

# The Principles of Humane Experimental Technique

**W.M.S. Russell and R.L. Burch**

## CHAPTER 6

### REDUCTION

Many laws regulate variation, some few of which can be dimly seen, and will... be briefly discussed.

#### **The Design and Analysis of Experiments**

The science of statistics has been connected historically with three large-scale human activities: biological research, insurance, and gambling. The connection with biology began with Galton and Pearson at the turn of the century. The great progress in the first half of the present century has been associated especially with life insurance and with two branches of biology--experimental agriculture and the theory of genetics and evolution. Haldane in this country, Lotka and Sewall Wright in the United States, have all made important contributions, but preeminence in this field must be accorded to Sir Ronald Fisher. His great book on the Design of Experiments (1942) is still a classic, and more than anyone else he is responsible for bringing statistical methods into experimental biology. Today statistical methodology is a large and flourishing science, of which a substantial part is concerned with experimental technique. Statistical methods began to be introduced into bioassay by such pioneers as Gaddum and Trevan (cf. e.g. the latter's obituary by Buttle, 1956). By the thirties, the subject was expanding rapidly. Application to bioassay problems began to be made on a systematic basis, notably by teams sponsored by the M.R.C., and especially by Emmens in a series of important reports to that body. Emmens himself is responsible for one of the clearest and simplest assays (1948). More recently, this subdivision of the subject is being vigorously carried forward by Finney and his associates, and elaborate treatment of the more complex problems is provided in Finney's texts (1952, 1955). In large-scale bioassay work, statistical methods of a sort are regularly used by now. But there has been a certain lag, and some available tests have probably not been exploited to the full even in research immediately after their provision. Hume (1957b, c) has cited Fisher's exact two-by-two test as an instance of this. Time is money for a commercial firm, and simplicity of procedure, both in design and computation, is

crucial in practice. The increasing availability and cheapness of computing machines may be of service here.

Every time any particle of statistical method is properly used, fewer animals are employed than would otherwise have been necessary. The whole subject has twice been surveyed by Hume (1947a, 1957b, c) from the human point of view, and we shall here mention only a few cardinal points.

Failure to make some of the planned observations is a common misadventure in many experimental procedures. Statisticians are justly indignant if asked to cope with the results of bad design. Of course it is an elementary principle for any experimenter, not himself a statistician, to seek advice *before* experimenting, though this may cease to apply in assay work once a routine has been established. But statisticians are more indulgent to unavoidable accidents, and some of them have to complete the planned observations (see, for instance, Sampford, 1952). The alternative to salvage (e.g. to a mode of analysis which allows for the lost observations) is repetition of the experiment with more animals. This branch of the subject is therefore an important means of reduction.

For reduction purposes, as we have noted, the statistical method has a key property--it specifies the minimum number of animals needed for an experiment. This statement needs qualification. It certainly is always possible, in accordance with the arbitrary but workable concept of significance level, to decide *after the event* whether enough animals have been used. This saves needless repetition, and where, as sometimes in bioassay, workers are familiar with the amount of variation to be expected, a number found to give significant results can be fixed upon for regular practice. Exact treatments of the problem of choosing the right number *in advance* on the basis of experience are limited in scope so far. The problem is discussed and guidance supplied by Hume in the papers mentioned. But unexpected variation is liable to arise from time to time in species used in bioassay, and repetition of assays may then become necessary.

There is already available a technique which may be helpful here--that of sequential analysis, chiefly developed in relation to quality control in nonbiological industry. This is a method of conducting experiments in the stages.

"The determination to terminate the experiment depends, at each stage, on the results of the observations previously made. A merit of the sequential method... is that test procedures can be constructed which require, on the average, a substantially smaller number of observations than equally reliable test procedures based on a predetermined number of observations" (Wald, 1947)

--for determination has to allow for more residual variation than may actually arise. This method was called to the attention of doctors a few years ago as a useful one for clinical research (Annot., *B.M.J.*, 1954b). It is readily applicable to bioassay. Hormone preparations made and used by one of us (W.M.S.R.) were assayed in this way (Russell, 1954) with the help of a statistically experienced colleague (D. Michie). We have heard that the method is already in use in one large pharmaceutical laboratory. Since it was initially designed in a quality control context, this mode of analysis seems ready-made for batch-testing for toxicity.

Toxicity testing, as usual, is the scene of some confused thought, which may be delaying the exploitation of statistical methods. We have not infrequently heard the opinion expressed that, while you cannot have too much uniformity in bioassay, in toxicity tests you need a thoroughly heterogeneous mass of animals, and plenty of them. The physician, it is argued, is going to deal with patients with a very wide range of sensitivities to a given toxic action. There is a vague feeling that since this variation is quite uncontrolled, that of test animals ought to be uncontrolled, too. It is a sort of high-fidelity argument, this time applied to the properties of populations.

On this subject Hume has written clearly and concisely (1957c):

"The fallacy consists in supposing that in order to obtain a broad inductive basis a heterogeneous stock should be used. It would be as if you were to estimate the value of a pocketful of silver by counting the coins as coins, without sorting the sixpences, shillings, and half-crowns. The proper procedure is, of course, to use several different homogeneous samples, by using a plurality of pure lines (or preferably F1 crossbreeds), and to allow for the variance between samples; for otherwise the experimenter deprives himself of the possibility of making a relatively precise estimate of the error (Fisher, 1942)."

A great variety of relatively pure lines of laboratory animals is now available, with many known physiological differences between them (Elizabeth Russell, 1955); nor is selective breeding the only way of producing several stocks, each of them uniform but different from the others in respect of some physiological property, such as sensitivity to a toxic effect.

This example raises the most fundamental principle of statistical method from the biologist's point of view. The analysis of variance (the basic statistical tool in bioassay and most other experimental contexts) depends for its success on the *isolation* of as many as possible of the sources of variance, and all designs are constructed on this basis. Variance contributed by each isolated factor can then be assessed by comparison with residual, uncontrolled variance, which should be as small as we can possibly make it. The mere segregation of differences between individual animals has

greatly increased the precision of many assays, and the elimination of differences between litters has made possible yet further increase (Emmens, 1948). Strict randomization must be fed into the system in appropriate ways. Since (as Freud observed in another context) the human brain is supremely inefficient as a generator of random series, this is by no means a casual procedure. Randomness is, of course, a purely relative term, meaning the absence of systematic relation to the problem that interests us (Russell, in press, b, c). But strict randomization procedures ensure that, along any one line of variation, only the factor we are varying (or whose variation we are passively observing) is working systematically. They enable us to ascribe the residual variance to a large number of unknown competing variables. The smaller we can make this residuum, the more information we derive from the experiment. The problem is essentially that of improving a noisy channel of communication.

Where differences between individuals and litters are concerned, we do not directly control the variance. We simply control its intervention in the experimental results. But statistical methods are not intended to absolve us from deliberate control of variance-reducing factors. If, for instance, we can literally remove any unwanted source of variance, we reap our reward at once in smaller residual variance, greater precision, and hence fewer experimental animals.

Often we can begin by controlling large groups of variables, about whose composition we know virtually nothing, in simple practical ways, by manipulating a *blanket* variable which contains the whole group (Russell *et al*, 1954). For instance, we can so design an experiment as to isolate variation due to residence of the animals in different cages. There is still a matter of design alone. But if systematic variation along this line turns up in the analysis, it may provide a clue for real control. To take a trivial example, the cage groups may be found to vary systematically with the distance of their cages from a source of light or heat, which we can then deliberately adjust.

Statistical methods, then, enable us to take the fullest advantage of our capacity directly to control the factors causing variation between animals, or within one animal at different times. This control may take the form of simply reducing variation (increasing uniformity). Or, as in the toxicity examples, it may mean harnessing the controlled variation for our own purposes. In either instance, statistical methods are invaluable, but they cannot replace the understanding and deliberate control of the factors causing variance in physiological responses. The systematic quest for this control is a very recent development, to which we now turn.