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

(I)	Y 1	W 2	Y 3	W 4	(II)	Y 1	Y 2	Y 3	Y 4
	Y 5	W 6	W 7	Y 8		W 5	W 6	W 7	W 8

    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 advantage been 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 measure 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, μ_i is the population mean for treatment i, μ is the population mean, α_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, μ_i, and variance, σ^2 or $e_{ij} \overset{iid}{\sim} N(0, \sigma^2)$ i.e. independently and identically normally distributed with mean 0 and constant variance σ^2. Also $\sum \alpha_i = 0$,

Estimation of the Parameters

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

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

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

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

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

Impose the constrain $\sum_i \alpha_i = 0$

$$\Rightarrow rt\hat{\mu} = \sum y_{ij}$$

$$\therefore \hat{\mu} = \frac{\sum \sum y_{ij}}{rt} = \overline{y_{..}}$$

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

$$-2\sum_{j}(y_{ij} - \hat{\mu} - \hat{\alpha}_i) = 0$$

$$\sum_{j} y_{ij} - \sum_{j}\hat{\mu} - \sum_{j}\hat{\alpha}_i = 0$$

$$\sum_{j} y_{ij} - r\hat{\mu} - r\hat{\alpha}_i = 0$$

$$\hat{\alpha}_i = \frac{\sum_{j} y_{ij}}{r} - \hat{\mu}$$

$$= \bar{y}_{i.} - \bar{y}_{..}$$

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

        Treatments

	1	2...	T	
1	y_{11}	$y_{12} \cdots$	y_{t1}	$y_{.1}$

2	y_{12}	y_{22} $\cdots$	y_{t2}	$y_{.2}$
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
r	y_{1r}	y_{2r} $\cdots$	y_{tr}	$y_{.r}$
	$y_{1.}$	$y_{2.}$ $\cdots$	$y_{t.}$	$y_{y.}$

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,

$$SST = \sum_{i=1}^{t}\sum_{j=1}^{ni}(y_{ij} - \bar{y}..)^2 = \sum\sum y_{ij}^2 - \frac{y_{..}^2}{N}$$

Sum of squares due to treatment SSB

$$SSB = \sum_{i=1}^{t}\sum_{j=1}^{ni}(\bar{y}_{ii.} - \bar{y}..)^2 = \sum_{i=1}^{t}n_i(\bar{y}_{i.} - \bar{y}_{..})^2 = \sum_{i=1}^{t}\frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N}$$

Sum of Square due to Error, SSE

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

$$C.F = \frac{y_{..}^2}{N} = \text{correcting factor}$$

$$\Rightarrow SST = SSB + SSE$$

$$i.e. SSE = SST - SSB$$

ASSIGNMENT

- Show that:

$$\sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = \sum_i \sum_j (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2$$

ANOVA TABLE

Source of variation	Degree of freedom	Sum of Squares	Means squares	F-ratio
Between treatment	t-1	SSB	$MSB = \dfrac{SSB}{t-1}$	$Fc = \dfrac{MSB}{MSE} = Fc$
Error	N-t	SSE	$MSE = \dfrac{SSE}{N-t}$	
Total	N-1	SST		

If there are no differences in the effect of the treatment Fc follows the F-distribution. Hence, if $F_c > F_T$ where F_T 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

$H_0: \mu_1 = \mu_2 = \cdots = \mu_t = \mu$

$H_1:$ atleast one $\mu_i \neq \mu$

Or $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_t = 0$

$H_1:$ atleast one $\alpha_i \neq 0$

Reject H_0 if $F_C > F_T$

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

Scheffee etc.

LSD = tSED

$$SED = \sqrt{MSE \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

$$= \sqrt{\frac{2MSE}{r}}, \text{ for } n_i = n_j = r$$

If the observe 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.

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

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.

A	10	15	8	12	15		
B	14	8	31	15			
C	17	16	14	15	17	15	18
D	12	15	17	15	16	15	

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.

A	14.3	11.6	11.8	14.2
B	20.7	21.0		
C	32.6	32.1	33.0	
D	24.3	25.2	24.8	

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

A	24	18	18	29	22	17	15
B	46	39	37	50	44	45	30
C	32	20	26	41	36	28	27

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:

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$$
$$i = 1, 2, ..., t \text{ and } j = 1, 2, ..., r$$

where y_{ij} is the observed value for block j of treatment i, μ is the population mean, α_i is the effect of treatment i, β_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,
- $e_{ij} \stackrel{i\,i\,d}{\sim} N(0, \sigma^2)$
- $\sum\alpha_i = 0, \sum\beta_j = 0,$

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 I to t and assign the t-treatment at random to the t-unit within each block following any randomization scheme.

DATA STRUCTURE

Blocks

Treatment	1	2	3 … r	Total
1	y_{11}	y_{12}	$y_{13} \cdots y_{1r}$	$y_{1.}$
2	y_{21}	y_{22}	$y_{23} \cdots y_{2r}$	$y_{2.}$
3	y_{31}	y_{32}	$y_{33} \cdots y_{3r}$	$y_{3.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	
t	y_{t1}	y_{t2}	$y_{t3} \cdots y_{tr}$	$y_{t.}$
Total	y_{11}	y_{12}	$y_{.3} \cdots y_{.r}$	$y_{..}$

Analysis of Variance

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

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

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

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

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

ANOVA TABLE

Source	Df	SS	MS	F
Block	r – 1	SS_{BLOCK}	$MSB = \dfrac{SSB}{r-1}$	$\dfrac{MSB}{MSE} = F_{BLOCK}$
Treatment	t – 1	SS_{Trt}	$MST = \dfrac{SST}{t-1}$	$\dfrac{MST}{MSE} = F_{Trt}$
Error	(r-1)(t-1)	SSE	$MSE = \dfrac{SSE}{(r-1)(t-1)}$	
Total	rt – 1	SS_{Total}		

Hypothesis

$$H_{10}: \beta_1 = \beta_2 = \cdots = \beta_r = 0$$
$$H_{11}: \text{atleast one } \beta_j \neq 0$$
$$\text{Or } H_{20}: \alpha_1 = \alpha_2 = \cdots = \alpha_t = 0$$
$$H_{21}: \text{atleast one } \alpha_i \neq 0$$

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 $F_{BLK} > F_{T1}$

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 $LSD = tSED = t\sqrt{\dfrac{2MSE}{r}}$

Coefficient of variation

$$CV = \sqrt{\dfrac{MSE}{overall\ mean}} \times 100$$

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

$$y_{ij} = \mu + \alpha_i + \beta_j + a_{ij} \qquad \begin{array}{l} i = 1, \dots, t \\ j = 1, \dots, r \end{array}$$

$$x = \mu + \alpha_i + \beta_j + e_{ij} \; ,$$

$$a_{ij} = (x - \mu - \alpha_i - \beta_j) \; ,$$

$$S = e_{ij}^2 = (x - \mu - \alpha_i - \beta_j)^2$$

$$\frac{ds}{dx} = 2(x - \mu - \alpha_i - \beta_j) \; ,$$

$$\Rightarrow \hat{x} - \hat{\mu} - \hat{\alpha}_i - \beta_j = 0 \quad ,$$

$$\mu = y_{..} = \frac{y_{..}}{N} = \frac{G_0 + x}{N}$$

$$\alpha_i = \frac{y_{i.}}{r} - \frac{y_{..}}{N}$$

$$= \frac{T_0 + x}{r} - \frac{G_0 + x}{N}$$

$$\beta_j = \frac{y_{.j}}{t} - \frac{y_{..}}{N}$$

$$= \frac{B_0 + x}{r} - \frac{G_0 + x}{N}$$

$$\Rightarrow \hat{x} - \frac{G_0 + \hat{x}}{rt} - \left(\frac{T_0 + \hat{x}}{r} - \frac{G_0 + \hat{x}}{rt}\right),$$

$$- \left(\frac{B_0 + \hat{x}}{t} - \frac{G_0 + \hat{x}}{rt}\right) = 0$$

$$\Rightarrow rt\hat{x} - t_{T_0} - t\hat{x} - rB_0 - r\hat{x} + G_0 + \hat{x} = 0$$

$$\Rightarrow r(rt - t - r + 1)\hat{x} = t_{T_0} + rB_0 - G_0$$

$$\Rightarrow r(r - 1)(t - 1)\hat{x} = t_{T_0} + rB_0 - G_0$$

$$\therefore \hat{x} = \frac{t_{T_0} + rB_0 - G_0}{(t-1)(r-1)}$$

where G_0 is the grand total excluding the missing value.

- T_0 is the total observed value for the treatment that contained the missing value.
- 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. $R.E = \dfrac{(r-1)E_b + r(t-1)E_e}{(rt-1)E_e} \times 100$

Where E_b is the block mean square and the E_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.

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

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.

	V_1	V_2	V_3	V_4	V_5	
I	2.73	1.73	3.00	2.13	0.27	9.86
II	*	2.27	1.53	2.40	0.27	6.47
III	2.33	2.00	2.13	1.87	0.10	8.43
IV	2.53	2.00	2.00	1.47	0.13	8.13
	7.59	8.00	8.66	7.87	0.77	32.89

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

A (7.62)　　C (*)　　　E (7.46)　　D (7.17)　　B (8.14)

E (7.68)　　B (8.15)　　C (7.73)　　D (7.57)　　A (800)

C (7.47)　　E (7.20)　　A ((7.93)　B (7.87)　　D (7.80)

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:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \delta_k + e_{ijk}$$

where $i, j, k = 1, 2, \ldots, t$

where y_{ijk} is the observed value from row j and column k receiving treatment i.

- μ is the overall mean

- α_i is the effect of treatment i

- β_j is the effect of row j

- δ_k is the effect of row k

- e_{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.

$$e_{ijk} \overset{iid}{\sim} N(0, \sigma^2)$$

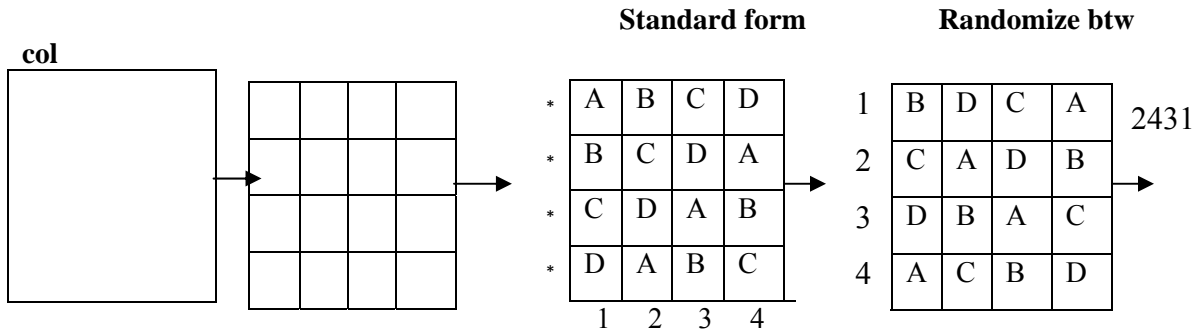$\Sigma \alpha_i = 0$, $\Sigma \beta_j = 0$ and $\Sigma \delta_k = 0$

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 x 4 LS

col

Standard form

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

1 2 3 4

Randomize btw

1	B	D	C	A
2	C	A	D	B
3	D	B	A	C
4	A	C	B	D

2431

Randomize between row

1	C	A	D	B
2	D	B	A	C
3	A	C	B	D
4	B	D	C	A

2341

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

Source	Df	SS	MS	F
Rows	$t-1$	SSR	MSR	MSR/MSE
Columns	$t-1$	SSC	MSC	MSC/MSE
Treatments	$t-1$	SS_{Trt}	MS_{Trt}	MS_{Trt}/MSE
Error	$(t-1)(t-2)$	SSE	MSE	
Total	t^2-1	**SST**		

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

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

Where Er, 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.

I   II   III   IV

I

II

III

IV

1) $\begin{array}{c} R.E\ (RCBD,ROW) \\ \{using\ column\ as\ block\ in\ RCBD\} \end{array} = \frac{t_r + (t-1)E_e}{tE_e} \times 100$

2) $\begin{array}{c} R.E\ (RCBD,Column) \\ \{using\ row\ as\ block\ in\ RCBD\} \end{array} = \frac{e_c + (t-1)E_e}{tE_e} \times 100$

Example: Suppose we have $R.E(RCBD,ROW) = 87\%$ $R.E(RCBD,Col) = 394\%$. 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

$$\hat{x} = \frac{t(R_0 + C_0 + T_0) - 2G_0}{(t - 1)(t - 2)}$$

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)

			Column			
Row	1	2	3	4	5	Total
1	B: 257	E: 230	A: 279	C: 287	D: 202	1255
2	D: 245	A: 283	E: 245	B: 280	C: 260	1313
3	E: 182	B: 252	C: 280	D: 246	A: 250	1210
4	A: 203	C: 204	D: 227	E: 193	B: 259	1086
5	C: 231	D: 271	B: 266	A: 334	E:339	1440

Total	1118	1240	1297	1340	1309	6304

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.

	I	**II**	**III**			**I**	**II**	**III**
I	A_1	B_3	C_2		I	C_3	A_2	B_1
II	B_2	C_1	A_3		II	A_1	B_3	C_2
III	C_3	A_2	B_1		III	B_2	C_1	A_3

Randomize between rows

	I	**II**	**III**
	A_2	B_1	C_3

B₃   C₂   A₁

C₁   A₃   B₂

Randomize between columns

T = 4   A, B, C. D

	I	II	III	IV
I	A_1	C_4	B_3	D_2
II	B_2	D_3	A_4	C_1
III	C_3	A_2	D_1	B_4
IV	D_4	B_1	C_2	A_3

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

I	II	III	IV	V
A_1	B_3	C_5	D_2	E_4
B_2	C_4	D_1	E_3	A_5
C_3	D_5	E_2	A_4	B_1
D_4	E_1	A_3	B_5	C_2
E_5	A_2	B_4	C_1	D_3

MODEL

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

where i, j, k, l = 1, 2, ... , t

where y_{ijk} is the observed value from row *j* and column *k* receiving treatment *i*.

μ is the grand mean

α_i is the effect of treatment *i*

β_j is the effect of row j

δ_k is the effect of column k

τ_l is the effect of subscript factor l.

e_{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

$$e_{ijkl} \overset{iid}{\sim} N(0, \sigma^2)$$

$\sum \alpha_i = 0, \ \sum \beta_j = 0, \ \sum \delta_k = 0$ 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 petition into five components i.e. the row, column, subscript factor, treatment and error.

The sum of square are obtain in the usual form i.e.

$$SST = \sum y_{ijkl}^2 - \frac{y_{....}^2}{N} \qquad N = t^2$$

$$SS_{Trt} = \frac{\sum y_{i...}^2}{t} - \frac{y_{....}^2}{N}$$

$$SSR = \frac{\sum y_{.j..}^2}{t} - \frac{y_{....}^2}{N}$$

$$SSC = \frac{\sum y_{..k.}^2}{t} - \frac{y_{....}^2}{N}$$

$$SS_{Subscript} = \frac{\sum y_{...l}^2}{t} - \frac{y_{....}^2}{N}$$

$$SSE = SST - SS_{TRT} - SSR - SSC - SS_{Subscript}$$

ANOVA TABLE

Source	Df	SS	MS	F
Rows	t-1	SSR	MSR	MSR/MSE
Column	t-1	SSC	MSC	MSC/MSE
Subscript	t-1	$SS_{Subscript}$	$MS_{Subscript}$	$MS_{Subscript}/MSE$
Treatments	t-1	SS_{Trt}	MS_{Trt}	MS_{Trt}/MSE
Error	(t-1) (t-3)	SSE	MSE	
Total	t^2-1	**SST**		

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.

Order	AREA 1	2	3	4	5	Total
1	A_1 (3.5)	B_3 (4.2)	C_5 (6.7)	D_2 (6.6)	E_4 (4.1)	25.1
2	B_2 (8.9)	C_4 (1.9)	D_1 (5.8)	B_3 (4.5)	A_5 (2.4)	23.5
3	C_3 (9.6)	D_5 (1.7)	E_2 (2.7)	A_4 (3.7)	B_1 (6.0)	25.7
4	D_4 (10.6)	E_1 (10.2)	A_3 (4.6)	B_5 (3.7)	C_2 (5.1)	34.1
5	E_5 (3.1)	A_2 (7.2)	B_4 (4.0)	C_1 (3.3)	D_3 (3.5)	21.1
Total	35.6	27.2	23.8	21.8	21.1	129.5

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.

Breeds

	I	**II**	**III**	**IV**
1	A_1 (5.9)	B_3 (4.2)	C_4 (10.2)	D_2 (6.6)
2	B_2 (8.9)	A_4 (4.5)	D_3 (6.0)	C_1 (3.0)
3	C_3 (9.6)	D_1 (5.8)	A_2 (7.2)	B_4 (4.6)
4	D_4 (10.5)	C_2 (4.1)	B_1 (3.5)	A_3 (6.7)

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 '*ab'* treatment combinations.

A	**B**
Maize variety	**Fertilizer Rate**
a $\quad a_0$ White	b_0 $\quad 0kg/ha$
$\quad\quad a_1$ Yellow	b $\quad b_1$ $\quad 30kg/ha$
	$\quad b_2$ $\quad 60kg/ha$
	$\quad b_3$ $\quad 90kg/ha$
	$\quad b_4$ $\quad 120kg/ha$

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.

A	B
Maize Variety	Fertilizer Rate
a_0 White	b_0 0 kg/ha
a_1 Yellow	b_1 30 kg/ha

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

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

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

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

Consider the following result

Mean Responses

	b_0	b_1		b_0	b_1		b_0	b_1
a_0	20	30		20	30		20	40
a_1	40	50		30	50		50	30
	(i)			(ii)			(iii)	

$m.e \ of \ A = \dfrac{1}{2}(20 + 20) = 20$

$m.e \ of \ B = \dfrac{1}{2}(10 + 10) = 10$

$m.e \ of \ A = \dfrac{1}{2}(20 + 20) = 20$

$AB = \dfrac{1}{2}(20 - 20) = 0$

$= \dfrac{1}{2}(10 - 10) = 0$

$m.e \ of \ A = \dfrac{1}{2}(10 + 20) = 15$

$m.e \ of \ B = \dfrac{1}{2}(10 + 20) = 15$

$AB = \dfrac{1}{2}(20 - 10) = 5$

$m.e \ of \ A = \dfrac{1}{2}(30 - 10) = 10$

$m.e \ of \ B = \dfrac{1}{2}(-20 + 20) = 0$

$AB = \dfrac{1}{2}(-10 - 30) = -20$

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 $a_0b_0, a_0b_1, a_1b_0, a_1b_1$.

a_1b_0	a_0b_0	a_1b_0
a_1b_1	a_0b_1	a_0b_1

a_1b_0	a_0b_0	a_0b_1
a_1b_1	a_1b_1	a_0b_1

	a_0b_0	a_1b_1	a_0b_0
RCBD	a_0b_1	a_1b_0	a_1b_1

	a_0b_0	a_0b_1	a_1b_0
CRD	a_1b_1	a_1b_0	a_0b_0

MODEL CRD

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$$

$$i = 1, \ldots, a; \quad j = 1, \ldots, b; \quad k = 1, \ldots, r$$

where y_{ijk} is the observed value from the kth unit corresponding to level i of factor A and the jth level of B

μ is the grand mean

α_i is the effect of level i of factor A

β_j is the effect of the jth of factor B

$(\alpha\beta)_{ij}$ is the effect of interaction between α_i and β_j.

e_{ijk} is the random error component.

MODEL RCBD

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + T_k + e_{ijk}$$

Where T_k is the effect of the kth block

$$\Sigma \alpha_i = 0, \quad \Sigma \beta_j = 0 \quad \Sigma(\alpha\beta)_{ij} = 0$$

H_0: $\alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$

H_1: atleast one $\alpha_i \neq 0$

H_{20}: $\beta_1 = \beta_2 = \cdots = \beta_b = 0$

H_{21}: atleast 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.

$$SST = \sum_{u} y_{iu}^2 - \frac{y_{..}^2}{abr}$$

$$SSA = \frac{\sum_i y_{i.}^2}{br} - \frac{y_{..}^2}{abr}$$

$$SSB = \frac{\sum_j y_{.j.}^2}{ar} - \frac{y_{..}^2}{abr}$$

$$SSAB = \frac{\sum_{ij} y_{ij.}^2}{br} - \frac{y_{..}^2}{abr}$$

-SSA- SSB

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

ANOVA TABLE

Source	Df	SS	MS	F
A	a-1	SSA	MSA	MSA/MSE
B	b-1	SSB	MSB	MSB/MSE
AB	(a-1) (b-1)	SSAB	MSAB	$MSAB/MSE$
ERROR	ab(r-1)	SSE	MSE	
Total	abr-1	SST		

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

Type	15°F		70°F		125°F		
A	130 155		34 40		20 70		
	74 180		80 675		82 58		
	539		229		230		998
B	150 188		126 122		25 70		
	159 126		106 115		58 46		
	623		469		198		1290
C	1382 110		174 120		96 104		
	168 160		150 139		82 60		
	576		583		342		1501
	1738		1281		770		3789

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

Car brand	A		B		C	
1	5	7	12	8	16	12
	11	9	4	6	8	10
2	5	6	9	2	8	13
	4	2	2	7	14	9
3	2	4	9	12	6	7
	6	7	8	7	8	5

4   6   5   9   7   14 11
      8   11   3   4   12 10