THE ECONOMIC EFFICIENCY OF EXPERIMENTATION[1]

by

D. J. FINNEY F. R. S.[2]

1. Introduction

1.1. Until 35 years ago, the responsibility of the statistician in research was thought to consist only in the analysis of experimental data as a preliminary to their interpretation. The concept of experimental design was practically unknown. From about 1923, a series of remarkable papers by R. A. FISHER revolutionized this outlook. The wealth of detailed knowledge of particular designs that derived from FISHER's work is of great theoretical interest and of major importance to the satisfactory conduct of experiments; perhaps even more important for the effective use of experimental resources, however, is the recognition of two general principles that emerge from the theory and practice of design:

(i) The design of an experiment in great measure determines the form of statistical analysis appropriate to the results.

(ii) The success of an experiment in answering the questions that interest the experimenter, without excessive expenditure of time and resources, depends largely upon right choice of design.

1.2. In this paper, rather than discuss technical details of particular experimental designs, I have chosen to consider more general topics relating to the optimal specification of an experimental design for a particular purpose. A great variety of different questions arises, and I can do no more than outline their scope and give additional attention to one or two that I find especially interesting. Most of the questions cannot be formalized as well as those on the pure structure of designs, and answers usually require intimate knowledge of the particular field of application. Too often, therefore, they are settled solely by the experimenter, who may ask the statistician to suggest designs only within a fairly rigid specification of his own choosing. For the statistician to be sole arbiter would be equally unsatisfactory, for he might fail to see the full implications of his recommendations in the ultimate interpretation of the research. The great need is for close collaboration between experimenter and statistician. As a statistician, I naturally concentrate on the statistical point of view, but I know how easily this can lead me into impracticable proposals and how frequently compromise between a theoretical ideal and practical convenience is essential.

1.3. I am convinced that the statistician can and should make as valuable a contribution to the general planning of an experiment or of a whole programme

[1] Read at the Biometric Symposium 7. 9. 1959 (Budapest).

[2] University of Aberdeen.

of experiments as he does, perhaps with greater mathematical technicality, to details of design relating to the allocation of treatments to plots and to the statistical analysis of experimental records. In doing so, he will need to exercise considerable tact, for an experimenter who has not previously experienced benefits from consulting a statistician about all phases of his research may not unreasonably distrust, and even resent, criticism of his choice of treatments or of the general scope of his experiments by one who is not a specialist in that particular field of science.

2. Examples of Topics

2.1. Under the narrower heading of experimental design, the statistician is accustomed to discuss particular designs, such as Latin squares, balanced incomplete blocks, and factorial designs, as well as procedures of randomization, methods for the construction of solutions of combinatorial problems, confounding, and the analysis of variance and covariance and associated computational techniques. More recently, there have been important developments in sequential experimentation, and the methods of Box and others for the exploration of optimal conditions have perhaps done more to extend the understanding and use of experimental design than anything else published in the last 20 years. Much has to be decided before statistician or experimenter resolves on the use of one particular design. Too often, the important decisions are taken without conscious thought, and without realization of the consequences that they may have for the precision of results obtained in respect of important issues or even for the relevance of the results to the main subject of inquiry.

2.2. As examples of some of the topics that tend to be neglected in any purely formal presentation of experimental design, I list the following:

(i) Use of controls.
(ii) Number of treatments to be tested.
(iii) Number of different factors in a factorial experiment.
(iv) Number of levels of a factor.
(v) Number of replicates of treatment combinations.
(vi) Number of similar experiments on different sites, at different times, or on different sources of material.
(vii) Definition of plots or other units to which experimental treatments are to be applied (shape and dimensions of field plots, species, age and other characteristics of experimental animals, use of repeated observations on same plot at intervals of time).
(viii) Choice of suitable design.
(ix) Specification of measurements to be made and techniques for making them.
(x) Use of concomitant variates for improvement of precision.
(xi) Relation between successive experiments on different phases of an experimental programme.

2.3. It would be easy to speak at length on each one of these topics, yet still only to have discussed aspects familiar by experience to one statistician. Although this might have some value as illustration, even though it did not amount to a comprehensive account, the time required would be excessive. I have therefore chosen a few special situations that will exemplify the statistician's line of argument rather than pretend to be comprehensive.

3. Internal and External Economy

3.1. Considerations that may broadly be described as economic are inseparable from any discussion of the relative merits of alternative experimental procedures. Although I have no intention of suggesting that scientific research is to be evaluated merely according to conventional economic principles, in any experimental science there will naturally be a desire to employ resources effectively in relation to the purpose of the research. The manner in which these resources are measured will vary from one situation to another, and in practice emphasis may have to be placed on *one* component of the resources because of the impossibility of measuring *all* on the same scale. For example, an able scientist is likely to have more ideas for investigations than he himself or those who work under him can use, and his chief concern may be economy of time, in order that he shall spend long enough on any one project to extract adequate and precise information but not so long that the rate of increase of knowledge is too small to justify continuation. In other circumstances the main factor to be considered may be the quantity of materials consumed, the area of land occupied, or the number of animals used. Whatever the limiting factor or factors may be, they can conveniently be regarded as measuring *costs ;* I do not imply that cost must be assessed on a monetary scale, although such a scale is often the simplest way of expressing several components of cost commensurably.

3.2. Under the heading of the *internal economy* of experimentation may be grouped matters relating to the optimal planning of experiments for a specified total cost. Whether the research is a disinterested inquiry into natural phenomena or a problem of immediate practical importance to some industrial or biological technology, the aim of maximizing the returns from an experiment relative to its cost is reasonable, and in broad general terms much can be done by preliminary study of alternative plans. Difficulty arises when the aims of the research are not si nply the estimation of one or two clearly defined quantities, because then the maximization cannot be seen in exact mathematical terms. "To estimate the bactericidal potency of X, a new drug, relative to an established standard, S" may not be easy, especially if a wide range of conditions is to be examined; nevertheless, the efficiency of alternative experimental plans can be discussed in some detail, the aim being maximum precision of estimation relative to the total number of test cultures or other experimental units to be expended. The efficiency of a project "To relate the chemical structures of a group of compounds to their bactericidal activities" could not be discussed in the same detail: the various conceivable finding sare not simply related to the magnitude of experimentation, and the values of alternative outcomes cannot be expressed on a scale commensurable with that of cost. When emphasis is to be placed on the general advancement of knowledge rather than on the acquisition of a particular item of information, an investigator can seldom state in advance which aspects of his experiment are of primary interest; even more rarely can he list the conclusions to which he might be led and assign values to them.

3.3. In the past, statisticians have paid considerable attention to the internal economy of experimentation, but only in recent years has realization grown that it is pr)per, and indeed desirable, for them also to advise on what the total cost of a research project ought to be in relation to the value of results

expected from it. Whether or not justification for the amount of experimentation undertaken should be sought on these lines depends upon the nature of the project, and here the contrast between fundamental and technological research is intensified. FISHER ([16], Chapter 4) has stated essentially the same contrast in relation to significance tests and inductive reasoning on the one hand, decision and acceptance procedures on the other. He emphasized that the use of significance tests and fiducial distributions has the object of summarizing the evidence currently available from scientific experiments and of aiding the advancement of science through the accumulation of knowledge; the procedure is quite distinct from that appropriate to the taking of an irrevocable decision that concerns some matter of policy or recommending one of several alternative courses of action. "An important difference is that Decisions are final, while the state of opinion derived from a test of significance is provisional, and capable, not only of confirmation, but of revision. An acceptance procedure is devised for a whole series of cases. No particular thought is given to each case as it arises, nor is the tester's capacity for learning exercised. A test of significance on the other hand is intended to aid the process of learning by observational experience." It is fairly easy, logically though perhaps not practically, to devise a cost function for the loss incurred if an experiment leads to the recommendation to farmers of an inferior variety of wheat instead of the best; one cannot reasonably talk in quantitative terms about the loss consequent upon failing to discover a new fundamental principle of science, or upon a belief in a wrong theory.

3.4. None the less, many experiments and programmes of research are conducted for specific purposes that permit advance formulation of the value of the results. Although this is especially true in technological research, instances arise even in pure science where matters of technique must be explored as a preliminary to the main project; for example, alternative methods of sampling an insect population or of counting blood cells may need to be compared with a view to accepting one as standard in future fundamental research, the relative merits being assessed on the basis of accuracy per unit cost with costs possibly measured in time rather than in money. I shall today be concerned more with problems of scientific technology, and the manner in which the planning of experimentation should be influenced by economic considerations that go beyond merely the most efficient utilization of specified resources.

3.5. If the purpose of an experiment or groups of experiments is to estimate a parameter whose magnitude is of importance to some sector of the national economy, other things being equal any increase in the size or number of experiments will improve the precision of the estimate (Examples are mentioned in Section 6 below). In so far as an increase in precision enables policy based upon the estimate to be more correctly formulated, the loss to the economy consequent upon the policy being less beneficial than that based upon the unknown true value of the parameter will be reduced. Hence any increase in precision can be assigned a value on the scale of costs, which may be set against the cost of additional experimentation. Study of the optimal conditions, defined as maximizing the net gain, relates to the *external economy* of the system. The connexion with the theory of decision functions is close, though the approach is somewhat different.

3.6. I may summarize by saying that discussion of the internal economy of an investigation usually turns upon maximizing the precision of an estimate

(or the average precision of several estimates) for a specified cost, or minimizing the cost of attaining a specified precision. Other problems may occasionally be considered under this head. Discussion of the external economy, on the other hand, is usually concerned with evaluating the disadvantages to the whole economy in which the investigation is set that will result from imprecise estimates, and balancing the gains from reducing the imprecision against the cost of the extra experimentation necessary in order to achieve this reduction.

4. Number of Factors

4.1. Many experiments are planned with inadequate appreciation of what factorial design, confounding, and fractional replication can achieve. If an experiment intended for the study of two or three factors can have others incorporated, without seriously increasing the size of the experiment or seriously reducing the replication, additional information may be obtainable for a negligible additional labour or cost.

4.2. Consideration of experiments on factors at 2 levels enables the argument for additional factors to be forcefully expressed. An experiment on one factor would seldom be regarded as adequately replicated unless it had about 8 plots of each level. If two such experiments on the same subject-matter but with different factors were contemplated, 32 plots would be involved; by combining these into one 2^2 experiment, however, the same number of plots could be used to give information on the interaction of the factors as well as on the average of main affects. Randomized blocks of 4 might be used, instead of blocks of 2, but any loss in precision from the larger blocks would be offset by the greater replication (now sixteen-fold) on each factor. In this experiment, 21 out of 31. d. f. are used in the estimation of error variance and only 3 are assigned to treatment comparisons, a somewhat extravagant provision for error since the advantage of extra error degrees of freedom decreases markedly after the first 10. It would be an unimaginative experimenter who could not think of additional factors highly relevant to his investigation. Introduction of a third factor and arrangement as 2^3 in 4 blocks of 8 involves some loss in precision from the larger blocks, but now 7 d. f. are measuring treatment comparisons and 21 d. f. estimate error: still only one-quarter of the effort of the experimenter is spent directly on comparing treatments and three-quarters on assigning a measure of precision to the comparisons. Without change of block size, one or two more factors can be introduced, by confounding (3-factor and higher interactions only) and using remaining high order interactions for estimating error. This leads to either 10 d. f. for treatments (main effects and 2-factor interactions in a 2^4) and 18 d. f. for error or 15 d. f. for treatments (2^5 design) and 13 d. f. for error. Thus saturation of the experiment with factors, so making it a single replicate, still leaves almost half the effort expended on error estimation. Fractional replication provides a method of supersaturation; this is of restricted value in so small an experiment, but, if one 2-factor interaction can be sacrificed, a one-half replicate of 2^6 is possible with the same blocks of 8 plots, and now the allocation of d. f. is 20 for treatment main effects and 2-factor interactions, 8 for error.

4.3. FISHER [15] has shown how to allow for the inevitable imprecision of the sample estimate of variance in assessing the information available from an experiment. If an experiment has an error mean square s^2 with f degrees

of freedom, the precision of a contrast estimated as a difference between the means of two sets of r plots is

$$(1) \qquad \frac{r(f+1)}{2s^2(f+3)}.$$

An index of information for the whole experiment, with main effects and 2-factor interactions all regarded as of equal value, might be defined as this quantity multiplied by the number of such treatment effects that can be estimated (FINNEY, [9]). Table 1 shows values of this index for the six experiments discussed above. Even allowing that the first requires only half as many plots as the others and that the first two have smaller blocks, the advantages of 2^5 and 2^6 designs appear substantial. Of course, this index should not be employed uncritically, as it makes no distinction in value between main effects and interactions nor does it take account of the occasional need for estimation of 3-factor or higher order interactions.

Table 1

Information on main effects and 2-factor interactions obtainable fom 2^n experiments on 32 plots or less

No. of factors n	No. of plots	Block size	Replication r	Error d. f. f	Main effects and 2-factor interactions	Index of information
1	16	2	8	7	1	$3.2/s^2$
2	32	4	16	21	3	$22.0/s^2$
3	32	8	16	21	6	$44.0/s^2$
4	32	8	16	18*	10	$72.4/s^2$
5	32	8	16	13*	15	$105.0/s^2$
6	32	8	16	8*	20	$130.9/s^2$

* Here the error, wholly or in part, is estimated from interactions

4.4. The preceding paragraphs are not to be read as uncompromising advocacy of multifactorial single or fractional replicates for every purpose. The statistician must know the potentialities of factorial design and must put clearly before the experimenter a statement of the gains to be expected from variants of the original proposals; the experimenter has the last word. Limited experience and capabilities of the staff responsible for executing and recording experiments, or the restriction of interest to one or two well-defined questions, occasionally make very simple designs preferable to those that on paper appear more informative. Despite the merits of *successful* multifactorial fractional replicates, a simpler design that gives trustworthy results on a narrower front is preferable to a scheme too ambitious for the circumstances of the experiment. Often the inclusion of 5 factors in an experiment instead of the 2 initially suggested involves only slight increases in the labour of performing the experiment and of analysing the results; if the experiment has to be conducted under great stress, or with the aid of staff unaccustomed to anything other than the simplest experiments, the success of the whole may be jeopardized by the risk of mistakes being made, and 2^2 may be a wiser

choice than 2^5. Although the statistician may be continually looking for opportunities of inserting as many factors as possible into experiments, he should not let the arguments that led to Table 1 be his only guide, but should temper his ambition with discretion.

5. Use of Concomitant Variates

5.1. In most experimental designs, block constraints are important, and one of the duties of a statistician is to assess the best system of blocking. Although much depends upon experience of the field of research, any logical classification of experimental units that is possible before the experiment begins and that is in the least likely to be associated with the final "yield" measurements deserves consideration as a basis for blocking. Often the number of choices is greater than the experiment can accommodate, even though recourse be had to Latin square and other designs that use two or more block systems simultaneously. The statistician learns to be suspicious of statements by experimenters that differences between days of experimentation or between alternative supplies of a reputedly standard material are negligible, but instead tests them whenever possible in records of former experiments; not infrequently, he finds that the efficiency of comparisons between treatments would be much improved if experiments were balanced in respect of them.

5.2. When a choice must be made between alternative classifications, other things being equal some preference should be given to qualitative characters as a basis for blocking, as quantitative characters can often be dealt with by covariance analysis. If an experiment is conducted in the belief that it does not matter which of four different persons is responsible for a certain operation (or which of four different supplies of a nominally standard drug is used), so that no attempt is made to balance the design over the four, and subsequently obvious differences appear, salvage can still be effected. One person, say P_1, can be taken as standard; three dummy variates can be defined such that the first takes the value 1 for all yields from P_2 and zero for P_1, P_3, P_4, a second is 1 for P_3 only and zero for the others, and a third is 1 for P_4 only. Covariance analysis on all the dummy variates simultaneously then adjusts all comparisons of treatment means to estimated equivalence in respect of P_1, P_2, P_3, P_4; the sum of squares removed from error by regression on the three variates reduces to the ordinary sum of squares between persons if the design is in fact perfectly balanced over the four. Not only is this process laborious, but estimation of the covariance adjustments inevitably involves some loss of information that would not have occurred if balance had been perfect. Hence the advantage lies with using such a classification as a basis for blocks if this be possible.

5.3. A quantitative character, such as initial weight of an animal or yield of a plot in any pre-experimental period, may be such that the relation of final yield to it (apart from treatment and block effects) is a regression of fairly simple type, perhaps linear or quadratic. Covariance analysis using this one concomitant variate (and its square if the regression is quadratic) then secures that comparisons of adjusted treatment means are made on terms of equivalence in respect of the concomitant. None the less, as KEMPTHORNE [19] has emphasized, strict validity of the procedure depends upon the exact truth of the regression model, an assumption that is avoided if groups of plots exactly or approximately equal in respect of the quantitative character are

given the status of block constraints. OUTHWAITE and RUTHERFORD [20] made an instructive re-examination of data first reported and analysed by FEDERER and SCHLOTTFELDT [7]. An experiment on plant growth was arranged in randomized blocks, each block being a single line of plots and successive blocks being parallel lines. Inspection of the yields suggested a trend along the line of the blocks, and OUTHWAITE and RUTHERFORD eliminated the trend by a multiple covariance analysis, using position of a plot in the block as one concomitant variate and successive integer powers of this value as other concomitants; by continuing to a polynomial of degree 7, they were able to eliminate all differences corresponding to the average positions of the 8 treatments within blocks, a procedure similar to the use of dummy variates mentioned in the preceding paragraph. Nevertheless, they found that the errors of estimation inherent in the covariance adjustments constituted about a 15% loss of information relative to a design in which Latin square constraints were used to balance treatments in respect of positions within the blocks. Here is an instance in which a quantitative (positional) character would have been better used as a basis for blocking than in a covariance analysis. Undoubtedly blocking is to be preferred when practicable, and covariance should be regarded as a device for use when the experiment cannot hold enough block constraints, when an unsuspected source of variation is seen to be associated with some character during or at the end of the experiment (the character itself having been assessed before treatments were applied or being for other reasons independent of the treatments), or for the adjustment of mistakes in design.[3]

5.4. One way of using a quantitative character as a basis for blocking is to stratify the plots according to the value of the character. For example, randomized blocks for t treatments might consist of the t plots with the highest values, the t next highest, and so on. Table 2 shows the design of an experiment on the comparison of 5 inocula of a plant virus in which this idea was extended to a Latin square (COX and COCHRAN, [2]); five plants formed the primary blocks (rows), but the relative sizes of five leaves on each plant were used as a secondary character for an orthogonal set of blocks (columns). A similar method of control can be used in the absence of any primary blocking system. Suppose that 5 treatments are to be tested on 25 experimental units, or plots, for each of which a concomitant variate has been measured before the experiment begins. The values of the concomitant can be listed in order of magnitude from highest to lowest, and the treatment of each plot then determined by writing successive rows of a 5×5 Latin square alongside this list; the square of Table 2 would give the series of treatments

A, E, D, C, B, C, D, A, B, E, B, ... C, E, A.

Analysis proceeds exactly as for a Latin square. The enthusiast for eliminating every shred of an effect of the independent variate can make covariance adjustments as well if he wishes! Papers by COX [1] and FELDT [8] are relevant to the whole of this discussion.

[3] This last is of course something that should not occur, but methods are required for dealing with it. An example closely allied to the present discussion has been discussed by FINNEY and COPE [14].

Table 2

Allocation of five virus inoculations, using five leaves on each of five plants in a randomized Latin square design

Plant no.	Relative size of leaf 1	2	3	4	5
I	A	E	D	C	B
II	C	D	A	B	E
III	B	C	E	A	D
IV	E	A	B	D	C
V	D	B	C	E	A

5.5 A modified form of balance has sometimes been used, especially in animal experiments where perhaps initial weight is the quantitative character. Instead of random allocation to treatments, the sets of animals allotted to different treatments are so selected as to make all treatments have mean initial weights as nearly equal as possible. Quite apart from the subjective influences that enter as soon as true randomization is set aside, this procedure inflates variation *within* treatments at the expense of variation *between* treatments. Tests of significance and assessments of standard errors may be seriously biased unless the statistical analysis uses the covariance analysis that the design was presumably specially devised to avoid! The theory has been discussed elsewhere (FINNEY, [10]); this design appears to have no merits.

6. Number of Experiments

6.1 YATES [21] has discussed an elementary approach to the external economy of estimating the optimal rate of application of fertilizer to a particular crop; his aim is to sample a region in which the crop is grown by placing experiments at randomly distributed sites. If all experiments are to be of one simple standard pattern, the cost of a series of n can be expressed approximately as

$$A(n) = c + an, \tag{2}$$

where c represents fixed expenditure on the whole programme and a is the additional cost per experiment. Write ξ for the rate of fertilizer application per unit area that would maximize the net benefit, under the assumption that all farmers in the region adopt this rate. The yields in the n experiments can be averaged so as to estimate a response curve, from which X, an estimate of ξ, can be obtained. If this response curve, the regression of yield on amount of fertilizer per unit area, were quadratic, the reduction in net gain resulting from use of X instead of the unknown true value ξ would be proportional to $(\xi - X)^2$; for any other regression relation, the reduction in yield will still be of this form to the first order. Although $(\xi - X)^2$ is unknown, its expectation, $V(X)$, can be estimated from the experiments, and will be inversely proportional to n, say

$$V(X) = v/n. \tag{3}$$

The loss in yield per unit area from future use of X instead of ξ therefore has expectation $\lambda\, v/n$.

6.2 If the estimate X is used as a basis of fertilizer practice for a total area T, the expected total loss of crop from errors of estimation of ξ is

$$L(n) = \lambda v T/n. \tag{4}$$

If $A(n)$ has been measured on a scale of equivalent value of crop, the optimal n can be determined by minimizing $A(n) + L(n)$, and is

$$n^* = (\lambda v T/a)^{1/2}. \tag{5}$$

When instead of n^* experiments fn^* are performed (where f is either less or greater than 1), the expected net returns from application of the results of the research are reduced by

$$\frac{(f-1)^2 a n^*}{f}, \tag{6}$$

an amount which is small when f is near to 1. Consequently, if experimental resources have to be allocated between several different research programmes, each may be made somewhat smaller than its own optimal without serious loss to the whole economy. The theory can be modified in order to take account of the estimate X being used on crops for a number of future years.

6.3. The advantage to be gained by increasing n has from one point of view been underestimated. Knowledge of response curves for very many sites within the region might enable the region to be subdivided into districts within each of which the response-potential of the crop was more homogeneous. If this were followed by separate recommendations on rate of fertilizer application for the several districts, the total crop yield should be greater than if one average value were recommended throughout.

6.4. Greater logical and mathematical complexities appear when the need is to make a choice between two alternative procedures in the best possible way, instead of to estimate an optimal on a continuous scale. One problem of this kind, relevant to situations that occur in a variety of technological fields, has been studied by GRUNDY, HEALY and REES [17], GRUNDY REES and HEALY [18].

6.5. Suppose that the desirability of making a change in some standard process of technology is under investigation (e. g. the use of a new synthetic plastic for a particular type of electrical insulation, or the incorporation of a certain hormone into the diet of young pigs). A unit experiment has been performed, from which the increase in the amount or quality of production by the new process relative to that by the old is estimated to be x with variance σ^2; x, of course, may be positive or negative and the variance will be supposed based on enough degrees of freedom for it to be taken as the population value. Then x is an estimate of a population parameter, θ representing the expected improvement in production attributable to the new process. A decision on whether or not to adopt the new process ought to depend upon whether or not $W\theta$, a measure of the advantage gained from the change (assumed proportional to θ), exceeds a, the costs inherent in making the change (costs of new equipment, loss of production during a period of change, additional labour requirements until a new routine is working satisfactorily, and so on). The investigator has two reasonable alternatives open to him:

(i) Calculate $Wx - a$, and recommend the adoption or rejection of the new process according as this is positive or negative;

(ii) Recognize that evidence is inadequate and conduct a further n units of experimentation so as to obtain a second estimate, y, of θ with variance σ^2/n. Then calculate

$$\frac{W(x + ny)}{1 + n} - a,$$

and recommend adoption or rejection according as this is positive or negative.

6.6. If the cost of the additional experiments is proportional to their number, being a say, the expected gain from rule (ii) is a function of θ, x and n:

$$Q(\theta, x, n) = (W\theta - a)P - an, \tag{7}$$

where P is the probability that the new process is adopted. Of course, Q may be negative, for example if n is taken excessively large. Suppose now that x and y are normally and independently distributed about their mean θ. The condition that rule (ii) leads to adoption of the new process may be written

$$y > -\frac{x}{n} + \frac{a(n+1)}{Wn};$$

the probability of this is

$$P = \Phi\left[\{nW\theta + Wx - a(n+1)\}/\sigma W n^{\frac{1}{2}}\right], \tag{8}$$

where

$$\varphi(z) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}z^2} \tag{9}$$

and

$$\Phi(z) = \int_{-\infty}^{z} \varphi(t)\, dt\,. \tag{10}$$

Moreover, (8) can also be regarded as applicable under the conditions of rule (i), for as $n \to 0$, $P \to 1$ or 0 according as $(Wx - a)$ is positive or negative.

6.7. The investigator must decide what value of $n(\geq 0)$ he will use. If his decision is to be in some sense optimal from the economic point of view presumably it should be based only upon the behaviour of the function in (7). Now in $Q(\theta, x, n)$, x is known from the first experiment, but θ is unknown, so that the obvious course of maximizing Q is not available; n must be chosen solely as a function of x and of the parameters of costs and variability. One principle that has found favour in problems of this kind is that of the *minimax*, which involves choosing n as a function of x in such a way as to minimize the maximum loss that can occur through the value of θ being unfavourable to the course of action that is eventually adopted. This loss is measured relative to the value of an immediate correct decision, $(W\theta - a)$ or 0 according as $(W\theta - a)$ is positive or negative; the loss is the difference between this amount and $Q(\theta, x, n)$, and the minimax value of n is that which minimizes the maxi-

mum of the loss function with respect to θ. Minimax estimation is known to have certain optimal properties, but GRUNDY and his colleagues remark that they "are not aware of any necessity for preferring the minimax solution to all alternatives in practical problems". In this problem, the minimax method is mathematically intractable; they propose instead to choose n so as to maximize a value of $\bar{Q}$ averaged with respect to θ, employing for this purpose the fiducial distribution of θ on the evidence of the first experiment, that is to say a normal distribution of mean x and variance σ^2. The average is therefore

$$(11) \qquad \overline{Q}(x, n) = \int_{-\infty}^{\infty} Q(\theta, x, n)\, \varphi\left(\frac{\theta - x}{\sigma}\right) \sigma^{-1}\, d\theta$$

$$= \int_{-\infty}^{\infty} (Wz\sigma + Wx - \alpha)\, \Phi\left[zn^{\frac{1}{2}} + \frac{(Wx - \alpha)(n+1)}{\sigma W n^{\frac{1}{2}}}\right] \varphi(z)\, dz - an$$

by substitution from (7) and (8) and the transformation $\theta = z\sigma + x$. Some manipulation of standard integrals then leads to

$$(12) \qquad \overline{Q}(x, n) = (Wx - \alpha)\, \Phi(u) + \sigma W \left(\frac{n}{n+1}\right)^{\frac{1}{2}} \varphi(u) - an\,,$$

where

$$(13) \qquad u = \frac{(Wx - \alpha)}{\sigma W} \cdot \left(\frac{n+1}{n}\right)^{\frac{1}{2}}.$$

Note that, as $n \to 0$, $\bar{Q}$ tends to $(Wx - \alpha)$ or 0 according as $(Wx - \alpha)$ is positive or negative, and so gives correctly $\bar{Q}(x, 0)$.

6.8. It is convenient to regard n as measured on a continuous scale, so making it the ratio of the amount of experimentation recommended for the second stage to that already undertaken in the determination of x. Differentiation of $\bar{Q}(x, n)$ with respect to n shows that, as n increases from zero, $\bar{Q}$ first decreases and may thereafter either continue to decrease or attain first a minimum, then a maximum, and subsequently decrease steadily. Practical importance attaches to the absolute maximum of $\overline{Q}$ for $n \geq 0$; the recommendation will be to take an immediate decision ($n = 0$) if $\overline{Q}(x, 0)$ is not exceeded by $\overline{Q}(x, n)$ for any $n > 0$, but to use the value of n corresponding to the maximum of $\bar{Q}$ if this exists and exceeds $\bar{Q}(x, 0)$. GRUNDY *et al* have studied the conditions for an absolute maximum of $\bar{Q}$ and have constructed a nomogram for determining the optimal n.

6.9. Enough has been said to illustrate the character of this theory. The authors have developed the mathematics in greater detail, have given tables to show how the recommendations operate for particular values of x and the various parameters, and have compared the performance of their method with that of several alternatives. As might be expected, the function $\bar{Q}(x, n)$ is flat in the neighbourhood of its maximum, so that little harm comes

of taking n at some distance from its maximizing value. Moreover (except when $Wx - a = 0$), the behaviour of $\bar{Q}$ in the neighbourhood of $n = 0$ is such as to ensure that either $n = 0$ is recommended or the recommended additional amount of experimentation is fairly large; clearly a small amount is of little use as it can neither confirm nor controvert the evidence of the quantity $(Wx - a)$ effectively and is therefore almost sure to be economically disadvantageous.

7. Relation between Successive Experimental Phases

7.1. A major research project will seldom involve only one experiment or one group of contemporaneous experiments. Different aspects of the research are likely to involve different schemes of experimentation, related by the fact of being part of one project. Important questions arise in connexion with the allocation of effort between these phases; one class of problems of considerable statistical interest is that in which the "treatments" on which experiments are to be conducted in one phase depend upon the outcome of an earlier phase. I shall discuss two examples of the efficient planning of selection experiments, in which an initially large number of treatments are to be subjected to experiment, as a result of which the best performers will be selected for a second stage of experimentation.

7.2. I first consider a basic problem of plant breeding for yield improvement, although the mathematical model that I adopt is necessarily somewhat crude; I have discussed elsewhere (FINNEY, [11], [12]) numerous practical points that need to be taken into account. By crossing established varieties, plant breeders can produce large numbers of new seedlings of a crop species; the great majority of these will prove to be useless, but some may deserve perpetuation as the foundations of new varieties. Suppose that each year a *cohort* of N potentially new varieties of a crop is ready to begin its programme of yield testing, that testing is to continue over k successive years, and that at the end of this time a proportion π of the varieties is to be passed forward as "successes". The number of varieties of the cohort retained under test will be reduced from year to year, so that in year r all survivors from year $(r-1)$ will be grown in a field trial and the fraction P_r of these showing the best yields (without any reference to tests of significance) will be retained for a further year. Account must be taken of the possibility that N may be so large as to make the testing of all in the first year inexpedient, and instead P_0N might be randomly selected from the cohort, the remaining fraction $(1-P_0)$ being discarded without test. Clearly the P_r are subject to the constraint

$$P_0P_1P_2 \ldots P_{k-1}P_k = \pi. \tag{14}$$

7.3. In any one year, survivors of k different cohorts will be under test, each at a different stage of testing. Under stable conditions the experimenter, will have a total area A available for field trials in each calendar year, the site perhaps changing from year to year although its area is constant. This he must subdivide so as to allot an area A_r to the $P_0P_1 \ldots P_rN$ survivors of the cohort now in its rth year of testing $(r = 1,2, \ldots k)$, where

$$A_1 + A_2 + \ldots + A_{k-1} + A_k = A. \tag{15}$$

The problem of optimal planning is to determine values of P_0, the P_r, and the A_r, subject to (14), (15), so as to maximize the expected yields of the πN varieties finally selected.

7.4. The N varieties may be regarded as a sample from an infinite population in which the distribution of expected yields is of specified form, and only a normal distribution will be considered here; CURNOW [3], [4] has obtained some results pertaining to other forms of distribution. The field trial at stage r will estimate the expected yields with an error that may reasonably be assumed normal; moreover, this error will decrease as the number of varieties in stage r is decreased or as A_r is increased, on account of changes in plot size and replication that are made possible. As an approximation, ε^2, the error variance will be assumed to be expressible as

$$\varepsilon^2 = \frac{\gamma P_0 P_1 \ldots P_{r-1} A \sigma^2}{A_r}, \tag{16}$$

where γ is a constant and σ^2 is the variance of the distribution of expected yields.

7.5. Mathematical analysis of this model for one-stage selection is then quite simple. The mean yield of the πN varieties selected can be shown to have an expectation that exceeds the general mean by an amount

$$G = \sigma Z_1 / P_1 (1 + \gamma P_0)^{\frac{1}{2}}, \tag{17}$$

where Z_1 is the ordinate to the standardized normal frequency function (mean 0, variance 1) corresponding to a single-tail probability P_1. Under condition (14) this is maximized by taking P_1 as the solution of

$$Z\,2P + \pi\gamma = 2TP\,(P + \pi\gamma),$$

where T is the abscissa (unit normal deviate) and Z the ordinate corresponding to P. Hence P_1 is a function of $\pi\gamma$ alone, and $P_0 = \pi/P_1$. Equation (18) can lead to a value of P_1 smaller than π, which is an indication that G, the gain in mean yield, would have been greater if N had been greater but that the best procedure now is to take $P_1 = \pi$, $P_0 = 1$. More commonly, P_1 will be greater than π, and an initial random discard of $N(1-P_0)$ varieties will be advantageous. The gain can be quite considerable; for example, if $\gamma = 5$ and $\pi = 0.01$, reasonable practical values, the optimal P_1 is 0.063, and initial reduction of the N varieties to $0.16\,N$ will increase G by 35% as compared with using $P_1 = \pi$ on all the original varieties; if π had been 0.1, the gain from using the optimal procedure, a random discard to reduce N to $0.56\,N$ followed by $P_1 = 0.18$, would have been only 5% of the value of G for $P = 0.1$.

7.6. As soon as more than one stage has to be considered, the mathematics increase in complexity, and even for two-stage selection little has yet been achieved beyond numerical study of one or two particular cases. However, these are quite revealing. As might be expected, an initial discard is much less important, since its function of reducing the total number of varieties to a number that can be tested with reasonable precision on the available land is in part performed by the first stage of selection. If π is very small, a value of P_0 different from unity may be advantageous, but a gain of 10% or more relative to the best that can be achieved with $P_0 = 1$ is exceptional. Only the

case of $P_0 = 1$, therefore, has been studied in any detail. For any specified values of γ and π, a selection programme is then completely defined by choice of P_1 and A_1, since $P_2 = \pi/P_1$, $A_2/A = 1 - A_1/A$. If two rectangular axes are taken as scales of P_1 (better, $\log P_1$) and A_1/A, the quantity G can be evaluated for many different pairs of points in the diagram; G is again defined to be the excess of the expected mean yield of the πN selected varieties over the general mean of all varieties, but its symbolic expression is much more cumbrous than (17) and is laborious to compute. Contours of equal G can be inserted on the diagram, and these will surround a single maximum at the optimal combination of P_1 and A_1. A full study of $\gamma = 1$, $\pi = 0.01$, supported by a lesser set of computations for $\gamma = 10$, $\pi = 0.01$ and argument from the run of the results suggests a simple broad generalization: under a wide range of values of γ and π corresponding roughly to practical conditions, the maximization of G is achieved approximately by

$$P_1 = P_2 = \pi^{\frac{1}{2}}, \qquad A_1 = A_2 = \tfrac{1}{2} A\,. \tag{19}$$

These certainly do not give the true maximum, but the surface relating G to P_1 and A_1 is very flat near its maximum and the loss from the approximation is negligible. When π exceeds 0.1, this two-stage selection is scarcely any better than one-stage; as π decreases the advantage of the extra stage becomes greater, although it appears that the gain relative to the optimal one-stage procedure is always much smaller than that for optimal one-stage relative to one-stage with $P_0 = 1$. CURNOW [5] has made some calculations relating to three-stage selection, which support the view that

$$\left.\begin{aligned} P_0 &= 1, \\ P_1 = P_2 &= P_3 = \pi^{1/3}, \\ A_1 = A_2 &= A_3 = A/3, \end{aligned}\right\} \tag{20}$$

approximate to the optimal conditions and also indicate that the further gain relative to two-stage is small. The generalization of this simple rule to k stages is obvious; although the deviation of the approximate values of the P_r and A_r from the true maximizing values may increase with increasing k, the difference between the mean expected yields for those varieties selected according to the rule and those selected according to a true optimal procedure is almost certainly always small.

7.7. CURNOW [3] and I [13] have made some progress with study of the external economy associated with this problem of selection of crop varieties. For one-stage selection, the mathematics are relatively simple and curves can be constructed from which the optimal policy can be derived; for two-stage the analysis is more complex. Despite the weaknesses of the model, this work does give some guidance on the amount of effort that can most advantageously be spent on varietal selection.

7.8. Instead of discussing this, I shall conclude with an account of another selection problem, that of the "screening" of chemical compounds for possible therapeutic activity. During the manufacture of pharmaceutical and other chemical products, large numbers of different compounds may be made; any one may be of value in the treatment of a particular disease, but the proportion of active compounds will be exceedingly small and only empirical

test can determine whether or not a compound is active. A cure for cancer may already exist as a compound synthesized for an entirely different research or industrial purpose and since put aside as of no further immediate interest, yet untried in a field where it will prove immensely valuable.

7.9. Here again, there is need for preliminary tests with low replication, on the basis of which compounds whose therapeutic activity is very slight can be rejected, followed by tests with higher replication as a second stage; again the possibility of a sequence of several stages can be considered. In this work, however, a better approximation to reality than can be given by a continuous distribution of expected yields may be provided by a discrete distribution with an exceedingly small proportion of active compounds (assumed of equal value) and a complementary large proportion of compounds with zero activity.

7.10. DAVIES [6] has made a beginning in the development of statistical theory appropriate to this situation. Necessarily he restricted himself to the supposition that the compounds tested can be regarded as randomly selected from a large population. Although non-statistical considerations must be taken into account, a plan of screening derived from a theory based on randomness is likely to be useful as a first approximation to an optimal procedure. DAVIES remarked that "When an active compound has been found, it is usual to make a number of compounds of similar value in the hope of improving the activity. This is what we call 'following a lead'. When a lead is being followed the test is no longer a random one, and the considerations on which to base the design of the test are then different".

7.11. For many purposes, the original distribution of activity may be taken as having only two values, 1 and 0, with probabilities δ and $(1 - \delta)$ where δ is usually small (possibly of the order of 0.01, 0.001, or even smaller). Any screening procedure will involve testing compounds, perhaps by giving each to one or more animals, measuring the activity, and selecting those for which the mean activity so measured exceeds a specified value; the measurement will be subject to experimental error, and so will not be restricted to the values 1 and 0. Inevitably some of those selected will be *false positives*, that is to say compounds with true value 0 that pass the test by the chances of experimental error, and equally inevitably some true positives will fail to be selected. DAVIES suggested as a suitable criterion for an optimal scheme that, when the total number of compounds is reduced by selection to a fixed proportion, π, of itself, the proportion of truly active compounds should be maximized (or the proportion of false positives minimized) relative to the total effort expended. For this initially binomial distribution of true activity, the criterion coincides with that of maximizing the true mean activity; for a more general distribution, a criterion based upon the proportion of compounds that exceed a certain level of true activity appears to be preferable to one based upon the mean.

7.12. Suppose that the rules of selection adopted are such that compounds with true activities 0, 1 have probabilities θ, θ' respectively of being selected. In order to conform to the requirement that the intensity of selection shall be π, these must satisfy

$$\pi = \theta\,(1 - \delta) + \theta'\,\delta. \tag{21}$$

The proportion of active compounds amongst all compounds selected is then

$$\psi = \theta'\,\delta/\pi\,, \tag{22}$$

and ψ is also the mean activity of those selected. The testing will involve giving doses of the compounds to experimental units, these perhaps being members of one of the usual species of laboratory animal. If the procedure has more than one stage, the number of animals used may vary from one compound to another, and in particular may differ in average between active and inactive compounds. If the expectations of the numbers of animals used for compounds of activity 0, 1 are m, m' respectively,

$$\overline{m} = m(1 - \delta) + m' \delta \tag{23}$$

is the average number of animals per compound. The expenditure in testing one compound can perhaps be satisfactorily represented as

$$A = a\overline{m} + c, \tag{24}$$

where a is the cost per animal used and c the cost of the compound itself. The aim will therefore be to maximize ψ subject to π and A being fixed, and with δ unknown but presumed very small. These expressions are easily generalized to integrals if the original distribution of activity is continuous.

7.13. One possibility would be to adopt a k-stage testing programme analogous to that discussed earlier in this Section. If $k = 2$, computations can be expeditiously performed with the help of existing tables of the normal distribution. Suppose that the first stage assigns r_1 animals to each compound; the observed effect of any compound has a mean x whose variance is σ^2/r_1, where σ^2 is the variance per animal. All compounds are rejected for which $x < X$, and this amounts to selection of a proportion P_1, so reducing an initial N compounds to $P_1 N$. At the second stage, r_2 animals are assigned to each remaining compound and the mean effect, y, has variance σ^2/r_2. All compounds are rejected for which the mean over both stages is less than Y, that is to say

$$\begin{aligned} r_1 x + r_2 y &< (r_1 + r_2)\, Y, \\ y &< [(r_1 + r_2)\, Y - r_1 x]/r_2, \end{aligned} \tag{25}$$

and this amounts to selection at the second stage of a proportion P_2, where

$$P_1 P_2 = \pi. \tag{26}$$

Here X and Y are quantities still to be determined, subject to (26).

7.14. If normality of distribution can be assumed for x and y, the probability that a compound is selected at the first stage is

$$\theta_1, \theta_1' = \left(\frac{r_1}{2\pi\sigma^2}\right)^{\frac{1}{2}} \int_{\frac{(X-\mu)\sqrt{r_1}}{\sigma}}^{\infty} e^{-\frac{1}{2}t^2} dt, \tag{27}$$

where μ takes the values 0, 1 for the two values of the true activity. The probability that a compound that goes forward from the first stage is selected at the second stage is the probability that y should exceed $[(r_1 + r_2)\, Y - r_1 x]/r_2$ given that x exceeds X, and this can be shown to be

$$\theta_2, \theta_2' = H\left[\frac{(X-\mu)\sqrt{r_1}}{\sigma}, \frac{(Y-\mu)\sqrt{(r_1+r_2)}}{\sigma}, \sqrt{\left(\frac{r_1}{r_1+r_2}\right)}\right], \tag{28}$$

where $H[\xi, \eta, \varrho]$ symbolizes the standardized bivariate normal integral:

$$(29) \qquad H[\xi, \eta, \varrho] = \frac{1}{2\pi(1-\varrho^2)} \int_{\xi}^{\infty} \int_{\eta}^{\infty} \exp \cdot [-(u^2 - 2\varrho uv + v^2)/2(1-\varrho^2)]\, du dv$$

and again $\mu = 0, 1$ are to be substituted. Then in the notation of the earlier part of this Section,

$$(30) \qquad \begin{aligned} \theta &= \theta_1 \theta_2 , \\ \theta' &= \theta_1' \theta_2' , \end{aligned}$$

and

$$(31) \qquad \begin{aligned} m &= r_1 + \theta_1 r_2 , \\ m' &= r_1 + \theta_1' r_2 . \end{aligned}$$

7.15. The investigation of optimal conditions is much simplified by making use of the fact that δ is very small, so that, from (21), θ is almost equal to π. Particular conditions can be examined by choosing values for θ_1 and r_1 (an integer). If σ is known with sufficient accuracy from previous experience, tables of the normal integral can be used to give X from (27), and also to give θ_1' by use of the other value of μ. Hence approximately θ_2 is found as

$$\theta_2 = \pi/\theta_1.$$

Corresponding to any trial value of r_2, Y can be determined from (28) by use of tables of the bivariate normal integral, and (28) also gives θ_2'. From (30) θ' is obtained. Because δ is small, the mean number of animals per compound will scarcely differ from m, or $(r_1 + \theta_1 r_2)$. Finally

$$(32) \qquad \frac{\theta'}{A} \doteqdot \frac{\theta'}{a(r_1 + \theta_1 r_2) + c}$$

is evaluated, which by (21), (22) is proportional to the frequency of true positives per unit cost amongst the selected compounds. Repetition of the cycle of calculation for alternative values of r_1, r_2, θ_1, enables a search for the maximum of (32) to be conducted.

7.16. Enough has been said to indicate Davies's approach to this problem, although the account given is only a simplified version of his. He gave a numerical illustration of the calculations in connexion with the screening of compounds that might increase survival time amongst mice infected with tuberculosis. He compared both the theory and the particular practical application of the two-stage test with a full sequential procedure, and found the latter to be considerably more efficient in respect of frequency of positives selected per unit cost. However, a sequential method is sometimes inconvenient in practice, and he suggested that a three stage selection scheme might be a satisfactory compromise; as far as is known, no research has been done on this.

(Received June 2, 1959.)

REFERENCES

[1] Cox, D. R.: "The use of a concomitant variable in selecting an experimental design". *Biometrika* **44** (1957) 150—158.

[2] Cox, G. M.—Cochran, W. G.: "Designs of greenhouse experiments for statistical analysis". *Soil Science* **62** (1946) 87—98.

[3] Curnow, R. N.: *The Statistical Theory of Selection* Ph. D. Thesis — (1959) University of Aberdeen.

[4] Curnow, R. N.: "The consequences of errors of measurement for selection from certain non-normal distributions". *Bulletin de l'Institut International de Statistique* **37** (1960a).

[5] Curnow, R. N.: (1960b) (paper in preparation).

[6] Davies, O. L.: "The design of screening tests in the pharmaceutical industry". *Bulletin de l'Institut International de Statistique* **36** (1958) 226—241.

[7] Federer, W. T.—Schlottfeldt, C. S.: "The use of covariance to control gradients in experiments". *Biometrics* **10** (1954) 282—290.

[8] Feldt, L. S.: "A comparison of the precision of three experimental designs employing a concomitant variable". *Psychometrika* **23** (1958) 335—353.

[9] Finney, D. J.: "The statistician and the planning of field experiments". *Journal of the Royal Statistical Society*, A **119** (1956) 1—27.

[10] Finney, D. J.: "Stratification, balance, and covariance". *Biometrics* **13** (1957) 373—386.

[11] Finney, D. J.: "Plant selection for yield improvement". *Euphytica* **7** (1958) 83—106.

[12] Finney, D. J.: "Statistical problems of plant selection". *Bulletin de l'Institut International de Statistique* **36** (3) (1958) 242—268.

[13] Finney, D. J.: "A simple example of the external economy of varietal selection". *Bulletin de l'Institut International de Statistique* **37** (1960).

[14] Finney, D. J.—Cope, F. W.: "The statistical analysis of a complex experiment involving unintentional constraints". *Biometrics* **12** (1956) 345—368.

[15] Fisher, R. A.: *The Design of Experiments* (6th edition). London (1951). Oliver & Boyd.

[16] Fisher, R. A.: *Statistical Methods and Scientific Inference*. London (1956). Oliver & Boyd.

[17] Grundy, P. M.—Healy, M. J. R.—Rees, D. H.: "Economic choice of the amount of experimentation. *Journal of the Royal Statistical Society* B**18** (1956) 32—55.

[18] Grundy, P. M.—Rees, D. H.—Healy, M. J. R.: "Decision between two alternatives — How many experiments? "*Biometrics* **10** (1954) 317—323.

[19] Kempthorne, O. *The Design and Analysis of Experiments*. New York (1952).1 Wiley & Sons.

[20] Outhwaite, A. D.—Rutherford, A. A.: "Covariance analysis as an alternative to stratification in the control of gradients. *Biometrics* **11** (1955) 431—440.

[21] Yates, F.: "Principles governing the amount of experimentation in developmenta work". *Nature* **120** (1952) 138.

A KÍSÉRLETEZÉS RENTABILITÁSA

D. J. FINNEY

Kivonat

A statisztikus feladata manapság nemcsak abban áll, hogy a kísérleti eredményeket analizálja, hanem hogy a kísérletek tervezésében is segítséget nyújtson. Szükséges annak felismerése, hogy a szűkebb értelemben vett kísérleti tervezés (design) csak egyik szempontja a tágabb értelemben vett kísérleti tervezésnek és nem szabad, hogy a statisztikus részvétele csak a tervezés leszűkített kombinatorikus problémáira szorítkozzék. Ahhoz, hogy a kísérletező erőfeszítései a lehető leghatásosabbak legyenek, kell, hogy szoros együttműködés legyen a statisztikus és kísérletező tudós között.

Az előadásban szó volt a kísérletezésnek mind a belső, mind a külső ökonómiájáról, tehát arról a kérdésről, hogyan kell a kísérletet úgy tervezni, hogy a rendelkezésre álló időt, anyagot, munkát a legnagyobb hatásfokkal használjuk fel, továbbá arról, hogy mekkora lehetőségeknek kell rendelkezésre állni ahhoz, hogy a kísérlet a legjobban szolgálhassa a kísérlet célját. Az alapvető kutatásoknál gyakran csak a belső ökonómiát érdemes vizsgálni, de sok technológiai kísérletnél a külső tényezőket is tekintetbe kell venni. A második fejezetben felsoroltunk egy-két olyan különleges kérdést, melyeknek megvitatása szükséges.

A 4—7. fejezetek a tervezés egyes speciális problémáinak elméletéről és gyakorlatáról szólnak. Először a vizsgált faktorok optimális számának meghatározásáról esett szó. A szerző véleménye szerint minőségileg és mennyiségileg is az a legelőnyösebb, ha olyan sok faktort vonunk egyszerre be a vizsgálatba, amennyit csak megenged a kísérlet mérete. A második kérdéscsoport a parcellákra és más kísérleti egységekre vonatkozó concomitans információ lehetőségének problémája. Ennek viszonylagos előnye akkor mutatkozik meg, amikor a homogén blokkok képzéséhez használjuk fel, vagy amikor az eredményeket a concomitanssal végzett kovariancia analízis segítségével korrigáljuk.

A harmadik probléma arról szól, hogy miként lehet meghatározni egy adott tényező tisztázására irányuló kísérletezés szükséges mennyiségét. Az egyik fajtája ennek az, amikor meg kell határozni e hasonló kísérletek optimális számát olyan esetben, amikor az a célunk, hogy praktikus használatra megállapítsuk valamely faktor legjobb szintjét (pl. a vetés alatt álló területegységre jutó műtrágya mennyiségét). Ez matematikailag sokkal egyszerűbb, mint amikor olyan kísérletek optimális számát kell meghatároznunk, melyek alapján gyakorlati döntést kell hoznunk abban, hogy két alternatív és élesen különböző eljárás közül melyiket válasszuk (pl. két különböző diéta). Az előadás érintette e második fajta problémát is olyan esetekre, mikor van bizonyos előzetes információnk és a kísérletezőnek az eddigi adatok alapján, vagy azonnal kell döntenie, vagy meg kell határoznia a döntéshez szükséges további kísérletek mennyiségét.

Végül a szerző két példával illusztrálja az egymást követő lépcsőkből álló kísérletezések optimális tervezésének kérdését. Az egyik a termesztett növények varietásainak szelekciójával foglalkozik, amikor is az egymást követő idényben végeznek termőföldi kísérleteket és minden idényben a legjobbaknak egy hányadát választják ki a következő idényben történő folytatólagos kísérlethez. Ha a varietások kezdeti száma és az utolsó idény végén is megtartottak száma továbbá a teljes kísérletezési terület adott, a szelekciós frakciók arányát és az egyes stádiumokhoz szükséges földterületek arányát kell megállapítani. Láthatóan az „egyenlő frakciók minden évben és az egész területnek az egyenlő elosztása" adja azt a sémát, mely általában közel áll az optimálishoz. A második példa arról szól, hogy igen nagy számú kémiai preparátumot szűrünk azért, hogy megkeressük azokat, melyeknek therapiás értékük van. A kezdeti populáció valószínűleg igen kevés aktív készítményt tartalmaz és bizonyára nagyszámú hatástalant is. A laboratóriumi állatokkal végzett standard próbákkal különbséget teszünk a pozitívok és negatívok között és (éppúgy, mint a varietás szelekciónál) célszerű szukcesszív lépcsőket alkalmazhatunk. Ilyenkor célunk az, hogy az utolsó lépésben is megtartott készítmények (melyeket azután természetesen sokkal komolyabb próbáknak

kell alávetni) a lehető legnagyobb hányadát tartalmazzák az aktív készítményeknek. Azonban ezen cél elérésekor tekintetbe kell venni a szelekciós program árát, amit elsősorban a felhasználásra szánt állatok száma határoz meg.

О РЕНТАБЕЛЬНОСТИ ЭКСПЕРИМЕНТИРОВАНИЯ

D. J. FINNEY

Резюме

Задача статистика в наше время состоит не только в анализе результатов экспериментов, но и в оказании помощи при планировании экспериментов. Необходимо сознавать, что планирование экспериментов в узком смысле (design) есть лишь одна из точек зрения планирования экспериментов в более широком смысле и нельзя допустить, чтобы участие статистика ограничивалось бы съуженными комбинаторными проблемами планирования. Для того, чтобы усилия экспериментатора были наиболее плодотворными, необходимо тесное сотрудничество между статистикам и экспериментирующим ученым.

В докладе речь шла как о внутренней, так и о внешней экономии экспериментирования, стало быть о том, как следует планировать эксперимент так, чтобы использовать имеющиеся время, материал, работу с наибольшей выгодой, далее о том, какими возможностями надо располагать, чтобы эксперимент наилучшим образом служил интересам общества. В основных исследованиях часто имеет смысл исследовать лишь внутренного экономию, но при многих технологических экспериментах необходимо принимать во внимание и внешние факты. Во второй главе было приведено несколько особых вопросов, которые необходимо обсудить.

В главах 4—7 говорится о теории и практике некоторых специальных проблем планирования. Сначало говорилось об определении оптимального числа исследуемых факторов. По мнению автора и качественно и количественно наиболее выгодно одновременно вовлекать в эксперимент так много факторов, сколько позволяет размер эксперимента. Вторая группа вопросов — проблема возможности конкомитантной информации, относящейся к делянкам и другим экспериментальным единицам. Относительная выгода этого выявляется тогда, когда это применяется к образовыванию однородных блоков или когда результаты исправляются с помощью ковариантного анализа с конкомитантам.

Третья проблема говорит о том, как можно определить необходимое количество экспериментов для определения данного фактора. Один из видов этого имеет место, когда надо определить оптимальное число этих аналогичных экспериментов в таком случае, когда наша цель заключается в определении наилучшего уровня некоторого фактора для практического применения (например, количества минеральных удобрений на единицу площади посева). Это математических значительно проще, чем определение оптимального количества таких экспериментов, на основании которых нужно практически решить, какой из альтернативных и резко различных методов выберать (например, две различные диэты). Доклад касался и этой проблемы

второго рода для случаев, когда имеется некоторая предварительная информация и экспериментатору на основании имеющихся данных нужно либо немедленно решать либо определить количество экспериментов, необходимых для решения.

Наконец, автор на двух примерах иллюстрирует вопрос об оптимальном планировании экспериментов, состоящих из последовательных ступеней. Один из них занимается селекцией вариантов выращиваемых растений, когда в последовательных сезонах производятся эксперименты на обработанной земле и в каждом сезоне выберается некоторая часть наилучших для продолжения эксперимента в следующем сезоне. Если начальное число вариантов, число удержанных и в конце последнего сезона и полная площадь экспериментирования даны, то необходимо определить соотношение селекционных фракций и соотношение площадей земли, необходимых для отдельных стадий. Очевидно «одинаковые фракции в каждом году и равномерное распределение всей площади» дают схему, которая обычно близка к оптимальной. Второй пример говорит о том, что фильтруется очень большое число химических препаратов для определения тех, которые имеют терапическую ценность. Начальная популяция, вероятно, содержит очень мало активных изделий и очень много недействительных. С помощью лабораторных стандардных проб, произведенных на животных, различаем положительные и отрицательные и (также, как и в случае вариетатной селекции) можем применять целесообразные последовательные ступени. В таком случае цель заключается в том, чтобы изделия, удержанные при последнем шаге (которые затем, естественно, нужно подвергнуть значительно более серьёзным экспериментам) содержали бы как можно большую часть активных изделий. Однако, при достижении этой цели необходимо принять во внимание цену селекционной программы, которая определяется в первую очередь числом используемых животных.