DESIGN AND BLOCKING CONSIDERATIONS IN EMPIRICAL INVESTIGATIONS

by

W. T. Federer

BU-548-M February 1975

ABSTRACT

Basic definitions, sampling and population structures, statistical inferences, and properties of statistical designs are presented. Two simple experiment designs, the completely randomized and the randomized complete block, are discussed. Two models of variation arising from the addition of treatments to the experimental units are given; they are the additive and multiplicative models. The effects on expectations of mean squares in an analysis of variance are indicated. Some considerations on blocking or stratification are discussed together with the relationship of the above two designs to a simple random sample survey design and a cluster-simple random sample survey design.

In the Mimeo Series of the Biometrics Unit, Cornell University.

DESIGN AND BLOCKING CONSIDERATIONS IN EMPIRICAL INVESTIGATIONS

by

W. T. Federer

BU-548-M* February 1975

1. INTRODUCTION

ANY study or investigation involving the collection of recorded observations contains a procedure or design for collecting the observations. The design associated with a set of recorded observations may have or may not have taken into account all possible uses to which the observations will be subjected. Types of investigations involving collection of observations are:

observational study -- the kind of investigation which involves the collection of observations as a part of an operation, such as, for example, medical, dental, traffic violation, census, weather, student, Dairy Herd Improvement, economic, maintenance and repair, and many other records. These observations may be put to a variety of uses, but these uses were not taken into account in designing the observational study. The observations made were included as part of the operation without any thought as to future usage, such as, for example, a student's Ph.D. dissertation.

survey -- a planned collection of observations on all or a part of the individuals in a universe for a specific purpose. When survey records are used for other than the intended purpose and the specified universe, the survey becomes an observational study.

experiment -- a planned investigation of a phenomenon or of phenomena which may or may not occur in any real world population and a planned collection of observations associated with the entities in the investigation. Considerable

---

* In the Mimeo Series of the Biometrics Unit, Cornell University.

control by the experimenter over the kinds and types of observations is possible in experiments, but not in surveys.

Now, Statistics, the subject, is concerned with

(i) the design (planning) of an investigation,

(ii) the summarization of facts from investigations, and

(iii) the inferences drawn from the investigation relative to specified population parameters.

These three parts of Statistics are inextricably interwoven, and investigators must be concerned with all three parts. In investigations, it is unreal to ignore (i) or to use a definition such as "I assume my sample to be representative of the specified population and therefore will use these statistical procedures." The members of a given Statistics class could be assumed to be representative of the people in the State of New York with respect to their eating habits, but almost everyone would recognize that the assumption is not true. The important question is not what one ASSUMES, but what is TRUE in the REAL world. Hence, before meaningful and realistic statistical inferences can be made it is necessary to KNOW the sampling structure and the nature of the population. Statistical design relates to these topics. The required sampling structure to make the desired statistical inferences is determined through statistical design. It cannot be ignored, defined, or assumed if statistical inferences are to be made from the facts in an investigation.

2. DEFINITIONS

In order to limit misunderstanding, it is important to define a number of concepts. Statistical design encompasses the following items:

(i) Variables and populations -- Before a sampling procedure can be visualized or comprehended, it is essential that a complete description be made of the variables of interest, their patterns of variation or their distributions, and the particular population (universe) facts desired. The nature of the variation and the goals of the investigation should be as precisely and completely specified as possible.

(ii) Measurements and measuring instruments -- The method of measuring, the measuring instrument, and the properties of the measurements with respect to repeatability and bias should be described in detail as they relate to the goals of the study. For example, time to occurrence of an event, say death, as opposed to whether or not the event occurred, are two quite different measurements and can be used in very different ways.

(iii) Treatment design -- The treatment design constitutes the selection of treatments (entities of interest such as nutritional diets, methods of teaching, brands of light bulbs, types of vaccines, etc.) to be used in an experiment. It must be such as to attain the objectives of the investigation. Adequate points of reference (controls) need to be included. For example, it is impossible to determine if a vaccine is effective if none of the test individuals have the disease. Hence, a control is required to determine the incidence of a disease or other event in a population.

(iv) Experiment design -- The experiment (or experimental) design is the sampling arrangement of the treatments in an experiment. It should be one allowing contrasts among treatment parameters to be estimated with high repeatability and without bias.

(v) Sequential design -- A sequential design is one in which observations are obtained sequentially and in which the next treatment to be observed is determined

by previous results in the experiment. It should be one producing minimum average sample size and with a small variance among sample sizes.

(vi) Sample survey design -- A sample survey design refers to the method of selecting units or entities in a survey, with the probability of selecting a unit being known or unknown. It should be one allowing estimation of population parameters of interest with high repeatability and without bias.

(vii) Model building design -- The process of model building involves formulation of model-obtaining observations to test the model-testing adequacy of the proposed model-reformulation of model-reexperimenting-retesting-reformulation-etc. until an adequate model has been devised. The process or method of selecting observations is known as a model building design. Growth curve models, epidemiological models, Mendelian inheritance models, growth and decay models, and many others have been formulated, both with and without random error components.

(viii) Determination of sample size -- Sample size determinations may be made on the basis of finances and personnel available, of personal considerations, of available material, or of statistical considerations, such as having d degrees of freedom associated with the error variance, the standard error of a mean equal a specified per cent of the mean, the probability of a correct ranking of treatments equal to a specified percentage, or any of a number of other statistical criteria. The goal here is to use procedures which minimize sample size.

(ix) Principles and properties of statistical designs -- Such principles of statistical design as blocking, replication, randomization, orthogonality, confounding, sensitivity, balance, and others need much more study before their nature, properties, and consequences will become fully apparent. Such properties as variance-optimality, connectedness (when all contrasts are estimable), equi-replication, etc. are receiving the attention of some statistical researchers.

We know enough about many of these concepts to make use of them in statistical design.

An observation is a fact about a specified unit and may be recorded in the form of a symbol, number, picture, graph, adjective, or some other appropriate distinguishing symbol. For statistical computations these symbols are converted to numbers. If an observation is a fact from which a conclusion may be drawn, we call the observation a datum. One should note that a datum may be a number, but that a number may or may not be a datum. To illustrate, take the numbers 3, 6, 7. These are just numbers and nothing else. However, if we know that 3, 6, and 7 are the average number of headaches per month incurred by 100 patients using three headache preventative methods, A, B, and C, respectively, in an experiment comparing headache preventatives, these numbers would represent data.

A population is defined in terms of the units or elements constituting the population. Under a sampling procedure, an element of the population is called a sampling unit. (Note that one or many observations may be obtained on each sampling unit (s.u.).) A probability sample is one in which the probability of selecting an s.u. is known; otherwise, the sample is defined to be a nonprobability sample. A simple random sample is a probability sample, and is one in which the probability of selection of all possible samples of size s is equal; or alternatively, it is one in which any s.u. has an equal and an independent chance of being selected in a sample of size s.

Inherent in the idea of a frequency distribution of observations is the idea that the distribution has a location parameter such as a mean (or center of gravity) or a median (a central value of all observations). Then, any single observation may be viewed as:

an observation = (a location parameter) + (a deviation of the observation from the parameter).

We may call the location parameter an assignable cause and the deviation from a randomly selected observation an unassignable cause. In general then, we could use a variation model of the form:

an observation = (assignable causes) + (nonassignable causes).

Note that a more appropriate model might be:

an observation = (assignable causes) x (nonassignable causes).

Most statistical procedures considered by statistics classes are of the former (additive) type, rather than the multiplicative type. There are many other models possible.

3. STRATIFICATION (GROUPING) IN POPULATIONS

In order to conceptualize the distributional nature of a population in which there is a grouping into subpopulations, let us consider an example. A litter is the group of individuals born to a dam (mother) after a pregnancy. Litters vary in size and sex. For a particular species, let us consider all possible litters of size s (specified) and of the same sex (specified). For any given litter there are many possible genotypes for each individual. Hence, for each litter the population mean is determined by the genetic constitution of the parents. For a given set of parents the location parameter will be an unknown specified value. There will be a distribution of possible genotypes for any set of parents. The members of any litter may (under certain circumstances) be considered to be a simple random sample from the population. Likewise, the parents of single pregnancies with specific parameters may be considered to constitute a population of litter parameters. Random mating of genotypes results

in a statistical distribution of the litter mean parameters. Hence, there is one overall population composed of a very large number of subpopulations. A population could (under an additive model) have variation of the following nature:

an observation = (population mean) + (deviation of subpopulation mean from population mean) + (a deviation from the subpopulation mean).

Such a structure as the above contains a natural nesting of subpopulations within a population.

As a second example, consider heights of individuals 26 years old for all people living in the U. S. on February 12, 1975. We know that there is a sex difference. Hence, the population of heights may be subdivided into two subpopulations on the basis of sex.

If a simple random sample (srs) is drawn from every subpopulation, the sampling design is known as a stratified-simple random sample. If a simple random sample of subpopulations is made and a srs is drawn from each of the selected subpopulations, the sample design is known as a cluster-simple random sample. In the first example, it would be possible to take a cluster-srs but not a stratified-srs. The latter design could be used in the second example described above, involving only two subpopulations.

Stratification, blocking, or grouping of s.u.'s into subpopulations which have a smaller variance than the population of ungrouped s.u.'s, is a technique that has wide usage in experimentation and in survey work. The more one knows about the variation, the more it is possible to block the s.u.'s into subpopulations which have relatively small variation within subpopulations. In this way one can block the variation among s.u.'s into an assignable cause, variation among subpopulation means, and a nonassignable cause relating to variation among s.u.'s

within the blocks or strata and for which no assignable cause of their variation is known.

Batches of material, greenhouse benches, individuals of same weight, sex, age, and ethnic backgrounds, chicks from same hatch, geographically close areas of land, and food processed on the same day using standardized procedures represent methods of grouping experimental material in order to reduce the variation among s.u.'s within groups.

In a cluster-srs design, we define the subpopulation as the primary sampling unit (psu) and the s.u.'s within a subpopulation as the secondary sampling unit (ssu). Obviously, adding nesting and subsampling would require the need for additional definitions (e.g., tertiary sampling units (tsu), etc.).

4. COMPLETELY RANDOMIZED EXPERIMENT DESIGN

The concept of a completely randomized experiment design and its relationships to resulting statistical inferences need to be fully comprehended by an investigator using such a statistical design. In the first place, the experimenter has a treatment design consisting of v treatments (v=1,2,$\cdots$). It is desired to compare the treatment responses (means, variances, regressions, etc.) in a specified population. Hence, the second item to be considered and defined is the population to which statistical inferences are to be made. The third item is to describe the nature of variation in the population under consideration. Suppose that the variation of an individual response, Y_j, may be described as:

$$Y_j = (\text{population mean } \mu) + (\text{a random error component } \epsilon_j) = \mu + \epsilon_j, \qquad (4.1)$$

where the ϵ_j are identically and independently distributed with a common variance σ^2_ϵ and μ is an unknown population parameter.

If this represents the variation model prior to adding any treatments, then when a treatment i (i=1,2,$\cdots$,v) is applied to a randomly selected s.u., the resulting response could be of the form:

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} = \mu + (\mu_i - \mu) + \epsilon_{ij} = \mu_i + \epsilon_{ij} , \qquad (4.2)$$

where $\tau_i = \mu_i - \mu$, μ_i is the mean of the population of responses resulting from the application of the i^{th} treatment to every s.u. in the population, and the ϵ_{ij} are identically and independently distributed with mean zero and variance σ^2_ϵ. Note also, that the effect of applying treatment i to an s.u. could be multiplicative rather than additive as above. In this case,

$$Y_{ij} = \tau_i(\mu+\epsilon_{ij}) = \mu^*_i + \epsilon^*_{ij} , \qquad (4.3)$$

where μ^*_i is the population mean resulting from applying treatment i to every s.u. in the population and ϵ^*_{ij} are identically and independently distributed with mean zero and variance $\sigma^2_{\epsilon i}$. Equations (4.2) and (4.3) appear similar symbolically, but the underlying model is quite different. Statistical inferences must take this into consideration. The type of treatment effect on responses could also affect the conditions of independence and identicality of distributions.

Now for the fourth item to be considered. A completely randomized experiment design consists of selecting v simple random samples of size r_i from the population and allocating the r_i units to the i^{th} treatment.

If model (4.2) (or 4.3) holds, then an analysis of variance on the $r_1 + r_2 + \cdots + r_v = r_.$ observations would be of the form:

Source of variation	d.f.	Expected value (E[ ]) of the mean square	
		Under (4.2)	Under (4.3)
Total	$r_.$	-	-
Correction for the mean	1	-	-
Among treatments	$v - 1$	$\sigma_\epsilon^2 + f(\tau_i)$	$\sigma_2^2 + f(\tau_i)$
Within treatments	$r_. - v$	σ_ϵ^2	σ_1^2
Within treatment 1	$r_1 - 1$	σ_ϵ^2	$\sigma_{\epsilon 1}^2$
Within treatment 2	$r_2 - 1$	σ_ϵ^2	$\sigma_{\epsilon 2}^2$
⋮			
Within treatment v	$r_v - 1$	σ_ϵ^2	$\sigma_{\epsilon v}^2$

where $\sigma_1^2 = \sum_{i=1}^{v}(r_i-1)\sigma_{\epsilon i}^2/(r_.-v)$, $\sigma_2^2 = \sum_{i=1}^{v}\sigma_{\epsilon i}^2(1-\frac{r_i}{r_.})/(v-1)$, and $f(\tau_i$ = treatment effects) = $E[\Sigma r_i\tau_i^2 - (\Sigma r_i\tau_i)^2/r_.]$. When the r_i are equal, $\sigma_1^2 = \sigma_2^2$. Note that model (4.3) automatically induces unequal variances for each treatment if the null hypothesis, H_o: $\mu_1 - \mu_2 = \cdots = \mu_v$, is not true.

The estimated error variance for any particular treatment sample mean is obtained as variation among s.u.'s treated alike. All r_i s.u.'s receiving treatment i are treated alike in that they all receive the i^{th} treatment, and before applying the i^{th} treatment they all came from the same distribution. In the case of model (4.2), the treatment effect is additive and does not affect the variation among the s.u.'s. Hence, to obtain an estimate of σ_ϵ^2, we use the within treatments mean square, which is a pooled estimate of individual mean squares for each treatment.

In the above, a treatment was applied to an s.u. It was not applied to a part of an s.u. nor to a group of s.u.'s. We define an <u>experimental unit</u> to be

the smallest unit to which one treatment is applied. If a treatment is applied to r_i experimental units (e.u.'s) we say that there are r_i replicates on the i^{th} treatment. If the r_i are all equal, we say that the design is an equi-replicated design. In the discussion thus far, the size of the s.u. and of the e.u. are identical, but this need not be the case in that a single replicate of a single treatment could be a group of s.u.'s (e.g., 25 chicks in one pen could be the e.u. for one replicate of one diet). In order to obtain a valid estimate of an error variance, we need to consider variation among e.u.'s treated alike. One should distinguish between observational units (o.u.) and e.u.'s in that an o.u. is the smallest unit on which an observation or recording is made. In some instances an o.u. and an e.u. are the same, but in many cases they are not.

To illustrate a frequently occurring situation which appears to be a completely randomized experiment design but in reality is not, consider the following. Suppose we draw a simple random sample of v e.u.'s allocating one to each of v treatments. Then, we take m measurements or observations on each of the v treatment e.u.'s. Here the o.u. is one measurement. Given that equation (4.2) holds, the response or yield equation then becomes:

$$Y_{ilh} = \mu + \tau_i + \epsilon_{il} + \delta_{ilh} \qquad (4.4)$$

where i=1,2,$\cdots$,v, j=1, h=1,2,$\cdots$,m measurements, and the δ_{ijh} are taken to be identically and independently distributed with mean zero and variance σ^2_δ. The analysis of variance table is of the form:

Source of variation	d.f.	Expected value of mean square
Total	mv	-
Correction for mean	1	-
Among treatments	v-1	$\sigma_\delta^2 + m\sigma_\epsilon^2 + f(\tau_i)$
Within treatments	v(m-1)	σ_δ^2

Since the computations are the same as in the previous analysis of variance table, many individuals (statisticians included) proceed incorrectly with F-tests of the same hypothesis, compute confidence intervals, etc. This illustrates the necessity of knowing the statistical design.

Three examples of the above situation which are frequently encountered are:

(i) In the comparison of v cooking procedures, all replications of one treatment are made from one batch and on one day; the variation among o.u.'s within treatments, within batch, and within day is wrongly used as an estimate of the error variance for treatment differences.

(ii) v greenhouse flats are used for v different treatments with p plants per flat. The variation among plants within flats is wrongly used as an estimate of the error variance (note this could be an overestimate if competition were present, and an underestimate if the flat to flat variation exceeds zero) for contrasts of treatment means.

(iii) m plants are measured on each of v strains, but all m plants of one strain are in one location. The variation among plants within strain and location is wrongly used to estimate the error variance of differences in strain means.

5. RANDOMIZED COMPLETE BLOCK EXPERIMENT DESIGN

When the population of e.u.'s can be allocated to subpopulations in such a way that the variation within subpopulations is smaller than among subpopulations, a comparison of the v treatments on e.u.'s drawn from a subpopulation would have a smaller variance than if the e.u.'s were selected without regard to the subpopulation structure. Given two statistical designs, the one yielding a smaller variance of a treatment contrast is said to be more efficient than the second. It is more efficient in that more replicates are required for the second design in order to obtain the same error variance for a given contrast.

The randomized complete block experiment design is one in which a cluster-srs of e.u.'s is drawn with v (or more) units in every cluster and in which the v treatments are randomly allocated to the v e.u.'s within each selected cluster (block). A different randomization is required in every block. The response equation before any treatments are applied is taken to be

$$Y_{ij} = \mu + \beta_j + \epsilon_{ij} = \mu + (\mu_{.j}-\mu) + \epsilon_{ij} = \mu_{.j} + \epsilon_{ij} , \qquad (5.1)$$

where the $\mu_{.j}$ is the subpopulation mean for the j^{th} block, the β_j are identically and independently distributed with mean zero and variance σ^2_β, the ϵ_{ij} are identically and independently distributed with mean zero and variance σ^2_ϵ, and the β_j and ϵ_{ij} are independent. Now if the treatment effects are additive, regardless of the e.u. selected for the i^{th} treatment within a subpopulation, the yield equation would be

$$\begin{aligned} Y_{ij} = \mu_{ij} + \epsilon_{ij} = \mu_{i.} + \mu_{.j} - \mu + \epsilon_{ij} &= \mu + (\mu_{i.}-\mu) + (\mu_{.j}-\mu) + \epsilon_{ij} \\ &= \mu + \tau_i + \beta_j + \epsilon_{ij} \end{aligned} \qquad (5.2)$$

where $\mu_{i.}$ is the treatment mean over all subpopulations (blocks), $(\mu_{i.}-\mu) = \tau_i$ = the deviation of the i^{th} treatment mean from the population mean, and the other symbols are as defined in equation (5.1).

Note that in experimental situations encountered in practice, many models are possible. The above one is so frequently presented in statistical writings that one obtains a biased view of the real world (something like only listening to the CBS news or reading the New York Times). Note also that a yield equation similar to (4.3) may fit many experimental situations; it would be of the form

$$Y_{ij} = \tau_i(\mu+\beta_j+\epsilon_{ij}) = \mu_i^* + \tau_i\beta_j + \epsilon_{ij}^* \quad . \tag{5.3}$$

This form would involve a multiplicative effect (interaction) between blocks and treatments and unequal variances for the different treatments given that $\tau_i \neq 1$, a constant, for every i.

In setting up a randomized complete block experiment design, the following steps are involved:

1. Selection of the v treatments to be compared in the forthcoming experiment.

2. Definition and description of the population to which inferences are to be made, including subpopulation structure and variation.

3. Selection of a simple random sample of b subpopulations (blocks).

4. Selection of a simple random sample of v e.u.'s from each of the b selected subpopulations.

5. A separate random allocation of the v e.u.'s within each block to the v treatments.

Under the above structure (5.2), the expected values of the mean squares in the analysis of variance are:

Source of variation	d.f.	Expected value of mean square
Total	bv	-
Correction for mean	1	-
Among blocks	b-1	$\sigma_\epsilon^2 + v\sigma_\beta^2$
Among treatments	v-1	$\sigma_\epsilon^2 + f(\tau_i)$
Blocks x treatments = residual	(b-1)(v-1)	σ_ϵ^2

Even if the yield equation is of the form

$$Y_{ij} = \mu + \tau_i + \beta_j + \beta\tau_{ij} + \epsilon^*_{ij} ,$$

where $\beta\tau_{ij}$ is an interaction effect of treatments and blocks (differences between treatment responses depends upon the block selected). Let $\epsilon_{ij} = (\beta\tau_{ij}+\epsilon^*_{ij})$, then $\sigma^2_\epsilon = \sigma^2_{\epsilon*} + \sigma^2_{\beta\tau}$ in the above analysis of variance table.

The variance of a sample treatment mean $\bar{y}_{i.}$ is $V(\bar{y}_{i.}) = (\sigma^2_\epsilon + \sigma^2_\beta)/r$. The variance of a difference between two treatment means is $V(\bar{y}_{i.}-\bar{y}_{i'.}) = 2\sigma^2_\epsilon/r$, for $i \neq i'$, and hence does not depend upon block to block variation.

Many statistical writings consider the blocks to be a census rather than a srs of subpopulations. Serious consideration should be given to the kinds of statistical inference possible for this situation. Is it realistic to consider inferences only to those subpopulations which have been selected, i.e., the so-called fixed-effects case? It is highly doubtful if any experimenter would want to confine his inferences solely to the blocks used in an experiment. If the blocks represent a factor such as sex, they should be treated as another factor rather than as a stratification variable. It should be noted that a treatment variable is often confused with a stratification variable in practice. This can lead to difficulties in interpretation and in statistical inferences.

6. OTHER CONCEPTS OF STATISTICAL DESIGN

There are many principles and properties associated with statistical design. (See pages 100-104, 108-109, and 131-136 of the first reference listed below.) The completely randomized (cr) and the randomized complete block (rcb) experiment designs may have such properties as equi-replication, orthogonality, and equi-variance of differences between treatment means (also called variance-balanced). An orthogonal design has the properties that

(i) the relative proportion of times a treatment occurs in a block (or other stratification) is constant for all blocks (or other strata),

(ii) the computations remain simple,

(iii) it is variance-balanced for equi-replicated designs and when models of the form (5.2) hold, and

(iv) it is the most efficient equi-replicated design.

Two of the simplest statistical designs, rc and rcb designs, have been considered. There are many other designs. Additional orthogonal experiment designs are the latin square and split-plot designs. There are many nonorthogonal statistical designs.

Additional discussion of the ideas in this paper may be found in the following reference:

Federer, W. T. [1973]. STATISTICS AND SOCIETY. Marcel Dekker, Inc., New York. Chapters III to VI.

Additional references, which may be beyond the comprehension of most beginning students of statistics, are:

Anderson, V. L. and MacLean, R. A. [1974]. DESIGN OF EXPERIMENTS. Marcel Dekker, Inc., New York. Chapters 3 to 5.

Cox, D. R. [1958]. PLANNING OF EXPERIMENTS. John Wiley and Sons, Inc., New York. Chapters 1 to 5.

Note that the term balance has many usages in statistical literature. One use is to call equi-replicated designs balanced designs. Another use is to call orthogonal designs balanced designs. Still other individuals refer to balanced designs which are both orthogonal and equi-replicated. The term variance-balanced has been used above. There are many more contexts in which the term balance is used. In order to avoid confusion, the user should always define what is meant by the term.