COURSE CODE: STS 352

COURSE TITLE: EXPERIMENTAL DESIGN 1

NUMBER OF UNIT: 2 UNITS

COURSE DURATION: TWO HOURS PER WEEK.

COURSE COORDINATOR: MR G.A. DAUDU

LECTURER OFFICE LOCATION: AMREC

COURSE CONTENT:
Basic concepts of experimentation, Completely randomized design, Randomised complete block design, Latin Square Design, Graeco Latin Square Design, Simple factorial Design

COURSE REQUIREMENTS:
This is a compulsory course for all statistics students. Students are expected to have a minimum of 75% attendance to be able to write the final examination.

READING LIST:
1.) Statistical Design and Analysis of Experiments by P.W.M. John.
2.) Experimental Designs by Cochran and Cox.
3.) Designs and Analysis of Experiments for Biology and Agric. Students by Oyejola, B.A.
4.) Statistical Methods by Snedecor and Cochran.
5.) Statistical Procedures for Agricultural Research by Gomez and Gomez.

LECTURE NOTES

Introduction
An experiment involves the planning, execution and collection of measurements or observations.

Examples of simple experiment
1. Comparison of two teaching methods
2. Comparison of two varieties of maize

The difference among experimental units treated alike is called experimental error, this error is the primary basis for deciding whether an observed difference is real or
just due to chance. Clearly every experiment must be designed to have a measure of the experimental error.

**Definitions**

**Experimental Unit/plot**

This is the smallest to which a treatment is applied, and on which an observation is made e.g. an animal bird, an object, a cage, a field plat and so on.

- Definition of a unit depend on the objective of the experiment.

**Factors**

These are distinct types of condition that are manipulated on the experimental unit e.g. age, group, gender, variety, fertilizer and so on.

**Factor Levels**

Different mode of the presence of a factors are called factor levels.

- Factors can be quantitative or qualitative.

**Treatments**

Each specific combination of the levels of different factors is called the treatment.

**Replication**

These are the numbers of experimental units to which a given treatment is applied.

**MAIN ASPECT OF DESIGNING EXPERIMENT**

a. Choose the factor to be studied in the experiment and the levels of each factor that are relevant to the investigation.

b. Consider the scope of inference and choose the type of experimental unit on which treatment are to be applied.

c. From the perspective of cost and desired precision of inference, decide on the number of units to be used for the experiment.

d. Finally, and most important, determine the manner in which the treatments are to be applied to the experimental units (i.e. design of the experiment).

**PRINCIPLES OF EXPERIMENTAL DESIGN**
There are three basic statistical requirements for a good experiment:

- Randomization
- Replication
- Local Control or Blocking

1. **RANDOMIZATION:** This is the process by which it is ensured that each treatment has an equal chance of being assigned to any experimental unit e.g.

- Suppose two maize varieties, Yellow (Y) and White (W) are to be compared using four experimental units for each variety.

  ![Layouts](https://www.unaab.edu.ng/images/layouts.png)

  In layout (II) if the field has fertility gradient so that there is a gradual productivity from top to bottom. Then the white variety will be at an advantage being in a relatively more fertile area hence, the comparison within the variety would be biased in favour of variety “W”.

  A better layout is obtained by randomization as shown in layout (I).

2. **REPLICATION:** Each treatment being applied to more than one experimental unit. Experimental error can be measured only if there are replications. Also the more the experimental units used for each treatment, the lower would be the standard error for the estimate for treatment effect and hence, the more precise the experiment. Precision is the measurement of how close the observed values are to each other.

3. **BLOCKING OR LOCAL CONTROL:** This is the process of grouping together experimental units that are similar and assigning all treatments into each group or block separately and independently. This allows the measurement of variation among blocks which can be removed from the experimental error. Blocking is therefore one of the measures for reducing or minimizing experimental error. The
ability of detecting existing or real differences among treatments increase as the size of the experimental error decreases.

COMpletely RANDOMIZE DESIGN (CRD)

Introduction
A CRD is a design in which the treatments are assigned completely at random so that each experimental unit has equal chance of receiving any one treatment. Any difference among the experimental units receiving the same treatment is considered to be experimental error.

Model:
\[ y_{ij} = \mu_i + e_{ij} \]
\[ = \mu + \alpha_i + e_{ij} \]
\[ i = 1, 2, ..., t \text{ and } j = 1, 2, ..., r \]

Where \( y_{ij} \) is the observed value for replicate \( j \) of treatment \( i \), \( \mu_i \) is the population mean for treatment \( i \), \( \mu \) is the population mean, \( \alpha_i \) is the effect of treatment \( i \) and \( e_{ij} \) is the experimental error resulting from replicate \( j \) of treatment \( i \).

**Assumption:** \( y_{ij} \) are assumed normally distributed about the mean, \( \mu_i \), and variance, \( \sigma^2 \) or \( e_{ij} \sim N(0, \sigma^2) \) i.e. independently and identically normally distributed with mean 0 and constant variance \( \sigma^2 \). Also \( \sum \alpha_i = 0 \),

**Estimation of the Parameters**

\[ y_{ij} = \mu + \alpha_i + e_{ij} \]
\[ e_{ij} = y_{ij} - \mu - \alpha_i \]

\[ S = \sum_{ij} e_{ij}^2 = \sum_{ij} (y_{ij} - \mu - \alpha_i)^2 \]

\[ \frac{dS}{d\mu} = -2 \sum_i \sum_j (y_{ij} - \mu - \alpha_i) \]

\[ \Rightarrow \quad -2 \sum_i \sum_j (y_{ij} - \hat{\mu} - \alpha_i) = 0 \]

\[ \sum_i \sum_j y_{ij} - \sum_i \sum_j \hat{\mu} - \sum_i \sum_j \alpha_i = 0 \]

\[ \sum_i \sum_j y_{ij} - rt \hat{\mu} - r \sum_i \alpha_i = 0 \]

Impose the constrain \( \sum \alpha_i = 0 \)

\[ \Rightarrow \quad rt \hat{\mu} = \sum_i \sum_j y_{ij} \]

\[ \therefore \quad \hat{\mu} = \frac{\sum_i \sum_j y_{ij}}{rt} = \bar{y}_{..} \]

\[ \frac{dS}{d\alpha_i} = -2 \sum_j (y_{ij} - \mu - \alpha_i) \]
\[
-2 \sum_j (y_{ij} - \hat{\mu} - \hat{\alpha}_i) = 0
\]
\[
\sum_j y_{ij} - \sum_j \hat{\mu} - \sum_j \hat{\alpha}_i = 0
\]
\[
\sum_j y_{ij} - r\hat{\mu} - r\hat{\alpha}_i = 0
\]
\[
\hat{\alpha}_i = \frac{\sum_j y_{ij}}{r} - \hat{\mu}
\]
\[
= \bar{y}_{i.} - \bar{y}_{..}
\]

Randomization Procedure

1. Determine the total number of experimental units or plots (N) where N = rt with r being the number of replications and t the number of treatments.

2. Assign a plot number to each experimental unit in any convenient manner consecutively 1 to N.

3. Assign the treatments to the experimental units by any chosen randomization scheme e.g. using table of random numbers, random number generator, drawing of lots and so on.

Data Structure

<table>
<thead>
<tr>
<th>Treatments</th>
<th>1</th>
<th>2</th>
<th>\ldots</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$y_{11}$</td>
<td>$y_{12}$</td>
<td>\cdots</td>
<td>$y_{1t}$</td>
</tr>
</tbody>
</table>
Analysis of Variance

The total variation in CRD is partitioned into two sources of variation i.e. variation due to treatment and variation due to the error. The relative size of the two variations is used to indicate whether the observed difference among the treatment means is significant or due to chance, the treatment difference is said to be significant if the treatment variation is significantly larger than the experimental error.

Total sum of squares, SST,

\[ SST = \sum_{i=1}^{t} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2 = \sum y_{ij}^2 - \frac{\bar{Y}^2}{N} \]

Sum of squares due to treatment SSB

\[ SSB = \sum_{i=1}^{t} \sum_{j=1}^{n_i} (\bar{y}_{ii} - \bar{y}_{..})^2 = \sum n_i (\bar{y}_{ii} - \bar{y}_{..})^2 = \sum n_i \frac{Y_{ii}^2}{N} - \frac{\bar{Y}^2}{N} \]

Sum of Square due to Error, SSE

\[ SSE = \sum_{i=1}^{t} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2 \]

\[ C.F = \frac{\bar{Y}^2}{N} = \text{correcting factor} \]

\[ \Rightarrow SST = SSB + SSE \]

\[ i.e. SSE = SST - SSB \]

ASSIGNMENT
Show that:
\[ \sum_{i} \sum_{j} (y_{ij} - \overline{y}_{i})^2 = \sum_{i} \sum_{j} (y_{ij} - \overline{y}_{.})^2 + \sum_{i} \sum_{j} (y_{ij} - \overline{y}_{.i})^2 \]

**ANOVA TABLE**

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degree of freedom</th>
<th>Sum of Squares</th>
<th>Means squares</th>
<th>F-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between treatment</td>
<td>t-1</td>
<td>SSB</td>
<td>MSB = ( \frac{SSB}{t-1} )</td>
<td>( Fc = \frac{MSB}{MSE} = Fc )</td>
</tr>
<tr>
<td>Error</td>
<td>N-t</td>
<td>SSE</td>
<td>MSE = ( \frac{SSE}{N-t} )</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>N-1</td>
<td>SST</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If there are no differences in the effect of the treatment \( Fc \) follows the \( F \)-distribution. Hence, if \( Fc > FT \) where \( FT \) is the table value from the \( F \)-Table with \( t – 1 \) and \( N – t \) degrees of freedom at a given significance level, then the effect are said to be significantly different

- \( H_0: \mu_1 = \mu_2 = \ldots = \mu_t = \mu \)
- \( H_1: \text{at least one } \mu_i \neq \mu \)

Or \( H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_t = 0 \)
- \( H_1: \text{at least one } \alpha_i \neq 0 \)

Reject \( H_0 \) if \( Fc > FT \)

**COMPARISON OF MEANS**

If a significant result is declared then there is need to identified the mean that are different and this can be done using multiple comparison of means such as

- LSD – Least Significant Difference
- DMRT – Duncan’s Multiple Range Test
- Turkey
- Scheffée etc.

LSD = tSED
If the observed difference between any two means is greater than the LSD value then those two means are said to be significantly different.

**COEFFICIENT OF VARIATION**

This is a measure of precision of the estimates obtained from the data. It is also used to assess the quality of the management of an experiment. A low coefficient of variation indicates high precision of estimate or efficient management of the experiment.

\[
CV = \sqrt{\frac{MSE}{\text{Grand mean overall}}} \times 100
\]

Example: In an effort to improve the quality of recording tapes, the effect of four kinds of coating A, B, C, D on the reproducing quality of sound are to be compared. The measurements of sound distortion are given below.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>14</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>8</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>31</td>
<td>14</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>15</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>15</td>
<td>18</td>
<td>15</td>
</tr>
</tbody>
</table>

Recommend the best coating for the sound production.

**ADVANTAGE OF CRD**

1. The design is very flexible
2. The statistical analysis is simple
3. It has high degrees of freedom relative to other designs
4. It is best for small experiments

DISADVANTAGE OF CRD
Design is very inefficient if units are not homogenous.

ASSIGNMENT
1. Analyze the following data from a field experiment with four treatments using 1% significance level. Carryout mean comparison if necessary. How good is the management of the experiment.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>14.3</td>
<td>11.6</td>
<td>11.8</td>
<td>14.2</td>
</tr>
<tr>
<td>B</td>
<td>20.7</td>
<td>21.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>32.6</td>
<td>32.1</td>
<td>33.0</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>24.3</td>
<td>25.2</td>
<td>24.8</td>
<td></td>
</tr>
</tbody>
</table>

2. Three fertilizer sources A, B, C, were each applied to seven plots chosen at random in a field of carrot. Analyze the data using 5% significance level. Carryout mean comparison if necessary. How efficient was the management of the experiment.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>24</td>
<td>18</td>
<td>18</td>
<td>29</td>
<td>22</td>
</tr>
<tr>
<td>B</td>
<td>46</td>
<td>39</td>
<td>37</td>
<td>50</td>
<td>44</td>
</tr>
<tr>
<td>C</td>
<td>32</td>
<td>20</td>
<td>26</td>
<td>41</td>
<td>36</td>
</tr>
</tbody>
</table>

RANDOMISED COMPLETE BLOCK DESIGN (RCBD)
Introduction

The design is used when the experimental unit can be grouped such that the number of units in a group is equal to number of treatments. The groups are called blocks or replicates and the purpose of grouping is to have units in a group as homogeneous as possible so that observed differences in a group are mainly due to treatments. Variability within group is expected to be lower than variability between groups. Since the number of units per block equal the number of treatments, the blocks are of equal size hence, the design is a complete block design. The primary purpose of blocking is to reduce the experimental error by eliminating the known sources of variability.

Model:

\[ y_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \]

\[ i = 1, 2, ..., t \quad \text{and} \quad j = 1, 2, ..., r \]

where \( y_{ij} \) is the observed value for block \( j \) of treatment \( i \), \( \mu \) is the population mean, \( \alpha_i \) is the effect of treatment \( i \), \( \beta_j \) is the effect of block \( j \) and \( e_{ij} \) is the experimental error resulting from block \( j \) of treatment \( i \).

Assumption:

- block and treatment effect are additive,
- \( e_{ij} \sim \mathcal{N}(0, \sigma^2) \)
- \( \sum \alpha_i = 0, \sum \beta_j = 0, \)

Estimation of Parameters

A procedure similar to that used in CRD can be utilized here to obtain the desired estimates.

Randomization and Layout

The randomization process for randomised complete block design is applied separately and independently to each of the block.

- Divide the experimental area into \( r \)-blocks.
- Sub-divide the block into \( t \)-experimental units. Where \( t \) is the number of treatments.
Number the plot consecutively from 1 to $t$ and assign the $t$-treatment at random to the $t$-unit within each block following any randomization scheme.

### DATA STRUCTURE

<table>
<thead>
<tr>
<th>Blocks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment</strong></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>$t$</td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
</tbody>
</table>

### Analysis of Variance

The total variation is partitioned into the variation due to blocks, variation due to treatments, and variation due to error. i.e.

$$
SS_{Total} = \sum \sum (y_{ij} - \bar{y}_{.})^2 = \sum \sum y_{ij}^2 - \frac{\bar{y}^2}{N}
$$

$$
SS_{Trt} = \sum \sum (y_{i.} - \bar{y}_{.})^2 = \sum \frac{y_{i.}^2}{r} - \frac{\bar{y}^2}{N}
$$

$$
SS_{Block} = \sum \sum (\bar{y}_{ij} - \bar{y}_{.})^2 = \sum \frac{y_{i.}^2}{t} - \frac{\bar{y}^2}{N}
$$

$$
SSE = SS_{Total} - SS_{Trt} - SS_{Block}
$$
### ANOVA TABLE

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block</td>
<td>r - 1</td>
<td>SS_Block</td>
<td>MSB = $\frac{SS_B}{r - 1}$</td>
<td>$\frac{MSB}{MSE} = F_{Blok}$</td>
</tr>
<tr>
<td>Treatment</td>
<td>t - 1</td>
<td>SS_Treatment</td>
<td>MST = $\frac{SS_T}{t - 1}$</td>
<td>$\frac{MST}{MSE} = F_{Trt}$</td>
</tr>
<tr>
<td>Error</td>
<td>(r-1)(t-1)</td>
<td>SSE</td>
<td>MSE = $\frac{SS}{(r - 1)(t - 1)}$</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>rt - 1</td>
<td>SS_Total</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Hypothesis**

$H_{10}: \beta_2 = \beta_3 = \cdots = \beta_r = 0$

$H_{11}: \text{at least one } \beta_j \neq 0$

Or $H_{20}: \alpha_1 = \alpha_2 = \cdots = \alpha_s = 0$

$H_{21}: \text{at least one } \alpha_i \neq 0$

Comparing the calculated F-ratios to the table F-value at a given significance level, we decide to reject or fail to reject the null hypothesis. i.e. Reject $H_{10}$ if $F_{Blok} > F_{T1}$

Reject $H_{20}$ if $F_{Trt} > F_{T2}$

### COMPARISON OF MEANS

Use LSD to compare the treatments if the F-ratio is found to be significant.

where $LSD = t \cdot SED = t \cdot \sqrt{\frac{MSE}{r}}$

Coefficient of variation

$$CV = \frac{MSE}{\text{overall mean}} \times 100$$

### CAUSES OF MISSING VALUES AND THEIR ESTIMATIONS

A missing data can occur whenever a valid observation is not available for any one of the experimental units, occurrence of missing data result in two major problems i.e. loss of information and non applicability of standard analysis of variance.
COMMON CAUSES OF MISSING DATA include:

1. When intended treatment is not applied i.e. improper treatment.
2. When experimental plants are destroyed probably due to poor germination, physical damage, pest damage etc. This causes total or high percentage of the plants in a plot to be destroyed such that no meaningful observation can be made on the plot.
3. Loss of harvested sample: This may result from the fact that some plant characters cannot be conveniently recorded either in the field or immediately after harvest due to some other process required. Hence some samples will be lost between the time of harvesting and actual recording of data.
4. This happens after data have been recorded and transcribed generally referred to a illogical data. The value may be too extreme as a result of misread observation or incorrect transcription.

ESTIMATION OF MISSING VALUE FROM RCBD

Let \( x \) be the missing value

\[
\begin{align*}
Y_{ij} & = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \quad i = 1, \ldots, I, \\
\alpha_i & = \mu - \mu - \alpha_i - \beta_j, \\
\sigma_{ij} & = \left( \mu - \mu - \alpha_i - \beta_j \right)^2.
\end{align*}
\]

\[
\frac{\partial^2}{\partial x^2} = 2(x - \mu - \alpha_i - \beta_j),
\]

\[
\Rightarrow \hat{x} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j = 0,
\]

\[
\begin{align*}
\gamma_i &= \frac{\sum Y_i - \bar{Y}}{N} = \frac{\sum \sigma_{ij}}{N} \\
\alpha_i &= \frac{\sum Y_i - \bar{Y}}{N} = \frac{\sum \sigma_{ij}}{N} \\
\beta_j &= \frac{\sum Y_j - \bar{Y}}{N} = \frac{\sum \sigma_{ij}}{N}
\end{align*}
\]
\[ \hat{\mathbf{c}} = \mathbf{T}_0 - \mathbf{B}_0 = \left( \frac{1}{r} \mathbf{T}_0 - \mathbf{B}_0 \right) \]

\[ = 0 \]

\[ = r \left( t - t_0 \right) \hat{\mathbf{c}} = t_0 + r \mathbf{B}_0 - \mathbf{G}_0 \]

\[ = r \left( r - 1 \right) \left( t - 1 \right) \hat{\mathbf{c}} = t_0 + r \mathbf{B}_0 - \mathbf{G}_0 \]

\[ \hat{\mathbf{c}} = \frac{t_0 + r \mathbf{B}_0 - \mathbf{G}_0}{(r - 1)(t - 1)} \]

where \( \mathbf{G}_0 \) is the grand total excluding the missing value.

- \( \mathbf{T}_0 \) is the total observed value for the treatment that contained the missing value.
- \( \mathbf{B}_0 \) is the total observed value for the Block that contained the missing value.

**Note that:** the degree of freedom must be adjusted by the number of missing values i.e. reduce the number of degrees of freedom by the number of missing values.

**ADVANTAGES OF RANDOMIZED COMPLETE BLOCK DESIGNS**

1. A reduction of experiment error due to blocking is expected.
2. Estimation of missing value is easy to compute.
3. The ANOVA is also easy to compute.

**DISADVANTAGES OF RCBD**

1. Not best for large number of treatments.
2. More tasking in the execution of the design than the CRD.
3. Missing value can create problem especially in estimation and non formal analysis.
4. The precision will be affected due to missing values.

**RELATIVE EFFICIENCY**

Blocking maximizes the difference among blocks. Hence it is necessary to examine how much is gained by the introduction of blocking into the design. The magnitude of the
reduction in the experimental error due to blocking over the CRD can be obtained by computing relative efficiency. 

\[ R.E = \frac{(r-1)S_B + (r-1)S_E}{(r-1)S_E} \times 100 \]

Where \( S_B \) is the block mean square and the \( S_E \) is the error mean square.

1. **Example:** In an experiment to examine the respond of maize to nutrient fertilizer application. Six treatments were used in four blocks. Analyze the data and recommend the appropriate fertilizer rate.

<table>
<thead>
<tr>
<th></th>
<th>F_1</th>
<th>F_2</th>
<th>F_3</th>
<th>F_4</th>
<th>F_5</th>
<th>F_6</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0.42</td>
<td>0.46</td>
<td>0.60</td>
<td>2.63</td>
<td>2.81</td>
<td>3.39</td>
<td>10.31</td>
</tr>
<tr>
<td>II</td>
<td>0.22</td>
<td>0.42</td>
<td>0.86</td>
<td>2.39</td>
<td>2.82</td>
<td>2.50</td>
<td>9.22</td>
</tr>
<tr>
<td>III</td>
<td>0.31</td>
<td>0.58</td>
<td>0.83</td>
<td>1.68</td>
<td>1.95</td>
<td>2.50</td>
<td>7.85</td>
</tr>
<tr>
<td>IV</td>
<td>0.33</td>
<td>0.55</td>
<td>0.89</td>
<td>2.22</td>
<td>2.81</td>
<td>2.10</td>
<td>8.9</td>
</tr>
<tr>
<td>TOTAL</td>
<td>1.28</td>
<td>2.01</td>
<td>3.18</td>
<td>8.92</td>
<td>10.40</td>
<td>10.49</td>
<td>36.28</td>
</tr>
</tbody>
</table>

2. The following data are yield of groundnut in a variety trial involving five varieties of groundnut using four replications in randomized complete block design. The data has one missing value. Analyze the data and make your recommendation.

<table>
<thead>
<tr>
<th>V_1</th>
<th>V_2</th>
<th>V_3</th>
<th>V_4</th>
<th>V_5</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>2.73</td>
<td>1.73</td>
<td>3.00</td>
<td>2.13</td>
</tr>
<tr>
<td>II</td>
<td>*</td>
<td>2.27</td>
<td>1.53</td>
<td>2.40</td>
</tr>
<tr>
<td>III</td>
<td>2.33</td>
<td>2.00</td>
<td>2.13</td>
<td>1.87</td>
</tr>
<tr>
<td>IV</td>
<td>2.53</td>
<td>2.00</td>
<td>2.00</td>
<td>1.47</td>
</tr>
<tr>
<td></td>
<td>7.59</td>
<td>8.00</td>
<td>8.66</td>
<td>7.87</td>
</tr>
</tbody>
</table>

**ASSIGNMENT**

In an experiment to test the effect of five level of potash (ABCDE) on the yield of cotton the following strength indices were obtain as given below. One of the data point is missing. Analyze the data and compare the mean of the five level of potash and make necessary recommendation. How effectively was the experiment carried out. Was there any gain in precision in using RCBD over CRD
LATIN SQUARE DESIGN

Introduction
The major feature of the latin square design is its capacity to simultaneously handle two known sources of variation among the experimental units. These are commonly referred to as row blocking and column blocking. It is ensured that every treatment occurs only once in each row and once in each column. Hence the variation due to row and column can be estimated and removed from the experimental error.

Note that: the presence of row and column blocking also constitute a restriction. This is due to the requirement that all treatment appear in each row and in each column. This is only satisfied if the number of replications equal the number of treatments. Hence, for large number of treatments, the design is not practicable. Also when the number of treatments is small the degrees of freedom for the error becomes too small for the error to be reliably estimated, the design is therefore not generally, widely adopted.
Model:
\[ y_{ijk} = \mu + \alpha_i + \beta_j + \delta_k + \epsilon_{ijk} \]
where \( i, j, k = 1, 2, \ldots, t \)
where \( y_{ijk} \) is the observed value from row \( j \) and column \( k \) receiving treatment \( i \).
- \( \mu \) is the overall mean
- \( \alpha_i \) is the effect of treatment \( i \)
- \( \beta_j \) is the effect of row \( j \)
- \( \delta_k \) is the effect of row \( k \)
- \( \epsilon_{ijk} \) is the random error component for row \( j \) and column \( k \) receiving treatment \( i \).

Assumptions:
The model is completely additive i.e. there is no interaction between the rows, columns and treatments.
\[ \epsilon_{ijk} \overset{d}{\sim} N(0, \sigma^2) \]
\[ \sum_i \alpha_i = 0, \quad \sum_j \beta_j = 0 \quad \text{and} \quad \sum_k \delta_k = 0 \]

Estimation of Parameters
A procedure similar to that used in CRD can be utilized here to obtain the desired estimates.

RANDOMIZATION PROCEDURE
1. Obtain a square field partitioned into \( t \) rows and \( t \) columns.
2. Arrange the treatment into the unit in a standard form.
3. Randomize between the columns
4. Randomize between the rows

Example: Consider an experiment with four treatments to be compared using latin square design i.e. \( 4 \times 4 \) LS
**ANALYSIS OF VARIANCE**

The total variation is partitioned into components for row, column, treatment and error. The sum of squares are obtained in the usual form.

**ANOVA TABLE**

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rows</td>
<td>t-1</td>
<td>SSR</td>
<td>MSR</td>
<td>( \frac{MSR}{MSE} )</td>
</tr>
<tr>
<td>Columns</td>
<td>t-1</td>
<td>SSC</td>
<td>MSC</td>
<td>( \frac{MSC}{MSE} )</td>
</tr>
<tr>
<td>Treatments</td>
<td>t-1</td>
<td>SS_{Trt}</td>
<td>MS_{Trt}</td>
<td>( \frac{MS_{Trt}}{MSE} )</td>
</tr>
<tr>
<td>Error</td>
<td>(t-1) (t-2)</td>
<td>SSE</td>
<td>MSE</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>t^2-1</td>
<td>SST</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**BLOCKING EFFICIENCY**

The efficiency of both row and column blockings in a latin square design indicate the gain in precision relative to either the CRD or RCBD.
**RELATIVE EFFICIENCY OF LSD TO CRD**

Relative efficiency of a latin square design as compared to CRD is given by

\[
R.E(CRD) = \frac{E_r + E_c + (t-1)E_e}{t+1} \times 100
\]

Where \(E_r\), \(E_c\), \(E_e\) are the mean squares row, column, and error respectively with \(t\) as the number of treatment. For an \(R.E = 325\%\) it indicates that the use of LSD is estimated to increase the experimental precision by 225\% while if the R.E is less than 100\% means that there is no gain.

**RELATIVE EFFICIENCY LSD TO RCBD**

Relative efficiency of latin square design as compared to RCBD can be computed in two ways i.e. when rows are considered as blocks and when columns are considered as blocks of the RCBD.

\[
\begin{align*}
1) \quad R.E(RCBD,\text{ROW}) & = \frac{E_r + (t-1)E_e}{tE_e} \times 100 \\
2) \quad R.E(RCBD,\text{Column}) & = \frac{E_c + (t-1)E_e}{tE_e} \times 100
\end{align*}
\]

**Example:** Suppose we have \(R.E(RCBD,\text{ROW}) = 87\%\) \(R.E(RCBD,\text{Col}) = 994\%\).

Result indicate that, the additional column blocking by use of latin square design is estimated to have increased the precision over that of the RCBD with row as block by 294\%. However, the additional row blocking in the LS Design did not increase the precision over the RCBD with column as blocks. Hence for this experiment the Randomized Complete Block Design with column as blocks would have been as efficient as the LS Design.

**MISSING VALUE ESTIMATION**

Missing value in the latin square experiment can be estimated as follows
\[ \hat{e} = \frac{\bar{t}(R_0 + C_0 + T_0) - 3G_0}{(t-1)(t-2)} \]

Where the \( R_0, C_0, T_0 \) are the total of the row, column and treatment respectively that contain the missing observation. Again one degree of freedom is subtracted from both total and error degrees of freedom in the case of one missing value.

**ADVANTAGES AND DISADVANTAGES**

1. The elimination of two sources of variation often lead to a smaller error mean square than would be obtained by use of CRD and RCBD.
2. ANOVA is simple
3. Missing values can easily be handled

Disadvantages.

1. Assumption of no interaction between different factors may not hold.
2. Unlike the CRD and RCBD, the number of treatments is restricted to the number of replications. Hence it is limited in application.
3. For large number of treatments such as \( t \geq 12 \), the square becomes too large and does not remain homogeneous.
4. For small number of treatments such as \( t \leq 3 \), degrees of freedom for the error is usually too small for any meaningful comparison or conclusion.
5. A square field is often required for the design and this may not be practicable.

**Example:** The following show the field layout and yield of a 5 x 5 latin square experiment on the effect of spacing on yield of millet, the spacing are: A(2cm), B(4cm), C(6cm), D(8cm) and E(10cm)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B: 257</td>
<td>E: 230</td>
<td>A: 279</td>
<td>C: 287</td>
<td>D: 202</td>
<td>1255</td>
</tr>
<tr>
<td>2</td>
<td>D: 245</td>
<td>A: 283</td>
<td>E: 245</td>
<td>B: 280</td>
<td>C: 260</td>
<td>1313</td>
</tr>
<tr>
<td>3</td>
<td>E: 182</td>
<td>B: 252</td>
<td>C: 280</td>
<td>D: 246</td>
<td>A: 250</td>
<td>1210</td>
</tr>
<tr>
<td>4</td>
<td>A: 203</td>
<td>C: 204</td>
<td>D: 227</td>
<td>E: 193</td>
<td>B: 259</td>
<td>1086</td>
</tr>
<tr>
<td>5</td>
<td>C: 231</td>
<td>D: 271</td>
<td>B: 266</td>
<td>A: 334</td>
<td>E: 339</td>
<td>1440</td>
</tr>
</tbody>
</table>
GRAECO LATIN SQUARE DESIGN (GLSD)

Introduction
In the Graeco latin square design, the treatments are grouped into replicates in three different ways: this triple grouping is to eliminate from the error, three sources of blocking of blocking variation. Recall the earlier design:

- CRD: No blocking variation
- RCBD: Single blocking variation
- Latin Square: Double blocking variation
- Graeco latin: Tripple blocking variation

Thus GLSD provide more opportunity than the other designs in the reduction of error through skillful planning. The experimental unit should be arranged and the experiment conducted so that differences in the three directions represent major sources of variation.

RANDOMIZATION
Arrange the rows and the columns independently at random including the treatment. Assign the subscript at random to their respective classification. The treatment and subscript must appear once in a column and once in a row, and must appear together only once. The layout can be difficult to design but like in the latin square design, once we obtain a standard form then we can randomize between the columns and between the rows.

<table>
<thead>
<tr>
<th>I</th>
<th>II</th>
<th>III</th>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>A1</td>
<td>B3</td>
<td>C2</td>
<td>C3</td>
<td>A2</td>
</tr>
<tr>
<td>II</td>
<td>B2</td>
<td>C1</td>
<td>A3</td>
<td>A1</td>
<td>B3</td>
</tr>
<tr>
<td>III</td>
<td>C3</td>
<td>A2</td>
<td>B1</td>
<td>B2</td>
<td>C1</td>
</tr>
</tbody>
</table>

Randomize between rows

<table>
<thead>
<tr>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2</td>
<td>B1</td>
<td>C3</td>
</tr>
</tbody>
</table>
B_3  C_2  A_1  
C_1  A_3  B_2

Randomize between columns

T = 4  A, B, C, D

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>A_1</td>
<td>C_4</td>
<td>B_3</td>
<td>D_2</td>
</tr>
<tr>
<td>II</td>
<td>B_2</td>
<td>D_3</td>
<td>A_4</td>
<td>C_1</td>
</tr>
<tr>
<td>III</td>
<td>C_3</td>
<td>A_2</td>
<td>D_1</td>
<td>B_4</td>
</tr>
<tr>
<td>IV</td>
<td>D_4</td>
<td>B_1</td>
<td>C_2</td>
<td>A_3</td>
</tr>
</tbody>
</table>

T = 5  A, B, C, D, E

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_1</td>
<td>B_3</td>
<td>C_5</td>
<td>D_2</td>
<td>E_4</td>
<td></td>
</tr>
<tr>
<td>B_2</td>
<td>C_4</td>
<td>D_1</td>
<td>E_3</td>
<td>A_5</td>
<td></td>
</tr>
<tr>
<td>C_3</td>
<td>D_5</td>
<td>E_2</td>
<td>A_4</td>
<td>B_1</td>
<td></td>
</tr>
<tr>
<td>D_4</td>
<td>E_1</td>
<td>A_3</td>
<td>B_5</td>
<td>C_2</td>
<td></td>
</tr>
<tr>
<td>E_5</td>
<td>A_2</td>
<td>B_4</td>
<td>C_1</td>
<td>D_3</td>
<td></td>
</tr>
</tbody>
</table>

**MODEL**

\[ y_{ijkl} = \mu + \alpha_i + \beta_j + \delta_k + \tau_l + e_{ijkl} \]

where  \( i, j, k, l = 1, 2, \ldots, t \)

where \( y_{ijkl} \) is the observed value from row \( j \) and column \( k \) receiving treatment \( i \).

\( \mu \) is the grand mean

\( \alpha_i \) is the effect of treatment \( i \)
\( \beta_j \) is the effect of row \( j \)
\( \phi_k \) is the effect of column \( k \)
\( \tau_l \) is the effect of subscript factor \( l \).
\( \epsilon_{ijkl} \) is the random error component

**Assumptions:**
The model is completely additive i.e. no interaction between the row, the column, the subscript factor and the treatment
\[ \epsilon_{ijkl} \sim N(0, \sigma^2) \]
\[ \sum \alpha_i = 0, \sum \beta_j = 0, \sum \phi_k = 0 \text{ and } \sum \tau_l = 0 \]

**Estimation of Parameters**
A procedure similar to that used in CRD can be utilized here to obtain the desired estimates.

**STATISTICAL ANALYSIS**
The total variation is partitioned into five components i.e. the row, column, subscript factor, treatment and error.

The sum of squares are obtained in the usual form i.e.

\[ \text{SST} = \sum y^2 \frac{r}{N} - \frac{\sum y^2 \cdot \sum r}{N} \]
\[ \text{SSR} = \sum y^2 \frac{r}{t} - \frac{\sum y^2 \cdot \sum r}{N} \]
\[ \text{SSR} = \sum y^2 \frac{t}{r} - \frac{\sum y^2 \cdot \sum t}{N} \]
\[ \text{SSR} = \sum y^2 \frac{t}{r} - \frac{\sum y^2 \cdot \sum t}{N} \]
\[ \text{SSSubscript} = \sum y^2 \frac{t}{r} - \frac{\sum y^2 \cdot \sum t}{N} \]
\[ SSE = SST - SS_{TRT} - SSR - SS_{C} - SS_{Subscript} \]

**ANOVA TABLE**

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rows</td>
<td>t-1</td>
<td>SSR</td>
<td>MSR</td>
<td>MSR/MSR</td>
</tr>
<tr>
<td>Column</td>
<td>t-1</td>
<td>SSC</td>
<td>MSC</td>
<td>MSC/MSC</td>
</tr>
<tr>
<td>Subscript</td>
<td>t-1</td>
<td>SS_{Subscript}</td>
<td>MS_{Subscript}</td>
<td>MS_{Subscript}/MSE</td>
</tr>
<tr>
<td>Treatments</td>
<td>t-1</td>
<td>SS_{Trt}</td>
<td>MS_{Trt}</td>
<td>MS_{Trt}/MSE</td>
</tr>
<tr>
<td>Error</td>
<td>(t-1) (t-3)</td>
<td>SSE</td>
<td>MSE</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>t^2-1</td>
<td>SST</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note the followings:**

1. The number of replications equals the number of treatments hence for large number of treatments, the design is not practicable \((t > 12)\).
2. The experimental error is likely to increase with the size of the square.
3. Small square provide only a few degree of freedom for the error.
4. Experimental units are difficult to balance conveniently in all the three groupings.

**Example**

Consider the following data which was obtain from an experiment to study the sampling error, the data consist of five samplers (A – E) being controlled for order of sampling, area of sampling and qualification of sampler. Analyse the data.

<table>
<thead>
<tr>
<th>Order</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A1 (3.5)</td>
<td>B3 (4.2)</td>
<td>C5 (6.7)</td>
<td>D2 (6.6)</td>
<td>E4 (4.1)</td>
<td>25.1</td>
</tr>
<tr>
<td>2</td>
<td>B2 (8.9)</td>
<td>C4 (1.9)</td>
<td>D1 (5.8)</td>
<td>B3 (4.5)</td>
<td>A5 (2.4)</td>
<td>23.5</td>
</tr>
<tr>
<td>3</td>
<td>C3 (9.6)</td>
<td>D5 (1.7)</td>
<td>E2 (2.7)</td>
<td>A4 (3.7)</td>
<td>B1 (6.0)</td>
<td>25.7</td>
</tr>
<tr>
<td>4</td>
<td>D4 (10.6)</td>
<td>E1 (10.2)</td>
<td>A3 (4.6)</td>
<td>B5 (3.7)</td>
<td>C2 (5.1)</td>
<td>34.1</td>
</tr>
<tr>
<td>5</td>
<td>E5 (3.1)</td>
<td>A2 (7.2)</td>
<td>B4 (4.0)</td>
<td>C1 (3.3)</td>
<td>D3 (3.5)</td>
<td>21.1</td>
</tr>
<tr>
<td>Total</td>
<td>35.6</td>
<td>27.2</td>
<td>23.8</td>
<td>21.8</td>
<td>21.1</td>
<td>129.5</td>
</tr>
</tbody>
</table>
ASSIGNMENT
The data below is obtained from an experiment using Graeco Latin Square Design with four diet, (A, B, C, D). breed I, II, III, IV weight group {1, 2, 3, 4} feed concentration \{i \, ii \, iii \, iv\}. Is there any significant difference between the diets. If any, compare them and make necessary recommendation. Also comment on the management of the experiment.

<table>
<thead>
<tr>
<th>Breeds</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A_1 (5.9)</td>
<td>B_3 (4.2)</td>
<td>C_4 (10.2)</td>
<td>D_2 (6.6)</td>
</tr>
<tr>
<td>2</td>
<td>B_2 (8.9)</td>
<td>A_4 (4.5)</td>
<td>D_3 (6.0)</td>
<td>C_1 (3.0)</td>
</tr>
<tr>
<td>3</td>
<td>C_3 (9.6)</td>
<td>D_1 (5.8)</td>
<td>A_2 (7.2)</td>
<td>B_4 (4.6)</td>
</tr>
<tr>
<td>4</td>
<td>D_4 (10.5)</td>
<td>C_2 (4.1)</td>
<td>B_1 (3.5)</td>
<td>A_3 (6.7)</td>
</tr>
</tbody>
</table>

SIMPLE FACTORIAL EXPERIMENT

Introduction
Factorial experiments are used in the study of the effects of two or more factors. In factorial experiments, all the possible combinations of the level of the factors make up the treatments. For example, if there are two factor A, B each with ‘a’ and ‘b’ levels respectively, then we have ‘\text{ab}’ treatment combinations.

\[
\begin{array}{cc}
A & B \\
\text{Maize variety} & \text{Fertilizer Rate} \\
a_0 \text{ White} & \begin{array}{c}
b_0 \text{ 0 kg/ha} \\
b_1 \text{ 30 kg/ha} \\
b_2 \text{ 60 kg/ha} \\
b_3 \text{ 90 kg/ha} \\
b_4 \text{ 120 kg/ha}
\end{array} \\
a_1 \text{ Yellow} & \\
\end{array}
\]

There are 10 treatment combinations. Also factorial experiments allow us to investigate the interaction between the factors. That is, how the levels of a factor perform in the presence of the levels of another factor. We are able to answer the question on how the responses to one factor were affected by another.
In analyzing data from a factorial experiment, we would be interested in the main effect and the interaction effect of a factor. The main effect of a factor is a measure of the change in response in the level of a factor averaged over all levels of the other factors. For example, let two factors A and B be at two levels $a_0, a_1$ and $b_0, b_1$ respectively with treatment combinations $a_0b_0, a_0b_1, a_1b_0, a_1b_1$. The main effect of A is a measure of change in A from $a_0$ to $a_1$ averaged over the two levels of B.

i.e. At level $b_0$ of B: the simple effect of A is $a_1b_0 - a_0b_0$.

Similarly, at level $b_1$ of B: the simple effect of A is $a_1b_1 - a_0b_1$.

Main effect of A = $\frac{1}{2r} \left[ (a_1b_1 - a_0b_1) + (a_1b_0 - a_0b_0) \right]$

The $r$ represent the replication, where each treatment total response is from $r$ units.

Also at level $a_0$ of A: simple effect of B is $a_0b_1 - a_0b_0$.

Similarly, at level $a_1$ of A simple effect of B is $a_1b_1 - a_1b_0$

Thus

Main effect of B = $\frac{1}{2r} \left[ (a_1b_1 - a_0b_1) + (a_1b_0 - a_0b_0) \right]$

Each effect of a factor at a given level of the other factor is known as simple effect. The interaction effect is the differential response to one factor in combination with varying levels of a second factor. That is, an additional effect due to the combined influence of two or more factors. For example, interaction between A and B (AB) is estimated as the difference between two simple effects. $(a_1b_1 - a_0b_1)$ and $(a_1b_0 - a_0b_0)$

i.e. the interaction effect $AB = \frac{1}{2r} \left[ (a_1b_1 - a_0b_1) + (a_1b_0 - a_0b_0) \right]$

$BA = \frac{1}{2r} \left[ (a_1b_1 - a_0b_1) + (a_1b_0 - a_0b_0) \right]$

Consider the following result

<table>
<thead>
<tr>
<th>Mean Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>b_0</td>
</tr>
<tr>
<td>a_0</td>
</tr>
<tr>
<td>a_1</td>
</tr>
</tbody>
</table>

(i) (ii) (iii)
From above illustration (i) shows the case of no interaction, (ii) shows the case of mild interaction and (iii) shows the case of strong interaction.

**LAYOUT**

Any of the earlier design discussed can be used, in particular, the RCBD. The treatment combinations are assigned to each block randomly.

For example, consider the case of two factors A and B, each at two levels $a_0, a_1$ and $b_0, b_1$ respectively. The treatment combinations are $a_0b_0, a_0b_1, a_1b_0, a_1b_1$. 

\[
\begin{align*}
    m.e \ of \ A &= \frac{1}{2}(20 + 20) = 20 \\
    m.e \ of \ B &= \frac{1}{2}(10 + 10) = 10 \\
    m.e \ of \ A &= \frac{1}{2}(20 + 20) = 20 \\
    AB &= \frac{1}{2}(20 - 20) = 0 \\
    &= \frac{1}{2}(10 - 10) = 0 \\
    m.e \ of \ A &= \frac{1}{2}(10 + 20) = 15 \\
    m.e \ of \ B &= \frac{1}{2}(10 + 20) = 15 \\
    AB &= \frac{1}{2}(20 - 10) = 5 \\
    m.e \ of \ A &= \frac{1}{2}(30 - 10) = 10 \\
    m.e \ of \ B &= \frac{1}{2}(-20 + 20) = 0 \\
    AB &= \frac{1}{2}(-10 - 30) = -20
\end{align*}
\]
### MODEL CRD

\[
y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha \beta)_{ij} + \epsilon_{ijk}
\]

where \(y_{ijk}\) is the observed value from the \(k^{th}\) unit corresponding to level \(i\) of factor A and the \(j^{th}\) level of B.

- \(\mu\) is the grand mean.
- \(\alpha_i\) is the effect of level \(i\) of factor A.
- \(\beta_j\) is the effect of the \(j^{th}\) of factor B.
- \((\alpha \beta)_{ij}\) is the effect of interaction between \(\alpha_i\) and \(\beta_j\).
- \(\epsilon_{ijk}\) is the random error component.

### MODEL RCBD

\[
y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha \beta)_{ij} + T_k + \epsilon_{ijk}
\]

Where \(T_k\) is the effect of the \(k^{th}\) block.

\[
\sum \alpha_i = 0, \quad \sum \beta_j = 0, \quad \sum (\alpha \beta)_{ij} = 0
\]

- \(H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_t = 0\)
- \(H_1: \text{at least one } \alpha_i \neq 0\)
- \(H_{20}: \beta_1 = \beta_2 = \ldots = \beta_r = 0\)
- \(H_{21}: \text{at least one } \beta_j \neq 0\)
- \(H_{30}: (\alpha \beta)_{ij} = 0\)
- \(H_{31}: (\alpha \beta)_{ij} \neq 0\) \{there is interaction between A and B\}

### ANALYSIS OF VARIANCE
The total variation is partitioned into that due to factor A, factor B, the interaction AB and the error i.e.

$$SST = SSA + SSB + SSAB + SSE$$

where the SST, SSA, SSB, SSAB and SSE are obtain in the usual form i.e.

$$SST = \sum_{i} \sum_{r} \left( y_{ir} - \frac{\sum_{i} y_{ir}^{2}}{abr} \right)$$

$$SSA = \sum_{r} \sum_{i} \left( y_{ir} - \frac{\sum_{i} y_{ir}^{2}}{abr} \right)$$

$$SSB = \sum_{i} \sum_{r} \left( y_{ir} - \frac{\sum_{i} y_{ir}^{2}}{abr} \right)$$

$$SSAB = \sum_{r} \sum_{i} \left( y_{ir} - \frac{\sum_{i} y_{ir}^{2}}{abr} \right)$$

$$-SSA - SSB$$

$$SSE = SST - SSA - SSB - SSAB$$

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>a-1</td>
<td>SSA</td>
<td>MSA</td>
<td>MSA/MSE</td>
</tr>
<tr>
<td>B</td>
<td>b-1</td>
<td>SSB</td>
<td>MSB</td>
<td>MSB/MSE</td>
</tr>
<tr>
<td>AB</td>
<td>(a-1) (b-1)</td>
<td>SSAB</td>
<td>MSAB</td>
<td>MSAB/MSE</td>
</tr>
<tr>
<td>ERROR</td>
<td>ab(r-1)</td>
<td>SSE</td>
<td>MSE</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>abr-1</td>
<td>SST</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Example**

An engineer designing a battery for use in a device that would be subjected to some extreme variation in temperature has three types of plate materials to use. He decided to test the plate materials under three temperature settings ($15^0F$, $70^0F$, $125^0F$) to see their effect on the life of a battery. Four test runs are to be made at each treatment combination. Test the effect of temperature and plate material and their possible interaction on the battery life.
Temperature

<table>
<thead>
<tr>
<th>Type</th>
<th>$15^0$F</th>
<th>$70^0$F</th>
<th>$125^0$F</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>130 155</td>
<td>34 40</td>
<td>20 70</td>
</tr>
<tr>
<td></td>
<td>74 180</td>
<td>60 675</td>
<td>82 58</td>
</tr>
<tr>
<td></td>
<td>539</td>
<td>229</td>
<td>230 998</td>
</tr>
<tr>
<td>B</td>
<td>150 188</td>
<td>126 122</td>
<td>25 70</td>
</tr>
<tr>
<td></td>
<td>159 126</td>
<td>106 115</td>
<td>58 46</td>
</tr>
<tr>
<td></td>
<td>623</td>
<td>469</td>
<td>198 1290</td>
</tr>
<tr>
<td>C</td>
<td>1382 110</td>
<td>174 120</td>
<td>96 104</td>
</tr>
<tr>
<td></td>
<td>168 160</td>
<td>150 139</td>
<td>62 60</td>
</tr>
<tr>
<td></td>
<td>576</td>
<td>583</td>
<td>342 1501</td>
</tr>
<tr>
<td></td>
<td>1738</td>
<td>1281</td>
<td>770 3789</td>
</tr>
</tbody>
</table>

ASSIGNMENT

Three types of tyres are to be compared using four different brand of cars. The threading on the tyres are measured after a period of use. Below is the rescaled data. Are the tyres significantly different? Also does the performance of the tyres depend on the brand of the car?

Tyres

<table>
<thead>
<tr>
<th>Car brand</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>