The identity matrix

When my daughter Margherita was attending primary school, she was asked by her teacher to describe her parents' jobs. She wrote that her mother was a biologist working with bacteria, microscopes, Petri dishes and pipettes, whereas her father was a kind of biologist working with numbers. After all these years, I remain deeply convinced that with this ingenuous, childish answer she aptly defined those who work in the field of biostatistics, or medical statistics or biometry, i.e. in the field of "quantitative biology" to use Kenneth Mather's expression (1). All those who call themselves biostatisticians, or something of that kind, can be regarded as quantitative biologists, where the term biology is used in a broad sense, embracing psychology, sociology, ecology, anthropology, ethology and, of course, of all the branches of medicine and the health sciences.

The term biometry brings us to the very heart of scientific-experimental activity: the devising of methods for measuring phenomena, so as to be able to distil, from vague qualitative impressions, a well-defined quantitative representation (i.e. a model) of certain, important aspects of the real (for the objectivist) or experiential (for the constructivist) world. The process of modelling begins with the formulation of a promising hypothesis (e.g. air pollution increases the risk of cardiovascular events); this is followed, in this case, by the choice of the pollutants to be monitored (oxides – SO_x, NO_x, CO, CO_2, volatile organic compounds, particulate), the determining of how they will be measured (the instruments and methods), as well as when and where (sampling over time and space), and the definition of the cardiac events to be monitored (e.g. all fatal and non-fatal cardiovascular events, only events requiring hospitalisation); we then make some assumptions on the relationship between the risk of an event and air pollution (e.g. the risk is proportional to the concentration of pollutants, the random term follows a Poisson distribution), and finally sketch an outline of the expected results. A biostatistician must always remember that unless he grasps the biological essence of the study being planned, all his knowledge of theoretical statistics will be completely useless. None of the complicated and sophisticated techniques currently used by "trendy" statisticians (e.g. all the computer-intensive procedures) is as important as the ability to understand the dynamics underlying a complex system (the brain-intensive method). As Malcolm C. Rorty (2) wrote "The statistical work in astronomy must be done primarily by astronomers, and in biology by biologists, and in each science by those who are trained primarily in that science and only secondarily in the statistical method." A biostatistician should be a biologist or, at least, think biologically.

"Karma police, arrest this man, he talks in maths. He buzzes like a fridge, he's like a detuned radio" (3). These words, taken from a famous song, well express the distaste that biologists and medical doctors feel for mathematics. If you think that this quotation is too frivolous, let me cite Wladyslaw Z. Billewicz (4) "The statistician must be able to reduce all the concepts he wants to put across to plain English, even if this involves drastic simplications (sic!). Failing that he may find himself presented with experiments executed without previous consultation and being asked 'to produce means and standard deviations, you know, just enough to satisfy the editor'." A biostatistician should always choose the simplest model. Of course, simple does not mean simplified (i.e. unrealistic), or unduly parsimonious. A simple model is a model that is easy to explain and that can easily be translated from formulas (however complicated) into biological concepts that shed light on some mystery of the experiential world. Every time a new, more complicated model is introduced, the biostatistician should show its benefit/cost ratio, justifying, in terms of the gain in insight, its loss of simplicity. Conversely, every time a new, simpler model is introduced, the biostatistician should explain that the gain in simplicity outweighs the loss of insight, although in my experience this is never the case.

This fledgling journal should give priority to papers dealing with methods, i.e. with ways of tackling problems that may be encountered by many of those working in the area of biostatistics and clinical epidemiology, rather than papers presenting results or techniques. Results should be published in specialised journals – there are plenty of those! As for techniques, it is necessary to be wary of the casual use of ones that are superficially learned rather than deeply comprehended, as this may have unpredictable and inexplicable consequences when the so-called biostatistician knows little of the experiential world that he is supposed to be describing or even predicting.

I have tried to show that method means much more than two or three pages of mathematical formulas (let's leave those to the mathematicians), distribution functions (let's leave those to theoretical statisticians) and computerised algorithms (let's leave those to computer scientists), and that excessive focus on techniques often blocks our view of the method (μέθοδος), in other words of the way (ὁδός) and what lies beyond (μετὰ) – the destination.

Silvano Milani

PS

"What does the title of this editorial mean, Mr Holmes?" "It is simplicity itself, my dear Watson. A single scalar may not be enough to describe the identity of the biostatistician."

References

1. Mather K. The Elements of Biometry. London: Methuen & Co 1967: 1-7.
2. Rorty MC. Statistics and the scientific method. JASA 1931; 26: 1-10.
3. Yorke TE, Greenwood JRG, O'Brien EJ, Greenwood CC, Selway PJ. Karma Police. London: Parlophone Records 1997.
4. Billewicz WZ. Statistics in medical investigations. The Statistician 1934; 14: 121-136.