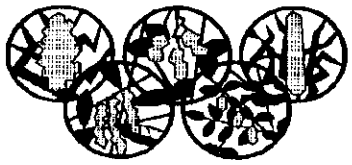
Research Planning and Data Handling

Compiled by

B. Diwakar and D.L. Oswalt



Skill Development Series no. 13

ICRISAT

Human Resource Development Program

International Crops Research Institute for the Semi-Arid Tropics
Patancheru, Andhra Pradesh 502 324, India

1992

Human Resource Development Program

Publications in this Skill Development Series (SDS) are issued semiformally for limited distribution to program participants, colleagues, and collaborators. Copies may be requested by the SDS number. Constructive criticism from readers is welcomed by: Program Leader, Human Resource Development Program, ICRISAT.

Acknowledgments

Information has been taken from published and unpublished reports. This publication should not be cited as a reference.

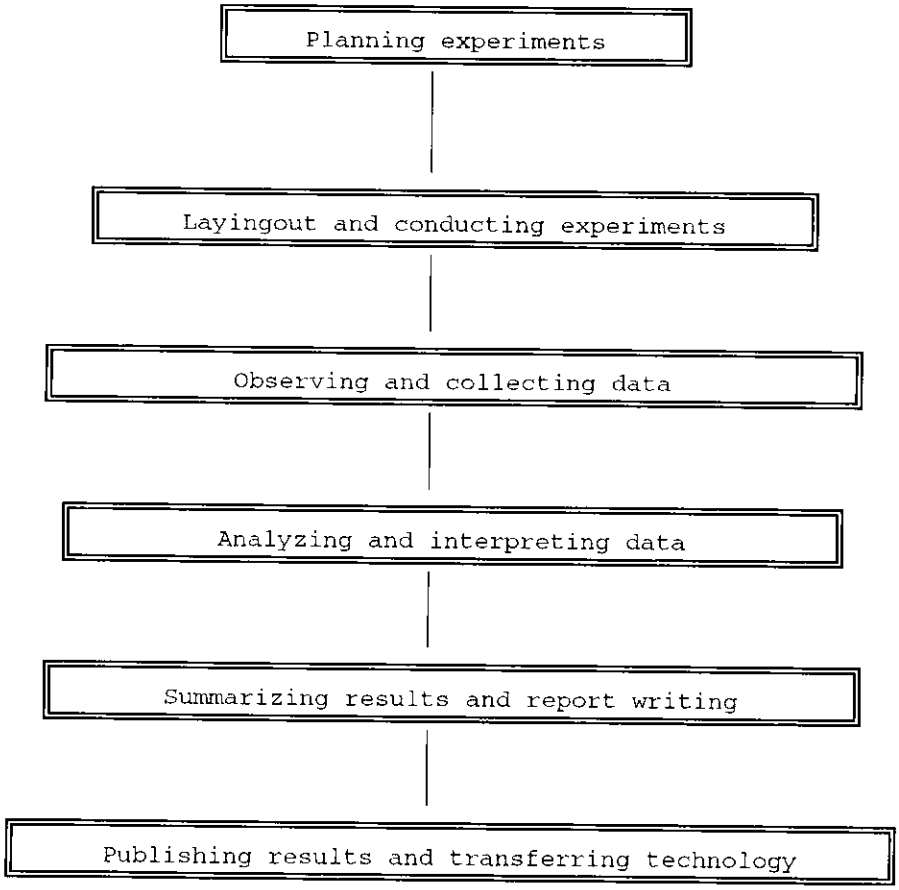
Comments, suggestions, and encouragement from Dr K.V.S. Rao and ICRISAT staff for compiling this document are gratefully acknowledged. Thanks to S.V. Prasad Rao, P. Chenchaiah, and program office staff for computerizing this manuscript.

Contents

Introduction

The objective of all agricultural research scientists is to organize research operations so that the findings can be used to improve agricultural productivity and sustainability. In agricultural research a scientist identifies solutions to problems through experimentation.

Research can be broadly defined as a systematic inquiry into a subject to identify and utilize new facts or principles. The procedure for research is generally known as the scientific method which, although difficult to define precisely, usually involves the following steps.



Planning Experiments

It is important to consider each det
following steps are involved:

- o Define the problem
- o Review literature
- o State objectives
- o Select treatments
- o Choose an experimental design
- o Determine the number of replications
- o List and procure the experimental materials
- o Consider data type and precision of experiment
- o Provide adequate facilities and funds
- o Determine procedure and potential for technology utilization

Refinement of Techniques and Selection of Material

Faulty techniques may increase experimental error and bias treatment effects. Potential pitfalls leading to faulty techniques need to be identified at the time of planning the experiment and managed appropriately (MP 1.)

A good technique should:

1. Control external influences so that all treatments are comparably affected.
2. Prevent gross errors.
3. Uniformly apply treatments.
4. Devise suitable and unbiased measurements of treatment influences.

For most applied research in agriculture, it is important to use the kinds of materials that will be used in actual production.

Stating the Objectives

The first activity in planning an experiment is to be clear and specific about the objectives of the experiment. For example, the problem may be one of assessing the value of weed control by a new technique. The questions involved in meeting the objectives could be:

a. Is weed control essential to increase the crop yields?
b. If weed control is useful, is hand weeding required or are machines or herbicides required to do the job efficiently?
c. If hand weeding is to be employed, how many weedings and at what stage(s) of the crop?
d. If herbicides are to be used, then what herbicide(s), at what rate, and when to apply?

These relevant issues must be settled before the experiment is initiated, to avoid the possibility that data from the experiment are found to be inadequate in scope. Therefore, it is essential that the experimenter precisely define the objectives specifying all details (MP 2).

Selection of Treatments

Careful selection of treatments is important in achieving the objectives and to increase the precision of the experiment. For example, while studying the effect of herbicides, fertilizers, fungicides, or insecticides, it is more useful to determine how the experimental units respond to increasing rates of treatment material, than to decide whether or not increasing rates are significantly different. Thus, a proper series of rates will make it possible to plan tests of significance that are more selective than merely comparing adjacent means in an array. The primary objective of such experiments should

be to determine a response curve with a range of rates going well beyond the optimum. This might appear to require a trial with a large number of treatments, but this need not be the case. Four or five well-spaced rates will suffice to determine the shape of the response curve, within the region of interest. The optimum rate of nitrogen is expected to be around 80 or 100 kg ha^{-1}, suitable treatments for a trial might be 0, 60, 120, and 180 or 0, 50, 100, 150, and 200 kg N ha^{-1}. Thus rates in equal or multiple increments within the expected range of the response are most efficient in establishing equations for a rate:response curve. Experiments where two or more types of treatments are tested simultaneously (factorial experiments) considerable improvement can be made in the precision of comparison of levels of each factor as well as for their interactions.

Before experimenting with a number of treatments in an elaborate trial, it is frequently advisable to try the treatments on a set of observation plots. Such preliminary trials often reveal the gross unsuitability of some treatments under field conditions or the possible difficulties that might be involved in the application of the treatments.

Choice of Experimental Design

While planning an experiment, the research worker should pay particular attention to ensure the adoption of an appropriate design. A standard design is always sound. The appropriate design depends largely on the number and nature of the proposed treatments. Thus, if many treatments are to be tested in a factorial scheme, it would generally be more efficient to adopt a confounded design than a simple randomized-block design. If many genotypes are to be tested it would be desirable to use an incomplete block design. In experiments involving a combination of treatments where one of the factors requires a large-size plot, like irrigation, it might be advisable to adopt a split plot design with irrigation treatments assigned to main plots. If both the factors need large plots for effective application of treatments, like irrigation and land configuration treatments, a strip plot design will be more efficient.

One needs to consider the available resources while choosing an experimental design. Complex designs (lattice designs) can be successfully utilized when skilled field assistants are available. In the absence of such help, mistakes might invalidate the analysis and make the results difficult, laborious or impossible. The randomized-block design possesses a remarkable simplicity and flexibility in its layout and statistical analysis. It should be preferred under circumstances where the successful employment of a more complex design is problematic owing to the lack of resources. Uniformity-trial data indicated that for a plot size of 0.004 ha to 0.008 ha, the randomized-block design may be adopted with up to 20 treatments without appreciable loss of efficiency (Panse and Sukhatme 1989).

A standard design is sound when the design chosen is appropriate for the experiment. Sometimes a standard design is arbitrarily modified to suit the special requirements of an experiment and then design mistakes are liable to occur in the process, resulting in faulty data or conclusions. Faults of the design may be discovered when it is too late, either after the experiment is started or even when the data are to be analyzed. A practice which safeguards against the use of a faulty design is to calculate a skeleton analysis of variance for the proposed experimental design. This will help to reveal and correct potential defects in the design, and thus avoid unforeseen difficulties in the analysis and interpretation of experimental data.

Number of Replications

Replications (blocks) refer to the number of sets of treatments. Replications will provide a measure of the validity of conclusions drawn from an experiment.

Fewer replications are needed in a homogenous field. The greater the number of replications, the lesser the chance errors. To determine the number of replications (MP 3) for an experiment consider:

- The inherent variability of the experimental materials.
- The experimental design.
- The number of treatments.
- The degree of precision desired.

Layingout and Conducting Experiments

Selection of an Experimental Site

Adjacent plots, sown simultaneously with the same variety and treated uniformly, will differ in all characters measured quantitatively. The causes for these differences are numerous, but the most obvious and probably the most important, are soil heterogeneity, history, and management.

The major features that magnify differences due to soil heterogeneity are (Gomez and Gomez 1984):

a) **Slopes.** Soil nutrients are soluble in water and tend to move to lower areas. Thus fertility gradients are more pronounced in sloping areas. Therefore, an ideal experimental site is one that has no slope. However, an area with a uniform and gentle slope having predictable fertility gradients can be preferred if a leveled area is not available.

b) **Field history.** Areas that were sown to different crops, with different fertilizer levels, various management practices, fallow areas, alleys, etc., are sources of additional soil heterogeneity. It is necessary to standardize the field to obtain uniform fertility by sowing the area with a cover crop, for at least one season, and to equalize the nutrients before conducting an experiment. Alternatively, the fertility gradient can be determined by uniformity trials and proper block arrangement avoids some delay in use of field plots (MP 4).

c) **Graded areas.** To reduce the unevenness in the field or to layout the field for proper drainage and uniform irrigation, grading is done by removing the top soil from elevated areas and spreading it in the lower areas of the field. This operation results in an uneven depth of surface soil and at times exposes the subsoil. Such a field should be avoided. If this is not possible, the pattern of soil heterogeneity should be assessed through uniformity trials so that blocking for uniform soil conditions can be accomplished.

d) **The presence of large trees, buildings, and other structures.** These influence the surrounding areas by shading of neighboring cropped areas and root competition by trees. Therefore, avoid these areas for experimentation.

e) **Unproductive site.** One should be able to grow a good crop on the area before one can plan a successful experiment. A field with very poor or problem soils should not be selected unless it is an experiment designed specifically to evaluate such conditions.

Measuring Soil Heterogeneity

The soil heterogeneity in a field can be measured and plotted by conducting a uniformity trial (MP 4) A uniformity trial is based on the premise that uniform soil, when cropped uniformly, will produce a uniform crop and the measurable differences in crop performance if present are due to the soil heterogeneity.

Competition effects

The interdependence of adjacent plants or plots for solar energy, soil nutrients, moisture, sprays, mechanical injury, etc., is commonly referred to as competition effects. Competition effects between plants within a plot should be kept uniform to:

a. ensure that the plant response really represents the conditions being tested
b. reduce experimental and sampling errors, both of which are likely to increase with variation among plants within a plot.

Some types of competition effects are:

a. **Varietal competition.** Different varieties sown in adjacent plots will be subjected to different environments depending upon their location relative to adjacent plots. The plants near the perimeter generally experience the effects of varietal competition. Tall or tillering genotypes compete more than short and nontillering types.

b. **Fertilizer competition.** When adjacent plots received different rates of fertilizers, plants that receive a higher rate tend to be more vigorous and more competitive. Secondly, seepage of water from fertilized plots may spread fertilizer to the root zone of the adjacent plot. The net advantage usually favors the plot receiving smaller applications of fertilizers.

c. **Pathways or alleys**: Plants adjacent to pathways or alleys have less competition and more space to grow than those in the center of the plot.

d. **Missing hills or plants in the row.** Even the most careful researcher cannot be assured of a complete stand for all plots in an experiment. Poor germination, insect or disease damage, physical mutilation, and so on, may cause death of a hill or a few plants in a row. The plants surrounding a space in the row are exposed to less competition than the other plants in the row.

Some methods for controlling competition effects are given in MP 5.

Plot layout

One of the most common errors committed in field experiments is in measuring plot dimensions. For plots with a length of 10 m, an error of 0.1 m or 0.5 m is not visible. But because even a small error in plot dimension can greatly affect the experimental results, it is important to double-check plot dimensions as plots are laidout in the field.

Confusion sometimes occurs in making decisions regarding the borders of a plot when crops are sown in rows. For example, in a 4-row plot with 75 cm space between rows, the width of the plot is 3 m, with the starting and ending points as shown in Fig. 1.

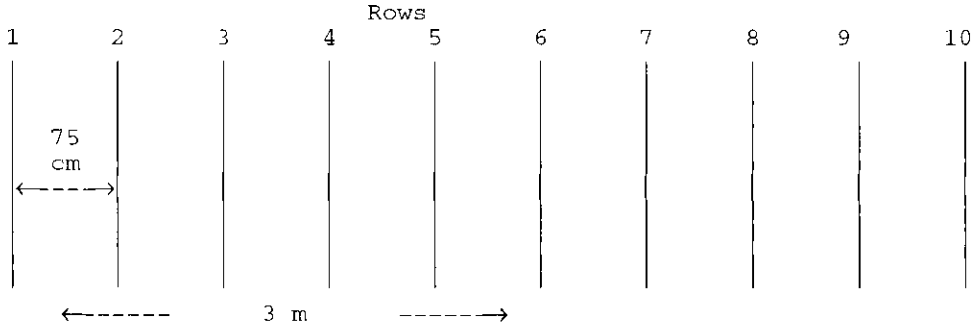


Figure 1. Boundaries of plot width in a 4-row plot with 75 cm space between rows.

Please note that the plot width is not measured from the first to the last row of the plot. Instead, the width of the plot includes half of the space between rows on each side of the plot.

Errors in row spacing are reflected as errors in plot measurements. For example, a crop was sown with rows 75 cm apart. A plot size of four-5 m rows would measure 3 m × 5 m. If the spacing between rows in one plot was not correct (80 cm), then a plot size of 3.2 × 5.0 m would result. In such a case, yield determinations would significantly be affected by the difference in the sown area.

If anomalies in row spacing are not identified and rectified, one of the following alternatives need to be chosen:

a. Harvest only rows which are rightly spaced.
b. If the discrepancy in row spacing is not large, harvest the whole plot and mathematically adjust the plot yield considering the area of normal and anomalous plot.

If a mathematical correction is to be made, the corrected plot yield is computed (Gomez and Gomez 1984) as:

$$Y = \frac{a}{b} \times y$$

Where: a = intended plot area, b = actual plot area, Y = corrected plot yield, y = noncorrected plot yield

Labeling

Mistakes in plot labeling can occur during seed or fertilizer packeting and plot marking for treatment. A wrong label can be on the seed packet or attached to the plot. Such errors are easily detected and rectified by a shrewd experimenter.

Conducting the experiment

A number of precautions are necessary during the progress of the experiment. The principle objective is to provide uniform conditions to all plots. The lack of uniform conditions among similar experimental plots contribute to the experimental error.

a) **Sowing.** Oversowing seed by 25%–50% may be required to ensure that enough seedlings emerge to establish the required plant population. It is important that the same seed lots are uniformly used in all similar plots or treatments. The seeds should be sown on the same day completing all of one replication at a time. To ensure uniform germination of seeds, compact the soil around the seed immediately after sowing.

b) **Thinning and gap filling.** Thinning should be done within 6 to 10 days after seedling emergence to avoid unevenness in the growth of the crop. Care must be taken, while removing the excess seedlings, to avoid disturbing the remaining plants. Seedlings should be left as equally spaced as possible. A 2 m stick can be placed along the row and the seedlings in excess of the calculated number can be removed. The most vigorous seedlings of similar size are to be retained. Diseased or damaged and weak seedlings may be removed while thinning. To reduce bias the same person should thin all plots in a replication and should finish the whole replication within a day. Transplanting seedlings in vacant places is not recommended because the transplants are usually weaker and less productive than the normal plants. A researcher has two alternatives in dealing with missing plants. One is to manipulate the thinning process so that the areas adjacent to the vacant space have more plants or to ignore the gaps at the seedling stage and adjust the number of plants at harvest (MP 5), or adjust plot size at harvest (Refer above).

c. **Fertilizer application.** The fertilizers must be applied uniformly in the experimental area. When applied by hand, subdivide the experimental area into smaller units to enable a more uniform application to smaller areas. This can be done by fertilizing each row of a plot separately. The usual procedure is to weigh or measure the fertilizers required row^{-1}, and uniformly spread it on each row by hand or with a calibrated machine.

d. **Cultural operations.** All cultural operations (plowing, weeding, earthing-up and plant protection) should be done uniformly for all plots, one replication at a time, unless they are treatments themselves. Otherwise, the experimental error is inflated due to differences in cultural operations among the blocks (replications). The experiment must be protected from damage by animals, birds, trespassers, etc. Damaged or disturbed plots will be rejected and such plots are considered as missing plots for data analysis (MP 17).

e. **Off-type plants and volunteer plants.** A month after sowing, some exceptionally tall or vigorous plants may be detected. These off-type plants may be from the seeds of a previous crop or mutants. These plants cannot be treated as normal plants. Secondly, these plants cannot simply be ignored as they already have affected surrounding plants. Their removal creates missing hills or gaps in the row and thus affect the surrounding plants. Therefore, they are normally allowed to mature. Just before harvest, such plants are counted and removed. The yield of the plot is then computed as

$$Y = \frac{(a + b)}{a} \times y$$

Where: Y = corrected plot yield
y = actual grain mass from normal plants in the plot
b = number of off-types

This correction assumes that the competition effects provided by the off-types to the surrounding plants are similar to those of the normal plants (Gomez and Gomez, 1984).

f. **Errors in data collection.** It is a good practice to review the data collected, immediately after completing the measurement, to detect any unusually high or unusually low readings. The doubtful figures need to be checked by again evaluating the specific plot. In determining the data, the produce should be kept until the analysis of data are completed. Another source of error is the differential effect among individuals taking the measurements. The number of persons involved in the data collection must be kept at minimum and the same person should evaluate all plots in one replication. It is important to double-check row dimensions and plot size as plots are harvested and avoid mislabeling data by counter-checking as both can reduce error.

g. **Field book and observation notes.** The field book should be so designed that all data are entered directly, by pencil, so that all statistical analysis of the data can be calculated directly from it or only one data transcription is needed for all analyses. This will save time, money, and the chance for errors.

To properly assess experimental data, chronological notes should be recorded in the field book related to emergence, growth of the crop, incidence of pests and diseases, differential behavior of treatments, abnormal events, rainfall data, labor utilization, scheduled events, etc. It should also contain the field plan.

Observing and Collecting Data

Sampling

In agricultural experiments the principal character studied is the yield or productivity. A variety of other observations such as rate of growth, flowering date, damage by pests and diseases, etc. are also made. In recording these ancillary observations it is generally necessary to resort to sampling since it is not practicable to take observations on every plant in each experimental plot.

Sampling is a procedure for selecting a fraction of a population to estimate the total population. For example, measure plant height of 10 of the 200 plants in the plot or for tillering count 1 m^2 of the 15 m^2 plot. An appropriate sample provides an estimate, or a sample value, that is close to the value that would have been obtained had all plants in the plot been measured. The difference between a sample value and its plot value constitutes a sampling error. Thus a good sampling technique would give small sampling errors (MP 6).

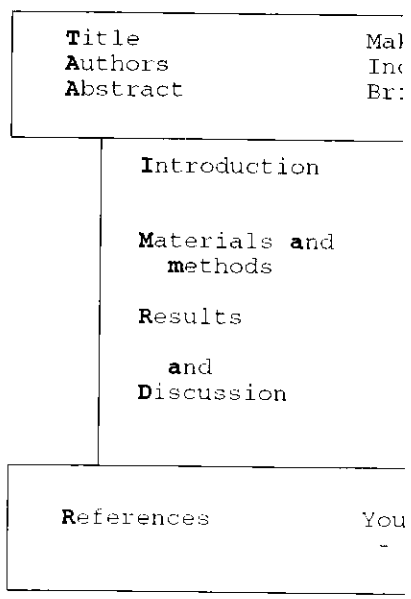
Recording Observations

Basically the observations to be recorded (MP 7) depend on the evaluation of treatment effects on the growth factors (MP 8), yield components (MP 9), and yield. The plans must include a schedule for data collection that provides the facility to obtain timely and unbiased evaluation for each identified observation.

Publication of research
disclosing sufficient informat
assess the observations, to re
intellectual processes.

Organization of a Scientific P

A scientific paper is primaril
organization of facts is the k
highly stylized with distinct
paper should have in order, it
and Discussion, and Conclusion
and there are a number of peri
and evidence. A format is sho



The other principal ingre
appropriate language that commu
distinct meaning. The best Eng

Choosing the Journal

The choice depends on the natur
journals which publish in that
refresh your memory is to scan
usually easy to determine, on t
might publish papers in your fi

Analyzing and Interpreting Data

Data Analysis

The data collected in an experiment on plant characters (variables) help in evaluating the effect of treatments by analyzing the variability between treatments based on the adopted experimental design.

Various forms of plot arrangements to accommodate the requirements of particular problems have been evolved and are known as experimental designs. The underlying principle of all these designs is the same, namely that they seek to provide by means of randomization and replication an unbiased comparison of treatments against their standard errors and aim at reducing these errors with the help of replication and management (MP 10). Potential designs are; 't' – test of significance (MP 11), completely randomized block (MP 12), randomized complete block (MP 13), latin square (MP 14), or factorial experiments using a randomized-complete block design (MP 15).

Problem Data

Before analyzing the data for an experimental design, one should make sure that the data follow normal distributions. If the data do not follow a normal distribution, then direct analysis of data are not valid as the tests of significance used to identify the significance between the treatments are based on the assumptions about the data. If the data do not conform to these assumptions, such an analysis may cause researchers to reach conclusions that are not justified. They may also overlook important conclusions that would be reached if the data were properly analyzed.

The assumptions on which an analysis of variance is based are:

1. The error terms are randomly, independently, and normally distributed.
2. The variances of different samples are homogeneous.
3. Variances and means of different samples are not correlated.
4. The main effects are additive.

Any drastic departure from one or more of the above assumptions must be corrected by appropriate data transformation (MP 16) before the analysis of variance is applied.

Missing Plots

It will be observed that in all experimental designs, every treatment is repeated the same number of times (usually once) in each replication. Nevertheless, the observations for plots are lost or so affected by some extraneous cause that it would not be proper to regard them as normal experimental observations. Such plot data that must be omitted from the analysis are 'missing plots'. Rejection of values should never be due to appearance (too high or too low), but only when evidence shows that the values are affected by some accidental or external factor(s).

The statistical analysis of incomplete observations when one or more plots are missing is necessarily complicated owing to the disturbance in the initially symmetrical distribution of plots. The arithmetical procedures for a randomized complete-block design and for latin square design are shown in MP 17.

Publication of research results, as a scientific paper, helps in disclosing sufficient information on the research study to enable peers to assess the observations, to repeat the study, and to evaluate scientific and intellectual processes.

Organization of a Scientific Paper

A scientific paper is primarily an exercise in organization, and a good organization of facts is the key to good writing. A scientific paper is highly stylized with distinct and clearly evident components. Each scientific paper should have in order, its Introduction, Material and Methods, Results and Discussion, and Conclusions. This is only the core of the paper, however, and there are a number of peripheral items that provide additional information and evidence. A format is shown below (Farr 1985).

Title	Make it brief and suitable for indexing
Authors	Include full addresses
Abstract	Briefly describe the problem and its solution
Introduction	What is the problem? - Define your parameters
Materials **a**nd **m**ethods	How did you do it? - Enable others to copy
Results **a**nd **D**iscussion	What did you find? - Present representative data What does it all mean? - Discuss your results; don't rearrange them.
References	Your authority for statements made - make sure that they are accurate and with complete details.

The other principal ingredient of a scientific paper is the use of appropriate language that communicates effectively, clearly, and in words of distinct meaning. The best English is that which conveys most in least words.

Choosing the Journal

The choice depends on the nature of your work; you must identify those journals which publish in that subject area. A good way to get started or to refresh your memory is to scan a recent issue of ***Current Contents***. It is usually easy to determine, on the basis of journal titles, which journals might publish papers in your field.

Summarizing Results and Report Writing

Reporting results is an essential part of the research process. The scientist must not only 'do' science but also must 'write' science. The research scientist must provide a written document showing what was done, why it was done, how it was done, and what was the inference. Successes and failures must be described in the report in an honest and fully explanatory manner.

Parts of a Report

A report has the following parts.

a. Title page: Usually carries
   - title of report and subtitle
   - address of the individual and affiliated organization
   - study number and date of release of report.

b. Foreword: This introduces the reader with the purpose of the study and also answers, questions like what agency requested it? What methods were employed in the study? How the information from the study is useful and its utility.

c. Summary of findings: This enables a busy reader to know the results of the study quickly. The summary should be very brief and precise.

d. Recommendations and implications of the study.

e. Table of contents: This helps the reader to identify parts of the report.

f. Body of the report: This should be very well organized with a full sequence of the proceedings of the study.

g. Appendix: This provides information supplementary to the text at the end of the report.

Style of Writing

The value of the report depends on the style of writing. The three golden rules for writing reports are:

- o Use simple, everyday language.
- o Use complete, direct, and positive statements.
- o Be brief.

Good report writing depends on clear thinking. Therefore, spend time in evolving the logic, sequence, and key points of the study before starting the study and before starting to write. Good report planning requires assessment of the audience and decisions about what to include and what to omit to convey the messages from the research study (Linton 1954, and Ward 1959).

Publishing Results and Technology Transfer

A scientific experiment, no matter how spectacular the results, is not complete until the results are published or utilized. In fact, the corner stone of the philosophy of science is based on the fundamental assumption that original result must be published, only then can new scientific knowledge be authenticated and added to the existing database. Whether or not one wholly subscribes to the 'publish or perish' adage, there is no question that the goal of scientific research is publication. Scientists are measured, and recognized (or remain unknown), by their publications.

MP 1. Preresearch Review and Management

Factors confounding response estimations related to the objectives that could alter research activities and data collection which should be considered while reviewing the literature and writing the materials and methods **before** initiating an experiment.

	Managed?		
A. Related to location	Yes	No	Solution
Direction of winds, storms, erosion, flooding			
Appropriate temperature and humidity range			
Appropriate air and potential for water drainage			
Light direction, light duration, and light quality			
Potential damage by people, animals, trees, roads, pollution			
Previous management history of area			
Representation of area for utilization of findings			
Accessibility for data collection and treatments			
Similar previous studies in the same location			

	Managed?		
B. Related to soil	Yes	No	Solution
Parent material variations and their limiting nutrients and texture relative to estimated responses, and surface configuration			
Profile water holding capacity, water drainage potential, air exchange rate, organic matter variations, pH change and variability in nutrient availability, mineralization rates, expected limitations in availability of each nutrient, leaching potential, drainage limitations, structure and texture variability, restrictions to root growth and development			
Representation of production area			
Protection from flooding or erosion			
Accessibility for management and observation			

C. **Related to resources**	Managed? Yes	No	Solution
Timely availability of water from rainfall or supplemental sources, power, technical staff, labor, transport, supplies, and financial support			
Availability of all essential nutrients and residual toxic ions from management within physiological development limits to assure accurate collection of data to estimate treatment responses			
Adequate manpower skilled in timely unbiased observation and management			
Techniques and tools to manage soil aeration, density of population, application of treatments, weeds, prevention of mechanical damage, spacing, spraying at varying stages of physiological development, insects, diseases, and other pests within the range of the canopy expressions			

D. **Related to seed**	Managed? Yes	No	Solution
Variations in source of seed lot for a genotype, seed size, quality, and age; related to depth of sowing, rate of sowing, and potential seedling vigor			
Procedures available to reduce variation related to variable placement to facilitate differences due to seed size, density, seed moisture content, soil moisture differences, due to duration of sowing operations, soil moisture contact with seed, human variation in skill, attitude, individual bias, and well-being related to sowing			
Adequate quantities of uniform quality in the pure and viable seed			

E. Related to equipment	Managed? Yes	No	Solution
Providing timely and uniform soil preparations			
Provision for variable depth and placement of seed with uniform contact of soil with seed			
Provision to apply fertilizers and sprays uniformly with variable plot rates in relation to soil surface area or plant surface area			
Provision of adequate harvesting, threshing, drying, and storage facilities to accommodate small and large plot produce			

F. Related to essential data	Managed? Yes	No	Solution
Identification of independent responses related to objectives			
Anticipated range in expected responses			
Precision required to indicate definitive differences			
Calculated number of replications and locations or seasons required to identify and validate critical treatment differences			
Criteria for data validation or cancellation			
Magnitude of data collection, storage, and processing as related to utilization of conclusions and their impact			
Determination of analytical techniques and procedures			

MP 2. Research Plan Checklist

Crop: Soil type:
Title: _____
Objective:
_____
Experimental design: _____
No.of replications: _____
Treatments:

Description	Symbol used

Rows plot^{-1}: _____ Length of row: _____ m Row spacing _____ cm
Gross area of plot: _____ m^2; Net area of plot to be harvested _____ m^2

Seed requirement:

Variety/ies						
100-seed mass (g)						
Desired population (000 ha^{-1})						
Seed rate kg ha^{-1}						
Seed rate g plot-1						
Seed rate g row^{-1}						
Amount of seed required for the experiment						

General application of fertilizer:

Rate of application ha^{-1}:

	N	P	K
Basal:			
Top dressing:			

	Name	Analysis (%)
Complex fertilizer (CF)		
Straight fertilizer (SFI)		
Straight fertilizer (SF2)		
Straight fertilizer (SF3)		

For basal application:

	CF	SF1	SF2	SF3
kg ha^{-1}				
g plot^{-1}				
g row^{-1}				

For top dressing:

	SF1	SF2	SF3
kg ha^{-1}			
g plot^{-1}			
g row^{-1}			

Fertilizer calculation for a fertilizer experiment:

Nitrogen (basal application):

Fertilizer: _____ Analysis: _____

	T1	T2	T3	T4	T5	T6
Rate of N ha^{-1}						
Fertilizer kg ha^{-1}						
Fertilizer g $plot^{-1}$						
Fertilizer g row^{-1}						

Phosphorus (basal application):

Fertilizer: _____ Analysis: _____

	T1	T2	T3	T4	T5	T6
Rate of P ha^{-1}						
Fertilizer kg ha^{-1}						
Fertilizer g $plot^{-1}$						
Fertilizer g row^{-1}						

Potassium (basal application):

Fertilizer: _____ Analysis: _____

	T1	T2	T3	T4	T5	T6
Rate of K ha^{-1}						
Fertilizer kg ha^{-1}						
Fertilizer g $plot^{-1}$						
Fertilizer g row^{-1}						

Nitrogen (top dressing):

Fertilizer: _____ Analysis: _____

	T1	T2	T3	T4	T5	T6
Rate of N ha^{-1}						
Fertilizer kg ha^{-1}						
Fertilizer g $plot^{-1}$						
Fertilizer g row^{-1}						

Seed treatment required (name of chemical, quantity, etc.):

_____

_____

Any special requirements like gypsum for groundnut, herbicides, etc.

_____

_____

Summary of supplies required (e.g., seed, fertilizers, chemicals, paper packets, seed packets, tags, pegs, etc.):

Item	Quantity	Item	Quantity

Experimental plan

Name: _____ Country: _____
Crop: _____ Soil type: _____

Title: _____
Date of sowing: _____

PLAN

R-VI						
R-V						
R-IV						
R-III						
R-II						
R-I						

Treatment 1: _____
Treatment 2: _____
Treatment 3: _____
Treatment 4: _____
Treatment 5: _____
Treatment 6: _____

Note: For making a detailed plan use a plan with row numbers.

MP 3. Identifying the Number of Replications

Two procedures for calculating the number of replications for an experiment conducted in a particular field are described below.

Procedure 1. This method takes into consideration, the amount of variability of experimental material and field (which is measured in terms of coefficient of variation (CV) and standard error of means (SEM)). To calculate 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 one regards as of practical importance, we can use the formula (Jeffers 1978):

$$N = (CV \div SEM)^2$$

Where N = No. of replications
CV = Coefficient of variation (%)
SEM = Standard error of mean (±)

Example: Coefficient of variation(%) = 12
Standard error of mean (±) = 6
No.of replications = $(12 \div 6)^2 = 4$

Procedure 2. If CV and SEM are not known, then the number of replications can be arrived at using the principle that the precision of treatment comparisons increases if the experimental error is kept to the minimum. The experimental error can be kept to the minimum by providing more degrees of freedom for the experimental error. In other words, a lower number of degrees of freedom for experimental error results in enlarged experimental error. Based on this principle, the number of degrees of freedom for error should not be less than 15 (not less than 10 in any case).

When 't' treatments are replicated 'r' times in a randomized-block design, the error is based on (t-1)(r-1) degrees of freedom which should not be less than 15.

No. of treatments	2	3	4	5	6
No. of replications	(10 to 15)	(6 to 9)	(4 to 6)	(4 to 5)	(4+)

Increasing either number of replications or plot size can improve precision, but the improvement achieved by doubling plot size is almost always less than the improvement achieved by doubling replications (Warren 1982).

There is currently ample room for argument about where to stop in increasing replications to avoid bias. While the identification of a stopping point seems to await further research, it seems unlikely that many researchers would elect to use more then 10 to 20 replications even if the efficiency is more than for fewer replications of larger plots. Eventually, the prospect of summarizing very large numbers of observations is apt to curtail the use of increased replications even when it is more efficient (Warren 1982).

MP 4. Conducting Uniformity Trials

Uniformity trials involve sowing an experimental site with a single crop variety and applying all cultural and management practices as uniformly as possible. The sown area is subdivided into small strips or plots of similar sizes referred to as basic units and the grain yield is recorded from the basic units. Yield differences between these basic units are taken as a measure of the area's soil heterogeneity. The smaller the basic unit, the more detailed is the measurement of soil heterogeneity.

Several types of analyses are available to evaluate the pattern of soil heterogeneity based on uniformity trials. The soil productivity contour map is a simple but informative presentation of the soil fertility variation of an area. The map (Fig. 1) describes graphically the productivity level of an experimental site at ICRISAT Center.

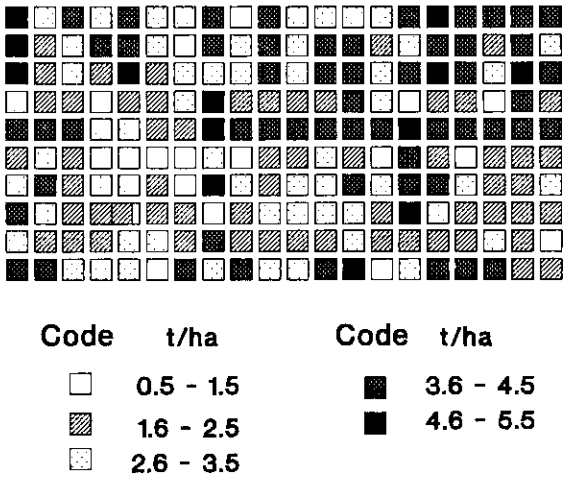


Figure 1. Map of responses.

Coping with soil heterogeneity

Three options that are commonly used are the proper choice of plot size and shape, block size and shape, and number of replications.

a. **Plot size and shape.** The difference in soil fertility between plots within a block is reflected in the experimental error. Thus the smaller this difference is, the smaller the experimental error. Hence, the choice of plot size and shape determines the differences in soil productivity from plot to plot within a block and consequently can reduce the experimental error. When large-size plots are used, higher costs are involved. Though soil variability effects are smaller, the precision decreases with an increase in plot size. Hence, the plot size that an experimenter should aim for is one that balances precision with cost and is commonly referred to as the optimal plot size.

In a uniformity trial conducted at ICRISAT Center, plots of 3m × 3m were harvested and the yields were recorded. The yield of individual 9 m^2 plots of pearl millet ranged from 0.5 to 5.5 t ha^{-1} (Fig. 1). To identify the optimal plot size from this uniformity trial, yield data were computed for different plot sizes by combining the basic 3m x 3m units. The yield data for various plot sizes were statistically analyzed. (Please note that for purposes of analysis of data, the plots were arbitrarily assigned to replications (horizontal strips) and treatments (vertical strips)). Analysis of variance of the data showed several alternate plot sizes, 6m × 3m or 3m × 6m or 9m × 6m, etc., with coefficients of variation (CV's) ranging from 7 to 14% (Table 1.).

Table 1. Significance of treatments/replications and coefficient of variation for different plot sizes.

Length		Width			
		3 m	6 m	9 m	12 m
3 m	a)[1]	Sig	NS	NS	Sig
	b)	Sig	Sig	Sig	Sig
	c)	14	9	10	4
6 m	a)	Sig	NS	Sig	Sig
	b)	Sig	Sig	Sig	Sig
	c)	11	7	5	3
9 m	a)	Sig	NS	Sig	Sig
	b)	Sig	Sig	Sig	Sig
	c)	9	4	4	3
12 m	a)	Sig	NS	Sig	Sig
	b)	Sig	Sig	Sig	Sig
	c)	9	4	3	3

1. a) Significance between treatments at $P = 0.05$
   b) Significance between replications at $P = 0.05$
   c) Coefficient of variation (%)

The conclusions drawn from this uniformity trial were (i) soil fertility in this field was highly sporadic with identified uniform areas - both high as well as low fertility as reflected in the yield (ii) the smaller the plot size, the higher was the CV%, (iii) 6m x 9 m plot was optimum.

As long as plot to plot interference can be avoided, reduced plot size can be translated into increased precision or accommodation of larger treatment sets. Even when crop, genotype, and region are kept relatively constant, there is ample evidence that no single plot size is universally best. There is always a distribution of plot sizes that are best or nearly best (Warren 1982).

Once the optimal plot size is determined, the choice of plot shape is governed by the following considerations:

- Long and narrow plots should be used for areas with distinct fertility gradient, with the length of the plot parallel to the fertility gradient of the field.
- Plot should be as square as possible whenever the fertility pattern of the area is spotty or not known, or when border effects are large. (Gomez and Gomez 1984).

The following facts need to be considered while choosing the plot size and shape.

- Small plots may yield undependable results. Small errors in small plots can often be greatly magnified when computed to hectare estimates.
- Unnecessarily large plots waste time and resources.
- Plots must be wide enough to permit more border rows when necessary, e.g., for destructive plant sampling.
- Plot size and shape are also influenced by the type of experiments, e.g., tillage experiments or irrigation experiments as compared to varietal or fertilizer experiments.

b. **Block size and shape.** An experimental unit refers to the unit of experimental material to which a treatment is applied. The term 'plot' is synonymous with experimental unit. A block refers to the strip of experimental field containing the different experimental units (plots). The primary purpose of blocking is to reduce experimental error by eliminating the contribution of known sources of variation among experimental units. This is done by grouping the experimental units into blocks such that variability within each block is minimized. Because, the variation within a block becomes a part of the experimental error. Block arrangement is most effective when the experimental area has a predictable pattern of variability, e.g., soil heterogeneity, direction of insect migration, slope of the field in water stress studies.

Block size is related to the plot size chosen, the number of treatments to be tested and the experimental design used.

A proper shaping of the block helps to reduce the productivity differences among plots within a block so that most of the soil variability is accounted for variability between blocks. In other words, the plots within a block will be sown on more homogenous area. The guidelines are:

- Orient the blocks to maximize the differences between them.
- If the fertility gradient is unidirectional, long and narrow blocks should be used so that their length is perpendicular to the fertility gradient (Fig. 2).

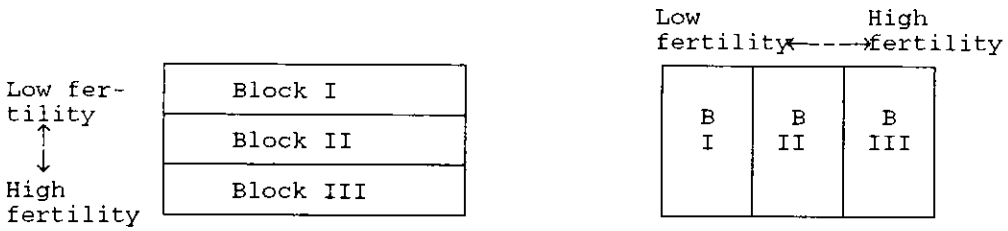


Figure 2. Determining the direction of blocks as per the soil fertility gradient.

- If the fertility gradient is in two directions in the field (by length as well as breadth), then blocks are to be considered in both the directions. This is achieved by following latin square design which accounts for the row variation (breadthwise of the field) and column variation (lengthwise of the field) (Fig. 3).

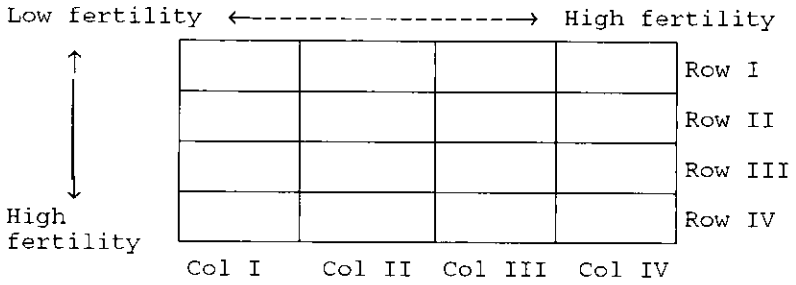


Figure 3. Identifying rows and columns for two-way soil fertility gradients.

- When the fertility gradient is not known or fertility differences are sporadic, use square blocks.

- If the number of treatments is very large and an uniform area within a block cannot be attained, an incomplete block design like lattice design may be used (Gomez and Gomez, 1984).

In the analysis of variance, a significant 'F' test for the variation between blocks indicates that the block arrangement in reducing the experimental error.

MP 5. Controlling Competition Effects

a. An obvious solution is to exclude the outside plants which have not been exposed to the complete surroundings from the plot measurements. The number of rows to be discarded on each side of the plot depends upon the size of competition effects. Generally, for the narrow spaced crops like rice, groundnut, pulse crops like mung bean, cowpea, etc., the two outer rows on either side of a plot are discarded for data collection. For crops like sorghum, pearl millet, maize, etc., where plant spacing is wider than 50 cm, discarding one outer row is usually sufficient.

b. Adjacent plots are sown to genotypes of fairly similar morphology or are subjected to similar fertilizer rates. In genotype trials, grouping genotypes fairly homogeneous in competition ability and using a group-balanced-block design can be followed. In fertilizer trials, treatments can be tested along with other factors in a factorial experiment. To minimize fertilizer competition, grouping plots having the same fertilizer rate as a main plot, in a split plot design, will help.

c. A stand correction for each crop using an appropriate correction factor could be calculated using a formula described by Gomez and Gomez (1984). For maize the factor is 0.6 assuming that the yield compensation will be about 60%. The following is the formula for stand correction in maize.

$$Wa = \left(1 + \frac{0.6m}{n}\right) Wf$$

Where Wa = adjusted panicle mass
Wf = panicle mass from n hills or plants
m = no. of missing hills or plants
n = no. of hills or plants harvested

Where there is a large variation in the yield advantage between genotypes and different management practices, the required correction factor is too large to be practical in the above procedure of mathematical correction (Gomez and Gomez 1984). Alternatively, discard all plants immediately adjacent to a missing hill or missing plants in the row and harvest only those that are naturally competitive. For computation of yield, only the area of harvested plants is to be used. This procedure is possible when a large number of plants row^{-1} or $plot^{-1}$ are available.

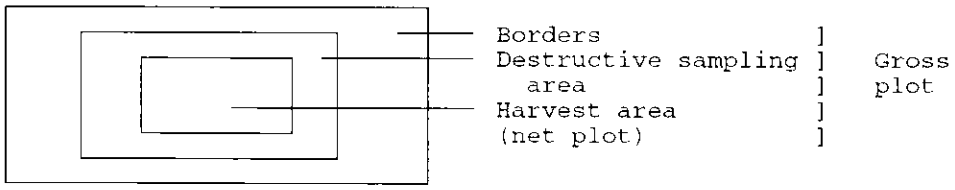
If the stand count is not 100% due to treatment effect, then the above stand correction adjustment should not be made.

MP 6. Sampling Methods

A sampling method specifies

a. the sampling unit — the unit upon which measurements are to be made
b. the method of selecting the sampling units from a plot
c. the sample size — the number of sampling units to be taken from each plot.

Sampling units are to be selected at random. When the sampling procedure requires the destruction of sample for measurements like leaf area, laboratory analysis for plant nutrient status, etc. separate the sample area from the rest of the plot. A common practice is to leave an area (2 to 3 rows of crop) in the center of the plot for harvesting and use the surrounding areas (rows of crop) for sampling.



The choice of sampling unit depends to some extent on considerations of convenience and the nature of crop. If plant growth is such that individual plants are distinct as for example pigeonpea, cotton, tobacco, etc., a plant may serve as the sampling unit. With crops like groundnut, pearl millet, chickpea, sorghum, etc., where the crop is grown in continuous rows (close spacing) or individual plants cannot be easily distinguished, a meter of the row or a suitable fraction may be taken as the sampling unit.

MP 7. Rounding and Reporting Numbers

Precision and accuracy are often used synonymously, but in a statistical sense, they are different. Precision refers to the magnitude of the difference between two treatments that an experiment is capable of detecting at a given level of significance. Accuracy refers to the closeness with which a particular measurement can be made.

At the time data are collected, they should be examined for out-of-line figures, and all such entries rechecked to prevent possible errors. There is enough variation in biological data without allowing more to creep-in through avoidable mistakes. Whenever possible, original records should be collected in an organized way to avoid recopying.

In taking measurements on experimental units it is seldom worthwhile to record figures to a number place less than one-fourth of the standard deviation (SD) unit^{-1}. If the SD is 6.96 kg unit^{-1}, then one-fourth of the SD is 6.96 ÷ 4 = 1.74. As the first place is in the one's position, data can be recorded to the closest kilogram. If SD were 2.5 kg unit^{-1}, 2.5 ÷ 4 = 0.625, the first place is the tenth position, and data could be recorded to the closest tenth of a kilogram.

The instrument used for weighing and measuring need not be more accurate than required by the precision of the experiment. For example, if a series of weighings are to be made and rounded off to the closest kilogram, the scale used can be in kilogram units rather than divisions of a kilogram.

It is not wrong to carry more digits than the variability of the data justify, and with modern data processing equipment this can be done easily, but in reporting results superfluous digits should be dropped and rounded to the place indicated by taking one-fourth of the standard error of a mean (SEM). If the SD unit^{-1} is 6.96 kg and each treatment mean is based on five replications, SEM = 6.96 ÷ 5 = 3.11 and 3.11 ÷ 4 = 0.68, indicating that means should be rounded to one decimal place. In doing an analysis of variance, it is best to carry the full number of figures obtained from the uncorrected sum of squares.

MP 8. Measurement of Growth Factors

a) **Plant height**. Before flowering, the height of the plant is measured from the surface of the soil (ground level) touching the stem to the base (or auricle) of the top most fully open leaf. After flowering, it is the distance from ground level to the tip of the inflorescence (panicle) or top leaf.

b) **Tiller number**. Tiller number is the number of tillers unit^{-1} area or the number of tillers plant^{-1}. At harvest, tillers can be separated into productive (having panicles) and nonproductive tillers (absence of panicles).

c) **Leaf area and leaf area index**. The leaf area refers to the leaf surface and the leaf area index (LAI) is the area of the leaf surface (unit area)$^{-1}$ of land surface.

These observations can be recorded from intact plants or after removing the plants from the field (destructive sampling). The leaf area is sometimes measured using a representative leaf amongst the leaves on the plant. For example, the 3rd leaf from the top was found to provide an estimate of the leaf area of the plant in case of sorghum and pearl millet.

After identifying the plants at random, the length and maximum width of each leaf or representative leaf are measured and the leaf area is computed on the length-width method.

Calculation of leaf area: $K \times L \times W$ where K is the "adjustment factor", L is length, and W is the width. The value of K varied with the shape of the leaf which in turn is affected by the variety, nutritional status, and growth stage of the leaf. The following values of K have been suggested.

Crop	K value	
Sorghum	0.747	(Stickler et al 1961)
Pearl millet	0.723	(Singh et al 1970)
Maize	0.733	(McKee 1964)

Leaf area index (LAI) is computed as:

$$LAI = \frac{\text{Total leaf area of n plants (cm}^2\text{)}}{\text{Area of land covered by n plants (cm}^2\text{)}}$$

d) Dry matter accumulation by plants. The plants identified at random from outside the net plot area are cut at the base at the time of sampling and the plants are chopped into small bits, dried in a glasshouse (not in the open sun) and then in a hot-air oven at 75°C to constant dry mass.

MP 9. Measuring Yield Components

Data on yield components help in understanding the productivity of a crop. Some important yield contributing characters are length and width of panicle, number of panicles, spikelets panicle^{-1}, percentage of filled spikelets, number of seeds panicle^{-1}, and seed mass (test weight). A good statistical relationship exists between yield components and yield.

Calculating crop yield based on yield contributing characters:

Sorghum:

Yield plant^{-1} = (panicles plant^{-1}) × (spikelets panicle^{-1})

× (no. of seeds spikelets^{-1}) × (seed mass)

Groundnut:

Yield plant^{-1} = (mature pods plant^{-1}) × (seeds pod^{-1}) × (kernel mass)

Pigeonpea or Chickpea:

Yield plant^{-1} = (pods plant^{-1}) x (seeds pod^{-1}) x (seed mass)

From the yield plant^{-1}, one can compute the yield ha^{-1} based on the number of plants harvested ha^{-1}.

Problem solving

What is the yield plant^{-1} and yield ha^{-1} of pearl millet given the following:

Spacing plant^{-1} = 75 cm x 10 cm

2 panicles plant^{-1}, 800 seeds panicle^{-1}, and 1.2 g (100 seeds)$^{-1}$.

Worksheet

Step 1. Yield plant^{-1} = 2 panicles plant^{-1} × 800 seeds panicle^{-1}

× 0.012 g seed^{-1} = 19.2 g plant^{-1}.

Step 2: Plant density ha^{-1} = 10 000 m^2 ÷ (0.75 × 0.10 m plant^{-1})
= 133 333 plants ha^{-1}

Step 3: Yield ha^{-1} = (Yield plant^{-1}) × (Plant density)

$$= \frac{19.2 \text{ g plant}^{-1} \times 133\ 333 \text{ plants ha}^{-1}}{1000 \text{ g kg}^{-1}} = 2560 \text{ kg ha}^{-1}.$$

Measuring yields

a) Grain yield. Grain yield refers to the mass of cleaned and dried grain to constant moisture (usually to 8 to 12% moisture) harvested from a unit area. The grain yield is usually expressed either in kg ha^{-1} or t ha^{-1}.

b) Stover yield. The harvested plants in each plot are bundled and dried in the field for about ten days to a constant moisture content.

Derived data

Examples of data are harvest index and shelling percentage.

$$\text{Harvest Index} = \frac{\text{Dry matter yield}}{\text{Grain yield}}$$

$$\text{Shelling (\%)} = \frac{\text{Mass of kernel} \times 100}{\text{Mass of pods}}$$

MP 10. Steps in the Analysis of Variances

Step 1. Tabulate the data, compute treatment totals, replication totals, grand total, grand mean, and treatment means.
Note: Grand total = Treatments total = Replications total

Step 2. Correction factor (CF) =
$(\text{Grand total})^2 \div$ No. of observations which gave the grand total

Step 3. a. Total sum of squares (SS) = $\Sigma(\text{Ind. data point})^2 - CF$

b. Treatment SS = $\dfrac{(\Sigma T_I)^2 + (\Sigma T_2)^2 + \ldots + (\Sigma T_i)^2}{\text{Replications (r)}} - CF$

c. Replication SS = $\dfrac{(\Sigma RI)^2 + (\Sigma RII)^2 + \ldots + (\Sigma Ri)^2}{\text{Treatments (t)}} - CF$

d. Error SS = Total SS - (Replication SS + Treatment SS)

Step 4. Replication variance = Rep. SS ÷ (r-1)
Treatment variance = Treat. SS ÷ (t-1)
Error variance = Error SS ÷ (r-1) (t-1)

Step 5. 'F' value for Reps. = Replication variance ÷ Error variance
'F' value for treats. = Treat. variance ÷ Error variance

Step 6. Compare calculated 'F' values with table 'F' values and conclude
F calculated > F table value - 'F' is significant.
F calculated < F table value - 'F' is not significant.

Step 7. If 'F' test is significant, compute critical difference (CD), or least significant difference (LSD).

Step 8. a. Standard error (SE) = $\sqrt{\text{Error variance}}$

b. Standard error of a mean (SEM) = $\dfrac{\sqrt{\text{Error variance}}}{\sqrt{\text{No. of observations which gave the mean value}}}$

c. Standard error of difference between two treatment means (SED) = SEM × $\sqrt{2}$ or

$$\frac{\sqrt{\text{Error of variance}} \times \sqrt{2}}{\sqrt{\text{No. of observations which gave the mean value}}}$$

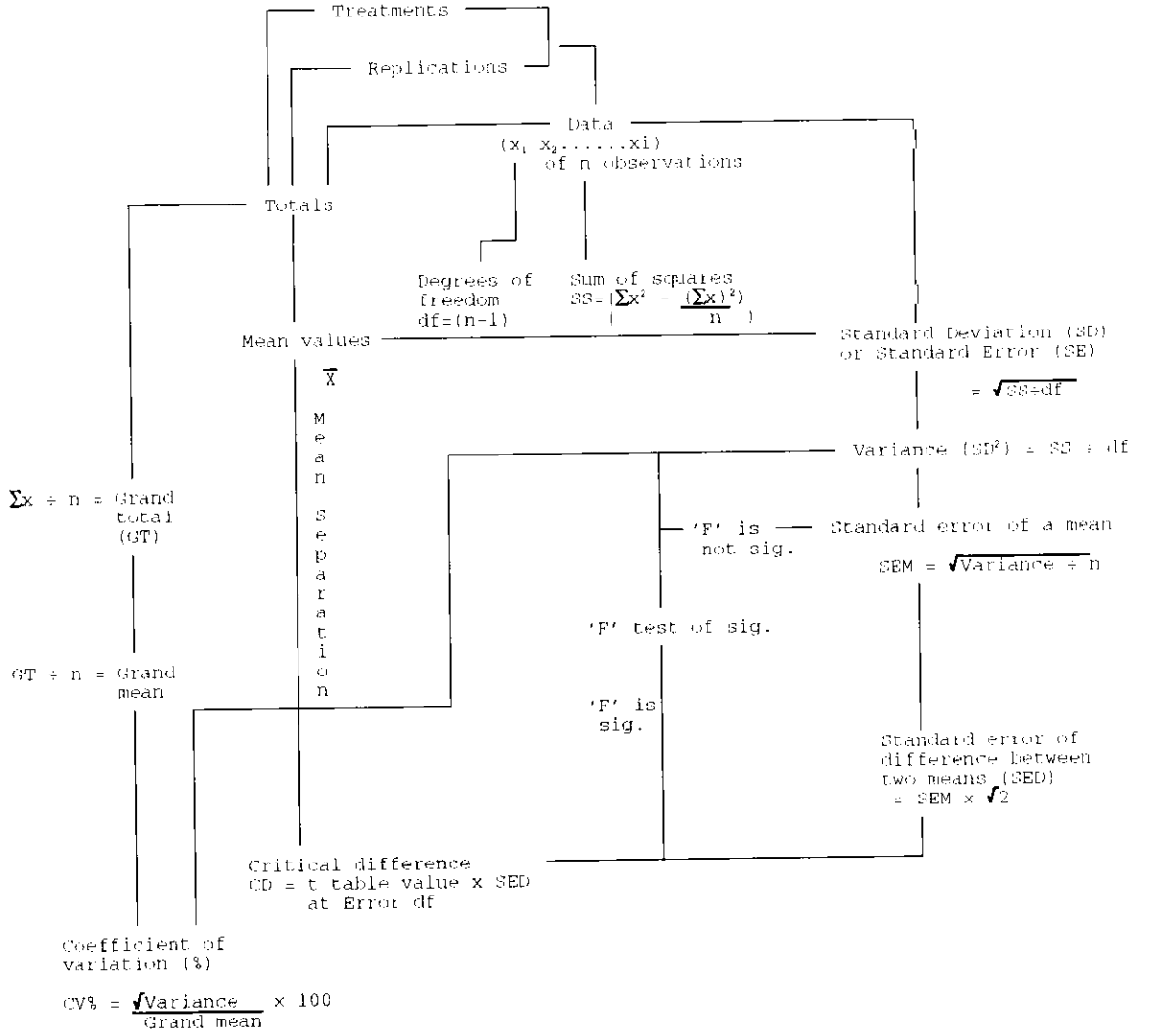
d. Critical difference (CD) =
't' value from 't' table × SED at error df for $P=0.05$

Step 9. Arrange the treatment means in descending order and rank the treatments using alphabetical notations comparing them with the CD value.
a. If the difference between any two treatment means is greater than the CD value, then the two treatment effects are different.

b. If the difference between any two treatment means is less than the CD value, then the two treatment effects are not different.

Step 10. Coefficient of variation (CV%) = $\dfrac{\sqrt{\text{Error variance}} \times 100}{\text{Grand mean}}$

Interrelationships Among Statistical Parameters



MP 11. 't' — Test of Significance

Problem

The following mean yields were recorded from two sorghum cultivars tested at ten locations in India. Find whether a measurable difference exists in the performance of these two cultivars;

a. assuming the samples are independent
b. assuming the samples could be paired.

	Locations									
Cultivars	1	2	3	4	5	6	7	8	9	10
	Mean yield (t ha^{-1})									
CSH 5	4.15	4.02	4.25	4.26	4.10	3.22	3.19	3.60	3.68	3.21
SPH 224	3.61	3.47	3.59	3.63	3.62	4.08	3.73	4.17	4.22	3.89

Worksheet

Location	Mean yield (t ha^{-1}) CSH 5 (x_1)	SPH 224 (x_2)	Difference (d) ($x_1 - x_2$)	d^2
1	4.15	3.61	0.54	0.2916
2	4.02	3.47	0.55	0.3025
3	4.25	3.59	0.66	0.4356
4	4.26	3.63	0.63	0.3969
5	4.10	3.62	0.48	0.2304
6	3.22	4.08	-0.86	0.7396
7	3.19	3.73	-0.54	0.2916
8	3.60	4.17	0.57	0.3249
9	3.68	4.22	-0.54	0.2916
10	3.21	3.89	-0.68	0.4624

	Independent 't' test		Paired 't' test
Sum of x	27.68	38.01	Σd = -0.33
Mean ($\Sigma x \div 10$)	3.768	3.801	d = 0.033
Sum of sq.	143.75	145.13	3.7671
CF = $(\Sigma x)^2 \div 10$	141.98	144.48	0.0109
$SD^2 = \frac{\text{Sum of sq.} - CF}{x-1}$	0.1966	0.0722	0.4173
$SD = \sqrt{SD^2}$	0.44	0.26	0.65
Pooled SD	0.35		
$SEM = SD^2 \div \sqrt{n}$	0.1106		0.206
$SED = SEM \times \sqrt{2}$	0.157	-	-
't' = $(x_1 - x_2) \div SED$	0.21		't' = d ÷ SEM 0.16
't' Table for 18 df at $P = 0.05$ =	2.10		't' Table for 9 df at $P = 0.05$ = 2.26
't' test	Not significant		Not significant
Conclusions:	CSH 5 and SPH 224 did not differ in their yield potential.		No yield differences between CSH 5 and SPH 224 tested over 10 locations.

MP 12. Completely Randomized-block Design

This design, the simplest type, is set-up by assigning treatments at random to a previously determined set of experimental units. The design is the most efficient in solutions in which there is little variability among the units associated with position in the experimental area, age, vigor, or other identifiable sources. It is flexible with regard to the physical arrangement of the experimental units, maximizes the degrees of freedom for estimating experimental error, and minimizes the 'F' value required for statistical significance. A disadvantage is that there are often identifiable sources of variation among the experimental units so that other designs, when skillfully employed, usually are capable of reducing the estimates of experimental error, which makes it possible to detect smaller, significantly different treatment effects. Hence this design is mostly used for pot culture or greenhouse experiments and rarely used for field experiments.

Layout. The randomization of treatments could be done by using random number tables or by drawing lots. The steps involved for randomization of 6 treatments with 4 replications are:

Step 1. Prepare 24 identical pieces of paper and divide them into 6 groups, each group with 4 pieces of paper. Label each piece of paper of the same group with the same letter corresponding to a treatment. Uniformly fold each of the 24 labeled pieces of paper, mix them thoroughly, and place them in a container. For this example, the treatments are labeled as A, B, C, D, E, and F.

Step 2. Draw one piece of paper at a time, without replacement and with constant shaking of the container after each draw to mix its content. Allot the treatments to plots as the label indicates after each draw. For example, with the first draw, if the label indicated is D, allot treatment D to the first plot. Similarly in the second draw, if the label indicated is A, assign treatment A to the second plot. This procedure is repeated till all the 24 plots are assigned with the treatments.

IV	E	B	D	C	B	F
III	B	D	A	F	C	E
II	C	A	D	C	E	F
Rep I	D	A	B	E	F	A

Figure 1. Complete randomization of treatments

Please note that treatment A has appeared twice in replication I and is not appearing in replication IV. This is allowed in this experimental design as it is assumed that the experimental units (plots) are homogeneous.

Problem: An experiment was conducted to identify the response of sorghum (CSH 9) to phosphorus applications in a Vertisol field using a completely randomized-block design with the treatments replicated 4 times.

Treatments:

P0 = No phosphorus
P1 = 5 kg P ha^{-1}
P2 = 10 kg P ha^{-1}
P3 = 15 kg P ha^{-1}
P4 = 20 kg P ha^{-1}
P5 = 25 kg P ha^{-1}

The data on grain yield are reported below from this experiment. Analyze the data and draw the conclusions.

Treatments	Sorghum grain yield (t ha^{-1})					
	I	II	III	IV	Treat-ments Total	Treat-ment means
P0	0.89	1.91	0.62	1.67	5.09	1.27
P1	2.67	1.71	1.20	1.31	6.89	1.72
P2	4.00	5.11	2.80	3.10	15.01	3.75
P3	3.58	4.00	2.78	3.47	13.83	3.46
P4	3.78	3.27	3.49	3.07	12.61	3.15
P5	3.45	4.60	4.62	4.94	17.61	4.40
Replications Total	18.37	19.60	15.51	17.56	71.04	

Analysis:

Step 1. CF = $(71.04)^2 \div 24 = 210.29$

Step 2. Treatment SS =

$$\frac{(5.09)^2 + (6.89)^2 + (15.01)^2 + (13.83)^2 + (12.61)^2 + (17.61)^2}{4} - CF$$

$= 239.77 - 210.28 = 29.49$

Step 3. Total SS = $(0.89)^2 + (2.67)^2 + \ldots\ldots\ldots + (4.94)^2 - CF$
$= 248.83 - 210.28 = 38.55$

Step 4. Error SS = Total SS - Treat. SS = 38.55 - 29.49 = 9.06

Step 5: Analysis of Variance Table

Source	df	SS	MSS	F	F(Table) 5%	1%
Treatments	5	29.49	5.89	11.79**	2.77	4.25
Error	18	9.06	0.50			
Total	23	38.55				

Step 6. A significant 'F' test indicates true differences in yield between phosphorus levels.

Step 7. To find the differences between P levels, calculate LSD as

LSD = 't' value (from Table at Error df for P = 0.05) × SED

Step 8. SED = $\sqrt{(\text{Error MSS} \times 2)} \div \sqrt{\text{Reps}}$

$= \sqrt{(0.50 \times 2)} \div \sqrt{4} = 0.5$

Step 9. 't' value from 't' table at 18 df for P = 0.05 is 2.101

LSD = 2.101 × 0.5 = 1.05

Step 10. For comparison of yield at different P levels, arrange the grain yield for P levels in decreasing order and rank them using the LSD value. Treatment means with the same letter indicate no significant difference between the two means.

Treatment	Mean yield (t ha^{-1})
P5	4.40 A
P2	3.75 AB
P3	3.46 AB
P4	3.15 B
P1	1.72 C
P0	1.27 C

Step 11. Conclusion: Sorghum cv CSH 9 has responded only up to 15 kg of P ha^{-1}.

Step 12. Coefficient of variation = $\sqrt{}$Error MSS x 100 ÷ Grand mean

= $\sqrt{}$0.5 x 100 ÷ 2.96 = 23.89 or 24%

MP 13. Randomized Complete-block Design

In this design the treatments are assigned at random to a group of experimental units (plots) called the block or replication. The objective is to keep the variability as small as possible among the experimental units within a block. If there are no block differences, this design will not contribute to precision in detecting treatment differences and hence one could use the completely randomized-block design.

Layout. A block (replication) should consist of plots that are as uniform as possible. Blocks can be kept compact by placing the plots, usually long and narrow in shape, close together. As the block size increases, so does the within-block variability.

When a fertility or productivity gradient is expected within the experimental area, blocks should be laid across the gradient and plots within a work laid parallel to the gradient as below.

Low fertility ↑ ↓ High fertility

IV	E	B	F	D	A	C
III	F	E	C	A	D	B
II	D	E	B	A	C	F
Rep. I	C	A	D	E	F	B

Figure 1. Six treatments replicated four times in a randomized complete block design.

After identifying uniform blocks, the treatments are assigned at random to the plots within each block, with a separate randomization being made for each block. For example, the six treatments are labeled as six pieces of paper and then draw one piece of paper at a time. The piece of paper drawn once is not replaced in the next draw. This way, one can assure that the same treatment will not appear again in the same block. After assigning six treatments in the first block, the procedure is repeated to randomize the treatments in the remaining blocks, one block at a time. Please note that each block is complete with all six treatments.

It is worthwhile, at this point, to emphasize the major difference between a completely randomized-block design (CRBD) and a randomized complete-block design (RCBD). Randomization in the CRBD is done without any restriction, but for the RCBD, all treatments must appear in each block.

Example. An experiment was laid out to compare 6 groundnut varieties for yield potential, using 4 replications in a randomized-block design.

Pod yield (t ha^{-1})

	Varieties	Replications I	II	III	IV	Treat. total	Treat. mean
1	ICGS 26	3.13	3.33	3.15	3.27	12.88(4)	3.22
2	ICGS 67	1.90	2.55	2.33	1.69	8.47	2.12
3	ICGS 12	1.96	2.04	1.56	2.02	7.58	1.89
4	ICGS 6	3.96	3.07	3.19	3.32	13.54	3.38
5	JL 24	1.84	1.27	1.67	1.56	6.34	1.58
6	ICGS 44	2.62	3.29	3.37	3.04	12.34	3.08
Rep. Total		15.41 (6)	15.55	15.29	14.90	61.15 (24)	

Analysis:

Step 1. $CF = \frac{(61.15)^2}{24} = 155.80$

Step 2. Total SS = $[(3.13)^2 + (1.90)^2 + ... + (3.04)^2] - CF = 13.56$

$$\text{Rep.SS} = \frac{(15.41)^2 + (15.55)^2 + (15.29)^2 + (14.90)^2}{6} - CF = 0.04$$

$$\text{Treat.SS} = \frac{(12.88)^2 + ... + (12.34)^2}{4} - CF = 11.92$$

Step 3. Error SS = Total SS - (Rep.SS + Treat.SS)

= 13.56 - (0.04 + 11.92) = 1.60

Step 4. Analysis of Variance Table

Source	df	SS	MSS	F	F (Table) 5%	1%
Reps.	3	0.04	0.013	0.12	3.29	5.42
Var.	5	11.92	2.384	22.33**	2.90	4.56
Error	15	1.60	0.107			
Total	23	13.56				

Step 5. 'F' test indicates that the pod yield is significantly different between the varieties.

Step 6. To find out the differences between varieties regarding the pod yield, compute LSD.

LSD = 't'(value from table at Error df for P = 0.05) × SED.

Step 7. $SED = \sqrt{(\text{Error MSS} \times 2)} \div \sqrt{\text{Reps}}$

$= \sqrt{(0.107 \times 2)} \div \sqrt{4} = 0.231$

LSD = 2.131 × 0.231 = 0.492

Step 8. Arrange the pod yield of varieties in a descending order and rank them using LSD value.

Var.	Mean yield (t ha^{-1})	
ICGS 6	3.38	A
ICGS 26	3.22	A
ICGS 44	3.08	A
ICGS 67	2.12	B
ICGS 12	1.89	BC
JL 24	1.58	C

Step 9. Interpretation: ICGS 6, ICGS 26 and ICGS 44 gave similar pod yield. JL 24 gave the lowest pod yield.

Step 10. Coefficient of variation = $(\sqrt{\text{Error MSS}} \times 100) \div \text{Grand mean}$

= $(\sqrt{0.107} \times 100) \div 2.58 = 12.8\%$

MP 14. Latin Square Design

This design evaluates two known sources of variation among experimental units. Hence this design is useful for field trials in which the experimental area has two fertility gradients running perpendicular to each other or has a unidirectional fertility gradient but also has residual effects from previous trials. The latin square design is useful for insecticide field trials where the insect migration has a predictable direction that is perpendicular to the dominant fertility gradient of the experimental field. Similarly this design could be adopted for greenhouse trials in which the experimental pots are arranged in straight lines, perpendicular to the glass or screen walls, such that the differences among rows of pots and the distance from the glass wall (or screen wall) are expected to be the two major sources of variability among the experimental pots (Gomez and Gomez 1984).

Conditions for the Design

1. Each treatment appears once and only once in each row and once per column.

2. The number of replications must be equal to the number of treatments, which makes it possible to remove or distribute the environmental variation uniformly over the existing variation, and tries to minimize it, therefore we can find real estimates or differences among treatments.

Advantages

a. It distributes the treatments uniformly over the existing variations.
b. It is good for conditions where we have a lot of environmental variation.
c. The analysis is simple.
d. It is more efficient than the randomized-complete block (RCB) at detecting real differences and similarities.

Disadvantages

a. The number of replicates must be equal to the number of treatments. So one cannot have many treatments because it will increase the number of replicates.
b. It is difficult to calculate when there are missing values.

Therefore, the latin square design has certain serious limitations. It is only available for a restricted number of treatments. If the number of treatments are large, the field layout becomes difficult covering too much area to be very efficient. On the other hand, this design is not suitable for an experiment with a few treatments. For example, with three treatments, it provides only two degrees of freedom (df) for error, and so is capable of detecting only very large differences and with less precision than can be obtained with four treatments.

Thus the latin square design is usually limited to experiments involving five to eight treatments.

Source of variation	df	df with	2 × 2 T	2 × 2 R	3 × 3 T	3 × 3 R	4 × 4 T	4 × 4 R	5 × 5 T	5 × 5 R	6 × 6 T	6 × 6 R	7 × 7 T	7 × 7 R
Rows	r-1		1		2		3		4		5		6	
Columns	r-1		1		2		3		4		5		6	
Treatments	t-1		1		2		3		4		5		6	
Error	(r-1)(r-2)		0		2		6		12		20		30	
Total	t^2-1		3		8		15		24		35		48	

Layout

In this design there are as many replications as there are treatments. The experimental area is divided into blocks equal to the number of treatments in both the direction. The two-directional blocking is commonly referred to as row blocking and column blocking. The blocks are divided into plots such that there are as many plots in each row block as there are in each column block. The plots are then assigned to the various treatments such that every treatment occurs only once in each row block and once in each column block. This can be done by randomized numbers. In the statistical tables by Fisher and Yates, sets of randomized numbers for 4 × 4 to 7 × 7 latin squares are given. The following layout for a 5 × 5 latin square illustrates the procedure.

	V	D	E	B	A	C
	IV	E	A	C	B	D
	III	A	B	D	C	E
	II	B	C	D	E	A
Row	I	C	D	A	E	B
Col		I	II	III	IV	V

Analysis

Example

The details of a sorghum experiment to compare the methods of nitrogen application are given below:

Analyze the data and draw your conclusions. Compare the efficiency of the design adopted with that of randomized-block design by taking the rows as blocks, and taking the columns as blocks.

1. Design: Latin square design 2. Plot size: $3m^2$

3. Treatments 100 kg N ha^{-1} applied as follows:

A - Ammonium sulfate applied in single dose.
B - Ammonium sulfate applied in two doses.
C - Urea applied in single dose.
D - Urea applied in tow doses.
E - No fertilizer application.

Plot yields (g)

Row	Column C1	C2	C3	C4	C5	
R5	D 82	B 55	E 26	C 78	A 63	Row Total 304
R4	E 26	C 57	B 47	A 45	D 78	253
R3	B 71	A 46	D 82	E 57	C 88	344
R2	A 59	E 43	C 80	D 90	B 77	349
R1	C 92	D 91	A 66	B 88	E 61	398
Column total	330	292	301	358	367	1648

Step 1: Get the totals.

a. Row totals = $\sum$Row 1 + ... + $\sum$Row 5
= 398 + 349 + 344 + 253 + 304 = 1648

b. Column totals = $\sum$Col 1 + $\sum$Col 2 + ... + $\sum$Col 5
= 330 + 292 + 301 + 358 + 367 = 1648

c. Treatment totals = $\sum A + \sum B + \sum C + \sum D + \sum E$
= 279 + 338 + 395 + 423 + 213 = 1648

Step 2: Correction factor (CF) = $\frac{(GT)^2}{T^2} = \frac{(1648)^2}{5^2}$ = 108636.16

Step 3: Calculate sum of squares (SS).

a. Total SS = $\sum$(Ind. data point)^2 - CF

= $(82^2 + 55^2 + ... + 61^2)$ - 108636.16 = 9283.84

b. Row SS = $\frac{(\sum R1)^2 + ... + (\sum R5)^2}{5}$ - CF

= $\frac{(398)^2 + (349)^2 + ... + (304)^2}{5}$ - 108636.16 = 2357.04

c. Column SS = $\frac{(\sum \text{Col } 1)^2 + ... + (\sum \text{Col } 5)^2}{5}$ - CF

= $\frac{(330)^2 + (292)^2 + ... + (367)^2}{5}$ - 108636.16 = 887.44

d. Treatment SS = $\frac{(\sum A)^2 + ... + (\sum E)^2}{5}$ - CF

= $\frac{(279)^2 + ... + (213)^2}{5}$ - 108636.16 = 5845.44

e. Error SS = Total SS - (Row SS + Col SS + Treat. SS)

= 9283.84 - (2357.04 + 884.44 + 5845.44) = 193.92

Step 4: Analysis of variance.

Source	df	SS	MSS	'F'	Table 'F' 0.05	0.01
Rows	4	2357.04	589.26	36.46**		
Columns	4	887.44	221.86	13.73**	3.26	5.41
Treatments	4	5845.44	1461.36	90.43**		
Error	12	193.92	16.16			
Total	24	9283.84				

Step 5: Inference: There are significant differences between rows, columns, and methods of nitrogen application as the calculated 'F' values are higher than the Table 'F' values.

Step 6: To find the differences between methods of nitrogen application, calculate LSD.

LSD = 't' value (from Table at Error df for P=0.05) × SED

Step 7: SED = $\sqrt{(\text{Error MSS}) \times 2} \div \sqrt{5}$

$= \sqrt{16.16 \times 2} \div \sqrt{5} = 2.54$

't' value at 12 df for P=0.05 = 2.179

Step 8: LSD = 2.179 × 2.54 = 5.53

Step 9: For comparison of methods of nitrogen application arrange the plot yields in decreasing order and rank them using the LSD value.

Treatment	Plot yield g ($3m^{-1}$)
D	84.6 a
C	79.0 b
B	67.6 c
A	55.8 d
E	42.6 e

Step 10: Conclusions:

a. Nitrogen application significantly increased sorghum yields.
b. Urea application was significantly better than ammonium sulfate application.
c. Application of fertilizers in two doses significantly improved sorghum yields.
d. Urea applied in two doses was the best method of nitrogen application.

Step 11: Coefficient of variation = $\sqrt{\text{Error MSS}} \times 100 \div$ Grand mean

$= \sqrt{16.16} \times 100 \div 65.92 = 6\%$

Step 12: Efficiency of latin square design over RBD.

a. Taking Rows as blocks

$$\text{MSS of error of RBD} = S^2RBD = \frac{ncSc^2 + (nt+ne)Se^2}{nc+nt+ne}$$

where nc = df for columns
nt = df for treatments
ne = df for error
Sc^2 = MSS of columns
Se^2 = MSS of error

$$S^2RBD = \frac{4 \times 221.86 + (4 + 12) \times 16.16}{4 + 4 + 12} = 57.3$$

$$\text{Efficiency} = \frac{\text{MSS of error of RBD}}{\text{MSS of error of Latin Square}} = \frac{S^2RBD}{Se^2}$$

$$= 57.3 \div 16.16 = 3.54$$

Taking columns as blocks

$$\text{MSS of error of RBD} = S^2RBD = \frac{nr\ Sr^2 + (nt+ne)Se^2}{nr + nt + ne}$$

Where nr = df for rows
Sr^2 = MSS of rows

$$= \frac{4 \times 589.26 + (4 + 12) \times 16.16}{4 + 4 + 12} = 130.78$$

$$\text{Efficiency} = S^2RBD \div Se^2 = 130.78 \div 16.16 = 8.09$$

Conclusion: Since efficiency is greater than one, the latin square design is more efficient than randomized-block design in this problem.

Summary

The latin square method has the advantage of reducing environmental variations by distributing the treatment into rows and columns. The total variance is divided into four known components, i.e., rows, columns, treatments, and error, while in the case of randomized-block design the variance was divided into three known sources of variation, i.e., blocks, treatments, and error. Therefore, the df, SS, and MSS are small in case of latin square making it more sensitive to detect the significance of difference.

MP 15. Factorial Experiment Using RCBD

To study the effect of plant densities on the performance of three sorghum types, an experiment was conducted at ICRISAT Center during the 1979 rainy season. The details of the experiment are given below:

Treatments. All possible combinations of 3 levels of population densities and 3 different sorghum types.

Population densities

P1 = 100 000 plants ha^{-1}
P2 = 150 000 plants ha^{-1}
P3 = 200 000 plants ha^{-1}

Sorghum genotypes

V1 = CSH 1
V2 = CS 3541
V3 = CSH 6

Plot size = 7 m x 3 m

Replications: 4

The following data on grain yield was recorded. Analyze the data and draw the conclusions from the analysis.

Yield of sorghum grain (100 kg ha^{-1}) from 3^2 factorial experiment

Varieties	Population		Replications				Total
			1	2	3	4	
V1	P1	31	20	18	21	90(4)	
	P2	35	25	20	23	103	
	P3	41	32	25	25	123	
V2	P1	40	36	35	38	149	
	P2	35	33	31	35	134	
	P3	28	28	23	30	109	
V3	P1	35	36	25	28	124	
	P2	37	43	42	35	157	
	P3	48	52	52	40	192	
Rep. Total		300 (9)	305	271	275	1181 (36)	

Analysis

Step 1. $CF = (1181)^2 \div 36 = 139461 \div 36 = 38743.36$

Step 2. a. Total SS = 41371 - 38743.36 = 2627.64

b. Rep SS = (350991 ÷ 9) - 38743.36 = 255.64

Step 3. From the original data, prepare a two-way table between varieties and population densities.

	P1	P2	P3	Total
V1	90 (4)	103	123	316 (12)
V2	149	134	109	392
V3	124	157	192	473
Total	363 (12)	394	424	1181 (36)

Step 4. a. Total SS of this two way table
= 40691.25 - 38743.36 = 1947.89

b. Variety SS = 39770.75 - 38743.36 = 1027.39

c. Pop. SS = 38898.42 - 38743.36 = 155.06

d. Var. SS x Pop. SS = Total SS - (Var. SS + Pop.SS)
= 1947.89 - (1027.39 + 155.06) = 765.44

Step 5. Analysis of Variance table:

Source	df	SS	MSS	'F' ratio	F (Table) 5%	F (Table) 1%
Replications	3	255.64	85.21	4.82**	3.01	4.72
Varieties	2	1027.39	513.69	29.07**	3.40	5.61
Population	2	155.06	77.53	4.38*	3.40	5.61
Var x Pop	4	765.44	191.36	10.83**	2.78	4.22
Error	24	424.11	17.67			
Total	35	2627.64				

Step 6. Inference: The variation between varieties, population densities and interaction effects are significant.

Step 7. a. For main effects:

SED = $\sqrt{(2 \times 17.67) \div 12}$ = 1.72

CD = t at 24 df x SED = 2.06 x 1.72 = 3.54

Factor	Mean yield	Conclusion
V1	26.33 A	
V2	32.67 B	V3 V2 V1
V3	39.42 C	

b. As the other factor (varieties) is also at three levels, the previous SED and CD can be used for comparison.

Factor	Mean yield	Conclusion
P1	30.25 A	
P2	32.83 AB	P3 P2 P1
P3	35.33 B	

Step 8. For interaction: Var. x Pop.

SED = $\sqrt{(2 \times 17.67) \div 4}$ = 2.97

CD = t at 24 df x SED = 2.06 x 2.97 = 6.12

Mean yields

	P1	P2	P3
V1	22.50	25.75	30.75
V2	37.25	33.50	27.25
V3	31.00	39.25	48.00

c. Comparison of population densities with each sorghum variety

V1	P3 A	P2 AB	P1 B
V2	P1 A	P2 A	P3 B
V3	P3 A	P2 B	P1 C

d. Comparison of varieties at each population density

P1	V2 A	V3 B	V1 C
P2	V3 A	V2 A	V1 B
P3	V3 A	V1 B	V2 B

Step 9. **Summary Conclusions**

a. CSH 6 was found to be best amongst the varieties followed by CS 3541. CSH 1 had the lowest yields.

b. The highest population density, 150 000 plants ha^{-1} resulted in significantly higher yield than 100 000 plants ha^{-1}.

c. The increased yields with higher population densities were observed only with hybrids CSH 6 and CSH 1.

d. The better performance of CSH 6 was noted only at the highest population densities.

MP 16. Data Transformation

Variance heterogeneity is the most common condition in experimental data that violates one or more of the assumptions of the analysis of variances. For data with heterogenous variances, the specific type of variance heterogeneity must be identified before selecting an appropriate transformation procedure.

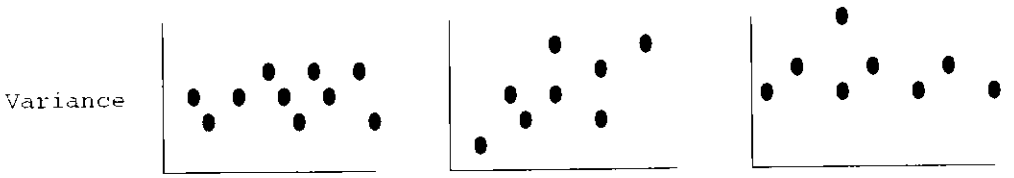
The following is a simplified procedure for detecting the presence and the type of variance heterogeneity.

Step 1. For each treatment, calculate the variance and the mean across replications. For ease of computation, the range can be used in place of the variance.

Sept 2. Plot a scatter diagram between the mean values and the variances (or ranges). The number of points in the scatter diagram equals the number of treatments.

Step 3. Observe the scatter diagram for any pattern of relationship between the means and the variances (or ranges).

The following diagrams illustrates three possible outcomes of such observations.



Data transformation is the most appropriate remedial measure for data where the variance and the mean are functionally related. With this procedure, the original data are converted into a new scale resulting in a new data set which satisfy the condition of homogeneity of variance.

The appropriate data transformation depends on the specific type of relationship between the variance and the mean (Table 1).

Table 1. Data transformation.

Relationship	Type of transformation	Observations
SD is proportional to mean	Logarithmic	Whole numbers that cover a wide range e.g., insects plot^{-1}, egg masses plant^{-1}, etc.
Variance (SD^2) is proportional plot $^{-1}$, to mean	Square-root	1. Data obtained in rare events e.g., infested plants, weeds plot^{-1}, etc. 2. Percentage data with range between 0 % to 30% or 70% to 100%.
Variances tend to be to be small at the the two ends of the range of values (close to zero and 100%), but larger in the middle (around 50%).	Arc-sine or angular transformation	Proportions, decimal fractions, or percentages.

Rules with regard to percentage data:

a. For percentage data within a range of 30% to 70%, no transformation is needed.

b. For percentage data within the range of either 0 to 30% or 70 to 100%, but not both, the square-root transformation should be used.

c. For percentage data that do not follow the ranges specified in either rule **a** or **b**, the arc-sine transformation is appropriate.

Caution: Percentage data derived from count data should be clearly distinguished from other types of percentage data, like threshing percentage, protein percentage, shelling percentage, nitrogen percentage, etc. which are not derived from count data.

MP 17. Missing Plot Technique

I. Randomized-block Design

Replications	V_1	V_2	V_3	V_4	V_5	V_6	Total
1	16.00	15.75	23.00	15.50	16.00	12.50	98.75
2	17.50	16.00	22.25	17.50	15.50	18.00	106.75
3	20.50	14.25	19.75	Y	13.00	17.50	(85.00)
4	16.00	10.25	16.25	14.75	15.50	13.50	86.25
Total	70.00	56.25	81.25	(47.75)	60.00	71.50	386.75

Step 1. Estimate the missing value of Y by using the formula

$$Y = \frac{rR + tT - G}{(r-1)\ (t-1)}$$

Where r = No. of replications = 4
R = Total of remaining units in the block where the missing unit appears = 85.00
t = No. of treatments = 6
T = Total of missing treatment in other blocks = 47.75
G = Grand Total (without the missing value) = 386.75

$$Y = \frac{(4 \times 85) + (6 \times 47.75) - 386.75}{3 \times 5} = \frac{239.75}{15} = 15.98$$

Step 2. Substitute the value of Y (15.98) and find the new treatment total, block total, and grand total. Then proceed with the analysis as done for usual randomized-block design.

Correction factor = $(402.73)^2 \div 24 = 6757.97$

Total SS = $\sum Y^2_{ij} - CF = 214.58$

Treatment SS = $\frac{(70.00)^2 + \ldots\ldots + (71.50)^2}{4} - CF = 101.88$

Replication SS = $\frac{(98.75)^2 + \ldots\ldots + (86.25)^2}{6} - CF = 51.72$

Step 3. Find the adjusted value of treatment sum of squares.

Note: This is to be done, when the treatment effect appears only to be just significant. If the effect is not significant or highly significant this adjustment is not necessary.

For this, we deduct the following value from the treatment SS:

$\frac{(R + tT - G)^2}{t(t-1)(r-1)^2}$ Where R, t, T, G are the same as in the previous equation to estimate Y.

R = 85.00, t = 6, T = 47.75, and G = 386.75

Substituting the values:

$$\frac{[85.00 + (6 \times 47.75) - 386.75]^2}{6 \times 5 \times (3)^2} = \frac{(-15.25)^2}{270} = 0.86$$

Adjusted value of treatment SS = 101.88 - 0.86 = 101.02

Analysis of variance table:

Source	df	SS	MSS	'F' ratio
Replication (r-1)	3	51.72	17.24	3.95
Treatment (t-1)	5	101.02	20.20	4.63
Error (r-1)(t-1)-1	14	60.98	4.356	
Total (rt-2)	22	214.58		

Tabular value of 'F' (3, 14) at 5% level = 3.34
Tabular value of 'F' (5, 14) at 5% level = 2.96

Inference: As the table value of 'F' (5, 14) is less than calculated 'F' value, the varieties are significantly different from one another.

Step 4. To find standard error of difference:

a. Standard error of difference between missing treatment and any of the treatments.

V_4 vs any other treatment

$$= \sqrt{S_e^2\left(\frac{2}{r} + \frac{t}{r(r-1)(t-1)}\right)}$$

$$= \sqrt{4.356\left(\frac{2}{4} + \frac{6}{4 \times 3 \times 5}\right)}$$

$$= \sqrt{2.6136} = 1.61$$

b. Standard error of difference between any two other non-missing treatments.

$$\sqrt{2S_e^2 \div r} = \sqrt{(2 \times 4.356) \div 4} = 1.48$$

Technique of Analysis of RBD when more than one treatment is missing

V_4 is missing in replication III
V_5 is missing in replication IV

First we take the average of available values and substitute that value for V_4. We can use the formula to get the value of V_5. Then taking this value of V_5 use the formula to get the value of V_4. This procedure is repeated until consecutive values of V_4 obtained by using the different methods agree. Similarly calculate V_5.

Use these values and proceed with the analysis.

II. Latin square design

(5 treatments and 5 replications)

Yield ($t\ ha^{-1}$)	A	B	C	D	E
	8.08	7.18	8.42	6.98	8.59
	4.91	6.61	..	4.34	6.11
	4.25	5.96	3.32	4.44	5.60
	5.79	5.59	6.86	4.02	4.21
	5.89	3.46	2.38	5.81	6.14
Treat. totals	28.89	28.80	20.98	25.59	30.65

Row totals: $R_1 = 39.22$, $R_2 = 21.92$, $R_3 = 23.57$, $R_4 = 26.47$, $R_5 = 23.68$

Column totals: $C_1 = 25.29$, $C_2 = 30.84$, $C_3 = 24.51$, $C_4 = 29.31$, $C_5 = 24.96$

Grand total: 134.91 134.91

Estimate the missing value Y:

$$Y = \frac{r\ (R + C + T) - 2G}{(r-1)\ (r-2)}$$

Where r = Number of rows or treatments,
R, C, and T are row, column, and treatment total respectively, all having the missing value (without that)

r = 5, R = 21.97, C = 24.51, T = 20.98, G = 134.91

Substituting the values:

$$Y = \frac{5\ (21.97 + 24.51 + 20.98) - (2 \times 134.91)}{4 \times 3} = 5.62$$

Missing value Y = 5.62.

Replacing the missing value Y with 5.62, find the grand total and other totals and analyze the data as if it is a latin square design with all the values.

Grand total = 140.53
Correction factor = $(140.53)^2 \div 25 = 789.95$

a. Total SS $= \sum\sum Y^2_{ij} - CF = 60.54$

b. Row SS $= \frac{(39.22)^2 + ... + (23.68)^2 - CF}{5} = 33.32$

b. Treat. SS $= \frac{(28.89)^2 + ... + (30.65)^2 - CF}{5} = 3.19$

c. Column SS $= \frac{(25.29)^2 + ... + (24.96)^2 - CF}{5} = 6.17$

Analysis of variance:

Source		df	SS	MSS	'F' ratio
Row	(r-1)	4	33.32	8.33	5.14
Column	(r-1)	4	6.17	1.54	0.95
Treatment	(r-1)	4	3.19	0.80	0.49
Error	(r-1)(r-2)-1	11	17.86	1.62	
Total	$(r^2 - 2)$	23	60.54		

Tabular value of 'F' 4, 11 at 5% level = 3.36

Inference: As the 'F' ratio is not significant, the treatment effects are not significant.

Standard error of difference

1. Comparing between the missing plot treatment and any other treatment, say A and C,

$$e=\sqrt{S_e^2\left(\frac{2}{r}+\frac{1}{(r-1)(r-2)}\right)}$$

$$=\sqrt{1.62\left(\frac{2}{5}+\frac{1}{4\times3}\right)}$$

$$=\sqrt{0.78}=0.88$$

2. Standard error of difference between other non-missing treatments:

$$\sqrt{2S_e^2 \div r} = \sqrt{(2 \times 1.62) \div 5} = 0.805$$

When there are more than 2 missing plots, with a latin square design the method of estimating the missing value is the same as we have seen in the case of randomized-block design with identical conditions.

Effective number of replications

A. Latin square:

First find the average number of replicates for each treatment.

Each treatment = 1 replication if the other treatment is present in both the corresponding row and column; 2/3 when the other treatment is missing either in row or column and is 1/3 when the other treatment is missing both in row and column, when the treatment in question itself is missing, the replication is zero.

Row

Note: * = missing plots

1	C	D	E	B	A	F
2	D	E	B	*A*	F	C
3	E	*B*	A	F	C	D
4	B	*A*	F	C	D	E
5	A	F	C	D	E	*B*
6	F	C	D	E	*B*	A
Col	1	2	3	4	5	6

Effective number of replications for A

1st row and col. A and B are present so 1 is replication
2nd row and col. A is absent so 0 is replication
3rd row and col. A is present but B is absent so 1/3 is replication
4th row and col. A is absent so 0 is replication
5th row and col. A is present but B is absent so 1/3 is replication
6th row and col. A is present but B is absent so 1/3 is replication

So, total No. of effective replications for A = 2

and SE of difference = $\sqrt{S^2_e\ (1/rA + 1/rB)}$
where rA and rB are effective replications of A & B respectively.

B. **Randomized-block Design:**

Each treatment is given 1 replication if the other treatment is also present: (t-2) ÷ (t-1) if the other treatment is absent and zero when the treatment in question is itself missing.

References

Fisher , R.A. and Yates, F. 1963. Statistical Tables for Biological, Agricultural, and Medical Research. London, UK: Oliver and Boyd.

Farr, A.D. 1985. Science writing for beginners. London, UK: Blackwell Scientific Publication.

Gomez, K.A., and **Gomez, A.A.** 1984. Statistical procedures for agricultural research. 2nd edn. New York, USA: John Wiley & Sons.

Jeffers, J.N.R. 1978. Design of Experiments. Institute of Terrestrial Ecology. Natural Environmental Research Council, Cambridge, UK.

Linton, C.D. 1954. How to write reports. New York, USA: Harper and Row.

McKee, G.W. 1964. A coefficient for computing leaf area in hybrid corn. Agronomy Journal 56: 240-241.

Panse, V.G., and **Sukhatme, P.V.** 1989. Statistical methods for agricultural workers. 4th edn. New Delhi, India: Indian Council of Agricultural Research.

Singh, D., Singh, V., and **Chandel, S.R.S.** 1970. Determination of leaf area in bajara (*Pennisetum typhoides* S & H). Science and Culture 36:46-47.

Stickler, F.C., Wearden, S., and **Pauli, A.W.** 1961. Leaf area determination in grain sorghum. Agronomy Journal 53:187-188.

Ward, W.B. 1959. Reporting agriculture. Ithaca, New York, USA: Cromstock.

Warren, J.A. 1982. Folklore and plot technique, Part I - contentions for the common man. ICRISAT mimeo (Limited distribution).

Evaluations

Select the most appropriate answer and check the correct answer at the end of the booklet.

1. Before writing a proposal for a field experiment, it is necessary to
   a) determine the size and shape of the plots.
   b) state the objectives of the experiment.
   c) choose an experimental design.
   d) develop collaborative arrangements with all concerned.

2. When the objectives of the experiment have been stated, the next step is
   a) to locate an experimental field.
   b) to purchase necessary equipment.
   c) to write precise questions that the experiment is expected to answer.
   d) to select a suitable design.

3. The treatments consist of crop varieties, and materials or methods to be tested. They should be
   a) useful in actual crop production.
   b) suitable to the soil type.
   c) representative of the population.
   d) representative of the sample.

4. The set of treatments should include
   a) all possible entries.
   b) a naturally defined control.
   c) a few more entries than the minimum number.
   d) some entries rejected in earlier experiments.

5. The levels in any one factor should change
   a) in small amounts.
   b) in a random manner.
   c) by a constant amount or in a constant ratio.
   d) as advised by the boss.

6. The experimental site should be
   a) selected by the research station manager.
   b) representative of the soil type and location.
   c) near a road and easily accessible.
   d) near a water source for irrigation.

7. The experiments should always be positioned
   a) in a north-south direction.
   b) on high ground to avoid drainage problems.
   c) away from trees to avoid their root effects.
   d) as close to housing as possible for better supervision.

8. The experimental plots should
   a) be small to accommodate many treatments.
   b) be large to improve precision.
   c) be large enough to prevent influences of adjacent plots.
   d) always be rectangular in shape.

9. Improvement in precision can be achieved by
   a) increasing plot size.
   b) increasing number of replications.
   c) increasing number of replications after a minimum plot size is selected.
   d) having large plots as well as a large number of replications.

10. The experimental area should be free from
    a) residual effects of the previous crop.
    b) stubbles of previous crops.
    c) fertility differences.
    d) environmental pollution.

11. The plot size is defined by
    a) the nature of the experimental materials or treatments.
    b) type and kind of the fertilizer used as a basal rate.
    c) the nature of irrigation provided.
    d) the intensity of the plant protection coverage.

12. Blocks (replications) should be arranged in
    a) any manner.
    b) relation to the fertility gradient.
    c) relation to the slope of the land.
    d) consultation with the farm manager.

13. Replications are
    a) repetitions of treatments.
    b) repetitions of sets of treatments over locations.
    c) randomly arranged sets of treatments.
    d) repetitions of sets of treatments over seasons.

14. All the experimental plots should be
    a) of varying shape and size among replications.
    b) of the same size but not necessarily same shape.
    c) of the same size and shape.
    d) as decided by the Director of Research.

15. Competition/interaction between two plots (tall vs dwarf lines) of the experiment can be adjusted by
    a) placing them away from each other.
    b) increasing the space between plots or surrounding each plot by a buffer zone.
    c) putting them at random.
    d) suitable analytical procedures.

16. Pathways are left (usually 1 m) between the treatments because
    a) we can avoid snakes in the fields.
    b) we can move small tools and equipment easily from one experiment to another.
    c) we can move conveniently to take notes of observations.
    d) it is a requirement for all good experimental layouts.

17. We need to know the area of a field because
    a) it is customary for all scientists.
    b) we can precisely estimate the yields of crops or other data.
    c) we can locate replications in different areas of the field.
    d) we can estimate the pest and disease incidence.

18. A variable is defined as
    a) anything that varies quantitatively.
    b) anything that varies qualitatively.
    c) any character which changes such as color, size, or shape of flower petals or grain yield.
    d) any character that changes occasionally.

19. An ammonium sulfate trial on a lime deficient soil may not yield dependable data. Therefore, the scope of the experiment presumes
    a) a more thorough satisfaction of the statistical aspects.
    b) availability of supplies and hardware.
    c) availability of staff to handle the data produced.
    d) a knowledge of the agricultural background of the problem.

20. If the optimum rate of nitrogen is expected to be 80 to 100 kg ha^{-1}, the suitable rates for a trial would be (kg N ha^{-1})
    a) 0,20,40,60,80,100,120,140.
    b) 60,80,100,120,140.
    c) 60,120,180.
    d) 0,50,100,150,200.

21. The choice of treatments using farm yard manure (FYM) (0.5% N) and ammonium sulphate (20% n) could be
    a) 16 t of FYM compared with 250 kg ammonium sulphate.
    b) 10 t of FYM compared with 250 kg ammonium sulphate.
    c) not comparable as FYM has other important elements such as P and K.
    d) 10 t of FYM compared with 250 kg ammonium sulphate in split applications.

22. If a large number of treatments are to be tested the more efficient design will be
    a) a randomized block.
    b) a latin square.
    c) a confounded design.
    d) a paired-plot design.

23. The objectives and null hypotheses to be tested should be written
    a) after the data collection was over.
    b) after the layout of the experiment was over.
    c) before starting the layout.
    d) at the time of writing the research plan/proposal.

24. Uniformity in field operations is essential to reduce experimental error. To get a uniform stand where drill sowing is common, one should
    a) adopt drill sowing in the experimental plots.
    b) use hand sowing to increase uniformity.
    c) divide the seed for each row length and adopt drill sowing.
    d) divide the seed for each row length and adopt hand sowing.

25. All cultivation operations for each replication (sowing, weeding, etc.) should be conducted
    a) on the same day.
    b) when labor is available.
    c) for half day at a time.
    d) when convenient for the scientist.

26. A uniformity trial should be conducted to know
    a) the fertility gradient in the experimental field.
    b) how to bring the field to the minimum cropping value.
    c) the slope of the field.
    d) the nutrient deficiencies in the soil

27. If the experimental material cannot be accommodated in uniform plots or fields then the
    a) experimental material should be divided into blocks with the least variation.
    b) experimental material should be located randomly in the field.
    c) field should be rejected.
    d) field should be made uniform.

28. Label each experimental plot with a number or descriptive symbol to
    a) keep proper record linking it with the experimental plan and to record results.
    b) check with field labels.
    c) check data recorded by the assistants.
    d) inform the public visiting the experiments.

29. If land or experimental material (seed) is not sufficient for the number of replications required to give significant differences of practical importance,
    a) it is not worth doing the experiment.
    b) we should manage with whatever we have.
    c) we can conduct the experiment half this year, half next year.
    d) we should reject some of the treatments and redesign the experiment.

30. The variables to be observed and recorded should be decided
    a) by the supervisor.
    b) at the time of developing the proposal and objectives.
    c) as the crop grows in the field.
    d) when time is available.

31. A small plot cannot be reduced indefinitely because
    a) of border effects.
    b) plant establishment is poor in small plots.
    c) of the small number of plants for observation in small plots.
    d) it is not physically possible to work in the plot.

32. The treatments and controls are allocated to the plots of the experiment
    a) by an explicit randomizing procedure.
    b) by a casual distribution.
    c) by a rotation.
    d) on the advice of elders.

33. While measuring the breadth of the plot, it is necessary to include a distance equal to
    a) average row width x number of rows.
    b) one row spacing beyond each of the extreme rows.
    c) two row spacing beyond each of the extreme rows.
    d) one and a half row spacing.

34. When spacings of 30, 45, and 80 cm are compared for the same crop or for intercropping trials,
    a) the breadth of the net plot should be the same.
    b) the breadth of the net plot could be different as an equal number of rows $plot^{-1}$ is important.
    c) the length of row can be varied to obtain the same size net plot.
    d) uniformity in plot shape and size is not necessary.

35. The number of replications in an experiment is based on the need to match the size of differences likely to be detected as significant with the size of differences one considers to be of practical importance. This can be estimated by the formula
(N=no. of replications, C=CV, S=SE of means)
    a) $N = (C \times S)^2$.
    b) $N = (C \div S)^2$.
    c) $N = (S)^2 \div (C)$.
    d) $N = (C \times S) \div (S \times C)$

36. If the estimated CV in an experiment is 8% and the SE of mean is 4, what is the minimum recommended number of replications for the treatments?
    a) 4. b) 1. c) 16. d) 2.

37. Provision of borders for an experimental plot is accomplished
    a) by leaving one or more rows or a narrow strip of the crop not harvested.
    b) by leaving some plants at both ends.
    c) by raising other crops (safflower, chickpea, or chillies) in the row or margins of the plot.
    d) by leaving a clean cultivated area around the plot.

38. Errors due to soil heterogeneity can be reduced by
    a) adjustments in data from poor fertility areas.
    b) proper layouts, randomization, and replication.
    c) adding a fixed percentage to all plot yields.
    d) harvesting only a small area with a good crop stand and yield.

39. Errors due to faulty technique such as use of inaccurate scales, observer's bias, mechanical errors in totalling, etc., can be eliminated by
    a) entrusting the jobs to assistants.
    b) training the workers.
    c) working for a limited time and taking more rest.
    d) using automatic recording equipment.

40. The chance or random errors cannot be attributed to any cause. They can be reduced by
    a) suitable statistical analytical techniques.
    b) assuming them as natural to biological research and ignoring them.
    c) consulting the field workers to make adjustments.
    d) altering the data by a constant amount.

41. On randomization, if you detect apparently unfortunate coincidences,
    a) you can rerandomize any part of the allocation.
    b) you can accept them as they occur.
    c) you can accept them as they occur if there is no variation in the location or experimental material which has not been incorporated into the design of the experiment.
    d) you cannot accept such coincidences.

42. Randomization is done to
    a) satisfy the statistician.
    b) prevent experimenter's bias in sampling.
    c) satisfy the superiors.
    d) deal with missing plot yields.

43. The randomized-block design is
    a) useful only for crop improvement trials.
    b) less efficient being too simple.
    c) has the advantage of being robust and easy to analyze.
    d) useful for agronomic trials.

44. The row and column designs (like latin square) help to
    a) study several factors at a time.
    b) layout treatments in separate blocks.
    c) remove the effects of one-way fertility trend.
    d) remove the effects of two-way fertility gradients.

45. Before developing an experimental design, one should have an estimate of the precision likely to be achieved (like CV and CD) through
    a) a pilot experiment run earlier.
    b) consultation with statisticians.
    c) use of random tables.
    d) use of table of 't' value.

46. In experiments involving a combination of treatments (irrigation and crop varieties), the more efficient design will be a
    a) block design. b) completely randomized block.
    c) split plot design. d) randomized-block design.

47. If a large number of varieties are to be tested; it would be desirable to use
    a) an incomplete block design (lattice).
    b) a paired plot design.
    c) a split plot design.
    d) a randomized-block design.

48. In varietal trials, the border plants near pathways are usually
    a) more vigorous.
    b) less vigorous.
    c) above average.
    d) below average.

49. In fertilizer experiments, the effects of the treatment may extend
    a) beyond the border of the plots.
    b) to the center of the plots.
    c) away from the experimental area.
    d) along the border near irrigation channels.

50. When the requisite skills, resources, and training to conduct a more complex experiment are not available, one should
    a) select a complex design for greater efficiency.
    b) select a simple randomized-block design.
    c) select a lattice design.
    d) not attempt to conduct field experiments.

51. All experimental data should be recorded
    a) directly in a rough note book.
    b) directly in a permanent record.
    c) in a perforated note book.
    d) in a pocket size diary.

52. If different observers were employed to record data; or assessments are made on different days, or at different times of the day, any unwanted variation in assessment is absorbed by
    a) taking averages to be more accurate.
    b) planning to use different blocks or replications of the experiment.
    c) applying suitable statistical treatment.
    d) applying adjustments to data.

53. In the field note book, one should provide for
    a) rough work sheets.
    b) recording unexpected changes or effects.
    c) extra marginal space for recording names of persons for giving work credits.
    d) time of arrival and departure, description of work done, etc., by the scientist.

54. The measurements recorded in an experiment should be checked to see whether
    a) they are meaningful and relevant to the objectives.
    b) we can make corrections to get meaningful information.
    c) they are in the expected direction or trend.
    d) changes can be made to obtain expected trends.

55. When crop varieties do not mature at the same time
    a) all the plots should be harvested for uniformity with early varieties.
    b) plots should be harvested when they mature.
    c) all the plots should be harvested after the late varieties also mature.
    d) plots should be harvested with adjustments for bird damage, shedding grain or immature seed.

56. If plant growth is such that individual plants cannot be easily distinguished, data on sampling unit for ancillary observations (like rate of growth, flowering date, pest and disease incidence, etc.) may be obtained by
    a) measuring a 1 m row or a suitable fraction of it.
    b) separating the plants along with tillers and numbering them for random selection.
    c) taking a few plants in each continuous row by careful selection.
    d) taking the control for observation on one or two rows.

57. Observation plots are preliminary trials to evaluate
    a) gross suitability of the material with reference to local weather, pests, and diseases.
    b) accurately the performance of new lines in a local environment.
    c) the reaction of the farmers to a new material.
    d) gross suitability of the material in a local environment.

58. The mean is defined as the
    a) average yield for the plot.
    b) average yield for the experiment.
    c) sum of all observations divided by their number.
    d) sum of all observations divided by one less than their number.

59. The formula for the standard deviation of a sample is ('x' = observation value, $\bar{X}$ = mean, 'n' = number of observations)

    a) $\sum X^2 \div n$
    b) $\sqrt{\sum(X-X)^2 \div n}$
    c) $\sqrt{\sum(X-X)^2 \div (n-1)}$
    d) $\sqrt{\sum X^2 \div n}$

60. If $(X-\bar{X}) = d$ in the above formula, the standard error of mean is obtained by
    a) $\sum d^2 \div n(n-1)$
    b) $\sqrt{\sum d^2 \div n(n-1)}$
    c) $\sqrt{\sum d^2 \div n-1}$
    d) $\sqrt{\sum d \div n(n-1)}$

61. The formula for deriving standard error of difference between two means (SED) is

    a) $\sqrt{\sum d^2 \div n(n-1)}$
    b) $\sqrt{\sum d^2 \div n(n-1)} \times \sqrt{2}$
    c) $\sqrt{\sum d^2 \div n(n-1)} \times \sqrt{2}$
    d) $\sqrt{\sum d \div n(n-1)} \times \sqrt{4}$

62. The formula for obtaining 't' is (D is the difference of two sample means)
    a) $t = D \div SED$
    b) $t = D \times SED$
    c) $t = D \times \sqrt{SED}$
    d) $t = D - SED$

63. Degrees of freedom is
    a) one less than the number of observations for each source of variation.
    b) the several alternatives for recording observations.
    c) the several alternatives for analyzing data.
    d) two less than the number of treatments x no. of replications.

64. The formula for determining the correction factor (CF) is (X = the total of n observations; 'n' = number of observations)
    a) $CF = \sum X^2 \div n$
    b) $CF = (\sum X \div n)^2$
    c) $CF = (\sum X)^2 \div n$
    d) $CF = \sum X \div n^2$

65. The mean square is obtained by dividing the
    a) treatment sum of squares by sum of squares for blocks.
    b) total sum of squares by degrees of freedom for error.
    c) sum of squares by appropriate degrees of freedom.
    d) total sum of squares by total degrees of freedom.

66. The formula for obtaining the coefficient of variation (CV) is (SD = standard deviation, $\bar{X}$ = mean, SE = standard error)
    a) $CV = SD \times 100$
    b) $CV = (SD \div \bar{X}) \times 100$
    c) $CV = SE \times 100$
    d) $CV = (SE \times \bar{X}) \times 100$

67. The least significant difference (LSD) is derived from (t=value of t from tables, SED is SE of difference between two means)
    a) $LSD = t \times SED$
    b) $LSD = t - SED$.
    c) $LSD = t \pm SED$.
    d) $LSD = t - SED$.

68. Tests of significance are usually applied using probability (P) levels
a) P = 0.05 and 0.10. b) P = 0.25 and 0.10.
c) P = 0.05 and 0.01. d) P = 0.50 and 0.10.

69. In tests of significance the meaning of P = 0.05 is that
a) 5 out of 100 events do not occur purely by chance.
b) 5 out of 100 events may occur due to chance.
c) 5% of the events occur purely by treatment effects.
d) 5% of the events do not occur purely by treatment effects.

70. The coefficient of variation (CV) decreases steadily (though not proportionately)
a) as the plot size increases.
b) as the plot size decreases.
c) irrespective of the plot size if the shape is correct.
d) irrespective of the size and shape of the plot.

71. An increase in the number of replications, with a reduced plot size leads to a
a) higher LSD. b) less precise comparison of treatments.
c) higher CV. d) more precise comparison of treatments.

72. When the data of an experimental plot are lost through accidents or mishaps
a) the experiment should be rejected.
b) the treatment should be rejected.
c) estimates of the missing plot data may be calculated by using approved statistical procedure.
d) past years data may be used for substitution.

73. Data as percentages (like pest incidence) can be analyzed
a) as any other numerical data.
b) after transformation.
c) after rounding to the nearest ten.
d) after conversion into decimals.

74. The methods of analysis of data should be determined and the equipment needed are to be obtained
a) when analytical work commences.
b) after the data are collected.
c) at the time of writing the research proposal.
d) at the time of laying out the experiment.

75. Scores (1 to 5 or 1 to 9) for qualitative observations such as pest or disease incidence (slight, medium, severe, very severe, etc.)
a) can be analyzed in the usual manner (ratio scale).
b) are in the ordinal scale and therefore cannot be analyzed in the usual manner.
c) can be analyzed only by a statistician.
d) can be pooled and analyzed.

76. Using the 'F' test of significance, the observed differences indicate significance, if the observed 'F' ratio (at P=0.05 or 0.01 level) is
a) twice the table value.
b) less than the table value.
c) equal to the table value.
d) more than the table value.

77. If the 'F' test is significant for blocks, it may be due to the fertility gradient or other variations. The treatment data should be
a) rejected. b) accepted conditionally.
c) accepted. d) further modified.

78. An analysis of variance is adopted when data from an experiment follow a
a) poison distribution. b) poison and normal distribution.
c) normal distribution. d) binomial distribution.

79. Homogeneity of variances imply a
    a) close relationship among all the treatment variances.
    b) functional relationship among treatment variances and means.
    c) large variation among treatment means.
    d) insufficient number of replications.

80. The most common symptom of experimental data that violates one or more of the assumptions of the analysis of variance is
    a) variance heterogeneity.
    b) variance homogeneity.
    c) additive effects of variances.
    d) nonuniformity in experimental field.

81. The means and variances are not correlated in data which follow a
    a) poison distribution.
    b) normal distribution.
    c) binomial distribution.
    d) poison, normal and binomial distribution.

82. A situation in which the effect of a treatment is to increase the yield by a certain percentage or proportion is identified as
    a) a treatment interaction effect.
    b) an additive treatment effect.
    c) a multiplicative treatment effect.
    d) an associate treatment effect.

83. Variance heterogeneity in a data set could be identified by plotting a scatter diagram between
    a) standard deviation and variance values.
    b) variance and mean values.
    c) variance and range values.
    d) standard deviation and range values.

84. Data transformation are appropriate when the
    a) variance is equal to the mean value.
    b) variance and means are not correlated.
    c) standard deviation is proportional to mean value.
    d) standard deviation is equal to the mean value.

85. Logarithmic transformation is adopted when the
    a) variance is proportional to the mean value.
    b) standard deviation is proportional to the mean value.
    c) standard deviation is equal to the mean value..
    d) variance is proportional to standard deviation.

86. The square-root transformation is followed when the
    a) variance is proportional to the mean value.
    b) standard deviation is proportional to the mean value.
    c) standard deviation is equal to the mean value.
    d) variance is proportional to the standard deviation.

87. The following data (number of weeds per sq.m) were collected from a weed control trial:

Treatment	Rep I	II	III	IV
NW	202	180	92	142
HW1	55	102	80	26
RW	2	0	5	10
Atrazine	18	14	20	2

What type of transformation do you suggest to analyze the data?
    a) Square-root transformation.
    b) Arc sine transformation.
    c) Logarithmic transformation.
    d) None of the above.

88. In a disease resistance trial, the following percentages of disease infestation were observed:

Treatment	Rep I	II	III	IV
A	38	42	39	46
B	51	48	56	62
C	32	68	40	70

What type of transformation do you suggest to analyze the data?
a) Square-root transformation.
b) Arc sine transformation.
c) Logarithmic transformation.
d) No transformation is required.

89. The following are the average number of eggs plant^{-1} recorded in an experiment on shoot fly:

Treatment	Rep I	II	III	IV
X	2	8	6	0
Y	1	0	0	2
Z	6	4	5	8

What type of transformation do you suggest to analyze the data?
a) Square-root transformation
b) Arc sine transformation
c) Logarithmic transformation
d) No transformation is required

90. The shelling percentage of groundnut varieties were recorded as:

Treatment	Rep I	II	III	IV
Var 1	57	62	68	56
Var 2	48	50	48	52
Var 3	72	75	70	68

What type of transformation do you suggest to analyze the data?
a) Square-root transformation.
b) Arc sine transformation.
c) Logarithmic transformation.
d) No transformation is required.

Correct responses to the questions.

1. b); 2. c); 3. a); 4. b); 5. c); 6. b); 7. c); 8. c); 9. c);
10. a); 11. a); 12. b); 13. a); 14. c); 15. b); 16. c); 17. b);
18. c); 19. d); 20. d); 21. b); 22. c); 23. d); 24. c); 25. a);
26. a); 27. a); 28. a); 29. d); 30. b); 31. a); 32. a); 33. b);
34. b); 35. b); 36. a); 37. d); 38. b); 39. d); 40. a); 41. a);
42. b); 43. c); 44. d); 45. a); 46. c); 47. a); 48. a); 49. a);
50. b); 51. b); 52. a); 53. b); 54. a); 55. b); 56. a); 57. a);
58. c); 59. c); 60. c); 61. b); 62. a); 63. a); 64. c); 65. c);
66. b); 67. a); 68. c); 69. b); 70. b); 71. d); 72. c); 73. b);
74. c); 75. b); 76. d); 77. c); 78. c); 79. a); 80. a); 81. b);
82. c); 83. b); 84. c); 85. c); 86. a); 87. a); 88. d); 89. c);
90. d).