Design and Analysis of Experiments

M N DAS
N C GIRI

Design and Analysis of Experiments

M.N. DAS
Central Water Commission, New Delhi

N.C. GIRI
University of Montreal, Quebec, Canada

WILEY EASTERN LIMITED
New Delhi Bangalore Bombay Calcutta

WILEY EASTERN LIMITED
4835/24 Ansari Road, Daryaganj, New Delhi 110002
4654/21 Daryaganj, New Delhi 110002
16 Resthouse Crescent, Bangalore 560001
Abid House, Dr Bhadkamkar Marg, Bombay 400007
40/8 Ballygunge Circular Road, Calcutta 700019

ISBN 0 85226 158 6

Published by Ravi Acharya for Wiley Eastern Limited, 4835/24 Ansari Road, Daryaganj, New Delhi 110002 and printed by Abhay Rastogi at Prabhat Press, 20/1 Nauchandi Grounds, Meerut 250002. Printed in India.

Preface

The material presented in the book is the product of the experience gained by the authors while offering courses on design and analysis of experiments to graduate and post-graduate students and applied workers in the Institute of Agricultural Statistics, New Delhi and Department of Mathematics, University of Montreal, Canada. The book has been written to suit the academic persons including teachers and students and the applied workers who have to apply statistical principles for the collection and interpretation of experimental data. Almost all the commonly adopted designs have been included in the book. Most of the advanced techniques and methodology available for designing and analysis of experiments are included here together with discussion of elementary concepts and preliminary treatments of different topics. A number of new concepts and alternative methods of treatment of several topics have been presented. Simple and convenient methods of construction of confounded symmetrical and asymmetrical factorial designs, alternative methods of analysis of missing observations, orthogonal latin squares, designs for bio-assays and weighing designs are some of the examples. The book contains in all nine chapters including a chapter on basic statistical methods and concepts. A chapter on designs for bio-assays required for pharmaceutical investigations and response surface designs has also been added.

Complicated mathematical treatment has been avoided while presenting the results in the different chapters. No emphasis has been given on combinatorics while presenting the methods of construction of various designs. More emphasis has been laid on common sense, experience and intuition while introducing the topics, providing proofs of main results and discussing extension and application of the results.

The book can serve as a textbook for both the graduate and post-graduate students in addition to being a reference book for applied workers and research workess and students in statistics.

The authors acknowledge gratefully the facilities provided by the University of Montreal for writing the book.

M.N. DAS
N.C. GIRI

May 1979

DESIGN AND ANALYSIS OF EXPERIMENTS

Contents

CHAPTER 1

Concepts of Experiments: Design and Analysis

1.1 Design of Experiments and Collection of Data

Experimentation and making inferences are twin essential features of general scientific methodology. Statistics as a scientific discipline is mainly designed to achieve these objectives. It is generally concerned with problems of inductive inferences in relation to stochastic models describing random phenomena. When faced with the problem of studying a random phenomenon, the scientist, in general, may not have complete knowledge of the true variant of the phenomenon under study. A statistical problem arises when he is interested in the specific behaviour of the unknown variant of the phenomenon.

After a statistical problem has been set-up, the next step is to perform experiments for collecting information on the basis of which inferences can be made in the best possible manner.

The methodology for making inferences has three main aspects. First, it derives methods for drawing inference from observations when these are not exact but subject to variation. As such, the inferences are not exact but probabilistic in nature. Second, it specifies methods for collection of data appropriately so that the assumptions for the application of appropriate statistical methods to them are satisfied. Lastly, techniques for proper interpretation of results are devised. There has been a great advance in the derivation of statistical methods applicable to various problems under different assumptions. A good coverage of these methods is available in Fisher (1953), Giri (1976) and Scheffe (1959). A good deal of work has also been done in the field of data collection and interpretation techniques. The topic of the present book, viz., design and analysis of experiments falls in the sphere of data collection and interpretation techniques. The other main topic in this regard is the theory of sample surveys. Though the theories of sample surveys and design of experiments are both concerned with data collection techniques, they serve different purposes. The theory of sample surveys has the objective of deriving methods for collection of samples of observations from a population which exists in its own way such that the sample can adequately represent and accurately interpret the population. In the case of experimental data no such population exists in its own way. What exists is a problem and the data have, so to say, to be manufactured by proper experimentation so that an answer to the problem can be inferred from the data. Creation of controlled conditions is the main characteristic feature of experimentation. It is the design of an experiment which specifies the nature of control over the operations in the experiments. Proper designing is necessary also to ensure that the assumptions required for appropriate interpretation of the data are satisfied. Designing is necessary, moreover, to increase accuracy and sensitivity of the results.

Data obtained without regard to the statistical principles cannot lead to valid inferences. They can, no doubt, be analyzed, but the results obtained from them need not hold true subsequently in situations other than those in which they were collected. For example, if two varieties of a crop are to be compared with regard to their yield performance by conducting an experiment, and a particular variety which the experimenter for some personal reasons, say, wants to favour, is allotted to the better plots, then the statistical principles are violated in the experiment and the data collected from the experiment cannot be validly interpreted. Their interpretation may show the favoured variety more promising. But the same result may not be obtained in future when the variety does not receive a favoured treatment. It is, therefore, necessary that the data are collected by adopting proper designs so that they can be validly interpreted. For further reading on this topic Fisher (1953), Kempthorne (1952), and Federer (1955) may be consulted.

1.2 Experiments and their Designs

As already stated an experiment starts with a problem, an answer to which is obtained from interpretation of a set of observations collected suitably. For this purpose a set of experimental units and adequate experimental material are required. Equal sized plots of land, a single or a group of plants, etc. are used as experimental units for agricultural experiments. For animal husbandary experiments animals, animal organs, etc. form the experimental units. Again, for industrial experiments machines, ovens and other similar objects form the experimental units.

The problems are usually in the form of comparisons among a set of treatments in respect of some of their effects which are produced when they are applied to the experimental units. A general name 'treatment' will be given throughout the book to denote experimental material among which comparison is desired by utilizing the effects which are produced when the experimental material is applied to the experimental units.

For example, in agricultural experiments different varieties of a crop, different fertilizer doses, different levels of irrigation, different combinations of levels of two or more of the above factors, viz. variety, irrigation, nitrogen fertilizer, date of sowing, etc. may constitute the treatments.

Given a set of treatments which can provide information regarding the objective of an experiment, a design for the experiment defines the size and number of the experimental units, the manner in which the treatments are allotted to the units and also the appropriate type and grouping of the experimental units. These requirements of a design ensure validity, interpretability and accuracy of the results obtainable from an analysis of the observations. These purposes are served by the principles of (i) randomization which defines the manner of allocation of the treatments to the experimental units, (ii) replication which specifies the number of units to be provided for each of the treatments and (iii) error control which increases the precision by choosing appropriate type of experimental units and also their grouping.

1.3 Methodology for Making Inference

The basis for making statistical inference is one or more samples of observations from one or more variables. These observations are required to satisfy certain assumptions. Some of the assumptions are that the observations should belong to a population having some specified probability distribution, and that they should be independent. The distribution which is usually assumed is the normal distribution because most of the variables encountered in nature are found to have this distribution. Such distributions involve certain unknown quantities called parameters which differ from variable to variable. The main purpose of statistical inference is, (i) to estimate such parameters by using the observations, and (ii) to compare these parameters among themselves again by using the observations and their estimates. The methodology dealing with the first part of the inference has developed into what is known as *theory of estimation* and that for the second part into methods of *testing of hypothesis*. The *maximum likelihood* method of estimation and the *least squares* method of estimation are the two more important methods of estimation. The least squares method of estimation gives the same estimate as that of the maximum likelihood method of estimation under normality assumption. This method has been discussed in Section 1.7. The details about these methods are available in publications on statistical methodology cited in Section 1.1.

For testing of hypothesis, first certain hypothesis involving the parameters is made. The hypotheses are of comparative nature and depend on the type of inference problems. For example, let there be two varieties of a crop denoted by v_1 and v_2, there being n_1 observations of yield on v_1 from plots each of a given size and n_2 observations on v_2 from similar plots. Further, let the observations of the ith variety ($i=1, 2$) have the normal distribution

$$\frac{1}{\sigma_i\sqrt{2\pi}}\, e^{-(y-\mu_i)^2/2\sigma_i^2}$$

where y denotes the variable of yield of the crop, μ_i is the mean yield and σ_i^2 is the variance of the ith variable.

By using the two samples of observations, μ_i and σ_i^2 can be estimated by adopting an appropriate method of estimation. In regard to testing of hypothesis we have two problems of comparison—one being the comparison between σ_1 and σ_2 and other between μ_1 and μ_2. For the purpose of comparison an hypothesis of the type $\sigma_1=\sigma_2$ or $\mu_1=\mu_2$ is made. This type of hypothesis which specifies that the difference between any pair of a number of similar parameters is zero, is known as the *null hypothesis*.

Next, a statistic, that is, a function of the observations, is defined so as to suit the objective of the problem. This is called the *test statistic* as it is used to test the tenability or otherwise of the hypothesis.

Let us imagine a very large number of independent samples of observations from a population under investigation. From each such sample a value of a statistic can be obtained. Thus, there will be as many values of the statistic as the number of samples. We can now think of a probability distribution of these values of the statistic. Such a distribution is called the *sampling distribution* of the statistic.

For testing of hypothesis first the sampling distribution of a test statistic is theoretically derived assuming that the hypothesis regarding the parameters of the original population of the observations is correct. From the sampling distribution it is possible to evaluate the probability of occurrence of the value of the test statistic lying in given ranges of values. For example, one can evaluate the probability of assuming by a test statistic, T all values which are greater than a given value, say, T_0 or less than a value T_0' or which lie between T_0 and T_0'. Usually, for specific test statistic tables are prepared showing different values of T_0 and the corresponding probability or vice versa. Such tables are available in Fisher and Yates (1942). Evidently, the larger T_0 deviates from its central or most expected value, the less is the probability of getting samples which provide T greater than T_0. For testing of hypothesis, usually two values of T_0 are important, one of these corresponds to 5 per cent probability of having samples giving T greater than T_0 when the hypothesis is correct and the other corresponds to 1 per cent probability of having samples giving T greater than T_0. We may call these two values of T_0 as $T_{.95}$ and $T_{.99}$ respectively or in general $T_{1-\alpha}$ corresponding to α per cent probability. These are called respectively the 5 per cent, 1 per cent and α per cent values of significance. We have said earlier that for testing of hypothesis the value of a suitable test statistic is calculated from the observations. If this value of the statistic is greater than its tabulated value at 5 per cent level of significance, the probability of getting such samples as has given the value is less than 5 per cent if the hypothesis is correct. This, in other words, means that the correctness of the hypothesis is greatly doubtful. When the sampling distribution of a statistic is symmetrical, a similar conclusion is also possible if T is less than $T_{.05}$, as the sample value of T is too small in such cases. Hence, in such a situation the hypothesis is rejected at 5 per cent level of significance. This shows that we are likely to reject a correct hypothesis in 5 per cent of the samples. If, again, the calculated value of the test statistic is greater than its tabulated value at 1 per cent level of significance, the probability of getting such samples as has given the value is less than 1 per cent if the hypothesis is correct. In this case the hypothesis is rejected at 1 per cent level of significance. Evidently, rejection at 1 per cent level of significance carries more definite information than rejection at 5 per cent level of significance, because when we reject an hypothesis at 1 per cent level we are likely to reject a correct hypothesis for 1 per cent samples while in the case of 5 per cent level rejection we are likely to reject a correct hypothesis in 5 per cent cases. Thus, in the 1 per cent case we are likely to be in less error by rejecting an hypothesis.

Though we have discussed the test procedure with reference to 5 per cent and 1 per cent levels of significance, it is not necessary that testing should be restricted to only these two levels. Tables are available for testing at other levels of significance if so required. For example, in a situation where more definite information is required, a test at 0.1 per cent level of significance can be applied.

In relation to design and analysis of experiments, two test statistics, viz. (i) *t*-test for testing the significance of the difference between two linear estimates, and (ii) *F*-test or variance ratio test for testing equality of two variances, are usually applied. Some of the hypotheses tested by these tests are discussed below:

Case I: One sample from a normal population with mean μ. When the problem is to know if μ differs from a given or conjectured value μ_0, the hypothesis $\mu=\mu_0$ is made. The test statistic to be applied in this case is called one sample *t*-test and is taken as

$$t=\frac{\bar{y}-\mu_0}{\sqrt{s^2/n}}$$

where $\bar{y}$ is the sample mean based on n observations from a normal population and s^2 is the error mean square given by $(\Sigma(y-\bar{y})^2/(n-1)$ and has $(n-1)$ degrees of freedom (d.f.). The concept of degrees of freedom has been discussed in a later section.

Here, t is said to have the degrees of freedom of s^2, that is, $n-1$. The 5 per cent or 1 per cent values of t depend on the d.f. of t and are tabulated for different values of d.f. Usually, t-table is prepared by showing the 5 per cent level of significance as $t_{.975}$. This is done in consideration of the fact that t has a symmetrical distribution and can be both positive and negative as the mean of the t-distribution is zero. Actually, $t_{.025}$ is negative and the probability of t being less than $t_{.025}$ is .025. Thus, the 5 per cent significance value of t is a value, say, t_0' such that the probability of having samples to give values of t exceeding t_0' is 2.5 per cent and that of having samples to give values of t less than $-t_0'$ is 2.5 per cent so that the probability of t being greater than t_0 or less than $-t_0$ is 5 per cent. This type of test which takes into account the probability of occurrence of extreme values of a test statistic on either direction of its mean value is known as two-tailed test. In such cases the alternatives to the hypothesis, viz. $\mu \neq \mu_0$ do not specify the nature of difference, that is, whether the difference should be only positive or only negative. The hypothesis which specify the direction like $\mu = \mu_0$, the alternative being $\mu > \mu_0$, leads to one-tailed test. In such tests the 5 per cent level of significance corresponds to $t_{.05}$. Though we have discussed the test with reference to 5 per cent level of significance, the same considerations and procedures apply to tests at other levels of significance.

Case 2: Two samples from two normal populations having means μ_1 and μ_2 and a common variance. Setting the problem to test if the estimates of mean from the two samples are homogeneous, the hypothesis $\mu_1 = \mu_2$ is made. The test statistic t is defined as

$$t = \frac{\bar{y}_1 - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $\bar{y}_1$ and $\bar{y}_2$ are the means of the two samples of sizes n_1 and n_2 respectively,

$$s^2 = \frac{\sum (y_{1i} - \bar{y}_1)^2 + \sum (y_{2i} - \bar{y}_2)^2}{n_1 + n_2 - 2},$$

y_{1i} and y_{2i} denoting observations from the two samples.

Here, the d.f. of t is $n_1 + n_2 - 2$. The table of t is the same for both one sample and two sample tests. There are other situations where t-test is applied. Full discussion on the topic has been given in Snedecor (1946), Goulden (1952) and other books on statistical methods.

Case 3: Variance ratio or *F*-test. Let mean squares, s_1^2 and s_2^2 be two estimates of variance σ^2 given by

$$s_1^2 = \frac{\sum (y_i - \bar{y})^2}{n_1 - 1} \quad \text{and} \quad s_2^2 = \frac{\sum (x_j - \bar{x})^2}{n_2 - 1}$$

obtained from two samples $y_1, y_2, \ldots, y_{n_1}$ and $x_1, x_2, \ldots x_{n_2}$ from two normal populations having variances, σ_1^2 and σ_2^2 respectively. In order to test if the mean squares s_1^2 and s_2^2 can be considered to be of equal order, the hypothesis $\sigma_1^2 = \sigma_2^2$ is made. It is tested by defining the test-statistic, F given by

$$F = \frac{s_1^2}{s_2^2} \text{ with } n_1 - 1 \text{ and } n_2 - 1 \text{ d.f.}$$

F-table has been so prepared that F is always greater than 1. Thus, F has to be taken as either s_1^2/s_2^2 or s_2^2/s_1^2, whichever is greater than one. If, however, it is intended to test if one, say, s_1^2 is greater than s_2^2, then the ratio should be s_1^2/s_2^2 and no test is required if $s_1^2 < s_2^2$. We shall encounter this type of situation mostly in this study.

It is not always necessary that s_1^2 and s_2^2 should be calculated from two independent samples. What is necessary is that s_1^2 and s_2^2 should be two independent mean squares.

For details of this topic the reader may consult Lehmann (1959).

1.4 Three Principles of Designs of Experiments

We have seen that randomization, replication and error control are the three main principles of designs of experiments. The roles they play in data collection and interpretation are discussed below.

Randomization

After the treatments and the experimental units are decided the treatments are allotted to the experimental units at random to avoid any type of personal or subjective bias which may be conscious or unconscious. This ensures validity of the results. It helps to have an objective comparison among the treatments. It also ensures independence of the observations which is necessary for drawing valid inference from the observations by applying appropriate statistical techniques.

We shall see subsequently that depending on the nature of the experiment and the experimental units, there are various experimental designs. Each design has its own way of randomization. We shall, therefore, discuss the procedure of random allocation separately while describing each specific design. However, for a thorough discussion on the subject the reader may see Fisher (1942), Kempthorne (1952) and Ogawa (1974).

Replication

If a treatment is allotted to r experimental units in an experiment, it is said to be replicated r times. If in a design each of the treatments is replicated r times, the design is said to have r replications. Replication is necessary to increase the accuracy of estimates of the treatment effects. It also provides an estimate of the error variance which is a function of the differences among observations from experimental units under identical treatments. Though, the more the number of replications the better it is, so far as precision of estimates is concerned, it cannot be increased indefinitely as it increases cost of experimentation. Moreover, due to limited availability of experimental resources too many replications cannot be taken.

The number of replications is, therefore, decided keeping in view the permissible expenditure and the required degree of precision. Sensitivity of statistical methods for drawing inference also depends on the number of replications. Sometimes this criterion is used to decide the number of replications in specific experiments. A more detailed discussion of this topic is deferred till a discussion of experimental error is done. The principle of error control will also follow the same discussion.

1.5 Experimental Error and Interpretation of Data

After the observations are collected they are statistically analysed so as to obtain relevant information regarding the objective of the experiment. As we know, the objective is usually to make comparisons among the effects of several treatments when the observations are subject to variation. Such comparisons are made by the technique of *analysis of variance*. It will be seen subsequently that inference is drawn through this technique by comparing two measures of variation, one of which arises due to uncontrolled sources of variation, called the error variation and the other includes variation due to a controlled set of treatments together with the variation due to all the uncontrolled causes of variation contributing to the error variation.

For example, let there be six plots of land denoted by P_1, P_2, P_3, P_4, P_5 and P_6. The first three plots receive one treatment, say, t_1 and the last three, another treatment, t_2. Suppose, further, that plots P_1 and P_4 receive one level of irrigation, P_2, P_5, another level and P_3, P_6, a third level. Let y_1, y_2, y_3, y_4, y_5 and y_6 denote the observations on a character recorded from the above six plots in that order. Then the comparison y_1-y_2 which denotes the variation in the observations from the first two plots, does not contain any component of variation due to the treatments as both of them receive the same treatment. But the comparison is not free from the effect of the other controlled factor viz. irrigation as P_1 received one level of irrigation while P_2 received another level. Hence, this comparison by itself does not contribute to error variation, But the comparison $(y_1-y_2)-(y_4-y_5)$ is, evidently, free from the variability caused by both the controlled factors, viz. treatment and irrigation. Hence, such comparisons which contain contributions due to uncontrolled factors like soil fertility and management variation which were not specified the plots, build up error variance. The actual measure of error variance is a function of the squares of all such comparisons. The procedure of obtaining it has been discussed in the next section. There are, again, some other concepts of experimental error which we shall discuss at appropriate places.

Determination of Number of Replications

Error variance provides a measure of precision of an experiment, the less the error variance the more is the precision. Once a measure of error variance is available for a set of experimental units, the number of replications needed for a desired level of sensitivity can be obtained as below.

Given a set of treatments an experimenter may not be interested to know if two treatments differ in their effects by less than a certain quantity, say, d. In other words, he wants an experiment which should be able to differentiate two treatments when they differ by d or more.

As discussed in the previous section the significance of the difference between two treatments is tested by t-test where

$$t=\frac{\bar{y}_i-\bar{y}_j}{\sqrt{2s^2/r}}$$

Here, $\bar{y}_i$, and $\bar{y}_j$ are the arithmatic means of two treatment effects each based on r replications, s^2 is a measure of the error variation.

Given a difference d, between two treatment effects such that any difference greater than d should be brought out as significant by using a design with r replications, the following equation provides a solution of r:

$$t_0=\frac{|d|}{\sqrt{2s^2/r}},$$

where t_0 is the critical value of the t-distribution at the desired level of significance, that is, the value of t at 5 or 1 per cent level of significance read from the table. If s^2 is known or is based on a very large number of observations, made available from some pilot pre-experiment investigation, then t is taken as the normal variate. If s^2 is estimated with n degrees of freedom (d.f.) then t_0 corresponds to n d.f.

When the number of replications is r or more as obtained above, then all differences greater than d are expected to be brought out as significant by an experiment when it is conducted on a set of experimental units which has variability of the order of s^2.

For example, in an experiment on wheat crop conducted in a seed farm in Bhopal, India to study the effect of application of nitrogen and phosphorus on yield a randomized block design with three replications was adopted. There were 11 treatments two of which were (i) 10 lb/acre of nitrogen (ii) 20 lb per acre of nitrogen. The average yield figures for these two applications of the fertiliser were 1438 and 1592 lbs/acre respectively and it is required that differences of the order of 150 lb/acre should be brought out significant. The error mean squares (s^2) was 12134.88. Assuming that the experimental error will be of the same order in future experiments and t_0 is of the order of 2.00, which is likely as the error d.f. is likely to be more than 30 as there are 11 treatments, we have all the information to obtain the numbers of replications r from the relation:

$$t_0=\frac{|d|}{\sqrt{2s^2/r}}$$