Assumptions in the Analysis of Variance

1. Error terms are randomly, independently, and normally distributed
   \[ e_{ij} \sim \mathcal{N}(0, \sigma^2) \]

2. Variances of different samples are homogenous

3. Variances and means of different samples are not correlated, i.e. are independent

4. The main effects are additive

Additivity of Main Effects - RCBD

\[
\begin{align*}
X_{11} &= \bar{X}. + t_1 + b_1 + e_{11} \\
- (X_{21} &= \bar{X}. + t_2 + b_2 + e_{21}) \\
X_{11} - X_{21} &= t_1 - t_2 + e_{11} - e_{21}
\end{align*}
\]

Difference between 2 treatments in RCBD is the same in all blocks and is influenced only by the residual effects

Method of Least Squares

The mathematical model describes the components of a design.

\[
X_{ij} = \bar{X}. + t_i + b_j + e_{ij}
\]

\[ e_{ij} = X_{ij} - (\bar{X}. + t_i + b_j) \]

Best estimator of population parameters is obtained by minimizing SSError

\[
\text{SSError} = \sum e_{ij}^2 = \sum (X_{ij} - (\bar{X}. + t_i + b_j))^2
\]

The minimum is found by the method of Least Squares. Differentiate \( \sum e_{ij}^2 \) with respect to each unknown in turn and set the derivative to 0:

\[
\sum (X_{ij} - (\bar{X}. + t_i + b_j)) = 0
\]

\[
\sum X_{ij} = \sum (\bar{X}. + t_i + b_j)
\]

This is the Normal Equation for an RCBD. It is used to obtain the best estimates of the population parameters.
RCBD experiment with 3 treatments and 2 blocks

<table>
<thead>
<tr>
<th></th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>B Totals</th>
<th>B Means</th>
<th>Dev. from mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>3</td>
<td>9</td>
<td>6</td>
<td>18</td>
<td>6</td>
<td>+1</td>
</tr>
<tr>
<td>B2</td>
<td>5</td>
<td>1</td>
<td>6</td>
<td>12</td>
<td>4</td>
<td>-1</td>
</tr>
<tr>
<td>T Totals</td>
<td>8</td>
<td>10</td>
<td>12</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T Means</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Dev. from mean</td>
<td>-1</td>
<td>0</td>
<td>+1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RCBD with 3 treatments and 2 blocks in dot notation

<table>
<thead>
<tr>
<th>Bj</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>B Totals</th>
<th>B Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>X_{11}=3</td>
<td>X_{21}=9</td>
<td>X_{31}=6</td>
<td>X_{1}=18</td>
<td>\overline{X}_{1} = 6</td>
</tr>
<tr>
<td>B2</td>
<td>X_{12}=5</td>
<td>X_{22}=1</td>
<td>X_{32}=6</td>
<td>X_{2}=12</td>
<td>\overline{X}_{2} = 4</td>
</tr>
<tr>
<td>T Totals</td>
<td>X_{1}=8</td>
<td>X_{2}=10</td>
<td>X_{3}=12</td>
<td>X_{.}=30</td>
<td></td>
</tr>
<tr>
<td>T Means</td>
<td>\overline{X}_{1} = 4</td>
<td>\overline{X}_{2} = 5</td>
<td>\overline{X}_{3} = 6</td>
<td>\overline{X}_{.} = 5</td>
<td></td>
</tr>
</tbody>
</table>

Normal Equations

For one observation, t_{i}b_{j}, the expected value is

\hat{X}_{ij} = \overline{X}_{.} + t_{i} + b_{j}

For the treatment 1 total, T_{1}

\hat{X}_{11} = \overline{X}_{.} + t_{1} + b_{1}  \\
\hat{X}_{12} = \overline{X}_{.} + t_{1} + b_{2}

T_{1} = X_{1.} = 8 = 2\overline{X}_{.} + 2t_{1} + b_{1} + b_{2}

For the block 1 total, B_{1}

B_{1} = X_{.1} = 18 = 3\overline{X}_{.} + t_{1} + t_{2} + t_{3} + 3b_{1}

Best Estimators of the Effects of Population Parameters
From the normal equation for the $T_1$ total, the best estimator for treatment 1 effect, $t_1$, is

$$T_1 = 2\bar{X}_.. + 2t_1 + b_1 + b_2$$

Effects are recorded as deviations from the mean, so

$$b_1 + b_2 = 0$$

$$T_1 = 2\bar{X}_.. + 2t_1$$

$$t_1 = \frac{2t_1}{2} = \bar{X}_1. - \bar{X}_.$$ 

Similarly, the best estimator for block $j$ is $b_j = \bar{X}_i. - \bar{X}_.$

Measure of Variance

From the mathematical model

$$e_{ij} = X_{ij} - (\bar{X}_.. + t_i + b_j)$$

$$e_{ij} = X_{ij} - \bar{X}_{ij} = \text{Observed} - \text{Expected}$$

Use best estimators to calculate the expected values for observations

$$\bar{X}_{ij} = \bar{X}_.. + t_i + b_j = \bar{X}_.. + (\bar{X}_i. - \bar{X}_.) + (\bar{X}_j. - \bar{X}_.)$$

Error components are the differences between observed and expected values

$$e_{11} = 3 - \left[5 + (4 - 5) + (6 - 5)\right] = -2$$

$$e_{12} = 5 - \left[5 + (4 - 5) + (4 - 5)\right] = +2$$

$$e_{21} = 9 - \left[5 + (5 - 5) + (6 - 5)\right] = +3$$

$$e_{22} = 1 - \left[5 + (5 - 5) + (4 - 5)\right] = -3$$

$$e_{31} = 6 - \left[5 + (6 - 5) + (6 - 5)\right] = -1$$

$$e_{32} = 6 - \left[5 + (6 - 5) + (4 - 5)\right] = +1$$

Estimated Errors

<table>
<thead>
<tr>
<th></th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>Bj</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>-2</td>
<td>+3</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>B2</td>
<td>+2</td>
<td>-3</td>
<td>+1</td>
<td>0</td>
</tr>
<tr>
<td>T1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The errors can be squared and summed to obtain the sum of squares for error

$$\text{SSErr} = \sum e_{ij}^2 = (-2)^2 + 2^2 + \ldots + 1^2 = 28$$
This is the same result as is obtained more efficiently in the ANOVA.

\[
df \text{ Error} = df \text{ Total} - df \text{ Trt} - df \text{ Block} = 5 - 2 - 1 = 2
\]
Equivalently
\[
df \text{ Error} = (r - 1)(t - 1) = (2-1)(3-1) = 2
\]

\[
\text{MSE} = \frac{28}{2} = 14 = s^2 \text{ or variance}
\]
\text{MSE} \text{ is the unexplained or random variability}