

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 economic way. If the experiment is designed properly keeping in mind the question, then the data generated is valid and proper analysis of the data provides the valid statistical inference.

experimental units:

For conducting an experiment, the experimental material is divided into smaller parts and each of them is referred to as an experimental unit.

Example: Subjects in a drug testing experiment are experimental units.

Treatment

Different objects or procedures which are to be compared in an experiment are called treatments.

Important principles to conduct for design of experiments

① Blinding: For example, in experiments involving placebo, there is often a placebo effect that occurs when an unrelated subject reports an improved symptom. To account for the placebo effect, use the blinding, a technique in which the subjects do not know whether he or she has received a treatment or a placebo.

② Randomization:

The principle of randomization involves allocation of treatments to experimental units at random to avoid any bias in the experiment resulting from the influence of some extraneous unknown factors that may affect the experiment.

Random assignment is needed for the following reasons.

- a) It eliminates the systematic bias
- b) It is needed to obtain a representative sample from the population.
- c) It helps in distributing the unknown variation due to confounded variables throughout the experiment and breaks the confounding influence. (May 31)

① Recap: Design of experiment

① Blinding:

② Randomization:

If the randomization process is such that every experimental unit has an equal chance of receiving each treatment, it is called a complete randomization.

③ Replication: In the replication principle, any treatment is repeated multiple times to obtain a reasonable estimate of the parameters with higher precision.

④ Local control: Local control is a broad term used to reduce experimental errors. For example, if the experimental units are divided into different groups such that each group is homogeneous then more local control is achieved.

complete and incomplete block designs

In most of the experiments, to achieve local control, experimental units are grouped into blocks having more or less identical characteristics

to remove the blocking effect from the experimental error. Such designs are termed as block designs.

The number of experimental units in a block is called the block size.

If, size of ~~the~~ block = number of treatments and each treatment in each block is assigned randomly, then it is called a full replication and the design is called as complete block design.

When the number of treatments is very large it is difficult to create homogeneous blocks with respect to some characteristics.

With large number of experimental units within a block, the cost of the experiment will grow. In such cases, every block will receive a subset of all treatments. This is called incomplete block design.

- ① Completely randomized Design (CRD)
- ② Randomized Block Design (RBD)
- ③ Latin square design (~~LSD~~)
- ④ ~~Incom~~ Incomplete block design (IBD)

Completely randomized design

The CRD is the simplest design. Suppose there are τ treatments which are to be compared.

- ① All experimental units are same and no division or grouping among them exists.
CRD is only limited to controlled lab experiments.
- ② Design is entirely flexible in the sense that any number of treatments or replications can be used.

Procedure: Let the τ treatments are numbered $1, 2, \dots, \tau$ and n_i be the number of replications required for the i th treatment so that $\sum_{i=1}^{\tau} n_i = n$. (n = total number of experimental units)

- ① Select n_1 units out of n randomly and apply treatment 1 to those n_1 units.
- ② Select $\textcircled{1} n_2$ out of $(n-n_1)$ units and apply treatment 2 to those n_2 units.
- ③ Continuing in this way, apply $\textcircled{2}$ treatment τ to ~~n_τ~~ n_τ units.

Disadv: ① No local control on blocking effect is introduced. Any sources of variation other than the treatments would be considered ~~as~~ variation due to unknown error.

Analysis:

y_{ij} : individual measurement of j th experimental unit for i th treatment, $i=1, \dots, r$; $j=1, \dots, n_i$.

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \tau^2)$$

$$\sum_{i=1}^r n_i \alpha_i = 0 \quad \text{under this } \cancel{\text{constr}} \text{ constraint,}$$

$$\hat{\mu} = \bar{y}_{..}, \quad \bar{y}_{..} = \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} y_{ij}$$

$$\hat{\alpha}_i = \bar{y}_{i..} - \bar{y}_{..}$$

$$y_{ij} = \bar{y}_{..} + (\bar{y}_{i..} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i..})$$

$$SST = \sum_{i=1}^r \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2$$

$$SSR = \sum_{i=1}^r n_i (\bar{y}_{i..} - \bar{y}_{..})^2$$

$$SSE = \sum_{i=1}^r \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i..})^2 = y^T (I - P) y$$

<u>Source of variation</u>	<u>Df</u>	<u>SS</u>	<u>MS</u>	<u>F</u>
treatments	$r-1$	SSR	$MSR = \frac{SSR}{r-1}$	$\frac{MSR}{MSE}$
errors	$n-r$	SSE	$MSE = \frac{SSE}{n-r}$	

in the test statistic to test if $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_r = 0$

Randomized block design

In CRD there was no way to identify variation due to heterogeneous blocks. To identify more sources of variations. (3) Here is the layout.

- ① Group the experimental materials into blocks of size b units ($b = \text{no. of treatments}$).
- ② Blocks are constructed such that the experimental units within a block are relatively homogeneous and resemble each other more than units from other blocks.

(3)

		blocks						
		1	2	3	4	\dots	b	
Treatments	1	y_{11}	y_{12}	-	-	-	-	y_{1b}
	2	y_{21}	-	-	-	-	-	y_{2b}
	3	-	-	-	-	-	-	-
	4	y_{41}	-	-	-	-	-	y_{4b}
	\vdots	-	-	-	-	-	-	-

(5)

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \quad \text{④ } \epsilon_{ij} \sim N(0, \sigma^2)$$

$i=1, \dots, v; \quad j=1, \dots, b$

$$\underline{Y} = \underline{x} \otimes \underline{\gamma} + \underline{\varepsilon}, \quad \underline{\gamma} = (\mu, \alpha_1, \dots, \alpha_v, \beta_1, \dots, \beta_b)$$

$$\underline{x} = \begin{bmatrix} \frac{1}{\sqrt{b}} & \frac{1}{\sqrt{b}} & 0 & 0 & \underline{I}_b \\ 0 & \frac{1}{\sqrt{b}} & \dots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & 0 & \underline{I}_b & \vdots \\ \frac{1}{\sqrt{b}} & 0 & 0 & \underline{I}_b & \vdots \end{bmatrix}$$

~~Rank~~

$$\text{rank } \otimes(\underline{x}) = b + v - 1$$

no. of columns

$$\text{of } \underline{x} = b + v + 1$$

We need two constraints to uniquely estimate all parameters.

$$\sum_{i=1}^v \alpha_i = 0, \quad \sum_{j=1}^b \beta_j = 0 \quad [\text{constraints}]$$

$$\min_{\mu, \alpha_i, \beta_j} \sum_{i=1}^v \sum_{j=1}^b (y_{ij} - \mu - \alpha_i - \beta_j)^2$$

$$S = \sum_{i=1}^v \sum_{j=1}^b (y_{ij} - \mu - \alpha_i - \beta_j)^2$$

$$\frac{\partial S}{\partial \mu} = 0 \Rightarrow \sum_{i=1}^v \sum_{j=1}^b (y_{ij} - \mu - \alpha_i - \beta_j) = 0$$

$$\Rightarrow \sum_{i=1}^v \sum_{j=1}^b (y_{ij} - \mu) = 0 \Rightarrow \hat{\mu} = \frac{1}{b} \sum_{j=1}^b \sum_{i=1}^v y_{ij} = \bar{y}_{..}$$

$$\frac{\partial S}{\partial \alpha_i} = 0 \Rightarrow \sum_{j=1}^b (y_{ij} - \mu - \alpha_i - \beta_j) = 0$$

$$\Rightarrow \sum_{j=1}^b (y_{ij} - \mu - \alpha_i) = 0 \Rightarrow \alpha_i b = \sum_{j=1}^b (y_{ij} - \bar{y}_{..})$$

$$\Rightarrow \hat{\alpha}_i = \frac{1}{b} \sum_{j=1}^b y_{ij} - \bar{y}_{..} = \bar{y}_{i..} - \bar{y}_{..}$$

$$\frac{\partial S}{\partial \beta_j} = 0 \Rightarrow \hat{\beta}_j = \bar{y}_{..j} - \bar{y}_{..}, \quad \bar{y}_{..j} = \frac{1}{n} \sum_{i=1}^n y_{ij}$$

$$y_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\epsilon}_{ij}$$

$$= \bar{y}_{..} + (\bar{y}_{i..} - \bar{y}_{..}) + (\bar{y}_{..j} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i..} - \bar{y}_{..j} + \bar{y}_{..})$$

$$X = \begin{bmatrix} 1_b & 1_b & 0 & \dots & 0 & I_b \\ 1_b & 0 & 1_b & \dots & 0 & I_b \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1_b & 0 & \dots & 0 & I_b \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1_b & I_b \\ \hline x_0^* & & & & & \\ \hline & x_1^* & & & & \\ \hline & & x_2^* & & & \end{bmatrix}$$

$$\textcircled{1} \quad P_{X_0^*} y = \left(\begin{array}{c} \bar{y}_{..} \\ \vdots \\ \bar{y}_{..} \end{array} \right) \quad \textcircled{2}$$

b.J. stimmt

$$(P_{X_1^*} - P_{X_0^*}) y = \left(\begin{array}{c} \bar{y}_{i..} - \bar{y}_{..} \\ \bar{y}_{i..} - \bar{y}_{..} \\ \bar{y}_{..j} - \bar{y}_{..} \\ \vdots \\ \bar{y}_{..j} - \bar{y}_{..} \end{array} \right)$$

back ①

$$(P_{X_2^*} - P_{X_1^*}) \underline{y} = \begin{pmatrix} \bar{y}_{11} - \bar{y}_{..} \\ \bar{y}_{12} - \bar{y}_{..} \\ \vdots \\ \bar{y}_{1b} - \bar{y}_{..} \\ \bar{y}_{11} - \bar{y}_{..} \end{pmatrix} \text{ } v \text{ times}$$

$$\underline{y} = P_{X_0^*} \underline{y} + (P_{X_1^*} - P_{X_0^*}) \underline{y} + (P_{X_2^*} - P_{X_1^*}) \underline{y} + (I - P_{X_2^*}) \underline{y}$$

$$\underline{y}' \underline{y} = \underline{y}' P_{X_0^*} \underline{y} + \underline{y}' (P_{X_1^*} - P_{X_0^*}) \underline{y} + \underline{y}' (P_{X_2^*} - P_{X_1^*}) \underline{y} + \underline{y}' (I - P_{X_2^*}) \underline{y}$$

By Cochran's theorem, ncp1

$$\underline{y}' \underbrace{P_{X_0^*} \underline{y}}_{\sim \chi^2(\text{rank}(P_{X_0^*}))} = \chi^2(1)$$

$$\underline{y}' \underbrace{(P_{X_1^*} - P_{X_0^*}) \underline{y}}_{\sim \chi^2(\text{rank}(P_{X_1^*} - P_{X_0^*}))} \stackrel{\text{ncp2}}{\sim} \chi^2(\chi^2(\text{rank}(P_{X_1^*} - P_{X_0^*}))) = \chi^2(v-1)$$

$$\underline{y}' \underbrace{(P_{X_2^*} - P_{X_1^*}) \underline{y}}_{\sim \chi^2(\text{rank}(P_{X_2^*} - P_{X_1^*}))} \stackrel{\text{ncp3}}{\sim} \chi^2(\chi^2(\text{rank}(P_{X_2^*} - P_{X_1^*}))) = \chi^2(b-1)$$

$$\underline{y}' \underbrace{(I - P_{X_2^*}) \underline{y}}_{\sim \chi^2(b^2 - (v-1) - (b-1) - 1)} \otimes \\ = b^2 - v - b + 1 \\ = (b-1)(v-1)$$

$$y_{ij} = \bar{y}_{..} + (\bar{y}_{i..} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i..} - \bar{y}_{.j} + \bar{y}_{..})$$

$$\underline{y}' P_{X_0^*} \underline{y} = \text{#} \parallel P_{X_0^*} \underline{y} \parallel^2 = b^2 \bar{y}_{..}^2$$

$$\underline{y}' (P_{X_1^*} - P_{X_0^*}) \underline{y} = \parallel (P_{X_1^*} - P_{X_0^*}) \underline{y} \parallel^2 = b \sum_{i=1}^v (\bar{y}_{i..} - \bar{y}_{..})^2$$

back ②

$$\underline{y}' (\mathbf{P}_{X_2^{\perp}} - \mathbf{P}_{X_1^{\perp}}) \underline{y} = \|(\mathbf{P}_{X_2^{\perp}} - \mathbf{P}_{X_1^{\perp}})\underline{y}\|^2 = \sum_{j=1}^b v_j (\bar{y}_{\cdot j} - \bar{y}_{\cdot \cdot})^2$$

$$\underline{y}' (\mathbf{I} - \mathbf{P}_{X_2^{\perp}}) \underline{y} = \sum_{i=1}^d \sum_{j=1}^b (y_{ij} - \bar{y}_{i \cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot \cdot})^2$$

$$H_0: \alpha_1 = \dots = \alpha_d = 0$$

$$H_0: \beta_1 = \dots = \beta_b = 0$$

back ③