DESIGN OF EXPERIMENTS (STAT-564,M.Sc.(Agril.Stat) II Semester 2019-20

Design of experiment means how to design an experiment in the sense that how the observations or measurements should be obtained to answer a query in a valid, efficient and economical way. The designing of the experiment and the analysis of obtained data are inseparable. If the experiment is designed properly keeping in mind the question, then the data generated is valid and proper analysis of data provides the valid statistical inferences. If the experiment is not well designed, the validity of the statistical inferences is questionable and may be invalid. It is important to understand first the basic terminologies used in the experimental design.

EXPERIMENTAL UNIT:

For conducting an experiment, the experimental material is divided into smaller parts and each part is referred to as an experimental unit. The experimental unit is randomly assigned to treatment is the experimental unit. The phrase “randomly assigned” is very important in this definition.

EXPERIMENT:

A way of getting an answer to a question which the experimenter wants to know.

TREATMENT

Different objects or procedures which are to be compared in an experiment are called treatments.

SAMPLING UNIT:

The object that is measured in an experiment is called the sampling unit. This may be different from the experimental unit.

FACTOR

A factor is a variable defining a categorization. A factor can be fixed or random in nature. A factor is termed as a fixed factor if all the levels of interest are included in the experiment. A factor is termed as a random factor if all the levels of interest are not included in the experiment and those that are can be considered to be randomly chosen from all the levels of interest

REPLICATION:

It is the repetition of the experimental situation by replicating the experimental unit.

EXPERIMENTAL ERROR:

The unexplained random part of the variation in any experiment is termed as experimental error. An estimate of experimental error can be obtained by replication.

TREATMENT DESIGN

A treatment design is the manner in which the levels of treatments are arranged in an experiment. Suppose some varieties of fish food is to be investigated on some species of fishes. The food is placed in the water tanks containing the fishes. The response is the increase in the weight of fish. The experimental unit is the tank, as the treatment is applied to the tank, not to the fish. Note that if the experimenter had taken the fish in hand and placed the food in the mouth of fish, then the fish would have been the experimental unit as long as each of the fish got an independent scoop of food.

DESIGN OF EXPERIMENT:

One of the main objectives of designing an experiment is how to verify the hypothesis in an efficient and economical way. In the contest of the null hypothesis of equality of several means of normal populations having the same variances, the analysis of variance technique can be used. Note that such techniques are based on certain statistical assumptions. If these assumptions are violated, the outcome of the test of a hypothesis then may also be faulty and the analysis of data may be meaningless. So the main question is how to obtain the data such that the assumptions are met and the data is readily available for the application of tools like analysis of variance. The designing of such a mechanism to obtain such data is achieved by the design of the experiment. After obtaining the sufficient experimental unit, the treatments are allocated to the experimental units in a random fashion. Design of experiment provides a method by which the treatments are placed at random on the experimental units in such a way that the responses are estimated with the utmost precision possible.

PRINCIPLES OF EXPERIMENTAL DESIGN:

There are three basic principles of design which were developed by Sir Ronald A. Fisher. (i) Randomization (ii) Replication (iii) Local control

(i) Randomization

The principle of randomization involves the allocation of treatment to experimental units at random to avoid any bias in the experiment resulting from the influence of some extraneous unknown factor that may affect the

experiment. In the development of analysis of variance, we assume that the errors are random and independent. In turn, the observations also become random. The principle of randomization ensures this. The random assignment of experimental units to treatments results in the following outcomes. a) It eliminates systematic bias. b) It is needed to obtain a representative sample from the population. c) It helps in distributing the unknown variation due to confounded variables throughout the experiment and breaks the confounding influence. Randomization forms a basis of a valid experiment but replication is also needed for the validity of the experiment. If the randomization process is such that every experimental unit has an equal chance of receiving each treatment, it is called complete randomization.

(ii) Replication:

In the replication principle, any treatment is repeated a number of times to obtain a valid and more reliable estimate than which is possible with one observation only. Replication provides an efficient way of increasing the precision of an experiment. The precision increases with the increase in the number of observations.

(iii) Local control (error control)

The replication is used with local control to reduce the experimental error. For example, if the experimental units are divided into different groups such that they are homogeneous within the blocks, then the variation among the blocks is eliminated and ideally, the error component will contain the variation due to the treatments only. This will, in turn, increase the efficiency.

Complete and incomplete block designs:

In most of the experiments, the available experimental units are grouped into blocks having more or less identical characteristics to remove the blocking effect from the experimental error. Such design is termed as block designs. The number of experimental units in a block is called the block size. If size of block = number of treatments and each treatment in each block is randomly allocated, then it is a full replication and the design is called a complete block design. In case, the number of treatments is so large that a full replication in each block makes it too heterogeneous with respect to the characteristic under study, then smaller but homogeneous blocks can be used. In such a case, the blocks do not contain a full replicate of the treatments. Experimental

designs with blocks containing an incomplete replication of the treatments are called incomplete block designs.

Completely randomized design (CRD)

The CRD is the simplest design. Suppose there are v treatments to be compared. All experimental units are considered the same and no division or grouping among them exist. In CRD, the v treatments are allocated randomly to the whole set of experimental units, without making any effort to group the experimental units in any way for more homogeneity. Design is entirely flexible in the sense that any number of treatments or replications may be used. The number of replications for different treatments need not be equal and may vary from treatment to treatment depending on the knowledge (if any) on the variability of the observations on individual treatments as well as on the accuracy required for the estimate of individual treatment effect. Example: Suppose there are 4 treatments and 20 experimental units, then - the treatment 1 is replicated, say 3 times and is given to 3 experimental units, - the treatment 2 is replicated, say 5 times and is given to 5 experimental units, - the treatment 3 is replicated, say 6 times and is given to 6 experimental units and - finally, the treatment 4 is replicated [20-(6+5+3)=]6 times and is given to the remaining 6 experimental units.

All the variability among the experimental units goes into experimented error. CRD is used when the experimental material is homogeneous. CRD is often inefficient. CRD is more useful when the experiments are conducted inside the lab. CRD is well suited for the small number of treatments and for the homogeneous experimental material.

Layout of CRD

Following steps are needed to design a CRD: Divide the entire experimental material or area into a number of experimental units, say n.¬ Fix the number of replications for different treatments in advance (for given total number of¬ available experimental units). No local control measure is provided as such except that the error variance can be reduced by¬ choosing a homogeneous set of experimental units.

The statistical model for CRD with one observation per unit $Yij = \mu + ti + eij$
μ = overall mean effect
ti = true effect of the ith treatment

eij = true effect of the jth unit receiving ith treatment

The ANOVA table of CRD can be displayed as Follows:

SV	d.f	S.S	M.S	Fcal	Ftab
Treatments	t-1	SST	MST	MST/MSE	
Error	subtraction	SSE	MSE		
Total	rt-1				

Randomized Block Design

If a large number of treatments are to be compared, then a large number of experimental units are required. This will increase the variation among the responses and CRD may not be appropriate to use. In such a case when the experimental material is not homogeneous and there are v treatments to be compared, then it may be possible to group the experimental material into blocks of sizes v units. Blocks are constructed such that the experimental units within a block are relatively homogeneous and resemble to each other more closely than the units in the different blocks. If there are b such blocks, we say that the blocks are at b levels. Similarly, if there are v treatments, we say that the treatments are at v levels. The responses from the b levels of blocks and v levels of treatments can be arranged in a two-way layout. The observed data set is arranged as follows:

Layout: A two-way layout is called a randomized block design (RBD) or a randomized complete block design (RCB) if, within each block, the v treatments are randomly assigned to v experimental units such that each of the v! ways of assigning the treatments to the units has the same probability of being adopted in the experiment and the assignment in different blocks are statistically independent. The RBD utilizes the principles of design - randomization, replication and local control .

Latin Square Design

The treatments in the RBD are randomly assigned to b blocks such that each treatment must occur in each block rather than assigning them at random over the entire set of experimental units as in the CRD. There are only two factors – block and treatment effects – which are taken into account and the

total number of experimental units needed for complete replication are bv where b and v are the numbers of blocks and treatments respectively. If there are three factors and suppose there are b, v and k levels of each factor, then the total number of experimental units needed for a complete replication are bvk. This increases the cost of experimentation and the required number of experimental units over RBD. In Latin square design (LSD), the experimental material is divided into rows and columns, each having the same number of experimental units which is equal to the number of treatments. The treatments are allocated to the rows and the columns such that each treatment occurs once and only once in each row and in each column. In order to allocate the treatment to the experimental units in rows and columns, we take help from Latin squares. Latin Square: A Latin square of order p is an arrangement of p symbols in 2 p cells arranged in p rows and p columns such that each symbol occurs once and only once in each row and in each column. For example, to write a Latin square of order 4, choose four symbols – A, B, C and D. These letters are Latin letters which are used as symbols. Write them in a way such that each of the letters out of A, B, C and D occurs once and only once in each row and each column. For example,

as A B C D
B C D A
C D A B
D A B C

This is a Latin square.

Factorial Experiments

Factorial experiments involve simultaneously more than one factor and each factor is at two or more levels. Several factors affect simultaneously the characteristic under study in factorial experiments and the experimenter is interested in the main effects and the interaction effects among different factors. First, we consider an example to understand the utility of factorial experiments. Example: Suppose the yield from different plots in an agricultural experiment depends upon (i) variety of crop and (ii) type of fertilizer. Both the factors are in the control of the experimenter. (iii) Soil fertility. This factor is not in the control of the experimenter. In order to compare different crop varieties - assign it to different plots keeping other factors like irrigation, fertilizer, etc. fixed and the same for all the plots. - The conclusions for this will be valid only for the crops grown under similar conditions with respect to the factors like fertilizer, irrigation etc. In order to

compare different fertilizers (or different dosage of fertilizers) - sow single crop on all the plots and vary the quantity of fertilizer from plot to plot. - The conclusions will become invalid if different varieties of the crop are sown. - It is quite possible that one variety may respond differently than another to a particular type of fertilizer. Suppose we wish to compare - two crop varieties – a and b, keeping the fertilizer fixed and - three varieties of fertilizers – A, B and C. This can be accomplished with two randomized block designs (RBD) by assigning the treatments at random to three plots in any block and two crop varieties at random.

If the number of levels for each factor is the same, we call it is a symmetrical factorial experiment. If the number of levels of each factor is not the same, then we call it as asymmetrical or mixed factorial experiment.

We consider only symmetrical factorial experiments.

Through the factorial experiments, we can study - the individual effect of each factor and - interaction effect.

Now we consider a 2^2 factorial experiment with an example and try to develop and understand the theory and notations through this example.

General notation for representing the factors is to use capital letters, e.g., A, B, C etc. and levels of a factor are represented in small letters. For example, if there are two levels of A, they are denoted as a_0 and a_1 . Similarly, the two levels of B are represented as b0 and b1 . An important point to remember is that the factorial experiments are conducted in the design of an experiment. For example, the factorial experiment is conducted as an CRD, RBD.or LSD.
Factorial experiments with factors at two levels (2^2 factorial experiment):

Suppose in an experiment, the different varieties (V0 and V1) and different doses of fertilizers (F0 and F1) is to be laid out, each factor consists of two levels. Then it is called 2X2 factorial experiment. This experiment has four treatment combinations which are as V0F0, V1F0,V0F1 and V1F1. The differential response of the Factor V in the presence of F0 which is equal to V1F0-V0F0. Similarly The differential response of the Factor V in the presence of F1 which is equal to V1F1-V0F1. These two effects are called simple effects. The mean of these two simple effects is called main effect of the factor V. In the same way we find the main effect of the factor F. The half of the difference between two simple effects from II to I will provide the

interaction effects of the factor VxF. Thus, the main effect and interaction effect along with ANOVA of the two factors V and F are given below:

Main effect of the factor V= $\frac{1}{2}$[(V-1)(F+1)

Main effect of the factor F= $\frac{1}{2}$[(V+1)(F-1)

Interaction effect of the factor VXF= $\frac{1}{2}$[(V-1)(F-1)

These effects can also be shown with the help of a following table:

Effects	V0F0	V1F0	V0F1	V1F1	Divisor
Mean	+	+	+	+	4
V	-	+	-	+	2
F	-	-	+	+	2
VXF	+	-	-	+	2

If this experiment is conducted in RCBD, then the ANOVA table can be displayed as Follows:

SV	d.f	S.S	M.S	Fcal	Ftab
Replication	r-1	SSR	MSR		
V	v-1	SSV	MSV	MSV/MSE	
F	f-1	SSF	MSF	MSF/MSE	
VXF	(v-1)(f-1)	SS(VF)	MS(VF)	MS(VF)/MSE	
Error	subtraction	SSE	MSE		
Total	rvf-1				

If calculated value of any effect is found to be larger than its tabulated value at 5% or 1% level of significance, the null hypothesis is rejected and thus, we conclude that the effects are found to be significant. The crical difference will be used to identify the significant and at par treatments as well as interactions.

This kind of exercise can be demonstrated for the number of increased factors and their increased number of levels.

Contrasts and Analysis of Variance

The main technique adopted for the analysis and interpretation of the data collected from an experiment is the analysis of variance technique that essentially consists of partitioning the total variation in an experiment into components ascribable to different sources of variation due to the controlled factors and error. Analysis of variance clearly indicates a difference among

the treatment means. The objective of an experiment is often much more specific than merely determining whether or not all of the treatments give rise to similar responses. For examples, a chemical experiment might be run primarily to determine whether or not the yield of the chemical process increases as the amount of the catalyst is increased. A medical experimenter might be concerned with the efficacy of each of several new drugs as compared to a standard drug. A nutrition experiment may be run to compare high fiber diets with low fiber diets. A plant breeder may be interested in comparing exotic collections with indigenous cultivars. An agronomist may be interested in comparing the effects of bio fertilisers and chemical fertilisers. An water technologist may be interested in studying the effect of nitrogen with Farm Yard Manure over the nitrogen levels without farm yard manure in presence of irrigation.

Contrast:
The linear combination of the treatment means or total whose sum of the coefficients is equal to zero. For instance T_1-$2T_2$+T_3 is a contrast.

Orthogonal contrast:
Two contrast C1 and C2 are said to orthogonal if the sum of their multiplication of the treatment means or total coefficients is equal to zero. For example A and B are two orthogonal contrast if A=1/2[(a-1)(b+1) and B=1/2[(a=1)(b-1).

Mutually orthogonal contrast: If in an experiment ,there are more than two orthogonal contrast exist, it is called mutually orthogonal contrast.For example in the factorial experiment of 2^2,all V,F and VXF are mutually orthogonal contrast.

Confounding:

If the number of factors or levels increase in a factorial experiment, then the number of treatment combinations increases rapidly. When the number of treatment combinations is large, then it may be difficult to get the blocks of sufficiently large size to accommodate all the treatment combinations. Under such situations, one may use either connected incomplete block designs, e.g., balanced incomplete block designs (BIBD) where all the main effects and interaction contrasts can be estimated or use unconnected designs where not all these contrasts can be estimated. Non-estimable contrasts are said to be confounded.

Types of Confounding:

(i) Complete Confounding:

If the same kind of interaction is confounded in all the replication, it is called complete confounding. For example in a 2^3 factorial experiment, the interaction ABC is confounded in all three replications which are given below:

Replication I

Block1	Block 2
(1)	a
ab	b
ac	c
bc	abc

Replication II

Block1	Block 2
a	ab
b	bc
c	ac
abc	(1)

Replication III

Block1	Block 2
(1)	a
ac	b
ab	c
bc	abc

The interaction ABC is confounded in all three replications.

(ii) Partial Confounding:

If the different kinds of interaction is confounded in different replication, it is called partial confounding.

Replication I

Block1	Block 2
(1)	a
ab	b
c	ac
abc	bc

Replication II

Block1	Block 2
ac	c
b	ab
(1)	a
abc	bc

Replication III

Block1	Block 2
ac	abc
c	bc
ab	a
b	(1)

The interactions AB, AC and BC are confounded in Replications I, II and III respectively.