

Experimental Design and Its Application

Experimental Design is a vast concept. In this article, we will learn about:

1. The various terminologies required to build the concept (experimental units, treatments, experimental error, blocks).
2. The three principles of Experimental Design i.e. Randomization, Replication, and Local Control
3. The 4 main Experimental Designs i.e. CRD, RCBD, LSD and Factorial Designs.

Experiments answer questions. Experiments help us develop mathematical relationships between the various experimental units, minimize errors and draw various conclusions.

Experiments can be controlled. Having that control allows us to make stronger inferences about the nature of differences that we see in the experiment. Whereas, observational data are not changed or modified by any attempt on the part of an experimenter. So, it is often difficult to assign cause and effect by studying observational data which is easier in the case of experimental data.

An experimental design is defined as an effective procedure to plan experiments such that the data can be analyzed so as the results yield objective and valid conclusions.

In the words of Allen L. Edwards, "The Experimental design is called a randomized group design. The essential characteristic of this design is that subjects are randomly assigned to the experimental treatments or vice versa."

TERMINOLOGY

Treatments, units and assignment method specify the experimental design.

1. Experimental Units

Various objects of comparison in a comparative experiment. It is the factor about which the experimenter makes inferences.

Many experiments are conducted to establish the effect of one or more variables (independent) on a response (dependent variable). The independent variables are factors or treatments, such as different ways of packaging merchandise, different advertisement channels and so on. The values of a response are supposed to reflect the effects of different treatments.

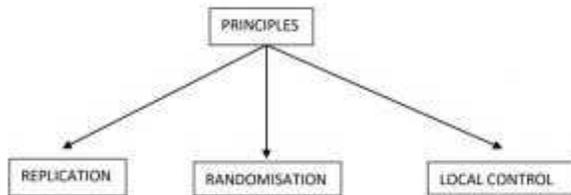
3. Experimental Error

It is the random variation present in all experimental results. Different experimental units give different responses to the same treatment. Also, if we apply the same treatment again and again to an experimental unit, we often get different results every time. Such variations are due to random or chance factors beyond human control.

4. Blocks

The term block comes from the agricultural heritage of experimental design where a large block of land was selected for the various treatments, that had uniform soil, drainage, sunlight, and other important physical characteristics. Homogeneous clusters improve the comparison of treatments by randomly allocating levels of the treatments within each block. These blocks are amongst themselves.

THE THREE PRINCIPLES



Replication: Although randomization helps to ensure that treatment groups are as similar as possible, the results of a single experiment, applied to a small number of objects or subjects, should not be accepted without question. Randomly selecting two individuals from a group of four and applying a treatment with “great success” generally will not impress the public or convince anyone of the effectiveness of the treatment.

To improve the significance of an experimental result, replication, the repetition of an experiment on a large group of subjects, is required. If a treatment is truly effective, the long-term averaging effect of replication will reflect its experimental worth.

If it is not effective, then the few members of the experimental population who may have reacted to the treatment will be negated by the large numbers of subjects who were unaffected by it. Replication reduces variability in experimental results, increasing their significance and the confidence level with which a researcher can draw conclusions about an experimental factor

Randomization: As it is generally extremely difficult for experimenters to eliminate bias using only their expert judgment, the use of randomization in experiments is common practice. In a randomized experimental design, objects or individuals are randomly assigned (by chance) to an experimental group. Using randomization is the most reliable method of creating homogeneous treatment groups, without involving any potential biases or judgments.

Local Control: The process of reducing the experimental 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. Local control also increases the efficiency of the design.

COMPLETELY RANDOMISED DESIGN or CRD

| | | | |
|---|---|---|---|
| A | B | D | A |
| C | D | B | C |
| B | A | D | C |

A completely randomized design (CRD) is one where the treatments are assigned completely at random so that each experimental unit has the same chance of receiving any one treatment. For the CRD, any difference among experimental units receiving the same treatment is considered as an experimental error.

Hence, CRD is appropriate only for experiments with homogeneous experimental units, such as laboratory experiments, where environmental effects are relatively easy to control. For field experiments, where there is generally large variation among experimental plots in such environmental factors as soil, the CRD is rarely used.

It is the simplest of all designs. The Local Control principle is missing in this design.

Layout

Let us suppose, we have g treatments to compare and N units to use in our experiment. For a completely randomized design:

1. Select sample sizes n_1, n_2, \dots, n_g with $n_1 + n_2 + \dots + n_g = N$.
2. Choose n_1 units at random to receive treatment 1, n_2 units at random from the $N - n_1$ remaining to receive treatment 2, and so on.

This is best suited for the experiments with a small number of treatments.

RANDOMIZED COMPLETE BLOCK DESIGN or RCBD

The randomized complete block design (RCBD) is one of the most widely used experimental designs in forestry research. The design is especially suited for field experiments where the number of treatments is not large and there exists a conspicuous factor based on which homogeneous sets of experimental units can be identified. The primary distinguishing feature of the RBD is the presence of blocks of equal size, each of which contains all the treatments.

In a block design, experimental subjects are first divided into homogeneous blocks before they are randomly assigned to a treatment group. If, for instance, an experimenter had reason to believe that age might be a significant factor in the effect of a given medication, he might choose to first divide the experimental subjects into age groups, such as under 30 years old, 30-60 years old, and over 60 years old.

Then, within each age level, individuals would be assigned to treatment groups using a completely randomized design. In a block design, both control and randomization are considered.

Layout

Let us consider five treatments A, B, C, D, and E each replicated four times. We divide the whole experimental area into four relatively homogeneous blocks and each block into five units. Treatments are then allocated at random to the plots of a block, fresh randomization being done for each block.

| | | | | | |
|-----------|---|---|---|---|---|
| Block I | A | E | B | D | C |
| Block II | E | D | C | B | A |
| Block III | C | B | A | E | D |
| Block IV | A | D | E | C | B |

In general, for v treatments each being replicated r times, we will have r blocks and each block will be divided into v units. So, we will have $rv=N$ experimental units in the field.

LATIN SQUARE DESIGN OR LSD

A Latin square is a table filled with n different symbols in such a way that each symbol occurs exactly once in each row and exactly once in each column. Here are two examples.

| | | |
|---|---|---|
| 1 | 2 | 3 |
| 2 | 3 | 1 |
| 3 | 1 | 2 |

| | | | |
|---|---|---|---|
| A | B | D | C |
| B | C | A | D |
| C | D | B | A |
| D | A | C | B |

A Latin square design is a method of placing treatments so that they appear in a balanced fashion within a square block or field. Treatments appear once in each row and column. Replicates are also included in this design.

- Treatments are assigned at random within rows and columns, with each treatment once per row and once per column.
- There are equal numbers of rows, columns, and treatments.
- Useful where the experimenter desires to control variation in two different directions.

- **Layout**

Suppose we have 4 treatments A, B, C and D, then it means that we have Number of treatments = Number of Rows = Number of Columns = 4
In general, if we have v treatments, then we will need the same number of rows and columns.

- There will be v^2 be experimental units in the field. The v treatments are then allocated at random to these rows and columns in such a way that every treatment occurs once and only once in each row and each column.
- For example, if we are interested in studying the m types of fertilizers on the yield of a certain variety of wheat, it is customary to conduct the experiments on a squared field with m^2 experimental units of equal area and to associate treatments with different fertilizers and row and column effects with variations in fertility of soil.

FACTORIAL DESIGN

Suppose an investigator is interested in examining three components of a weight loss intervention. The three components are

1.
 1. Keeping a food diary (yes or no) denoted by F;
 2. Increasing activity (yes or no) denoted by A;
1. Home visit (yes or no) denoted by H.

The investigator plans to manipulate each of these components experimentally. Thus, each becomes an independent variable.

The investigator plans to use a factorial experimental design. In factorial experiments, the effects of several factors of variation are studied and investigated simultaneously, the treatments being all the combinations of different factors under study. In these experiments, an attempt is made to estimate the effects of each of the factors and also the interaction effects, i.e., the variation in the effect of one factor as a result to different levels of other factors.

In the above example, each independent variable is a factor in the design. As there are three factors and each factor has two levels, this is a $2 \times 2 \times 2$, or 2^3 , factorial design. This design will have =8 different experimental conditions.

The notation used to denote factorial experiments conveys a lot of information. When a design is denoted by a 2^3 factorial, this identifies the number of factors (3); how many levels each factor has (2); and how many experimental conditions there are in the design ($2^3 = 8$).

Similarly, a 2^5 design has five factors with three levels, and $2^5 = 32$ experimental conditions; and a 3^2 design has two factors, each with three levels, and $(3^5 = 32)$ experimental conditions; and a degree has two factors each with a different number of levels. A 2^3 design has five factors-four with two levels and one with three levels and has $16 \times 3 = 48$ experimental conditions.

If the number of combinations in a full factorial design is too high to be logistically feasible, a fractional factorial design may be done, in which some of the possible combinations (usually at least half) are omitted.

Efficiency in Experimental design:

In order to compare two experimental designs, we require some kind of information for the basis of comparison. We compare the two designs by taking the amount of information with respect to designs.

We know that the amount of information in any design is reciprocal of mean square error. The relative efficiency of RCBD over CRD is equal to the ratio of amount of information in RCBD to the amount of information in CRD. i.e.

Relative efficiency of RCBD over CRD = $\frac{1}{E_1} / \frac{1}{E_2}$ where E_1 and E_2 are the error mean square for RCBD and CRD respectively.

If the relative efficiency is found to be more than 100 percent, then the excess is known as the gain in efficiency due to RCBD. When the error d.f. is less than 20, the relative efficiency has to be corrected by multiplying it by the precision of the factor. The precision factor can be taken as

$$\frac{(n_2 + 1)(n_1 + 3)}{(n_1 + 1)(n_2 + 3)}$$

where n_1 and n_2 are the d.f. for error in CRD and RCBD respectively.

But the error mean square in CRD can be determined using the formula as

$$\text{Error mean square in CRD} = \frac{nrE_r + (nt + ne)E_e}{nr + nt + ne}$$

where nr = d.f. for replication, E_r = replication mean square, nt = d.f. for treatment, ne = d.f. for error and E_e = error mean square

In order to have the efficiency of LSD over RCBD and CRD, we have to consider the types of blocks. If the LSD has been conducted as RCBD taking columns as blocks, then it is known as column blockings and if it has been as RCBD taking rows as blocks, then it is known as row blockings. Hence in case of column blockings, the estimate of error mean square for RCBD is given by

$$\text{EMS (RCBD)} = \frac{nrEr + (nt + ne)Ee}{nr + nt + ne}$$

Similarly, in case of row blocking, we have

EMS (RCBD) = where the symbols have their usual meanings.

The relative efficiency of LSD over RCBD is then given by

RE = [(Amount of information in LSD)/(Amount of information in RCBD)]X100 When the error d.f. is less than 20, the precision factor has to be considered. It is given by

$$\text{PF} = \frac{(ne^* + 1)(ne + 3)}{(ne + 1)(ne^* + 3)}$$

where ne^* is the error d.f for LSD and ne is d.f for RCBD.

Similarly, the relative efficiency of LSD over CRD is then given by

$$\text{EMS (CRD)} = \frac{nrEr + ncEc + (nt + ne)Ee}{nr + nc + nt + ne}$$

In the precision factor formula, the number of degrees of freedom for CRD will be

$$ne^* = nr + nc + ne$$

Analysis of Variance (ANOVA)

In some decision-making situations, the sample data may be divided into various groups i.e. the sample may be supposed to have consisted of k-sub samples. There are interest lies in examining whether the total sample can be considered as homogenous or there is some indication that sub-samples have been drawn from different populations. So, in these situations, we have to compare the mean values of various groups, with respect to one or more criteria.

The total variation present in a set of data may be partitioned into a number of non-overlapping components as per the nature of the classification. The systematic procedure to achieve this is called **Analysis of Variance (ANOVA)**. With the help of such a partitioning, some testing of hypothesis may be performed.

Initially, Analysis of Variance (ANOVA) had been employed only for the experimental data from the Randomized Designs but later they have been used for analyzing survey and secondary data from the Descriptive Research.

Analysis of Variance may also be visualized as a technique to examine a dependence relationship where the response (dependence) variable is metric (measured on interval or ratio scale) and the factors (independent variables) are categorical in nature with a number of categories more than two.

Example of ANOVA

Ventura is an FMCG company, selling a range of products. Its outlets have been spread over the entire state. For administrative and planning purpose, Ventura has sub-divided the state into four geographical-regions (Northern, Eastern, Western and Southern). Random sample data of sales collected from different outlets spread over the four geographical regions.

Variation, being a fundamental characteristics of data, would always be present. Here, the total variation in the sales may be measured by the squared sum of deviation from the mean sales. If we analyze the sources of variation in the sales, in this case, we may identify two sources:

- Sales within a region would differ and this would be true for all four regions (within-group variations)
- There might be impact of the regions and mean-sales of the four regions would not be all the same i.e. there might be variation among regions (between-group variations).

So, total variation present in the sample data may be partitioned into two components: between-regions and within-regions and their magnitudes may be compared to decide whether there is a substantial difference in the sales with respect to regions. If the two variations are in close agreement, then there is no reason to believe that sales are not same in all four regions and if not then it may be concluded that there exists a substantial difference between some or all the regions.

Here, it should be kept in mind that ANOVA is the partitioning of variation as per the assignable causes and random component and by this partitioning ANOVA technique may be used as a method for testing significance of difference among means (more than two).

Types of Analysis of Variance (ANOVA)

If the values of the response variable have been affected by only one factor (different categories of single factor), then there will be only one assignable reason by which data is sub-divided, then the corresponding analysis will be known as **One-Way Analysis of Variance**. The example (Ventura Sales) comes in this category. Other examples may be: examining the difference in analytical-aptitude among students of various subject-streams (like engineering graduates, management graduates, statistics graduates); impact of different modes of advertisements on brand-acceptance of consumer durables etc.

On the other hand, if we consider the effect of more than one assignable cause (different categories of multiple factors) on the response variable then the corresponding analysis is known as **N-Way ANOVA** ($N \geq 2$). In particular, if the impact of two factors (having multiple categories) been considered on the dependent (response) variable then that is known as **Two-Way ANOVA**. For example: in the Ventura Sales, if along with geographical-regions (Northern, Eastern, Western and Southern), one more factor 'type of outlet' (Rural and Urban) has been considered then the corresponding analysis will be Two-Way ANOVA. More examples: examining the difference in analytical-aptitude among students of various subject-streams and geographical locations; the impact of different modes of advertisements and occupations on brand-acceptance of consumer durables etc.

Two-Way ANOVA may be further classified into two categories:

- **Two-Way ANOVA with one observation per cell:** there will be only one observation in each cell (combination). Suppose, we have two factors A (having m categories) and B (having n categories), So, there will be $N = m \cdot n$ total observations with one observation (data-point) in each of $(A_i B_j)$ cell (combination), $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$. Here, the effect of the two factors may be examined.
- **Two-Way ANOVA with multiple observations per cell:** there will be multiple observations in each cell (combination). Here, along with the effect of two factors, their interaction effect may also be examined. Interaction effect occurs when the impact of one factor (assignable cause) depends on the category of other assignable cause (factor) and so on. For examining interaction-effect it is necessary that each cell (combination) should have more than one

observations so it may not be possible in the earlier Two-Way ANOVA with one observation per cell.

Conceptual Background

The fundamental concept behind the Analysis of Variance is “**Linear Model**”.

X_1, X_2, \dots, X_n are observable quantities. Here, all the values can be expressed as:

$$X_i = \mu_i + e_i$$

Where μ_i is the true value which is because of some assignable causes and e_i is the error term which is because of random causes. Here, it has been assumed that all error terms e_i are independent distributed normal variate with mean zero and common variance (σ^2).

Further, true value μ_i can be assumed to be consist of a linear function of t_1, t_2, \dots, t_k , known as “effects”.

If in a linear model, all effects t_j 's are unknown constants (parameters), then that linear model is known as “**fixed-effect model**”. Otherwise, if effects t_j 's are random variables then that model is known as “**random-effect model**”.

One-Way Analysis of Variance

We have n observations (X_{ij}), divided into k groups, A_1, A_2, \dots, A_k , with each group having n_j observations.

Here, the proposed fixed-effect linear model is:

$$X_{ij} = \mu_i + e_{ij}$$

Where μ_i is the mean of the i^{th} group.

General effect (grand mean): $\mu = \sum (n_i \cdot \mu_i) / n$

and additional effect of the i^{th} group over the general effect: $\alpha_i = \mu_i - \mu$.

So, the linear model becomes:

$$X_{ij} = \mu + \alpha_i + e_{ij}$$

$$\text{with } \sum_i (n_i \alpha_i) = 0$$

The least-square estimates of μ and α_i may be determined by minimizing error sum of square $(\sum_i \sum_j e_{ij}^2) = \sum_i \sum_j (X_{ij} - \mu - \alpha_i)^2$ as:

$X_{..}$ (combined mean of the sample) and $X_{i.}$ (mean of the i^{th} group in the sample).

So, the estimated linear model becomes:

$$X_{ij} = X_{..} + (X_{i.} - X_{..}) + (X_{ij} - X_{i.})$$

This can be further solved as:

$$\sum_i \sum_j (X_{ij} - X_{..})^2 = \sum_i n_i (X_{i.} - X_{..})^2 + \sum_i \sum_j (X_{ij} - X_{i.})^2$$

Total Sum of Square = Sum of square due to group-effect + Sum of square due to error

or

Total Sum of Square = Between Group Sum of square + Within Group Sum of square

$$TSS = SSB + SSE$$

Further, Mean Sum of Square may be given as:

$$MSB = SSB/(k-1) \text{ and } MSE = SSE/(n-k),$$

where $(k-1)$ is the degree of freedom (df) for SSB and $(n-k)$ is the df for SSE.

Here, it should be noted that SSB and SSE added up to TSS and the corresponding df's $(k-1)$ and $(n-k)$ add up to total df $(n-1)$ but MSB and MSE will not be added up to Total MS.

This by partitioning TSS and total df into two components, we may be able to test the

hypothesis:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

H_1 : Not all μ 's are same i.e. at least one μ is different from others.

or alternatively:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k = 0$$

H_1 : Not all μ 's are zero i.e. at least one μ is different from zero.

MSE has always been an unbiased estimate of σ^2 and if H_0 is true then MSB will also be an unbiased estimate of σ^2 .

Further MSB/σ^2 will follow Chi-square (χ^2) distribution with $(k-1)$ df and MSE/σ^2 will follow Chi-square (χ^2) distribution with $(n-k)$ df. These two χ^2 distributions are independent so the ratio of two Chi-square (χ^2) variate $F = MSB/MSE$ will follow variance-ratio distribution (F distribution) with $(k-1), (n-k)$ df.

Here, the test-statistic F is a right-tailed test (one-tailed Test). Accordingly, p-value may be estimated to decide about reject/not able to reject of the null hypothesis H_0 .

If H_0 is rejected i.e. all μ 's are not same then rejecting the null hypothesis does not inform which group-means are different from others, So, Post-Hoc Analysis is to be performed to identify which group-means are significantly different from others. Post Hoc Test is in the form of multiple comparison by testing equality of two group-means (two at a time) i.e. $H_0: \mu_p = \mu_q$ by using two-group independent samples test or by comparing the difference between sample means (two at a time) with the least significance difference (LSD)/critical difference (CD)

$$= t_{\text{error-df}} * MSE / (1/n_p + 1/n_q)^{1/2}$$

If observed difference between two means is greater than the LSD/CD then the corresponding Null hypothesis is rejected at alpha level of significance.

Assumptions for ANOVA

Though it has been discussed in the conceptual part just to reiterate it should be ensured that the following assumptions must be fulfilled:

1. The populations from where samples have been drawn should follow a normal distribution.
2. The samples have been selected randomly and independently.
3. Each group should have common variance i.e. should be homoscedastic i.e. the variability in the dependent variable values within different groups is equal.
4. Errors should follow a normal distribution with mean zero and variance σ^2 .
5. The effects are additive.

It should be noted that the Linear Model used in ANOVA is not affected by minor deviations in the assumptions especially if the sample is large.

ANOVA vs T-test

We employ two-independent sample T-test to examine whether there exists a significant difference in the means of two categories i.e. the two samples have come from the same or different populations. The extension to it may be applied to perform multiple T-tests (by taking two at a time) to examine the significance of the difference in the means of k-samples in place of ANOVA. If this is attempted, then the errors involved in the testing of hypothesis (type I and type II error) can't be estimated correctly and the value of type I error will be much more than alpha (significance level). So, in this situation, ANOVA is always preferred over multiple independent samples T-tests.

Two-Way Analysis of Variance (Two-Way ANOVA)

Here, the value of the dependent variable (response variable) may be impacted by two assignable causes (factors). For example: in the Ventura Sales along with geographical regions (Northern, Eastern, Western and Southern), termed as Factor "A", we want to

examine the impact of the type of outlet (Rural and Urban), termed as Factor “B”, on the mean-sales of the outlets.

Here, the proposed fixed-effect Linear Model is:

$$X_{ijk} = \mu_{ij} + e_{ijk}$$

Where, μ_{ij} is the true-value of the $(i, j)^{\text{th}}$ cell and e_{ijk} is the error term. Error term is assumed to be independently normally distributed with zero mean and common variance.

μ_{ij} is further decomposed as:

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$$

So, the fixed effect Linear Model becomes:

$$X_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk}$$

Where,

μ = Overall Mean Value

α_i = Effect of the Factor A_i

β_j = Effect of the Factor B_j

$\alpha\beta_{ij}$ = Interaction-Effect of the Factor $(A_i B_j)$

e_{ijk} = Error Term

The Least-square estimates may be determined by minimizing error sum of square $(\sum_{i,j,k} e_{ijk})^2$.

Accordingly, the Analysis of Variance is based on the following relation:

$$\sum_{i,j,k} (X_{ijk} - X_{...})^2 = np \sum_i (X_{i..} - X_{...})^2 + mp \sum_j (X_{.j.} - X_{...})^2 + p \sum_{i,j} (X_{ij.} - X_{i..} - X_{.j.} + X_{...})^2 + \sum_{i,j,k} (X_{ijk} - X_{ij.})^2$$

Total SS = SS due to Factor A + SS due to Factor B + SS due to Interaction of A & B + SS due to Error.

or

$$TSS = SSA + SSB + SS(AB) + SSE$$

By partitioning the variation into the above components, we are able to test following hypotheses:

$$H_{01}: \mu_1 = \mu_2 = \dots = \mu_m = 0 \text{ (No Effect of Factor A)}$$

$$H_{02}: \mu_1 = \mu_2 = \dots = \mu_n = 0 \text{ (No Effect of Factor B)}$$

$$H_{03}: \mu_{ij} = 0 \text{ for all } i \text{ and } j \text{ (Interaction-Effect is absent)}$$

Accordingly, the Mean Sum of Squares given by:

$$MSA = SSA / (m-1)$$

$$MSB = SSB / (n-1)$$

$$MS(AB) = SS(AB) / (m-1)(n-1)$$

$$MSE = SSE / mn(p-1)$$

and the Variance Ratios:

$$F_A = MSA / MSE \quad \sim \text{F distribution with } (m-1), mn(p-1) \text{ df}$$

$$F_B = MSB / MSE \quad \sim \text{F distribution with } (n-1), mn(p-1) \text{ df}$$

$$F_{AB} = MS(AB) / MSE \quad \sim \text{F distribution with } (m-1)(n-1), mn(p-1) \text{ df.}$$

Accordingly, p-values may be estimated to decide about reject/not able to reject the three null hypotheses H_{01} , H_{02} and H_{03} respectively.

