Statistics Training Guide

Experimental Design

Analysis of Variance

Dolores R. Ledesma
Biometrician

AVRDC – The World Vegetable Center
East and Southeast Asia
Bangkok, Thailand

34th International Vegetable Training Course
Vegetable for Sustainable Development
Kasetsart University Kamphaeng Saen Campus
Nakhon Pathom, Thailand
December 2015

Contents

EXPERIMENTAL DESIGN

I What is Experimental Design?

Experimental design is a procedure of planning experiments so that the data obtained can be analyzed to arrive at valid results and unbiased conclusions. Planning an experiment is important because experimental units (or experimental plots) are not uniform. Uncontrolled variations can result in data with so much "noise", causing the experimental results to be unclear or biased, and therefore not valid. The objective of experimental design is to reduce the impact of uncontrolled variations and validate the results of the statistical analysis. It has three important components: replication, randomization, and error control

II Basic Statistical Concepts

Experimental unit is the experimental material to which a treatment is applied. The following are some examples: a plot of land, a single leaf, a single plant, 24 fruits on two trays, 3 plants in a pot, 12 seeds in a petri dish, etc.

Test factor (or simply factor) is a variable that the experimenter varies in an experiment. If the experiment is conducted to compare the yield of ten varieties then "variety" is a factor. Other examples are age of seeds (1, 2, 3 months), shelter (w/, w/o), or types of mulch (w/o mulch, w/plastic, w/straw).

Level is a quantity or aspect of a given factor. Enclosed in parentheses in the paragraph on Test factor above are examples of levels of the listed factors.

Treatment is a level of a factor or specific combinations of factors and levels applied to the experimental units to measure the effect of its application. In single factor studies, such as the regional yield trials, the treatments are the test varieties.

Example:

Fresh Market Tomato	**Grafting**
FMTT904 (V1)	With
FMTT957 (V2)	Without
FMTT962 (V3)	

Treatment effect is the expected increase or decrease in response with the application of a particular treatment.

Example:

Variety		Varietal effect on yield
Recommended	6 t/ha	varietal effect is the "increase" in yield of the hybrid over the check
Check	3 t/ha	

Precision refers to the closeness of the measurement to the average. It is achieved through the following:

1. Experimental units which are as uniform as practicable
2. Careful conduct of all operations before and during the experiment
3. Replication

Standard deviation is a measure of spread or extent to which observations vary in a population. It gives a small value if the observations cluster closely about the mean and a large one if they are spread widely. Below are two sets of data with a common mean but different variability:

								Mean	Standard deviation
Set I	7	8	8	9	10	10	11	**9**	**1**
Set II	1	2	3	9	14	16	18	**9**	**7**

Both sets have a mean of 9 but Set I has observations concentrated around the mean while Set II has observations spread more into the tails. The difference in the levels of variation of the two sets is reflected on their standard deviations. We can say that Set II has a larger variation compared with Set I.

Standard error is also a measure of dispersion. It measures the variation of sample means instead of individual observations. Suppose that there are 1000 plants in the field shown in Fig. 1. If we obtain 10 random samples of 5 plants each and measure mean plant height of the five plants in each of the 10 random samples, we would have a series of 10 mean values of plant height. These 10 mean values also vary from sample to sample. The variation of the means is called "standard deviation of the means" or what is commonly known as "standard error".

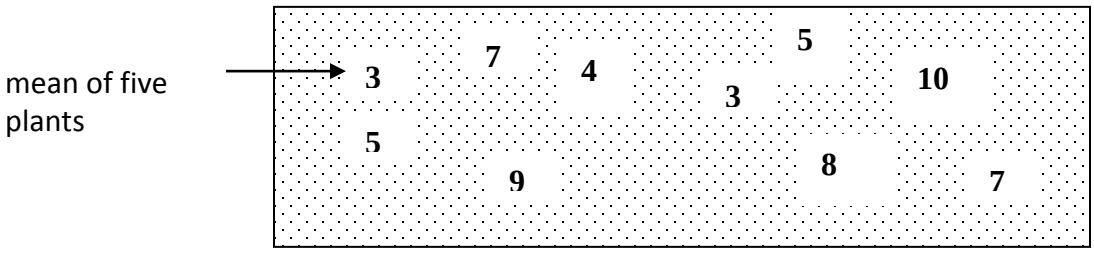


Fig. 1

If you wish to describe the spread of values in the sample, use standard deviation. If you wish to provide information about the reliability of the sample mean, use standard error which describes the variation among individual sample means. It measures how close the average of your samples is to the real mean. Standard error depends on the sample size. The standard error of the mean decreases as the sample size, ***n***, increases**.** If you have a large sample, you would expect your average to be pretty close to the real population mean. If the standard error is large relative to the mean, say 5 ± 4.7 (mean=5 and standard error=4.7), we should not attach too much importance to the sample mean because we are likely to obtain a quite different mean value if we repeated the sample measurements. In journal articles, some papers use standard deviations (SD) to describe the distribution of variables, but others give the standard errors (SE) of the means of the variables.

Bias is the difference in the expected value and the true value of an estimate and results from subjective assignment of treatments to experimental units, subjective scoring, and/or measurements using badly calibrated instruments.

Coefficient of Variation (CV) is a measure of the reliability of an experiment. It expresses the experimental error as a percentage of the mean. The higher the CV the lower the precision associated with the comparison of treatment means.

III Experimental Error

Experimental error is not an error in the sense of being wrong. It results from the natural variation that exists among experimental units or random variations in the procedures used in an experiment. Experimental error is defined as the variability or differences in the experimental units that have the same treatments. In a variety trial, it is the differences among plots that were planted with the same variety. It provides a basis for determining the chances that the observed differences among the varieties are real or not. Steps can be taken to minimize the effect of variation if its cause is known but it can never be totally eliminated. There always will be some variation among experimental units that cannot be controlled.

The presence of inherent variability (or variation) in any experiment is a reality a researcher has to constantly deal with in his quest for new scientific discovery or technological innovations. Whether big or small, there's always going to be natural variation in what is being measured. If variation did not exist, statistical analyses will not be needed.

Take a simple experiment conducted in the field by a researcher who wanted to compare the yields of two varieties of tomato. He planted two tomato varieties, A (recommended) and B (traditional) in two plots, as shown in Fig.2.

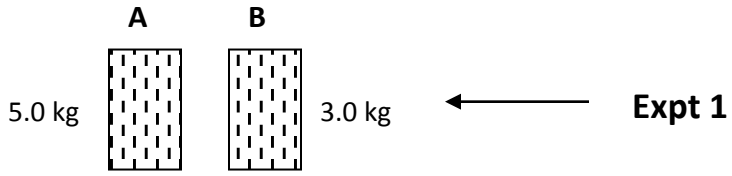


Fig. 2

The researcher applied the same cultural practices on both plots. Yield was measured and compared. Recommended variety A gave higher yield (5kg). Based on this result can 'A' be recommended to the farmers? How can the researcher and the farmers be assured that the yield difference observed between the two plots was due to varietal difference and not due to random variations occurring in the field where the plots were located?

Because of some doubts expressed on results of Expt 1, the researcher replanted the same two plots, both with variety A (Fig.3).

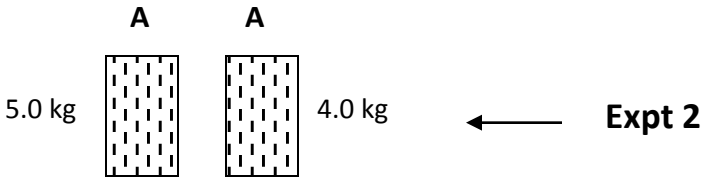


Fig. 3

The researcher found that the two plots in Expt 2 which were planted to exactly the same recommended variety A, and treated identically, also gave different yields (5.0 vs 4.0 kg). This indicates that besides variety, other factors such as soil fertility, soil moisture, disease and insect incidences, etc, also affect yield. This means that the observed difference in yields of A and B plots in Expt 1 could also be due to the natural variations present in the experimental plots, and not due alone to variety used. It is clear from the illustration that any observed difference between any two varieties could be due to confounding effects of various factors, controlled and noncontrolled. How then can the researcher separate the real effect of variety from the effect of other sources of variation?

Let

D = observed yield difference between A and B plots in Expt 1 (varietal effect)

E = observed yield difference between two A plots in Expt 2 (experimental error)

The yield difference between the A and B plots is "significant", only if this difference is "greater" than the expected outcome if the two plots were both planted to A. That is, we can conclude that the yield difference between A and B is significant, and not due to chance (i.e., not due to other sources of variation), if **D** is "substantially" larger than **E.** How "substantial" the difference is can only be decided by criteria provided by a statistical test which involves both D and E.

Since the experimental error is a deciding factor in declaring whether a varietal difference is real or just due to chance, the experiment should be designed so that a measure of experimental error can be obtained. **Only when treatments are replicated, that the experimental error can be measured**.

IV Replication

Replication is the independent application of a treatment to more than one experimental unit. It provides a more precise measure of treatment effects by providing an average of individual observations. It provides a basis for estimating experimental error (also called "noise"), which is needed to determine if varietal differences (also called "signal") are big enough to infer that they are real.

In an experiment where varieties are being evaluated, replication means that each variety has been planted to more than one plot. No test plots are identical and therefore, data collected will vary. The purpose of replication is to allow us to make a more accurate estimate of the varietal performance even though there are uncontrolled variations in the experiment.

Suppose we planted two tomato varieties, a recommended (V1, unshaded) and a local variety (V2, shaded) as shown in Fig. 4, with five plants for each variety. We want to know which of the two varieties will give higher yield.

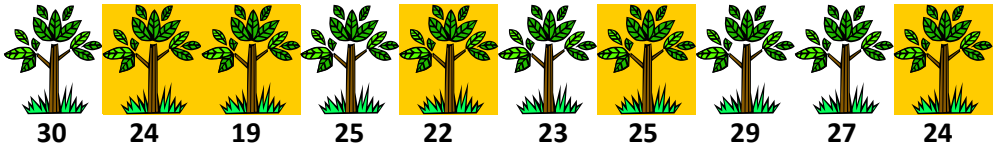


Fig. 4

After harvest, we compare the two varieties. The five recommended V1 plants (unshaded) gave 30, 25, 23, 29, and 27 g/plant with mean=26.8 g, and the V2 local plants (shaded) gave 24, 19, 22, 25, and 24 g/plant with mean=22.8 g. The statistical analysis showed that the recommended variety has significantly higher yield. Without replication we would not be able to analyze the data statistically because not all of the recommended plants gave higher yields than the local varieties.

Suppose we had only two plants (Fig. 5*)*

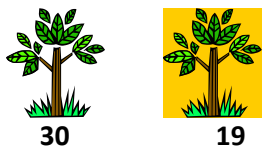


Fig. 5

If the recommended plant gave 30 kg and the local variety, 19 kg, then it does not seem very hard to say that the recommended variety has a better yielding capacity than the local check, as it is expected and consistent with the result in Fig 4.

But suppose the recommended variety had 23 kg and the local variety had 25 (see Fig. 6), we might conclude that the recommended variety does not seem to offer a yield advantage over the local variety, and probably reject it.

23 **25**

Fig. 6

Since varieties in both Fig. 5 and Fig. 6 lack replications, there is no measure of experimental error against which the observed differences can be tested. If the recommended variety was truly higher yielding, the average effect over replications will reflect its real worth. If it is not, the few experimental plots which gave high yields will be negated by the plots with low yields, as illustrated in Fig. 4. It is, therefore, not valid to make any conclusion with only a single observation.

How many replications are needed? There is no hard and fast rule to follow on number of replications to use. In general, the greater the variation expected, the greater the number of replications required. Most often, however, the amount of variation is not known at the start of the experiment. The use of at least three replicates is always suggested, but four or five is better. Found in the Appendix are formulas on how to compute the required number of replications.

Oftentimes, compromises have to be made on the number of treatments and number of replications to include since the size of most experiments is limited by budget or space, etc. Suppose there are only 12 available plots in the field, the combination of number of varieties and replications that can be accommodated is as follows:

Varieties	Replications
6	2
4	3
3	4
2	6

How do you choose from among the four combinations? The number of replications required depends on the variation expected and the degree of precision desired by the researcher. In this example, using more replications would limit the number of

treatments to test but the results are expected to be more reliable and it is better than having unreliable results because of insufficient number of replications.

Pseudo-replication. Replication in the statistical sense refers to different measurements from **independent** experimental units. Consider the illustration in Fig. 7. A researcher is evaluating the yield performance of 4 new varieties of pepper. He sows seeds on each of 4 plots, one variety per plot. At harvest he divided each plot into 3 subunits. The yield in each subunit was measured and labeled correspondingly, as rep1, rep2, and rep3. Are these true replications?

Clearly, this is a case of pseudo-replication. The subunits are not independent of one another. Here, the experimental unit is a plot and the 3 subunit measurements taken are only samples and not replicate values.

At harvest

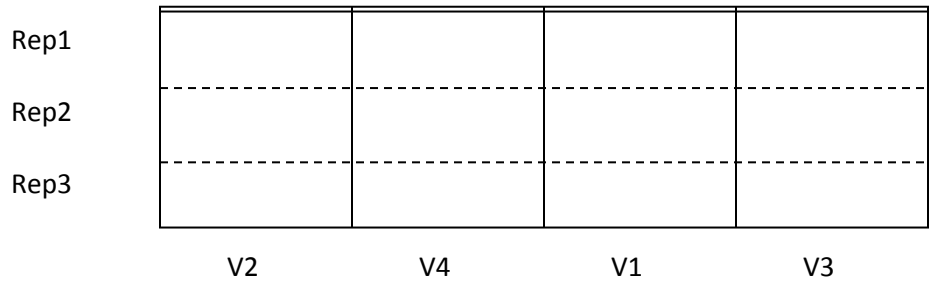


Fig. 7

V Randomization

Randomization is a way of assigning treatments to the experimental units such that each experimental unit has an equal chance of receiving any of the treatments. It avoids biased assignment of treatments. It helps equalize known, unknown, and unpredictable influences on the experiment and allows us to obtain a valid estimate of experimental error. All possible systematic biases cannot be anticipated nor avoided, the rule therefore is to **randomize all experiments.**

Illustration of the importance of randomization. Suppose we want to compare the yield of two sweet pepper varieties. We use five replications (each replicate represented by a plant in this example) and plan to lay out the varieties following either Layouts A, B, or C in Fig. 8. Which layout is correct?

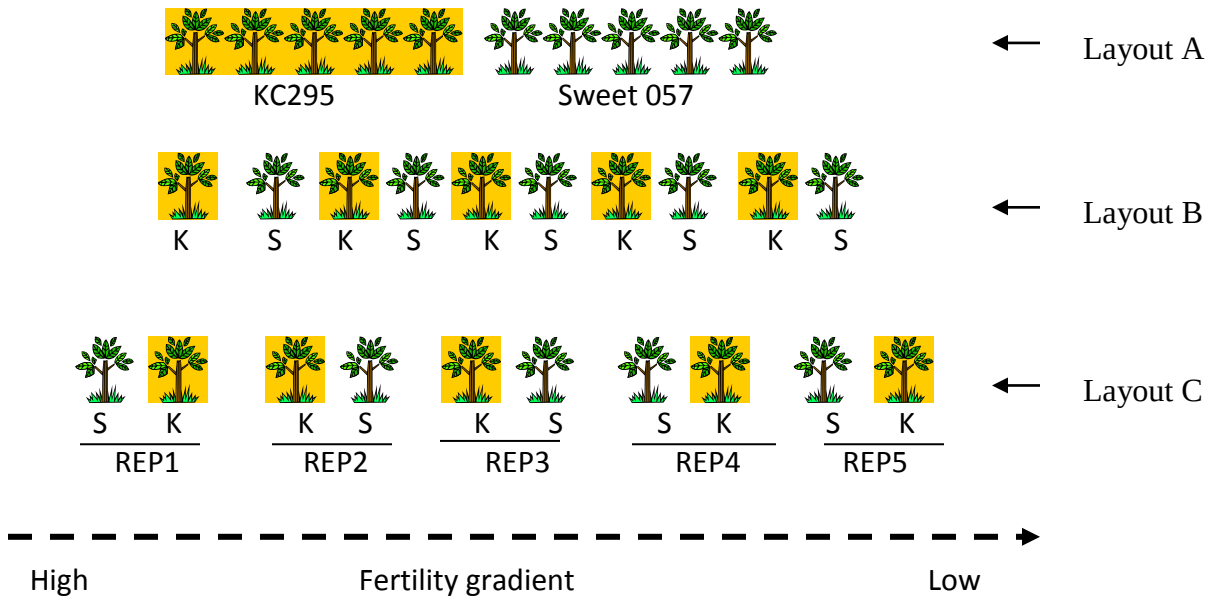


Fig. 8

Neither Layout A nor Layout B is statistically sound. The varieties in Layout A are not randomized, while in Layout B, the varieties are systematically arranged. With the gradual reduction of fertility from left to right, variety KC295 is favored in both layouts because plots assigned to it are located in the portion of the field where fertility is relatively higher. The yield difference between KC295 and Sweet 057 could be due to the difference in the soil fertility level, and not due to the yielding capacity of the varieties. To avoid such bias, the varieties should be assigned at random (Layout C) so that the comparison of the varieties will not be biased in favor of any particular one.

There are many ways to randomize varieties but the simplest is to write the numbers on individual pieces of papers, mix the slips of papers, and then select the slips one at a time without looking, to decide the order of varieties. The table of random numbers, widely available in most Statistics books, can also be used. Many statistical packages now include randomization and field layouting modules. **Never arrange treatments in an order that is merely convenient**. Keep in mind that you can never have a perfect randomization. On rare occasions that you happen to obtain a systematic pattern on your first try, do not hesitate to randomize again.

An example of a systematic pattern to watch out for is shown in Fig. 9:

Layout of an experiment with 6 varieties and 3 reps

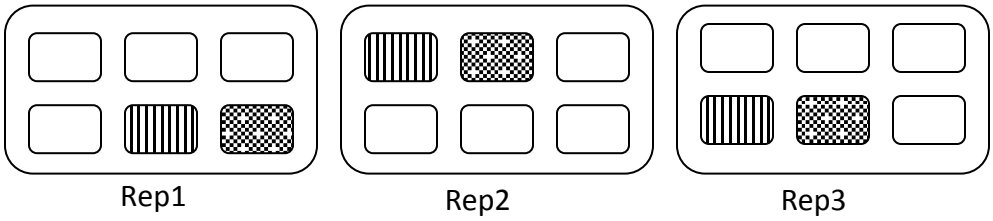


Fig. 9

Note that in all 3 replications, varieties LINE and DOTS are consistently positioned next to each other. Suppose that variety "LINE" is naturally much taller than the other 5 varieties such that it can slightly shade the variety next to it. Variety DOTS would then be unfairly and consistently disadvantaged if this randomization was used because DOTS is always planted next to the tall variety. We want to avoid such biased arrangement because the position of a variety relative to another variety may affect their performances.

VI Error Control

There are two kinds of variation in a variety trial:

1. the variation that can be accounted for in the statistical analysis, such as the variation due to varieties *(varietal effect)*
2. the inherent variation in the experimental units which may be due to factors other than varieties, such as soil heterogeneity, pest and disease incidences, cultural practices, environment, sampling procedures, and other unexplained variations *(experimental error).*

Experimental error is the "deciding factor" in concluding whether an observed varietal difference is <u>real</u> and not just due to chance. We should therefore always strive to control the 2nd type of variation the best way we can. The more homogenous the experimental units are, the smaller the experimental error which results in higher precision and greater chances of detecting differences among test varieties. It is impossible to have complete control over all the sources of variation in the field but it is possible to minimize them.

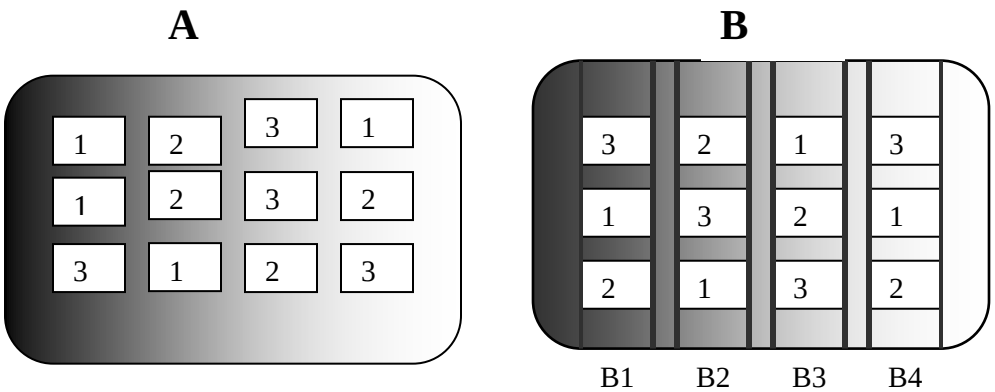
VII Blocking

There are several ways of controlling inherent sources of variation. The most effective is "proper blocking" which refers to the assignment of a group of plots or experimental units into blocks such that the plots or experimental units within a block are as

homogenous as possible. By blocking, we can measure block-to-block variation and then remove that variation from the experimental error, and consequently, reduce its magnitude.

In variety trials, a block would usually contain a set of all varieties. For example, if there are four varieties, each block will have four plots or experimental units, one for each variety. This is what is called a "complete block." A complete block represents one replication. If there are four replications then there are four blocks. The terms "replication" and "block" are commonly used interchangeably, but they are not one and the same and they serve different purposes. The treatments are assigned randomly and independently within each block.

If you can identify or anticipate sources of variation, blocking is an effective tool in reducing experimental error. The illustration in Fig. 10 shows a field with nonhomogenous experimental units. The shaded side has a history of waterlogging. If you assign the varieties completely at random in A, there will be large confounding variations in your data. One known possible source of variation, aside from that due to varieties, is the non-uniform moisture level in the area. Differences due to this source of variation can confound the effect of variety, in which case, any observed difference between varieties is not only due to variety, but could also be due to the effect of excess water on the varieties assigned to the waterlogged plots. This extraneous effect could mask real varietal effect and give unreliable results. Because you know that the possible source of variation is water gradient, you can block the experimental units by using it as the blocking criterion.



Varieties are replicated but **not** blocked

Varieties are replicated and blocked

Fig. 10

The objective of blocking is to make variation from plot to plot within blocks as small as possible while maximizing the variation among blocks. Thus, always orient your blocks perpendicular to the source of variation (B) to maximize the differences among blocks and to minimize the within-block differences.

Variation can be reduced by maintaining uniformity of environmental factors in all plots. The most important ones to be always concerned about, aside from soil characteristics,

are nonuniform conduct of cultural practices, nonuniform disease pressure and insect incidences, plant population, nonuniform measurement techniques, etc. When we standardize measurement techniques we become confident that the measurements are being done in a consistent way which could lead to reduced experimental error (lower noise).

To further minimize variation, consider each block as a unit and do field operations such as transplanting or harvesting by block. Don't start or move to another block before you have completed the other block. Suppose that you have 10 varieties with four replications laid out in four blocks. Each block with 10 plots is one replication. Let's say that due to time constraint, only 25 plots can be harvested in one day. After you have completed the 20 plots on the first two blocks, do not harvest 5 plots on the third block just because there's still time to do it. Rather, harvest two blocks (20 plots) on the first day and the other two blocks (next 20 plots), the following day. This would allow the removal of the variation due to "day" from the experimental error. This should be practiced for all other field operations and not just in harvesting.

If field operations are handled by more than one person, as is the usual case, assign only one person to one or two blocks, but not different persons in one block. In this way, variation due to differences in the way people handle experimental materials and procedures can be removed from the experimental error.

When we use standard measurement techniques we become confident that the measurements are being done in a consistent way which could also lead to reduced experimental error (lower noise).

How to block. A logical sequence in the process of blocking starts with first finding out if there is something (soil fertility, soil type, weed incidence, drainage, shadow, etc) which is not uniform throughout the experimental site that may influence varietal effect. Next is determining its direction. This may not be easy unless a uniformity trial has been conducted previously on that site. For example, there is a moisture gradient forming a slope running along the north-south direction (dark to light shade in Fig. 10), you may use this as basis for blocking.

The orientation of the block should be perpendicular to the source of variation (Fig. 11A). However, if the sources of variation are not known, compact or "squarish" block (Fig. 11B) should be used because plots that are closer are expected to be more alike than those that are farther apart. The use of long and narrow blocks should be avoided when the conditions of soil and other environmental factors are unknown.

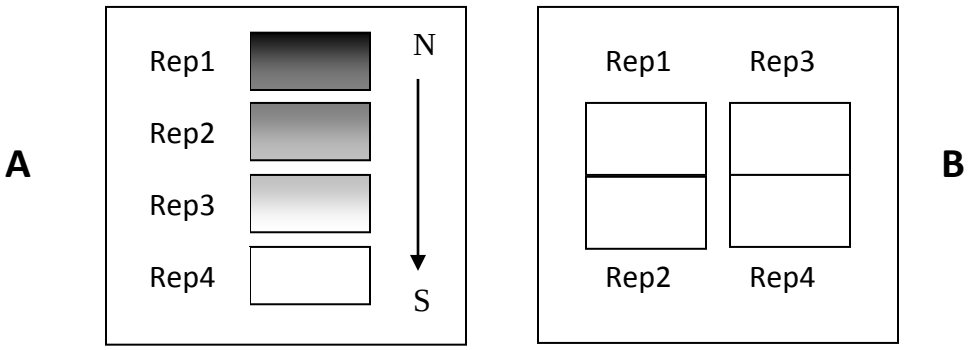


Fig. 11

Greenhouse experiments. Due to absence of sources of variation such as soil heterogeneity, moisture or fertility gradients, some experimenters believe that blocking is not essential in experiments conducted in greenhouses. One will be surprised that in some greenhouse experiments, the variations are as large as, if not higher, than variations in field experiments. The variations found in the greenhouses may not come from the same sources as those in the field, nevertheless, they are present and could mask treatment effects just as much, if not more, than the variations in the fields.

For example, in a greenhouse experiment, the experimental units (pots) are placed on a bench along the east side of the greenhouse. Some pots are exposed to full sunlight, some are half-exposed (gray-shaded), while others are totally unexposed (black-shaded). Assigning treatments completely at random as shown in Fig. 12 results in nonhomogeneous experimental units and thus, large variation.

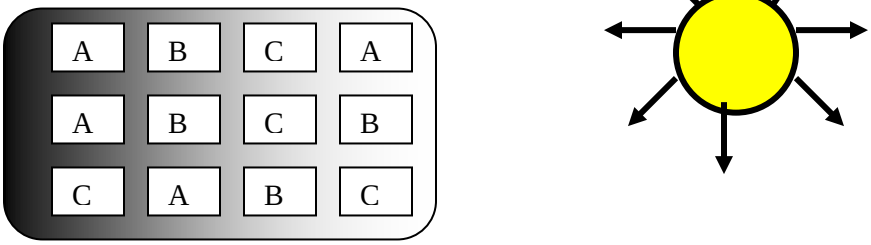


Fig. 12

Blocking, in this case, is essential to remove the variation due to nonuniform light from the experimental error. Use the direction of light source as basis for blocking and orient blocks perpendicular to the light source as shown in Fig 13. In this way, within each block (column), experimental units receive the same amount of light and unexplained variations are minimized. Differences among blocks are larger but can be removed from the experimental error.

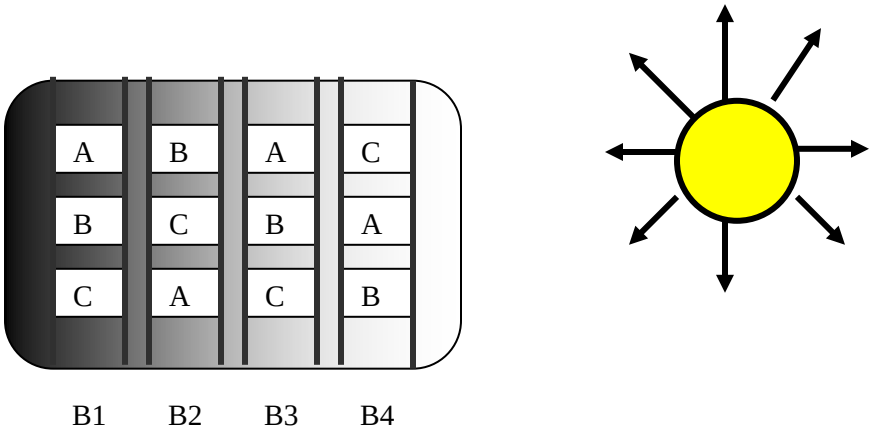


Fig. 13

VIII Types of Experiments

There are two types of experiments as*:*

1. **Single factor experiments** - only one factor varies while all others are kept constant.

   Examples:
   a) A fertilizer trial where several types of fertilizers are evaluated
   b) A variety trial where newly-developed varieties are evaluated
   c) An insecticide trial where several insecticides are evaluated

The results of a single-factor experiment are applicable only to the particular level at which other factors are fixed in the trial. Because of this constraint, single factor experiments are often criticized for their limited application. For example, the results of the experiment involving different organic fertilizers to assess its effect on yield of chili pepper using only one chili pepper variety are applicable only to this particular variety. The effect of the organic fertilizers may be different if other varieties are used.

2. **Multi-factor (or Multi-location) experiments** – In multi- factor experiments, the treatments consist of all possible combinations of selected levels of two or more factors. In multi-location experiments, single factor varietal trials are conducted separately in several locations.

Examples/*contoh*:

a) A 3x2 factorial experiment involving 3 varieties of tomato and 2 grafting treatments conducted in one location

Trt No	Variety	Growth regulator
1	V1	Grafted
2	V1	Non grafted
3	V2	Grafted
4	V2	Non grafted
5	V3	Grafted
6	V3	Non grafted

b) A 2x2x2 factorial experiment involving 2 shelter types, 2 tomato varieties, and 2 types of irrigation

Trt No	Shelter	Variety	Irrigation
1	S1	V1	I1
2	S1	V1	I2
3	S1	V2	I1
4	S1	V2	I2
5	S2	V1	I1
6	S2	V1	I2
7	S2	V2	I1
8	S2	V2	I2

In a multi-factor or factorial experiment, several factors are considered simultaneously and it is possible to test not only the effect of each factor but also the changes in its effect when the levels of the other factors vary. Such changes are known as "interaction effects". See illustration of interaction between variety and fertilizer type in Fig. 14.

IX Concept of Interaction

- Interaction effect between two factors can be measured only if the factors are tested together in the same experiment

- When interaction is present, the effect of a factor changes as the level of the other factor varies

- When interaction is absent, the effect of a factor does not vary when the levels of the other factor changes

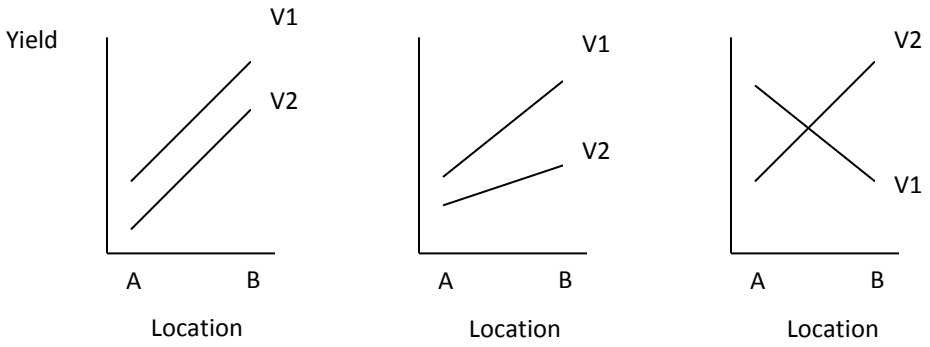


Fig. 14

a) No interaction

b) Interaction is difference in amount of response

c) Interaction is difference in direction of response

Fig. 14 shows the presence and absence of GxE interaction between variety and location. Without GxE interaction (a), varietal effect is the same over the two environments (locations). That is, V1 gives consistently higher yield than V2, whether grown in Location A or in Location B. When GxE interaction is present, varietal effect varies with the location where the varieties are grown. The difference maybe in the amount of varietal response to the environment. In (b), for example, V1 and V2 both respond positively with increased yield in both locations, but the increase in V1 is larger than that in V2; while in (c), the location has an inverse effect on yield. The yield of V2 increased in Location B, while the yield of V1 decreased.

X Factorial Treatments

Oftentimes, you would hear or read that the experiment was conducted using a "factorial design". This is misnomer because such an experimental design does not exist. The term "factorial" only refers to the way the treatments are arranged and not to an experimental design. When two or more factors are involved in an experiment, the combinations of all the levels of these factors are referred to as "factorial treatments". If the factorial treatments are tested in an experiment conducted using RCBD, then we say it is a "factorial experiment in RCBD".

The term "factor" refers to the variables that vary in the experiment, such as variety, types of fertilizer, insecticide, planting density, tillage, etc. "Levels" of a factor are various quantities or aspects of a given factor. In a factorial experiment, all levels of one factor are paired with all the levels of the other factor to form the treatments. The total

number of treatments in a factorial experiment is the product of all the levels of the factors involved in the experiment.
Illustration:

A. Two factors, for example, NPK (with and without) and variety (V1 and V2) give a 2x2 ("two by two") factorial treatments, as follows:

T1 = V1, without NPK (V_1N_0)
T2 = V1, with NPK (V_1N_1)
T3 = V2, without NPK (V_2N_0)
T4 = V2, with NPK (V_2N_1)

B. Three factors, say, *variety* (V1, V2, V3), *fertilizer* (organic, inorganic), and *planting date* (early, late) give a 3x2x2 ("three by two by two") factorial treatments, as follows:

T1=V1, organic, early	**T5**=V2, organic, early	**T9** =V3, organic, early
T2=V1, organic, late	**T6**=V2, organic, late	**T10**=V3, organic, late
T3=V1, inorganic, early	**T7**=V2, inorganic, early	**T11**=V3, inorganic, early
T4 = V1, inorganic, late	**T8**=V2, inorganic, late	**T12**=V3, inorganic, late

Another example: In an experiment to evaluate the yield performance of chili pepper hybrids, management practices such as fertilizer application, irrigation and drainage, and pest management and all other cultural practices were kept constant. The yields of the six hybrids were found to be significantly different, with two of them showing great promise. But since the results of single-factor experiments are expected to hold true only under the specific levels maintained for the other factors, there is no way we can generalize the results to cover the other levels not included in the experiment. – Questions we then ask may be: *Would we get the same resuls if we grow them using different fertilizer regimes? Would the two promising hybrids perform as well?* -- We won't know the answers to these questions, unless we conduct a factorial experiment with hybrids and fertilizer as test factors. Because of the limited range over which results can be generalized based on single-experiments, factorial experiments are now generally used in agricultural research experiments.

In a 2x2 factorial experiment involving 2 fungicide treatments (with and without) and 2 compost tea (with and without), there are a total of 4 factorial treatments, and it is possible to:

(1) compare the chili pepper yield corresponding to the 2 fungicide treatments

(2) compare the chili pepper yield corresponding to the 2 compost tea treatments

(3) determine if the relative effect of the fungicide treatments changes with compost tea treatments applied, or vice versa.

Conducting two separate single-factor experiments involving each of the factors can provide information for (1) and (2), but (3) is only possible by testing both factors in the experiment at the same time.

Adding more levels to a factor quickly increases the number of treatments so include only the levels which are important or practical. For example, using three fungicide treatments and three compost tea treatments, instead of two, would give 3x3=9 factorial treatments, more than double the original number.

XI Groups of Experimental Designs

The two common groups of experimental designs are the following:

a. Single-factor experiments:

1 **C**ompletely **R**andomized **D**esign (CRD)
1 **R**andomized **C**omplete **B**lock **D**esign (RCBD)
2 Latin Square Design
3 Lattice Designs

b. Multi-factor experiments:

1 CRD
2 RCBD
3 Split plot
4 Split-split plot
5 Split-split-split plot

XII Completely Randomized Design (CRD)

CRD is the simplest design. The treatments are replicated but not blocked, rather, they are assigned to experimental units completely at random. Any difference among experimental units receiving the same treatment is considered as experimental error. If homogeneity of the whole experimental area can be assured, there is no need for blocking and CRD can be used. Unfortunately, there is a slim chance that any field would be uniform enough to make the use of blocks unnecessary, so a CRD is rarely used for field trials. CRD is more commonly used in laboratory and greenhouse experiments where environments are believed to be relatively homogenous, although blocking is also often necessary even in the more controlled greenhouse environment.

Illustration: Layout for an experiment in CRD with five treatments and four replications (Fig. 15). There are 5trts x 4reps = 20 experimental units. The treatments are assigned completely at random so each experimental unit has an equal chance of receiving any of the treatments. The treatments in CRD do not have to be replicated the same number of times. Treatments can have unequal number of replicates.

T1	**T4**	**T3**	**T5**
T3	**T2**	**T5**	**T2**
T4	**T1**	**T1**	**T3**
T1	**T5**	**T4**	**T2**
T2	**T4**	**T3**	**T5**

Fig. 15 Layout in CRD, 5 treatments and 4 replications**.**

XIII Randomized Complete Block Design (RCBD)

RCBD is the most widely used design for varietal trials and generally, in agricultural field experiments. It is characterized by blocks, each of which contains a complete set of treatments, one plot for each treatment. The experimental area is divided into blocks and plots are allocated within each block. Each variety must be included once in each block and assigned to each plot at random, independently and separately for each block. RCBD is most effective if patterns of nonhomogeneity or potential sources of variation in the field can be predicted. But even if this information is not available, this design can still be used by making the blocks as square as possible. It is the variabilities among experimental units, that are not due to the source of variation used as blocking criterion, that contribute to experimental error. In RCBD, the variability among blocks can be measured and removed from the experimental error, resulting in higher precision. RCBD can be used for for single-factor or multi-factor (factorial) experiments.

A field layout for single factor experiment in RCBD is shown in Fig.16:
a) Field for an experiment in RCBD with five varieties and four replications b) The field is divided into four blocks, each representing one replication. c) Each block is then divided into five plots. d) varieties are randomly assigned to plots in each block. A different randomization scheme should be used for each block, and each variety should appear once in each block.

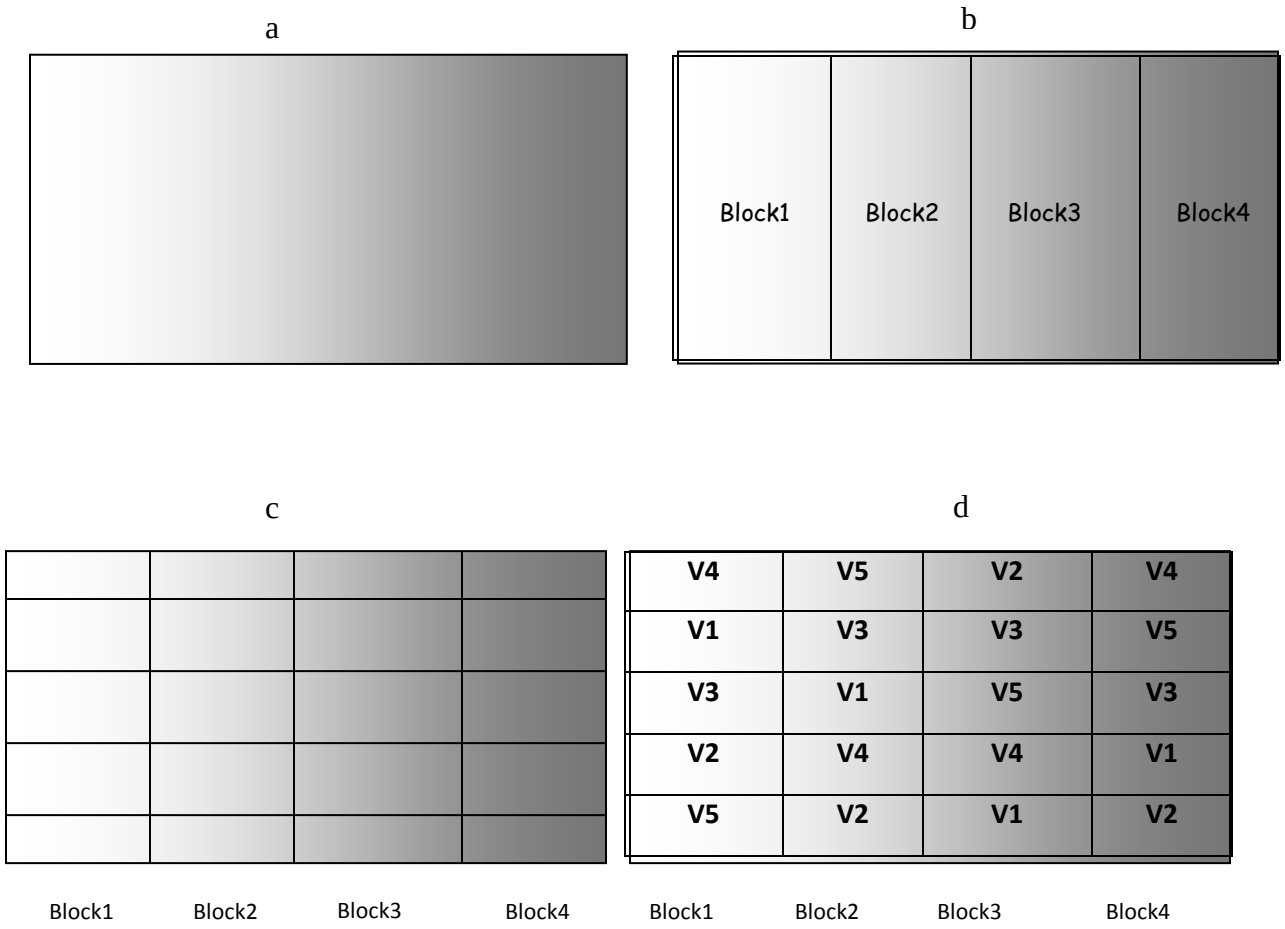


Fig. 16

XIV Factorial Experiments

A factorial experiment is one in which the effects of multiple factors are investigated simultaneously. The treatments consist of all combinations that can be formed from the different factors. An experiment with 4 factors each with 2 levels would result in 2x2x2x2=32 factorial treatments. A factorial experiment can be conducted using RCBD and many other designs. As in single-factor experiments in RCBD, the treatments are assigned randomly to the pool of experimental units, with an equal number of units (reps) for each treatment.

An illustration of field layout for a two-factor experiment in RCBD is shown in Fig. 17

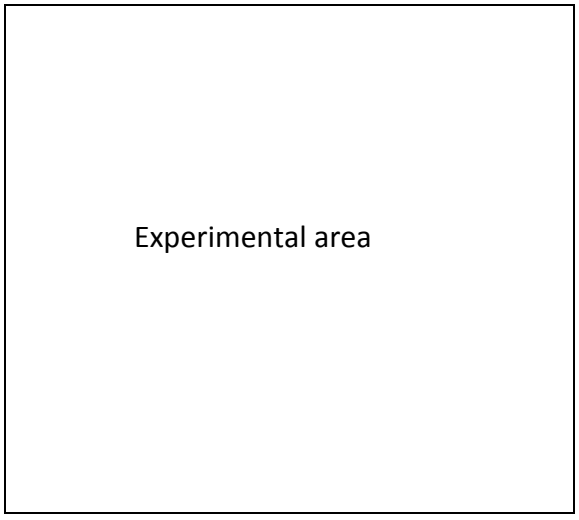
Treatments:
T1: V1, manure 30t/ha - V1M1
T2: V1, manure 20t/ha - V1M2
T3: V1, manure 10t/ha - V1M3

T4: V2, manure 30t/ha - V2M1
T5: V2, manure 20t/ha - V2M2
T6: V2, manure 10t/ha - V2M3
T7: V3, manure 30t/ - V3M1
T8: V3, manure 20t/ha - V3M2
T9: V3, manure 10t/ha - V3M3

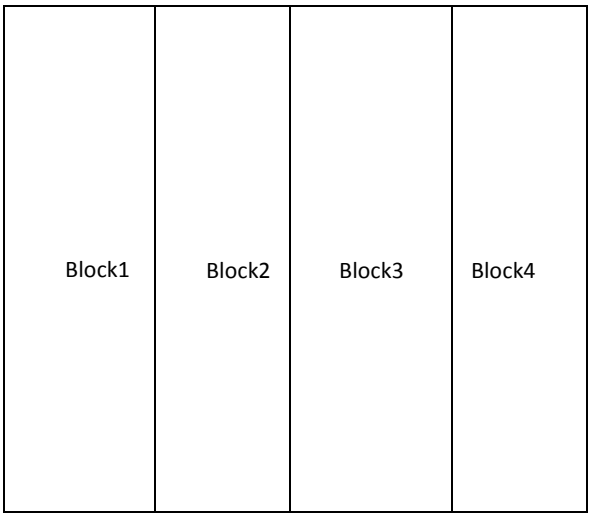
This is a 3x3 factorial experiment in RCBD. There are two factors. The first factor is 3 tomato varieties (V1, V2, and V3), and the second factor is rate of manure application with three levels (10, 20, and 30t/ha). The different combinations of the levels of these two factors form the 3x3=9 factorial treatments in this experiment.

The RCBD can be used to test these factorial treatment combinations. The arrangement of the treatments is basically the same whether it is a single-factor or a factorial experiment. That is, each treatment (single or factorial) is assigned at random once within each block. The field layout in Fig.17 is constructed following steps "a to d" for single factor experiment in RCBD discussed previously.

a) experimental area



b) experimental area is divided into 4 blocks (reps)



c) each block is divided equally into 9 plots

Block1 Block2 Block3 Block4

d) each treatment (a combination of V and M) is assigned at random separately and independently in each block

V3M1	**V2M2**	**V3M3**	**V3M1**
V2M2	**V1M1 *(T1)***	**V2M2**	**V1M3**
V2M2	**V1M2**	**V1M2**	**V2M2**
V1M1 *(T1)*	**V3M1**	**V3M1**	**V3M3**
V1M2	**V2M1**	**V1M3**	**V2M1**
V2M1	**V2M3**	**V2M1**	**V1M1 *(T1)***
V3M3	**V3M3**	**V2M3**	**V2M3**
V3M2	**V3M2**	**V3M2**	**V3M2**
V1M3	**V1M3**	**V1M1 *(T1)***	**V1M2**

Block1 Block2 Block3 Block4

Fig 17. Field layout for a multi-factor experiment in RCBD with 3x3=9 factorial treatment combinations

XV Split Plot Design

The split plot designs (SP) are used in experiments involving factorial treatments in which one of the factors, possibly because of the nature of the experimental materials or operations involved, requires bigger plots.

For example, a researcher might be interested in knowing the effect of four tillage practices (T1, T2, T3, T4) on yield of three crop varieties (V1, V2, and V3). For efficient application of the tillage treatments, larger plots are needed. Varieties, on the other hand, do not need large plots. This is an example of two factors to be tested in the same experiment having different required plot sizes (Fig. 18). Although RCBD can be used in this experiment, its requirement to have all the factorial combinations assigned at random, using the same plot size for all, is not convenient for this experiment. SP should be used instead. The tillage factor levels are assigned to the larger plots called "mainplots" and varieties, to the smaller plots within each mainplot which are called "subplots".

In split plot design, the randomization is accomplished by first assigning the mainplot factor at random to the mainplots, separately for each replication. The subplot factor

levels are then randomly assigned to the subplots within each mainplot, using separate randomization for each mainplot.

Because there are two sizes of plots, there are two experimental errors in split plot design, one for each size of plot. The subplot factor and its interaction with the mainplot factor are generally associated with smaller experimental error. They are, therefore, estimated with precision higher than that for the mainplot factor which, due to its larger size of plot, has lesser degree of precision. In effect, the precision for the mainplot factor is sacrificed to have greater precision for the subplot factor. This feature of the SP should be taken into consideration when deciding to use it. Assign the more important factor to the subplot for greater precision and the less important one to the mainplot. If both factors are equally important, use RCBD. The mainplot treatments in a split plot design are measured with less precision than that in RCBD.

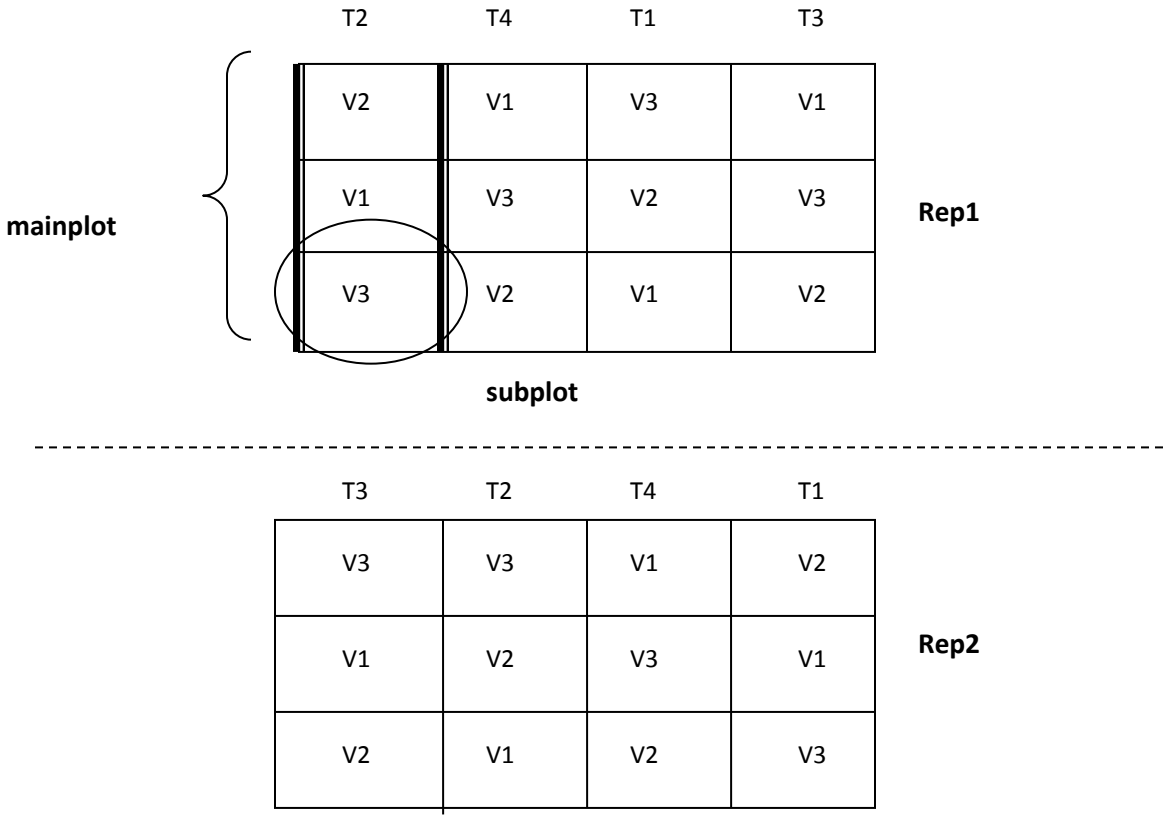


Fig 18. Field layout for a split plot design with 4 tillage practices as mainplot treatments and 3 varieties as subplot treaments

XVI Variety trials/Multi-location trials

Varieties differ in many aspects – yielding capacity, adaptation, quality, pest/disease resistance, nutrients, etc. The objective of a variety trial is to identify varieties that

perform consistently and exceptionally better than others. Any type of experimental design may be used for a variety trial. The most common is the randomized complete block design (RCBD). The result of a variety trial conducted in a single location is expected to hold true only in that location since no variety performs well in all environments. The performance of a variety is affected by environmental conditions prevailing in the locations where it is planted, such as rainfall, temperature, soil, insect/disease, crop management, etc. Hence, a variety with outstanding performance in one location may perform poorly in another location.

Breeders are interested in varieties with consistent superior performance based on the desired traits over a wide range of environments. Environment may refer to location, season, year, or a combination of these factors. To evaluate the adaptation of varieties in differing environments, multi-location trials are conducted. In multi-location trials, the performance of crop cultivars can be evaluated by comparing their means across locations. This is carried out by performing a combined analysis of variance and comparing the variety means across locations.

XVII Genotype x Environment Interaction

When the genotype and environment interact the test varieties will fail to give the same relative performance for a given trait in all environments. An environment may represent a location, a season, or year, or a combination of two or three of these factors. See illustrations in Fig. 19a, 19b, and 19c.

a)

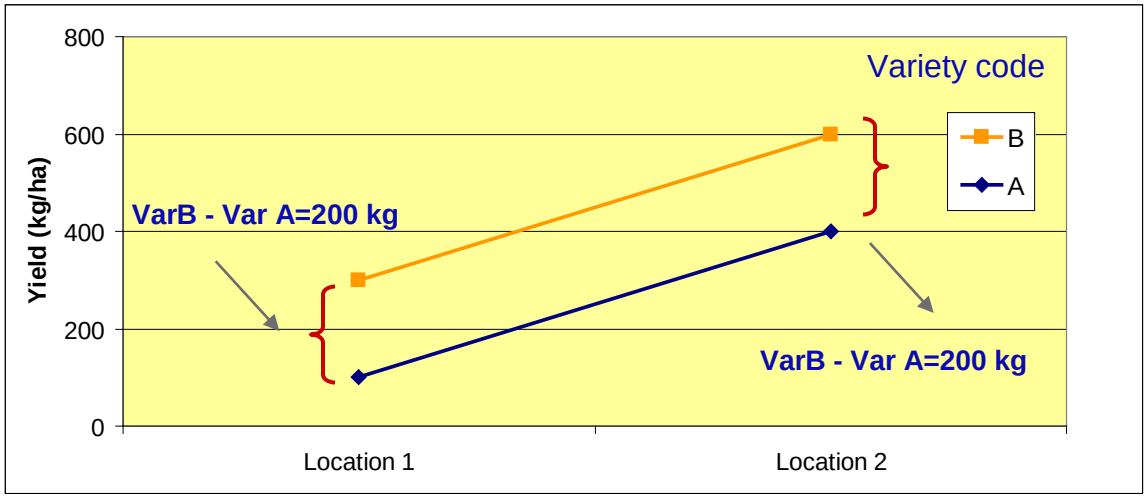


Fig.19a: No interaction varieties have the same relative performance for a given trait (say, yield) in both locations; ranking of variety is unchanged between locations.

b)

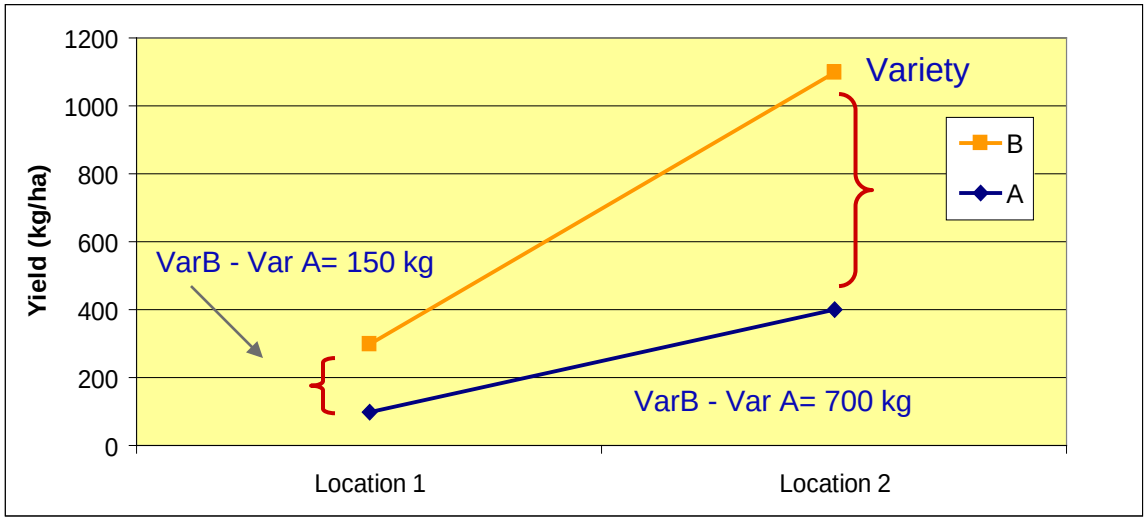


Fig. 19b. With interaction - (Case 1): Yield of varieties varies greatly between locations; ranking of variety is unchanged between locations.

c)

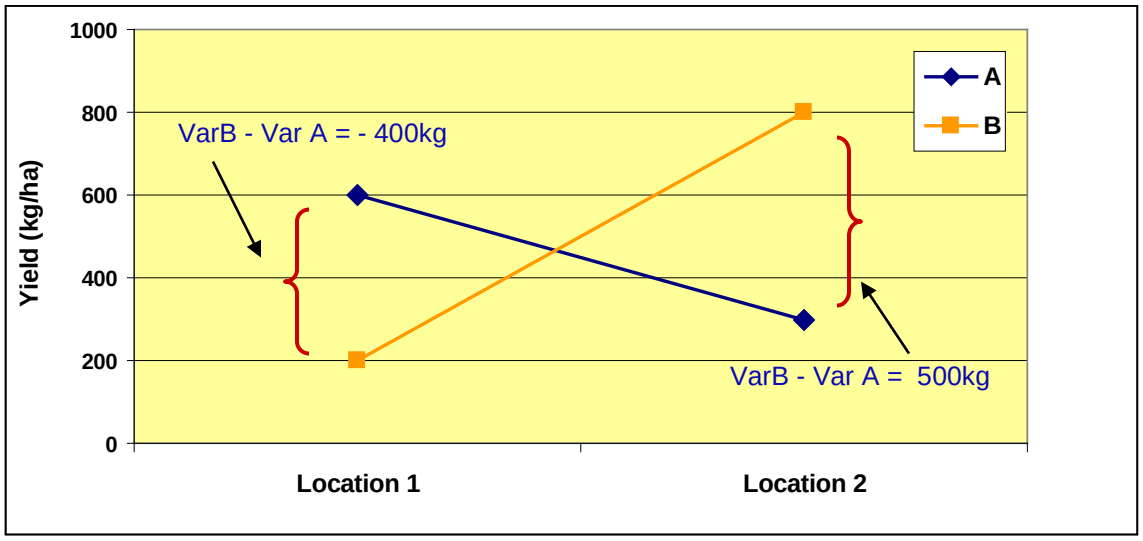


Fig. 19c. With interaction - (Case 2): Yield of varieties varies greatly between locations; ranking of variety changes between locations.

ANALYSIS OF VARIANCE

The purpose of the statistical analysis that follows after data have been collected is to provide answers to questions and objectives of the experiment. If the experiment is well-planned based on statistically valid procedures and the mean comparison to be performed is decided well ahead and taken into consideration in the planning stage, data analysis will not be a complicated process.

The analysis of variance (ANOVA) is a procedure that partitions the total variation (TSS) into different sources based on the layout of the experiment in Fig .20. Each source of variation is tested for significance. The principle behind experimental design and ANOVA is to identify the sources of variation and formulate the proper tests to compare them.

Y1 ***T1***	Y2 ***T4***	Y3 ***T2***	Y4 ***T3***	Y5 ***T5***
Y6 ***T3***	Y7 ***T1***	Y8 ***T5***	Y9 ***T2***	Y10 ***T4***

Rep 1 (Block1)

Rep 2 (Block2)

Fig. 20 Layout of an experiment with 5 varieties, 2 reps

Yield is measured on each of the 10 experimental plots and denoted as:

$$Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7, Y_8, Y_9, Y_{10}$$

where Y1-Y5 are yields of treatments T1, T4 ,T2, T3, and T5 respectively, in Rep I, and Y6-Y10 are yields of treatments T3, T1 ,T5, T2, and T4 respectively, in Rep II.

Case1:

$Y_1 = Y_2 = Y_3 = Y_4 = Y_5 = Y_6 = Y_7 = Y_8 = Y_9 = Y_{10} =$ ' 4 '

(because yields of all plots are equal to 4, no variation exists)

Case2:

$Y_1 = 2, Y_2 = 3, Y_3 = 4, Y_4 = 1, Y_5 = 3, Y_6 = 5, Y_7 = 3, Y_8 = 6, Y_9 = 4, Y_{10} = 2$

(because yields of plots vary, variation exists)

Because of the inherent variation present in the fields where experiments are conducted, even if the same treatment is applied and all other factors are kept constant, variation in yields will still be present. The total variation among the 10 yield values in Case2 can be measured, and is called the Total Sum of Squares (TSS). TSS is due to many sources. Most of the sources of variation can be identified, and its contribution to TSS measured through statistical procedures. In Case1, TSS=0 because no variation was observed. In Case2, TSS≠0 because variations were observed among the Y values.

The total variation in the data set (Fig.21), as measured by TSS in RCBD, can be partitioned by ANOVA into:

a) Sum of squares due to treatments (TrSS) is a measure of differences among treatments
b) Sum of squares due to blocks (BSS) is a measure of differences between block (or reps)
c) Sum of squares due to experimental error (ESS) is a measure of differences among replications within a treatment. We may, symbolically, write the relationship as:

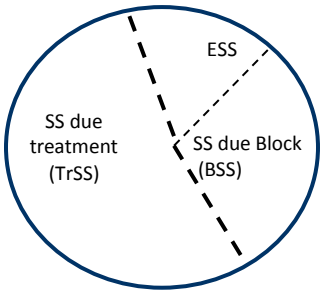


Fig. 21 Partitioning scheme for TSS in RCBD

We may, symbolically, write the relationship as:

$$TSS = TrSS + BSS + ESS$$

Variations within treatments (experimental error)

Within T1:	Between Y1 (rep1) and Y7 (rep2)
Within T2:	Between Y3 (rep1) and Y9 (rep2)
Within T3:	Between Y4 (rep1) and Y6 (rep2)
Within T4:	Between Y2 (rep1) and Y10 (rep2)
Within T5:	Between Y5 (rep1) and Y8 (rep2)

Variations among treatments (treatment effects)

Treatment mean yields: T1 Mean(Y1,Y7) vs T2 Mean(Y3,Y9) vs T3 Mean(Y4,Y6) vs T4 Mean(Y2,Y10) vs T5 Mean(Y5,Y8)

Variations among blocks (block effects)

Block (Rep) mean yields: Block1 Mean (Y1, Y2, Y3, Y4, Y5) vs Block2 Mean(Y6, Y7, Y8, Y9, Y10)

The different *sources of variation* are dictated by the experimental design used in the experiment. Without blocking, both BSS and ESS become part of the experimental error as shown in the partitioning scheme of the sources of variation in CRD (Fig. 22).

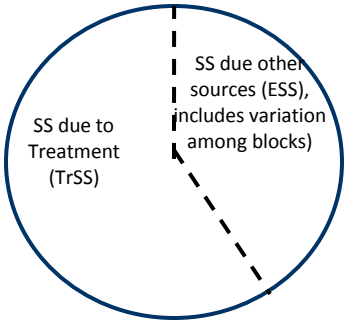


Fig. 22 Partitioning scheme for TSS in CRD

In RCBD, it is possible to remove the portion due to variation among blocks (BSS) from the experimental error, resulting in much reduced experimental error. The smaller the experimental error, the higher is the precision.

There are three sources of variation (SV) in RCBD (Table 1): a) the variation due to treatments, b) the variation due to blocks (replications), and c) the random variations due to unknown sources (called "experimental error").

Table 1. ANOVA format for RCBD (single factor), with t treatments and *r* replications

Randomized Complete Block Design (RCBD)

SV	df	SS	MS	F
Replication	r-1	BSS	BMS = BSS / (r-1)	BMS / EMS
Treatment	t-1	TrSS	TrMS = TrSS / (t-1)	TrMS / EMS
Expt. Error	(r-1)(t-1)	ESS	EMS = ESS / (r-1)(t-1)	
Total	rt-1	TSS		

Most statistical software provide straightforward ANOVA table outputs, but require that you know before hand (1) what design you are using, (2) what formulas are used to test the hypotheses, (3) and how to generate those tests.

To ensure valid and appropriate analyses, the ANOVA table format with the sources of variation and degrees of freedom should be formulated at the time of design and before the experiment is executed.

To check the appropriateness of your analysis you should generate the ANOVA table for your experiment based on the experimental design you are using. The sums of squares and degrees of freedom should add up and F-tests should use the proper error terms.

There are three sources of variation, based on RCBD among the 5tx2r = 10 observations in the field trial example above: 1) the variation due to treatment (which may be indicated as "TrMS"), 2) the variation due to blocks (BMS), and 3) the random variations due to unknown sources (EMS or "experimental error"). The ratio between the two variations TrMS/EMS will indicate whether the observed mean differences are due to real treatment effects or due to chance (Table1). The ratio BMS/EMS tests whether the blocking used is effective in reducing experimental error.

The test of hypothesis of "no treatment difference" can, therefore, be performed by comparing TrMS and EMS. If there are no differences in treatment means, TrMS and EMS should be very similar. Otherwise, we suspect that the observed differences are due to treatments.

XVIII Statistical Hypothesis

Before selecting the appropriate *experimental design* within which the different treatment conditions will be evaluated, the research hypotheses based on the objectives of an experiment should first be formulated.

The experimental objectives should be written clearly and in the order of priority as many important decisions in the experiment are arrived at based on the objectives. The choice of treatments and the experimental design to be used, the data to be collected, the data analysis to be performed, and the presentation of results all depend on the objectives of the study.

For example, the major problems in one area are related to soil deficiency -- high salinity, low soil pH, poor soil structure, NPK deficiency, low soil nutrient availability, and in another area it is insect and disease incidences. Based on this knowledge important research objectives can be identified in each location. With clear objectives the treatments can then be identified, the experimental design chosen, and the measurements to be undertaken ascertained. The data analysis to be performed even before the actual experiment is conducted can be determined to ensure that the objectives can be satisfied.

Generally, the objectives involve determining the effect of treatments on the expected response, such as yield, plant height, resistance to pest or disease, etc. After the objectives have been identified, these should be translated in the form of hypotheses which can be tested statistically.

Research hypotheses are questions of interest to the researcher based on what is known about a particular theory or phenomenon. The next stage in experimentation is the translation of the different research hypotheses into a set of *treatment conditions* which can be proven as true or false using some statistical tests, (such as the F-test), through ANOVA. There are two types of statistical hypotheses:

1. Null hypothesis (H_O) - states that no differences exist among treatment means /
2. Alternative hypothesis (H_a) - contradicts the null hypothesis /

Examples:

a) Objective: To determine the effect of organic fertilizer application on yield of chili pepper

Treatments: lime/manure and compost/manure

Null hypothesis: Ho: $Y_{LM} = Y_{CM}$ (No difference between yields of lime/manure and compost/manure)

Alternative hypothesis: Ha: $Y_{LM} \neq Y_{CM}$

If the F-test in the ANOVA based on the ratio TrMS/EMS is significant, we reject "Ho" and conclude that the yield Y_{LM} is significantly different from Y_{CM} (maybe higher or lower). If the F-test is not significant, we cannot reject the "Ho". It does not prove, however, that the treatments are the same. It only means that the test was not able to detect the difference between treatments.

b. Objective: To assess the yield performance of three pepper hybrids
Treatments: Three pepper hybrids

Null hypothesis: Ho: Y1 = Y2 = Y3 (The three pepper hybrids have equal yields)
Alternative hypothesis: Ha: $Y1 \neq Y2 \neq Y3$

c. Objective: To evaluate the effect of fertilizer application on yield and other yield parameters of tomato

Null hypothesis: Ho: $Yw = Yw/o$ (Yields are equal whether or not fertilizer is applied)
Alternative hypothesis: Ha: $Yw \neq Yw/o$

In variety trials, the stated objectives as listed in the protocol should be directly related to the advancement criteria that have been set. Based on these criteria, decision can be made on whether or not to advance a test variety. Typically, the performance of the test variety is evaluated relative to a check variety. So both the test varieties and check varieties needed to make the assessments should be tested together in the trial. Clear objectives associated with your advancement criteria will help you determine which varieties to include in your trial and what data need to be collected.

XIX Single-factor experiment in RCBD

The following data on fruit length (cm) of pepper were measured from an experiment laid out in RCBD with 3 manure treatments and 4 replications (Table2). The objective of the experiment is to determine the type of manure treatment which would result in longer fruit length. The null hypothesis is Ho: $T_1 = T_2 = T_3$.

Table 2. Data on fruit length of chili pepper

Manure treatments (t/ha)	Rep	Fruit length (cm)
T1	1	9.8
T1	2	10.2
T1	3	10.4
T1	4	9.6
T2	1	12.8
T2	2	13.1
T2	3	13.9
T2	4	14.5
T3	1	16.2
T3	2	15.9
T3	3	18.1
T3	4	17.1

The ANOVA (Table 3) for the data found in Table 2 can be performed using CROPSTAT.

Table 3. Example – ANOVA based on RCBD

ANOVA for fruit length (cm), 3 manure treatments and 4 reps

SV	df	SS	MS	F	Pr>F
Rep	3	2.8800	0.9600	2.55 ^{ns}	0.1514
Treatment	2	93.2317	46.6158	124.03**	<0.0001
Error	6	2.255	0.3758		
Total	11	98.3667			

C.V. = 4.6 %

******** *significant at P<0.01;* ^{ns} *– not significant*

From Table 3, TrMS = 46.6158
BMS = 0.9600
EMS = 0.3758

Fc for treatment effect = TrMS/EMS = 46.6158/0.3758 = 124.03
Fc for rep effect = BMS/EMS = 0.9600/0.3758 = 2.55

The probability (P-value) in the ANOVA is a measure used to decide whether the null hypothesis is true or false. If true, the null hypothesis is to be accepted, and if false, the null hypothesis is to be rejected. The smaller the P-value the more confident we will be in rejecting the null hypothesis and declaring as significant the mean differences. Since Pr>F = 0.0001 for treatment is less than 0.01, we reject the null hypothesis and conclude that the mean fruit lengths of the 3 treatments are significantly different at 1% level.

A significance level denoted by "P<0.01" means that there is less than 1% probability (i.e., less than 1 in 100) that any observed differences among the treatment means could occur by chance, and are not due to real treatmentl effects, while "P<0.05" means that the probability is less than 5%, (or less than 5 in 100). No matter what the experimental design is, the general criteria listed in Table 4 are used in determining the significance of treatment effects.

Table 4. Criteria for significance

Condition	**Varietal Effect**
$P < 0.01$	**
$0.01 \leq P < 0.05$	*
$P \geq 0.05$,	ns

If the experimental error is large relative to the variation due to treatment, the Pr>F for treatment effect will be greater than 0.05. It indicates the failure of the test to detect significant differences, and the null hypothesis of no treatment mean differences, (Ho: T1=T2=T3), cannot be rejected. However, even if the F-test is nonsignificant, it does not mean that all treatments are the same. There are three possibilities why the F-test is not significant --- 1) the experimental error is large and treatment differences are large; 2) the experimental error is large and treatment differences are small; 3) the experimental error is small and treatment differences are small. Whenever nonsignificant test is obtained, the experimental error and the treatment differences should be examined to determine if the trial is worth conducting all over again. The trial can be repeated if case (1) above is observed. Focus in the repeat experiment should be directed toward reducing experimental error so that the precision and power of the test are increased.

The CV stands for "coefficient of variation" and should always be presented in the analysis. It indicates the degree of precision of the experiment. It tells us how reliable

the experiment is. Since the CV expresses the variation as a percentage of the mean, the lower the CV, the higher is the reliability of the experiment.

The CV is computed as:

$$CV\ (\%) = \left(\frac{\sqrt{EMS}}{\bar{X}} \right) \times 100$$

There is no hard and fast rule on what to consider a high or low CV value. The magnitude of the CV depends on the crop, type of experiment, and the character measured. Only a researcher who has considerable experience working with a crop would have a good judgment on the acceptability of the CV value.

Proper blocking minimizes differences among plots within a block and maximizes differences among blocks. To determine if the goal of blocking is satisfied, we examine the significance of the rep effect in Table 3. The Pr>F for the rep effect (P=0.1514) is greater than 0.05 which indicates that differences among rep means are not significant. The nonsignificant rep effect implies that the differences among blocks have not been maximized and therefore, the blocking used was not effective in reducing experimental error. A new blocking criterion or orientation should be considered in the next trial if the same experimental area is going to be used.

The data on disease incidence in Table 5 come from a greenhouse varietal screening trial. The ANOVA results are shown in Table 6. The objective was to evaluate the different pepper varieties for resistance to late blight. The experiment used RCBD with three replications.

Table 5. Data on disease incidence

Variety No.	Rep 1	Rep 2	Rep 3
066	0.20	0.32	0.60
067	0.48	0.55	0.78
1367	0.37	0.40	0.87
375	0.45	0.60	0.75
385	0.17	0.25	0.40
535	0.05	0.10	0.30
631A	0.08	0.15	0.22
631B	0.36	0.47	0.67
743	0.20	0.30	0.48

Tabel 6. ANOVA RCBD

SV	df	SS	MS	F	Pr>F
Replication (B)	2	0.4325	0.2162	53.20 **	<0.0001
Variety (V)	8	0.7780	0.0973	23.93**	<0.0001
Error	16	0.0650	0.0041		
Total	26	1.2755			

*C.V. = 16.3 %, ** significant at P<0.01*

From Table 6, VMS = 0.0973
BMS = 0.2162
EMS = 0.0041

Fc for Varietal effect = VMS/EMS = 0.0973/0.0041 = 23.93
Fc for Replication effect = BMS/EMS = 0.2162/0.0041 = 53.20

Since Pr>F for variety is less than 0.01, we can reject the null hypothesis, (Ho: V1 = V2 =…..= V9) and conclude that the mean disease incidences of the nine pepper varieties are significantly different at 1% level. The replication effect is also highly significant which indicates that blocking in this case was effective in reducing experimental error.

XX Factorial experiment in RCBD

In a factorial experiment laid out in RCBD, the source of variation accounted for by treatment sum of squares (TrSS) is further subdivided into components due to variations in each of the factors involved in the factorial experiment, and their interactions (Table7).

Table 7. Format ANOVA RCBD (multi factor)

Randomized Complete Block Design (RCBD)

SV	df	SS	MS	F
Replication	r-1	BSS	BMS = BSS / (r-1)	BMS / EMS
Treatment	t-1	TrSS	TrMS = TrSS / (t-1)	TrMS / EMS
Factor1 (N)	(n-1)	NSS	NMS = NSS/(n-1)	NMS / EMS
Factor2 (V)	(v-1)	VSS	VMS = VSS/(v-1)	VMS / EMS
N x V	(n-1)(v-1)	NVSS	NVMS = NVSS/(n-1)(v-1)	NVMS / EMS
Error	(r-1)(t-1)	ESS	EMS = ESS / (r-1)(t-1)	
Total	rt-1	TSS		

The following data on total yield (t/ha) of chili pepper (Table 8) were measured from an experiment laid out in RCBD with 3 rates of NPK (N) and 3 chili pepper varieties (V) as treatments, with 4 replications. The objective of the experiment was to determine the effect of NPK and varieties on yield. Based on the objective, the statistical hypotheses are:

Ho: $N_1 = N_2 = N_3$, $V_1 = V_2 = V_3$, NxV interaction = 0

Ha: $N_1 \neq N_2 \neq N_3$, $V_1 \neq V_2 \neq V_3$, NxV interaction $\neq 0$

Table 8: Data on total yield (t/ha) of chili pepper *using 3 different rates of NPK*

Variety	NPK	rep	Total yield (t/ha)
1	30	1	20.47
1	30	2	20.67
1	30	3	21.08
1	30	4	22.00
1	20	1	26.43
1	20	2	24.67
1	20	3	18.51
1	20	4	22.00
1	10	1	16.04
1	10	2	14.67
1	10	3	12.08
1	10	4	15.00
2	30	1	39.34
2	30	2	29.08
2	30	3	29.11
2	30	4	35.00
2	20	1	17.23
2	20	2	14.09
2	20	3	11.95
2	20	4	15.09
2	10	1	39.59
2	10	2	36.27
2	10	3	19.38
2	10	4	34.10
3	30	1	33.51
3	30	2	31.77
3	30	3	29.12
3	30	4	28.98
3	20	1	31.13
3	20	2	32.46
3	20	3	26.96
3	20	4	30.21
3	10	1	29.88
3	10	2	35.60
3	10	3	23.24
3	10	4	25.56

The ANOVA for data in Table 8 is shown in Table 9.

SV	df	SS	MS	F	Pr > F
Replication	3	235.6637	78.5546	7.32**	0.0012
Treatment	8	1729.7008	216.2126	20.15**	<0.0001
NPK (N)	2	681.5225	340.7612	31.76**	<0.0001
Variety (V)	2	201.5795	100.7897	9.39**	0.0010
N x V	4	846.5988	211.6497	19.73**	<0.0001
Error	24	257.5093	10.7296		
Total	35	2222.8737			

C.V. = 12.9%

Tabel 9. ANOVA for RCBD with factorial treatments consisting of 3 NPK treatments and 5 varieties, in 4 replications.

From Table 9, NMS = 340.7612
VMS = 100.7897
NVMS = 211.6497
EMS = 10.7296

Fc for main effect of NPK = NMS/EMS = 340.7612/10.7296 = 31.76
Fc for main effect of variety = VMS/EMS = 100.7897/10.7296 = 9.39
Fc for NxV interaction effect= NVMS/EMS = 211.6497/10.7296 = 19.73

Since Pr>F are < 0.01 for all effects, we reject the null hypotheses (Ho: $N_1= N_2=N_3$, $V_1=V_2=V_3$, NxV interaction = 0) and conclude that the main effect of NPK, the main effect of variety, and their interaction are all highly significant (P<0.01). Since NxV interaction is significant, the NPK means can only be compared separately for each variety; and similarly the variety means can only be compared at each rate of NPK applied.

The replication effect is also highly significant which indicates that blocking was effective in reducing experimental error.

XXI Analysis for Split Plot Design

The analysis based on split plot design is done in two stages: the mainplot analysis and the subplot analysis. The ANOVA format for the split plot design is shown in Table 10. As in RCBD with factorial treatments, the analysis provides information on the main effect of individual factors as well as the interaction effect between these two factors.

The first-stage analysis is done on the mainplot factor. The significance of the block effect and that of the main factor is tested using the error associated with the mainplot (EaMS). The second-stage analysis involves the main effect of the subplot factor and its interaction with the mainplot factor. Both effects are tested using the error associated with the subplot (EbMS).

Table 10. ANOVA format for split plot design

SV	df	SS	MS	F
Replication	r-1	RSS	RMS = RSS / r-1	RMS / EaMS
Mainplot (A)	a-1	ASS	AMS = ASS / a-1	AMS / EaMS
Error (a)	(r-1) (a-1)	EaSS	EaMS = EaSS / (r-1)(a-1)	
Subplot (B)	b - 1	BSS	BMS = BSS / b-1	BMS / EbMS
A x B	(a - 1)(b - 1)	ABSS	ABMS = ABSS / (a-1)(b-1)	ABMS / EbMS
Error (b)	a (r-1) (b-1)	EbSS	EbMS = EbSS / a (r-1)(b-1)	
Total	rab-1	TSS		

The yield data shown in Table 11 come from an experiment laid out in split plot design with variety as mainplot factor and grafting (w/ and w/out) as subplot factor, in four replications.

Table 11. Data from a trial based on split plot design

Variety	Grafting	Rep	Yield (t/ha
TLCV15	Grafted	1	39.3
TLCV15	Nografted	1	14.5
FMT847	Grafted	1	40.4
FMT847	Nografted	1	39.9
CHT501	Grafted	1	26.2
CHT501	Nografted	1	6.2
TLCV15	Grafted	2	42.5
TLCV15	Nografted	2	25.1
FMT847	Grafted	2	41.6
FMT847	Nografted	2	16.8
CHT501	Grafted	2	25.9
CHT501	Nografted	2	4.1
TLCV15	Grafted	3	37.0
TLCV15	Nografted	3	21.0
FMT847	Grafted	3	41.1
FMT847	Nografted	3	43.5
CHT501	Grafted	3	26.8
CHT501	Nografted	3	0.2
TLCV15	Grafted	4	43.4
TLCV15	Nografted	4	5.7
FMT847	Grafted	4	41.5

FMT847	Nografted	4	16.1
CHT501	Grafted	4	32.2
CHT501	Nografted	4	13.0

The result of the ANOVA (Table 12) indicates that both the main effects of grafting (G) and variety (V) are significant, at P<0.01 and P<0.05, respectively, while the interaction effect between variety and grafting (VxG) is not significant. Note that variety and rep effects are tested against EaMS while the main effect of grafting and its interaction with variety are tested against EbMS. Note also that there are two CV values in a split plot design: the CV(a) for the mainplot and CV(b) for the subplot, which indicate the levels of precision associated with the corresponding factors. The CV(b) is, generally, smaller than CV(a) which reflects the higher level of precision usually associated with the subplot factor.

Table 12. ANOVA based on Split Plot design for tomato yield (t/ha), with variety as mainplot factor and grafting as subplot factor.

SV	df	SS	MS	F	Pr > F
Replication	3	35.3367	11.7789	0.15 ^{ns}	0.9231
Variety (V)	2	1373.6108	686.8054	9.01 *	0.0156
Error (a)	6	457.5758	76.2626		
Grafting (G)	1	2238.8017	2238.8017	39.83 **	0.0001
V x G	2	161.6308	80.8154	1.44 ^{ns}	0.2872
Error (b)	9	505.9375	56.2153		
Total	23	4772.8933			

C.V.(a) = 32.5% ; C.V. (b) = 27.9%

**,* significant at P< 0.01 and P<0.05, respectively

^{ns} not significant

From Table 12, Mainplot MS (variety) = 686.8054
Subplot MS (grafting) = 2238.8017
Interaction MS (variety x grafting) = 80.8154
BlkMS (rep) = 11.7789
EaMS = 76.2626
EbMS = 56.2153

Fc for Rep =	BlkMS/EaMS = 11.7789/76.2626 = 0.15
Fc for Variety =	Mainplot MS/EaMS = 686.8054/76.2626 = 9.01
Fc for Grafting =	Subplot MS/EbMS = 2238.8017/56.2153 = 39.83
Fc for V x G =	Interaction MS/EbMS = 80.8154/56.2153 = 1.44

Since Pr>F for grafting (0.0001) is less than 0.01 and Pr>F for variety (0.0156) is less than 0.05, we reject the null hypotheses, (Ho: G1 = G2 and Ho: V1 = V2 = V3), for the main effects of the two factors and conclude that the yields of the three tomato varieties are significantly different at P<0.05 and the effect of grafting is significant at P<0.01. For the VxG interaction, the P>F = 0.2872 is greater than 0.05 indicating that it is not significant. The nonsignificant VxG interaction means that the effect of grafting on yield is the same for all three varieties and/or varietal effect is the same whether the plants are grafted or not. The P>F for replication effect (0.9231) is greater than 0.05 and also not significant which indicates that blocking was not effective in reducing experimental error.

XXII Analysis of Data from Multi-location Variety Trials.

Before performing a combined analysis of data from different locations (Table 13), an ANOVA should be run for each location and the experimental errors examined for heterogeneity. Depending on the results of these preliminary steps, the combined analysis can be carried out.

Table 13. SV and DF for individual and combined ANOVA

RCBD 6 varieties, 4 reps, 2 locations

Individual analysis

	Degrees of freedom (DF)	
Source of variation (SV)	Location1	Location2
Rep	3	3
Variety	5	5
Error	15	15
Total	23	23

Combined analysis over 2 locations

SV	DF
Location (L)	1
Reps w/in location	6
Variety (V)	5
V x L	5
Pooled error	30
Total	47

XXIII COMPARISON OF MEANS

The ANOVA is not the final step in data analysis. To properly interpret the statistical results, the table of means based on significant main or interaction effects of treatments, including the standard errors of the means are needed. The table of means

rather than the ANOVA table is oftentimes the one required and presented in report write-ups.

When the ANOVA is performed and the null hypothesis is rejected, then it can be concluded that the treatment means under consideration are significantly different. The ANOVA result, however, does not say which means are significantly different, and which are not. After ANOVA, one then proceeds to comparison of means using any of the methods of comparing means.

There are several methods of comparing means

a) Planned pairwise comparison - usually performed for specific pairs of treatments identified before the start of the experiment

   Example:
   - Control vs each of the fertilizer/starter solution treatments
   - Traditional variety vs each of the recommended varieties
   - No N application vs different N-rates applied

The Least Significant Difference (LSD) provides a single LSD value, at a prescribed level of significance, which serves as a boundary between significant and nonsignificant difference between any pair of treatment means. It is recommended for use only when the F-test is significant. The LSD is computed as follows:

Any pair of varieties with mean difference greater than the LSD value is declared significantly different at the specified level of significance. There is no need to manually compute the LSD value as most statistical software provides this as output.

$$\mathbf{LSD}_{\alpha} = (t_{\alpha})\ \text{SED}$$

SED=standard error of the difference

$$\text{SED} = \sqrt{\frac{2\ (\text{EMS})}{r}}$$

$t_{\alpha, n}$ = the tabular t value at α level of significance, EMS is the error mean square, and n = error degrees of freedom from the ANOVA table.

Example:

Table 14. Mean comparison using LSD

Organic fertilizer	No of reps	Mean yield *(t/ha)*	Mean difference
Compost (C)	3	41.2	
Manure (M)	3	40.6	C-M = 41.2 – 40.6 = 0.6 ^{ns}
Lime (L)	3	27.8	C-L = 41.2 – 27.8 = 13.4 **
			M-L = 40.6 – 27.8 = 12.8 **
LSD (0.05) = 3.9			
(0.01) = 5.6			

** Significant at P<0.01 by LSD; ^{ns} not significant

Mean differences which are greater than the LSD values are statistically significant. Since the differences 'C-L' (13.4) and 'M-L''(12.8) are greater than LSD(0.01)=5.6, these differences are declared significant at P<0.01. The difference 'C-M' (0.6) is considered not significant because it is less than LSD(0.05)=3.9 (Table 14).

Several methods of multiple comparisons such as the Duncan's multiple range tests (DMRT), Tukey's multiple-comparison method, Bonferroni, Scheffe', etc., are generally available in most statistical packages and are used when all possible pairs of treatments are compared to identify pairs which are significantly different.

Example:
- Comparison of the yield performance of new hybrid varieties
- Comparison of resistance to bacterial wilt of different tomato cultivars

b) Unplanned pairwise comparison – performed when each pair of treatments is compared to identify pairs which are significantly different

Several methods of multiple comparisons such as the Duncan's multiple range tests (DMRT), Tukey's multiple-comparison method, Bonferroni, Scheffe', Keuls, Newman, Tukey, etc., are generally available in most statistical packages and are used when all possible pairs of treatments are compared to identify pairs which are significantly different.

Example:
- Comparison of the yield performance of new hybrid varieties
- Comparison of resistance to leaf blight of different tomato cultivars
- Comparison of several herbicides in controlling weeds

The Duncan's Multiple Range Test (DMRT) is applicable to an unplanned pair comparison. It is useful in experiments that require the comparison of all possible pairs

of treatment means. Unlike LSD, DMRT does not use a single value with which to compare all differences between pairs of means to declare significance. In DMRT, the means are arranged in an array from highest to lowest (or lowest to highest), and varying comparison values -- larger for pairs of means which are farther apart in the array -- are used to compare all possible pairs of means.

In Table 15, the mean yields of the different varieties have been grouped by DMRT in

Variety	No. of reps	Mean Yield [1/]	DMRT Grouping
Andulus	3	99.0	a
PV2	3	87.3	b
BlueStar	3	83.7	b
PV3	3	83.7	b
9852-191	3	75.7	bc
PBC438	3	69.0	c
PV1	3	68.3	c
PBC843	3	63.7	c
PV5	3	49.0	d
PV4	3	47.3	d

several overlapping and non-overlapping groups, with means arranged in descending order. Means with common letters are not significantly different at 5% or 1% levels of significance. Andulus had the highest mean yield and the only variety assigned the letter “a”, which indicates that it is significantly higher than the yields of all other test varieties. The mean yields of PV2, BlueStar, PV3, and 9852-191 are not significantly different from each other because they all have the common letter “b”, but are significantly higher than the rest of the varieties except 9852-191 which has yield not significantly different from PBC438, PV1, and PBC843 because of their common letter, “c”. PV4 and PV5, both with common letter ”d”, gave yields that are not significantly different from each other, but significantly lower than the rest of the varieties tested.

What method to use is a matter of choice. Multiple range tests are generally not recommended for mean comparison when treatments are structured. To compare treatments with structure, it is best to use LSD and orthogonal contrasts based on the partitioning of treatment sum of squares (PSS). *PSS is not covered in this course.*

REFERENCES

Box, G.E.P., W.G Hunter and J.S. Hunter.1978.Statistics for Experimenters, An Introduction to Design, Data Analysis, and Model Building. John Wiley and Sons. USA

Gomez, K. A. and A. A. Gomez. 1984. Statistical Procedures for Agricultural Research. John Wiley and Sons, Ltd. New York.

Little, T. M. and F Jackson Hills. 1978. Agricultural Experimentation. John Wiley and sons, Ltd. New York.

Mead, R. and R. N. Curnow.1983. Statistical Methods in Agriculture and Experimental Biology. J.W. Arrowsmith, Ltd. Bristol, U.K.

Ott, L. R. and M. Longnecker. 2001. Statistical Methods and Data Analysis. Wadsworth Group. USA.

Pearce, S. C. 1983. The Agricultural Field Experiment- A statistical Examination of Theory and Practice. John Wiley and Sons, Ltd. New York.

Petersen, R. G. 1984. Agricultural Field Experiments. Marcel Dekker, Inc. New York.

Scott, M. E. and H. D. Delaney. 2003. Designing Experiments and Analyzing data. Lawrence Erlbaum Assoc Inc Dimensions. USA

Sheskin, D.J. 2004. Handbook of parametric and Nonparametric Statistical Procedures. 3rd ed. Chapman & Hall/CRC Press. Florida, USA.

Snedecor, G. W. and W. G. Cochran. 1973. Statistical Methods. The Iowa State University Press. Ames, Iowa, USA

Steel, R. G. D. and James H.Torre. 1960. Principles and Procedures of Statistics with special reference to the Biological Sciences. McGraw-Hill Book Company, Inc. USA.

Weiss, N.A. 2005. Introductory Statistics. 7th ed. Pearson Education, Inc. USA.

DETERMINATION OF THE NUMBER OF REPLICATIONS [1]

For a given experiment, the number of replications required depends on (a) the level of experimental error expected to be encountered, and (b) the degree of precision desired; both of which are based on the most important response character (say, yield).

The magnitude of an experimental error can be represented by the coefficient of variation (cv); and the degree of precision represented by either the standard error of the treatment mean (s.e. $\bar{x}$) or the standard error of the treatment difference (s.e.d.)

For a RCB design, the two standard errors are computed as:

$$\text{s.e.}\,\bar{x} = \sqrt{\frac{s^2}{r}}$$

and

$$\text{s.e.d.} = \sqrt{\frac{2s^2}{r}}$$

where s^2 = error mean square in the ANOVA and r = no. of replications

Since cv = $\frac{100s}{\bar{x}}$, the standard error of the mean can be written as:

$$\text{s.e.}\,\bar{x} = \frac{(cv)(\bar{x})}{100\sqrt{r}}$$

or

$$r = \left[\frac{(cv)(\bar{x})}{100(s.e.\bar{x})}\right]^2 \quad \ldots\ldots\ldots\ldots (1)$$

where $\bar{x}$ = grand mean / *dimana* $\bar{x}$ *= rata-rata keseluruhan*

And, the standard error of the treatment difference can be written as

$$\text{s.e.d.} = 2\left[\frac{(cv)(\bar{x})}{100(r)}\right]^2$$

or

$$r = \left[\frac{(cv)(\bar{x})}{100(s.e.d.)}\right] \quad \ldots\ldots\ldots\ldots\ldots (2)$$

Example 1. For an experiment with a RCB design in which the expected cv of yield is 10%, the mean yield level is 6 t/ha. The researcher wishes to have the standard error of a treatment mean of not more than 0.3 t/ha; the number of replications needed can be computed, using equation (1) above, as follows:

$$r = \left[\frac{(10)(6)}{100(0.3)}\right]^2 = 4$$

where cv = 10 %; $\bar{x}$ = 6 t/ha; and s.e.$\bar{x}$ = 0.3 t/ha

Example 2. Assuming the same experiment as in Example 39 but the researcher wishes to prescribe the degree of precision in terms of least significant difference (LSD) that can be detected between any pair of means of 0.8 t/ha. In this case the s.e.d. value should first be approximated as:

$$\text{s.e.d} = \frac{LSD}{2} = \frac{0.8}{2} = 0.4$$

From equation (2)

$$r = 2\left[\frac{(10)(6)}{100(0.4)}\right]^2 = 4.5 \text{ or } 5$$

Source: Biometrics Unit, IRRI