

JARA NASEER.

Manual For

Biometric Techniques For Agricultural Research In Pakistan

Using

STATISTICA

Azhar Mahmood

0333-6566310

Department of Mathematics & Statistics

University of Agriculture

Faisalabad

TABLE OF CONTENTS

1. ROLE OF STATISTICS IN RESEARCH	1
1.1. SCIENTIFIC METHOD	1
1.2. TYPES OF REASONING	2
1.2.1. DEDUCTIVE	2
1.2.2. INDUCTIVE	2
1.3. APPLICATIONS OF STATISTICS IN RESEARCH	4
2. INTRODUCTION	4
2.1. SOME USEFUL DEFINITIONS	5
2.2. MEANING OF DESIGN OF AN EXPERIMENT	5
2.3. STATISTICAL DESIGN	5
2.4. FUNCTIONS OF A STATISTICAL DESIGN	5
2.5. BASIC PRINCIPLES OF STATISTICALLY SOUND DESIGN	5
2.5.1. REPLICATION	6
2.5.1.1. HOW MANY REPLICATIONS?	7
2.5.2. RANDOMIZATION	7
2.5.2.1. WHY RANDOMIZE?	8
2.5.3. LOCAL CONTROL	8
2.5.3.1. GROUPING	8
2.5.3.2. BLOCKING	9
2.5.3.3. BALANCING	10
2.6. SUMMARY	12
3. STRATEGIES-DESIGN OF EXPERIMENTS	12
3.1. STATING THE OBJECTIVES	12
3.2. DEFINING THE POPULATION ABOUT WHICH INFERENCES ARE TO BE MADE	12
3.3. SELECTION OF EXPERIMENTAL TREATMENTS	12
3.4. PLOT SHAPE AND SIZE	13
3.5. NUMBER OF REPLICATIONS	13
3.6. LAYOUT OF THE EXPERIMENT	14
3.7. RANDOMIZATION	15
3.8. RECORDING OF RESULTS	15
3.9. PLANNING FOR ANALYSIS	16
3.10. THE FINAL (AND MOST IMPORTANT) QUESTION	16
4. COMPLETELY RANDOMIZED DESIGN	18
4.1. INTRODUCTION	18
4.1.1. ADVANTAGES OF COMPLETELY RANDOMIZED DESIGN	18
4.1.2. DISADVANTAGES OF COMPLETELY RANDOMIZED DESIGN	18
4.2. INTERPRETATION OF THE RESULTS OF ANOVA	19
4.2.1. INSIGNIFICANT F-RATIO FOR TREATMENTS	19
4.2.2. F-RATIOS THAT ARE LESS THAN UNITY	19
4.2.3. COEFFICIENT OF VARIATION	20
4.3. INFERENCE AFTER ANOVA	20
4.3.1. MULTIPLE COMPARISON TESTS	20
4.3.2. CONTRASTS	21
4.3.3. RESPONSE CURVES: WHEN TREATMENTS ARE DIFFERENT LEVELS OF QUANTITATIVE FACTOR	22
4.4. PRESENTATION OF RESULTS	22
4.5. THE RELATION BETWEEN CRD AND T-TEST	23
5. RANDOMIZED COMPLETE BLOCK DESIGN	27
5.1. INTRODUCTION	27
5.1.1. ADVANTAGES OF RANDOMIZED COMPLETE BLOCK DESIGN	27

5.1.2. DISADVANTAGES OF RANDOMIZED COMPLETE BLOCK DESIGNS	26
5.2. RELATIVE EFFICIENCY OF RCBD AS COMPARED TO CRD	31
5.3. MISSING OBSERVATION IN RCBD	32
6. LATIN SQUARE DESIGN	35
6.1. INTRODUCTION	35
6.1.1. ADVANTAGES OF LATIN SQUARE DESIGN	35
6.1.2. DISADVANTAGES OF LATIN SQUARE DESIGN	37
6.2. RELATIVE EFFICIENCY OF LS DESIGN AS COMPARED TO RCBD	38
6.3. MISSING OBSERVATION IN LATIN SQUARE DESIGN	38
7. GRAU-3-LATIN SQUARE DESIGN	ERROR! BOOKMARK NOT DEFINED.
7.1. INTRODUCTION	ERROR! BOOKMARK NOT DEFINED.
8. FACTORIAL EXPERIMENTS	40
8.1. INTRODUCTION	43
8.2. INTERPRETING THE RESULTS FROM A FACTORIAL EXPERIMENT	44
8.3. ADVANTAGES	44
8.4. DISADVANTAGES	45
8.5. TYPES	46
8.6. FACTORIAL EXPERIMENT UNDER CRD	49
8.7. FACTORIAL EXPERIMENT UNDER RCBD	52
8.8. FACTORIAL EXPERIMENT UNDER LATIN SQUARE DESIGN	54
8.9. GENERAL ANOVA TABLES (FOR FACTORIAL EXPERIMENTS)	55
9. SPLIT PLOT DESIGN	55
9.1. INTRODUCTION	56
9.1.1. THE SPLIT-PLOT DESIGN IS DESIRABLE IN THE FOLLOWING SITUATIONS	57
9.1.2. STANDARD ERROR OF THE MEAN DIFFERENCE FOR EACH OF THE FOUR TYPES OF PAIR COMPARISONS IN A SPLIT-PLOT DESIGN	64
9.1.3. GENERAL ANOVA FOR 2 AND 3 FACTORS SPLIT PLOT DESIGN	65
10. SPLIT BLOCK OR STRIP PLOT DESIGN	65
10.1. INTRODUCTION	65
11. NESTED OR HIERARCHICAL DESIGNS	71
11.1. INTRODUCTION	71
12. INCOMPLETE BLOCK DESIGNS	74
12.1. INTRODUCTION	74
12.1.1. BALANCED INCOMPLETE BLOCK DESIGNS (BIBD)	75
12.1.2. PARTIALLY BALANCED INCOMPLETE BLOCK DESIGNS (PBIBD)	76
12.2. BALANCED LATTICE DESIGN	77
12.2.1. ADVANTAGES OF LATTICE DESIGN	77
12.2.2. DISADVANTAGES OF LATTICE DESIGN	77
12.3. PARTIALLY BALANCED LATTICE DESIGN (DOUBLE, TRIPLE AND QUADRUPLE LATTICE DESIGN)	78
12.4. RANDOMIZATION	78
12.5. REPEATABLE DESIGN	78
12.6. CHOICE AMONG DIFFERENT DESIGNS	79
12.7. ANALYSIS OF BALANCED LATTICE DESIGN	80
13. GENERALIZED LATTICE DESIGN (OR ALPHA DESIGN)	84
13.1. INTRODUCTION	84
13.1.1. EFFICIENCY FACTOR	88
13.2. ROW-COLUMN DESIGNS	89
13.3. FRACTIONAL DESIGNS	89

14. PROCEDURE TO PERFORM DIFFERENT ANALYSIS OF EXPERIMENTAL DESIGN USING STATISTICA	90
14.1. CRD.....	91
14.2. RCBD.....	91
14.3. LATIN SQUARE DESIGN.....	92
14.4. FACTORIAL EXPERIMENT UNDER CRD.....	92
14.5. FACTORIAL EXPERIMENT UNDER RCBD.....	92
14.6. SPLIT PLOT EXPERIMENT UNDER CRD.....	93
14.7. SPLIT PLOT EXPERIMENT UNDER RCBD.....	93
14.8. STRIP PLOT DESIGN.....	94
14.9. NESTED DESIGN.....	95
14.10. INFERENCES BEYOND ANOVA.....	95
14.10.1. MULTIPLE COMPARISON TESTS.....	95
14.10.2. CONTRASTS.....	95
14.10.3. ORTHOGONAL POLYNOMIAL CONTRASTS.....	96
15. REFERENCES.....	97
16. APPENDIX.....	97
16.1. Z TABLE.....	97
16.2. PERCENTAGE POINT OF CHI-SQUARE DISTRIBUTION.....	98
16.3. PERCENTAGE POINT OF STUDENT'S T-DISTRIBUTION.....	99
16.4. PERCENTAGE POINT OF F DISTRIBUTION.....	99
16.4.1. 10 PERCENT POINTS OF F DISTRIBUTION.....	100
16.4.2. 5 PERCENT POINTS OF F DISTRIBUTION.....	101
16.4.3. 1 PERCENT POINTS OF F DISTRIBUTION.....	101

1. ROLE OF STATISTICS IN RESEARCH

Research according to Webster is "studious inquiry or examination i.e critical and exhaustive investigation or experimentation having for its aim the discovery of new facts and their correct interpretation".

It aims at

- Revising accepted conclusions, theories or laws in the light of newly discovered facts.
- Continued search for knowledge and understanding

Therefore research is an inquiry into the nature of, the reasons for and the consequences of any particular set of circumstances, whether these circumstances are experimentally controlled or recorded just as they occur.

1.1. SCIENTIFIC METHOD

It is a pursuit to achieve a systematic interrelation of facts, through experimentation, observation and logical arguments from accepted postulates, and a combination of these three in varying proportions. Thus research and scientific method are closely related, if not one and the same thing.

Statistics supplies a kit of tools that can be extremely valuable in research. This course will present these tools and methods of how, when and why to apply these would be discussed. For a research worker there is a lot to offer in statistics and an intelligent research worker is always making use of statistics in planning, analyzing and interpreting the results of research.

The word statistics is used in a variety of different ways. When used in singular sense it is synonymous with data. When used as a plural to word statistic it refers to quantities calculated from sample observations. Also in singular sense Statistics is a science which deals with:

- Collecting and summarizing data
- Designing experiments and surveys
- Measuring the magnitude of variation in both experimental and survey data
- Estimating population parameters and providing various measures of the accuracy and precision of these estimates
- Testing hypotheses about populations and
- Studying relationships among two or more variables

In short statistics and research go hand in hand in pursuit of new knowledge.

1.2. TYPES OF REASONING

1.2.1. DEDUCTIVE

It is the determination of the consequences involved in a set of premises cannot therefore itself add to knowledge of natural phenomena but can be used to aid the verification of theories based on a set of premises.

1.2.2. INDUCTIVE

It is the process of generalization, of drawing general conclusions from a particular set of instances being a fair sample of the population about which we make statements. Induction results lack certainty and contain an element of error or probable ness.

The theory of statistics which is entirely deductive provides the basis of inductive inference/process.

Role of Statistics in the scientific method uses three functions, namely description, analyses and prediction.

Description deals with the reduction in the mass of data, e.g. ages of all people in Pakistan to a small set of numbers as possible. It may be deductive if the data is for the whole population or inductive if the data is collected from only part of the population.

When we are dealing with samples and using induction then we are more in need of analyses. We want to make comparisons of populations on the basis of information collected from samples drawn independently. We want to estimate the characteristics of the samples and this requires use of analysis.

Description and analysis help to make predictions. So we see there is a strong link between all three functions of statistics and we need to understand this linkage to make use of it.

1.3. APPLICATIONS OF STATISTICS IN RESEARCH

- Improved analytical tools in agricultural and biological sciences
- Agronomists conduct endless number of experiments to determine differences among yields of various crop varieties, effects of fertilizers and best methods of cultivation etc. for better yields for higher productivity.
- Poultry breeding, animal breeding and animal nutrition etc.

In the past it was thought that statistics has no place in the so called "exact sciences" such as chemistry, physics and many branches of engineering. However for all laboratory

experience the application of statistics improves quality of research. Even the chart of elements showing the atomic weights uses weighted average weights of the atomic weights of the individual isotopes of the given element, the weights being the frequency of occurrence of the elements in a normal or naturally occurring mixture.

- Statistical applications are nowadays very common in fields like meteorology and astronomy. Data used for forecasts is given statistical treatment before it can be used for forecasting weather.
- Statistics is now playing an important role in engineering like study of heat transfer through insulating material/units time, performance guarantee testing programs, production control, and inventory control, standardization of fits and tolerances of machine parts.
- Agricultural engineering combines the practices of engineering and agriculture. Statistics help evaluating the performance of machines at different settings.
- Social sciences have also wide applications of statistics. Survey methods are used to evaluate public opinion or behavior to different policies of the Government. In economics statistical applications are almost indispensable.
- Evenly statistics has been used to identify criminals through DNA testing to help law and order authorities in the area of criminology.

This can be an unending list. Some other areas of importance are public health studies, epidemiology, demography, biological assay, psychology, education, sociology and various areas of home economics.

The subject matter of statistical applications is not only confined to scientific world but has applications in arts too.

- It has been used as aid in determining the authorship of certain manuscripts by analyzing the length of sentences.
- Authenticity of paintings has also been established by analyzing the frequency of brush strokes.

2. INTRODUCTION

2.1. SOME USEFUL DEFINITIONS

1. An *experiment* is a planned inquiry to discover new facts, or to confirm or deny the results of previous investigations.
2. A *treatment* is a procedure whose effect on the experimental material is to be measured.
3. A particular class of related treatments is often called *factor*.
4. The states of a factor, i.e., the treatments within the class, are called the *levels* of the factor.
5. An *Experimental Unit (EU)* is the piece of experimental material to which one trial of a single experiment is applied.
6. A sampling unit is that fraction of the experimental unit on which the effect of the treatment is measured.
7. A group of homogenous experimental units is called a *block*.
8. An *experimental design* is a set of rules by which the treatments to be used in an experiment are assigned to the experimental units.
9. If treatments are assigned to a set of units in such a way that every unit is equally likely to receive any treatment, the assignment is said to be *random*.
10. When a treatment appears more than once in an experiment, the treatment is said to be *replicated*.
11. *Experimental error* is the variation among experimental units, which have been treated alike.

Variation comes from two main sources. Firstly there is the inherent variability that exists in the experimental material to which treatments are applied. Second, there is the variation which results from any lack in uniformity in the physical conduct of the experiment.

One of the major goals of a sound research procedure is to reduce experimental error and thus increase precision. Experimental error can be reduced in the following ways:

- Blocking
- Increase the number of replication
- Proper use of efficient statistical design

- Concomitant variable
- Size and shape of experimental units.

2.2. MEANING OF DESIGN OF AN EXPERIMENT

It is the planning of an experiment in such a way that the information that would be collected is relevant to the problem under investigation. Therefore, design of an experiment is a complete sequence of steps taken ahead of time to ensure that the appropriate data will be obtained in a way that allows an objective analysis leading to valid inferences with regard to the set hypotheses.

2.3. STATISTICAL DESIGN

A statistical design is a set of statistical rules for allocating the given number of experimental units/plots to the given number of treatments.

2.4. FUNCTIONS OF A STATISTICAL DESIGN

1. Disentangle treatment differences/variation from variation/differences among experimental units/plots; say to separate wheat from chaff.
2. To provide unbiased estimates of treatment differences.
3. Measure uncertainty in the estimates of treatment differences in terms of Standard Deviation (SD), Standard Error of estimates (SE) etc.

2.5. BASIC PRINCIPLES OF STATISTICALLY SOUND DESIGN

1. Replication of treatments
2. Randomization of treatments
3. Local control of variation among experimental units

2.5.1. REPLICATION

It is the number of distinct non-overlapping experimental units assigned to "t" treatments. No experiment should be conducted without complying with this basic principle. This is the most important ingredient of an experiment.

Replication is must as it helps to:

1. Estimate experimental error

In the field experiment there is always an uncontrolled variable in addition to the main variable (variable of interest) which influences the values of the main variable. This uncontrolled variable is called concomitant variable or simply covariate.

2. Give precise estimates of treatment effects as any effects due to uncontrolled factors go into experimental error and only treatment effects remain in the estimates of the treatment effects.

NOTE/CAUTION

Multiple observations taken from one experimental unit only or over certain time interval are not replications. These do not provide a valid estimate of experimental error. If multiple observations are taken from different parts of experimental unit they are called samples and their variation results in sampling error.

2.5.1.1. HOW MANY REPLICATIONS?

The number of replications in an experiment depends upon

1. Resources available (This often dominates everything else).
2. Magnitude of experimental error.
3. Size of treatment differences which is biologically important.
4. A small difference to be detected requires more replications than a large difference.

If we know Coefficient of Variation (CV) and Standard Error of Mean (SEM) from some previous experiment of a similar type we can use their relationship to get an idea about the number of replications for a new experiment which is planned.

If no such information exists then we should use the principle that at least 10-15 degrees of freedom are available for estimating σ^2 . This means that error degrees of freedom in any case should not be less than 10. This gives something like the following for replicates

No. of Treatments	2	3	4	5	6
No. of Replicates	10-15	6-9	4-6	4-5	4

NOTE:-As a rough estimate use as many number of replication for which error degree of freedom is not less than 8 and number of replication for which error degree of freedom greater than 20 is wastage of resources. In case of later situations try to add some more treatments or other factors to have more beneficial use of the resources. This is like asking more questions.

2.5.2. RANDOMIZATION

The allocation of treatments to the experimental units in such a way that each unit has an equal chance of receiving any treatment is called randomization.

1. Randomization is necessary in order to eliminate any personal bias in the allocation of treatments to the experimental units.
2. Every treatment gets the same chance to go to any experimental unit/plot.

2.5.2.1. WHY RANDOMIZE?

1. To validate statistical analysis (statistical reasoning: Independence of errors).
2. To protect against bias in estimates of treatment differences (Scientific reason).

Arrangement of treatment should not intentionally combine with any systematic pattern among plots/units to produce systematic error in treatment differences.

EXAMPLE

Experimental Units	1	2	3	4	5	6	7	8
Variety	A	B	A	B	A	B	A	B

Suppose there is a fertility trend among experimental units from left to right. Then B always fall in a relatively less fertile land as compared to A, therefore, B is in a disadvantageous position altogether. It is very rare that randomization will results in the above systematic allocation. Watchful eyes can even sometimes detect a bad randomization and go for randomizing once more

EXAMPLE

Ten measurements each are taken on two machines for checking their speed of making a complicated calculation. First 10 measurements are taken of machine A and the 10 measurements are taken on machine B. Suppose that the operator taking measurements makes less errors and is becoming expert with every measurement taking operation. Then one machine is at disadvantage as compared to the other, so bias creeps in the treatment estimates. Therefore, we must randomize except when there are compelling practical reasons not to, and recognize the limitations of inferences resulting from not randomizing.

2.5.3. LOCAL CONTROL

It refers to amount of balancing, blocking and grouping of experimental units that is employed in the adopted statistical design. The purpose of local control is to make the experimental design more efficient. It makes any test of significance more sensitive or test procedure more powerful. This increase in efficiency or sensitivity or power occurs because a proper use of local control will reduce the magnitude of the estimate of experimental error.

2.5.3.1. GROUPING

This means placing of a set of homogenous experimental units into homogenous groups in order that the different groups may be subjected to different treatments. These groups may be subjected to different number of experimental units.

EXAMPLE

A pharmaceutical company is investigating the comparative effects of three proposed compounds (A, B, and C). Experiment is to be conducted on rats by injecting compound and recording reaction. We have a litter consisting of 11 rats available for this experiment. We randomly assign rats to three groups and inject compounds A, B, and C. The three groups have rats 4, 4 and 3. The compounds A, B, and C are given to groups G_1 , G_2 , and G_3 respectively.

2.5.3.2. BLOCKING

This means allocation of the experimental units in such a way that the units within a block are relatively homogenous, while blocks may differ from one another. That is, using the researcher's prior knowledge about the experimental units or experimental situation. He uses this knowledge to control variation and comes up with a more efficient design.

EXAMPLE

Let us assume that 12 rats are available and pedigree shows that 6 are from litter X, 3 are from litter Y and 3 from litter Z. Since it may well be expected that rats in the same litter will perform more nearly alike than rats from different litters (due to inherited characteristics), it would seem natural to form three blocks. The first block would contain the 6 rats from litter X, the second would contain the 3 rats from litter Y, and the third would contain the 3 rats from litter Z. The three treatments (A, B, and C) would then be assigned at random to the rats within blocks. Since each rat is subjected to only one treatment, the block containing 6 rats

would undoubtedly end up with 2 rats seeing treatment A, 2 seeing treatment B, and 2 seeing treatment C. The other two blocks would have single rats seeing each treatment.

We can see here that there are three blocks but each treatment is repeated 4 times. So it is not necessary and important to note that blocks and replicates are the same.

2.5.3.3. BALANCING

This will mean obtaining of the experimental units, the grouping, the blocking, and the assignment of treatments to the experimental units in such a way that a balanced configuration results. We may have little or no balance, partial balance, approximate balance or complete balance in any particular design.

Important points for grouping plots/experimental units into blocks such that:

- Plots within a block are as like as possible.
- Plots in different blocks are as different as possible.
- Bad plots in one block. Good plots in another block.
- Each treatment gets each kind of plot.
- Estimates of treatment differences are more accurate because the systematic variation between blocks is removed from these estimates.
- The experimental error now remaining consists of plot to plot variation within blocks which due to randomization of treatments to plots within block, is effectively random in magnitude and direction.
- The magnitude of this experimental error is smaller than what it would have been in the absence of blocking, provided blocking has been properly and effectively implemented.
- Treatments are compared under more nearly equal conditions because comparisons are made within blocks of uniform units.
- It can sometimes increase the information from an experiment. Blocks need not be placed at the same location or run at the same time. By placing blocks at different locations, for example, a wider variety of conditions can be sampled with a given experiment.

2.6. SUMMARY

In a broader sense the usual experimentation takes place such that we have a number of different treatments, one of which is applied to each experimental unit. We observe the response of each treatment over a period of time and measure the response as decided before the start of the experiment. Our main objective is to separate the treatment effects from uncontrolled variation among units. Once the treatments, the experimental units, and the nature of observations have been decided the principal requirement of the experiment are:

- Experimental units receiving different treatments should differ in no systematic way from one another.
- Experimental error should be reasonably small and this should be achieved with as few units as possible in order to reduce the cost of the experiment.
- The conclusions should have wide range of validity.
- The results should be achieved with the simplest possible design with desired precision.
- A proper statistical analysis of the data should be possible without making any unrealistic assumptions.

We adhere to the principles of experimental design as our objective is to increase accuracy in conducting an experiment, either to reduce the width of the interval estimates or to decrease the size of the differences for significance. We get increased accuracy as the standard error of a treatment mean, S_Y , decreases. There are a number of ways we can increase accuracy:

- Increase the size of the experiment by increasing replications or by including more treatments. We should be careful as this can also be self defeating as increasing the size may result in heterogeneous experimental units.
- Proper selection of treatments may lead to use of factorial treatment structure which had a feature of hidden replication for some comparisons. This can help to improve the precision and stability of the estimates.

- Grouping the experimental units into homogeneous blocks is a well established technique and must be used where prior information about the type and nature of experimental units is available.
- Refinement of the technique is also a useful method of improving accuracy of the experiment
- Measure on a concomitant variable where available can be used through analysis of covariance to improve the accuracy of the experiments.

Once the experiment has been completed, the data obtained are subjected to a procedure called analysis of variance. This is a systematic way of partitioning the total variation into components each of which is associated with a source of variation. One of the important sources of variation is the variation among the experimental units treated alike. This is a measure of experimental error which provides the basis for interval estimates and significance tests and unless this is available there is no use of doing experiments.

3. STRATEGIES-DESIGN OF EXPERIMENTS

Below given is a check list of questions which the scientist should address to himself to become a good designer of experiments and a research.

3.1. STATING THE OBJECTIVES

1. Have you stated clearly and explicitly the objectives of the experiment and the reasons for undertaking it?
2. Have you translated these objectives into precise questions that the experiment can be expected to answer?

3.2. DEFINING THE POPULATION ABOUT WHICH INFERENCES ARE TO BE MADE

1. Have you defined carefully the population about which you are seeking to make inferences from the results of the experiment?
2. Is the site or location of the experiment representative of that defined population?
3. Is the experimental material to be used in the experiment, e.g. plants, animals, soil, water, etc., representative of the defined population?

3.3. SELECTION OF EXPERIMENTAL TREATMENTS

1. Have the experimental treatments been defined sufficiently precisely for them to be applied correctly by the experimenter or by those wishing to repeat the experiment, and are they realistic?
2. If the "treatments" consist of species, varieties, or strains of organisms, are they representative of some defined population of organisms?
3. Can the experimental treatments be expressed as "factors", that is as groups of treatments at two or more levels?
4. If so, can all combinations of factors be achieved and are these combinations realistic?
5. Is the number of levels within each factor restricted to two or three?
6. If not, is there any real advantage in using more than three levels to determine the shape of the response curve?
7. Do the levels of any one factor change by a constant amount or in a constant ratio?
8. If not, is there a good reason for departing from linear relationships or relationships which can be made linear by an appropriate transformation?

9. Is the number of factorial combinations so large that there would be some advantage in considering only some of those combinations, perhaps sequentially?
10. Is there a naturally defined control treatment which should be included in the experiment?

3.4. PLOT SHAPE AND SIZE

1. Is the plot size for the experiment defined by the nature of the experimental material or the site?
2. If not, will the proposed plot size enable the treatments to be applied and allow the desired records to be made?
3. Is the plot shape defined by the nature of the experimental material or treatments?
4. If not, will the proposed plot shape enable the treatments to be applied and allow the desired records to be made?
5. Are the experimental plots all of the same size and shape?
6. If not, are you aware of the problems that may be encountered during the analysis of the results of the experiment?
7. Is there likely to be interaction between the individual plots of the experiment?
8. Can this competition be reduced by increasing the space between plots or surrounding each plot by a buffer zone?
9. Are the plots of the experiment of the smallest size consistent with the other constraints?

3.5. NUMBER OF REPLICATIONS

1. Do you have any preliminary estimates of the precision likely to be achieved by the experiment (expressed as a coefficient of variation, for example)?
2. Is it possible to conduct a pilot experiment to determine the coefficient of variation likely to be encountered, and to test the experimental procedures?
3. Have you determined the size of the difference between treatment means which you would regard as of practical importance, if such a difference were to exist?
4. Have you calculated the number of replications that would be necessary to match the size of the differences likely to be detected as significant with the size of differences you regard as of practical importance?

(e.g. $n = (CV/S)^2$ where n = number of replications and S = standard error of mean)

5. Do the controls need to be replicated more or less frequently than the other treatments, in order to place greater emphasis on particular comparisons?

3.6. LAYOUT OF THE EXPERIMENT

1. Is it possible to divide the site of the experiment or the experimental material into blocks within each of which there will be less variation than over the experiment as a whole?
2. Is the size of these blocks sufficiently large to contain at least one plot of each treatment and controls?
3. Have you considered the advantages of robustness and ease of analysis of a randomized block design?
4. If the blocks are not large enough to contain at least one plot of each treatment and controls, is there some way of allocating the treatment replications so that the important comparisons are estimated with the greatest precision?
5. If the treatment comparisons are not orthogonal, do you know how the data can be analyzed, and will that analysis answer the questions the experiment is designed to pose?
6. Are there any regular trends across the experimental site or material? If so, are these trends in one or both directions?
7. Have you considered the use of row and column designs to remove the effects of one or two way trends?
8. Is there likely to be any advantage in the use of a split plot design?
9. If so, are the treatments applied to the sub-plots the ones for which the greatest precision is required?
10. Will confounding of treatment factors or interactions with block differences improve the efficiency of the design?
11. Have you planned to use the blocks of the experiments to absorb as much as possible of the extraneous variation in the execution and conduct of the experiment?
12. Is it possible that plots may be lost through accidents or mishaps?
13. If so, does your choice of experimental layout allow for a meaningful interpretation of the results?

3.7. RANDOMIZATION

1. Have the treatments and controls been allocated to the plots of the experiment by an explicit randomizing procedure?
2. Was a separate randomization carried out for each block or row of the experiment?
3. Were the constraints on the randomization correctly applied?
4. Were you tempted to re-randomize any part of the allocation of treatments and controls to plots because of apparently unfortunate coincidences?
5. If so, do you have some knowledge of variation in the site or experimental material which has not been incorporated into the design of the experiment?
6. Does a plan exist, showing the allocation of the treatments and controls to the individual plots?

3.8. RECORDING OF RESULTS

1. Does each plot of the experiment have a clear number or designation, linking it unambiguously to the plan of the experiment?
2. Have you defined the time intervals at which assessments of the experimental results are to be made?
3. Have you defined the variables or attributes to be counted or measured at each assessment?
4. If so, are the measurements meaningful and relevant to the objectives of the experiment?
5. Are any of the assessments to be made from samples of the experimental plot rather than from the whole plot?
6. If so, has the efficiency of the sampling been tested?
7. Are any of the assessments to be used as covariates to correct for unavoidable but measurable differences between the plots?
8. If so, will these assessments need to be made before any of the experimental treatments are applied, or can take any effect?
9. Have you planned to use the blocks or rows of the experiment to absorb any unwanted variation in assessment, e.g. different observers, assessments on different days or at different times of the day?

10. Have you designed a record form which will ensure that all assessments are complete and are recorded against the correct plot?
11. Have the assessors been trained to measure and count the variables or attributes efficiently and accurately?
12. Is there space on the record forms for observations to be recorded of unexpected changes or effects, and have the assessors been encouraged to look for these effects?

3.9. PLANNING FOR ANALYSIS

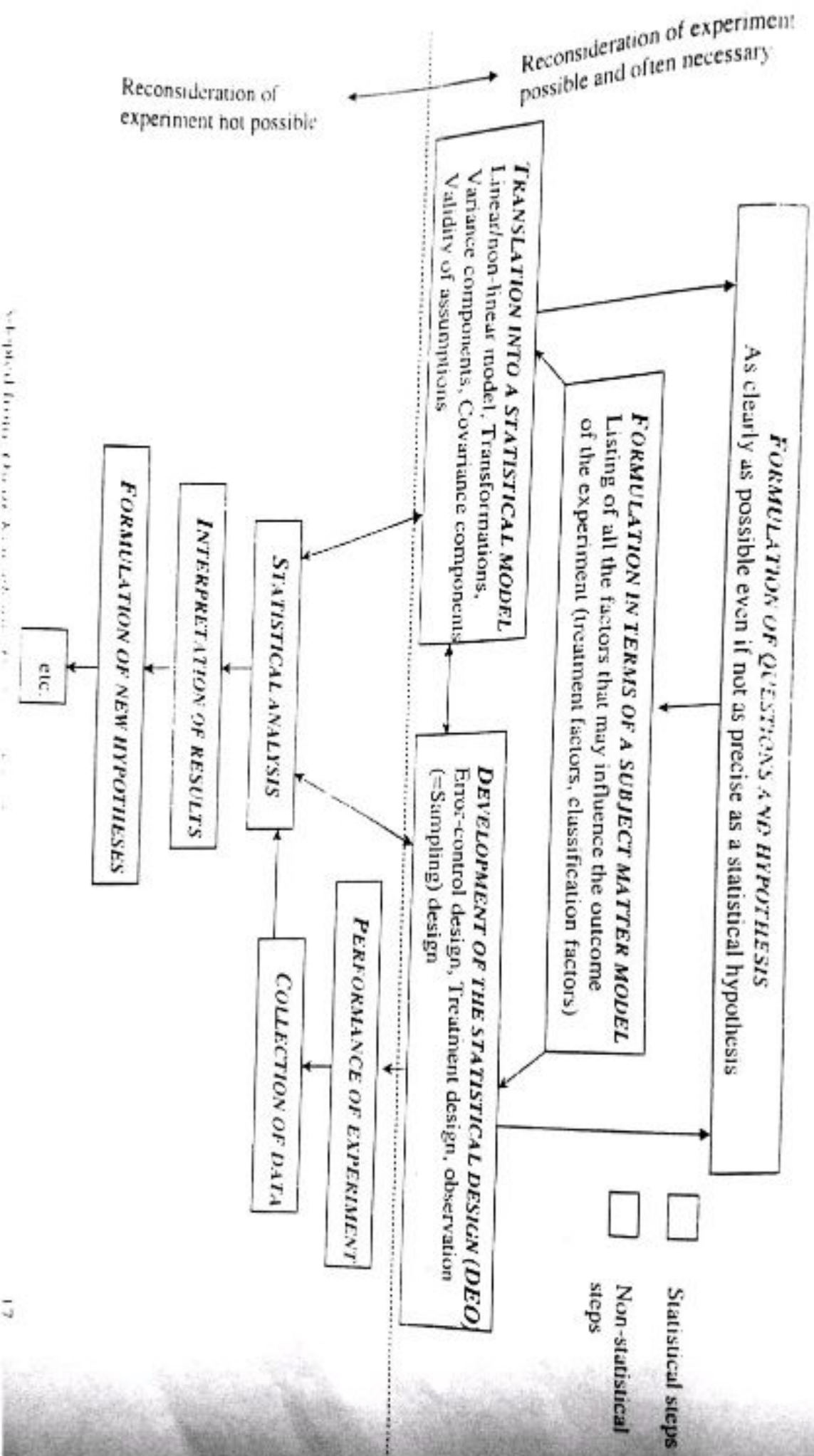
1. Have the hypotheses to be tested in the analysis of the results of the experiment, and their alternatives, been defined a priori?
2. Are these tests expressed, as far as possible, as null hypotheses?
3. Have any special contrasts to be tested or estimated in the analysis been defined in advance of a first inspection of the results of the experiment?
4. Do you understand the methods of analysis that will need to be used for this experiment and made arrangements for the computations to be done on a computer, or elsewhere?
5. If not, have you obtained advice from a qualified statistician on the analysis and interpretation of the results, preferably before starting on the experiment?

3.10. THE FINAL (AND MOST IMPORTANT) QUESTION

1. If you are in doubt about the purpose of any of the questions in this checklist, should you not obtain some advice from a statistician with experience of your field of research before continuing with the experiment?

There is usually little that a statistician can do to help you once you have committed yourself to a particular experimental design.

LOGICAL STEPS OF SCIENTIFIC EXPERIMENTATION



4. COMPLETELY RANDOMIZED DESIGN

4.1. INTRODUCTION

Completely Randomized Design is a design in which the treatments are assigned to the experimental units completely at random, that is the randomization is done without any restriction. The design is completely flexible, i.e., any number of treatments and of replications may be used. Moreover, the number of replications for any treatment needs not to be equal. A COMPLETELY RANDOMIZED DESIGN is considered to be most useful in situations where:

- The experimental units/material is homogeneous
- The experiments are small such as laboratory experiments
- Some experimental units are likely to be destroyed or fail to respond

4.1.1. ADVANTAGES OF COMPLETELY RANDOMIZED DESIGN

- The design is completely flexible, i.e., any number of treatments and of replications may be used. Moreover, the number of replications for any treatment needs not to be equal.
- It gives maximum degrees of freedom for error sum of squares as compared with the other designs for the same situation.
- The design is very simple and is easily laid out.
- The statistical analysis is very simple both for equal and unequal number of replications.
- If the data from some experimental units is missing, it does not complicate the analysis. The missing observations can be discarded without affecting the results of the experiment and efficiency of this design is not severely affected.

4.1.2. DISADVANTAGES OF COMPLETELY RANDOMIZED DESIGN

- The design is applicable only to a small number of treatments.
- The design is applicable only to homogeneous experimental material.
- If the experimental units are not homogeneous, then the use of this design gives the large experimental error as compared to some other designs, which use the

homogeneous experimental units as blocks and ultimately reduce the experimental error.

- There is possibility of entering the whole of the variation among the experimental units into the experimental error, as the randomization is not restricted in any direction.

4.2. INTERPRETATION OF THE RESULTS OF ANOVA

4.2.1. INSIGNIFICANT F -RATIO FOR TREATMENTS

Insignificant F test for treatment indicate the failure of the experiment to detect any difference among treatments. It does not, in any way, prove that all the treatment are the same, because the failure to detect treatment differences, based on the insignificant F test, could be the result of either a very small or no treatment differences or a very large experimental error, or both. Thus, whenever the F test is insignificant, the researcher should examine the size of the experimental error and the numerical differences among treatment means. If both values are large, the trial may be repeated and the efforts should be made to reduce the experimental error so that the differences among treatments, if any, can be detected. On the other hand, if both values are small, the differences among treatments are probably too small to be of any economic value and thus no additional trials are needed.

4.2.2. F -RATIOS THAT ARE LESS THAN UNITY

An F -value less than one is a valuable warning sign. Suppose that F -ratio with V_1 and V_2 degrees of freedom is so small that $F' = \frac{1}{F}$, with V_2 and V_1 degrees of freedom is significant

then it seems reasonable to reject the postulated statistical model.

If the statistical model is rejected because of significant F' value then the following steps should be taken:

1. The experimental procedure should be reviewed to see whether the assumptions of ANOVA are satisfied. For example, if the proper randomization was not employed, the validity of the independence assumption is doubtful.
2. If sufficient observations are available, then check the assumption of normality by plotting the data on normal probability paper or use computer software to draw

the normal plot and in the event of rejection of the hypothesis of homogeneity of variances look for an appropriate one on the data or response variable.

3. Check the assumption of homogeneity of variances by using Bartlett test
4. The underlying phenomenon should be restudied to see if the assumed linear model is a good approximation to real world situation. If the assumed model is rejected then develop a new model that better describes the observed data under investigation.

✓ 4.2.3. COEFFICIENT OF VARIATION

$$C.V. = \frac{\sqrt{MSE}}{\text{Grand Mean}} \times 100$$

The C.V indicates the degree of precision with which the treatments are compared and is a good index of the reliability of the experiment. The CV expresses the amount of experimental error relative to the size of the data. The small value of CV indicates that the results of the experiment are reliable.

The CV varies greatly with the nature of the experiment and the character measured, so there is no hard and fast rule for the acceptability of a particular CV value. The experiment is considered as reliable for any value of CV range from 3% to 20% (depending upon the nature of the experiment). For the laboratory experiments where experimental material is quite homogenous, less than 10% is an acceptable value of CV.

4.3. INFERENCE AFTER ANOVA

Def When F_{cal} for treatments is significant, it means that there is evidence that at least two means differ from each other. We can test the significance of difference between treatment means by constructing confidence intervals for difference between means. (If confidence interval contains 0 value then the two means are not significant i.e., no difference between two means). Alternatively, different methods have been developed for locating significant differences among treatment means after completing analysis of variance.

4.3.1. MULTIPLE COMPARISON TESTS

There are two types of pair comparisons

- **Unplanned Pair Comparisons** (Comparisons suggested by the data or arguing after the facts also called data snooping) in which no specific comparison is chosen in advance. Instead, every possible pair of treatment means is compared to identify pairs of treatments that are significantly different. Duncan's Multiple Range test (DMRT) is applicable to unplanned pair comparisons, and is very commonly used by biologists.
- **Planned Pair Comparisons** In which the specific pair of treatment was identified before the start of the experiment. The Least Significant Difference test (LSD) is the most appropriate test when the comparisons are meaningfully planned in terms of nature of the treatments. The planned comparison does not require the significance of the F test for treatments. Least significant Increase (LSI) is a valid test procedure for comparing different treatments with control.

LSD is not a valid test procedure for comparing all possible pairs of treatment means (unplanned comparisons), especially when the number of treatment is large. For the number of treatments less than 6 both LSD and DMRT can be used appropriately for unplanned comparisons provided that the treatments are not of quantitative nature. When treatments are of quantitative nature then fitting a suitable function is surely more appropriate method of analysis for describing the data.

Some other commonly used test procedures are:

- Least Tuckey's Honestly significant difference (HSD)
- Scheffe's test
- Student-Newman-Keuls' multiple range test
- Dunnett's test.

4.3.2. CONTRASTS

In planned experiments treatments are usually selected to answer specific questions posed by the objects of the experimenters. Single degree of freedom comparisons are the most informative method of extracting the information from a set of data beyond the analysis of variance when treatments have some specific structure. It is not necessary to perform over all test of significance prior to carrying out the single degree of freedom comparisons.

A **contrast** may be defined as a linear combination of treatments total or means such that

$$C_k = \sum_{i=1}^t C_{ki} Y_i \quad \text{or} \quad \sum_{i=1}^t C_{ki} \mu_i \quad \text{with the restriction that} \quad \sum_{i=1}^t n_i C_{ki} = 0$$

The significance of a single degree of freedom contrast may be tested by using t or F test.

Two contrasts among t treatments means are said to be orthogonal to each other if they utilize non-overlapping pieces of information from an experiment. In an experiment of t treatments there may be so many contrasts, but only $(t-1)$ contrasts will be orthogonal. Two contrasts are

said to be orthogonal if $\sum_{i=1}^t n_i C_{ki} C_{ji} = 0$

For orthogonal contrasts the following condition must hold

$$\text{Treatment SS} = \text{Contrast}_1 \text{SS} + \text{Contrast}_2 \text{SS} + \dots + \text{Contrast}_{t-1} \text{SS}$$

4.3.3. RESPONSE CURVES: WHEN TREATMENTS ARE DIFFERENT LEVELS OF QUANTITATIVE FACTOR

The technique of regression analysis can be used to make more complete and informative analyses of data in which the treatments are different levels of a single quantitative factor. If the treatments are of quantitative nature in such situations it is more sensible to investigate how the measured characteristic varies with the change in the level of the treatment. That is, we wish to gain some idea of the shape of the response curve so that an estimate may be made of the optimum level of the treatment. To see the relationship plot the treatment means against the levels of the quantitative factor, take some idea about the general shape of the response curve and then undertake a more rigorous analysis of the data.

One can use polynomial regression to develop a response curve for the data and if the treatments are equally spaced the target can be achieved by using orthogonal polynomials.

4.4. PRESENTATION OF RESULTS

Although ANOVA table is a convenient way of summarizing certain aspects of the analysis of the data but it tends to overemphasize tests of hypotheses and underemphasize estimation. For the better understanding of the results of the research work should:

- Always accompany an ANOVA table with tables of means, together with their standard errors

- Whenever possible, portray the results in graphical form as it is more comparable than looking at the numbers

4.5. THE RELATION BETWEEN CRD AND *t*-TEST

In a CRD involving only two treatments, to test the hypothesis $T_1 - T_2 = 0$ by using ANOVA is equivalent to test $\mu_1 = \mu_2$, where $\mu_1 = \mu + T_1$ and $\mu_2 = \mu + T_2$ by using *t* test.

✓ EXAMPLE:

The following table shows some of the results of an experiment on the effects of applications of sulphur in reducing scab disease of potatoes. The object in applying sulphur is to increase the acidity of the soil, since scab does not thrive in very acid soil. In addition to untreated plots which serve as a control, 3 amounts of dressing were compared—300, 600, and 900 lb. per acre. Both a fall and a spring application of each amount was tested, so that in all there were seven distinct treatments. The sulphur was spread by hand on the surface of the soil, and then diced into a depth of about 4 inches. The quantity to be analyzed is the "scab index". That is roughly speaking, the percentage of the surface area of the potato that is infected with scab. It is obtained by examining 100 potatoes at random from each plot, grading each potato on a scale from 0 to 100% infected, and taking the average.

FIELD PLAN AND SCAB INDICES FOR A COMPLETELY RANDOMIZED EXPERIMENT ON POTATOES							
F3	O	S6	F9	S6	S9	S3	F6
9	12	18	10	24	17	30	16
O	S3	F9	F6	S3	O	O	S6
10	7	4	10	21	24	29	12
F3	S9	F6	O	F6	S9	F3	F9
9	7	18	30	18	16	16	4
S3	O	S9	S6	O	F9	O	F3
9	18	17	19	32	5	26	4

Notation: F = Fall, S = Spring application, O = Control. The numbers 3, 6, and 9 are the amounts of sulphur in 100 lb. per acre.

Analyze the experiment and write a statistical description of the analyzed data highlighting the main finding.

RESULTS GROUPED BY TREATMENTS								
	O	F3	S3	F6	S6	F9	S9	
	12	30	9	30	16	18	10	17
	10	18	9	7	10	24	4	7
	24	32	16	21	18	12	4	16
	29	26	4	9	18	19	5	17
TOTAL	181	38	67	62	73	23	57	G = 501
MEAN	22.6	9.5	16.8	15.5	18.2	5.8	14.2	

Analysis:

Correction Factor = (Grand Total)² / Total no of observations

$$= (181+38+67+...+23+57)^2/32 = (501)^2/32 = 7843.781$$

$$\text{Total SS} = (12)^2 + (10)^2 + (24)^2 + ... + (16)^2 + (17)^2 - CF = 9936 - 7843.781 = 2095.219$$

$$\begin{aligned} \text{Treatment SS} &= \frac{(181)^2}{8} + \frac{(38)^2}{4} + \frac{(67)^2}{4} + \frac{(62)^2}{4} + \frac{(73)^2}{4} + \frac{(23)^2}{4} + \frac{(57)^2}{4} - CF \\ &= 8816.125 - 7843.781 = 972.345 \end{aligned}$$

$$\text{Error SS} = \text{Total SS} - \text{Treatment SS} = 2095.219 - 972.345 = 1122.874$$

$n-t$

ANOVA TABLE				
SOV	SS	DF	MSS	FCAL
TREATMENT	972.345	7-1 = 6	162.0575	3.61*
ERROR	1122.87	$n-t$ 32-7 = 25	44.9148	
TOTAL	2095.219	32-1=31		

$n-t$ $32-7=25$

*Significant at 5%.

$$C.V = \frac{\sqrt{MSE}}{\text{Grand Mean}} \times 100 = \frac{\sqrt{44.9148}}{15.6563} \times 100 = 42.82\%$$

STATISTICAL REPORT

The F-ratio for treatments is highly significant at the 1% level. From the treatment means it appears that all dressings had some beneficial effect, and that the fall application was more effective than the spring one (because the means for fall applications are less than that of spring application). There is little or no evidence that the higher dressings were more effective than the lowest dressing.

If the summary is conducted from this point of view, we should isolate and test two individual components of the treatments SS: (i) a component measuring the average effect of all dressings, (ii) a component comparing the fall and spring applications. The following computations are required to do the same

Average effect of sulphur: The total over all dressings is 320, representing 24 plots. Since the control total, 181, represents 8 plots, the comparison is $3(181) - 320 = 223$. The contribution to the sum of squares in the analysis of variance is $\frac{(223)^2}{96}$, or 518.

Fall versus spring application: The comparison is $(38+62+23-67-73-57) = -74$. It is obviously orthogonal to the previous comparison. The contribution to the sums of squares is $\frac{(74)^2}{24} = 228.2$.

The remaining four components of the treatments SS must represent comparisons among the levels of sulphur.

$$H: 3\mu_1 = \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6 + \mu_7$$

$$C1 = 3Y_1 - Y_2 - Y_3 - Y_4 - Y_5 - Y_6 - Y_7 = 3(181) - 38 - 67 - 62 - 73 - 23 - 57 = 223$$

We are attempting to compare the mean of 8 observations with the mean of 24 observations ($4+4+4+4+4$). It is necessary to adjust for the spurious weighting given by our comparison of treatment totals based on unequal numbers of observations. Since the smallest integer that may be divided evenly by both 8 and 24 is 24 [LCM(8, 24) = 24], we see that 3 ($24/8$) and 1 ($24/24$) are the indicated weights to be used if our comparison is to be unaffected by the differing numbers of observations associated with the various treatments.

$$H: \mu_2 + \mu_4 + \mu_6 = \mu_3 + \mu_5 + \mu_7$$

$$C2 = Y_2 + Y_4 + Y_6 - Y_3 - Y_5 - Y_7 = 38 + 62 + 23 - 67 - 73 - 57 = -74$$

CONTRAST	O	F3	S3	F6	S6	F12	S12	C	$\sum w_i c_i^2$	SSCj	VAR(Cj) = $MSE \times \sum w_i^2$
	181	38	67	62	73	23	57				
C1	3	-1	-1	-1	-1	-1	-1	223	$8 \times (3)^2 + 4 \times (-1)^2 + 4 \times (-1)^2 + 4 \times (-1)^2 + 4 \times (-1)^2 + 4 \times (-1)^2 + 4 \times (-1)^2$ = 96	518.0104	$44.9 \times 96 = 4310.4$
C2	0	1	-1	1	-1	1	-1	-74	$6 \times 4 = 24$	228.1667	$44.9 \times 24 = 1077.6$

$$\begin{aligned} \text{Control vs Sulphur SS} &= \frac{(181)^2}{8} + \frac{(320)^2}{24} - \frac{(501)^2}{32} \\ &= 8361.79167 - 7843.78125 = 518.01 \end{aligned}$$

$$\begin{aligned} \text{Fall vs Spring SS} &= \frac{(123)^2}{12} + \frac{(197)^2}{12} - \frac{(320)^2}{24} \\ &= 4494.83 - 4266.67 = 228.163 \end{aligned}$$

$$\begin{aligned} \text{Among Fall application SS} &= \frac{(38)^2}{4} + \frac{(62)^2}{4} + \frac{(23)^2}{4} - \frac{(123)^2}{12} \\ &= 1454.25 - 1260.75 = 193.5 \end{aligned}$$

$$\text{Among spring application SS} = \frac{(67)^2}{4} + \frac{(73)^2}{4} + \frac{(57)^2}{4} - \frac{(197)^2}{12}$$

$$= 3266.75 - 3234.083 = 32.67$$

ORTHOGONAL POLYNOMIAL CONTRASTS FOR FALL APPLICATION

	F3	F6	F12	$C_i = \sum C_{ij} Y_j$	$\sum C_{ij}^2$	SS_{C_i}
CONTRASTS	38	62	23			
LINEAR	-1	0	1	-15	2	28.125
QUADRATIC	1	-2	1	-63	6	165.375
TOTAL						193.500

ORTHOGONAL POLYNOMIAL CONTRASTS FOR SPRING APPLICATION

	S3	S6	S12	$C_i = \sum C_{ij} Y_j$	$\sum C_{ij}^2$	SS_{C_i}
CONTRASTS	67	73	57			
LINEAR	-1	0	1	-10	2	12.500
QUADRATIC	1	-2	1	-22	6	20.167
TOTAL						32.667

ANOVA TABLE

SOV	SS	DF	MSS	F _{CAL}
TREATMENT	972.345	6	162.058	3.61*
CONTROL VS. SULPHUR	518.010	1	518.010	11.54**
FALL VS. SPRING APPLICATION	228.167	1	228.167	5.08*
AMONG FALL APPLICATION	193.500	2	96.750	2.15
LINEAR	28.125	1	28.125	0.63
QUADRATIC	165.375	1	165.375	3.68
AMONG SPRING APPLICATION	32.667	2	16.335	0.36
LINEAR	12.500	1	12.500	0.28
QUADRATIC	20.167	1	20.167	0.45
ERROR	1122.87	25	44.915	
TOTAL	2095.219	31		

The average reduction in scab due to sulphur is significant at the 1% level, while the superiority of the fall application is also significant. Differences among the levels show no sign of significance.

The conclusions might be phrased as "The application of sulphur produced a significant decrease in the scab index, the average being 22.6 for the untreated plots, 16.4 $\{=(16.8+18.2+14.2)/3\}$ for plots with the spring application, and 10.2 $\{=(9.5+15.5+5.8)/3\}$ for plots with the fall application. The fall application proved significantly better than the spring application. There was no indication that the higher levels of dressing were more effective than the lowest level."

5. RANDOMIZED COMPLETE BLOCK DESIGN

5.1. INTRODUCTION

The completely randomized design is used when the experimental material is homogenous; but there are situations when there is relatively large variability in the experimental material and it is possible to make blocks in such a manner that the experimental units within a particular block are relatively homogenous. The design applied in such situation is randomized complete block design (RCBD). The treatments are assigned at random to the experimental units within each block, which means the randomization is restricted within blocks. The main objective of blocking is to reduce the variability among experimental units within a block as much as possible and to maximize the variation among blocks. Randomized complete block designs are commonly used in agriculture experiments when the fertility and other soil factors show variability across field. The blocks are made perpendicular to the variation present in the experimental material. If there are no differences among blocks, then RCBD would not contribute to improve precision in detecting differences among treatment means. The word "Complete" here indicates that each block contains the complete set of treatments.

5.1.1. ADVANTAGES OF RANDOMIZED COMPLETE BLOCK DESIGN

- The design is flexible, i.e., any number of treatments and of replication (but not less than 2) may be used.
- The statistical analysis is fairly simple when there are no missing observations.
- It is easy to adjust for missing observations.
- Grouping the experimental material controls the source of extraneous variation and hence the estimate of the experimental error is decreased.
- When the variability among blocks is large then precision increases because the sum of squares for blocks is extracted from the sum of squares for experimental error leading to smaller mean square error (MSE).
- Placing blocks under different conditions increases the scope of the experiment.
- It provides unbiased estimates of the means of the blocking factor.
- A Randomized Complete Block Design is generally more efficient than Complete Randomized Design subject to proper block orientation.

5.1.2. DISADVANTAGES OF RANDOMIZED COMPLETE BLOCK DESIGN

- It controls variability only in one direction
- It is not a suitable design when the number of treatment is very large or when the blocks are not homogeneous.
- With the increase in block size, within block variability increases which is not the aim of this design because our purpose is to reduce variability within the blocks and to maximize the among block variability.
- Missing observations in the data cause problem in the analysis. One or two missing observations can be handled easily but more missing observations affect the efficiency of the design.
- The degrees of freedom for error is less for Randomized Complete Block Design because of the formation of blocks as compared with the Completely Randomized Design. This loss can be dangerous when blocks effect is non-significant.

Example 4.1

The following table comes from an experiment in which four seed treatment were compared with no treatment (check) on soybean seeds. The data are the number of plants that failed to emerge out of 100 planted in each plot. The row criteria of classification are treatment and replications (blocks).

No of failures out of 100 planted Soybean seeds							
Treatments	Rep. (Block)					Total	Mean
	1	2	3	4	5		
Check	8	10	12	13	11	54	10.8
Arasan	2	6	7	11	5	31	6.2
Spargon	4	10	9	8	10	41	8.2
Swmesem. Jr.	3	5	9	10	6	33	6.6
Fermate	9	7	5	5	3	29	5.8
Total	26	38	42	47	35	188	7.52

Analysis:

$$\text{Correction Factor} = (\text{Grand Total})^2 / \text{Total no of observations}$$

$$= (188)^2 / 25 = 1413.76$$

$$\text{Total SS} = (8)^2 + (2)^2 + (4)^2 + \dots + (6)^2 + (3)^2 - \text{CF} = 1634 - 1413.76 = 220.24$$

$$\text{Block SS} = \frac{(26)^2}{5} + \frac{(38)^2}{5} + \frac{(42)^2}{5} + \frac{(47)^2}{5} + \frac{(35)^2}{5} - CF = 1463.6 - 1413.76 = 49.84$$

$$\text{Treatment SS} = \frac{(54)^2}{5} + \frac{(31)^2}{5} + \frac{(41)^2}{5} + \frac{(33)^2}{5} + \frac{(29)^2}{5} - CF = 1497.6 - 1413.76 = 83.84$$

$$\text{Error SS} = \text{Total SS} - \text{Block SS} - \text{Treatment SS} = 86.56$$

ANOVA TABLE

SOV	SS	DF	MSS	F _{cal}
Block	49.84	(b-1) = 4	12.46	
Treatment	83.84	(t-1) = 4	20.96	3.87*
Error	86.56	(b-1)(t-1) = 16	5.41	
TOTAL	220.24	(bnt-1) = 24		

$$CV = \frac{\sqrt{MSE}}{\text{Mean}} = \frac{\sqrt{5.41}}{7.52} = 30.93\%$$

STATISTICAL REPORT

Since *F*-ratio for treatments is significant so we conclude that germination of soybean seed is different for different treatments. A high value of *CV* indicates that there are other unknown factors that are responsible for variation in the data that needs to be controlled for improvement in the precision of the experiment.

Standard Errors:

Standard Error of a treatment mean

$$= \sqrt{\frac{MSE}{b}} = \sqrt{\frac{5.41}{5}} = 1.040$$

Standard Error of the difference between two treatment means = $\sqrt{\frac{2MSE}{b}} = \sqrt{\frac{2 \times 5.41}{5}} = 1.471$

As the treatments have some specific structure so we can perform more detail analysis to extract useful information beyond ANOVA table.

$$\text{Chemical Vs Check SS} = \frac{(134)^2}{20} + \frac{(54)^2}{5} - \frac{(188)^2}{25} = 67.24$$

$$\text{Among Chemicals SS} = \frac{(31)^2}{5} + \frac{(41)^2}{5} + \frac{(33)^2}{5} + \frac{(29)^2}{5} - \frac{(134)^2}{20} = 16.60$$

ANOVA TABLE

SOV	SS	DF	MSS	F _{cal}
Block	49.84	5-1 = 4	12.46	
Treatment	83.84	5-1 = 4	20.96	3.874*
Chemical Vs Check	67.24	2-1=1	67.24	12.429**
Among Chemicals	16.60	4-1=3	5.53	1.023
Error	86.56	24-8 = 16	5.41	
TOTAL	220.24	25-1=24		

STATISTICAL REPORT

The application of chemical treatments has significantly reduced the number of seed germination failure as compared to without application of chemical treatment. However all chemical treatments seem to have similar affect on the seed germination failures.

Example 4.2

In an experiment to compare the effect of four drugs A, B, C and D on the lymphocyte counts in mice a randomized block design with four mice from each of five litters was used, the litter being regarded as blocks. The lymphocyte counts (thousand per mm³ of blood) are given in the table below:

Litters(Blocks)		1	2	3	4	5	TOTAL	MEANS
Drugs	A	7.1	6.1	6.9	5.6	6.4	32.1	6.42
	B	6.7	5.1	5.9	5.1	5.8	28.6	5.72
	C	7.1	5.8	6.2	5.0	6.2	30.3	6.06
	D	6.7	5.4	5.7	5.2	5.3	28.3	5.66
TOTAL		27.6	22.4	24.7	20.9	23.7	119.3	5.96

ANALYSIS

Correction Factor = (Grand Total)² / Total no of observations = (119.3)² / 20 = 711.62

Total SS = (7.1)² + (6.7)² + ... + (5.3)² - CF = 720.51 - 711.62 = 8.89

Treatment SS = $\frac{(32.1)^2}{5} + \frac{(28.6)^2}{5} + \frac{(30.3)^2}{5} + \frac{(28.3)^2}{5} - CF = 713.47 - 711.62 = 1.85$

Block SS = $\frac{(27.6)^2}{4} + \frac{(22.4)^2}{4} + \frac{(24.7)^2}{4} - CF = 718.02 - 711.62 = 6.40$

Error SS = Total SS - Treatment SS - Block SS = 8.89 - 6.40 - 1.85 = 0.64

ANOVA TABLE

SOV	SS	DF	MSS	F _{cal}
BLOCK	6.40	4	1.600	30.19*
TREATMENT	1.85	3	0.616	11.59*
ERROR	0.64	12	0.053	
TOTAL	8.89	19		

$$C.V. = \frac{\sqrt{MSE}}{\text{Grand Mean}} = \frac{\sqrt{0.053}}{5.96} = 3.86\%$$

STATISTICAL REPORT

The four drugs have different effects on the lymphocyte counts in mice. As the value of CV is not very high so the results of the experiment are reliable. Using litters as blocks was quite effective in reducing the size of experimental error.

STANDARD ERRORS:

Standard Error of a treatment mean

$$= \sqrt{\frac{MSE}{b}} = \sqrt{\frac{0.053}{5}} = 0.103$$

Standard Error of the difference between two treatment means = $\sqrt{\frac{2MSE}{b}} = \sqrt{\frac{2 \times 0.053}{5}} = 0.146$

As the treatments are unstructured so we can apply multiple comparison tests to identify significantly different treatment pairs.

DMR TEST FOR ALL PAIR OF MEANS (UNPLANNED COMPARISONS)

STEP I:- Arrange all treatment means in ascending order

TREATMENTS	D	B	C	A
MEANS	5.66	5.72	6.06	6.42

STEP II:- Calculate DMR value by using the formula

$$DMR = r_{\alpha, p(Edf)} \left(\sqrt{\frac{MSE}{b}} \right)$$

Where $r_{\alpha, p(Edf)}$ Duncan's multiple range table values

MSE=Mean Square for Error

b= No of Blocks

P	2	3	4
$r_{\alpha, p(Edf)}$	3.08	3.23	3.33
DMR	0.318	0.333	0.343

STEP III:- Make Comparisons among all possible pairs of treatments means as

COMPARISON	DIFFERENCE BETWEEN MEANS	DMR VALUE	RESULT
A vs D	0.76	0.343	Significant
A vs B	0.70	0.333	Significant
A vs C	0.36	0.318	Significant
C vs D	0.40	0.333	Significant
C vs B	0.34	0.318	Significant
B vs D	0.06	0.318	Not-significant

D	B	C	A
5.66	5.72	6.06	6.42
c	c	b	a

Treatment means sharing common letter are insignificant.

5.2. RELATIVE EFFICIENCY OF RCBD AS COMPARED TO CRD

The primary purpose of blocking is to reduce experimental error by eliminating the contribution of known sources of variation among experimental units. One measure of the

effectiveness of blocking is the *F*-Ratio for blocks. Blocking is considered effective in reducing experimental error if *F*-Ratio for block is significant. An insignificant *F*-Ratio for blocks reveals that there is no difference among blocks, i.e., the blocks are reasonably homogenous as among block variation was not significant. On the other hand significant *F*-Ratio for blocks indicates that blocking has been effective in reducing experimental error and hence blocking was useful for increasing the efficiency of RCBD as compared to CRD.

A more precise measure of the efficiency of the RCBD relative to the CRD for a given situation is obtained by computing the relative efficiency (RE) as

$$RE = \frac{(b-1)MSB + b(t-1)MSE}{(bt-1)MSE}$$

Where *b*, *t* = Number of blocks and treatments in RCBD

MSB, MSE = Block and Error mean sum of squares from ANOVA of RCBD

If $RE > 1$, then the RCBD is more efficient than would have been a CRD in the same situation. The quantity $(RE-1) \times 100$ is the percent increase in efficiency resulting from the use of RCBD. The quantity $bx(RE)$ is the number of replication in CRD that would have been required to obtain the same precision as that obtained by using RCBD with '*b*' blocks.

For example 4.2, $RE = \frac{(b-1)MSB + b(t-1)MSE}{(bt-1)MSE} = \frac{(5-1)1.6 + 5(4-1)0.053}{(5 \times 4 - 1)0.053} = 7.14$

As $RE > 1$, so RCBD is more efficient than CRD in the same situation. The number of replication in CRD that would have been required to obtain the same precision as that obtained by using RCBD by using 5 blocks only is equal to $(5)(7.14) = 36$

5.3. MISSING OBSERVATION IN RCBD

During the conduct of experiment, it is possible that one or more observations are destroyed. This may happen when an animal become sick or dies but not as a result of the treatment, when rodents destroy a plot in a field trial or when a flask breaks in the laboratory. Occurrence of missing data results in two major difficulties—loss of information and non-applicability of the standard analysis of variance technique as treatments are no longer orthogonal to blocks as every treatment does not occur in every block as a result each block is not a complete block. In this situation, the missing observation is estimated by using

appropriate formula according to the experimental design used and the usual analysis of variance, with some slight modifications, is performed just as if the estimated observation was real data.

If there is only one missing observation then estimate the missing observation by using the formula.

$$\square \text{ STEP 1} \quad Y_m = \frac{bB + tT - G^*}{(b-1)(t-1)}$$

Where Y_m = Estimate of missing observation.

B, T = Sum of observed values of block and treatment having a missing observation

b, t = Number of blocks and treatments

G^* = Sum of observed values, i.e., sum of all the values except missing value

□ STEP 2 Subtract one from both the total and error degree of freedom

$$\square \text{ STEP 3} \quad \text{Calculate the amount of bias as: Bias} = \frac{[B - (t-1)Y]^2}{t(t-1)}$$

and then subtract this amount of bias from both Treatment SS and Total SS.

For more than one missing observation use iterative procedure.

Example

Consider in example 4.2, the mouse in litter 1 who received the drug A has died and data is not available for final analysis. Estimate the response for this mouse and then perform ANOVA.

Litters(Blocks)								
DRUGS		1	2	3	4	5	Totals	Corrected Total
	A	Y (7.53)	6.1	6.9	5.6	6.4	25.0	32.53
	B	6.7	5.1	5.9	5.1	5.8	28.6	28.6
	C	7.1	5.8	6.2	5.0	6.2	30.3	30.3
	D	6.7	5.4	5.7	5.2	5.3	28.3	28.3
TOTAL		20.5	22.4	24.7	20.9	23.7	112.2	
CORRECTED TOTAL		28.03	22.4	24.7	20.9	23.7		119.73

$$\hat{y} = \frac{5(20.5) + 4(25) - 112.2}{(5-1)(4-1)} = 7.53$$

Calculation for amount of bias

$$\text{Bias} = \frac{[20.5 - (4 - 1)7.53]^2}{4(4 - 1)} = 0.36$$

Adjusted ANOVA TABLE

SOV	DF	SS	MSS	Fcal
Treatment	4-1=3	2.26-0.36=1.89	0.632	13.16*
Block	5-1=4	7.25	1.810	
Error	12-1=11	0.53	0.048	
TOTAL	20-1-1=18	10.04-0.364=9.676		

Statistical software MSTATC can be used to estimate the missing value(s) in the data. The analysis can be completed by putting estimated value in the data. However if analysis is done by first estimating the missing value and then by putting the value in the data and then completing the analysis may result in an inappropriate ANOVA as adjustment of df for error is not accomplished by the program. If program is asked to estimate the missing value and complete the analysis in one go then data is properly analyzed and conclusions can be drawn from such analysis directly.

6. LATIN SQUARE DESIGN

6.1. INTRODUCTION

When the gradient of variation is found in two directions it becomes necessary to take account of both these two sources of variation simultaneously. This can be achieved by simultaneous blocking of experimental units in two mutually perpendicular directions, called rows and columns. The design used in these situations is called Latin square design. Since each row and each column is a complete block, so each treatment must appear once and only once in each row and each column. *control 2 way of variation*

In Latin square design, the number of treatments is equal to the number of rows and to the number of columns. The objective is to eliminate the variability due to rows and columns from the experimental error. It is advocated that these designs are useful when the number of treatments is between 5 and 10. When the number of treatments increases it becomes laborious to use this design.

6.1.1. ADVANTAGES OF LATIN SQUARE DESIGN

✓ A Latin square design reduces the error variance by controlling two source of nuisance variation. *N-S and E-W*

✓ The analysis without missing values is straightforward.

✓ A Latin square design is generally more efficient than a Randomized Complete Block Design. *RCBD*

Design.

✓ Each row and column is a complete replication.

The analysis of a LSD is simple & remains relatively simple with missing observations

6.1.2. DISADVANTAGES OF LATIN SQUARE DESIGN

✓ A Latin square design is less flexible than a Randomized Complete Block Design. It is *practical* practicable only for 5 to 10 treatments. When the number of treatments exceeds 10, the design is seldom used.

✓ For a small number of treatments, a Latin Square design does not provide a sufficient number of replicates to give a valid estimate of error as two few df are left to estimate error variance.

✓ When number of treatments is less than 4, the degrees of freedom for error is rather small thus validity of experimental error becomes questionable.

Replication in Latin Square design is costly.

In agricultural experiments, the land requirement is rigid, the actual layout in the field may be laborious and approach to the central most plots becomes difficult.

If there are missing observations in the experiment then the analysis becomes complicated.

Example

The following table shows the field layout and yield of a 5×5 Latin square experiment on the effect of spacing on yield of millet plants. ^{treatment} Five levels of spacing were used. The data on yield (grams/plot) was recorded and is given below.

Yields(G) of plots of millet arranged in a LS (Spacing in inches A. 2, B.4; C.6; D. 8; F.10)						
	Column					Total
Row	1	2	3	4	5	
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: 240	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: 338	1440
Total	1118	1240	1297	1340	1309	6304

Treatments						
	A	B	C	D	E	Total
Total	1349	1314	1262	1191	1188	6304
Mean	269.8	262.8	252.8	238.2	237.6	252.2

ANALYSIS

$$\text{Correction Factor} = (\text{Grand Total})^2 / \text{Total no of observations}$$

$$= (6304)^2 / 25 = 1589616.64$$

$$\text{Total SS} = (257)^2 + (245)^2 + \dots + (338)^2 - \text{CF} = 1626188 - 1589616.64 = 36571.36$$

$$\text{Row SS} = \frac{(1255)^2}{5} + \frac{(1313)^2}{5} + \frac{(1210)^2}{5} + \frac{(1086)^2}{5} + \frac{(1440)^2}{5} - \text{CF}$$

$$= 1603218 - 1589616.64 = 13601.36$$

$$\text{Column SS} = \frac{(1118)^2}{5} + \frac{(1240)^2}{5} + \frac{(1297)^2}{5} + \frac{(1340)^2}{5} + \frac{(1309)^2}{5} - \text{CF}$$

$$= 1595762.8 - 1589616.64 = 6146.16$$

$$\text{Spacing SS} = \frac{(1349)^2}{5} + \frac{(1314)^2}{5} + \frac{(1262)^2}{5} + \frac{(1191)^2}{5} + \frac{(1188)^2}{5} - CF$$

$$= 1593773.2 - 1589616.64 = 4156.56$$

$$\text{Error SS} = \text{Total SS} - \text{Rows SS} - \text{Column SS} - \text{Spacing SS} = 12667.28$$

ANOVA TABLE

SOV	SS	DF	MSS	F_{cal}
Rows	13601.36	4	3400.34	
Columns	6146.16	4	1536.54	
Spacing	4156.56	4	1039.14	0.98
Error	12667.28	12	1055.61	
TOTAL	36571.36	24		

F_{tab}

3.259

$$C.V. = \frac{\sqrt{MSE}}{\text{Grand Mean}} \times 100 = \frac{\sqrt{1055.61}}{252.16} = 12.88\%$$

STATISTICAL REPORT

The analysis indicates that yield of millet plants is same for all the row spacing under investigation. It does not prove that the yield from different row spacing really are the same but only shows that no differences could be observed on the basis of statistical evidence obtained from the given data. It may be that either there are really no differences or there are small differences which are masked due to the small sample sizes. A large value of mean square for rows relative to error indicates that horizontal blocking is useful in reducing the experimental error and hence helpful in improving the precision of the experiment.

$$\sqrt{\frac{MSE}{t}}$$

Standard Errors.

$$\text{Standard Error of a treatment mean} = \sqrt{\frac{MSE}{t}} = \sqrt{\frac{1055.607}{5}} = 14.53$$

Among all treatment means

$$\text{Standard Error of the difference between two treatment means} = \sqrt{\frac{2MSE}{t}} = \sqrt{\frac{2 \times 1055.607}{5}} = 20.9$$

Feasibility for this experiment (field experiment)

6.2. RELATIVE EFFICIENCY OF LS DESIGN AS COMPARED TO RCBD

a) Considering Column as Blocks

$$RE_c = \frac{\text{Mean Square Error} + (t-1)MSE}{(t)MSE} = \frac{3400.34 + 4 \times 1055.607}{5 \times 1055.607} = 1.44$$

RE_c

$$RE_c = \frac{MSR + (t-1)MSE}{(t)MSE}$$

The result indicates that there is considerable gain in efficiency due to introduction of row variable (i.e., horizontal blocking).

b) Considering Row as Blocks

$$R^2 = \frac{MSC + (t-1)MSE}{(t+1)MSE}$$

$$RE = \frac{MSC + (t-1)MSE}{(t+1)MSE} = \frac{1536.54 + 4 \times 1055.607}{5 \times 1055.607} = 1.09$$

The result indicates that the introduction of column variable (i.e., vertical blocking) is not much effective in reducing experimental error. In above situation, a RCBD with row as blocks would have been as efficient as a LS design.

$$\frac{MSR + MSC + (t-1)MSE}{(t+1)MSE}$$

c) Relative efficiency of LS design as compared to CRD

$$RE = \frac{MSR + MSC + (t-1)MSE}{(t+1)MSE} = \frac{3400.34 + 1536.54 + 4 \times 1055.607}{6 \times 1055.607} = 1.45$$

As $RE > 1$, so LS design in present situation is estimated to increase the experimental precision over CRD by 45%.

6.3. MISSING OBSERVATION IN LATIN SQUARE DESIGN

If there is only one missing observation then estimate the missing observation by using the formula

□ STEP 1
$$Y_m = \frac{t(R + C + T) - 2G^*}{(t-1)(t-2)}$$

Where Y_m = Estimate of the missing observation.

R, C, T = Sum of observed values of row, column and treatment having a missing observation

t = Number of treatments

G^* = Sum of observed values i.e., sum of all the values except missing value

□ STEP 2 Subtract one from both the total and error degree of freedom

□ STEP 3 Calculate the amount of bias as:
$$\text{Bias} = \frac{[G^* - R - C - (t-1)T]^2}{[(t-1)(t-2)]^2}$$

and then subtract this amount of bias from both Treatment SS and Total SS.

Example

In above example suppose that observation for row spacing D in column 5 and row 1 was not available for final analysis. Estimate that observation and then perform the ANOVA

Row	Column					Total	Corrected Total
	1	2	3	4	5		
1	B: 257	E: 230	A: 279	C: 287	D: 295.08	1053	1348.08
2	D: 245	A: 283	E: 245	B: 280	C: 260	1313	1313
3	E: 182	B: 252	C: 280	D: 246	A: 250	1210	1210
4	A: 203	C: 204	D: 227	E: 193	B: 259	1086	1086
5	C: 231	D: 271	B: 266	A: 334	E: 338	1440	1440
Total	1118	1240	1297	1340	1107	6102	
Corrected Total	1118	1240	1297	1340	1402		6397

Treatments						
	A	B	C	D	E	Total
Total	1349	1314	1262	989	1188	6102
Corrected Total	1349	1314	1262	1284	1188	6397

$$\hat{Y}_{*} = \frac{t(R + C + T) - 2G^{*}}{(t-1)(t-2)} = \frac{5(1053 + 1107 + 989) - 2(6102)}{4 \times 3} = 295.08$$

$$\text{Amount of Bias} = \frac{[G^{*} - R - C - (t-1)T]^2}{[(t-1)(t-2)]^2} = \frac{(6102 - 1053 - 1107 - 4 \times 989)^2}{(4 \times 3)^2} = 1.36$$

Adjusted ANOVA TABLE

SOV	DF	SS	MSS	Fcal
Row	4	14771.6	3692.9	0.95 ^{ns}
Column	4	9327.0	2331.7	
Treatment	4	2944 - 1.36 = 2942.64	735.66	
Error	12 - 1 = 11	8508.3	773.48	
TOTAL	25 - 1 - 1 = 23	35550.9 - 1.36 = 35549.54		

7. FACTORIAL EXPERIMENTS *To get more comprehensive detailed study*

7.1. INTRODUCTION

Factorial experiments permit the experimenter to evaluate the combined effect of two or more factors simultaneously. ^{Advantage} Information obtained from factorial experiments is more complete than that obtained from a series of single factor experiments in the sense that factorial experiments also evaluate interaction effect which is impossible in single factor experiments.

One disadvantage of factorial experiment is that it is difficult to handle large number of factors or levels because in this case size of the experiment becomes very large. For example a factorial experiment with 6 factors each at two levels requires 64 experimental units for single replicate of this experiment, for such experiment it is difficult to have large ^② amount of homogeneous experimental material. Also in the presence of higher order interactions the interpretation of results becomes more complex.

For example, many of the properties of the chemical substance H_2O (water) cannot be predicted from the properties of Oxygen and the properties of Hydrogen studied in isolation. Most of the properties of water are attributable to the effect of the interaction between oxygen and hydrogen. The compound formed by this interaction has properties, which are not given by simply adding the properties of oxygen to the properties of hydrogen.

Factor: A factor is a kind of treatment e.g., If diet is a factor, several diets may be used, if baking temperature is a factor, several temperatures will be used for baking. If pesticide is a factor, several concentrations may be tested on insects.

Level: The term level refers to the several treatments within any factor. In the above example, different baking temperatures are levels of the factor temperature.

NOTATIONS IN FACTORIAL EXPERIMENTS

Factors are denoted by capital letters e.g., A, B, C ...and levels of the factors are denoted by small letters and subscript number to show a level number. The corresponding small letters are used to indicate the number of levels of that particular factor.

For example if a factor spray to control insects has 3 levels then we may write these levels as s_1, s_2, s_3 and $s=3$ to indicate the number of levels of the factor spray.

In factorial experiments, the experimental treatments are formed by combining the levels of different factors e.g., if an experiment involves spray at three different concentrations to control insects and these are applied by using two methods then the $3 \times 2 = 6$ treatments are $s_1m_1, s_1m_2, s_2m_1, s_2m_2, s_3m_1, s_3m_2$

Simple Effects: A simple effect is the difference (in responses) between two levels of a factor for a certain level of other factor(s) i.e., it measures the variation among different levels of a factor for a specific level of other factor.

Main Effects: A main effect is the average of simple effects of a factor. So it measures variations among various levels of a factor. Main effect of a factor is denoted by the letter of that factor.

Interaction Effects: The interaction measures the change in response of different levels of a factor over the levels of other factors. For two levels it is the difference of simple effects.

Two factors are said to interact, when a change in one factor produces a different change in the response at one level of another factor than at other level of this factor.

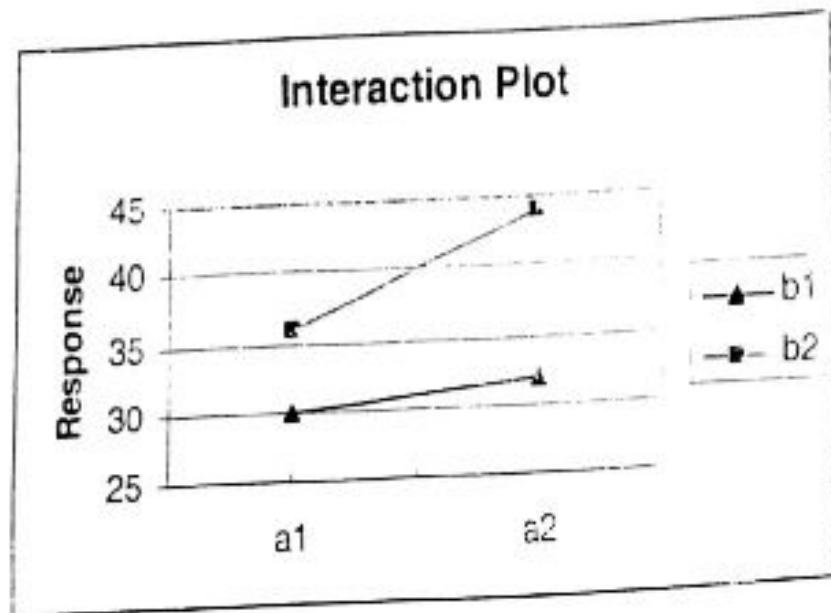
EXAMPLE

SIMPLE, MAIN AND INTERACTION EFFECTS					
	a1	a2	Simple effects of A	Main effect A	Interaction effect of A and B
b1	30	32	32-30=2	(8+2)/2=5	8-2=6
b2	36	44	44-36=8		
Simple effects of B	6	12			
Main effect of B	(6+12)/2=9				
Interaction effect of AB	12-6=6				

Main effect of factor A = Average of simple effects of factor A = 5

Main effect of factor B = Average of simple effects of factor B = 9

As the simple effects of factor A are different at different levels of factor B (2 and 8) so the two factors are *dependent* i.e., they interact with the other. The effects can be shown graphically as below. Had they been independent or there was no interaction we would have a parallel lines graph.



EXAMPLE

SIMPLE, MAIN AND INTERACTION EFFECTS					
	a1	a2	Simple effects of A	Main effect A	Interaction effect of A and B
b1	30	32	$32-30=2$	$(2+2)/2=2$	$2-2=0$
b2	36	38	$38-36=2$		
Simple effects of B	6	6			
Main effect of B	$(6+6)/2 = 6$				
Interaction effect of AB	$6-6 = 0$				

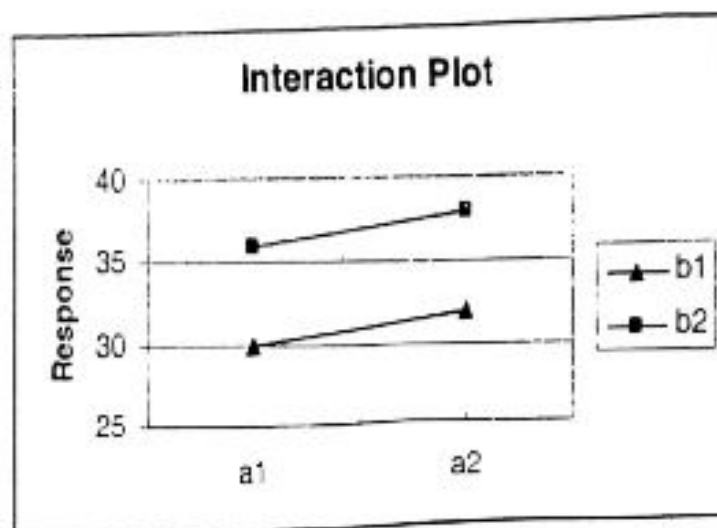
Main effect of factor A = Average of simple effects of factor A = 2

Main effect of factor B = Average of simple effects of factor B = 6

As the simple effects of factor A are same at different levels of factor B (2 & 2) so the two factors are independent i.e., they do not interact with each other.

In case of no interaction main effects are same as simple effects.

The effects can be shown graphically as below. Parallel lines show the absence of interaction between the two factors.



7.2. INTERPRETING THE RESULTS FROM A FACTORIAL EXPERIMENT

The major advantage of doing experiments when the treatment structure is of factorial nature is that you are able to study the interaction between different factors. This helps to get results from a single factorial experiment rather than conducting many single factor trials. The results from the data analysis in factorial experiments are presented differently as compared to single factor experiments. The way the means are presented depends on whether there is a significant interaction and if so, how significant.

1. If you see no evidence of interaction which means that the hypothesis of presence of interaction cannot be rejected as the $F_{\text{calculated}}$ for interaction is less than $F_{\text{tabulated}}$ then you base results on the significant main effect means. This suggests that since the factors are working independently therefore all information lies in the main effects. So we can compare main effect means for the factors which are tested in the experiment using appropriate standard error of mean or difference.
2. If you see evidence of interaction and reject the hypothesis of no interaction then you should look at the size of the F -ratio.
 - a. The interaction is significant but the F -ratio for interaction is much smaller than F -ratio for the main effects. In this situation present the two-way table of treatment means together with the overall means for each factor. Present the standard errors for the main effects as well as standard error for the difference between the two means in the body of the table.

- b. The interaction is significant but the F -ratio is of the same order or even greater than the F -ratio of the main effects. In this case the main effect of each factor is meaningless. Therefore you have to present a two-way table of treatment means without the overall means for each factor. Due to this factorial experiment is much more precise as compared to single factor experiments.

Advantage

Factorial experiments are very useful type of experiments as they help to conduct a single experiment instead of doing many experiments to understand complex phenomenon involving many factors.

✓ 7.3. ADVANTAGES

When the factors being studied in the experiment are independent, then there are two advantages of using factorial experiments.

1. All of the simple effects of a factor are equal to the main effects. Therefore the main effect is all that is needed to describe the function of a factor.
2. Each main effect is estimated as if the whole experiment was conducted for this factor alone. This phenomenon usually called hidden replication is a salient advantage of factorial structure.
3. Without factorial structure it is difficult or impossible to investigate interactions and the interpretation of factor effects may be misleading. Conversely when interactions are small or non-existent the main effect estimates are of general applicability because now we know that the effects of one factor are the same whatever the level of the other factors.

effects ★

✓ 7.4. DISADVANTAGES

To interpretation of result is difficult

There are two primary disadvantages of using factorial experiments.

1. When the number of factors increases the size of the experiment becomes very large which is difficult to handle. For example an experiment with eight factors each at two levels would need 256 experimental units for a single replication. This makes the experiment too large for limited resources which are normally available to the experimenter. Also to find uniform material for such an experiment is difficult.

2. As the number of factors increase and the factors are interacting with one another then it is difficult to interpret the results.

7.5. USES

Despite certain disadvantages the factorial experiments are useful in a number of situations:

1. In exploratory experiments where only aim is to decide which factors are important and which are not. In earlier stages of research when it is not clear what levels of the factors are to be applied factorial experiments are used to reach at suitable levels?
2. To study relationships among several factors, in particular to find the presence and magnitude of interactions among different factors.
3. In experiments where the goal is to make recommendations over a wide variety of conditions it is necessary to use factorial experiments.

7.6. FACTORIAL EXPERIMENT UNDER CRD

$$2 \times 4 \times 4 = 32$$

$$P = 2$$

Consider an agronomic experiment to assess the effects of date of planting (early or late) and type of fertilizer (none, Aero, Na, or K) on the yield of soybeans. Thirty-two homogeneous experimental plots were available. The treatments were assigned to the plots at random, subject only to the restriction that 4 plots be associated with each of the 8 treatment combinations. The data are given in the table below. Analyze the data and construct a table of means with appropriate standard errors and write a short report on the conclusions which can be drawn from this experiment. Draw any graphs you deem fit to explain the results.

$4 \times 4 \times 2$

$4 \times 8 = 32 = n = \text{No. of observations}$

Date of Planting	Fertilizer	Experimental Unit Within Treatments				Total
		1	2	3	4	
Early	Check	28.6	36.8	32.7	32.6	130.7
	Aero	29.1	29.2	30.6	29.1	118.0
	Na	28.4	27.4	26.0	29.3	111.1
	K	29.2	28.2	27.7	32.0	117.1
Late	Check	30.3	32.5	31.6	30.9	125.3
	Aero	32.7	30.8	31.0	33.8	128.3
	Na	30.3	32.7	33.0	33.4	129.4
	K	32.7	31.7	31.8	29.4	125.6
		126	127.5	127.4	128	985.8

$F \times R \times R$
 $4 \times 2 \times 4$

ANALYSIS

Correction Factor = $(G.T)^2 / (\text{Total no of observations})$

$$= (28.6 + 29.1 + 28.4 + \dots + 33.9 + 29.4)^2 / 32 = (985.8)^2 / 32 = 30368.80125$$

$$\text{Total SS} = (28.6)^2 + (29.1)^2 + \dots + (29.4)^2 - CF = 30529.34 - 30368.80125 = 160.53875$$

Two-way interaction table for Date of Planting and Fertilizer

$F \times R \times R$

Date of planting	Fertilizer				Total
	Check	Aero	Na	K	
Early	130.7	118.0	111.1	117.1	476.9
Late	125.3	128.3	129.9	125.6	508.9
Total	255.8	246.3	241.0	242.7	985.8

$$\text{Date of Planting SS} = \frac{(476.9)^2}{16} + \frac{(508.9)^2}{16} - CF = 30400.80125 - 30368.80125 = 32$$

Treatment = Fertilizer & spacing

$$\text{Fertilizer SS} = \frac{(255.8)^2}{8} + \frac{(246.3)^2}{8} + \frac{(241.0)^2}{8} + \frac{(242.7)^2}{8} - CF$$

$$= 30385.2025 - 30368.80125 = 16.40125$$

$F \times S \times R$

$$\text{Treatment SS} = \frac{(130.7)^2}{4} + \frac{(118.0)^2}{4} + \dots + \frac{(129.9)^2}{4} + \frac{(125.6)^2}{4} - CF$$

$$= 30455.595 - 30368.80125 = 86.79375$$

$$\text{Date of Planting} \times \text{Fertilizer SS} = \text{Treatment SS} - \text{Date of Planting SS} - \text{Fertilizer SS}$$

$$= 86.79375 - 32 - 16.40125 = 38.3925$$

$$\text{Error SS} = \text{Total SS} - \text{Treatment SS} = 160.53875 - 86.79375 = 73.745$$

ANOVA TABLE

SOV	SS	DF	MS	Fcal
Date of Planting	32.0	(a-1)=2-1 = 1	32.00	10.42**
Fertilizer	16.4	(b-1)=4-1 = 3	5.47	1.78
Date of Planting \times Fertilizer	38.4	(a-1)(b-1) = 3	12.80	4.17*
Error	73.7	ab(n-1) = 24	3.07	
TOTAL	160.5	32-1 = 31		

Highly significant
260
3.009
3.009
significant

$$CV = \frac{\sqrt{MSE}}{\text{Grand Mean}} \times 100 = \frac{\sqrt{3.07}}{30.81} = 5.69\%$$

$2 \times 4 (3)$
 $2 \times 4 (3)$
24

TABLE OF MEANS AND SE'S

Fertilizer	Date of Planting	
	Early	Late
Check	32.68 (0.88)	31.28 (0.88)
Aero	29.50 (0.88)	32.08 (0.88)
Na	27.78 (0.88)	32.48 (0.88)
K	29.28 (0.88)	31.40 (0.88)
Average	29.81 (0.84)	31.81 (0.84)

SE for fer = $\pm 0.88\%$ for date of planting = ± 0.64

STATISTICAL REPORT

The significant interaction indicates that the yield response due to the application of different fertilizer depends on whether the date of planting has been early or late. (Without application of any fertilizer the yield response almost remains the same at both early and late planting) For all type of fertilizer the late planting yields more than early planting and yield will be highest for application Na at late planting, which is almost same yield with application of Aero also at late planting.

The standard errors of the treatment means shown in the above table were calculated as:

Standard Errors:

$$SE(\text{Date of Planting}) = \sqrt{\frac{MSE}{bn}} = \sqrt{\frac{3.07}{4 \times 4}} = 0.44$$

date of planting statistically significant effect on yield were. fertilizer effect on yield was recorded as non-significant interaction/combine effect of fertilizer & date of planting showed significant effect on yield.

$$SE(\text{Fertilizer}) = \sqrt{\frac{MSE}{p \times R}} = \sqrt{\frac{3.07}{8}} = 0.62$$

$$SE(\text{Date} \times \text{Fertilizer}) = \sqrt{\frac{MSE}{n/kp}} = \sqrt{\frac{3.07}{4}} = 0.88$$

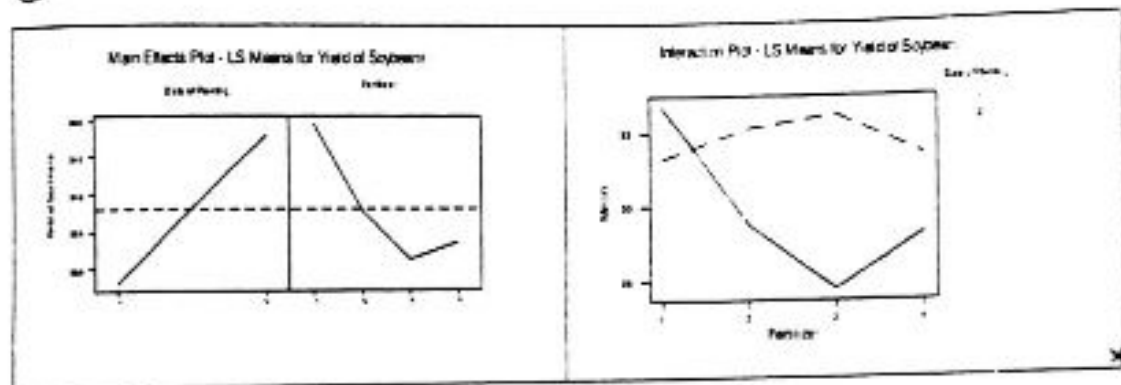
$$SE \text{ of difference between two values in body of table } \sqrt{\frac{2MSE}{n}} = \sqrt{2} \times SE(\text{Date} \times \text{Fertilizer})$$

$$= \sqrt{2} \times 0.88 = 1.24$$

SE of difference between two Date of Planting means

$$\sqrt{\frac{2MSE}{bn}} = \sqrt{2} \times SE(\text{Date}) = \sqrt{2} \times 0.44 = 0.62$$

6



$$SE \text{ b/w any } \uparrow \text{ two fertilizers} = \sqrt{\frac{2MSE}{8}} = \frac{2 \times 3.07}{8} = 0.87$$

$$SE \text{ b/w any two date of Planting} = \sqrt{\frac{2 \times 3.07}{16}} = 0.61$$

$$SE \text{ for date of planting \& fertilizer} = \sqrt{\frac{2MSE}{r}} = \sqrt{\frac{2 \times 3.07}{4}} = 1.24$$

(SE_{Day})

$$SE \text{ for date of planting means} = \sqrt{\frac{MSE}{F \times R}} = \sqrt{\frac{3.07}{4 \times 4}} = \sqrt{\frac{3.07}{16}} = 0.44$$

$$SE \text{ for fertilizer means} = \sqrt{\frac{MSE}{D \times P \times R}} = \sqrt{\frac{3.07}{2 \times 4}} = \sqrt{\frac{3.07}{8}} = 0.62$$

$$SE \text{ for date of planting \& fertilizer interaction} = \sqrt{\frac{MSE}{R}} = \sqrt{\frac{3.07}{4}} = 0.88$$

$$SE \text{ for any two date of planting means} = \sqrt{\frac{2 \times 3.07}{4 \times 4}} = \sqrt{\frac{6.14}{16}} = 0.62$$

$$SE \text{ for any two fertilizer means} = \sqrt{\frac{2 \times 3.07}{2 \times 4}} = \sqrt{\frac{6.14}{8}} = 0.88$$

$$SE \text{ for any two values of date of Planting \& fertilizer interaction} = \sqrt{\frac{2MSE}{R}} = \sqrt{\frac{2 \times 3.07}{4}} = \sqrt{\frac{6.14}{4}} = 1.24$$

7.7. FACTORIAL EXPERIMENT UNDER RCBD

An experiment was conducted on strawberries under cloches to investigate the responses of four varieties to three times of covering. A randomized block design was used, with four blocks and twelve treatment combinations.

$V \times C \times B = 4 \times 3 \times 4 = 48$

$V=4$
 $C=3$
 $B=4$

Treatment / experimental unit = $4 \times 3 \times 4 = 48$

Time of covering	Variety	Blocks				
		1	2	3	4	
Feb	V	10.2	10.1	12.1	12.3	44.7
	R	11.1	9.8	8.6	9.4	38.9
	F	6.8	9.5	9.5	10.3	36.1
	G	5.3	7.5	4.6	7.3	24.7
Mar	V	8.0	9.7	12.0	7.8	37.5
	R	9.7	7.9	10.3	11.2	39.1
	F	8.6	9.6	9.5	10.0	37.7
	G	3.4	4.2	7.3	7.6	22.5
Apr	V	2.0	6.1	4.8	6.7	19.6
	R	10.9	8.4	6.5	9.2	35.0
	F	2.2	4.9	4.4	3.6	15.1
	G	2.1	0.9	3.4	2.3	8.7
Totals		80.3	88.6	93.0	97.7	359.6

ANALYSIS

$$\text{Correction Factor} = (G.T)^2 / (\text{Total no of observations}) = \frac{13596^2}{48} = \frac{12931216}{48}$$

$$= (10.2 + 11.1 + 6.8 + \dots + 8.7)^2 / 48 = 2694.00$$

$$\text{Total SS} = (10.2)^2 + (11.1)^2 + \dots + (8.7)^2 - CF = 3133.16 - 2694.00 = 439.16$$

$$\text{Block SS} = \frac{(80.3)^2}{12} + \frac{(88.6)^2}{12} + \frac{(93.0)^2}{12} + \frac{(97.7)^2}{12} - CF$$

$$= 2707.7 - 2694.00 = 13.70$$

Two-way interaction table for times and varieties

Covering Time	Variety				Total
	V	R	F	G	
Feb	44.7	38.9	36.1	24.7	144.4
Mar	37.5	39.1	37.7	22.5	136.8
Apr	19.6	35.0	15.1	8.7	78.4
Totals	101.8	113.0	88.9	55.9	359.6

$$\text{Variety SS} = \frac{(101.8)^2}{12} + \frac{(113.0)^2}{12} + \frac{(88.9)^2}{12} + \frac{(55.9)^2}{12} - CF = 152$$

Usually field experiment can be checked in 5% level while in lab experiment 1% probability level is used

CV \rightarrow Normal lab experiment should have $< 10\%$
 \rightarrow for field 30/

$$\text{Covering Time SS} = \frac{(144.2)^2}{16} + \frac{(136.8)^2}{16} + \frac{(78.4)^2}{16} - CF = 163.01$$

$$\text{Variety} \times \text{Covering Time Subtable SS} = \frac{(44.7)^2}{4} + \frac{(38.9)^2}{4} + \frac{(8.7)^2}{4} - CF = 356.02$$

$$\text{Variety} \times \text{Covering Time SS} = \text{Variety} \times \text{Covering Time Subtable SS} - \text{Variety SS} - \text{Covering Time SS} \\ = 356.02 - 152.69 - 163.01 = 40.32$$

ANOVA TABLE

SOV	SS	DF	MS	Fcal
Block	13.70	(n-1) = 3	4.57	
Varieties	152.69	(a-1) = 3	50.90	24.2 ^{ns}
Covering Time	163.01	(b-1) = 2	81.50	38.8 [*]
Variety \times Covering time	40.32	(a-1)(b-1) = 6	6.72	3.2 [*]
Error	69.38	(ab-1)(n-1) = 33	2.012	
TOTAL	439.16	(abn-1) = 47		

Treatment
11-3-2-5

$$152.69 = 43.16 \quad \text{Total mean} = 356.02 - 152.69 = 203.33$$

CV field 18% or equal to 30 $C.V. = \frac{\sqrt{MSE}}{\text{Mean}} \times 100 = \frac{\sqrt{2.012}}{7.5} \times 100 = 18.91\%$ $P_{10} = 35.0$

STATISTICAL REPORT

As interaction is significant so varieties and covering date are not independent so present the two way table of means

Two-way interaction table for times and varieties mean

Covering Time	Variety				Means
	V	R	F	G	
Feb	11.2	9.7	9.0	6.2	9.0
Mar	9.4	9.8	9.4	5.6	8.5
Apr	4.9	8.8	3.8	2.2	4.9
Means	8.5	9.4	7.4	4.7	7.5

$$\frac{44.7}{4} = 11.2$$

Standard Errors:

SE of difference between any two values in body of the table = $\sqrt{\frac{2MSE}{4B}} = \sqrt{\frac{2 \times 2.102}{4}} = 1.03$

SE of difference between two variety means = $\sqrt{\frac{2MSE}{12}} = \sqrt{\frac{2 \times 2.102}{12}} = 0.59$

SE of difference between two covering time means = $\sqrt{\frac{2MSE}{16}} = \sqrt{\frac{2 \times 2.102}{16}} = 0.51$

$$SE_{\text{among } V \times \text{Time}} = \sqrt{\frac{MSE}{4}}$$

$$SE_{\text{among } V} = \sqrt{\frac{MSE}{T \times R}}$$

$$SE_{\text{among } T} = \sqrt{\frac{MSE}{B \times F}}$$

SE for V means

SE for T means

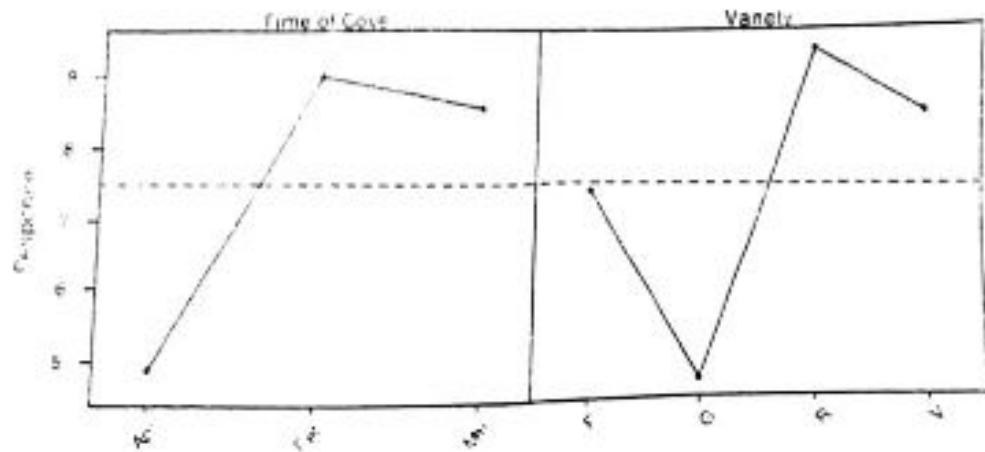
SE for VXT

SE for any two treat means

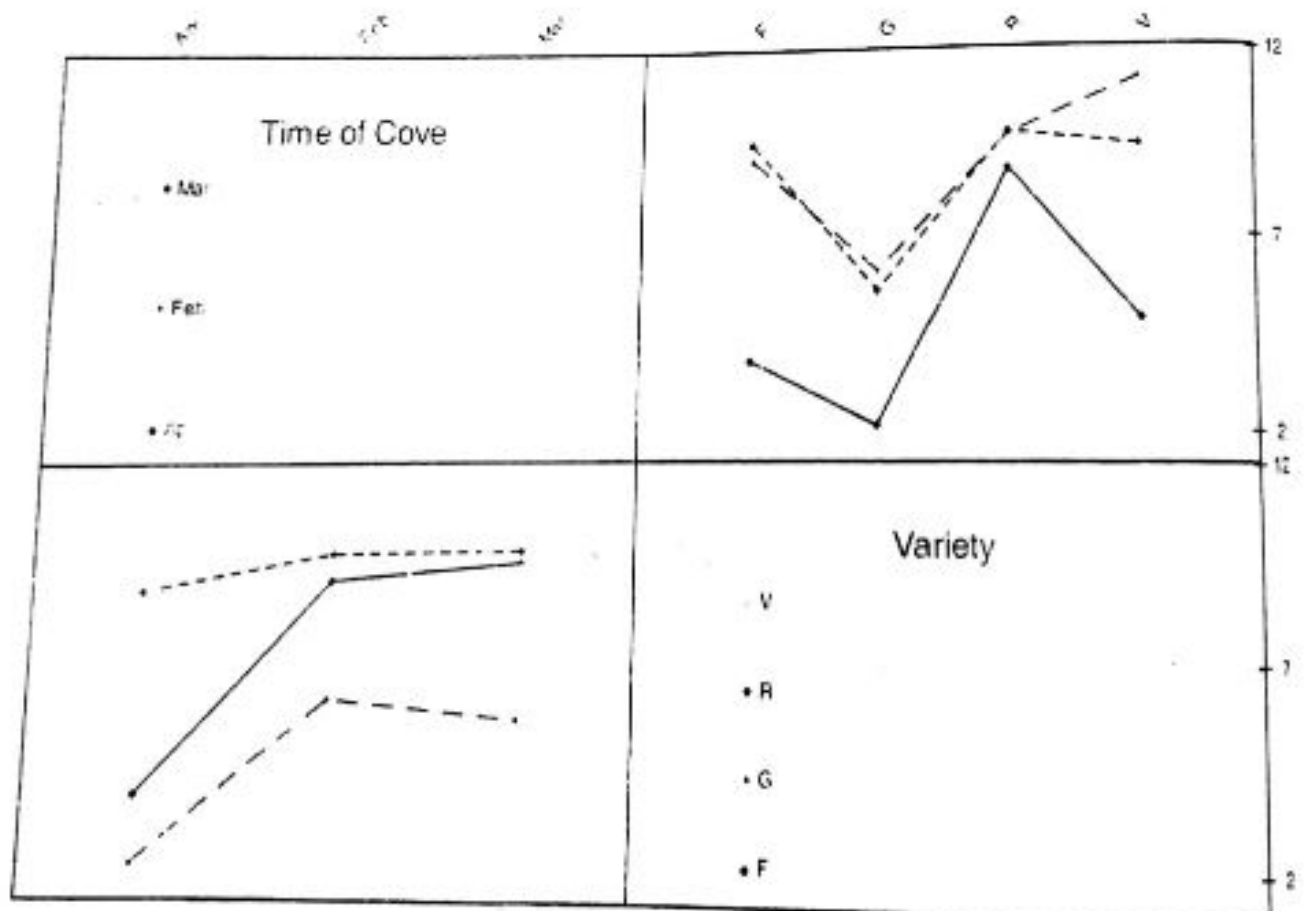
SE for any two time

SE for any two of T \times V

Main Effect Plot



Interaction Plot



For each variety the difference between the means for the first two covering times was not significant: For all varieties except R the third covering time gave lower yield than the other two times. For the first two covering time variety G was lower than the other three varieties while for the third covering time variety R gave a higher yield than the other three varieties

7.8. FACTORIAL EXPERIMENT UNDER LATIN SQUARE DESIGN

An engineer is studying methods for improving the ability to detect targets on a radar scope. Two factors she considers to be important are the amount of background noise or "ground clutter" on the scope and the type of filter placed over the screen. An experiment is designed using three levels of ground clutter and two filter types. The experiment is performed by randomly selecting a treatment combination (ground clutter level and filter type) and then introducing a signal representing the target into the scope. The intensity of this target is increased until the operator observes it. The intensity level at detection is then measured as the response variable. Because of operator availability, it is convenient to select an operator and keep him or her at the scope until all the necessary runs have been made. Furthermore, operators differ in their skill and ability to the scope. Consequently, it seems logical to use the operators as blocks. Six operators are randomly selected. Once an operator is chosen, the order in which the six treatment combinations are runs is randomly determined. Because of the setup time required, only six runs can be made per day. Thus, days become a second randomization restriction, resulting in the 6x6 Latin Square Design, as shown in the following table along with the intensity level measurements.

Day	Operator					
	1	2	3	4	5	6
1	A($f_1g_1 = 90$)	B($f_1g_2 = 106$)	C($f_1g_3 = 108$)	D($f_2g_1 = 81$)	F($f_2g_3 = 96$)	E($f_2g_2 = 88$)
2	C($f_1g_1 = 114$)	A($f_1g_2 = 96$)	B($f_1g_3 = 105$)	F($f_2g_1 = 83$)	E($f_2g_2 = 86$)	D($f_2g_3 = 84$)
3	B($f_1g_1 = 102$)	E($f_2g_2 = 90$)	F($f_2g_3 = 95$)	A($f_1g_2 = 92$)	D($f_2g_1 = 85$)	C($f_1g_3 = 104$)
4	E($f_2g_1 = 87$)	D($f_2g_2 = 84$)	A($f_1g_3 = 100$)	B($f_1g_2 = 96$)	C($f_1g_1 = 110$)	F($f_2g_3 = 91$)
5	F($f_2g_1 = 93$)	C($f_1g_2 = 112$)	D($f_2g_3 = 92$)	E($f_2g_2 = 80$)	A($f_1g_3 = 90$)	B($f_1g_1 = 98$)
6	D($f_2g_1 = 86$)	F($f_2g_2 = 91$)	E($f_2g_3 = 97$)	C($f_1g_2 = 98$)	B($f_1g_3 = 100$)	A($f_1g_1 = 92$)

In this table we have used the lowercase letters f_i and g_j to represent the i th and j th levels of filter type and ground clutter, respectively. That is, f_1g_2 represents filter type 1 and medium ground clutter.

Note that now six operators are required, so the number of treatment combinations in the 3x2 factorial design exactly equals the number of restriction levels. Furthermore, in this design, each operator would be used only once on each day. The Latin letters A, B, C, D, E, and F represent the 3x2 = 6 factorial treatment combinations as follows: A = f_1g_1 , B = f_1g_2 , C = f_1g_3 , D = f_2g_1 , E = f_2g_2 , F = f_2g_3 . The data analysis is given below:

ANALYSIS

Correction Factor = $(G.T)^2 / (\text{Total no of observations})$

$$= (90+106+102+...+98+92)^2 / 36 = (3396)^2 / 36 = 320356$$

$$\text{Total SS} = (90)^2 + (114)^2 + ... + (92)^2 - CF = 323154 - 320356 = 2798$$

Two-way interaction table for Ground Clutter and Filter Type:

Ground Clutter	Filter Type		Total
	1	2	
Low	560	512	1072
Medium	607	528	1135
High	646	543	1189
Total	1813	1583	3396

$$\text{Row/Day SS} = \frac{(563)^2}{6} + \frac{(568)^2}{6} + \frac{(568)^2}{6} + \frac{(568)^2}{6} + \frac{(565)^2}{6} + \frac{(564)^2}{6} - CF$$

$$= 320360.333 - 320356 = 4.33$$

$$\text{Col/Operator SS} = \frac{(572)^2}{6} + \frac{(578)^2}{6} + \frac{(597)^2}{6} + \frac{(530)^2}{6} + \frac{(561)^2}{6} + \frac{(557)^2}{6} - CF$$

$$= 320784 - 320356 = 428$$

$$\text{Grand Clutter SS} = \frac{(1072)^2}{12} + \frac{(1135)^2}{12} + \frac{(1189)^2}{12} - CF = 320927.5 - 320356 = 571.5$$

$$\text{Filter Type SS} = \frac{(1813)^2}{18} + \frac{(1583)^2}{18} - CF = 321825.44 - 320356 = 1469.44$$

$$\text{Treatment / Model SS} = \frac{(560)^2}{6} + \frac{(512)^2}{6} + \frac{(607)^2}{6} + \frac{(528)^2}{6} + \frac{(646)^2}{6} + \frac{(543)^2}{6} - CF$$

$$= 322523.667 - 320356 = 2167.667$$

$$\text{Ground Clutter} \times \text{Filter Type SS} = \text{Treatment SS} - \text{Ground Clutter SS} - \text{Filter Type SS}$$

$$= 2167.667 - 571.5 - 1469.44 = 126.767$$

$$\text{Error SS} = \text{Total SS} - \text{Treatment SS} - \text{Row SS} - \text{Col SS}$$

$$= 2798 - 2167.667 - 4.33 - 428 = 198.003$$

ANOVA TABLE

SOV	SS	DF	MS	Fcal
Row/Day	4.33	(ab-1) = 5	0.86	
Col/Operator	428.00	(ab-1) = 5	85.60	
Ground Clutter	571.50	(a-1) = 2	285.75	28.86**
Filter Type	1469.44	(b-1) = 1	1469.44	148.43**
Ground Clutter × Filter Type	126.77	(a-1)(b-1) = 2	63.385	6.40*
Error	198.00	(ab-1)(ab-2) = 20	9.90	
TOTAL	2798.00	(ab) ² -1 = 31		

STATISTICAL REPORT

The major difference observed was in the filter type with level one giving high level intensity measurement difference of around 13 points was observed from low to high levels which was statistically significant. The difference from low to high levels of ground clutter increased between 1 and 2 filter types from 48 to 103 which was the main reason for interaction effect to be significant.

$$2^4 = 16 \quad 2 \times 2 \times 2 \times 2 \quad \text{very imp}$$

7.9. GENERAL ANOVA TABLES (FOR FACTORIAL EXPERIMENTS)

First two columns of ANOVA table for 2 and 3 factors factorial experiments

CRD

2-Factor		3-factor	
S.O.V	D.F	S.O.V	DF
A	(a-1)	A	(a-1)
B	(b-1)	B	(b-1)
AB	(a-1)(b-1)	C	(c-1)
Error	ab(r-1)	AB	(a-1)(b-1)
		AC	(a-1)(c-1)
		BC	(b-1)(c-1)
		ABC	(a-1)(b-1)(c-1)
		Error	abc(r-1) $abcd(r-1)$
TOTAL	abr-1	TOTAL	abcr-1

RCBD

2-Factor		3-factor	
S.O.V	D.F	S.O.V	DF
Block	(r-1)	Block	(r-1)
A	(a-1)	A	(a-1)
B	(b-1)	B	(b-1)
AB	(a-1)(b-1)	C	(c-1)
Error	(ab-1)(r-1)	AB	(a-1)(b-1)
		AC	(a-1)(c-1)
		BC	(b-1)(c-1)
		ABC	(a-1)(b-1)(c-1)
		Error	(abc-1)(r-1) $abcd(r-1)$
TOTAL	abr-1	TOTAL	abcr-1

LS Design

2-Factor		3-factor	
S.O.V	D.F	S.O.V	DF
Row	(r-1)	Row	(r-1)
Column	(r-1)	Column	(r-1)
A	(a-1)	A	(a-1)
B	(b-1)	B	(b-1)
AB	(a-1)(b-1)	C	(c-1)
Error	(ab-2)(r-1)	AB	(a-1)(b-1)
		AC	(a-1)(c-1)
		BC	(b-1)(c-1)
		ABC	(a-1)(b-1)(c-1)
		Error	(abc-2)(r-1) $abcd(r-1)$
TOTAL	abr-1	TOTAL	abcr-1

A
 B
 C
 D
 AB
 AC
 AD
 BC
 BD
 CD
 ABC
 ABD
 ACD
 BCD
 ABCD
 E
 AE
 BE
 CE
 DE
 ABCE
 ABCE
 BCE
 BCE
 CDE
 ABCE
 ABCE
 BCDE
 ACDE
 ABCDE

8. SPLIT PLOT DESIGN

8.1. INTRODUCTION

In some multifactor designs, we may be unable to apply the set of all treatment combinations to the experimental units according to the randomization procedure appropriate to the CR, RCB or Latin Square designs. However, other randomization procedures are possible. One of the alternate randomizations gives rise to the split-plot design which is a special kind of Incomplete Block Design.

Split-plot designs are frequently used for factorial experiments. Such design may incorporate one or more of the CR, RCB or Latin Square designs. The underlying principle is that there are whole plots or whole units, in which levels of one or more factors are applied, then these whole plots are divided into subplots or subunits to which levels of one or more additional factors are applied. Thus each whole unit becomes a block for the subplot treatments.

The designs, in which the levels of one factor can be applied to large experimental units and the levels of the other factor to the subunits, are known as "split plot designs". In these designs, the levels of one factor are assigned at random to larger experimental units and the levels of the other factors are applied at random to the subunits within the larger experimental units. The subunits are obtained by dividing the large experimental units.

The split plot design involves assigning the levels of one factor to the main plots, which may be arranged on a CR, RCB or Latin Square design. The levels of the other factor are assigned to the subplots within each main plot. e.g., if there are three varieties and three fertilizers and we want more precision for the fertilizers, then with the Randomized Complete Block Design with three replications, the varieties are randomly assigned to the main plots within three blocks using a separate randomization for each. Then the levels of fertilizer are randomly assigned to the subplots within the main plots using a separate randomization in each main plot. The layout is shown in figure.

Block 1			Block 2			Block 3		
V_1	V_2	V_3	V_2	V_1	V_3	V_1	V_2	V_3
V_1f_1	V_2f_2	V_3f_3	V_2f_1	V_1f_2	V_3f_3	V_1f_3	V_2f_2	V_3f_1
V_1f_2	V_2f_3	V_3f_2	V_2f_2	V_1f_3	V_3f_1	V_1f_1	V_2f_3	V_3f_2
V_1f_3	V_2f_1	V_3f_1	V_2f_3	V_1f_1	V_3f_2	V_1f_2	V_2f_1	V_3f_3

Block within block and this is incomplete block.

The total number of treatments for this experiment are $3 \times 3 = 9$. From the layout it is clear that all the 9 treatments occur once in each block but within a block all levels of varieties occur together, so with respect to varieties, it is a randomized complete block with three varieties and three blocks. But it is an incomplete block as far as the full set of treatments is concerned. For this reason, split plot designs are also called Incomplete Block Designs.

The restriction imposed on randomization for assigning the levels of factors within a block give rise to two experimental errors for the split-plot design. One error is for the main plots and the other is for the subplots. The subplot error is smaller if proper layout is used. This is because it involves variability among closely spaced subplots within the main plots.

Mostly the whole plot error (E_a) is greater than subplot error (E_b) because the observations on the subunits of the same whole unit tend to be positively correlated and react more alike than subunits from different whole units. But if by chance $E_a < E_b$ then we consider both E_a and E_b as estimate of the same and consequently the two runs of squares can be pooled and divided by the pooled d.f. to obtain the estimate of σ^2 .

8.1.1. THE SPLIT-PLOT DESIGN IS DESIRABLE IN THE FOLLOWING SITUATIONS Use

- i. It may be used when the treatments associated with the levels of one or more of the factors require larger amount of experimental material in an experimental unit than do treatments for other factors. That the nature of factors may be such that levels of one factor require larger experimental units as compared to the levels of the other factor. For example, if the two factors are the sowing methods and nitrogen levels, then in two factor experiment the sowing methods require machinery, so they require large experimental units and the nitrogen levels can be applied to the smaller units.
- ii. The design may be used when an additional factor is to be incorporated in an experiment to increase its scope i.e. it may be that new treatments have to be introduced into an experiment which is already in progress.
- iii. From previous information, it may be known that larger differences can be expected among the levels of certain factors than among the levels of others.
- iv. The design is used where greater precision is desired for comparisons among certain factors than for others, greater precision may be required for levels of one factor as

compared to the levels of other factor e.g., if we want to compare two factors, varieties and fertilizers and more precision is required for fertilizers, then varieties should be placed in the larger units and the fertilizer treatments would be applied to the smaller units.

In summary, since in split-plot experiments variation among subunits is expected to be less than among whole units, the factors which require smaller amount of experimental material, or which are of major importance, or which are expected to exhibit smaller differences, or for which greater precision is desired for any reason, are assigned to the subplots. Note that in split plot designs, factors are of unequal importance.

8.1.2. STANDARD ERROR OF THE MEAN DIFFERENCE FOR EACH OF THE FOUR TYPES OF PAIR COMPARISON IN A SPLIT-PLOT DESIGN

TYPE OF PAIR COMPARISON		MEASURED AS	EXAMPLE	SE OF DIFFERENCE
NUMBER	BETWEEN			
1	Two main-plot means (averaged over all Subplot treatments)	$a_i - a_j$	$a_1 - a_2$	$\sqrt{\frac{2E_s}{bn}}$
2	Two main-plot means at the same or different sub-plot treatments	$a_i b_j - a_k b_l$ $a_i b_1 - a_j b_1$	$a_1 b_1 - a_2 b_1$ $a_1 b_2 - a_2 b_1$	$\sqrt{\frac{2\{(b-1)E_s + E_s\}}{bn}}$
3	Two sub-plot means (averaged over all main-plot treatments)	$b_i - b_j$	$b_1 - b_2$	$\sqrt{\frac{2E_s}{an}}$
4	Two sub-plot means at the same main-plot Treatments	$a_i b_j - a_i b_k$	$a_1 b_1 - a_1 b_2$	$\sqrt{\frac{2E_s}{n}}$

$E_s = MS_{Error}$, $E_s = MS_{Error}$, n = Number of Replications, a = Number of main-plots treatments, b = Number of sub-plot treatments.

EXAMPLE $V \times S \times R$
 $4 \times 3 \times 4$

A researcher was interested to compare 4 varieties of wheat at 3 seeding rates. An experiment was conducted in a Completely Randomized Design with four replicates each and with 4 varieties in main plots and 3 seeding rates in subplots. The data recorded about the plant height at maturity (cm) is as follows:

least important
main plot \times block

Subplot \times main plot / least important

$V=4$

$S=3$

$R=4$

Simp
mai

confounding.

$V \times S$
4 3 8
22 10

	V ₁				V ₂		
	S ₁	S ₂	S ₃		S ₁	S ₂	S ₃
r ₁	93.80	96.08	95.44	r ₂	94.16	98.64	95.72
r ₂	94.52	91.68	90.29	r ₃	95.91	101.12	99.36
r ₃	86.32	87.92	88.48	r ₄	94.20	93.92	91.20
r ₄	89.44	89.84	88.76		90.04	97.36	95.48
	364.08	365.52	362.97		374.31	363.64	369.00

	V ₃				V ₄		
	S ₁	S ₂	S ₃		S ₁	S ₂	S ₃
	85.04	90.12	91.76		98.80	95.64	101.92
	91.68	95.72	98.40		96.08	99.16	97.12
	92.28	85.56	90.04		91.80	96.92	99.36
	95.16	92.24	88.80		99.44	94.44	96.20

incomplete Block Design
Split Block Design

Analyze the data and present the ANOVA table

ANALYSIS

Correction Factor = $(G.T)^2 / (\text{Total no of observations})$

$$= (93.80 + 94.52 + \dots + 99.36 + 96.20)^2 / 48 = (4503.36)^2 / 48 = 422505.2352$$

$$\text{Total SS} = (93.80)^2 + (94.52)^2 + \dots + (96.20)^2 - CF = 423321.4234 - 422505.2352 = 816.1882$$

Two-way table for Varieties and Seeding Rates

Seeding Rate	Variety				Total	Mean
	V ₁	V ₂	V ₃	V ₄		
S ₁	364.08	374.31	364.16	386.12	1488.67	93.64
S ₂	365.52	391.04	363.64	386.16	1506.36	94.15
S ₃	362.97	381.76	369.00	394.60	1508.33	94.27
Total	1092.57	1147.11	1096.80	1166.88	4503.36	93.82
Mean	91.05	95.59	91.40	97.24	93.82	

$R \times V \times S$
 $4 \times 4 \times 3$

Two-way table for Varieties and Replicates

Replicate	Variety				Total
	V ₁	V ₂	V ₃	V ₄	
1	285.32	288.52	266.92	296.36	1137.12
2	276.49	296.39	285.80	292.36	1151.04
3	262.72	279.32	267.88	288.08	1098.00
4	268.04	282.88	276.20	290.08	1117.20
Total	1092.57	1147.11	1096.80	1166.88	4503.36

$$\text{Variety SS} = \frac{(1092.57)^2}{12} + \frac{(1147.11)^2}{12} + \frac{(1096.80)^2}{12} + \frac{(1166.88)^2}{12} - CF$$

$$= 422845.811 - 422505.2352 = 340.5758$$

$$\text{Rep} \times \text{V (Subtable Total SS)} = \frac{(1285.32)^2}{3} + \frac{(1276.49)^2}{3} + \dots + \frac{(288.08)^2}{3} + \frac{(290.08)^2}{3} - CF$$

$$= 423089.0716 - 422505.2352 = 583.8364$$

$$\text{Main Plot Error or Error}_1 \text{ SS} = \text{Rep} \times \text{V Subtable Total SS} - \text{Variety SS}$$

$$= 583.8364 - 340.5758 = 243.2606$$

$$\text{Seeding Rate SS} = \frac{(1488.67)^2}{16} + \frac{(1506.36)^2}{16} + \frac{(1508.33)^2}{16} - CF$$

$$= 422519.888 - 422505.2352 = 14.6528$$

$$\text{V} \times \text{S (Subtable Total SS)} = \frac{(364.08)^2}{4} + \frac{(365.52)^2}{4} + \dots + \frac{(386.16)^2}{4} + \frac{(394.60)^2}{4} - CF$$

$$= 422898.0521 - 422505.2352 = 392.8169$$

$$\text{Variety} \times \text{Seeding Rate SS} = \text{V} \times \text{S Subtable Total SS} - \text{Variety SS} - \text{Seeding Rate SS}$$

$$= 392.8169 - 340.5758 - 14.6528 = 37.5883$$

$$\text{Subplot Error or Error}_2 \text{ SS} = \text{Total SS} - \text{Variety} \times \text{Seeding Rate Subtable Total SS} - \text{Error}_1 \text{ SS}$$

$$= 816.1882 - 392.8169 - 243.2606 = 180.1107$$

ANOVA TABLE

SCV	SS	DF	MS	Fcal
Variety	340.5758	(a-1) = 3	113.5253	5.60*
Error ₁	243.2606	a(n-1) = 12	20.2717	
Seeding Rate	14.6528	(b-1) = 2	7.3264	0.98 ^{NS}
Variety × Seeding Rate	37.5883	(a-1)(b-1) = 6	6.2647	0.84 ^{NS}
Error ₂	180.1107	a(b-1)(n-1) = 24	7.5046	
TOTAL	816.1882	(abn-1) = 47		

$$CV_1 = \frac{\sqrt{MSE_1}}{\text{Mean}} \times 100 = \frac{\sqrt{20.2717}}{93.82} = 4.80\%$$

$$CV_2 = \frac{\sqrt{MSE_2}}{\text{Mean}} \times 100 = \frac{\sqrt{7.5046}}{93.82} = 2.91\%$$

STATISTICAL REPORT

As interaction is not significant so Variety and Seed rate are acting independently of each other. There is no difference in the yield due to different seeding rates. As main effect of Variety is significant so we present one way table of variety means only:

Variety	V ₁	V ₂	V ₃	V ₄
Mean	91.05	95.59	91.40	97.24

Variety V₄ produced the maximum yield where as V₁ produced the minimum yield which is not much different from yield of V₃. Thus seeding rates used did not produce significantly different yields.

Standard Errors

- SE of difference between two main plot means

$$\sqrt{\frac{2E_a}{bn}} = \sqrt{\frac{2 \times 20.2717}{3 \times 4}} = 1.838$$

- SE of difference between two main plot means at the same or different sub-plot treatments

$$\sqrt{\frac{2(b-1)E_b + E_a}{bn}} = \sqrt{\frac{2 \times (3-1)(7.5046) + 20.2717}{3 \times 4}} = 2.425$$

EXAMPLE (Note: This is the same example but considered as Split plot in RCBD for the purpose of demonstration of Analysis)

A researcher was interested to compare 4 varieties of wheat at 3 seeding rates. An experiment was conducted in a Randomized Complete Block Design with 4 varieties in main plots and 3 seeding rates in subplots. The data recorded about the plant height at maturity (cm) is as follows:

Block	V ₁			V ₂		
	S ₁	S ₂	S ₃	S ₁	S ₂	S ₃
1	93.80	96.08	95.44	94.16	98.64	95.72
2	94.52	91.68	90.29	95.91	101.12	99.36
3	86.32	87.92	88.48	94.20	93.92	91.20
4	89.44	89.84	88.76	90.04	97.36	95.48

Block	V ₃			V ₄		
	S ₁	S ₂	S ₃	S ₁	S ₂	S ₃
1	85.04	90.12	91.76	98.80	95.64	101.92
2	91.68	95.72	98.40	96.08	99.16	97.12
3	92.28	85.56	90.04	91.80	96.92	99.36
4	95.16	92.24	88.80	99.44	94.44	96.20

Analyze the data and complete the ANOVA table.

SOLUTION

Correction Factor = $(G.T)^2 / (\text{Total no of observations})$

$$= (93.80 + 94.52 + \dots + 99.36 + 96.20)^2 / 48 = (4503.36)^2 / 48 = 422505.2352$$

$$\begin{aligned} \text{Total SS} &= (93.80)^2 + (94.52)^2 + \dots + (96.20)^2 - CF \\ &= 423321.4234 - 422505.2352 = 816.1882 \end{aligned}$$

$$\begin{aligned} \text{Block SS} &= \frac{(1137.12)^2}{12} + \frac{(1151.04)^2}{12} + \frac{(1098.00)^2}{12} + \frac{(1117.20)^2}{12} - CF \\ &= 422639.568 - 422505.2352 = 134.3328 \end{aligned}$$

Two-way table for Varieties and Blocks

Block	Variety				Total
	V ₁	V ₂	V ₃	V ₄	
1	285.32	288.52	266.92	296.36	1137.12
2	276.49	296.39	285.80	292.36	1151.04
3	262.72	279.32	267.88	288.08	1098.00
4	268.04	282.88	276.20	290.08	1117.20
Total	1092.57	1147.11	1096.80	1166.88	4503.36

Two-way table for Varieties and Seeding Rates

Seeding Rate	Variety				Total	Mean
	V ₁	V ₂	V ₃	V ₄		
S ₁	364.08	374.31	364.16	386.12	1488.67	93.04
S ₂	365.52	391.04	363.64	386.16	1506.36	94.15
S ₃	362.97	381.76	369.00	394.60	1508.33	94.27
Total	1092.57	1147.11	1096.80	1166.88	4503.36	93.82
Mean	91.05	95.59	91.40	97.24	93.82	

CV → coefficient of variance describe the difference efficiency of the experiment. How much the result/work is precise. It should be in the range of 10-30%.

$$\text{Variety SS} = \frac{(1092.57)^2}{12} + \frac{(1147.11)^2}{12} + \frac{(1096.80)^2}{12} + \frac{(1166.85)^2}{12} - CF$$

$$= 422845.811 - 422505.2352 = 340.5758$$

$$\text{Block} \times \text{V Subtable Total SS} = \frac{(285.32)^2}{2} + \frac{(276.49)^2}{3} + \frac{(288.08)^2}{3} + \frac{(290.57)^2}{3} - CF$$

$$= 423089.0716 - 422505.2352 = 583.8364$$

$$\text{Main Plot Error or Error}_1 \text{ SS} = \text{Block} \times \text{Variety Subtable Total SS} - \text{Variety SS} - \text{Block SS}$$

$$= 583.8364 - 340.5758 - 134.3328 = 108.9277$$

$$\text{Seeding Rate SS} = \frac{(1488.67)^2}{16} + \frac{(1506.36)^2}{16} + \frac{(1508.33)^2}{16} - CF$$

$$= 422519.888 - 422505.2352 = 14.6528$$

$$\text{V} \times \text{S Subtable Total SS} = \frac{(364.08)^2}{4} + \frac{(365.52)^2}{4} + \frac{(386.16)^2}{4} + \frac{(394.60)^2}{4} - CF$$

$$= 422898.0521 - 422505.2352 = 392.8169$$

$$\text{Variety} \times \text{Seeding Rate SS} = \text{V} \times \text{S Subtable Total SS} - \text{Variety SS} - \text{Seeding Rate SS}$$

$$= 392.8169 - 340.5758 - 14.6528 = 37.5883$$

$$\text{Subplot Error or Error}_2 \text{ SS} = \text{Total SS} - \text{V} \times \text{S Subtable Total SS} - \text{Error}_1 \text{ SS} - \text{Block SS}$$

$$= 816.1882 - 392.8169 - 108.9277 - 134.3329 = 180.1107$$

ANOVA TABLE

SOV	SS	DF	MS	F cal
Block	134.3329	(n-1) = 3	44.7776	
Variety	340.5758	(a-1) = 3	113.5253	9.25
Error ₁	108.9277	(a-1)(n-1) = 9	12.1031	
Seeding Rate	14.6528	(b-1) = 2	7.3264	0.58
Variety × Seeding Rate	37.5883	(a-1)(b-1) = 6	6.2647	
Error ₂	180.1107	a(b-1)(n-1) = 24	7.5046	
TOTAL	816.1882	(abn-1) = 47		

$$CV_A = \frac{\sqrt{MSE_A}}{\text{Mean}} \times 100 = \frac{\sqrt{12.1031}}{93.82} \times 100 = 3.7\%$$

$$CV_B = \frac{\sqrt{MSE_B}}{\text{Mean}} \times 100 = \frac{\sqrt{7.5046}}{93.82} \times 100 = 2.91\%$$

STATISTICAL REPORT

As interaction is not significant so Variety and Seed rate are acting independently of each other. There is no difference in the yield due to different seeding rate. As main effect of Variety is significant so present one way table of variety means only.

Variety	V ₁	V ₂	V ₃	V ₄
Mean	91.05	95.59	91.40	97.24

Variety V₄ produce the maximum yield where is V₁ produces the minimum yield which is not much different from yield of V₃.

Standard Errors

SE of difference between two main plot means

$$\frac{\sqrt{2E_e}}{bn} = \frac{\sqrt{2 \times 12.1031}}{3 \times 4} = 0.41$$

SE of difference between two main plot means at the same or different sub-plot treatments

$$\frac{\sqrt{2(b-1)E_e + I_{e,c}}}{bn} = \frac{\sqrt{2 \times (3-1)(7.5046) + 12.1031}}{3 \times 4} = 0.54$$

8.1.3. GENERAL ANOVA FOR 2 AND 3 FACTORS SPLIT PLOT DESIGN

CRD

2-Factor		3-factor	
S.O.V	D.F	S.O.V	Df
A	$(a-1)$	A	$(a-1)$
E1	$a(r-1)$	E1	$a(r-1)$
B	$(b-1)$	B	$(b-1)$
AB	$(a-1)(b-1)$	AB	$(a-1)(b-1)$
E2	$a(b-1)(r-1)$	E2	$a(b-1)(r-1)$
		C	$(c-1)$
		AC	$(a-1)(c-1)$
		BC	$(b-1)(c-1)$
		ABC	$(a-1)(b-1)(c-1)$
		E3	$ab(c-1)(r-1)$
TOTAL	$abr-1$	TOTAL	$abcr-1$

RCBD

2-Factor		3-factor	
S.O.V	D.F	S.O.V	Df
Block	$(r-1)$	Block	$(r-1)$
A	$(a-1)$	A	$(a-1)$
E1	$(a-1)(r-1)$	E1	$(a-1)(r-1)$
B	$(b-1)$	B	$(b-1)$
AB	$(a-1)(b-1)$	AB	$(a-1)(b-1)$
E2	$a(b-1)(r-1)$	E2	$a(b-1)(r-1)$
		C	$(c-1)$
		AC	$(a-1)(c-1)$
		BC	$(b-1)(c-1)$
		ABC	$(a-1)(b-1)(c-1)$
		E3	$ab(c-1)(r-1)$
TOTAL	$abr-1$	TOTAL	$abcr-1$

LS Design

2-Factor		3-factor	
S.O.V	D.F	S.O.V	Df
Row	$(r-1)$	Row	$(r-1)$
Column	$(r-1)$	Column	$(r-1)$
A	$(a-1)$	A	$(a-1)$
E1	$(a-2)(r-1)$	E1	$(a-2)(r-1)$
B	$(b-1)$	B	$(b-1)$
AB	$(a-1)(b-1)$	AB	$(a-1)(b-1)$
E2	$a(b-1)(r-1)$	E2	$a(b-1)(r-1)$
		C	$(c-1)$
		AC	$(a-1)(c-1)$
		BC	$(b-1)(c-1)$
		ABC	$(a-1)(b-1)(c-1)$
		E3	$ab(c-1)(r-1)$
TOTAL	$abr-1$	TOTAL	$abcr-1$

9. SPLIT BLOCK OR STRIP PLOT DESIGN

9.1. INTRODUCTION

The Strip-plot design is specifically suited for a two-factor experiment in which the desired precision for measuring the interaction effect between the two factors is higher than that for measuring the main effect of either one of the two factors. This is accomplished with the use of three plot sizes:

1. *Vertical-strip plot* for the first factor—the *vertical factor*.
2. *Horizontal-strip plot* for the second factor—the *horizontal factor*.
3. *Interaction plot* for the interaction between the two factors.

The vertical-strip and the horizontal-strip plot are always perpendicular to each other. However, there is no relationship between their sizes, unlike the case of main plot and subplot of the split-plot design. The interaction plot is, of course, the smallest. Thus, in a strip-plot design, the degrees of precision associated with the main effects of both factors are sacrificed in order to improve the precision of the interaction effects.

In the Split Plot Design, the levels of second factor B are independently randomized within the subplots of each level of first factor A . But it may be that the levels of factor B are applied in strips across an entire replication of the factor A in main plots, this is called "Split Block or Strip Plot Design".

In the Strip Block design the levels of one factor are assigned to strip plots in one direction and the levels of the second factor to the strips perpendicular to the first strip. A separate randomization is done for each block for each factor A and B .

The design facilitates physical operations and increases the precision for estimation of AB interaction.

If there are three varieties and three fertilizers where the varieties are in main plots and the fertilizers in subplots, then the layout given below for one block indicates the difference between a Split Plot and a Split Block Design.

SPLIT PLOT		
V ₂	V ₁	V ₃
F ₃	F ₂	F ₁
F ₁	F ₁	F ₂
F ₂	F ₃	F ₃

SPLIT BLOCK		
V ₂	V ₁	V ₃
F ₂	F ₂	F ₂
F ₁	F ₁	F ₁
F ₃	F ₃	F ₃

STANDARD ERROR OF THE MEAN DIFFERENCE FOR EACH OF THE FOUR TYPES OF PAIR COMPARISON IN A SPLIT-BLOCK DESIGN

TYPE OF PAIR COMPARISON		MEASURED AS	EXAMPLE	SE OF DIFFERENCE
NUMBER	BETWEEN			
1	Two horizontal means (averaged over all vertical treatments)	$a_i - a_j$	$a_1 - a_2$	$\sqrt{\frac{2E_e}{bn}}$
2	Two horizontal means at the same vertical treatments	$a_i b_j - a_k b_j$	$a_1 b_1 - a_2 b_1$	$\sqrt{\frac{2\{(b-1)E_{a\cdot} + E_e\}}{bn}}$
3	Two vertical means (averaged over all horizontal treatments)	$b_i - b_j$	$b_1 - b_2$	$\sqrt{\frac{2E_{\cdot e}}{an}}$
4	Two vertical means at the same horizontal treatments	$a_i b_j - a_i b_k$	$a_1 b_1 - a_1 b_2$	$\sqrt{\frac{2\{(a-1)E_{\cdot b} + E_e\}}{an}}$

$E_a = MS_{Error(a)}$, $E_b = MS_{Error(b)}$, $E_{ab} = MS_{Error(ab)}$, n = Number of Replications
 a = Number of levels of horizontal factor, b = Number of levels of vertical factor.

EXAMPLE

A researcher was interested to determine the effect of three irrigation rates on three varieties of chickpea cultivars. He made three blocks and assigned three varieties to strips of three plots in each block. Irrigation levels were then given to strips of three plots across the variety plots at right angles. Separate randomization was used in each case. The average of nodules per plant of chickpea cultivars are recorded as

Block	V ₁			V ₂			V ₃		
	I ₁	I ₂	I ₃	I ₁	I ₂	I ₃	I ₁	I ₂	I ₃
1	20.5	25.0	33.1	24.0	43.8	44.4	20.6	25.0	28.4
2	24.2	31.0	38.5	27.5	38.3	50.7	16.5	32.0	34.1
3	19.2	23.7	36.8	25.7	37.1	45.9	15.5	20.8	30.7

Analyze the data and complete the ANOVA table.

SOLUTION

Two-way table for Varieties and Blocks

Block	Variety			Total
	V ₁	V ₂	V ₃	
1	78.6	112.2	74.0	264.8
2	93.7	116.5	82.6	292.8
3	79.7	108.7	67.0	255.4
Total	252.0	337.4	223.6	813.0

Two-way table for Varieties and Irrigation Rates

Irrigation Rate	Variety			Total	Mean
	V ₁	V ₂	V ₃		
I ₁	63.9	77.2	52.6	193.7	21.5
I ₂	79.7	119.2	77.8	276.7	30.7
I ₃	108.4	141.0	93.2	342.6	38.0
Total	252.0	337.4	223.6	813.0	
Mean	28.0	34.5	24.8		

Two-way table for Varieties and Irrigation Rates

Block	Irrigation Rate			Total
	I ₁	I ₂	I ₃	
1	65.1	93.8	105.9	264.8
2	68.2	101.3	123.3	292.8
3	60.4	81.6	113.4	255.4
Total	193.7	276.7	342.6	813.0

Correction Factor = $(G.T)^2 / (\text{Total no of observations})$

$$= (20.5 + 24.2 + \dots + 34.1 + 30.7)^2 / 27 = (813.0)^2 / 27 = 24480.3333$$

$$\text{Total SS} = (20.5^2 + (24.2)^2 + \dots + (30.7)^2) - CF = 26787.8200 - 24480.3333 = 2307.4867$$

$$\text{Block SS} = \frac{(264.8)^2}{9} + \frac{(292.8)^2}{9} + \frac{(255.4)^2}{9} - CF$$

$$= 24564.4489 - 24480.3333 = 84.1156$$

$$\text{Variety SS} = \frac{(252.0)^2}{9} + \frac{(337.4)^2}{9} + \frac{(223.6)^2}{9} - CF$$

$$= 25259.9689 - 24480.3333 = 779.6356$$

$$\text{Block} \times \text{V Subtable Total SS} = \frac{(78.6)^2}{3} + \frac{(93.7)^2}{3} + \dots + \frac{(82.6)^2}{3} + \frac{(67.0)^2}{3} - CF$$

$$= 25358.0933 - 24480.3333 = 877.7600$$

$$\text{Variety Error or Error}_V \text{ SS} = \text{Block} \times \text{Variety Subtable Total SS} - \text{Variety SS} - \text{Block SS}$$

$$= 877.7600 - 779.6356 - 84.1156 = 14.0088$$

$$\text{Irrigation Rate SS} = \frac{(193.7)^2}{9} + \frac{(276.7)^2}{9} + \frac{(342.6)^2}{9} - CF$$

$$= 25717.4822 - 24480.3333 = 1237.1489$$

$$\text{Block} \times \text{I Subtable Total SS} = \frac{(65.1)^2}{3} + \frac{(68.2)^2}{3} + \dots + \frac{(123.3)^2}{3} + \frac{(113.4)^2}{3} - CF$$

$$= 25844.4533 - 24480.3333 = 1364.1200$$

$$\text{Irrigation Error or Error}_I \text{ SS} = \text{Block} \times \text{Irrigation Subtable Total SS} - \text{Irrigation SS} - \text{Block SS}$$

$$= 1364.1200 - 1237.1489 - 84.1156 = 42.8555$$

$$\text{V} \times \text{I Subtable Total SS} = \frac{(63.9)^2}{3} + \frac{(79.7)^2}{3} + \dots + \frac{(77.8)^2}{3} + \frac{(93.2)^2}{3} - CF$$

$$= 26580.3933 - 24480.3333 = 2100.0600$$

$$\text{Variety} \times \text{Irrigation Rate SS} = \text{V} \times \text{I Subtable Total SS} - \text{Variety SS} - \text{Irrigation Rate SS}$$

$$= 2100.0600 - 779.6356 - 1237.1489 = 83.2756$$

$$\text{V} \times \text{I Error or Error}_{VI} \text{ SS} = \text{Total SS} - \text{V} \times \text{I Subtable Total SS} - \text{Error}_V \text{ SS} - \text{Error}_I \text{ SS} - \text{Block SS}$$

$$= 2307.4867 - 2100.0600 - 14.0088 - 42.8555 - 84.1156 = 66.4467$$

ANOVA TABLE

SOV	SS	DF	MS	Fcal
Block	84.1156	(n-1) = 2	42.0578	
Variety	779.6356	(a-1) = 2	389.8178	111.31 ^{***}
Error _v	14.0088	(a-1)(n-1) = 4	3.5022	
Irrigation rate	1237.1489	(b-1) = 2	618.5744	57.74 ^{***}
Error _i	42.8556	(b-1)(n-1) = 4	10.7139	
Variety × Irrigation Rate	83.2756	(a-1)(b-1) = 4	20.8189	2.51 ^{ns}
Error _{v.i}	66.4467	(a-1)(b-1)(n-1) = 8	8.3058	
TOTAL	2307.4867	(abn-1) = 26		

STATISTICAL REPORT

An insignificant interaction indicates that Variety and Irrigation Rate are acting independently of each other. There is difference in the yield due to different Varieties as well as due to different Irrigation Rate. Present the results in one way table of means for Varieties and Irrigation Rates separately.

Irrigation Rate	I ₁	I ₂	I ₃
Mean	21.52	30.74	38.03

The yield of chickpea cultivars is highest for 3rd level of Irrigation and this behaviour remains the same for all three varieties.

Variety	V ₁	V ₂	V ₃
Mean	28.00	34.49	24.84

Variety V₂ produces the highest yield and this remains true for all irrigation levels.

STANDARD ERROR OF THE MEAN DIFFERENCE FOR EACH OF THE FOUR TYPES OF PAIR COMPARISON IN A SPLIT-BLOCK DESIGN FOR CHICKPEA EXAMPLE.

TYPE OF PAIR COMPARISON		MEASURED AS	SE OF DIFFERENCE
NUMBER	BETWEEN		
1	Two horizontal means (averaged over all vertical treatments)	$a_i - a_j$	$\sqrt{\frac{2E_a}{bn}} = \sqrt{\frac{2 \times 3.5022}{3 \times 3}}$ $= 0.8822$
2	Two horizontal means at the same vertical treatments	$a_i b_j - a_k b_j$	$\sqrt{\frac{2\{(b-1)E_a + E_b\}}{bn}}$ $= \sqrt{\frac{2\{(3-1) \times 8.3058 + 3.5022\}}{3 \times 3}}$ $= 2.1142$
3	Two vertical means (averaged over all horizontal treatments)	$b_i - b_j$	$\sqrt{\frac{2E_b}{an}} = \sqrt{\frac{2 \times 10.7139}{3 \times 3}}$ $= 1.5430$
4	Two vertical means at the same horizontal treatments	$a_i b_j - a_i b_k$	$\sqrt{\frac{2\{(a-1)E_b + E_a\}}{an}}$ $= \sqrt{\frac{2\{(3-1) \times 8.3058 + 10.7139\}}{3 \times 3}}$ $= 2.4642$

$E_a = MS_{\text{Error}(a)}$, $E_b = MS_{\text{Error}(b)}$, $E_{ab} = MS_{\text{Error}(ab)}$, n = Number of Replications

a = Number of levels of horizontal factor, b = Number of levels of vertical factor.

10. NESTED OR HIERARCHICAL DESIGNS

10.1. INTRODUCTION

In certain multifactor experiments the level of one factor (e.g., factor B) are similar but not identical for different levels of another factor (e.g., factor A). Such an arrangement is called a *Nested* or *Hierarchical design*, with the levels of factor B nested under the levels of factor A .

Factor A is said to be crossed with another factor B if every level of factor A occurs with every level of factor B and vice versa. This is what is usually seen in the factorial experiments. Sometimes the factors may be nested inside of one another. Factor B is said to be nested with another factor A if the levels of B exist only within the levels of A . In the crossed classification the levels of factor B are identical whereas in the nested classification, the so called levels of B are similar. The terms *Hierarchical* is also used for nested classification.

The linear statistical model for the two stage nested design is

$$Y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \epsilon_{(ij)k} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, n \end{cases}$$

That is, there are a levels of factor A , b levels of factor B nested under each levels of factor A , and n replicates. The subscript $j(i)$ indicates that the j th level of factor B is nested under the i th level of factor A . It is convenient to think of the replicates as being nested within the combination of levels of A and B ; thus, the subscript $(ij)k$ is used for the error term. This is a balanced nested design since there are an equal number of levels of B within each level of A and equal number of replicates. Since every level of factor B does not appear with every level of factor A , there can be no interaction between A and B .

EXAMPLE

Four plants were taken at random, and then three leaves were randomly selected from each plant. From each leaf were taken two samples of 100 mg in which calcium was determined by microchemical methods.

Calcium Concentration (% Dry basis) in b=3 leaves from
Each of a=4 Turnip Plants, n=2 Determination/Leaf

Plant, i i= 1,2,...,a	Leaf, j j= 1,2,...,b	Determination	
		1	2
1	1	3.28	3.09
	2	3.50	3.45
	3	2.88	2.80
2	1	2.28	2.44
	2	1.87	1.92
	3	2.19	2.19
3	1	2.77	2.66
	2	3.74	3.44
	3	2.87	2.57
4	1	3.78	3.87
	2	4.07	4.12
	3	3.31	3.31

SOLUTION

Correction Factor = $(G.T)^2 / (\text{Total no of observations})$

$$= (3.28+3.09+\dots+3.31)^2/24 = (72.29)^2/24 = 217.7435$$

$$\text{Total SS} = (3.28)^2 + (3.09)^2 + \dots + (3.31)^2 - CF = 228.0139 - 217.7435 = 10.2704$$

Two-way table for Plants and Leaves

Leaf	Plant				Total
	1	2	3	4	
1	6.37	4.90	5.43	7.65	24.35
2	7.00	3.79	7.18	8.19	26.16
3	5.68	4.38	5.10	6.62	21.78
Total	19.05	13.07	17.71	22.46	72.29

$$\begin{aligned} \text{Plants SS} &= \frac{(19.05)^2}{6} + \frac{(13.07)^2}{6} + \frac{(17.71)^2}{6} + \frac{(22.46)^2}{6} - CF \\ &= 225.3039 - 217.7435 = 7.5604 \end{aligned}$$

$$\begin{aligned} \text{Plants} \times \text{Leaves Subtable SS} &= \frac{(6.37)^2}{2} + \frac{(4.90)^2}{2} + \dots + \frac{(6.62)^2}{2} - CF \\ &= 227.93405 - 217.7435 = 10.19055 \end{aligned}$$

$$\begin{aligned} \text{Leaves within Plants SS} &= \text{Plants} \times \text{Leaves Subtable SS} - \text{Plant SS} \\ &= 10.19055 - 7.5604 = 2.63015 \end{aligned}$$

$$\begin{aligned}
 \text{Error SS} &= \text{Total SS} - \text{Plants} \times \text{Leaves Subtable SS} \\
 &= \text{Total SS} - \text{Plants SS} - \text{Leaves within Plants SS} \\
 &= 10.2704 - 10.19055 = 0.07985
 \end{aligned}$$

ANOVA TABLE				
SOV	SS	DF	MS	Fcal
Plants	7.5604	(a-1) = 3	2.5201	$\frac{2.5201}{0.3288} = 7.66^*$
Leaves(Plants)	2.6302	a(b-1) = 8	0.3288	$\frac{0.3288}{0.0067} = 49.00^{**}$
Error	0.0799	ab(n-1) = 12	0.0067	
TOTAL	10.2704	(abn-1) = 23		

STATISTICAL REPORT

There is difference in calcium determination among leaves within plants as well as difference in determination from plant to plant.

11. INCOMPLETE BLOCK DESIGN

11.1. INTRODUCTION

The subject of Statistics deals with variability and methods how to deal with it. In the planning of an investigation, the items used to control variability are

1. Refinement of experimental technique.
2. Selection of homogeneous material and/or environment.
3. Grouping (blocking, stratifying) material into homogeneous subgroup (block, strata), to compare treatments in as similar conditions as possible.
4. Measurement of related variables and use of covariance.

There are many ways of blocking (arranging) the experimental units (EUs) in a comparative experiment with t treatments. If the sample of EUs is from a homogeneous population, then no blocking is required and a completely randomized experimental design of the t treatments randomly allotted to the t EUs is used. If homogeneous blocks of size t are available to accommodate all t treatments, a randomized complete block design is used.

Grouping of experimental units into homogenous units removes the among block variability from experimental error and hence it increases the power of detecting difference between treatments due to small experimental error. In a RCBD or Latin square design the experimental units are grouped into blocks and every treatment occurs once in each block, and hence all treatments are compared with equal precision.

In case of large number of treatments it is not possible to place all treatments in a single block due to availability of insufficient homogenous experimental units. In choice of work on animals, it will be desirable to compare treatments within litters, but the litter size will depend on the particular species and will often be such that it is impossible to include all treatments within a litter. Similar consideration apply to greenhouse pot experiments where the block is restricted to the width of the bench; to experiment on plant virus diseases, where the block consists of a small number of leaves on each plant; to cookery experiments where only a limited number of stoves are available. Even if a sufficient number of homogenous experimental units are available to include all treatments inside a block, their use may make the cost of the experiment unaffordable. As generally it is true that the precision increases as

block size decreases, so smaller blocks are preferable over larger blocks. One way to reduce the block size and hence increase the precision of the experiment is to perform complete replicate of full set of factorial treatments into incomplete blocks. The technique of arranging complete replicate of factorial treatments into incomplete blocks of size smaller than the number of factorial treatment combinations in one replicate is called Confounding. As a result of such arrangements of the treatments into different incomplete block we have to sacrifice information on some of the treatment contrast(s) in order to increase precision on other treatments.

However incomplete Block Designs (IBDs) were developed for experiments in plant breeding, where it is desired to make all comparisons among pairs of treatments with equal precision. Consequently a different method for reducing the size of block is employed as compared to method of reduction of block size in confounding. An additional advantage of the use of IBD is the reduction in the experimental cost.

In IBD there are two types of error structure:

1. Error for intra block comparison
2. Error for inter block comparison

11.1.1. BALANCED INCOMPLETE BLOCK DESIGNS (BIBD)

When all treatment comparisons are equally important, the treatment combinations used in each block should be selected in a balanced manner, that is, any pair of treatment occurs together the same number of times as any other pair. Thus, a balanced incomplete block design is an incomplete block design in which any two treatments appear together an equal number of times. This property of balance ensures that all treatments will be compared with nearly the same precision even though the difference among the blocks may be large.

The parameters b , k , t , r , and λ , representing blocks, units in a block, treatments, replicates, and times treatments together in a block respectively, of a BIBD cannot be chosen arbitrarily. They must satisfy the relations: $bk = rt$, $r(k-1) = \lambda(t-1)$, $b \geq t$ and hence $r \geq k$. Even if the parameters satisfy these relations it is not always possible to arrange the treatments in blocks to get the corresponding design, i.e., the above relations are necessary for BIBD but not sufficient. If λ (the number of times (an integer) a treatment

occurs with each of the other treatments within an incomplete block is equal for all pair of treatments, the design is balanced and if $b = t$, the design is said to be symmetric BIBD.

Other relations:-

$$b_u = \frac{t!}{k!(t-k)!}, r_u = \frac{(t-1)!}{(k-1)!(t-k)!}, \lambda_u = \frac{(t-2)!}{(k-2)!(t-k)!}$$

For BIBD

$$b_u : r_u : \lambda_u$$

if λ is an integer then design is balanced

Example

How many blocks of size 3 and replicates are required for a BIBD of 9 treatments/varieties?

By using relation $b_u : r_u : \lambda_u = 84 : 28 : 7 = 12 : 4 : 1$ so 12 blocks (of size 3) and 4 replicates are required for BIBD

11.1.2. PARTIALLY BALANCED INCOMPLETE BLOCK DESIGNS (PBIBD)

BIBDs don't exist for all treatment combinations that the experimenter might wish to employ because of the constraint that λ be an integer can force either the number of blocks or the block size to be excessively large. For example, with eight treatments to be arranged in blocks of 3 plots each at least 56 blocks are required. Consequently for λ to be an integer every treatment has to be replicated at least 21 times. To overcome this difficulty, the experimenter can employ *Partially Balanced Incomplete Block Designs* (PBIBDs), with its use the number of replications for each treatment can be made much smaller as compared to a BIBD. In PBIBD, some pairs of treatments appear together λ_1 times, some pairs appear together λ_2 times and so on, and the remaining pairs appear together λ_m times. Pairs of treatments that appear together λ_i times are called i th-associates. The design is then said to have m associate classes.

PBIBD is a useful IB'D in situation where the experimenter requires greater precision for some specific treatment pairs than others. PBIBD are more difficult to analyze statistically as there are several different SEs.

11.2. BALANCED LATTICE DESIGN

A BIBD with k^2 treatments arranged in $(k+1)$ replicate with k blocks per replicate and k runs per block is called balanced lattice design. The special feature of balanced lattice as distinguished from other lattice is that every pair of treatment occurs once in the same incomplete blocks. Consequently all pair of treatments is compared with the same degree of precision. To obtain balance some restriction on number of treatments and number of blocks has to be placed. Balanced lattice design cannot be constructed for 36, 100, and 144 treatments.

As in balanced lattice design, every variety is compared with every other variety once and only once in the same incomplete blocks of size " k ". It is necessary to have $(k+1)$ replicates for k^2 varieties, $(k^2 + k + 1)$ replicates for k^3 varieties and so on.

11.2.1. ADVANTAGES OF LATTICE DESIGN

1. A large number of treatments may be compared within relatively small blocks, incomplete blocks.
2. They may be analyzed as RCBD or as a CRD depending upon whether or not the incomplete blocks are arranged in complete block. This procedure introduces a small bias in some designs and in some methods of analysis. For practical purposes, the amount of information from lattice is never less than for the comparable complete block design.

11.2.2. DISADVANTAGES OF LATTICE DESIGN

1. They are more complex computationally. This is especially true when missing plots occur or when covariance analysis is required.
2. Lattice designs are not available for all number of treatments.
3. The analysis becomes complex if the treatments are subject to different error variances.
4. The designs are more difficult to construct.

11.3. PARTIALLY BALANCED LATTICE DESIGN (DOUBLE, TRIPLE AND QUADRUPLE LATTICE DESIGN)

A lattice design with k^2 treatments arranged in rk blocks grouped into r replicates, where $r = 2, 3, 4$ is called a double, triple and quadruple Lattice design respectively. The advantage of simple Lattice is that it is available for 36, 100 and 144 number of treatments.

The double (simple) lattice is not appropriate for 9 and 16 treatments, as for such number of treatments the error degree of freedom is only 4 and 3 as compare to 5 and 16 degree of freedom for error in RCBD. The triple lattice is not appropriate for 9 treatments.

11.4. RANDOMIZATION

The randomization in IBD consists of three steps:

1. Use separate randomization scheme for blocks within each replicate
2. Use separate randomization scheme for allocation of treatments within each block
3. Allocate the treatment code to the treatments randomly.

11.5. RESOLVABLE DESIGN

An IBD in which each group of blocks contains each treatment exactly once is called a **RESOLVABLE DESIGN (RD)**. If incomplete blocks can be grouped in such a way that the group constitutes a complete replicate then IBD is called **RESOLVABLE DESIGN (RD)**. The advantage of RD is that if blocking is not effective in reducing experimental error then the design can be analyzed by considering it as RCBD. So Balanced Lattice is a resolvable design.

11.6. CHOICE AMONG DIFFERENT DESIGNS

When the number of treatments and the size of the blocks are fixed in advance by the conditions of the experiment then there is very little choice in selecting the appropriate IBD. However, in situations where the number of treatments and the block size can be varied then a design which can be arranged in separate replicates is preferable to one which cannot, and a balanced design is preferable to a partially balanced design.

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Rep	4	1074.325	1074.325	268.581	51.64	0.000
Treat	15	1388.750	1163.469	76.898	14.79	0.000
Block(Rep)	15	246.844	246.844	16.456 (E _b)	3.16	0.001
Error	45	234.031	234.031	5.201 (E _c)		
Total	70	2943.950				

Source	DF	SS	MS
Replications	4	1074.325	268.581
Blocks within Replications (Adj.)	15	246.844	16.456
Component B	15	246.844	16.456
Treatments (Unadj.)	15	1388.750	92.583
Intra Block Error	45	234.031	5.201
Randomized Complete Block Error	60	480.875	8.015
Total	79	2943.950	37.265

As $16.45 > 5.20$ so blocking is effective

*(Blocks within Replications (Adj.) - Intra Block Error) SS

The Efficiency of the Experiment Relative to Randomized Complete Block = 131.60

$A = 0.0428$ $MST_{(adj)} = 94.13$ $MSE' = 6.09$ $F = 15.46$

Variance of Means in Same Block 2.435989

LSD at .01 level 4.197811

LSD at .05 level 3.143543

Adjusted Treatment Means

Treatments	1	2	3	4	5	6	7	8
Adjusted Mean	25.677	21.119	25.742	26.300	31.828	20.357	19.956	23.381

Treatments	9	10	11	12	13	14	15	16
Adjusted Means	20.822	29.783	30.019	28.973	30.824	32.270	29.891	23.460

ANALYSIS OF PARTIALLY BALANCED LATTICE DESIGN:

$$SSR = \frac{1}{r} \sum R^2 - CF \quad SST = \frac{1}{r} \sum T^2 - CF \quad SSB = \frac{1}{br} \sum W^2 - CF$$

$$CF = \frac{(GT)^2}{(K^2 + K^2)}$$

S.O.V	DF	SS	MSS
Replication	k-1	SSR	L_{r-1}
Treatments (Unadj.)	k^2-1	SST	L_1
Block (Adj.)	r(k-1)	SSB	L_b
Intra-Block Error	(rk-k-1)(k-1)	SSE	L_e
Total	bk-1=rkk-1=rk^2-1		

1. Please remember that: If $E_b \leq E_e$ then blocking is not effective, i.e., there is no gain in precision due to blocking. In this situation ignore the restriction on blocking and analyze the data as RCBD considering replicates as blocks.

For RCBD, $SSE_{(RCBD)} = (SSB + SSE)$ $df_{RCBD} = df_{block} + df_{intra-block} = (k^2-1)(r-1)$

$$MSE_{(RCBD)} = SSE_{(RCBD)} / df_{RCBD}$$

2. If $E_b > E_e$, then blocking is effective, i.e., there is gain in precision due to blocking. In this case adjusted SST is computed for test of significance for treatment:

$$A = \frac{(E_1 - E_r)}{k(r-1)(E_b)}$$

$$SST_{(adj)} = SST - Ak(r-1) \left(\frac{r}{(r-1)(1+Ak)} SSB_{(unadj)} - SSB \right)$$

$$\text{where } SSB_{(unadj)} = \frac{1}{k} \sum \sum B^2_{ij} - CF - SSR$$

$$MST_{(adj)} = SST_{(adj)} / (1-1)$$

For test of significance, the effective MSE is

$$MSE = (1 + rkA / (k+1)) MSE$$

$$F = MST_{(adj)} / MSE \quad \text{with } (k^2 - 1), (k-1)(rk-k-1) \text{ df}$$

Pair wise comparison

$$\text{Adjusted treatment mean} = (\bar{Y}_j) = \frac{T_j}{r}$$

$$SE(\bar{Y}_j) = \sqrt{\frac{MSE}{r}} \quad SE(\bar{Y}_j - \bar{Y}_{j'}) = \sqrt{\frac{2(MSE)}{r}}$$

For more correct variances

$$\text{The variance of difference between treatments in the same block } V(\bar{d}) = \frac{2E'_r}{r} [1 + (r-1)A]$$

$$\text{The variance of difference between treatments in the different block } V(\bar{d}) = \frac{2E'_r}{r} [1 + rA]$$

Relative efficiency of LATTICE design as compare to RCBD

$$RE = \frac{MSE_{(RCBD)}}{MSE} \times 100$$

IF $RE > 100$, then LATTICE design is more precise than RCBD

12. GENERALIZED LATTICE DESIGN (OR ALPHA DESIGN)

12.1. INTRODUCTION

Patterson and Williams (1976) devised a new class of incomplete block designs called *Alpha Designs*. Alpha designs are in some respects a generalization of Dr. Frank Yates' original lattice designs. The main advantage of alpha designs is flexibility as they are available whenever the number of varieties V is a multiple of block size k , and they can be easily adapted even when it is not.

These early alpha designs were aimed primarily at controlling variation down the columns of plots in the field. This is often adequate when plots are long and narrow. Patterson and Hunter (1983) have demonstrated the value of alpha designs in such circumstances in terms of gain in efficiency.

Example

The genotypes are split into k sets of s genotypes per set. For 28 genotypes in blocks of four plots, $s = 7$ and we would consider the four sets of seven genotypes. For the first replicate we choose the seven blocks to be

- i. The first genotype of each set.
- ii. The second genotype of each set.
-
- iii. The seventh genotype of each set.

Block	Replicate 1			
	Set			
	1	2	3	4
I	1	8	15	22
II	2	9	16	23
III	3	10	17	24
IV	4	11	18	25
V	5	12	19	26
VI	6	13	20	27
VII	7	14	21	28
Array	0	0	0	0

Then, as for the second and subsequent replicates of a lattice design, we construct the blocks in each replicate so that

- i. Each block includes one genotype from each set, and
- ii. In each replicate the genotypes in each block have not occurred together in a block in any previous replicate.

Replicate II				
Block	Set			
	1	2	3	4
VIII	1	9	17	26
IX	2	10	18	27
X	3	11	19	28
XI	4	12	20	22
XII	5	13	21	23
XIII	6	14	15	24
XIV	7	8	16	25
Array	0	1	2	4

Set 1 is unchanged, set 2 is moved up one position, set 3 is moved up two positions and set 4 is moved four positions. The blocks for the second replicate are formed from the first, second, ..., seventh positions in the cycled sets.

Replicate III				
Block	Set			
	1	2	3	4
XV	1	11	21	27
XVI	2	12	15	28
XVII	3	13	16	22
XVIII	4	14	17	23
XIX	5	8	18	24
XX	6	9	19	25
XXI	7	10	20	26
Array	0	3	6	5

For the third replicate set 2 is moved up three places from its original position, set 3 six places and set 4 five places. The choice for the numbers of places by which each set is moved are made so that each pair of sets occurs in different relative positions in each replicate (e.g. for sets 3 and 4, in the first replicate genotype 15 occurs with genotype 22, in the second replicate with genotype 24 and in the third with genotype 26).

The Alpha designs are most effective when the block size, k , is less than the square root of the number of genotypes, g , and hence less than s . It is possible to define and construct Alpha designs with $k > s$ but these designs cannot avoid genotype pairs occurring together in a block in more than one replicate.

If the number of genotypes, g , is not an exact multiple of the block size, k , the method of construction of Alpha designs can still be used in a modified form. First we construct a design for the smallest value g^1 , which is greater than the number of genotypes and which is also an exact multiple of k ($s = g^1/k$). The surplus (dummy) genotypes are all included in the last set of s genotypes. This ensures that these unwanted genotypes will never occur together in a block in the constructed design. For the actual design we simply omit the surplus genotypes, with the result that in each replicate there is a mixture of block sizes, k and $(k-1)$.

Suppose that we want a design with 18 genotypes using four replicates with block size 4 (or smaller). We set up the initial array of four sets of five genotypes, with arrays (0, 1, 0, 0), (0, 1, 4, 2), (0, 2, 3, 4), and (0, 3, 2, 1).

Replicate I					Replicate II				
Set					Set				
Block	1	2	3	4	Block	1	2	3	4
I	1	6	11	16	VI	1	7	15	18
II	2	7	12	17	VII	2	8	11	19
III	3	8	13	18	VIII	3	9	12	20
IV	4	9	14	19	IX	4	10	13	16
V	5	10	15	20	X	5	6	14	17
Array	0	0	0	0		0	1	4	2

Replicate I					Replicate II				
Set					Set				
Block	1	2	3	4	Block	1	2	3	4
XI	1	8	14	20	XVI	1	9	13	17
XII	2	9	15	16	XVII	2	10	14	18
XIII	3	10	11	17	XVIII	3	6	15	19
XIV	4	6	12	18	XIX	4	7	11	20
XV	5	7	13	19	XX	5	8	12	16
Array	0	2	3	4		0	3	2	1

The resulting design after omitting the genotypes 19 and 20 is given below in an unrandomized form.

Replicate I					Replicate II				
Set					Set				
Block	1	2	3	4	Block	1	2	3	4
I	1	6	11	16	VI	1	7	15	18
II	2	7	12	17	VII	2	8	11	
III	3	8	13	18	VIII	3	9	12	
IV	4	9	14		IX	4	10	13	16
V	5	10	15		X	5	6	14	17
Array	0	0	0	0		0	1	4	2

Replicate III					Replicate IV				
Set					Set				
Block	1	2	3	4	Block	1	2	3	4
XI	1	8	14		XVI	1	9	13	17
XII	2	9	15	16	XVII	2	10	14	18
XIII	3	10	11	17	XVIII	3	6	15	
XIV	4	6	12	18	XIX	4	7	11	
XV	5	7	13		XX	5	8	12	16
Array	0	2	3	4		0	3	2	1

There are choices to be made in determining the patterns of rotations for the sets for each new replicate. Some rotation patterns will provide more efficient designs than others. Basic generating arrays, for producing the rotation patterns, are given by Patterson et al. (1978) for optimal Alpha designs. A computer program, ALPHA+ (Williams and John, 1993), is also available to provide optimal designs.

12.1.1. EFFICIENCY FACTORS

There is wide variation in quality between all possible alpha designs and care has to be taken in choosing one for a particular trial. There are many ways of measuring the quality of incomplete block designs but the measure of most relevance where the interest is in comparing all pairs of varieties in the efficiency factor (E) is defined as

$$E = \frac{\text{Average variance of difference in RCBD}}{\text{Average variance of difference in ICBD}}$$

Assuming the error mean squares are the same in both cases. The value of E lies between 0 and 1 and its difference from one measures the loss from confounding by using incomplete blocks. In practice one expects that this loss is more than compensated for by reductions in error means squares from using small blocks. Indeed, experience with cereal trials in the UK has shown net relative efficiencies averaging 1.4 to 1.5; in other words the benefit of alpha designs is roughly equivalent to having an extra replicate in a three replicate randomized complete block design.

Patterson and Williams (1976) used E as one criterion in building up a catalogue of high efficient alpha designs for up to 100 varieties and two to four replicates. They proved their worth in practice but by the early 1980's the demand for bigger designs, and the wider availability of microcomputers, lead to development of a computer algorithm which searches for and automatically generates alpha designs with high efficiency factors (Patterson and Patterson, 1984). To judge the effectiveness of the search, E is compared with theoretical upper bounds (Williams and John, 1993). The algorithm can normally produce an alpha design with 1% of an upper bound for any resolvable design with the same parameters.

12.2. ROW-COLUMN DESIGNS

Alpha designs are useful when plots are long and narrow. However when the plots are squarer in shape then designs which allow for both row and column variation can be more effective. Recent development in design construction (Nguyen and Williams, 1993) have shown how alpha designs can be used to produce efficient row-column designs, i.e., where the plots of each replicate are laid out in a rectangular array made up of S columns (blocks) and k rows (plots per block). Row-Column designs should be of particular benefit in situations where it is natural to organize plots within a replicate in a rectangular arrangement as is the case in many forestry variety trials.

The properties of row-column designs are determined by those of the two component block designs, one with blocks given by the rows and the other with blocks given by the columns, as well as by the way the components are put together. A study of the potential value of row-column designs in 60 barley variety trials with plot sizes in the range 4×0.6 , 5×1.2 , and 2×1.5 metres has been done by Robinson *et al* (1988). They found that the median estimated efficiency for row-column designs was 1.66 compared with 1.41 for column design.

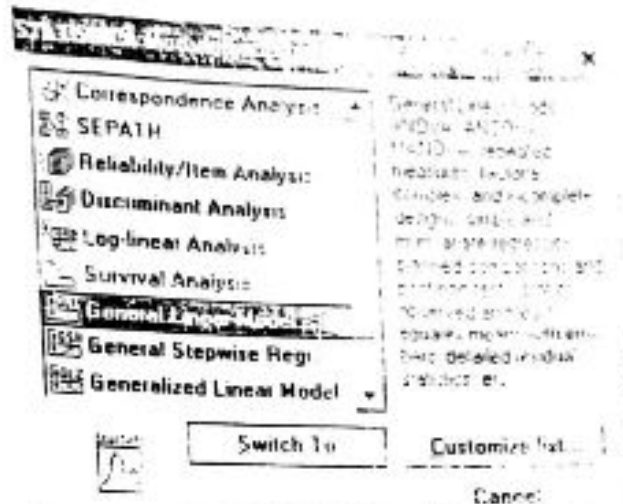
12.3. LATINIZED DESIGNS

Often in experimental layouts, replicates are placed next to each other so that the columns (blocks) of each replicate form long columns running down the replicates. Then there is a need to ensure that a variety occurs only once in each column. Latinized designs have this property and can be used in conjunction with alpha designs and row-column designs. An example of a situation where Latinized designs are useful has been given by Williams (1986).

13. PROCEDURE TO PERFORM DIFFERENT ANALYSIS OF EXPERIMENTAL DESIGN USING STATISTICAL ANALYSIS

13.1. CRD

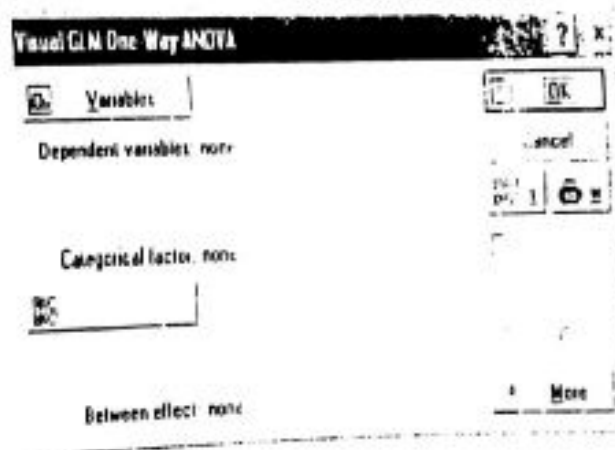
- Switch to General Linear Model (GLM)



- Open the file containing the data set (assumed that the data file already exist)
- Analysis > Resume Analysis > One-Way ANOVA



- Click OK Tab
- Click Variables Tab



- Enter Variable Number or Name of Dependent (or Response) Variable in the Dependent variable list.
- Enter Variable Number or Name of Categorical predictor (factor {treatment}) variable in the Categorical predictors (factors) list.

Select dependent variables and a categorical predictor

1-YLD 2-TREAT	1-YLD 2-TREAT	OK
		Cancel

Select All Spread Zoom Select All Spread Zoom

Dependent variable is: Categorical predictor (factor):

- Click OK twice, the following dialog box will appear

Visual GLM Results - Analysis 1

Summary | Comp | Means | Prof | Resids | Matr | Print

<input checked="" type="checkbox"/> All effects/G	<input checked="" type="checkbox"/> All effects
<input checked="" type="checkbox"/> Univar. results	<input checked="" type="checkbox"/> Cell statistics
<input checked="" type="checkbox"/> Between effects	<input checked="" type="checkbox"/> Whole model R
<input checked="" type="checkbox"/> Design terms	<input checked="" type="checkbox"/> Assumptions
<input checked="" type="checkbox"/> Coefficients	<input checked="" type="checkbox"/> Estimate
<input checked="" type="checkbox"/> Test	

Alpha values

Conf. limits: .950

Sign. level: .050

☒ Minimize ☒ Modify ☒ Close

☒ Update ☒ Auto hide

- Click Cell statistics Tab to see different statistics
- Click All effects/G Tab to see Error sums
- Click Univar. results to see ANOVA Table

13.2. RCB

- Switch to General Linear Model (GLM)
- Open the file containing the data set (assumed that: the data file already exist)
- Analyze > Resume Analysis > Main effects ANOVA
- Click QV Tab
- Click Variables Tab
- Enter Variable Number or Name of Dependent (or Response) Variable in the Dependent variable list.
- Enter Variable Numbers or Name of Categorical predictors (factors {block and treatment}) variables in the Categorical predictors (factors):

- Click Ok twice
- Click Cell statistics Tab to see different statistics
- Click All effects/G Tab to see Error bars
- Click Univar. results to see ANOVA Table

13.3. LATIN SQUARE DESIGN

- Switch to General Linear Model (GLM)
- Open the file containing the data set (assumed that the data file already exist)
- Analysis > Resume Analysis > Main effects ANOVA
- Click OK Tab
- Click Variables Tab
- Enter Variable Number or Name of Dependent (or Response) Variable in the Dependent variable list:
- Enter Variable Numbers or Name of Categorical predictors (factors) (row, column and treatment) variables in the Categorical predictors (factors):
- Click Ok twice
- Click Cell statistics Tab to see different statistics
- Click All effects/G Tab to see Error bars
- Click Univar. results to see ANOVA Table

13.4. FACTORIAL EXPERIMENT UNDER CRD

- Switch to General Linear Model (GLM)
- Open the file containing the data set (assumed that the data file already exist)
- Analysis > Resume Analysis > Factorial ANOVA
- Click OK Tab
- Click Variables Tab
- Enter Variable Number or Name of Dependent (or Response) Variable in the Dependent variable list:
- Enter Variable Numbers or Name of Categorical predictors (factors) variables in the Categorical predictors (factors):
- Click Ok twice
- Click Cell statistics Tab to see different statistics
- Click All effects/G Tab to see Error bars
- Click Univar. results to see ANOVA Table

13.5. FACTORIAL EXPERIMENT UNDER RCBD

- Switch to General Linear Model (GLM)
- Open the file containing the data set (assumed that the data file already exist)
- Analysis > Resume Analysis > Factorial ANOVA
- Click OK Tab
- Click Variables Tab
- Enter Variable Number or Name of Dependent (or Response) Variable in the Dependent variable list:

Categorical predictors (factors):

- Click Ok
- Click More option
- Click Exit to syntax editor option
- Modified the DESIGN statement as
DESIGN=BLOCK+A|B
Note that A|B is the abbreviated form of A+B+A*B
- Click OK (Run)
- Click Cell statistics Tab to see different statistics
- Click All effects/G Tab to see Error bars
- Click Univar. results to see ANOVA Table

13.6. SPLIT PLOT EXPERIMENT UNDER CRD

- Switch to General Linear Model (GLM)
- Open the file containing the data set (assumed that: the data file already exist)
- Analysis > Resume Analysis > Nested design ANOVA
- Click OK Tab
- Click Variables Tab
- Enter Variable Number or Name of Dependent (or Response) Variable in the Dependent variable list:
- Enter Variable Numbers or Name of Categorical predictors (factors) variables in the Categorical predictors (factors):
- Click Ok
- Click More option
- Click Random factors option
- Select **Replicate** as random factor
- Click Ok
- Click Exit to syntax editor option
- Modified the DESIGN statement as
DESIGN=A+REPLICATE(A)+B+A*B
- Click OK (Run)
- Click Cell statistics Tab to see different statistics
- Click All effects/G Tab to see Error bars
- Click Univar. results to see ANOVA Table

13.7. SPLIT PLOT EXPERIMENT UNDER RCBD

- Switch to General Linear Model (GLM)
- Open the file containing the data set (assumed that: the data file already exist)
- Analysis > Resume Analysis > Factorial ANOVA
- Click OK Tab
- Click Variables Tab
- Enter Variable Number or Name of Dependent (or Response) Variable in the Dependent variable list:

- Enter Variable Numbers or Name of Categorical predictors (factors):
- Click Ok
- Click More option
- Click Random factors: option
- Click **Block** as random factor
- Click Ok
- Click Exit to syntax editor option
- Modified the DESIGN statement as
DESIGN=BLOCK+A+BLOCK*A+B+A*B
- Click OK (Run)
- Click Cell statistics Tab to see different statistics
- Click All effects/G Tab to see Error bar
- Click Univar. results to see ANOVA Table

13.8. STRIP PLOT DESIGN

- Switch to General Linear Model (GLM)
- Open the file containing the data set (assumed that the data file already exists)
- Analysis > Resume Analysis > Factorial ANOVA
- Click OK Tab
- Click Variables Tab
- Enter Variable Number or Name of Dependent (or Response) Variable in the
Dependent variable list
- Enter Variable Numbers or Name of Categorical predictors (factors):
- Click Ok
- Click More option
- Click Random factors: option
- Click **Block** as random factor
- Click Ok
- Click Exit to syntax editor option
- Modified the DESIGN statement as
DESIGN=BLOCK+A+BLOCK*A+B+BLOCK*B+A*B
- Click OK (Run)
- Click Cell statistics Tab to see different statistics
- Click All effects/G Tab to see Error bar
- Click Univar. results to see ANOVA Table

13.9. NESTED DESIGN

- Switch to General Linear Model (GLM)
- Open the file containing the data set (assumed that the data file already exists)
- Analysis > Resume Analysis > Nested design ANOVA
- Click OK Tab
- Click Variables Tab

- Enter Variable Number or Name of **Dependent** (or Response) Variable in the Dependent variable list:
- Enter Variable Numbers or Name of Categorical predictors (factors) variables in the Categorical predictors (factors):
- Click Ok
- Click More option
- Click Random factors: option
- Click it as random factor
- Click Ok
- Click PK (Run)
- Click Sell statistics Tab to see different statistics
- Click All effects/G Tab to see Error bars
- Click Univar. results to see ANOVA Table

13.10. INFERENCES BEYOND ANOVA

13.10.1. MULTIPLE COMPARISON TESTS

- Click Comp. Tab
- Select the required factor to be tested in **Effect** dialog box
- Select Specify post-hocs for obs. means
- Select the appropriate Multiple Comparison Test

13.10.2. CONTRASTS

- Click Comp. Tab
- Select the required factor to be tested in **Effect** dialog box
- Select Specify contrasts for LS means
- Enter the coefficients of Contrasts in relevant Column

13.10.3. ORTHOGONAL POLYNOMIAL CONTRASTS

- Click Comp. Tab
- Select the required factor to be tested in **Effect** dialog box
- Select Specify contrasts for LS means
- Select polynomial Tab
- Select the desired degree of Polynomial