3 RANDOMIZED COMPLETE BLOCK DESIGN (RCBD)

- The experimenter is concerned with studying the effects of a single factor on a response of interest. However, variability from another factor that is not of interest is expected.
- The goal is to control the effects of a variable not of interest by bringing experimental units that are similar into a group called a “block”. The treatments are then randomly applied to the experimental units within each block. The experimental units are assumed to be homogeneous within each block.
- By using blocks to control a source of variability, the mean square error (MSE) will be reduced. A smaller MSE makes it easier to detect significant results for the factor of interest.
- Assume there are \(a\) treatments and \(b\) blocks. If we have one observation per treatment within each block, and if treatments are randomized to the experimental units within each block, then we have a randomized complete block design (RCBD). Because randomization only occurs within blocks, this is an example of restricted randomization.

3.1 RCBD Notation

- Assume \(\mu\) is the baseline mean, \(\tau_i\) is the \(i^{th}\) treatment effect, \(\beta_j\) is the \(j^{th}\) block effect, and \(\epsilon_{ij}\) is the random error of the observation. The statistical model for a RCBD is

\[
y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \quad \text{and} \quad \epsilon_{ij} \sim IIDN(0, \sigma^2). \tag{6}
\]

- \(\mu, \tau_i (i = 1, 2, \ldots, a)\), and \(\beta_j (j = 1, 2, \ldots, b)\) are not uniquely estimable. Constraints must be imposed. To be able to calculate estimates \(\hat{\mu}, \hat{\tau}_i, \text{and} \hat{\beta}_j\), we need to impose two constraints.
- Initially, we will assume the textbook constraints: \(\sum_{i=1}^{a} \tau_i = 0\) and \(\sum_{j=1}^{b} \beta_j = 0\).
- These are not the default SAS constraints (\(\tau_a = 0, \beta_b = 0\)) or R constraints (\(\tau_1 = 0, \beta_1 = 0\)).

- Applying these constraints, will yield least-squares estimates

\[
\hat{\mu} = \hat{\tau}_i = \hat{\beta}_j
\]

where \(\bar{y}_i\) is the mean for treatment \(i\), and \(\bar{y}_j\) is the mean for block \(j\).

- Substitution of the estimates into the model yields:

\[
y_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j + \epsilon_{ij}
\]

where \(e_{ij} = \tilde{e}_{ij}\) is the residual of an observation \(y_{ij}\) from a RCBD. The value of \(e_{ij}\) is

\[
e_{ij} = y_{ij} - (\bar{y}_i - \bar{y}_j) - (\bar{y}_j - \bar{y}_\cdot) - \bar{y}_\cdot =
\]

- The total sum of squares (\(SS_{total}\)) for the RCBD is partitioned into 3 components:

\[
\sum_{i=1}^{a} \sum_{j=1}^{b} (y_{ij} - \bar{y}_\cdot)^2 = \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{y}_i - \bar{y}_\cdot)^2 + \sum_{j=1}^{b} \sum_{i=1}^{a} (\bar{y}_j - \bar{y}_\cdot)^2 + \sum_{i=1}^{a} \sum_{j=1}^{b} (y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}_\cdot)^2
\]

\[
= b \sum_{i=1}^{a} (\bar{y}_i - \bar{y}_\cdot)^2 + a \sum_{j=1}^{b} (\bar{y}_j - \bar{y}_\cdot)^2 + \sum_{i=1}^{a} \sum_{j=1}^{b} (y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}_\cdot)^2
\]

\[
= b \sum_{i=1}^{a} + a \sum_{j=1}^{b} + \sum_{i=1}^{a} \sum_{j=1}^{b}
\]

OR \(SS_{Total} = SS_{Trt} + SS_{Block} + SS_E\)
Alternate formulas to calculate $SS_{Total}$, $SS_{Trt}$ and $SS_{Block}$.

$$SS_{Total} = \sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij}^2 - \frac{y^2}{ab}$$

$$SS_{Trt} = \sum_{i=1}^{a} \frac{y_{i}^2}{b} - \frac{y^2}{ab}$$

$$SS_{Block} = \sum_{j=1}^{b} \frac{y_{j}^2}{a} - \frac{y^2}{ab}$$

$$SS_E = SS_{Total} - SS_{Trt} - SS_{Block}$$

where $\frac{y^2}{ab}$ is the correction factor.

### 3.2 Cotton Fiber Breaking Strength Experiment

An agricultural experiment considered the effects of $K_2O$ (potash) on the breaking strength of cotton fibers. Five $K_2O$ levels were used (36, 54, 72, 108, 144 lbs/acre). A sample of cotton was taken from each plot, and a strength measurement was taken. The experiment was arranged in 3 blocks of 5 plots each.

<table>
<thead>
<tr>
<th>Block</th>
<th>36</th>
<th>54</th>
<th>72</th>
<th>108</th>
<th>144</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.62</td>
<td>8.14</td>
<td>7.76</td>
<td>7.17</td>
<td>7.46</td>
<td>$y_1 = 38.15$</td>
</tr>
<tr>
<td>2</td>
<td>8.00</td>
<td>8.15</td>
<td>7.73</td>
<td>7.57</td>
<td>7.68</td>
<td>$y_2 = 39.13$</td>
</tr>
<tr>
<td>3</td>
<td>7.93</td>
<td>7.87</td>
<td>7.74</td>
<td>7.80</td>
<td>7.21</td>
<td>$y_3 = 38.55$</td>
</tr>
<tr>
<td>Totals</td>
<td>23.55</td>
<td>24.16</td>
<td>23.23</td>
<td>22.54</td>
<td>22.35</td>
<td>$y_\cdot = 115.83$</td>
</tr>
</tbody>
</table>

Treatment Means

- $\bar{y}_1 = 7.850$
- $\bar{y}_2 = 8.053$
- $\bar{y}_3 = 7.743$
- $\bar{y}_4 = 7.513$
- $\bar{y}_5 = 7.450$

Block Means

- $\bar{y}_1 = 7.630$
- $\bar{y}_2 = 7.826$
- $\bar{y}_3 = 7.710$

Grand Mean

- $\bar{y} = 7.723$

Uncorrected Sum of Squares = $\sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij}^2 = 2685.5151$

Correction factor = $\frac{y^2}{ab} = 115.83^2/15$

$$\sum_{i=1}^{a} \frac{y_{i}^2}{b} = \frac{23.55^2 + 24.16^2 + 23.23^2 + 22.54^2 + 22.35^2}{3} = \frac{2685.5151}{3} =$$

$$\sum_{j=1}^{b} \frac{y_{j}^2}{a} = \frac{38.15^2 + 39.13^2 + 38.55^2}{5} = \frac{4472.6815}{5} =$$

$SS_{Total} = 895.6183 - 894.4393 =

$SS_{Trt} = 895.1717 - 894.4393 =

$SS_{Block} = 894.5364 - 894.4393 =

$SS_E = 1.1790 - 0.7324 - 0.0971 =

### Analysis of Variance (ANOVA) Table

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>d.f.</th>
<th>Mean Square</th>
<th>F Ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_2O$ lbs/acre</td>
<td>.18311</td>
<td></td>
<td>.0404</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blocks</td>
<td>.04856</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>.043685</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Test the hypotheses $H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = \tau_5 = 0$ versus $H_1 : \tau_i \neq 0$ for some $i$.

- The test statistic is $F_0 = 4.1916$.
- The reference distribution is $F(a-1, (a-1)(b-1)) = F(4, 8)$.
- The critical value is $F_{.05}(4, 8) = \ldots$
- The decision rule is to reject $H_0$ if the test statistic $F_0$ is greater than $F_{.05}(4, 8)$.

Is $F_0 > F_{.05}(4, 8)$? Is \( \ldots \)

- The conclusion is to reject $H_0$ and conclude that

SAS Output for the RCBD Example

**ANOVA RESULTS FOR STRENGTH BY TREATMENT**

*The GLM Procedure*

**Dependent Variable: strength**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>6</td>
<td>0.82956000</td>
<td>0.13826000</td>
<td>3.16</td>
<td>0.0677</td>
</tr>
<tr>
<td>Error</td>
<td>8</td>
<td>0.34948000</td>
<td>0.04368500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>14</td>
<td>1.17904000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R-Square</th>
<th>Coeff Var</th>
<th>Root MSE</th>
<th>strength Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.703589</td>
<td>2.706677</td>
<td>0.209010</td>
<td>7.722000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>k2O</td>
<td>4</td>
<td>0.73244000</td>
<td>0.18311000</td>
<td>4.19</td>
<td>0.0404</td>
</tr>
<tr>
<td>block</td>
<td>2</td>
<td>0.09712000</td>
<td>0.04856000</td>
<td>1.11</td>
<td>0.3750</td>
</tr>
</tbody>
</table>

| Parameter | Estimate | Standard Error | t Value | Pr > |t| |
|-----------|----------|----------------|---------|------|---|
| Intercept | 7.43800000000 | 0.14278072 | 52.09 | <.0001 |
| k2O 36 | 0.40000000000 | 0.17065560 | 2.34 | 0.0471 |
| k2O 54 | 0.60333333333 | 0.17065560 | 3.54 | 0.0077 |
| k2O 72 | 0.29333333333 | 0.17065560 | 1.72 | 0.1240 |
| k2O 108 | 0.06333333333 | 0.17065560 | 0.37 | 0.7202 |
| k2O 144 | 0.00000000000 |  |  |  |
| block 1 | -0.08000000000 | 0.13218926 | -0.61 | 0.5618 |
| block 2 | 0.11600000000 | 0.13218926 | 0.88 | 0.4058 |
| block 3 | 0.00000000000 |  |  |  |

**Note:** The XX matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter ‘B’ are not uniquely estimable.
3.3 SAS Code for Cotton Fiber Breaking Strength RCBD

```sas
DM 'LOG'; CLEAR; OUT; CLEAR;
OPTIONS NODATE NONUMBER LS=76;

ODS GRAPHICS ON;
ODS PRINTER PDF file='C:\COURSES\ST541\RCBD.PDF';

******************************************;
*** A RANDOMIZED COMPLETE BLOCK DESIGN ***;
******************************************;

DATA in; INPUT k2O block strength @@;
CARDS;
36 1 7.62 36 2 8.00 36 3 7.93
54 1 8.14 54 2 8.15 54 3 7.87
72 1 7.76 72 2 7.73 72 3 7.74
108 1 7.17 108 2 7.57 108 3 7.80
144 1 7.46 144 2 7.68 144 3 7.21

PROC GLM DATA=in PLOTS = (ALL);
CLASS k2O block;
MODEL strength = k2O block / SS3 SOLUTION;
MEANS block;
MEANS k2O / TUKEY CLDIFF LINES;
ESTIMATE 'K2O=36' K2O 4 -1 -1 -1 -1 / DIVISOR=5;
ESTIMATE 'K2O=54' K2O -1 4 -1 -1 -1 / DIVISOR=5;
ESTIMATE 'K2O=72' K2O -1 -1 4 -1 -1 / DIVISOR=5;
ESTIMATE 'K2O=108' K2O -1 -1 -1 4 -1 / DIVISOR=5;
ESTIMATE 'K2O=144' K2O -1 -1 -1 -1 4 / DIVISOR=5;
TITLE 'ANOVA RESULTS FOR STRENGTH BY TREATMENT';
RUN;
```

### Output

#### Comparisons significant at the 0.05 level are indicated by ***.

<table>
<thead>
<tr>
<th>k2O Comparison</th>
<th>Difference Between Means</th>
<th>Simultaneous 95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>54 - 36</td>
<td>0.2033</td>
<td>0.3862 0.7929</td>
</tr>
<tr>
<td>54 - 72</td>
<td>0.3100</td>
<td>-0.2796 0.8996</td>
</tr>
<tr>
<td>54 - 108</td>
<td>0.5400</td>
<td>-0.0496 1.1296</td>
</tr>
<tr>
<td>54 - 144</td>
<td>0.6033</td>
<td>0.0138 1.1929 ***</td>
</tr>
<tr>
<td>36 - 54</td>
<td>-0.2033</td>
<td>-0.7929 0.3862</td>
</tr>
<tr>
<td>36 - 72</td>
<td>0.1067</td>
<td>-0.4829 0.6962</td>
</tr>
<tr>
<td>36 - 108</td>
<td>0.3367</td>
<td>-0.2529 0.9262</td>
</tr>
<tr>
<td>36 - 144</td>
<td>0.4000</td>
<td>-0.1896 0.9896</td>
</tr>
<tr>
<td>72 - 54</td>
<td>-0.3100</td>
<td>-0.8996 0.2796</td>
</tr>
<tr>
<td>72 - 72</td>
<td>0.2300</td>
<td>-0.3196 0.8196</td>
</tr>
<tr>
<td>72 - 108</td>
<td>0.2933</td>
<td>-0.2962 0.8829</td>
</tr>
<tr>
<td>72 - 144</td>
<td>0.5400</td>
<td>-1.1296 0.0496</td>
</tr>
<tr>
<td>108 - 54</td>
<td>-0.3367</td>
<td>-0.9262 0.2529</td>
</tr>
<tr>
<td>108 - 72</td>
<td>-0.2300</td>
<td>-0.8196 0.3596</td>
</tr>
<tr>
<td>108 - 144</td>
<td>0.0633</td>
<td>-0.5262 0.6529</td>
</tr>
<tr>
<td>144 - 54</td>
<td>0.6033</td>
<td>-1.1929 0.0138 ***</td>
</tr>
<tr>
<td>144 - 36</td>
<td>-0.4000</td>
<td>-0.9896 0.1896</td>
</tr>
<tr>
<td>144 - 72</td>
<td>-0.2933</td>
<td>-0.8829 0.2962</td>
</tr>
<tr>
<td>144 - 108</td>
<td>-0.0633</td>
<td>-0.6529 0.5262</td>
</tr>
</tbody>
</table>

#### Means with the same letter are not significantly different.

<table>
<thead>
<tr>
<th>Tukey Grouping</th>
<th>Mean</th>
<th>N</th>
<th>k2O</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>8.0533</td>
<td>3</td>
<td>54</td>
</tr>
<tr>
<td>A</td>
<td>7.8500</td>
<td>3</td>
<td>36</td>
</tr>
<tr>
<td>A</td>
<td>7.7433</td>
<td>3</td>
<td>72</td>
</tr>
<tr>
<td>A</td>
<td>7.5133</td>
<td>3</td>
<td>108</td>
</tr>
<tr>
<td>B</td>
<td>7.4500</td>
<td>3</td>
<td>144</td>
</tr>
</tbody>
</table>
3.4 Restrictions on Randomization

- Two common reasons for blocking:
  1. The experimenter has multiple sets of experimental units that are homogeneous within sets but are heterogeneous across sets. This typically occurs when there is not a sufficient number of homogeneous experimental units available to run a CRD leading the experimenter to form groups of units that are as homogeneous as possible.
  2. The experimenter has time constraints that do not allow a CRD to be run within a continuous period of time that would ensure uniformity of experimental conditions. Under these circumstances, blocks take the form of a time unit (such as a day).

- For a RCBD, there is one restriction on randomization. Randomization is restricted to randomly assigning the treatments to the experimental units within each block.

- In their Design of Experiments text, Anderson and McLean (A&M) introduce a random component called a restriction error into the traditional RCBD model to present a more realistic picture of the experimental situation. This approach will be useful later when we have multiple restrictions on randomizations (e.g., split-plot designs).

- Essentially, we’re saying there must be a different error structure between a completely randomized design and a design that has a restriction on randomization. And, because there is a different error structure, there must be differences in the model and the analysis.

- Thus, A&M suggest that the traditional model

\[ y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \]  

should include a term indicating where the restriction on randomization occurred. That is:

\[ y_{ijk} = \mu + \tau_i + \beta_j + \delta_j + \epsilon_{ij} \]  

where \( \mu, \tau_i, \) and \( \beta_j \) are the same in (8) as in (7), \( y_{ij} \) is the response from the \( i \)th treatment in block \( j \) for the \( k \)th randomization, and \( \delta_j \) is the restriction error associated with the \( j \)th block.

- We also assume \( \delta_j \sim N(0, \sigma_\delta^2) \), and each \( \delta_j \) is completely confounded with the \( j \)th block effect.

### Comparison of CRD and RCBD ANOVA Tables

<table>
<thead>
<tr>
<th>Source</th>
<th>CRD with 2 model effects</th>
<th>EMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>( \beta_j )</td>
<td>( b - 1 )</td>
</tr>
<tr>
<td>Treatments</td>
<td>( \tau_i )</td>
<td>( a - 1 )</td>
</tr>
<tr>
<td>Error</td>
<td>( \epsilon_{ij} )</td>
<td>( (a - 1)(b - 1) )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>RCBD from A&amp;M</th>
<th>EMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>( \beta_j )</td>
<td>( b - 1 )</td>
</tr>
<tr>
<td>Restriction Error</td>
<td>( \delta_j(k) )</td>
<td>0</td>
</tr>
<tr>
<td>Treatments</td>
<td>( \tau_i )</td>
<td>( a - 1 )</td>
</tr>
<tr>
<td>Error</td>
<td>( \epsilon_{ijk} )</td>
<td>( (a - 1)(b - 1) )</td>
</tr>
</tbody>
</table>

where \( \phi(\beta) \) is a function of \( \beta_1, \ldots, \beta_b \) if blocks are fixed or \( \phi(\beta) = \sigma_\delta^2 \) if blocks are random.
In both the fixed and random block cases, the ANOVA F-tests associated with treatment effects are identical. You use 
\[ F_0 = \frac{MS_{trt}}{MS_E} \] to test 
\[ H_0 : \tau_1 = \cdots = \tau_a = 0 \quad \text{against} \quad H_1 : \text{not all of the } \tau_i \text{s are equal} \] (9)

The EMS for the RCBD indicates that the correct denominator EMS for testing for a significant block effect (either fixed or random) is the EMS for the restriction error. The problem is that this is not estimable from the data.

Under these circumstances, the test of the hypothesis involving the combination of the block effects and the restriction error in (10) would be appropriate to test for a ‘general’ blocking effect.

The statistic \( F = \frac{MS_{blocks}}{MS_E} \) is actually a test of

\[ H_0 : \sigma^2_\delta + \phi(\beta) = 0 \quad \text{against} \quad H_1 : \sigma^2_\delta + \phi(\beta) \neq 0 \] (10)

Note that even if \( \beta_1 = \beta_2 = \cdots = \beta_b = 0 \) (fixed) \( \text{or} \) \( \sigma^2_\beta = 0 \) (random) is true, we still have the restriction error in the EMS which prevents it from matching the error EMS = \( \sigma^2 \).

Because of the restriction on randomization, A&M claim that there is no F test for blocks. That is, there is no test for \( H_0 : \sigma^2_\beta = 0 \) if blocks are random and no test for \( H_0 : \beta_1 = \beta_2 = \cdots = \beta_b \) if blocks are fixed.

Fortunately this is not a problem because most of the time the experimenter is only interested in whether or not blocking had been effective in reducing the \( MS_E \) for improved testing of the effects of the treatment of interest.

### 3.5 Example of an Analysis With and Without Blocks

Three different disinfecting solutions are being compared to study their effectiveness in stopping the growth of bacteria in milk containers. The analysis is done in a laboratory, and only three trials can be run on any day. Because days could represent a potential source of variability, the experimenter decides to use a randomized block design with days as blocks. Observations are taken for four days. The inside of the milk containers are covered with a certain amount of bacteria. The response is the percentage of bacteria remaining after rinsing the container with a disinfecting solution.

<table>
<thead>
<tr>
<th>Day</th>
<th>Solution 1</th>
<th>Solution 2</th>
<th>Solution 3</th>
<th>Solution 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13</td>
<td>22</td>
<td>18</td>
<td>39</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>24</td>
<td>17</td>
<td>44</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>22</td>
</tr>
</tbody>
</table>

The data were analyzed assuming two different models. The first model does not include blocks. The second model includes blocks. The SAS analysis for both models is on the next page. Here are important results:
Without blocks  With blocks

\begin{align*}
R^2 \\
MSE \\
p-value
\end{align*}

- Note that we would fail to reject $H_0$ if blocks were not in the model because there is large variability across blocks ($MS_{day} = 368.97$).

- If the $SS_{day} = 1106.92$ and $df_{day} = 3$ is pooled with the the $SS_E = 41.83$ and $df_E = 6$ in the model with days (blocks), then it forms the $SS_E = 1158.75$ and $df_E = 9$ for the model without days (blocks).

SAS Code for RCBD Analyses With and Without Blocks

```sas
DM 'LOG;CLEAR;OUT;CLEAR';
ODS GRAPHICS ON;
* ODS PRINTER PDF file='C:\COURSES\ST541\RCBD2.PDF';
OPTIONS NODATE NONUMBER LS=76 PS=54;

*********************************************;
*** RCBD ANALYSES WITH AND WITHOUT BLOCKS ***;
*********************************************;
DATA IN;
DO solution = 1 TO 3;
DO day = 1 TO 4;
  INPUT growth @@; OUTPUT;
END; END;
LINES;
13 22 18 39 16 24 17 44 5 4 1 22
;
*********************************************;
*** RUN AN ANOVA WITH SOLUTION ONLY, NO DAY BLOCKS ***;
*********************************************;
PROC GLM DATA=IN;
  CLASS solution;
  MODEL growth = solution / ss3;
TITLE 'RCBD WITHOUT DAYS (BLOCKS) IN THE MODEL';

*********************************************;
*** RUN AN ANOVA WITH DAYS AS BLOCKS ***;
*********************************************;
PROC GLM DATA=IN;
  CLASS day solution;
  MODEL growth = solution day / ss3;
TITLE 'RCBD WITH DAYS (BLOCKS) IN THE MODEL';
RUN;
```
RCBD Without Days as Blocks

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2</td>
<td>703.500000</td>
<td>351.750000</td>
<td>2.73</td>
<td>0.1182</td>
</tr>
<tr>
<td>Error</td>
<td>9</td>
<td>1158.750000</td>
<td>128.750000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>11</td>
<td>1862.250000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square Coeff Var Root MSE growth Mean
0.377769 60.51630 11.34681 18.75000

Source    DF     Type III SS Mean Square F Value Pr > F
solution  2  703.500000 351.750000 2.73   0.1182

RCBD Without Days as Blocks

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>5</td>
<td>1810.416667</td>
<td>362.083333</td>
<td>41.91</td>
<td>0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>6</td>
<td>51.833333</td>
<td>8.638889</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>11</td>
<td>1862.250000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square Coeff Var Root MSE growth Mean
0.972166 15.67573 2.939199 18.75000

Source    DF    Type III SS Mean Square F Value Pr > F
solution  2  703.500000 351.750000 40.72  0.0003
day       3  1106.916667 368.972222 42.71  0.0002
3.6 Type I vs Type III Analyses

- Without the /ss3 option in the MODEL statement, SAS will contain two ANOVA tables: ANOVA for Type I sum of squares and ANOVA for Type III sum of squares.

- If there are no missing observations, the Type I and Type III analyses are identical.

- If there are missing observations, the Type I and Type III analyses are different. To see how they differ we will first look at the Type I analysis.

3.6.1 Type I Analysis

- The Type I analysis is based on sequentially fitting the data to the model one factor at a time. It is often referred to as the **sequential sum of squares method**.

- For the RCBD there are two possibilities that I will refer to as
  - Version 1 (V1) when fitting treatments before blocks.
  - Version 2 (V2) when fitting blocks before treatments.

- Let $RSS_i$ be the error sum of squares ($SS_E$) after fitting the model in the $i^{th}$ step.

- The steps for determining the ANOVA SS for V1 are:
  1. Fit $y_{ij} = \mu + \epsilon_{ij}$ and obtain $RSS_1 = SS_{total}$.
  2. Fit $y_{ij} = \mu + \tau_i + \epsilon_{ij}$ and obtain $RSS_2 = SS_E$ for the model with treatments only.
  3. Fit $y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$ and obtain $RSS_3 = SS_E$ for the model with treatments and blocks.

- The steps for determining the ANOVA SS for V2 are:
  1. Fit $y_{ij} = \mu + \epsilon_{ij}$ and obtain $RSS_1 = SS_{total}$.
  2'. Fit $y_{ij} = \mu + \beta_j + \epsilon_{ij}$ and obtain $RSS_2^* = SS_E$ for the model with blocks only.
  3. Fit $y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$ and obtain $RSS_3 = SS_E$ for the model with blocks and treatments.

- The ANOVA sum of squares for V1 and V2 are summarized in the following table:

<table>
<thead>
<tr>
<th>Step</th>
<th>V1 Source</th>
<th>Fit</th>
<th>df</th>
<th>Type I SS for V1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Total</td>
<td>$\mu$</td>
<td>$N - 1$</td>
<td>$RSS_1$</td>
</tr>
<tr>
<td>2</td>
<td>Treatment</td>
<td>$\tau_i$</td>
<td>$a - 1$</td>
<td>$R(\tau</td>
</tr>
<tr>
<td>3</td>
<td>Blocks</td>
<td>$\beta_j$</td>
<td>$b - 1$</td>
<td>$R(\beta</td>
</tr>
<tr>
<td>3</td>
<td>Error</td>
<td>$\epsilon_{ij}$</td>
<td>$N - a - b + 1$</td>
<td>$RSS_3$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step</th>
<th>V2 Source</th>
<th>Fit</th>
<th>df</th>
<th>Type I SS for V2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Total</td>
<td>$\mu$</td>
<td>$N - 1$</td>
<td>$RSS_1$</td>
</tr>
<tr>
<td>2'</td>
<td>Blocks</td>
<td>$\beta_j$</td>
<td>$b - 1$</td>
<td>$R(\beta</td>
</tr>
<tr>
<td>3</td>
<td>Treatment</td>
<td>$\tau_i$</td>
<td>$a - 1$</td>
<td>$R(\tau</td>
</tr>
<tr>
<td>3</td>
<td>Error</td>
<td>$\epsilon_{ij}$</td>
<td>$N - a - b + 1$</td>
<td>$RSS_3$</td>
</tr>
</tbody>
</table>
• In V1, the quantity \( R(\tau|\mu) \) is called the **reduction in SS due to** \( \tau \) **adjusted for** \( \mu \) and \( R(\beta|\tau,\mu) \) is called the **reduction in SS for** \( \beta \) **adjusted for** \( \tau \) and \( \mu \).

• In V2, the quantity \( R(\beta|\mu) \) is called the **reduction in SS due to** \( \beta \) **adjusted for** \( \mu \) and \( R(\tau|\beta,\mu) \) is called the **reduction in SS for** \( \tau \) **adjusted for** \( \beta \) and \( \mu \).

### 3.6.2 Type III Analysis

• The Type III analysis is referred to as the **marginal means** or the **Yates weighted squares of means** analysis.

• For a RCBD, the Type III \( SS_{trt} \) and \( SS_{blocks} \) are computed using the following procedure:

  1. Fit the model with treatments only: \( y_{ij} = \mu + \tau_i + \epsilon_{ij} \). Then \( RSS_2 = SS_E \) for this model.
  2. Fit the model with blocks only: \( y_{ij} = \mu + \beta_j + \epsilon_{ij} \). Then \( RSS^*_2 = SS_E \) for this model.
  3. Fit the model \( y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \). Then \( RSS_3 = SS_E \) and \( RSS_1 = SS_{total} \) for the model with both treatments and blocks.

<table>
<thead>
<tr>
<th>Step</th>
<th>Source</th>
<th>Fit</th>
<th>df</th>
<th>Type III SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Total</td>
<td>( \mu )</td>
<td>( N - 1 )</td>
<td>( RSS_1 )</td>
</tr>
<tr>
<td>2</td>
<td>Treatment</td>
<td>( \tau_i )</td>
<td>( a - 1 )</td>
<td>( R(\tau</td>
</tr>
<tr>
<td>3</td>
<td>Blocks</td>
<td>( \beta_j )</td>
<td>( b - 1 )</td>
<td>( R(\beta</td>
</tr>
<tr>
<td>1</td>
<td>Error</td>
<td>( \epsilon_{ij} )</td>
<td>( N - a - b + 1 )</td>
<td>( RSS_3 )</td>
</tr>
</tbody>
</table>

• If any \( y_{ij} \) values are missing, then \( SS_{trt} + SS_{blocks} + SS_E \neq SS_{total} \) for a Type III analysis.

### 3.6.3 RCBD Analysis with a Missing Observation

See the example in Section 3.3 for the description of the experiment. Suppose \( y_{23} \) was missing from the RCBD. The RCBD data table is:

<table>
<thead>
<tr>
<th>Solution</th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13</td>
<td>22</td>
<td>18</td>
<td>39</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>24</td>
<td>.</td>
<td>44</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>22</td>
</tr>
</tbody>
</table>

• Let us examine the Type I and Type III sums of squares. The next page contains the SAS output.

• The top of the next page contains the Type I (V1) sum of squares and the bottom of the page contains the Type I (V2) sum of squares. Note the difference in sums of squares, mean squares, F-statistics, and p-values for the Type I analyses.

• The reason for the difference between the V1 and V2 Type I sum of squares is that a Type I analysis is sequential so the order in which terms enter the model is important.

• The Type III analysis is the same for both analyses Type III sums of squares are not calculated sequentially. That is, the order in which terms enter the model is not important.

• The following page contains the two analyses with only one effect in each model. I included these analyses so you can see how \( RSS_2 \) and \( RSS^*_2 \) are calculated.
ANOVA RESULTS: (MODEL WITH SOLUTION THEN DAY)

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>5</td>
<td>1811.575758</td>
<td>362.315152</td>
<td>38.27</td>
<td>0.0005</td>
</tr>
<tr>
<td>Error</td>
<td>5</td>
<td>47.333333</td>
<td>9.466667</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>10</td>
<td>1858.909091</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>solution</td>
<td>2</td>
<td>790.909091</td>
<td>395.454545</td>
<td>41.77</td>
<td>0.0008</td>
</tr>
<tr>
<td>day</td>
<td>3</td>
<td>1020.666667</td>
<td>340.222222</td>
<td>35.94</td>
<td>0.0008</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>solution</td>
<td>2</td>
<td>670.500000</td>
<td>335.250000</td>
<td>35.41</td>
<td>0.0011</td>
</tr>
<tr>
<td>day</td>
<td>3</td>
<td>1020.666667</td>
<td>340.222222</td>
<td>35.94</td>
<td>0.0008</td>
</tr>
</tbody>
</table>

ANOVA RESULTS: (MODEL WITH DAY THEN SOLUTION)

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>5</td>
<td>1811.575758</td>
<td>362.315152</td>
<td>38.27</td>
<td>0.0005</td>
</tr>
<tr>
<td>Error</td>
<td>5</td>
<td>47.333333</td>
<td>9.466667</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>10</td>
<td>1858.909091</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>day</td>
<td>3</td>
<td>1141.075758</td>
<td>380.358586</td>
<td>40.18</td>
<td>0.0006</td>
</tr>
<tr>
<td>solution</td>
<td>2</td>
<td>670.500000</td>
<td>335.250000</td>
<td>35.41</td>
<td>0.0011</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>day</td>
<td>3</td>
<td>1020.666667</td>
<td>340.222222</td>
<td>35.94</td>
<td>0.0008</td>
</tr>
<tr>
<td>solution</td>
<td>2</td>
<td>670.500000</td>
<td>335.250000</td>
<td>35.41</td>
<td>0.0011</td>
</tr>
</tbody>
</table>

So where did $RSS_2$ and $RSS^*_2$ come from?
\( \text{RSS}_2 \) is the \( SSE \) for the model with only treatments and no blocks.

**ANOVA RESULTS FOR THE MODEL WITH SOLUTION (TREATMENTS) ONLY**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2</td>
<td>790.909091</td>
<td>395.454545</td>
<td>2.96</td>
<td>0.1090</td>
</tr>
<tr>
<td>Error</td>
<td>8</td>
<td>1068.000000</td>
<td>133.500000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>10</td>
<td>1858.909091</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R-Square</th>
<th>Coeff Var</th>
<th>Root MSE</th>
<th>growth Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.425469</td>
<td>61.10405</td>
<td>11.55422</td>
<td>18.90909</td>
</tr>
</tbody>
</table>

**ANOVA RESULTS FOR THE MODEL WITH DAYS (BLOCKS) ONLY**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>solution</td>
<td>2</td>
<td>790.909091</td>
<td>395.4545455</td>
<td>2.96</td>
<td>0.1090</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>solution</td>
<td>2</td>
<td>790.909091</td>
<td>395.4545455</td>
<td>2.96</td>
<td>0.1090</td>
</tr>
</tbody>
</table>

\( \text{RSS}_2 \) is the \( SSE \) for the model with only blocks and no treatments.

**ANOVA RESULTS FOR THE MODEL WITH DAYS (BLOCKS) ONLY**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3</td>
<td>1141.075758</td>
<td>380.358586</td>
<td>3.71</td>
<td>0.0696</td>
</tr>
<tr>
<td>Error</td>
<td>7</td>
<td>717.833333</td>
<td>102.547619</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>10</td>
<td>1858.909091</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R-Square</th>
<th>Coeff Var</th>
<th>Root MSE</th>
<th>growth Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.613842</td>
<td>53.55403</td>
<td>10.12658</td>
<td>18.90909</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>day</td>
<td>3</td>
<td>1141.075758</td>
<td>380.358586</td>
<td>3.71</td>
<td>0.0696</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>day</td>
<td>3</td>
<td>1141.075758</td>
<td>380.358586</td>
<td>3.71</td>
<td>0.0696</td>
</tr>
</tbody>
</table>

All of these calculations are done automatically in the RCBD analyses for the two models on the previous page.
Type I SS (V1) Summary

| $RSS_1$ = 1858.91 | $R(\mu) = RSS_1$ | $= 1858.91$ |
| $RSS_2$ = 1068.00 | $R(\tau|\mu) = RSS_1 - RSS_2$ | $= 790.91$ |
| $RSS_3$ = 47.33 | $R(\beta|\tau, \mu) = RSS_2 - RSS_3$ | $= 1020.67$ |

Type I SS (V2) Summary

| $RSS_1$ = 1858.91 | $R(\mu) = RSS_1$ | $= 1858.91$ |
| $RSS_2^*$ = 717.83 | $R(\beta|\mu) = RSS_1 - RSS_2^*$ | $= 1141.08$ |
| $RSS_3$ = 47.33 | $R(\tau|\beta, \mu) = RSS_2^* - RSS_3$ | $= 670.50$ |

Type III SS Summary

| $RSS_1$ = 1858.91 | $R(\mu) = RSS_1$ | $= 1858.91$ |
| $RSS_3$ = 47.33 | $R(\beta|\tau, \mu) = RSS_2^* - RSS_1 = 1020.67$ |
| $RSS_2$ = 1068.00 | $R(\tau|\beta, \mu) = RSS_2 - RSS_1 = 670.50$ |

DM 'LOG; CLEAR; OUT; CLEAR;';
ODS GRAPHICS ON;
ODS PRINTER PDF file='C:\COURSES\ST541\RCBDMISS.PDF';
OPTIONS NODATE NONUMBER;

********************************************************************************
*** RCBD WITH A MISSING OBSERVATION ***
********************************************************************************
DATA IN;
  DO solution = 1 TO 3;
  DO day = 1 TO 4;
    INPUT growth @@; OUTPUT;
  END; END;
CARDS;
13 22 18 39 16 24 . 44 5 4 1 22
;  
********************************************************************************
*** RUN AN ANOVA WITH SOLUTION APPEARING FIRST ***
********************************************************************************
PROC GLM DATA=IN;
  CLASS solution day;
  MODEL growth = solution day;
  TITLE 'ANOVA RESULTS (SOLUTION THEN DAY)';

********************************************************************************
*** RUN AN ANOVA WITH DAY APPEARING FIRST ***
********************************************************************************
PROC GLM DATA=IN;
  CLASS day solution;
  MODEL growth = day solution;
  TITLE 'ANOVA RESULTS (DAY THEN SOLUTION)';

********************************************************************************
*** RUN AN ANOVA WITH SOLUTION ONLY ***
********************************************************************************
PROC GLM DATA=IN;
  CLASS solution;
  MODEL growth = solution;
  TITLE 'ANOVA RESULTS (SOLUTION ONLY)';

********************************************************************************
*** RUN AN ANOVA WITH DAY ONLY ***
********************************************************************************
PROC GLM DATA=IN;
  CLASS day;
  MODEL growth = day;
  TITLE 'ANOVA RESULTS (DAY ONLY)';
RUN;
R code for RCBD with missing value

# ANOVA for RCBD with missing observation

strength <- c(13,22,18,39,16,24,NA,44,5,4,1,22)
solution <- c(1,1,1,2,2,2,3,3,3,3,3,3)
day <- c(1,2,3,4,1,2,3,4,1,2,3,4)

f1 <- aov(strength~factor(day)+factor(solution))
summary(f1)
f2 <- lm(strength~factor(day)+factor(solution))
summary(f2)

R output for RCBD with missing value

> summary (f1)

            Df Sum Sq Mean Sq F value Pr(>F)
factor(day)  3 1141.1  380.4  40.18  0.000636 ***
factor(solution) 2  670.5  335.2  35.41  0.001116 **
Residuals      5  47.3   9.5
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
1 observation deleted due to missingness

> summary(f2)

Coefficients:          Estimate Std. Error t value Pr(>|t|)
(Intercept)            15.333      2.206   6.952   0.000946 ***
factor(day)2            5.333      2.512   2.123   0.087176 .
factor(day)3            1.667      2.901   0.575   0.590479
factor(day)4           23.667      2.512   9.421   0.000227 ***
factor(solution)2        3.000      2.432   1.233   0.272264
factor(solution)3     -15.000      2.176  -6.895   0.000983 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.077 on 5 degrees of freedom
(1 observation deleted due to missingness)

Multiple R-squared: 0.9745,   Adjusted R-squared: 0.9491

F-statistic: 38.27 on 5 and 5 DF,  p-value: 0.0005468
3.6.4 Type I vs Type III Hypotheses

- Because of differences between Type I and Type III SS, there will be differences in the hypotheses associated with the F-tests (assuming the restriction on randomization is ignored).

- Let \( \mu_{ij} = \mu + \tau_i + \beta_j \) be the \( i^{th} \) treatment, \( j^{th} \) block mean.

**Hypotheses for Type III and Type I (V2) Sum of Squares**

\[
H_0 : \bar{\mu}_1 = \bar{\mu}_2 = \cdots = \bar{\mu}_a.
\]

\[
H_1 : \mu_{i*} \neq \bar{\mu}_i \text{ for some } i \neq i^* \text{ and } \bar{\mu}_i = \left( \sum_{j=1}^{b} \mu_{ij} \right) / b.
\]

**Hypotheses for Type I (V1) Sum of Squares**

\[
H_0 : 1/ \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij} \mu_{1j} = 1/ \sum_{i=1}^{a} \sum_{j=1}^{b} n_{2j} \mu_{2j} = \cdots = 1/ \sum_{i=1}^{a} \sum_{j=1}^{b} n_{aj} \mu_{aj}
\]

\[
H_1 : 1/ \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij} \mu_{ij} \neq 1/ \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij*} \mu_{ij*} \text{ for some } i \neq i^*.
\]

where \( n_i \) = the number of nonmissing \( y_{ij} \) values for the \( i^{th} \) treatment, and \( n_{ij} = 1 \) if \( y_{ij} \) is not missing and \( n_{ij} = 0 \) if \( y_{ij} \) is missing.

- The Type III hypotheses are comparing the treatment means average across the blocks (and are the ones I want to test.) Therefore I recommend using the p-values from a Type III analysis.

- If there are no missing \( y_{ij} \) values, the Type I and Type III hypotheses are the same.

### 3.7 RCBD Normal Equations

- For model \( y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \), the error is \( \epsilon_{ij} = y_{ij} - \mu - \tau_i - \beta_j \)

- Substituting in estimates produces the residual \( \hat{\epsilon}_{ij} = \epsilon_{ij} = y_{ij} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j \).

- Goal: Find \( \hat{\mu}, \hat{\tau}_i, \) and \( \hat{\beta}_j \) that minimize \( L \):

\[
L = \sum_{i=1}^{a} \sum_{j=1}^{b} \hat{\epsilon}_{ij}^2 = \sum_{i=1}^{a} \sum_{j=1}^{b} (y_{ij} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j)^2
\]

- Solution: Solve the normal equations

\[
\frac{\partial L}{\partial \hat{\mu}} = -2 \sum_{i=1}^{a} \sum_{j=1}^{b} (y_{ij} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j) = 0
\]

\[
\frac{\partial L}{\partial \hat{\tau}_i} = -2 \sum_{j=1}^{b} (y_{ij} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j) = 0 \text{ for } i = 1, 2, \ldots, a
\]

\[
\frac{\partial L}{\partial \hat{\beta}_j} = -2 \sum_{i=1}^{a} (y_{ij} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j) = 0 \text{ for } j = 1, 2, \ldots, b
\]
• After distributing the sum and then simplifying, we get:

\[ \begin{align*}
(i) \quad y_i &= ab\hat{\mu} + b \sum_{i=1}^{a} \hat{\tau}_i + a \sum_{j=1}^{b} \hat{\beta}_j \\
(ii) \quad y_i &= b\hat{\mu} + b\hat{\tau}_i + \sum_{j=1}^{b} \hat{\beta}_j \quad \text{for } i = 1, 2, \ldots, a \\
(iii) \quad y_{ij} &= a\hat{\mu} + \sum_{i=1}^{a} \hat{\tau}_i + a\hat{\beta}_j \quad \text{for } j = 1, 2, \ldots, b
\end{align*} \]

• For (i), (ii), and (iii), there is a total of \(1 + a + b\) equations. If you sum the \(a\) equations in (ii), you get (i). If you sum the \(b\) equations in (iii), you also get (i). Thus, the rank is \(a + b - 1\) which implies that \(\mu\) and each \(\tau_i\) and \(\beta_j\) are not uniquely estimable. To get estimates of \(\mu\) and each \(\tau_i\) and \(\beta_j\), we must impose 2 constraints. We will use \(\sum_{i=1}^{a} \tau_i = 0\) and \(\sum_{j=1}^{b} \beta_j = 0\).

• Substitution of these constraints into (i), (ii), and (iii) yields

\[ \begin{align*}
(1) \quad ab\hat{\mu} &= y_i \\
(2) \quad b\hat{\mu} + b\hat{\tau}_i &= y_i \\
(3) \quad a\hat{\mu} + a\hat{\beta}_j &= y_{ij}
\end{align*} \]

• Then, from (1), we have

\[ \hat{\mu} = \frac{y_i}{ab} \]

• Substitution of \(\hat{\mu} = \overline{y}_i\) in (2) yields:

\[ b\overline{y}_i + b\hat{\tau}_i = y_i \quad \rightarrow \quad \overline{y}_i + \hat{\tau}_i = \overline{y}_i \quad \rightarrow \quad \hat{\tau}_i = \]

• Substitution of \(\hat{\mu} = \overline{y}_i\) in (3) yields:

\[ a\overline{y}_i + a\hat{\beta}_j = y_{ij} \quad \rightarrow \quad \overline{y}_i + \hat{\beta}_j = \overline{y}_i \quad \rightarrow \quad \hat{\beta}_j = \]

3.8 **Matrix Forms for the RCBD**

**Example:** The goal is to determine whether or not four different tips produce different readings on a hardness testing machine. The machine operates by pressing the tip into a metal test coupon, and from the depth of the resulting depression, the hardness of the coupon can be determined. The experimenter decides to obtain four observations for each tip. Four randomly selected coupons (blocks) were used and each tip (treatment) was tested on each coupon. The data represent deviations from a desired depth in 0.1 mm units:

<table>
<thead>
<tr>
<th>Type of Tip</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of Coupon</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>
• Model: 
  \( y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \) for \( i = 1, 2, 3, 4 \) and \( j = 1, 2, 3, 4 \)

  \( \epsilon_{ij} \sim N(0, \sigma^2) \quad \beta_j \sim N(0, \sigma^2_\beta) \)

• Assume (i) \( \sum_{i=1}^{4} \tau_i = 0 \) and (ii) \( \sum_{j=1}^{4} \beta_j = 0 \). If we estimate \( [\mu, \tau_1, \tau_2, \tau_3, \beta_1, \beta_2, \beta_3] \), we can then estimate \( \tau_4 = -\tau_1 - \tau_2 - \tau_3 \) from (i) and \( \beta_4 = -\beta_1 - \beta_2 - \beta_3 \) from (ii).

\[
X = \begin{bmatrix}
1 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & -1 & -1 & -1 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & -1 & -1 & -1 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & -1 & -1 & -1 \\
1 & -1 & -1 & -1 & 1 & 0 & 0 \\
1 & -1 & -1 & -1 & 0 & 1 & 0 \\
1 & -1 & -1 & -1 & 0 & 0 & 1 \\
1 & -1 & -1 & -1 & -1 & -1 & -1
\end{bmatrix}
\]

\[
y = \begin{bmatrix}
-2 \\
-1 \\
1 \\
5 \\
-1 \\
3 \\
4 \\
-3 \\
-1 \\
0 \\
1 \\
2 \\
7
\end{bmatrix}
\]

\[
X'X = \begin{bmatrix}
16 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 8 & 4 & 4 & 0 & 0 & 0 \\
0 & 4 & 8 & 4 & 0 & 0 & 0 \\
0 & 4 & 4 & 8 & 0 & 0 & 0 \\
0 & 0 & 0 & 8 & 4 & 4 \\
0 & 0 & 0 & 4 & 8 & 4 \\
0 & 0 & 0 & 4 & 4 & 8
\end{bmatrix}
\]

\[
(X'X)^{-1} = \frac{1}{16} \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 3 & -1 & -1 & 0 & 0 \\
0 & -1 & 3 & -1 & 0 & 0 \\
0 & -1 & -1 & 3 & 0 & 0 \\
0 & 0 & 0 & 0 & 3 & -1 \\
0 & 0 & 0 & 0 & -1 & 3 \\
0 & 0 & 0 & 0 & -1 & -1
\end{bmatrix}
\]

\[
X'y = \begin{bmatrix}
20 \\
-12 \\
-11 \\
-17 \\
-22 \\
-21 \\
-9
\end{bmatrix}
\]

\[
(X'X)^{-1}X'y = \frac{1}{16} \begin{bmatrix}
20 \\
-36 + 11 + 17 \\
-12 - 33 + 17 \\
-66 + 21 + 9 \\
22 - 63 + 9 \\
22 + 21 - 27
\end{bmatrix}
\]

\[
= \frac{1}{16} \begin{bmatrix}
20 \\
-8 \\
-4 \\
-28 \\
-36 \\
-32 \\
16
\end{bmatrix} = \begin{bmatrix}
5/4 \\
-2/4 \\
-1/4 \\
-7/4 \\
-9/4 \\
-8/4 \\
4/4
\end{bmatrix} = \begin{bmatrix}
\bar{y}_1 - \bar{y}_4 \\
\bar{y}_2 - \bar{y}_4 \\
\bar{y}_3 - \bar{y}_4 \\
\bar{y}_1 - \bar{y}_3 \\
\bar{y}_2 - \bar{y}_3 \\
\bar{y}_3 - \bar{y}_2
\end{bmatrix} = \begin{bmatrix}
\hat{\mu} \\
\hat{\tau}_1 \\
\hat{\tau}_2 \\
\hat{\tau}_3 \\
\hat{\beta}_1 \\
\hat{\beta}_2 \\
\hat{\beta}_3
\end{bmatrix}
\]

Thus, \( \hat{\tau}_4 = -\hat{\tau}_1 - \hat{\tau}_2 - \hat{\tau}_3 = 10/4 \) and \( \hat{\beta}_4 = -\hat{\beta}_1 - \hat{\beta}_2 - \hat{\beta}_3 = 13/4 \)
Alternate Approach: Keeping $a + b + 1$ Columns

\[
X = \begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1
\end{bmatrix}
\quad y = \begin{bmatrix}
-2 \\
-1 \\
1 \\
5 \\
-1 \\
-2 \\
-1 \\
3 \\
4 \\
-3 \\
0 \\
-1 \\
2 \\
2 \\
1 \\
5 \\
7 \\
0
\end{bmatrix}
\]

\[
X'X = \begin{bmatrix}
16 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 \\
4 & 5 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
4 & 1 & 5 & 1 & 1 & 1 & 1 & 1 & 1 \\
4 & 1 & 1 & 5 & 1 & 1 & 1 & 1 & 1 \\
4 & 1 & 1 & 1 & 5 & 1 & 1 & 1 & 1 \\
4 & 1 & 1 & 1 & 1 & 5 & 1 & 1 & 1 \\
4 & 1 & 1 & 1 & 1 & 1 & 5 & 1 & 1 \\
4 & 1 & 1 & 1 & 1 & 1 & 1 & 5 & 1 \\
4 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 5
\end{bmatrix}
\quad (X'X)^{-1} = \frac{1}{16} \begin{bmatrix}
3 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\
-1 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4
\end{bmatrix}
\]

\[
X'y = \begin{bmatrix}
20 \\
3 \\
4 \\
-2 \\
15 \\
-4 \\
-3 \\
9 \\
18
\end{bmatrix}
\quad (X'X)^{-1}X'y = \begin{bmatrix}
5/4 \\
-2/4 \\
-1/4 \\
-7/4 \\
10/4 \\
-9/4 \\
-8/4 \\
4/4 \\
13/4
\end{bmatrix}
= \begin{bmatrix}
\hat{\mu} \\
\hat{\tau}_1 \\
\hat{\tau}_2 \\
\hat{\tau}_3 \\
\hat{\tau}_4 \\
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4
\end{bmatrix}
\]