Assumptions in the Analysis of Variance

1. Error terms are randomly, independently, and normally distributed

$$e_{ij} \approx iN(0, \sigma^2)$$

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

$$\begin{array}{l} X_{11} = \overline{X}.. \; + t_1 + b_1 + e_{11} \\ \underline{-(X_{21} = \overline{X}.. \; + t_2 + b_2 + e_{21})} \\ X_{11} - X_{21} = t_1 - t_2 + e_{11} - e_{21} \end{array}$$

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.

$$X_{ij} = \overline{X}.. \; + t_i + b_j + e_{ij}$$

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

Best estimator of population parameters is obtained by minimizing SSError

$$\text{SSError} = \sum e_{ij}^2 = \sum (X_{ij} - (\overline{X}.. \; + t_i + b_j))^2$$

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:

$$\sum (X_{ij} - (\overline{X}..+t_i + b_j)) = 0$$

$$\sum X_{ij} = \sum (\overline{X}..+t_i + b_j)$$

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

	T1	T2	T3	B totals	B means	Dev. from mean
B1	3	9	6	18	6	+1
B2	5	1	6	12	4	-1
T totals	8	10	12	30		
T means	4	5	6		5	
Dev. from mean	-1	0	+1			

RCBD with 3 treatments and 2 blocks in dot notation

	T_i				
B_j	T1	T2	T3	B Totals	B Means
B1	$X_{11}=3$	$X_{21}=9$	$X_{31}=6$	$X_{.1}=18$	$\overline{X}_{.1}=6$
B2	$X_{12}=5$	$X_{22}=1$	$X_{32}=6$	$X_{.2}=12$	$\overline{X}_{.2}=4$
T Totals	$X_{1.}=8$	$X_{2.}=10$	$X_{3.}=12$	$X_{..}=30$	
T Means	$\overline{X}_{1.}=4$	$\overline{X}_{2.}=5$	$\overline{X}_{3.}=6$		$\overline{X}_{..}=5$

Normal Equations

For one observation, t_1b_1, the expected value is

$$\hat{X}_{11} = \overline{X}_{..} + t_1 + b_1$$

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

$$b_1 + b_2 = 0$$

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

$$t_1 = \frac{T_1}{2} - \bar{X}_{..} = \bar{X}_{1.} - \bar{X}_{..}$$

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

Measure of Variance

From the mathematical model

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

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

Use best estimators to calculate the expected values for observations

$$\hat{X}_{ij} = \bar{X}_{..} + t_i + b_j = \bar{X}_{..} + (\bar{X}_{i.} - \bar{X}_{..}) + (\bar{X}_{.j} - \bar{X}_{..})$$

Error components are the differences between observed and expected values

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

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

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

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

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

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

Estimated Errors

	T1	T2	T3	$B_{.j}$
B1	-2	+3	-1	0
B2	+2	-3	+1	0
$T_{i.}$	0	0	0	

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

$$SSError = \sum e_{ij}^2 = (-2)^2 + 2^2 \ldots + 1^2 = 28$$

This is the same result as is obtained more efficiently in the ANOVA.

df Error = df Total - df Trt - df Block = 5 - 2 - 1 = 2
Equivalently
df Error = (r - 1)(t - 1) = (2-1)(3-1) = 2

MSError = 28/2 = 14 = s^2 or variance
MSError is the unexplained or random variability