FRACTIONAL REPLICATION IN INDUSTRIAL RESEARCH

CUTHBERT DANIEL
NEW YORK CITY

1. Introduction

It is interesting to compare the current state of industrial experimentation with the situation in experimental agriculture twenty years ago. The distribution of statistical knowledge among workers in the two fields seems to be about the same. The two groups would agree that their problems are of great complexity; that many factors are operative; that no detailed analytic theory of process operation is available; and that in addition to the effects of the "known" factors, large amounts of unpredictable variation are present.

Industrial experimentation, at least in the process industries, differs strikingly from agricultural work in the speed with which experimental results can be obtained. Experimental units are often successive runs on the same piece of equipment, but even quite large designs rarely take more than a few weeks to complete. A further difference, I believe, lies in the number of dependent variables that must be measured or calculated. The important properties of a type of rayon, or of a new kind of steel, or of an improved cake-mix, cannot be summarized by a single number. More usually, from five to ten dependent variables are required. Finally, the usefulness of blocks of homogeneous experimental material appears to be less in industrial work, but this appearance may be due to our not knowing how to specify blocks.

Industrial experimentation may be viewed as a continuum extending all the way from slight modifications of existing operation to the development of entirely new processes. Balanced designs of the type to be discussed are easier to use near the "existing process" end of this scale, but experience is rapidly being gained with the development of new processes.

Factors whose first-order effects are large, say of the order of 3 or 4 standard deviations, are usually discovered by making a few runs. The symbol σ_R will be used for the standard deviation of runs made under "fixed" conditions. If we must estimate a first-order effect of the order of σ_R with say a 95% confidence interval of half-width $\sigma_R/2$, then

$$1.96\sigma_R\sqrt{\frac{4}{N_R}} = \tfrac{1}{2}\sigma_R \tag{1}$$

from which N_R, the number of statistically independent runs required, is roughly 64. This number is conservative for two reasons. Chemists and plant engineers are likely to underestimate their σ_R, often by a factor of one-half. In the second place, since a considerable number of effects are usually measured in the same experiment, the usual 95% level of confidence is often not high enough. To reach the 99% level (per effect), roughly $(2.58/1.96)^2$ or 1.7 times as many runs must be made as at the 95% level.

At least 64 runs are required, then, half at each of two levels of a particular factor,