STAT 485/859: Design and Analysis of Experiments
(TR 13:00-14:15 CL 312)
4. Randomized Blocks, Latin Squares, and Related Designs

Dianliang Deng, College West 307.22

Department of Mathematics and Statistics

University of Regina, September, 2019

4.1 The Randomized Complete Block Design

In any experiment, variability arising from nuisance can affect the results.
Nuisance Factor: a design factor that has an effect on the response, but we are not interested in
There are three situations:
I. Unknown and uncontrollable: Randomization is a design technique to guard against such a “lurking” nuisance factor.
II. Known but uncontrollable: Analysis of covariance.
III. Known and controllable: Blocking technique.

Example:

Reconsider the hardness testing experiment. Suppose now that we wish to determine whether or not four different tips produce different readings. There is only one factor - tip type - and a completely randomized single-factor design would consist of randomly assigning each one of $4 \times 4 = 16$ runs to an experimental unit.

We would like to make the experiment error as small as possible. A design that would accomplish this requires the experimenter to test each tip once on each of four coupons. This design is called a **randomized complete block design (RCBD)**. The word “Complete” indicates that each block (coupon) contain all the treatments. This design can remove the variability of blocks (coupons).

Table 4.1 Randomized Complete Block Design for the Hardness Testing Experiment

Test Coupon (BLOCK)			
1	2	3	4
Tip 3	Tip 4	Tip 2	Tip 1
Tip 1	Tip 3	Tip 4	Tip 2
Tip 4	Tip 2	Tip 1	Tip 3
Tip 2	Tip 1	Tip 3	Tip 4

4.1.1 Statistical Analysis of the RCBD

The statistical model for RCBD is :

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}, i = 1, 2, ..., a; j = 1, 2, ..., b.$$

where μ: overall mean
τ_i: Effect of the ith treatment
β_j: Effect of the jth block
ϵ_{ij}: $NID(0, \sigma^2)$.
and

$$\sum_{i=1}^{a} \tau_i = 0, \quad \sum_{j=1}^{b} \beta_j = 0$$

This is a fixed effect model.

In an experiment involving the RCBD, we are interested in testing the equality of the treatments. Thus, the hypotheses are

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_d, \quad H_1 : \text{ at least one } \mu_i \neq \mu_j$$

and the equivalent hypotheses are

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_d = 0, \quad H_1 : \text{ at least one } \tau_i \neq 0$$

Next we define some notation:

$$y_{i.} = \sum_{j=1}^{b} y_{ij}, \qquad i = 1, 2, ..., a$$

$$y_{.j} = \sum_{i=1}^{a} y_{ij}, j = 1, 2, ..., b$$

$$y_{..} = \sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij} = \sum_{i=1}^{a} y_{i.} = \sum_{j=1}^{b} y_{.j}$$

and

$$\bar{y}_{i.} = y_{i.}/b, \quad \bar{y}_{.j} = y_{.j}/a, \quad \bar{y}_{..} = y_{..}/N$$

We may express the total corrected sum of squares as

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{..})^2 = b\sum_{i=1}^{a}(\bar{y}_{i.}-\bar{y}_{..})^2 + a\sum_{j=1}^{b}(\bar{y}_{.j}-\bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})^2$$

This is the fundamental ANOVA equation fr the RCBD. Further,

$$\underbrace{SS_T}_{ab-1} = \underbrace{SS_{Treatments}}_{a-1} + \underbrace{SS_{Blocks}}_{b-1} + \underbrace{SS_E}_{(a-1)(b-1)}$$

Moreover, we have

$$E(MS_{Tretments}) = \sigma^2 + \frac{b\sum_{i=1}^{a}\tau_i^2}{a-1}$$

$$E(MS_{Blocks}) = \sigma^2 + \frac{a\sum_{j=1}^{b}\beta_j^2}{b-1}$$

and

$$E(MS_E) = \sigma^2$$

where $MS_{treatment}, MS_{Blocks}, MS_E$ are the mean squares which are obtained from the corresponding sums of squares $SS_{Treatment}, SS_{Blocks}, SS_E$.

Therefore, to test the equality of treatment means, we would use the test statistic

$$F_0 = \frac{MS_{Treatment}}{MS_E}$$

which is distributed as $F_{a-1,(a-1)(b-1)}$ if the null hypothesis is true. The quantities can be computed in the columns of a spreadsheet. Alternatively, computing formulas can be expressed in terms of treatment and block totals.

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij} - \frac{y_{..}^2}{N}$$

$$SS_{treatments} = \frac{1}{b} \sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{N}$$

$$SS_{Blocks} = \frac{1}{a} \sum_{j=1}^{b} y_{.j} - \frac{y_{..}^2}{N}$$

and

$$SS_E = SS_T - SS_{Treatments} - SS_{Blocks}$$

Table 4-2 Analysis of Variance for a Randomized Complete Block Design

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Treatments	$SS_{\text{Treatments}}$	$a-1$	$\dfrac{SS_{\text{Treatments}}}{a-1}$	$\dfrac{MS_{\text{Treatments}}}{MS_E}$
Blocks	SS_{Blocks}	$b-1$	$\dfrac{SS_{\text{Blocks}}}{b-1}$	
Error	SS_E	$(a-1)(b-1)$	$\dfrac{SS_E}{(a-1)(b-1)}$	
Total	SS_T	$N-1$		

Example 4.1 A medical device manufacturer produces vascular grafts (artificial veins).

Table 4.3 Randomized Complete Block Design for the Vascular Graft Experiment

Extrusion Pressure(PSI)	Batch of Resin (Block) 1	2	3	4	5	6	Treatment Total
8500	90.3	89.2	98.2	93.9	87.4	97.9	556.9
8700	92.5	89.3	90.6	94.7	87.0	95.8	550.1
8900	85.5	90.8	89.6	86.2	88.0	93.4	533.5
9100	82.5	89.5	85.6	87.4	78.9	90.7	514.6
Block Total	350.8	359.0	364.0	362.2	341.3	277.8	2155.1

Table 4.4 Analysis of Variance for the Vascular Graft Experiment

Source of Variation	Sum of Squares	Degree of Freedom	Mean Square	F_0	P-Value
Treatment (Extrusion Pressure)	178.17	3	59.39	8.11	0.0019
Blocks (Batches)	192.25	5	38.45		
Error	109.89	15	7.33		
Total	480.31	23			

Multiple Comparison If the treatment in an RCBD are fixed, and the analysis indicates a significant difference in treatment means, the experimenter is usually interested in multiple comparison to discover which treatment means differ.

Treatment Means (Adjusted, If Necessary)

	Estimated Mean	Standard Error
1–8500	92.82	1.10
2-8700	91.68	1.10
3-8900	88.92	1.10
4-9100	85.77	1.10

Treatment	Mean Difference	DF	Standard Error	t for H_0 Coeff=0	Prob > |t|
1 vs 2	1.13	1	1.56	0.73	0.4795
1 vs 3	3.90	1	1.56	2.50	0.0247
1 vs 4	7.05	1	1.56	4.51	0.0004
2 vs 3	2.77	1	1.56	1.77	0.0970
2 vs 4	5.92	1	1.56	3.79	0.0018
3 vs 4	3.15	1	1.56	2.02	0.0621

We can also use the graphical procedure to compare mean yield at the four extrusion pressures.

4.1.2 Model Adequacy Checking

We have previously discussed the importance of checking the adequacy of the assumed model, that is we need to check the assumptions: Normality, unequal error variance by treatment or black, block-treatment interaction.

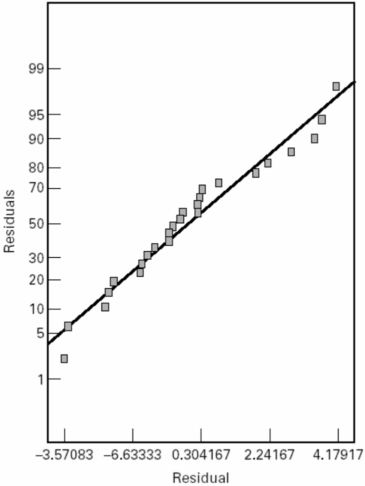

Figure 4-4 Normal probability plot of residuals for Example 4-1.

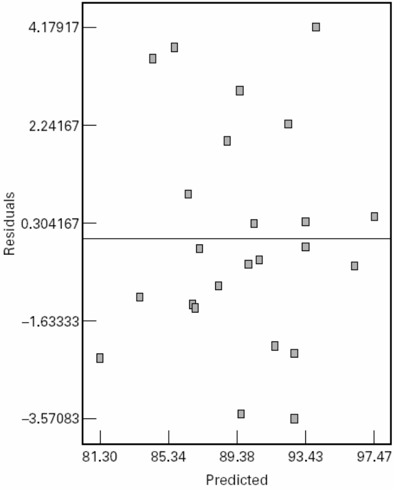

Figure 4-5 Plot of residuals versus $\hat{y}_{ij}$ for Example 4-1.

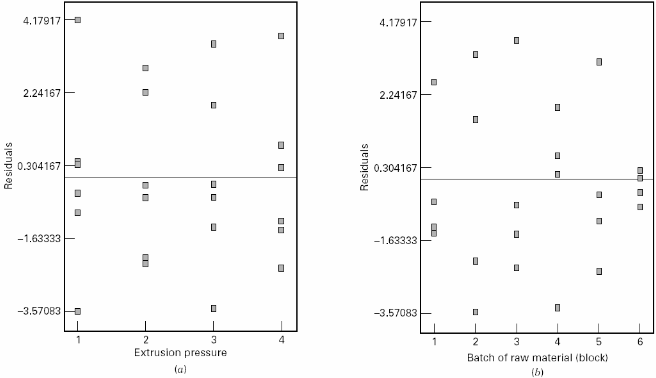

Figure 4-6 Plot of residuals by extrusion pressure (treatment) and by batches of resin (block) for Example 4-1.

4.1.3 Some Other Aspects of the Randomized Complete Block Design

Additivity of the Randomized Block Model The linear statistical model

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

is completely additive.
Similarly, interactions between treatments and blocks can occur when the response is measured on the wrong scale. Thus, a relationship that is multiplicative in the original units, say

$$E(y_{ij}) = \mu\tau_i\beta_j$$

is linear of additive in a log scale:

$$\ln E(y_{ij}) = \ln\mu + \ln\tau_i + \ln\beta_j$$

or

$$E(y_{ij}^*) = \mu^* + \tau_i^* + \beta_j^*$$

Treatments and/or blocks as random effects

Choice of Sample Size. The OC curve approach can be used to determine the number of blocks to run.

$$\Phi^2 = \frac{b\sum_{i=1}^{a}\tau_i^2}{a\sigma^2}$$

Estimating Missing Value When using the RCBD, sometimes an observation in one of the blocks is missing. A missing observation results in that the treatments are no longer orthogonal to blocks, that is, every treatment does not occur in every block.

Two approaches: One is an approximate analysis; other is an exact analysis. Suppose that the observation y_{ij} is missing. Denote it by x. In general, let $y'_{..}, y'_{i.}, y'_{.j}$ denote the corresponding quantities for the data with one missing value. Note that

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{1}{b}\sum_{i=1}^{a}\left(\sum_{j=1}^{b} y_{ij}\right)^2 - \frac{1}{a}\sum_{j=1}^{b}\left(\sum_{i=1}^{a} y_{ij}\right)^2 + \frac{1}{ab}\left(\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}\right)^2$$

or

$$SS_E = x^2 - \frac{1}{b}(y'_{I.} + x)^2 - \frac{1}{a}(y'_{.j} + x)^2 + \frac{1}{ab}(y'_{..} + x)^2 + R$$

Therefore we can obtain

$$x = \frac{ay'_{i.} + by'_{.j} - y'_{..}}{(a-1)(b-1)}$$

In table 4.3, we assume that the observation y_{24} is missed. We find that $y'_{2.} = 455.4, y'_{.4} = 267.5$, and $y'_{..} = 2060.4$ and thus,

$$x = y_{24} = \frac{4(455.4) + 6(267.5) - 2060.4}{(3)(5)} = 91.08.$$

Table 4.7 Approximate Analysis of Variance for Example 4.1 with One Missing Value

Source of Variation	Sum of Squares	Degree of Freedom	Mean Square	F_0	P-Value
Extrusion Pressure	166.14	3	55.38	7.63	0.0029
Batches of raw material	189.52	5	37.90		
Error	101.70	14	7.26		
Total	457.36	23			

4.1.4 Estimating Model Parameters and the General Significance Test

For the RCBD model,

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}, \qquad \begin{cases} i = 1, 2, ..., a \\ j = 1, 2, ..., b \end{cases}$$

we have the following normal equations:

$$\begin{array}{llr}
\mu: & ab\hat{\mu} + b\hat{\tau}_1 + b\hat{\tau}_2 + \cdots + b\hat{\tau}_a + a\hat{\beta}_1 + a\hat{\beta}_2 + \cdots + a\hat{\beta}_b = & y_{..} \\
\tau_1: & b\hat{\mu} + b\hat{\tau}_1 \qquad\qquad + \hat{\beta}_1 + \hat{\beta}_2 + \cdots + \hat{\beta}_b = & y_{1.} \\
\tau_2: & b\hat{\mu} + \quad b\hat{\tau}_2 \qquad + \hat{\beta}_1 + \hat{\beta}_2 + \cdots + \hat{\beta}_b = & y_{2.} \\
\tau_a: & b\hat{\mu} + \qquad b\hat{\tau}_a + \hat{\beta}_1 + \hat{\beta}_2 + \cdots + \hat{\beta}_b = & y_{2.} \\
\vdots & \qquad\qquad \vdots & \vdots \\
\beta_1: & \hat{\mu} + \hat{\tau}_1 + \hat{\tau}_2 + \cdots + \hat{\tau}_a + a\hat{\beta}_1 \qquad\qquad = & y_{.1} \\
\beta_2: & \hat{\mu} + \hat{\tau}_1 + \hat{\tau}_2 + \cdots + \hat{\tau}_a + \quad a\hat{\beta}_2 \qquad = & y_{.2} \\
\vdots & \qquad\qquad \vdots & \vdots \\
\beta_b: & \hat{\mu} + \hat{\tau}_1 + \hat{\tau}_2 + \cdots + \hat{\tau}_a + \qquad a\hat{\beta}_b = & y_{.b}
\end{array}$$

With two constraints

$$\sum_{i=1}^{a} \tau_i = 0, \qquad \sum_{j=1}^{b} \beta_j = 0$$

we have the least square estimates for the parameters as follows

$$\hat{\mu} = \bar{y}_{..}, \hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}, \hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}, \begin{cases} i = 1, 2, ..., a \\ j = 1, 2, ..., b \end{cases}$$

and

$$\hat{y}_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j = \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..}$$

The general regression significance test can be used to develop the analysis of variance for the RCBD. The reduction in the sum of squares for fitting the full model is

$$R(\mu, \tau, \beta) = \hat{\mu} + \sum_{i=1}^{a} \hat{\tau}_i y_{i.} + \sum_{j=1}^{b} \hat{\beta}_j y_{.j} = \sum_{i=1}^{a} \frac{y_{i.}^2}{b} + \sum_{j=1}^{b} \frac{y_{.j}^2}{a} - \frac{y_{..}^2}{ab}$$

with $a+b-1$ degrees of freedom, and the error sum of squares is

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - R(\mu,\tau,\beta) = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$$

with $(a-1)(b-1)$ degrees of freedom.

To test the hypothesis $H_0 : \tau_i = 0$, the reduce d model is

$$y_{ij} = \mu + \beta_j + \epsilon_{ij}$$

and thus the reduction for fitting this model is

$$R(\mu,\beta) = \sum_{j=1}^{b} \frac{y_{.j}^2}{a}$$

which has b degrees of freedom. Therefore, the sum of squares due to $\{\tau_i\}$ after fitting μ and $\{\beta_j\}$ is

$$R(\tau|\mu,\beta) = R(\mu,\tau,\beta) - R(\mu,\beta) = \sum_{i=1}^{a} \frac{y_{i.}^2}{b} - \frac{y_{..}^2}{ab}$$

with $a-1$ degrees of freedom.

4.2 The Latin Square Design Suppose that an experimenter is studying the effects of five different formulations of a rocket propellant used in aircrew escape systems on the observed burning rate. Each formulation is mixed from a batch of raw material that only large enough for five formulations to be tested. Further, the formulations are prepared by several operators, and there may be substantial differences in the skills and experience of the operators. Thus, it would seem that there are **two nuisance factors**: batches of raw material and operator. The appropriate design for this problem consists of testing each formulation exactly once in each batch of raw material and for each formulation to be prepared exactly once by each of five operators. The resulting design is called a **Latin Square Design.**

1. These designs are used to simultaneously control (or eliminate) two sources of nuisance variability
2. A significant assumption is that the three factors (treatments, nuisance factors) do not interact
3. If this assumption is violated, the Latin square design will not produce valid results
4. Latin squares are not used as much as the RCBD in industrial experimentation

Table 4-8 Latin Square Design for the Rocket Propellant Problem

Batches of Raw Material	Operators				
	1	2	3	4	5
1	$A = 24$	$B = 20$	$C = 19$	$D = 24$	$E = 24$
2	$B = 17$	$C = 24$	$D = 30$	$E = 27$	$A = 36$
3	$C = 18$	$D = 38$	$E = 26$	$A = 27$	$B = 21$
4	$D = 26$	$E = 31$	$A = 26$	$B = 23$	$C = 22$
5	$E = 22$	$A = 30$	$B = 20$	$C = 29$	$D = 31$

Page 140 shows some other Latin squares Table 4-13 (page 140) contains properties of Latin squares Statistical analysis?

- This is a 5×5 Latin Square design
- Page 139 shows some other Latin squares
- Table 4-12 (page 142) contains properties of Latin squares
- Statistical analysis?

The statistical (effects) model for a Latin square design is

$$y_{ijk} = \mu + \alpha_i + \tau_j + \beta_k + \epsilon_{ijk} \quad \begin{cases} i = 1, 2, ..., p \\ j = 1, 2, ..., p \\ k = 1, 2, ..., p \end{cases}$$

where τ_j is the jth treatment effect, α_i is the ith row effect, β_k is the kth effect. The statistical analysis (ANOVA) is much like the analysis for the RCBD.

$$\underbrace{SS_T}_{p^2-1} = \underbrace{SS_{Rows}}_{p-1} + \underbrace{SS_{Columns}}_{p-1} + \underbrace{SS_{Treatments}}_{p-1} + \underbrace{SS_E}_{(p-1)(p-2)}$$

Source of Variation	Sum of Squares	Degree of Freedom	Mean Squares	F_0
Treatments	$SS_{Treatments} = \frac{1}{p}\sum_{j=1}^{p} y_{.j.}^2 - \frac{y_{...}^2}{N}$	$p-1$	$\frac{SS_{treatment}}{p-1}$	$F_0 = \frac{MS_{treatme}}{MS_E}$
Rows	$SS_{Rows} = \frac{1}{p}\sum_{i=1}^{p} y_{i..}^2 - \frac{y_{...}^2}{N}$	$p-1$	$\frac{SS_{Rows}}{p-1}$	
Columns	$SS_{Columns} = \frac{1}{p}\sum_{k=1}^{p} y_{..k}^2 - \frac{y_{...}^2}{N}$	$p-1$	$\frac{SS_{Column}}{p-1}$	
Error	SS_E(by substraction)	(p-1)(p-2)	$\frac{SS_E}{(p-1)(p-2)}$	
Total	$SS_T = \sum_i \sum_j \sum_k y_{ijk}^2 - \frac{y_{...}^2}{N}$	$p^2 - 1$		

Example 4.3 Consider the rocket propellant problem. The design for this experiment is 5×5 Latin square. After coding by subtracting 25 from each observation, we have the data in Table 4.10.

Table 4.10 Coded Data for the Rocket Propellant Problem

Batches of	Operator					
Raw Material	1	2	3	4	5	$y_{i..}$
1	A=-1	B=-5	C=-6	D=-1	E=-1	-14
2	B=-8	C=-1	D=5	E=2	A=11	9
3	C=-7	D=13	E=1	A=2	B=-4	5
4	D=1	E=6	A=1	B=-2	C=-3	3
5	E=-3	A=5	B=-5	C=4	D=6	7
$y_{..k}$	-18	18	-4	5	9	$10 = y_{...}$

Latin Letter	A	B	C	D	E
Treatment Total	$y_{.1.} = 18$	$y_{.2.} = -24$	$y_{.3.} = -13$	$y_{.4.} = 24$	$y_{.5.} = 5$

Table 4.11 Analysis of Variance for the Rocket Propellant Experiment

Source of Variation	Sum of Squares	Degree of Freedom	Mean Square	F_0	P-value
Formulation	330.00	4	82.50	7.73	0.0023
Batches of raw material	68.00	4	17.00		
Operator	150.00	4	37.50		
Error	128.00	12	10.67		
Total	676.00	24			

As in any design problem, the experimenter should investigate the adequacy of the model by inspecting and plotting the residual, which is

$$e_{ijk} = y_{ijk} - \hat{y}_{ijk} = y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + 2\bar{y}_{...}$$

If one observation in a Latin square is missing, it is can be estimated by

$$y_{ijk} = \frac{p(y'_{i..} - y'_{.j.} - y'_{..k}) - 2y'_{...}}{(p-2)(p-1)}$$

Replication of Latin Squares A disadvantage of small Latin squares is that they provide a relatively small number of error degrees of freedom. When small Latin squares are used, it is frequently desirable to replicate them to increase the error degrees of freedom.

Crossover Designs and Designs Balanced for Residual Effects.

Occasionally, one encounters a problem in which time periods are a factor in the experiment. In general, there are p treatments to be tested in p time periods using np experimental units. For example, a human performance analysis is studying the effect of two replacement fluids on dehydration in 20 subjects. The experimenter has the subjects who took fluid A take fluid B and those who took fluid B take fluid A. This design is called a **crossover design**.

Latin Squares: A Crossover Design

	I		II		III		IV		V		VI		VII		VIII		IX	
S	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
P1	A	B	B	A	B	A	A	B	A	B	B	A	A	B	A	B	A	B
P2	B	A	A	B	A	B	B	A	B	A	A	B	B	A	B	A	B	A

Analysis of Variance for the Crossover Design

Source of Variation	Degrees of Freedom
Subjects (column)	19
Periods (rows)	1
Fluids (letters)	1
Error	18
Total	39

4.3 The Graeco-Latin Square Design

Consider a $p \times p$ Latin square, and superimpose on it a second $p \times p$ Latin square in which the treatments are denoted by Greek letters. If the two squares when superimposed have the property that each Greek letter appears once and only once with each Latin letter, the two Latin squares are said to be orthogonal, and the design obtained is called a Graeco-Latin square.

Table 4.17 4×4 **Graeco-Latin Square Design**

	Column			
Row	1	2	3	4
1	$A\alpha$	$B\beta$	$C\gamma$	$D\delta$
2	$B\delta$	$A\gamma$	$D\beta$	$C\alpha$
3	$C\beta$	$D\alpha$	$A\delta$	$B\gamma$
4	$D\gamma$	$C\delta$	$B\alpha$	$A\beta$

The statistical model for the Graeco-Latin square design is

$$y_{ijkl} = \mu + \beta_i + \tau_j + \omega_k + \Psi_l + \epsilon_{ijkl}, \begin{cases} i = 1, 2, ..., p \\ j = 1, 2, ..., p \\ k = 1, 2, ..., p \\ l = 1, 2, ..., p \end{cases}$$

where y_{ijkl} is the observation in row i, column l for Latin letter j and Greek letter k.

Table 4.18 Analysis of Variance for a Graeco-Latin Square Design

Source of Variation	Sum of Squares	Degrees of Freedom
Latin letter treatments	$SS_L = \frac{1}{p}\sum_{j=1}^{p} y_{.j..}^2 - \frac{y_{....}^2}{N}$	$p-1$
Greek letter treatments	$SS_G = \frac{1}{p}\sum_{k=1}^{p} y_{..k.}^2 - \frac{y_{....}^2}{N}$	$p-1$
Rows	$SS_{Rows} = \frac{1}{p}\sum_{i=1}^{p} y_{i...}^2 - \frac{y_{....}^2}{N}$	$p-1$
Column	$SS_{Columns} = \frac{1}{p}\sum_{l=1}^{p} y_{...l}^2 - \frac{y_{....}^2}{N}$	$p-1$
Error	SS_E	$(p-3)(p-1)$
Total	$SS_T = \sum_i \sum_j \sum_k \sum_l y_{ijkl}^2 - \frac{y_{....}^2}{N}$	p^2-1

Example 4.4 Suppose that in the rocket propellant experiment of Example 4.3 an additional factor, test assemblies, could be of importance. Let there be five test assemblies denoted by $\alpha, \beta, \gamma, \delta$ and ϵ.

Table 4.19 Graeco-Latin Square Design for the Rocket Propellant Problem

Batches of Raw Material	Operator 1	2	3	4	5	$y_{i...}$
1	$A\alpha = -1$	$B\gamma = -5$	$C\epsilon = -6$	$D\beta = -1$	$E\delta = -1$	-14
2	$B\beta = -8$	$C\delta = -1$	$D\alpha = 5$	$E\gamma = 2$	$A\epsilon = 11$	9
3	$C\gamma = -7$	$D\epsilon = 13$	$E\beta = 1$	$A\delta = 2$	$B\beta = -4$	5
4	$D\delta = 1$	$E\alpha = 6$	$A\gamma = 1$	$B\epsilon = -2$	$C\beta = -3$	3
5	$E\epsilon = -3$	$A\beta = 5$	$B\delta = -5$	$C\alpha = 4$	$D\gamma = 6$	7
$y_{...l}$	-18	18	-4	5	9	$y_{....} = 10$

Greek Letter	α	β	γ	δ	ϵ
Treatment Total	$y_{..1.} = 10$	$y_{..2.} = -6$	$y_{..3.} = -3$	$y_{..4.} = -4$	$y_{..5.} = 13$

Table 4.20 Analysis of Variance for the Rocket Propellant Problem

Source of Variation	Sum of Squares	Degree of Freedom	Mean Square	F_0	P-value
Formulation	330.00	4	82.50	10.00	0.0033
Batches of raw material	68.00	4	17.00		
Operator	150.00	4	37.50		
Test assemblies	62.00	4	15.50		
Error	66.00	8	8.25		
Total	676.00	24			

4.4 Balanced Incomplete Block Design In certain experiments using randomized block designs, we may not be able to run all the treatments combinations in each block. The randomized block design in which every treatments is not present in every block is called the randomized incomplete block design.

When all treatments comparisons are equally important, the treatment combinations used in each block should be selected in balanced manner. **A balanced incomplete block design (BIBD)** is an incomplete block design in which any two treatments appear together an equal number of times. Suppose that there are a treatment and each block can hold exactly $k(k < a)$ treatments.

4.4.1 Statistical Analysis of the BIBD

As usual, we assume that there are a treatments and b blocks and each block contains k treatments, that each treatment occurs r times in the design, and that there are $N = ar = bk$ total observations. Further, the number of times each pair of treatments appears in the same block is

$$\lambda = \frac{r(k-1)}{a-1}$$

If $a = b$, the design is said to be symmetric. Note that the parameter λ must be and integer.

The statistical model for the BIBD is

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon$$

The total variability in the data is expressed by the total corrected sum of squares:

$$SS_T = \sum_i \sum_j y_{ij}^2 - \frac{y_{..}^2}{N}$$

Total variability may be partitioned into

$$SS_T = SS_{treatments(adjusted)} + SS_{Blocks} + SS_E$$

where the sum of squares for treatments is adjusted to separate the treatment and the block effects. Note that

$$SS_{Blocks} = \frac{1}{k} \sum_{j=1}^{b} y_{.j}^2 - \frac{y_{..}^2}{N}$$

The adjusted treatment sum of squares is

$$SS_{Treatments(adjusted)} = \frac{k \sum_{i=1}^{a} Q_i^2}{\lambda a}$$

where Q_i is the adjusted total for the ith treatment, which is computed as

$$Q_i = y_{i.} - \frac{1}{k} \sum_{j=1}^{b} n_{ij} y_{.j}, \qquad i = 1, 2, ..., a$$

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

Note that $\sum_{i=1}^{a} Q_i = 0$. $SS_{Treatments(adjusted)}$ has $a-1$ degrees of freedom. And thus

$$SS_E = SS_T - SS_{Treatments(adjusted)} - SS_{Blocks}$$

and has $N-a-b+1$ degrees of freedom. The statistic for testing the equality of the treatment effects is

$$F_0 = \frac{MS_{Treatments(adjusted)}}{MS_E}$$

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Treatment (adjusted)	$\frac{k\sum_{i=1}^{a} Q_i^2}{\lambda a}$	$a-1$	$\frac{SS_{Treatments(adjusted)}}{a-1}$	$\frac{MS_{Treatments(}}{MS_E}$
Blocks	$\frac{1}{k}\sum_{j=1}^{b} y_{.j}^2 - \frac{y_{..}^2}{N}$	$b-1$	$\frac{SS_{Blocks}}{b-1}$	
Error	SS_E	$N-a-b+1$	$\frac{SS_E}{N-a-b+1}$	
Total	$\sum\sum y_{ij}^2 - \frac{y_{ij}^2}{N}$	$N-1$		

Example (4.5)

Consider the following data:

Table 4.21 Balanced Incomplete Block Design for Catalyst Experiment

Treatment Catalyst)	Block(Batch of Raw Material 1	2	3	4	$y_{i.}$
1	73	74	–	71	218
2	–	75	67	72	214
3	73	75	68	–	216
4	75	–	72	75	222
$y_{.j}$	221	224	207	218	$y_{..} = 870$

This is a BIBD with $a = 4, b = 4, k = 3, r = 3, \lambda = 2$ and $N = 12$. Thus, we have

$$SS_T = \sum_i \sum_j y_{ij}^2 - \frac{y_{..}^2}{12} = 81.00, SS_{Blocks} = \frac{1}{3}\sum_{j=1}^{4} y_{.j}^2 - \frac{y_{..}^2}{12} = 55.00$$

Further we have

$$Q_1 = (218) - \frac{1}{3}(221 + 224 + 218) = -9/3, Q_2 = -7/3, Q_3 = -4/3, Q_4 = 20/3$$

$$SS_{Treatments(adjusted)} = \frac{3[(-9/3)^2 + (-7/3)^2 + (-4/3)^2 + (20/3)^2]}{(2)(4)} = 22.75$$

Now we have $SS_E = 81.00 - 55.00 - 22.75 = 4.25$ and $F_0 = \frac{22.75/3}{3.25/5} = 11.66$ and p-value is 0.0107.

If orthogonal contrasts are employed, the contrasts must be made on the adjusted treatment totals, the $\{Q_i\}$ rather than the $\{y_{i.}\}$. The contrast sum of squares is

$$SS_C = \frac{k\left(\sum_{i=1}^{a} c_i Q_i\right)^2}{\lambda a \sum_{i=1}^{a} c_i^2}$$

The standard error of an adjusted treatment effect is

$$S = \sqrt{\frac{kMS_E}{\lambda a}}$$

Sometimes we would like to test the block effect. To do this, SS_T can be partitioned as

$$SS_T = SS_{Treatments} + SS_{Blocks(adjusted)} + SS_E$$

and

$$Q'_j = y_{.j} - \frac{1}{r}\sum_{i=1}^{a} n_{ij} y_{i.}, j = 1, 2, ..., b$$

$$SS_{Blocks(adjusted)} = \frac{r\sum_{j=1}^{b}(Q'_j)^2}{\lambda b}$$

The BIBD in Example 4.5 is symmetric. Therefore,

$$Q_1' = (221) - \frac{1}{3}(218 + 216 + 222) = 7/3, Q_2' = 24/3, Q_3' = -31/3, Q_4' = 0$$

and

$$SS_{Blocks(adjusted)} = \frac{2[(7/3)^2 + (24/3)^2 + (31/3)^2 + (0)^2]}{(2)(4)} = 66.08$$

Also, $SS_{Treatments} = 11.67$ The statistic for testing the equality of the block effects is

$$F_0 = \frac{MS_{Blocks(adjusted)}}{MS_E} = \frac{66.08/3}{3.25/5} = 33.90$$

and p-value is 0.0010. Notice that the sums of squares associated with the mean squares do not add to the total sum of squares, that is,

$$SS_T \neq SS_{Treatments(adjusted)} + SS_{Blocks(adjusted)} + SS_E$$

This is a consequence of the nonorthogonality of treatments and blocks.

4.4.2 Least Squares Estimation of the Parameters Consider the estimation of the parameters in the BIBD model. The least squares normal equations are

$$\mu: \quad N\hat{\mu} + r\sum_{i=1}^{a}\hat{\tau}_i + k\sum_{j=1}^{b}\hat{\beta}_j = y_{..}$$

$$\tau_i: \quad r\hat{\mu} + r\hat{\tau}_i + \sum_{j=1}^{b} n_{ij}\hat{\beta}_j = y_{i.}, \quad i = 1, 2, ..., a$$

$$\beta_j: \quad k\hat{\mu} + \sum_{i=1}^{a} n_{ij}\hat{\tau}_i + k\hat{\beta}_j = y_{.j}, \quad j = 1, 2, ..., b$$

Imposing $\sum \hat{\tau}_i = \sum \hat{\beta}_j = 0$, we have $\hat{\mu} = \bar{y}_{..}$. Further, we have

$$rk\hat{\tau}_i - r\hat{\tau}_i - \sum_{j=1}^{b}\sum_{p\neq i}^{a} n_{ij}n_{pj}\hat{\tau}_p = ky_{i.} - \sum_{j=1}^{b} n_{ij}y_{.j}$$

After simplification, we have

$$\lambda a\hat{\tau}_i = kQ_i, i = 1, 2..., a$$

and thus,

$$\hat{\tau}_i = \frac{kQ_i}{\lambda a}, \quad i = 1, 2, ..., a$$