STATISTICS, REALITY, TRUTH

Giovanni A. Barbieri
Istat (Statistics Italy)
Via Cesare Balbo 16, I-00184 Rome, Italy
barbieri@istat.it

Among its fundamental principles, the European statistical law requires reliability, "meaning that statistics must measure as faithfully, accurately and consistently as possible the reality that they are designed to represent". The reference to reality is necessary if one wants to avoid statistics to be just one of the many narratives competing for attention. On the other hand, the statement exposes statistics to a risk of naïf positivism, as it was in Quételet's time. The formulation itself is ambiguous: one thing is to represent, another to measure. How can we avoid being smashed between the anvil and the hammer? Do we need an epistemology for official statistics? And – even more crucially – how can we communicate the scientific principles at the base of official figures?

Stanco dell'infinitamente piccolo e dell'infinitamente grande,
lo scienziato si dedicò all'infinitamente medio (Ennio Flaiano)
Tired of the infinitely small and infinitely large,
the scientist devoted himself to the infinitely medium

BACKGROUND

The European statistical law (OJEU, 2009; OJEU, 2015), before establishing a legal framework for the development, production and dissemination of European statistics, has three articles (under the title "General provisions") devoted to the subject matter and scope of the Regulation itself, and to the relevant definitions and principles.

Among the latter – a very short list of six – one finds *reliability*, "meaning that statistics must measure as faithfully, accurately and consistently as possible the reality that they are designed to represent and implying that scientific criteria are used for the selection of sources, methods and procedures."

Shortly below, the same article states that those statistical principles "are further elaborated in the Code of Practice" for the national and community statistical authority (ESSC, 2011). The Code itself, when dealing with "accuracy and reliability" in its 12th Principle, makes just a fleeting reference to the link between reliability and reality ("European Statistics accurately and reliably portray reality") and the proposed indicators deal with documenting the production process, errors and revision practices.

None the less, the rationale for this choice of the European legislator is clear enough: try and avoid to put statistics in the same class as the many "narratives" competing for the attention of the public. It remains to be understood whether the expected result has been achieved, or whether this extemporaneous excursion in epistemology brings official statistics into even more insidious waters.

KINDS OF REALISM

There are many ambiguities in the formulation of the principle of reliability.

The first is the confusion between "measuring" and "representing": is statistics "designed" to represent reality? Is the representation done through measurement? Through measurement alone? Who is the designer, anyway?

The second has to do with the adverbs used to qualify the measurement process itself: "faithfully, accurately and consistently". Since they are not synonyms, one should assume that each of them contributes to putting statistics as close as possible to reality.

Alain Desrosières discussed the issue of reality and statistics more than once and especially in a paper titled "How Real are Statistics?" (Desrosières, 2001). The problem he is dealing with is different from the one we are discussing here, but there are many contacts.

The point Desrosières makes is that the very notion of *reflecting reality* "implies an intrinsic difference between an object and its statistics." On the other hand, another way of stating

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

the relationship between statistics and reality asks the former to "approximate reality as closely as possible", thus reducing the difference from an ontological (unbridgeable) gap to a mere question of measurement error.

Metrological realism

The second approach has a long tradition in statistics and leads directly to naïf positivism. At the end of the Eighteenth Century, instruments were so imprecise and measurement conditions so varied that scientists (astronomers in the first place) had a problem in determining the "real" position of a celestial body. The difficulty was solved by mathematicians such as Gauss, Laplace and Legendre, through the introduction of the method of ordinary least squares and the concept of a "normal distribution". The solution is also associated with the idea of a "law of large numbers" conceived (in the formulation of Desrosières) as "an operator for the transformation and transition from the world of observation to the world of generalisation, extrapolation and forecasting."

It should be clear now why this way of thinking tastes like positivism and even neo-platonism. Once you accept the idea that astronomical measurements are just approximations to the real position of a star, and that all measures are affected by errors, the temptation to think that the different manifestations of characters in a population and their variability are a sort of measurement error, hiding the "real" observed unit, is immediate.

This is not surprising at all. It is well-known that scientific progress also, if not exclusively, happens by enlarging the scope of models through metaphors (Lakoff and Johnson, 1980).

Following this line of thought, Adolphe Quételet – an astronomer before being a statistician – proposed the concept of *homme moyen*, In Quételet's conception, l'*homme moyen* is not an abstraction, but something real, submitted to a series of imprecise measurements.

> Si l'homme moyen était parfaitement déterminé, on pourrait, comme je l'ai fait observer déjà, le considérer comme le type du beau; et tout ce qui s'éloignerait le plus de ressembler à ses proportions ou à sa manière d'être constituerait les difformités et les maladies; ce qui serait dissemblable, non seulement sous le rapport des proportions et de la forme, mais ce qui sortirait encore des limites observées, serait monstruosité. (Quételet 1835)

After noting that this metrological realism was extended to the social sciences through the sampling method, Desrosières summarises the point as follows: "statistical units are 'homogeneous' (but the definition of the term is ambiguous […]; the distributions of the variables studied do not diverge too significantly from the normal curve; and the law of large numbers can be applied." Thus, according to this sort of realism, statistical measurement "reflects an underlying macrosocial reality, revealed by those computations."

We can laugh at the naïvety of Quételet's remark, but we should also consider how influential this approach is even now. Just peel away that patina of Nineteenth-Century scientism satirised by Charles Dickens (1854), and you find Max Weber's (1904) ideal-types and the ubiquitous (and dangerous) idea that one can forecast the behaviour of a member of a group once one knows the characteristics of its average member.

But if metrological realism has these dangers, what are the alternatives? How to avoid the dire straits between metrological realism and social constructivism?

Social constructivism/constructionism

To say it bluntly, and a bit simplistically, the danger of social constructivism/ constructionism (Berger & Luckmann 1966) is that – at its extreme – it denies reality itself. One should admit that no serious author states (at least in print) that everything is socially constructed. Even the infamous Derrida (1967) quotation ("There is nothing outside the text") is at best a wrong translation ("il n'y a pas de hors-texte", i. e. "there is no out-of-context").

Nevertheless, the notion of "reflecting reality" – as anticipated above – implies a hiatus between reality and its representation, leading at one extreme to the dissolution of objective reality itself:

> Realism and relativism represent two polarised perspectives on a continuum between objective reality at one end and multiple realities on the other. […] Adopting a realist position ignores the way the researcher constructs

> interpretations of the findings and assumes that what is reported is a true and faithful interpretation of a knowable and independent reality. Relativism leads to the conclusion that nothing can ever be known for definite, that there are multiple realities, none having precedence over the other in terms of claims to represent the truth about social phenomena. (Andrews, 2012)

Relativism is a disaster for official statistics: if reality is objective and absolute, it is possible to represent and measure it "faithfully, accurately and consistently" and to expose unfaithful (and eventually inaccurate and inconsistent) representations as false, or fake. However, if there are multiple realities, each a (different) construct, no reality check makes sense. Simply, there is no single reality against which to perform a litmus test. Every representation is faithful in itself, and every measure is accurate.

AN EPISTEMOLOGY FOR OFFICIAL STATISTICS?

If positivistic naïf realism and postmodernistic narratives are both detrimental, is there another way to build the reliability of official statistics?

First of all, is the gap between statistics and its object really unbridgeable? One has just to look at the history of philosophy to see that the reflection on the nature of the hiatus between things and mind has gone on for centuries and has been one of the major drivers of the history of thought. The attempts of reconciliation between an unattainable reality out there (*Das Ding an Sich*, in Kantian terminology) and our mental toolkits has not necessarily led to one of the extremes described above (simplistic realism or extreme relativism) and not even to one of the many incarnations of Cartesian dualism (*res cogitans* vs *res extensa*). On the contrary, arguably the difficulties it implies have not hindered but fostered scientific discoveries.

In fact, we know now that the act of observation (but the same we may say about perception at one extreme and understanding at the other) affects both the observer and what is observed: after interacting, the "thing in itself" is not in itself anymore. Reality is unattainable as such but within reach of observation and scientific inquiry.

On the other hand, observation itself is not for free. Maxwell's demon (Maxwell, 1871/2001) can counteract the second law of thermodynamics by discriminating between slow gas molecules (a small door is kept closed, leaving them in the original chamber) and fast ones (the door is opened, thus moving them to a second chamber). Because faster molecules are hotter, one room warms up, and the other cools down, but at the cost of a "bit" of information for every decision. Entropy and information are two sides of the same coin (Shannon, 1948; Brillouin, 1956, Landauer, 1961).

Statistics and models

The connection between Maxwell's *Gedankenexperiment* and statistics is very strong, via mechanical statistics and Boltzmann. Statistics has at its core a specific answer and a specific method to cope with the difficulties presented above: model building.

Statistics as a science has the object of studying the collective phenomena susceptible of quantitative description and measurement, where collective phenomena are those that occur, with different intensities, into a plurality of subjects or agents. When Maxwell and Boltzmann introduced statistical methods and models in thermodynamics and gas physics, their justification was that the systems to study have a huge number of components. Following the trajectories of each particle (micro scale) is not only useless, but it hinders the comprehension of the system as a whole (macro scale).

This strategy – which is properly statistic – is valid when the many single agents are similar enough to be virtually indistinguishable. Otherwise, microfounded analyses are more appropriate: but these are statistical models anyway.

On the role and usefulness of models, there is a vast literature and even anecdotes and folklore. There is the celebrated short story by J. L. Borges (1975, but originally published in 1946) on the map of the empire: "In that Empire, […] the Cartographers Guilds struck a Map of the Empire whose size was that of the Empire, and which coincided point for point with it. The following Generations […] saw that that vast map was Useless […]"). There is the one-liner attributed to Einstein ("Models should be as simple as possible, but not simpler"). Even if he never wrote exactly this concise phrase, he said something to this effect in the conference in "On the Method of Theoretical Physics" in Oxford (Einstein, 1934): "It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience.")

But why should a model be parsimonious? The answer is once again in Maxwell's thought experiment: because models are costly. Because perception, observation and understanding are costly. On the other hand, understanding implies reaching out for the far-away object (the nearly unattainable reality) and putting it to work in our "system of understanding".

In my opinion, the best clarification of the role of models in statistics, as a scientific tool, comes from an often misquoted George Box. He returned to the issue at least twice, with a slightly different wording and emphasis:

> Since all models are wrong the scientist cannot obtain a "correct" one by excessive elaboration. On the contrary following William of Occam he should seek an economical description of natural phenomena. Just as the ability to devise simple but evocative models is the signature of the great scientist so overelaboration and overparameterization is often the mark of mediocrity. (Box, 1976)
>
> Now it would be very remarkable if any system existing in the real world could be *exactly* represented by any simple model. However, cunningly chosen parsimonious models often do provide remarkably useful approximations. […].
>
> For such a model there is no need to ask the question "Is the model true?". If "truth" is to be the "whole truth" the answer must be "No". The only question of interest is "Is the model illuminating and useful?" (Box, 1979).

Choices and procedures

The reference to "usefulness" does not cause particular problems in discussing "statistics and science" (the specific object of Box's remark) but opens new questions when we look at the other inseparable meaning of statistics: a tool for democracy, a common language informing public debate. The questions at the centre of Desrosières (2010) are those "about the realism of measures, […] pertinence, precision, the sociological salience of the methodological tools"; in other words, those establishing the usefulness of statistic models by reaching a consensus among citizens, experts and stakeholders.

Although he admits that the notion of model encapsulates all the aspects he treats separately, Desrosières finds convenient to single out three steps: quantification (the making of numbers); the use of numbers as variables; and the inscription of variables in full-fledged models.

Quantification has here a meaning different from *measurement*: even if the latter is itself not straightforward once one abandons metrological realism (see above), the former implies expressing by numbers something that before used to be expressed only by words. As a consequence, quantification is always the result of negotiations and conventions.

The move *from numbers to variables* is a process of abstraction (from the poor to poverty, from the unemployed to unemployment, and so on). It is a move from the individual to the collective (and as such it lies at the heart of the statistical approach), but it is also a move from quantification proper (the realm of statisticians and methodologists) to processing and analysis (the turf of subject-matter experts). Variables imply the formalisation of relations (with other variables, although behind them there are also *social* relations) and a pragmatic intent (diagnosis, aims and means, evaluation of achievements and results). It should be clear that this move is crucial in establishing the impartiality of official statistics, and how the citizens perceive it.

Modeling brings matters one step further. Relations are inscribed in a formally consistent architecture. The pragmatic intent is completed by retroaction and feedback loops so that the indicators resulting from the model influence the behaviour of the various agents involved.

CONCLUSIONS

So we are left with the final question: how can we communicate the scientific principles at the base of official figures?

I do not even take into consideration the temptation not to communicate the aspects that could be perceived as shortcomings (the unattainability of the "whole truth", uncertainty, the conventional nature of crucial processes such as quantification, the making of variables and modelisation).

Without the ambition of being exhaustive, I fix here just a few points.

Numeracy and data literacy

Let us define *numeracy* as the ability to reason and to apply numerical concepts (including include number sense, operation sense, computation, measurement, geometry, probability and statistics). *Data literacy*, on the other hand, is the ability to understand and reason with data, or arguments that use data, and to apply this competence in deliberations and decisions affecting personal, social and political life.

In this context, data literacy is central to the mission of official statistics, and numeracy is a set of tools and competencies for data literacy. Both are necessary, but I envisage the first as a task for school's curricula, while the latter should be prominent in the activities of statistical offices.

Beware of storytelling

In his address to the UK Royal Statistical Society, its president (Spiegelhalter, 2017) said that deliberate fabrication ('fake news') is not the main issue because one can debunk them by a combination of fact checking, crowdsourcing on social media, and algorithms. A much bigger risk – he continues – is "manipulation and exaggeration through inappropriate interpretation of 'facts' that may be technically correct but are distorted by what we might call 'questionable interpretation and communication practices'."

Storytelling is often the vector of these practices. Moreover, storytelling – according to Desrosières (2010) – is at odds with sound statistical procedures:

> The notion of "making a variable" is orthogonal to that of "making a narrative" used in the historical sciences. In fact, we can imagine a cross-table with individuals in the rows, and a standardised collection of variables about these individuals in the columns. The making of narrative means reading along the rows of the table, whereas the making of variables means reading in columns, and then the confrontation between these columns via mathematical statistics.

Access to experts

It has been a commonplace, for decades already, to affirm that general use statistical data presented in tables, once the staple product of statistical offices, are by now a commodity, while tailor-made outputs and analyses, aimed at particular segments of our customer base, are the specialities. But enabling technologies, competition and the availability of specialised tools are raising the bar once again, commodifying what were once specialities.

The strength of statistical offices rests on its human capital. A clue of this potential is in the gap between UK citizens trusting the accuracy of official figures (78%) and those trusting the Office for National Statistics (90%) (NCSR, 2017). The experts with the ability to put data and metadata together, giving insights on the meaning and the usefulness of official statistics, can trigger a virtuous circle, where producers and users coevolve in their capacity of using data. To give free and easy access to this expertise should be a priority.

Trust and trustworthiness

Statistics: a matter of trust was the title of the Green Paper presented by UK's Prime Minister in 1998. Of course, it is. However, if you want the trust of your users, you need to be trustworthy yourself.

According to O'Neill (2013), "trust is the response, trustworthiness is what we have to judge". To be trustworthy one needs to demonstrate competence, honesty and reliability. Plus, one has to "make oneself vulnerable", providing the means for others to check whether one is trustworthy.

How does this apply to official statistics? "There appear to be two main ways of ensuring that trustworthiness can be properly assessed: training audiences in critical appraisal, and encouraging platforms dedicated to response and 'calling-out'" (Spiegelhalter, 2017).

Critical appraisal, in particular, should question the analysis itself (internal validity: are the numbers trustworthy?), its interpretation (external validity: are the conclusions drawn trustworthy?), and its communication (the 'spin': is the source trustworthy? What is it really telling me?) (Spiegelhalter ,2017).

REFERENCES

Andrews, T (2012). What is social constructionism? *Grounded Theory Review*, *11*(1), 39-49.

Berger, P. L. & Luckmann, T. (1966). *The social construction of reality: A treatise in the sociology of knowledge*. Garden City, NY: Anchor Books.

Borges, J. L. (1975). *A universal history of infamy*. London: Penguin Books.

Box G. (1976). Science and statistics. *Journal of the American Statistical Association*, *71*(356), 791-799.

Box G. (1978). Robustness in the strategy of scientific model building. In Launer, R. L. & Wilkinson, G. N., *Robustness in statistics* (pp. 201-236). New York, NY: Academic Press.

Brillouin L. (1956). *Science and information theory*. New York, NY: Academic Press.

Derrida J. (1967). *Of grammatology, Part II Introduction to the "Age of Rousseau"* (pp. 158-59 & 163). Baltimore, MD: Johns Hopkins University Press.

Desrosières, A. (2001). How real are statistics? Four possible attitudes. *Social Research*, *68*(2), 339-355.

Desrosières, A. (2010). A politics of knowledge-tools – The case of statistics. In: Sangolt, L. (Ed.), *Between enlightenment and disaster* (pp. 111-129). Brussels: P.I.E. Peter Lang.

Dickens, C. (1854). *Hard Times*. Available from www.gutenberg.org.

Einstein, A. (1934). On the method of theoretical physics, The Herbert Spencer lecture, delivered at Oxford, June 10, 1933. *Philosophy of Science*, *1*(2), 163-169.

European Statistical System Committee (2011). *European statistics code of practice*. Retrieved from http://ec.europa.eu/eurostat/documents/3859598/5921861/KS-32-11-955-EN.PDF/5fa1ebc6-90bb-43fa-888f-dde032471e15

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: The University of Chicago Press.

Landauer, R. (1961). Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development. International Business Machines*, *5*(3), 183-191.

Maxwell, J. C. (2001). *Theory of heat.* New York, NY: Dover. Original work published in 1871.

National Centre for Social Research (2017). Public confidence in official statistics. London: NCSR. Retrieved from http://natcen.ac.uk/our-research/research/public-confidence-in-official-statistics

Official Journal of the European Union (2009). Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics… Available from eur-lex.europa.eu.

Official Journal of the European Union (2015). Regulation (EU) 2015/279 of the European Parliament and of the Council of 29 April 2015 amending the Regulation (EC) No 223/2009 on European statistics. Available from eur-lex.europa.eu.

O'Neill, O. (2013). What we don't understand about trust. Retrieved from www.ted.com/talks/onora_o_neill_what_we_don_t_understand_about_trust/transcript

Quételet A. (1835). *Sur l'homme et le développement de ses facultés: ou essai d'une physique sociale, Tome second*, pp. 266-267. Retrieved from books.google.fr.

Shannon C. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27,* 379-423 & 623-656.

Spiegelhalter D. (2017). Trust in numbers. Address delivered on June 28th, 2017. Retrieved from www.rss.org.uk/Images/PDF/events/2017/President%27s-address-June-2017.pdf

Weber, M. (1904). Die "Objectivität" sozialwissenschaftlicher und sozialpolitischer Erkenntnis. *Archiv für Sozialwissenschaft und Sozialpolitik*, *XX*, 22-87.