Chapter – One

Introduction

Contents

INTRODUCTION

Two unbiased estimators are sitting in a bar, having a few beers. The first one says, "How do you like being married?" The second one says, "It's okay, but you lose a degree of freedom!"

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. In research, a scientist identifies solution to problems through experimentation. The designing of 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.

Design of experiments (DOE) is defined as a branch of applied statistics that deals with planning, conducting, analyzing, and interpreting controlled tests to evaluate the factors that control the value of a parameter or group of parameters.

DOE is a powerful data collection and analysis tool that can be used in a variety of experimental situations. DOE is a systematic method to determine the relationship between factors affecting a process and the output of that process.

Experiments are performed by investigators in virtually all fields of inquiry, usually to discover something about a particular process or system.

Literally, an *experiment* is a test. More formally, we can define experiment as a test or series of tests in which purposeful changes are made to the input variables of the system so that we may observe and identify the reasons for the changes that may be observed in the output response.

In General experiments are used to study the performance of the process or systems, and more formally the process of experiment is represented by the diagram shown below. We usually visualize the process as a combination of machines, methods, people and other recourses that transform input variable into output (results).

Some of the input variables $x_1, x_1, \ldots, x_p$ are controllable, whereas other variables $z_1, z_2, \ldots, Z_q$ are uncontrollable (although they may be controllable for purposes of a test).

- **Controllable factors**, or x factors, are those input parameters that can be modified in an experiment or process.
- **Uncontrollable factors**, or z factors, are those input parameters that cannot be changed.

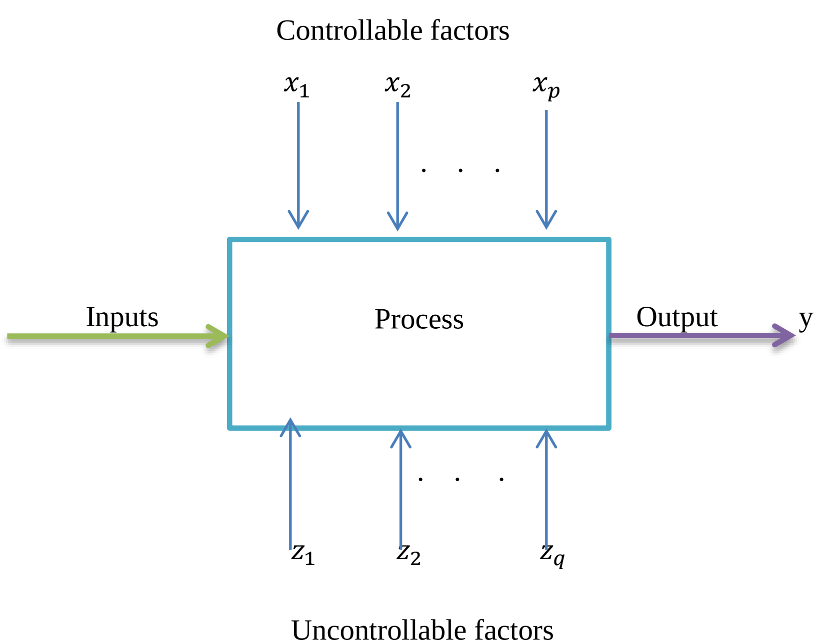

Figure1.1: General Model for experiment

Goals of experimentations are:

- ✓ To provide valid comparison of the effects of treatments.
- ✓ To provide valid information about the relationship between variables of interest.
- ✓ To isolate effects of each input variable.
- ✓ To determine effects of interactions.
- ✓ To determine magnitude of experimental error.

1.1 Strategy of Experimentation

Usually, an objective of the person conducting the experiment, called the **experimenter**, is to determine the influence that these factors have on the output response of the system.

The process of conducting experiment is called ***experimentation***.

The general approach of planning and conducting the experiment is known as ***strategy of experimentation*.**

Some of the strategies of experimentation are:

- ✓ ***Best-Guess Approach***: This strategy is used to select an arbitrary combination of factors. Based on the outcome of the current test, selection of combination of factors will be continued almost indefinitely.
    - Despite its applicability, it has two disadvantages
    - o If the first guess doesn't produce the desired result, we need to select another combination of factors and this may continue for a number of times without any guarantee of success.
    - o If the first Best - Guess produced the desired or acceptable result, then the researcher may stop experimentation though there is no guarantee that the *best* solution has been obtained.

      This strategy of experimentation is used most frequently by engineers and scientists.
- ✓ ***One-factor-at-a-time***: This strategy begins with selection of starting point, then successively varying each factor over its range with the other factors held constant at the base line level.

After all tests are performed, a series of graphs are constructed showing how the response variable is affected by varying each factor with all other factors held constant.

- The disadvantage of this strategy is that it doesn't consider the *interaction effect* or doesn't take into consideration the joint effect of two or more factors on the response variable.

✓ ***Factorial Experiment***: This is the experimental strategy in which factors are varied together, instead of one at a time and considers the interaction as well.

1.2. Some Typical Applications of Experimental Design

Design of experiment (DOE) is widely applicable in different scientific disciplines so as to alleviate the problems.

Some application of experimental design in any field of study includes:

- Evaluation and comparison of basic design configuration (Structure).
- Evaluation of material alternative.
- Selection of the design parameters (factors) …etc.

○ **Generally experimental designs are used:**

✓ To provide estimates of a treatment effects or differences among treatment effects.
✓ To provide an efficient way of testing hypothesis about the response to treatments.
✓ To assess the reliability of estimates and assumptions.
✓ To estimate the variability of the experimental material.
✓ To increase precision by eliminating extraneous external source of variation from the comparisons of interest.
✓ To provide a systematic, and efficient pattern of conducting an experiment.

1.3 Guidelines for Designing Experiments

Every statistical investigation has its own steps (stages) that a researcher must follow so as to reach at valid or sound conclusion. Despite the nature of a specified investigation, there are five stages in any statistical investigations: problem formulation, data collection, classification and organization, presentation, analyses and interpretation.

- ❖ **To use the statistical approach in designing and analyzing an experiment follows the following guidelines:**

1. **Recognition of and statement of the problem**: This is the first stage of DOE, in this stage we have to develop all ideas about the objective of the experiment. It is helpful to prepare a list of specific problems or questions that are to be addressed by the experiment. It is also important to keep the overall objectives in mind.
2. **Selection of the response variable**: are variable really provides useful information about the process under study of experiment. The *response* is the experimental outcome or observation.

3. Choice of factors, levels and ranges: when considering the factors that may influence in the Performance of a process or the system. Factors may be *quantitative* and *qualitative*.

The experimenter usually discovers that these factors can be classified as either ***potential design factors*** or ***nuisance factors***.

- ❖ ***Potential design factors:*** are those factors that the experimenter may wish to vary in the experiment. This can be classified in three some important factors:
- *Design factors:* are the factors actually selected for the study in the experiment.
- *Held-constant factors:* are variables that may exert some effect on response, they will be held at a specific level.
- *Allowed to vary factors:* are variables that vary unit to unit and rely randomization to balance out any material or experimental unit effect.
- o ***Nuisance factors***: are factors that may have large effects and but not be interested in them in the context of the present experiment. This factor classified as controllable and uncontrollable (noise factors**).**

  A controllable nuisance factor is one whose levels may be set by the experimenter.

4. **Choice of experimental design**: Choice of design involves the consideration of sample size (number of replicates), the selection of a suitable run order for the experimental trials, and the determination of whether or not blocking or other randomization restrictions are involved.
5. ***Perform the Experiment:*** To avoid confusion and to eliminate potential problems of running the wrong combination of factor levels in a multifactor, plan of the experiment

should be put on a separate piece of paper and given to the personnel performing the experiment.

When running the experiment, it is vital to monitor the process carefully to ensure that everything is being done according to the plan. Errors in the process at this stage usually destroy the experimental validity.

6. Statistical analysis of the data: The process of extracting relevant information from the summarized data. Statistical methods should be used to analyze the data.

Graphical methods, estimations and hypothesis testing are very useful in analyzing data from a designed experiment. It is helpful to present the results of many experiments in terms of an empirical model interpretation, residual analysis and model adequacy checking are also important analysis techniques.

7. Conclusion and recommendation: the experimenter must draw practical conclusions about results and recommend a course of action.

1.4 Basic Principles and Terminology

- There are three Fundamental / Basic principles of **experimental design** , those are :-
- ***Replication,***
- ***Randomization***
- ***Blocking.***

An *experimental unit* is a generic term that refers to a basic unit as material, animal, person and machine to which treatment is applied.

Experimental error is a measure of the variation, which exists among observations on experimental units treated alike.

i. Replication. A situation where a treatment appears more than once in an experiment, it is said to be replicated.

Generally ***Replication*** means repetition of a basic experiment without changing any factor settings or it refers to the numbers of experimental units that receive the same treatment.

It enables the estimation of the magnitude of experimental error (i.e., the error variance). Against which the differences among treatments are judged.

Increasing the number of replications, or replicates, decreases the variance for detecting differences in treatments.

❖ **The functions of Replication are:**

1. It provides an estimate of experimental error because it provides several observations on experimental units receiving the same treatment. For an experiment on which each treatment appears only once, no estimate of experimental error is possible and when there is no method of estimating the experimental error, there is no way to determine whether observed differences indicate the real differences or due to inherent variability.
2. It improves the precision of an experiment: As the number of replicates increase, the estimates of population means as observed treatment means become closer to the true value.
3. It increases the scope of inference and conclusion of the experiments: Field experiments are normally repeated over years and locations because conditions vary from year to year and location to location. The purpose of replication in space and time is to increase the scope of inference. The results of an experiment are applicable only to conditions that are similar to that condition.

Why **Replicate**? :- To reduce the effect of uncontrolled variation (i.e., increase precision) and to quantify uncertainty.

ii.Randomization. It should be applied to the allocation of units to treatments, the order in which the treatments are applied in performing the experiment and the order in which the responses are measured.

Randomization is the cornerstone underlying the use of statistical methods in experimental design.

By randomization we mean that both the allocation of the experimental material and the order in which the individual runs or trials of the experiment are to be performed are randomly determined.

Randomization is the design technique when the nuisance factor is known but uncontrollable.

Generally, we define a **nuisance factor** as a design factor that probably has an effect on the response, but we are not interested in that effect.

✓ Randomization reduced the unwanted influence of subjective judgment in treatment allocation.

- ✓ Assigning the treatments to the experimental units in such a way that any unit has equal chance to receive any treatment (at a random), i.e. every treatment should have an equal chance of being assigned to any experimental units.
- ✓ Moreover, randomization ensures *validity* of the estimate of the experiment error and provides a basis for inference in analyzing the experiments.

Purpose / Objective of randomization

- **To eliminate bias**: randomization ensures that no treatment is favored or discriminated against the systematic assignment to units in a design.
- **To ensure independence among the observations**. This is necessary to provide valid significance tests and interval estimates.

NB: Randomization is usually done by using tables of random numbers or by drawing cards, coins or lots.

iii.Blocking, Blocking is a design technique used to improve the precision with which comparisons among the factors of interest are made.

Often blocking is used to reduce or eliminate the variability transmitted from nuisance factors; that is, factors that may influence the experimental response but in which we are not directly interested.

When the nuisance source of variability is **known** and **controllable**, a design technique called **blocking** which can be used to systematically eliminate its effect on the statistical comparisons among treatments.

If **blocking** is effective, it should be applied to remove the block-to-block variation. Randomization can be applied to the assignments of treatments to units within the blocks. Generally, a block is a set of relatively homogeneous experimental conditions.

- ❖ **Some Basic term Definition:**

Response Variable: Variable measured in experiment (outcome, y).

Factors: Input variables that can be changed or independent variables.

Levels: Specific values of factors (inputs) or the value assumed by a factor in an experiment.

Interaction: Effect of one input factor depends on level of another input factor.

Confounding: A concept that basically means that multiple effects are tied together into one parent effect and cannot be separated.

Fixed Effects Model: If the treatment levels are specifically chosen by the experimenter, then conclusions reached will only apply to those levels.

Random Effects Model: If the treatment levels are randomly chosen from a population of many possible treatment levels, then conclusions reached can be extended to all treatment levels in the population.

Treatment: It is an amount of material or a method that is to be tested in the experiment such as Crop varieties, insecticides, feedstuffs, fertilizer rates, method of land preparation, irrigation frequency, etc.

Experimental unit: It is an object on which the treatment is applied to observe an effect, *e.g.* cows, plot of land, petri-dishes, pots, etc.

Experimental error: is the variation in the responses among experimental units (e.u.'s) which are assigned the same treatment, and are observed under the same experimental conditions.