Chapter 1

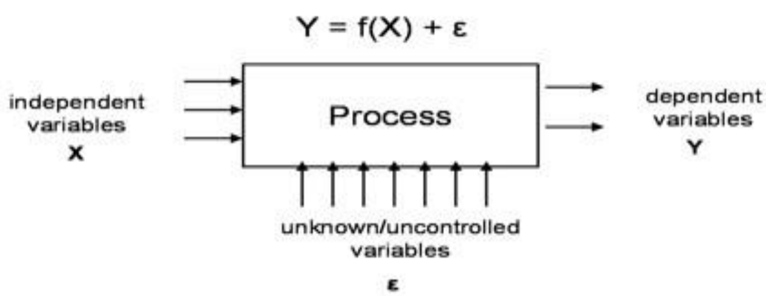
Introduction

Experimental Design

Design of Experiments (DOE) is also referred to as Designed Experiments or Experimental Design - all of the terms have the same meaning.

Data for statistical studies are obtained by conducting either experiments or surveys. Experimental design is the branch of statistics that deals with the design and analysis of experiments. The methods of experimental design are widely used in the fields of agriculture, medicine, biology, marketing research, and industrial production.

Schematic View of experimental Design



Components of Experimental Design

Factor/Independent/Explanatory Variable

In an experimental study, variables of interest are identified. One or more of these variables, referred to as the factors of the study, are controlled so that data may be obtained about how the factors influence another variable referred to as the response variable, or simply the response. **Factor,** which has originated from the field of agriculture. For example, to determine the effect of sunlight, irrigation, and fertilizer on crop production, we typically say or assume that sunlight is a **factor**, irrigation is a **factor**, fertilizer is a **factor**, etc. on the crop production. **Factor** name has become more popular for the **independent variable** in design of experiments.

Response/Dependent Variable

Dependent variable is known as the response variable in design of experiments. The word response originated from the agriculture field too. Is the plant responding to a fertilizer? Therefore, the word response is used for the dependent variable. Therefore, the response makes more sense when some treatments (or combination of treatments or the combinations of the levels of the factors) are applied to experimental units, and the response is observed and measured.

Factor levels

To study the effect of a factor, a couple of conditions or settings of the factor is applied on the experimental units and the output is observed. Each of these conditions or setting is known as **level**. For example, to determine the effect of a fertilizer, a couple of **levels** of fertilizer is applied and the output is observed. The levels of fertilizer could be set by varying the amount and strength of the fertilizer.

Factor levels are the "values" of that factor in an experiment. **For example**, in the study involving color of cars, the factor car color could have four levels: red, black, blue and grey. In a design involving vaccination, the treatment could have two levels: vaccine and placebo.

Types of factors

- **Experimental Factors:** These are factors that you can specify (and set the levels) and then assign at random as the treatment to the experimental units. Examples would be temperature, level of an additive fertilizer amount per acre, etc.
- **Classification Factors:** These can't be changed or assigned, these come as labels on the experimental units. The age and sex of the participants are

classification factors which can't be changed or randomly assigned. But you can select individuals from these groups randomly.

Quantitative vs. Qualitative Factors

Quantitative Factors

You can assign any specified level of a quantitative factor. Factors are measured on a numerical scale. Examples: percent or pH level of a chemical.

Qualitative Factors

These factors have categories which are different types. Examples might be species of a plant or animal, a brand in the marketing field, gender, - these are not ordered or continuous but are arranged perhaps in sets.

Examples:

1. In a study of the effects of colors and prices on sales of cars, the factors being studied are color (qualitative variable) and price (quantitative variable).

2. In an investigation of the effects of education on income, the factor being studied is education level (qualitative but ordinal).

Fixed Vs Random Factors

Fixed Factor: Levels of the factor are fixed: the levels observed in the study represent the actual levels of experimental interest for example gender(male, female).

Random Factor: Levels of the factor are random: the levels observed in the study represent a sample from all the possible levels for example subjects, school.

Treatment/ Treatment Combinations

The word treatment in the design of experiment can simply be thought of as the medical treatment. For example, if a patient is given a treatment of a medicine, he/she is on a particular treatment.

- In a single factor study, a treatment corresponds to a factor level; thus the number of treatments equals the number of different factor levels of that factor.
- In a multi-factor study, a treatment corresponds to a *combination of factor levels across different factors;* thus the number of all possible treatments is the product of the number of factor levels of different factors. Assume that a human comfort study uses two levels of the temperature factor and two levels of the humidity factor, which results in four treatment combinations (low-low, low-high, high-low, and high-high levels of the temperature and the humidity).

Experimental unit (Experimental plot)

The smallest division of the experimental material to which we apply the treatment and can make the observation on it is called experimental unit or experimental plot.

What are Experimental Errors?

Ø The variation in responses (results) caused by the extraneous factors is termed as experimental errors.

Ø The experimental errors may arise due to:

$ The inherent variability in the experimental material to which the treatments are applied.

$ The lack of uniformity in the methodology of conducting the experiment.

$ Lack of representatives of the sample to the population under study.

Principles of Experimental Design

- Professor Ronald A. Fisher pioneered the design of experiments in statistics.
- In his classic book entitled '*The Design of Experiments*' deals with many statistical experimental designs and its applications.
- According to Fisher, a good experimental design should:

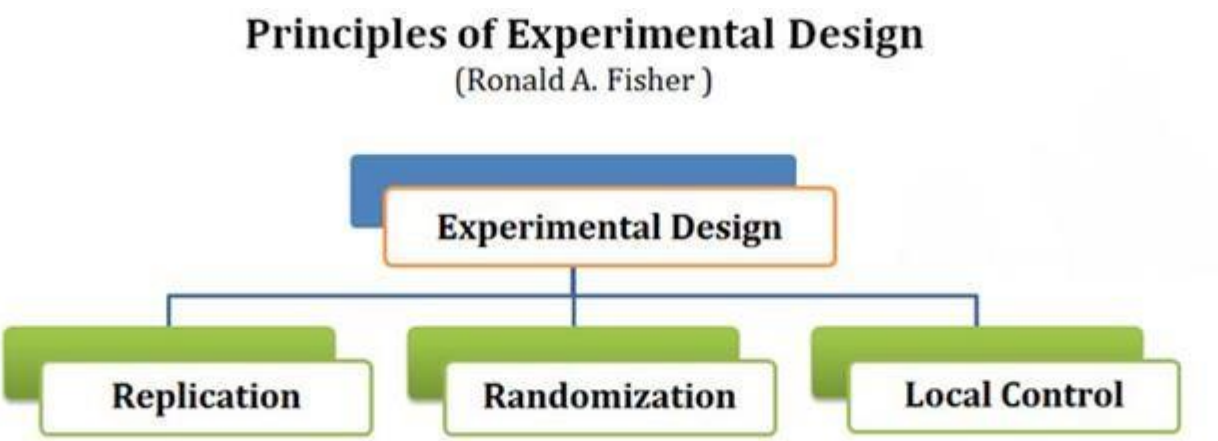
  (A). Increase the efficiency of design

  (B). Reduce the experimental **errors**

- The increased efficiency and reduced experimental errors in experimental designs are achieved by THREE basic principles.

The principles of Experimental Design

1. Replication
2. Randomization
3. Local-Control(Error control)



(1). Replication

- The repetition of the treatments under the investigation is called replication.
- Single treatment does not produce variations in the results.
- Replication of treatments increases the reliability of the estimates.
- It helps to reduce the experimental errors.
- However, replication alone has a limited role in increasing the efficiency of the design.

(2). Randomization

- When all the treatments have an equal chance of being allocated to different experimental units is called randomization.
- Randomization helps to eliminate the bias of any form.
- In the absence of 'replication', the randomization will NOT be effective.
- Replication and Randomization together form the foundation stone in the success of an experimental design.

(3). Local-control

- The process of reducing the experimental errors by providing the relatively heterogeneous experimental areas into homogenous units is called Local-control.
- The local-control will increase the efficiency of the experimental designs.
- Local-control can be used to reduce the extraneous errors.
- Reduction of extraneous errors reduced 'experimental errors'.

Different Types of Experimental Designs

I. Single-Factor Experiments

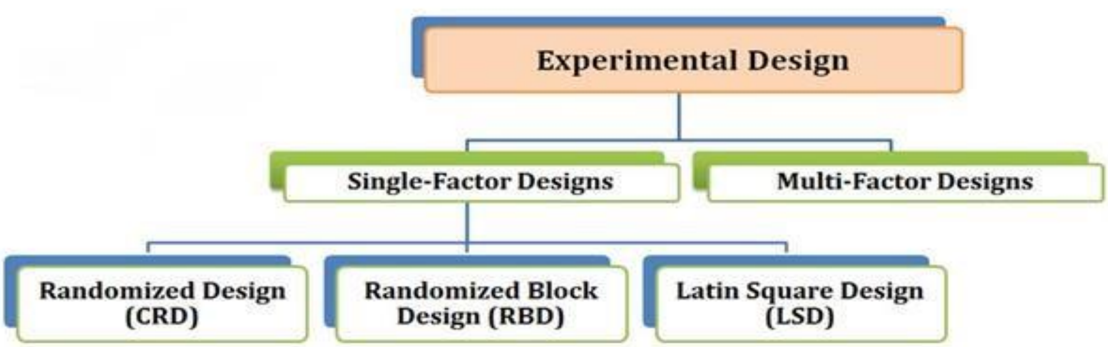
II. Multi-Factor Experiments

(I). Single-Factor Experiments:

- Single factor experiments are those experiments in which only a single factor varies while all others are kept constant.
- Here the treatments consist exclusively of the different levels of the single variable factor.
- All other factors are applied uniformly to all plots.
- Examples of Single-Factor Experimental Designs:
  - (1) Completely Randomized Design (CRD)
  - (2) Randomized Block Design (RBD)
  - (3) Latin-Square Design (LSD)

(II). Multi-Factor Experiments

- Multi-factor experiments are also called as factorial experiments.
- They are used in the experiments where the effects of more than one factor are to be determined.
- A multi-factor experimental design is used to study a problem that is affected by a large number of factors.



Types of Experimental Designs

(1). Completely Randomized Design

Ø It is commonly called as CRD.

Ø CRD is a statistical experimental design where the treatments are assigned completely at random so that each treatment unit has the same chance of receiving any one treatment.

Ø In CRD, any difference among experimental units receiving the same treatment is considered as an experimental error.

Ø CRD is applicable only when the experimental material is homogenous (Example: Homogenous soil condition in the field).

Ø Usually in the field, the soil will be HETEROGENOUS.

Ø Thus, CRD is not a preferable method in field experiments.

Ø CRD is generally applicable to the lab experimental conditions.

Ø In labs, the environmental conditions can be easily controlled.

Ø The concept of 'Local-control' is not used in CRD.

Advantages of CRD

Ø CRD is easy to understand and calculate the variance.

Ø The number of replications can vary from treatment to treatment.

Ø CRD has high flexibility and thus any number of treatments can be used.

Ø Simple statistical analysis is required in the analysis of CRD.

Ø CRD provides maximum number of degree of freedom.

Disadvantages of CRD

Ø CRD can be applied only to homogenous experiments.

Ø The principle of 'Local-control' is not used in CRD.

Designing a simple CDR experiment

An agriculture scientist wants to study the effect of 4 different fertilizers (A,B,C,D) on corn productivity. Four replicates of the 4 treatments are assigned at random to the experimental units

Treatments:types of fertilizer (A,B,C,D)

Experimental unit: Plant (corn)

Dependent Variable(Response): Production of corn

steps

1. Label the experimental units with number 1 to 16.
2. Find 16 three digit random numbers from the random numbers table.
3. Rank the random numbers from the smallest to the largest.
4. Allocate treatment A to the first 4 experimental units, treatment B to the next 4 experimental units and so on .

We will blindly pick a starting location from the table

Table of Random Numbers

36518	36777	89116	05542	29705	83775	21564	81639	27973	62413	85652	62817	57881
46132	81380	75635	19428	88048	08747	20092	12615	35046	67753	69630	10883	13683
31841	77367	40791	97402	27569	90184	02338	39318	54936	34641	95525	86316	87384
84180	93793	64953	51472	65358	23701	75230	47200	78176	85248	90589	74567	22633
78435	37586	07015	98729	76703	16224	97661	79907	06611	26501	93389	92725	68158
41859	94198	37182	61345	88857	53204	86721	59613	67494	17292	94457	89520	77771
13019	07274	51068	93129	40386	51731	44254	66685	72835	01270	42523	45323	63481
82448	72430	29041	59208	95266	33978	70958	60017	39723	00606	17956	19024	15819
25432	96593	83112	96997	55340	80312	78839	09815	16887	22228	06206	54272	83516
69226	38655	03811	08342	47863	02743	11547	38250	58140	98470	24364	99797	73498
25837	68821	66426	20496	84843	18360	91252	99134	48931	99538	21160	09411	44659
38914	82707	24769	72026	56813	49336	71767	04474	32909	74162	50404	68562	14088
04070	60681	64290	26905	65617	76039	91657	71362	32246	49595	50663	47459	57072
01674	14751	28637	86980	11951	10479	41454	48527	53868	37846	85912	15156	00865
70294	35450	39982	79503	34382	43186	69890	63222	30110	56004	04879	05138	57476
73903	98066	52136	89925	50000	96334	30773	80571	31178	52799	41050	76298	43995
87789	56408	77107	88452	80975	03406	36114	64549	79244	82044	00202	45727	35709

From our starting point we will creating three digit numbers **553 408 031 278 839**

098 151 688 722 228 062 065 427 283 516 692

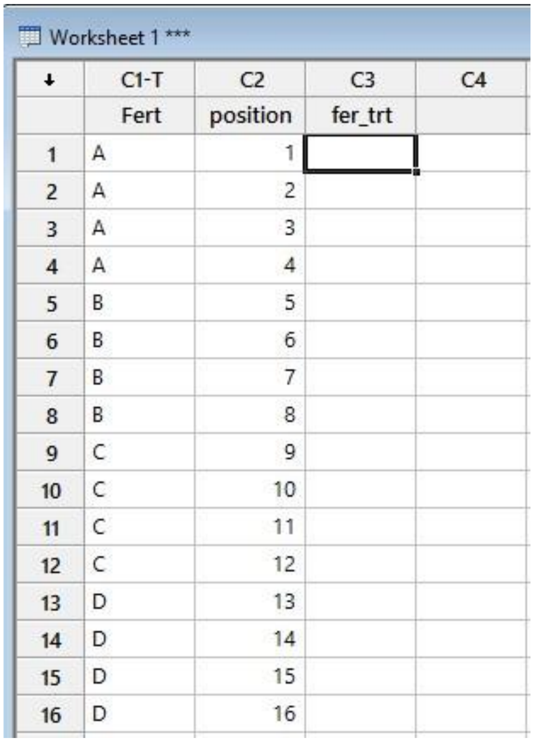
Random	Ranking	treatment
553	12	A
408	9	A
031	1	A
278	7	A
839	16	B
098	4	B
151	5	B
688	13	B
722	15	C
228	6	C
062	2	C
065	3	C
427	10	D
283	8	D
516	11	D
692	14	D

The following table shows the plan of experiment with the treatments that have been allocated to experimental units according to CRD.

Treatments	**A**	12	9	1	7
	B	16	4	5	13
	C	15	6	2	3
	D	10	8	11	14

Using Minitab

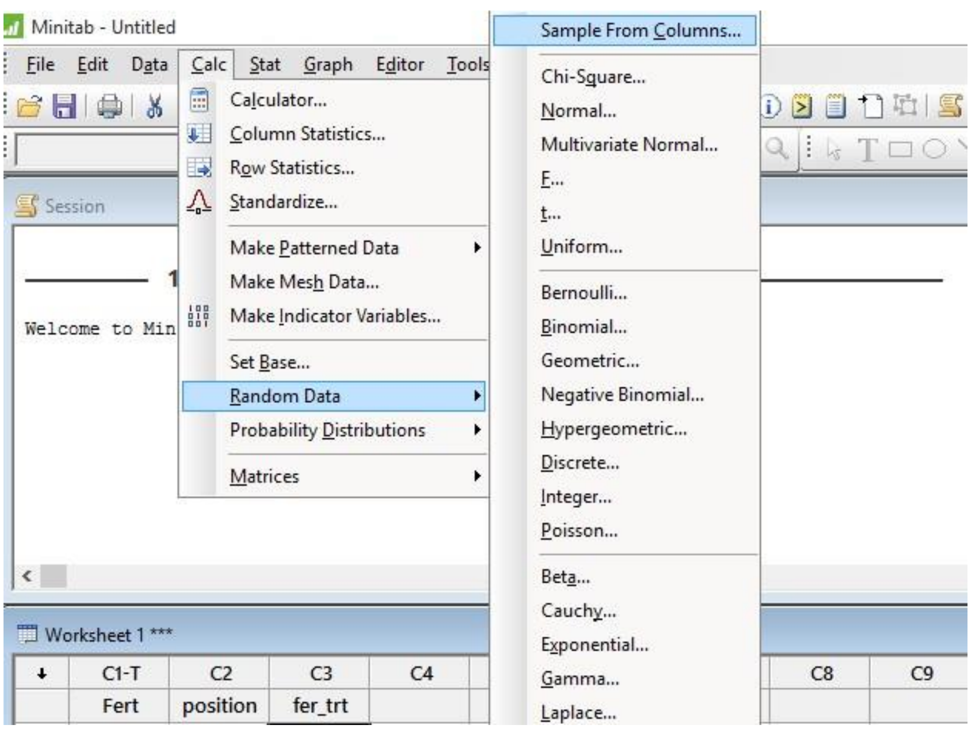
In Minitab, I list the treatment levels and replications to use (order not important), along with a position number 1 to N.



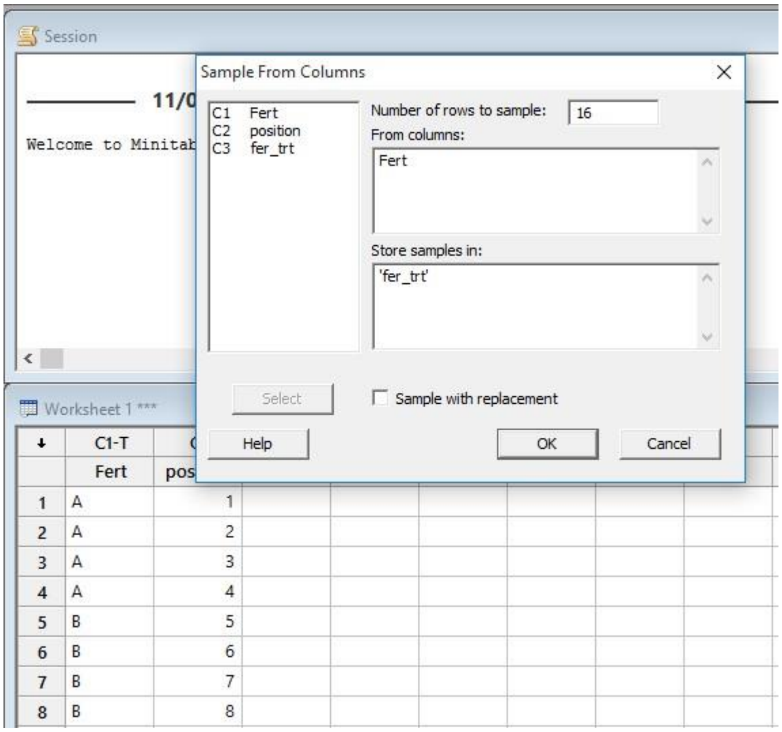
Worksheet 1 ***

↓	C1-T	C2	C3	C4
	Fert	position	fer_trt	
1	A	1		
2	A	2		
3	A	3		
4	A	4		
5	B	5		
6	B	6		
7	B	7		
8	B	8		
9	C	9		
10	C	10		
11	C	11		
12	C	12		
13	D	13		
14	D	14		
15	D	15		
16	D	16		

Then create a column heading for the assigned treatment levels (I used 'Fert_trt' for a heading for column 3).Choose Calc > Random Data > Sample from Columns:



Which brings up this dialog box:



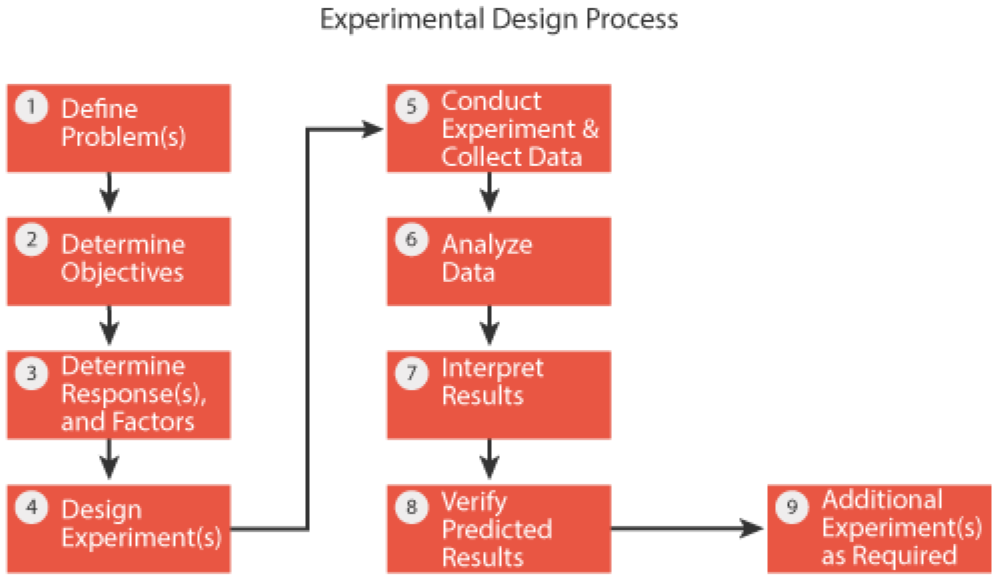
Be sure that you do NOT have the Sample With Replacement box checked. This gives us the treatment levels to apply to each position on the greenhouse bench:

Worksheet 1 ***

↓	C1-T	C2	C3-T
	Fert	position	fer_trt
1	A	1	B
2	A	2	C
3	A	3	C
4	A	4	D
5	B	5	C
6	B	6	D
7	B	7	D
8	B	8	C
9	C	9	A
10	C	10	B
11	C	11	A
12	C	12	A
13	D	13	A
14	D	14	B
15	D	15	D
16	D	16	B

Experimental Design Process

The flow chart below illustrates the experiment design process:



Quiz

MCQ

1. CRD can be used with

   (a) Equal replication (b) unequal replication

   (c) Equal and unequal replication (d) single replication

2. When there are 5 treatments each replicated 4 times the total number of experimental plots will be

   (a)5 (b) 4 (c) 9 (d)20

True or False

CRD can be adopted only when the experimental material is homogenous.( )

- Mention any two advantages of CRD?