Unit - II Design of Experiments

Definition $\frac{S_E^2}{S_F^2}$

The Design of Experiment may be defined as "the logical construction of the experiments in which the degree of uncertainty with which the inference is drawn may be well defined."

Phases of Experimental Design

* Experimental or planning phase
  * Statement of problem
  * choice of response or dependent variables
  * Selection of factors to be varied.
  * Choice of level of these factors
    (a) Quantitative or qualitative
    (b) Fixed or random.
  * How factor levels are to be combined?

* Design Phase
  * Number of observations to be

* Order of experimentation.

* Method of randomization to be used.

* Mathematical model to describe the experiment.

* Hypothesis to be tested.

Analysis phase

* Data collection and processing

* computation of test statistics

* Interpretation of results for the experimenter.

Terminology In Expermental Designs 5M

Experiment

An experiment is a device or a means of getting an answer to the problem under consideration. There are two categories:

* Absolute

* comparative

Absolute Experiments

(i) Absolute experiments consist in determining the absolute value of some characteristics like (i) obtaining the average intelligence quotient of a group of people.

(ii) findind the correlation coefficient between two variables in a bivariate distribution.

Comparative Experiments

Comparative Experiments are designed to compare the effect of two or more objects on same population characteristics. Ex:- comparison of different manures or fertilizers different kinds of varieties of a crop, different cultivation processes, different pieces of land in a field experiment, or different diets or medicines in a dietary or medical experiment respectively.

Treatments 5M

various objects of comparison in a comparative experiment are termed as treatments. Ex:- In field experimentation different fertilizers or different varieties of crop or different methods of cultivation are the treatments.

Experimental Unit 5M

The smallest divison of the experimental material to which we apply the treatments and on which we make observations on the variable under study, is termed as experimental unit. Ex:- In feild experiments the plot of 'land' is the experimental unit. In other experiments, unit may be a patient in a hospital, a lump of dough or a batch seeds.

Blocks 5M

In agricultural experiments, most of the times we divide the whole experimental unit (field) into relatively homogenous sub-groups or strata. These strata which are more uniform amongst themselves than the field as a whole, are known as blocks.

yield

The measurement of the variable under study on different experimental units (Ex:- plots, in field experiments) are termed as yields.

Experimental Error

In field experimentation, it is a common experience that the fertility gradient of the soil does not follow any systematic pattern but behaves in an erratic fashion. Experience tells us that even if the same treatment is used on all the plots, the yields would still vary due to the differences in soil fertility. [Such variation from plot to plot, which is due to random (or chance or non assignable) factors beyond human control, is spoken of as experimental error.] It may be pointed out that the term 'error' used here is not synonymous with 'mistake' but is a technical term which includes all types of extraneous variations due to.

(i) The inherent variability in the experimental material to which treatments are applied.

(ii) The lack of uniformity in the methodology of conducting the experiment or in other words failure to standardise the experimental technique and

(iii) lack of representativeness of the sample to the population under study

Uniformity Trials

Uniformity trials enable us to have an idea about the fertility variation of the field. By uniformity trial, we mean a trial in which the field (experimental material) is divided into small units (plots) and the same treatment is applied on each of the units and their yields are recorded.

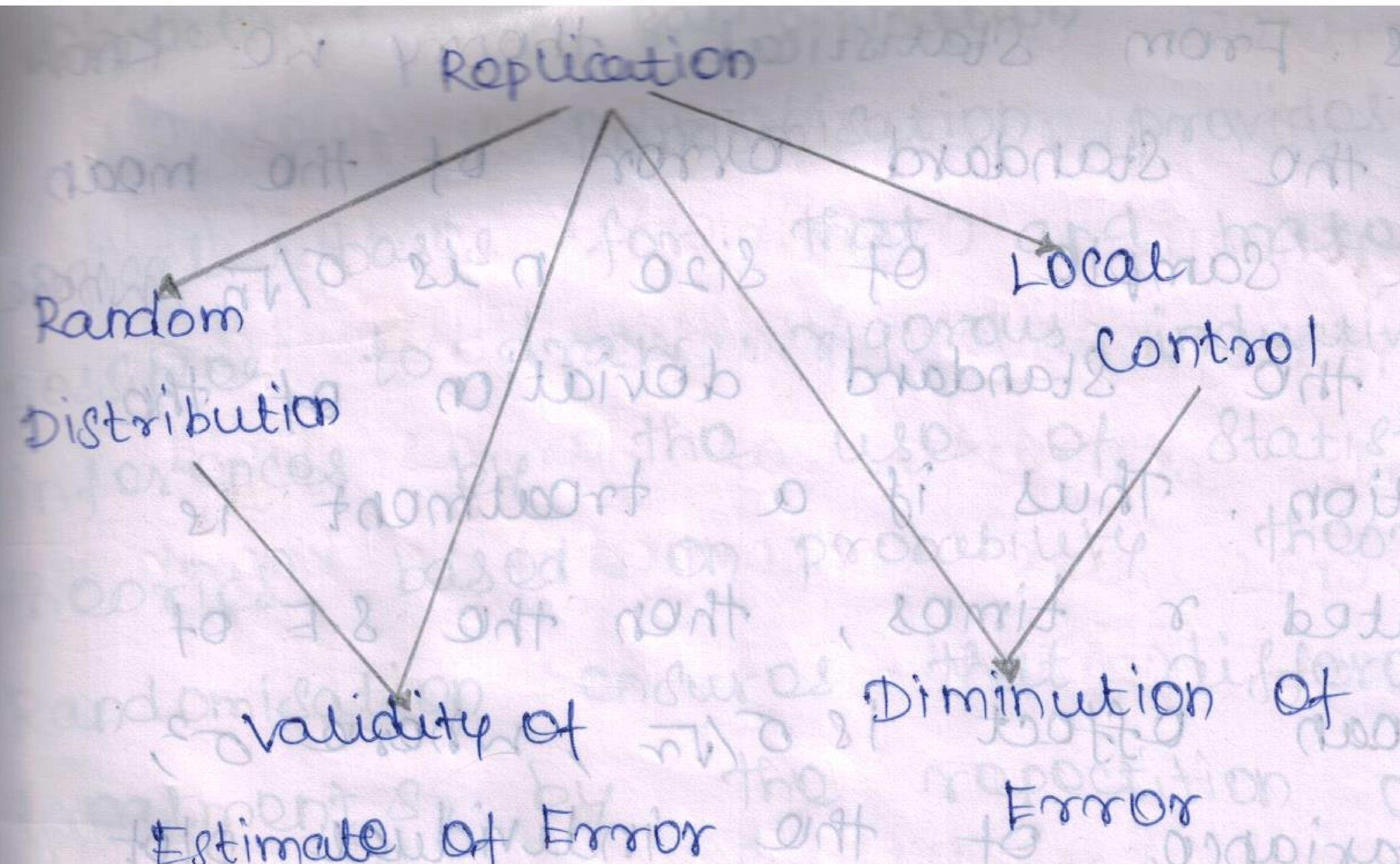
Principles of Experiments

For the validity of statistical analysis and enhancing the precision of the experiments, three basic principles

(i) replication

(ii) randomisation

(iii) Local control



Replication

Replication means the repetition of the treatments under investigation. An experimenter resorts to replication in order to average out the influence of the chance factors on different experiment units. Thus, the repetition of treatments results in more reliable estimate than is possible with a single observation.

Advantages of Replication

At the first instance replication serves to reduce experimental error and thus enables us to obtain more precise estimates of the treatment effects. From statistical theory we know that the standard error of the mean of a sample of size n is $\sigma/\sqrt{n}$, where σ is the standard deviation of the population. Thus if a treatment is replicated r times, then the S.E of its mean effect is $\sigma/\sqrt{r}$, where σ^2, the variance of the individual plot is estimated from the 'error variance' Thus "the precision of the experiment is inversely proportional to the square root of the replications."

Randomisation

In the absence of the prior knowledge of the variability of the experimental material, this objective is achived through 'randomisation', a process of assigning the treatments to various experimental unit in a purely chance manner.

Objectives of Randomisation

Randomisation provides a logical basis for that and makes it possible to draw rigorous inductive inferences by the use of statistical theories based on probability theory. Randomisation ensures that different treatments, by the repetition of the experiment, on the average are subject to equal environmental effect. Randomisation eliminates bias in any form.

Local control

In addition to the principles of replication and randomisation discussed earlier, the experimental error can further be reduced by making use of the fact that neighbouring areas in a field are relatively more homogeneous than those widely spread. In order to separate the soil fertility effects from the experimental error, the whole experimental area (field) is divided into homogeneous groups (blocks) row-wise or column-wise (one way elimination of fertility gradient c.f

The process of reducing the experiment-al error by dividing the relatively heterogeneous experimental area (field) into homogeneous blocks (due to physical contiguity as far as field experiments are concerned) is known as Local control.

Completely Randomised Design (CRD) 5M

[The simplest and most flexible design is the completely randomised design.]

[In this design the experimental units are allotted at random to the treatments, so that every unit gets the same chance of receiving every treatment. In addition the units should be processed in random order at all subsequent stages in the experiment where this order is likely to affect the results. Also in this design treatments are allocated at random to the experimental units over the entire experimental material.

* Let us suppose that we v treatments the i^{th} treatment being replicated r_i times, or $i = 1, 2, \ldots, v$. Then the whole

$n = \sum r_i$ experimental units and the treatment material is are distributed completely at random over the units subject to the condition that the i^{th} treatments occurs r_i times Randomisation assures the extraneous factors do not continually influence one treatment.

[In general, equal number of replications for each treatment should be made.] Also a table of random numbers is to be used to assign the units to the treatments.

Advantages 5M

(i) It is easy to layout the design

(ii) It results in the maximum use of the experimental units since all the experimental materials can be used.

(iii) It allows complete flexibility as any number of treatments and replicates may be used. The number of replicates, if desired, can be varied from treatment to treatment.

(iv) The statistical analysis is easy even if the number of replicates are not the same for all treatments or if the experimental errors differ from treatment to treatment.

(v) The relative loss of information due to missing data is smaller in comparison with any other design and they do not pose any problem in carrying out the standard analysis of data.

(vi) It provides the maximum number of degrees of freedom for the estimation of the error variance, which increases the sensitivity or the precision of the experiment for small experiments, (i.e) for experiments with small number of treatments.

Disadvantages 5M

(i) In certain circumstances, the design suffers from the disadvantages of being inherently less informative than other more sophisticated layouts.

This usually happens if the experimental material is not homogeneous.

(ii) Since randomisation is not restricted in any direction to ensure that the units receiving one treatment are similar to those receiving the other treatment, the whole variations among the experimental units is included in the residual variance. This makes the design less efficient and results in less sensitivity in detecting significant effects. ~~As such~~

(iii) As such CRD is seldom used in field experimentation, where due to the fertility gradient of the soil the whole experimental material, viz, field is not homogeneous and it is better to use more efficient designs like Randomised Block Design (R.B.D) or Latin Square Design (L.S.D) etc. discussed in § 6.6 and § 6.7, etc.

Applications

(i) complety randomised is most useful in laboratory technique and methodological studies, e.g in physics, chemistry or cookery, in chemical and biological experiments, in some green house studies, etc. where either the experimental material is homogeneous or the intrinsic variability between units can be reduced.

(ii) C.R.D is also recommended in situations where an appreciable fraction of units is likely to be destroyed or fail to respond.

Statistical Analysis of C.R.D

Statistical analysis of a C.R.D is analogous to the ANOVA for a one way classified data for fixed effect model, the linear (assuming various effects to be additive) becomes.

1. Null Hypothesis

Here we want to test the equality of the population means. Hence the Null Hypothesis is given,

$H_0 : \mu_1 = \mu_2 = \dots = \mu_v = \mu$

2. Alternative Hypothesis

$H_1 = \mu_1 \neq \mu_2 \neq \dots \neq \mu_v = \mu$

We know that $\bar{y}_i$ = mean of the i^{th} class

$$= \frac{\sum_{j=1}^{r_i} y_{ij}}{r_i} \quad (i = 1, 2, \dots v)$$

$$\bar{y}_{..} = \text{overall mean} = \frac{1}{n} \sum_{i=1}^{v} \sum_{j=1}^{r_i} y_{ij} = \frac{1}{n} \sum_{i=1}^{v} r_i \bar{y}_{i.}$$

Total Sum of Squares

$T.S.S = S.S.E + S.S.T$

$$S_T^2 = T.S.S = \sum_{i=1}^{v} \sum_{j=1}^{r_i} (y_{ij} - \bar{y}_{..})^2$$

$$= \sum_{i=1}^{v} \sum_{j=1}^{r_i} (y_{ij} - \bar{y}_{i.} + \bar{y}_{i.} - \bar{y}_{..})^2$$

$$= \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 + \sum_i r_i (\bar{y}_{i.} - \bar{y}_{..})^2 + 2\left[\sum_i \left\{ (\bar{y}_{i.} - \bar{y}_{..}) \sum (y_{ij} - \bar{y}_{i.}) \right\}\right]$$

$$= \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 + \sum r_i (\bar{y}_{i.} - \bar{y}_{..})^2 \quad \text{--- (3)}$$

$$\left[\because \sum_{j=1}^{r_i} (y_{ij} - \bar{y}_{i.}) = 0 \right]$$

From (3) we get get

$$S_E^2 = S.S.E = \sum_i \sum_j (y_{ij} - y_{i.})^2$$

Degrees of freedom

The degrees of freedom is for total sum of square (S_T^2) is $(n-1)$

The degrees of freedom for treatment sum of square is (S_t^2) is $(v-1)$

The D.F is error sum of square is (S_E^2) is $(n-v)$

Mean sum of square :-

The M.S.S due to treatments

$$= \frac{S_t^2}{v-1} = s_t^2$$

M.S.S. due to error $= \frac{S_E^2}{n-1} = s_E^2$

ANOVA Table for C.R.D

The ANOVA table is one way classified given below.

Source of variance	d.f	Sum of squares	Mean sum of squares	variance Ratio
Treatments	$v-1$	S_T^2	$s_t^2 = \frac{S_T^2}{(v-1)}$	$F_T = \frac{s_T^2}{s_E^2}$
Error	$n-v$	S_E^2	$s_E^2 = \frac{S_E^2}{(n-v)}$	
Total	$n-1$	$S_T^2 + S_E^2$		

Conclusion:

IF $F_T \le F_{(v-1),(n-v)}$; we accept Ho otherwise reject Ho