DESIGN OF EXPERIMENTS

by

JOHN WISHART

Institute of Statistics
Mimeo Series #15
For distribution to classes
Lectures of Spring 1949
Design of Experiments

Lecture I

Fundamental Principles

The object of any experiment, whether this is of a field or laboratory character, is to provide the research worker with an answer to the question which he has put to Nature. Since the answer generally comes out in numerical terms, we are brought up immediately against the analysis of numerical data, and since this is definitely more than just a simple exercise in arithmetic in view of the fact that the data resulting from such experiments as the research worker may make are subject to errors of one sort or another, the problem then acquires a statistical character, and it is then necessary to have some knowledge of elementary statistical analysis. The principles of the statistical method, therefore, should be regarded as part of the set of fundamental principles that are germane to the study of experimental design. But there is more to it than this. There was certainly a time when the aid of the statistician was only called on when the experiment was completed, and had resulted in a lot of numerical data which were not clearly analyzable by the non-mathematician. It is a fact that in 1919 R. A. Fisher was called on to take a post at Rothamsted Experimental Station in Harpenden, England, some 70 or 80 years after an important series of agronomic experiments had been laid down by Lawes and Gilbert. It had become known by this time that there was a theory of errors that could be used to elicit what was of value from a mass of numerical material, and so the aid of the mathematician was called on at this time to render more certain, it was hoped, the findings of the agronomist, or at any rate to separate the reasonably certain results from others which might be more dubious. Thus was the Statistical Department at Rothamsted born. But Fisher was not content to do merely as he was told. He recognized that the limitations of the original design might prevent certain facts from over clearly emerging, and so was led to formulate certain principles of design, together with a golden rule that the particular methods likely to be used in the end in the
statistical analysis of the resulting data should be kept in mind from the beginning and used to determine the most efficient plan for the experiment before it was started. The statistician is wanted, for example, before the experiment is started as well as after the practical part of it is completed, and he may even be useful while it is going on.

It will be the object of this course to help you to acquire a working technique for setting up experiments, for dealing with difficulties which may arise in their conduct, and for analyzing the resulting data. The course will not be mathematical in the sense of requiring the full justification of every step, and the formal proof of every proposition. On the other hand, we should recognize that we are dealing with concepts which are fundamentally mathematical in their nature, and with logical principles and steps which are by themselves quite a good training in the principles of the inductive method. It will be assumed that the student is familiar, at least in an elementary way, with the theory of statistical methods, but it is proposed to begin with a brief recapitulation of the methods normally used in working out tests of significance, chiefly in order to stress the simplifying nature of the assumptions made by the mathematician in reaching his beautiful results, and therefore to call attention to the fact that such assumptions must hold, at any rate approximately, if the standard results are to be applied without modification.

First, then, let us fix ideas by deciding to talk in the language of the agronomic investigator, both in order to study what our experimental material is, and also because it is in this field that the subject grew up and has made the greatest advances. In other fields the same basic principles apply. We shall, therefore, concern ourselves with the field experiment. This means that we attempt to test, under practical farming conditions, the effects of varying the varieties of the crops we are studying, or the systems of manuring, or of cultivation,
or perhaps the soil or the season. Before considering how this is done, let us consider the kind of measure that will come out of the experiment. This may be a yield measure, i.e. of grain or of "stern" (cereal crops) or of roots or of "tops" (root crops), or it may be a number determination, e.g. number of plants, or of tillers, or it may be a derived quantity, such as yield of dried produce in pasture, or of sugar from a beet crop, or we may be concerned with other derived quantities that may be obtained from a portion of the data by chemical analysis, such as the nitrogen percentage in barley, or the sugar percentage in beet. These are the quantities to consider in relation to the statistical assumptions. It is convenient to use \( x \) for such a quantity, whatever it is, and also convenient to use a single descriptive term for it, so if I speak of \( x \) as a "yield", you will remember that it may often be something else.

Now what have we learned from our study of statistical theory? Briefly this is as follows: the natural variability that is present in our experimental material is such that we do not go far wrong in assuming that a set of measurements \( x \) of the same underlying entity vary, due to observational errors, around some true value characteristic of that measure, and which we may call \( n \) to distinguish both from the single \( x \) and from the mean of a finite set of \( x \)'s. Further that such variation is normal in character, meaning that the Normal or Gaussian law operates, under which the distribution of frequency in the "population" is described by a curve centered at \( n \), and such that the logarithm of the ordinate of the curve at a distance \( \delta \) from \( n \) is less than that of the ordinate at the mean by a quantity proportional to the square of \( \delta \). If the assumption is exactly correct, the true population may be simply specified with the aid of two parameters only, one being the true value \( n \) of the variate \( x \) (the mean value of \( x \) or \( E(x) \)) and the other being the mean value of \( (x - n)^2 \), known as the variance and denoted by \( \sigma^2 \).

The first is a constant localizing the center of the distribution of the values that can occur on both sides of this center, due to errors, and the second is a measure
of the amount of dispersion that takes place round this center. If the assumption is only approximately correct, we may have to look at other measures such as that measuring asymmetry, for example, and also see what would be the effect of such measures not being zero on the various tests of significance.

Let us now consider the various elementary propositions of statistical theory of which we shall make use in analyzing experiments.

(1) Provided the finite series of measures \(x_1, x_2, \cdots, x_p\) constitute a random sample of data from a Normal population, then the statistic \(\bar{x}\) defined as

\[
\frac{x_1 + x_2 + \cdots + x_p}{p} \quad \text{or} \quad \frac{\sum (x)}{p}
\]

i.e. the arithmetic mean of the sample, is an unbiased estimate of \(m\) and varies round \(m\) in a normal distribution whose mean is \(m\) and whose variance is \(\sigma^2/p\) or whose standard derivation S.D. is \(\sqrt{\sigma^2/p}\). We have assumed that \(x_1 \cdots x_p\) are of equal "weight", otherwise weight adjustments are needed in calculating \(\bar{x}\). In giving the above result we are appealing to a thought process under which we regard \(\bar{x}\) as a single member of some population, call it \(x_1\), and then imagine a fresh sample of size \(p\) from the same population to yield another mean \(\bar{x}_2\) and so on. In the limit of this process we generate an infinite population of values of \(\bar{x}\), and it is the mean value of all these \(\bar{x}\)'s which can be shown to be equal to \(m \left[ E(\bar{x}) = m \right] \). This is what we mean by \(\bar{x}\) being an unbiased estimate. But \(E(\bar{x} - m)^2\) must depend on \(p\), and in fact works out as \(\sigma^2/p\), i.e. the variance is less than that of the population from which we started. This is intuitively sound, for in averaging \(p\) \(x\)s to produce a first \(\bar{x}\) we have obviously smoothed out some of the errors in the original population of measures.

We often express this proposition by saying that \(\bar{x}\) is \(N(m, \sigma / \sqrt{p})\), denoting that \(\bar{x}\) follows a normal distribution with mean and standard deviation shown in the bracket. Now it is always possible to reduce such a variate to
standard measure, as it is called. If we take the deviation of \( \bar{x} \) from the
mean \( \bar{m} \), and put this on a scale on which the S.D. is one unit of scale
(equivalent to dividing \( \bar{x} - \bar{m} \) by the S.D.) we get a variate which is \( N(0,1) \).
Calling this variate \( u \), another way of expressing the result (1) is to say
that
\[
\frac{\bar{x} - \bar{m}}{\bar{\sigma}/\sqrt{p}} \quad \text{is} \quad N(0,1) \quad \text{if} \quad \frac{X - \bar{m}}{\bar{\sigma}} \quad \text{is} \quad N(0,1)
\]
\( \bar{x} \) being \( \sum (x)/p \)

Deductions as to the value to be put on comparison of different values of \( u \)
are complicated by the fact that two unknowns are present, namely \( m \) and \( \bar{\sigma} \),
so without more ado we shall pass on to the more practicable proposition.

(2) If, in addition to calculating \( \sum (x)/p \) as an estimate of \( m \), we work out

\[
s^2 = \frac{\sum (x - \bar{x})^2}{p - 1}
\]
as an estimate of \( \sigma^2 \),

then

\[
t = \frac{\bar{x} - \bar{m}}{s/\sqrt{p}} \quad \text{with} \quad n = p - 1
\]
follows a different distribution law to that of \( u \) in (1), even when \( x - m/\bar{\sigma} \)
is \( N(0,1) \), due to the fact that \( \bar{\sigma} \) is replaced by an estimate \( s \) obtained
from the sample. It is, of course, the Student t-distribution and depends
upon a parameter \( n \) (the number of degrees of freedom). This number is in
the present case \( p - 1 \). The real value of this result lies in the fact that,
given \( n \) and the observational values which determine for us \( p, \bar{x} \) and \( s \), we
can calculate \( t \) and then determine from its distributional table the
probability of getting a departure from \( n \) as great as, or greater, than \( t \)
have in this sample, the sample being assumed to come from the given
population. Now if this probability is very small some one or another of the
assumptions will not fit the facts, and the normal inference is that our sample
then comes from a population whose mean is not \( \mu \), but something else. Take, for example, the plant breeder who knows from long experimentation the value \( \mu \) of the yield of a standard variety. He then sets out a number of plots, \( x_1 \ldots x_p \), of a new variety, bred, let us say, in the hopes of greater yield. If \( t \) is found to be significant on consulting tables, because it is clear that the probability \( p \) in the table is less than, say, 0.05 or 0.01, the inference is that the true yield of the new variety \( \mu' \), say, is different from \( \mu \). The plant breeder hopes that \( \mu' \) will be larger than \( \mu \), but it might be smaller, and so the two-sided test is usually made, i.e. the test picked up on using the table of \( t \) as normally printed, the original form being Fisher's.

A variant of this test occurs when we do not have an \( n \) with which to compare \( \bar{x} \). Everything else being known, including the probability point \( t_{0.05} \) (or \( t_{0.01} \)), a simple calculation gives values \( n_1 \) and \( n_2 \) equidistant from and on either side of \( \bar{x} \), and thus determining confidence limits within which the true \( \mu \) of the population, a sample from which gave the value \( \bar{x} \), may be expected to lie.

Experiments are usually, however, of the comparative type. The experimenter may not have a long run of trials on a standard variety, and in any case the particular characteristics of the season in which he tests his new variety may not be normal so that the standard, if tested at the same time, may come out higher or lower than the average. So it is good sound experimental practice to set up an experiment in which both the new and standard varieties are tested side by side. In general there may be no need to have the number of plots the same for each (later, we shall consider the most satisfactory forms of design). Let us suppose, then, that we have two independent samples, one of \( p \) values \( x_1 \ldots x_p \) with arithmetic mean \( \bar{x} \) and estimated variance \( \sum (x - \bar{x})^2/(p - 1) \) and the other of \( q \) values \( y_1 \ldots y_q \) with arithmetic mean \( \bar{y} \) and estimated variance \( \sum (y - \bar{y})^2/(q - 1) \), then if the two samples come from the same normal population \( N(\mu, \sigma^2) \) our third statistical result is
(3) \[ t = \frac{\bar{x} - \bar{y}}{s \sqrt{\frac{1}{p} + \frac{1}{q}}} \]

where \[ s^2 = \frac{\sum (x - \bar{x})^2 + \sum (y - \bar{y})^2}{p - 1 + q - 1} \]

follows the same t-distribution with D.F. \( n = p + q - 2 \). Everything here in \( t \) is now calculable from the two samples and if \( t \) turns out to be significant, we usually take it that the true means are not equal to one another, but must be some numbers \( n_1 \) and \( n_2 \) different from one another. If the \( x \)'s represent the new variety and the \( y \)'s the standard, and if \( t \) is positive, we could be held to have proved, as far as probability considerations and a single experiment enable us to prove anything, that the new variety had beaten the standard under the conditions in which the experiment had been carried out.

It is, of course, true that if the means are different, equal to \( n_1 \) and \( n_2 \), for example, and if the \( \sigma^2 \)'s can be assumed equal, then we still have a t-distribution if the numerator be written \( \bar{x} - \bar{y} - (n_1 - n_2) \). So the above expression could be used to calculate confidence limits for \( n_1 - n_2 \). If, however, the \( \sigma^2 \) are different, we run into difficulties of another kind, with which we shall not concern ourselves here.

Next, there is no need to confine ourselves to a comparison of two varieties only. If we want to compare a number of varieties with one another, and possibly also with a standard, it is obviously best to grow them all together under the same general conditions. Provided the samples are all independent, any one sample could be compared with any other by the foregoing test, but this is tedious and does not always lead to clear-cut results if the samples are small in size. A composite test is, however, possible. We want a test of whether the set of means, \( n_1, n_2, \ldots, n_k \), say, are the same or different, and we have knowledge of \( \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k \), together with the
variation that takes place round these means in the separate samples. This is where the analysis of variance first comes in as an obvious generalization of the t test for 2 samples. As in (3) we can have

\[ s^2 = \frac{\sum_1 + \sum_2 + \cdots + \sum_k}{n_1 + n_2 + \cdots + n_k} \]

Where the \( \sum \)'s are sums of squares of deviations from sample means for the k samples and \( n_1 \ldots n_k \) are the respective numbers of D.F. (number of observations in each sample minus 1). We can also obtain an estimate of variance from \( \bar{x}_1 \bar{x}_2 \ldots \bar{x}_k \). If the m's are all equal

\[ s_1^2 = \frac{\sum p(\bar{x} - \bar{x})^2}{(k - 1)} \]

will estimate the same \( \sigma^2 \) that the above \( s^2 \) does, where the sample sizes are \( p_1 \ldots p_k \), the means \( \bar{x}_1 \ldots \bar{x}_k \), and \( \bar{x} \) is the weighted mean \( p_1\bar{x}_1 + \cdots + p_k\bar{x}_k/(p_1 + \cdots + p_k) \). But if the m's are different the second estimate will be inflated on that account, because \( \bar{x}_1 \ldots \bar{x}_k \) will be estimating different means. So we need a test of whether two independent estimates of variance agree with one another, in which case they may be said to be estimating the same \( \sigma^2 \), or differ significantly, in which case they will be estimating different \( \sigma^2 \)'s. Usually this will come about through \( s_1^2 \) being a good deal larger than \( s^2 \), pointing to a possible explanation in that \( \bar{x}_1 \ldots \bar{x}_k \), or at any rate some of them, estimate different population means. Now this is a slightly more complicated form of a simple test for the significance of the difference between the estimates of variance \( s_1^2 \) and \( s_2^2 \) from two quite independent samples, e.g., as in case (3), and the test is that

\[ F = \frac{s_1^2}{s_2^2} \]

subject to the condition that \( s_1^2 \) and \( s_2^2 \) are independent estimates of a variance \( \sigma^2 \), with D.F. \( n_1 \) and \( n_2 \), has a known Variance-Ratio distribution, and the Snedecor or Fisher-Yates tables can be entered with the given \( n_1 \) and \( n_2 \) to answer the question whether the condition holds or whether \( s_1^2 \) and \( s_2^2 \) are
significantly different, in which case they are estimating different $\sigma^2$'s.

The simple form of this test gives us a supplementary test for case (3), because if we can show that $\sum (x - \bar{x})^2/p - 1$ and $\sum (y - \bar{y})^2/q - 1$ are not significantly different, we are strengthened in our conclusion that a significant $t$ points to the means $m_1$ and $m_2$ being different. In the more complicated form, the same test (4) answers the question of whether the set of $k$ means $\bar{X}_1 \ldots \bar{X}_k$ are significantly different or not, provided we can assume that $\sum (\bar{X}_1 - \bar{X})^2/p_1 - 1$, $\sum (\bar{X}_2 - \bar{X})^2/p_2 - 1$, etc. are not significantly different.

We can always test this point, for there is a homogeneity of variances test, due to Bartlett, which can be used. I shall not pause to describe this, because it can only be used for the case I have been describing. When we get, for example, to randomized blocks and other forms of design it is not normally possible to verify this assumption. Fortunately, for the kind of variations in treatment that usually arise in agricultural experiments, differences in means are not often accompanied by significant changes in variance. In other words there is often a useful stability of variance that justifies us in using the simple tests we have described in all our analyses.
Lecture 2

My last lecture was maybe a bit mathematical for some of you, but you need not worry if you do not know all the details of the tests of significance I described. What I was trying to do was to call attention to the two chief tests you should know about, namely the t-test for the difference between two means, and the F-test for the ratio of 2 variances, because these are basic to the analysis of data from field experiments. What I had particularly in mind was the assumed conditions under which these mathematical laws operate, namely that the primary data should be normal, in addition to which the two samples whose means are compared by the t test should be independent. The F-test, also, applies to two independent samples whose variance estimates are to be compared, but in analysis of variance problems we have another and more complicated case. Here we have any number of independent samples, and the two things we work out are (a) the estimated variance of the means of those samples ($s_1^2$) and (b) a suitable average of all the separate estimates of variance that are provided by the samples ($s^2$). Now if a sample comes from a normal distribution its mean is distributed independently of its variance. So you see that the normal assumption is necessary to show that $s_1^2$, which is made up of the means, is independent of $s^2$, which is made up of the separate variances. Under those conditions $s_1^2/s^2$ follows the F-distribution, and the table tells us whether $s_1^2$ is significantly greater than $s^2$ or not. If so it means that the set of means we have been comparing are not homogeneous, i.e., do not all estimate the same mean. We are therefore entitled to take it as proved that some of the samples have yielded higher than others. We can then go back to test (3) to compare individual means by the t-test.

So much by way of recapitulation of standard statistical results. I had more mathematics but it will be postponed. Let us now consider the experimental procedure, and see how far the conditions which apply to our statistical tests are met with in the experimental variable $x$, taking $x$ to be, say, the yield of an agricultural plot. This gives us the opportunity to establish the basic principles
of experimental design. The study began with uniformity trials. For example in England, Morcor and Hall (1911) obtained 500 separate measurements of the yield of a small area of wheat by planting the crop on a uniform 1 acre area at Rothamsted, dividing at harvest into 500 small portions. Wood and Stratton (1912) did a similar experiment at Cambridge. "Student" and Udny Yule were associated with this work as statisticians. Now the determination of the frequency distribution of the 500 wheat yields of the Morcor and Hall data is a favorite practical exercise in Cambridge. If we ignore the relation of these plot yields to one another on the field, and merely group in the normal way of forming a frequency distribution, it is possible to show that this fits very well to the normal curve.

<table>
<thead>
<tr>
<th>lb</th>
<th>2.7</th>
<th>2.9</th>
<th>3.1</th>
<th>3.3</th>
<th>3.5</th>
<th>3.7</th>
<th>3.9</th>
<th>4.1</th>
<th>4.3</th>
<th>4.5</th>
<th>4.7</th>
<th>4.9</th>
<th>5.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>freq.</td>
<td>4</td>
<td>15</td>
<td>20</td>
<td>47</td>
<td>63</td>
<td>78</td>
<td>88</td>
<td>69</td>
<td>59</td>
<td>35</td>
<td>10</td>
<td>8</td>
<td>4</td>
</tr>
</tbody>
</table>

There have been many similar cases, and it seems to be fairly well established that, at any rate so far as yield is concerned, normality (or at any rate approximate normality) is the rule. One of our basic conditions in applying our tests appears, therefore, to be satisfied. On the other hand, if we look more closely at the raw data, with the yields set out on their appropriate plots in a plan of the field, we see at once that neighboring plots are by no means independent of one another. There is, no doubt, a random element in the variation so far as some items are concerned, e.g. errors in weighing, losses due to weather and bird and insect depredations, etc., but the major feature appears to be a pronounced fertility drift, to a certain extent irregular, but sufficiently uniform to show patches of the field where the yields tend to be high together, and others where they are low together and so on. Contour maps of fertility have often been constructed which look not unlike the geographical contour maps of somewhat hilly country. There is a certain amount of correlation between the yields of adjacent plots in the sense that the neighbor of a high yielding plot tends on the whole
to be high rather than low yielding. At this stage we must remember our
independence condition for the measures constituting a statistical sample.
It is clear that in the present case we could not take, say, 10 adjacent
plots and count this lot of data as a statistical random sample. Its mean
would not give us an unbiased estimate of the true yield of a unit plot over
the entire area, and its variance would underestimate $\sigma^2$, if $\sigma^2$ is thought
of as the variance of a unit plot yield taken over the whole area of one acre,
(e.g. an estimate from the complete frequency distribution). But ignoring for
the time being the question whether 1/500 acre is a suitable unit to take for a
plot or not, it can easily be demonstrated that if we pick out $p$ plots by a
process of random selection, calling the yields $x_1x_2 \cdots x_p$, then we do secure

$$\bar{x} = \frac{1}{p} \sum (x)$$

an unbiased estimate of the true yield $\theta$, and one which has variance $\sigma^2/p$ (also $\sigma^2$ can be estimated by $\frac{\sum (x - \bar{x})^2}{(p - 1)}$).

Furthermore if we select a second sample of $q$ values, again by purely random
methods, we get in $\bar{y} = \frac{1}{q} \sum (y)$ a second unbiased estimate of $\theta$, with variance $\sigma^2/q$, and independent of the first. The $t$-test already described can then apply
to the difference between $\bar{x}$ and $\bar{y}$. In a uniformity trial such a test tells us
little, except that it enables us to verify that non-significant values of $t$ for
the most part occur (or that significant differences only turn up in the
expected proportion depending on the level of significance chosen), but this
procedure suggests that if we want to set up an experiment in which, say, a
variety $A$ is to be compared for yield performance with a variety $B$, we should at
the very least choose from the area a number of plots at random to assign to $A$,
and similarly a number (not necessarily the same number) at random to assign to $B$.

We have here been led quite naturally to invoke two important principles
governing experimental procedure. One is that already touched on, namely,
randomization, which ensures that certain conditions will hold under which the
standard statistical tests of significance can be made. The other is that of
replication. We do not want to have to set up a uniformity trial in every case prior to conducting a real experiment, and even if we did the essential conditions (e.g. season, etc.) would have altered before the real trial came on. Thus at the trial proper we are faced with the fact that the variance of yield for the treatment (or treatments) in question is unknown, and has to be estimated. This cannot be done from a single observation (for which D.F. = 0). With 2 we can have an estimate with 1 D.F., and estimates of greater precision can be got when the D.F. are higher, i.e. with, as we say, increased replication. It sounds a common-place today to say that repetitions of plots having the same treatment are necessary if any estimate at all of the error is to be calculated (in order, of course, that the reality or otherwise of the yield differences for different varieties or treatments may be demonstrated by statistical tests), but it is only about 25 years since such a principle had to be firmly stated by R.A. Fisher. Prior to this field experiments were usually of a very simple character, with at most duplication of plots of the same treatment in the same year. It is true that there was yearly repetition in the case of the Rothamsted long-range experiments, but it took a Fisher to point out that this was not satisfactory since it failed to eliminate bias in individual permanent plots.

Replication, then, is needed in order to provide us with an estimate of the error that our comparisons are likely to be subject to, the randomization coming in with this replication to ensure that the error estimates are valid, as well as to secure unbiased estimates of the treatment means. But there is a further use for replication. Individual plot errors have over and over again been noted as so high that there would be little chance of demonstrating the reality of the differences between single plots differently treated, even if the $\sum 2's$ were known. To some extent, of course, the plot error may be reduced by choosing the most suitable size and shape of plot commensurate with economical experimentation. The figures of the Mercer and Hall experiment, which have often been quoted,
illustrate this very well. Starting with the 1/500 acre plots, we can aggregate unit plots (sometimes in more than one way) to form larger areas. In each case we can work out the standard deviation (the precision decreasing as the number of observations decreases), and to compare the results, since the yield estimated gets larger with increasing aggregation, we can calculate the S.D. as a percentage of the mean yield (this is the statistician's coefficient of variation). Morcor and Hall gave the following figures

<table>
<thead>
<tr>
<th>No. of plots in block</th>
<th>area (acres)</th>
<th>S.D. as % of (C.V.) mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/500</td>
<td>11.7</td>
</tr>
<tr>
<td>2</td>
<td>1/250</td>
<td>10.0</td>
</tr>
<tr>
<td>4</td>
<td>1/125</td>
<td>8.9</td>
</tr>
<tr>
<td>({10})</td>
<td>1/50</td>
<td>({6.3}) Two different ways of aggregating</td>
</tr>
<tr>
<td>({10})</td>
<td>1/25</td>
<td>7.8</td>
</tr>
<tr>
<td>20</td>
<td>1/10</td>
<td>5.7</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>5.1</td>
</tr>
</tbody>
</table>

The C.V. may be observed to fall off, with increasing size of plot, but not nearly so quickly as \(1/\sqrt{p}\), if \(p\) is written for the number of plots in the block. This is, of course, because the plots are aggregates of neighboring units and not random samples. For random samples of size 4 (1/125 area plots) we would expect a S.D. of \(11.7/\sqrt{4} = 5.9\), but we get 8.9.

It was calculations of this kind which led to recommendations as to size of plot, i.e. that \(\frac{1}{40}\) acre or so also suitable for cereals (for larger plots the reduction of error was held not to compensate sufficiently for the extra trouble and expense of dealing with a larger area). There have also been studies on the best shape of plot, which have led to a generally accepted rule that long and fairly narrow plots are to be preferred to square plots of the same area, chiefly for the reason that their centers are closer together.

Let us suppose, therefore, that we have got our size and shape of plot fixed, with a very good idea that our error (S.D. of one plot yield) is of the order of say 8% (usual for cereals). Now we might be comparing treatments or varieties in which a small difference in yield of, say 5%, might represent a paying
proposition, and be worth advocating, if the reality of the difference could be demonstrated, i.e. if we could be reasonably sure (or statistically sure) that it would come out that way again on repetition. A significant difference (see table of t) is a little over twice the standard error of this difference with a reasonable number of D.F. (more if the D.F. are small). This means in the above case that the S.E. difference should not be more than about 2\(\frac{1}{2}\), or the S.E. for each treatment not more than say \(1\frac{1}{2}\) (2 + \(\frac{1}{2}\) \(\sqrt{2}\)). But the S.E. for each plot is 8, so that it is quite hopeless to expect to show up the reality of a 5% increase with single plots. But multiple plots have already been shown to be necessary for the calculation of the error. This will then mean that individual treatment plots will be replaced by treatment means (say of p plots each) in which case \(\bar{S}\) has been reduced to \(\bar{S} / \sqrt{p}\). In the above case about 28 replications (\(\sqrt{5.3}\)) would reduce an 8% S.E. to 1\(\frac{1}{2}\). This, of course, is far and away above what is usually practiced, but the illustration is given to show that replication is needed, not only to estimate the error, but also to reduce it (according to the \(\bar{S} / \sqrt{p}\) formula). It may also show that it is not only to please the statistician that experiments with much replication are advocated, but to serve the very practical end of demonstrating the validity of differences in which the practical agronomist is interested.

To complete our scheme there is one more principle of importance in improving the accuracy of field experiments. Before touching on this, let us see how with our present knowledge we could set up an experiment on the Mercer and Hall one acre field to test, say, 5 varieties. First we could select for each plot an area of 1/50 acre. This is just an aggregation of 10 of the original unit plots, and according to our table should give us 50 plot sites which, if treated uniformly, might be expected to give us 50 yields averaging about 40 lbs. (this is 10 times the mean of the quoted frequency distribution.) and having a S.D. of between 6 and 8% of the mean, i.e. something of the order of 2.4 to 3.2 lbs. Examination of
The actual yields of these 10 plot aggregates shows that the lowest value is 33.3 lbs. and the highest 43.4 lbs., so that the range is 10.1, between three and four times what the S.D. is expected to be. In fact the mean of the 50 plots comes out at 39.5 lbs. and the variance 6.26 lbs., so that the S.D. is $\sqrt{6.26} = 2.50$ lbs. (this was the lower of the two cases quoted in the table).

Now if we are to compare 5 varieties on such an area we can have 10 replications of each. So the game is to choose 10 plots by a process of random selection and allot to treatment A. Similarly with B, C, D, E. Each set of 10 can then be regarded as a reasonably unbiased sample of the whole area, so that, had the treatments all been the same the 5 means would have varied round 39.5 with S.D. $\sigma / \sqrt{10}$, which would be approximately $\sqrt{6.26} = 0.79$. A set of data worked out in this way gave a lowest mean of 38.79 and a highest 40.04, the range being 1.25. This S.D. of the 5 means furnishes us with a yardstick with which to measure the reality or otherwise of the differences between the 5 means had they belonged to different treatments. Roughly speaking we can take it that a difference of about three times the above 0.79 (say 2.3) in any two of the means would be an unlikely result to reach by chance, and would therefore be accepted as significant in the sense of demonstrating that the higher yielding of the two varieties being compared was definitely better than the other.

But it will be noted that in constructing this yardstick we have used as our estimate of the S.D. the figure calculated from the 50 plot yields under uniform treatment. In a concrete case in which five different varieties were used, we shall expect, if the varieties yield differently, that the 50 plot yields will yield a S.D. which is too high in its application to any one of the varieties. Our yard stick is no use to us if it is going to expand as the differences we measure get bigger, so we must modify our procedure. This is quite easy to do.

For each separate set of 10 plots refer to the same variety, and a calculation of the form $\frac{1}{9} \sum (x - \bar{x})^2$ for each will give an estimate of the variance unaffected by
variety differences. An average of the five possible estimates gives a mean
estimate \( s^2 \) for the variance of a single plot determination with \( 9 \times 5 = 45 \) d.f.
Then \( s^2/10 \) will be the estimated variance of the mean of 10 plots and so
\( s/\sqrt{10} \) will be our yardstick standard error.

Instead, however, of merely seeing roughly whether certain pairs of variety
means differ by more than three times this error we can get a more precise test,
on the basis of our (4), by working out

\[
s^2 = \frac{\sum (T - \bar{T})^2}{10 \times 4} = \frac{\sum(T - \bar{T})^2}{\sum T^2} - \frac{9}{5} \left( \frac{T^2}{1} - \frac{9 \bar{T}^2}{5} \right)
\]

as the estimated variance calculated from the variety means. The totals are

\( T_1 T_2 \cdots T_5 \), with mean \( \bar{T} \), and because a total of 10 items has a variance 10
times that of a single item, we must divide \( \sum (T - \bar{T})^2 \) by 10 to get the
"Sum of Squares" for varieties. Because there are 5 of these we divide by 4,
the number of D.F., to get the variance estimate.

Then \( F = s^2/\sigma^2 \) with \( f_1 = 4, f_2 = 45 \), and this test furnishes us with a
single measure of whether the 5 variety means are spread out more than we should
expect if they had been means of 5 random samples of the same variety. It is
generally recommended that this test be done first, because it can generally give a
fairly firm decision. Given a positive result we then look at the fairly obvious
pairs with the yardstick \( s/\sqrt{10} \).
Design of Experiments - Lecture 3

What we did at the end of the last lecture will be recognised as an analysis of variance appropriate to a single degree of classification. Set out in tabular form we have:

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>D.F.</th>
<th>Sum of squares</th>
<th>Mean square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between sample means</td>
<td>4</td>
<td>$\sum (T - \bar{T})^2/10$</td>
<td>$s_1^2$</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>$\sum_1 (x - \bar{x})^2$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>$\sum_2$</td>
<td></td>
</tr>
<tr>
<td>Within samples</td>
<td>945</td>
<td>$\sum_3 \ldots + \sum_5$</td>
<td>$s^2$</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>$\sum_4$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>$\sum_5$</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td>$\sum_1^{50} (x - \bar{x})^2$</td>
<td></td>
</tr>
</tbody>
</table>

Note that the total $\sum (x - \bar{x})^2$, which in the Mercer and Hall uniformity trial was the quantity 306.89, which on dividing by 49 gives the mean square or estimated variance 6.26, is equal to the sum of the two components which have been calculated in the way indicated. This is a necessary circumstance, and can be used to check the calculations. When the two sums of squares are divided by the D.F. (4 and 45) we get, in an example which has been worked out, $s_1^2 = 3.46$, $s^2 = 6.51$. $s_1^2$ is numerically less than $s^2$, and the two are obviously not significantly different, without troubling to use the F-test. This is, of course, because there are no treatment differences in the experiment; we merely compared 5 random samples out of the uniformity trial data. In a real experiment in which the 5 samples comprise different varieties or treatments $s_1^2$ will certainly contain a component of random variation which will tend to be of the same order as $s^2$, but if there are real differences in treatment yield it will also contain a component on this account, and will thus
in the aggregate be greater than $s^2$. With $f_1 = 4$, $f_2 = 45$ the significant
$F$ at the 5 percent point is just under 2.6, and we would therefore not hold
it as established that there were real treatment differences unless $F$ was 2.6
or greater. The reason for this is that in the absence of real differences
$s'^2$ is just as likely to be larger than $s^2$ as it is to be smaller (as in
the case we illustrated) and so we take $F$ at a point which makes it very unlikely
that the difference was due to chance.

Broadly speaking, the methods which have been described represent a
technique of experimentation which pays attention to the necessary principles,
and uses statistical methods to provide a test of homogeneity of the set of
sample means derived from the plots given over to the different treatments.
Such a technique, without further trimmings, can be used without more ado
in many laboratory experiments where the natural variability of the experimental
material is sufficiently under control to make further refinement unnecessary.

But I began by saying that I was going to concentrate on agricultural
field trials, which brings me to my third main principle. It has been already
stated that the errors in field trials can be high in relation to treatment
differences that represent a paying proposition to the practical man and
I gave an instance of how, with a fairly average plot standard error for a
cereal crop, something like 25 replications of the plots of the various
treatments would be needed to demonstrate the reality of a 5% difference in
yield. What we have been illustrating is the choosing of 5 lots of 10 each
out of the 50 Mercer and Hall plots. What $s^2$ estimated there was the variance
of one plot as judged by variation over the whole field. But if we look at
the 50 plots as set out in the field (Exercise 1) we see that they reflect
the same distribution of fertility that I referred to in talking of the 500 ultimate plot yields. That is, there are groups of plots which are above the average yield, and groups below the average. Now it is good experimental practice to have the treatments we are to compare as like to one another as possible in conditions other than the factor on which we require information. Such a factor, for example, might be a varietal one. The adoption of this practice makes it easier for us to accept a yield difference as being due to the factor which has been varied as between the groups, since other things, broadly speaking, have been equalised.

Suppose, then, we divide up the 50 plots into 10 groups or blocks of 5 plots each, choosing the block this size because we are interested in comparing 5 varieties. Over any single block the soil variation will certainly be much less than over the area of the field as a whole, and thus within the block we have in general sufficiently uniform conditions to render possible a fairly precise comparison between the 5 varieties, if we assign one to each plot of the block. Such assignment must, of course, be done at random to preserve one of our main general principles. If repeated on all the other 9 blocks, we end up, as before, with 5 samples, each of 10 plots, but this time one plot (for each variety) has been chosen from each block. In one sense, of course, this implies a restriction on the completely random arrangement previously described, but so long as we take due account of what we have done in the statistical analysis of the data we are all right. By balancing the blocks, i.e. by seeing to it that all varieties are represented once in each block, we have in fact removed from the errors of our comparisons that part of the variability over the area that is due to the differences between block totals. We do not need to have different treatments assigned within the blocks to see what has happened.
We do not even need to assign dummy letters, A, B, C, D, E as we did before. A legitimate first decomposition of the total sum of squares would be into the two parts (a) between blocks (b) within blocks, as the following analysis of variance shows for the data of Exercise 1:

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>D.F.</th>
<th>S.S.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between blocks</td>
<td>9</td>
<td>190.53</td>
<td>21.17</td>
</tr>
<tr>
<td>Within blocks</td>
<td>40</td>
<td>116.36</td>
<td>2.91</td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td>306.89</td>
<td>6.26</td>
</tr>
</tbody>
</table>

An F-test on the ratio 21.17/2.91, with 9 and 40 D.F., shows that a very significant amount of the variability has been eliminated by keeping our treatment comparisons within blocks. In fact, with this type of design, our variance per plot has been reduced from 6.26 to 2.91, so that the accuracy of the experiment has been more than doubled i.e. less than half the number of replications will be required to demonstrate the significance of the same difference. I gave a very rough calculation earlier on the number of replications that would be required to demonstrate a difference between the mean yields of two treatments of 5%. Let us repeat this on the present data:

Mean yield of the 1/50 acre plots = 40
5% of this = 2

(a) S.E. per plot on complete randomization = \( \sqrt{6.26} = 2.50 \)
S.E. for mean of n plots = \( 2.50/\sqrt{n} \)

Significant difference between 2 means = \( \frac{2.50}{\sqrt{n}} \times \sqrt{2} \times t_{49} \)
= \( \frac{7.5}{\sqrt{n}} \) approximately

Equate this to 2, when we have \( n = 3.75^2 = 14 \)

(b) S.E. per plot within blocks = \( \sqrt{2.91} = 1.71 \)

Other things being the same, except for a slight change from \( t_{49} \)
to $t_{40}$, for a significant difference of $2$.

We require $n^2 = 14 \times 2.91/0.26 = 6.5$

Now it so happens that we were able to fit in 10 replications of our 5 treatments on to the 50 plots available. With the first arrangement i.e. complete randomization, this is now seen to be not enough to enable us to detect a difference of $5\%$; but with the second it is actually more than enough, and we see, therefore, that we have a very precise type of design by arranging the area in blocks.

We are here invoking a further principle, that designated by Fisher "local control". Its effect can be seen at its easiest if we have only 2 treatments to compare, for if we mark off our plots in pairs and assign the treatments A and B at random to the plots of the pair, we have only to work out the difference $d = A - B$ for each pair and average the d's to give us a measure of the comparison we wish to make. The error of this comparison is reduced because, broadly, the individual d's will be unaffected by the levels of block fertility, i.e. by the quantities $A + B$. A possible bias due to fertility drift has already been disposed of by assigning the plots of the pair at random.

With more than 2 treatments per block, the same considerations apply, because the differences $A - B$, $A - C$, etc. that we are looking for are made up of individual differences of this sort between the plots of a single block, and these are relatively little affected by the level of block fertility. What I am describing is, of course, the method of randomized blocks, as to the details of which we shall return. At the present time what we can say is that the fertility variations over the area of a field can at any rate be approximated to by a series of table-lands, each extending over the area of a block, and because in examining differences between treatments we are only concerned in averaging the differences that occur from the mean of the
particular table-land representing the block, and have no concern with the broad change between the fertility of one block and the next, we are bound to have a more precise type of trial by introducing the block concept on top of the ordinary randomized arrangement of the plots of the different treatments. The limitation to the efficiency of this arrangement is the size of the block. With too many treatments being compared in the same experiment, the block becomes rather large, and the table-land concept of the block fertility becomes less and less an approximation to the facts. The Latin square design, to which we shall come, is also an illustration of the principle of local control, and, within its limitations as regards the number of treatments which can be accommodated, can be even more efficient than the method of randomized blocks.

It will be convenient at this stage to summarize what we have learned of the principles of good experimental design with the aid of the diagram which Fisher used to have hanging up on his wall at Rothamsted.

\[ \text{I} \quad \text{Replication} \]
\[ \text{II} \quad \text{Random Distribution} \]
\[ \text{III} \quad \text{Local Control} \]
\[ \text{Validity of Estimate of Error} \]
\[ \text{Reduction of error} \]

Replication is at the base of the structure. On the one hand it provides directly an estimate of the experimental error and, indirectly through random distribution, ensures the validity of that estimate; on the other hand it leads to a reduction of the error, directly by operation of the formulae $\sigma / \sqrt{p}$, and indirectly through the operation of a system of local control, i.e., by using methods such as randomized blocks or the Latin square.
The process of randomization.

The safest way to randomize the treatments in an experiment is to make use of a table of random numbers. This consists of a table of, usually, 2 figure numbers, which may be combined, if desired, to give 3 figure, 4 figures, etc. numbers, and which has been constructed by choosing the digits by a random process designed to avoid bias. Let me give three simple instances.

(1) Suppose you are asked to assign 10 of the MacCor and Hall plots to treatment A, 10 to B etc. ... 10 to E on a complete randomization basis. Starting at any chosen position in the random number table, and reading systematically in any direction, forward, backward, up or down, we get a series of numbers as shown below. Read off the remainder on mentally dividing this number by 5, counting the remainder as 5 if it is a multiple of 5. Then assign A, B, C, D, E to the remainders 1, 2, 3, 4, 5. By the time you have 10 of any one letter, ignore this remainder in the subsequent search through the table. (In a really complete randomization we would accept the (variable) numbers of A's, B's, etc. provided by the random number table, and the experiment would be a valid one, but it is usual, and accords with common sense, to have equal replication for all treatments.)

Random number: 17 26 39 70 88 19 6 40 69 17 29 47 20 50 67 19 97 74 34 86
Remainder: 2 1 4 5 3 4 1 5 4 2 4 2 5 5 2 4 2 4 4 1
Treatment: B A D E C D A E D B D B E E B D B D D A etc.

(2) If you are asked to divide the 50 plots into 10 blocks of 5 and assign the treatments A -- E at random within the blocks (randomized blocks design) then you follow the same procedure, but, having accepted the first appearance of a letter in a block, you should ignore the occurrence of the same remainder so long as you are dealing with that block. By the time you have 4 remainders you are finished with that block, because the fifth must follow. Thus:
Random number 53 74 23 99 67 61 32 28 69 84 94 62 67 86 24 98 33 41 19 95 47 etc.
Remainder 3 4 - - 2 1 | 2 3 4 - - - 1 | 4 3 - 1 - 5 | 2
Treatments C D B A E | B C D A E | D C A E B | B etc.

(3) As a third case, suppose we wish to test 6 treatments (A — F) in
8 replications on any area. We should make up 8 compact blocks, each to
contain 6 plots. To give us proper randomization we should ignore the
occurrence of the numbers 97, 98, 99, 00 (which is regarded as 100) in the
random number table, because up to 96 all remainders are dividing by 6 will have
occurred an equal number of times, but not beyond this. So we shall have the
following (e.g.)

Random number 98 83 71 94 22 59 97 50 99 52 08 52 85 08 40 87 80 61 65 etc.
Remainder x 5 - 4 - - x 2 x - - - 1 - - 3 | 2 1 5
Treatments E D B A C F | B A E etc.
1. A fully randomized experiment.

Let there be \( r \) plots of each of \( t \) treatments. Let \( x \) denote the individual plot yield (\( tr \) values altogether) and \( T \) the treatment total of \( r \) plots (\( t \) values altogether). Then, treated as a single set of \( tr \) yields, with grand total \( G = \sum \limits _{1}^{t} (T) \) and mean \( \bar{x} = G/(tr) \), we have for the sum of squares of deviations about the mean:

\[
\text{Total S.S.} = \sum (x - \bar{x})^2 = \sum \limits _{1}^{t} (x^2) + \sum \limits _{2}^{t} (x^2) + \cdots + \sum \limits _{t}^{t} (x^2) - G^2/tr
\]

which may be written, by putting in and taking out again quantities of the form \( T^2/r \):

\[
\text{Total S.S.} = \frac{1}{r} \left\{ \frac{t^2}{r} + T_2^2 + \cdots + T_t^2 - G^2/t \right\}
\]

\[
+ \sum \limits _{1}^{t} (x^2) + \sum \limits _{2}^{t} (x^2) + \cdots + \sum \limits _{t}^{t} (x^2)
\]

\[
- \frac{T_1^2}{r} - \frac{T_2^2}{r} - \cdots - \frac{T_t^2}{r}
\]

\[= T + W, \text{ say}\]

where \( T \) is the whole of the first line, easily seen to be the sum of squares of the deviations of the \( t \) treatment totals from the mean total \( \bar{T} = G/t \), divided by \( r \). \( T \) is the S.S. between the treatments, and has \( e_1 = t - 1 \) D.F. because derived from a sample of \( t \) observations. (By writing \( T_1 = r\bar{x}_1 \), where \( \bar{x}_1 \) is the mean of the first treatment, and so on for the others, we can see that \( T \) may alternatively be written as \( r \) times the S.S. of deviations of the \( t \) treatment means from the general mean \( \bar{x} \).)

\( W \) is obviously a sum of \( t \) parts, the first part, \( \sum (x^2) - \frac{T_1^2}{r} \), for example, being the S.S. of deviations of the \( r \) plot values from their mean \( \bar{x}_1 \), and thus having \( r - 1 \) D.F. The whole of \( W \) may be expressed as the S.S. within the
treatments, and altogether it will have \( e_2 = \sum (r - 1) = t(r - 1) \) D.F.

The two parts \( T \) and \( W \) are distributed independently of one another on the assumption of normality of the primary data, and the analysis of variance table takes the form:

<table>
<thead>
<tr>
<th>Variation</th>
<th>D.F.</th>
<th>S.S.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between treatments</td>
<td>( e_1 = t - 1 )</td>
<td>( T )</td>
<td>( T/e_1 )</td>
</tr>
<tr>
<td>Within treatments</td>
<td>( e_2 = t(r - 1) )</td>
<td>( W )</td>
<td>( W/e_2 )</td>
</tr>
<tr>
<td>Total</td>
<td>( tr - 1 )</td>
<td>( T + W )</td>
<td></td>
</tr>
</tbody>
</table>

Then \( F \) is the ratio of \( T/e_1 \) to \( W/e_2 \), and we enter the \( F \)-table with \( e_1 \) and \( e_2 \) D.F. to see what relation \( F \) bears to the values tabled for 5% and 1% significance. If larger than one or other of these, we declare that there are significant differences between the treatment totals (or means). To examine these, we set them out together with their standard error (S.E.)

Now \( s^2 = \frac{W}{e_2} \) is the estimated variance of a single plot yield. Thus \( rs^2 \) is the estimated variance of a treatment total or \( s^2/r \) is the estimated variance of a treatment mean.

The table then takes either of the two forms below, the second of which consists simply of the numbers in the first divided by \( r \).

(1) treatment 1 2 3 \( \cdots \) \( t \) Mean S.E.
   totals \( T_1 \) \( T_2 \) \( T_3 \) \( \cdots \) \( T_t \) \( \bar{T} \) \( s/\sqrt{r} \)

(2) treatment 1 2 3 \( \cdots \) \( t \) Mean S.E.
   means \( \bar{x}_1 \) \( \bar{x}_2 \) \( \bar{x}_3 \) \( \cdots \) \( \bar{x}_t \) \( \bar{x} \) \( s/\sqrt{r} \)

Broadly, we look for differences between two treatment totals (or means) greater than three times the S.E. given in the table. More exactly, the S.E. of the difference between any two \( T \)'s, \( T_1 - T_2 \), for example, is \( s/\sqrt{2r} \), and so \( T_1 - T_2/2s/\sqrt{2r} \) has to be greater than the value given in Students' \( t \)-table.
for $e_2$ D.F. and the level of significance adopted. To test the difference of means by the formula \( \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{2}{r}}} \) is, of course, the same thing. It is useful to construct a least significant difference (l.s.d.) to use as a yardstick for rapid picking out of the significant differences. This yardstick will be obtained for the t totals $T_1$, $T_2$, \ldots, $T_6$, by multiplying the S.E. $s \sqrt{\frac{r}{r}}$ by \( \sqrt{2} \) multiplied by the Student value from the table of t appropriate to $e_2$ D.F.

For example, if $e_2 = 20$, this value is 2.09 for the 5% point. If, then, the treatment totals are arranged in numerical order it is easily possible to identify the total which significantly exceeds the first tabulated, then the one which is significantly greater than the second, etc. Usually the criterion of three times the S.E. of the T's is good enough. This kind of analysis should be used with common sense, because we must recognize that we are dealing with statistical variables, and even in the complete absence of real effects it must be recognized that the T's will range over several times the S.E., depending on the number of T's.

The first two parts of Exercise 1 illustrate calculations of the kind just given, and the second in particular will give practice in picking out significant differences. A complete analysis of an experiment of this kind will contain (a) a plan of the experimental layout; (b) the individual plot yields, which can be conveniently given on the plan itself, or alternatively be tabulated together with a key to the plots of the plan; (c) the analysis of variance; (d) a summary table of total or mean yields together with their standard error; and (e) a verbal statement of the significant conclusions which have emerged. In connection with (d) it is best to supplement the table with the yields expressed in the common agricultural units, e.g. bushels or cut. per acre, because this conveys more to the agronomist, and also enables him to assess the comparative results of different trials. Such figures are, of course, obtained by calculating a single quantity by which all the total or mean yields must be multiplied.
Method of Randomized Blocks

The basic principles of this design have already been described. Having decided on the right size of plot, we require to construct a compact block large enough to contain all the treatments under test. Replication then consists in having a number of such blocks, and in each block the order of the treatments is decided by use of random numbers as previously described. Suppose there are \( t \) treatments and \( r \) replicates. Then, as we shall shortly see, there will be

\[ (t - 1)(r - 1) \text{ D.F.} \]

for the estimation of error variance, and we should see to it that, not only is the whole area under experiment chosen to be as uniform as possible, but also that the number of blocks is such as to give enough D.F. to enable the error to be determined with some approach to precision. Thus the error D.F. should not be less than 10 or 12. With only 2 treatments, this requires about 12 blocks as a minimum; with 4 treatments 5 replications at least will be needed. In theory, when the number of treatments rises to 12 or more, 2 blocks should be enough, but it has been my experience that there is too much scope here for something to go wrong. We have already seen that with a large block there is a greater chance of unevennesses in fertility tending to bias some of the treatments. It is not enough to have a sufficient number of D.F. to estimate the error variance with some approach to precision. We must remember also that our conclusions will rest on the differences that emerge between the estimated means of the populations representing the various treatments. Although those means will all be unbiased, there will be a greater tendency to large departures from the true means with small samples than with large ones. It is sometimes forgotten that a single sample of 2 cannot estimate a mean with much precision, for the S.E. of the sample mean will be estimated as \( s/\sqrt{2} \). The D.F. for error determine the precision with which \( s \) estimates \( \sigma \), but it is the amount of replication which determines the size of \( s/\sqrt{r} \). Experiments with only 2 or 3 blocks are not, therefore, as a rule to be strongly recommended. On the other hand
4 replications can usually be depended upon to be reasonably satisfactory. While, therefore, we should determine \( r \), knowing \( t \), so that \((r - 1)(t - 1)\) is of the order of 10 or 12 at least, we would do well to put a minimum value of 4 upon \( r \), and aim at securing four-fold replication wherever we can. The only conditions under which this procedure can be modified are where two or more interacting series of treatments are included in one and the same experiment, and we can be reasonably sure that "interaction" is without much physical significance.

Now for the statistical analysis. The yield or other data, normally written down on a plan of the experiment so that each number is associated with a treatment within a block, are conveniently transferred to a row and column arrangement on paper, where rows stand for treatments and columns for blocks (or vice versa). We then get an arrangement like the one shown below:

<table>
<thead>
<tr>
<th>Blocks</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>( \cdots )</th>
<th>( r )</th>
<th>Treatment Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( x )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( T_1 )</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( T_2 )</td>
</tr>
<tr>
<td>Treatments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( T_3 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( \cdots )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( \cdots )</td>
</tr>
<tr>
<td>( t )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( T_t )</td>
</tr>
<tr>
<td>Block totals</td>
<td>( B_1 )</td>
<td>( B_2 )</td>
<td>( B_3 )</td>
<td>( \cdots )</td>
<td>( B_r )</td>
<td>( G )</td>
</tr>
</tbody>
</table>

Where we want to identify an individual plot we shall call it \( x \). The totalling by rows has produced \( t \) treatment totals, \( T_1, T_2, \ldots, T_t \), similarly the column totals are \( B_1, B_2, \ldots, B_t \), while finally the grand total, denoted by \( G \), is the sum of either all the \( B_i \)'s or all the \( T_i \)'s (both should be summed to provide a check).
Let the correction factor $G^2/rt$ be denoted by $c$. Then we have first to calculate a total $SS$
\[ S = \sum (x^2) - c \]
where the summation is over all $rt$ plot values. The D.F. are $rt - 1$. We then calculate a blocks component
\[ B = \sum (B^2)/t - c \]
with D.F. $r - 1$, the summation being over the $r$ block totals. Secondly we calculate a treatments component
\[ T = \sum (T^2)/r - c \]
with D.F. $t - 1$, the summation being over the $t$ treatment totals. $B$ and $T$ are part of $S$, and the remainder, namely $S - B - T$, is called the Error $SS$, and may be denoted by $E$. It has the D.F.
\[ rt - 1 - (r - 1) - (t - 1) = (r - 1)(t - 1) \]
It will be seen that $B + E = W$ of the completely randomized design. Because of the arrangement in blocks $B$ may be separately calculated, and may in many cases form a substantial part of $W$, thus reducing considerably the error estimate of variance. The theory behind what we are doing is that we suppose the yield of any plot to be compounded of three parts, one depending on the mean fertility level of the block, one depending upon the particular treatment assigned to the plot, and a residual component, positive or negative, measuring the random error, and assumed to be normally distributed about zero with a certain variance, estimated by the Error M.S. The Analysis of Variance table is as follows:

<table>
<thead>
<tr>
<th>Variation</th>
<th>D.F.</th>
<th>S.S.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>blocks</td>
<td>$r - 1$</td>
<td>$E$</td>
<td>$E/(r - 1)$</td>
</tr>
<tr>
<td>treatments</td>
<td>$t - 1$</td>
<td>$T$</td>
<td>$T/(t - 1)$</td>
</tr>
<tr>
<td>error</td>
<td>$(r - 1)(t - 1)$</td>
<td>$E$</td>
<td>$E/{ (r - 1)(t - 1) } = s^2$</td>
</tr>
<tr>
<td>Total</td>
<td>$rt - 1$</td>
<td>$S$</td>
<td></td>
</tr>
</tbody>
</table>
A measure of our success in local control will be whether
\[ s^2 = E \left\{ (r - 1)(t - 1) \right\} \] is considerably less than \((B + E)/t(r - 1)\) or not, which can be measured by taking
\[ F = \frac{B}{r - 1} \frac{E}{(r - 1)(t - 1)} \]
with \(r - 1\) and \((r - 1)(t - 1)\) D.F. To see whether significant differences have emerged between the treatments we test
\[ F = \frac{T}{t - 1} \frac{s^2}{r} \]
with \(t - 1\) and \((r - 1)(t - 1)\) D.F. If significant, we take it as established that some treatments have yielded better than others, and can draw up the summary table:

<table>
<thead>
<tr>
<th>Treatments</th>
<th>1</th>
<th>2</th>
<th>...</th>
<th>t</th>
<th>Mean</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>total</td>
<td>(T_1)</td>
<td>(T_2)</td>
<td>...</td>
<td>(T_t)</td>
<td>(\bar{T})</td>
<td>(s \sqrt{r})</td>
</tr>
</tbody>
</table>

and use \(3s \sqrt{r}\) as a least significant difference. Even if the treatments \(F\) is not significant it may emerge on further examination of the treatments S.S. that part of the effect is significant. We shall return to the methods of examining this question.

**Method of the Latin Square**

The main difficulty here is to determine the appropriate design consistent both with the randomization principle and the particular conditions. First we must have \(r\), the number of replications, equal to \(t\), the number of treatments, and we shall choose to express this number by the letter \(r\). The plots should be arranged in an \(r \times r\) array, i.e. having \(r\) rows and \(r\) columns, and we want the treatments assigned so that any particular treatment, \(T_1\), for example, shall occur only once in each row, and only once in each column, of the array. The individual plot may have any suitable size and dimensions, in which case the whole area will have the same comparative dimensions, and will not normally be exactly square.
To illustrate the choice of a random Latin square, let us take the simple case of $4 \times 4$, while recognizing that, according to our rules, this is normally too small because, as we shall see, it only gives us $6$ D.F. for error. Restricting ourselves at first to the narrower range of squares which all have the letters ABCD in that order along the first row and down the first column, we can see that there are four of these:

\[
\begin{array}{cccc}
ABCD & ABCD & ABCD & ABCD \\
BADC & BDCD & BDAC & BADC \\
CDAB & CDAB & CDBA & CDAB \\
DCBA & DABC & DCBA & DCBA \\
1 & 2 & 3 & 4
\end{array}
\]

Numbers 1-3 belong to a single transformation set, while number 4 belongs to a different set. By a process of random selection choose one of these numbers, say 2. Then randomize all rows other than the first, obtaining, say

\[
\begin{array}{cccc}
ABCD & DABC & CDAB & BCD4 \\
\end{array}
\]

Finally assign the four treatments $T_1, T_2, T_3$ and $T_4$ at random to the letters $A, B, C, D$, when we might get

\[
\begin{align*}
T_2 & T_3 T_1 T_4 \\
T_4 & T_2 T_1 T_3 \\
T_1 & T_4 T_3 T_1 \\
T_3 & T_1 T_4 T_2
\end{align*}
\]

as the field arrangement. This one has been chosen out of the $24 \times 6 \times 4 = 576$ different possible Latin squares of this size. With larger squares it is only necessary to write down one typical square from each transformation set, so long as we know the number of squares occurring in each set. By a random choice we then pick on a set specimen, giving due weight to the numbers in each set. We should then re-arrange all rows and all columns at random, and finally assign the treatments at random to the letters of the square. For the specimen sets, and for further information respecting the randomization procedure, see "Statistical Tables,"
by Fisher and Yates, pp. 15 and 60.

The field design is now determined, and a plan should be prepared to identify the treatments to be imposed on the plots, and on which to record the numerical results, e.g., yield measures. This is already a row-column array, and on the same paper it is convenient to total rows and columns, recording these as below:

\[
\begin{array}{c|cccc|c}
\text{Rows} & 1 & 2 & 3 & \cdots & r & \text{Row Totals} \\
\hline
1 & x & & & & R_1 \\
2 & & & & & R_2 \\
3 & & & \cdots & & R_3 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\text{r} & & & & & R_r \\
\text{Column Totals} & C_1 & C_2 & C_3 & \cdots & C_r & G \\
\end{array}
\]

We must then pick out the replicates of each treatment, and record separately the treatment totals, as below:

\[
\text{Treatment totals:} \quad | \begin{array}{cccc|c}
T_1 & T_2 & T_3 & \cdots & T_r & G \\
\end{array}
\]

As a check it should be ascertained that all the R's, all the C's, and all the T's, separately add up to G, the grand total.
The statistical analysis then follows. The correction factor is 
\[ c = \frac{G^2}{r^2}. \]
The total S.S. is 
\[ S = \sum (x^2) - c, \]
with \( r^2 - 1 \) D.F., where the summation is over all \( r^2 \) plot values. We then go on to calculate the row S.S.
\[ R = \sum \frac{(r^2)}{r - c}; \]
the column S.S.
\[ C = \sum \frac{(C^2)}{r - c}; \]
and the treatments S.S.
\[ T = \sum \frac{(T^2)}{r - c}; \]
all these have \( r - 1 \) D.F., and the summations are over the \( r \) row, column or treatment totals respectively. \( R, C \) and \( T \) are all part of \( S \), and the Error S.S. is
\[ E = S - R - C - T \]
having
\[ r^2 - 1 - 3(r - 1) = (r - 1)(r - 2) \text{ D.F.} \]
This time it is \( R + C + E \) which is equal to \( W \) of the completely randomized design, and it is because of the special arrangement in rows and columns according to the Latin Square pattern that we expect to get \( W \) very considerably reduced by elimination of the \( R \) and \( C \) components, and may thus expect a very much reduced error estimate of variance. The theoretical assumption is that the yield of any plot is made up of four parts, three of them depending upon row, column and treatment, while the fourth is the residual component measuring the random error.

The Analysis of Variance Table is as follows.

<table>
<thead>
<tr>
<th>Variation</th>
<th>D.F.</th>
<th>S.S.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>rows</td>
<td>( r - 1 )</td>
<td>( R )</td>
<td>( R/(r - 1) )</td>
</tr>
<tr>
<td>columns</td>
<td>( r - 1 )</td>
<td>( C )</td>
<td>( C/(r - 1) )</td>
</tr>
<tr>
<td>treatments</td>
<td>( r - 1 )</td>
<td>( T )</td>
<td>( T/(r - 1) )</td>
</tr>
<tr>
<td>error</td>
<td>( (r - 1)(r - 2) )</td>
<td>( E )</td>
<td>( E/\frac{1}{2}(r - 1)(r - 2)s^2 )</td>
</tr>
<tr>
<td>Total</td>
<td>( r^2 - 1 )</td>
<td>( S )</td>
<td></td>
</tr>
</tbody>
</table>

An examination of the F composed of either \( R/(r - 1)s^2 \) or \( C/(r - 1)s^2 \) with \( (r - 1) \) and \( (r - 1)(r - 2) \) D.F. will indicate the degree of our success in local control, while for treatments we have \( F = T/(r - 1)s^2 \), with the same D.F. The summary table, and the conclusions, are as for the method of randomized blocks.
The two methods which have been described provide the main groundwork for the field experiment. Other methods to be described are elaborations on this main basis. As we have seen, randomized blocks are the more flexible of the two, for they permit a judicious amount of replication to be carried out, an amount which does not increase beyond the normal resources of the experimenter as his number of treatments increases. The limit is the size of the block, which should not be too large. With the Latin Square, the size $5 \times 5$ is about the lowest size to give reasonable precision to the error determination, for with this size $\text{D.F.} = 4 \times 3 = 12$. There is also an upper limit of size, and it is only rarely that we should wish to go beyond about $8 \times 8$. This is not only because of the amount of replication needed. The assumptions may tend to be invalidated for long rows and columns. Let us illustrate a case which might conceivably happen, and in which the Latin Square would be an unfortunate choice. Suppose that the distribution of fertility is of the character shown in the diagram, proceeding uniformly from high to low, and low to high.

\[
\begin{array}{ccc|c|c}
\text{High} & & \text{Low} \\
\hline
\text{Low} & & \text{High} \\
\end{array}
\]

The field is imagined as divided by the broken lines into 9 areas, those not marked high or low being average. It is clear that the row totals will be all average, and likewise for columns, so that no reduction of error M.S. will come about through removal of row or column differences. On the other hand, had the area been divided into the four blocks shown by the unbroken lines, and the arrangement been one in randomized blocks, it is obvious that the blocks component would have removed a considerable part of the heterogeneity, for two of them will be above average and two below.
So far we have been content to express our treatments S.S. as a single quantity, measuring the effect by comparison with error, and saying that individual comparisons of treatment means can be made with the aid of the appropriate S.E. calculated from the error M.S., when the F-test of the analysis of variance has given a significant result. But it is sometimes a distinct advantage to break up the treatments S.S. into component parts. For the present we shall consider that only a single factor is involved, leaving it for later consideration how to break up the S.S. when multiple factors come in, as when, for example, we might wish to test various amounts of nitrogenous and phosphatic fertilizers, in all combinations.

The broad rule here is to do the break-up according to the imposed treatments, and not merely because some of the totals are high, and others low. Such break-up is not often necessary if the F-test has given a significant result, because then the use of the S.E. as a yardstick is usually enough. But cases do occur where the F is not significant, and yet reason suggests that something can be learned from the experiment. For example we might be comparing a number of varieties, bred for high yielding, but have included a control treatment consisting of a standard variety. Or we may be testing the effect of nitrogen on a crop, applying this in equivalent amounts, in different chemical forms, and may include a control plot with no nitrogen. In these two cases it may be observed that all the new varieties, or all the nitrogenous treatments, give somewhat higher yields than control, but do not differ much among themselves. With, say, 6 treatments in all, the treatment M.S. for 5 D.F. may not be large enough to show up as significant, yet the single D.F. component of this for the control against all the others together may by itself be significant. Even if F for treatments as a whole is significant, we shall still normally be interested in whether the new varieties, or the different forms of nitrogenous treatments, differ significantly in yield.
Let the treatment totals be $T_1$, $T_2$, \ldots, $T_t$, $T_1$ being the yield of the "control" plot. The treatments S.S., with $r$ replicates, is

$$\frac{1}{r} \left[ \sum_{i=2}^{t} T_i^2 + T_2^2 + \cdots + T_t^2 - \frac{G^2}{t} \right]$$

and the part within brackets may be written as the sum of two parts:

(a) Control v. others $$T_1^2 + \frac{(G - T_1)^2}{t - 1} - \frac{G^2}{t}$$

(b) Within others $$\sum_{i=2}^{t} T_i^2 - \frac{(G - T_1)^2}{t - 1}$$

It will be noticed that $(G - T_1)^2/(t - 1)$ in the first line has become the correction factor in the second line, where only $t - 1$ of the treatments are dealt with. This analysis is a special case of that in which we divide a total S.S. into parts "between group means" and "within groups". Line (a) has 1 D.F. and line (b) the remaining $t - 2$. When both are divided by $r$, and then by the D.F., we have the basis for two values of $F$, on dividing by the error M.S. $S^2$. If most of the variation consists of the difference in yield between the control treatment and the mean of the others, then the 1 D.F. effect may be significant even if the $F$ for the whole of treatments is not significant.

This procedure does not only apply to the comparison of one treatment with the mean of the others. The treatments could, in fact, be divided up into any number of groups, irregular as regards their size, if this corresponded with the facts of the treatment allocation.

Another case is that where, for example, the treatments might consist of 0, 1, 2 etc. dressings of a fertilizer, and where in the result there appeared to be a moderate rise in yield, but not enough for treatments M.S. as a whole to be significant. Here we may appeal to regression theory, regarding the yield as the dependent variable $y$ and the dressings as the independent variable $x$. In
general the dressings might not be equally spaced, as in our example. Regression theory tells us that the fitted straight line to the set of y's has regression coefficient

\[ b = \frac{\sum y(x - \bar{x})}{\sum (x - \bar{x})^2} \]

and that the component of the S.S. \( \sum (y - \bar{y})^2 \) due to regression is

\[ \frac{\left[ \sum y(x - \bar{x}) \right]^2}{\sum (x - \bar{x})^2} \]

with 1 D.F. The remaining D.F. go with \( \sum (y - \bar{y})^2 \), the S.S. of the deviations of the y's round the fitted line. If this break-up be applied to the treatment totals \( T_1, T_2, \text{etc.} \) of r plots each, then we divide both components by r. It may then happen that the 1 D.F. effect, or the linear component, which tests for the reality of a slope in the fitted regression line, i.e. tests the hypothesis that \( \beta = 0 \), where \( \beta \) is the true regression coefficient, \( b \) being the sample estimate, gives a significant F on comparison with the error M.S., in which case we may be held to have demonstrated that the fertilizer treatment has been instrumental in improving the yield, even although the increases from zero dose to 1 dose, from 1 to 2, etc. are too small to be individually significant.
The calculations become especially simple if the doses proceed by equal steps. Take, for example, the case of 5 treatment totals. We may write the x's as -2, -1, 0, 1 and 2, with mean zero and S.S. 10.

\[
\begin{array}{ccc}
T_1 & T_2 & T_3 & T_4 & T_5 \\
-2 & -1 & 0 & 1 & 2 \\
\end{array}
\]

Then the 1 D.F. component due to regression will be

\[
\frac{(2(T_5 - T_1) + (T_4 - T_2))^2}{10}
\]

which is divided by r, the number of replicates, for comparison with the error M.S. = $s^2$.

The desire of the agronomist in such cases is, by varying the size of the dose, to determine its optimum value. He may therefore deliberately go too far, and produce a condition of ill balance, resulting in the yield ceasing to rise, and even falling. In such a case the curve of yield will not be linear, and we may be interested in seeing whether it may be significantly demonstrated as non-linear, or in determining where it reaches its highest point. With the doses proceeding in equal steps this may be easily done by the orthogonal polynomial procedure. (In other cases, i.e. when the observations are unequally weighted and possibly unequally spaced, the problem can also be solved, but the computations become lengthy.) To illustrate such a case, let us take 5 yields of barley grain, in cwt. per acre, reported from a Woburn rotation experiment in 1931, where 0, 1, 2, 3 and 4 dressings of phosphate were given on top of standard dressings of nitrogen and potash. The linear, quadratic etc. coefficients are taken from Fisher and Yates' Statistical Tables: Table XXIII, $n^1 = 5$. 
With 5 points to fit we can go up to the quartic and fit perfectly. Alternatively we can stop at any point we like. The divisor shown on the right is the sum of the squares of the coefficients shown on the same line. Now the total of the 5 numbers is 122.2, and so the S.S., for 4 D.F., is \(20.9^2 + \ldots + 24.8^2 - 122.2^2/5 = 18.372\). Let us see how this is decomposed. The sum of products (S.P.) of yield and coefficient, for the linear term, is 7.2 (i.e. \(20.9(-2) + 26.4(-1) + \ldots\)). The S.S. for this component, for 1 D.F., is then 7.2^2/10 = 5.184. For the quadratic component the S.P. is -9.4 and the S.S. due to this -9.4^2/14 = 6.311. For the cubic component the two numbers are 5.1 and 2.601, and for the quartic component -17.3 and 4.276. We then have the following decomposition of the treatments S.S.:
We see that in this process we have, starting with the mean 24.44, reproduced the original 5 yields. Normally, we should not trouble to go as far as this with only a few yields; it is evident, for example, from the quadratic that non-linearity, may exist, although it has not yet been tested for, and we can even determine, either from a graph or from the equation to the fitted quadratic, that the maximum is reached at about 2½ doses. To go further in fitting is merely to twist the curve to accommodate itself to the errors of the individual determinations.

This is an example of working out orthogonal components, which essentially consists in breaking up a sum of squares into independent parts with single degrees of freedom. It is based on the following simple algebraic considerations:

If we have \( r \) values of \( x \), namely \( x_1, x_2, \ldots, x_r \), and write

\[
X = a_1 x_1 + a_2 x_2 + \ldots + a_r x_r
\]

then if the \( x \)'s all come from the same normal population, with mean \( m \) and S.D. \( \sigma \), and the \( a \)'s are constant coefficients, it is not difficult to see that the mean of \( X \) is

\[
(a_1 + a_2 + \ldots + a_r) m
\]

and that the variance of \( X \) is

\[
(a_1^2 + a_2^2 + \ldots + a_r^2) \sigma^2.
\]

<table>
<thead>
<tr>
<th>Line 1 x</th>
<th>2</th>
<th>-1</th>
<th>-2</th>
<th>-1</th>
<th>2</th>
<th>(-9.4 + 14 = -0.67)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(-1.34)</td>
<td>+0.67</td>
<td>+1.34</td>
<td>+0.67</td>
<td>(-1.34)</td>
<td></td>
</tr>
<tr>
<td>Quadratic</td>
<td>21.66</td>
<td>24.39</td>
<td>25.78</td>
<td>25.83</td>
<td>24.54</td>
<td></td>
</tr>
<tr>
<td>Line 1 x</td>
<td>-1</td>
<td>2</td>
<td>0</td>
<td>-2</td>
<td>1</td>
<td>(5.1 + 10 = 0.51)</td>
</tr>
<tr>
<td></td>
<td>(-0.51)</td>
<td>1.02</td>
<td>0</td>
<td>(-1.02)</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>Cubic</td>
<td>21.15</td>
<td>25.41</td>
<td>25.78</td>
<td>24.81</td>
<td>25.05</td>
<td></td>
</tr>
<tr>
<td>Line 1 x</td>
<td>1</td>
<td>-4</td>
<td>6</td>
<td>-4</td>
<td>1</td>
<td>(-17.3 + 70 = -0.247)</td>
</tr>
<tr>
<td></td>
<td>(-0.25)</td>
<td>+0.99</td>
<td>-1.48</td>
<td>+0.99</td>
<td>-0.25</td>
<td></td>
</tr>
<tr>
<td>Quartic</td>
<td>20.9</td>
<td>26.4</td>
<td>24.3</td>
<td>25.8</td>
<td>24.8</td>
<td></td>
</tr>
</tbody>
</table>
Incidentally, for the special case of all the a's equal to 1 this gives us the previously quoted result that a sample total T has variance $r \sigma^2$. In another case, if we put all the a's equal to $1/r$, we see that the sample mean $\bar{x}$ has variance $\sigma^2/r$.

At present we are interested in a case where the sum of the squares of the a's is equal to unity. In such a case the variance of $X$ is $\sigma^2$, i.e., is equal to the variance of the component x's. If we have another linear function

$$Y = b_1 x_1 + b_2 x_2 + \ldots + b_r x_r$$

then, not only is the variance of $Y \sigma^2$ if the sum of the squares of the b's is unity, but also $X$ and $Y$ will be uncorrelated if the sum of products

$$a_1 b_1 + a_2 b_2 + \ldots + a_r b_r$$

is zero. This means that the two functions are statistically independent if the x's come from a normal population, as is usually assumed. We speak of $X$ and $Y$ as orthogonal functions. Further functions of this character can be constructed, for example

$$Z = c_1 x_1 + c_2 x_2 + \ldots + c_r x_r$$

under the conditions $c_1^2 + c_2^2 + \ldots + c_r^2 = 1$, $a_1 c_1 + a_2 c_2 + \ldots + a_r c_r = 0$, $b_1 c_1 + b_2 c_2 + \ldots + b_r c_r = 0$. Writing $X$, $Y$ and $Z$ as $X_1$, $X_2$ and $X_3$, we may proceed with constructing linear functions of this character until we get to $X_r$, when we have the result that

$$X_1^2 + X_2^2 + \ldots + X_r^2 = X_1^2 + x_2^2 + \ldots + x_r^2.$$  

The last example was a case of this kind. First we had

$$20.9^2 + 26.4^2 + 24.3^2 + 25.8^2 + 24.8^2 = 3004.94.$$  

The correction factor $122.2^2/5 = 2986.568$ is itself the square of a function $X_1$ orthogonal to the other four, as can easily be verified. The others are $X_2$, $X_3$, $X_4$ and $X_5$, and therefore

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 = 2986.568 + 18.372$$  

$$= 3004.94.$$
equal to the sum of the squares of the 5 original x's.

To show the orthogonality property, note that \( X_1 \) is a linear function of the 5 x's, with coefficients

\[
\begin{align*}
1 & \quad 1 & \quad 1 & \quad 1 & \quad 1 \\
\text{divided by} & \quad \sqrt{5}.
\end{align*}
\]

On the other hand \( X_2 \) has the coefficients

\[
\begin{align*}
-2 & \quad -1 & \quad 0 & \quad 1 & \quad 2 \\
\text{divided by} & \quad \sqrt{10}.
\end{align*}
\]

In both cases the sum of squares of coefficients is unity, when we take into account the divisor. Also the sum of products of coefficients is

\[
\begin{align*}
-2 & \quad -1 & \quad +0 & \quad +1 & \quad +2 = 0 \quad \text{(divided by \( \sqrt{50} \)).}
\end{align*}
\]

Similarly \( X_3 \) has the coefficients

\[
\begin{align*}
2 & \quad -1 & \quad -2 & \quad -1 & \quad 2 \\
\text{divided by} & \quad \sqrt{14}
\end{align*}
\]

and we not only see that the sum of squares here is unity, but also the sum of products with both the coefficients of \( X_1 \) and of \( X_2 \) is zero. And so on to the end.

Certain special cases of the above procedure are of interest in experimental design. The simplest of all occurs when there are only 2 observations \( x_1 \) and \( x_2 \).

Take

\[
\begin{align*}
X_1 & = \frac{1}{\sqrt{2}}(x_1 + x_2) \\
X_2 & = \frac{1}{\sqrt{2}}(x_1 - x_2).
\end{align*}
\]

These are orthogonal functions, and therefore

\[
\begin{align*}
x_1^2 + x_2^2 & = X_1^2 + X_2^2 \\
& = \frac{1}{2}(x_1 + x_2)^2 + \frac{1}{2}(x_1 - x_2)^2
\end{align*}
\]

which can be verified by simple algebra. But \( \frac{1}{2}(x_1 + x_2)^2 \) is the correction factor for a sample of 2. Subtracting from both sides we therefore have for the S.S. in a sample of 2 the very simple expression

\[
\frac{1}{2}(x_1 - x_2)^2
\]

with 1 D.F. Given only 2 observations, all we have to do is to take half the square of their difference to get the S.S. Incidentally we have by this means been able to
express a sum of squares of deviations of 2 quantities from their common mean as a single square, corresponding to the single degree of freedom available.

In a field experiment we sometimes have only two treatments. If the yield totals for these are $T_1$ and $T_2$, and the number of replicates is $r$ the S.S. for treatments, from the above, is easily seen to be calculated as

$$\frac{(T_1 - T_2)^2}{2r}$$

that is, as the square of the difference between the treatment totals, divided by the total number of plots.

For 3 observations $x_1$, $x_2$ and $x_3$, a useful formula for the S.S. is

$$\frac{1}{3}[(x_1 - x_2)^2 + (x_2 - x_3)^2 + (x_3 - x_1)^2]$$

that is, it is the mean of the squares of the three differences that can be worked out from the observations. But note that this is not an example of an orthogonal break up, for it consists of 3 squares, and yet has only 2 D.F. The coefficients of $x_1$, $x_2$ and $x_3$ in the three parts which are squared do not satisfy the orthogonality conditions. For orthogonality we should have to take different coefficients. One way of doing this is as follows:

<table>
<thead>
<tr>
<th></th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>Divisor</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>$\sqrt{3}$</td>
</tr>
<tr>
<td>b</td>
<td>1</td>
<td>-1</td>
<td></td>
<td>$\sqrt{2}$</td>
</tr>
<tr>
<td>c</td>
<td>1</td>
<td>1</td>
<td>-2</td>
<td>$\sqrt{5}$</td>
</tr>
</tbody>
</table>

then $x_1^2 + x_2^2 + x_3^2 = \frac{1}{3}(x_1 + x_2 + x_3)^2 + \frac{1}{2}(x_1 - x_2)^2 + \frac{1}{6}(x_1 + x_2 - 2x_3)^2$.

Taking the correction factor over to the other side we therefore have the S.S. for a sample of 3 equal to

$$\frac{1}{2}(x_1 - x_2)^2 + \frac{1}{6}(x_1 + x_2 - 2x_3)^2$$

with 2 D.F. Here again we have been able to show the S.S. as the sum of a number
of independent squares equal to the number of D.F. The first term is the same as that obtained from a sample of 2; the second comes in with the occurrence of a third observation. It should be noted that this way of expressing the S.S. is not unique, but it is often useful practically, for computation purposes. Note also that if in the above table we read $x_3$, $x_1$, $x_2$ for $x_1$, $x_2$, $x_3$ the use of the same a, b, c coefficients, in the new order, gives us the S.S. expressed in terms of the linear and quadratic components with the x's corresponding to evenly graduated amounts, say, of a fertilizer dressing.

If this idea of orthogonal split can be grasped, it will explain why a treatments S.S. can be broken up in more than one way. Going back to our two examples, let us consider just three treatments. In the first we had a control plot together with two others, considered, for the sake of argument, as being two forms of nitrogen in equivalent amounts. What we did was to separate off the 1 D.F. S.S. for control vs. others, leaving the remainder to express the variation within the others. With only 3 treatments these parts have 1 D.F. each, and correspond to the orthogonal functions

\[
\begin{array}{cccc}
  x_1 & x_2 & x_3 & \text{Divisor} \\
  2 & -1 & -1 & \sqrt{6} \\
  0 & -1 & +1 & \sqrt{2}
\end{array}
\]

the S.S. being

\[
\frac{(2x_1 - x_2 - x_3)^2}{6} \quad \text{and} \quad \frac{(x_3 - x_2)^2}{2}.
\]

Added together these give $\sum(x - \bar{x})^2$ for the sample of 3. (For simplicity we have left out the first orthogonal function which determines the mean.)

In the second case we had graded amounts of fertilizer added, starting with a control plot. Here we separated off a 1 D.F. S.S. for linear component, and were able to go on calculating quadratic, cubic etc. components. With 3 treatments there
will only be a single 1 D.F. component left after calculating the linear component, and this will be the quadratic, the orthogonal functions being

\[
\begin{array}{cccc}
-1 & 0 & 1 & \sqrt{2} \\
1 & -2 & 1 & \sqrt{6}
\end{array}
\]

and the S.S. are

\[
\frac{(x_3 - x_1)^2}{2} \quad \text{and} \quad \frac{(x_1 - 2x_2 + x_3)^2}{6}.
\]

The particular problem will usually suggest what is the appropriate split to make. If there is any doubt as to whether the different parts of the S.S. are additive and independent the check should be made to see whether the multipliers satisfy the orthogonality conditions.
In the case of 4 treatments we can illustrate another variation, thus furnishing an introduction to the idea of introducing two factors into an experiment. Suppose we wish to test the effect of adding nitrogen (N) in some form and also potash (K). Choosing certain units of dressing we get a good symmetrical arrangement by having a control plot, one having N, one having K, and one having both, according to the scheme:

<table>
<thead>
<tr>
<th></th>
<th>No N</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>No K</td>
<td>$x_1$</td>
<td>$x_2$</td>
</tr>
<tr>
<td>K</td>
<td>$x_3$</td>
<td>$x_4$</td>
</tr>
</tbody>
</table>

This scheme shows that there are two levels of application of N, namely none and some, and also two levels of application of K. In all combinations these provide for $(2 \times 2)$ or $(2)^2 = 4$ treatment-combinations in all. An experiment may therefore be laid out in randomized blocks, or in a Latin square, in which each block (or row) contains 4 plots, to which the 4 treatment-combinations are assigned at random. The analysis of the experiment proceeds in the usual way until we get to the treatments S.S. with 3 D.F. This may now be decomposed to show separately the effects of N and K, and it is here that the orthogonal split comes in. The above table shows yields $x_1$, $x_2$, $x_3$ and $x_4$ for the four combinations as if these were single plots. In an actual experiment, of course, they would be replaced by $T_1$, $T_2$, $T_3$ and $T_4$, the totals of r replicates. What we are concerned with now is the decomposition of the treatments S.S.

$$\sum (x - \bar{x})^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2 - \frac{1}{4}(x_1 + x_2 + x_3 + x_4)^2.$$  

Looking at the table we see that the left-hand column comprises two values $x_1$ and $x_3$ which are both without N, while in the right-hand column we have $x_2$ and $x_4$, both with N. A comparison of $x_1 + x_3$ with $x_2 + x_4$ will therefore indicate whether there has been a response to N on the average of the no K and K plots. It is true that we
could simply compare $x_1$ and $x_2$ for the effect of $N$ in the absence of $K$, and $x_3$ and $x_4$ for the effect of $N$ in the presence of $K$, but the other comparison will have greater precision, being based on double the number of plots, and will convey the same information, provided there is no vital change in the response to $N$ as between no $K$ and $K$ (or no vital change in the response to $K$ as between no $N$ and $N$ - the two things are equivalent).

Now let us see what an orthogonal split would give:

<table>
<thead>
<tr>
<th>Treatments</th>
<th>00</th>
<th>NO</th>
<th>OK</th>
<th>NK</th>
<th>Divisor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values</td>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_3$</td>
<td>$x_4$</td>
<td></td>
</tr>
<tr>
<td>$x_1$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>$\sqrt{4}$</td>
</tr>
<tr>
<td>$x_2$</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>$\sqrt{4}$</td>
</tr>
<tr>
<td>$x_3$</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>$\sqrt{4}$</td>
</tr>
<tr>
<td>$x_4$</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>$\sqrt{4}$</td>
</tr>
</tbody>
</table>

The coefficients are exceedingly simple, being either +1 or -1, and you may readily verify that the orthogonality conditions are satisfied.

We then have

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2.$$

Where

$$X_1^2 = \frac{1}{2}(x_1 + x_2 + x_3 + x_4)^2$$

$$X_2^2 = \frac{1}{2}\left\{(x_2 + x_4) - (x_1 + x_3)\right\}^2$$

$$X_3^2 = \frac{1}{2}\left\{(x_3 + x_4) - (x_1 + x_2)\right\}^2$$

$$X_4^2 = \frac{1}{2}\left\{(x_1 + x_4) - (x_2 + x_3)\right\}^2$$

We, therefore, have the treatments S.S. $\Sigma(x - \bar{x})^2$, with 3 d.f., which is got by subtracting $X_1^2$ from both sides, equal to the sum of 3 components, with 1 d.f. each. Remembering our earlier formula for a S.S. in a sample of 2 as a single square, we see that $X_2^2$ is the S.S. for the column component of the table (effect of $N$),
while $X^2_3$ is the S.S. for the row component (effect of K). In fact the situation is like any row-column lay-out, in which the total S.S. is divided into row and column components and a residual. In the present case the residual is $X^2_4$, and will thus measure the deviations from an additive law assumed for the N and K effects, i.e. the assumption that the effect of N is the same in the presence of and absence of K, and vice versa. It will be noticed that $X^2_4$ is calculated from the difference of the diagonal totals of the plots. Note that if $x_1 \ldots x_4$ are totals of r replicates, the quantities $X^2_2$, $X^2_3$, and $X^2_4$ should be divided by r.

The three squares, $X^2_2$, $X^2_3$, and $X^2_4$ are independent, and have 1 D.F. each. We have therefore decomposed the treatments S.S. into those 3 parts in such a way that separate F-tests can be made by taking the ratio of each in turn with the error M.S. Instead, therefore, of having a single F to say whether treatments as a whole are significant, we have separate tests to say whether

(a) N has given a significant effect on average of no K and K.

(b) K has given a significant effect on average of no N and N.

(c) N and K operate independently, or whether there is "interaction" between them.

The term "interaction" has here been introduced, and may be simply defined in the present instance. Let us consider a case where the yields are as follows:

<table>
<thead>
<tr>
<th></th>
<th>No N</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>No K</td>
<td>100</td>
<td>110</td>
</tr>
<tr>
<td>K</td>
<td>120</td>
<td>150</td>
</tr>
</tbody>
</table>

We see that there has been a gain of 10 due to N in the absence of K, and a gain of 20 due to K in the absence of N. With a strictly additive law operating we should expect 130 for the plots having both N and K. To find that the result is 150 implies a departure from the additive law, in the direction of the addition to yield being greater than the sum of the parts due to N and K separately. Sometimes we find the
opposite effect, namely that the NK yield comes out to be less than 130. In both cases we should say that N and K interacted significantly if the $x_4^2$ component was significant. Looking at the form of $x_4^2$ we see that the difference of the cross-sums in the above table is $250 - 230 = 20$, whereas had the NK plot been 130 the difference would be zero. A difference significantly above or below zero would be evidence of interaction, and would thus furnish information in regard to the effect of the fertilizers that would not be gained had two separate experiments been laid out, testing for N and K separately. We see, therefore, the advantage of introducing the factors to be tested into one and the same experiment. Not only do we get this added information (which is just as useful if interaction is not significant as when it is), but we can generally economize on number of plots. For example, a $(2^2)$ experiment on the factors N and K could be run with 5 replications and give $3 \times 4 = 12$ D.F. for error, whereas single experiments to test N and K separately would require 10-12 replications each unless we were prepared to assume that a single error term could be calculated by pooling the interaction terms from the two separate experiments. Even so we should want at least 6 replications in each.

It should be clear without much further explanation how such an experiment as has been suggested above can be analyzed. Give the 4 treatments 00, NO, OK and NK, in, say, r randomized blocks, a first analysis on the ordinary pattern will give component S.S. for blocks (r - 1 D.F.), treatments (3 D.F.) and error (3(r - 1) D.F.). The treatments S.S. will come from the four treatment totals $T_1$, $T_2$, $T_3$ and $T_4$ (the subscripts corresponding to those of the $x$'s in the above table) by the calculation

$$\frac{1}{r}(T_1^2 + T_2^2 + T_3^2 + T_4^2) - \frac{G^2}{4r}$$

This is then broken up into three components, shown below together with the relevant S.S.

$$N \left\{ (T_2 + T_4) - (T_1 + T_3) \right\}^2/4r$$

$$K \left\{ (T_3 + T_4) - (T_1 + T_2) \right\}^2/4r$$
Since each component has 1 D.F. the above quantities also represent the M.S. for the component parts, and testing for significance consists in determining the F for the ratio of each to the Error M.S. \( s^2 \), the D.F. being 1 and 3\((r - 1)\).

The S.E. appropriate to the treatment totals \( T_1, T_2, T_3 \) and \( T_4 \) may be placed with these totals in a summary table, but there is no need for further examination of differences with the aid of their S.E.'s. The reason for this is that an F test, with 1 and \( n \) D.F. is equivalent to a t test for \( n \) D.F.; in fact \( F = t^2 \) in this special case. To illustrate this let us choose one of the comparisons, say \( N \) vs. No \( N \). We have

\[
\begin{array}{ccc}
\text{Total of No } N & \text{Total of } N & \text{Difference} \\
T_1 + T_3 & T_2 + T_4 & d = (T_2 + T_4) - (T_1 + T_3)
\end{array}
\]

The variance of any \( T \) is \( rs^2 \), and therefore of the sum of 2 \( T \)'s is \( 2rs^2 \). It follows that the variance of \( d \), which is the difference of two sums, each of 2 \( T \)'s, must be \( 4rs^2 \). Thus we shall have

\[
t = \frac{d}{s \sqrt{4r}} \quad \text{with} \quad \text{D.F.} = 3(r - 1).
\]

But F for this comparison is \( d^2/(4rs^2) \), and is thus seen to be exactly \( t^2 \).

The calculation has been explained in this way because most of you will be familiar with the formula for the variance of the sum or difference of two values, the variance for each of which is \( \sigma^2 \). But we have extended our theoretical knowledge a little beyond this point now. Recalling our formula for the variance of a linear function of \( x \)'s, we see that in the formula

\[d = -T_1 + T_2 - T_3 + T_4\]

the coefficients are either -1 or +1, and the sum of squares is 4. Thus the estimated variance of \( d \) must be 4 times that of any \( T \), i.e. it must be \( 4rs^2 \).
To compare 2r of the plots against the other 2r for the effect of N, i.e. to use all the data, is more efficient than to compare N with no N separately in the presence and absence of K. The same applies to the other effects, and as long as we look at them all, we see that we have got all that can be abstracted from the data, with the maximum precision.

Factorial Designs. 2 Factors.

The case we have been dealing with is the simplest case of introducing two factors into an experiment, and we have seen the value of having all combinations present, i.e. of having an NK plot as well as the other three. In general, it should be stated that if we are interested in testing for 2 factors, which might be anything we please, e.g. two fertilizer treatments, or a combination of varieties with fertilizer dressings, or dressings with methods of cultivation, etc. we get more information, and at less cost, with a combined experiment of the factorial design than by laying down experiments to test the factors separately.

We should now consider more complex cases. The next in order, for example, would be one in which one factor was present at 2 levels and the other at t levels. We might, for example, apply a number of fertilizer dressings to each of 2 varieties of a crop. We would then have a (2 x t) design, the total number of treatments to go into a block being 2t. The treatment totals, for all blocks, may be arranged in a (2 x t) table, as below, in which we denote the one factor by the letters A and B, and the other by the numerals 1, 2, 3, ... t.

<table>
<thead>
<tr>
<th>Treatments</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>...</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Varieties</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>T</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>S_A</td>
</tr>
<tr>
<td>B</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>S_B</td>
</tr>
<tr>
<td>Sum</td>
<td>S_1</td>
<td>S_2</td>
<td>S_3</td>
<td>...</td>
<td>S_t</td>
</tr>
<tr>
<td>Diff.</td>
<td>D_1</td>
<td>D_2</td>
<td>D_3</td>
<td>...</td>
<td>D_t</td>
</tr>
</tbody>
</table>

For any block, the total treatment is given by the sum of the sums of the separate rows (A and B) and the difference between the totals for the two groups is given by the difference between the sums of the rows.
In this special case it is useful to form the sums for A and B, calling them $S_1, S_2$, etc. and also the differences $A - B$ (or $B - A$), calling these $D_1, D_2$, etc. We also have the row sums $S_A$ and $S_B$, the grand total $S = S_A + S_B$ and the total difference $S_A - S_B$.

Ignoring the division by $r$, the number of replicates, which will apply right through, we may now analyze the treatments $S.S.$, which is the sum of the squares of the $2t$ values of $T$, less the correction factor $S^2/2t$, into three components:

\[
\begin{align*}
\text{D.F.} & \\
\text{(a) treatment with $t$ levels} & \frac{1}{2}(S_1^2 + S_2^2 + \ldots + S_t^2) - S^2/2t & t - 1 \\
\text{(b) treatment with 2 levels} & D^2/2t & 1 \\
\text{(c) interaction} & \frac{1}{2}(D_1^2 + D_2^2 + \ldots + D_t^2) - D^2/2t & t - 1
\end{align*}
\]

It is easy to see by adding up that this is correct, from the earlier formula $x_1^2 + x_2^2 = \frac{1}{2}(x_1 + x_2)^2 + \frac{1}{2}(x_1 - x_2)^2$, which applies column by column to the above table. The $S.S.$ in (a), when divided by $r$ and then by $(t - 1)$, will give a M.S. for the effect of the treatment with $t$ levels, i.e., it measures the differences due to this treatment on the average of $A$ and $B$, and this when divided by the error M.S. $s^2$ will give an $F$ with $t - 1$ and $(r - 1)(2t - 1)$ D.F. for testing this effect (assuming the arrangement to be one in randomized blocks).

The $S.S.$ in (b), when divided by $r$, will, because it has only 1 d.f., gives an M.S. for the difference between $A$ and $B$, on the average of all levels of the other treatment, and this when divided by $s^2$ will yield an $F$ with 1 and $(r - 1)(2t - 1)$ D.F.

If the $S.S.$ in (c) be treated exactly like that in (a) it will yield an $F$, with the same D.F., which will measure whether the response to the treatment with $t$ levels has been the same with $A$ as with $B$ (except for random fluctuations
brought about by experimental error). In other words, we can test whether interaction is significant or not. This is another case where we can see exactly what is meant by interaction, for $F$ will only be large if the $D$'s differ a great deal among themselves, and this will come about only if the response to the treatment with $t$ levels varies as between $A$ and $B$.

Arithmetically, the calculation of the $(c)$ component is exactly like that of the $(a)$ component, except that we work on differences instead of sums, and you would do well to accustom yourselves to this kind of computation, which will apply wherever there is a factor present at only 2 levels. It should be noted, also, that this case is like the previous one in that all three components can be directly calculated by simple methods, thus furnishing a check on the computations.

The general case of 2 factors will be one in which there are, say, $k$ levels of one treatment and $l$ levels of another. This means that we have to accommodate $t = kl$ treatments in the block to allow for all combinations of treatments, and then have, say, $r$ blocks. We may speak of this as a $(k \times l)$ design. The treatment-combinations will, of course, be arranged at random within the block, but when the treatment totals, for all blocks, have been worked out, they may conveniently be arranged in a row-column table, as below:

```
   A  \\
B   1 2 3 4

1 T1' T1' T1' T1' 
2 T2' T2' T2' T2'
. . . . . . . . .
. . . . . . . . .
G Tl' Tk' Tk' Tk' Tk'
```

```
T1 T2 ....... Tk G
```
Now this is very like the old arrangement for a single factor of treatment, and blocks, in a row-column arrangement, and the analysis is exactly the same. That is, the treatments S.S. with \( t - 1 \) D.F. is broken up into three components:

(a) the A set, with \( k - 1 \) D.F., having as S.S.

\[
\frac{1}{r} \left( T_1^2 + T_2^2 + \ldots + T_k^2 \right) - \frac{G^2}{rt}
\]

(b) the B set, with \( l - 1 \) D.F., having as S.S.

\[
\frac{1}{rk} \left( T_1'{}^2 + T_2'{}^2 + \ldots + T_l'{}^2 \right) - \frac{G^2}{rt}
\]

(c) the interaction of the A and B sets, with D.F. \((k - 1)(l - 1)\), and calculated by subtracting the (a) and (b) components from the total treatments S.S.

The divisor \( r \) occurs throughout because the treatment totals in the \((k \times l)\) table are totals of \( r \) replicates. Otherwise the calculation is exactly as described earlier. On dividing by the D.F. we have three M.S. instead of one, and therefore three values of F, which we may call \( F_1 \), \( F_2 \) and \( F_3 \). \( F_1 \) tests for the significance of yield differences among the \( k \) levels of the A set, on the average of the B set, i.e. it measures differences between \( T_1 \), \( T_2 \) \ldots \( T_k \); \( F_2 \) similarly tests for yield differences in the \( l \) levels of the B set, on the average of the A set; while \( F_3 \) measures the interaction of the two sets.
F₃ may or may not be significant. If it is, this will mean that the response to treatment in the A set is different at different levels of the B set. For summary purposes, therefore, we shall require the \( t = k \bar{X} \) treatment totals to be set out, as above, together with their S.E. \( s/\sqrt{r} \) in the ordinary way. If, however, \( F₃ \) is not significant we can consider that the data may be usefully summarized by giving the row and column totals only, together with their S.E.'s. That for the A set is \( s/\sqrt{r} \), and for the B set \( s/\sqrt{r}. \) This means incidentally that the two effects are measured with considerably enhanced precision, for the figures have the precision attached to the means of \( r \bar{X} \) and \( r \bar{Y} \) observations respectively, whereas the real replication was only \( r \)-fold.

We stated in (c) that the interaction S.S. was obtained by difference. It may, however, be obtained directly if we are prepared to do a little computation. The following simple example illustrates the point, and throws light, incidentally, on the meaning of the interaction as a measure of the deviations from an additive law of effects due to the two factors separately.

The numbers to be analyzed are:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>64</td>
<td>68</td>
<td>54</td>
<td>62</td>
<td>62</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>65</td>
<td>69</td>
<td>60</td>
<td>66</td>
<td>65</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>72</td>
<td>70</td>
<td>57</td>
<td>61</td>
<td>65</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>67</td>
<td>69</td>
<td>57</td>
<td>63</td>
<td>64</td>
</tr>
</tbody>
</table>

The table has been bordered by the means, instead of the totals, and the general mean is 64. (For ease of illustration numbers were taken so as to give whole number means.)

**Calculation of Interaction S.S.** The entry 64 is in the row whose mean is 62, 2 less than the general mean, and in the column whose mean is 67, 3 more than the general...
mean. On the additive law we should therefore expect the cell to contain a number +3 - 2, i.e. 1 more than the general mean, namely 65. A similar calculation for all cells shows that the numbers in the first row are all two less than their column means, while the numbers in the second and third rows are one greater. We thus have the set of theoretical numbers:

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>67</td>
<td>55</td>
<td>61</td>
</tr>
<tr>
<td>68</td>
<td>70</td>
<td>58</td>
<td>64</td>
</tr>
<tr>
<td>68</td>
<td>70</td>
<td>58</td>
<td>64</td>
</tr>
</tbody>
</table>

and therefore the set of residuals:

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>-3</td>
<td>-1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>-1</td>
<td>-3</td>
</tr>
</tbody>
</table>

These residuals should add to zero. Their sum of squares works out at 48, and this is the S.S. due to interaction, with 3 x 2 or 6 D.F. To show this another way, subtract 64 from all 12 original numbers, and sum the squares of the deviations. This gives a total S.S. of 324. Then for factor A subtract 64 from the column means, obtaining the deviations 3, 5, -7 and -1. Three times the sum of squares of these numbers gives 252, the S.S. for A with 3 D.F. For factor B we have the deviations -2, 1 and 1, and four times their sum of squares is 24, the S.S. for B with 2 D.F. We then have for interaction, by difference, 324 - 252 - 24 = 48, with 6 D.F., exactly as obtained above.

Going back to the ordinary arrangement of a single set of treatments in randomized blocks, it should now be clear that the error S.S. is, in the present language, the interaction between treatments and blocks, i.e. a measure of the way in which the differences between the treatment yields vary in the different blocks. Because this second factor, blocks, is not a second set of treatments, we can readily accept the fact that the interaction in this case measures the random error variation. We
also see from the above how it could be directly calculated.

**More than 2 Factors.**

From the theoretical point of view there is obviously no reason why we should not introduce 3 or more factors into an experiment. Provided the factorial design is followed, i.e., if arrangements are made for all combinations to be present, we can expand the former analysis of the treatments S.S., and measure separately the direct effects of the single factors and the various interactions between them. No new principles are involved except for the introduction of the idea of higher-order interactions. Suppose that in an experiment testing Nitrogen at four levels (N) with three varieties (effect V) we had wanted as well to test for the presence and absence of P. This would give us a (4 x 3 x 2) design, in which each block would contain 12 plots without P and 12 with P. All treatments would, of course, be randomized within the block. In general, with an (a x b x c) design, we can set out a two-way table showing two of the factors (totalling for the third) in 3 different ways (ab), (ac) and (bc). By carrying out the calculations described in the two-factor case we shall obtain, in addition to the marginal calculations which measure the effects of the factors singly, three different interactions which we can describe as the (ab), (ac) and (bc) interactions. These will normally be obtained by difference, i.e., by working out the total S.S. for each 2-way table and subtracting the marginal components. There are, of course, 6 margins with three 2-way tables, but they are not all different. With the (ab) table, for example, the marginal effects will be those of (a) and (b), while with the (ac) table the effect (a) will appear again in one margin, while the other will be (c).

These three interactions will be called first-order interactions, because we have only so far considered the factors in pairs, and the calculation will show that we have not yet accounted for the whole of the treatments S.S. In the case chosen above for illustration a sketch analysis of variance to show the D.F. will be as follows:
The first-order interactions have 3 x 2, 3 x 1 and 2 x 1 D.F., and after accounting for all these as well as the three direct effects, we see that we have 6 D.F. left. In working out the S.S. we can also see that we have not accounted for the whole of the total S.S., with 23 D.F. It is the remainder of the S.S. with the residual 6 D.F. which is conventionally described as the second-order interaction. It is always a little hard to explain what it really measures, but we may perhaps make it clear in this way. Choose one of the first-order interactions, that between N and V for example. Instead of working out a measure of this in the usual way from the whole of the data, let us assume that two 2-way tables have been set up for the NV comparison, one for all the plots without P and one for all the plots with P. The interaction NV can then be measured for each of the tables. If this interaction is approximately the same for both, apart from sampling errors, i.e. if N and V appear to interact in the same way in the absence of and in the presence of P, then there will be no second-order interaction, i.e. this M.S. component will not be significant when compared with error. But if N and V interact differently in the two cases, this will be indicated by the second-order interaction being significant. It is not often, in practice, that second or still higher order interactions show up as significant, so that it is usually possible to summarize the results of such complex experiments in relatively simple terms. We shall see later, also, that we can draw on experience of the general non-significance of such interactions to deal with

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>6</th>
<th>Second-order interaction NVP.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>2</td>
<td></td>
<td>Direct effects</td>
</tr>
<tr>
<td>P</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NV</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP</td>
<td>3</td>
<td></td>
<td>First-order interactions</td>
</tr>
<tr>
<td>VP</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remainder</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
experiments in which the real replication is perhaps inadequate, and so economize on
the number of plots.

In the general case, with factors at a, b and c levels, the 3 direct effects
will have a - 1, b - 1 and c - 1 D.F.; the three first-order interactions will have
(a - 1)(b - 1), (a - 1)(c - 1) and (b - 1)(c - 1) D.F., while the single second-order
interaction will have (a - 1)(b - 1)(c - 1) D.F. Algebraical addition shows that all
this adds up to abc - 1, the total D.F. The same general considerations apply to the
cases of 4 and more factors, with 4 factors, for example, there will be 4 direct ef-
facts, 6 first-order interactions, 4 second-order interactions and 1 third-order
interaction, these numbers being the binomial coefficients of \((1 + 1)^4\).

As a particular case let us consider a \((2^3)\) design. Suppose that we test
three fertilizers, N, P and K at 2 levels each, i.e. presence and absence. This
gives us eight treatment combinations to put into the block, one with no dressing at
all, three for the single dressings of N, P and K, three for the 2-factor combina-
tions NP, NK and PK, and one for all three dressings applied. The decomposition of
the treatments S,S. can be accomplished, as in the simpler case of the \((2^2)\) design,
by finding the appropriate 8 orthogonal components, using the table below. In this
table the treatment combinations have been indicated by small letters in order to
reserve the corresponding capital letters for the effects. This is in line with
using \(x_1\), \(x_2\), etc. for the primary variates and \(X_1\), \(X_2\) etc. for the orthogonal linear
functions. For a reason which will be made clear shortly, the order of the columns
is such as to exhaust the combinations of n and p before introducing the third factor
ek. As the coefficients are all +1 or -1, it will suffice to indicate these by means
of the signs + and - only, and since the divisor will be \(\sqrt{8}\) in all cases, there is
no need to record this for each line.
With a little practice it is fairly easy to see what the coefficients must be. In the first line they are all +. In the second line a + is put under any treatment combination with n in it, and a - under the others. Similarly for the third line (+ going with p) and the fifth line (+ going with k). In the fourth line we remember our diagonal rule, also the fact that in determining the NP interaction we add together the plots with and without k. Thus in the first four columns we put + with ooo and npo, and - with the other two. The same pattern is then repeated with the second four columns. The pattern in the 6th and 7th lines follow from similar considerations. Lastly, we note that the NPK second-order interaction is a measure of whether the first-order NP interaction is the same without k as with k. It may therefore be measured by retaining the signs of the fourth line so far as the last four columns are concerned, and reversing them for the first four columns. A check should always be made to see that the sum of products of any two lines is zero. This condition could, indeed, give us the signs for the last line if we found it difficult to determine them in any other way.

<table>
<thead>
<tr>
<th>Effect</th>
<th>ooo</th>
<th>noo</th>
<th>opo</th>
<th>npo</th>
<th>oek</th>
<th>nok</th>
<th>opk</th>
<th>npk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>N</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>P</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>NP</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>K</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>NK</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>PK</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>NPK</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>
Assembling the treatment totals (of, say, \( r \) replicates) in this way, they are then squared and divided by \( 8r \) to give the single D.F. components of the treatments sum of squares. There is a systematic way of doing the calculations without constructing the above table, which is given and illustrated by Yates in his Bulletin No. 35 of the Soil Bureau, entitled "The Design and Analysis of Factorial Experiments." The method, incidentally, makes use of the fact that in the case of single D.F. effects \( F \) is equal to \( t^2 \), so that there is no necessity to square. Let us work through Yates's example, in which the above treatments \( n, p \) and \( k \) are replaced by \( n, k \) and \( d \) (dung) and the figures operated on are the treatment totals of 4 replicates from a randomized blocks experiment on potatoes (plots 1/60 acre).

<table>
<thead>
<tr>
<th>Totals</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>425</td>
<td>851</td>
<td>3712</td>
<td>2331</td>
</tr>
<tr>
<td>n00</td>
<td>426</td>
<td>2321</td>
<td>6159</td>
<td>333</td>
</tr>
<tr>
<td>OkO</td>
<td>1118</td>
<td>2679</td>
<td>86</td>
<td>2271</td>
</tr>
<tr>
<td>nkO</td>
<td>1203</td>
<td>3480</td>
<td>247</td>
<td>105</td>
</tr>
<tr>
<td>Odd</td>
<td>1283</td>
<td>1170</td>
<td>2987</td>
<td></td>
</tr>
<tr>
<td>nOD</td>
<td>1396</td>
<td>85</td>
<td>801</td>
<td>161</td>
</tr>
<tr>
<td>Okd</td>
<td>1673</td>
<td>113</td>
<td>84</td>
<td>-669</td>
</tr>
<tr>
<td>nkD</td>
<td>1807</td>
<td>134</td>
<td>21</td>
<td>-63</td>
</tr>
</tbody>
</table>

\[
\sqrt{r s^2} = \sqrt{4 \times 347} = 37.2
\]
\[
\sqrt{8rs^2} = \sqrt{32 \times 347} = 105.4
\]

The original treatment totals have an estimated variance of \( rs^2 \), where \( r \) is the number of replications = 4, and \( s^2 \) is the error M.S., calculated in the ordinary way as 347. The columns headed (1), (2) and (3) represent a series of additions and subtractions. To form column (1) add the numbers in the preceding column, in pairs; thus 425 + 426 = 851, 1118 + 1203 = 2321 etc.
This provides the first four numbers. To obtain the second set of four numbers, subtract the numbers in the preceding column, in pairs, taking care always to subtract the same way round. Thus 426 - 425 = 1, 1203 - 1118 = 85 etc.
To form column (2) perform exactly the same operations on column (1) that were performed on the numbers in the totals column, i.e. first add in pairs and then subtract in pairs. Finally operate on the numbers in column (2) in the same way to obtain the final numbers in column (3). In the result we have a number of linear functions of the eight treatment totals, with coefficients either +1 or -1, which correspond to the eight orthogonal functions. Instead of squaring and dividing by 8, we can compare the numbers in column (3) directly with the common S.E. appropriate to them all, which is \( \sqrt{8n_r^2} = 105.4 \). The first number, 9331, has been underlined to indicate that it is merely the grand total of all plots, from which the correction factor can be obtained; we are not further concerned with it. The seven remaining numbers represent the single D.F. components of the treatment effects. Each number, divided by its S.E. 105.4, will be a \( \pm \) with 21 D.F., this last number being appropriate to an 8 treatment, 4 block experiment. The L.S.D.'s may be calculated as 219 for the 5 percent level, 295 for the 1 percent level and 403 for the 0.1 percent level. It follows that the N effect (333) is significant at 1 percent, and K, D and KD are significant at the 0.1 percent level. As this last contains the only significant interaction, it is of interest to see how it arises. Yates gives the following mean yields in tons per acre, averaging nitrogen and no nitrogen:

<table>
<thead>
<tr>
<th></th>
<th>No Potash</th>
<th>Potash</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Dung</td>
<td>2.84</td>
<td>7.78</td>
<td>5.31</td>
</tr>
<tr>
<td>Dung</td>
<td>8.97</td>
<td>11.65</td>
<td>10.31</td>
</tr>
<tr>
<td>Mean</td>
<td>5.90</td>
<td>9.72</td>
<td>7.81</td>
</tr>
</tbody>
</table>
The margins show the significant rises in yield due to Potash and Dung respectively. They interact in the following way; without Dung the rise due to Potash was 4.94 tons, but with Dung present it was only 2.68 tons. Alternatively we could look at the thing the other way round and say that the rise due to Dung was 6.13 tons in the absence of Potash, but only 3.27 tons when Potash was present.

Yates's method is merely a systematic way of constructing the several orthogonal quantities. To make it operative we should arrange the yields in the order shown, or in a similar order, such that the zero dressing is followed by any two of the single dressings and then by these dressings combined; the third dressing then follows, followed by its combinations with the other two in the order in which they were first taken. Lastly we have the combination of all three dressings taken together.

As a second particular case take the \((3^3)\) design, another of the popular favorites. Usually this applies when three standard fertilizers, such as N, P, and K, are tested at three levels of each i.e. 0, 1 and 2 dressings, in order that information may be available not only on whether the fertilizers is effective, but also on a suitable level of dressing to be used. Considering all combinations we shall have 27 treatment-combinations. If these be randomized in \(r\) blocks the only new problem concerns the division of the treatments \(S.S\) into its component parts. Following out the general theory, the decomposition will be as follows:

<table>
<thead>
<tr>
<th></th>
<th>D.F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>2</td>
</tr>
<tr>
<td>P</td>
<td>2</td>
</tr>
<tr>
<td>K</td>
<td>2</td>
</tr>
<tr>
<td>NP</td>
<td>4</td>
</tr>
<tr>
<td>NK</td>
<td>4</td>
</tr>
<tr>
<td>PK</td>
<td>4</td>
</tr>
<tr>
<td>NPK</td>
<td>3</td>
</tr>
<tr>
<td>Total treatments</td>
<td>26</td>
</tr>
</tbody>
</table>
You should be familiar by now with the nature of the calculations which will be required to effect the decomposition of the treatments S.S. so far as is indicated by this table. What is now of interest is whether a further decomposition, down even to single D.F., is possible. For various reasons such a decomposition is sometimes desired. For one thing, we may desire to know whether the response to any one of the fertilizers is linear, or whether there is any evidence, so far as three points can tell us, of non-linearity, as would emerge, for example, if a law of diminishing returns were operative. By this we mean that the response to the second unit of dressing might not be so great as that due to the first. For another, we might have an experiment with an inadequate amount of replication, and would then be interested in seeing whether any internal evidence was available from the differently treated plots to help us to estimate the random error. In an extreme case we might have no replication at all, and yet, as we shall see, something can be done to estimate the error.

It will be simpler for a time to go back to a (3^2) design, and so we shall first consider that we have 9 treatment-combinations, the factors being N at three levels in association with K, also at three levels. To lend concreteness to the case, we shall consider the following yields, in tons per acre, taken from a Rothamsted experiment on potatoes.

<table>
<thead>
<tr>
<th></th>
<th>ON</th>
<th>1N</th>
<th>2N</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>OK</td>
<td>8.7</td>
<td>11.8</td>
<td>13.9</td>
<td>34.4</td>
</tr>
<tr>
<td>1K</td>
<td>9.6</td>
<td>12.6</td>
<td>13.0</td>
<td>35.2</td>
</tr>
<tr>
<td>2K</td>
<td>9.9</td>
<td>12.1</td>
<td>13.6</td>
<td>35.6</td>
</tr>
<tr>
<td>Total</td>
<td>28.2</td>
<td>36.5</td>
<td>40.5</td>
<td>105.2</td>
</tr>
</tbody>
</table>

A first decomposition of the numbers in the above table, into marginal effects and interactions, yields the following results:
As the treatments consist of 0, 1 and 2 dressings of fertilizer, we may apply the orthogonal multipliers used earlier to separate the S. S. for 2 D. F. of both N and K into linear and quadratic components, according to the scheme:

<table>
<thead>
<tr>
<th>level of dressing</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>linear (l)</td>
<td>-1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Quadratic (q)</td>
<td>1</td>
<td>-2</td>
<td>1</td>
</tr>
</tbody>
</table>

These multipliers are applied to the marginal components, themselves totals of 3 values. For example, in the case of N we have 40.5 - 28.2 = 12.3, and so the linear component of the S. S. is $12.3^2/6 = 25.21$. The denominator 6 consists of the divisor 2 (the sum of squares of -1 and +1), multiplied by 3 since each figure is a total of 3 experimental values. Again we have $28.2 + 40.5 - 2(36.5) = -4.3$, and so the quadratic component of the S. S. is $4.3^2/18 = 1.03$. We note that these components have 1 D. F. each, and that their sum is 26.24, as we had formerly. A similar calculation will decompose the S. S. due to K.

The interaction is now seen to be decomposable into four components. There will be the interaction $(lq)$, also $(lq)$, $(qk)$ and finally $(qq)$. The first measures the way in which the linear components of the effects of N and K interact with one another; the second measures the interaction of the linear term of N with the quadratic term of K, and so on. These interactions may be determined by using the following four small tables of multipliers:
The construction of those tables is easily verified. In the case of $(\ell \ell)$, for example, we can place -1, 0, 1 along the last row and down the last column, and this determines the remaining coefficients. In the case of $(\ell q)$ we start with -1 0 1 in the first row, and then, because the first term in the first column is already -1, we reverse the usual signs taken with the quadratic multipliers. The remaining values are then determined. It will be noticed that the second row of numbers is 2 0 -2, and here the usual linear multipliers have been doubled as well as having their signs changed.

The $(q \ell)$ multipliers can be obtained from those in $(\ell q)$ by turning the scheme of numbers through a right angle. Finally the $(qq)$ set starts with 1 -2 1 in both the first row and first column, and then the other numbers are filled in to preserve these ratios. In each case the divisor is the sum of the squares of the numbers in the table.

The tabular numbers are then taken as multipliers of the corresponding yield values. Take the case of $(\ell \ell)$, where 5 of the numbers are zero. If we add 13.9 and 9.9 and subtract the sum of 8.7 and 13.6 we get 1.5. Then the S.S. for the $(\ell \ell)$ component of the interaction is $1.5^2/4 = 0.56$. And so on for the others. The following table shows under the heading 'c' the sum of products of yield and coefficient, and under 'd' the appropriate divisor.

Finally the component of the treatments S.S. is shown under 'c^2/d'.

<table>
<thead>
<tr>
<th></th>
<th>$(\ell \ell)$</th>
<th>$(\ell q)$</th>
<th>$(q \ell)$</th>
<th>$(qq)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0  -1</td>
<td>-1  0  1</td>
<td>-1  2  -1</td>
<td>1  -2  1</td>
</tr>
<tr>
<td>0</td>
<td>0  0  0</td>
<td>2  0 -2</td>
<td>0  0  0</td>
<td>-2  4  -2</td>
</tr>
<tr>
<td>-1</td>
<td>0  1</td>
<td>-1  0  1</td>
<td>1  -2  1</td>
<td>1  1  -2  1</td>
</tr>
</tbody>
</table>

Divisor 4  12  12  36
In an actual experiment these components of S.S. would be compared in

\[
\begin{array}{cccc}
\text{N} & \{l\} & 12.3 & 6 & 25.21 & 1 \\
\{q\} & 4.3 & 18 & 1.03 & 1 \\
\text{K} & \{l\} & 1.2 & 6 & 0.24 & 1 \\
\{q\} & -0.4 & 18 & 0.01 & 1 \\
\text{NK} & \{lq\} & 1.5 & 4 & 0.56 & 1 \\
\{qq\} & 2.1 & 12 & 0.37 & 1 \\
\{q\} & 0.3 & 12 & 0.01 & 1 \\
\{qq\} & 3.5 & 36 & 0.34 & 1 \\
\hline
\text{Total} & & 27.77 & & \\
\end{array}
\]

In an actual experiment these components of S.S. would be compared in turn with the error M.S. by the F-test. In addition to testing for the linear and quadratic parts of the main effects due to N and K, we usually find that if an interaction such as NK is significant, it is generally only the \(lq\) component of such interaction which is significant when the S.S. has been decomposed in this way. It is therefore usually claimed that the remaining components of interaction could be used to determine an estimate of the error M.S. if one were not otherwise available. For example, if the yields we are at present discussing were yields of 9 single plots laid out to test N and K at 3 doses each, but without any replication, we might argue that, because of the unlikelihood of the existence of \(lq\), \(ql\) and \(qq\) interactions, an error M.S. could be constructed by working out \((0.37 + 0.01 + 0.34)/3 = 0.24\), and assigning this 3 D.F. It would then be clear that the linear effect of N, for which F is a little over 100, is the only significant effect (for \(n_1 = 1, n_2 = 3\) the 5 percent point is 10.1, and the 1 percent point 34.1). The next largest effect, namely the quadratic effect of N, leads to an F a little over 4, which is not significant. This would not, however, be a very good experiment, since 3 D.F. for error is too small to give the desired precision.

Returning to the \(3^3\) design, we might have a 27 plot experiment for the 27 combinations of N, P and K at three levels each, but no replication. In
such a case the best we could do would be to work out the \((\bar{X}_\bar{Y})\) components of the \(NP\), \(NK\) and \(PK\) interactions, and lump in the remaining 3 sets of 3 D.F. each with the whole of the second-order interaction \(NPK (3 \text{ D.F.})\), thus giving us an error based on 17 D.F.. The sketch analysis of variance would then be as follows:

\[
\begin{array}{c|c}
\text{D.F.} & \\
N & 2 \\
P & 2 \\
K & 2 \\
NP(\bar{X}\bar{Y}) & 1 \\
NK(\bar{X}\bar{Y}) & 1 \\
PK(\bar{X}\bar{Y}) & 1 \\
\text{Error} & 17 \\
\text{Total} & 26 \\
\end{array}
\]

To determine the S.S., the first three lines are determined in the ordinary way. These S.S. can, of course, be divided into their two component parts for the linear and quadratic effects. The next three lines are determined as indicated above. Thus, for \(NP(\bar{X}\bar{Y})\), we should add up all the plots having 0, 1 and 2 dressings of \(K\), for each combination of \(N\) and \(P\), and so get a 3x3 table in which each entry was the sum of three yields. The multipliers would be the numbers in the square

\[
\begin{pmatrix}
1 & 0 & -1 \\
0 & 0 & 0 \\
-1 & 0 & 1 \\
\end{pmatrix}
\]

i.e. we would add the yields of the 2\(N\)(no \(P\)) and 2\(P\)(no \(N\)) plots, and subtract from the sum of the yields of the no \(N\) no \(P\) and 2\(N\)2\(P\) plots. This difference would then be squared and divided by 3x4, 3 because we are dealing with totals of 3 yields and 4 as the sum of squares of the four 1's in the above table.

Similarly for \(NK(\bar{X}\bar{Y})\) we should add up for \(P\) and form the two-way \(NK\) table of totals, then operate in the same way with the above multipliers. Finally, the same operation determines \(PK(\bar{X}\bar{Y})\). Since the total S.S. for treatments can be determined directly from the 27 yields, the "error" M.S. is...
obtained by difference, without further calculation. In this case we have a satisfactory degree of precision for error determination, within the limits of our assumptions, and in fact experiments have been set up in this way, with no replication. A further refinement would be to separate off the S.S. for the 1 D.F. (L L) component of the second-order interaction NPK as a check that this interaction is of no importance, and this would still leave 16 D.F. for error. The calculation would be made by applying the coefficients

\[
\begin{bmatrix}
-1 & 0 & 1 \\
0 & 0 & 0 \\
1 & 0 & -1
\end{bmatrix}
\]

to the (3x3) NF table for OK, and the reverse coefficients

\[
\begin{bmatrix}
1 & 0 & -1 \\
0 & 0 & 0 \\
-1 & 0 & 1
\end{bmatrix}
\]

to the similar table for 2K (the coefficients for 1K being all zero). It is therefore very easily made. In fact, the tables of coefficients for all 26 components can be easily worked out, but will not be given here.

There are, however, disadvantages in this type of design. For one thing, we have no real certainty that what we measure in the 17 (or 16) D.F. is error uninflated by real effects. Furthermore, the 27 plots will normally cover a lot of ground, and there is no provision for elimination of soil heterogeneity as in the case of the ordinary replicated and randomized trial with small blocks. The first difficulty is often met by having a second replication, if not a third, so that a direct estimate of error is possible; the second is met by the process of confounding, to which we shall come.

Returning now to general considerations, we have seen that when a number of factors are present in one and the same experiment, we not only have a first break-up of the total S.S. into component parts such as blocks, treatments and error, but also one of those components (treatments in the case illustrated) can itself be broken up. The important thing is to see that
in both cases the parts are orthogonal to one another, because it is this which ensures that the parts are additive and independent. A further illustration, but of a different kind, occurs where two or more sets of plots are laid out side by side comprising the same treatments. It was stated earlier that a single Latin square of size 4x4 was not very good. A very efficient experiment comparing four treatments can, however, be set out by having two Latin squares side by side. The Latin squares should independently be chosen by the usual random process, and we end up with eight replications of each of the four treatments, while so far as soil heterogeneity is concerned we see that we get much of this eliminated through rows and columns for each of the squares separately. To analyse the data note that a first break-up will be into the two parts (a) between squares and (b) within squares, while in (b) the two squares can at first be considered separately, as the following analysis shows:

\[
\begin{array}{c|c|c}
\text{Between squares} & \text{Within square I} & \text{Within square II} \\
\hline
\text{D.F.} & 1 & 15 \\
\hline
\text{Total} & 31 \\
\end{array}
\]

Now the part "within square I" represents the data of square I considered by itself, and this can in the usual way be divided into rows, columns, treatments, and error, with 3, 3, 3 and 6 D.F.. Similarly for square II. We then see that so far as soil heterogeneity is concerned we have eliminated a part of the total S.S. accounting for 13 D.F. (1 between squares, and 6 each for the rows and columns of the squares). We have two different estimates of treatment variation, with 3 D.F. each, total 6. But if we now think of the treatment totals as set out in a 4x2 table, thus:

\[
\begin{array}{c|cccc}
\text{Treatments} & 1 & 2 & 3 & 4 \\
\hline
\text{Squares} & \text{I} & \text{II} \\
\hline
\text{I} & \text{I} & \text{I} & \text{I} & \text{I} \\
\text{II} & \text{II} & \text{II} & \text{II} & \text{II} \\
\end{array}
\]
we see that an analysis of the eight totals can be made into row differences (1 D.F.) which is the difference between squares already accounted for; column differences (3 D.F.), which is the real treatment effect as obtained from the squares taken together; and the interaction between treatments and squares. Since the squares are side by side and the treatments are likely to react equally on each, this interaction will be a measure of random error, and can be added to the two S.S., with 6 D.F. each, which already measure the error of the separate squares. We thus end up with the following sketch analysis of variance:

<table>
<thead>
<tr>
<th></th>
<th>D.F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elimination of heterogeneity (a)</td>
<td>13</td>
</tr>
<tr>
<td>Treatments</td>
<td>(b)</td>
</tr>
<tr>
<td>Error</td>
<td>(c)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
</tr>
</tbody>
</table>

The calculation in (a) consists in working out rows and columns for both squares, and adding the square of the difference between square totals divided by 32. So far as (b) is concerned, this is the usual calculation on the four treatment totals of 8 plots each. The error (c) may then be obtained by difference from the total S.S., or alternatively by calculating its component parts. We see that the error D.F. have risen to 15.
Design of Experiments - Lecture 10

Split-plot Experiments.

In the factorial designs we have been discussing, all combinations of the factors were randomized over a replication, and we saw how a single error term (that obtained from interaction of total treatments and replications), was applicable to all comparisons, whether these were the direct effects of the various factors or the interactions between them. But a second factor may be introduced into a single factor experiment, or a third factor into a two-factor experiment, in another way. Thus, suppose a design has already been fixed, and the experiment is under weigh, comparing, let us say, a number of varieties, and it is then desired to test their manurial requirements. Without disturbing the set-up it is possible to divide each plot into a number of sub-plots, to which the different manurial treatments are applied at random. If done in this way for all plots we end up with one series of treatments within another. The experiment might be carried out the other way round. Thus the main plots could carry the manurial treatments, but be set out in rows which are planted down to different varieties. Such a device is particularly useful when some of the treatments to be tested are of such a nature that they necessitate large plots, whereas to have plots this size for all treatments would take up too large an area to be economical. Two examples may be cited from current work on the University Farm at Cambridge. In view of the nature of the soil a good deal of attention is being paid to the establishment of leys, and two problems are (1) the best composition of the ley and (2) its treatment during establishment. In a particular case 4 seeds mixtures were sown under wheat one year. One contained Italian Rye Grass, one a Red Clover, one had both and one had neither. The following year it was desired to graze half the plots and to cut the remainder for hay. Finally in the third year the experimental determination consisted in recording the weight of clover (green), in the May cut (this was actually
done by weighing the total produce and estimating the clover percentage by sampling. Now we need relatively large areas to enfold for grazing, also for haying if the ley is not to be cut by hand. So it was decided in the first instance to lay out eight large blocks, dividing each into 4 plots for the seeds mixture comparisons. In the second year each pair of blocks was taken as a unit, and a random choice was made as to which block (as a whole) should be grazed and which cut for hay. In the third year the individual plots were harvested.

Another example concerned the effect of dunging an area of beans, in which it was desired to contrast deep and shallow ploughing. A ploughing treatment requires a large plot, and it was therefore arranged to have 6 randomized blocks of two large plots each, for the ploughing comparison, but to divide up each plot into 3 sub-plots, and to assign to these at random the treatments none, 10 and 20 loads of dung per acre.

It is treatments like grazing or cultivation, or sometimes liming, or even varieties if sown by drill over a relatively small total experimental area, that would be applied to the main plots, while the sub-plots are suitable for variable artificial fertilizer treatments spread by hand, or for varieties sown in single rows by hand.

The total number of combinations of treatments is the same as in the fully randomized type of trial. In our first example there are 8 combinations of seeds mixtures and haying-grazing. In fact, in view of the seeds mixtures chosen we see that we really have a $(2^3)$ design. In the second example there are 6 combinations of ploughing and dunging. But all comparisons are not now made with the same degree of precision. In the first example the seeds mixtures are being compared between sub-plots grouped together in one main plot, and we are concerned with differences between these. Since the main plot is only part of the block, there will be less variation throughout its area due to soil
fertility differences than there will be due to this cause between the main plots of a block. We shall therefore expect to find that the seeds mixtures can be compared with greater precision than the grazing-haying treatments. This factor, indeed, should be taken into account in deciding which shall be the main plot and which the sub-plot comparisons. Generally we may take it that there is a loss of information on the main plot comparisons, and a gain on the sub-plot comparisons, as compared with the corresponding experiment in which the treatments were fully randomized in all combinations. This is a warning against resorting too freely to the split-plot design except in cases where the effects of the whole-plot treatments are not of primary importance, or where there are serious practical difficulties in doing anything else. As we shall see, the measurement of the significance of interaction between the two sets of treatments is based on the same error M.S. as the sub-plot comparisons, so that it is often good enough to resort to the split-plot method if it is interactions, rather than the main effect of the whole-plot treatments, that we are after.

The statistical analysis of such experiments is not really much more complicated than that of the ordinary factorial design. Let us think of a randomized blocks design, with r replicates, in which there are a different main treatments, and that the main plots, ar in number, are each divided into b parts to which sub-plot treatments are assigned. We may first consider that we have abr yields to analyze, divided into ar groups each of size b.

A first sketch analysis of variance gives

<table>
<thead>
<tr>
<th></th>
<th>D.F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Plots</td>
<td>ar - 1</td>
</tr>
<tr>
<td>Within plots, i.e.</td>
<td></td>
</tr>
<tr>
<td>Between sub-plots</td>
<td>ar(b - 1)</td>
</tr>
<tr>
<td>Total</td>
<td>abr - 1</td>
</tr>
</tbody>
</table>
The S.S. for the first line is obtained by calculating the S.S. of the plot total yields and dividing by b, since each figure will be a total of the yields of b sub-plots. The second S.S. is obtained either by difference from the total S.S., or by working out the S.S. for each set of b sub-plot yields and adding up for all plots.

The further analysis of the first S.S. is formally the same as in any randomized block trial, for the plots are all alike in respect to the sub-plot treatments. So we divide into blocks, main plot treatments and error, the only difference being that all three components are divided by b, since their total has already been so divided. This device, in fact, while not altering the relative values of the components, nor the tests which can be made on this part of the analysis alone, has reduced all comparisons to the sub-plot basis, which is a practical convenience. So far we have

\[
\begin{array}{c|c}
\text{Blocks} & r - 1 \\
\text{Main plot treatments} & a - 1 \\
\text{Main plot error} & (a - 1)(r - 1) \\
\hline
\text{Total between plots} & \frac{a(r - 1)}{b - 1} \\
\end{array}
\]

Turning now to the S.S. within main plots, we note that it may not be homogeneous if there are real differences between the sub-plot treatments, or if these interact significantly with the main-plot treatments. Formally the break-up of the S.S. proceeds as follows:

\[
\begin{array}{c|c}
\text{Sub-plot treatments} & b - 1 \\
\text{Interaction} & (a - 1)(b - 1) \\
\text{Sub-plot error} & a(b - 1)(r - 1) \\
\hline
\text{Total within plots} & \frac{a(r - 1)}{b - 1} \\
\end{array}
\]

The easiest way to calculate the S.S. is to set out the ab treatment totals in a 2-way table exactly as in a factorial design analysis, and work
out the two direct effects and their interaction. The direct effect due to main plot treatments goes into the first of the above two analyses, while the remaining two components go into the second. Finally the sub-plot error may be obtained by difference from the total within plots, or from the total S.S.

Putting the two parts together and using \( A \) and \( B \) as general terms for the two sets of treatments, we have the following sketch analysis of variance:

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>( r - 1 )</td>
</tr>
<tr>
<td>( A )</td>
<td>( a - 1 )</td>
</tr>
<tr>
<td>Error (i)</td>
<td>( (a - 1)(r - 1) )</td>
</tr>
<tr>
<td>( B )</td>
<td>( b - 1 )</td>
</tr>
<tr>
<td>( AB )</td>
<td>( (a - 1)(b - 1) )</td>
</tr>
<tr>
<td>Error (ii)</td>
<td>( a(b - 1)(r - 1) )</td>
</tr>
<tr>
<td>Total</td>
<td>( abr - 1 )</td>
</tr>
</tbody>
</table>

We end up with two error M.S. The \( A \) treatments are tested by taking \( F \) as the ratio of the \( A \) M.S. to the error (i) M.S., Similarly the \( B \) treatments, and the \( AB \) interaction, are tested against the error (ii) M.S.,

Effectively, therefore, we end up with a similar analysis of variance to the ordinary factorial form, but with the single error, having \( (ab - 1)(r - 1) \) D.F., split up into two parts, error (i), with \( (a - 1)(r - 1) \) D.F., being used for the main plot treatments and error (ii), with \( a(b - 1)(r - 1) \) D.F., being used for the sub-plot treatments and the interaction of the two sets of treatments. For summary purposes standard errors may be calculated. The matter is very simple if the interaction is not significant. If the error (i) and error (ii) M.S. are denoted by \( s_1^2 \) and \( s_2^2 \), then the \( A \) treatment means, of \( br \) plot values each, have the S.E. \( s_1/\sqrt{br} \), while the \( B \) treatment means, of \( ar \) values each, have the S.E. \( s_2/\sqrt{ar} \). If the interaction is significant we may wish to compare individual treatment means, of \( r \) plot values each. Set these out in a 2-way table in which the \( A \) treatments go along the rows, and the \( B \) treatments down the columns. In any comparison which involves the average
B effects and the interaction AB, but does not involve a mean A effect, i.e.,
comparisons which are in the same column, the appropriate S.E. is $s_p \sqrt{r}$.
Normally it will not be necessary to make comparisons between values in the
body of the table which include any component of the mean A differences, and
so it is normally sufficient to present the above 3 S.E.'s with the summary
table. Should, however, comparisons be wanted between means which are not
in the same column, I believe that a proper value for the S.E. of each of
these can be obtained from considerations such as the following. Since $s_2^2$
is an estimate of $b \sigma_2^2 + \sigma_2^2$, where $\sigma_2^2$ is the variance of main plot
values and $\sigma_2^2$ that of the sub-plot-values, we may estimate $\sigma_2^2$ by
$s_1^2 - s_2^2/b$, and it would appear therefore that the required estimated variance
of the means (of r values each) to be compared is

$$
\frac{1}{r} \left( s_2^2 + \frac{(s_1^2 - s_2^2)}{b} \right)
= \frac{1}{br} s_1^2 + \frac{b - 1}{br} s_2^2
$$

The S.E. is the square root of this value, and to be on the safe side the
significant value of t to be taken should be that appropriate to the D.F.
of $s_1$ (the smaller of the two). But this formula requires to be checked.

The analysis takes on an especially simple form when the main plot is
split into 2 parts only, a thing which is often done. Here we take advantage
of the short-cut calculations already exemplified with a (2 x a) design.
In fact, if the sums and differences of adjoining sub-plot yields are found,
a calculation on the sums yields the main plot part of the analysis, while a
parallel calculation on the differences gives the rest, except that in the
latter case there is no component for blocks. Imagine that the split of the
main plot is into manured and unmanured halves. Let x denote yield of the
unmanured half, and y the yield of the manured half. Further let $x + y = t$,
$x - y = d$. There will be ar values each of t and d. Let $T$ denote the sum
of the t's over all plots, let $T_1, T_2, \ldots T_a$ denote the main plot treatment totals and $T'_1, T'_2, \ldots T'_r$ the block totals. Use a similar notation, but with $D$ in the place of $T$, for the corresponding quantities involving the $d$'s. Then parallel calculations are made on the t's and d's as follows:

1. Total S.S. between main plots (ar - 1 D.F.)
   \[
   \frac{1}{2}(t_1^2 + t_2^2 + \ldots + t_{ar}^2) - T^2/2ar
   \]

2. Total S.S. within main plots (ar D.F.)
   \[
   \frac{1}{2}(d_1^2 + d_2^2 + \ldots + d_{ar}^2)
   \]

3. S.S. between blocks (r - 1 D.F.)
   \[
   (T'_1^2 + T'_2^2 + \ldots + T'_r^2)/2a - T^2/2ar
   \]

4. S.S. between main plot treatments (a - 1 D.F.)
   \[
   (T_1^2 + T_2^2 + \ldots + T_a^2)/2r - T^2/2ar
   \]

5. Interaction S.S. (a - 1 D.F.)
   \[
   (D_1^2 + D_2^2 + \ldots + D_a^2)/2r - D^2/2ar
   \]

6. S.S. for sub-plot treatments (1 D.F.)
   \[
   D^2/2ar
   \]

Given the above, the error sums of squares are readily obtained by difference. The main plot error S.S. is (1) - (3) - (4), while the sub-plot error S.S. is (2) - (5) - (6).
Efficiency

$s_1^2$ is an estimate of a variance necessarily greater than that of which $s_2^2$ is an estimate. This does not mean that $s_1^2$ is always greater than $s_2^2$, though it will usually happen so. It is possible that, as estimates, $s_1^2$ might be abnormally low and $s_2^2$ abnormally high. What is true is that the whole-plot comparisons are less precise than subplot comparisons involving the same number of plots. We may use our analysis of variance to calculate the single error that would have been appropriate had all ab treatment combinations been completely randomized over the block. What we do is to replace the various treatments mean squares by the appropriate error mean squares, so that the table would become

<table>
<thead>
<tr>
<th></th>
<th>D.F.</th>
<th>S.S.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>r - 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole plots</td>
<td>r(a - 1)</td>
<td>r(a - 1)$s_1^2$</td>
<td>$s_1^2$</td>
</tr>
<tr>
<td>Sub-plots</td>
<td>ar(b - 1)</td>
<td>ar(b - 1)$s_2^2$</td>
<td>$s_2^2$</td>
</tr>
<tr>
<td>Total within blocks</td>
<td>r(ab - 1)</td>
<td>$s_1^2 + ar(b - 1)s_2^2$</td>
<td>$s_2^2$</td>
</tr>
</tbody>
</table>

The derived error M.S. is $s^2$, calculated as shown above. We can then assert that the precision of the A comparisons would have been increased by complete randomization in the ratio $s_1^2 : s^2$, while the precision of the B effects and of the AB interaction would have been decreased in the ratio $s_2^2 : s^2$. $s^2$ is evidently the appropriately weighted mean of $s_1^2$ and $s_2^2$.

This split-plot method of analysis can be extended to deal with a further split within the first. Thus if to the former general design, with a main plot treatments and a split for b sub-1 treatments, we add a further split of the sub-1 plots for c sub-2 treatments, we can make a first broad analysis of variance...
by dividing the total S.S. into 3 parts, according to the scheme:

(i) between main plots \( \frac{D.F._1}{\text{ar} - 1} \)
(ii) between sub1-plots within main plots \( \frac{D.F._2}{\text{ar} \cdot (\text{b} - 1)} \)
(iii) between sub2-plots within sub1-plots \( \frac{D.F._3}{\text{ab} \cdot (\text{c} - 1)} \)

Total \( \frac{D.F._4}{\text{aber} - 1} \)

The further decomposition of (i), (ii), and (iii), will be shown separately:

(i) blocks \( \frac{D.F._5}{\text{r} - 1} \)

\[ A \]
\[ \text{Error (i)} \]
\[ \frac{D.F._6}{(\text{a} - 1)(\text{r} - 1)} \]
\[ \text{ar} - 1 \]

(ii) \( \frac{D.F._7}{\text{b} - 1} \)

\[ B \]
\[ \text{AB} \]
\[ \text{Error (ii)} \]
\[ \frac{D.F._8}{(a - 1)(b - 1)} \]
\[ a\cdot(b - 1)(r - 1) \]

Total \( \frac{D.F._9}{\text{ar} \cdot (\text{b} - 1)} \)

(iii) \( \frac{D.F._{10}}{\text{c} - 1} \)

\[ C \]
\[ \text{AC} \]
\[ \text{BC} \]
\[ \text{ABC} \]
\[ \text{Error (iii)} \]
\[ \frac{D.F._{11}}{(a - 1)(c - 1)} \]
\[ (b - 1)(c - 1) \]
\[ (a - 1)(b - 1)(c - 1) \]
\[ \text{ab} \cdot (\text{c} - 1)(r - 1) \]

Total \( \frac{D.F._{12}}{\text{ab} \cdot (\text{c} - 1)(r - 1)} \)

If the break into (i), (ii) and (iii) be made by the ordinary series of calculations, and if then the blocks and total treatments S.S. be worked out, the latter being decomposed, with the aid of the a x b x c array of treatment totals, into the three direct effects, the three first-order interactions and the second-order interaction, we have everything we need for computing the S.S. for the above table.

The calculation of S.E.'s should likewise present no difficulty. If the error M.S.'s are denoted by \( s^2_{11}, s^2_{22}, \) and \( s^2_{33}, \) we should note that \( s^2_{33} \) is an estimate of \( \sigma^2_{s2}, \) while \( s^2_{22} \) is an estimate of \( \sigma^2_{s1} + \sigma^2_{s2}, \) and \( s^2_{11} \) is an estimate of \( \sigma^2_{m} + \sigma^2_{s1} + \sigma^2_{s2}, \) where \( \sigma^2_{m} \) is defined as before, while \( \sigma^2_{s1} \) and \( \sigma^2_{s2} \) are the sub1 and sub2 treatment variances.

As a last illustration we shall take the case where the main plot is divided one way into strips for one set of sub-plot comparisons (e.g. in horizontal strips) and at the same time divided the other way (e.g. in vertical strips) for another
different set of sub-plot comparisons. The principle of the method of analysis here may be seen most simply by imagining that there are no main plot treatments at all. Let there be p plots set out over an area. Divide each horizontally into b strips, assigning plot by plot, at random, the b levels of treatment B to those strips. Independently of this divide the plots vertically into c strips, again assigning at random, plot by plot, the c levels of a treatment C. The difference between this lay-out and the last one (apart from the absence of an A set of treatments applied to the main plots) is that there the sub-treatments formed a split within each one of the sub-treatments, which were themselves a split of the main plot. In the present case, the two sets of treatments are at the same level, so to speak, but are interwoven. We obviously cannot treat this experiment as if the B and C sets were randomized in all combinations over the plots, even if it is true, which it is, that all combinations of B and C are present p times on the experimental area.

There will be three parts to the analysis apart from the differences between plots, and three error M.S.'s, and these parts are shown separately in the sketch analysis of variance which follows.

<table>
<thead>
<tr>
<th></th>
<th>D.F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between plots</td>
<td>p - 1</td>
</tr>
<tr>
<td>B</td>
<td></td>
</tr>
<tr>
<td>Error (B)</td>
<td>( b - 1 )</td>
</tr>
<tr>
<td></td>
<td>( (b - 1)(p - 1) )</td>
</tr>
<tr>
<td>Between rows within plots</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>( c - 1 )</td>
</tr>
<tr>
<td>Error (C)</td>
<td>( (c - 1)(p - 1) )</td>
</tr>
<tr>
<td>Between columns within plots</td>
<td></td>
</tr>
<tr>
<td>BC</td>
<td>( p(c - 1) )</td>
</tr>
<tr>
<td>Error (BC)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( (b - 1)(c - 1) )</td>
</tr>
<tr>
<td></td>
<td>( (b - 1)(c - 1)(p - 1) )</td>
</tr>
<tr>
<td>Interaction within plots</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>( p(b - 1)(c - 1) )</td>
</tr>
<tr>
<td></td>
<td>( bcp - 1 )</td>
</tr>
</tbody>
</table>
The calculations should present no difficulty. The first S.S., between plots, is calculated in the usual way from the plot totals. Within each plot the S.S. can be decomposed into rows, columns and interaction. The sum of all the row components gives the total of the next section of the analysis, while the sum of all the column components gives the total for the section following. Lastly, the sum of all the interaction components gives the total of the last section. We have now to separate off from these sections the components due to mean B and C effects and their interaction. This is best done by constructing the \((b \times c)\) table for totals of all plots and analyzing as has been done for the separate plots. We then find that we have three error S.S. obtained by difference. If the error M.S. are denoted, in order, by \(s_1^2\), \(s_2^2\), and \(s_3^2\), then it is clear that the direct B effect has to be tested against \(s_1^2\), the direct C effect against \(s_2^2\) and finally the interaction BC against \(s_3^2\). The appropriate \(t\) must be used in calculating the S.E.'s of the various means.

To impose a set of main plot treatments \(A\), say at \(a\) levels, on to this experiment is just to divide up the \(p\) main plots into \(r\) sets of \(a\), as we do when we design a randomized blocks experiment. \(p\) is then the product of \(a\) and \(r\), and \(r\) is the number of blocks. The allotment of the \(a\) levels of treatment \(A\) is done, of course, at random within blocks, independently of the treatments B and C. The S.S. between plots is then divided into blocks, A and error (A) in the normal way. This break-up enables a further break-up to be made in the remaining sections of the analysis of variance table, and this gives an opportunity of examining all the interactions of \(A\), \(B\), and \(C\) with one another. For example, so far as \(A\) and \(B\) are concerned, we have:
between plots

\[
\begin{align*}
\text{blocks} & : r - 1 \\
A & : a - 1 \\
\text{Error (A)} & : \frac{(a - 1)(r - 1)}{ar - 1}
\end{align*}
\]

between rows within plots

\[
\begin{align*}
B & : b - 1 \\
AB & : \frac{(a - 1)(b - 1)}{ar(b - 1)} \\
\text{Error (B)} & : \frac{a(b - 1)(r - 1)}{ar(b - 1)}
\end{align*}
\]

A similar analysis, taking "between columns" along with "between plots", will provide the basis for testing $C$ and $AC$ against an error ($C$). Finally we shall have

\[
\begin{align*}
\text{blocks} & : r - 1 \\
A & : a - 1 \\
\text{Error (A)} & : \frac{(a - 1)(r - 1)}{ar - 1}
\end{align*}
\]

interaction within plots

\[
\begin{align*}
BC & : \frac{(b - 1)(c - 1)}{ar(b - 1)(c - 1)} \\
ABC & : \frac{(a - 1)(b - 1)(c - 1)}{ar(b - 1)(c - 1)} \\
\text{Error (BC)} & : \frac{a(b - 1)(r - 1)}{ar(b - 1)(c - 1)}
\end{align*}
\]

and we see that not only BC, but also the second-order interaction ABC, must be tested against the third error M.S.

The analysis has not been written down as a whole in order to emphasize that after the first stage there are a series of three parallel developments, rather than, as with the successive sub-division of plots described earlier, a steady progression downwards.
Sampling

Sampling methods have, I believe, a lecture course to themselves, but it is appropriate, at this stage of a course on the design of experiments, to take up the subject from the point of view of the modifications that have to be made in the usual statistical analysis of experimental data to take account of cases where sampling is resorted to. Hitherto we have spoken as if a final determination such as yield was the only variable to be examined, and have assumed that the produce of the whole plot was available (except that in the last lecture the weight of green clover was stated to be a measure derived from (a) the total weight of the crop and (b) the percent clover in that crop, estimated by sampling). But there is a great deal more besides yield to take account of, and it is a wise experimenter who visits his plots at frequent intervals and takes observations on them. Some of these observations, such as number of plants germinating, number of tillers produced by various dates, etc., are amenable to statistical analysis. Such counts or measurements made during growth may be very valuable in helping to explain subsequent differences in yield, or in showing how yield is synthesized. Developmental studies can play an important part, for example, in studying the effect of applying artificial fertilizers to crops, particularly in respect to time of application.

But now we come to the point that it is next to impossible to count all the plants or all the tillers from all plots of a field experiment on, say, wheat. We must resort to sampling, by which we mean selecting in some fashion part only of the produce of the plot and producing a count for that part only, in the hope that this will be sufficiently representative of the total behaviour for reliable deductions to be made from the part. Another necessity for sampling arises from the fact that measures such as chemical constitution of the plants, or botanical analysis, involve the destruction of the material examined, and so only a portion can be taken, the remainder being left to grow on. Even in the case of yield
we sometimes have need of sampling. The plots may be too small for the use of ordinary farm implements, and yet, especially at outlying stations, resources for accurate harvesting of small plots may be absent. The best that can then be done is to sample the plots and drive back to the experiment station with the produce.
Suppose that we want a numerical determination of some characteristic of our material, e.g. tiller number, yield, % of dry matter, or of nitrogen, or of sugar, by sampling a proportion of the whole. The object is to obtain as close an estimate as we can of a measure that would be known exactly had the produce of the whole plot been counted, weighed or analyzed. I say exactly, in the sense that what was on the plot would be exactly determined if we measured the whole of it. But the ordinary measure of what is present on the plot is already known to be inexact in another sense. It is subject to various errors, of a major nature in things like soil fertility and of a minor nature in regard to weighing and instrumental etc. errors.

To take a measure from part only of the plot and to use this measure for the whole is to introduce another source of error, that due to sampling. For example, if one-tenth of the plot has been counted for number of tillers, the resulting figure, when multiplied by 10, will not be equal to the total number of tillers on the plot. If the yield of this one-tenth were determined, no matter how carefully, ten times this yield would not be the yield of the whole plot. The % dry matter in a pasture crop determined by weighing a few cuts, then drying and weighing again, would not furnish the exact % of dry matter present in the whole plot.

Our object is to obtain as close an estimate as we can by sampling of the measure which would be exactly determined if the whole plot had been taken instead of part. The problem is therefore similar to that always facing the statistician. In the ordinary way he postulates an ideal population of measures and assumes his set of observed measures to be a statistical sample from the population. His measures should be without bias, i.e. his $\bar{x}$, the sample mean, should be such that its expected value is $m$, the true mean; also he expects to be able, by comparing his different $x$ values, to estimate the degree of variability to which his measures are subject, i.e. to estimate $\sigma^2$ by means of $s^2$. He gets his information by seeing
to it that his sample is a random one. On top of this there are the usual rules to be followed for determining good designs, as in agricultural field experimentation generally.

There is a small degree of difference in sampling plots, in that from one point of view the population which is being sampled is a finite one, namely what is present on the plot, but this does not create a serious difficulty, especially as the sample is usually a relatively small part, say 10%, of the whole. The practical man, as well as the statistician, soon finds himself laying down one rule, namely that the sample should be representative, i.e. that it should give an unbiased estimate of the whole plot measure. Methods have to be devised for securing this representativeness, and this brings in, not only consideration of what fraction shall be taken, but also the pattern of selection. For example, a number of parts evenly distributed over the plot is likely to provide a more representative sample than a similar total amount bulked at one point only. A further rule is that since there will be error due to sampling, an element of random choice should enter into the taking of the sample if the size of the sampling error is to be measured. Some people would say that since the sampling error is included with the other errors when the data of the experiment, obtained by sampling, are analyzed, we do not specially want to know what the sampling component is, any more than we want to estimate other components, such as that due to soil. But, unless we have a tried method which is known from past experience to keep the sampling error down to a permissible limit, we need some method of controlling the sampling error, and if necessary reducing it, in the same sense that special techniques exist in field experiments for reducing the soil error. We cannot make much progress here unless we are able to assess the sampling error apart from the other components of error.

Consider first the sampling of a single plot. Let the sample be made up of p equal-sized sampling units or parts, which are selected at random from the bulk of the material at our disposal. Let the measures obtained from these sampling units
be denoted by \( x_1, x_2, \ldots, x_p \). These now form a statistical sample, and the arithmetic mean \( \bar{x} \) will furnish our best estimate of the true value \( m \) of the measure in a sampling unit of this size. The variation of the \( x \)'s around \( \bar{x} \) leads by the ordinary formula to an estimate \( s^2 \) of the variance of the measures, and the estimated S.E. of \( \bar{x} \) will then be \( s/\sqrt{p} \). This quantity \( s/\sqrt{p} \) we shall speak of as the sampling S.E., or simply as the sampling error, and its square will be called the sampling variance. If as an additional refinement we want to take account of the finite size of the "population" being sampled, then if the sampling unit is one \( n \)th part of the plot, the sampling variance will be \( s^2/p - s^2/n \), so that we may express the sampling error as \( s\sqrt{(1-f)/p} \), where \( f \) is the fraction \( p/n \) of the plot which is sampled. This is not very different from \( s/\sqrt{p} \) where \( f \) is a small fraction. We see at once that for a given size of sampling unit, the larger the number of sampling units taken the smaller is the error.

So long as we are considering a single plot only, we see that \( p \) should be of a reasonable size, say 10-12 at least, if the true sampling error is to be estimated by \( s/\sqrt{p} \) with the same precision as a plot error is normally estimated from the usual randomized designs. To have \( p \) so high incidentally keeps the sampling error low because of the occurrence of \( \sqrt{p} \) in the denominator. The question now arises as to how these \( p \) sampling units are to be selected. We may go ahead and determine an area, one \( p \)th of the whole, and make a random selection of \( p \) such areas. The area may be a rectangular piece of a pasture or cereal crop, or part of a row (or of 2 or 3 adjoining rows) of a root crop. Alternatively, it is sometimes convenient to measure an area of this size selected from a larger area, as would happen, for example, if 4 sub-sections in a pasture or cereal crop were selected so that spaces occurred between them according to some regular pattern, or if instead of measuring the produce from a single length of, say, 4 feet of row of a root crop, we were to
examine four separate foot-lengths, each separated by, perhaps, a foot. It is convenient to have a name for these parts or sub-sections of a sampling-unit, and we shall call them units. The object, of course, is to secure greater representativeness, which is done by avoiding the correlations that may exist between adjacent small areas, and by spreading the sampling units over a wider area. The point that must be borne in mind is that while it is permissible to use a regular pattern for the units, no attempt should be made to calculate the sampling error from the variation of individual units. The produce from the units should be cut and bulked, and the required measurement made from the sampling-unit as a whole. It goes without saying that while it is the sum total of all sampling-units which constitutes the sample, the measure for which is taken as the measure for the plot (multiplying, if we like, by n/p to obtain the plot measure if it is a count of plants, or of tillers, or a yield determination), we can get no information on the size of the sampling error unless we have separate determinations for each sampling-unit. They are the parts which have been selected at random, and therefore it is from them that we get the unbiased estimate of the plot determination, and also an estimate of the sampling error associated with that determination.

Sometimes it is advantageous to eliminate some of the heterogeneity within the plot by dividing it into sections and arranging to have an equal number of sampling units within each section. This is like introducing blocks into a field experiment. For example, if p = 12 we might divide the plot into 4 equal sections and choose 3 sampling-units at random from each section. In addition the sampling-unit may be made up of units as just indicated. The analysis of the set of p x's will, however, be different. The total S.S. will be divided into a part between sections, with 3 D.F., and the remainder, being the S.S. within sections, having 2 x 4 = 8 D.F.

It is this remainder which will furnish the estimate of sampling error, and the consequent reduction of error through elimination of the variability between sections may well compensate for the loss of 3 D.F.
So much by way of indicating the principles so far as sampling a single plot is concerned. When we turn to the data of a field experiment, of whatever character, the principles are the same, but the number of random sampling-units which require to be assessed separately may be much reduced, although their size will have to be increased to maintain the same fraction of the total experimental area. With q plots of p sampling-units each, we can have a first division of the total S.S. for the pq sampling-units into a part between plots, having \( q - 1 \) D.F., and a part within plots, having \( q(p - 1) \) D.F. The former is a measure of the total variation between plot samples (which are aggregates of sampling-units) and is therefore the part which can be further decomposed, depending on the design of the experiment that has been set up on those plots. For example if the q plots are given over to testing \( t \) treatments in \( r \) randomized blocks, so that \( q = rt \), the decomposition of the measure obtained for each plot, whether this be the mean of all the sampling-units on that plot, or their total, or their total multiplied by a factor to give the measure that corresponds with the whole plot area, will be into the usual components for blocks \( (r - 1) \) D.F.\) treatments \( (t - 1) \) D.F.\) and plot error \( (r - 1)(t - 1) \) D.F.\), this last, of course, including its component of sampling error. The latter part, i.e. that within plots, having \( q(p - 1) \) D.F., is a simple measure of the variation of sampling-units within plots, and will thus furnish us with our measure of the sampling error without further decomposition. At this stage we can, in fact, detect a formal analogy with the split-plot design. The place of the sub-plots carrying different treatments is taken by the sampling-units, which are not differently treated, and thus the former "within plots" analysis into sub-plot treatments, interaction with main-plot treatments and sub-plot error, is no longer applicable, and we have one component only, measuring the sampling error. Working on the basis of the sampling-unit as our unit of measurement we can have a formal analysis of variance as below.
<table>
<thead>
<tr>
<th></th>
<th>D.F.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>$r - 1$</td>
<td></td>
</tr>
<tr>
<td>Treatments</td>
<td>$t - 1$</td>
<td></td>
</tr>
<tr>
<td>Plot error</td>
<td>$(r - 1)(t - 1)$</td>
<td>$s_1^2$</td>
</tr>
<tr>
<td>Sampling error</td>
<td>$rt(p - 1)$</td>
<td>$s_2^2$</td>
</tr>
<tr>
<td>Total</td>
<td>$rt(p - 1)$</td>
<td></td>
</tr>
</tbody>
</table>

On the other hand, if we were to divide each plot into $u$ sections, allotting $v$ sampling-units to each, so that $uv = p$, a part would be separated off from sampling error due to the variation between sections, and it is the remainder that would furnish an estimate of the sampling error, thus:

<table>
<thead>
<tr>
<th></th>
<th>D.F.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between sections</td>
<td>$rt(u - 1)$</td>
<td></td>
</tr>
<tr>
<td>Within Sections (i.e. sampling error)</td>
<td>$rtu(v - 1)$</td>
<td>$s_2^2$</td>
</tr>
<tr>
<td>Total within plots</td>
<td>$rt(uv - 1)$</td>
<td></td>
</tr>
</tbody>
</table>

Now we stated that when we had a number of plots the number of independent sampling-units that had to be taken from each could be reduced. The reason for this is that the sampling error M.S. in both of the above tables is a composite of the variation between sampling-units taken from all plots, and since, with a set sampling scheme, there is no reason to expect the true sampling error to vary from plot to plot, we are able to get a single estimate of sampling error with a precision equivalent to $rt(uv - 1) = q(p - 1)$ D.F. in the one case and $q(uv - 1)$ D.F. in the other. If $q$, the number of plots in the experiment, is 10 or more, as it is pretty certain to be (otherwise the plot error could not be determined with much precision) we see that $p$ could be as low as 2 in the first case, and $v$ need certainly be no greater than 2 in the second. From this has arisen the quite common practice of taking duplicate sampling-units only from the plots, or at any rate from plot sections. We note from our concluding section on split-plot designs that with
2 sampling units per plot, a calculation on the sums of the sampling units will give us the analysis of the variation between plots into its usual components, and thus enable us to test for treatment differences, while half the sum of squares of the differences between sampling units will give us the S.S. for the sampling error line of the analysis. If the sampling unit measures be denoted by \( x \) and \( y \), and \( p = 2 \), \( q \) being the number of plots (not sectionally divided), then \( \frac{\sum (x - y)^2}{2q} \) will be the sampling M.S. and \( \frac{\sum (x - y)^2}{4q} \) the sampling variance. The sampling error, written in general as \( s^2 / \sqrt{p} \), will be the square root of the sampling variance.

The fact, however, that we have reduced our sampling units to 2 makes it all the more important that they should each be representative of the plot. This is usually achieved by widening the net-work of systematically arranged units that go to make up the sampling unit. To illustrate from a root-crop experiment on, say, potatoes or sugar-beet, if it were desired to sample 10% of the plots we might choose two numbers at random out of the first 20 numbers, say 7 and 18. A first sampling-unit would consist in bulking the seventh root from the beginning of the first row and every 20th root thereafter as the plot rows are traversed. To obtain the second sampling-unit a similar procedure would be gone through, beginning with the 18th root. The sampling units would then satisfy the condition of independence. A fresh random allocation of two numbers should, of course, be made for every plot.

Now go back to the analysis of variance table, and consider the following extract from it:

<table>
<thead>
<tr>
<th>D.F.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plot error</td>
<td>((t - 1)(r - 1))</td>
</tr>
<tr>
<td>Sampling error</td>
<td>(r(t - 1))</td>
</tr>
</tbody>
</table>

\(s^2_2\) is an estimate of \(\frac{\sigma^2}{2}\), the true variance of a sampling-unit, and \(s^2_1\) is an estimate of \(p \sigma^2 + \frac{\sigma^2}{2}\), where \(\frac{\sigma^2}{2}\) is the plot variance. But the estimated
sampling variance (i.e., the variance of the means of p sampling units) is \( s^2 / p \) and the estimated plot variance (also of the mean of p sampling units) is \( s_1^2 / p \). This latter figure, which corresponds to the estimated variance of a single plot in an ordinary experiment conducted without sampling, except that it now contains a component of sampling error, is therefore an estimate of \( \sigma^2 + s^2 / p \), and so we see that plot variance and sampling variance are additive. This gives us an idea of the extent to which our plot error will be increased by sampling. Suppose, for example, we expect a plot error of 10 per cent without sampling, and work a sampling process within a sampling error of 5 per cent. The aggregate plot error will then be \( \sqrt{100 + 25} = \sqrt{125} = 11.2 \) per cent. In terms of variance the ratio 100:125 or 80 per cent is a measure of the efficiency of conducting the experiment by a sampling process as compared with a full-scale operation. Normally we would be satisfied with this degree of efficiency, taking into account the amount of labor which has been saved.
Confounding

We have seen that there are very great advantages attendant on using factorial designs, in that information is available from the experiment not only on the main effects of the several factors included but also on the way in which these interact with one another. But there is a corresponding disadvantage. The enemy of precise and accurate statistical assessment of experimental data is heterogeneity of the material. In agricultural field experimentation we have seen how this can be materially reduced by proper design. Using randomized blocks, which are balanced with one another in the sense that all treatment combinations are present in each block, we are able to eliminate, both from our comparisons and from the error which measures them, that part of the heterogeneity which is represented by the differences between block totals. Any heterogeneity which is present within the blocks will, however, disturb the comparisons, and render the experiment less precise. We began, therefore, with a desideratum that the blocks should be reasonably small and compact. But we abandon this satisfactory state of affairs when we introduce more and more factors into one and the same experiment. The \(2^3\) design involves no more than 8 plots per block, but a \(3^3\) design brings in 27 plots. If each of these is, say, 1/50 acre, the area of the block is upwards of half an acre, and in an area this size there is plenty of room for heterogeneity to manifest itself.

Attention was early directed, therefore, to the question of whether something could be done to reduce heterogeneity even at the cost of sacrificing some of the information which the experiment is capable of providing. There were examples of such designs even in the very early days when randomized experiments were being developed. Since then the subject has been studied more systematically. Use is made of the principle of what R. A. Fisher called "confounding".

In the ordinary factorial type of design the information that comes out is of several kinds. There are the effects of single factors, which are always of interest, so that one would not willingly sacrifice information of this kind. Then there are
the interactions between pairs of factors. Now this is valuable information, of a kind that would not be available had the factors been tested in separate experiments. Usually, therefore, we would wish to preserve information on first-order interactions. The position is different when we come to interactions involving three or more factors, i.e., second and higher order interactions. These are often of little experimental value, either because the experimenter is fairly confident before he starts that they are unimportant quantitatively, or because there is little prospect of information on these interactions being utilized in practice. If, then, we are ready to sacrifice information of this kind as the price for securing increased precision on the more direct comparisons, we may resort to confounding. This consists in increasing the number of blocks, or groups, of relatively homogeneous material, beyond the number of replications, in the experiment, so that each complete replication may occupy two or more blocks. We want to be able to eliminate heterogeneity by taking out, in the Analysis of Variance, the whole of the variation between blocks. This can be done by arranging the treatment-combinations in block sets within each replication in such a fashion that the differences between block totals will involve the interactions which we are prepared to sacrifice, while at the same time seeing to it that the other effects involve only comparisons of plots within the blocks.

A very simple example, although of no practical utility, may be constructed by using a \((2^2)\) design, and this will serve to illustrate the method before we go on to consider more practical cases. Let us consider the testing of nitrogen and potash at two levels each, the treatment combinations being 00, n0, Ok and nk, which for convenience will be denoted by treatments 1, 2, 3 and 4. In the ordinary way these will be randomized in a replication consisting of 4 plots, and the experiment will consist of a number of such replications. Suppose we were prepared to sacrifice information on the NK interaction in order to improve the precision of the N and K direct effects. This could be done by putting treatments 1 and 4
together in one block, and 2 and 3 in the other. Let us suppose that there are 5 replications of these four treatments. We now have 10 blocks, 2 in each replication, and each has 2 plots. The S.S. for blocks, with 9 D.F., will involve differences between the 2 blocks in each replication as well as the differences between replications. The first block will have as its total the sum of treatments 1 and 4, while the other will consist of the sum of 2 and 3. Now the sum of 1 + 4 compared with the sum of 2 + 3 brings in the contribution to the NK interaction that is provided by that replication. The interaction is therefore confounded in the sense that when this calculation is made for all blocks we cannot measure the mean interaction directly, for included with it will be a component of heterogeneity arising from block fertility differences. In all replications, 1 and 4 should occur together in one block, and 2 and 3 in the other, but of course the allocation of a pair of treatments to one or other of the blocks of a replication is made at random, and likewise the allocation of the two treatments within the block.

On the other hand the N effect is obtained by comparing the total of 1 and 3 with that of 2 and 4. Using $t_i$ for treatment total, we require $t_2 + t_4 - t_1 - t_3$, which may be expressed as $(t_4 - t_1) + (t_2 - t_3)$. This is made up of a series of differences between the plots within the blocks, and is therefore unaffected by fertility differences between blocks. The N effect is unconfounded and may be measured. Similarly for the K effect, which is measured by $t_3 + t_4 - t_1 - t_2$, which may be expressed as $(t_4 - t_1) + (t_3 - t_2)$. This also only involves differences within blocks. To analyze such an experiment the four treatment totals are prepared in the ordinary way, but only the two direct effects are calculated. Then the blocks S.S.S. is calculated, and finally the error by difference from the total. The sketch A.V. is as follows:
This is all that need be done, but it is of interest to see how the analysis is made up. In the ordinary way, there would be 4 D.F. for blocks, three single D.F. for N, K and NK, and $4 \times 3 = 12$ D.F. for error. Now the error D.F. are made up of 4 each for interaction of blocks, with N and K, and 4 for interaction of blocks and NK. Only the first two of these, totalling 8 D.F., produce the error in the modified design, while blocks (4 D.F.), NK (1 D.F.) and their interaction (4 D.F.) lead to the total of 9 D.F. representing the full blocks component of the modified analysis. In fact, to the ordinary 4 D.F. for blocks we have added 5, one each for the interaction component calculated from each pair of blocks.

There is an analogy here with the split-plot design. Had we divided each replication into 2 main plots (the present blocks) for nitrogen and no nitrogen, i.e. had one consisted of treatments 1 and 3 and the other of treatments 2 and 4, and then split each main plot into 2 sub-plots for potash and no potash, the sketch A, V. would have been:

<table>
<thead>
<tr>
<th>D.F.</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>replications</td>
<td>4</td>
</tr>
<tr>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>error (i)</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>9</td>
</tr>
<tr>
<td>K</td>
<td>1</td>
</tr>
<tr>
<td>NK</td>
<td>1</td>
</tr>
<tr>
<td>error (ii)</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
</tr>
</tbody>
</table>
This is formally the same as the confounded analysis we have just carried out, except that here it is the main effect of N which has been confounded with blocks, and which can therefore only be measured against an error which represents the interaction of replications and nitrogen. In the confounded design, the two main plots within a replication contain different sub-treatments, so to speak, but so arranged that it is the interaction NK which is confounded. Normally we would not wish to confound a main effect, and the reasons for the split-plot design were explained at the time. Just as we have an error (i) above for the N effect, admitted to be of poor precision, so we could in the confounded design work out an error for measuring the NK interaction by means of the interaction of replications and NK, provided we have a number of replications. This, however, would normally be of poor precision, and since we have started by being prepared to sacrifice the comparisons which have been confounded, we do not as a rule bother with this part of the analysis.

The following simple example will illustrate the case we have just been discussing. Suppose there are 3 replications, each divided into 2 blocks, and that the numbers for analysis are as on the field plan below:

<table>
<thead>
<tr>
<th>block totals</th>
<th>00</th>
<th>0k</th>
<th>n0</th>
<th>nk</th>
<th>nk</th>
<th>nk</th>
<th>0k</th>
</tr>
</thead>
<tbody>
<tr>
<td>totals</td>
<td>16</td>
<td>19</td>
<td>23</td>
<td>21</td>
<td>20</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>rep. totals</td>
<td>35</td>
<td></td>
<td>44</td>
<td></td>
<td></td>
<td></td>
<td>41</td>
</tr>
<tr>
<td>treatment</td>
<td>18</td>
<td></td>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td>39</td>
</tr>
<tr>
<td>totals</td>
<td>5</td>
<td>10</td>
<td>9</td>
<td>11</td>
<td>12</td>
<td>14</td>
<td>120</td>
</tr>
</tbody>
</table>
\[ c = \frac{120^2}{12} = 1200 \]
\[ 5^2 + 11^2 + 9^2 + \cdots + 10^2 = 1290 \]
\[ \text{S.S. (total)} = 90 \]
\[ 16^2 + 19^2 + \cdots + 21^2 = 2428 \]
\[ \text{divide by 2} = 1214 \]
\[ \text{subtract c} = -1200 \]
\[ \text{S.S. (blocks)} = 14 \]

**Analysis of Variance**

<table>
<thead>
<tr>
<th></th>
<th>D.F.</th>
<th>S.S.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>5</td>
<td>14</td>
<td>2.8</td>
</tr>
<tr>
<td>N</td>
<td>1</td>
<td>33.3</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>1</td>
<td>40.3</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>4</td>
<td>2.3</td>
<td>0.583</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>11</strong></td>
<td><strong>90</strong></td>
<td></td>
</tr>
</tbody>
</table>

Further analysis for illustrative purposes only.

**Blocks**
\[ 35^2 + 44^2 + 41^2 = 4842 \]
\[ \text{divide by 4} = 1210.5 \]
\[ \text{subtract c} = -1200 \]
\[ \text{between reps.} = 10.5 \ (2 \text{ D.F.}) \]

**within reps**
\[ \frac{1}{6}(3^2 + 2^2 + 1^2) = 3.5 \ (3 \text{ D.F.}) \]
\[ \frac{1}{14} = 0.071 \]
Note: Within reps contains NK, calculated as

\[(18 + 39 - 31 - 32)^2/12 = 3.0 \quad (1 \text{ D.F.})\]

The remainder, 0.5, is interaction of NK and reps. (2 D.F.)

\[
\begin{align*}
\text{Error} & \\
(10 + 11 - 5 - 9)^2/4 &= 12.25 \\
(11 + 14 - 7 - 12)^2/4 &= 9 \\
(10 + 14 - 6 - 11)^2/4 &= 12.25 \\
\end{align*}
\]

Less N

\[
\begin{align*}
33.5 \\
33.3
\end{align*}
\]

Interaction of N and reps. 0.16 (2 D.F.)

\[
\begin{align*}
(9 + 11 - 5 - 10)^2/4 &= 6.25 \\
(12 + 14 - 7 - 11)^2/4 &= 16 \\
(11 + 14 - 6 - 10)^2/4 &= 20.25 \\
\end{align*}
\]

Less K

\[
\begin{align*}
42.5 \\
40.3
\end{align*}
\]

Interaction of K and reps. 2.16 (2 D.F.)

\[
\begin{align*}
\text{Sum of interactions of N and K with reps} &= 2.3 \quad (4 \text{ D.F.})
\end{align*}
\]
A design which is of considerable utility is where there are three factors a, b and c in the \(2^3\) design, and each replication is divided into two blocks of 4 plots each, in such a fashion that the second-order interaction ABC is confounded. You will remember that with this lay-out it was possible to divide up the total treatments S.S., with 7 D.F., into 7 constituent parts, with 1 D.F. each and that the calculation in each case involved comparing the total of one-half of the plots with the total of the other half, in accordance with a schematic table which was given. The last line of that table (Lecture 8 p. 6) was as follows:

\[
\begin{tabular}{cccccc}
000 & 000 & 000 & 000 & 000 & 000 \\
abc & + & + & + & + & + \\
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8
\end{tabular}
\]

For convenience we have added a numbering system to the treatment-combinations. To determine ABC we work out the treatment totals for 2+3+5+8 and 1+4+6+7, and then subtract, square the difference and divide by the total number of plots. This gives us the required clue for the division of a replication into two blocks. One block will contain the treatments 2, 3, 5 and 8, i.e. the plots having a only, b only, c only and all three combined; while the second block will contain 1, 4, 6 and 7 i.e. the plots having no dressing, together with the three which have two of the treatments in combination.

We must therefore randomize in each replication with this restriction in mind. We divide the first set of 8 plots into 2 blocks of 4, assigning at random to the first block either the set 1, 4, 6, 7 or the set 2, 3, 5, 8. Within each block the set of 4 treatments is allocated at random to the 4 plots. This procedure is repeated with the other replications. The analysis of the resulting measures proceeds very simply. Let us suppose by way of illustration
that there are 5 replications, i.e. 10 blocks. The total S.S. (39 D.F.) is first worked out, then the blocks S.S. (9 D.F.) and finally the S.S. for the unconfounded treatment effects (6 D.F.). The error S.S. is then found by difference. The D.F. are 39-9-6 = 24, which can be checked by noting that it is the product of 4 and 6, the D.F. for replications and unconfounded treatments. We can proceed by finding the total treatments S.S. (7 D.F.) and then subtracting the confounded ABC effect (1 D.F.) instead of building up the 6 separate parts, but it is usually of interest to do the complete break-down of the treatments S.S.

The above is a particular case of the general system of confounding in powers of 2, by which is meant that when s factors are present at 2 levels each, so that there are \(2^s\) different treatments, we may divide up each replication into a number \(2^b\) of blocks in order to introduce confounding. There will then be \(2^s-b\) plots in each block. For example, with 6 factors there will be \(2^6\) or 64 treatment-combinations, and if we arrange to have 4 blocks in each replication there will be 16 plots in each block. \(s\) is 6 and \(b\) is 2 and \(s-b = 4\). The number of comparisons to be confounded will be one less than the number of blocks i.e. \(2^b-1\), or in this case 3. Now we naturally want to choose high-order interactions to confound, as being of little practical interest, but we have to be careful.

With 6 factors there will be 6 direct effects, A, B, C, D, E, F; 15 first-order interactions AB, AC, ... EF; 20 second-order interactions ABC, ABD, ... DEF; 15 third-order interactions ABCD, ABCE, ... CDEF; 6 fourth-order interactions ABCDE, ... BCDEF; and 1 fifth-order interaction ABCDEF. Now there is an easily demonstrated rule that if any two interactions be chosen to confound, the third is automatically determined by throwing together the letters of the two already chosen and suppressing any letters they have in common. Thus if we chose ABC and ADE we should necessarily also confound BCDE, which can be described as the interaction of ABC and ADE. Now this means that we could not confound at one and the same time the fifth-order interaction ABCDEF and two chosen fourth-
order interactions, for these would not combine in the way indicated. If we chose ABCDE and BCDEF then we must also have AF confounded, and this we would wish to avoid, since it is a first-order interaction. But we could choose ABC and DEF, two second-order interactions, and then we would also have their interaction ABCDEF confounded. The confounded interactions could be confined to third-order by choosing, for example, ABCD and CDEF, when the third would be ABEF. The particular factors chosen for inclusion in the experiment will always help to determine our choice.

A useful rule is that, without confounding any interaction of less than 3 factors, we may use any number of factors which is less than the number of plots in a block. So if we have 8 plots in a block we may go as far as 7 factors, while we could go up to 15 factors in blocks of 16 plots. In the last case illustrated \((2^3)\), 4 was 3 and 1 was 1 so that there were \(2^3-1 = 4\) plots in the block. This was just the right number for going up to 3 factors without confounding a first-order interaction, and as we saw the second-order interaction ABC was the only one to be confounded in this set-up. This agrees with the formula \(2^1 - 1 = 1\). With only one comparison confounded, it is easy to write down the contents of the two complementary blocks in a replication.

But when there is a group of interactions to be confounded it would be tedious to work out the block contents by successive sub-division. A general method may, however, be given. Note first that in the \((2^3)\) case the treatments which went with the control in one block were ab, ac and bc, which have an even number of letters in common with the comparison abc which was confounded. A set such as these form a group since if ab and ac possess this property their interaction, namely bc, must also possess it. To generalise this, we begin by finding the set of treatments which goes with the control by looking in succession for s-b treatment symbols which have the property, and then writing down all their
interactions, giving us the first "intra-block sub-group", which will be of order $2^{s-b}$. We may then turn to another block containing a treatment that was not in the first, and all we have to do then is write down the interactions of this treatment with the treatments of the first sub-group. And so on for all blocks.

Let us illustrate with a $(2^5)$ design in which we shall divide up each replication into $2^2$ blocks each consisting of $2^3$ plots. $s$ is 5 and $b$ is 2, so that $s-b = 3$. We require to choose $2^2-1$ or 3 interactions for confounding. If we choose $ABC$ and $ADE$ then we must necessarily choose their interaction $BCDE$. This completes the choice. We now require in succession the treatments having an even number of letters in common with both $ABC$ and $ADE$. So the contents of one block will consist of the treatments in column I of the table below, in which the symbol (1) denotes the control.

<table>
<thead>
<tr>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>a</td>
<td>b</td>
<td>d</td>
</tr>
<tr>
<td>bc</td>
<td>abc</td>
<td>c</td>
<td>bcd</td>
</tr>
<tr>
<td>dc</td>
<td>ade</td>
<td>bde</td>
<td>e</td>
</tr>
<tr>
<td>abd</td>
<td>bd</td>
<td>ad</td>
<td>ab</td>
</tr>
<tr>
<td>abc</td>
<td>bc</td>
<td>ae</td>
<td>abde</td>
</tr>
<tr>
<td>acd</td>
<td>cd</td>
<td>abcd</td>
<td>ac</td>
</tr>
<tr>
<td>ace</td>
<td>cde</td>
<td>abcde</td>
<td>acde</td>
</tr>
<tr>
<td>bcd</td>
<td>abed</td>
<td>cde</td>
<td>bce</td>
</tr>
</tbody>
</table>

To obtain another block start with a treatment, e.g. (a), that is not in the first. If this be put at the head of column II, the remaining treatments are easily derived by operating with (a) on the treatments of the first column, other than (1), by the rule for finding interactions already explained in terms of effects (capital letters). Thus (a) taken with (bc) gives (abc), while with (abd) it gives (bd). And so on. We start a third series with (b), which has not yet been used, and operate with it on column I to produce the required set (column III). As this set contains (c) we must pass to (d), in conjunction with column I, to derive the last set (column IV).
It will be noticed that all 32 treatment-combinations are now present in four sets. For the first replication of the experimental area we allot a column of symbols at random to each block, then within the blocks we allot at random the eight treatment-combinations to the plots of the block. This procedure is repeated with the other replications.

To analyse, suppose there were 3 full replications, so that the experiment consisted of 96 plots. The total S.S. has 95 D.F. There are 12 blocks from which a block S.S. is obtained with 11 D.F. Next add up the 3 replicates for each treatment-combination, setting out the totals in the form of the above table with four columns I-IV. From this table of 32 totals work out the total treatments S.S., with 31 D.F., and subtract the component with 3 D.F., for the column differences, leaving the unconfounded treatments S.S. with 28 D.F.

The error S.S. (56 D.F.) is then found by difference. It is a good exercise to show that a contrast between columns I + IV and columns II + III gives the ABC interaction, while the ADE interaction is given by a similar contrast between columns I + III and columns II + IV. The remaining contrast, namely that between columns I + II and columns III + IV, gives the BCDE interaction. All three are determined together when we take out the column differences (3 D.F.). We have not yet decomposed the treatments S.S., but that is easily done by means of the ordinary rules, certainly so far as direct effects and first-order interactions is concerned. Yates in "Design and Analysis of Factorial Experiments" shows the detailed calculations for a single replication of such an experiment. Here he uses all the unconfounded second-, third- and fourth-order interactions (D.F. = 16 - 3 = 13) as his measure of error (para. 7 p. 27).

There are many possibilities of this sort. Fisher, in "The Design of Experiments", para. 45, illustrates a case of including 7 factors at two levels, with the 128 treatment-combinations in a replication being divided into 16 blocks.
of 8 plots each. \( s \) is 7 and \( b \) is 4, so as \( 2^4 - 1 = 15 \), there are no fewer than 15 interactions to be confounded. He lists a set of these, together with the contents of the first block, from which the other blocks may be generated. An advantage of having a scheme as detailed as this is that it is easy to pass from any satisfactory solution to one involving one factor fewer, by deleting all symbols of the subgroup confounded which involve any chosen letter, thus halving its order, at the same time deleting the same letter from the symbols in which it occurs in the intra-block subgroup. These changes introduce no new interaction to be confounded, and do not reduce the order of those already selected, so that if none involve less than 3 factors in the initial solution, this will still be true of the new solution. From Fisher's solution may be obtained solutions for 6, 5 and 4 factors in blocks of 8 plots; the one we have illustrated is, in fact, a special case of his. On the other hand, if we wanted blocks of 4 or 16 plots, we should have to make a fresh start with the new size of block. We cannot, for example, obtain from Fisher's case the \((2^3)\) design in 2 blocks of 4 plots that we have previously illustrated.

It is impossible in a short course to go into full details of the possible designs, and we shall content ourselves with a short table in which specimen necessary interactions to be confounded are indicated.

<table>
<thead>
<tr>
<th>No. of treatments</th>
<th>No. of Plots per block</th>
<th>Confounded interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 2^3 = 8 )</td>
<td>4</td>
<td>ABC</td>
</tr>
<tr>
<td>( 2^4 = 16 )</td>
<td>{ 4 8 }</td>
<td>AB, ACD, BCD</td>
</tr>
<tr>
<td>( 2^5 = 32 )</td>
<td>8</td>
<td>ABC, ADE, BCDE</td>
</tr>
<tr>
<td>( 2^6 = 64 )</td>
<td>{ 8 16 }</td>
<td>ABC, CDE, ADF, BEF, ABDE, BCDE, ACF, ABCD, ABEF, CDEF</td>
</tr>
</tbody>
</table>

It will be left as an exercise for the student to decide the make-up of the blocks in the cases not hitherto considered; for example, he should check that in the
(2^4) case, allowing 8 plots per block (ABCD confounded) the contents of the 2 blocks are as in the two lines below:

(1) ab ac bc ad bd cd abcd
    a  b  c  abc d  abd acd bcd

Such a design will enable an experimenter to cope with 4 factors, each at two levels, and yet keep his blocks down to 8 plots. To bring them down to 4 plots would mean sacrificing one first-order interaction.
So far we have considered the make-up of a single replicate and have regarded the replicates as simple repetitions of one another, except for re-randomization of contents. But instead we can resort to partial confounding by arranging to change the interactions to be confounded as we pass from one replicate to another. If we do so we can arrange that partial information, at least, is available on all interactions. To illustrate in a single case, we saw that in the \((2^3)\) design with ABC confounded, the two blocks of a replication would contain \((1), ab, ac, bc and a, b, c, abc\) respectively. If, on the other hand, AB were confounded, the two blocks would contain \((1), ab, c, abc and a, b, ac, bc\) respectively. In a two-replicate experiment we could use the first arrangement in Replicate I, and the second in Replicate II. This would give us half-information on the interactions AB and ABC, and full information on the others.

If all interactions of the same order are of equal interest, it may be undesirable to confound some of them more fully than others. In such cases, provided the necessary amount of replication can be afforded, a balanced design can be used, incidentally making the computations simpler. By a balanced design we mean utilising partial confounding with enough replications to ensure that all interactions of the same order are confounded to the same extent. For example, we have just indicated the set-up in a replicate of a \((2^3)\) design which would confound AB. A second replicate could confound AC, and a third BC, and as this exhausts the possibilities of selecting two letters out of three, a balanced design would consist of these 3 replicates and would give \(2/3\) information on each of the first-order interactions. As an exercise I have chosen an experiment of this kind to which a fourth replication, with ABC confounded, has been added. The resulting design is still a balanced one, but this time there will be \(3/4\) information, both on all first-order interactions and on the second-order interaction.
The design (given in our table) for $2^4$ treatments in blocks of 4 plots, confounding $AB$ as well as two second-order interactions, shows that six replicates would be required for a balanced design. These would confound, in succession, $AB$, $AC$, $AD$, $BC$, $BD$ and $CD$, as well as their complementaries out of the second-order interactions. We would get $5/6$ information on each of the first-order interactions, since only one out of 6 is confounded in each replicate, and $1/6$ information on each of the second-order interactions; for since with $AB$ goes $ACD$ and $BCD$, two out of the four second-order interactions will appear in each replicate, and it can easily be seen that any of the four will appear in three replicates and not in the other three. It will be of interest to write down the complete lay-out in this case. At the head of each replication the interactions confounded in that replication are stated (the capital letters), while inside the rectangle the treatments to be allotted to each of the four blocks are indicated by small letters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AB, ACD, BCD</strong></td>
<td><strong>AC, ABD, BCD</strong></td>
</tr>
<tr>
<td>(1) abc</td>
<td>(1) a b c</td>
</tr>
<tr>
<td>cd</td>
<td>bd</td>
</tr>
<tr>
<td>acd</td>
<td>abd</td>
</tr>
<tr>
<td>bcd</td>
<td>d</td>
</tr>
<tr>
<td>ab</td>
<td>bc</td>
</tr>
<tr>
<td>ac</td>
<td>ad</td>
</tr>
<tr>
<td>abd</td>
<td>bcd</td>
</tr>
<tr>
<td>bd</td>
<td>ad</td>
</tr>
<tr>
<td>ac</td>
<td>ab</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rep. III</th>
<th>Rep. IV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AD, ABC, BCD</strong></td>
<td><strong>BC, ABD, ACD</strong></td>
</tr>
<tr>
<td>(1) abd</td>
<td>(1) a b c</td>
</tr>
<tr>
<td>bc</td>
<td>bd</td>
</tr>
<tr>
<td>abc</td>
<td>abd</td>
</tr>
<tr>
<td>ac</td>
<td>ad</td>
</tr>
<tr>
<td>bc</td>
<td>cd</td>
</tr>
<tr>
<td>ab</td>
<td>ac</td>
</tr>
<tr>
<td>ad</td>
<td>bc</td>
</tr>
<tr>
<td>abc</td>
<td>bcd</td>
</tr>
<tr>
<td>bc</td>
<td>ad</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BD, ABC, ACD</strong></td>
<td><strong>CD, ABC, ABD</strong></td>
</tr>
<tr>
<td>(1) abc</td>
<td>(1) a b c d</td>
</tr>
<tr>
<td>ac</td>
<td>ab</td>
</tr>
<tr>
<td>c</td>
<td>b</td>
</tr>
<tr>
<td>abc</td>
<td>a bd</td>
</tr>
<tr>
<td>acd</td>
<td>abcd</td>
</tr>
<tr>
<td>ab</td>
<td>bd</td>
</tr>
<tr>
<td>ad</td>
<td>bc</td>
</tr>
<tr>
<td>abc</td>
<td>bcd</td>
</tr>
<tr>
<td>acd</td>
<td>abcd</td>
</tr>
<tr>
<td>abd</td>
<td>bd</td>
</tr>
<tr>
<td>ab</td>
<td>bc</td>
</tr>
</tbody>
</table>
**Statistical Analysis:**

With complete confounding it has already been explained that all we have to do is calculate the total and the blocks S.S. in the usual way, then work out the total treatments S.S., subtracting from it the confounded components, represented by the differences between the blocks compounded from all replicates. The error S.S. is then found by difference. Of course a little more work is involved if the single D.F. treatment S.S. (i.e. the unconfounded effects) are required, but this is straightforward. In the case of partial confounding, we calculate separately the S.S. for each comparison which is partially confounded, using for this purpose only the replications in which it is unconfounded. Thus in Exercise 7, while the treatment totals of all four replicates are used to determine the direct effects of A, B and C, we subtract replicate II before calculating AB, replicate III before calculating AC and replicate I before calculating BC. The second-order interaction ABC is got from the total of the first three replicates. To complete the analysis all we need is the blocks S.S., with 7 D.F., calculated in the usual way.

So far we have restricted ourselves to designs of the \(2^8\) type in considering confounding, but the method is fairly general, and a large number of designs have been worked out where the factors are at 3, 4, etc. levels, alone or in conjunction with each other, or with factors at 2 levels only. We shall choose for illustration the \(3^3\) design, in which the 27 treatment-combinations in each replicate are divided into 3 blocks of 9 plots each. This will round off our former consideration of this design, which is a popular one because it enables us to deal at three levels, instead of two, with three factors, such as the main manurial dressings N, P and K. It will serve, also, to indicate a different confounding method.
The treatments will be indicated by a three-figure number in which the figures are 0, 1 and 2, and their order is that of the factors a, b and c. Thus 021 will indicate the treatment a at zero level, b at level 2 and c at level 1. There are obviously 3×3×3 or 27 treatment symbols needed to cover all combinations. There would be no difficulty if all 27 treatments were completely randomized over a block of 27 plots, or if only the factors a and b were present, yielding 3×3 or 9 treatment-combinations to cover a block of 9 plots. What we want to do is to break up the 27 combinations into 3 sets of 9, so as to form 3 blocks out of a replication, in such a fashion that only the second-order interaction ABC is affected by confounding. We note first that blocks of 9 plots are needed if all the first-order interactions are not to be confounded. We next note that a set of 9 treatments present in the same block must be such that the 3 levels of each factor (0, 1 and 2) must each be represented by 3 plots, and that the 9 combinations of each pair of factors (ab, ac and bc) must be represented by one plot each. A possible arrangement for a single block is therefore represented by the Latin square,

\[
\begin{array}{ccc}
 & b & \\
 l. & 0 & 1 & 2 \\
a & 0 & 0 & 1 & 2 \\
1 & 1 & 2 & 0 \\
2 & 2 & 0 & 1 \\
\end{array}
\]

in which rows represent levels of treatment a, columns represent levels of treatment b, and the numbers within the square represent levels of treatment c. It is easily seen that such a set of 9 combinations satisfies the above conditions.

For the other 2 blocks we must similarly choose sets of 9 treatment-combinations such that altogether the whole 27 are contained within a replication, and such that the confounding will only affect ABC. This is done by cyclic substitution in the above squares, i.e. by replacing 0 by 1, 1 by 2 and 2 by 0. We get in succession
These three squares define the contents of the three blocks of a replication, and it is easy to verify that all 27 combinations are present, and that the conditions attached to each set of 9 are satisfied. It is not so easy to explain just exactly what has been confounded, but perhaps at this stage it will be enough to say that, out of the 8 D.F. for second-order interaction ABC, a particular 2 are represented by the differences of the above three block totals and have therefore been confounded. The remaining 6 D.F. of the ABC interaction are unconfounded.

In setting up a \( (3^3) \) design of this sort, starting with Latin square number 1, we would allot the squares 1, 2 and 3 at random to the three blocks of a replication, and then within each block would assign the nine treatment-combinations which belong to that block at random. For additional complete replications the same procedure is repeated. This is on the assumption that the second-order interaction ABC is negligible, so that we can sacrifice completely the chosen part of it. If not negligible, it is desirable to modify the procedure in a way which will shortly be explained. To summarize, taking the Latin squares 1, 2 and 3 in turn, and reading along the rows, row by row, we have the following allocation of treatment-combinations to be randomized within the three blocks of a replication:
Statistical Analysis

We have already explained how a single replication of a \((3^3)\) factorial design could be analysed by making an error out of all the interactions except the \((L L)\) parts of \(AB, AC\) and \(BC\). This left 17 D.F. for error. If the present arrangement be adopted we can have some elimination of soil heterogeneity between blocks, and the procedure is to calculate the blocks S.S., with 2 D.F., in the ordinary way and subtract from the \(ABC\) interaction, leaving an error term with altogether 15 D.F.

If, however, the experiment is replicated, there is no need to obtain an error from the treatment interactions. With 2 replicates, for example, each fully randomized, we should have the total D.F. of 53 broken up into 1 for replicates, 26 for treatments and 26 as a direct measure of error. The 26 for treatments consist of 2 each for \(A, B\) and \(C\); 4 each for \(AB, AC\) and \(BC\); and 8 for \(ABC\). If now each replicate is broken up into three blocks as described, all we have to do is to set out the block totals for the replicates in a 3x2 table as under:

<table>
<thead>
<tr>
<th>Reps.</th>
<th>I</th>
<th>II</th>
<th></th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
<td>c</td>
<td>a</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The S.S. for this table, with 5 D.F., gives the total of block differences for eliminating soil heterogeneity. We then work out the treatments S.S., with 26 D.F., breaking this up into its parts, of which the last will be the second-
order interaction, with 8 D.F.. From the above (3x2) table is then calculated

the blocks S.S., with 2 D.F., from the column totals, i.e. from block totals
added for the two replications. Note that Block 1 in both replications refers
to the blocks based on Latin square 1, and so on for 2 and 3. This S.S. represents
the confounded part of the ABC S.S. and is therefore subtracted to leave an
unconfounded ABC effect, with 6 D.F.. The error S.S. may now be found by
difference from the total S.S. (53 D.F.), and the complete sketch analysis of variance
is as follows:

<table>
<thead>
<tr>
<th></th>
<th>D.F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>blocks</td>
<td>5</td>
</tr>
<tr>
<td>A</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
</tr>
<tr>
<td>AB</td>
<td>4</td>
</tr>
<tr>
<td>AC</td>
<td>4</td>
</tr>
<tr>
<td>BC</td>
<td>4</td>
</tr>
<tr>
<td>ABC (unconf.)</td>
<td>6</td>
</tr>
<tr>
<td>Error</td>
<td>24</td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
</tr>
</tbody>
</table>

If we have three replicates instead of 2, the analysis will be the same in
form, except that there will be 8 D.F. for blocks and 48 D.F. for error. With
four replicates the D.F. for blocks are 11, and for error 72. And so on.
Design of Experiments - Lecture 16

If, now, the second order interaction ABC is not assumed negligible, it is preferable to resort to partial confounding in order to get some information, at least, on the whole of the ABC interaction, instead of only on three-quarters of it. To do this, we note that there are altogether 12 Latin squares of size 3x3, and that the three we have illustrated form set no. I, being generated from one another, as shown, by cyclic substitution. The first square, no. 1, is shown below as a typical member of set I, and with it are shown typical members of sets II, III and IV. (The a and b borders are to be understood as going along with all of them).

<table>
<thead>
<tr>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2</td>
<td>0 2 1</td>
<td>0 1 2</td>
<td>0 2 1</td>
</tr>
<tr>
<td>1 2 0</td>
<td>2 1 0</td>
<td>2 0 1</td>
<td>1 0 2</td>
</tr>
<tr>
<td>2 0 1</td>
<td>1 0 2</td>
<td>1 2 0</td>
<td>2 1 0</td>
</tr>
</tbody>
</table>

So far we have used set I, which confounded a particular 2 of the 8 D.F. for ABC. But we could equally well have used set II, and confounded a different 2 out of the 8 D.F.. Similarly for set III or set IV. The four sets are independent of one another, and in fact the ABC interaction S.S. can be shown to be made up of 4 parts (i) the S.S. for the differences of the block totals in a set-up derived by using the three squares of set I, (ii) a similar S.S. for the block differences derived from the three squares of set II, (iii) the blocks S.S. from the three squares of set III and (iv) the blocks S.S. from set IV.

A partially confounded experiment can therefore be devised in 2 replicates, by using one of the above four sets in one replication, and a second in the other. If we choose I and II, we shall add to the make-up of the first replication, already given in detail, the following allocation of treatment-combinations in the second replication:
With such an experiment there are 4 out of the 8 D.F. for ABC which are unconfounded; while of the remainder 2(I) are confounded in the first replicate and 2(II) in the second (the brackets are used to distinguish these particular components). The make-up of the analysis of variance can be settled by writing it out separately for the two replications and then combining, as is done in the following table:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$AB$</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>$AC$</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>$BC$</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>ABC (unconf.)</td>
<td>${4$</td>
<td>${4$</td>
</tr>
<tr>
<td>Blocks (conf. ABC)</td>
<td>$2(II)$</td>
<td>$2(I)$</td>
</tr>
<tr>
<td>Reps. Error</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Total 26 26 53

Single factors

1st order interactions

unconfounded $ABC$

partially conf. ABC
Statistical Analysis. From Rep. I we may calculate the total S.S. (26 D.F.) and subtract blocks (2(I) D.F.), leaving the S.S. within blocks (24 D.F.). This will contain a component (2(II) D.F.) for the part of ABC that is confounded in Rep. II, and this is calculated separately by making up three new totals corresponding to the block totals of rep. II. Similar calculations made from Rep. II (in the last case with the aid of the set-up for Rep. I) determine blocks (2(II) D.F.), the S.S. within blocks (24 D.F.) and the component (2(I) D.F.) for the part of ABC that is confounded in Rep. I. We now form the 27 treatment totals for the combined replications. In the ordinary way we determine the S.S. for A, B, C, AB, AC and BC (altogether 18 D.F.). To this we add the S.S. for the partially confounded parts of ABC (2(II) + 2(I) D.F.), and then calculate the S.S. for the unconfounded part of ABC (4 D.F.). This last calculation can be made in two parts by making up three totals for reps. I + II with the aid of (a) the set III and (b) the set IV of Latin squares. We have now accounted for 26 out of the 48 D.F. for the total S.S. within blocks in replications I and II, and the difference provides the error S.S., with 22 D.F.

This description of the analysis follows the way in which the A.V. was built up above from the two replications, and we have accounted for everything except the S.S. for replications, easily calculated (if required) from the 2 rep. totals. Other systematic methods of conducting the analysis have been described. See, for example, Yates, The Design and Analysis of Factorial Experiments p. 42. It is recommended that a suitable example be worked through from start to finish so that you may familiarize yourselves with the working procedure.

It is left as an exercise for the student to make up the analysis for three replications, in the third of which the set III of Latin squares is used. He will find that the six component D.F. for two replications change into 6, 12, 2, 6 (2(I) + 2(II) + 2(III)), 8 and 46 respectively, totalling 80. If a fourth replication be added, using the set IV of Latin squares, the design becomes a
balanced one, for all four parts of the \( ABC \) interaction are confounded in one replication and unconfounded in the other three, so that 3/4 information is obtained on the \( ABC \) interaction. The sketch analysis of variance is briefly set out below:

<table>
<thead>
<tr>
<th>D.F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single factors</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>1st order interactions</td>
</tr>
<tr>
<td>( ABC ) (part. conf.)</td>
</tr>
<tr>
<td>Blocks</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>Error</td>
</tr>
<tr>
<td>70</td>
</tr>
<tr>
<td>Total</td>
</tr>
<tr>
<td>107</td>
</tr>
</tbody>
</table>

Normally, of course, the first two lines will be broken up to show the separate effects.

To complete the story we shall add here the blocks of treatments for replications III and IV, corresponding to set III and set IV of the Latin squares:

<table>
<thead>
<tr>
<th>Block 7</th>
<th>Block 8</th>
<th>Block 9</th>
<th>Block 10</th>
<th>Block 11</th>
<th>Block 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>a b c</td>
<td>a b c</td>
<td>a b c</td>
<td>a b c</td>
<td>a b c</td>
<td>a b c</td>
</tr>
<tr>
<td>0 0 0</td>
<td>0 0 0</td>
<td>0 0 2</td>
<td>0 0 0</td>
<td>0 0 1</td>
<td>0 0 2</td>
</tr>
<tr>
<td>0 1 1</td>
<td>0 1 2</td>
<td>0 1 0</td>
<td>0 1 2</td>
<td>0 1 0</td>
<td>0 1 1</td>
</tr>
<tr>
<td>0 2 2</td>
<td>0 2 0</td>
<td>0 2 1</td>
<td>0 2 1</td>
<td>0 2 2</td>
<td>0 2 0</td>
</tr>
<tr>
<td>1 0 2</td>
<td>1 0 0</td>
<td>1 0 1</td>
<td>1 0 1</td>
<td>1 0 2</td>
<td>1 0 0</td>
</tr>
<tr>
<td>1 1 0</td>
<td>1 1 1</td>
<td>1 1 2</td>
<td>1 1 0</td>
<td>1 1 1</td>
<td>1 1 2</td>
</tr>
<tr>
<td>1 2 1</td>
<td>1 2 2</td>
<td>1 2 0</td>
<td>1 2 2</td>
<td>1 2 0</td>
<td>1 2 1</td>
</tr>
<tr>
<td>2 0 1</td>
<td>2 0 2</td>
<td>2 0 0</td>
<td>2 0 2</td>
<td>2 0 0</td>
<td>2 0 1</td>
</tr>
<tr>
<td>2 1 2</td>
<td>2 1 0</td>
<td>2 1 1</td>
<td>2 1 1</td>
<td>2 1 2</td>
<td>2 1 0</td>
</tr>
<tr>
<td>2 2 0</td>
<td>2 2 1</td>
<td>2 2 2</td>
<td>2 2 0</td>
<td>2 2 1</td>
<td>2 2 2</td>
</tr>
</tbody>
</table>

This form of design may be extended to \( (3^4) \) in blocks of 9 plots, or in general to \( (3^n) \) in blocks of \( 3^{n-1} \) or \( 3^{n-2} \). But perhaps enough has been said to indicate the possibilities. For further special cases see R. A. Fisher: The Design of Experiments, and for a fairly exhaustive treatment of the various possible cases, see F. Yates: The Design and Analysis of Factorial Experiments.
In examining the results of such experiments, with the aid of the various possible standard errors, the ordinary methods are followed so far as direct effects and first-order interactions are concerned. It will not ordinarily be necessary to prepare a summary table to include all three factors (for $3^3$); if, however, such are required certain corrections are required for the average yields because of the confounding of parts of the second-order interaction with the blocks. For these adjustments, and for the calculation of the standard errors of the various differences of adjusted yields, see Yates: "The Design and Analysis of Factorial Experiments" (pp. 46-47 for $3^3$).

**Gain in Accuracy from Confounding**

In dealing with the split-plot design we gave certain formulae which enabled us to measure the efficiency of the design in contrast to one based on full randomization within whole blocks. Similar formulae may now be given to measure the efficiency of confounding. Our object has been to decrease the experimental error, and so it is generally useful to test how successful we have been. What we do is to make an estimate of the experimental error $E_r$ that would exist had the experiment been laid out in completely randomized replications. Let $E_b$ be the M.S. for blocks, and $E_e$ that for error, in the confounded design. With only one replication the estimate of $E_r$ is

$$E_r = \frac{(n_b E_b + n_e E_e)}{(n_b + n_e)}$$

where $n_b$ is the no. of D.F. for blocks and $n_b + n_e$ is the total D.F. in the experiment. $E_r$ may then be directly compared with $E_e$ in any way we please. We are assuming that the interactions which are confounded with blocks are negligible.

With complete confounding in more than one replication, but again assuming the confounded interactions to be negligible, we define $E_b$ as the M.S. for blocks within replications and $n_b$ as the corresponding D.F.. But this time $n_b + n_e$ is
the total D.F. in the experiment less the D.F. for replications. With these conventions the above formula applies. It also applies in partial confounding if all the confounded comparisons are assumed negligible.

With some extra trouble it is possible to avoid making the assumption that the confounded interactions are negligible. The way to do this is to remove treatment effects from the M.S. for blocks within replications. Let us use the letter u for "unadjusted" and the letter a for "adjusted". Then blocks within reps. (u) plus treatments (a), two quantities which appear in the A.V. (provided reps. is taken out of total blocks), must be equal to blocks within reps. (a) plus treatments (u). We must therefore calculate the unadjusted treatments S.S. and subtract from the sum of the two components we have selected from the A.V.

The M.S. calculated from blocks within reps. (a) is denoted by $E_b(a)$. Then a quantity $E_b^1$, which will replace $E_b$ in the formula for $E_r$, is obtained from the formula

$$E_b^1 = \left\{ rE_b(a) - E_e \right\} / (r-1)$$

where $r$ is the number of replications. This method will apply to all partially confounded designs, and $n_b$ and $n_e$ are as previously defined.

**Designs for Large Numbers of Varieties**

The method of confounding has been found useful for factorial designs in order to reduce the effective size of the block, for elimination of fertility differences, by confounding interactions which are unlikely to be of physical significance. But another set of trials are those conducted by the plant breeder, who may be concerned solely to differentiate between the yields and other characters of his varieties, and may have a large number of these. The same
difficulties in regard to soil heterogeneity will occur here as with a factorial
design if the number of varieties to be included makes the replications large
in size, and special methods have been devised to cope with this situation.
The subject is a large one, and there are many forms of design. We shall illustrate
by talking first about lattice designs, and later about other forms of incomplete
block designs. Broadly, the position in such cases is that we can imagine a
factorial arrangement as being super-imposed upon the varieties to be compared,
in order to utilise the principles of confounding. But since all varieties are
of equal standing, so to speak, in the experiment, it will not matter, in the
artificial factorial design imposed upon them, whether it is main effects or
high-order interactions which are confounded. Some confounding has to take place
if the replications are to be divided into a number of incomplete blocks, and it
is the nature of the confounding that takes place which determines the design.
Where possible the confounding should be balanced in order to secure equality
of the varieties in regard to the various "effects" that can be measured.

A square lattice is a design in which \( p^2 \) varieties, forming one complete
replication, are divided into \( p \) blocks of \( p \) plots each. It is also possible to
have a rectangular lattice using \( p \times q \) varieties, which extends the range of
the design, but here we shall restrict ourselves to square lattices, which
without ambiguity may simply be called lattices. To illustrate in a very simple
case take \( p = 2 \). We want to arrange the varieties 1, 2, 3 and 4 so that there are
two blocks per replication. Consider the Latin square

\[
\begin{array}{cc}
A & B \\
B & A \\
\end{array}
\]

for which the 3 D.F. (4-1) are broken up into 3 orthogonal components, with 1
D.F. each, one for rows, one for columns and one for "treatments", i.e. the
difference between letters. Associate this square with the 4 varieties, thus:
By taking rows, we divide up the 4 varieties into 2 blocks, namely the block 1+2 and the block 3+4. This is one possible arrangement of a replication: in a line of 4 plots we decide at random whether 1 and 2, or 3 and 4, will go on the first two plots of the 4, and the two blocks mentioned then become the two incomplete blocks of this replication. Within the blocks we then decide whether 1 or 2, or 3 or 4, will come first.

A second grouping arises from taking columns, so that another replication will contain the blocks 1+3 and 2+4. The Simple Lattice design is one consisting of the above two replications, the one having a grouping by rows to decide the blocks, and the other a column grouping. This applies to any number of varieties which is a perfect square. With \( p^2 \) varieties all we have to do is arrange them in a square \( p \times p \) table, which decides the make-up of the row and column groupings.
Design of Experiments - Lecture 17

When the number of varieties is not very large, it is desirable to repeat the above pattern, which is done as a whole. As a first stage, therefore, we would pass from 2 to 4 replications, with a duplication of the grouping by rows and similarly a duplication of the grouping by columns. On the experimental area, laid out to accommodate 4 complete replications, we would decide which two replications were to have the grouping by rows, when the others would be assigned the grouping by columns. This being done, the arrangement of the blocks within replications is done at random, and likewise the arrangement of the plots within the block.

Returning now to the case of 4 varieties, a third grouping will arise by taking the letters of the Latin square. In this case a replication will contain the blocks 1 + 4 and 2 + 3. This comparison will be recognized as the interaction of rows and columns, as would emerge if we had had a $2^2$ factorial design. The three groupings, by rows, by columns and by letters, are orthogonal to one another. This exhausts the possibilities, for we have used up all three of the D.F. of the 2 x 2 Latin square. A design using three replications, grouping by rows, columns and letters, is therefore balanced for the case of 4 varieties, and is then called a balanced lattice. Such a design can, of course, be duplicated to provide 6 replications in all.

But with 9, or 16, or 25 etc. varieties the position is different. With $p^2 = 9$, $p = 3$, for example, there are four orthogonal comparisons (2 D.F. each) which can be obtained by writing down the Graeco-Latin square

\[
\begin{array}{ccc}
A \alpha & B \beta & C \gamma \\
B \gamma & C \alpha & A \beta \\
C \beta & A \gamma & B \alpha
\end{array}
\]

which may be obtained by putting together the two orthogonal 3 x 3 Latin squares given in Table XVI of Fisher and Yates: "Statistical Tables." This is a square which has the Latin square property for the Greek as well as the Latin letters;
in addition each Greek letter appears once with each Latin letter. The four groupings are by rows, columns, Latin letters and Greek letters.

If we use 3 replications with 9 varieties, choosing groupings by rows, columns and Latin letters, we only use three out of the possible 4 groupings. Such a design is called a **triple lattice**, a name which will apply to any such three-replicate experiment where \( p \) is greater than 2 (for with 4 varieties a triple lattice is a balanced lattice). A triple lattice may be duplicated to provide twice the number of replications. When all 4 groupings are used in a 4-replicate experiment with 9 varieties, this in turn becomes a balanced lattice.

With \( p^2 = 16, p = 4 \), the 15 D.F. are divided into 5 comparisons of 3 D.F. each. These correspond to groupings by rows, columns, Latin letters, Greek letters in square (1), and Greek letters in square (2), of the two possible Graeco-Latin squares shown below:

\[
\begin{array}{cccc}
A \alpha & B \beta & C \gamma & D \delta \\
B \gamma & A \delta & D \alpha & C \beta \\
C \delta & D \gamma & A \beta & B \alpha \\
D \beta & C \alpha & B \delta & A \gamma
\end{array}
\]

\[
\begin{array}{cccc}
A \alpha & B \beta & C \gamma & D \delta \\
B \beta & A \alpha & D \gamma & C \delta \\
C \delta & D \beta & A \gamma & B \alpha \\
D \gamma & C \delta & B \alpha & A \beta
\end{array}
\]

These are obtained from the three orthogonal 4 x 4 Latin Squares given in Fisher and Yates: Table XVI by (1) taking I with II and (2) taking I with III. Note that the Greek letters in square (2) are obtained from those in square (1) by permuting the rows other than the first, the Latin letters meantime being held unchanged.

It is therefore possible to have up to 5 replications of 16 varieties with each replication corresponding to a different grouping. 2 replicates gives a simple lattice, 3 a triple lattice, 4 a quadruplo lattice and 5 a balanced lattice.

In general one may work out similar problems in full for a number of varieties which is the square of a prime number, or of a power of a prime. In other cases
e.g., with 36 varieties, one cannot proceed beyond a triple lattice, for there is no 6 x 6 Graeco-Latin square. Table XVI of Fisher and Yates lists the orthogonal squares which are to be used, up to 9 x 9. With small numbers of varieties a balanced design may not give enough replication, and the arrangement is then repeated one or more times, with fresh randomization. With large numbers a balanced design would involve too many plots, and the advantages of such a design are usually sacrificed in order to lay down a simple, triple or quadruple lattice, or a repeated simple or triple, as the case may be.

Before giving specific examples of the make-up of the lattices, it is desirable, in order to distinguish between different kinds of lattices, to mention briefly certain variations into which we shall not have time to go. First note that it would be possible to array a replication of 9 varieties in a 3 x 3 square pattern instead of in 3 blocks of three, thus enabling elimination of both rows and columns to take place within the replication, on the Latin square analogy, instead of just block differences. Now we have seen that there are 4 possible groupings of the 9 varieties. We may therefore arrange one replication so that rows correspond to one grouping and columns to a second, and then introduce the third and fourth groupings to decide the rows and columns of a second replication. Thus with 2 replications each grouping is confounded once and once only in the way indicated, the design is balanced, for equal information is obtained on every pair of D.F., and all varietal comparisons are made with equal accuracy.

Such a design is known as a lattice square, a name which must be distinguished from square lattice. With p a suitable odd number, the number of replications required for a balanced design is \( \frac{1}{2}(p + 1) \), whereas a balanced lattice would require p + 1 replications. Even when p is small it may pay to choose a lattice square and then duplicate or triplicate it.

The other variation is where the number of varieties is a perfect cube, and we wish to arrange \( p^3 \) varieties into blocks. We saw how a simple lattice was a
2 replication arrangement, the blocks in one corresponding to the rows, and in
the other to the columns, of the square \( p \times p \) array. A \( p \times p \times p \) array will be a
three-dimensional or cubic array, and could be arranged in blocks in three ways
representing to three sets of lines or planes in a cube which are mutually at
right angles to one another. What we then get is a three-replicate design called a
cubic lattice, which will be the simplest of the three-dimensional designs, increased
replication in which will eventually lead to a balanced design. Other variations
are possible, but the important thing for the student to grasp is the difference
between the lattice (which is usually square, and may be simple, triple, etc., or
balanced) and the lattice square on the one hand, and the cubic lattice on the
other (or the two combined). If he understands the lattice, and is able to analyze
the results of lattice experiments, he will probably be able to worry out for
himself more complex arrangements.

**Layouts in simple cases.** \( p = 2 \). Let the varieties be numbered 1, 2, 3 and 4.

<table>
<thead>
<tr>
<th>rep I</th>
<th>rep II</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 2</td>
<td>1 4</td>
</tr>
<tr>
<td>4 1</td>
<td>3 2</td>
</tr>
</tbody>
</table>

Specimen simple lattice

<table>
<thead>
<tr>
<th>Block</th>
<th>Block</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2</td>
<td>3 4</td>
</tr>
</tbody>
</table>

Balanced lattice - to above add

<table>
<thead>
<tr>
<th>Block</th>
<th>Block</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 4</td>
<td>3 1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>rep III</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 4</td>
</tr>
<tr>
<td>3 1</td>
</tr>
</tbody>
</table>

\( p = 3 \). Number the varieties from 1 to 9.

<table>
<thead>
<tr>
<th>Rep I</th>
<th>Rep II</th>
<th>Rep III</th>
<th>Rep IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 4 7</td>
<td>1 9 2</td>
<td>2 5 8</td>
<td>4 5 2</td>
</tr>
<tr>
<td>1 6 8</td>
<td>7 3 8</td>
<td>4 7 1</td>
<td>8 9 6</td>
</tr>
<tr>
<td>3 5 9</td>
<td>4 6 5</td>
<td>9 3 6</td>
<td>3 1 7</td>
</tr>
</tbody>
</table>

Blocks 1-3 Blocks 4-6 Blocks 7-9 Blocks 10-12
Reps I and II give simple lattice, reps I-III a triple lattice and reps I-IV a balanced lattice.

A specimen lattice square will be

```
Rep I     Rep II
\begin{array}{ccc}
6 & 4 & 5 \\
1 & 2 & 3 \\
8 & 9 & 7 \\
\end{array}
\begin{array}{ccc}
1 & 4 & 7 \\
9 & 3 & 6 \\
5 & 8 & 2 \\
\end{array}
```

Rep I confounds rows and Latin letters, rep II columns and Greek letters.

\[ p = 4 \]. Number the varieties from 1 to 16.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9  2</td>
<td>15  4</td>
<td>16  4</td>
<td>11  16</td>
<td>16  15</td>
</tr>
<tr>
<td>11  1</td>
<td>7  16</td>
<td>6  10</td>
<td>8  3</td>
<td>2  10</td>
</tr>
<tr>
<td>10  3</td>
<td>11  8</td>
<td>1  13</td>
<td>13  10</td>
<td>9  8</td>
</tr>
<tr>
<td>12  4</td>
<td>3  12</td>
<td>11  7</td>
<td>2  5</td>
<td>7  1</td>
</tr>
<tr>
<td>6  14</td>
<td>1  14</td>
<td>5  14</td>
<td>9  12</td>
<td>5  6</td>
</tr>
<tr>
<td>8  16</td>
<td>13  2</td>
<td>2  8</td>
<td>6  7</td>
<td>4  12</td>
</tr>
<tr>
<td>7  13</td>
<td>5  10</td>
<td>15  9</td>
<td>4  1</td>
<td>14  13</td>
</tr>
<tr>
<td>5  15</td>
<td>9  6</td>
<td>12  3</td>
<td>15  14</td>
<td>11  3</td>
</tr>
</tbody>
</table>

Blocks 1-4 Blocks 5-8 Blocks 9-12 Blocks 13-16 Blocks 17-20

Reps I and II give simple lattice; reps I-III a triple lattice; reps I-IV a quadruple lattice and reps I-V a balanced lattice.

A balanced set of lattice squares for 25 varieties is given in Yates "Design and Analysis, etc." p. 88.
Example of cubic lattice. 3³ varieties, numbered 1-27, in 9 blocks of 3 plots.

Arrangement before randomization of blocks and block contents.

<table>
<thead>
<tr>
<th>Rep I</th>
<th>Rep II</th>
<th>Rep III</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>17</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>18</td>
<td>11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Blocks 1-9</th>
<th>Blocks 10-18</th>
<th>Blocks 19-27</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A balanced design would require 13 replications.

**Statistical Analysis.** A full analysis is a somewhat involved procedure, especially when an attempt is made to recover inter-block information. But under certain circumstances this is not necessary so far as the F-test is concerned. An outline sketch will be given in a particular case which derives from the analysis previously given in the case of partially confounded designs. The object is to show the make-up. In practice, the analysis proceeds differently.

Let us consider the case of 9 varieties in 3 blocks of 3 plots each per replication. The 8 D.F. for varieties split up into four pairs, which may be designated 2₁, 2₂, 2₃, 2₄, and which derive from the four ways (rows, columns, Latin letters, Greek letters) of taking the varieties in 3 blocks of 3. With a particular replication one of these effects is confounded with blocks. Consider the **simple lattice**, where the 2₁ effect is confounded in the first replication and the 2₂ effect in the other, and where therefore the “interaction” of rows and columns, with 4 D.F. (2₃ + 2₄) is unconfounded in both. The make-up of the D.F. for the separate replications, and for the two together, is indicated in the table below:
In the final column we see that we have a total of 8 D.F. for varieties (a), i.e. varieties adjusted for blocks, together with 4 D.F. for blocks (u), i.e. blocks within reps unadjusted for varieties, also reps (1 D.F.), and finally 4 D.F. for error, this latter coming from the interaction of the completely unconfounded variety effects (4 D.F.) with reps (1 D.F.). This has determined the constitution of the error component, but, as we shall see later, a simple lattice will normally be analyzed differently, and more simply. But one point may be made now. We saw in the last lecture that from the point of view of discussing the efficiency of confounding, it was desirable to calculate blocks (a), i.e. adjusted for varieties, by subtracting varieties (u) from the sum of varieties (a) and blocks (u). The same need arises here when we wish to recover inter-block information. If, then, we analyze as shown, we have only to calculate varieties (u), with 8 D.F., in the ordinary way, to determine blocks (a). An example will shortly be given of the normal computational procedure.

To improve the precision of the error determination, a simple lattice such as this may be duplicated, in which case the make-up, conducted on the above lines, will be as below:

<table>
<thead>
<tr>
<th></th>
<th>Rep I</th>
<th>Rep II</th>
<th>Rep I + Rep II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Varieties (unconf.)</td>
<td>(2_{II} ) 6</td>
<td>(2_I ) 6</td>
<td>part. conf. (2_{II} ) 4</td>
</tr>
<tr>
<td>Blocks</td>
<td>(2_I )</td>
<td>(2_{II} )</td>
<td>unconf. (2_I ) 4</td>
</tr>
<tr>
<td>Reps.</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Error</td>
<td>-</td>
<td>-</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>8</td>
<td>17</td>
</tr>
</tbody>
</table>

In the final column we see that we have a total of 8 D.F. for varieties (a), i.e. varieties adjusted for blocks, together with 4 D.F. for blocks (u), i.e. blocks within reps unadjusted for varieties, also reps (1 D.F.), and finally 4 D.F. for error, this latter coming from the interaction of the completely unconfounded variety effects (4 D.F.) with reps (1 D.F.). This has determined the constitution of the error component, but, as we shall see later, a simple lattice will normally be analyzed differently, and more simply. But one point may be made now. We saw in the last lecture that from the point of view of discussing the efficiency of confounding, it was desirable to calculate blocks (a), i.e. adjusted for varieties, by subtracting varieties (u) from the sum of varieties (a) and blocks (u). The same need arises here when we wish to recover inter-block information. If, then, we analyze as shown, we have only to calculate varieties (u), with 8 D.F., in the ordinary way, to determine blocks (a). An example will shortly be given of the normal computational procedure.

To improve the precision of the error determination, a simple lattice such as this may be duplicated, in which case the make-up, conducted on the above lines, will be as below:
The variety totals for all four reps enable us to get out the S.S. (for

$2_II$ and $2_I$ D.F.) for the partly confounded parts, leaving a residue for interaction of these components with reps (4 D.F.) which goes to error. Equally, the

"interaction" in the combined table (4 D.F.) can be calculated, leaving a residue with 4 D.F. to go to error. The error is then made up of these two residues, plus the S.S. already determined from the two pairs of reps (4 D.F. each), and has thus 16 D.F. in all.

Here, again, varieties (u) can easily be determined from the 9 variety totals, and blocks (a) thus determined by difference. As we shall see from an example, such an experiment may be first treated as a completely randomized one by absorbing blocks (a) into the error. In such a case all we have to do is calculate S.S. for varieties (u), reps and total, and obtain error by difference. If varieties are clearly significant on the F-test, or if no reasonable reduction of error would make them significant anyhow, we need go no further with the analysis of variance. But if the varieties M.S. is not significant, although large compared to error, this may be because the error is inflated by block differences which it is desirable to remove. The procedure to follow is then as indicated in the example which will now be worked through.
The theory behind the analysis of simple and triple lattice experiments is given in Research Bulletin 281 of the Iowa State College Agricultural Experiment Station (1940). It is fully illustrated by an experiment on 81 varieties of corn, of the duplicated simple lattice pattern in one case, and of the duplicated triple lattice pattern in another. To make an easy example I have chosen the first 9 varieties only, numbered 1-9, and have made them up into blocks of rows (group x) and by columns (group y). Replicates 1 and 2 are duplicates in group x and replicates 3 and 4 are duplicates in group y. The data are assembled below:

### Table 1.

<table>
<thead>
<tr>
<th>Group x</th>
<th>Rep 1</th>
<th></th>
<th>Rep 2</th>
<th></th>
<th>Block totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19.5</td>
<td>23.3</td>
<td>24.1</td>
<td></td>
<td>66.9</td>
</tr>
<tr>
<td>2</td>
<td>23.7</td>
<td>24.0</td>
<td>25.8</td>
<td></td>
<td>73.5</td>
</tr>
<tr>
<td>3</td>
<td>24.5</td>
<td>27.1</td>
<td>21.2</td>
<td></td>
<td>72.8</td>
</tr>
</tbody>
</table>

\[ \text{Block totals} = 213.2 \]

\[ \text{S.S. (8 D.F.)} = 40.9089 \]

<table>
<thead>
<tr>
<th>Group y</th>
<th>Rep 3</th>
<th></th>
<th>Rep 4</th>
<th></th>
<th>Block totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20.0</td>
<td>28.5</td>
<td>30.4</td>
<td></td>
<td>78.9</td>
</tr>
<tr>
<td>2</td>
<td>25.6</td>
<td>25.2</td>
<td>32.4</td>
<td></td>
<td>83.2</td>
</tr>
<tr>
<td>3</td>
<td>25.6</td>
<td>28.4</td>
<td>25.7</td>
<td></td>
<td>79.7</td>
</tr>
</tbody>
</table>

\[ \text{Block totals} = 241.8 \]

\[ \text{S.S. (8 D.F.)} = 102.6200 \]

\[ \text{S.S. (8 D.F.)} = 79.4800 \]

\[ \text{S.S. between reps (3 D.F.)} = 397.8342 \]

\[ \text{Total S.S. within reps (32 D.F.)} = 295.1089 \]

\[ \text{Total S.S. (35 D.F.)} = 602.9431 \]
Table 2. Totals for groups x and y.

<table>
<thead>
<tr>
<th>Group x</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>row totals</th>
<th>Group y</th>
<th>1</th>
<th>4</th>
<th>7</th>
<th>row totals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>45.9</td>
<td>54.3</td>
<td>51.2</td>
<td>151.4</td>
<td></td>
<td>48.3</td>
<td>57.8</td>
<td>60.4</td>
<td>166.5</td>
</tr>
<tr>
<td></td>
<td>168.6</td>
<td>50.7</td>
<td>57.0</td>
<td>65.0</td>
<td></td>
<td>172.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>165.6</td>
<td>60.1</td>
<td>63.2</td>
<td>58.6</td>
<td></td>
<td>181.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col. totals</td>
<td>149.9</td>
<td>173.1</td>
<td>162.6</td>
<td>485.6</td>
<td></td>
<td>159.1</td>
<td>178.0</td>
<td>184.0</td>
<td>521.1</td>
</tr>
</tbody>
</table>

Table 3. Variety totals

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>94.2</td>
<td>105.0</td>
<td>111.3</td>
<td>109.5</td>
<td>115.9</td>
<td>121.2</td>
<td>112.7</td>
<td>124.9</td>
<td>112.0</td>
<td>1006.7</td>
</tr>
</tbody>
</table>

Variety S.S. (8 D.F.) = 159.7854

Blocks within Reps.

Here the S.S. can be computed from (i) the differences between the corresponding block totals, in Reps 1 and 2 for group x, and in Reps 3 and 4 for group y, and (ii) the sums of those same corresponding block totals. The Blocks S.S. will be the sum of these two components, but in the case of (ii) it will be necessary to adjust for the fact that the blocks are composed of different varieties. We therefore make the following calculations:

Component (i). From Table 1, group x, the block total differences are 17.6, 21.6 and 20.0, total 59.2. For group y they are 8.7, 6.3 and 22.5, total 37.5. The S.S. of deviations of these sets (divided by 6) are 1.3511 and 25.48, yielding 26.8311 for component (i) S.S. (4 D.F.)

Component (ii). In Table 2, the three differences between row totals (group x) and column totals (group y), which contain the same varieties, are -7.7, -9.4 and -18.4, total -35.5. The corresponding differences between row totals (group y) and column totals (group x) are 16.6, -0.4 and 19.3, total 35.5. The S.S. of deviations of these sets (divided by 12) are 5.5106 and 19.0105, yielding 24.5211 for component (ii) S.S. (4 D.F.).
The error S.S. is obtained by difference, and we have the following Analysis of Variance.

<table>
<thead>
<tr>
<th></th>
<th>S.S.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicates</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Component (i)</td>
<td>4</td>
<td>26.8311</td>
</tr>
<tr>
<td>Component (ii)</td>
<td>4</td>
<td>24.5211</td>
</tr>
<tr>
<td>Blocks (eliminating varieties)</td>
<td>8</td>
<td>51.3522</td>
</tr>
<tr>
<td>Varieties (ignoring blocks)</td>
<td>8</td>
<td>159.7856</td>
</tr>
<tr>
<td>Error (intra-block)</td>
<td>16</td>
<td>83.9711</td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td>602.9431</td>
</tr>
</tbody>
</table>

The first (approximate) test is to treat the experiment as one in randomized complete blocks. Putting Blocks S.S. with Error S.S. we have a M.S. of 5.6385, with 24 D.F. Dividing into 19.9732 we find F to be 3.54, significant at the 1 percent point. In view of this it is not necessary to go further. Should, however, F have failed to show significance, the further procedure would have been as follows: Use the row totals in Table 2 to obtain a S.S. for component (ii) of blocks (unadjusted for varieties), i.e. work out the S.S. between rows in group x (dividing by 6), giving 28.1377, and similarly for group y, giving 20.0133. The total (for 4 D.F.) is 48.1510, which we may call $B_u$. The S.S. for component (ii), adjusted, was 24.5211, and this we call $B_a$. We now require the weighting factor

$$\lambda = \frac{w - w'}{w + w'}$$

where $w = 1/E = 0.19054$

and $w' = 3/4B - E = 0.14686$

(for B and E see above A.V. table). So $\lambda = 0.12946$.

Then the total change in S.S. for varieties (a) to adjust for block effects is

$$- \lambda \left[ (1 + \frac{w'}{w})B_u - B_a \right] = -7.8638$$

Thus the S.S. for varieties (a) is 159.7856 - 7.8638 = 151.9218, giving a M.S. of 18.9902. Comparing this with E gives an F of 3.62, with 8 and 16 D.F.

To obtain corrected variety means, multiply the six differences calculated above under Blocks (component (ii)) by $-\lambda /12$, i.e. by -0.0108. Border a table
of the unadjusted variety means with these corrections \( c_x \) and \( c_y \).

<table>
<thead>
<tr>
<th>Variety means (u)</th>
<th>( c_x )</th>
<th>Variety means (a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3</td>
<td>23.55 26.25 27.82</td>
<td>1 2 3</td>
</tr>
<tr>
<td>4 5 6</td>
<td>27.38 28.98 30.30</td>
<td>4 5 6</td>
</tr>
<tr>
<td>7 8 9</td>
<td>28.18 31.22 28.00</td>
<td>7 8 9</td>
</tr>
</tbody>
</table>

\[
c_y = -0.179 + 0.004 - 0.208
\]

To any value in the table on the left add the corrections in the same row and column, obtaining variety means (a) as on right.

The variance of the difference between two variety means in the same block is

\[
\frac{E}{3} \left( \frac{-u}{w + \bar{w}} + 1 \right)
\]

and between two variety means in different blocks is

\[
\frac{E}{6} \left( \frac{4u}{w + \bar{w}} + 1 \right).
\]

Without appreciable error, the mean of these two results can usually be used for all comparisons, namely

\[
\frac{E}{4} \left( \frac{2u}{w + \bar{w}} + 1 \right).
\]

These formulae only apply to the special case illustrated. The mathematical theory, including general formulae, is outlined on page 45 onward of Research Bulletin 281 already referred to.

Other Balanced Incomplete Block Designs.

A feature common to the lattice designs we have been studying is that each replication consists of a number of incomplete blocks. For example, with 9 varieties, we saw how one replication could have blocks consisting of 1 + 2 + 3, 4 + 5 + 6, 7 + 8 + 9; another would contain 1 + 4 + 7, 2 + 5 + 8, 3 + 6 + 9; a third 1 + 6 + 8, 2 + 4 + 9, 3 + 5 + 7; and a fourth 1 + 5 + 9, 2 + 6 + 7, 3 + 4 + 8.
With all four replications the design is balanced. Every pair of treatments will be found to occur once, and once only, in the same block. We now consider other cases, with different numbers of treatments, and of plots per block, where balanced designs can be constructed. We have had the case of \( p^2 \) varieties in \( p(p+1) \) blocks of \( p \) plots each, derivable from the rows, columns, and \( p - 1 \) orthogonal Latin squares of size \( p \times p \). With this case goes another in which \( p^2 + p + 1 \) varieties can be arranged in \( p^2 + p + 1 \) blocks, each of \( p + 1 \) plots.

All we have to do here is to arrange \( p^2 \) of the varieties in the usual way, then add the same new variety to each of the blocks of the first replication, a second to each of the blocks of rep 2, and so on, finally adding a last block consisting of the extra \( p + 1 \) varieties. For example, with \( p = 2 \), we have for four varieties replications, I, II and III, each of 2 blocks, as follows:

<table>
<thead>
<tr>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

From this we derive the following seven incomplete blocks containing varieties 1-7, in which every pair of varieties will occur once, and once only, in the same block:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1</th>
<th>3</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1</th>
<th>4</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
<td>7</td>
</tr>
</tbody>
</table>

| 5 | 6 | 7 |

In the same way our illustration of 9 varieties in 12 incomplete blocks, 3 to each replication, could be turned, by the addition of varieties 10, 11, 12 and 13, into an experiment with 13 blocks each consisting of 4 varieties. In all such cases the numbers should be assigned at random to the varieties, as also the \( p^2 + p + 1 \) sets to the successive blocks, within which the \( p + 1 \) varieties should also be randomized.

This is not, however, the only way to enlarge the field of incomplete block designs. The subject has been studied by a number of authorities, and a list of
known solutions is given in Tables XVII - XIX of Fisher and Yates, Statistical Tables. The introduction to those Tables (p. 17 on) reviews the position, and illustrates the analysis of such experiments.

Let $k$ be the number of experimental units per block, less than $v$, the number of varieties. Then if every two varieties occur together in the same number ($\lambda$) of blocks, block differences can be eliminated in a simple manner. Let $r$ be the number of replicates, and $b$ the number of blocks. Then the number of experimental units (plots) is $rv = bk$, while $\lambda = r(k - 1)/(v - 1)$, (unity in the cases so far considered). We could take all possible combinations of the $v$ varieties, $k$ at a time, but this would lead to a large number of replicates, if $v$ is large, and attention has therefore been directed to finding suitable arrangements without too much replication. For example, with 3 replicates we can have 4 varieties in either 6 or 4 blocks, or 7 varieties in 7 blocks. With 4 replicates we can have 5 varieties in 10 or 5 blocks, 7 varieties in 7 blocks, 9 in 12 blocks or 13 in 13 blocks. And so on. One can study 16 or 21 varieties in 5 replicates, and as many as 31 in 6 replicates.

When $b = v$ it is possible, at least in the cases of practical importance, to arrange the order of the varieties in the blocks so that each variety occurs once in each position in the block; it is then possible to eliminate experimental differences between these positions, each of which comprises a complete replication. Such a scheme gives rise to what are known, from the discoverer, as the Youden Squares. Suppose we wished to test 13 treatments on 13 plants, each with 4 corresponding leaves. We could assign the material so that each treatment occurs once on each type of leaf. At the same time the 13 plants constitute randomized incomplete blocks of 4 treatments each, in which each treatment occurs on the same plant once with every other treatment. Similarly a varietal trial of 13 varieties in blocks of 4 could be arranged on the ground in a $4 \times 13$ rectangle of plots so that both row and column differences can be eliminated from the varietal comparisons.
To obtain such an experimental arrangement, we make up suitable blocks of 4 numbers out of the numbers 1-13, assign these blocks to the plants, or on the ground, at random, at the same time assigning the experimental treatments (or the varieties) at random to the numbers 1-13.
Block differences may be entirely eliminated from the varietal comparisons, or alternatively the information given by these comparisons may be combined with that provided by the intra-block comparisons, with weights depending on the relative accuracy of the two sets of comparisons (as illustrated in the Iowa Corn experiment).

Let $V_s =$ sum of yields of variety $s$

$T_s =$ sum of block totals of blocks containing variety $s$

$G =$ total yield of all plots.

The estimates of varietal differences derived from intra-block comparisons are obtained from

$$Q_s = V_s - T_s / k$$

Then the differences of the quantities $Q_s / e$ give the actual differences in units of the total yield of the $r$ replicates, where $e$ is the efficiency factor $v \lambda / rk$.

They should sum to zero. Their error variance is $r / ew$ where $w$ is the reciprocal of the intra-block error variance.

Similarly the estimates of varietal differences derived from the inter-block comparisons are given by the differences of

$$rT_s / (r - \lambda)$$

in units of the total yield of $r$ replicates. Their error variance is $kr^2 / (r - \lambda) w'$, where $w'$ is the reciprocal of the error variance of the inter-block comparisons, in units of a single plot.

The most efficient estimates of the varietal differences are given by the quantities

$$Y_s = V_s + \mu W_s$$

where

$$W_s = (v - k)V_s - (v - 1)T_s + (k - 1)G$$
and

\[ \mu = \frac{w - w'}{wv(k - 1) + w'(v = k)} \]

The error variance of the Y's is

\[ \frac{kr(v - 1)}{wv(k - 1) + w'(v - k)} \]

The Analysis of Variance is shown in the table below. Only the left-hand side (i) of the S.S. is required if recovery of inter-block information is not undertaken. In such cases there is no need to sub-divide the blocks S.S. In the Table S.S. stands for sum of squares of deviations of the quantity following.

<table>
<thead>
<tr>
<th>(i)</th>
<th>D.F.</th>
<th>S.S.(i)</th>
<th>S.S.(ii)</th>
<th>(ii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Blocks (u)</td>
<td>b - 1</td>
<td>S.S.(blocks) k</td>
<td>(3)</td>
<td>Blocks (a)</td>
</tr>
<tr>
<td>2. Varietal Component</td>
<td>v - 1</td>
<td>(\frac{S.S.T}{k(r - \lambda)})</td>
<td>(\frac{S.S.W}{rv(v - k)(k - 1)})</td>
<td>Varietal Component</td>
</tr>
<tr>
<td>3. Remainder</td>
<td>b - v</td>
<td>(1)</td>
<td>(1)</td>
<td>Remainder</td>
</tr>
<tr>
<td>4. Varieties (a)</td>
<td>v - 1</td>
<td>(\frac{S.S.Q}{r_e})</td>
<td>(\frac{S.S.V}{r})</td>
<td>Varieties (u)</td>
</tr>
<tr>
<td>5. Intra-block error</td>
<td>rv - v - b + 1</td>
<td>(2)</td>
<td>(2)</td>
<td>Intra-block error</td>
</tr>
<tr>
<td>6. Total</td>
<td>rv - 1</td>
<td>S.S.(plots)</td>
<td>S.S.(plots)</td>
<td>Total</td>
</tr>
</tbody>
</table>

**Notes.** S.S.(i). Calculations have to be made in lines 1 and 2, when the S.S. for line 3 is obtained by difference. It is denoted by (1). Further calculations in lines 4 and 6 enable intra-block error to be obtained by difference. It is denoted by (2).

S.S.(ii). (1) and ( ), as calculated for S.S.(i), are transferred across. The calculation in line 2 then enables Blocks (a) to be obtained by addition. It
is denoted by (3). Further calculation in line 4 gives varieties (u), and then by subtraction from the total we may verify the calculation by reproducing the intra-block error (2).

If the M.S. for intra-block error is E and the M.S. (ii) for Blocks (a), with \( b - 1 \) D.F., is B, then \( w \) and \( w' \) are estimated from the equations

\[
\begin{align*}
  w &= \frac{1}{E} \\
  w' &= \frac{r(r-1)}{k(b-1)B - (v-k)E}
\end{align*}
\]

\( w' \) may be taken as equal to \( w \) if \( B \) is less than \( E \).

This general structure of analysis also applies to lattice designs provided the S.S. for blocks is taken to be that within replications, i.e., the S.S. for replications \((r-1)\) D.F.) has to be eliminated from the S.S. for blocks. It will be remembered that this was what we did. The formula for \( w' \) then becomes

\[
\begin{align*}
  w' &= \frac{r-1}{rB - E}
\end{align*}
\]

In a Youdon Square the S.S. corresponding to the \( k - 1 \) D.F. for the \( k \) series (representing leaf position, etc.) must likewise be eliminated.

Fisher and Yates give a worked-out example on pp. 20-21.

**Analysis of Covariance**

There is only time to discuss one aspect of the analysis of covariance, and that briefly. But it is at any rate fitting that we should wind up a course which has been largely concerned with the working out of designs which shall reduce the experimental error as much as possible, by referring to still another method of achieving this object. The need for any randomized design arises from the fact that similar plots over even a uniform area are very variable in their fertility.

If we are to have accurate treatment comparisons we should not only have an adequate amount of replication, but should also choose a suitable design, of which many variants have been presented. With, for example, the randomized blocks design, we
are tacitly assuming that the area covered by the block is reasonably uniform, although the mean fertility levels may differ as much as we please. If the blocks are not uniform, that fact will be reflected in the size of the new error term. Now sometimes we have additional information which helps us to see how variable the plots within the blocks are. The earliest cases considered were long-range experiments on perennial crops like tea and rubber, where it was usual to have data for the yields under uniformity trial conditions before the experiment proper was undertaken. On the assumption that there were no pronounced fertility changes by the time the experiment was under weigh, the yields of one or more preliminary years would provide useful ancillary information; a method was then sought which would take this information into account in order to render the comparisons in the experimental year more precise.

Even without previous uniformity trials, ancillary information of this kind is sometimes available. Suppose we are testing manurial dressings, applied in the spring, for their effect on autumn-sown wheat. If a count of numbers of plants be made at a time when they may be considered to be well established, but before any dressings are applied, these numbers will provide ancillary information bearing on fertility differences between plots. Another example may be cited from the domain of animal experimentation.

It was our experience that it was difficult to keep the errors of such trials satisfactorily low, in spite of simulating the randomized blocks design in agronomy trials by assigning the experimental treatments to groups of animals belonging to the same brood, or litter (which then constituted a block) and even arranging (in nutrition studies) to have them fed individually. One of the difficulties is that if we study the effect of treatments by measuring the growth rates of the animals, they are sure to be different in initial weight at the start of the experiment (before differential feeding starts), and are therefore likely to differ in growth-rate from this cause alone. Initial weights, which are available
anyhow, provide, none the less, ancillary information of the kind we have been
talking about, and may be taken into account to improve the accuracy of the growth-
rates under treatment.

In general, then, we see that, in addition to the experimental variable which
we have for analysis, and which we may call y, we may have another value x,
related to the same plot and unaffected by the treatment given to the plot. This
is the simplest case - in general we may have a number of such values x₁, x₂, etc.
corresponding to each y. Now we may expect to find the precision of our comparison
improved provided (a) x is reasonably highly correlated with y, after allowing
for the fact that the y's may differ through being differentially treated, and
also for the fact that both the x's and the y's may differ in the different block
aggregates: and (b) we are justified in "correcting for x", i.e., adjusting the
values of y, for the different treatments, to correspond to equal values of x
(a variable unaffected by those treatments).

It is here that regression analysis comes in. We regard y as the dependent
variable, and x as the variable influencing y (generally called, by an unfortunate
terminology, the independent variable). Let me recall to you that if for a
set of homogeneous data consisting of the sample pairs, x₁, y₁; x₂, y₂; ..., xₚ, yₚ,
we postulate that there is a linear regression relationship, then we fit a line

\[ Y = a + b(x - \bar{x}) \]

and find that \( a = \bar{y} \), while \( b \), the estimated regression coefficient, is given by

\[ b = \frac{Sy(x - \bar{x})}{S(x - \bar{x})^2} \]

A test of the significance of \( b \) is made by separating off from the total S.S.
of the y's that part due to the regression (1 D.F.), namely

\[ \left( \frac{Sy(x - \bar{x})}{S(x - \bar{x})^2} \right)^2 \]
leaving a part, which is in fact \( S(y - \bar{y})^2 \), for the S.S. for the deviations of the \( y \)'s from the fitted regression line (with \( p - 2 \) D.F.). A comparison of the M.S. from these two parts by the F-test (with 1 and \( p - 2 \) D.F.) will indicate whether the \( b \) is significant or not.

This, then, is the first stage. But the data of the agricultural experiment are heterogeneous. If, however, we parallel the analysis of variance of the \( y \)'s (separating, shall we say, into blocks, treatments and error) with a similar analysis of variance of the \( x \)'s, and also with a similar division of the Sum of Products of the deviations of the \( x \)'s and \( y \)'s from their respective means (called S.P.), we are in a position to calculate \( b \) from the error components, and this will be the appropriate \( b \) to use. Let \( E_x \) be the error S.S. for \( x \) (with \( n \) D.F.) and \( E_y \) that for \( y \). Furthermore let \( E_{xy} \) be the error S.P. Then we shall have

\[
b = \frac{E_{xy}}{E_x}
\]

and the test of significance will be obtained from the following table:

<table>
<thead>
<tr>
<th>D.F.</th>
<th>S.S.</th>
<th>M.S.</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>regression</td>
<td>1</td>
<td>( \frac{E_{xy}}{E_x} )</td>
<td>( R )</td>
</tr>
<tr>
<td>deviations</td>
<td>( n - 1 )</td>
<td>( E_y - \frac{E_{xy}^2}{E_x} )</td>
<td>( D )</td>
</tr>
<tr>
<td>Total</td>
<td>( n )</td>
<td>( E_y )</td>
<td></td>
</tr>
</tbody>
</table>

\( R \) and \( D \) have been used to denote the results of dividing the two S.S. by the respective D.F. F has 1 and \( n - 1 \) D.F., and if significant we may proceed with the next stage of the calculation. If not, the analysis may proceed on \( y \) in the ordinary way without taking account of \( x \), since the additional calculations may not be worthwhile.

In the above analysis a total S.P. will be calculated as the sum of all \( xy \) products, less the correction factor \( \frac{G_x G_y}{p} \), where \( G_x \) and \( G_y \) are the grand totals.
(of p values) of the x and y series. This S.P. is decomposed in exactly the same way as a S.S., merely substituting products for squares at every stage. For example, the treatments S.P. will be the sum of products of corresponding treatment totals, divided by r, the number of replications, less the correction factor already given.

The next steps are based upon the known test for the significance of the difference between two regression coefficients determined from independent samples. What we need is a test of whether treatment means differ significantly after each has been adjusted, on the basis of the calculated "error" regression, for the value of its x mean. We are finished with components of the analysis such as blocks (or rows and columns if the design is a Latin square), and concentrate, in the analysis of variance and covariance, on treatments (or part of treatments in certain cases, as when the design is factorial) and error. Let the table be:

<table>
<thead>
<tr>
<th></th>
<th>D.F.</th>
<th>S.S. (x)</th>
<th>S.P. (xy)</th>
<th>S.S. (y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatments</td>
<td>n₁</td>
<td>Tₓ</td>
<td>Tₓy</td>
<td>Tᵧ</td>
</tr>
<tr>
<td>error</td>
<td>n₂</td>
<td>Eₓ</td>
<td>Eₓy</td>
<td>Eᵧ</td>
</tr>
<tr>
<td>treatments + error</td>
<td>n₁ + n₂</td>
<td>Tₓ + Eₓ</td>
<td>Tₓy + Eₓy</td>
<td>Tᵧ + Eᵧ</td>
</tr>
</tbody>
</table>

We now subtract from Tᵧ + Eᵧ, and from Eᵧ, the S.S. due to the regression belonging to these lines, and subtract, thus:

<table>
<thead>
<tr>
<th></th>
<th>S.S. (y)</th>
<th>S.S. (R)</th>
<th>Difference</th>
<th>D.F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatments + error</td>
<td>Tᵧ + Eᵧ</td>
<td>(\frac{(Tₓy + Eₓy)^2}{Tₓ + Eₓ})</td>
<td>Tᵧ + Eᵧ - (\frac{(Tₓy + Eₓy)^2}{Tₓ + Eₓ})</td>
<td>n₁ + n₂ - 1</td>
</tr>
<tr>
<td>error</td>
<td>Eᵧ</td>
<td>(\frac{Eₓy^2}{Eₓ})</td>
<td>Eᵧ - (\frac{Eₓy^2}{Eₓ})</td>
<td>n₂ - 1</td>
</tr>
</tbody>
</table>

difference  Tₐ  n₁
\( T_a \) has been used to denote that it is the adjusted \( S.S. \) for treatments. Dividing
by \( n_1 \) we get the \( M.S. \) for adjusted treatments. The \( M.S. \) from the error line may be
denoted by \( s^2 \). Then the ratio of these \( M.S. \) is \( F = T_a / n_1 s^2 \), with \( n_1 \) and \( n_2 - 1 \) D.F.
If significant, we can conclude that there are treatment differences over and above
those which might have been due to the corresponding values of \( x \) being unequal.

For summary table purposes we require the adjusted treatment means. Let
\( \bar{x}_1, \bar{y}_1; \bar{x}_2, \bar{y}_2; \ldots \bar{x}_t, \bar{y}_t \) be the means for the \( x \) and \( y \) series, where there are
\( t \) treatments, \( \bar{x} \) and \( \bar{y} \) being the general means. Let \( b \) be the error regression
coefficient \( \frac{E_{xy}}{E_x} \). Then the adjusted treatment means are
\( \bar{y}_1 - b(\bar{x}_1 - \bar{x}) \),
\( \bar{y}_2 - b(\bar{x}_2 - \bar{x}), \ldots \bar{y}_p - b(\bar{x}_p - \bar{x}) \). We may calculate the estimated variance
of the difference between any two of these adjusted means. For example, the
difference between the first two is

\[ \bar{y}_1 - \bar{y}_2 - b(\bar{x}_1 - \bar{x}_2) \]

for which the estimated variance is

\[ s^2 \cdot \frac{1}{2} + \frac{(\bar{x}_1 - \bar{x}_2)^2}{E_x} \]

The first term of this expression is the usual expression for the variance of the
difference between two independent means of \( r \). The second part is the variance
of the second part of the difference, and will vary according to the values of \( \bar{x}_1 \),
\( \bar{x}_2 \), etc. Sometimes the correction is very small, and the ordinary formula can then
be used with reasonable safety.

With more than one adjusting variate we enter the realm of partial and multiple'
regression analysis, but apart from the calculations required to obtain the \( S.S. \) due
to regression, the analysis is the same in principle as the case which has been
discussed, certainly so far as the \( F \) test is concerned.
Design of Experiments - Exercise 1

The yields of wheat (grain) in pounds from 50 sections of Mercer and Hall's 1 acre field are as follows (arranged as on field plan).

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>41.1</td>
<td>42.5</td>
<td>40.3</td>
<td>38.5</td>
<td>36.6</td>
<td></td>
</tr>
<tr>
<td>41.8</td>
<td>40.5</td>
<td>38.3</td>
<td>40.2</td>
<td>38.0</td>
<td></td>
</tr>
<tr>
<td>40.4</td>
<td>41.9</td>
<td>37.8</td>
<td>40.0</td>
<td>39.5</td>
<td></td>
</tr>
<tr>
<td>37.8</td>
<td>42.4</td>
<td>41.8</td>
<td>34.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40.4</td>
<td>42.0</td>
<td>36.7</td>
<td>41.8</td>
<td>38.8</td>
<td></td>
</tr>
<tr>
<td>39.4</td>
<td>42.7</td>
<td>38.2</td>
<td>39.7</td>
<td>38.5</td>
<td></td>
</tr>
<tr>
<td>42.8</td>
<td>42.2</td>
<td>38.1</td>
<td>38.0</td>
<td>40.2</td>
<td></td>
</tr>
<tr>
<td>41.6</td>
<td>40.2</td>
<td>35.5</td>
<td>33.3</td>
<td>37.6</td>
<td></td>
</tr>
<tr>
<td>41.8</td>
<td>41.4</td>
<td>40.1</td>
<td>34.0</td>
<td>38.1</td>
<td></td>
</tr>
<tr>
<td>43.4</td>
<td>43.1</td>
<td>42.1</td>
<td>34.5</td>
<td>38.5</td>
<td></td>
</tr>
</tbody>
</table>

(1) Make your own completely randomized selection of 5 non-overlapping samples of 10 values each from the above data. Calling the samples A, B, C, D and E, analyse the total sum of squares into the two parts (a) between samples (b) within samples. Carry out the F-test to compare the two estimates of variance.

(2) Using the samples selected in (1) add 5 to all plots of A, subtract 5 from all plots of B, add 2 to all plots of C and subtract 2 from all plots of D. Then repeat the analysis as in (1).

(3) Construct 10 blocks out of the above table by dividing each column half-way down. Allot the letters A to E at random within each block and add and subtract 5 and 2 as in (2). Then analyse as a randomized block trial.
A standard variety (A) of barley was compared with 3 new varieties (B, C, and D), bred for high yield, in a randomized block experiment, according to the plan set out below. Each plot was 1/40 acre in size, and the numbers in the plots denote the yields in pounds. Carry out the appropriate statistical analysis, and say what can be learned from it (a) on the yielding capacities as a whole of the new varieties in relation to that of A, and (b) concerning the comparative yielding capacities of the new varieties. Give an effective summary table in which the yields are expressed in hundredweights per acre, together with their standard error.

<table>
<thead>
<tr>
<th>Blocks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>C 68</td>
</tr>
<tr>
<td>D 87</td>
</tr>
<tr>
<td>A 57</td>
</tr>
<tr>
<td>B 72</td>
</tr>
</tbody>
</table>

The sum of the squares of the 20 numbers in this table is 124,270.
An experiment on the use of nitrogenous fertilizers on wheat was arranged as a 5 x 5 Latin square, each plot being 1/40 acre in size. The control plot (having no fertilizer) is denoted by 0; S marks the plots which received a single dressing of sulphate of ammonia in March, while SS marks the plots which received the same total dressing, but in 6 monthly dressings from November to April. Plots which received cyanamide in October to an equivalent amount (in nitrogen) as the others are marked C, while D marks plots which received half their dressing as cyanamide and half as dicyanodiamide. The plan is given below, the numbers denoting the yields of the plots in pounds. Conduct the statistical analysis of the data, measuring the significance of the effect of applying a nitrogenous dressing, of whatever kind; also see if you can determine, by a statistical test, whether some forms of dressing are more effective than others. Set out a full summary table with your conclusions.

<table>
<thead>
<tr>
<th>D</th>
<th>SS</th>
<th>O</th>
<th>C</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>72.2</td>
<td>55.4</td>
<td>36.6</td>
<td>67.9</td>
<td>73.0</td>
</tr>
</tbody>
</table>

The sum of squares of the 25 numbers in the table is 113,574.73
An experiment was laid out to test the effect of nitrogenous and potassic fertilizers on asparagus. There were eight randomized blocks; O means no dressing, N stands for a dressing of nitrogen, K for a dressing of potash, and NK for the combined dressing. The yield in lb. per plot is shown in the plan together with the treatment symbol.

<table>
<thead>
<tr>
<th></th>
<th>NK</th>
<th>K</th>
<th>O</th>
<th>N</th>
<th>K</th>
<th>O</th>
<th>NK</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.0</td>
<td>16.2</td>
<td>14.6</td>
<td>12.7</td>
<td>13.0</td>
<td>12.7</td>
<td>11.3</td>
<td>10.3</td>
<td></td>
</tr>
<tr>
<td>10.2</td>
<td>12.8</td>
<td>13.8</td>
<td>13.9</td>
<td>13.3</td>
<td>15.2</td>
<td>12.1</td>
<td>11.5</td>
<td></td>
</tr>
<tr>
<td>14.7</td>
<td>9.3</td>
<td>8.8</td>
<td>9.0</td>
<td>11.4</td>
<td>10.4</td>
<td>11.7</td>
<td>9.3</td>
<td></td>
</tr>
<tr>
<td>10.8</td>
<td>8.9</td>
<td>8.3</td>
<td>10.3</td>
<td>9.1</td>
<td>10.5</td>
<td>8.2</td>
<td>13.5</td>
<td></td>
</tr>
</tbody>
</table>

Carry through the statistical analysis in full, and prepare the necessary summary tables at the end to show the significant effects, together with their standard errors.

The crude sum of squares of the 32 plot yields amounts to 1729.48.
Lettuce Experiment

The following data are taken from a single replication of a larger experiment on the manurial requirements of lettuce. The fertilizers used were nitrogenous, phosphatic, and potassic, each in 0, 1 and 2 dressings, as indicated in the columns headed \( n \), \( p \), and \( k \).

The data for analysis consist of the numbers of lettuce plants emerging on germination. Carry out the statistical analysis according to the method outlined in class, i.e. isolating the main effects of the fertilizers and the double linear \((n,p)\) components of the first-order interactions, using the remaining S.S. to estimate error. As a supplementary exercise isolate the triple linear \((n,p,k)\) component of the second order interaction.

<table>
<thead>
<tr>
<th>npk</th>
<th>No.</th>
<th>npk</th>
<th>No.</th>
<th>npk</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>112</td>
<td>37</td>
<td>020</td>
<td>47</td>
<td>222</td>
<td>15</td>
</tr>
<tr>
<td>121</td>
<td>31</td>
<td>002</td>
<td>33</td>
<td>201</td>
<td>42</td>
</tr>
<tr>
<td>010</td>
<td>44</td>
<td>221</td>
<td>15</td>
<td>012</td>
<td>41</td>
</tr>
<tr>
<td>022</td>
<td>25</td>
<td>122</td>
<td>29</td>
<td>210</td>
<td>24</td>
</tr>
<tr>
<td>100</td>
<td>39</td>
<td>011</td>
<td>38</td>
<td>000</td>
<td>52</td>
</tr>
<tr>
<td>211</td>
<td>28</td>
<td>212</td>
<td>27</td>
<td>111</td>
<td>28</td>
</tr>
<tr>
<td>202</td>
<td>35</td>
<td>200</td>
<td>38</td>
<td>120</td>
<td>30</td>
</tr>
<tr>
<td>001</td>
<td>37</td>
<td>101</td>
<td>31</td>
<td>102</td>
<td>50</td>
</tr>
<tr>
<td>220</td>
<td>26</td>
<td>110</td>
<td>32</td>
<td>021</td>
<td>29</td>
</tr>
</tbody>
</table>
Design of Experiments: Exercise 6

Pasture Experiment (Split-Plot)

The seeds mixtures denoted by A, B, C and D below were sown under wheat in 1938, the treatments being randomized over the plots of each block. In the summer of 1939 Blocks, I, IV, VI and VII were grazed and Blocks II, III, V and VIII were cut for hay (each pair of blocks was taken as a unit and a random choice made as to which block should be grazed and which cut for hay). The table below gives the estimated weights, in pounds, of clover (green) in the May 1940 cut. Analyse these weights fully and draw attention to the significant results.

A - Mixture including Italian Ryegrass
B - " " " L. F. Red Clover
C - " " " both I.R.G and L.F.R.C.
D - " " " neither I.R.G nor L.F.R.C.

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>8.7</td>
<td>15.2</td>
<td>14.8</td>
<td>16.3</td>
</tr>
<tr>
<td>B</td>
<td>14.1</td>
<td>12.4</td>
<td>22.4</td>
<td>25.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>VIII</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>22.2</td>
<td>21.9</td>
<td>16.6</td>
<td>19.6</td>
</tr>
<tr>
<td>B</td>
<td>21.5</td>
<td>20.6</td>
<td>13.6</td>
<td>13.3</td>
</tr>
</tbody>
</table>

The sum of the 32 weights in the above table is 527.4, and the sum of their squares is 9410.18. The blocks were in one long line in order I - VIII, but have been shown above for convenience in 2 lots.
Design of Experiments - Exercise 7

The field plan below is that of a greenhouse experiment on soybeans conducted in Iowa, the figures being yields of soybean hay in ounces per plot, obtained by clipping and weighing. The design is \((2^3)\) factorial, the contrasts being

\[
\begin{align*}
\text{a - variety I vs variety II} \\
\text{b - poor soil (diluted with sand) vs good soil (compost added)} \\
\text{c - dry soil vs wet soil}
\end{align*}
\]

The first-order interactions BC, AB and AC are confounded, in that order, in replications 1, 2 and 3, and the second-order interaction ABC is confounded in replication 4. Carry out the statistical analysis, and summarize the significant results. You may use the fact that the Total S.S. (31 D.F.) is 3605.97.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>BC</td>
<td>II</td>
<td>II'</td>
<td>III</td>
</tr>
<tr>
<td>(1)</td>
<td>b</td>
<td>ab</td>
<td>bc</td>
<td>a</td>
</tr>
<tr>
<td>7</td>
<td>24</td>
<td>36</td>
<td>31</td>
<td>28</td>
</tr>
<tr>
<td>abc</td>
<td>ac</td>
<td>bc (1)</td>
<td>ac</td>
<td>c</td>
</tr>
<tr>
<td>39</td>
<td>31</td>
<td>19</td>
<td>26</td>
<td>24</td>
</tr>
<tr>
<td>a</td>
<td>c</td>
<td>abc</td>
<td>b</td>
<td>ab (1)</td>
</tr>
<tr>
<td>30</td>
<td>21</td>
<td>41</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>bc</td>
<td>ab</td>
<td>c</td>
<td>a</td>
<td>bc abc</td>
</tr>
<tr>
<td>27</td>
<td>39</td>
<td>30</td>
<td>33</td>
<td>26</td>
</tr>
</tbody>
</table>
Exercise 8

The attached sheet gives the yields of sugar beet obtained from 108 plots of a $\left(3^3\right)$ experiment, laid out in four replications, with each replication divided into three blocks. Reps. W, X, Y and Z correspond to reps. IV, III, I and II of the lecture course. M denotes that the fertilizer dressings (N, P, K) were put on in 1, 2 and 3 applications; N denotes that 3 levels (1, 2 and 3) of nitrogen were used; and (pk) denotes that 3 levels (0, 1 and 2) of a combined dressing of phosphate and potash were used. The block numbers, (1), (2) and (3), correspond to the three Latin Squares of a cyclic set, but do not necessarily correspond to the numbers used in the lecture course.

You may attempt a full scale analysis of the entire experiment. Alternatively, if not sure of the way to do this, analyze replications W and X only, according to the method outlined in the lecture course.
### Replication W

<table>
<thead>
<tr>
<th>MN(pk)</th>
<th>Yield</th>
<th>MN(pk)</th>
<th>Yield</th>
<th>MN(pk)</th>
<th>Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>111</td>
<td>15.7</td>
<td>311</td>
<td>20.2</td>
<td>321</td>
<td>16.9</td>
</tr>
<tr>
<td>132</td>
<td>16.7</td>
<td>121</td>
<td>17.6</td>
<td>232</td>
<td>17.9</td>
</tr>
<tr>
<td>322</td>
<td>15.4</td>
<td>231</td>
<td>16.6</td>
<td>122</td>
<td>15.8</td>
</tr>
<tr>
<td>310</td>
<td>16.2</td>
<td>222</td>
<td>18.0</td>
<td>330</td>
<td>12.8</td>
</tr>
<tr>
<td>331</td>
<td>15.5</td>
<td>332</td>
<td>18.4</td>
<td>220</td>
<td>18.6</td>
</tr>
<tr>
<td>120</td>
<td>16.5</td>
<td>210</td>
<td>18.3</td>
<td>311</td>
<td>16.9</td>
</tr>
<tr>
<td>230</td>
<td>17.7</td>
<td>112</td>
<td>21.5</td>
<td>312</td>
<td>18.4</td>
</tr>
<tr>
<td>221</td>
<td>16.9</td>
<td>130</td>
<td>19.9</td>
<td>110</td>
<td>17.8</td>
</tr>
<tr>
<td>212</td>
<td>17.4</td>
<td>320</td>
<td>19.4</td>
<td>211</td>
<td>17.6</td>
</tr>
</tbody>
</table>

Total: 146.0 169.9 152.7 468.6

### Replication X

<table>
<thead>
<tr>
<th>MN(pk)</th>
<th>Yield</th>
<th>MN(pk)</th>
<th>Yield</th>
<th>MN(pk)</th>
<th>Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>330</td>
<td>13.1</td>
<td>111</td>
<td>16.4</td>
<td>112</td>
<td>13.3</td>
</tr>
<tr>
<td>110</td>
<td>14.5</td>
<td>331</td>
<td>13.7</td>
<td>310</td>
<td>16.0</td>
</tr>
<tr>
<td>311</td>
<td>12.2</td>
<td>210</td>
<td>14.3</td>
<td>321</td>
<td>15.2</td>
</tr>
<tr>
<td>220</td>
<td>12.2</td>
<td>312</td>
<td>15.8</td>
<td>120</td>
<td>15.7</td>
</tr>
<tr>
<td>212</td>
<td>11.3</td>
<td>221</td>
<td>16.7</td>
<td>311</td>
<td>16.3</td>
</tr>
<tr>
<td>132</td>
<td>11.4</td>
<td>232</td>
<td>16.0</td>
<td>211</td>
<td>16.7</td>
</tr>
<tr>
<td>321</td>
<td>9.7</td>
<td>320</td>
<td>15.8</td>
<td>222</td>
<td>14.6</td>
</tr>
<tr>
<td>322</td>
<td>10.3</td>
<td>122</td>
<td>15.6</td>
<td>320</td>
<td>12.2</td>
</tr>
<tr>
<td>121</td>
<td>12.9</td>
<td>130</td>
<td>18.5</td>
<td>332</td>
<td>16.7</td>
</tr>
</tbody>
</table>

Total: 107.7 142.8 137.2 387.7

### Replication Y

<table>
<thead>
<tr>
<th>MN(pk)</th>
<th>Yield</th>
<th>MN(pk)</th>
<th>Yield</th>
<th>MN(pk)</th>
<th>Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>230</td>
<td>15.8</td>
<td>310</td>
<td>15.6</td>
<td>221</td>
<td>16.4</td>
</tr>
<tr>
<td>110</td>
<td>13.6</td>
<td>130</td>
<td>15.9</td>
<td>120</td>
<td>15.7</td>
</tr>
<tr>
<td>331</td>
<td>16.8</td>
<td>122</td>
<td>16.4</td>
<td>210</td>
<td>15.2</td>
</tr>
<tr>
<td>312</td>
<td>16.3</td>
<td>332</td>
<td>17.0</td>
<td>311</td>
<td>16.7</td>
</tr>
<tr>
<td>222</td>
<td>14.8</td>
<td>321</td>
<td>13.7</td>
<td>112</td>
<td>13.5</td>
</tr>
<tr>
<td>312</td>
<td>13.2</td>
<td>220</td>
<td>14.4</td>
<td>232</td>
<td>14.4</td>
</tr>
<tr>
<td>211</td>
<td>16.7</td>
<td>111</td>
<td>17.9</td>
<td>131</td>
<td>16.5</td>
</tr>
<tr>
<td>121</td>
<td>17.0</td>
<td>212</td>
<td>15.1</td>
<td>322</td>
<td>14.6</td>
</tr>
<tr>
<td>320</td>
<td>15.1</td>
<td>231</td>
<td>14.2</td>
<td>330</td>
<td>13.8</td>
</tr>
</tbody>
</table>

Total: 139.3 142.2 136.8 416.3

### Replication Z

<table>
<thead>
<tr>
<th>MN(pk)</th>
<th>Yield</th>
<th>MN(pk)</th>
<th>Yield</th>
<th>MN(pk)</th>
<th>Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>131</td>
<td>13.2</td>
<td>310</td>
<td>12.6</td>
<td>120</td>
<td>12.7</td>
</tr>
<tr>
<td>320</td>
<td>11.8</td>
<td>211</td>
<td>12.2</td>
<td>330</td>
<td>11.5</td>
</tr>
<tr>
<td>332</td>
<td>12.9</td>
<td>232</td>
<td>11.4</td>
<td>321</td>
<td>10.4</td>
</tr>
<tr>
<td>110</td>
<td>15.9</td>
<td>331</td>
<td>14.0</td>
<td>132</td>
<td>16.2</td>
</tr>
<tr>
<td>212</td>
<td>14.9</td>
<td>322</td>
<td>13.4</td>
<td>222</td>
<td>16.4</td>
</tr>
<tr>
<td>311</td>
<td>17.2</td>
<td>121</td>
<td>15.9</td>
<td>111</td>
<td>17.3</td>
</tr>
<tr>
<td>221</td>
<td>17.3</td>
<td>130</td>
<td>15.5</td>
<td>312</td>
<td>16.5</td>
</tr>
<tr>
<td>230</td>
<td>17.2</td>
<td>112</td>
<td>16.7</td>
<td>210</td>
<td>16.7</td>
</tr>
<tr>
<td>122</td>
<td>17.0</td>
<td>220</td>
<td>15.7</td>
<td>231</td>
<td>17.7</td>
</tr>
</tbody>
</table>

Total: 137.4 127.4 135.4 400.2

1672.8