212

DEVELOPMENT OF THE THEORY OF EXPERIMENTAL DESIGN

by R. A. Fisher
University of Cambridge (Great Britain)

At the time, now about thirty-five years ago, when first I came to study statistical methods, nothing was further from my thoughts, or from those of my contemporaries, than that the art of experimental design would ever come to be, as it now surely is, an integral part of the subject. The horizon was indeed filled with problems of *fitting*, or as we should now say, of estimation; and the two primary tests of significance, Pearson's χ^2 test, and Student's test, though certainly not widely understood, were already available, though in the case of χ^2 still in a form that ignored the degrees of freedom. The mathematical theory of all these aspects of statistical methodology was frankly in chaos. No step forward was possible without concurrently clearing up a number of errors and misapprehensions, only indirectly connected with the matter in hand.

The line of statistical thought which led most directly to the problems of experimental design, was undoubtedly the Theory of Estimation, devoted to the study of the distributional properties, in appropriately chosen sets of samples, of the various possible estimates, or estimators, which may be used to estimate the unknown parameters in the problem. Even though at this stage of discussion no thought was given to the manner in which the data were obtained, but these data were considered as an immutable part of the problem, yet it gradually became clear that

Proceedings of the International Statistical Conferences, (Washington), 3: 434-439, (1947).

absolute criteria existed by which one estimate might be judged, without ambiguity, to be better than another, and that in the Method of Maximum Likelihood the practical worker possessed a tool providing estimates which even in theory could not be improved upon.

In the course of this work it was shown that one valid method for the comparison of estimates was supplied by the amount of information relevant to the unknown parameters, this quantity being a scalar for a single parameter, and a symmetrical matrix in general. It was rigorously demonstrated that the amount of information extracted in the process of estimation could never exceed the quantity supplied by the data.

This finding, combined with the practical fact that directly available processes of computation would extract almost always a very large fraction of the total available, shifted the *moral* balance. The statistician was no longer responsible for the accuracy or precision of the results of his labors. His business became much less like that of a conjurer who is expected to work wonders, and more like that of a chemist who undertakes to assay the proportion of gold in a sample of ore. He need not be ashamed if the assay is low, or elated with pride if it is high. Consequently, provided he used available methods of proved competence, and did his arithmetic right, the statistician had no longer any responsibility for the value, or for the worthlessness, of his estimates. The weight of his responsibility was thrown back on to the process by which the data had come into existence. So, armed with the concept of amount of information as a practical tool, statisticians came to study what forms of experiment, or what types of observational programs would yield the most information for a given expenditure in time, money, and labor.

The purpose of these introductory remarks of mine will have been served if I have made clear the logical necessity of this step in the history of the subject, by which interest shifted from the problem of minimizing loss of information in the processes of statistical reduction, to the problem of maximizing the amount of information created by a well planned expenditure of experimental resources. There is but one further point I need make.

The only modification of the art of experimentation, as developed on statistical principles, which does not flow directly from the requirement to maximize the amount of information, obtainable with limited resources, consists in the introduction into the experimental procedure of what is called *randomization*. The theory of estimation presupposes a process of random sampling. All our conclusions within that theory rest on this basis; without it our tests of significance would be worthless. In order to justify the conclusions of the theory of estimation, and the tests of significance as applied to counts or measurements arising in the real world, it is logically necessary that they too must be the results of a random process. In controlled experimentation it has been found not difficult to introduce explicit and objective randomization in such a way that the tests of significance are demonstrably correct. In other cases we must still act in the faith that Nature has done the randomization for us, as for example, when we conduct a series of variety trials over a period of five years, and draw our conclusions *as if* the weather experienced in these years were really a random sample of five out of an infinite population of seasons determined by the climate prevailing at the site of experimentation. In all cases, however, we now recognize randomization as a postulate necessary to the validity of our conclusions, and the modern experimenter is careful to make sure that this postulate is justified.

My few remarks, Ladies and Gentlemen, are aimed at displaying the seed and roots of this young plant; later speakers will I hope tell us something of its flowers and fruit.

Résumé

L'auteur expose le développement des idées ayant conduit, via la Théorie de l'Estimation, à l'étude des plans d'expériences.

On s'approfondit d'abord sur le problème de minimiser la perte d'information dans le procès de réduction d'un ensemble d'observations, plus tard sur le problème de rendre maximum la valeur informative résultant d'une utilisation aussi efficient que possible des ressources disponibles. L'on a introduit aussi dans la procédure expérimentale l'exigence de fortuité.