this attitude (various use of the term "always", and they find a one-hour class "long"), and those at the bottom, who answer with some reservation (use of terms like "sometimes").

As the graphic is the result of an approximation, verification is necessary. To do this, five answer terms are selected that express educational goodwill and are located to the left of the graphic: being late is bad, never do homework with the radio on, school affairs are prepared for the next day before dinner, follow the advice of the teacher for homework, one-hour classes seem short.

To verify that this combination is not abusive, here is a simple count of the population of respondents indicating the number of persons who have these five terms, those who have four, and so on, down to those who have none. The count gives the following distribution:

Number of "educational goodwill" answer terms used

Total	0	1	2	3	4	5
N=1865	402	662	556	197	44	4
%= 100	21.6	35.5	29.8	10.6	2.4	0.2

The result is both counter-intuitive and reassuring. Of course we expected to encounter an "overdone" attitude with the presence to five terms, but there are only four such respondents. However, 44 with four school terms of good will can be characterized as belonging to what can no longer qualify as a specified group, such as the four "overdone" respondents, but as an abstract type. If we consider the distribution in the other direction, it is clear that those who have no such terms do not belong in this abstract type, just as those who only have one such term. Where is the border? Is it with three terms; in this case, 197 + 44 + 4 or 245 individuals, representing 13.1 percent of the total population? For only two terms, we add 556 and the group now represents 42.9 percent of the total. The choice of membership with at least three terms is justified by the fact that in the former case the group designates a rather small number (with three terms of at least five

¹ "Evidence" which must however be checked of course because, for this method, this "evidence" is an approximation of the data, since only the first two factors are taken into consideration and there are many more factors.

possible terms), which is not the case for the latter case that doesn't discriminate much within this population.

What is clear is that the "type" shown by the graphic should not to be taken as a pure type but as a "sketch" or what Max Weber called an "ideal-type"; that is to say, a set of individuals grouped together by features that make sense together (which we have just seen "make sense together").

This "stylization" of reality, is it useful? One can go further and use it in more detailed analyses that use the possession of this type (three terms at least) as a function of various characteristics that can be tested with logistic regression on discrete data in a procedure called "other things being equal". Here I quote the words used in the survey we are citing:

So we want to explain now belonging to the type «educational goodwill" takes into account the effect of several factors: nationality, distinguishing between those of North African origin and other; sex; social class divided into "lower" (workers, pensioners, without a trade, unemployed or non-response), and the rest (farmers, artisans, merchants, business executives, intellectual professions, middle management, employees) summarized in "upper". Finally, we add the level of mathematics divided between a level described as "good" and the rest (average, fair, poor), described as "not good".

The results are:

Term to be explained: Educational goodwill

Reference Situation: NBon Masc Inf Autr 14.3

Marginal Effects

	Odds-ratio	Percent	Significant at
Good	0.62	-5.0	0.000 ***
Female	1.23	2.7	0.038 **
Upper	0.80	-2.6	0.015 **
Magh	2.39	14.2	0.000 ***

Taking separately the specific effect of each characteristic ("other things being equal"), we see that being a good student (in mathematics) does not favor "educational goodwill" since the effect is negative (significant effect like all the others). A good student is easily distanced himself from the educational institution, is not "scholarly" and does not feel obliged to respect the advice of his teachers (or parents).

The effect is positive for being a female: here we find the behavior often attributed to girls who are said to more "respectful of instructions" than boys. Other things being equal, being a female increased the proportion of those with "educational goodwill".

For the "upper" category (so-called upper because it is mainly characterized by the fact of not being in the "worker" or the precarious categories), there is a negative effect which means that these students, who are not in the "lower" category, feel less of a need to accommodate academic standards. They have a causal attitude toward academic standards.

On the other hand, having parents of North African nationality has a very strong effect on the proportion of "educational goodwill" school which accordingly is increased by more than 10 percent. Obviously, for this subset of the population, obedience to academic standards is seen as something important, regardless of academic success, sex or social class. In the migratory intention of these families, school is seen as a means of social progress and academic standards are seen as something to be respected by the children. As we have clearly seen, not all the children of these

families play the game, but playing this game is very strongly encouraged in this category of nationality.

So we can say in conclusion that in parallel with the desire for a serious education, which had already been noted by Vallet and Caille (1995), academic goodwill is part of the migration project for children of families of North African nationality. However, as demonstrated more recently by Stéphane Beaud (2002), confidence in the educational system is not necessarily rewarded, since the systematic refusal to let young people of North African origin to follow paths of learning and training tends to condemn them to remain in the school system where they can be seen as prisoners (Cibois, 2002).

The results given here are significant at different levels, but this does not prevent us – and it's often done in articles of reference – to verify the results obtained by other methods to ensure their stability.

To conclude on the use of statistics in sociology, we must note that their use is part of a wider problematic. When one uses data from the Education Panel, you already know something about the academic situation of immigrant populations. A sociologist's research on the question is part and parcel of long-term and cumulative results.

A DIALOG BETWEEN REALISMUS AND CONSTRUCTIVUS

There remains the epistemological problem: upon what does the sociologist base his or her interpretation? Does "academic goodwill" belong to reality or is it the result of a constructionist approach? There are arguments both ways that I will stylize as a dialogue between two positions (ideal-types, of course) held by two speakers: *Realismus* and *Constructivus*.

Realismus: The school generates attitudes and expectations that are real because the child who does not respect them is punished in a way that the child does not question their reality. Thus, study the internalization of these practices by children and by their families is simply photographing that reality.

Constructivus: If an observed type of behavior is based on actual practices, the observation of the type, photographing it, if you will, is the result of a construction which we have followed the winding progress and remains based on a scaffolding that some could described as belonging to the subjectivity of the researcher, his or her research interests, and even their beliefs in politics and society. Indeed, investigating the cross between willingness at school and membership in the world of immigration is a specifically marked concern at the social level as belonging to what might be called a "leftist political goodwill" (not say "politically correct"). The researcher could look all different things and projecting his or her attention on the issue that concerns them, thus making them exist socially. In other political and social environments, this question would not even exist and we can therefore say that it is socially constructed by the researcher's interest.

Realismus: You mix up perspectives. I grant you that the focus of a photographic lens is determined by the one who holds the camera. The cliché is the result of a cross between a concern but also of a reality that was preexistent.

Constructivus: Are you so sure? Just because you isolate certain features in your viewfinder doesn't mean that they exist independently of your focus on them. School sanctions for not showing the right attitude only define an "attitude" because you group them together for the needs of your demonstration. The inaccuracy of your ideal type shows that by the way. You have isolated practices that can be linked to other worlds, that of the school, that of the family, that of recreation, each universe with its own logic.

Realismus: But don't these specific logics that you refer to have their own separate existence? Social facts are "things", as Durkheim said, which are just as present as that tree I see or that the table on which I work.

Constructivus: Your table is typically of a social construct of the "carpentry" type and your tree is often an artificial construction, an ornamental plant in a park, which would not exist alone in nature.

Realismus: The tree in front of me belongs to a scientifically-established species.

Constructivus: A species is also a categorization, varying at every moment as shown in Evolution and whose current definition is pragmatic just as that of today's authoritative specialist, E. Mayr (1963 : 19) who defines a species as a "groups of actually or potentially interbreeding natural populations which are reproductively isolated from such groups". A species is a local and temporal categorization which includes individuals with similarities that allow interbreeding; it is not based on "objective" morphological characteristics as originally with Cuvier. It is individuals who carry the genetic material that, by the way, also has a certain variability.

Realismus: So you deny the reality of scientific categorizations?

Constructivus: Absolutely not, but even everyday language is well aware of the priority of a point of view. Let the tree problem again and consider a forest in a natural state, and in a thought experiment, look at isolated trees, perceived individually. Why can they be individualized? That is because they are on a human scale, which shows the importance of the point of view. In a meadow, the blades of grass can similarly be individualized just like the trees on a human scale. But if we consider the prairie as a whole, it is for this reason of scale. Moreover, when we fly over a forest, we have the same reaction and talk about woods, groves, patches of forest, or forests, without feeling the need for a more accurate analysis. Does a tree exist alone? No more than a blade of grass. Either can exist without an ecological environment which assumes the existence of soil, water and carbon dioxide. According to the point of view, we can consider it alone if we are interested in its qualities for construction, or in its environment, if we are interested in its growth. If we are reluctant to deny its individual existence, let's put ourselves at the level of the blade of grass where the change of scale would facilitate the intellectual task. The vocabulary also becomes explicit: the grass is a collective and to individualize it is necessary to grasp the blade, the part of the collective. The tree is an individual and its collective is the forest. It's concern or intention, the human interest leading categorization that is fundamental, not the reality that accepts all points of view. As for the individual, be it a tree or a blade of grass, it can just as well be analyzed according to its components, which also exist, as well as grouped in a collectives that also exists.

Realismus: You admit that reality exists, which reassures me for practical life and for material reality. What would be artificial would be a point of view according to the objectives of the person who made the categories. I agree for the table or even for the tree, but what about for the intimate structure of matter? Atoms and molecules are not the views of the mind.

Constructivus: Physicists who think about these questions – for example Espagnat (1994) or Bitbol (1998) – are more skeptical, the first refers to a "veiled reality", and the second of a "blinding proximity of reality", which in both cases shows that the reality is not that easily within our reach ...

CONCLUSION

The dialogue could go on for a long time. It is important to note that once thinking has shifted toward its own functioning – that is, when a *concept* itself becomes a subject of thought – the question of its existence independent of who's doing the thinking can be asked. For Plato, the concept has an independent existence, since if we acquire a concept, it's because we call into memory pre-existing *ideas*. This realism of ideas was only challenged latter during the Middle Ages by nominalism, which opposed the Thomistic synthesis of truth as *adequatio rei et intellectus*: this adequacy of intelligence and reality continues today to be the spontaneous

philosophy of scientists. A theory has to "stick to reality", it can't "go against the facts", facts are "hard and inescapable".

This position is untenable: how can there be identity between reality and intelligence, if not through a representation by a concept or by a theory. In this case, the theory is not reality, but the "thought of" reality, brought back to the framework of our intelligence. That science produces are models of reality; that is to say, concepts that make hypotheses about the behavior of reality. Experience, by comparing the model and reality, will tell us whether or not the model was more or less correct.

My conclusion is that as soon as we think about a problem, we must abandon "the spontaneous realism of the scientist" which is a pragmatic approach which simply means we believe in what we do. It is naive to ignore the current philosophical debate on this issue and to stick to a simple good old common sense. Ordinary reality does indeed exist, but can be cut in several ways depending on the point of view taken. As for the reality of the scientific process, the further we go from the ordinary world, the more it becomes problematized (but not *problematic* in the sense of being doubtful). As emphasized by Ian Hacking (1999), in his last chapter, even the stones, symbols of hardness, rigidity, have a history, as seen in the case of dolomite where the living things perhaps play a role.

REFERENCES

- Beaud S (2002) *Les enfants de la démocratisation : 80% au bac, et après ?* (Children of the Democratization: 80% Earn a Baccalaureate, and Afterwards?). Paris: La Découvert.
- Benzécri JP (ed.) (1973) *L'analyse des données, I, La taxinomie, II, L'analyse des correspondances* (Data Analysis, I, Taxonomy, II, Correspondence Analysis). Paris: Dunod.
- Bitbol M (1998) *L'aveuglante proximité du réel* (The Blinding Proximity of Reality). Paris: Champs/Flammarion.
- Boudon R. (1967) *L'analyse mathématique des faits sociaux* (The Mathematical Analysis of Social Facts). Paris: Plon.
- Boudon R (1971) *Les mathématiques en sociologie* (Mathematics in Sociology). Paris: Presses Universitaires de France.
- Cibois P (2002) La bonne volonté scolaire. Expliquer la carrière scolaire d'élèves issus de l'immigration (Academic Goodwill - Explaining the School Careers of Pupils from Immigrant Families). In: Blanchard P and Ribémont T (Eds), *Méthodes et outils des sciences sociales. Innovation et renouvellement* (Methods and Tools of Social Science - Innovation and Renewal). Paris: L'Harmattan, Cahiers politiques, 111-26.
- Espagnat B (1994) *Le réel voilé* (Veiled Reality). Paris: Fayard.
- Guilbaud GT (1952) *Les théories de l'intérêt général et le problème logique de l'agrégation* (Theories of the General Interest and the Logical Problem of Aggregation). *Économie appliquée* 5: 501-51.
- Hacking I (1999) *The Social Construction of What?* Cambridge (Mass.): Harvard University press.
- Maitre J (1972) *Sociologie religieuse et méthodes mathématiques* (Religious Sociology and Mathematical Methods). Paris: Presses Universitaires de France.
- Mayr E (1963) *Animal species and evolution*. Cambridge (Mass.): Belknap press.
- Vallet LA and Caille JP (1995) Les carrières scolaires au collège des élèves étrangers ou issus de l'immigration (The School Careers of Foreign or Immigrant Grade School Students). *Education et Formations* 40 : 5-14.