STATISTICS AND EXPERIMENTAL DESIGN

WORKING MANUAL

Roger Petersen

Technical Manual 11 En

International Center for Agircultural Research in the Dry Areas (ICARDA), Box 5466, Aleppo, Syria

ICARDA 1985, reprinted 1987, 1988

This document was reported to AGRIS and is included in the data base under reference number 87-000106

Contents

DEFINITIONS OF TERMS

1. A measurable characteristic of an individual is called a variable. Examples of variables are yields of wheat plots, weights of animals, heights of barley plants and numbers of kernels per head. There are two kinds of variables:

   1. Discrete, or count, variables take only integral values for example, number of diseased plants in a plot .

   2. Continuous, or measured, variables which can take any value over a small range. An example is the weight of grain from a plot of wheat.

2. A single measurement of a variable is called an observation of that variable. An example is the yield of one plot of barley in an experiment.

3. The observations in a particular set, or group, of observations are called the data in that set. All of the yields from a single wheat experiment are the data from that experiment.

4. A population consists of all possible observations of a variable. The limits of a population should be carefully defined, but once these limits are set any observation within the limits is a member of the population. For example, a variable might be the yield of a hectare of wheat. The population might then be defined as wheat yields on all of the hectares in Syria on which wheat is grown. The yield of any hectare of wheat in Syria would be a member of the population.

5. A sample is a set of observations taken from a population. If yields of wheat per hectare in Syria are the population wheat yields per hectare in Aleppo province could be a sample from the population.

6. A parameter is a summary number used to describe a population. For a given population a parameter is a constant, fixed value. For example, the average yield per hectare would be a parameter of the population of wheat yields in Syria.

7. A <u>statistic</u> is a summary number used to describe a sample. For a given population a statistic is a variable because its value will change from one sample to another within the population.

8. An <u>experiment</u> is a planned investigation to discover new facts or to confirm or deny the results of previous investigations. For example, to find the most suitable wheat variety for the rainfed area of Syria we could conduct an experiment which included a large number of varieties and hope to be able to select the one which is best.

9. A <u>treatment</u> is a procedure whose effect on the experimental material is to be measured. In a variety trial each variety would be a different treatment. The addition of nitrogen fertilizer to a plot would be a treatment in an agronomic experiment. In some experiments doing nothing at all might be one of the treatments.

10. <u>An experimental unit</u> is the piece of experimental material on which one treatment is applied. In a variety trial the plot would be the experimental unit, while in a grazing trial a single pasture might be the unit.

11. A <u>sampling unit</u> is the fraction of the experimental unit on which the effect of the treatment is measured. If the four center rows of a six row plot are harvested for yield the four rows would be the sampling unit.

12. An <u>experimental design</u> is a set of rules by which the treatments to be used in an experiment are assigned to the experimental units.

13. When a treatment appears more than once in an experiment, the treatment is said to be <u>replicated</u>.

14. A group of uniform experimental units is called a <u>Block</u>. In many agricultural experiments each treatment is assigned once, and only once, in each block and, hence, one replication of the set of all treatments occurs in a block. For this reason a block is often referred to as a replication in agriculture. This is not strictly correct because in some experimental designs the units are not grouped into blocks even th ough each treatment occurs more than once.

In other designs there are fewer units in each block than there are treatments in the experiment.

15. The variable which is measured on an experimenal unit is often called the <u>yield</u>. The weight of grain on a plot, for example, would be the yield of that plot.

BASIC STATISTICAL COMPUTATIONS

Symbols and Subscripts : Suppose we have a list of numbers:
3, 5, 4, 3, 2, 6, 5, 8, 7, 4

We use a letter, such as x, y, z, to stand for any number in the list. We indicate a particular number in the list by putting a subscript on the symbol for the number. If x stands for any number in the above list x_i would stand for the number in the position designated by the subscript i. for this list:

$x_1 = 3$	$x_6 = 6$
$x_2 = 5$	$x_7 = 5$
$x_3 = 4$	$x_8 = 8$
$x_4 = 3$	$x_9 = 7$
$x_5 = 2$	$x_{10} = 4$

We usually use the letter n to stand for the number of values in the list. For the above list n=10. We also sometimes use the letters m,p,r to stand for the count of numbers in a list.

We can choose any symbol to designate the entries in a list, and we can use any letter to indicate the count of numbers in the list. For example, it would be equally correct to use y; for a number in the above list, and p to stand for the count. In this case $y_4 = 3$, $y_6 = 6$, and p= 10.

Often, in statistics, we have to deal with a set of numbers arranged in rows and columns. An example is the array of numbers

2	5	3	2
1	4	2	5
3	4	1	2

This is an array with three rows and four columns. We can use subscripted letters to stand for any number in the array.

For example we might use x_{ij} for this purpose. In this symbol the first subscript is the number of the row in which the number is found, while the second subscript designates the column. For the above array

$$x_{11} = 2, \quad x_{12} = 5, \quad x_{13} = 3, \quad x_{14} = 2$$

$$x_{21} = 1, \quad x_{22} = 4, \quad x_{23} = 2, \quad x_{24} = 5$$

$$x_{31} = 3, \quad x_{32} = 4, \quad x_{33} = 1, \quad x_{34} = 2$$

<u>Summation notation</u> : One of the most frequently used arithmetic operations in statistics is finding the sum of a list or array of numbers. We indicate this operation using summation notation. Suppose we have the list

$$x_1, x_2, x_3, x_4, x_5, x_6$$

where x_i stands for any of the 6 entries in the list. We can indicate the total of all of the numbers by the symbol

$$\sum_{i=1}^{6} x_i = x_1+x_2+x_3+x_4+x_5+x_6$$

this symbol means : start with the number whose subscript is one, add to it the number whose subscript is two, and continue adding numbers, increasing the subscript each time by one, until the number whose subscript is six has been added. As an example, suppose we have the list

4, 1,43, 2, 5, 2, 3

and let y_j stand for any number in the list. For this list the total number of entries is n=8. We can indicate the sum of all of the numbers in the list by the symbol

$$\sum_{j=1}^{8} y_j$$

this symbol stands for the following operation :

$$\sum_{j=1}^{8} y_j = y_1+y_2+y_3+y_4+y_5+y_6+y_7+y_8$$

substituting the numbers for their symbols we have

$$\sum_{j=1}^{8} y_j = 4+1+4+3+2+5+2+3 = 24$$

A common operation in statistical analysis is to find the sum of the squares of a list of numbers. We can indicate this operation using summation notation. If x_i stands for any number in a list of n numbers, the sum of squares of all of the numbers is symbolized

$$\sum_{i=1}^{n} x_i^2 = x_1^2 + x_2^2 + \ldots + x_n^2$$

Given the list, symbolized by x_i :

1, 2, 6, 3, 4

the sum of squares of these numbers is

$$\sum_{i=1}^{5} x_i^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 = 1^2 + 2^2 + 6^2 + 3^2 + 4^2 =$$

$$1+4+36+9+16 = 66$$

be careful to distinguish between the symbol

$$\sum_{i=1}^{n} x_i^2 \quad \text{and the symbol} \quad \left(\sum_{i=1}^{n} x_i\right)^2.$$

The first symbol says to square the numbers before they are added, while the second says to first find the sum then square the sum. For the above list we have

$$\sum_{i=1}^{5} x_i^2 = 1^2 + 2^2 + 6^2 + 3^2 + 4^2 = 66$$

$$\left(\sum_{i=1}^{5} x_i\right)^2 = (1+2+6+3+4)^2 = (12)^2 = 144$$

We can also use summation notation to indicate various sums of numbers in a two-way array. Suppose that x_{ij} is the symbol for any of the numbers in the following array

2	2	4	3	5
3	1	2	2	4
4	5	3	1	2

This array has p= 3 rows and r= 5 columns. To indicate the sum of all of the numbers we would use

$$\sum_{i=1}^{3}\sum_{j=1}^{5} x_{ij} = \sum_{i=1}^{3} (x_{i1}+ x_{i2}+ x_{i3}+x_{i4}+ x_{i5})$$

$$= (x_{11}+x_{12}+x_{13}+x_{14}+x_{15}) + (x_{21}+x_{22}+x_{23}+x_{24}+x_{25}) + (x_{31}+x_{32}+x_{33}+x_{34}+x_{35}) .$$

Substituting the numbers for their symbols we have

$$\sum_{i=1}^{3}\sum_{j=1}^{5} x_{ij}= (2 +2 +4 +3 +5) + (3 +1 +2 +2 +4) + (4 +5 +3 +1 +2) = 16 +12+ 15 = 43$$

Similarly,

$$\sum_{i=1}^{3}\sum_{j=1}^{5} x_{ij}^2 = (2^2+2^2+4^2+3^2+5^2) + (3^2+1^2+2^2+2^2+4^2) +$$

$$(4^2+5^2+3^2+1^2+ 2^2) = (4+ 4+ 16+ 9+ 25) +(9+ 1+ 4+ 4+ 16)+$$

$$(16+ 25+ 9+ 1+ 4) = 50+ 34+ 55 = 139$$

Note the difference between $\sum_{i=1}^{3}\sum_{j=1}^{5} x_{ij}^2$, and

$$\sum_{i=1}^{3} (\sum_{j=i}^{5} x_{ij})^2= (2+ 2+4+3+5)^2+ (3+1+2+2+4)^2 + (4+5+3+1+2)^2= (16)^2+(12)^2+(15)^2=$$

$$256+144+225 = 625$$

And

$$\left(\sum_{i=1}^{3}\sum_{j=1}^{5} x_{ij}\right)^2 = \left[(2+2+4+3+5) + (3+1+2+2+4) + (4+5+3+1+2)\right]^2$$

$$= \left[16+12+15\right]^2 = (43)^2 = 1849$$

BASIC STATISTICS

In every statistical analysis three types of statistics are almost always computed. These are :

Means, Variances and Standard Errors of the means.

Means : The mean of a set of n observations, $x_1, x_2, \ldots x_n$ is the sum of the observations divided by the number of observations. The mean is usually symbolized by

$$\bar{x} = (x_1+x_2+\ldots+x_n)/n.$$

We express the mean in summation notation as

$$\bar{x} = \left(\sum_{i=1}^{n} x_i\right)/n$$

Given the set of observations, symbolized by x_i :

3, 6, 2, 5, 4, 3

we see that, for this set, n=6. The mean, $\bar{x}$, of the set is

$$\bar{x} = \left(\sum_{i=1}^{6} x_i\right)/n = (x_1+x_2+x_3+x_4+x_5+x_6)/6$$

$$= (3+6+2+5+4+3+)/6 = 23/6 = 3.83$$

Variances : The variance of a set of n observations, x_1, $x_2, \ldots, x_n$, is the sum of squares of the difference between the observations and their mean divided by one less than the number of observations.

The variance is usually symbolized by s^2 (s-squared) using summation notation, the defining formula for the variance is

$$s^2 = \sum_{i=1}^{n} (x_i-\bar{x})^2/(n-1)$$

Note that it can be shown that

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2/n$$

which is often an easier form to use in computer

We should examine this expression in detail :

$\sum_{i=1}^{n} x_i^2 = x_1^2 + x_2^2 + \ldots + x_n^2$. This term is called the Raw, or Uncorrected, sum of squares.

$(\sum_{i=1}^{n} x_i)^2/n = (x_1 + x_2 + \ldots + x_n)^2/n$. This term is called the Correction term, symbolized c.t.

The raw sum of squares minus the correction term is called Sum of squares, symbolized SS.

The divisor of s^2, n-1, is called the Degree of Freedom symbolized by d.f.

From this we see that the following relationships hold :

$$s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2/(n-1) = \left[\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2/n \right] /(n-1)$$

$$= \left[\text{Raw ss} - \text{c.t.} \right] /(n-1) = SS/d.f.$$

Suppose we compute the variance, s^2, for the set of numbers, 3, 6, 2, 5, 4, 3, for which we computed the mean :

1. $\sum_{i=1}^{6} x_i = 3+6+2+5+4+3 = 23$
2. $C.T. = (\sum_{i=1}^{6} x_i)^2/n = (23)^2/6 = 529/6 = 88.17$
3. $\sum_{i=1}^{6} x_i^2 = 3^2+6^2+2^2+5^2+4^2+3^2 = 99.00$ = Raw S.S.
4. SS = Raw ss - C.T. = 99.00-88.17= 10.83
5. s^2 = SS/d.f. = SS/(n-1) = 10.83/(6-1)= 2.17

Standard Errors : The Standard error of the mean of n observations, $x_1, x_2, \ldots, x_n$, is the square root of the variance of the observations divided by the number of observations. The standard error of the mean is usually symbolized by $s_{\bar{x}}$. By definition we have

$$s_{\bar{x}} = \sqrt{s^2/n}$$

For the set of data we have been examining we have $s^2 = 2.17$ and $n=6$. The standard error is

$$s_{\bar{x}} = \sqrt{s^2/n} = \sqrt{2.17/6} = \sqrt{.36} = 0.60$$

EXERCISES

For the following exercises let

$\{x_i\}$ = 2, 1, 6, 3, 3, 4, 7, 6, 8, 2

$\{y_j\}$ = 6, 2, 5, 3, 8, 2, 8

(Z_{ij}) =

2	4	1	3	5
1	1	3	6	2
5	2	5	1	1
3	2	5	1	4

1. What is the numerical value of

$x_1 =$ ________

$x_3 =$ ________

$x_{10} =$ ________

$x_{12} =$ ________

2. What is the num erical value of

$Z_{14} =$ ________

$Z_{35} =$ ________

$Z_{44} =$ ________

3. What is the numberical value of

$$\sum_{i=1}^{10} x_i = \_\_\_\_\_\_\_\_$$

$$\sum_{i=1}^{10} x_i^2 = \_\_\_\_\_\_\_\_$$

$$\left(\sum_{i=1}^{10} x_i \right)^2 = \_\_\_\_\_\_\_\_$$

$$\sum_{i=1}^{7} x_i = \_\_\_\_\_\_\_\_$$

4. What is the numerical value of

$$\sum_{i,j=1}^{6} x_i y_j = \_\_\_\_\_\_\_\_$$

$$\sum_{i=1}^{4} \sum_{j=1}^{5} Z_{ij} = \_\_\_\_\_\_\_\_$$

$$\sum_{i=1}^{4} \sum_{j=1}^{5} Z^2_{ij} = \_\_\_\_\_\_\_\_$$

$$\sum_{i=1}^{4} \left(\sum_{j=1}^{5} Z_{ij} \right)^2 = \_\_\_\_\_\_\_\_$$

$$\sum_{j=1}^{5} \left(\sum_{i=1}^{4} Z_{ij} \right)^2 = \_\_\_\_\_\_\_\_$$

$$\left(\sum_{i=1}^{4} \sum_{j=1}^{5} Z_{ij} \right)^2 = \_\_\_\_\_\_\_\_$$

5. For the set $\{x_i\}$

$\bar{x}$ = ____________________

SS = ____________________

df = ____________________

s^2 = ____________________

$s_{\bar{x}}$ = ____________________

6. For the set $\{y_j\}$

$\sum_{j=1}^{7} y_j$ = ________________

$\bar{y}$ = ________________

C.T. = ______________

$s_{\bar{y}}$ = ______________

EXPERIMENTAL DESIGN

We have taken a brief look at the basic tools of statistics. We are now ready to see how these tools can be used in the design and analysis of agricultural experiments. To do this we can take a look at an example : suppose we have a wheat breeder who wants to compare the yield of a new variety with that of an old variety. He has two basic objectives in making the comparison. The first is to answer the question : is there difference in yield between the two varieties ? The second objective, which is related to the first, is to estimate the size of the difference.

As might be expected, the answer to the first question involves a test of hypothesis. We set up the hypothesis that there is no difference in yield and hope that the data lead us to reject this hypothesis. We meet the second objective by computing an estimate of the difference, preferably an Interval Estimate. Almost all agricultural experiments are conducted for one or both purposes : testing of hypothesis and estimation of differences in the effects of different treatments, and to obtain information on why the treatments behave as they do. The role of experimental design in this process is to provide efficient and precise information to meet these experimental objectives.

Suppose the wheat breeder decided to conduct an experiment to compare the new variety with the old. He could plant one variety on one plot and the other variety on another plot. At harvest time he could then observe the difference in yields between the two plots. There is one obvious drawback to this procedure the breeder would have no way to determine how much of his observed difference was due to a true difference between varieties and how much was due to the natural variation found in all biological material. This random variation among plots, or experimental units, treated alike is called Experimental error . This does not mean that mistakes have been made in conducting the experiment. Rather it is due to Biological variation, Soil variation, Variation in technique, etc. The breeder must have a measure of experimental error if he wants to test the difference between varieties or to compute an interval estimate of the difference. To measure experimental error the breeder must repeat or replicate, each variety more than onein his experiment.

Replication serves a number of purposes in an experimental design :

1. It provides an estimate of experimental error because it provides several observations on experimental units receiving the same treatment.

2. It increases precision by reducing standard errors. Recall that a confidence interval estimate of a mean is given by

$$L(\mu) = \bar{x} \pm t\alpha\sqrt{s^2/n},$$

where n is the number of observations, or replications. As n increases the width of the confidence interval decreases. Hence the estimate of μ becomes more precise.

3. It can broaden the base for making infrenece. As replication is increased a wider variety of units can be brought into the experiment, and the results will apply over a wider variety of conditions.

Replication is not the only factor which must be considered in designing an experiment. Suppose the breeder had arranged his experiment in the field in the following way :
(N = new variety, O= old variety)

N	O	N	O	N	O	N	O	N	O

High ——→——→——→——→——→——→Low
Fertility

Now, suppose that there was a fertility gradient in the field ranging from high at one end to low at the other. In each pair of plots the new variety has been placed on the higher fertility level. A difference between the new and the old may be partly a variety difference and partly a fertility difference. Placement of the plots is said to be biased in favor of the new variety. We would like to eliminate this bias by arranging the treatments so that no treatment is consistently favored by being placed under the best conditions in the experiment. To do this we use a process called Randomization.

By randomization we mean that treatments are assigned to the experimental units in such a way that any unit is equally likely to receive any treatment. In some experimental designs randomization is restricted in certain ways, but in no design is it completely eliminated. We will discuss the randomization procedure as we discuss the individual designs.

There are a couple of purposes for randomization in an experimental design :

1. To eliminate bias. Randomization insures that no treatment is favored or discriminated against by its assignment to plots in the design.
2. To assure independence among the observations. This is necessary to provide valid significance tests and interval estimates.

We need to consider one additional feature of a good experimental design. This is the feature called Local control or Blocking. Under this procedure we arrange the experimental material into groups, or blocks of more or less uniform experimental units. The treatments are then assigned at random to the units, or plots, within the blocks. As an example, suppose we wanted to conduct an experiment in a field in which there is two types of soil. We could have one block of plots on one soil type and another block of plots on the other. Treatment comparisons would then be made on the same soil types, and differences between soil types would not be a factor in the comparisons. There are several reasons for blocking :

1. It can increase the precision of an experiment. Differences among blocks are removed from experimental error in the analysis of the results.
2. Treatments are compared under more nearly equal conditions because comparisons are made within blocks of uniform plots.
3. It can sometimes increase the information from an experiment. Blocks need not be placed at the same location. By placing blocks at different locations a wider variety of conditions can be sampled by an experiment.

In summary, then, a good experimental design has elements of the following characteristics :

1. Replication
2. Randomization
3. Local control, or Blocking.

Field Layout

In constructing the field plan for the experiment there are a number of questions which need to be considered. Among these are :

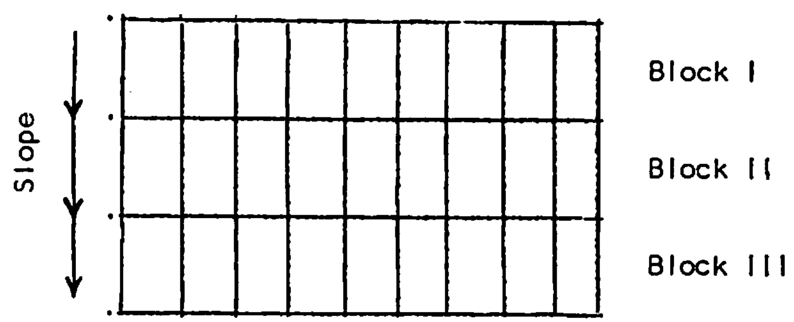
1. Are there any gradients in slope, fertility, drainage, etc. in the field being considered ?

2. Is there anything, such as trees, buildings, wind breaks, etc., which might cause different results in one part of the field than in another ?

3. How are the treatments to be applied, and how are yields to be determined ?

4. Is it necessary to be concerned about border effects ?

If there are gradients in the field a general rule is that the plots should be grouped into blocks. The blocks should be rectangular with only one or two rows of plots, and should be placed perpendicular to the gradient. The plots should be long and narrow placed so that the long direction is parallel to the gradient. To illustrate this, suppose an experiment is to be conducted on a sloping field. The field plan would appear as



On the other hand, if the field appears to be fairly uniform, or if there is no pattern to the variability there would be no basis for blocking, and they probably should not be used. Under these conditions square plots are usually more efficient than rectangular plots. In general, the more variable is the field the larger should be the plot.

If possible the experiment should be located at a distance from buildings, wind breaks, and other things which might affect the results. If this cannot be done plots which might be affected should be grouped into one block and the other plots into other blocks.

Consideration should be given, in the field arrangement to the equipment to be used to apply the treatments and to measure yield. Space must be left at the ends of the plots so that the equipment can be operated without affecting other plots. If space is too limited to permit this an experimental design should be chosen which will permit the equipment to be operated over several plots before turns or changes are required.

Plants at the edges of plots behave differently than do plants in the center of plots. This is because plot edges recieve different amounts of sunlight, different amounts of moisture, different competition from plants in adjoining plots, etc., than do the centers of the plots. Such differences are called edge, or border, effects. Border effects must be accounted for when measuring plot yield. It is usually assumed that yields at the center of the plot are more typical of what happens in practice than are yields at the border. Hence yields are usually measured only on the center of the plot. This may be accomplished by removing the borders before yield is measured, or, for row-seeded crops, measuring yield only on the center rows of the plots. In any case plots must be big enough to permit measurement of yields free of border effects.

Successful Experimentation

There are number of steps which should be taken in planning and conducting a successful agricultural experiment :

1. Define the problem - state the problem clearly and concisely. If the problem cannot be clearly defined there is little chance that it can be solved.

2. State the objectives- this may be in the form of questions to be answered, hypotheses to be tested, or effects to be estimated. When there is more than one objective, list them in order of importance.

3. Critically analyze the problem and the objectives - what is the present status of knowledge concerning the problem ? Will the experiments add to this knowledge? How do the objectives bear on the solution of the problem.

*4. Select the treatments - treatments should be used whose evaluation will answer the objectives of the experiment.

*5. Select the experimental material - the material should be representative of the population on which you wish to test your treatments and make your inferences.

*6. Select an experimental design - as a general rule chose the simplest design which will give you the precision you require.

7. Select the experimental unit. For field experiments with plants this means deciding on the size and shape of the plots.

8. Control the effects of adjacent plots on each other - this is usually done by plot borders and by randomization.

9. Decide on the data to be collected - the data should properly evaluate the treatments in line with the objectives. Additional data should be taken to explain why the treatments perform as they do.

*10. Outline the statistical analysis and summary of results - write out the analysis of variance table including sources of variation and degrees of freedom, along with planned F tests. Outline the tables to be used to summarize the results.

At this point the plan should be reviewed by your colleagues and by a biometrician. They may have ideas on points you might have overlooked, and on ways in which the experiment might be improved.

11. Conduct the experiment - follow your experimental plan. If errors occur either correct them or make note of them so they can be taken into consideration in the analysis. Avoid fatigue in collecting the data. Immediately check observations which seem out of line. If necessary to copy data, check the copied figures against the originals.

*12. Analyze the data and interpret the results. Follow the analysis you have outlined, and conduct the planned significance tests. Interpret your results in light of the experimental conditions and previously established facts. Don't accept a statistically significant result if it appears to be out of line, but investigate the matter further.

*13. Prepare a report of the research - "The job is not finished until the paper work is done". There is no such thing as a negative result. Lack of significance may indicate that there is no real difference among the treatments used.

*Points at which the biometrician may be of particular assistance.

(The above steps were adapted from "Statistical Methods for Agricultural Research" by T.M. Little + F.J. Hill)

COMPLETELY RANDOMIZED DESIGN

Characteristics : The completely randomized design for p treatment has rp plots. Each of the p treatment is assigned at random to r of the plots.

Example : A wheat breeder wanted to compare the yields of six new varieties : 1, 2, 3, 4, 5, and 6. He ran an experiment using a completely randomized design replicated four times. (For this trial p= 6, r= 4). At harvest time he measured the yield on each plot. The field plan and yields (T/ha) were as follows : (variety numbers are circled)

① 1.51	⑤ 1.16	① 1.49	③ 1.43	② .60	⑤ 1.22
② .90	③ 1.46	④ 1.30	④ 1.20	① 1.54	④ 1.33
⑥ .98	② .74	⑤ 1.16	⑥ .90	① 1.55	③ 1.26
④ 1.26	③ 1.28	⑥ .76	⑤ 1.12	② .66	⑥ .82

Calculations for statistical analysis :

1. Make a table of yields, treatment totals, and treatment means :

In general :

Treatment	Yield				Total	Mean
1	y_{11}	y_{12}	...	y_{1r}	T_1	$\bar{y}_1$
2	y_{21}	y_{22}	...	y_{2r}	T_2	$\bar{y}_2$
.	.	.	...	.	.	.
.	.	.	...	.	.	.
.	.	.	...	.	.	.
p	y_{p1}	y_{p2}	...	y_{pr}	T_p	$\bar{y}_p$
				Sum	G	$\bar{\bar{y}}$

Where : y_{ij} = jth yield on the i th treatment

$Ti = \sum_{j=i}^{r} y_{ij}$ = Sum of yields on the i th treatment

$\bar{y}i = T_i/_r$ = Mean yield on the i th treatment

$G = \sum_{i=1}^{p} T_i$ = Grand total of all yields

$\bar{\bar{y}} = G/_{rp}$ = Grand mean of all yields.

In our numerical example we have :

Variety	Yield (T/Ha)				Total	Mean
1	1.51	1.49	1.54	1.55	6.09	1.52
2	.60	.90	.74	.66	2.90	.72
3	1.43	1.46	1.26	1.28	5.43	1.36
4	1.30	1.20	1.33	1.26	5.09	1.27
5	1.16	1.22	1.16	1.12	4.66	1.16
6	.98	.90	.76	.82	3.46	.86
				Sum	27.63	1.15

Sample computations :

$$T_1 = \sum_{j=1}^{4} y_{1j} = 1.51 + 1.49 + 1.54 + 1.55 = 6.09$$

$$\bar{y}_1 = T_1/4 = 6.09/\,4 = 1.52$$

$$G = \sum_{i=1}^{6} T_i = 6.09+2.90+5.43+5.09+4.66+3.46 = 27.63$$

$$\bar{\bar{y}} = G/\,(4)(6) = 27.63\,/24 = 1.15$$

II. Fill in a table of preliminary ANOVA computations :
In general :

(1) Source of variation	(2) Number of Totals squared	(3) Observations per total	(4) Sum of $(\text{Total})^2$	(5) Raw ss (4)/(3)
Correction	1	rp	G^2	$CT=G^2/rp$
Total	rp	1	$\sum_i \sum_j y_{ij}^2$	$\sum_i \sum_j y_{ij}^2$
Treatment	p	r	$\sum_{i=j}^{p} T_i^2$	$\sum_{i=j}^{p} T_i^2 / r$

For the Numerical example we have

(1) Source of Variation	(2) Number of Totals squared	(3) Observations per total	(4) Sum of $(\text{Total})^2$	(5) Raw ss (4)/(3)
Correction	1	24	763.4169	CT=31.8090
Total	24	1	33.7709	33.7709
Treatment	6	4	134.5783	33.6446

Sample Computations :

$G^2 = (27,63)^2 = 763.4169$

$CT = G^2/rp = 763.4169/24 = 31.8090$

$\sum_i \sum_j y_{ij}^2 = 1.51^2 + 1.49^2 + \ldots + .26^2 + .82^2 = 33.7709$

$\sum_{i=1}^{6} T_i^2 = 6.09^2 + 2.90^2 + \ldots + 3.46^2 = 134.5783$

$\sum_i T_i^2)/r = 134.5783/4 = 33.6446$

III. Complete the analysis of variance, ANOVA , table.

ANOVA

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Total	rp-1	SSTOT		
Treatment	p-1	SST	MST	FT
Error	p(r-1)	SSE	MSE	

% CV =

Standard Error =

Calculations :

1. Except for the "Error" line, entries in the first three columns of the ANOVA table are taken from the preliminary table :

   a- Source of variation = (1) with "correction" omitted.

   b- Degrees of freedom = (2) minus 1

   c- Sum of squares = (5) minus "CT"

2. For "Error" line

   a- Degrees of Freedom = Total d.f. - treatment d.f.

   = p(r-1)

   b- SSE = SSTOT - SST

3. Mean squares (not computed for total)

   a- Divide sum of squares by degrees of freedom on same line.

4. F_t = MST / MSE

5. % CV = $(\sqrt{MSE} / \bar{\bar{y}})\ 100$

6. Standard error = $\sqrt{MSE/r}$

For the numerical example we have :

ANOVA

Source	d.f.	S.S.	M.S.	F
Total	23	1.9619		
Variety	5	1.8356	.3671	52.44
Error	18	.1263	.0070	

% CV = 7.3 %

Standard Error = .0418

Sample Computations :

a- "Sources" from ⓵

b- Degrees of freedom = ⓶ - 1
   Total, 24 - 1 = 23
   Treatment, 6 - 1 = 5
   Error, 23 - 5 = 18 = 6(4-1)

c- Sum of squares = ⓹ - CT
   SSTOT = 33.7709 - 31.8090 = 1.9619
   SST = 33.6446 - 31.8090 = 1.8356
   SSE = SSTOT - SST = 1.9619 - 1.8356 = .1263

d- Mean square = SS/d.f.
MST = 1.8356 / 5 = .3671
MSE = .1263 / 18 = .0070

e- F_T = MST / MSE = .3671 / .0070 = 52.44

f- % CV = $(\sqrt{MSE} / \bar{\bar{y}})100 = (\sqrt{.0070}/1.15)100$= 7.3 %

g- Standard Error $= \sqrt{MSE / r} = \sqrt{.0070/4}$ = .0418

Significance Test : F_T in the analysis of variance table is used to test the significance of differences among treatment means. We look in the F table for the 5 % and 1% F values using the column for p-1 degrees of freedom and the row for p(r-1) degrees of freedom.
If F_T is larger than the 1% F the defferences are " Highly Significant ". If F_T is smaller than the 1% F but larger than the 5% F the differences are "Significant" . If F_T is smaller than the 5% F the differences are not significant.

For the numerical example look in the column for p-i = 6-1 = 5 d.f. and the row for p(r-1) = 6(4-1) = 18 d.f. The 5% F is 2.77 and the 1%F is 4.25. Since 52.44 is greater than 4.25 the differences among the mean yields of the six wheat varieties are highly significant.

Presentation of Results : The results of the statistical analysis may be presented in a table of means, with their standard error and a statement of the significance of the differences.

For the numberical example the results are summarized as follows:

Mean yield (T/ha) of six new wheat varieties

Variety	1	2	3	4	5	6	Standard error
Mean yield**	1.52	.72	1.36	1.27	1.16	.86	.042

** Differences are significant at the 1 % level

Exercise: An agronomist wanted to determine the effect of five weed control chemicals; A, B, C, D, and E, on the germination of barley. He obtained 20 large pots of soil and planted 100 barley seeds in each pot. He then applied each chemical to four pots, selected at random from the 20. He placed the pots on a bench in the Green House, and after 14 days he counted the number of seeds which did not germinate. The layout and the number which did not germinate are as follows

A	D	D	E	C
10	5	9	7	10
B	C	B	B	C
6	9	7	11	8
A	D	E	E	A
12	10	5	5	13
E	C	B	D	A
3	10	5	6	11

Analyze this data from a completely randomized design.

ANOVA work sheet for completely randomised design

Preliminary Computations

(1) Source of Variation	(2) Number of Totals squared	(3) Observations Per total	(4) Sum of $(\text{total})^2$	(5) Raw SS (4)/(3)
Correction	1			CT=
Total Treatment		1		

ANOVA

Source	df	SS	MS	F
Total				
Treatment				
Error				

% cv =

Standard Error =

RANDOMIZED BLOCK DESIGN

<u>CHARACTERISTICS</u> : The Randomized Block Design for p treatments has rp plots arranged into r blocks (groups of plots) with p plots in each block. Each of the p treatments is assigned at random to one plot in each block.

For example: an agronomist wanted to determine the effect of spacing between rows on the number of tillers per m^2 produced by barley. He chose four row spacings 15 cm, 20cm, 25cm and 30cm as his treatments. The field in which the experiment was conducted varied in soil fertility from one side to the other. To remove the effect of this difference he used a Randomized Block Design with five blocks. At maturity he counted the number of tillers per m^2 in each plot. The field lay-out and the tiller count are shown in the following plan:

← Fertility →

BLOCK	I	II	III	IV	V
	(1) 205	(2) 178	(4) 180	(1) 210	(2) 185
	(4) 172	(1) 197	(1) 215	(3) 182	(4) 183
	(3) 164	(3) 177	(2) 192	(2) 200	(3) 190
	(2) 170	(4) 161	(3) 172	(4) 164	(1) 223

Calculations for statistical Analysis :

1. Make a table of yields arranged by block and treatment. Compute block totals, and treatment totals and means :

	Block					
Treatment	1	2	...	r	Total	Mean
1	y_{11}	y_{12}		y_{1r}	T_1	$\bar{y}_1$
2	y_{21}	y_{22}	...	y_{2r}	T_2	$\bar{y}_2$
.	.	.		.	.	.
.	.	.		.	.	.
.	.	.		.	.	.
p	y_{p1}	y_{p2}	...	y_{pr}	T_p	$\bar{y}_p$
Total	B_1	B_2	...	B_r	G	$\bar{\bar{y}}$

Where :

y_{ij} = Yield of the ith treatment in the jth block.

$T_i = \sum_{j=1}^{r} y_{ij}$ = Sum of yields on the ith treatment.

$\bar{y}_i = Ti/r$ = Mean yield on the ith treatment.

$B_j = \sum_{i=1}^{p} y_{ij}$ = Sum of yields on the jth block.

$G = \sum_{i=1}^{p} T_i = \sum_{j=1}^{n} B_j$ = Grand Total of all yields.

(N.B: A good check of calculations is to see that both treatment totals and block totals sum to the grand total).

$\bar{\bar{y}}$ = G/rp = Grand mean of all yields.

Example :

For our numberical data we have

	BLOCK						
Treatment	1	2	3	4	5	Total	Mean
1	205	197	215	210	223	1050	210.0
2	170	178	192	200	185	925	185.0
3	164	177	172	182	190	885	177.0
4	172	161	180	164	183	860	172.0
Total	711	713	759	756	781	3720	186.0

Sample Calculations :

$$T_1 = \sum_{j=1}^{5} y_{1j} = 205+197+215+210+223 = 1050$$

$$\bar{y}_1 = T_1/r = 1050 / 5 = 210.0$$

$$B_1 = \sum_{i=1}^{4} y_{i1} = 205+170+164+172 = 711$$

$$G = \sum_{i=1}^{4} T_i = 1050+925+885+860 = 3720$$

$$= \sum_{j=i}^{5} B_j = 711+713+759+756+781 = 3720$$

$$\bar{\bar{y}} = G/ rp = 3720 /(5)(4) = 186.0$$

2. Fill in a table of Preliminary ANOVA computations :

① Source of Variation	② Number of Totals squared	③ Oservations per Total	④ Sum of $(\text{Total})^2$	⑤ Raw SS ④/③
Correction	1	rp	G^2	$CT=G^2/rp$
Total	rp	1	$\sum_i \sum_j yij^2$	$\sum_i \sum_j yij^2$
Block	r	p	$\sum_j Bj^2$	$(\sum_j Bj^2)/p$
Treatment	p	r	$\sum_i Ti^2$	$(\sum_i Ti^2)/r$

Example :

For the numerical data the preliminary ANOVA computations are :

Source of Variation	Number of Totals squared	Observations Per Total	Sum of $(\text{Total})^2$	Raw SS ④/③
Correction	1	20	13,838,400	CT=691,920
Total	20	1	697,884	697,884
Block	5	4	2,771,468	692,867
Treatment	4	5	3,480,950	696,190

Sample Computations :

$r = 5,\ p=4,\ rp = 20$

$G^2 = (3720)^2 = 13,838,400$

$CT = G^2/rp = 13,838,400/20 = 691,920$

$\sum_i \sum_j y_{ij}^2 = 205^2+197^2+\ldots+164^2+183^2 = 697,884$

$$\sum_j B_j^2 = 711^2+713^2+759^2+756^2+781^2 = 2,711,468$$

$$(\sum_j B_j^2)/p = 2.771,468/4 = 692,867$$

$$\sum_i T_i^2 = 1050^2+925^2+885^2+860^2 = 3,480,950$$

$$(\sum_i T_i^2)/r = 3.480,950/5 = 696,190$$

3. Complete the analysis of variance, ANOVA, table :

ANOVA

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Total	rp-1	SSTOT		
Block	r-1	SSB	MSB	F_B
Treatment	p-1	SST	MST	F_t
Error	(r-1)(p-1)	SSE	MSE	

% CV =

Standard Error =

Calculations :

- Except for the "Error" line, entries in the first three columns of the ANOVA table are taken from the table of Preliminary Computations :

  a- Sources of variation = (1) with "Correction" omitted.

  b- Degrees of freedom = (2) minus 1

  c- Sum of squares = (5) minus "CT"

- For the "Error" line

  a- Degrees of freedom = Total d.f.-Block d.f.-Treatment d.f. = (r-1)(p-1)

  b- SSE = SSTOT - SSB - SST

- Mean squares (not computed for total)

  a- divide the sum of squares by the degrees of freedom on the same line.

- F column

  a- F_B = MSB/ MSE

  b- F_T = MST/ MSE

- % CV = $\left(\sqrt{MSE}/\bar{\bar{y}}\right)100$

- Standard error = $\sqrt{MSE/\ r}$

  Example : For the numerical data given above we have :

ANOVA

Source	d.f.	SS	MS	F
Total	19	5,964		
Block	4	947	236.75	3.80
Treatment	3	4,270	1,423.33	22.86
Error	12	747	62.25	

% CV = 4.2%

Standard Error = 3.53

Sample Computations :

- Sources of Variation = (1)

- Degrees of Freedom = (2) - 1
  Total, rp - 1 = 20 - 1 = 19
  Block, r - 1 = 5 - 1 = 4
  Treatment, p - 1 = 4 - 1 = 3
  Error, (r-1)(p-1) = (5-1)(4-1) = 12 = 19-4-3

- Sums of Squares, (5) - CT

  Total = 697,884 - 691,920 = 5,964 = SSTOT
  Block = 692,867 - 691,920 = 947 = SSB
  Treatment = 696,190 - 691,920 = 4,270 = SST
  Error = 5,964-947 - 4,270 = 747 = SSE

- Mean squares, SS/df
  MSB = 947/4 = 236.75
  MST = 4,270/ 3 = 1,423.33
  MSE = 747 / 12 = 62.25

- F, MS/MSE
  F_B = 236.75/62.25 = 3.80
  F_T = 1,423.33/62.25 = 22.86

- $\% \ CV = (\sqrt{MSE}/\bar{\bar{y}})\ 100$

  $= (\sqrt{62.25}/186.0)\ 100 = 4.2\%$

- Standard Error $= \sqrt{MSE/r}$

  $= \sqrt{62.25/5} = 3.53$

SIGNIFICANCE TESTS

F_T may be used to test the significance of the differences among the treatment means. We look in the 5% and 1% F table to find the F value in the column headed p-1 degrees of freedom and the row for (r-1)(p-1) degrees of freedom. If F_t is greater than the 1% F the differences are termed highly significant; if F_T is greater than the 5% F but smaller than the 1% F the differences are termed "Significant", If F_T is smaller than the 5% F the differences are not considered significant.

F_B gives us information about whether blocking was effective in increasing the accuracy of the experiment. If F_B is greater than the 5% F for r - 1 AND (r - 1)(p - 1) degrees of freedom then blocking increased the accuracy.

For the present example at p-1 = 3 and (r-1)(p-1) = 12 d.f. the 1% F is 5.95 and the 5% F is 3.49. Since 22.86 is greater than 5.95 the differences among the treatment means are highly significant. For blocks the 5% F at r-1= 4 and (r-1)(p-1) =12 d.f. is 3.26. Since 3.80 is somewhat larger than 3.26 there is an indication that the accuracy of the experiment was increased by blocking.

Presentation of Results

The results of the statistical analysis can be summarized in a table of treatment means together with their standard error and a statement about the significance of the differences.

The numerical example might be summarized in the following way:

Mean number of tillers per M^2 in barley sown at different row spacings.

Spacing	15 cm	20 cm	25 cm	30 cm	Standard Error
Mean**	210.0	185.0	177.0	172.0	3.53

** Significantly different at the 1% level.

Exercise

An experiment was conducted to measure the differences in yield among six varieties of wheat. Because the experimental site was sloping a randomized block design was used to eliminate the effect of slope. The field plan and yields are as follows :

						Block
① 2208	③ 2750	④ 2844	② 1615	⑤ 2250	⑥ 3034	I
③ 2844	① 2156	④ 2396	⑤ 2823	⑥ 2354	② 2854	II
⑥ 2033	③ 1908	⑤ 2000	④ 1575	① 2050	② 2117	III
⑤ 2650	① 1958	③ 2358	② 2458	⑥ 1967	④ 2475	IV

Conduct a statistical analysis of the results of this experiment.

ANOVA work sheet for Randomized Block Design :-

1. Preliminary Computations

①	②	③	④	⑤
Source of Variation	Number of Totals squared	Observations Per Total	Sum of (Total)^2	Raw SS ④/③
Correction	1			CT=
Total Block Treatment		1		

2. Analysis of Variance

Source	d.f.	SS	MS	F
Total				
Block				
Treatment				
Error				

FACTORIAL EXPERIMENTS - TWO FACTORS

CHARACTERISTICS

In a Factorial Experiment the treatments consists of combinations of two or more levels (quantities) of the two or more of the factors of production. For a complete factorial set of treatments each level of each factor occurs together with each level of every other factor. The total number of treatments is the product of the number of levels of all of the factors. The treatments are usually applied to plots in a randomized block design, although any other design might be used.

For example suppose an agronomist wanted to see how Nitrogen and Potassium fertilizer, both alone and in combination with each other, would affect the yield of barley. Suppose he wanted to use three levels of Nitrogen: None (N_o), 25 kg/ha (N_1), and 50 kg/ha (N_2) and two levels of potassium: None (Ko) and 25 kg/ha (K_1). He could use a factorial experiment with two factors:

Factor A= Nitrogen at 3 levels : N_o, N_1, N_2

Factor B= Potassium at 2 levels : K_o, K_1.

The factorial set would contain 3x2 = 6 treatments:

1. No Ko	3. N_1 Ko	5. N_2 Ko
2. No K_1	4. N_1 K_1	6. N_2 K_1

Suppose he conducted the experiment in a Randomized block design with three blocks. The field plan and yield might be as follows :-

BLOCK I		II		III	
N_1K_1 2.29	N_2Ko 1.75	NoK_1 1.26	N_1K_1 1.60	NoKo 1.22	N_2Ko 2.13
N_2K_1 2.17	NoK_1 1.52	N_2Ko 1.95	N_1Ko 1.61	N_2K_1 1.82	N_1K_1 1.80
NoKo 1.52	N_1Ko 1.55	NoKo 1.47	N_2K_1 1.88	NoK_1 1.67	N_1Ko 1.62

Calculations for Statistical Analysis : Two Factors

Factor A at a levels, factor B at b levels. Randomized block design with r blocks. (In this example factor A is Nitrogen and factor B is Potassium. a=3, b=2, r=3) Let

y = yield on the ith level of factor A, jth level of factor B in the kth block.

Tables of Totals:

1. Make a table of totals arranged by factor A and factor B:

Factor A	FACTOR B 1	2	...	b	Sum
1	$T_{11.}$	T_{12}	...	T_{1b}	A_1
2	$T_{21.}$	$T_{22.}$	...	T_{2b}	A_2
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
a	T_{a1}	T_{a2}	...	T_{ab}	A_a
Sum	B_1	B_2	...	B_b	G

Where

$T_{ij} = \sum_{k=1}^{r} y_{ijk}$ = Sum of yields on i th level of A, jth level of B, summed over blocks.

$A_i = \sum_{j=1}^{b} T_{ij}$ = Sum of all yields on ith level of A.

$B_j = \sum_{i=1}^{a} T_{ij.}$ = Sum of all yields on jth level of B.

$G = \sum_{i=1}^{a} A_i = \sum_{j=1}^{b} B_j$ = Grand total of all yields.

II. Make a table of block totals:

Block	1	2	. . .	r	Sum
Sum	R_1	R_2	. . .	R_r	G

In which

$$R_k = \sum_{i=1}^{a} \sum_{j=1}^{b} y_{ijk} = \text{Sum of all yields in the kth block.}$$

$$G = \sum_{k=1}^{r} R_k = \text{Grand total of all yields.}$$

In our example we have :

I. Nitrogen X Potassium totals

Nitrogen	Potassium 0	Potassium 1	Sum
0	4.21	4.45	8.66
1	4.78	5.69	10.47
2	5.83	5.87	11.70
Sum	14.82	16.01	30.83

Sample calculations:

$$T_{11} = \sum_{k=1}^{3} y_{11k} = 1.52 + 1.47 + 1.22 = 4.21$$

$$A_1 = \sum_{j=1}^{2} T_{1j} = 4.21 + 4.45 = 8.66$$

$$B_1 = \sum_{i=1}^{3} T_{i1} = 4.21 + 4.78 + 5.83 = 14.82$$

$$G = \sum_{i=1}^{3} A_i = \sum_{j=1}^{2} B_j = (8.66+10.47+11.70)$$

$$= (14.82+16.01) = 30.83$$

II. Block totals

Block	I	II	III	Sum
Sum	10.80	9.77	10.26	30.83

Sample Calculations :

$$R_1 = \sum_{i=1}^{3} \sum_{j=i}^{2} y_{ij1} = 1.52+1.52+1.55+2.29+1.75+2.17= 10.80$$

$$G = \sum_{k=1}^{3} R_k = 10.80+9.77+10.26 = 30.83$$

III. Fill a table of preliminary ANOVA computations :

① Source of Variation	② Number of Totals Squared	③ Observations Per Total	④ Sum of $(\text{Total})^2$	⑤ Raw SS ④ / ③
Correction	1	rab	G^2	$CT = G^2/rab$
Total	rab	1	$\sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{r} y_{ijk}^2$	$\sum_{i} \sum_{j} \sum_{k} y_{ijk}^2$
Block	r	ab	$\sum_{k=1}^{r} R_k^2$	$(1/ab) \sum_{k} R_k^2$
A	a	rb	$\sum_{b=1}^{a} A_i^2$	$(1/rb) \sum_{i} A_i^2$
B	b	ra	$\sum_{j=1}^{b} B_j^2$	$(1/ra) \sum_{j} B_j^2$
AB	ab	r	$\sum_{i=1}^{a} \sum_{j=1}^{b} T_{ij.}^2$	$(1/r) \sum_{i} \sum_{j} T_{ij.}^2$

For the example we find

① Source of Variation	② Number of Totals Squared	③ Observations Per Total	④ Sum of $(\text{Total})^2$	⑤ Raw SS ④/③
Correction	1	18	950.4889	52.8049
Total	18	1	54.2673	54.2673
Block	3	6	317.3605	52.8934
N	3	6	321.5065	53.5844
K	2	9	475.9525	52.8836
NK	6	3	161.1969	53.7323

Sample Calculations : (4)

Correction, $G^2 = (30.83)^2 = 950.4889$

Total, $\sum_i \sum_j \sum_k y_{ijk}^2 = 2.29^2 + 1.75^2 + \ldots + 1.62^2 = 54.2673$

Block, $\sum_k R_k^2 = 10.80^2 + 9.77^2 + 10.26^2 = 317.3605$

N, $\sum_i A_i^2 = 8.66^2 + 10.47^2 + 11.70^2 = 321.5065$

K, $\sum_j B_j^2 = 14.82^2 + 16.01^2 = 475.9525$

NK, $\sum_i \sum_j T_{ij}^2 = 4.21^2 + 4.45^2 + \ldots + 5.87^2 = 161.1969$

IV. Complete the Analysis of variance, ANOVA, table :

ANOVA

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Total	rab - 1	SSTOT		
Block	r - 1	SSR	MSR	F_R
A	a - 1	SSA	MSA	F_A
B	b - 1	SSB	MSB	F_B
AB	(a-1)(b-1)	SSAB	MSAB	F_{AB}
Error	(r-1)(ab-1)	SSE	MSE	

% CV =

Calculations: (First three columns from preliminary table)

1. Except for the "Error" line,
   Sources of Variation = ① with "Correction" omitted

2. Except for "Error" and "AB"
   Degrees of freedom = ② minus 1.
   a. Degrees of Freedom for "AB" = (a-1)(b-1)
   b. Degrees of Freedom for "Error" = (r-1)(ab-1)
      = Total d.f.- Block d.f.- A d.f.- B d.f.- AB d.f.

3. Sums of squares: Except for "AB" and "Error"
   Sum of squares = ⑤ - "CT"
   a. SS AB = ⑤ -CT - SSA - SSB
   b. SSE = SSTOT - SSR -SSA - SSB - SSAB

4. Mean squares: (Not computed for Total)
   a. Divide the sum of squares by the degrees of Freedom on the same line in the ANOVA table

5. F column: Divide the other mean squares by MSE
   a. F_R = MSR/MSE
   b. F_A = MSA/MSE
   c. F_B = MSB/MSE
   d. F_{AB} = MSAB/MSE

6. % CV = $(\sqrt{MSE} \,/\, \bar{\bar{y}})\ 100$, where $\bar{\bar{y}}$ = G/rab

For the numerical example we have :

ANOVA

Source	d.f.	SS	MS	F
Total	17	1.4624		
Block	2	0.0885	0.0443	0.99
N	2	0.7795	0.3898	8.74
K	1	0.0787	0.0787	1.76
NK	2	0.0692	0.0346	0.77
Error	10	0.4465	0.0446	

% CV = 12.3%

Sample Calculations:

1. Source of Variation - From (1), Preliminary table.

2. Degrees of Freedom - (2) minus 1
   AB d.f. = (a-1)(b-1) = (3-1)(2-1) = 2
   Error d.f.= (r-1)(ab-1) = (3-1)(6-1) = 10
   = 17 - 2 - 2 - 1 - 2 = 10

3. Sum of Squares - (5) minus CT
   SSTOT = 54.2673 - 52.8049 = 1.4624
   SSR = 52.8934 - 52.8049 = 0.0885
   SSA = 53.5844 - 52.8049 = 0.7795
   SSB = 52.8836 - 52.8049 = 0.0787

   SSAB = (5) - CT - SSA - SSB
   = 53.7323 - 52.8049 - 0.7795 - 0.0787 = 0.0692

   SSE = SSTOT - SSR - SSA - SSB - SSAB
   = 1.4624 - 0.0885 - 0.7795 - 0.0787 - 0.0692=0.4465

4. Mean squares - SS/ d.f.

   MSR = 0.0885 / 2 = 0.0443
   MSA = 0.7795 / 2 = 0.3898
   MSB = 0.0787 / 1 = 0.0787
   MSAB = 0.0692 / 2 = 0.0346
   MSE = 0.4465 /10 = 0.0446

5. $F = MS / MSE$

$$F_R = 0.0443 / 0.0446 = 0.99$$
$$F_A = 0.3898 / 0.0446 = 8.74$$
$$F_B = 0.0787 / 0.0446 = 1.76$$
$$F_{AB} = 0.0346 / 0.0446 = 0.77$$

6. $\% CV = (\sqrt{MSE} / \bar{\bar{y}}) 100, \quad \bar{\bar{y}} = G/rab = 30.83/18 = 1.7128$

$$\% CV = (\sqrt{.0446} / 1.7128) 100 = 12.3 \%$$

Significance Tests

The value in the F column may be used to test the significance of the factors used in the experiment. We look in the 5% and 1% F tables to find the F value in the column headed by the degrees of freedom for the factor being tested and the row for the error degrees of freedom. If the F in the ANOVA table is larger than the 1% F the factor is "highly significant". If it is larger than the 5% F but smaller tna the 1% F the factor is "significant". If it is smaller than the 5% F it is " not significant".

F_A is used to test the factor A means, F_B is used to test the factor B means, and F_{AB} is used to test the differences among the individual treatment combinations. F_R provides an approximate test of the effectiveness of blocking in reducing the experimental error.

In the present example at 2 and 10 d.f. the 1% F is 7.56 and the 5% F is 4.10. Since 8.74 is larger than 7.56 the differences between Nitrogen means are highly significant. And since 0.77 is smaller than 4.10 the AB effect is not significant. Also, 0.99 is smaller than 4.10 which means that blocking was not effective in reducing error. At 1 and 10 d.f. the 1% F is 10.04 and the 5% F is 4.96. Since 1.76 is smaller than 4.96 the difference between Potassium means is not significant.

Standard Errors: Two kinds of standard errors are computed: 1. Standard errors of the means, and 2. standard errors of the differences between means.

1. Standard Errors of means, $s_{\bar{y}}$
   a. Factor A means, $s_{\bar{y}A}$
      $s_{\bar{y}A} = \sqrt{MSE / rb}$
   b. Factor B means, $s_{\bar{y}B}$
      $s_{\bar{y}B} = \sqrt{MSE / ra}$
   c. Individual AxB means, $s_{\bar{y}AB}$
      $s_{\bar{y}AB} = \sqrt{MSE / r}$

2. Standard errors of differences between means, $s_{\bar{d}}$
   a. Difference between two factor A means, $s_{\bar{d}A}$
      $s_{\bar{d}A} = \sqrt{2\ MSE / rb}$
   b. Difference between two factor B means, $s_{\bar{d}B}$
      $s_{\bar{d}B} = \sqrt{2\ MSE / ra}$
   c. Difference between AxB means, $s_{\bar{d}AB}$
      $s_{\bar{d}AB} = \sqrt{2MSE / r}$

For the numerical example the standard errors are:
1. Means
   a. Nitrogen means, $s_{\bar{y}A} = \sqrt{0.0446 / 6} = 0.0862$
   b. Potassium means, $s_{\bar{y}B} = \sqrt{0.0446 / 9} = 0.0704$
   c. N X K Means, $s_{\bar{y}AB} = \sqrt{0.0446 / 3} = 0.1219$
2. Differences between means
   a. Two N Means, $s_{\bar{d}A} = \sqrt{(2)(0.0446)/6} = 0.1219$
   b. Two K Means, $s_{\bar{d}B} = \sqrt{(2)(0.0446)/9} = 0.0996$
   c. Two N x K means, $s_{\bar{d}AB} = \sqrt{(2)(0.0446)/3} = 0.1724$

Presentation of Results :

The presentation of results depends on the significance of the factors. If the AB effect is significant the results are given in a two-way table of means with the standard error, $s_{\bar{y}AB}$ as follows:

Mean yield for combinations of factors, A and B.

	Factor B			
Factor A	1	2		B
1	$\bar{y}_{11}$	$\bar{y}_{12}$	. . .	$\bar{y}_{1b}$
2	$\bar{y}_{21}$	$\bar{y}_{22}$	. . .	$\bar{y}_{2b}$
.	.	.		.
.	.	.		.
.	.	.		.
a	$\bar{y}a1$	$\bar{y}a2$	. . .	$\bar{y}_{ab}$

Standard error $=\sqrt{MSE/r}$

Where $\bar{y}_{ij} = T_{ij}./r$

If the AxB effect is not significant but one or the other of the factors is significant the results are given in one-way tables of means for the significant factors:

Mean yields at various levels of factor A

Factor A	1	2	. . . .	a	Standard Error
Mean	$\bar{y}_1$	$\bar{y}_2$	. . .	$\bar{y}_a$	$s\bar{y}_A = \sqrt{MSE/rb}$

Where $\bar{y}_i = A_i / rb$

Mean yields at various levels of factor B

Factor B	1	2	. . .	b	Standard Error
Mean	$\bar{y}_1$	$\bar{y}_2$	. . .	$\bar{y}_b$	$s_{\bar{y}b} = \sqrt{MSE /ra}$

Where $\bar{y}_j = B_j / ra$

In the present example the A x B effect was not significant, nor was the Potassium factor. The Nitrogen factor was highly significant. We can summarize the results in the following table :

Table I. Mean yield of barley at different level of added Nitrogen

Nitrogen (Kg/Ha)	0	25	50	Standard Error :
Mean yield (T/Ha)	1.44	1.74	1.95	0.09

Exercise

A barley breeder wanted to determine the effect of row spacing and planting date on the yield of barley. He conducted a factorial experiment with the following factors:

Factor A. Spacing: S_1= 15cm, S_2= 20cm, S_3= 25cm, S_4= 30cm.

Factor B. Planting Date: D_1= Early, D_2= Late.

The experiment was run using a randomized block design with three blocks. The field lay-out and yield (kg/ha) were as follows:

I		II		III	
S_2D_2 3308	S_4D_1 3883	S_1D_2 2537	S_3D_1 2611	S_3D_2 2487	S_2D_1 2362'
S_3D_2 3642	S_1D_2 2867	S_4D_2 3546	S_1D_1 2315	S_1D_1 1454	S_3D_1 2458
S_2D_1 2966	S_3D_1 3142	S_3D_2 2759	S_2D_1 2610	S_2D_2 2470	S_4D_1 2583
S_1D_1 2417	S_4D_2 4058	S_2D_2 2685	S_4D_1 2792	S_4D_2 2662	S_1D_2 2162

Conduct a statistical analysis of these results.

SPLIT - PLOT EXPERIMENTS

Characteristics

Split-Plot Experiments are factorial Experiments in which the levels of one factor are assigned at random to large plots (whole plots, main Plots) within blocks. The large plots are then divided into small plots (Split-Plots, Sub-Plots) and the levels of the second factor are assigned at random to the small plots within the large plots.

For example, an agronomist wanted to study the effect of three irrigation treatments and three Nitrogen fertilizer treatments on the yield of barley. The irrigation treatments require large plots, while the Nitrogen treatments, which can be spread by hand, can be applied to small plots. He used the following factors:

Factor A = Irrigation: I_1= None, I_2= Once, I_3= Twice

Factor B = Nitrogen: N_1= None, N_2= 25 Kg/Ha, N_3= 50 Kg/Ha

The experiment was run in four blocks. The field lay-out and yield (Kg/Plot) were

I

N_2 24	N_1 8	N_1 20
N_3 25	N_2 14	N_3 33
N_1 18	N_3 23	N_2 35

Irrigation:

I_2 I_1 I_3

II

N_1 14	N_2 20	N_1 23
N_3 21	N_3 24	N_2 31
N_2 9	N_1 20	N_3 38

I_1 I_2 I_3

BLOCK:-

III

Irrigation	I_1	I_3	I_2
	N_1 11	N_3 36	N_3 22
	N_3 18	N_1 22	N_2 24
	N_2 11	N_2 35	N_1 16

IV

Irrigation	I_3	I_1	I_2
	N_2 30	N_3 20	N_1 9
	N_1 23	N_1 11	N_3 29
	N_3 36	N_2 20	N_2 20

Calculations for Statistical Analysis:

Assume that factor A at a levels is assigned to the whole plots, and factor B at b levels is assigned to the split-plots. Assume there are r blocks. Let

y_{ijk} = Yield of the i th level of factor A, jth level of factor B in the kth block.

1. Make tables of totals as follows :

   a. Factor A by block totals

Factor A	Block 1	2	. . .	r	Sum
1	$T_{1.j}$	$T_{1.2}$	. . .	$T_{1.r}$	A_1
2	$T_{2.1}$	$T_{2.2}$	. . .	$T_{2.r}$	A_2
⋮	⋮	⋮		⋮	⋮
a	$T_{a.1}$	$T_{a.2}$	. . .	$T_{a.r}$	A_a
Sum	R_1	R_2	. . .	R_r	G

Where

$T_{i.k} = \sum_{j=1}^{b} y_{ijk}$ = Sum of yields in ith level of Factor A in the kth block summed over levels of factor B.

$A_i = \sum_{k=1}^{r} T_{i.k}$ = Sum of all yields on ith level of A.

$R_k = \sum_{i=1}^{a} T_{i.k}$ = Sum of all yields in the kth block.

$G = \sum_{i=1}^{a} A_i = \sum_{k=1}^{r} R_k$ = Grand total of all yields.

b. Factor A by factor B totals

Factor A	Factor B 1	2	. . .	b	Sum
1	$T_{11.}$	$T_{12.}$	. . .	$T_{1b.}$	A_1
2	$T_{21.}$	$T_{22.}$	. . .	$T_{2b.}$	A_2
⋮	⋮	⋮		⋮	⋮
a	$T_{a1.}$	$T_{a2.}$	. . .	$T_{ab.}$	A_a
Sum	B_1	B_2	. . .	B_b	G

in which

$T_{ij.} = \sum_{k=1}^{r} y_{ijk}$ = Sum of yields on ith level of factor A, jth level of factor B summed over blocks.

$B_j = \sum_{i=1}^{a} T_{ij.}$ = Sum of all yields on j th level of B_j and Ai, G are as before.

In the numberical example these tables are :

a. Factor A x Block

Block

Irrigation	I	II	III	IV	Sum
1	45	44	40	51	180
2	67	64	62	58	251
3	88	92	93	89	362
Sum	200	200	195	198	793

Sample Calculations :

$$T_{1.1} = \sum_{j=1}^{3} y_{1j1} = 8 + 14 + 23 = 45$$

$$A_1 = \sum_{k=1}^{4} T_{1.k} = 45 + 44 + 40 + 51 = 180$$

$$R_1 = \sum_{i=1}^{3} T_{i.1} = 45 + 67 + 88 = 200$$

$$G = \sum_{i=1}^{3} A_i = \sum_{k=1}^{4} R_k = 180 + 251 + 362 = 200 + 200 + 195 + 198 = 793$$

b. Factor A x Factor B

Nitrogen

Irrigation	1	2	3	Sum
1	44	54	82	180
2	63	88	100	251
3	88	131	143	362
Sum	195	273	325	793

Sample Calculations:

$$T_{11.} = \sum_{k=1}^{4} y_{11k} = 8 + 14 + 11 + 11 = 44$$

$$B_1 = \sum_{i=1}^{3} T_{i1.} = 44 + 63 + 88 = 195 .$$

2. Fill in a table of preliminary ANOVA computations:

(1) Source of Variation	(2) Number of Totals Squared	(3) Observations Per Total	(4) Sum of $(\text{Total})^2$	(5) Raw SS (4) / (3)
Correction	1	rab	G^2	$CT = G^2/rab$
Total	rab	1	$\sum_i \sum_j \sum_k y_{ijk}^2$	$\sum_i \sum_j \sum_k y_{ijk}^2$
Block	r	ab	$\sum_k R_k^2$	$(1/ab) \sum_k R_k^2$
A	a	rb	$\sum_i A_i^2$	$(1/rb) \sum_i A_i^2$
Error (A)	ra	b	$\sum_i \sum_k T_{i.k}^2$	$(1/b) \sum_k T_{i.k}^2$
B	b	ra	$\sum_j B_j^2$	$(1/ra) \sum_j B_j^2$
Ab	ab	r	$\sum_i \sum_j T_{ij.}^2$	$(1/r) \sum_i \sum_j T_{ij.}^2$

For the numberical example this table is

(1) Source of Variation	(2) Number of Totals Squared*	(3) Observations Per Total	(4) Sum of $(Total)^2$	(5) Raw SS (4)/(3)
Correction	1	36	628,849	CT= 17,468.02
Total	36	1	19,911	19,911.00
Block	4	9	157,229	17,469.89
I	3	12	226,445	18,870.42
Error (A)	12	3	56,733	18,911.00
N	3	12	218,179	18,181.58
IN	9	4	78,643	19,660.75

* In this example : r=4, a=3, b=3

Sample calculations : (4)

- Correction, $G^2 = (7931)^2 = 628\ 849$
- Total, $\sum_i \sum_j \sum_k y_{ijk}^2 = 24^2 + 8^2 + \ldots + 20^2 = 19{,}911$
- Block, $\sum_k R_k^2 = 200^2 + 200^2 + 195^2 + 198^2 = 157{,}229$
- I, $\sum_i A_i^2 = 180^2 + 251^2 + 362^2 = 226{,}445$
- Error (A), $\sum_i \sum_k T_{i.k}^2 = 45^2 + 44^2 + \ldots + 89^2 = 56{,}733$
- N, $\sum_J B_j^2 = 195^2 + 273^2 + 325^2 = 218{,}179$
- IN, $\sum_i \sum_j T_{ij.}^2 = 44^2 + 54^2 + \ldots + 143^2 = 78{,}643$

3. Complete the analysis of variance, ANOVA, table!!

Source of Variation	Degree of Freedom	Sum of Squares	Mean Square	F
Total	rab - 1	SSTOT		
Block	r - 1	SSR	MSR	F_R
A	a - 1	SSA	MSA	F_A
Error (A)	(r-1)(a-1)	SSEA	MSEA	
B	b - 1	SSB	MSB	F_B
AB	(a-1)(b-1)	SSAB	MSAB	F_{AB}
Error (AB)	a(r-1)(b-1)	SSEAB	MSEAB	

Calculation

Entries in the first three columns are taken from the prelimi-nary table.

- Sources of variation : (1) with "correction" omitted and "error (AB)" added.
- Degrees of freedom : (2) minus 1 except error (A) d.f.= (r - 1)(a - 1)
  AB d.f. = (a - 1)(b - 1)
  Error (AB) d.f. = a(r - 1)(b - 1)
- Sums of squares : (5) minus "CT" except
  SSEA = (5) - SSR - SSA - CT
  SSAB = (5) - SSA - SSB - CT
  SSEAB = SSTOT - SSR - SSA - SSEA - SSB - SSAB
- Means squares (not computed for total).
  Divide the sum of squares by the degrees of freedom on the same line in the ANOVA table.
- F column. Entries are computed as follows :

$$F_R = MSR/MSEA$$

$$F_A = MSA/MSEA$$

$$F_B = MSB/MSEAB$$

$$F_{AB} = MSAB/MSEAB$$

The ANOVA table for the numerical example is

ANOVA

Source	d.f.	SS	MS	F
Total	35	2,442.98		
Block	3	1.87	.62	.10
I	2	1,402.40	701.20	108.71
Error(A)	6	38.71	6.45	
N	2	713.56	356.78	30.62
IN	4	76.77	19.19	1.65
Error(AB)	18	209.67	11.65	

Sample Calculations :

- Sums of squares

  SSTOT = ⑤ - CT = 19,911 - 17,468.02 = 2,442.98
  SSR = ⑤ - CT = 17,469.89 - 17,468.02 = 1.87
  SSA = ⑤ - CT = 18,870.42 - 17,468.02 = 1,402.40
  SSEA = ⑤ - CT - SSR - SSA = 18,911.00 - 17,468.02 - 1.87 - 1,402.40 = 38.71
  SSB = ⑤ - CT = 18,181.58 - 17,468.02 = 713.56
  SSAB = ⑤ - CT - SSA - SSB = 19,660.75 - 17,468.02 - 1,402.40 - 713.56 = 76.77
  SSEAB = SSTOT - SSR - SSA - SSEA - SSB - SSAB
  = 2.442.98 - 1.87 - 1,402.40 - 38.71 - 713.56 - 76.77 = 209.67

- Means squares = SS/d.f.

- F values

  F_R = MSR/MSEA = 1.87 / 6.45 = .10

  F_A = MSA / MSEA = 701.20 / 6.45 = 108.71

  F_B = MSB / MSEAB = 356.78 / 11.65 = 30.62

  F_{AB} = MSAB / MSEAB = 19.19 / 11.65 = 1.65

4. Significance Tests

Entries in the F column are used to test the significance of the factorial effects. We can look in the 5% and 1% F tables in the column headed by the degrees of freedom for the effect being tested, and the row for the degrees of freedom for the mean square in the denominator of the F ratio. If the F in the ANOVA table is larger than the 1% F the effect is termed highly significant; if it is larger than the 5% F but smaller than the 1% F the effect is termed significant; if it is smaller than the 5% F the effect is not considered significant. F_R provides an approximate test for the effectiveness of blocking in reducing error.

In the example given above the 5% F at 2 and 6 d.f. is 5.14 while the 1% F is 10.92. Since 108.71 is larger than 10.92 the effect of irrigation is highly significant. At 2 and 18 d.f. the 5% F is 3.55 and the 1% F is 6.01. Since 30.62 is larger than 6.01 the nitrogen effect is considered highly significant. At 4 and 18 d.f. the 5% F is 2.93 and the 1% F is 4.58. Since 1.65 is smaller than 2.93 the irrigation by nitrogen effect is not significant. At 3 and 6 d.f. the 5% F is 4.76. Since 0.10 is less than 4.76 blocking was not effective in reducing experimental error.

5. Standard Error

There are two types of standard error: standard errors of means, and standard errors of differences between means.

a. Standard errors of means : $s_{\bar{y}}$

i) Factor A means, $s_{\bar{y}A}$

$$s_{\bar{y}A} = \sqrt{MSEA / rb}$$

ii) Factor B means, $s_{\bar{y}B}$

$$s_{\bar{y}B} = \sqrt{MSEAB / ra}$$

iii) AxB individual means, $s_{\bar{y}B}$

$$s_{\bar{y}B} = \sqrt{MSEAB / r}$$

b. Differences between means, $s_{\bar{d}}$

i) Difference between two A means, $\bar{y}_{a2} - \bar{y}_{a1}$

$$s_{\bar{d}A} = \sqrt{2\ MSEA\ /\ rb}$$

ii) Difference between two B means, $\bar{y}_{b2} - \bar{y}_{b1}$

$$s_{\bar{d}B} = \sqrt{2\ MSEAB\ /\ ra}$$

iii) Difference between two B means at the same level of A, $\bar{y}_{a1b2} - \bar{y}_{a1b1}$

$$s_{\bar{d}AB} = \sqrt{2\ MSEAB\ /\ r}$$

iv) Difference between two A means at same or different level of B, $\bar{y}_{a2b2} - \bar{y}_{a1b2}$ or $\bar{y}_{a2b2} - \bar{y}_{a1b1}$

$$s_{\bar{d}} = \sqrt{2\left[(b-1)\ MSEAB + MSEA\right]\ /\ rb}$$

For the numerical example the standard errors are :

a. Means

i) Irrigation means

$$s_{\bar{y}A} = \sqrt{MSEA/\ rb} = \sqrt{6.45\ /\ (4)(3)} = .73$$

ii) Nitrogen Means

$$s_{\bar{y}B} = \sqrt{MSEAB\ /\ ra} = \sqrt{11.65\ /\ (4)(3)} = .99$$

iii) Irrigation & Nitrogen means

$$s_{\bar{y}AB} = \sqrt{MSEAB\ /\ r} = \sqrt{11.65\ /\ 4} = 1.71$$

D. Differences between means

i) Two irrigation means

$$s_{\bar{d}A} = \sqrt{2MSEA / rb} = \sqrt{(2)(6.45)/(4)(3)} = 1.04$$

ii) Two nitrogen means

$$s_{\bar{d}B} = \sqrt{2MSEAB / ra} = \sqrt{(2)(11.65)/(4)(3)} = 1.39$$

iii) Two nitrogen means at same level of irrigation

$$s_{\bar{d}AB} = \sqrt{2MSEAB / r} = \sqrt{(2)(11.65)/4} = 2.41$$

iv) Two irrigation means at same or different level of nitrogen

$$s_{\bar{d}} = \sqrt{2\left[(b-1)\ MSEAB + MSEA\right]/rb}$$

$$= \sqrt{2\left[(3-1)(11.65) + 6.45\right]/(4)(3)} = 2.23$$

6. Presentation of results

As in the ordinary factorial experiment, the presentation of results depends on the outcome of the significance tests. If the AB effect is significant the results are summarized in a two-way table of AxB means as follows:

Mean yield at various combinations of factors A and B

Factor B

Factor A	1	2	. . .	b
1	$\bar{y}_{11}$	$\bar{y}_{12}$	. . .	$\bar{y}_{1b}$
2	$\bar{y}_{21}$	$\bar{y}_{22}$	. . .	$\bar{y}_{2b}$
.	.	.		.
.	.	.		.
.	.	.		.
a	$\bar{y}_{a1}$	$\bar{y}_{a2}$		$\bar{y}_{ab}$

Standard error = $\sqrt{MSEAB / r}$

Where

$\bar{y}_{ij} = T_{ij.}/ r$

If AB is not significant the results are presented in one-way tables of means for the significant A or B factor, or both. We have

Mean yield at various levels of factor A

Factor A	1	2	. . .	a	Standard error
Mean	$\bar{y}_{1.}$	$\bar{y}_{2.}$	. . .	$\bar{y}_{a.}$	$s_{\bar{y}A} = \sqrt{MSEA / rb}$

In which

$\bar{y}_{i.} = A_i / rb$

If factor B is significant we use

Mean yield at various levels of factor B

Factor B	1	2	. . .	b	Standard error
Mean	$\bar{y}_{.1}$	$\bar{y}_{.2}$	. . .	$\bar{y}_{.b}$	$s_{\bar{y}B} = \sqrt{MSEAB / ra}$

Where

$\bar{y}_{.j} = B_j / ra$

In the present example the irrigation x nitrogen (IN) effect is not significant. However both the I effect and the N effect are highly significant. We can summarize the result in two tables of means:

Table I. Mean Yield of Barley (Kg/Plot) under Three Irrigation Treatments.

Irrigation	None	Once	Twice	Standard Error
Mean	15.00	20.92	30.17	.73

Table 2. Mean Yield of Barley (Kg/Plot) under Three Nitrogen Treatments.

Nitrogen	None	25 Kg/Ha	50 Kg/Ha	Standard error
Mean	16.25	22.75	27.08	.99

Exercise

An experiment was designed to test the effect of previous treatment and nitrogen fertilizer on the yield of wheat. Three blocks were divided into three plots each and received the following treatments

(Factor A) :-

C_1 = Fallow, C_2 = Melons, C_3 = Lentils.

In the late summer the plots were split and the split-plots received the following treatments

(Factor B) :-

N_1 = No Nitrogen, N_2 = 50 Kg/Ha Nitrogen.

The area was then uniformly seeded to wheat, and yield determined the following summer. The field plan and yield (Kg/Plt) were as follows :

	C_3		C_1		C_2	
I	N_2 25.3	N_1 21.0	N_1 15.5	N_2 22.2	N_1 13.8	N_2 19.3

	C_1		C_3		C_2	
II	N_1 15.0	N_2 24.2	N_1 22.7	N_2 24.8	N_2 18.0	N_1 13.5

	C_2		C_1		C_3	
III	N_2	N_1	N_2	N_1	N_1	N_2
	20.5	13.2	25.4	15.2	22.3	28.4

Conduct a statistical analysis of the results of this experiment.

STRIP - PLOT EXPERIMENTS

Characteristics :

Strip-Plot Experiments are a special case of Split-Plot Experiments. In Strip-Plot trials the a levels of factor A are assigned at random to strips of plots within the r blocks, using a different randomization in each block. The b levels of factor B are then assigned at random to strips of plots, within the blocks, which are at right angles to the factor A strips.

For example, an agronomist wanted to measure the effect of two Autumn or Fall tillage treatments and three spring tillage treatments on the yield of wheat. Because of the size of machinery involved the spring treatments were applied to strips of plots at right angles to the Autumn treatments. He used the following treatments :

Factor A = Autumn(fall) tillage : F_1= Chisel, F_2= Subsoil

Factor B = Spring tillage S_1= Plow, S_2= Sweep, S_3=Offset disk.

The experiment was run in three blocks. The field layout and yields (Kg/Hectare) were

I	Plow S_1	Disk S_3	Sweep S_2
Chisel F_1	312	315	278
Subsoil F_2	318	222	267

II	Disk S_3	Plow S_1	Sweep S_2
Subsoil F_2	334	374	296
Chisel F_1	314	350	286

III	Disk S_3	Sweep S_2	Plow S_1
Subsoil F_2	298	228	384
Chisel F_1	312	309	361

Calculations for Statistical Analysis

Assume that there are a levels of factor A, b levels of Factor B. and r blocks. Let

y_{ijk} = Yield of the ith level of Factor A, jth level of factor B in the kth block.

1. Make Tables of Totals as follows:

a) Factor A by block totals

Factor A	Block 1	2	. . .	r	Sum
1	$T_{1.1}$	$T_{1.2}$	. . .	$T_{1.r}$	A_1
2	$T_{2.1}$	$T_{2.2}$	. . .	$T_{2.r}$	A_2
⋮	⋮	⋮		⋮	⋮
a	$T_{a.1}$	$T_{a.2}$	. . .	$T_{a.r}$	A_a
Sum	R_1	R_2	. . .	R_r	G

Where $T_{i.k} = \sum_{j=1}^{b} y_{ijk}$ = Sum of yields on ith level of A in the kth block summed over levels of B.

$A_i = \sum_{k=1}^{r} T_{i.k}$ = Sum of all yields on ith level of A.

$R_k = \sum_{i=1}^{a} T_{i.k}$ = Sum of all yields in the kth block.

$$G = \sum_{i=1}^{a} A_i = \sum_{k=1}^{r} R_k = \text{Grand total of all yields.}$$

b) Factor B by block Totals :-

Block

Factor B	1	2	. . .	r	Sum
1	$T_{.11}$	$T_{.12}$	. . .	$T_{.1r}$	B_1
2	$T_{.21}$	$T_{.22}$	. . .	$T_{.2r}$	B_2
⋮	⋮	⋮		⋮	⋮
b	$T_{.b1}$	$T_{.b2}$		$T_{.br}$	B_b
Sum	R_1	R_2	. . .	R_r	G

In which $T_{.jk} = \sum_{i=1}^{a} y_{ijk}$ = Sum of yields on the jth level of B in the kth block summed over levels of A.

$B_j = \sum_{k=1}^{r} T_{.jk}$ = Sum of all yields on jth level of B_j and R_k, G are as before.

c) Factor A by Factor B totals

Factor B

Factor A	1	2	. . .	b	Sum
1	$T_{11.}$	$T_{12.}$	. . .	$T_{1b.}$	A_1
2	$T_{21.}$	$T_{22.}$	. . .	$T_{2b.}$	A_2
⋮	⋮	⋮		⋮	⋮
a	$T_{a1.}$	$T_{a2.}$	. . .	$T_{ab.}$	A_a
Sum	B_1	B_2	. . .	B_b	G

Where $T_{ij.} = \sum_{k=1}^{r} y_{ijk}$ = Sum of yields on ith level of A, jth level of B summed over blocks and A_i, B_j, G are as before.

For the numerical example these tables are

a) Autumn tillage (A) by block

Block

Autumn(Fall)	I	II	III	Sum
F_1 Chisel	905	950	982	2837
F_2 Subsoil	807	1004	910	2721
Sum	1712	1954	1892	5558

Sample Calculations :

$$T_{1.1} = \sum_{j=1}^{3} y_{1j1} = 312 + 315 + 278 = 905$$

$$A_1 = \sum_{k=1}^{3} T_{1.k} = 905 + 950 + 982 = 2837$$

$$R_1 = \sum_{i=1}^{2} T_{i.1} = 905 + 807 = 1712$$

$$G = \sum_{i=1}^{2} A_i = \sum_{k=1}^{3} R_k = 2837 + 2721 = 1712 + 1954 + 1892 = 5558$$

b) Spring tillage (B) by block

Block

Spring	I	II	III	Sum
S_1 Plow	630	724	745	2099
S_2 Sweep	545	582	537	1664
S_3 Disk	537	648	610	1795
Sum	1712	1954	1892	5558

Sample Calculations :

$$T_{.11} = \sum_{i=1}^{2} y_{i11} = 312 + 318 = 630$$

$$B_1 = \sum_{k=1}^{3} T_{.1k} = 630 + 724 + 745 = 2099$$

c) Fall tillage (A) by Spring tillage (B) Totals

Spring

Autumn(Fall)	Plow S_1	Sweep S_2	Disk S_3	Sum
Chisel F_1	1023	873	941	2837
Subsoil F_2	1076	791	854	2721
Sum	2099	1664	1795	5558

Sample Calculations :

$$T_{.1.} = \sum_{k=1}^{3} y_{11k} = 312 + 350 + 361 = 1023$$

2. Fill in a table of preliminary ANOVA calculations :

① Source of Variation	② Number of Totals Squared	③ Observations Per Total	④ Sum of $(\text{Total})^2$	⑤ Raw SS ④/③
Correction	1	rab	G^2	$CT = G^2/rab$
Total	rab	1	$\sum_i \sum_j \sum_k y_{ijk}^2$	$\sum_i \sum_j \sum_k y_{ijk}^2$
Block	r	ab	$\sum_k R_k^2$	$(1/ab) \sum_k R_k^2$
A	a	rb	$\sum_i R_i^2$	$(1/rb) \sum_i R_i^2$
Error (A)	ra	b	$\sum_i \sum_k T_{i.k}^2$	$(1/b) \sum_i \sum_k T_{i.k}^2$
B	b	ra	$\sum_j B_j^2$	$(1/ra) \sum_j B_j^2$
Error (B)	rb	a	$\sum_j \sum_k T_{.jk}^2$	$(1/a) \sum_j \sum_k T_{.jk}^2$
AB	ab	r	$\sum_i \sum_j T_{ij.}^2$	$(1/r) \sum_i \sum_j T_{ij.}^2$

Source of Variation	Number of Totals Squared*	Observations Per Total	Sum of $(\text{Total})^2$	Raw SS ④/③
Correction	1	18	30,891,364	CT= 1,716,187
Total	18	1	1,748,880	1,748,880
Block	3	6	10,328,724	1,721,454
F	2	9	15,452,410	1,716,934
Error (A)	6	3	5,173,214	1,724,405
S	3	6	10,396,722	1,732,787
Error (B)	9	2	3,480,592	1,740,296
FS	6	3	5,206,912	1,735,637

* $r = 3$, $a = 2$, $b = 3$

Sample Calculations : (4)

Correction, $G^2 = (5558)^2 = 30,891,364$

Total, $\sum_i \sum_j \sum_k Y_{ijk} = 312^2 + 315^2 + \ldots + 361^2 = 1,748,880$

Block, $\sum_k R_k 2 = 1712^2 + 1954^2 + 1892^2 = 10,328,624$

F $\sum_i A_i 2 = 2837^2 + 2721^2 + 15,452,410$

Error (A) $\sum_i \sum_k T_{i.k} 2 = 905^2 + 950^2 + \ldots + 910^2 = 5,173,214$

S $\sum_j B_j 2 = 2099^2 + 1664^2 + 1795^2 = 10,396,722$

Error (B) $\sum_j \sum_k T_{.jk} 2 = 630^2 + 724^2 + \ldots + 610^2 = 3,480,592$

FS $\sum_i \sum_j T_{ij.} 2 = 1023^2 + 873^2 + \ldots + 854^2 = 5,206,912$

3. Complete the analysis of variance, ANOVA, table :

ANOVA

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Total	rab - 1	SSTOT		
Block	r - 1	SSR	MSR	F_R
A	a - 1	SSA	MSA	F_A
Error (A)	(r-1)(a-1)	SSEA	MSEA	
B	b - 1	SSB	MSB	F_B
Error (B)	(r-1)(b-1)	SSEB	MSEB	
AB	(a-1)(b-1)	SSAB	MSAB	F_{AB}
Error (AB)	(r-1)(a-1)(b-1)	SSEAB	MSEAB	

Computations :

Entries in the first three colums are taken from the preliminary table.

- Source of variation : (1) with "Correction" omitted and "Error (AB)" added
- Degrees of Freedom : (2) minus 1 except

  Error (A) d.f. = (r-1)(a-1)

  Error (B) d.f. = (r-1)(b-1)

  AB d.f. = (a-1)(b-1)

  Error (AB) d.f.= (r-1)(a-1)(b-1)

- Sums of squares : (5) minus "CT" except

  SSEA = (5) - SSR - SSA - CT

  SSEB = (5) - CT - SSR - SSB

  SSAB = (5) - CT - SSA - SSB

  SSEAB = SSTOT - SSR - SSA - SSEA - SSB - SSEB - SSAB

- Mean squares : (Not computed for total)

  Divide each sum of squares by the degrees of freedom on the same line in the ANOVA table.

- F column :

  F_R = MSR / MSEA

  F_A = MSA / MSEA

  F_B = MSB / MSEB

  F_{AB} = MSAB / MSEAB

The ANOVA table for the numerical example is

ANOVA

Source	d.f.	SS	MS	F
Total	17	32,693		
Block	2	5,267	2,633	2.39
Autumn/Fall	1	747	747	0.68
Error (A)	2	2,204	1,102	
Spring	2	16,600	8,300	14.82
Error (B)	4	2,242	560	
FS	2	2,103	1,052	1.19
Error (AB)	4	3,530	882	

Sample calculations :

- Sum of Squares

SSTOT = ⑤ - CT = 1,748,880 - 1,716,187 = 32,693

SSR = ⑤ - CT = 1,721,454 - 1,716,187 = 5,267

SSA = ⑤ - CT = 1,716,934 - 1,716,187 = 747

SSEA = ⑤ - CT - SSR - SSA

= 1,724,405 - 1,716,187 - 5,267 - 747 = 2,204

SSB = ⑤ - CT = 1,732,187 - 1,716,187 = 16,600

SSEB = ⑤ - CT - SSR - SSB

= 1,740,296 - 1,716,187 - 5,267 - 16,600 = 2,242

SSAB = ⑤ - CT - SSA -SSB

= 1,735,637 - 1,716,187 - 747 - 16,600 = 2,103

SSEAB = SSTOT - SSR - SSA - SSEA - SSB - SSEB -SSAB

= 32,693-5,267-747-2,204-16,600-2,242-2,103= 3,530

- Mean squares = SS/df

- F values

F_R = MSR / MSEA = 2,633 / 1,102 = 2.39

F_A = MSA / MSEA = 747 / 1,102 = 0.68

F_B = MSB / MSEB = 8,300 / 560 = 14.82

F_{AB} = MSAB / MSEAB = 1,052 /882 = 1.19

4. Significance tests :

The significance of the A, B and AB effects are tested with the F values in the ANOVA table. First look in the 5% and 1% F tables in the column for the degrees of freedom of the effect under test and the row for the degrees of freedom for the mean square in the denominator of the F ratio. If the F in the ANOVA table is larger than the 1% F the effect is highly significant; if it is larger than the 5% F but smaller than the 1% F the effect is significant; if it is smaller than the 5% F the effect is not significant. F_R provides an approximate test of the effectiveness of blocking in reducing error.

In the numerical example, at 1 and 2 d.f. the 5% F is 18.51 and the 1% F is 98.50. At 2 and 4 d.f. the 5% F is 6.94 and the 1% F is 18.00. Since 14.82 is larger than 6.94 but smaller than 18.00 the spring treatments are significant. None of the rest of the effects have F values larger than the 5% F, so none of them are significant.

5. Standard Errors :

a. Standard errors of means : $s_{\bar{y}}$

- Factor A means, $s_{\bar{y}A}$

$$s_{\bar{y}A} = \sqrt{MSEA / rb}$$

- Factor B means, $s_{\bar{y}B}$

$$s_{\bar{y}B} = \sqrt{MSEB} / ra$$

- AxB individual means, $s_{\bar{y}AB}$

$$s_{\bar{y}AB} = \sqrt{MSEAB / r}$$

b. Differences between means, $s_{\bar{d}}$

- Difference between 2 A means, $\bar{y}A2 - \bar{y}A1$

$$s_{\bar{d}A} = \sqrt{2\ MSEA\ /\ rb}$$

- Difference between 2 B means, $\bar{y}_{B2} - \bar{y}_{B1}$

$$s_{\bar{d}B} = \sqrt{2\ MSEB\ /\ ra}$$

- Difference between 2 A means at the same level of B, $\bar{y}_{A2B2} - \bar{y}_{A1B2}$

$$s_{\bar{d}AB} = \sqrt{2\left[(b-1)\ MSEAB + MSEA\right]/rb}$$

- Difference between 2 B means at the same level of A, $\bar{y}_{A2B2} - \bar{y}_{A2B1}$

$$s_{\bar{d}BA} = \sqrt{2\left[(a-1)MSEAB + MSEB\right]/ra}$$

- Difference between 2 A means at different levels of B, $\bar{y}_{A2B2} - \bar{y}_{A1B1}$

$$s_{\bar{d}} = \sqrt{2\left[(ab-a-b)\ MSEAB + a\ MSEA + b\ MSEB\right]/rab}$$

For the present numerical example the standard errors are:

a. Means

- Fall treatment means

$$s_{\bar{y}_A} = \sqrt{MSEA\ /\ rb} = \sqrt{1102\ /(3)(3)} = 11.06$$

- Spring treatment means

$$s_{\bar{y}_B} = \sqrt{MSEB\ /\ ra} = \sqrt{560\ /(3)(2)} = 9.66$$

- Fall X Spring means

$$s_{\bar{y}_{AB}} = \sqrt{MSEAB\ /\ r} = \sqrt{882\ /\ 3} = 17.15$$

b. Differences between means

- Two Fall means

$$s_{\bar{d}_A} = \sqrt{2\ MSEA\ /\ rb} = \sqrt{(2)(1102)/(3)(3)} = 15.65$$

- Two Spring means

$$s_{\bar{d}B} = \sqrt{2\ MSEB\ /\ ra} = \sqrt{(2)(560)/(3)(2)} = 13.66$$

- Two Fall means with same Spring treatment

$$s_{\bar{d}_{AB}} = \sqrt{2\left[(b-1)MSEAB + MSEA\right]\ /rb}$$
$$= \sqrt{2\left[(3-1)(882) + 1102\right]\ /\ (3)(3)} = 25.24$$

- Two Spring means with same fall treatment

$$s_{\bar{d}_{BA}} = \sqrt{2\left[(a-1)MSEAB + MSEB\right]\ /ra}$$
$$= \sqrt{2\left[(2-1)(882) + 560\right]\ /(3)(2)} = 21.92$$

- Two Fall means with different Spring treatment

$$s_{\bar{d}} = \sqrt{2\left[(ab-a-b)\ MSEAB + MSEA + bMSEB\right]\ /rab}$$
$$= \sqrt{2\left[(2\times3-2-3)(882)+(2)(1102)+(3)(560)\right]\ /(3)(2)(3)}$$
$$= 23.01$$

6. Presentation of Results :

The procedure follows that of the split-plot. First look at the AB effect. If it is significant the results are summarized in a two-way table of A x B means:

Mean yield at various combinations of factors A&B

Factor B

Factor A	1	2	. . .	b
1	$\bar{y}_{11}$	$\bar{y}_{12}$	. . .	$\bar{y}_{1b}$
2	$\bar{y}_{21}$	$\bar{y}_{22}$	. . .	$\bar{y}_{2b}$
⋮	⋮	⋮		⋮
a	$\bar{y}_{a1}$	$\bar{y}_{a2}$	. . .	$\bar{y}_{ab}$

$$\text{Standard error} = \sqrt{MSEAB / r}$$

Where $\bar{y}_{ij} = T_{ij.} / r$

If the AB effect is not significant the results are summarized in one-way tables of means for the significant A or B factors. If A is significant we have :-

Mean yield at various levels of Factor A.

Factor A	1	2	. . .	a	Standard error
Mean	$\bar{y}_{1.}$	$\bar{y}_{2.}$	. . .	$\bar{y}_{a.}$	$s\bar{y}_A = \sqrt{MSEA / rb}$

Where $\bar{y}_{i.} = A_i / rb$

If Factor B is significant the summary table is

Mean Yield at various levels of Factor B.

Factor B	1	2	. . .	b	Standard Error
Mean	$\bar{y}_{.1}$	$\bar{y}_{.2}$	. . .	$\bar{y}_{.b}$	$s\bar{y}_B = \sqrt{MSEB / ra}$

In which $\bar{y}_{.j} = B_j / ra$

In the present numerical example the only significant effect was that of the spring treatments. We can summarize the results as follows :

Table I. Yield of Wheat (Kg/Hectare) under several Spring Tillage Treatments.

Spring Tillage	Plow	Sweep	Disk	Standard error
Mean	349.8	277.3	299.2	9.66

EXERCISE

An Agronomist wanted to measure the effect of row spacing and planting date on the yield of durum. He decided to use the following factors :

Factor A = Row Spacing: S_1 = 15cm, S_2 = 25cm, S_3 = 35cm.

Factor B = Planting Date : D_1 = Early, D_2 = Late

To make the conduct of the experiment easier row spacing were assigned to strips of two plots each within the blocks, while planting dates were assigned to strips of three plots at right angles to row spacings. There were three blocks. The field plan and yields (Kg/Plot) were

I

	S_3	S_1	S_2
D_1	56	32	49
D_2	67	54	58

II

	S_1	S_3	S_2
D_2	52	72	64
D_1	38	62	50

III

	S_2	S_1	S_3
D_2	63	54	68
D_1	54	44	51

Conduct a statistical analysis of the results of this experiment.

Augmented Designs
for
Preliminary Yield Trials

Plant Breeding activities at ICARDA includes the development and screening of a large number of new selections. As a part of the screening process the new selections are evaluated for yield in a preliminary yield trial. In the past these trials consisted of single rows of the new selections along with rows of one or more standard, or check, varieties placed systematically throughout the trial. The new varieties were evaluated, subjectively, by comparing their yield with that of a nearby check. Since the new selections were not replicated it was not possible to perform a valid statistical analysis of their yields.

In an effort to put yield analysis on a more sound statistical basis the Food Legume Imprvement Program and the Cereal Crops Improvement Program have adopted the use of " Augmented Designs " for some of their preliminary yield screening trials. These designs were developed by Federer (1956, 1961) and described by Federer and Ragavarao (1975). Their purpose is the evaluation, including statistical analysis, of a large number of new selections.

Design PLAN

The basic design plan is to divide the experimental area into a number of blocks of test plots. With augmented designs three or more check varieties are assigned at random to plots within each block, while the remaining plots in each block are assigned to the new selections under test. Although the check varieties are replicated the new selections are not. They are assigned at random to plots throughout the blocks. Yields of the new selections are adjusted for block differences, which are measured by the check varieties which occur in every block.

Blocks need not all contain the same number of plots. The trial is most efficient, however, if block size is the same for all blocks. Block size is determined by the number of blocks, b, the number of check varieties, c, and the number of new selections, v. If the block size is constant (same number of plots in each block) the following definitions and relationships hold :

c = number of check varieties per block

v = number of new selections

b = number of blocks

$n = v/b$ = number of new selections per block.

$p = c + n$ = number of plots per block.

$N = bc + v = b\,(c + n)$ = total number of plots.

The total number of blocks is determined by the need to have at least 10 degrees of freedom for error in the analysis of the yield data. This, in turn, is determined by the number of check varieties (c) used in the trial. In the analysis of variance of the check varieties the experminetal error has $(b-1)(c-1)$ degrees of freedom. As a result the number of blocks (b) must be such that the following inequality holds : $b > \left[10/(c-1) + 1\right]$. For example, with four check varieties :

$$b > \left[10/(4-1) + 1\right]$$
$$b > 4.33$$

The minimum number of blocks would be five. Each block would contain five or more plots depending on the number of new selections to be tested.

In constructing the design the checks should be randomly assigned to plots within each block. Little is lost, however, if one check is systematically assigned to, say, the first plot in each block. The other c-1 checks are assigned at random to c-1 of the remaining plots in each block. The v new selections are then assigned to the remaining bn plots in the trial.

For example, suppose we want to evaluate 24 new selections and use three checks with one systmeatically assigned to the first plot in each block. We would require a minimum of $b > \left[10/(3-1) + 1\right] = b > 6$

Assume we choose to use 6 blocks. Then

$c = 3$: A, B, C

$v = 24$: 1, 2, ..., 24

$b = 6$

$n = v/b = 24/6 = 4$

$p = c + n = 3 + 4 = 7$

$N = bc + v = (6)(3) + 24 = 42$

The field plan might appear as follows :

BLOCK					
I	II	III	IV	V	VI
A	A	A	A	A	A
13	17	21	2	B	19
8	9	C	B	12	C
B	C	15	10	5	20
C	24	B	C	16	B
18	B	1	3	6	4
7	11	23	22	C	14

ANALYSIS

The first step in the analysis is to construct a two-way table of check yields, totals, and means :

CHECK VARIETY	BLOCK 1	2	3	...	b	TOTAL	MEAN
1	x_{11}	x_{12}	x_{13}	...	x_{1b}	C_1	$\bar{x}_1$
2	x_{21}	x_{22}	x_{23}	...	x_{2b}	C_2	$\bar{x}_2$
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
c	x_{c1}	x_{c2}	x_{c3}	...	x_{cb}	C_c	$\bar{x}_c$
TOTAL	B_1	B_2	B_3	...	B_b	G	M

In this table

x_{ij} = Yield of the i^{th} check in the j^{th} block.

$B_j = \sum_i x_{ij}$ = sum of all checks in the j^{th} block.

$C_i \; = \; \sum_j x_{ij}$ = sum of all yields of the i^{th} check.

$G \; = \; \sum_j B_j \; = \sum_i C_i$ = Grand total of all check yields.

$\bar{x}_i \; = C_i/b$ = mean yield of the i^{th} check.

$M \; = \sum_i \bar{x}_i$ = G/b = sum of the check means.

The next step is to compute an dajustment factor, r_j, for each block. This is computed as :

$$r_j = (1/c)(B_j - M)$$

NOTE : as a check on the computation, $\sum_j r_j = 0$

A table of the actual yields of the new selections, and the yields adjusted for the effect of the block in which the new selection was grown can now be constructed :

		YIELD	
SELECTION	BLOCK	OBSERVED	ADJUSTED
1		y_{1j}	$\hat{y}_1$
2		y_{2j}	$\hat{y}_2$
.		.	.
.		.	.
.		.	.
v		y_{vj}	$\hat{y}_v$

Where :

y_{ij} = Yield of the i^{th} new selection (in the j^{th} block)

$\hat{y}_i$ = $y_{ij} - r_j$ = Adjusted yield of the i^{th} new selection (adjusted for block effect)

An estimate of experimental error which can be used to compute standard errors and L.S.D.'s is most easily obtained using an analysis of variance of the check yields. The format for this is :

ANOVA

SOURCE	d.f.	SS	MS
Total	bc - 1	SSTot	
Blocks	b - 1	SSB	
Cheecks	c - 1	SSC	
Error	(b-1)(c-1)	SSE	MSE

The entries in the ANOVA Table are computed as follows:

$$SSTot = \sum_i \sum_j x_{ij}2 - G^2/bc$$

$$SSB = (1/c) \sum_i B_j2 - G^2/bc$$

$$SSC = (1/b) \sum_i C_i2 - G^2/bc$$

$$SSE = SSTot - SSB - SSC$$

$$MSE = SSE/(b-1)(c-1)$$

NOTE : The ANOVA in this table is simply a randomized block ANOVA on the check yields.

An estimate of the experimental error is given by s^2 = MSE.

There are a number of kinds of differences to be computed in an augmented design. These differences and their variances are :

1. Difference between the means of two check varieties :

   $$\text{Variance} = 2\text{MSE}/b_c^{\ 2}$$

2. Difference between adjusted yields of two new selections in the same block :

   $$\text{Variance} = 2\text{MSE} = s_b^{\ 2}$$

3. Difference between adjusted yields of two new selections in different blocks:

   $$\text{Variance} = 2\text{MSE}/(c+1)/c = s_v^{\ 2}$$

4. Difference between the adjusted yield of a new selection and a check mean :

   $$\text{Variance} = \text{MSE}\ (b+1)(c+1)/bc = s_{vc}^{\ 2}$$

5. Average variance of the difference between adjusted yields of two new selections:

   $$\text{Variance} = \text{MSE}\ (2c+1)/c = s_a^{\ 2}$$

NOTE :An L.S.D. based on the average variance, $s_a^{\ 2}$, is satisfactory for comparing adjusted yields of two new selections in most cases.

Least significant differences (L.S.D.'s) may be computed using the variances, given above, in the following way :

$$\text{L.S.D.} = t_a \sqrt{s^2}$$

Where

t_a = the 100a% (5% or 1%) two-tailled t with (b-1) (c-1) degrees of freedom

s^2 = the variance of the differnce for which the L.S.D. is being computed.

4. NUMERICAL EXAMPLE

A cereal breeder wanted to conduct a preliminary yield trial on 30 new selections of Durum developed for use in the A rainfall zone of northern Jordan. He wanted to compare the new selections against three standared varieties : 1. ST = Stork, 2. CI = Cimmaron, and 3. WA = Waha. He had only enough seed of the new selections to plant a single 2.5 meter row of each, so he decided to use an augmented design. Since he wanted to include c = 3 standard varieties he required a minimum of :

$$b > \left[10/(c-1) + 1\right]$$

$$b > \left[10/(3-1) + 1\right]$$

$$b > 6$$

blocks to have sufficient degrees of freedom for estimating experimental error. Using six blocks he has a design with the following design characteristics :

1. Number of checks, c = 3 :ST, CI, WA
2. Number of new selections, v = 30 :(1), (2), ..., (30)
3. Number of blocks, b = 6
4. Number of new selections per block, n = v/b = 30/6 = 5
5. Number of plots per block, p = c + n = 3 + 5 = 8.
6. Total number of plots, N = bc + v = (3)(6)+30=48.

Suppose that the field plan, after randomization, and the grain yields (kg/ha) were as given in the following plan :

I		II		III	
Sel.	Yield	Sel.	Yield	Sel.	Yield
(14)	2405	CI	3023	(18)	2603
(26)	2855	(4)	3018	ST	2260
CI	2592	(15)	2477	(27)	2857
(17)	2572	(30)	2955	CI	2918
WA	2608	WA	2477	(25)	2825
(22)	2705	ST	3122	(5)	2065
(13)	2391	(24)	2783	WA	3107
ST	2972	(3)	3055	(28)	1903

IV		V		VI	
Sel.	Yield	Sel.	Yield	Sel.	Yield
(9)	2268	(2)	1055	(29)	2915
(6)	2148	(21)	1688	(7)	3265
CI	2940	ST	1315	CI	3483
WA	2850	WA	1625	(1)	3013
(20)	2670	CI	1398	WA	3400
(11)	3380	(10)	1293	(12)	2385
(23)	2770	(8)	1253	ST	3538
ST	3348	(16)	1495	(19)	3643

To begin the analysis a table of yields, totals, and means of the standard varieties (checks) is constructed. For this set of data we have the following :

BLOCK

Variety	I	II	III	IV	V	VI	Total	Mean
Stork	2972	3122	2260	3348	1315	3538	16555	2759.17
Cimmaron	2592	3023	2918	2940	1398	3483	16354	2725.67
Waha	2608	2477	3107	2850	1625	3400	16067	2677.83
Total	8172	8622	8285	9138	4338	10421	48976	2720.89
							M	8162.67

The block adjustment factor, r_j, is $r_j = (1/c)(B_j - M)$ in which, for this analysis, c = 3 and M = 8162.67. These factors for the six blocks in this trial are :

Block	I	II	III	IV	V	VI	
r_j	3.11	153.11	40.78	325.11	-1274.89	752.78	0.00

The observed yields, y_{ij}, of the new selections and the adjusted yields, $\hat{y}_i*$, are

$$*\hat{y}_i = y_{ij} - rj$$

Sel.	Block	OBS. y_{ij}	ADJ. $\hat{y}_i$	Sel.	Block	OBS. y_{ij}	ADJ. $\hat{y}_i$
1	6	3013	2260.2	16	5	1495	2769.9
2	5	1055	2329.9	17	1	2572	2568.9
3	2	3055	2901.9	18	3	2603	2562.2
4	2	3018	2864.9	19	6	3643	2890.2
5	3	2065	2024.2	20	4	2670	2344.9
6	4	2148	1822.9	21	5	1688	2962.9
7	6	3265	2512.2	22	1	2705	2701.9
8	5	1253	2527.9	23	4	2770	2444.9
9	4	2268	1942.9	24	2	2783	2629.9
10	5	1293	2567.9	25	3	2825	2784.2
11	4	3380	3054.9	26	1	2855	2851.9
12	6	2385	1632.2	27	3	2857	2816.2
13	1	2391	2387.9	28	3	1903	1862.2
14	1	2405	2401.9	29	6	2915	2162.2
15	2	2477	2323.9	30	2	2955	2801.9

An analysis of variance is now done on the yields of the standard (check) varieties. This gives an estimate of the experimental error for the trial. The ANOVA for this set of data is

ANOVA

Source	d.f.	SS	MS
Total	17	7,899,564	
Blocks	5	6,968,486	
Checks	2	20,051	
Error	10	911,027	91,103

$$cv = (\sqrt{MSE}/\bar{\bar{x}})\ 100 = (\sqrt{91,103}/2720.89)100 = 11.1\%$$

MSE = 91,103 is the sample estimate of the experimental error for this trial.

The variances of the different types of comparison are:

1. Difference between the means of two check varieties

   $s_c2 = 2MSE/b = (2)(91103)/6 = 30,368$

2. Difference between adjusted yields of two selections in the same block

   $s_b2 = 2MSE = (2)(91103) = 182,206$

3. Difference between adjusted yields of two selections in different blocks

   $s_v2 = 2MSE\ (c+1)/c = (2)(91103)(4)/3 = 242,941$

4. Difference between an adjusted selection yield and a check mean

   $s_{vc}2 = (b+1)(c+1)\ MSE/bc = (7)(4)(91103)/(6)(3) = 141,716$

5. Average variance of the difference between two adjusted selection yields

   $s_a2 = (2c+1)\ MSE/c = \left[(2)(3)+1 \right] (91103)/3 = 212,574$

It is possible to compute an L.S.D. for each of the preceding comparisons .. However, the most useful comparisons are the comparison of a new selection with the mean of a check variety and comparisons among the adjusted yields of the new selections. The 5% L.S.D.'s for these comparisons are computed as follows :

1. For both L.S.D.'s the required t_a is the 5% two-sided t with $(b-1)(c-1) = (6-1)(3-1) = 10$ d.f. This value is $t_{0.05(10)} = 2.228$
2. The 5% L.S.D. for comparing an adjusted selection yield with the mean yield of a check is $LSD_{0.05} = t_{0.05(10)} \sqrt{s_{vc}^2} = 2.228 \sqrt{141,716} = 838.7$
3. The 5% L.S.D. for comparing two adjusted selection yields may be computed using the average variance. This L.S.D. is $LSD_{.05} = t_{.05(10)} \quad s_a^2 = 2.228 \sqrt{212,574} = 1,027.2$

The results of this trial can be summarized for the cereal breeder in the form of a "Report of Results". An example of this type of report is given below.

REPORT OF RESULTS

A preliminary yield trial was conducted to evaluate thirty new selections of Durum for possible use in the A rainfall zone of northern Jordan. The new selections were compared against three standard varieties, Stork, Cimmaron, and Waha, using an augmented design with six blocks of eight plots each. The mean yields of the standard varieties and the yields, adjusted for block differences, of the new selections are presented in Table I.

Table I. Mean yields of standard varieties and adjusted yield of new selections (kg/ha) arranged in order of decreasing Yields.

Rank	Sel.	Yield	Rank	Sel.	Yield	Rank	Sel.	Yield
1	11	3054.9	12	Cimm	2725.7	23	13	2387.9
2	21	2962.9	13	22	2701.9	24	20	2344.9
3	3	2901.9	14	Waha	2677.8	25	2	2329.9
4	19	2890.2	15	24	2629.9	26	15	2323.9
5	4	2864.9	16	17	2568.9	27	1	2260.2
6	26	2851.9	17	10	2567.9	28	29	2162.2
7	27	2816.2	18	18	2562.2	29	5	2024.2
8	30	2801.9	19	8	2527.9	30	9	1942.9
9	25	2784.2	20	7	2512.2	31	28	1862.2
10	16	2769.9	21	23	2444.9	32	6	1822.9
11	Stork	2759.2	22	14	2401.9	33	12	1632.2

c.v. = 11.1% s^2 = 91,103

Although the adjusted yields of ten of the new selections are greater than the yield of the highest check, Stork, none of these yields is significantly (5%) higher. On the other hand two of the new selections, 6 and 12, have a yield which is significantly (5%) less than the yield of the lowest check, Waha. Among the new selections the top 25 form a group in which none of the adjusted yields are significantly different.

REFERENCES

Federer, W.T. 1956. Augmented (or Hoonuiaku) Designs. Hawiian Planters' Record, 55 :191-208.

Federer, W.T. 1961. Augmented Designs with one-way Elimination of Heterogeneity. Biometrics, 17 : 447-473.

Federer, W.T. and D. Ragavarao, 1975. On Augmented Designs. Biometrics, 31 : 29-35.