Introduction to Design of Experiments

1.1 Introduction

It may be emphasized in the beginning itself that experimental design is first about agriculture, animal science, biology, chemistry, industry, education, etc. and then about Statistics and Mathematics. In fact, experimental design forms the backbone of agricultural sciences; it is an integral component of every research endeavour in agricultural sciences. To design a good experiment the researcher first needs to outline questions to be answered or needs one or more well defined hypotheses. Some examples of typical questions or hypotheses are

(i) How does the feed formulation affect the body weight of animals?

(ii) Which variety of crop species would be good for particular region?

(iii) Does the date of sowing affect the crop yield?

(iv) How does the water availability and its quality influence the crop yield?

(v) How does greenhouse gases emission influence the global warming?

(vi) Does the use of pesticide in crops affect the health of farmers as well as the people consuming the produce?

(vii) Do the micronutrients and minerals influence the productivity of crops?

(viii) Are the resource conservation technologies counterproductive?

(ix) How do altering manure management strategies at livestock operations or animal feeding practices control the methane emission?

It is hard to define a design of experiment, because it is a form of art along with the science. It may be borne in mind that no experiment could be the ultimate one. A good experiment would be one that allows testing what the researcher wants to test and exercises control over everything else. In that sense, a good experiment is one that estimates the effects that the researcher is interested in and simultaneously minimizes controls or eliminates confounding factor(s). A confounding factor is also at times called the nuisance factor. It potentially distorts the data. This factor is sitting hidden in a model and affects the variable being studied, but is not known or acknowledged.

An example would be a study of nutrients like nitrogen, phosphorous, potash and sulphur on the yield of wheat. If the minerals like zinc and manganese in the soil are likely to be present along with the nutrients, and the study measures only nutrients but not the minerals, the study may find that the nutrients do affect the yield of wheat which may or may not be true. The presence of minerals in the soil might also be affecting the yield. If this confounding factor is

identified early enough, adjustments can be made so that the confounding does not destroy the results or introduces bias in results.

Another example could be a study of feed formulation on body weight of animals. If the initial body weights of the animals are likely to be markedly different and the study measures only the periodic body weights of the animals after giving them feeds, the study may find that the feeds do affect the body weight of animals. But this may or may not be true. The initial body weights of animals might also be affecting the final bodyweight of the animals. If the experiment does not take care of this confounding factor, it may influence the results.

In planning any experiment, the experimenter needs to decide

(a) What conditions to study or what are the treatments, e.g., feed formulation, nutrients, irrigations, pesticides, varieties of a crop, resource conservation technologies, dates of sowing, etc.?

(b) What is the experimental material on which the experiment is to be conducted, e.g., animals, human beings, plots in a field, pots in a glasshouse, birds in a pen, tissues in a laboratory, trees, branches of a tree, leaf position on a tree, etc.? To be more specific, experimental material is actually a collection of subjects or units, or plots, etc. and is termed as experimental units or simply units.

(c) What measurements to make or what are the responses and how to measure these accurately and correctly, e.g., yield of crop, body weight of animal, milk yield, number of eggs layed, percentage of plants infected by disease, etc.? Response also denotes the measurable outcome as a result of application of treatments on the experimental units.

In any planned experiment, there are four major sources of variability. These are

(a) Variability due to the conditions under study or the treatments. This variability is desirable and is in fact a deliberate attempt of the researcher to create this variability.

(b) Variability in the experimental units. This variability is unwanted and undesirable but needs to be accounted for. Generally this variability is overlooked by the researchers.

(c) Variability in the measurement process or measuring the response. This part of the variability is unwanted and undesirable. We shall assume throughout that this variability is not present.

(d) Variability absolutely unaccounted for, unwanted and undesirable. The reason for this part of the variability is unknown to the experimenter.

Since it will be assumed that the variability in the measurement process or measuring the response is absent and the response obtained is the true, accurate and correct response of the treatment applied, there will be in fact three major sources of variability to reckon with, *viz.*, (a), (b) and (d).

Looking at the requirements of planning an experiment and the various sources of variability in the planned experiment, the thinking with respect to subject matter (agricultural, biological, industrial etc.) and statistical thinking is needed to reach for a good experimental design. In order to give a concrete form to this thinking, a strong interaction between the researcher and the statistician is absolutely essential for planning and executing an experiment. We begin with three important principles of a designed experiment.

1.2 Principles of design of experiments

There are three basic principles of designing an experiment namely *randomization, replication and local control (blocking)*. These techniques are discussed briefly in the sequel.

1.2.1 Randomization

Randomization means random assignment of conditions to study or treatments to the subjects or experimental material (in fact experimental units), without an obvious plan, prior to start of the experiment. Randomization converts unplanned, systematic variability into planned, chance-like variability. An analytical reason in support of randomization is that essentially it ensures observations generated to be independent and hence the statistical tools used for analysis of observations gathered become applicable. This is more important for the use of test statistic like Snedecor's F and Student's t in hypothesis testing, wherein a pre-condition is that the observations are independent and are identically distributed as normal variate. This is the major concern of randomization.

Randomization also serves the following purposes:

A random assignment of conditions to study or treatments to experimental units ensures that no experimental unit or no treatment received any favour in the beginning of the experiment. Randomization prevents systematic and subjective biases from being introduced into the experiment by the experimenter. In other words, randomization controls the experimenter bias. Lack of a random assignment of experimental units or subjects leaves the experimental procedure open to experimenter bias. It ensures that subjects or experimental units that are favoured or are adversely affected by unknown sources of variation are those "selected using chance device or random permutation" and not systematically selected. For example, in an initial varietal trial of a crop improvement programme, a breeder may assign his or her new strain of experimental crop to the parts of the field that look the most fertile to promote his or her strain; or a nutritionist may assign newly developed feed formulation to healthy and well growing animals to promote a favourite feed. The preferred variety or formulation may then appear to give better results no matter how good or bad it actually is.

Lack of random assignment can also leave the procedure open to systematic biases. Presence of systematic errors in an experiment makes the comparisons among treatments biased, no matter how precise measurements are or how many experimental units are used.

Consider an experiment involving response of four feed formulations to influence the growth in terms of body weight of animals. Suppose that the four feeds, *viz.*, A, B, C, D are given

to 12 animals. Each feed is observed on 3 animals. Without randomization experimenter would take 3 observations on feed one administered to three animals; then on feed two; then on feed three; and then on feed four, i.e., the order of the feeds given to 12 animals are A, A, A, B, B, B, C, C, C, D, D, D. This order might be perfectly satisfactory but could equally well prove to be disastrous. It may be possible that the first three observations on feed A arise from animals that have no disease, the next three observations from feed B arise from animals that acquired foot and mouth disease recently, while the next three observations from feed C arise from animals that are suffering from bloat and the last three observations on feed D arise from animals that are suffering from mastitis. Obviously, the response to feed A would be more pronounced than that to feeds B, C or D, whereas in fact feed A may actually not be better than any of these feeds. Similarly the response to feed C may be more pronounced than that of feeds B or D. It is quite likely that the experimental conditions might favour a particular feed. Order A, B, C, D, A, B, C, D, A, B, C, D, A, B, C, D might help to solve the problem, but it does not eliminate it completely.

Consider this experiment to study the influence of four feed compositions on growth of animals. The total number of ways in which 12 animals can be assigned to 4 feeds so that 3 animals are assigned to each feed is

$$\frac{12!}{3!3!3!3!} = 369{,}600.$$

A random assignment can lead to order A, A, A, B, B, B, C, C, C, D, D, D or A, B, C, D, A, B, C, D, A, B, C, D, A, B, C, D with probability 1/369,600. The probability is indeed infinitesimal, almost zero. Even though such arrangements can happen in a proper randomization, but to avoid such a thing happening purposefully, one must resort to a proper randomization.

Having said this, randomization has at times its limitations also because randomized experiments violate ethical standards and so cannot be adopted in practice in some situations. To make the exposition clear, an example is considered from clinical trials on human beings.

Suppose that a researcher wants to investigate the abortion–breast cancer hypothesis, which postulates a causal link between induced abortion and the incidence of breast cancer. A hypothetical controlled experiment starts with subjects (pregnant women) and divides them randomly into treatment group (receiving induced abortions) and control group (bearing children). Regular cancer screenings are conducted for women from both groups. Such an experiment would always run counter to common ethical principles. It would also suffer from various confounds and sources of bias, *e.g.*, it would be impossible to conduct it as a blind experiment.

The published studies investigating the abortion–breast cancer hypothesis generally start with a group of women who already have received abortions. Membership in this "treated" group is not controlled by the investigator: the group is formed after the "treatment" has been assigned.

Consider another study in which a researcher wants to compare some phenotypic traits among animals of three different breeds. In this case the experimenter starts with a number of animals, some animals belong to breed one, some belong to breed two and rest of them

belong to breed three. The researcher collects observations on the phenotypic traits of interest. In this experiment, it is not possible to assign breed to an animal at random because the animals already are of a particular breed. In other words, the assignment of breeds to animals is not under the control of the experimenter.

In view of this, design of experiments or experimental design is the design of all information-gathering exercises where variation is present, whether under the full control of the experimenter or not. The latter situation is usually called an observational study, and would be beyond the scope of this book. We shall, henceforth, focus on randomized experiments.

1.2.2 Replication

Replication is the repetition of the conditions of study or treatments under investigation to different experimental units, be it animals or pots or plots in a field, or position of leaf on a plant. Replication intends to increase the size of the experiment.

Replication enables the experimenter to obtain a valid estimate of the experimental error. Estimate of experimental error permits statistical inference; for example, performing tests of significance or obtaining confidence interval, etc. If there is no replication, then the researcher would not be able to estimate the experimental error. And as will be seen in the later Chapters, it is against this estimated experimental error the null hypotheses are tested.

Consider an example where two levels of Nitrogen as A = 30 kg/ha and B = 60 kg/ha are applied to wheat crop. The interest of study is to see how nitrogen influences the yield of wheat. In experiment 1, there are four plots available and each level of nitrogen is applied to two plots randomly. The plots receiving the same level of nitrogen are expected to give the same response. The difference gives the experimental error. In experiment 2, there are six plots and each level of nitrogen is applied to three plots randomly. The yield in kg per plot is given in bracket.

Experiment 1	A (31.5)	B (30.6)	B (28.2)	A (32.8)		
Experiment 2	B (26.8)	A (31.7)	A (33.4)	B (28.6)	A (32.9)	B (27.9)

In experiment 1, the experimental error can be estimated as

$$\frac{(31.5-32.8)^2+(30.6-28.2)^2}{2}=\frac{7.45}{2}=3.725 \quad .$$

This can also be estimated as

$$(31.50-32.15)^2+(32.80-32.15)^2+(30.60-29.4)^2+(28.20-29.40)^2$$
$$=0.4225+0.4225+1.4400+1.4400=3.725$$

Here the average yield from A is $\frac{31.5+32.8}{2}=32.15$ and the average yield from B is

$\frac{30.6+28.2}{2}=29.40$.

In this experiment, the experimental error is 3.725.

In experiment 2, the experimental error can be estimated as

$$\frac{(31.7-33.4)^2+(31.7-32.9)^2+(33.4-32.9)^2+(26.8-28.6)^2+(26.8-27.9)^2+(28.6-27.9)^2}{3}$$

$$=\frac{9.52}{3}=3.173$$

This can also be estimated as

$$(31.700-32.667)^2+(33.400-32.667)^2+(32.900-32.667)^2+(26.800-27.767)^2$$
$$+(28.600-27.767)^2+(27.900-27.767)^2=3.173$$

Here the average yield from A is $\frac{31.7+33.4+32.9}{3}=32.667$ and the average yield from B is $\frac{26.8+28.6+27.9}{3}=27.767$

In this experiment, the experimental error is 3.173.

Increasing the size of the experiment or increasing the replication also helps to increase the precision of estimating the pairwise differences among the treatment effects. This is so because with the increase in the size of the experiment, the experimental error reduces. As can be seen from the example above, the experimental error in experiment 2 reduces from that in experiment 1.

It may be emphasized here that replication is different from repeated measurements. Suppose that the four animals are each assigned to a feed and a measurement is taken on each animal. The result is four independent observations on the feed. This is replication. On the other hand, if one animal is assigned to a feed and then measurements are taken four times on that animal, the measurements are not independent. We call them repeated measurements. The variation recorded in repeated measurements taken at the same time reflects the variation in the measurement process, while variation recorded in repeated measurements taken over a time interval reflects the variation in the single animal's responses to the feed over time. Neither reflects the variation in independent animal's responses to feed. We need to know about the latter variation in order to generalize any conclusion about the feed so that it is relevant to all similar animals.

Generally speaking, all the treatments should be replicated same number of times. In that case the total number of experimental units is a scalar multiple of the number of treatments. In case the total number of experimental units is not a scalar multiple of the number of treatments, then the replication of treatments should be as equal as possible. In other words, the replications of treatments should not differ by more than one. For instance, if the number of treatments

is 7 and the number of experimental units is 24, then 4 treatments may be replicated three times and three treatments may be replicated 4 times. But there might occur some experimental situations where some treatments may need to be replicated more number of times than the other treatments and the difference in replications is more than one. In fact, there do occur experimental situations where some treatments are not replicated because not enough material is available for replication. There are other experimental situations where even a single complete replication is not experimented because of the resource constraint and economy. There will be occasions to refer to all such situations later in the book (both parts I and II).

1.2.3 Local control or blocking

Experimental conditions under which an experiment is run should be representative of those to which the conclusions of the experiment are to be applied. For inferences to be broad in scope, experimental conditions should be rather varied. Unfortunate consequence of increasing scope of experiment is an increase in variability of response. Blocking is a technique that is often used to help deal with this problem

As mentioned earlier, one source of variability is the experimental material or experimental units. Local control or blocking is a technique to account for the variability in response because of the variability in the experimental units. To block an experiment is to divide the experimental units into groups or blocks of similar units in such a way that the observations in each block are collected under relatively similar experimental conditions. If blocking is done well, the comparisons of two or more treatments are made with more precision than similar comparisons from an unblocked design.

It may be mentioned that the blocking is advantageous if the variability within the groups or blocks is as small as possible and between groups or blocks is as large as possible.

In feeding trials litters of the same animal can form natural blocks. Similarly, animals with similar body weights can also form blocks; animals with genetic similarity can also form blocks; animals with same age can also be a criterion for forming the blocks; animals with same lactation number or stage can be another consideration for forming blocks. Fertility gradient in field experiments can be a way of forming blocks. In this case, the blocks are formed perpendicular to the fertility gradient. Salinity levels in the field could also be a criterion for forming blocks in field experiments. Age of the trees in horticultural experiments could be a source of variability and trees of same age can form natural blocks. The soil depth may be another criterion of blocking. In hilly areas, terraces may be taken as natural blocks.

From practical considerations, the contiguous experimental units should form blocks. But sometimes it may so happen that the homogeneous experimental units may not be contiguous. In that case blocks formed are irregular in shape. It is indeed possible that the blocks may not have same number of experimental units. If we force the blocks to be of same size, then again variability may creep in and the purpose of blocking is defeated.

There can be more than one source of variability in the experimental material. If there are two sources of variability in the experimental units, then recourse is made to forming blocks in two directions, called rows and columns. The conditions under study or the treatments are applied to the cells at the intersection of rows and columns. Row-column designs are also useful for the situations, wherein, the fertility gradient is along the diagonal in the field. Sometimes, the two blocking systems may be nested one within another. There may be larger blocks and within each larger block there are smaller blocks, called sub-blocks. The treatments are applied to the sub-blocks within larger blocks. There may be another type of experimental situation where within the larger blocks, rows and columns are formed. The treatments are applied to the cells within each larger block.

When there is blocking, then the randomization of conditions to study or treatments to experimental units changes. The exact randomization will be described in the respective chapters.

1.3 Brief history of design of experiments

The statistical principles underlying design of experiments were pioneered by R. A. Fisher in the 1920s and 1930s at Rothamsted Experimental Station, an agricultural research station around forty kilometres north of London. Fisher had shown the way on how to draw valid conclusions from field experiments where nuisance variables such as temperature, soil conditions, and rainfall are present. He had shown that the known nuisance variables usually cause systematic biases in results of experiments and the unknown nuisance variables usually cause random variability in the results and are called inherent variability or noise. He introduced the concept of analysis of variance (ANOVA) for partitioning the variation present in data (a) due to attributable factors, and (b) due to chance factors. The methodologies he and his colleague Frank Yates developed are now widely used. Their methodologies have a profound impact on agricultural sciences research.

Though the experimental design was initially introduced in an agricultural context, the method has been applied successfully in the industry since the 1940s. George Box and his co-workers developed experimental design procedures for optimizing chemical processes, particularly response surface designs for chemical and process industries. W. Edwards Deming taught experimental designs to Japanese scientists and engineers in the early 1950s at a time when Japanese products were considered to be of poor quality. Genichi Taguchi, a Japanese engineer, suggested a number of techniques using orthogonal arrays. Taguchi coined the concept of robust parameter design and process robustness. Around 1990, Six Sigma, a new way of representing continuous quality improvement came into existence. Six sigma employs a technique that uses statistics to make decisions based on quality and feedback loops and is widely used by many large manufacturing companies. Design of experiments is considered an advanced method in the Six Sigma programs.

Recently, experimental designs are also being used in clinical trials. This evolved in the 1960s when medical advances were previously based on unreliable data. For example, doctors used to examine a few patients and publish papers based on such data. The biases resulting

from these kinds of studies became known. This led to a move toward making the randomized double-blind clinical trial the standard for approval of any new product, medical device, or procedure. The scientific application of the valid designing and analysis following proper statistical methods became very important in clinical trials.

More recently the experimental design techniques have started gaining popularity in the area of computer-aided design and engineering using computer/simulation models including applications in manufacturing industries.

1.4 Some preliminaries

In the context of design of experiments, some widely used terminologies including those discussed earlier are now defined in the sequence.

The term conditions to study or *treatments* is used to denote the different objects, methods or processes among which comparison is made. Some examples of treatments are different kinds of fertilizer in agronomic experiments, different irrigation methods or levels of irrigation, different fungicides in pest management experiments and doses of different drugs or chemicals in laboratory experiments, different varieties of crops, different pesticides, grazing systems for animals, different tree species in agro-forestry experiments, different concentrations of a solute in chemical experiments, etc.

A *control* treatment is a standard treatment that is used as a baseline or basis of comparison for the other treatments. This control treatment might be the treatment which is currently in use, or it might be a no treatment at all. For example, a study of new pesticides could use a standard pesticide as a control treatment, or an experiment involving fertilizers may have one treatment as no fertilizers at all. In clinical trials, a control treatment is generally a placebo.

Experimental units are the subjects or objects on which the treatments are applied. For example, plots of land receiving fertilizer, groups of animals receiving different feeds, or batches of chemicals receiving different temperatures, pots in glasshouse experiments, Petri dishes or tissues to culture bacteria or micro-organisms in laboratory experiments, etc.

Responses are measurable outcomes, which are observed after applying a treatment to an experimental unit. Alternatively, the response is what we measure to find out what happened in the experiment. In an experiment, there may be more than one response. Some examples of responses are grain yield or straw yield, nitrogen content in plants or biomass of plants, quality parameters of the produce, percentage of plants infested by disease, weight gain by animals, etc.

Factors are the variables whose influence on a response variable is being studied in the experiment. If only one factor is being studied in an experiment then such an experiment is called a single factor experiment. If more than one factor is being studied simultaneously in an experiment, then such an experiment is called multi-factor or factorial experiment. The term factor is commonly used in the case of factorial experiments. For example, temperature and concentration of chemicals in a chemical experiment are two factors, Nitrogen, Phosphorus and Potassium fertilizers are three factors in an agronomic experiment, dose and time of application of a chemical formulation are two factors in a laboratory experiment.

The term *factor levels* or a simply *levels* is used to denote the values or settings that a factor takes in a factorial experiment. For example, doses of a nitrogenous fertilizer as 0 kg/ha, 30 kg/ha, 80 kg/ha are three levels of the fertilizer, 10°C, 20°C, 30°C are three levels of temperatures in a chemical experiment, 10%, 20%, 30%, 40% concentration of a solute in a solution are four levels in a laboratory experiment, presence of polythene sheet on the surface of soil or its absence could be two levels of a practice in water management study.

Treatment combination or *level combination*: In factorial experiments, the set of values for all factors in a trial is called treatment combination or level combination. For example, if in a chemical experiment, there are two factors *viz.*, temperature and concentration and both these factors have three levels each as 10°C, 20°C, 30°C and 10%, 20%, 30%, respectively, then total number of treatment combinations is $3 \times 3 = 9$ and these 9 combinations are (10°C, 10%); (10°C, 20%); (10°C, 30%); (20°C, 10%); (20°C, 20%); (20°C, 30%); (30°C, 10%); (30°C, 20%); (30°C, 30%). These combinations are in fact 9 treatments. We can label the 9 treatments as 1, 2, 3, 4, 5, 6, 7, 8, 9. The association is the following: 1 ~ (10°C 10%); 2 ~ (10°C, 20%); 3 ~ (10°C, 30%); 4 ~ (20°C, 10%); 5 ~ (20°C, 20%); 6 ~ (20°C, 30%); 7 ~ (30°C, 10%); 8 ~ (30°C, 20%); 9 ~ (30°C, 30%).

Conversely, if there are 9 treatments and these 9 treatments can be thought of as combination of levels of two factors, both having 3 levels each, then the same association can be used to convert the treatments into treatment combinations.

Application of a treatment combination to an experimental unit is called a *run* or a *design point* in factorial experiments.

An *observational unit* is a unit on which the response variables are measured. Observational units are often the same as experimental units, but this may not be true always. The mistake of confusing observational unit with experimental unit leads to pseudo-replication as discussed in a paper by Hurlbert (1984). For example, consider an experiment to investigate the effects of ultraviolet (UV) levels on the growth of smolt. The experiment is conducted in two tanks where one tank receives high levels of UV light and the other tank receives no UV light. Fish are placed in each tank and at the end of the experiment growths of the individual fish are measured. In this experiment, the tanks are the experimental units but the observational units are the smolts. The treatments, presence and absence of UV light, are applied to the tanks and not to individual fish but a whole group of fish are simultaneously exposed to the UV radiation. Here any tank effect is completely confounded with the treatment effect and cannot be separated. Another example is that inorganic fertilizers are applied to plots in a field containing some plants. At the time of harvest, all the plants in the plot are not harvested. Only a sample of plants is harvested. In this case once again the plot is the experimental unit to which fertilizers are applied but the observational units are the plants sampled.

A *treatment contrast* or simply a *contrast* is a linear function of treatment effects such that the sum of the coefficients is zero. For instance, if $\tau_1, \tau_2, \cdots, \tau_v$ denote the v treatment effects, then $p_1\tau_1 + p_2\tau_2 + \cdots + p_v\tau_v$ is a contrast if and only if $p_1 + p_2 + \cdots + p_v = 0$. A big advantage of contrast is that one can make all the possible pairwise treatment comparisons. It also enables to make any

other comparison among treatment effects. For details the reader may see Chapter 3. A contrast is said to be elementary contrast if and only if only two of the coefficients are non-zero while all other coefficients are zero. $\tau_1 - \tau_2$, $\tau_i - \tau_l$, etc. are elementary contrasts. Other contrasts could be $2\tau_i - \tau_l - \tau_u$ or $\tau_i - \tau_l - \tau_u + \tau_w$.

1.5 Factorial experiment

There has been a description of treatments in Section 1.4. There has also been a description of factors and treatment combinations. It may be emphasised that the treatments in any experiment may either be unstructured or structured. Unstructured treatments are actually levels of a single factor in a single factor experiment. In these experiments, the interest is in making all the possible pairwise treatment comparisons. At times comparisons between subsets of treatments or among treatments within subgroups also form a part of the hypotheses to be tested. These experiments are generally conducted as unblocked design, block design or a row-column design or a nested design depending upon the problem to be solved and the nature of the experimental material.

On the other hand, the treatments may be structured in the sense that there are several factors and each factor has several levels. The treatments in this case are the level combinations of all the factors. The interest of the researcher is in estimating the factorial effects comprising of main effects and the interaction effects rather than making all the possible pairwise treatment comparisons or subgroups testing. The treatment sum of squares in this case is partitioned into main effects and interaction effects sum of squares. Otherwise, the experiment once again is conducted using an unblocked design, a block design or a row column design or a nested design as one would have used in case of unstructured treatments. There are no special designs for running factorial experiments. However, treatment structure or their fraction may be obtained based on availability of resources and objectives of the experiment. An incomplete block design in factorial experiment may be obtained in such a way that the desired factorial effects are estimated with more precision by sacrificing information on factorial effects of less interest, particularly the higher order interactions.

If there are several factors it is always advantageous to study them simultaneously rather than studying them separately. Suppose that there are two factors A and B, A having three levels and B having four levels. Let the levels of the two factors be denoted by 0, 1, 2 and 0, 1, 2, 3, respectively. The association between the 12 treatment combinations and the treatments is the following:

1 ~ (0, 0); 2 ~ (0, 1); 3 ~ (0, 2); 4 ~ (0, 3); 5 ~ (1, 0); 6 ~ (1, 1); 7 ~ (1, 2); 8 ~ (1, 3); 9 ~ (2, 0); 10 ~ (2, 1); 11 ~ (2, 2); 12 ~ (2, 3). The analysis of 12 treatments run in two replications as a completely randomized design (CRD) or an unblocked design is

Source	**DF**
Treatments	11
Error	12
Total	23

On the other hand if it is known that the treatments are structured as two factors with three and four levels, respectively, the same design can be analysed as factorial experiment run in a CRD with two replications. In that case the treatments can be partitioned into main effects and interaction effects as explained below.

Source	DF
Treatments	11
Main Effect A	2
Main Effect B	3
Interaction A*B	6
Error	12
Total	23

Another advantage of using a factorial experiment is the following: If one looks carefully at the design in the example, each treatment combination appears twice in the completely randomized design (CRD) because the replication is two. However, if one looks at the replications of levels, then the levels 0, 1, 2 of factor A are replicated eight times each. Similarly the levels 0, 1, 2, 3 of factor B are replicated six times each. This replication of levels within the replication of treatment combinations is known as hidden replication. It is because of this hidden replication that some comparisons are made with higher precision in factorial experiments. These experiments have another advantage that these allow to study the interaction effects. If an experiment is conducted separately for each factor, the interaction effects cannot be estimated. Moreover, to achieve the same precision as in factorial experiment, the replications would have to be large. For example, if factor A with three levels is conducted as CRD, to have 12 degrees of freedom for error, one needs to have 5 replications making a total of 15 observations. Similarly, if the factor B is run as a CRD, then to have 12 degrees of freedom for error, the replication should be four making a total of 16 observations. So the total number of observations becomes 31, but the interaction effect cannot be estimated. On the other hand a factorial experiment requires only 24 observations to have 12 degrees of freedom for error and allows estimation of interaction effect also. This means that running an experiment separately for each factor would result into an increase in the cost of the experimentation and interaction effects would have to be sacrificed. But factorial experiments have an advantage that not only the cost is reduced, the interaction effects are estimable and can be studied. Further the hidden replication in factorial experiments leads to an improved precision of the factorial effects.

1.6 Variability in the experimental data

The data generated through designed experiments exhibit a lot of variability. In Section 1.1 there was a mention of various type of variability in the data generated. The variability may be wanted, desirable, unwanted, undesirable but is controllable in the sense that it can be accounted for. There is also some more variability, unwanted, undesirable and uncontrollable. The reason for its presence is unknown. In an example in Section 1.2.2, it has been seen that even the experimental units (plots) subjected to the same treatment also give rise to different observations, thus creating variability. These plots are expected to give same response, but

actually the responses are different; reasons unknown. The statistical methodologies, in particular the theory of linear estimation and analysis of variance, enable us to partition the total variability in the data into two major components. The first major component comprises of that part of the total variability to which we can assign causes or reasons. The second component comprises of that part of the total variability to which we cannot assign any cause or reason. This variability arises because some factors are unidentified as a source of variation. Even after careful planning of the experiment, this component is always present and is known as *experimental error*. The observations obtained from experimental units identically treated are useful for the estimation of this experimental error. Ideally one should select a design that will give experimental error as small as possible. There is, though, no rule of thumb to describe what amount of experimental error is small and what amount of it can be termed as large. A popular measure of the experimental error is the percent Coefficient of Variation (CV). Generally the researcher desires the CV to be small, though there is no degree of smallness defined.

The explainable part of the total variability again has two major components. One major component is the conditions to study or the treatments. This part of the variability is wanted or desirable. There is always a deliberate attempt on the part of the experimenter to create variability by the application of several treatments. So in every designed experiment treatments are one component that cause variability. The other component of the explainable part of variability is the experimental units. This variability is unwanted and undesirable. The factors that cause this variability are called nuisance factors. This part of the variability is accounted for by using the principle of local control. Before planning the experiment, the experimenter must have a complete knowledge about the experimental units on which the experiment would be conducted and the sources of variability in the experimental units. If this variability is substantial and is not accounted for by proper designing of experiment, then this component would sit in the experimental error and make it unduly large. The end result would be a bad experiment. As mentioned earlier in Section 1.2.3, there could be many ways of accounting for the variability due to experimental units. The remedy will depend upon the sources and nature of the factors causing variability in the experimental units. As a matter of fact, the way to account for the variability in the experimental units will dictate what type of design is to be used. Many a time, depending upon practical constraints, a naive design may be the best design.

As-a-matter-of-factly many designs have been evolved in the literature depending upon how the variability present in the experimental units is taken care of and how the treatments are allocated to the experimental units or how the randomization is done. If the experimental units are homogeneous and do not exhibit considerable variability, then the treatments are applied randomly to all the experimental units assuming that all the experimental units are uniform. Such designs are known as zero-way elimination of heterogeneity designs or completely randomized designs (CRD) and will be dealt with in detail in Chapter 2. On the contrary, if the variability present in the experimental units is sizeable, then forming groups called blocks containing homogeneous experimental units can account for this variability if the variability in the experimental units is due to one nuisance factor only. As opposed to the allotment of treatments randomly to all the experimental units in a CRD, the treatments in this case are allotted randomly to the experimental units within each block. Such designs are termed as one-

way elimination of heterogeneity setting designs or the block designs. The most common block design is the randomized complete block (RCB) design which is also considered in Chapter 2. If there are two sources of variability in the experimental units, then the experimental units are grouped into arrays, called rows and columns, and the intersection of rows and columns, called cells are the experimental units. The treatments are allocated to the cells. For the randomization purpose, first the rows are randomized and then the columns are randomized. There is no randomization of treatments possible within rows and/or within columns. Such designs are called row-column designs or two-way elimination of heterogeneity designs. A special class of these designs is the Latin square designs and will be studied in Chapter 2.

Generally in experimentation, the number of treatments is large. For large number of treatments, the blocks become large if one has to apply all the treatments in a block, as desired by the RCB design. It may then not be possible to maintain homogeneity among plots within a block and the basic purpose of forming blocks is defeated. The intra block variance or the variance per plot becomes large resulting in a large experimental error and thus a high value of coefficient of variation (CV). To overcome this problem, recourse may be made to an incomplete block design. A block design is said to be an incomplete block design if the design has at least one block that does not contain all the treatments. Some common incomplete block designs are balanced incomplete block (BIB) design, partially balanced incomplete block (PBIB) design including Lattice designs – square and rectangular, cyclic design, alpha design, etc. The concept of incomplete block design can also be extended to incomplete row and / or incomplete column designs. An example of incomplete row-column design is a Youden design or a Generalized Youden design or a Pseudo Youden design.

The unexplainable part of the variability, called the experimental error, is always present. But through controlled experimentation, it is always possible to control this component of variability. It is desirable that this component is as small as possible. This part, therefore, can be controlled by proper designing of an experiment. This means that the design should be such that it accounts for all the sources of variability in the experimental units. If the experimenter fails to control the variability in the experimental units through proper designing, then the experimental error can be controlled by a very useful and important statistical technique called analysis of covariance. This would be dealt with in Chapter 4.

1.7 Shape and size of experimental units

In agricultural field experiments, often plots in fields are used as experimental units. One important issue in this context is the shape and size of the plots and their arrangement. Some general considerations for plot arrangements are given in the sequence.

i) The experimental area should be as uniform as possible. Uneven sites may lead to high error.

ii) Plots should be either rectangular or square and equal in area.

iii) The orientation of the plots should be same, for example, the longer side of the rectangular plots should be parallel to each other.

iv) Uniformity trials may be conducted to get optimum shape and size of the plots. Uniformity trial involves growing a particular crop on a field or piece of land with uniform conditions. All sources of variation except that due to native soil differences, are kept constant. At the time of harvest the entire field is divided into smaller units of same size and shape and the produce from each such unit is recorded separately. The smallest the basic units, the more detailed are the measurements of soil heterogeneity.

v) It may not be economically feasible to conduct a uniformity trial. Even time constraint may be prohibitive in the conduct of a uniformity trial. If soil parameters are known, then these can be used for formation of plots and blocks. At times, the residuals obtained from a previous designed experiment conducted at that place may be used as covariate in the analysis of data generated.

1.7.1 Determining optimum size of plots

In the sequel are described some methods for understanding soil fertility variation/plot size.

i) *Fertility contour map*: An approach to describe the heterogeneity of land is to construct the fertility contour map. This is constructed by taking the moving averages of yields of unit plots and demarcating the regions of same fertility by considering those areas, which have yield of same magnitude. This approach of describing the variation in fertility has been adopted by large number of workers in India and abroad. Fertility contour map can also be developed using the soil parameters in the observed samples obtained from the experimental area.

ii) *Maximum Curvature Method*: In this method basic units of uniformity trials are combined to form new units. The new units are formed by combining columns, rows or both. Combination of columns and rows is done in such a way that no columns or rows are left out. For each set of units, the coefficient of variation (CV) is computed. A curve is plotted by taking the plot size (in terms of basic units) on X-axis and the CV values on the Y-axis of graph sheet. The point at which the curve takes a turn, *i.e.*, the point of maximum curvature is located by inspection. The value corresponding to the point of maximum curvature will be optimum plot size.

iii) *Fairfield Smith's Variance Law*: Smith (1938) suggested an empirical relation between variance and plot size. Smith developed an empirical model representing the relationship between plot size and variance of mean per plot. This model is given by the equation

$$v_x = \frac{v_1}{x^b}$$

$$or \log v_x = \log v_1 - b \log x$$

where x is the number of basic units in a plot, is the variance of mean per plot of x units, is the variance of mean per plot of one unit and b is the characteristic of soil and measure of correlation among contiguous units. If $b = 1$, $v_x = \frac{v_1}{x}$ and the units making up the plots of x units are not correlated at all. If $b = 0$, $v_x = v_1$ and the units making up the plots of x units are perfectly correlated and hence there is no gain due to larger size of plots.

Generally b lies between 0 and 1. The values of v_1 and b are determined by least squares method.

This law can further be used for arriving at an optimum plot size. Smith recommended the cost function $C = C_1 + xC_2$, where C_1 is overhead cost which is independent of plot size and C_2 is the consideration of cost by a unit increase in the plot size. Optimum value of plot size is the one which minimizes the cost per unit of information *viz.* $(C_1 + xC_2)$. Once b is estimated from uniformity trial data, the optimum size of plot can be obtained using the following formula

$$x_{opt} = \frac{bC_1}{(1-b)C_2}.$$

Here it must be mentioned that the value of x_{opt} is some multiple of the basic plot size. For example, if $x_{opt} = 2$, then it means that the optimum plot size is twice the basic plot size used in the uniformity trial for estimating b.

1.7.2 Shape of the plots

Shape of plots in agricultural field experiments should be decided after taking care of following points:

i) Crop to be grown

ii) Convenience of planting and harvesting crop

iii) Ability to use machineries (if machineries are going to be used)

iv) Presence or absence of fertility gradient

v) Variation in soil depth

1.8 Determination of number of replications

A very important question that needs to be answered by the experimenter is about the number of replications to be used in a design. Although the answer largely depends upon the resources available, there are some scientific reasons also that help in determining the optimum replication number. The following points should be kept in mind while determining number of replications of the treatments.

i) The foremost important consideration in the determination of replication number is that there should be adequate error degrees of freedom. As far as possible, there should be about 12 degrees of freedom for error. The reason is not far to seek. The error mean square sits in the denominator of the test statistic to be used for testing the null hypothesis. If one looks at the tables of Snedecor's F, the value below 12 degrees of freedom is very high and very variable. So small variations in treatment effects will not be detected significant for smaller degrees of freedom for error. On the other hand, the table values of Snedecor's F stabilize after 12 degrees of freedom. So in order to be able to capture small variations, the error degrees of freedom should be at least 12. On the other hand, the error degrees of freedom should not be unduly large. It would be wastage of resources to spend large

degrees of freedom for estimating experimental error. It may be seen that if there are more number of treatments then lesser number of replications are required to ensure same number of error degrees of freedom.

ii) Availability of resources and precision required: Number of replications should be determined in such a way that the experiment can be conducted with the available resources namely labour, cost, time, experimental material, etc. and it should be able to achieve desired precision of comparisons among treatments. The smaller the differences that are desired to be detected between treatment means or effects, the more is the number of replications needed. Sometimes it may not be possible to obtain desired precision with available resources and there may be a need of a trade-off between available resources and desired precision level either by sacrificing precision or by increasing available resources.

iii) Type of experimental material: Generally homogenous experimental units require less number of replications and heterogeneous experimental units require more number of replications of the treatments.

iv) Manageability of the experiment: It should also be kept in mind that the experimenter should be able to manage to conduct the experiment well. This entails that number of treatments, their replications and number of experimental units should not be very large, otherwise it may lead to a poorly managed experiment.

We describe below a method due to Cochran and Cox (1964) to obtain number of replications of treatments. In conducting an experiment, the experimenter may be interested to detect difference of at least, say d, between two treatment effects. Let the two treatment means be $\bar{t}_1$ and $\bar{t}_2$. The significance of difference between two treatment effects is tested using Student's t statistic given by

$$t = \frac{|\bar{t}_1 - \bar{t}_2|}{\sqrt{2s_e^2 / r}}$$

where s_e^2 is the measure of error variation and r is the number of replications for both the treatments. If the experimenter wants to detect a difference of at least d between the two treatment effects, then the t-statistic should come significant at desired level of significance α and the corresponding t-statistic would be given by

$$t_\alpha = \frac{|d|}{\sqrt{2s_e^2 / r}}$$

where t_α denotes the critical value of t_α distribution at level of significance α. From the above equation one can get the number of replications as

$$r \leq \frac{2s_e^2 t_\alpha^2}{d^2}.$$