DISCUSSION OF REPORTS ON CLOUD SEEDING EXPERIMENTS

FRANK YATES
ROTHAMSTED EXPERIMENTAL STATION, HARPENDEN

1. General

The work on cloud seeding which has been described in this Symposium is new to me. I find it extremely gratifying to encounter such well designed and carefully conducted experiments in which the principles of randomization have been scrupulously adhered to.

A few of the speakers appear to think that there is some conflict between a "statistical" experiment, fulfilling the requirements of experimental design, and an experiment which because it ignores these requirements can be conducted with greater freedom. I think this is a misconception. Any experiment dealing with variable material should have built into it the appropriate random elements. If this is not done, confusion and false conclusions will ensue, and there will be endless arguments as to what the results really mean.

The real contrast, I think, is between a series of detailed short term experiments on single clouds and the like, undertaken solely with the object of gaining a better understanding of the physical processes involved, and long term experiments designed to determine the benefit, if any, resulting from some already determined method of seeding (or alternative methods). It should not be thought, however, that the latter type of experiment is incapable of throwing light on the underlying physical processes. If sufficiently objective and refined physical measurements are taken it may be very revealing.

Detailed short term experiments must, however, be properly designed, with appropriate randomization. Physicists, accustomed in much of their laboratory work to handling material with little variability, often neglect this requirement. The head of our own physics department at Rothamsted, Dr. Penman, talks of "getting the Fisher out of physics." I think he is mistaken, as, I may say, do the Rothamsted biologists.

2. Concomitant observations

One reason, I think, why there is this conflict between physicists and statisticians is that it is insufficiently realized that much more can often be elucidated from an experiment than is obtained by the mere comparison of crude means. Given concomitant observations on various physical and meteorological phenomena the results can be examined to see whether, singly or jointly, they influence the results. One standard and convenient method of doing this is by covariance analysis, but in simple experiments of the type we are considering subdivision of the results into groups depending on one or more concomitant variates may be equally, or more, revealing. Professor Neyman has given us one example of this in the subdivision of the results of the Swiss experiment by forecasters.

I think that as statisticians we can help considerably here. Although the principles of the methods are well known, their application to extensive experimental results requires a certain statistical finesse, and the publication of a few good examples will be very useful to practical experimenters.

I should, of course, emphasize that if the apparent influence of many concomitant variates is investigated a few are bound by chance to turn out significant. This never worries me, but the necessary warnings should be given and confirmation sought from other experiments. This provides a strong argument for the catalogue of experiments that Professor Neyman is hoping to establish.

Whether examinations of this kind on the present experiments will reveal much that is useful I would not know, but I think they are worth trying. It does seem, however, imperative to try and get a better understanding of the physical processes involved, and for this, detailed short term experiments may be necessary.

3. Long term effects

These to me are very puzzling, particularly as no one seems able to put forward any really convincing physical explanation. I can only offer two suggestions. All experiments now completed or being conducted should be examined for these effects, and the evidence assembled as a whole so it can be critically appraised. Second, future experiments should be so designed that they will provide direct evidence on such effects. Also, I think, the physical behavior of silver iodide after seeding should be studied further to see if any physical explanation can be found.

4. Estimates of effect of seeding

Most of the papers presented at the Symposium gave ratios of the rainfall in the seeded and unseeded areas (or similar ratios) but no estimates of the standard errors of such ratios were given. I strongly agree with Professor Bradley that such standard errors are desirable. We are not only concerned with testing whether seeding has any effect, but also with the accuracy of the estimated effect, even if this is not significantly different from zero. Once standard errors are given it is possible to combine estimates from different experiments, or from different parts of the same experiment, in whatever way appears appropriate, without reference back to the original data.

A few words on methods of estimation. Once the variance law has been established, a suitable transformation can be used. It has been rightly observed, however, that the means of transformed data, if transformed back, are biased. A suitable procedure for removing this bias, based on maximum likelihood, is outlined in [1]. This procedure appears to be not well known. I would be interested to see it applied to this material.

The use of a transformation implies that the data conform, at least approximately, to the associated variance law. An alternative approach is to work with ratios directly. As an example we may take the case of rainfall on two areas, one of which is selected at random for seeding at each test. Instead of taking the ratio of the seeded to the control rainfall, we may conveniently consider the ratio of the difference (seeded-control) to the mean of seeded and control areas. Denoting the values of these for each test by y and x, respectively, we can plot y against x, and fit a regression $y = \beta x$ passing through the origin.

If the variance of y is proportional to x the efficient estimate of β is given by

$$b = \frac{Sy}{Sx}, \tag{4.1}$$

and the estimate of the variance of b by

$$V(b) = \frac{S[(y - bx)^2/x]}{(n - 1)Sx}. \tag{4.2}$$

If the variance law is not known, we can still use the above estimate of b, which has the advantage of being what the practical man wants, but the above variance estimate is suspect. Instead we may use the estimate given by

$$V(b) = \frac{S(y - bx)^2}{(n - 1)(Sx)^2/n}, \tag{4.3}$$

which is unbiased (at least approximately) whatever the variance law, but which will give a relatively inaccurate estimate of error.

It may be noted that the use of the mean of control and seeded areas as the denominator of the ratio avoids the difficulty of zero values in the control or seeded areas or both.

The above methods are clearly capable of adaptation and refinement, and are I think, worth further exploration. One simple way of seeing how they behave in practice is to make parallel computations on the same sets of data, using one or more suitable transformations and the methods outlined above. I hope some comparisons of this kind will be published.

REFERENCE

[1] R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural, and Medical Research*, London and Edinburgh, Oliver and Boyd, 1957 and 1963 (5*th* and 6*th* ed.).