DESIGN AND ANALYSIS OF EXPERIMENTS

B.M Njuguna

2022-10-29

Contents

Introduction

This paper is based on advanced experimental design for scientific Studies. The various formulas used in different designs and experiments are as outlined below.

1. One Way Classification Model

This is basically the Analysis of Covariance(ANCOVA). Here single factor experiment with single covariate is considered. The model is as follows;

$$Y_{ij} = \mu + \tau_i + \beta(x_{ij} - \bar{x}..) + \epsilon_{ij}$$

where; i = 1, 2, ..., t and j = 1,2, ... r

Y_{ij}- is the jth response variable taken under the ith treatment

μ - overall mean

x_{ij} - The measure of covariate corresponding to Y_{ij}

$\bar{x}..$ - The mean of x_{ij} value

β - The linear regression coefficient of Y_{ij} and xij

The notations used are;

$$S_{yy} = \sum_{i=1}^{t}\sum_{j=1}^{r}(Y_{ij} - \bar{Y}..)^2 = \sum_{i=1}^{t}\sum_{j=1}^{r}Y_{ij}^2 - \frac{Y..^2}{tr}$$

$$S_{xx} = \sum_{i=1}^{t}\sum_{j=1}^{r}(X_{ij} - \bar{x}..)^2 = \sum_{i=1}^{t}\sum_{j=1}^{r}X_{ij}^2 - \frac{X..^2}{tr}$$

$$S_{xy} = \sum_{i=1}^{t}\sum_{j=1}^{r}(Y_{ij} - \bar{Y}..)(X_{ij} - \bar{x}..) = \sum_{i=1}^{t}\sum_{j=1}^{r}Y_{ij}X_{ij} - \frac{Y..X..}{tr}$$

$$T_{yy} = \sum_{i=1}^{t}\sum_{j=1}^{r}(Y_{i.} - \bar{Y}..)^2 = \frac{\sum_{i=1}^{t}Y_{i.}^2}{r} - \frac{Y_{..}^2}{tr}$$

$$T_{xx} = \sum_{i=1}^{t}\sum_{j=1}^{r}(X_{i.} - \bar{X}..)^2 = \frac{\sum_{i=1}^{t}X_{i.}^2}{r} - \frac{X_{..}^2}{tr}$$

$$\frac{\sum_{i=1}^{t}Y_iXi}{r} - \frac{Y_{..}X..}{tr}$$

$$E_{yy} = S_{yy} - T_{yy}$$

$$E_{xx} = S_{xx} - T_{xx}$$

$$E_{xy} = S_{xy} - T_{xy}$$

The statistical analysis is;

LSE of μ is; $\hat{\mu} = \bar{Y}..$

Then;

$$\hat{\beta} = \frac{E_{xy}}{E_{xx}}$$

$$SSE = E_{yy} - \frac{E_{xy}^2}{Exx}$$

SSE usually have t(r-1)-1 degrees of freedom

Now suppose we use to test $\tau_i = 0$. Then under Null hypothesis, the reduced model will be;

$$Y_{ij} = \mu + \beta(x_{ij} - \bar{x}.. + \epsilon_y$$

Then;

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}}$$

And;

$$SSE' = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$$

note SSE is smaller than SSE', where $SSE' - SSE$ is a reduction in sums of squares due to τ_i. Therefore for testing $\tau_i = 0$, the test statistic is;

$$F_{calc} = \frac{(SSE' - SSE)/(t-1)}{SSE/(t(r-1)-1)}$$

We test it against $F_{(t-1),t(r-1)-1,\alpha}$. The ANOVA table is as follows;

...1	...2	SS and Product	...4	...5	Adjustment for Regression	...7
Source of Variation	df	X	XY	Y	Y	df
Treatment	t-1	Txx	Txy	Tyy	NA	NA
Error	t(r-1)	Exx	Exy	Eyy	SSE	t(r-1)-1
Total	tr-1	Sxx	Sxy	Syy	SSE'	tr-2
Adjusted Treatment	NA	NA	NA	NA	SSE'-SSE	t-1

2. Nested Designs

i. Two - Stage Nested Designs

The statistical model is;

$$Y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \epsilon_{(ij)k}$$

where, i = 1,2,...,a, j = 1,2,...,b, and k = i,2,...r

μ - overall mean

τ_i - Effect of the ith factor A

$\beta_{j(i)}$ - Effect of the jth factor B nested under factor A

$\epsilon_{(ij)k}$ - Random error term.

The sums of squares are partitioned as follows;

$$SS_{total} = SS_A + ss_{B(A)} + SS_{Error}$$

They are calculated as follows;

$$SS_{Total} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r} Y_{ijk}^2 - \frac{(Y...)^2}{abr}$$

$$SS_A = \frac{\sum_{i=1}^{a} Y_{i..}^2}{br} - \frac{(Y...)^2}{abr}$$

$$SS_{B(A)} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij.}^2}{r} - \frac{\sum_{i=1}^{a} Yi..^2}{br}$$

$$SS_{Error} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r} Y_{ijk}^2 - \frac{\sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij.}^2}{r}$$

The ANOVA table is as follows;

Source of Variation	df	SS	MS
A	$a-1$	SS_A	MS_A
$B(A)$	$a(b-1)$	$SS_{B(A)}$	$MS_{B(A)}$
$Error$	$ab(r-1)$	SS_{Error}	MS_{Error}
$Total$	$abr-1$	SS_{Total}	

The appropriate statistic for testing the effect of factor A and B depends on whether the levels of A and B are fixed or random. The Expected Mean Squares in the two stage nested design are as follows;

$E(MS)$	A - fixed B- fixed	A - fixed B - Random	A - Random B - Random
$E(MS_A)$	$\frac{\sigma^2+br\sum\tau_i^2}{a-1}$	$\sigma^2 + r\sigma_\beta^2 + \frac{br\sum\tau_i}{a-1}$	$\sigma^2 + r\sigma_\beta^2 + br\sigma_\tau^2$
$E(MS_{AB})$	$\sigma^2 + \frac{r\sum_i\sum_j\beta_{j(i)}^2}{a(b-1)}$	$\sigma^2 + r\sigma_\beta^2$	$\sigma^2 + r\sigma_\beta^2$

$E(MS)$	A - fixed B- fixed	A - fixed B - Random	A - Random B - Random
$E(MS_E)$	σ^2	σ^2	σ^2

The testing of hypothesis is as follows;

The testing of hypothesis is as follows;

1. **when A is fixed and B is random**

To test $H_0 : \tau_i = 0$ vs $H_1 : \tau_i \neq 0$ the test statistic is;

$$F_{calc} = \frac{MS_A}{MS_{B(A)}}$$

We reject H_0 is $F_{calc} > F_{a-1,a(b-1),\alpha}$

To test $H_0 : \sigma^2_\beta = 0$ vs $H_1 : \sigma^2_\beta \neq 0$, the test statistic is;

$$F(calc) = \frac{MS_{B(A)}}{MS_{Error}}$$

We reject H_0 is $F_{calc} > F_{a(b-1),ab(r-1),\alpha}$

2. **A and B fixed**

To test $H_0 : \tau_i = 0$ vs $H_1 : \tau_i \neq 0$ the test statistic is;

$$F_{calc} = \frac{MS_A}{MS_{Error}}$$

To test $H_0 : \beta_{j(i)} = 0$ vs $H_1 : \beta_{j(i)} \neq 0$, the test statistic is;

$$F(calc) = \frac{MS_{B(A)}}{MS_{Error}}$$

3. **A and B random**

To test $H_0 : \sigma^2_\tau = 0$ vs $H_1 : {\sigma^2}_\tau \neq 0$ the test statistic is;

$$F_{calc} = \frac{MS_A}{MS_{B(A)}}$$

We reject H_0 is $F_{calc} > F_{a-1,a(b-1),\alpha}$

To test $H_0 : \sigma^2_\beta = 0$ vs $H_1 : \sigma^2_\beta \neq 0$, the test statistic is;

$$F(calc) = \frac{MS_{B(A)}}{MS_{Error}}$$

We reject H_0 is $F_{calc} > F_{a(b-1),ab(r-1),\alpha}$

ii. Three Stage Nested Design

The model is written as;

$$Y_{ijkl} = \mu + \tau_i + \beta_{j(i)} + \gamma_{k(ij)} + \epsilon_{(ijk)l}$$

where; i=1,2,...a, j = 1,2,...,b, k = 1,2,...c and l = 1,2,...r

μ - overall mean

τ_i - Effect of the ith factor A

$\beta_{j(i)}$ - Effect of the jth factor B nested under factor A

$\gamma_{k(ij)}$ - Effect of the kth factor B nested under factor A and B

$\epsilon_{(ij)k}$ - Random error term.

The sums of squares are partitioned as follows;

$$SS_{Total} = SS_A + SS_{B(A)} + SS_{C(B)} + SS_{Error}$$

$$SS_{Total=\sum_i \sum_j \sum_k \sum_l Y_{ijkl}^2} - \frac{Y_{....}^2}{abcr}$$

$$SS_A = \frac{\sum_i Y_{i...}^2}{bcr} - CT$$

$$SS_{B(A)} = \frac{\sum_i \sum_j Y_{ij..}^2}{cr} - \frac{\sum_i Y_{i...}^2}{bcr}$$

$$SS_{C(B)} = \frac{\sum_i \sum_j \sum_k Y_{ijk.}^2}{r} - \frac{\sum_i \sum_j Y_{ij..}^2}{cr}$$

$$SS_{Error} = \sum_i \sum_j \sum_k \sum_l Y_{ijkl}^2 - \frac{\sum_i \sum_j \sum_k Y_{ijk.}^2}{r}$$

The ANOVA Table is as follows;

Source of Variation	df	SS	MS
A	$a-1$	SS_A	MS_A
$B(A)$	$b-1$	$SS_{B(A)}$	$MS_{B(A)}$
$C(A)$	$ab(c-1)$	$SS_{C(B)}$	$MS_{C(B)}$
$Error$	$abc(r-1)$	SS_{Error}	MS_{Error}
$Total$	$abcr-1$	SS_{Total}	

3. Balanced Incomplete Block Design

In some experiments, the number of treatments is large, the adoption of RCBD therefore may result in an increase of error variance due to large block size. Here we use the BIBD. The parameters of BIBD are t, b, r, k, λ. For an experiment to be a BIBD, it must satisfy the following two conditions;

1. $bk = rt$
2. $\lambda(t-1) = r(k-1)$, this implies that; $\lambda = \frac{r(k-1)}{t-1}$. Note that λ is the number of times each pair of treatment appear or occur together in the same block.

The model can be written as;

$$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$$

where; i = 1,2,...,t, j = 1,2,...,r

μ - overall mean

τ_i - Effect of the ith treatment

β_j - Effect of the jth block

ϵ_{ij} - Random error term

The sums of squares are partitioned as follows;

$$SS_{Total} = SS_{Block} + SS_{treat(adj)} + SS_{Error}$$

$$SS_{Total=\sum_i \sum_j Y_{ij}^2} - \frac{Y_{..}^2}{N = (r \times torb \times k)}$$

$$SS_{Block} = \frac{\sum_{j=1}^{k} Y_{ij}^2}{k} - \frac{Y_{..}^2}{N}$$

$$SS_{treat(adj)} = k\frac{\sum_{i=1}^{r} Q_I}{\lambda t}$$

where Q_i is the adjusted total for the ith treatment computed as;

$$Q_i = Y_{i.} - \frac{1}{k}\sum_j n_{ijY_{.j}}$$

i = 1,2,...,t

Then; $n_{ij} = 1$ if treatment i appears in block j and $n_{ij} = 0$ otherwise.

The ANOVA table is as follows;

Source of Variation	df	SS	MS
$Block$	$b-1$	SS_{Block}	MS_{Block}
$Treat(adj)$	$t-1$	$SS_{treat(adj)}$	$MS_{treat(adj)}$
$Error$	$N-t-b+1$	SS_{Error}	MS_{Error}
$Total$	$N-1$	SS_{Total}	

To test $H_o : \tau_i = 0$ vs $H_1 : \tau_i \neq 0$, the test statistic is;

$$F_{calc} = \frac{MS_{treat(adj)}}{MS_{Error}}$$

We reject H_0 if $F_{calc} > F_{t-1,N-t-b+1}$

4. Partially Balanced Incomplete Design.

BIBD do not exist for all combination of parameters you might with to employ. PBIBD are designs whereby the number of times a given treatment occur will vary from treatment to treatment where some pairs of treatment appear together λ_1 times and others appear λ_2 times. Note $\lambda_1 > \lambda_2$. This is the simplest of the PBIBD which had two associate classes; λ_1 and λ_2. For a design to be PBIBD, it must satisfy the following two conditions;

$$\sum_{i=1}^{2} n_i = t - 1, n_1 + n_2 = t - 1$$

$$\sum_{i=1}^{2} \lambda_i n_i = r(k-1), \lambda_1 n_1 + \lambda_2 n_2 = r(k-1)$$

How to determine P^1_{jk}, you pick any two pair of treatments that are of first associate. Then to determine P^2_{jk}, you pick any two pairs of treatments which are second associate.

Lattice Designs

Consider a balanced incomplete design with $t = k^2$ treatments arranged in $b = k(k+1)$ blocks with k runs per blocks and $r = k + 1$ replicates. Such a design is called a **balanced lattice**.

1. A design for k^2 treatments in $2k$ blocks of k runs with 2 replicates is called a **simple lattice**.
2. A lattice design with $t = k^2$ treatments in $3k$ blocks grouped into 3 replicates, $r = 3$ is called a **triple lattice design**.
3. A lattice design for $t = k^2$ treatments in $4k$ blocks arranged in 4 replicates, $r = 4$ is called a **quadruple lattice**.
4. A lattice design for $t = k^3$ treatments in k^2 blocks of k runs is called a **cubic lattice design**.
5. A lattice design for $t = k(k+1)$ treatments in $k+1$ blocks of size k is called a **rectangular lattice design**.