The first element in table (3) is the number of means involved in the range of means from $S_3$ (14.3) to $S_1$ (8). These three means are 14.3, 10 and 8.

Between $S_3$ and $S_2$, two means are involved i.e. 14.3 and 100. When only one mean is involved as it is the case between $S_1$ and $S_2$, it is ignored since we will not be interested in comparing $S_1$ with $S_2$.

iii. The df for error is 19 from the ANOVA table. This value is paired up with each range obtained in (ii) above, and the corresponding Duncan’s tables.

Thus, the table values (at 0.05 level of significance) are:

\[(\text{range}, d_{\text{error}}) = \text{table value}\]

- \[(3, 19) = 3.11\]
- \[(2, 19) = 2.96\]

iv. The standard error (S.E) is computed as

\[\text{S.E} = \left\lfloor \frac{\text{MSE}}{r} \right\rfloor\]

Where MSE = \(\text{MSE}_{\text{error}}\) from the ANOVA table

\[r = \text{Number of observation per treatment}\]

Due to the unequal number of observations per treatment an approximate value of \(r\) is computed as:

\[\frac{\sum r_i}{n}\]

Where \(n\) = Number of observations

\[r_i = \text{The number of observations for the } i\text{th treatments}\]

\[6.3158\]
\[ SE = \frac{3.31 \times 1.4462}{S} = 4.4977 \]

5. The standard references:
   \[ 3.11 \times 1.4462 = 4.4977 \]
   \[ 2.96 \times 1.4462 = 4.2808 \]

6. Testing the mean difference
   \[ S - S_1 = 6.3 \text{ vs } 4.4977 \]
   \[ S - S_2 = 3.425 \text{ vs } 4.288 \text{ ns} \]
   \[ S - S_3 = 2.875 \text{ vs } 4.2808 \text{ ns} \]

This result can be presented as follows:

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>S_3</td>
<td>14.3  a</td>
</tr>
<tr>
<td>S_1</td>
<td>10.875 ab</td>
</tr>
<tr>
<td>S_2</td>
<td>8.0   b</td>
</tr>
</tbody>
</table>

Mean with the same letters are not significantly different at 0.05 level significance. This result is the same as we have for the LSD. The interpretation of the result is also the same as for the LSD procedure.

**HOMEWORK**

Hybrid seeds of *Tectona grandis* from four different sources are being compared with a view to selecting the best source from which seeds could be procured for distribution to Taungya farmers.
in J4, Ogun state. The four sources are Ondo, Ore, Ile-Ife and India. The hybrid from each source was planted in five plots within the Tuangya farm I J4 and the following yields were obtained.

## HYBRID SEED SOURCE

<table>
<thead>
<tr>
<th>HYBRID SEED SOURCE</th>
<th>Ondo</th>
<th>Ore</th>
<th>Ile-Ife</th>
<th>India</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>17.43</td>
<td>16.60</td>
<td>14.80</td>
<td>21.24</td>
</tr>
<tr>
<td></td>
<td>19.54</td>
<td>16.38</td>
<td>16.22</td>
<td>21.29</td>
</tr>
<tr>
<td></td>
<td>17.67</td>
<td>17.01</td>
<td>21.30</td>
<td>20.14</td>
</tr>
<tr>
<td></td>
<td>22.58</td>
<td>15.82</td>
<td>20.84</td>
<td>19.43</td>
</tr>
<tr>
<td></td>
<td>19.36</td>
<td>16.57</td>
<td>19.27</td>
<td>25.05</td>
</tr>
</tbody>
</table>

a) Perform an analysis of variance to see if there is any difference among yield of hybrid seeds of *Tectona grandis* from four sources.

b) Use the LSD procedure to test for differences among the means.

c) Which source would you recommend and why?

### RANDOMIZED COMPLETE BLOCK DESIGN (RCBD)

This is a design used when the experimental units are not homogenous, and this can be allocated to groups or blocks such that the variation among blocks is maximized while the variation within any particular block is minimized. Here, it is ensured that the number of units within a block is equal to the number of treatments being investigated.

If, for instance, we are using four treatments, then each block or group of experimental units will be divided into four. The treatments are then assigned randomly to the units within each block.
E.g. in an experimental conducted to study the effects of six different diets on rats. If the experimenter feels that the responses will be influenced by the age of the rats, he will be influenced by the age of the rats, he will not want his age effect to interfere with the analysis of the diet effect. In order to control this extraneous source of variability, he can group the rats into age classes and then assign the diets at random to the rats within each age class. The age classes are the blocks in this case. In RCBD, each treatment appears an unequal number of times in each block and each block contains all treatments.

ADVANTAGES OF THE RCBD

1) Grouping the experimental units into blocks yield more precise results with the CRD
2) There is no restriction on the number of blocks, and treatments that can be used.
3) The statistical analysis is simple. If the data from individual units are missing, they can be estimated using appropriate estimation formulae.
4) If the experimental error is heterogeneous, an unbiased error for testing any specific combination of the treatment means can still be obtained.

DISADVANTAGES OF THE RCBD

When the variation among experimental units within a block is large, a large error term results. This occurs when the number of treatments is large, thus it may not be possible too secure sufficiently uniform groups of units for blocks.

1) Name of design: Randomized complete block design (RCBD)
2) Model: $Y_{ij} = \mu + B_i + T_j + e_{ij}$
Where: \( Y_{ijk} \) = individual observation i.e. observation for the \( j_k \) treatment in the \( i_k \) block

\( \mu \) = General mean

\( B_i \) = effect of the \( i_k \) block

\( T_j \) = effect of the \( j_k \) treatment

\( e_{ijk} \) = experimental error

3) Assumption: As for CRD. In addition, the block effects are assumed to be additive and there is no interaction between the blocks and the treatments

4) Lay-out:

\[
\begin{array}{cccc}
B_1 & B_2 & B_3 & B_4 \\
T_{1.3} & T_{1.2} & T_{2.3} & T_{1.9} \\
T_{1.4} & T_{1.4} & T_{1.6} & T_{1.4} \\
T_{1.5} & T_{1.2} & T_{1.7} & T_{1.5} \\
\end{array}
\]

5) Hypotheses: The same procedure for CRD but the \( H_0 \) and \( H_A \) should be for treatment and block.

6) Computation: The same procedure for CRD except SS block which is included. SS error will be the difference of SS total, SS block and SS treatment.

7) Construct the ANOVA table

<table>
<thead>
<tr>
<th>SV</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block</td>
<td>4-1</td>
<td>SS_a</td>
<td>( \frac{SS_a}{b-1} )</td>
<td>( \frac{ms_B}{ms_e} )</td>
</tr>
</tbody>
</table>
8) Conclusion: Compare the f-calculated with the f-tabulated. This could be done for the
treatment or the block. Mainly, the source of variation is from the treatment, hence, it is
better to make inference from the treatment.
Using the example above as our guide.

1. Name: RCBD
2. Model: $Y_{ijk} = \mu + B_i + T_j + e_{ijk}$
3. Define each
4. Assumption: As stated
5. Lay-out:

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>9</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>7</td>
<td>3</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>9</td>
<td>16</td>
<td>18</td>
<td>55</td>
<td></td>
</tr>
</tbody>
</table>

5. Hypotheses: Treatments
   $H_0 = \text{No Significant difference in the treatment}$
6. Computation:

\[ CF = \frac{\sum Y_{ij}^2 - \sum \frac{Y_{ij}}{n}}{s} \]

\[ = \frac{295 - 252.08}{s} = 16.25 \]

\[ SS_{\text{total}} = \sum Y_{ij}^2 - CF = 252.25 - 252.08 = 0.17 \]

\[ SS_{\text{error}} = SS_{\text{total}} - SS_{\text{block}} - SS_{\text{treatment}} = 26.5 \]

7. ANOVA Table

<table>
<thead>
<tr>
<th>SV</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block</td>
<td>3</td>
<td>16.25</td>
<td>5.42</td>
<td>5.14 (P&lt;0.05)</td>
</tr>
<tr>
<td>Treatment</td>
<td>2</td>
<td>0.17</td>
<td>0.09</td>
<td>4.76 (P&lt;0.05)</td>
</tr>
<tr>
<td>Error</td>
<td>6</td>
<td>26.5</td>
<td>4.42</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>42.92</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Since the $F_{tab}$ is greater than the $F_{Calculated}$ at 0.05 and 0.01 levels of significance respectively, the $H_0$ is accepted. It is therefore concluded that there no significant difference in the treatments.

**LATIN SQUARE DESIGN**

Latin square design is a logical extension of the RCBD. It is used in situations where there is a two-way source of variation that cannot be isolated by one-way blocking alone. In such situations, there is difficulty in the choice between two alternative systems of block. Latin square design permits two sets of block constraints (rows and columns) to be used simultaneously. In Latin square design, the treatments are assigned at random subject to the restriction that each treatment appears but once in each row and once in each column. E.g. at J4 experimental station, some rows were ploughed to be used to test the rate of growth of *Terminalia superb* treated with three different fertilizers. Before planting, the soil chemist of the station discovered that fertility gradients existed both parallel to and at right angle to the ploughed rows. In this circumstances, the ploughed area will be blocked in two perpendicular directions so as to isolate the two-way fertility gradients. Since there are three treatments (i.e. fertilizers), the area should be divided such that there will be three blocks along the rows and three blocks along the column. The rows and the columns are randomized before the treatments are then allocated randomly to them. The allocation of the treatments should be such that each treatment appears exactly only once in each row and column. The design will be:

\[
\begin{array}{c|c|c}
R_1 & A & C \\
R_2 & & B \\
R_3 & & \\
\end{array}
\]
Where $R_1 - R_3$ are the rows

$C_1 - C_3$ are the columns

A – C are treatments (Note that each of these is randomized)

This is a 3 x 3 Latin square design i.e. there are 3 rows and columns and 3 treatments. It is a square because the number of rows equals the number of columns.

MODELS FOR LATIN SQUARE DESIGN

Model: $Y_{ijk} = \mu + r_i + c_j + T_k + \epsilon_{ijk}$

Where $Y_{ijk} = \text{individual observation}$

$\mu = \text{general mean}$

$r_i = \text{effect of the rows}$

$c_j = \text{effect of the columns}$

$T_k = \text{effects of treatments}$

$\epsilon_{ijk} = \text{experimental error}$

Assumptions: General assumption made for ANOVA.

In addition, the Latin Square design model assumes that there are no interactions between the rows and columns, rows and treatments, and between columns and treatments. In this design, any of the three effects (i.e. rows, columns and treatments may be fixed or random. Whichever is the same procedure for analysis is used.
Example

In a biometric investigation, four agro forestry tree species were treated each with four different nitrogenous fertilizers (A, B, C, and D) in successive weeks. To eliminate any possible effects due to the order in which the fertilizers were given, the experiment was arranged in latin square design form. After each weeks treatments, the nitrogen level of the tree species was measured (in standard units). The results obtained are as follows:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A9.93</td>
<td>B5.72</td>
<td>D5.83</td>
<td>C9.67</td>
</tr>
<tr>
<td>2</td>
<td>D6.69</td>
<td>A8.51</td>
<td>C10.00</td>
<td>B9.49</td>
</tr>
<tr>
<td>3</td>
<td>C5.89</td>
<td>D3.36</td>
<td>B3.55</td>
<td>A6.92</td>
</tr>
<tr>
<td>4</td>
<td>B3.59</td>
<td>C3.84</td>
<td>A4.00</td>
<td>D6.70</td>
</tr>
</tbody>
</table>

Carry out analysis of variance and test for significant differences among the treatments as well as among each of the two extraneous effects.

Solution

The name of the design is a 4 x 4Latin square design. The model and the assumption are as given earlier.

The layout

<table>
<thead>
<tr>
<th>Column week</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Row totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The above layout is for rows and columns. This will be followed by a separate layout for treatments.

Treatments (fertilizers)

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.93</td>
<td>5.72</td>
<td>9.67</td>
<td>5.83</td>
</tr>
<tr>
<td>8.51</td>
<td>9.49</td>
<td>10.00</td>
<td>6.67</td>
</tr>
<tr>
<td>6.92</td>
<td>5.84</td>
<td>5.89</td>
<td>3.36</td>
</tr>
<tr>
<td>4.00</td>
<td>3.59</td>
<td>3.84</td>
<td>6.70</td>
</tr>
<tr>
<td><strong>Treatments totals</strong></td>
<td><strong>29.36</strong></td>
<td><strong>22.35</strong></td>
<td><strong>29.40</strong></td>
</tr>
<tr>
<td><strong>Treatments means</strong></td>
<td><strong>7.34</strong></td>
<td><strong>5.59</strong></td>
<td><strong>7.35</strong></td>
</tr>
</tbody>
</table>

The hypotheses: There will be three sets of hypotheses. These are for the treatments (fertilizers); rows (trees) and columns (weeks).

TREATMENTS (FERTILIZERS)

- \( H_0 \): The effects of the four treatments are not significantly different
- \( H_A \): The effects of the four treatments are significantly different
ROWS (TREES)

H₀: There is no significant difference between the trees

H₁: There is significant difference between the trees

COLUMNS (WEEKS)

H₀: There is no significant difference between the weeks

H₁: There is significant difference between the weeks

Computation:

a) C.F = \frac{\sum (x_i - \bar{x})^2}{n-1} = 671.7168

b) SS\text{total} = (9.93)^2 + (6.67)^2 + \ldots + (6.70)^2 - \text{C.F} = 760.7845 - 671.7168 = 89.0677

c) SS_{rows} = \frac{(9.93)^2 + (6.67)^2 + \ldots + (6.70)^2}{6} - \text{C.F} = 50.7599

d) SS_{columns} = \frac{(9.93)^2 + (6.67)^2 + \ldots + (6.70)^2}{6} - \text{C.F} = 50.7599

\therefore SS_{columns} - C.F = 0.717168
\( e) \ SS_{\text{treatments}} = \frac{1}{3} \left( 671.7168 - 18.4242 \right) = 11.9946 \)

\( f) \ SS_{\text{error}} = SS_{\text{total}} - SS_{\text{rows}} - SS_{\text{columns}} - SS_{\text{treatments}} = 89.0677 - 50.7599 - 18.4242 - 11.9946 = 7.8890 \)

<table>
<thead>
<tr>
<th>SS</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F-cal</th>
<th>F-Tab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rows (trees)</td>
<td>3</td>
<td>50.7599</td>
<td>16.9200</td>
<td>12.87*</td>
<td>4.76</td>
</tr>
<tr>
<td>Columns (weeks)</td>
<td>3</td>
<td>18.4242</td>
<td>6.1414</td>
<td>4.67ns</td>
<td>4.76</td>
</tr>
<tr>
<td>Treatments (fert)</td>
<td>3</td>
<td>11.9946</td>
<td>3.9982</td>
<td>3.04ns</td>
<td>4.76</td>
</tr>
<tr>
<td>Error</td>
<td>6</td>
<td>7.8890</td>
<td>1.3148</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>15</td>
<td>89.0677</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

With respect to the rows, the \( H_0 \) is not accepted because F-calculated is greater than F-tabulated. It is therefore concluded that there are significant differences between the trees.

With respect to the columns, \( H_0 \) is accepted because F-calculated is less than F-tabulated. It is therefore concluded that there is no significant difference between weeks i.e. between the nitrogen level of the trees from week to week.

With respect to the treatments, the F-calculated is less than F-tabulated, hence, the \( H_0 \) is accepted i.e. no significant differences in the effect of the four different fertilizers.
However, the above result indicates that the significant differences in the nitrogen level of the trees are attributable to the significant differences between the trees themselves. The fertilizers and the weeks have significant effects on the nitrogen level of the trees.

MISSING VALUES

In RCBD: $M = \frac{S - T - B}{t-1}$

Where $M$ = an estimated of the missing value

$t$ = number of treatments in the experiment

$b$ = number of blocks in the experiment

$T$ = Sum of observation with the same treatments as the missing

$B$ = Sum of observation with the same block as the missing

$S$ = sum of all the actual observation

Once the estimate of the missing value has been computed the value is entered in the table with the observed values and the analysis of variance is performed as usual. Before constructing the ANOVA table, the following adjustments are to be made:

i. The df for error is reduced by 1

ii. The df for total is also reduced by 1

iii. The SS$_{tr}$ is reduced by $Z$

iv. The SS$_{tot}$ is also reduced by $Z$

$Z$ is the correction for bias

$Z$ is calculated using the following formula:

$Z = \frac{(T - B)^2}{T + B} - \frac{t(t-1)}{2(b-1)}$
After those adjustments, the ANOVA table is then constructed.

<table>
<thead>
<tr>
<th>SN</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>b – 1</td>
<td>SSb</td>
<td>MSb</td>
<td></td>
</tr>
<tr>
<td>Treatments</td>
<td>t – 1</td>
<td>(SST) – Z</td>
<td>MSt</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>(b – 1)(t – 1)</td>
<td>SSE</td>
<td>MSE</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>(bt – 1) – 1</td>
<td>(SST) – Z</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It should be noted that the df blocks and SS blocks are not adjusted at all. Also, the df treatments and the SSE are not adjusted. The SSE is not adjusted because the estimated missing value makes no contribution to it.

Example: The data below (with one missing observation) were collected from a field experiment involving five treatments in three blocks.

<table>
<thead>
<tr>
<th>TREATMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
</tr>
<tr>
<td>l o</td>
</tr>
<tr>
<td>c k s</td>
</tr>
<tr>
<td>I</td>
</tr>
<tr>
<td>I</td>
</tr>
<tr>
<td>II</td>
</tr>
<tr>
<td>III</td>
</tr>
</tbody>
</table>

M denotes the missing observation.