Design of Experiments

Table of Contents

INTRODUCTION

Experiment design is the process of planning and carrying out an experiment to get a reliable, useful, and cost-effective answer to a question by measuring the results. If an experiment is well-designed, it will produce valid data, and if the data is analysed correctly, we can draw reliable statistical conclusions.

It is a systematic and efficient way for scientists and engineers to study the relationship between multiple input variables (also called "factors") and key output variables (also called "responses"). It is a structured approach for collecting data and making discoveries. It is a good way to plan experiments so that the data can be analysed in a way that leads to valid and objective conclusions. Well chosen experimental designs maximise the amount of "information" that can be obtained for a given amount of experimental effort.

A proper experimental design serves as a road map for the study methods, allowing readers to understand how the data was obtained more clearly and, as a result, assist them in properly analysing the results.

When to use Experimental Design?

1) To determine whether a factor, or a collection of factors, has an effect on the response.
2) To determine whether factors interact in their effect on the response.
3) To model the behaviour of the response as a function of the factors.
4) To optimise the response.

Principles of Experimental Design

Every experiment adheres to these three fundamental guidelines:

a) *Randomisation*

Randomization is the key to a good experiment. This means giving treatments to experimental units at random so that every possible treatment assignment has the same chance.

- Obtaining a demographic sample that is representative
- Dividing up the experimental units into the various treatments at random, eliminates systematic bias.
- Contributes to the distribution of the unobserved variance caused by variables influencing both the independent and dependent variables throughout the experiment. Consequently, the errors become random and independent, which renders the observations random as well.

b) *Replication*

The validity of the experiment can be checked by duplicating the experimental unit so that the same conditions can be repeated multiple times. This gives a more accurate estimate of the experimental error.

- Experimentation is repeated by subjecting the same group of organisms to the same treatment a predetermined number of times in order to acquire a more robust and trustworthy estimate than that obtained from a single observation.
- Increasing the number of observations boosts the experiment's precision effectively. Suppose, for instance, the variance of a random variable x is σ^2. The variance of the sample mean, $\bar{\mathbf{x}}$, based on n observations is therefore (σ^2/n). Consequently, the variance of $\bar{\mathbf{x}}$ diminishes as n grows.

c) *Local Control*

Local control entails the grouping of identical experimental units into groups or blocks with the elimination of variation within the blocks in an effort to minimise experimental error.

- Aids in a more precise and efficient experiment
- Handles the extraneous sources of variation that are a factor in experimental error but are beyond the control of randomization and replication.
- This means that the error component should ideally only include the variation that can be attributed to the treatments themselves, namely the variance between blocks.

Methods of Experimental Design

1) Completely Randomised Design

Treatments are assigned randomly to experimental units. This ensures that each experimental unit has an equal chance of receiving a treatment. Any variation between experimental units receiving the same treatment is considered an experimental error. Consequently, this design only applies to experiments with homogeneous experimental units, such as laboratory experiments, where it is assumed that there are no uncontrolled variables.

<u>Advantages</u>

1) Simple layout
2) Complete flexibility. Any number of treatments and replications for each treatment can be tried.
3) The design provides the greatest number of degrees of freedom for experimental error.

<u>Disadvantages</u>

1) Rarely suitable for experiments because it is extremely challenging to find identical experimental scenarios. This design does not use the Principle of Local Control.
2) Comparatively less accurate than other designs

<u>Example</u>

Let's say a company that makes car wax has three different formulas and wants to find the best one. In order to test the quality of these waxes (Wax 1, Wax 2 and Wax 3), each wax is applied to a sample of 10 cars of the same model (total 30 cars). Now, each of these cars is subjected to multiple tests such as repeated car washes, scratch test, dust test, etc.

In this scenario, randomization was used to assign a wax to a car. Each car had an equal chance of getting one of the three waxes. Replication was employed by using cars of the same model to maintain homogeneity. The results of the tests would be analyzed collectively for 10 cars based on the wax applied. This highlights the use of local control as analyzing collective results can help in some elimination of variation from the results.

2) Randomised Block Design

To eliminate the effects of a few of the most important extraneous / nuisance factors in this situation, the idea of 'Blocking' is applied. The fundamental idea is to construct homogeneous blocks such that the extraneous factors are constant in them while the factor of interest is allowed to vary.

Advantages

1) Effectively handles non-homogeneous experimental material.
2) It is adaptable enough to support countless treatments, blocks, and replications.
3) The sample sizes for the various treatments do not have to be equal.
4) Smaller error variance as the Local Control principle makes that sure because of the homogeneous blocks and because of parting away some variance from the error variance due to the difference among blocks. Thus, this dominates over the Complete Block Design which has high experimental error due to high variability among experimental units.
5) Relatively easy statistical analysis even with the missing data.
6) If an entire treatment or a block needs to be dropped from the analysis for some reason, such as spoiled results, the analysis is not complicated thereby.

Disadvantages

1) Not suitable for a large number of treatments because the block size becomes too large. Because the prima facie idea of Randomised Block Design is based on the fact of reducing the variability within blocks, but with the increase of block size, we deviate from our basic setup.
2) It requires some strong assumptions more than that for a completely randomised design - like no interactions between treatments and blocks and constant variance from block to block. So, interactions between block and treatment effects increase error.

3) Generalised Randomised Block Design

When the nature of the interactions between blocks and treatments is of interest, more than one replicate is required in each treatment within a block. This design is called a Generalised Randomised Block Design.
Each treatment only occurs once in each block in RBD, making it impossible to test for a treatment-by-block interaction. However, GRBD (Generalised Randomised Block Design) permits replications of each treatment level within a block. The two factors, treatment, and block are also interchangeable in GRBD.
Smaller groups or blocks of experimental units often provide better homogeneity when they represent physical things. Because of this, we do not advocate employing a block design with more experimental units per block than the minimal x, where x is the number of levels of the treatment factor. However, in situations where the experimental runs represent trials rather than actual physical objects, greater block sizes may not always result in an increase in the variability of experimental units within a block, allowing for the speedy creation of experimental runs.

<u>Advantages</u>

1) When compared to RCBDs(Randomised complete block design) with more blocks, GRBD model designs provide more degrees of freedom for investigating treatment-effects.
2) Replication of the treatment inside each block, which enables the estimate and testing of an interaction term within a linear model without assuming any mistake.
3) Instead of using an RCBD with additional blocks while doing an experiment, if the experimenter wants to boost power, they may use a GRBD instead.

<u>Disadvantages</u>

1) Because there are numerous units in each block and each treatment must be administered to a number of units in each block, this process is quite tedious.
2) Because replications are used, experiment costs are higher.

4) Optimal Design

An optimal design optimises a numerical criterion that is usually related to variability or other statistical design properties and uses as input the number of runs, their causes and possible levels, the structure of the block (if any), and the assumed form of the relationship between the answer and factors.

Advantages

1) lowers the cost of experiments by figuring out the needed statistical model with fewer trials than before.
2) It can handle many different kinds of factors, like processes, mixtures, and discrete factors.
3) It can take care of treatments, i.e., factor levels that are continuous rather than discrete.
4) The design can be made better when the design space is limited, such as when the mathematical process space has settings of factors that are not possible in real life.

Disadvantages

Complexity is high because deciding on a good model and its corresponding criterion function requires a good understanding of both statistical theory and designing experiments in the real world.

5) Bayesian Experimental Design

Bayesian experimental design is based on Bayesian inference to interpret the observations and data collected during the experiment, which uses Bayes' theorem to update the probability for a hypothesis as new data or information becomes available. This makes it possible to take into account both any prior knowledge about the parameters to be determined and observational uncertainty.

The aim is to design an experiment so that the expected utility of outcome is maximised. The utility criterion chosen will decide the optimal experiment to conduct.

In simple terms, the Bayesian approach says that we can't be sure about what we believe, so we try to figure out the probability distribution of our belief/parameter. We try to estimate the probability distribution of the belief by taking into account what we already know and what we have seen. Simply put, we don't look at a single point estimate. Instead, we look at the probability distribution of the hypothesis given the data. The name for this is the "posterior probability."

Steps taken are:

1. Collect information about how the prior distribution looks like (not a necessity; we can use uninformed priors)
2. Deciding the decision boundary (for e.g. we want to 90% sure that version A is better than B)
3. Collecting the data
4. Calculating the posterior distribution for conversion rates of both the variants
5. Inference

Advantages

1) It lets us utilise prior information.
2) Utility functions can be tailored to specific experiments.
3) Which experiment design will inform the most about the model can be predicted before experiments in a laboratory are conducted.
4) It provides a convenient setting for a wide range of models, such as hierarchical models and missing data problems.
5) It provides interpretable answers, such as “the true parameter has a probability of 0.95 of falling in a 95% credible interval.”
6) It obeys the likelihood principle.
7) BED is also useful when the model raises specific physical questions that are answered by some experiments but not by others.

Disadvantages

1) This design can’t be used in real problems without using software as it involves various integrations and numerical optimization.
2) Choosing the prior requires a lot of skill, expertise and experience.
3) Posterior distributions can be heavily influenced by the priors.

4) It doesn't tell us how to select the prior. There is no correct way to choose.
5) High computation cost for models with large number of parameters

Example

We have a population of trees in a forest, from past experience it is known that 40% of trees are eucalyptus. A person goes into the forest to test this hypothesis and collects data for 300 trees, out of which he found only 20 trees of eucalyptus so p^ is just 0.067. In general following the frequentist approach statistician would say we reject the hypothesis, but the Bayesian approach would suggest, taking into consideration that trees may be in clusters and not spread homogeneously, that our result may not be correct, he would recommend a probability distribution of the population parameter with mean of 0.4.

6) Quasi-Experimental Design

A quasi-experiment is an empirical interventional research design in which randomization is not used to determine the causal effects of an intervention on a target population. The absence of randomization renders the experiment quasi-natural. This distinguishes between quasi-experimental designs and randomised experimental designs. Due to this, the internal validity, that is, the confidence you have that the causality is not due to outside factors, is not quite as strong as in an actual experiment as the confounding variables are inherent in the preselected elements.

Advantages

1) High external validity
2) High control over targeted hypotheses
3) Ability to be combined with other methodologies

Disadvantages

1) Low internal validity
2) Risk of inaccurate data
3) Risk of bias

Example

Consider a psychological health survey that you wish to conduct among mentally unstable patients for research of a drug that you are developing. But as per the ethics of the organisation, you cannot directly convey a survey among the patients about the treatment. So in such a scenario, we conduct a quasi experimental design approach. In this approach, we look at already collected data from the psychologists and do our survey according to that.

ANOVA Procedure for Randomised Block Design

Consider the following case where we apply ANOVA to an experiment conducted in the Randomised block design setting.

In this case, some students were first segmented into six categories based on their IQ, and each of these groups was further divided into three more groups corresponding to three different types of teaching methods. These students then were evaluated and the average scores of the tests corresponding to the IQ range and teaching method have been given to us in a table as shown below. Our aim is to check if there is a difference in the effectiveness of the method of teaching. In this test, we would take the alpha as 0.05.

	Teaching Method		
IQ	A	B	C
91-95	84	85	85
96-100	86	86	88
101-105	86	87	88
106-110	89	88	89
111-115	88	89	89
116-120	91	90	91

Our analysis starts by declaring the null hypothesis and the negation of the null hypothesis.

$$H0 = \mu 1 = \mu 2 = \mu 3$$

H1 = All means are not equal

Factors of Interest- Teaching methods (3)
Blocks- IQ groups (Six)

Observations are obtained from K independent groups. Within the i th group, there are ni observations of a single variable Y. We use Yij to represent the value of Y from the j th individual of the i th group.

For this, we are going to use the F-test and the test statistic is as follows.

$$F_{K-1,N-K} = \frac{MSB}{MSE}$$

Here MSB and MSE are given by the following equations.

$$MSB = \frac{1}{K-1}\sum_{i=1}^{K} n_i(\bar{Y}_i - \bar{Y})^2$$

$$MSE = \frac{1}{N-K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}_i\right)^2$$

The quantities MSB and MSE are often called the mean square between and mean square error, respectively.

There are other two types of mean which would use here, one is the overall mean, and the other is the block mean, and they are given by the following formulae.

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{K}\sum_{j=1}^{n_i} Y_{ij} \qquad \bar{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$$

The one on the left is the overall mean and the one on the right is the block mean.

The block means can be calculated very easily using a spreadsheet and the means are as follows.

Groups	Count	Sum	Average
Row 1	3	254	84.66666667
Row 2	3	260	86.66666667
Row 3	3	261	87
Row 4	3	266	88.66666667
Row 5	3	266	88.66666667
Row 6	3	272	90.66666667

Now we can calculate MSD and it is as following:
MSD = ⅕ * (3*9.3228 +3* 1.1095 + 3*0.5184 + 3*0.8961 + 3*0.8961 + 3*8.682)

MSD = 12.8555

MSE for this data can be also calculated, and I am directly writing down the calculated value below.
MSE = 0.6111

Hence, the F value would be;

F = 12.855/0.6111 = 21.0363

The corresponding p-value is 0.00001473449482, and it is less than 0.05.
Thus, we fail to reject H1.
Thus, there is no significant difference in the effectiveness of the three teaching methods.

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*

Between Groups	64.27777778	5	12.85555556	21.03636364	0.00001473449482	3.105875236
Within Groups	7.333333333	12	0.6111111111			
Total	71.61111111	17				

References

1. https://en.wikipedia.org/wiki/Bayesian_experimental_design
2. https://staffblogs.le.ac.uk/bayeswithstata/2015/08/07/bayesian-experimental-design-part-ii/
3. https://staffblogs.le.ac.uk/bayeswithstata/2015/07/31/bayesian-experimental-design/
4. https://www.sciencedirect.com/science/article/abs/pii/S0894177711001944
5. https://towardsdatascience.com/exploring-experimentation-in-bayesian-territory-8e2c40df77bd
6. https://cxl.com/blog/bayesian-frequentist-ab-testing/