

COURSE CODE:	FWM 310
COURSE TITLE:	FOREST AND WILDLIFE BIOMETRICS
NUMBER OF UNITS:	2 Units
COURSE DURATION:	Two hours per week

COURSE DETAILS:

Course Coordinator: DR A.C. ADETOGUN

Email:

Office Location:

Other Lecturers: DR.O.A JAYEOLA; MR. A. O. OLADOYE

COURSE CONTENT:

Practical concept in the design and analysis of experiments on tree crops and wildlife survey techniques as they relate to forestry problems. Processing of resource inventory and mensuration data for management purposed distribution, sampling and tests of hypothesis. Application of multivariate analysis to forestry and wildlife, basic techniques in survey sampling and design.

COURSE REQUIREMENTS:

. This is a compulsory course for all students in the University. In view of this, students are expected to participate in all the activities and have minimum of 75% attendance to be able to write the final examination.

READING LIST:

LECTURE NOTES

DESIGN AND ANALYSIS OF EXPERIMENT

Expel design and sample design.

Sample design is a method of choosing the required number of sample in a given population.

Exptal design is where identical or difficult subject.

An experimental unit is a unit of difference subject of the treatment is applied and he treatment is the factor whose effect is to be investigated or measured and compared the other treatment.

Experimental error is one of the most important in an experiment it is a variation among unit of experiment alike it is a measure of variation each exist among observation of experiment issue treated alike.

There are 2 main source of variation

1. Inherent variation that exist in the experimental material to each treatment are applied
2. Variation which result from lack of uniformity in the physical conduct of experiment e.g. unequal instrument unit. Like experimental error based on this sources of variation can thus be reduced by
 1. Handling the experimental materials so that the effects of inherent error may be reduced.
 2. Refusing experimental techniques

CRD

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

A	B	C	D
0%	10%	15%	20%

RCB --- introduce blocking so as to reduce the variation by randomization. The important thing is that within each block all of them have to be represented. There is no block where you have two treatments.

CRD

Model: $y_{ij} = \mu + T_j + \Sigma_{ij}$

A ₄	C ₂	A ₇	C ₂	B ₄
D ₄	B ₆	B ₅	D ₁	A ₂
C ₄	B ₇	E ₃	D ₆	E ₃
A ₃	C ₅	E ₂	C ₃	E ₄
D ₅	D ₄	A ₅	C ₄	B ₅

replicate i, 1, 5

Treatment j, 1, 5

$$SS \text{ total} = \sum y_{ij}^2 - CF = 4^2 + 2^2 + \dots + 25^2 - CF$$

$$Ss \text{ Treatment } \sum y_{.j}^2 / r^2 - CF = 473 - 416.16 = 56.862$$

$$= \frac{27^2 + 27^2 + 17^2 + 28^2 - CF}{5}$$

5

$$= 421.8 - 418.18 = 5.64$$

Error Ss = Total ss - Treatment ss

$$= 56.84 - 5.64 = 51.20$$

ANOVA TABLE

Source of variation	df	Ss	ms	F mst
Treatment	4	5.64	1.41	MSE
Error	20	51.20	2.5	0.564
Total	24	56.4		

$$F_{0.5}(4, 20) = 2.87$$

$$F_{0.1}(4, 20) = 4.43$$

F, calculated of 0.564 is < F tab of 2.87 at 5% probability level it means they there is no significant difference between the treatment i.e Treatment ABCDE are identical/there is no significant difficult in their effect of the fertilizer level. Therefore any of the treatment can be adopted.

RCB

$$\text{Model} = y_{ij} = \mu + T_j + B_i + E_{ij}$$

ANOVA TABLE

Source of variation	df	Ss	ms	F
Treatment	4	13.76	3.44	1.98
Block	4	22.16	5.54	3.04
Error	16	29.04	1.0	
Total	24	64.96		

$$\text{Total} = 25$$

$$\text{Treatment} = 5$$

$$\text{Block} = 5$$

F cal treatment = 1.89

F cal block = 3.04

F cal treatment has no significant different between treatment

F cal block has different between the block

Thus blocking is very significant important because it bring differences in the yield/block

LATIN SQUARE (3 way)

Ls =

Model = $y_{ijk} = \mu + R_i + C_j + T_k + E_{ijk}$

Source of variation	df	Ss	ms	F mst
Row	$r-1 = 4$	75.57	18.39	1.61
Column	$r-1 = 4$	583.51	145.89	13.12
Treatment	$r-1 = 4$	67.86	16.97	1.53
Error	$(r-1)(r-2) 12$	133.35	11.20	
Total	24			

$F_{0.5}(4, 12) = 3.26$

I	II	III	IV	V	
A	B	C	D	E	1
B	C	D	E	A	2
C	D	E	A	B	3
D	E	A	B	C	4
E	A	B	C	D	5

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

Tj BK

Even with the RCD there is still a problem, to solve this you introduce Latin square. Therefore the number of treatment must be equal to the number of row and the number of column must be equal to the number of treatment.

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

Tj Rk Cm

CRD can be used in comparison of two or more groups of treatment by one way analysis of variance (ANOVA). Anova is an arithmetic process for partitioning a total sum of square into components. CRD is useful when experimental units are homogenous.

CRD

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

RCB

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

D C A B

A B C D

B C D A

C D A B

D A B C

Significant difference

A α B

A α C

A α D

B α C

B α D

C α D

LAEST SIGNIFICANT DIFFERENCE (LSD)

This is to further find out the significant difference between at least 2 treatment. It is the option after our

1. F is significant
2. The number of treatment must not be more than 5
3. The experiment must be a control experiment and in effect get the difference between the control and other treatment. Relax the 2nd condition if it is a control experiment.

DMRT		Rice Yield				mean	d
bc	A	2537	2069	2104	1797	2127	
a	B	3366	2591	2211	2544	2678	
ab	C	2536	2459	2827	2385	2552	
bc	D	2387	2453	1556	2116	2120	

C E 1997 1679 1649 1859 1796
 Cd F 1796 1704 1904 13020 1681
 d G 1407 1516 1270 1071 1316

ANOVA TABLE

Source of variation	df	Ss	ms	F
Treatment	6	5.58m	931,196	9.82
Error	21	1.99m	94,773	
Total	27	7.57m		

F_{0.05} (6, 21) 2.57

F_{0.01} (6, 21) 3.81

Because the F cal is > F tab, there is a significant difference

$$LSD \alpha = t \alpha \sqrt{25^2 / r}$$

S² + error ms

T_α = F tab at dt level

r = number of replicate

$$LSD 0.05 = t_{0.05} (6, 21) = 2.08$$

$$= \frac{2.080 \times \sqrt{2 \times 94,773}}{4} = 453$$

$$LSD 0.01 = \frac{2.831 \times \sqrt{2 \times 94,773}}{4} = 616$$

Calculating difference 'd' i.e M-the control mean

d

A	811	
B	1362	
C	1236	
D	812	
E	480	
F	365	N.S at 0.05

We now compared the d of each weigh the LSD 0.05 and LSD 0.01

LSD 0.05 = 453 is significant with A-E

LSD 0.01 = 616 only 4 of the treatments are significant

Conclusion

DMRT: Dunnett Multiple Ratio Test

For the above experiment

	Sort		R	P	rp0.05	rp0.01	Rp(SrSxp)
B	2678	a	1	2	2.943	4.024	452.6
C	2552	a b	2	3	3.09	4.197	475.6
D	2128	b c	3	4	3.18	4.312	489.5
A	2127	b c	4	5	3.25	4.395	500.3
E	1796	c	5	6	3.30	4.495	507.97
F	1681	c	6	7	3.33	4.510	512.6
G	1361	d	7				

$$\text{Calculate SE} = \text{SX} = \sqrt{S^2/r} = \sqrt{\frac{94773}{4}} = 153.93$$

Shortest significant range

$$RP = rpSX$$

$$(1) 2678 - 512.6 = 2165$$

$$(2) 2552 - 508 = 2044$$

$$(3) 2128 - 500 = 1628$$

MSTATIC Statistic package: Group the treatment mean in decreasing order for the largest mean, subtract the shortest range of the highest P. Declare all means less than these significantly difference from the largest mean because it is only B and C that fall in these category it means they are similar that is, not significant.

Quick check is to subtract between 2 sot value and compare it with RP of the value between the 2, say B and C ----- $2678 - 2552 = 126$ compare $<$ RP of 2 ----- 452.6 since $126 < 452.6$. There is no significant difference between B and C

A and G = $2127 - 1316 = 811 > R4$ there is significant difference

A and G = $1796 - 1316 = 480 > R3$ there is significant difference

F and G = $1681 - 1316 = 365 < R2$ they are similar

AB NS

AC #

AD #

AD #

AE #

F #

AG NS

Diversity and similarity indices

This is an applied statistics

To measure the diversity of the species we can use (1) swimpson diversity index Σp_i

	$\frac{\Sigma ni(ni-1)}{N(N-1)}$	Σp_i
Eg	Zone 1	Zone 2
A	20	15
B	-	10
C	15	25
D	10	-
E	13	15
	4	4
	60	65

$$= \frac{\Sigma ni (ni-1)}{N(N-1)}$$

$$= \frac{(20 \times 19) + (15 \times 14) + (10 \times 9) + (15 \times 14)}{60 \times 59}$$

$$= 0.25$$

$$= \frac{(15 \times 14) + (10 \times 9) + (25 \times 24) + (15 \times 14)}{65 \times 64}$$

$$= 0.27$$

The higher the value the lower the diversity

Zone 1 is none diversity than zone 2

$$1/0.25 = 4$$

$$1/0.27 = 37$$

The diversity of swingson formular can be used

$$= \frac{N(N-1)}{\sum n(n-1)}$$

$$\sum p_i^2 = (20/60)^2 + (15/60)^2 + (10/60)^2 + (15/60)^2$$

$$= 0.33^2 + 0.25^2 + 0.11^2 + 0.25^2$$

Shamon index = H

$$i = -\sum (n_i/N) \log_e (n_i/N)$$

$$ii \sum p_i \log_e p_i$$

we can calculate evenness from the above

$$\text{Evenness} = H/\log_e S$$

S = number of species.

Similarity indices

Sorenson's similarity index

$$\frac{a}{a + b + c} \times 100$$

a = number of species common to both

b = number of species in site 1 but not in site 2

c = number of species in site 2 but not in site 2

$$3/3+1+1 \times 100 = 3/5 \times 100 = 60\%$$

$$\text{Simpson similarity index} = \frac{a}{\text{Min}(a+b)(a+c)}$$

$$3 = \frac{3}{4} \times 100 = 75\%$$

Min 14, 4

Sorrenson similarity can be used when one is sure that the sampling unit is the same e.g. when Zn 1 has 2 plots, Zn 2 has 3 plots e.t.c. siorrenson can not be used but swinmpson sin index can be used

A	20	15	5
B	-	10	-
C	15	25	-
D	10	-	20
E	15	15	5
F	10	5	-
G	2	-	4
H	3	-	-

Simpson sin index

$$= \frac{a}{\min(a+b)(a+c)}$$

$$a = 4$$

$$b = 3$$

$$c = 1$$

$$\text{simpson} = \frac{4}{7,5}$$

$$1 \text{ and } 2 = \frac{4}{5} = 80\%$$

$$1 \text{ and } 3 = \frac{4}{7,4} = \frac{4}{4} = 100\%$$

$$b = 3$$

$$c = 0$$

2 and 3 = a = 2

b = 3

c = 2

$2/5,4 = 2/4 = 50\%$

DESIGN AND ANALYSIS OF EXPERIMENTS

INTRODUCTION

An expert is a research tool used to discover something unknown or to test a principle of hypothesis. In other words, an expert is a planned inquiry to obtain new fact or to confirm or deny the results of previous experiments of the existing belief such that the inquiry will aid in making decision or recommendation.

TYPES OF EXPERTS

Experts can be broadly classified into two categories, namely absolute experts and comparative experts. The latter is conducted with the full aim of comparing the effects of several conditions on some phenomenon e.g. an experts may be conducted to determine the effects of some fertilizers on the growth rate of *Tectona grandis* while in absolute experts, the sole aim si to discover a new information which will either confirm or contradict an existing belief.

Experts in forestry and wildlife or nay other field, produce results which form the basis for making decisions. A forest pathologist may decide to apply a new fungicide on his nursery based on the reports from researchers that the fungicide is effective. In view of this, many important decisions are made based on the results of experts, it is of paramount importance, therefore, that experts, are well planned so as to produce valid results. Through careful planning and conduct of an experts, data are acquired which are then analyzed and interpreted with respect to the objectives of the experts.

The complete sequence of steps taken ahead of time to insure that the appropriate data will be obtained in a way that permits an objective analysis leading to valid inferences is referred to as the Experimental Design.

PROCEDURE FOR EXPERIMENTATION

The procedure for experimentation is usually determined by the nature of the experimentation and the objective of the study. The procedure to follow include:

- i. Define the problem: Every research begins with the identification of an existing problem which the researcher can reasonably hope to solve.
- ii. State your objective: The objective of the experimentation should be clearly stated. The objective may be in the form of question to be answered, the hypothesis to be tested or the effects to be estimated. The statement of objective should include an account of the area over which generalization are to be made or the population about which inferences are to be made.
- iii. Plan the experimentation: The experimentation should be planned with a view to achieving the objective. At this stage, adequate consideration should be given to the experimental materials, the treatments, the variables of interest, the number of replicates, and the data to be collected. The experimenter also make an outline of the statistical analysis to be carried out as well as possible summary tables or graph to be drawn.
- iv. Conduct the experimentation: The experimentation should be conducted carefully and objectively. It should be noted that no amount of statistical manipulation can make a bad experimentation give good results. The experimenter should, however, keep his mind open to the possibility of obtaining other useful information without prejudice to his main interest. The data collected at each stage of the experimentation should be clearly recorded and properly documented. The design must be kept as simple as possible. Efforts should be made to conserve time, money, personnel and experimental materials.

The scope of the experimentation should be large enough to provide a wide basis for inductive inference.

- v. Analyze the data and interpret the results: Care should be taken to avoid mistakes in copying data from the field note books. All calculations should be carefully checked. Simple tables and diagrams should be used when summarizing the results of statistical analysis. Correct decisions should be made concerning the hypotheses, and the interpretation of the results should be done in terms of the subject matter.
- vi. Prepare a complete, readable and correct report of the research: The report must tell the truth and must contain all the important information necessary for someone else to repeat the experimentation. The report must also meet standards in the scheme, arrangement, the tables, figures, nomenclature and bibliography.

COMMON TERM IN EXPERIMENTAL DESIGN

- 1) Treatment: This is a term used for any particular set of experimental conditions that will be imposed on an experimental unit within the confines of the chosen design e.g. an experimentation may be conducted to determine the effect of depth of seeding on the yield growth of *Gmelina arborea*. The several depths of seeding constitute effects of two experimental conclusions such as temperature and humidity, amount of fertilizer and depth of sowing are examined in an experimentation. Such treatment combined treatments are called treatment combinations.
- 2) Experimental unit: This refers to that material to which a single treatment is applied in one replication of the basic experimentation e.g. a plot of land, a batch of seeds, a group of animals etc.

3) Experimental error: This is a measure of the variation which exists among observations on experimental units treated alike. It describes the failure of the identically treated experimental units to yield identical results. Experimental error reflects.

- The errors of experimentation, observation and measurements
- The variation among experimental units, and
- The combined effects of all extraneous factors that could influence the characteristics under study but have not been singled out for attention in the current investigation.

Experimental error affects the accuracy of experimentations, and therefore the methods of reducing it are also means of increasing the accuracy of experimentations. Experimental error can be reduced in the following ways:

- a) Reducing the effects of inherent variability in the experimental material. This may be accomplished by
 - Using more homogenous experimental material
 - Taking additional measurements that provide more information on the experimental material
 - Using information provided by related varieties
 - Grouping the experimental units skillfully such that the units to which one treatment is applied are closely comparable with those to which another treatment is applied.
- b) Increasing the size of the experimentation either through the provision of more replicates or by the inclusion of additional treatments.
- c) Using more care in conducting the experimentation
- d) Using a more efficient experimental design.

BASIC PRINCIPLES OF EXPERIMENTAL DESIGN

There are three basic principles of experimental design namely randomization, replication and local control.

- 1) **Randomization:** This is a process by which the allocation of treatments to experimental units are done by means of some chance device. It ensures that treatments will not be continually favoured or handicapped in successive replications by extraneous sources of variation known as unknown. Thus, it assist in ‘averaging out’ the effects of extraneous factors that may or may not be present. The function of randomization is to ensure that the experimenter has a valid or unbiased estimate of experimental error. Apart from random designs, there are also non-random but selected fashion. The use of these designs makes it impossible to attach an unbiased measure of reliability to the inferences made. Systematic designs result in either underestimation or overestimation of experimental error, as well as inequality of precision in the various comparisons among the treatment means.
- 2) **Replication:** It refers to the repetition of the basic experimentation e.g. in testing the effects of three fertilizer on the growth of three tree species with three replication. The lay-out is as follows:

	A	B	C
I			
II			
III			

Rep I

	A	B	C
I			
II			
III			

Rep II

	A	B	C
I			
II			
III			

Rep III

When a treatment appears more than once in an experimentation, it is said to be replicated. Multiple readings for a treatment appearing once do not necessarily represent true replication. In other words, repeat measurement on the same experimental unit is not true replication. When each treatment appears only once in an experimentation, the experimentation is said to consist of a single replication or replicate. For such an experimentation, experimental error. Thus the observed differences could be explained as differences between the treatments or between the experimental units i.e. it cannot be determined whether observed differences indicate true differences or are due to inherent variation.

The functions of replication can be summarized as follows:

- It provides an estimate of the experimental error which serves as the basic unit of measurement for assessing the significance of observed differences or for determining the length of a confidence interval.
- It improves the precision of an experimentation by reducing the standard deviation of a treatment mean.
- It increases the scope of inference of the experimentation by selection and appropriate use of more variable experimental units.

However, the number of replicates for any experimentation depends on the degree of precision required, the experimental design, the number of treatments, the degree of variation in the experimental units, and funds and time available for the experimentation.

- 3) Local control: It refers to the amount of grouping, blocking and balancing of the experimental unit that is employed in the adopted statistical design. It makes the experimental design more efficient i.e. it makes the test of significance more sensitive. This is because the magnitude of the estimate of experimental error is reduced by proper use of local control.

- Grouping: It means placing a set of homogenous experimental units into groups in order that different treatments are applied to the different groups. The groups may even have different number of experimental units.
- Blocking: It means allocating the experimental units into blocks such that the units within a block are relatively homogenous while the units between the blocks are heterogeneous. In this case, all the treatments are applied within each block.
- Balancing: it is by obtaining the experimental units, the grouping, the blocking and the assignment of treatments to the units in such a way that a balance configuration results.

GENERAL LINEAR HYPOTHESES

STEPS IN DESIGNING AN EXPERIMENTAL

- Name the design
- State the statistical model
- State the assumptions
- Present the lay-out
- State the hypotheses i.e. tests of significance
- Do the estimation
- Draw inferences and make relevant conclusions

REFERS TO SINGLE FACTOR EXPERIMENTATIONS

COMPLETELY RANDOMIZED DESIGN (CRD) ONE WAY CLASSIFICATION

It is a design in which the treatments are allotted completely at random to the experimental units. Thus, every experimental unit has the same probability of receiving any treatment without any restriction. The design is useful only when the experimental units are essential homogenous e.g. given 4 fertilizers A, B, C, and D with which to test the H_0 that there are no differences among the effects of these fertilizers on the volume of wood produced by *Gmelina arborea*. let us further assume that there are 20 experimental plots available for the research. A good design method will place each fertilizer on an equal number of plots so that over estimate of mean effects fertilizer will have equal weight. We also insist that the fertilizers be assigned to the plots at random using table of random numbers or any random process.

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

ADVANTAGES OF THE CRD

- It is flexible in that any number of treatments and of replicates may be used
- The statistical analysis is simple even if the number of replicate varies with the treatments
- Simplicity of analysis is not lost if some experimental unit or entire treatments are missing or rejected
- The number of degrees of freedom for estimating experimental error is maximum.

DISADVANTAGES OF THE CRD

Since randomization is unrestricted, experimental error includes the entire variation among the experimental units except that due to treatments. In situations where there are variations among the units, CRD is inefficient.

- 1) Name of design: Completely randomized design (CRD)
- 2) Model: $Y_{ij} = \mu + t_j + e_{ij}$

Where:

Y_{ij} = Individual observation i.e. observation of j_{th} treatment in i_{th} plot

μ = General mean i.e. the population mean of all possible similar experimentations

t_j = Effect error containing all uncontrolled sources of variation.

- 3) Assumption: (I) The treatment effects are additive and fixed (no interaction) i.e.

$$\sum_{j=1}^t t_j = 0$$

(II) Experimental errors are randomly and independently and normally distributed above zero mean and with a common variance i.e. $e_{ij} \sim \text{NID}(0, \sigma^2)$.

- 4) Lay-out:

T_1	T_2	T_3	T_t
Treatment total $\sum T_1$	$\sum T_2$	$\sum T_3$		$\sum T_t$
Number of observation nT_1	nT_2	nT_3	nT_t
Treatment mean \bar{T}_1	\bar{T}_2	\bar{T}_3	\bar{T}_t

5) Hypotheses:

H_0 – No significance treatment differences

H_A – There are significant treatment differences

6) Computation: Estimate the correction factor

$$C.F = \frac{(\text{Grand total})^2}{\text{Total number of observation}}$$

$$\text{Total sum of square } (SS_{\text{total}}) = \sum Y_{ij}^2 - C.F$$

Where $\sum Y_{ij}^2$ is the summation of the square of each observation

$$\frac{\text{Treatment sum of square} = \sum T_1^2 + \sum T_2^2 + \dots + \sum T_t^2 - C.F}{(SS_{\text{trt}})} \quad \frac{\text{number of observation per treatment i.e. } nT}{\text{number of observation per treatment i.e. } nT}$$

Where $\sum T_1^2, \sum T_2^2, \dots, \sum T_t^2$ are the totals for treatments

T_1, T_2, \dots, T_t respectively and nT_1, nT_2, \dots, nT_t are the number of observations for treatment T_1, T_2, \dots, T_t respectively.

$$\text{error sum of square } (SS_E) = SS_{\text{total}} - SS_{\text{treatment}}$$

Degree of freedom (df)

- For total (df total) = Total number of observation – 1 or (r) (t) - 1
- For treatment (df treatments) = Number of treatments – 1 i.e. (t-1)
- For error (df error) = df total – df treatments or t(r-1)

Mean Squares (MS): These are obtained by dividing the SS treatments and SSerror by the df treatment and df error, respectively.

$$F - \text{Calculated (variance ratio)} = \frac{\text{MS treatments}}{\text{MS error}}$$

7) The analysis of variance table (ANOVA) table

Source of variation (S.V.)	Degrees of freedom (df)	Sum of squares (SS)	Mean squares (MS)	F-calculated
Treatments	$T - 1$	SS_{trt}	$SS_{trt}/t - 1$	MS_{trt}
Error	Df total – df trt	SS_E	SS_E/df_{error}	
Total	$(r) (t) - 1$ $n - 1$	SS_{total}		

8) Inference and conclusion

The F-tabulated will be obtained from the table at desired α or Pr level and it is usually at 0.05 or 0.01 level of significance. Taking the degrees of freedom for treatment and error as V_1 and V_2 respectively, the F-tabulated i.e. $F(V_1, V_2)$ will be checked from the table at the desired probability level.

CONCLUSION

Where F-calculated is greater than F-tabulated, the H_0 is rejected, thus it is concluded that there are significant treatment differences and vice-versa.

e.g. 4 spray treatments were applied completely at random in a teak plantation. 24 sampled plots were used and the yields per plot were recorded as follows:

$S_1(5)$	$S_2(2)$	$S_4(8)$	$S_4(2)$
$S_3(5)$	$S_1(3)$	$S_2(2)$	$S_1(2)$
$S_2(4)$	$S_3(6)$	$S_1(5)$	$S_3(3)$
$S_1(5)$	$S_1(2)$	$S_2(6)$	$S_1(7)$
$S_9(9)$	$S_3(9)$	$S_3(8)$	$S_3(3)$
$S_4(4)$	$S_2(7)$	$S_2(6)$	$S_4(5)$

Test the H_0 that there are no significant difference among the effect of these spray treatments on the yield of teak.

- i. Name of design: CRD
- ii. Model: $Y_{ij} = \mu + t_j + e_{ij}$

Define each

- iii. Assumptions: State them as shown in previous note
- iv. Lay out

	S ₁	S ₂	S ₃	S ₄
	5	4	5	9
	5	3	6	4
	3	7	9	8
	2	3	8	4
	2	6	3	2
	7	6	3	5
∑T	24	29	34	32
n	6	6	6	6
\bar{T}	4	4.83	5.67	5.33

v. Hypotheses:

H₀ – There is no significant difference among the effect of spray treatments

H_A – There is significant difference among the effect of spray treatments.

Compute the grand mean and the coefficient of variation curves

$$\text{Grand mean} = \frac{GT}{n}, \quad CV = \frac{\sqrt{MSE}}{\text{Grand Mean}} \times 100$$

The CV indicates the degree of precision with which the treatments are compared and is a good under of the reliability of the experimental. It expressed the experimental error as percentage of the mean, thus, the higher

vi. Computation:

$$\begin{aligned} \text{C.F.} &= \frac{(G.T)^2}{n} \text{ or } \frac{(\sum Y_{ij})^2}{n} = \frac{(5+5+3+\dots+5)^2}{24} = \frac{(119)^2}{24} \\ &= 590.0417 \end{aligned}$$

$$\begin{aligned}
SS_{\text{total}} &= \sum Y_{ij}^2 - C.F \\
&= 5^2 + 5^2 + 3^2 + \dots + 5^2 - 590.0417 = 110.9583
\end{aligned}$$

$$\begin{aligned}
SS_{\text{trt}} &= \frac{\sum T_j^2 - GF}{\text{No of Obs/Trt}} = \frac{24^2 + 29^2 + 34^2 + 32^2}{6} - 590.0417 \\
&= \frac{24^2 + 29^2 + 34^2 + 32^2}{6} - 590.0417 \\
&= \frac{3596}{6} - 590.0417 \\
&= 599.5 - 590.0417
\end{aligned}$$

$$\begin{aligned}
SS_{\text{error}} &= SS_{\text{total}} - SS_{\text{trt}} \\
&= 110.9583 - 9.4583 \\
&= 101.5000
\end{aligned}$$

vii. ANOVA Table

SV	df	SS	MS	F
Treatment	3	9.4583	$\frac{9.4583}{3} = 3.1528$	$\frac{3.1528}{5.075} = 0.6212$
Error	20	101.5000	$\frac{101.500}{20} = 5.075$	3
Total	23	110.9583		

F-tabulated (i.e. F(V₁, V₂))

F(3,20) = 3.10 and 4.94 for 0.05 and 0.01

viii. Levels of significance, respectively:

Since the F-tabulated is greater than the F-calculated, the H₀ is accepted. It is therefore concluded that there are no significant (P<0.05 or P<0.01) differences among the effects of the 4 spray treatments on the yield of Teak.

2) Forester conducted a preliminary study of the effects of 3 different sources of fertilizer on the growth rate of *Gmelina arborea* in the nursery. The results are as follows:

The CV value, the lower is the reliability of the experimentation. The CV value is generally placed below the analysis of variance table.

S ₁	S ₂	S ₃
8	11	12
7	10	12
8	10	13
9	9	9
	13	10
	12	12
	13	12
	9	18
		20

Are there difference in the effects of fertilizer sources?

i. Name of design: CRD with unequal number of observation per treatment.

ii. Model: $Y_{ij} = \mu + t_j + e_{ij}$

Define each

iii. Assumption: (i) Treatments effects are additive and fixed (i.e. no interaction) $\sum_{j=1}^t t_j =$

0

(ii) Experimental error are randomly, independently and normally distributed about zero mean and with a common variance i.e. $e_{ij} \sim \text{NID}(0, \sigma^2)$

iv. Lay-out:

	S ₁	S ₂	S ₃
	8	11	12
	7	10	12
	8	10	13
	9	9	9
		13	10
		12	12
		13	12
		9	18
			20
			25
ΣT	32	87	143
n	4	8	10
\bar{T}	8	10.875	14.3

v. Hypotheses:

H₀ – There is no significant difference

H_A – There is significant difference

vi. Computation:

$$C.F. = \frac{(\sum Y_{ij})^2}{n} = \frac{(262)^2}{22}$$

$$= 3120.1818$$

$$SS_{total} = \sum Y_{ij}^2 - C.F$$

$$= 8^2 + 9^2 + \dots \dots \dots 25^2 - 3120.1818$$

$$= 3498 - 3120.1818$$

$$= 377.8182$$

$$SS_{trt} = \frac{(\sum T_1^2)}{n} + \frac{(\sum T_2^2)}{n} + \frac{(\sum T_3^2)}{n} - C.F =$$

$$\left(\frac{(32)^2}{4} + \frac{(87)^2}{8} + \frac{(143)^2}{10} \right) - 3120.1818$$

$$= 32470.25 - 3120.1818$$

$$= 126.8435$$

$$SS_{error} = SS_{total} - SS_{trt}$$

$$= 377.8182 - 126.8435$$

$$= 250.975$$

vii. ANOVA Table

SV	df	SS	MS	F
Treatment	2	126.8435	6312.4216	4.8013
Error	19	250.975	13.2092	
Total	21	377.1882		

F-tabulated (i.e. $F(V_1, V_2)$)

$F(2,19) = 3.52$ and 4.94 at 0.05

$F(2,19) = 5.93$ and 4.94 at 0.01

Since the F-calculated is greater than the F-tabulated at either 0.05 or 0.01 level of significance, the null hypothesis is not accepted i.e. rejected. It is therefore concluded that there are significant ($P < 0.05$) differences in the effects of the sources of fertilizer on the growth rate of *Gmelina arborea* or in the effects of the fertilizer treatments. If 0.01 level of significance is used, then the conclusion will be that there are highly significance ($P < 0.01$) differences in th effects of the fertilizer treatments.

HOMEWORK

Members of African Giant rat (*Cricetomys gambianus*, Waterhouse) colony were fed with diets A, B, C and D. the purpose of the study was to determine which diet enehances the growth of the rats best. At the start of the experiment, 10, rats each of the same size were randomly allocated to each diet. We assumed there was no competition for food as there was more than enough for the rats.

Unfortunately, in the course of the experiments, some of the rats died. The results obtained were as shown below. The data represent the net gain in weight (gms) after two months.

A	B	C	D
1.3	2.4	0.9	3.1
1.5	1.5	1.1	3.2

2.0	2.8	1.2	2.5
2.1	1.7	0.4	4.0
2.5	3.5		3.4
3.4	3.8		3.5
	1.2		2.5
	1.4		2.8
	1.6		3.8
			4.2

Analyze the data and test for significant differences in the treatment effects.

FOLLOW-UP PROCEDURES OR POST-MORTEM ANALYSIS

When the H_0 is not accepted in the ANOVA procedure, it is possible to carry out further analysis so as to obtain more information. These further analyses are called the follow-up procedures or post-mortem analysis.

The post-mortem analysis used include:

- viii. Fishers' least significant difference (LSD)
- ix. Duncan's new multiple range test (DMRT)

LSD

When F is significant, the LSD or ordinary t-test for the difference between two means is applied to every pair of means. The LSD itself is based on a t-test, and it is given as:

$$\text{LSD} = t_{\alpha/2, d_{\text{error}}} \times \left(\sqrt{\frac{2MSE}{r}} \right) \text{ for equal number of observations per treatment, or}$$

LSD = $t_{\alpha/2, d_{\text{error}}} \times \left(\sqrt{\frac{2MSE}{r_1}} + \frac{MSE}{r_2} \right)$ for unequally replicated means i.e. treatments with unequal number of observations;

Where MSE = Mean square for error (from the ANOVA table)

r = Number of observations per treatment

r_1 and r_2 = Number of observations per treatments whose means are being compared.

t = Student's t-value (i.e. t-tabulated) from the t-table $\alpha/2$ implies a two-tailed test

any two means whose absolute difference exceeds the LSD value are declared significantly different.

When there are equal number of observations per treatment, a single LSD value is computed and used for the comparisons. However, when the treatment have unequal number of observations, different LSD values are computed for the comparisons

e.g. using example 2 above.

$$\begin{aligned} \text{LSD} &= t_{\alpha/2, d_{\text{error}}} \times \left(\sqrt{\frac{2MSE}{r_1}} + \frac{MSE}{r_2} \right) \\ &= 2.09 \left(\sqrt{\frac{13.2092}{r_1}} + \frac{13.2092}{r_2} \right) \end{aligned}$$

For comparing S_1 and S_3

$$\begin{aligned} \text{LSD} &= 2.09 \left(\sqrt{\frac{13.2092}{4}} + \frac{13.2092}{10} \right) \\ &= 4.4939 \end{aligned}$$

For comparing S_2 and S_3

$$\begin{aligned} \text{LSD} &= 2.09 \left(\sqrt{\frac{13.2092}{8}} + \frac{13.2092}{10} \right) \\ &= 3.6031 \end{aligned}$$

Re-arranging the means orderly, we have;

S_3 (14.3); S_2 (10.875); S_1 (8)

Subtracting the means other, we have;

$$S_3 - S_2 = 14.3 - 10.875 = 3.425 < 3.6031 \text{ ns}$$

$$S_3 - S_1 = 14.3 - 8 = 6.3 > 4.4939 *$$

The result of the LSD procedure are then presented as follows

Treatment	Mean
S_3	14.3 a
S_2	10.875 ab
S_1	8.0 b

Means with the same letters are not significantly different at 0.05 level of significance. From this result, the best source of fertilizer is S_3 . It led to high growth rate than other fertilizers.

DMRT

The DMRT is more conservative (i.e. less likely to detect real differences) than the LSD. This is because it involves several different critical differences in contrast to a single LSD value for equal number of observations per treatments.

Thus, it allows a higher rate of pairs of samples averages that are further apart when ordered by size. This conservative property of the DMRT however makes it less likely to indicate false differences.

Using example 2 above.

- i. The treatment means are arranged horizontally in decreasing order and vertically in ascending order.

		S ₃	S ₂	S ₁
		14.3	10.875	8.0
S ₁	8.0	6.3	2.875	-
S ₂	10.875	3.425	-	
S ₃	14.3	-		

- ii. The number of means involved in the range being compared are also obtained and put in a table similar to that shown in (1) above

	S ₃ (14.3)	S ₂ (10)	S ₁ (8)
S ₁ (8)	3	2	-
S ₂ (10)	2	-	
S ₃ (14.3)	-		

The first element in table (3) is the number of means involved in the range of means from S₃ (14.3) to S₁ (8). These three means are 14.3, 10 and 8.

Between S₃ and S₂, two means are involved i.e. 14.3 and 10. When only one mean is involved as it is the case between S₃ and S₂, it is ignored since we will not be interested in comparing S₃ with S₃.

- 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:

(range, d_{error}) = table value

(3, 19) = 3.11

(2, 19) = 2.96

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

$$S.E = \left(\sqrt{\frac{2MSE}{r}} \right)$$

Where MSE = MS_{Error} from the ANOVA table

r = Number of observation per treatment

Due to the unequal number of observations per treatment an approximate value of r is computed

as: $\frac{t}{\sum \frac{1}{r_i}}$

Where t = Number of observations

r_i = The number of observations for the i_{th} treatments

$$\frac{3}{\frac{1}{4} + \frac{1}{8} + \frac{1}{10}}$$

= 6.3158

∴ S.E = $\sqrt{\frac{13.2092}{6.3158}}$

= 1.4462

v. The standard references:

3.11 x 1.4462 = 4.4977

2.96 x 1.4462 = 4.2808

2.96 x 1.4462 = 4.2808

vi. Testing the mean difference

S₃ – S₁ = 6.3 vs 4.4977 *

S₃ – S₂ = 3.425 vs 4.28.8 ns

S₂ – S₁ = 2.875 vs 4.2808 ns

This result can be presented as follows:

Treatment	Mean
-----------	------

S ₃	14.3 a
S ₂	10.875 ab
S ₁	8.0 b

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

Ondo	Ore	Ile-Ife	India
17.43	16.60	14.80	21.24
19.54	16.34	16.22	21.29
17.67	17.01	21.30	20.34
22.58	15.82	20.84	19.43
19.86	16.57	19.27	25.05

- 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 to secure sufficiently uniform groups of units for blocks.

- 1) Name of design: Randomized complete block design (RCBD)

- 2) Model: $Y_{ijk} = \mu + B_i + T_j + e_{ijk}$

Where: Y_{ijk} = individual observation i.e. observation for the j_{th} treatment in the i_{th} block

μ = General mean

B_i = effect of the i_{th} block

T_j = effect of the j_{th} 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:

B_1	B_2	B_3	B_4
$T_1 3$	$T_3 3$	$T_2 3$	$T_1 9$
$T_3 4$	$T_2 4$	$T_3 6$	$T_2 4$
$T_2 5$	$T_1 2$	$T_1 7$	$T_3 5$

- 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

SV	df	SS	MS	F
Block	$b - 1$	SS_B	$\frac{SS_B}{b - 1} = MSB$	$\frac{MSB}{MSE}$
Treatment	$t - 1$	SS_{trt}	$\frac{SS_t}{b - 1} = MST$	$\frac{MST}{MSE}$
Error	$(b - 1)(t - 1)$	SSE	$\frac{SSE}{(b - 1)(t - 1)} = MSE$	
Total	$Bt - 1$	YSS		

- 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: $Y_{ijk} = \mu + B_i + T_j + e_{ijk}$
Define each
3. Assumption: As stated

4. Lay-out:

	3	3	3	9	18
	4	4	6	4	18
	5	2	7	5	19
Σ Block	12	9	16	18	55

5. Hypotheses: Treatments

H_0 = No Significant difference in the treatment

H_A = There are significant difference

6. Computation:

$$C.F = \frac{(\Sigma Y_{ijk})^2}{bt} = \frac{(55)^2}{4 \times 3} = \frac{(55)^2}{12}$$

$$= 252.08$$

$$SS_{total} = \Sigma Y_{ijk}^2 - C.F = 295 - 252.08 = 42.92$$

$$SS_{block} = \frac{(\Sigma B_1)^2 + \dots + (\Sigma B_4)^2}{t} - C.F = \frac{12^2 + 9^2 + 16^2 + 18^2 - 252.08}{3}$$

$$= 268.33 - 252.08$$

$$= 16.25$$

$$SS_{treatment} = \frac{(\Sigma T_1)^2 + \dots + (\Sigma T_4)^2}{b} - C.F = \frac{18^2 + 18^2 + 19^2 - 252.08}{4}$$

$$= 252.25 - 252.08$$

$$= 0.17$$

$$SS_{error} = SS_{total} - SS_{block} - SS_{trt}$$

$$= 42.92 - 16.25 - 0.17$$

$$= 26.5$$

7. ANOVA Table

SV	df	SS	MS	F
Block	3	16.25	5.4167	1.2264
Treatment	2	0.17	0.085	0.0193
Error	6	26.5	4.4167	
Total	11	42.92		

Treatment

Block

F-tab_(2, 6) 5.14 (P<0.05)

F_(3, 6) 4.76 (P<0.05)

F-tab_(2, 6) 10.92 (P<0.01)

F_(3, 6) 9.78 (P<0.01)

Since the F-tabulated is greater than the F-Calculated at 0.05 and 0.01 levels of significance respectively, the H₀ is accepted. It is therefore concluded that there no significant difference in the treatments.

LATIN SQAURE DESING

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:

R ₁	A	C	B
R ₂	C	B	A
R ₃	B	A	C
	C ₁	C ₂	C ₃

Where R₁ – R₃ are the rows

C₁ – C₃ are the columns

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

This is 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

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

Where Y_{ijk} = individual observation

μ = general mean

r_i = effect of the rows

C_j = effect of the columns

T_k = effects of treatments

\sum_{ijk} = 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 in 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:

	WEEK			
	1	2	3	4
1	A9.93	B5.72	D5.83	C9.67
2	D6.67	A8.51	C10.00	B9.49
3	C5.89	D3.36	B3.55	A6.92
4	B3.59	C3.84	A4.00	D6.70

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 4 Latin square design. The model and the assumption are as given earlier.

The layout

		Column week					
		1	2	3	4	Row totals	
Tree Sp	1	9.93	5.72	5.83	9.67	31.15	ΣR_1
	2	6.67	8.51	10.00	9.49	34.67	ΣR_2
	3	5.89	3.36	3.55	6.92	19.72	ΣR_3
	4	3.59	3.84	4.00	6.70	18.13	ΣR_4
Column total		26.08	21.43	23.38	32.78	103.67	
		ΣC_1	ΣC_2	ΣC_3	ΣC_4		

The above layout is for rows and columns. This will be followed by a separate layout for treatments.

Treatments (fertilizers)

	A	B	C	D
	9.93	5.72	9.67	5.83
	8.51	9.49	10.00	6.67
	6.92	3.55	5.89	3.36
	4.00	3.59	3.84	6.70
Treatments totals	29.36	22.35	29.40	22.56
Treatments means	7.34	5.59	7.35	5.64

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

TREATMENTS (FERTILIZERS)

H₀: 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

H_A: There is significant difference between the trees

COLUMNS (WEEKS)

H₀: There is no significant difference between the weeks

H_A: There is significant difference between the weeks

Computation:

$$\text{a) } C.F = \frac{(9.93 + 6.67 + \dots + 6.70)^2}{16} = \frac{(103.67)^2}{16}$$

$$= 671.7168$$

$$\text{b) } SS_{\text{total}} = (9.93)^2 + (6.67)^2 + \dots + (6.70)^2 - C.F$$

$$= 760.7845 - 671.7168$$

$$= 89.0677$$

$$\text{c) } SS_{\text{rows}} = \frac{(\sum R_1)^2 + \dots + (\sum R_4)^2}{\text{no of observations per row}} - C.F$$

$$\therefore = \frac{(31.15)^2 + \dots + (18.13)^2}{4} - 671.7168$$

$$= \frac{2889.9067}{4} - 671.7168$$

$$= 50.7599$$

$$\text{d) } SS_{\text{columns}} = \frac{(\sum C_1)^2 + \dots + (\sum C_4)^2}{\text{no of observation per coulumn}} - C.F$$

$$\therefore SS_{\text{columns}} = \frac{(26.08)^2 + \dots + (32.78)^2}{4} - 671.7168$$

$$= \frac{2760.5641}{4} - 671.7168$$

$$= 18.4242$$

$$\text{e) } SS_{\text{treatments}} = \frac{(29.36)^2 + \dots + (22.56)^2}{4} - 671.7168$$

$$= \frac{2734.8457}{4} - 671.7168$$

$$= 11.9946$$

$$\text{f) } SS_{\text{error}} = SS_{\text{total}} - SS_{\text{rows}} - SS_{\text{columns}} - SS_{\text{treatments}}$$

$$= 89.0677 - 50.7599 - 18.4242 - 11.9946$$

$$= 7.8890$$

SV	df	SS	MS	F-cal	F-Tab
Rows (trees)	3	50.7599	16.9200	12.87*	4.76
Columns (weeks)	3	18.4242	6.1414	4.67ns	4.76
Treatments (fert)	3	11.9946	3.9982	3.04ns	4.76
Error	6	7.8890	1.3148		
Total	15	89.0677			

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

$$\text{In RCBD: } M = \frac{tT + bB - S}{(t-1)(b-1)}$$

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 fro error is reduced by 1
- ii. The df for total is also reduced by 1
- iii. The SS_{trt} is reduced by Z
- iv. The SS_{total} is also reduced by Z

Z is the correction for bias

Z is calculated using the following formula:

$$Z = \frac{[B - (t-1)M]^2}{t(t-1)}$$

After those adjustments, the ANOVA table is then constructed.

SV	df	SS	MS	F
Blocks	b - 1	SSb	MSb	
Treatments	t - 1	(SSt) - Z	MSt	$\frac{MSt}{MSE}$
Error	(b - 1)(t - 1)	SSE	MSE	
Total	(bt - 1) - 1	(SST) - Z		

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.

TREATMENTS

B l o c k s		A	B	C	D	E
	I	2.62	2.67	1.04	1.88	1.74
	II	2.58	2.52	M	2.82	2.10
	III	2.65	1.92	0.91	2.65	2.78

M denotes the missing observation

Estimate the missing value and carry out the analysis of variance to test the effect of the treatments.

SOLUTION

The design is RCDB with one missing observation. The additive model and the assumption are as given for RCBD. Lay out before estimating the missing value.

	A	B	C	D	E	Blocks total	
BLOCKS	I	2.62	2.67	1.04	1.88	1.74	9.95
	II	2.58	2.52	M	2.82	2.10	10.02
	III	2.65	1.92	0.91	2.65	2.78	10.91
Treatments totals	7.85	7.11	1.95	7.35	6.62	30.88	
Treatments means	2.62	2.37	0.65	2.45	2.21		

To estimate the missing value:

$$\begin{aligned}
 M &= \frac{tT + bB - S}{(t-1)(b-1)} \\
 &= \frac{5(1.95) + 3(10.02) - 30.88}{(5-1)(3-1)} \\
 &= 1.1163 \text{ or } 1.12 \text{ to } 2 \text{ d.p}
 \end{aligned}$$

The correction for bias is:

$$Z = \frac{[B - (t-1)M]^2}{t(t-1)}$$

$$= \frac{[10.02 - (5-1)1.12]^2}{5(5-1)}$$

$$= 1.5346$$

The new lay-out (i.e. inserting the estimate of missing value is:

	A	B	C	D	E	Blocks total	
BLOCKS	I	2.62	2.67	1.04	1.88	1.74	9.95
	II	2.58	2.52	M	2.82	2.10	10.02
	III	2.65	1.92	0.91	2.65	2.78	10.91
Treatments totals		7.85	7.11	1.95	7.35	6.62	30.88
Treatments means		2.62	2.37	1.02	2.45	2.21	

The hypotheses:

H_0 : The treatments effects are not significantly difference

H_A : The treatments effects are not significantly difference

Computations:

$$C.F = \frac{(32.00)^2}{15}$$

$$= 68.2667$$

$$SS_{total} = (2.62)^2 + \dots + (2.78)^2 - C.F - t$$

$$= 74.5484 - 68.2667 - 1.5346$$

$$= 4.747$$

$$SS_{blocks} = \frac{(9.95)^2 + (11.14)^2 + (10.91)^2}{t=(5)} - C.F - Z$$

$$= 68.4260 - 68.2667$$

$$= 0.1593$$

$$\begin{aligned}
SS_{\text{trts}} &= \frac{(7.85)^2 + \dots + (6.62)^2}{b=3} - C.F - Z \\
&= 73.1488 - 68.2667 - 1.5346 \\
&= 3.3475 \\
SS_{\text{error}} &= SS_{\text{total}} - SS_{\text{block}} - SS_{\text{trts}} \\
&= 4.7471 - 0.1593 - 3.3475 \\
&= 1.2403
\end{aligned}$$

Other computations are done in the ANOVA table

ANOVA TABLE

SV	df	SS	MS	F
Blocks	2	0.1593	0.0796	
Treatments	4	3.3475	0.8369	5.3994
Error	7	1.2403	0.1550	
Total	13	4.7471		

Note: The total and df error have been reduced by 1

The F-tabulated for 4 and 7 df at 0.05 level of significance is 4.12. this value is less than the F-calculated, hence, the H_0 is not accepted. It is therefore concluded that the effects of the treatments are significantly different.

MISSING VALUE IN LATIN SQUARE DESIGN

$$M = \frac{M(R + C + T) - 2S}{(m-1)(m-2)}$$

Where M = Estimate of the missing value

- m = Number of rows or columns or treatments in the design (e.g. in a 4 x 4 Latin square, m = 4)
- R = Sum of observations in the same row as the missing value
- C = Sum of observation in the same column as the MV
- T = Sum of observation with the same treatment as the MV
- S = Sum of all the actual observations

To correct for bias

$$Z = \frac{(S - R - C - (m - 1) T)^2 - 2S}{(m - 1)^2 (m - 2)^2}$$

e.g., The following hypothetical data represent the yield (in tones per ha) of four cultivars of Khaya arranged in a 4 x 4 Latin square with one missing value. The four cultivars are A(KH4), B(KH20), C(KH29) and D(KS15).

COLUMNS

		1	2	3	4
ROWS	I	6.2 (C)	5.1 (D)	3.3 (B)	5.7 (A)
	II	5.1 (B)	3.8 (A)	M (C)	3.9 (D)
	III	5.4 (D)	4.8 (C)	6.9 (A)	4.6 (B)
	IV	3.2 (A)	2.9 (B)	5.2 (D)	5.4 (C)

Where M is the missing value. Estimate this MV and perform the analysis of variance to test for significant differences among the cultivars.

SOLUTION

The design is the Latin square design with one missing observation. The additive model and the assumption are given for Latin square design.

The lay-out (before estimating the missing value)

COLUMNS

		1	2	3	4	Row Total
ROWS	I	6.2	5.1	3.3	5.7	20.3
	II	5.1	3.8	M	3.9	12.8
	III	5.4	4.8	6.9	4.6	21.7
	IV	3.2	2.9	5.2	5.4	16.7
Column Totals		19.9	16.6	15.4	19.6	71.5

TREATMENTS

	A	B	C	D
	5.7	3.3	6.2	5.1
	3.8	5.1	M	3.9
	6.9	4.6	4.8	5.4
	3.2	2.9	5.4	5.2
Treatment Totals	19.6	15.9	16.4	19.6
Treatment Means	4.90	3.98	5.47	4.90

To estimate the missing value:

$$\begin{aligned}
 M &= \frac{M(R + C + T) - 2S}{(m-1)(m-2)} \\
 &= \frac{4(12.8 + 15.4 + 16.4) - 2(71.5)}{(4-1)(4-2)} \\
 &= 5.9
 \end{aligned}$$

The correction for bias is:

$$\begin{aligned}
 Z &= \frac{(S - R - C - (m-1)T)^2}{(m-1)^2(m-2)^2} \\
 &= \frac{[71.5 - 12.8 - 15.4 - (4-1)16.4]^2}{(m-1)^2(m-2)^2} \\
 &= 0.9669
 \end{aligned}$$

The new la-out (after estimating the MV) is:

COLUMN

		1	2	3	4	Row Total
ROWS	I	6.2	5.1	3.3	5.7	20.3
	II	5.1	3.8	M	3.9	12.8
	III	5.4	4.8	6.9	4.6	21.7
	IV	3.2	2.9	5.2	5.4	16.7
Column Totals		19.9	16.6	15.4	19.6	71.5

TREATMENTS

	A	B	C	D
	5.7	3.3	6.2	5.1
	3.8	5.1	M	3.9
	6.9	4.6	4.8	5.4
	3.2	2.9	5.4	5.2
Treatment Totals	19.6	15.9	16.4	19.6
Treatment Means	4.90	3.98	5.47	4.90

The hypotheses:

H_0 : There is no significant difference in the yield of the cultivars

H_A : There is no significant difference in the yield of the cultivars

Computations:

$$C.F = \frac{(77.42)^2}{16}$$

$$= 374.4225$$

$$SS_{total} = (6.2)^2 + \dots + (5.4)^2 - C.F - Z$$

$$= 394.12 - 374.4225 - 0.9669$$

$$= 18.7306$$

$$SS_{\text{rows}} = \frac{(20.3)^2 + \dots + (16.7)^2}{t=4} - C.F$$

$$= 3.4675$$

$$SS_{\text{columns}} = \frac{(19.9)^2 + \dots + (19.6)^2}{r=4} - C.F$$

$$= 2.9325$$

$$SS_{\text{trts}} = \frac{(19.6)^2 + \dots + (19.6)^2}{4} - C.F - Z$$

$$= 379.605 - 374.4225 - 0.9669$$

$$= 4.2156$$

$$SS_{\text{error}} = SS_{\text{total}} - SS_{\text{rows}} - SS_{\text{columns}} - SS_{\text{trts}}$$

$$= 4.7471 - 0.1593 - 3.3475$$

$$= 1.2403$$

Other computations are done in the ANOVA table

ANOVA TABLE

SV	df	SS	MS	F
----	----	----	----	---

Rows	3	3.4675	1.1558	
Columns	3	2.9325	0.9775	
Treatments	3	4.2156	1.4052	0.8658
Error	7	1.2403	0.1550	
Total	14	18.7306		

Note: The total and df error have been reduced by 1

The F-tabulated for 3 and 5 df at 0.05 level of significance is 5.41. This value is greater than the F-calculated, hence, the H_0 is not rejected. The conclusion therefore is that the yields of the four cultivars of khaya are not significantly different from one another.

Regression and correlation

When 2 quantity vary together cannot being functionally related. We may wish to measure degree of association which exist between them e.g. looking at relationship between height (h) and diameter (bh).

No inference are made about how they are connected, but are may nearly measure the way in which they move together

X (dbh)	y(Ht)
2	4
3	5
4	9
5	15

Three (3) types of trends are usually common in regression i.e positive, negative, and trend (o). if there is positive association it can be deduce that most if the data will be found in the 1st and 3rd

quadrant while a negative association will have its data in the 2nd and 4th quadrant but a number association will always distribute its data in all the 4 quadrant.

HE ^y	dbh ^x
5	10
6	8
8	15
12	12
9	6
8	9
10	12
13	5
6	10
<u>7</u>	<u>10</u>
72	97

$$\text{Statistically correlation } r = \frac{\text{spxy}}{\sqrt{\text{ssx} \cdot \text{ssy}}}$$

$$\text{Spxy} = \Sigma xy - \frac{\Sigma y \Sigma x}{n}$$

$$\text{Ssx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$$

$$\text{Ssy} = \Sigma y^2 - \frac{(\Sigma y)^2}{n}$$

Y	X	Xy	Y ²	X ²
5	10	50	25	100
6	8	48	36	64
8	15	120	64	225
12	12	144	144	144

7	6	42	49	36
8	9	72	64	81
10	12	120	100	144
13	5	65	169	25
6	10	60	36	100
7	10	70	49	100
$\Sigma = 72$	97	741		

$$r = \frac{42.6}{\sqrt{78.1 \times 57.6}}$$

$$= 0.63$$

r usually ranges from 0-1 it can neither be positive or negative or neutral i.e 0

r^2 = coefficient of variation. Determination

for the data

$$n = 10$$

$$\Sigma y = 72$$

$$\bar{y} = 7.2$$

$$\Sigma x = 97$$

$$\bar{x} = 9.7$$

$$\Sigma x^2 = 1019$$

$$\Sigma y^2 = 576$$

$$\Sigma xy = 741$$

$$r = \frac{s_{pxy}}{\sqrt{ssx \cdot ssy}}$$

$$s_{pxy} = \Sigma xy - \frac{\Sigma y \Sigma x}{n} \times 741 - \frac{72 \cdot 97}{10}$$

$$= 42.6$$

$$S_{sx} = \sum x^2 - \frac{(\sum x)^2}{n}$$

$$= 1019 - \frac{97^2}{10}$$

$$= 78.1$$

$$S_{sy} = \sum y^2 - \frac{(\sum y)^2}{n}$$

$$576 - \frac{72^2}{10}$$

$$= 57.6$$

Value of r cannot be said to be significant or not until the value are checked up in the stat table also the level of threshold should be known. i.e. $r(0.1)_{10}$ 0.1 is the 1% probability level and 10 stands for the two of observation.

Threshold determine time significant level of the result.

The probability level can either be 99% (1%) or 95% (5%).

$$= r(0.1)_{10}$$

$$r(0.5)_{10}$$

99% means the result will continue to be correct 99 times out of 100.

The 10 outside the bracket is the number of observation

$$r(0.1)_{10} = .794$$

$$r(0.5)_{10} = .648$$

This is effect means the result is not significant at 95% probability level.

Inference

If calculate is higher than Tab value then the statistic is significant and the reverse i.e if calculated vale is lower than tabulated value than the statistic is not significant

If null hypothesis says there is correlation between Ht and dbh and the result is not significant than the hull hypothesis rejected but if the statistic is significant then we say the null hypothesis is not rejected.

Coefficient of determination r^2

$R^2 \rightarrow$ this measure the proportion of variation of eh is associated Σ variation in X. if ranges from 0-1 and it must always be positive.

$$r = 0.63$$

$$r^2 = (0.63)^2$$

$$= 0.3969$$

$$= 0.397$$

What this one means is that 39% of variation in Y is attributable to variation in X

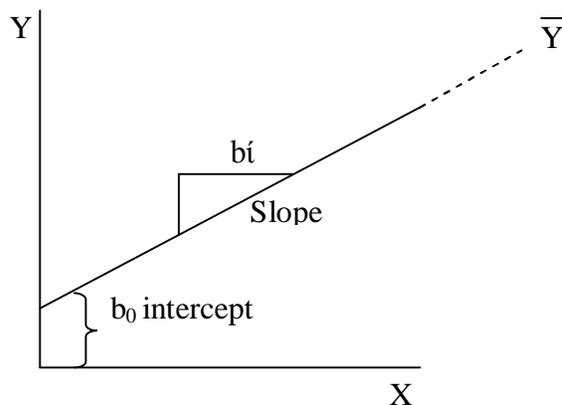
REGRESSION

If there are 2 variable X and Y and the live of relationship is needed then regression is always calculated. If there are only 2 variables then are talk of simple regression also there is always a dependent variable of an independent variable.

In case of Ht and dbh, dbh is easier to get and the Ht is the one we are interred in is the dependent variable while the dbh from which are want to evaluate Ht is the independent variable
Hence Ht = Y and dbh = X.

The known variable = X, the unknown is the Y

Single linear regression = $y = b_0 \pm b_1x$



Intercept is always present in a line that does not urgency from O, if a line start from O then the intercept is O or non existing.

$$Y = B_0 + B_1X + E_i$$

E_i = Error term

Y = estimated value of y for projection.

$$b_1 = \frac{SPXY}{SSX}$$

$$b_0 = \bar{y} - b_1x$$

from the previous data

$$b_1 = \frac{42.6}{78.1} = 0.545$$

$$b_0 = 7.2 - 0.545 \times 9.7$$

$$= 7.2$$

$$= 1.91$$

The equation then become $y = 1.91 + 0.545x$

If $X = 20$ then $Y = 1.91 + 0.545 \times 20$

Using computer to do this type of work an ANOVA table will be generated.

ANOVA

Source of variation	df	Ss	ms	F
Regression	1	23.22	23.22	5.4
Error	8	34.38	4.30	
Total	9	57.6		

Regress df = 1 because there are 2 variables

Ss regression = b_1 SPXY

$$Ss \text{ total} = SSY$$

$$b_1SPXY = 0.545 \times 42.6$$

$$= 23.21$$

$$SSY = 57.6$$

$$Ms \text{ regression} = \frac{SSr}{dfr}$$

$$MSE = \frac{SSE}{df}$$

$$F = \frac{MS}{MSE}$$

The essence of going to calculate F is to know whether the reg. line is significant or not.

$$\text{The threshold} = F_{\alpha 0.5}(1, 8) = 5.32$$

$$1 = \text{df of regression}$$

$$8 = \text{df of error}$$

$$F_{\alpha 0.1}(1, 8) = 11.3$$

Since E calculate (5.4) is $>$ Tab 5.32 at 95% than the statistic is significant.

But at 99% Tab 11.3 is $>$ 5.4 and thus the result is not significant.

REGRESSION MODELS

If there is no significant or else then transformation into 1st order polynomial $Y = b_0 + b_1X$ will be done and if it is still not significant we move to the 2nd order polynomials = $Y = b_0 + b_1X + b_2X^2$

$$\text{3rd} = Y = b_0 + b_1X + b_2X^2 + b_3X^3$$

2nd order = quadratic

3rd order = cubic

4th order = quartic

Log transformation

$$Y = b_0 + b_1 \log X \text{ ----- single log}$$

Log Y = b₀ + b₁ log X ----- double log

1st order

$$Y = a + b_1x$$

2nd $Y = a + b_1X + b_2X^2$

3rd $Y = a + b_1X + b_2X^2 + b_3X^3$

At every stage ANOVA table will be constructed and difficult F and r² will be found.

4th $Y = a + b_1X + b_2X^2 + b_3X^3 + b_4X^4$

IQ f age NL of a child

Y X Z

Model = Y = a + b₁X ± b₂Z₂ ---- multiple regression but not transformation.

Pr	Qw	L	Ec	Mc	U
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-

Predictor	coefficient	standard deviation	t-ratio	Probability
Constraint	-106	938	-1.14	0.299
QW	0.62	0.38	1.61	0.159
L	0.74	3.38	0.22	.835
EC	3.81	4.68	0.81	.447
Mc	-12.55	33.09	-0.38	.718
U	-43.92	51.56	-0.85	.427

$$r^2 = 55.1$$

ANOVA TABLE

Source	df	Ss	ms	F	P
Regression	5	3.208×10^{12}	7.617×10^{11}	1.77	0.323
Error	6	3.104×10^{12}	5.174×10^{11}		
Total	11	6.91×10^{12}			

Initially there are 1 dependent and 5 independent variable = 6 and $n = 12$

The most important things for the model is the coefficient column

Hence the model for Pr

$$= Pr = -106 + 0.62Qw + 0.74L + 3.81EC - 12.55 mc - 43.92u.$$

The other important thing is the P each of taken from 1 will give the level significant

$$= 1 - p$$

$$= 1 - 0.299 = 0.71$$

$$= 1 - 0.159 = 0.841$$

$$= 1 - 0.835 = 0.165$$

$$= 1 - 0.447 = 0.553$$

$$= 1 - 0.718 = 0.212$$

$$= 1 - 0.427 = 0.573$$

Conversing them to percentage are will then get the percentage significant level it will give them

help to know whether anyone is significant at 99% or 95%.

i.e $0.71 = 71\%$

$$0.841 = 84.1\%$$

$$0.165 = 16.5\%$$

$$0.553 = 55.3\%$$

$$0.212 = 21.2\%$$

$$0.573 = 57.3\%.$$

Thus none of this is significant at 99/95% level.

If r^2 is converted so r it become

$$r^2 = 55\% = 0.55$$

$$r = 0.742$$

Then check whether the regression is significant in the stat table

How to check multiple regression stat table check $n = 12$ under $k = 5$ (independent variable).

i.e $r_{0.5}(5, 12) = 0.886$ at $r_{0.1}(5, 12) = 0.938$. our r is not significant or either 95% and 99% level since r calculate = 0.742 while r tab = 0.8 and 0.9. if the tried since a model need to be developed.

P should be ranked in order of importance based on their figure. Neglecting the constraint then, step wise analysis should be done; this will be eliminating 5 variables

STEP WISE REGRESSION

Variable	1	2	3	4
Const	-106	-53	-33	-29
Qw	.65	0.62	0.67	0.72
L	1.9	1.5	1.9	-
Ec	0.9	1.0	-	-
Mc	-7	-	$r^2 = 48.61$	$r^2 = 43.06$
U	$r^2 = 49.66$	$r^2 = 49.32$	-	-

1st step equation

$$Pr = -106 + 0.65Qw + 1.9L + 0.9EC - 7mc$$

2nd

$$Pr = -55 + 0.62Qw + 1.5L + 1.0Ec$$

3rd

$$Pr = -33 + 0.67 Qw + 1.9L$$

4th

$$Pr = -29 + 0.72 Qw$$

Back to Anova table

$$P = 0.323$$

$$1 - p = 0.677 = 67.7\%$$

i.e it is not significant of either 95/99%

Considering

From table $F_{0.05}(5, 6) = 4.95$

df error = 6 t stat df error

Trqmt = 5 $F_{0.1}(5, 6) = 10.7$

F tab at 95% shows 4.95 while focal shows 1.47 i.e it is not significant at 95% not to talk of 99%

For transformation

r^2 for semilog = 52.1%

r^2 for Normal = 55%

r^2 for Double log = 33.1%.

semilog means conveying dependent variables as it is.

2nd order polynomials

$$Pr = a + Qw + Qw^2 + L + L^2 + Ec + Ec^2 + Mc + Mc^2 + U + U^2$$

