It will be assumed that anyone taking this course for credit has had a first course in probability and statistics. In particular, he should be familiar with the following notions.

1. The population

Any attempt at making a probabilistic prediction or a statistical inference rests on the assumption that we have a well-defined population, i.e. an aggregate of items each of which bears a mark (or several). Leaving aside certain mathematical difficulties, we may allow the number of items to be infinite and the marks to form either a discrete set or a continuous. In either event, we must know enough about the marks to be able to make a complete list of different marks (discrete case) or give a range (continuous). This might be called a population space.

It is essential that we be able to pronounce, for every item, that it either does or does not belong to the population. This is far from easy - witness the large amount of discussion about the rejection of observations.

Certainly, the specification of the population is an essential preliminary to everything else and should be approached with the greatest care.

2. Frequency distributions

Populations are described by frequency distributions. Standard methods of describing them are assumed known, e.g. density and distribution functions, as are the usual features of them, e.g. moments.

3. Randomly chosen samples

In practice, randomness in selecting samples must enter as an undefined element. What one should or can do to ensure or approximate randomness varies from one population to another. The objective, however, is clear. Randomness is involved to ensure that only the frequencies in the population dictate the outcome of the sampling. More precisely, randomness implies that the probabilities of the outcomes are equal to the corresponding frequencies in the population.

It is important to perceive that, without randomness, there can be no probabilistic predictions or statistical inference.

4. Probability and random variables

The point of view here adopted is that probability is defined only in situations where a well-defined population can be sampled randomly. Familiarity is assumed with the ideas associated with the following words: sample space, random variable, independence, mutual exclusiveness, conditional probability, expectation.

5. The following facts will be assumed known

a) E(sum) = sum (E) always

b) E(product) = product (E) for independent variables

c) Var(sum) = sum (Var) + 2 sum ($Covar$)

d) A linear function of independent normal variables is normal

e) The central limit theorem

f) Tests of significance, χ^2, t, F

g) Confidence limits.

The practice will be followed here of using symbols like x_1, x_2, ... to represent observations already made or yet to be made, or a variable x and also to represent random variables distributed according to the distribution of x.

Only one theorem, and a shallow one of that, is needed.

6. A set of independent random variables, x_1, x_2, ... x_n normally distributed with variance σ^2 and possibly different means, is transformed by an orthogonal linear transformation into set of independent normal variables, each with variance σ^2, to be called y_1, y_2, ... y_n. Only the means are changed

$$y_i = \Sigma a_{ij}\, x_j \qquad i = 1, \ldots n$$

$$\sum_j a_{ij}\, a_{kj} = \delta_{ik} \qquad i, k = 1, \ldots n$$

The fact that, under this transformation, $\sum_1^n y_i^2 = \sum_1^n x_i^2$, will be used frequently.

In what follows, the transformations to be used will all have the feature that the values of a_{1j} will be equal, hence equal to $1/\sqrt{n}$ It follows that $\sum_{j=1}^{n} a_{ij} = 0$, all $i > 1$. Then, if $x_i \sim N(\mu, \sigma^2)$, $y_1 = 1/\sqrt{n}\ \Sigma x_j = \sqrt{n}\ \bar{x}$. $E\, y_1 = \sqrt{n}\ \mu$ and for $i > 1$, $E\, y_i = \Sigma a_{ij}\, E\, x_j = \mu\ \Sigma a_{ij} = 0$.

7. The random variable $\chi^2_{(p)}$ can be defined as the sum of the squares of p independent variables, each $N(0,1)$. The variables y_i, $i \geq 2$, resulting from the transformation are independent $N(0,\sigma^2)$. Hence the

y_i/σ are $N(0,1)$ and $\sum_{i=2}^{n} y_i^2/\sigma^2 \sim \chi^2_{(n-1)}$. Also $\sum_{2}^{n} y_i^2 = \sum_{1}^{n} x_i^2 - y_1^2 = \sum_{1}^{n} x_i^2 - n\,\bar{x}^2 = \sum_{i} (x_i - \bar{x})^2$. Thus $\Sigma(x_i - \bar{x})^2/\sigma^2 \sim \chi^2_{(n-1)}$. We note, too, $E\,y_i^2 = \sigma^2$ $i > 2$ because $Var\ y_i = E\,y_i^2 = \sigma^2$. Hence $E\sum_{2}^{n} y_i^2 = E\left(\Sigma\left(x_i - \bar{x}\right)^2\right) = (n-1)\ \sigma^2$ and $s^2 = \Sigma\left(x_i - \bar{x}\right)^2/n-1$ is an unbiased estimator of σ^2.

8. Degrees of freedom

It is convenient and customary to say that a sample of n observations has n degrees of freedom. That is, the sample $(x_1, x_2, \ldots x_n)$, plotted as a point in n-space, may lie anywhere and is not constrained to any subspace. Under the transformation we have used, one d.f. has been given the task of isolating the mean, leaving the remaining n-1 d.f. to exhibit only the effects of error.

9. The analysis of variance distribution

Let $\chi^2_{(p)}$ and $\chi^2_{(q)}$ be independent. Then $F_{(p,q)} = \dfrac{\chi^2_{(p)}/p}{\chi^2_{(q)}/q}$ has a known and tabulated distribution. It is called the analysis of variance distribution, first announced to the world by R.A. Fisher, in 1924, in Toronto, at a meeting of the International Congress of Mathematicians and published in its Proceedings. Fisher prefered to use it in the form $z = \frac{1}{2}\log F$ and published suitable tables for z. Later, Snedecor tabulated F itself and introduced the symbol F for this purpose.

An earlier test function, $t_{(q)}$, is, from its definition related to $F_{(1,q)}$ by $t^2_{(q)} = F_{(1,q)}$.

Returning to the transformation, $(y_1 - \sqrt{n}\ \mu)^2 = \sigma^2\ \chi^2_{(1)}$ and

$y_2^2 + \ldots + y_n^2 = \sigma^2 \chi^2_{(n-1)}$, hence

$$\left(y_1 - \sqrt{n}\,\mu\right)^2 / \left[\left(y_2^2 + \ldots + y_n^2\right)/(n-1)\right] = F_{(1,\, n-1)}.$$

In terms of the x's, this F ratio is $n(\bar{x}-\mu)^2/s^2$, and $t_{(n-1)} = \sqrt{n}\,(\bar{x}-\mu)/s$, the familiar test function for testing μ or for calculating confidence limits for μ. Note that a two-sided test using $t_{(q)}$ is the same as a one-sided test using $F_{(1,q)}$

Two Samples

Suppose that a sample is drawn randomly from each of two populations, in circumstances in which the populations may be different in some respect. Let us say, though, that they are both normal, $\pi_1 : N(\mu_1, \sigma_1^2)$; $\pi_2 : N(\mu_2, \sigma_2^2)$.

Within this framework, there are two special situations both important.

1. $\sigma_1 = \sigma_2$, the question being the equality or otherwise of μ_1 and μ_2. This is usually the nature of an experimental situation.

2. The question at issue concerns the equality or otherwise of σ_1 and σ_2. Usually, we would not be concerned here also with the difference $\mu_1-\mu_2$.

The more general question of studying the difference $\mu_1-\mu_2$ without the restriction $\sigma_1 = \sigma_2$ (the Behrens-Fisher problem) will not be raised here, being rather remote from practical considerations.

1. Assume $\sigma_1 = \sigma_2 = \sigma$. Let the observations from π_1 and π_2 be $x_{i\alpha}$ $i = 1,2$ $\alpha = 1,2 \ldots n$. Our assumptions allow us to write

$$x_{i\alpha} = \mu_i + \varepsilon_{i\alpha} \qquad \varepsilon \sim N(0, \sigma^2)$$

Adopting an obvious symbolism, we can lay out an orthogonal transformation as follows.

	x_{11}	x_{12}	. . . .	x_{1n}	x_{21}	x_{22}	. . . .	x_{2n}	divisor
y_1	1	1	. . . .	1	1	1	. . . .	1	$\sqrt{2n}$
y_2	1	1	. . . .	1	-1	-1	. . . .	-1	$\sqrt{2n}$

y_3
.
.
.
y_{2n} orthogonal

Clearly y_1 is of no interest, but y_2 displays the difference between the two means, in addition to error.

As a first step in our enquiry, we may check whether y_2 does, in fact, contain more than error.

$$\frac{\left(y_2 - \sqrt{n/2}\ (\mu_1-\mu_2)\right)^2}{(y_3^2 + \ldots. + y_{2n}^2)/(2n-2)} \sim F(1,\ 2n-2)$$

If in this, we put $\mu_1 = \mu_2$ and find the resulting number too big to have come randomly from the F distribution, we have evidence that $\mu_1 \neq \mu_2$. This is the standard test of significance.

It is worthwhile to exhibit the required numbers for this test in a table, called an analysis of variance table.

source	degrees of freedom	sum of squares
between samples	1	y_2^2
within samples	$2n-2$	$y_3^2 + \ldots. + y_{2n}^2$

Let us write

$$T_i = \sum_{\alpha} x_{i\alpha} \qquad G = T_1 + T_2$$

The transformation is a useful instrument for displaying the structure of the data and for establishing the necessary distribution theory, but only occasionally is it a useful tool for computation. Rather, we will use it to derive computing rules, to be applied directly to the observations.

Let us write $T_i = \sum_{\alpha} x_{i\alpha}$, $T_1+T_2 = G$. Then $y_1 = G/\sqrt{2n}$ and

$$y_2^2 = (T_1 - T_2)^2/2n = (T_1^2 + T_2^2)/n - G^2/2n, \quad \sum_2^{2n} y_i^2 = \Sigma\Sigma x_{i\alpha}^2 - y_1^2 .$$

We then assemble these numbers in the table

source	df	SS
between	1	$(T_1 - T_2)^2/2n$ or $(T_1^2 + T_2^2)/n - G^2/2n$
within	$2n-2$	by subtraction
total	$2n-1$	$\Sigma\Sigma\, x_{i\alpha}^2 - G^2/2n$

The within-samples sum of squares calculated above by subtraction is easily seen to be $\Sigma\Sigma(x_{i\alpha} - \bar{x}_{i\cdot})^2 = (2n-2)\, s^2$.

Comments

If the two samples came as the result of an experiment - e.g. an attempt to compare two populations which arose through our doing two different things (i.e. treatments), our object, in the event that $\bar{x}_1 - \bar{x}_2$ differs from zero by more than error can reasonably account for, is to conclude that the difference between the treatments caused the observed difference between the responses. An experiment is carried out to make pronouncements about cause and effect. We are to envisage a causal system C, which may be quite elaborate, subject to our control in the sense that we can make changes in some of its variables and keep the others fixed. Then, we must have an effect system E which is observable, and may be simple (one variable) or more elaborate (several variables).

If we introduce a change in C and observe a change in E, we shall say that the change in C caused the change in E, provided:

(1) everytime the same change is introduced in C, the same change is observed in E.

(2) we can be sure that nothing else is responsible for the observed change in E.

These two conditions are simply the traditional tenents of science. (2) asserts the need for control and (1) asserts that unless a finding can be repeated it cannot have any claim to acceptance as a fact.

Evidently there are some difficulties here. (2) implies that we know all the variables in the causal system and have kept them under control, a rare situation indeed. Furthermore, observations in the *E* - system are inevitably afflicted with error, which can upset both the conditions which have been laid down.

It is the business of statistics to bring about some accommodation between what we would like to do and what we can do.

A first step in this direction is to develop the notion of what is to be meant by error. Speaking roughly and vaguely, error shows itself through the differences among the results when we try to do the same thing more than once. It is not intended that error include the consequences of mistakes and blunders or the effects of bias. Errors, as envisaged here, must behave as if they come randomly from some frequency distribution of errors, which have mean zero and a variance which does not change throughout the experiment. This construct has several implications about the conduct of experiments, which will be explored later.

Experiments

If what we call an experiment is intended to yield evidence about cause and effect, it follows that it is wholly concerned with changes, i.e. differences. Some of these differences will be caused by changes introduced into the causal system, obscured, of course, by errors. It is essential, then, that some provision be made to perceive what error contributions alone are capable of. An exercise that makes no provision for the definition and estimation of error cannot properly be called an experiment.

Contrasts

Any linear combination of the observations whose coefficients add to zero will be called a contrast. In the light of what has been said, the evaluation of the outcome of an experiment consists entirely of a study of contrasts, suitably chosen to correspond to changes introduced into the causal system.

This is the reason why we shall always include in the transformation on which we base our analysis, a component all of whose coefficients are equal. This ensures, through orthogonality, that all the other components must be contrasts.

To return to the question of two samples: the usual supposition is that we have two populations, with equal variances and possibly different means. A sample is selected randomly from each. In experimental situations, we have to start with a number of experimental units (animals, plot of land, etc.), presumably carefully chosen to be as much alike as possible.

Randomness now takes the form of random allocation of units to treatments. This has the effect of converting whatever differences there are among the units into a frequency distribution which is sampled randomly.

To perceive this point, think of an agricultural experiment carried out in a row, to keep the argument one-dimensional (or an experiment on a wire, or a rope). Let the length of the row be L and let x denote points along it.

Presumably, growing conditions will vary from point to point along the row, and will contribute to the differences among treatments that we will observe. If we could know the level of fertility (y) at each point x, we could use it to make corrections, but we could not talk about these fluctuations in terms of errors. For any designated value of y, call it Y, calculate the measure of x for which $y \leq Y$. Call this $F(Y)$. Then $F(Y)/L$ is the distribution function of a random variable, which can be sampled randomly by random selection of x.

Of course, the experiment will be carried out in intervals, not points and we will be using a grouped form of this distribution. The argument extends without change to higher dimensions.

If this curve depicting the variation in growing conditions has any persistant trends, it is easy to see that the resulting error distribution will be skewed. For this reason and others, blocking is usually worthwhile (see later).

In the two sample experiment, error enters, by definition, through differences among experimental units and is estimated by gathering up the differences among units treated alike, the within-samples sum of squares.

If the two samples have unequal numbers of observations, the orthogonal transformation must be changed if y_2 is to gather up the contribution of $\mu_1-\mu_2$.

	x_{11}	x_{12}	. . . .	x_{1n_1}	x_{21}	x_{22}	. . . .	x_{2n_2}	divisor
y_1	1	1	. . . .	1	1	1	. . . .	1	$\sqrt{n_1+n_2}$
y_2	n_2	n_2		n_2	$-n_1$	$-n_1$	. . . .	$-n_1$	$\sqrt{n_1 n_2 (n_1+n_2)}$
y_3									
.									
.					orthogonal				
.									
.									
$y_{n_1+n_2}$									

then $y_2 = \sqrt{\frac{n_1 n_2}{n_1+n_2}}\ (\mu_1-\mu_2)$ + error

y_3 . . . . $y_{n_1+n_2}$ exhibit error only.

The assumption of constant error variance

Generally speaking the assumption of constant variance is not to be accepted lightly. In experiments, though, that have been carefully planned and carried out, this assumption is usually satisfied fairly well. An exception might arise when the treatments have vastly different effects, but here no experiment is needed to establish this. A few special cases will be mentioned later.

Occasionally, the experiment may be carried out for the express purpose of finding out whether the treatments effect the error variance. The test of significance here is the two-sided

$$F = s_1^2/s_2^2 \sim F(n_1\text{-}1,\ n_2\text{-}1)$$

Three (or more) samples

	x_{11}	x_{12}	....	x_{1n_1}	x_{21}	x_{22}	....	x_{2n_2}	x_{31}	x_{32}	....	x_{3n_3}	divisor
y_1	1	1	....	1	1	1	....	1	1	1	....	1	$\sqrt{n_1+n_2+n_3}$
y_2	a	a	....	a	b	b	....	b	c	c	....	c	$\sqrt{n_1a^2+n_2b^2+n_3c^2}$
y_3	α	α	....	α	β	β	....	β	γ	γ	....	γ	$\sqrt{n_1\alpha^2+n_2\beta^2+n_3\gamma^2}$
y_4													
.													
.							orthogonal.						
.													
$y_{n_1+n_2+n_3}$													

We can write an orthogonal transormation as shown by choosing a, b, c, α, β, γ such that

$$n_1 a + n_2 b + n_3 c = 0$$

$$n_1 \alpha + n_2 \beta + n_3 \gamma = 0$$

$$n_1 a\alpha + n_2 b \beta + n_3 c \gamma = 0$$

then $y_2 = n_1 a\mu_1 + n_2 b\mu_2 + n_3 c\mu$ + error

We would, of course, want to choose a, b, c so that y_2 would display some desired contrast of μ_1, μ_2, μ_3 for example $a = 1/n_1$, $b = -1/n_2$, $c = 0$. Having done so α, β, γ are determined and lead to a y_3 that exhibits a linear function of μ_1, μ_2, μ_3 whose coefficients involve sample sizes. In general, this is of no interest (except, perhaps, in the context of regression). Therefore, in general, this experiment cannot be analyzed by means of an orthogonal transformation,

except in the case when the sample sizes are equal.

Even so, we can make choices of a, b, c, α, β, γ that satisfy the orthogonality conditions. The components y_4, y_5 .... $y_{n_1+n_2+n_3}$ will then, through orthogonality with y_1, y_2, y_3 have zero means and will therefore reflect error contributions only. The estimate of σ^2 will be $s^2 = \left(y_4^2 + ... + y_{n_1+n_2+n_3}^2\right) / \left(n_1+n_2+n_3-3\right)$. To see one way of computing this sum of squares, observe that the point of the transformation that deals with variation among samples may be written

	T_1	T_2	T_3
y_1	1	1	1
y_2	a	b	c
y_3	α	β	γ

or better

	$T_1/\sqrt{n_1}$	$T_2/\sqrt{n_2}$	$T_3/\sqrt{n_3}$	divisor
y_1	$\sqrt{n_1}$	$\sqrt{n_2}$	$\sqrt{n_3}$	$\sqrt{n_1+n_2+n_3}$
y_2	$\sqrt{n_1}a$	$\sqrt{n_2}b$	$\sqrt{n_3}c$	$\sqrt{n_1a^2+n_2b^2+n_3c^2}$
y_3	$\sqrt{n_1}\alpha$	$\sqrt{n_2}\beta$	$\sqrt{n_3}\gamma$	$\sqrt{n_1\alpha^2+n_2\beta^2+n_3\gamma^2}$

In this form, it displays an orthogonal transformation of the $T_i/\sqrt{n_i}$ into y's. Hence $y_1^2 + y_2^2 + y_3^2 = \sum_i \frac{T_i^2}{n_i}$ and the among samples sum of squares is

$$y_2^2 + y_3^2 = \sum_i \frac{T_i^2}{n_i} - \frac{G^2}{n_1+n_2+n_3}$$

Therefore, we can fill out the analysis of variance table as follows.

	df	ss
among samples	2	$\sum \frac{T_i^2}{n_i} - \frac{G^2}{N}$
within samples	$n_1 + n_2 + n_3 - 3$	by subtraction
total	$n_1 + n_2 + n_3 - 1$	$\sum_{i=1}^{3} \sum_{\alpha=1}^{n_i} x_{i\alpha}^2 - G^2/N$

$\left(N = n_1 + n_2 + n_3\right)$.

Orthogonal Experiments

An experiment which can be analyzed by means of an orthogonal transformation will be said to be an orthogonal experiment. Analysis, here, means the study of the contrasts the experiment was designed to explore. Two sample experiments are always orthogonal, but with three or more samples, orthogonality is ensured only with equal sample sizes.

It is usually a simple matter to layout the contrasts we wish to study and to calculate their standard errors. The use and interpretation of these values require care however (see next section).

It should be noted that there are penalties for non-orthogonality. The precision of the contrasts diminishes as the disparity among the sample sizes increases. Also, the comparisons we wish to study may not be independent.

The choice of components

Consider a three sample orthogonal experiment. The transformation is set up by choosing any three numbers which add to zero (a, b, c). Since only the ratios of the three numbers is required, there is only one free choice to be made. Then, numbers

α, β, γ are sought, such that

$$\alpha + \beta + \gamma = 0$$

$$a\alpha + b\beta + c\gamma = 0$$

The ratios of α, β, γ are thus completely determined. Any component orthogonal to y_1, y_2, y_3 must have coefficients that sum to zero within each sample and can therefore exhibit errors only.

The analysis of variance table is:

	df	*ss*
among samples	2	$y_2^2 + y_3^2$
within samples	$3n - 3$	$y_4^2 + \ldots + y_{3n}^2$
total	$3n - 1$	$\Sigma\ \Sigma\ x^2 - G^2/N$

This preliminary table should always be calculated, even though it does not exhibit everything we want. It does not require that choices of a, b, c, α, β, γ have been made, because the among samples sum of squares is the same for all choices.

In any truly experimental situation, the object of the exercise is to perceive certain well-defined contrasts which, in turn, dictate the choice of a, b, c and the other contrasts. It is not necessary that all the contrasts of interest can be put into an orthogonal transformation, because any contrast can be expressed as a linear combination of the set of orthogonal contrasts. Often the set of contrasts forms a hierarchy in which the order in which the contrasts are scrutinized is important.

An example

Suppose that treatment 1 is a standard and 2 and 3 are experimental variants of it. Then, the following choice of contrasts seems reasonable.

	T_1	T_2	T_3
y_1	1	1	1
y_2	-2	1	1
y_3	0	-1	1

y_3 inquires whether there is a difference between the two variants. If not, y_2 becomes reasonable to test whether their average differs from the standard. However, if y_3 is significantly large, y_2 becomes meaningless, because, as a matter of statistical decency, we never average things that differ through more than error. To do so is to distort and deceive.

Another example

Suppose that T_1 is some standard treatment, T_2 differs from it through the addition of one unit of some additive and T_3 through the addition of 2 units of the same additive. Then, a graph can be drawn, showing the dependence of the response on the amount of additive. Anything we may do by way of analysis can be thought of as an inquiry into the nature of the graph

	T_1	T_2	T_3
y_2	-1	0	1
y_3	-1	2	-1

Then, if the points lie on a straight line, $y_3 = 0$ and y_2 yields the slope of the line. Hence we first test y_3. If significantly different from zero, a test of significance of y_2 becomes useless and should not be made, because nothing depends on the significance or otherwise of y_2. Its value is useful, though, in writing the equation of the curve (a regression question).

If y_3 is not significantly different from zero, we conclude that a straight line will suffice and proceed to y_2 to study its slope. If y_2 is not significant, the additive has had no effect.

In both these examples, we might display the analysis of variance table, for our own purposes at least, in the form.

	df	*ss*	*df*	*ss*
among treatments	2	$y_2^2 + y_3^2$	1	y_2^2
			1	y_3^2
within treatments	$3n-3$			
total	$3n-1$			

In the second example, if we have more than three samples, the appropriate coefficients to use to display linear, quadratic , cubic.... components are not geometrically obvious. When the abscissae are equally spaced, as in the example, suitable coefficients are tabulated as values of orthogonal polynomials.

Experiments of the sort we have been discussing, in which the corresponding analysis of variance is between-within, with error displayed in the within-samples sum of squares are called completely randomized experimental plans.

Factorial Arrangements

Think of an experiment to compare two diets by feeding them to rats. Suppose we know or suspect that females will respond to the diets rather differently from males (or that we want to find out whether this is so). We would then think of comparing diets on both females and

males, equally often if orthogonality is to be maintained. We might then assign $4n$ animals randomly to the four "treatment" combinations, n to each.

In a sense, there is nothing new here. We have a completely randomized experiment with 4 samples (treatments)

	df
among	3
within	$4n - 4$
total	$4n - 1$

The thing that <u>is</u> new here is the nature or structure of the treatments, made up as they are by combining two <u>factors</u>, sex, diet in all possible ways. This dictates the comparisons we should make among the treatments.

We have three questions to be settled

1. Do the diets yield different responses?
2. Do the sexes yield different responses?
3. Is the observed difference between diets the same for the two sexes?

Corresponding to these questions, we have

	T_{11}	T_{12}	T_{21}	T_{22}	divisor
y_1	1	1	1	1	$\sqrt{4n}$
y_2	1	1	-1	-1	$\sqrt{4n}$
y_3	1	-1	1	-1	$\sqrt{4n}$
y_4	1	-1	-1	1	$\sqrt{4n}$

diet \ sex	1	2	
1	T_{11}	T_{12}	$T_{1.}$
2	T_{21}	T_{22}	$T_{2.}$
	$T_{.1}$	$T_{.2}$	G

y_2 displays the difference between diets, averaged over the sexes

y_3 displays the differences between the sexes, averaged over the diets (y_2 and y_3 are called main effects).

y_4 contrasts the diet difference between the sexes, or equivalently, the sex difference between the diets (y_4 is called an interaction component).

Of these components, y_4 is the most important and must be inspected first. If y_4 is significantly different from zero, we have at once, some conclusions.

1. the diets do yield different responses
2. the sexes do yield different responses
3. the difference between the diets is different in the two sexes or, equivalently, the difference between the sexes is different in the two diets. Therefore the averaging performed in y_2 and y_3 is unwarranted and these two components represent an attempt to answer a too-simple question. There is no occasion to test y_2 and y_3.

If y_4 is not significantly different from zero, the averaging in y_2 and y_3 is not forbidden and we may proceed to test them for significance.

It is to be noted that the component y_4 may be formed by a column by column multiplication of the coefficients of y_2 and y_3. Hence, we get the usual notation for y_4 i.e. sex × diet. (interaction)

It is easy to check that this kind of multiplication between two main effect components always yields an interaction component. There are also other ways of obtaining interaction components. These questions will come up again.

The final analysis of variance table takes the form

	df	
diets	1	y_2^2 [$= \left(T_{1.}^2 + T_{2.}^2\right)/2n - G^2/4n$]
sexes	1	y_3^2 [$= \left(T_{.1}^2 + T_{.2}^2\right)/2n - G^2/4n$]
diet × sexes	1	y_4^2 (or subtraction)
error	$4n - 4$	by subtraction
total	$4n - 1$	

We might say that when there is no $s \times d$ interaction, our two-dimensional question decomposes into two one-dimensional questions. The margins of our two-dimensional table contain all of the information. When an interaction exists, we still have one two-dimensional question. The body of the table is needed to properly display the information and the margins are not useful (except for computations) and can be misleading.

Note, if we have m sexes and n diets, then mn-1 d.f. among treatments would partition into

	df
sexes	$m - 1$
diets	$n - 1$
$s \times d$	$(m-1)(n-1)$
among treatments	$mn - 1$

The writing of models

In the experiment discussed above, we could use $x_{ij\alpha}$, $i = 1,2$, $j = 1,2$, $\alpha = 1,2,\ldots,n$ to represent an observation on diet i, sex j. We think of the variation among the x's in terms of $x_{ij\alpha} = \mu_{ij} + \epsilon_{ij\alpha}$, which corresponds to the preliminary analysis of variance table. Then, to take account of the structure of the treatments we may write $\mu_{ij} = \mu + \delta_i + \sigma_j + \tau_{ij}$, a tautology, of course, but one that exhibits the way in which we propose to interpret the data.

By adopting the convention that $\sum_i \delta_i = 0$, $\sum_j \sigma_j = 0$, $\sum_i \tau_{ij} = 0$, $\sum_j \tau_{ij} = 0$, we have the correct number of parameters and the number of each kind corresponds to the number of degrees of freedom.

It is instructive to substitute this structure into the components y_2, y_3, y_4. More on factorials later.

The completely randomized experiment is not often used. Usually, we know of reasons why our experimental units vary among themselves and can use this knowledge, in some measure, to circumvent its introduction into the contrasts and into error.

Blocking

Pursuing the biological example somewhat further, suppose it is known that animals from the same litter respond in a more uniform manner than do those from different litters. It follows that contrasts made within litters will be more precise than contrasts that involve differences among litters. To take advantage of this, if we can get an adequate number of suitable litters each of which supplies 2 males and 2 females, we could allocate randomly one male and one female of each litter to diet 1 and the others to diet 2. Then differences between diets, between sexes and their interaction are perceptible within litters.

If n litters are used, n-1 degrees of freedom will be used to display differences among litters, 3 degrees of freedom for treatment differences and, it should be obvious, $3(n-1)$ df for the interaction of treatments and litters. It may not be obvious, though, just where one would look to see the contributions of error only. We do not now have, as we had earlier, several observations which should be the same, apart from error. In this sense, there are no repetitions. On the other hand, something has been repeated, namely, each of our contrasts has been repeated n times, once in each litter. The differences among these contrasts, as we proceed from litter to litter, should provide a sensible definition of error, provided we use only litters that are not too different from one another, i.e. each should be an attempt to carry out the same trial as every other.

To say the same thing in another way: The litters should be chosen so as to ensure that there shall be no genuine interaction between litters and treatments. The litters × treatments interaction sum of squares will then reflect error only.

The responsibility for the selection of litters that meet the condition of no interaction with treatments must rest on the experimenter, who may properly be expected to have considerable knowledge to start with.

The litters, in this example, represent an attempt to group the experimental units into sets within which they are more uniform than they would be without the grouping. It follows that we would expect greater differences between sets than within sets. These sets are called blocks after an agricultural prototype. In the example, the blocks were large enough to receive all the contrasts, hence they are complete blocks. They may also be called replications, inasmuch as each block receives all the contrasts under study. The term replication is best used only in this sense, one trial involving all the contrasts, with blocking implied. It is best not to use the word to describe the repetitions in a completely randomized experiment.

The arrangement we have here, in which we have suitable (complete) blocks, with the treatments assigned randomly to the experimental units (sometimes called plots) within the blocks, is called a randomized (complete) block arrangement.

If we represent by x_{ijk} the observation on diet i, sex j, litter k, we have the following computational scheme.

	df	
blocks	n-1	$\Sigma T^2_{..k}/4 - G^2/4n$
treatments	3	$\Sigma\Sigma T^2_{ij..}/n - G^2/4n$
bl. × t.	$3(n-1)$	by subtraction
total	$4n-1$	$\Sigma\Sigma\Sigma x^2_{ijk} - G^2/4n$.

An orthogonal transformation corresponding to this arrangement could be

	x_{111}	x_{121}	x_{211}	x_{221}	. . . .	x_{11n}	x_{12n}	x_{21n}	x_{22n}
	1	1	1	1		1	1	1	1
bl.	a_1	a_1	a_1	a_1		a_n	a_n	a_n	a_n
d.	1	1	-1	-1		1	1	-1	-1
s.	1	-1	1	-1		1	-1	1	-1
d × s	1	-1	-1	1		1	-1	-1	1
bl × d	a_1	a_1	$-a_1$	$-a_1$		a_n	a_n	$-a_n$	$-a_n$
bl × s	a_1	$-a_1$	a_1	$-a_1$		a_n	$-a_n$	a_n	$-a_n$
bl×d×s	a_1	$-a_1$	$-a_1$	a_1		a_n	$-a_n$	$-a_n$	a_n .

Here $\Sigma a_k = 0$ and the a's represent the n-1 orthogonal choices that can be made.

One encounters, now and then, the view that replications should be tested. If significant, some doubt exists about the adequacy of the experiment. This is, of course, nonsense. The whole object in blocking is to arrange less variation within blocks than between. On the other hand, if the variation attributed to replications (or treatments) is huge, one might well be suspicious that the postulate of no replication × treatment interaction might be violated and a study of residuals might be in order.

Another, rather extreme, view shows itself occasionally: If the replications sum of squares is not significant, it may be pooled with the error sum of squares to augment the degrees of freedom allotted to error. This kind of self indulgence is intolerable. It is simply a way of saying: Choose the error you like best.

In the definition of error

Speaking generally and somewhat loosely, error shows itself when we try to do the same thing several times and come up with different results. In the completely randomized arrangement, we have a number of experimental units treated alike and perceive error in the differences among the results they yield. In the randomized block pattern, while we may not have several units treated alike, we have tried to repeat something, namely, the contrasts we are examining, which are repeated from replication to replication. The differences among these differences, i.e. the replications × treatments interaction, reflect error.

We could think of an arrangement in which both ways of displaying error enter. To use the same illustration as before, suppose that we have an adequate supply of litters which contain 4 males and 4 females. Two of the males chosen randomly, are given diet 1 and the other males diet 2; similarly for the females. If we use r litters, the analysis of variance reads:

	df
replications	$r-1$
experimental effects	3
rep. × exp. effects	$3(r-1)$
within pairs	$4r$.

The within pairs sum of squares reflects a primitive, inevitable, local sort of error which must enter into all the contrasts. The rep. × exp. effects, in addition to this local effort, may reflect other sources of error, arising from some lack of control from replication to

replication. It is likely, therefore, to run somewhat larger than the local error and is the proper error against which to test the experimental effects. It is therefore called experimental error, in contrast with the local sampling error.

Clearly, the provision for estimating sampling error is wasteful and would not be done unless we had reason to compare sampling and experimental error.

It is useful to think of the progression from the completely randomized to the randomized block as a restriction on randomness, in which certain degrees of freedom are withdrawn from error and given the job of isolating certain (potential) systematic effects, thus preventing them from entering into our contrasts and into error.

It should not be necessary to emphasize that the criteria for blocking must be based on dependable knowledge, presumably gained through prior experience. Irresponsible blocking can be costly and even calamitous.

Paired comparisons

An extreme form of the randomized block design arises when each block contains two observations, one on each of two treatments. This case is often discussed separately under the heading paired comparisons.

If x_{ij} is the observation on treatment i and in block j, then $d_j = x_{1j} - x_{2j}$ may be regarded as an observation on a population with mean $\mu_1 - \mu_2$ and variance $2\sigma^2$. The estimate of $2\sigma^2$ is

$$s^2 = \Sigma(d_j - \bar{d})^2/(r-1) \quad \text{and}$$

$$\frac{\bar{d} - (\mu_1 - \mu_2)}{s/\sqrt{r}} \quad \text{is} \quad t_{(r-1)} .$$

From this we get at once a test of significance for $\mu_1 - \mu_2 = 0$, or indeed any speculation about $\mu_1 - \mu_2$.

The only point of interest here is that the test can be seen to be unaffected if we drop the assumption of equal variances and independence. If x_{1j} and x_{2j} are observations on a bivariate normal with variances and covariance σ_1^2, σ_2^2, σ_{12} then the d_j may be regarded as an observation on a population with mean $\mu_1 - \mu_2$ and variance $\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}$. These considerations extend easily to those rare instances in which treatments may affect the variance, when we have more than two treatments. For example, we might study $x_{1j} + x_{2j} - 2x_{3j}$, or any other contrast we may wish to inspect.

The factorial arrangement

We have seen, up to this point, two types of experimental design, the completely randomized and the randomized block. The difference between them lies in the definition of error and the provision made to exhibit and estimate it.

The term factorial, on the other hand, refers to the particular character of the treatments. We envisage two or more factors each at several "levels". The treatment combinations are formed by combining each level of each factor with every level of every other factor. They may be tested in a completely randomized design, randomized block or any other. Our concern is with the interpretation of such treatment differences as may appear.

Only one situation will be discussed, that in which one, at least, of the factors has "levels" that can be specified numerically. Then, graphs may be plotted and the analysis amounts to an elucidation of these graphs.

The example discussed earlier may be adapted to make an illustration. If diet 1 is some standard diet and diet 2 is diet 1 plus one unit of some additive, the responses to the treatments (averages or totals) may be plotted. This graph will depict only the raw data, not any attempt at analysis or reduction.

If we think of curves joining the points (in this instance, they must be straight lines), the interaction component displays the difference in slope of the M and F lines. If this component is significant no further reduction is warranted.

If the interaction component is not significant, the main effects become meaningful and further reduction can be made.

This diminutive example is too small to exhibit some features of the analysis of factorial experiments. If we think of adding a third diet made up by adding two units of the additive, we get 6 totals or averages, which we could be plotted on a graph. We have then to ask: (1) are these curves parallel in the sense that one can be superposed on the other by a vertical displacement? If they are, (2) is the average curve straight or curved and (3) what is the distance between the two response curves?

The following transformation contains the answers to these questions. It is assumed that these contrasts are tested in some suitable design that supplies an estimate of error, and that each treatment is tested n times. Only the portion of the transformation that displays contrasts among the treatments is written.

		T_{11}	T_{12}	T_{13}	T_{21}	T_{22}	T_{23}	division
sexes	y_2	1	1	1	-1	-1	-1	$\sqrt{6n}$
diets L	y_3	-1	0	1	-1	0	1	$\sqrt{4n}$
diets Q	y_4	-1	2	-1	-1	2	-1	$\sqrt{12n}$
	y_5	-1	0	1	1	0	-1	$\sqrt{4n}$
	y_6	-1	2	-1	1	-2	1	$\sqrt{12n}$

		diet 1	2	3	
sexes	1	T_{11}	T_{12}	T_{13}	$T_{1.}$
	2	T_{21}	T_{22}	T_{23}	$T_{2.}$
		$T_{.1}$	$T_{.2}$	$T_{.3}$	G

	df	ss	
sexes	1	$y_2^2 = (T_{1.}^2 + T_{2.}^2)/3n - G^2/6n$	
diets	2	$y_3^2 + y_4^2 = \Sigma\, T_{.j}^2/2n - G^2/6n$	y_3^2 - from transformation; y_4^2 - by subtraction
$s \times d$	2	$y_5^2 + y_6^2$ by subtraction	y_5^2 - from transformation; y_6^2 - by subtraction

The coefficients in y_3 and y_4 come from a table of values of the orthogonal polynomials.

If neither of y_5^2 and y_6^2 is significant, we test y_4^2. If not significant, test y_3^2. These have to do with the shape of the average curve.

Although the examples used here are small, factorial arrangements with several factors each at several levels can be very large indeed, which can cause difficulties for several reasons. Several devices for circumventing these difficulties are available.

Incomplete blocks - confounding

The randomized block arrangement is, in a way, a standard and preferred design, to be used whenever possible. One reason why it may not be possible, particularly in large experiments, is that natural and sensible blocks are not large enough to accommodate all the treatments. In consequence, a replication will require two or more blocks. Each block is then said to be incomplete. The allocation of treatments to blocks and the consequences of various ways of doing so becomes an extensive and acutely important question. It will be pursued here for the case where the treatments arise from a factorial arrangement.

To adopt the earlier example, with two sexes and two diets, thus 4 treatments, let us say that litters of 4 are not available, and that litters of 2 are in good supply. Then, 2 litters will be required for one replication. Evidently, there are several ways in which diets can be introduced into this system. Let us pursue the consequences of some of these.

Think of using litters of 2 M and 2 F. We have still some choice about how diets are to be introduced.

Arrangement 1

	litter 1		litter 2		litter 3		litter 4	
	M	M	F	F	M	M	F	F
	diet	1	diet	2	diet	2	diet	1
sexes	1	1	-1	-1	1	1	-1	-1
diets	1	1	-1	-1	-1	-1	1	1
$s \times d$	1	1	1	1	-1	-1	-1	-1

Here, it is seen that all the treatment contrasts also gather up litter (block) differences. In these circumstances, it is said that treatments and litters are confounded. The only contrasts perceptible within litters are differences among animals treated alike, a sampling error, which is unsuitable as an error term for testing treatments. Obviously, this is an unacceptable arrangement.

Arrangement 2

	Rep 1				Rep 2			
	litter 1		litter 2		litter 3		litter 4	
	M	M	F	F	M	M	F	F
diets	1	2	1	2	1	2	1	2
s	1	1	-1	-1	1	1	-1	-1
d	1	-1	1	-1	1	-1	1	-1
$s \times d$	1	-1	-1	1	1	-1	-1	1

Here, we see that sexes and litters are confounded and that diets and $s \times d$ are not confounded with litters. This may well be an acceptable state of affairs and worth pursuing. The full transformation is given below.

	Rep 1				Rep 2			
	litter 1		litter 2		litter 3		litter 4	
	M	M	F	F	M	M	F	F
	1	2	1	2	1	2	1	2
reps	1	1	1	1	-1	-1	-1	-1
sexes	1	1	-1	-1	1	1	-1	-1
diets	1	-1	1	-1	1	-1	1	-1
$s \times d$	1	-1	-1	1	1	-1	-1	1
$r \times s$	1	1	-1	-1	-1	-1	1	1
$r \times d$	1	-1	1	-1	-1	1	-1	1
$r \times s \times d$	1	-1	-1	1	-1	1	1	-1

Inspection shows that three of the components, reps, sexes, $r \times s$ are confounded with blocks, the rest are not. The analysis splits into two parts, among blocks and within blocks.

	d.f.	ss
among blocks	3	$\frac{1}{2} \Sigma B_k^2 - G^2/8$
within blocks	4	by subtraction
total	7	$\Sigma\Sigma x^2 - G^2/8$

One or the other of two views may be suitable here. If, for any reason, we are content to ignore contrasts that have been confounded with blocks, we proceed with the within-blocks analysis. We could use the transformation to make these computations, but it is simpler to set up a sex-diet table.

diet

sex	1	2	
1	T_{11}	T_{12}	$T_{1\cdot}$
2	T_{21}	T_{22}	$T_{2\cdot}$
	$T_{\cdot 1}$	$T_{\cdot 2}$	G

sexes	$\frac{1}{4} \Sigma T_{i\cdot}^2 - G^2/8$
diets	$\frac{1}{4} \Sigma T_{\cdot j}^2 - G^2/8$
$s \times d$	by subtraction
treatments	$\frac{1}{2} \Sigma T_{ij}^2 - G^2/8$

within blocks	4	diets	1	
		$s \times d$	1	
		error	2	by subtraction

If we wish to inquire into sexes and if we can regard reps × sexes as error, we can separate

	reps	1	$\frac{1}{4}\Sigma T_r^2 - G^2/8$
among blocks 3	sexes	1	from earlier calculation
	$r \times s$ (error)	1	by subtraction

In this case, of course, with 1 df for error, this is ridiculous.

If we choose to use litters containing one male and one female, we might adopt the following arrangement

	litter 1		litter 2		litter 3		litter 4	
	M	F	M	F	M	F	M	F
	1	2	2	1	1	2	2	1
sexes	1	-1	1	-1	1	-1	1	-1
diets	1	-1	-1	1	1	-1	-1	1
$s \times d$	1	1	-1	-1	1	1	-1	-1

Here, $s \times d$ is confounded with blocks, sexes and diets are not.

One sometimes reads or hears that when one is faced with the need to sacrifice certain components in a factorial arrangement, through confounding with blocks, one should confound interactions. If the interactions to be confounded are all of high order, there is good reason for this, because there are good grounds for thinking that genuine interaction involving many factors (or more than 3 or 4?) rarely exist. Experience seems to confirm this expectation.

On the other hand, it may happen that it is preferable to confound a main effect than to sacrifice a low-order interaction.

The splitting of an experiment into two or more sections, each with its own error term, can occur in a variety of situations.

The split plot experiment

It happens not infrequently that some of the factors are such that different levels can be applied to the same experimental unit, whereas the various levels of other factors require different units.

If, in the small example used earlier, we compare, instead of diets, two treatments which can be applied to the same animal (say, applications on the skin, one on each side), still taking the view that the two sexes may yield different results and that litters exercise experimental control, we could get a number of litters, each made up of 1 male and 1 female apply both treatments to each animal. If we use r litters, the analysis of variance would read:

	d.f.		
		reps	r-1
among animals	$2r$ - 1	sexes	1
		$r \times s$	r-1 (error)
		treatments	1
		$s \times t$	1
within animals	$2r$	$r \times t$	$2(r-1)$ (error)
		$r \times s \times t$	

The experimental arrangement is structurally the same as the confounded arrangement discussed earlier. In this case, though, the main effect (sexes) is necessarily the one to be confounded with animals.

Another example, in agriculture, could arise in the following way. Let us say that a number of varieties (v) are to be tested in a randomized block with r replications. Let us say further that, in each

replication, each of the v plots is subdivided into t sub-plots, each of which is to receive, by random allocation, a different fertilizer treatment.

It is to be expected the variety contrasts will be exposed to larger errors than the fertilizer contrasts, in view of the larger plots used for varieties. The $r\ v\ t - 1$ d.f. will therefore be assembled as follows.

among large (main) plots			$rv - 1$
within main plots i.e.			
among small (sub) plots			$rv(t-1)$
within main plots			
		reps	$r - 1$
main plots $rv - 1$		varieties	$v - 1$
		$r \times v$	$(r-1)(v-1)$ (error)
		fertilizer	$t - 1$
sub plots $rv(t-1)$		$v \times f$	$(v-1)(t-1)$
		$r \times f$	$(r-1)(t-1)$
		$r \times v \times f$	$(r-1)(v-1)(t-1)$ (error)

The split plot arrangement is sometimes criticized because a main effect is confounded with the large main plots. This is, of course, true, but this is not necessarily a weakness and, in any event, it is often unavoidable.

The split plot arrangement does occur more often than one might expect. It is frequently the most natural arrangement and is sometimes quite inescapable.

We can, of course, have split-split ... plots. They are sometimes called nested designs.

More on factorial experiments

When several factors or several levels are used, the number of treatments can be large. The total number of tests required, especially when replicated, can be embarrassing and the control of heterogeneity can be difficult.

Several devices can be used here, each of them reflecting the hierarchy that exists among the contrasts and the expectation that interactions among several factors do not, in fact, exist.

The use of incomplete blocks, with which are confounded selected contrasts (preferably only high order interactions), has already been mentioned. This notion will be pursued further in a few special instances.

The 2^n factorial

Experiments of this sort, often with quite large numbers of factors, are frequently used for screening, to select certain factors for further study, to obtain qualitative information about the various factors and their interactions.

To make a start, think of a 2^3 experiment, with levels (a_0, a_1), (b_0, b_1), (c_0, c_1). There will be 8 treatments $a_i\ b_j\ c_k$, $i, j, k = 0, 1$. Another symbolism used here treats a_0, b_0, c_0 as if they represent absence of the factor (this need not actually be so) and a etc. as the presence of a given amount of the factor. Then, using (1) to be $a_0 b_0 c_0$, we get

$(1) = a_0\ b_0\ c_0$, $(a) = a_1\ b_0\ c_0$, $(b) = a_0\ b_1\ c_0$, $(c) = a_0\ b_0\ c_1$

$(ab) = a_1\ b_1\ c_0$, $(bc) = a_0\ b_1\ c_1$, $(ca) = a_1\ b_0\ c_1$ $(abc) = a_1\ b_1\ c_1$.

A transformation (of the treatment totals) to display the main effects and interactions shows that each of the components is a contrast between one set of four treatments and the other four. Indeed, these contrasts may be enumerated by setting up a generating function.

$1 = (a+1)(b+1)(c+1)$ $\qquad AB = (a-1)(b-1)(c+1)$

$A = (a-1)(b+1)(c+1)$ $\qquad BC = (a+1)(b-1)(c-1)$

$B = (a+1)(b-1)(c+1)$ $\qquad CA = (a-1)(b+1)(c-1)$

$C = (a+1)(b+1)(c-1)$ $\qquad ABC = (a-1)(b-1)(c-1)$

If, now, we raise the question of subdividing the blocks, to obtain smaller, incomplete blocks. Then, at least if the blocks are to be all of the same size, we must think either of 2 blocks of 4 or 4 blocks of 2.

If we think of 2 blocks, each to contain 4 treatments, the transformation shows at once the allocation of treatments to blocks that will confound any particular component. For example, if we wish to confound *ABC* with blocks, we put those treatments which receive a +1 in *ABC* into one block and those which receive a -1 into the other. Then, *ABC* becomes confounded with block 1 - block 2. Thus, in each replication we would allocate:

block 1 (a), (b), (c), (abc)

block 2 (1), (ab), (bc), (ca).

Allocation to units within blocks would, of course, be random. In some circumstances, we would also decide randomly which block get which set of treatments.

	(a)	(b)	(c)	(abc)	(1)	(ab)	(bc)	(ca)
A	1	-1	-1	1	-1	1	-1	1
B	-1	1	-1	1	-1	1	1	-1
C	-1	-1	1	1	-1	-1	1	1
AB	-1	-1	1	1	1	1	-1	-1
BC	1	-1	-1	1	1	-1	1	-1
CA	-1	1	-1	1	1	-1	-1	1
ABC	1	1	1	1	-1	-1	-1	-1

The same allocation of treatments to blocks could be reached without the transformation by writing out

$$ABC = (a-1)(b-1)(c-1) = abc - ab - bc - ca + a + b + c - 1$$

If we now think of dividing each replication into 4 blocks, each to receive 2 treatments, then at least three of the treatment contrasts must be confounded with blocks.

The arrangement

$$[(a) \quad (b)], \quad [(c) \quad (abc)], \quad [(1) \quad (ab)], \quad [(bc) \quad (ca)]$$

is seen to confound ABC, AB, C with blocks. The remaining components are computed within blocks.

It appears that we could choose any two of these components to be confounded with blocks and then the third would be determined. Some general considerations bear on this.

If we look at the rows of this transformation (8-tuples) as simply a set of elements, with multiplication defined as we have been using it, we see that the system has the following properties.

1. If two elements, R_1 and R_2, belong to the set, so does $R_1 R_2$.
2. The set contains an identity, 1, such that $R_1 1 = R_1$.

The set defines a group, under the operation we have called multiplication. The group may be specified by any four independent elements (generators).

We can speak of producing an "interaction" of any two elements R_1 and R_2 by multiplying them to get $R_1 R_2$.

By reverting to our original notation, we can compute the interaction of AB and ABC as $(AB)(ABC)$, which is seen from the transformation to be C. Now, for any element R, $R^2 = RR = 1$. In the above product, we could, acting as if the usual rules are applicable, write

$(AB)(ABC) = (AB)^2C$ or A^2B^2C which is simply C.

In the example we have been discussing, when we made the allocation into blocks that confounded C and ABC, their interaction, $(ABC)(C) = ABC^2 = AB$ was also confounded.

The elements C, AB, ABC form a sub-group of the 2^3 group.

With larger 2^n factorial, it is often possible to use confounding in which only high-order interactions are confounded. Plans are listed in such places as the Fisher and Yates tables and Cochran and Cox.

Double Confounding

R.A. Fisher, in his Design of Experiments, points out that, in some circumstances, one sub-group of treatment effects can be confounded with one type of heterogeneity and another subgroup with another. His example envisages a 2^7 factorial, i.e. 128 treatment combinations, to be applied (presumably) to the skins of cows. Each cow provides 8 sites, so that 16 cows are required for one replication. Each cow then becomes an incomplete block. The question is : can we find an incomplete block arrangement in which a controlled set of

contrasts is perceived also within sites?

To provide a simpler example, suppose we have a 2^3 factorial and that each cow provides 2 sites then 4 cows are needed for one replication.

We may select the sub-group A, BC, ABC to be confounded with cows, leading to the blocks $[(a), (abc)]$, $[(b), (c)]$, $[(ca), (ab)]$, $[(bc), (1)]$.

With respect to the other source of heteogeneity, sites, we may choose to try an independent sub-group, say AB, to confound with sites

$$AB = \{(abc) + (ab) + (c) + (1)\} - \{(bc) + (ca) + (a) + (b)\} .$$

The following transformation displays the consequences of this double confounding

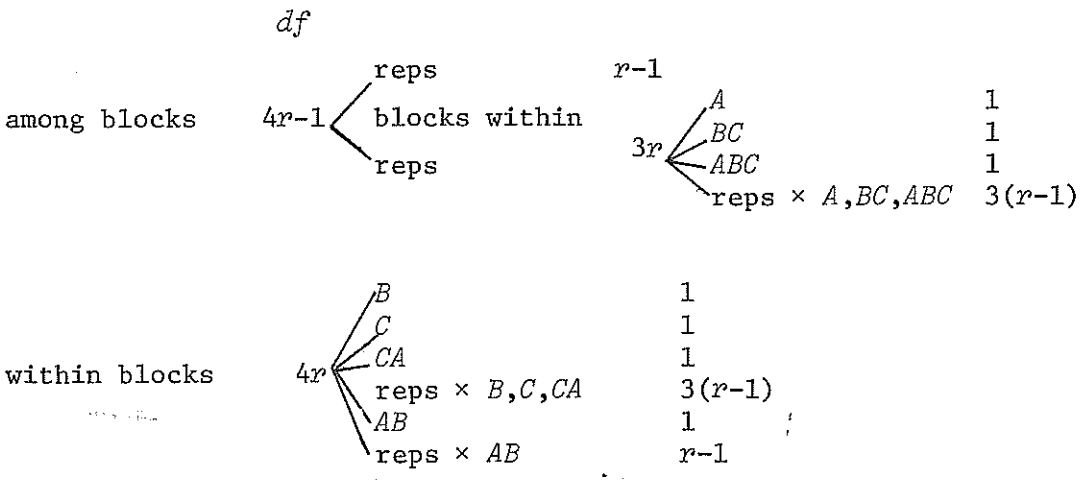
	Block 1		Block 2		Block 3		Block 4	
	Site 1	Site 2	Site 1	Site 2	Site 1	Site 2	Site 1	Site 2
	(a)	(abc)	(b)	(c)	(ca)	(ab)	(bc)	(1)
A	1	1	-1	-1	1	1	-1	-1
B	-1	1	1	-1	-1	1	1	-1
C	-1	1	-1	1	1	-1	1	-1
AB	-1	1	-1	1	-1	1	-1	1
BC	1	1	-1	-1	-1	-1	1	1
CA	-1	1	1	-1	1	-1	-1	1
ABC	1	1	1	1	-1	-1	-1	-1

We see that A, BC and ABC are confounded with blocks. AB is confounded with sites. The remaining components, B, C, CA are computed within blocks and within sites. Of course, these components have the structure blocks × sites, e.g. $B = (A)(AB)$, so one would need assurance that such interactions do not exist. In this example, replication would

be required to define and estimate error and, in any event, in this small example there is too much confounding to be generally tolerated. In Fisher's example, only interactions with several factors are confounded.

We are here approaching combinatorial questions of the sort encountered in latin squares.

If this plan is replicated r times, the analysis of variance would read, for those instances in which we would regard components confounded with blocks and sites to be not wholly lost and in which appropriate randomness has been introduced:



Partial Confounding

It may happen, particularly when the factorial is small, that we are unwilling to forego all information on any component through confounding it with blocks in every replication. For example, in the 2^3 factorial discussed earlier, conducted in two blocks per replication, we might prefer to confound ABC in one replication and another, say AB, in another. Such an arrangement is easily laid out.

The analysis of variance table may be calculated in a number of ways. The s.s. for *ABC* and *AB*, coming from only a portion of the data, are best computed from the transformation.

blocks	3	the dashes on *AB* and *ABC* are
A	1	simply reminders that they are
B	1	partially confounded.
C	1	
AB'	1	
BC	1	
CA	1	
ABC'	1	
error	5	
total	15	

	Rep 1 (*ABC* confounded)								Rep 2 (*AB* confounded)							
block	2	1	1	2	1	2	2	1	2	1	1	2	2	1	1	2
	(1)	(*a*)	(*b*)	(*ab*)	(*c*)	(*ca*)	(*bc*)	(*abc*)	(1)	(*a*)	(*b*)	(*ab*)	(*c*)	(*ca*)	(*bc*)	(*abc*)
*b*1	−1	1	1	−1	1	−1	−1	1	0	0	0	0	0	0	0	0
*b*1	0	0	0	0	0	0	0	0	−1	1	1	−1	−1	1	1	−1
reps	1	1	1	1	1	1	1	1	−1	−1	−1	−1	−1	−1	−1	−1
A	−1	1	−1	1	−1	1	−1	1	−1	1	−1	1	−1	1	−1	1
B	−1	−1	1	1	−1	−1	1	1	−1	−1	1	1	−1	−1	1	1
C	−1	−1	−1	−1	1	1	1	1	−1	−1	−1	−1	1	1	1	1
AB	1	−1	−1	1	1	−1	−1	1	0	0	0	0	0	0	0	0
BC	1	1	−1	−1	−1	−1	1	1	1	1	−1	−1	−1	−1	1	1
CA	1	−1	1	−1	−1	1	−1	1	1	−1	1	−1	−1	1	−1	1
ABC	0	0	0	0	0	0	0	0	−1	1	1	−1	1	−1	−1	1
rep × *A*	−1	1	−1	1	−1	1	−1	1	1	−1	1	−1	1	−1	1	−1
rep × *B*	−1	−1	1	1	−1	−1	1	1	1	1	−1	−1	1	1	−1	−1
rep × *C*	−1	−1	−1	−1	1	1	1	1	1	1	1	1	−1	−1	−1	−1
rep × *BC*	1	1	−1	−1	−1	−1	1	1	−1	−1	1	1	1	1	−1	−1
rep × *CA*	1	−1	1	−1	−1	1	−1	1	−1	1	−1	1	1	−1	1	−1
ABC	−1	1	1	−1	1	−1	−1	1	0	0	0	0	0	0	0	0
AB	0	0	0	0	0	0	0	0	−1	1	1	−1	−1	1	1	−1
reps	1	1	1	1	1	1	1	1	−1	−1	−1	−1	−1	−1	−1	−1
AB	1	−1	−1	1	1	−1	−1	1	0	0	0	0	0	0	0	0
ABC	0	0	0	0	0	0	0	0	−1	1	1	−1	1	−1	−1	1

Depending on the number of replications used, various other components could be partially confounded as well. If one were willing to carry out enough replications (in this instance, 7) one could arrange that each component is confounded with blocks in one replication only. This is the case of balanced confounding.

Hidden replication

In large factorials, provision is made for inspecting all interactions even though there are strong grounds for believing that they do not in fact exist. We might, therefore, to avoid replicating a large experiment, carry out only one replication, intending to estimate error using the components of interaction involving several factors. Fisher refers to this as hidden replication, not a felicitous title. As a principle for estimating error, it is clearly weaker than that based on genuine replication. It may involve some posterior picking and choosing of error components and needs some restraint to avoid picking the error one likes best. This is not, however, a condemnation of the whole idea. Without it, some experiments would demand so many observations that they would not be undertaken. Indeed, with very large factorials, the number of observations required for even one complete replication is too large and the question arises of carrying out only a subset of the factorial combinations. Considerations that bear on the selection of subsets are explored in the next section.

Fractional factorials (partial replication)

To perceive, in a small example the consequences of testing a subset of the factorial combinations, go back to one of the examples used in discussing confounding in incomplete blocks, say the one in which, with a 2^3 factorial, the *ABC* interaction is confounded with blocks.

If, now, we think of testing only the combinations in one of the blocks (either one), there will be, of course, only 3 d.f. for treatments and we see, from the transformation, that:

1. ABC has disappeared entirely. There is no place in this experiment where a contribution from ABC can be identified. We may say that ABC has been identified with 1 and write $ABC = 1$.

2. BC and A have become identified, i.e. confounded $A = BC$. Similiarly, $B = CA$ and $C = AB$. It is said that A and BC etc. are aliases for one another or that A and BC are aliased, not an attractive or necessary use of the word.

All of these confoundings can be deduced from $ABC = 1$, which is called an identifying relation. For example $A^2BC = A$ whence $BC = A$.

This particular arrangement would be intolerable except in highly special circumstances, owing to the confounding of main effects and first order interactions. In large factorials, however, it is possible to find fractions in which main effects and low-order interactions are confounded only with high-order interactions, which may be presumed not to exist. It may also be possible to arrange that some high-order interactions are confounded only with other high-order interactions and therefore reflect error only and provide an estimate of the error variance.

This approach to selecting fractions of factorials is based on the desire to retain orthogonality. Other criteria are sometimes used, as in the fitting of response surfaces, a topic that will not be pursued here.

Weighing designs

In some special circumstances we may be certain that interactions do not exist and fractional factorial designs, even of low order, can be

useful in cutting down the size of an experiment or in increasing its precision. A prototype of this situation is the weighing design. (Yates)

Suppose we have three objects, a, b, c whose weights are to be measured on a balance. Let us say also that a zero correction is needed, i.e. a reading with both scale pans empty (call it w_1). Then if each object is weighed individually, 4 readings will be needed. If W_a is the actual weight of a and w_a the measured weight of a.

$$w_a - w_1 = W_a + \text{error}$$

$$Var(w_a - w_1) = 2\sigma^2.$$

Let (a) stand for : object a is weighed

(ab) stand for : objects a and b weighed together etc.

(1) : both pans empty.

Then, as first pointed out by Yates, the four weighings (1), (ab), (bc), (ca) yield estimates of W_a, W_b, W_c with half the error variance as would be obtained from individual weighings.

	w_1	w_{ab}	w_{bc}	w_{ca}
	(1)	(ab)	(bc)	(ca)
A	-1	1	-1	1
B	-1	1	1	-1
C	-1	-1	1	1

$$A = -w_1 + w_{ab} - w_{bc} + w_{ca}$$

$$= -w_1 + (W_{ab} + w_1) - (W_{bc} + w_1) + (W_{ca} + w_1) + \text{error}.$$

Now $W_{ab} = W_a + W_b$ (no interaction), etc.

$$A = W_a + W_b - W_b - W_c + W_c + W_a + \text{error}.$$

$$= 2W_a + \text{error}$$

$$Var(A) = 4\sigma^2, \quad Var(A/2) = \sigma^2$$

$W_a = A/2$ + error and the variance with which W_a is estimated is σ^2, half that obtained with individual weighings.

Similarly W_b and W_c are estimated from $B/2$ and $C/2$.

We have, of course, used a ½ fraction of 2^3 factorial, of a sort that confounds main effects only with interactions. The complementary set of weighings could equally well be used, (a), (b), (c), (abc).

These two fractions could be combined if we put objects in both pans, distinguished by + and - .

	$(a) - (bc)$	$(b) - (ca)$	$(c) - (ab)$	$(abc) - (1)$
A	1	-1	-1	1

Then $W_a = A/4$ + error and the variance $a/4$ is $\sigma^2/4$.

Of course, the more often we weigh an object, the more precisely its weight will be estimated. What we have here is simply an economical way of doing so.

The Latin Square Experiment

As a small example, think of an agricultural trial in which three treatments are to be tested in three replications as shown

T_2	T_1	T_3
T_2	T_3	T_1
T_1	T_2	T_3

trend ↓

One could say that this arrangement is buttressed against trends in the direction of the arrow, or that the effects of such a trend have been blocked out, inasmuch as all the contrasts are perceptible within the same levels of the trend. A trend parallel to the blocks, on the other hand, would distort the contrasts and inflate the error.

It is reasonable, therefore, to ask whether it would be possible to gain control over both trends at the same time. That is, is it possible to arrange that the contrasts are made, not only within each row, but within each column as well. Evidently the requirement here is that each treatment appear, within each row and within each column, once and only once, (or at least, the same number of times, if orthogonality is to be maintained) with every other treatment.

The combinatorial question thus raised is easily answered. Enter T_1, T_2, T_3 in the first column, permute cyclically to get the second column and again to get the third. There is no suggestion here that this construction yields all possible arrangements that meet the two conditions. Indeed, it does not.

Such arrangements, called latin square arrangements, may be thought of as arising through an additional restriction on randomness compared with the randomized block design. It may not be obvious, here, just how full randomness, within the restraints, is to be obtained. Evidently, rows, columns and treatments may be permuted randomly without destroying the latin square properties. However, such rearrangements, starting from any given square, cannot yield all possible squares. A deeper inquiry, based on systematic counting of squares, reveals that all possible squares fall into sets (transformation sets) within which these permutations can put any square into any other, but not into a member of any other set. A succinct discussion of these questions, with rules for the introduction of randomness into latin square experiments, is given in the introduction of the Fisher and Yates tables.

It may not be obvious, either, that the latin square arrangement does what it is supposed to do, i.e. provide treatment and

error contrasts free from row and column differences. The transformation below shows that it does in fact do so. $x_{ij(k)}$ represents the observation in row i, column j on the treatment k assigned to this place by the latin square.

	$x_{11(1)}$	$x_{12(2)}$	$x_{13(3)}$	$x_{21(2)}$	$x_{22(3)}$	$x_{23(1)}$	$x_{31(3)}$	$x_{32(1)}$	$x_{33(2)}$
	1	1	1	1	1	1	1	1	1
rows	a_1	a_1	a_1	a_2	a_2	a_2	a_3	a_3	a_3
columns	b_1	b_2	b_3	b_1	b_2	b_3	b_1	b_2	b_3
treatments	c_1	c_2	c_3	c_2	c_3	c_1	c_3	c_1	c_2
error	d_1	d_2	d_3	d_3	d_1	d_2	d_2	d_3	d_1

$$\sum_{i=1}^{3} a_i = 0 \qquad \sum_{i=1}^{3} b_i = 0 \qquad \sum_{i=1}^{3} c_i = 0 \qquad \sum_{i=1}^{3} d_i = 0$$

Each selection of a's etc. can be made in two orthogonal ways.

It is seen that the treatment components sum to zero within each row and column and the error components sum to zero within each row, column and treatment.

Emerging from this discussion, but not dependent on it, is the existence of a pair of latin squares;

c_1	c_2	c_3		d_1	d_3	d_2
c_2	c_3	c_1		d_2	d_1	d_3
c_3	c_1	c_2		d_3	d_2	d_1

which have the property that, when one is superposed on the other, not only does each c and each d appear once in each row and once in each column, but each c is paired once with each d. Such squares are said to be mutually orthogonal. They have led, in this instance, to the

separation of the 4 d.f. for row × column interactions into 2 orthogonal sets of 2 d.f., one of which has been identified with treatments.

This leads to an important reservation covering the use of latin square arrangements. The sources of variation identified with rows and columns must not interact and, by symmetry, must not interact with treatments either. Rows and columns should, ideally, have the nature of replications. They need not, however, be simply topographical stratifications. They may be individuals, days, or whatever kind of blocking is required.

The same kind of discussion for n treatments in a $n \times n$ latin square yields an analysis of variance table:

	d.f.
rows	$n-1$
columns	$n-1$
treatments	$n-1$
error	$(n-1)\ (n-2)$
total	$n^2 - 1$

When a square, orthogonal to that used (sometimes called a graeco-latin square) exists, it could be used to split $n-1$ d.f. out of the error term and identified, if desired, with an additional source of systematic variation, leaving $(n-1)(n-3)$ d.f. for error.

When a complete set of orthogonal squares exists, i.e. $n-1$ squares, orthogonal to each other, the $(n-1)(n-2)$ d.f. may be split into $n-2$ sets of $n-1$ d.f., orthogonal to each other, some of which may be identified with other sources of systematic variation.

The uncovering of sets of orthogonal latin squares is thus seen to be an important combinatorial problem. Again, Fisher and Yates is a good reference.

Sets of latin squares

Often small latin squares do not provide a large enough experiment and must be replicated. If an $n \times n$ square is replicated r times (each with its own randomization), each square provides an analysis of variance.

	d.f.	s.s.
rows	$n - 1$	R_i
columns	$n - 1$	C_i
treatments	$n - 1$	
error	$(n-1)(n-2)$	E_i
total	$n^2 - 1$	

Combining these r analyses into one, we would have, except in special circumstances,

squares	$r - 1$	
rows within squares	$r(n-1)$	ΣR_i
columns within squares	$r(n-1)$	ΣC_i
treatments	$n - 1$	
squares × treatments	$(r-1)(n-1)$	
error within squares	$r(n-1)(n-2)$	ΣE_i
total	$rn^2 - 1$	

One could engage in some debate here about the proper error term, squares × treatments or pooled error within squares or a combination of both. In principle, squares × treatments is correct, but maybe one can be too stiff-minded here.

Cross-over designs (switch-back)

Suppose we have two types of desk calculator to be compared with respect to the speed in making a specified calculation, say the

calculation of the sum of squares of a given list of numbers. Let us say that we propose to test each machine 10 times. There are, of course, several plans we can think of:

1. Get 20 skilled operators and assign them randomly to the 20 tests. This is a completely randomized experiment.

2. Get 10 skilled operators and assign two tests to each, one using machine A, the other B. This is a randomized block design. Presumably, it would be decided randomly which machine each operator tests first.

3. Let us say, now, that it is deemed possible that there is a persistent difference between the first test and the second. If there is, this trend should not be allowed to enter into the comparison of the machines. The allocation of machines to order of test should not be left to randomness, but should be built in as a systematic effect. To maintain orthogonality, 5 of the operators, chosen randomly, would be given the A - B order, the others the B - A order. Note that this arrangement requires that the number of replications be a multiple of the number of treatments.

This arrangement is called a cross-over design. Clearly it identifies "order" with one of the operators × treatments components or, symmetrically, treatments with one of the operator × order components.

To perceive the structure by means of an orthogonal transformation, use 4 operators.

	operator 1		operator 2		operator 3		operator 4	
	first	second	1	2	1	2	1	2
	A	*B*	*A*	*B*	*B*	*A*	*B*	*A*
operators	a_1	a_1	a_2	a_2	a_3	a_3	a_4	a_4
machines	1	-1	1	-1	-1	1	-1	1
op. × mach.	a_1	$-a_1$	a_2	$-a_2$	$-a_3$	a_3	$-a_4$	a_4
order	1	-1	1	-1	1	-1	1	-1

"order" is seen to be component of operators × machines, obtained by choosing $a_1 = a_2 = 1$, $a_3 = a_4 = -1$.

If this arrangement is to succeed, it is necessary that the effect of order is the same for all operators.

The argument here is seen to have some of the features of the latin square and, indeed, we might think of the arrangement as a set of 2 × 2 latin squares sharing the same rows.

operator	1	2	3	4	5	6	7	8	9	10
row 1 (first)	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*
row 2 (second)	*B*	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*	*A*

The analysis of variance would read

operators	9
treatments	1
order	1
error	8
total	19

The 3^n factorial

The 3^n factorial, as does the 2^n factorial, admits a certain amount of "general" theory, which will not be pursued here. In particular, the 3^3 factorial has received considerable discussion, starting with R.A. Fisher (Design of Experiments), probably because it is elegant, in connection with the question of confounding in 3 blocks of 9 observations in each replication. One would hope, in such confounding that only two treatment d.f. need be confounded and that, with a suitable arrangement, these two d.f. would come from the three-factor interaction components.

If the two-factor interactions are to be perceptible within blocks, it is necessary that each ab pair, each bc pair and each ca pair must appear together in each block (the factors are denoted a_i, b_j, c_k, $i, j, k = 1, 2, 3$) thus, the enumeration of the combinations to go in one of the blocks comes down to a latin square arrangement,

	b_1	b_2	b_3
a_1	c_1	c_2	c_3
a_2	c_2	c_3	c_1
a_3	c_3	c_1	c_2

The other two blocks would be obtained from this by cyclical permutation of the symbols $c_1\ c_2\ c_3$.

An orthogonal transformation shows that, in fact, only two d.f. of the three factor interaction are confounded with blocks. Indeed the 8 d.f. for this interaction are split into 4 independent sets of 2 one set of which is confounded.

Another choice of the initial latin square on which this allocation is based leads to the confounding of another pair of these degrees of freedom. There are 4 different latin squares which cannot be put, one into another, by cyclical permutation of c_1, c_2, c_3.

1/3 of a 3^3 factorial and 1/9 of a 3^4

If only the combinations in one of the blocks are tested, we have a fractional factorial with truly horrendous confounding. If we focus on two factors, A and B say, we have

A	2	
B	2	
$A \times B$	4	C 2
		D 2

Now, using the two orthogonal squares, we can identify 2 of the $A \times B$ components with a third factor C and the other pair with a fourth factor D. Through symmetry we see that each main effect is confounded with a portion of each of the two-factor interactions. Reference to the transformation shows that each main effect is also confounded with parts of the higher-order interactions. Only in the most special of circumstances (i.e. no interactions) could the observations be interpreted with confidence.

Dummy comparisons

R.A. Fisher, Design of Experiments, Chapter 8, introduces two new notions, to which he attaches the terms dummy comparisons and interaction of quantity and quality. His arguments are carried by an agricultural experiment, in 5 replications, to study 4 sources of nitrogen; M_1, M_2, M_3, M_4, each at 3 levels of application; l_0, l_1, l_2.

	block 1									block 2									block 3								
	111	122	133	212	223	231	313	321	332	112	123	131	213	221	232	311	322	333	113	121	132	211	222	233	312	323	331
$A(2)$	a_1	a_1	a_1	a_2	a_2	a_2	a_3	a_3	a_3	a_1	a_1	a_1	a_2	a_2	a_2	a_3	a_3	a_3	a_1	a_1	a_1	a_2	a_2	a_2	a_3	a_3	a_3
$B(2)$	b_1	b_2	b_3	b_1	b_2	b_3	b_1	b_2	b_3	b_1	b_2	b_3	b_1	b_2	b_3	b_1	b_2	b_3	b_1	b_2	b_3	b_1	b_2	b_3	b_1	b_2	b_3
$C(2)$	c_1	c_2	c_3	c_2	c_3	c_1	c_3	c_1	c_2	c_2	c_3	c_1	c_3	c_1	c_2	c_1	c_2	c_3	c_3	c_1	c_2	c_1	c_2	c_3	c_2	c_3	c_1
AB, AC, BC	by multiplication																										
$ABC(8)$	d_1	d_3	d_2	d_3	d_2	d_1	d_2	d_1	d_3	d_2	d_1	d_3	d_1	d_3	d_2	d_3	d_2	d_1	d_3	d_2	d_1	d_2	d_1	d_3	d_1	d_3	d_2
	e_1	e_2	e_3	e_3	e_1	e_2	e_2	e_3	e_1	e_2	e_3	e_1	e_1	e_2	e_3	e_3	e_1	e_2	e_3	e_1	e_2	e_2	e_3	e_1	e_1	e_2	e_3
	f_1	f_3	f_2	f_2	f_1	f_3	f_3	f_2	f_1	f_2	f_1	f_3	f_3	f_2	f_1	f_1	f_3	f_2	f_3	f_2	f_1	f_1	f_3	f_2	f_2	f_1	f_3
	g_1	g_1	g_1	g_1	g_1	g_1	g_1	g_1	g_1	g_2	g_2	g_2	g_2	g_2	g_2	g_2	g_2	g_2	g_3	g_3	g_3	g_3	g_3	g_3	g_3	g_3	g_3

The analysis of variance would then normally read

replication	4		
treatment	11	sources	3
		levels	2
		$s \times l$	6
error	44		

We could lay out a transformation of the 11 treatment d.f., following the usual practice.

	$T_{M_1 l_0}$	$T_{M_1 l_1}$	$T_{M_1 l_2}$	$T_{M_2 l_0}$	$T_{M_2 l_1}$	$T_{M_2 l_2}$	$T_{M_3 l_0}$	$T_{M_3 l_1}$	$T_{M_3 l_2}$	$T_{M_4 l_0}$	$T_{M_4 l_1}$	$T_{M_4 l_2}$
sources (3)	a_1	a_1	a_1	a_2	a_2	a_2	a_3	a_3	a_3	a_4	a_4	a_4
levels (2)	p_1	p_2	p_3	p_1	p_2	p_3	p_1	p_2	p_3	p_1	p_2	p_3
$s \times l$ (6)	a_1p_1	a_1p_2	a_1p_3	a_2p_1	a_2p_2	a_2p_3	a_3p_1	a_3p_2	a_3p_3	a_4p_1	a_4p_2	a_4p_3

This transformation, as will be seen, is not correct.

Following Fisher, let us say that the level l_0 is, in fact, a zero application. Then, the combinations $M_1 l_0$, $M_2 l_0$, $M_3 l_0$, $M_4 l_0$ are not different and cannot enter into any component that displays differences among fertilizers. Indeed, the response curves are necessarily curves radiating from a point on the response axis. Clearly they cannot be parallel (the usual notion of interaction) and the meaning of interaction must be reconsidered in this context. This question will be addressed later.

Consider first the modification of the transformation to remove the "dummy comparisons".

	10	11	12	20	21	22	30	31	32	40	41	42
sources (3)	0	a_1	a_1	0	a_2	a_2	0	a_3	a_3	0	a_4	a_4
l_1	0	-1	1	0	-1	1	0	-1	1	0	-1	1
l_2	-2	1	1	-2	1	1	-2	1	1	-2	1	1
$s \times l_1$ (3)	0	$-a_1$	a_1	0	$-a_2$	a_2	0	$-a_3$	a_3	0	$-a_4$	a_4

Clearly, differences among fertilizers can be seen only at levels 1 and 2. Hence, there can be only 3 d.f. for interactions, instead of 6 and the reps × treatments (error) will lose 4 × 3 = 12 d.f. correspondingly. These 12 + 3 = 15 d.f. represent differences among dummies within reps.

The analysis of variance table now reads

reps 4

treatments 8 — sources 3; levels 2 (1, 1); $s \times l$ 3

error (between blocks) 32 (reps × treatments)

error (within blocks) 15 (sampling error)

The computation is obvious. Form a treatment table ignoring l_0:

	l_1	l_2	
M_1	T_{Ml}		T_M
M_2			
M_3			
M_4			

The two-way analysis yields

sources	3
levels	1
$s \times l$	3

The second levels d.f. is best computed from the transformation. The sum of the two error s.s. can be obtained by subtraction. If it is desired to separate the two kinds of error, the s.s. for error within blocks can be calculated directly and subtracted from the whole. Perhaps it is worth remarking that the computation yielding "sources" and that yielding $s \times l$ are of the same kind, the first conducted on $T_{M_i l_1} + T_{M_i l_2}$, the second on $T_{M_i l_2} - T_{M_i l_1}$, this is obvious from the transformation. We can write, if we wish:

$$\text{sources } 3 \quad \frac{1}{10}\Sigma\left[T_{M_i l_2} + T_{M_i l_1}\right]^2 - \frac{1}{40}\left\{\Sigma\left[T_{M_i l_2} + T_{M_i l_1}\right]\right\}^2$$

$$s \times l \quad 3 \quad \frac{1}{10}\Sigma\left[T_{M_i l_2} - T_{M_i l_1}\right] - \frac{1}{40}\left\{\Sigma\left[T_{M_i l_2} - T_{M_i l_1}\right]\right\}^2$$

The interaction of quantity and quality

If the four sources are identical in their action, except for possible differences in potency, they would yield identical response curves if they could be plotted against the amount of active ingredient instead of against the total bulk applied. We would, in such a situation, surely want to say that there is no interaction between sources and levels. The response curves are superposable with a suitable re-scaling. In this case, if two response curves differ, we would expect to find the difference increasing, with increasing levels. Following Fisher, taking $l_0 = 0$, $l_1 = 1$, $l_2 = 2$, we would

expect to find twice the difference at level 2 that we find a level 1. To study whether this is so (i.e. no interaction), we would inspect (difference in response at level 2) - 2 (difference in response at level 1), rather than (difference in response at level 2) - (difference in response at level 1), as would be done in the usual case to inquire into interactions.

The transformation needs modification to represent this altered definition of interaction. We can lay it out as follows

		$T_{M_1 l_1}$	$T_{M_1 l_2}$	$T_{M_2 l_1}$	$T_{M_2 l_2}$	$T_{M_3 l_1}$	$T_{M_3 l_2}$	$T_{M_4 l_1}$	$T_{M_4 l_2}$
levels	1	-1	1	-1	1	-1	1	-1	1
sources	3	a	$2a$	b	$2b$	c	$2c$	d	$2d$
$s \times l$	3	$-2a$	a	$-2b$	b	$-2c$	c	$-2d$	d

The $s \times l$ components have been laid out to reflect the proper interaction components and the s components have been made orthogonal to them. The sum of the s.s. for s and $s \times l$ is not changed.

The s.s. for sources can now be obtained from

$$\frac{1}{10} \Sigma \left(2T_{M_i l_2} + T_{M_i l_1} \right)^2 - \frac{1}{40} \left\{ \Sigma \left(2\, T_{M_i l_2} + T_{M_i l_1} \right) \right\}^2$$

and that for $s \times l$ by subtraction or directly by calculating

$$\frac{1}{10} \Sigma \left(T_{M_i l_2} - 2T_{M_i l_1} \right)^2 - \frac{1}{40} \left\{ \Sigma \left(T_{M_i l_2} - 2T_{M_i l_1} \right) \right\}^2$$

This argument seems a bit loose and, in any event, rather special. It is properly discussed in the context of biological assay.

This situation, in which the usual notion of interaction must be modified, can occur outside the field of bioassay and one should always be on guard. The thing to watch for is, of course, a set of response curves which must radiate from a point.

Variance components

Sometimes sampling is carried out in patterns which resemble the arrangements we have been discussing but with rather different objectives. Consider, for example, a process that produces a continuous flow of iron plates, all supposed to be identical but, in fact, varying from item to item as do all products of an industrial process. Let us say that we select randomly a set of plates and on each measure the thickness of k randomly chosen points. We could then calculate an analysis of variance table of the form

	d.f.
among plates	$n - 1$
within plates	$n(k-1)$

We can, of course, carry out a test of significance to decide whether there is more variation among plates than can be explained by the variation within plates. In any event, our concern would be to characterize this variation quantitatively, to describe a feature of the production process.

One way of making use of such data is to plot them on a control chart, with a view to checking on the fact that the plates come from a stable and unchanging population. The mechanics of this procedure will not be pursued here, but it will be assumed that the process is in control and that the plates, in so far as their average thickness is concerned, may be regarded as coming from a population of plates generated by the process. The objective of the sampling would be to characterize this population and to estimate parameters of this frequency distribution.

With the sampling envisaged here, we could think of an individual measurement as made up of three elements; the mean of the population, the

deviation from this of the average of the individual and the further deviation owing to the variation within the plates. Thus:

$$x_{i\alpha} = \mu + \pi_{i\alpha} + \varepsilon_{i\alpha} \qquad i = 1, \ldots, n \qquad \alpha = 1, \ldots, k$$

We will want to treat π and ε as random variables, hence the necessity for randomness in selecting the plates and in choosing the points within the plates. It will be assumed that π is $N(0, \sigma_\pi^2)$, ε is $N(0, \sigma^2)$ and that π and ε are independent, an important assumption that may not be met.

We wish, then, to estimate μ, σ_π^2 and σ^2. It seems obvious that μ is estimated by the overall average. It is obvious, too, that σ^2 will be estimated from the within-plates s.s. where no variation from π appears. The estimation of σ_π^2 requires some scrutiny.

Think of a suitable transformation that leads to the appropriate analysis of variance table. There will be n-1 components representing contrasts among plates. Each of them will be of the form

$$y = \sum_i a_i \sum_\alpha x_{i\alpha}, \qquad \sum_i a_i = 0$$

$$= \sum_i a_i \sum_\alpha (\mu + \pi_i + \varepsilon_{i\alpha})$$

$$= k \sum_i a_i \pi_i + \sum_i \sum_\alpha a_i \varepsilon_{i\alpha}$$

$$E(y) = 0$$

$$Ey^2 = Var\ y = k^2 \sum_i a_i^2 \sigma_\pi^2 + \sum_i \sum_\alpha a_i^2 \sigma^2 = k\sigma_\pi^2 + \sigma^2,$$

since $k \Sigma a_i^2 = 1$.

The s.s. among samples is the sum of n-1 such y^2. Therefore,

$$E\ \Sigma y^2/(n-1) = k\,\sigma_\pi^2 + \sigma^2$$

	d.f.	s.s.	m.s.	E.m.s.
among plates	n - 1		A	$k\,\sigma_\pi^2 + \sigma^2$
within plates	$n(k\text{-}1)$		W	σ^2

Then, equating m.s. and E.m.s., we get the estimate

$$\sigma_\pi^2 \sim \frac{A-W}{K}$$

Distribution problems for such estimates are difficult and are usually approached through approximations.

These ideas may be invoked with all the patterns in which observations may be taken. They become extremely troublesome when orthogonality is lacking. To extend this example somewhat, we may suppose that the plates come from various batches of steel and that r batches are sampled, n from each. Following the same kind of calculations, we get:

	d.f.	E.m.s.
among batches	r - 1	$nk\,\sigma_\beta^2 + k\,\sigma_\pi^2 + \sigma^2$
among plates	$r(n\text{-}1)$	$k\,\sigma_\pi^2 + \sigma^2$
(within batches)		
within plates	$rn\ (k\text{-}1)$	σ^2

One can, in principle, always regard the symbols in a model as random variables, provided the requisite randomness has been provided in the sampling. Thus we have fixed effect models (model 1), random effects models (model 2) and mixed models. This kind of thinking has run through, if only informally, some of the earlier discussions (split-plot) and bears on the selection of appropriate error terms.

Long term experiments, series or sets of experiments, experiments in which true replication is not possible.

As an example, we may think of an agricultural experiment to compare fertilizers (say), with a view to selecting a fertilizer which performs best over a large area or, as another possibility, an experiment carried out over a succession of years to pick out a fertilizer which performs best in all years. In either instance, we are confronted with a number of experiments, presumably each carried out properly with adequate replication to provide a reasonable estimate of error. This error is, however, local. It cannot be expected to apply over the whole set of experiments and, indeed, it may well vary from one experiment to another. There is no way in which we could think of replicating the whole set of experiments, thus defining a suitable error.

Let us discuss a situation in which p places are chosen. If we are seeking findings which are to apply to a whole region, these places should be chosen randomly from it. At each place, the same experiment is to be carried out, say a randomized block with t treatments and r replications. For each place, then, we will have an analysis of variance

	d.f.	s.s.	m.s.
replications	$r - 1$	R_i	
treatments	$t - 1$	T_i	
error	$(r-1)(t-1)$	E_i	$s_i^2 = E_i/((r-1)(t-1))$

Difficulty may arise when we try to combine these local analyses into one overall analysis. Presumably we would like to have an analysis of variance of the form:

	d.f.	s.s.
reps within places	$p(r-1)$	$\Sigma\ R_i$
places	$p - 1$	
treatments	$t - 1$	
places × treatments	$(p-1)(t-1)$	
error within places	$p(r-1)(t-1)$	$\Sigma\ E_i$

We can, of course, make these calculations, but the table may not be all that meaningful. It may be, for example, that the local errors are genuinely different and should not be pooled. This could be checked by Bartlett's test. This test proceeds as follows.

There are a estimates of variance, $s_i^2 \quad i = 1, \ldots, a$ based on f_i d.f. Compute

$$\bar{s}^{-2} = \sum_i f_i\ s_i^2 / \sum_i f_i$$

$$M = \left(\sum_i f_i\right) \log \bar{s}^{-2} - \sum_i f_i \log s_i^2$$

$$C = 1 + \frac{1}{3(a-1)} \left[\sum_i 1/f_i - \frac{1}{\Sigma f_i} \right]$$

Then, M/C is approximately $\chi^2_{(a-1)}$. The approximation is not good with very small f_i.

In any event, we would no doubt wish to compare the $p \times t$ interaction with the pooled error, to get some indication of this interaction.

If we can properly envisage a structure in which the average response to treatment in place i is

$$\bar{x}_{ij} = \mu + \pi_i + \tau_j + \mu_{ij} + \bar{\varepsilon}_{ij} \ , \ \sum_j \tau_j = 0$$

and if we can regard the μ_{ij} as independent random variables $N(0, \sigma_\mu^2)$ and independent of the ε's, we can compute the expected values of our mean squares as follows

	E.m.s.
treatments	$\sigma^2 + r\sigma_\mu^2 + \frac{rp}{t-1}\sum_j \tau_j^2$
places × treatments	$\sigma^2 + r\sigma_\mu^2$
local error	σ^2

This would dictate a test of significance for treatments using places × treatments as error. There may be real grounds for fears, though, that this model is too simple. It may well be that the $p \times t$ interaction is composed of heterogeneous components. If so, we can only separate the experiment into parts within which these interactions behave in a reasonably uniform fashion. This can be done by choosing a set of orthogonal contrasts among treatments and partitioning the $p \times t$ interactions correspondingly.

To make a diminutive example, take $p = 2$, $t = 3$

	place 1			place 2		
	t_1	t_2	t_3	t_1	t_2	t_3
p	1	1	1	-1	-1	-1
T_1	1	-1	0	1	-1	0
T_2	1	1	-2	1	1	-2
$p \times T_1$	1	-1	0	-1	1	0
$p \times T_2$	1	1	-2	-1	-1	2

We then have

	d.f.		d.f.
T_1	1	T_2	1
$p \times T_1$	1 (error)	$p \times T_2$	1 (error)

Clearly there can be no satisfactory general theory here.

The assumption of constant error variance - transformations

There are some situations in which it is known a priori that the assumption of constant variance is insupportable. For example, think of an experiment to compare a number of ways of planting seedlings, based on plots each of which receives n seedlings. The response will be the number or proportion of surviving seedlings.

Here we are dealing with binomially distributed variables and, strictly speaking, analysis of variance procedures are not applicable. However, we may use them as approximations, usually very good ones, in the same sense as the normal distribution approximates the binomial. However, in this instance, if the treatments do cause different proportions surviving, the variance will change with the treatment.

If each plot contains n seedlings, of which x survive, $E\left(\frac{x}{n}\right) = \pi$, $Var\left(\frac{x}{n}\right) = \pi(1-\pi)/n$. R.A. Fisher raised (and answered) the question: instead of analyzing $\frac{x}{n}$, can we seek some function of x/n which has constant variance, at least approximately? He showed that by replacing $p = x/n$ by $sin^{-1}\sqrt{p}$, the variance is reasonably constant except near $\pi = 0$ and $\pi = 1$.

This fact may be deduced in a direct fashion. Let t be an unbiased estimate of a parameter τ. Let $z = f(t)$ then

$$Var\ z = \sigma_t^2\ f'^2(\tau) + f'(\tau)\ f''(\tau)\ \mu_3 + \ldots .$$

Now, in the usual cases, σ_t^2 is of order $1/n$, μ_3 is of order $1/n^{3/2}$ etc. and to first order,

$$Var\ z = \sigma_t^2\ f'^2(\tau)$$

Applying this to the binomial, $Ep = \pi$, $Var\ p = \pi(1-\pi)/n$. We seek some function $z = f(p)$ such that the variance of z is constant

$$Var\ z = \sigma_p^2\ f'^2(\pi) = \text{constant},$$

i.e. $$f'^2(\pi) = cn/(\pi(1-\pi))$$

A solution of this differential equation is

$$f(\pi) = 2\sqrt{cn}\ sin^{-1}\sqrt{\pi}$$

When n is constant throughout the experiment, we may simply replace p by $sin^{-1}\sqrt{p}$.

A more careful inquiry shows that $p = 0$ and $p = 1$ introduce non-uniformity in variance. It has been suggested that the situation can be met by replacing $sin^{-1}\sqrt{0}$ by $sin^{-1}\sqrt{1/4n}$ and $sin^{-1}\sqrt{1}$ by $\pi/2 - sin^{-1}\sqrt{1/4n}$ (Bartlett).

In the limiting case, when the binomial is effectively Poisson, the above argument yields $z = \sqrt{x}$, x a count.

Not uncommonly, we encounter situations in which the coefficient of variation is reasonably constant, i.e. the standard deviation is proportional to the mean. The same argument applied to this case yields the transformation $z = log\ x$, where x is an observation, usually a measurement.

A particular case of this occurs in experiments conducted to compare variances. Suppose each portion of the experiment yields an estimate s_i^2 of σ_i^2 based on n d.f..

Assuming normality, we have $E\, s_{i\cdot}^2 = \sigma_{i\cdot}^2$, $Var\, s_{i\cdot}^2 = 2\sigma_{i\cdot}^4/n$. Thus $E(s_{i\cdot}^2) \propto \sqrt{Var(s_{i\cdot}^2)}$. Hence the advice: compute the logarithms of the variances and proceed with the analysis of variance of these logarithms.

1. Gains in Weight of Rats in 100 Days on a Stock Ration with various Amount of Gossypol Added

No Gossypol	.04%	.07%	.10%	.13%
228	186	179	130	154
229	229	193	87	130
218	220	183	135	130
216	208	180	116	118
224	228	143	118	118
208	198	204	165	104
235	222	114	151	112
229	273	188	59	134
233	216	178	126	98
219	198	134	64	100
224	213	208	78	104
220		196	94	
232			150	
200			160	
208			122	
232			110	
			178	

(a) Make an analysis of variance for within and between groups.

(b) Sketch a graph of the means.

(c) What should be the next step in the analysis?

2. In the following table are the amounts of fat absorbed by 48 mixes of doughnuts while being cooked. The object of the investigation is to learn whether the various fats are absorbed in significantly different amounts.

Grams of Fat Absorbed by Mixes of
24 Doughnuts during Cooking Period

Fat Number							
1	2	3	4	5	6	7	8
164	172	177	178	163	163	150	164
177	197	184	196	177	193	179	169
168	167	187	177	144	176	146	155
156	161	169	181	165	172	141	149
172	180	179	184	166	176	169	170
195	190	197	191	178	178	183	167

Make an analysis of variance for between and within fats ignoring the fact that the data in different rows were obtained on different days and show that the differences between fats is not significant. Then make an analysis of variance with between days, between fats and error and show that when the variation due to days is eliminated in this way from the error estimate, differences between fats are revealed.

3. The mean and its standard error are quoted below for the tensile strength of samples from each of two closely related timber species, one sample being tested from each tree. How would you proceed to investigate statistically the difference between the two species?

Number of trees	75	27
Mean tensile strength	27.5	32.4
Standard error of mean	2.27	3.15

4. In an investigation into the effect of deficiencies of trace elements on sheep, the following data were obtained:

Control	Cobalt	Copper	Cobalt + Copper
13.2	11.9	14.2	15.0
13.6	12.2	14.0	15.6
11.9	13.9	15.1	14.5
13.0	12.8	14.9	15.8
14.5	12.7	13.7	13.9
13.4	12.9	15.8	14.4

The animals selected were judged to be homogeneous and were allocated at random to the treatments. Carry out an analysis of variance and, further, make a suitable partition of the treatment sum of squares to identify the cause of the significant treatment effect.

5. An inter-laboratory test of a proposed method of determining percent iron was planned. A standard solution containing 2.950 percent iron was prepared, and distributed to two different laboratories. In laboratory I, the two analysts each made 4 independent determinations of percent iron; in laboratory II the three analysts each made 4 determinations.

Their observations were as follows:

Determinations of Percent Iron

	Lab. I		Lab. II		
Analyst	A	B	C	D	E
	2.743	2.873	3.155	2.905	3.045
	3.098	3.012	3.266	3.085	2.961
	2.921	2.966	2.954	2.891	2.974
	3.001	2.796	3.064	3.090	3.023

The experiment was intended to provide answers to the following questions:

(a) Is the proposed method biased, i.e. does it tend to give high (or low) results?

(b) Does one laboratory tend to give higher values than the other?

(c) Are there significant differences between the analysts of Lab I?

(d) Are there significant differences among the analysts of Lab II?

For this purpose, carry out an analysis of variance, partition the sum of squares appropriately, and make significance tests.

6. The data recorded in the following table represent measurements made during an investigation of the influence of annealing temperature upon the density of a high silica borosilicate glass. The annealing treatment was carried out at temperatures from 450 to 625° C. by 25° C intervals and was prolonged until constant density was reached. Tests were carried out on specimens cut from a sample plate, with a random assignment of two specimens to each temperature.

Temperature	450	475	500	525
Density	2.23636	2.23574	2.23544	2.23625
	2.23654	2.23493	2.23486	2.23522
	550	575	600	625
	2.23517	2.23712	2.23699	2.23891
	2.23508	2.23636	2.23797	2.23913

We wish to express density as a polynomial in temperature of degree not higher than 3.

(a) Sketch a graph of the observations.

(b) Do an analysis of variance, separating out orthogonal linear, quadratic, cubic and higher components.

(c) By successive tests, determine the degree of the polynomial which provides a suitable representation of the data.

(d) Estimate the coefficients of the polynomial and plot the curve.

Orthogonal polynomials, $n = 8$

ξ_1'	-7	-5	-3	-1	1	3	5	7
ξ_2'	7	1	-3	-5	-5	-3	1	7
ξ_3'	-7	5	+7	+3	-3	-7	-5	7

For $x = 1, 2, \ldots, 8$; $\xi_1' = 2x - 9$, $\xi_2' = x^2 - 9x + 15$,

$3\xi_3' = 2x^3 - 27x^2 + 103x - 99$.

7. A certain device for the armed services is produced by 10 manufacturers, and they each make 12 such items. The relative effectiveness, on an arbitrary scale, of each item, is given in the attached table. These numbers are, however, unknown and can only be obtained by destructive testing. The services are willing to destroy 10 percent in order to estimate the average effectiveness and the variability in effectiveness.

(a) Compute the mean and the variance for the output of each of the 10 manufacturers.

(b) Choose a random sample of 3 manufacturers. From the output of each of these 3 manufacturers, choose a random sample of 4 items. Tabulate the effectiveness data for this 10 percent sample of the whole population.

(c) Carry out an analysis of variance of the data in the sample.

Effectiveness

Manufacturer

I	II	III	IV	V	VI	VII	VIII	XI	X
15	74	64	48	77	50	1	71	26	64
85	15	10	23	40	100	51	1	75	40
47	100	11	7	69	55	35	76	39	90
13	25	35	71	26	56	87	0	19	40
10	52	49	6	55	96	33	1	26	20
5	82	9	75	6	50	67	42	78	26
65	79	37	0	50	60	89	5	79	18
59	60	68	26	53	39	68	37	89	31
31	53	70	6	31	10	5	11	21	36
91	50	44	15	59	46	46	2	17	44
63	72	75	0	59	35	73	48	35	21
89	99	73	46	72	20	47	55	100	46

8. Double base solid propellants for guns and rockets are a combination of nitrocellulose (NC) and nitroglycerine (NG). The combination is not always stable and, when unstable, some of the NG is slowly extruded from the solid grains and appears on the surface in the form of liquid drops. An experimental propellant has been subjected to various conditions to determine its stability.

Three batches of propellant were used for which the NG content was known accurately. The percentages, by weight, of NG in batches I, II and III were respectively, 8.50, 9.23, and 8.04. From each batch, four five-pound specimens were taken and put in storage under four temperature conditions: constant temperatures of -40^0 F, 70^0 F, and 120^0 F, and a cycling procedure in which the temperature was held alternately at the two extreme temperatures for 24-hr. periods. After 3 months, a few grains of propellant were taken from each specimen, the external liquid NG removed, and the percentage of combined NG remaining in the solid measured. The observations are tabulated below.

Percentage NG remaining in specimens

	Batch	Storage Temperature			
		-40 F	70 F	120 F	cycled
After 3 months	I	8.43	7.24	5.85	6.93
	II	9.01	7.77	6.56	8.12
	III	7.80	6.54	5.28	6.70
After 6 months	I	8.17	7.02	5.39	6.82
	II	9.22	7.50	6.00	7.53
	III	7.86	6.50	5.14	6.65
After 12 months	I	8.31	6.79	4.92	6.42
	II	8.90	7.28	5.52	7.24
	III	7.73	6.18	4.34	6.14

Over

Batches were not expected to interact with the other factors, so it seemed reasonable to use the second order interaction terms as an estimate of error.

(a) Set up an analysis of variance table which separates main effects, first order interactions, and error.

(b) (i) Test the hypothesis that batches do not interact with the other two factors. If not significant, would this result justify the assumption of no real second order interaction? If significant would the result cast doubt on the assumption?

(ii) Test the hypothesis that the average <u>reduction</u> in NG is the same for all batches.

(c) Make significance tests of the time - temperature interaction and, if necessary, of the main effects.

(d) It might be reasonable to assume that, in the period observed, NG is extruded at a constant rate, depending on temperature; i.e., for a fixed temperature, the absolute reduction in NG is proportional to the time elapsed. Can you construct a test of this?

(e) Summarize your conclusions about the stability of the propellant.

9. A 3^2 factorial experiment is carried out in a randomized block design with two replications. Calculate an appropriate analysis of variance with each of the sets of data given below and make the tests of significance that are required. (Assume that the levels of both a and b are equally spaced). Plot the graphs indicated by the analysis.

a)

	Rep. 1			Rep. 2		
	a_1	a_2	a_3	a_1	a_2	a_3
b_1	19.86	26.37	29.72	20.88	24.38	29.64
b_2	15.35	22.82	27.12	15.86	20.98	24.27
b_3	4.01	10.34	15.64	4.48	9.36	14.03

b)

	Rep. 1			Rep. 2		
	a_1	a_2	a_3	a_1	a_2	a_3
b_1	20.15	24.87	30.06	25.44	30.93	35.49
b_2	21.86	29.38	34.78	26.92	34.13	40.72
b_3	21.66	30.59	36.80	25.93	40.04	42.55

10. A 2^3 factorial in blocks of 4 units was replicated 4 times with all interactions partially confounded with blocks.

Rep I				Rep II			
Block 1		Block 2		Block 3		Block 4	
(abc)	31.13	(ab)	23.70	(abc)	28.14	(ab)	27.00
(a)	15.41	(ac)	18.45	(ac)	19.63	(bc)	24.51
(b)	21.72	(bc)	19.13	(b)	18.45	(a)	15.88
(c)	18.12	(1)	9.60	(1)	10.29	(c)	17.15

Rep III				Rep IV			
Block 5		Block 6		Block 7		Block 8	
(abc)	26.11	(ab)	27.63	(abc)	28.70	(ac)	18.18
(bc)	21.07	(ac)	20.69	(ab)	27.40	(bc)	19.75
(a)	13.71	(b)	20.98	(c)	19.96	(a)	11.57
(1)	13.63	(c)	16.67	(1)	14.33	(b)	18.60

Calculate the appropriate analysis of variance for this experiment and make the required tests of significance.

11. A 3^2 factorial experiment replicated twice in blocks of 3 units yielded the observations

Rep I

Block 1		Block 2		Block 3	
(a_1b_1)	19.86	(a_1b_2)	15.35	(a_1b_3)	4.01
(a_2b_2)	22.82	(a_2b_3)	10.34	(a_2b_1)	26.37
(a_3b_3)	15.64	(a_3b_1)	29.72	(a_3b_2)	27.12

Rep II

Block 4		Block 5		Block 6	
(a_1b_1)	20.88	(a_1b_2)	15.86	(a_1b_3)	4.48
(a_2b_3)	9.38	(a_2b_1)	24.38	(a_2b_2)	20.98
(a_3b_2)	24.27	(a_3b_3)	14.03	(a_3b_1)	29.64

Calculate the appropriate analysis of variance for this experiment and make the required tests of significance.

12. In a food research laboratory, an experiment was carried out on a cake icing. The best formulation had been determined, and the object of the experiment was to determine the tolerances of the product to ingredient variations (factors C through H) and to preparation variation, (water-level-AB). The variables were:

AB (water - 4 equally spaced levels), C (sugar type 1 - 2 levels), D (sugar type 2 - 2 levels), E (stabilizer #1 - 2 levels), F (stabilizer #2 - 2 levels), G (stabilizer #3 - 2 levels) and H (setting agent - 2 levels).

The response was a measure of the viscosity.

A 1/4 replicate of a 4×2^6 factorial was set up using -ACDFG and ABDEFH as defining contrasts. The 64 treatment combinations and the coded responses, were as follows:

Treatment combinations for a 1/4 replicate of a 4×2^6 factorial and the responses (in parentheses).

(1)	(26)	agh	(6)	bh	(43)	abg	(-3)
cg	(16)	ach	(10)	bcgh	(69)	abc	(-5)
dgh	(12)	ad	(13)	bdg	(45)	abdh	(-13)
cdh	(22)	acdg	(17)	bcd	(45)	abcdgh	(-4)
eh	(29)	aeg	(13)	be	(54)	abegh	(4)
cegh	(30)	ace	(17)	bceg	(54)	abceh	(5)
deg	(29)	adeh	(16)	bdegh	(43)	abde	(-2)
cde	(34)	acdegh	(16)	bcdeh	(67)	abcdeg	(-3)
fgh	(32)	af	(19)	bfg	(64)	abfh	(6)
cfh	(30)	acfg	(18)	bcf	(57)	abcfgh	(6)
df	(27)	adfgh	(29)	bdfh	(50)	abdfg	(6)
cdfg	(35)	acdfh	(22)	bcdfgh	(53)	abcdf	(7)
efg	(53)	aefh	(29)	befgh	(74)	abef	(8)
cef	(46)	acefgh	(21)	bcefh	(73)	abcefg	(13)
defh	(35)	adefg	(23)	bdef	(69)	abdefgh	(20)
cdefgh	(42)	acdef	(27)	bcdefg	(69)	abcdefh	(10)

(a) Tabulate the aliases of all main effects and 2-factor interactions in the design.

(b) Carry out an analysis of variance separating out components for each of the main effects, for the 2-factor interactions, and residual.

(c) Assuming that 3-factor and higher-order interactions are negligible, make the appropriate tests of significance.

(d) For the product to be acceptable, the average response must lie between 25 and 30. Give a rough indication of how much each of the ingredients and the water-level can be permitted to vary during production.

13. An experiment with shaped charges of a particular size and design was carried out to study the way in which mean penetration into armour plate depended on cone thickness. Later, the experiment was repeated 3 times, with different combinations of fuses and explosives to determine whether the dependence relation was affected by these factors. Unfortunately, the four replications were carried out at widely separated times with the result that a different batch of armour plate was used in each replication. Three cone thicknesses, proportional to the numbers 2, 3 and 4, were used. Two fuses and two explosives were tested.

Results were as follows

Penetration: Mean of 10 rounds. (Coded data)

Batch of Armour plate	Fuses	Explosive	Cone thickness		
			2	3	4
1	I	I	103	111	101
2	I	II	94	104	93
3	II	I	101	106	111
4	II	II	90	95	94

Note that both fuse and explosive comparisons are subject to both residual variation and to variation among batches of armour plate; but that cone thickness comparisons are subject only to residual variation. Assume that armour plate does not interact with the other factors.

Over

(a) Set up an analysis of variance table separating out the main effects and interactions of fuse, explosive and cone thickness.

(b) What do you use as estimates of (i) residual variation, and (ii) variation due to armour plate?

(c) Make significance tests to obtain answers to the following questions.

(i) Does cone thickness interact with fuse?

(ii) Does cone thickness interact with explosive?

(iii) Does cone thickness affect mean penetration? (On the average over the observed conditions.)

(iv) Is the difference in fuses negligible?

(v) Is the difference in explosives negligible?

14. The yield of a certain product of a chemical reaction is believed to depend on the temperature at which the reaction takes place. Discrepancies in the work of previous investigators have been attributed to minor differences in the apparatus used, and to differences in individual technique. Therefore, a Latin square design is chosen in which 4 men (rows) conduct the experiment at 4 temperatures (treatments) using 4 pieces of apparatus (columns).

The letters A, B, C, D in the Latin square below represent temperatures of 100^0, 200^0, 300^0, 400^0 respectively.

Treatment and Response, y

A	8.53	D	10.39	B	7.06	C	13.89
B	9.77	C	12.24	A	5.29	D	11.50
C	13.62	A	8.70	D	9.14	B	11.59
D	11.97	B	10.25	C	9.29	A	6.17

(a) Compute the mean yield for each treatment and plot against temperature.

(b) Set up an analysis of variance table with row, column, treatment and residual components.

(c) Partition the treatment sum of squares into 1 - d.f. components and, by successive testing, determine the degree of the polynomial that adequately represents the yield -- temperature relation. Estimate the polynomial.

(d) It is possible that there is on interaction between men and apparatus. Can you separate out a 1 - d.f. component from the residual in order to test this?

15. A biological assay of the digitalis-like principle of ouabain and other cardiac substances used a technique of slow infusion of a suitable dilution of the drug into an anaesthetized cat. When death occured, the total dose was measured and recorded. Three observers collaborated in a test of 12 drugs, A, B, ..., L. Each drug was tested on 12 cats. One observer could only test four cats per day, however, and to balance out possible observer and day-to-day differences, a 12 x 12 randomly-selected Latin square was used. The results are given in the table below. The columns represented the dates on which observations were made. The rows represent the first and second cat (a, b) tested by each of the observers (I, II, III) in the morning and afternoon of each day. The drug used is entered in each cell, together with the response to be analyzed: 100 x log dose in μg.

(a) Set up an analysis of variance table and test for differences among

(i) drugs, (ii) days, (iii) observers and (iv) between morning and afternoon.

(b) Use a procedure to compare drugs two at a time and state which pairs of drugs are significantly different.

Over

Tolerance of Cats for various cardiac Substances

			6/3	7/3	8/3	9/3	13/3	14/3	16/3	21/3	24/3	27/3	30/3	3/4
a.m.	I	a	I 525	J 273	B 315	L 557	H 189	G 228	F 54	K 473	D 165	E 254	A 193	C 358
a.m.	I	b	K 737	G 345	J 195	H 425	I 444	B 350	L 557	C 209	E 335	F 22	D 605	A 237
a.m.	II	a	B 293	L 427	G 371	C 413	D 515	J 307	K 446	E 301	H 266	A 368	F 198	I 515
a.m.	II	b	E 299	D 437	F 411	G 400	J 173	K 661	A 250	L 449	C 573	I 316	B 347	H 228
a.m.	III	a	C 601	K 398	A 400	B 502	F 385	L 443	I 329	D 394	G 444	H 377	J 307	E 247
a.m.	III	b	F 35	H 355	K 384	E 451	G 378	C 444	D 394	B 211	A 218	L 442	I 473	J 239
p.m.	I	a	J 376	C 540	E 350	K 512	A 501	I 674	H 256	F 126	B 253	G 336	L 373	D 132
p.m.	I	b	D 313	F 284	I 348	A 326	L 537	E 501	C 523	G 402	J 199	B 270	H 387	K 473
p.m.	II	a	A 309	B 294	C 446	D 336	E 349	F 283	G 322	H 377	I 477	J 305	K 650	L 580
p.m.	II	b	H 261	E 419	L 625	J 368	C 426	A 460	B 211	I 348	K 716	D 289	G 402	F 181
p.m.	III	a	G 363	I 606	D 651	F 360	K 453	H 336	J 185	A 205	L 437	C 632	E 337	B 167
p.m.	III	b	L 521	A 387	H 326	I 692	B 461	D 369	E 348	J 313	F 139	K 439	C 447	G 398

16. The table below gives the results of analyses of samples of cheese for moisture content.

Moisture Content of 2 Cheeses from each of 3 different Lots, determined 2 times

Cheese	Lot		
	I	II	III
1	39.02 38.79	35.74 35.41	37.02 36.00
2	38.96 39.01	35.58 35.52	35.70 36.04

Estimate the components of variance for Lots, Cheeses within lots, determinations within cheeses.

Supposing that the cost factors per lot, cheese and determinations are in the ratios 10 to 3 to 1, what is the most economical allocation of sampling effort to produce an estimate with variance 1.0907? What is the cost effective way of sampling if it is to cost no more than $100, if the 10, 3, 1 above are actual dollar costs.

17. In manufacturing certain articles by sand-casting, it is observed that the composition of the alloy used varies from one article to the next, and even within the same casting. An experiment is conducted to locate the source of variation and to estimate their effects. Four batches of alloy are mixed in the furnace: from each batch, four castings are made, and three specimens to be analyzed are taken from each casting.

The percentages of lead in the specimens are tabulated below:

Batch No.	Mold No.	% Lead		
1	11	6.8	6.0	4.8
	12	6.0	4.6	2.8
	13	6.3	4.9	3.0
	14	7.1	3.2	5.0
2	21	10.4	12.3	11.1
	22	12.4	12.4	12.2
	23	12.8	10.6	11.9
	24	13.3	13.8	13.4
3	31	8.6	8.8	9.5
	32	8.4	8.3	8.0
	33	7.7	7.0	8.3
	34	7.6	8.1	6.5
4	41	2.1	4.0	3.4
	42	4.9	4.8	6.2
	43	3.5	3.7	4.1
	44	5.7	2.4	2.4

Over

Assume the model

$$y_{ijk} = \mu + \alpha_i + \beta_{ij} + \gamma_{ijk}$$

$$i, j = 1, 2, 3, 4; \quad k = 1, 2, 3$$

$$\alpha_i : N(0, \sigma_\alpha^2), \quad \beta_{ij} : N(0, \sigma_\beta^2), \quad \gamma_{ijk} : N(0, \sigma_\gamma^2) .$$

Carry out an analysis of variance.

(a) Derive the expected mean squares for batches, molds and residual.

(b) Test the hypotheses

(a) $\sigma_\beta^2 = 0$, (b) $\sigma_\alpha^2 = 0$.

(c) Estimate the important variance components.

(d) Summarize your conclusions, interpreting them in terms of the physical situation. Can you make any recommendations or suggestions concerning the best way to go about reducing the variability of the final product?

18. A chemical paste is made in batches and put into casks. As a routine, three casks selected at random from each delivery were sampled, and the samples were kept for reference. It was desired to estimate the variability in the paste strength from cask to cask and from one delivery to another. Ten of the delivery batches were chosen at random and two analytical tests carried out on each of the 30 samples. In order to ensure that the tests were independent, all 60 strength determinations were carried out in a random order. The resulting data are given in the table below. (Delivery = batches)

% Paste Strength of Samples

Batch	Cask 1		Cask 2		Cask 3	
1	62.8	62.6	60.1	62.3	62.7	63.1
2	60.0	61.4	57.5	56.9	61.1	58.9
3	58.7	57.5	63.9	63.1	65.4	63.7
4	57.1	56.4	56.9	58.6	64.7	64.5
5	55.1	55.1	54.7	54.2	58.8	57.5
6	63.4	64.9	59.3	58.1	60.5	60.0
7	62.5	62.6	61.0	58.7	56.9	57.7
8	59.2	59.4	52.2	66.0	64.8	64.1
9	54.8	54.8	64.0	64.0	57.7	56.8
10	58.3	59.3	59.2	59.2	58.9	56.6

From this data estimate variances due to analytical error, variation between casks within batches and the variation between batches.

19. An experiment on the effectiveness of two treatments for a disease was carried out in each of four hospitals. In each hospital, 100 patients received treatment A, 100 received treatment B, and 100 served as controls.

Number of patients (out of 100) showing complete recovery within three months:

Hospital	Treatment A	Treatment B	Control
I	19	18	5
II	19	9	5
III	23	8	5
IV	6	15	11

We wish to make inferences about all hospitals of which the four participating in the experiment are a sample; but only about the particular treatments used. (Mixed model).

Transform the data to stabilize the variance, and make significance tests to get answers to the following questions:

(a) Do hospitals and treatments interact?

(b) Do hospitals affect a recovery rate at all?

(c) Are the treatments effective in increasing the recovery rate?

In each variance ratio used for a test, justify your choice of denominator variance, which is not necessarily the same in all cases.

(d) Obtain a 95 percent confidence interval for the difference in the proportion who recover between treatments A and B.

20. An inter-laboratory test to evaluate and compare methods for determining the proportion of nylon in mixtures.

Three different mixtures were used, all prepared in one of the laboratories to contain known proportions of nylon.

Mixture 1. Nylon and wool, 52.4% nylon.

Mixture 2. Nylon and viscose, 52.5% nylon.

Mixture 3. Nylon and cotton, 50.9% nylon.

Two methods of analyzing the mixtures, for the proportion of nylon in them, were tested, Method F, based on formic acid and Method H, based on hydrochloric acid.

Four laboratories, A, B, C, D participated and each provided two experienced analysts. Each analyst was given randomly selected duplicate pairs of each mixture to be tested by each method. Thus, he was given twelve mixtures, identified only by a code, and a preassigned random order in which they were to be tested. With eight analysts participating in the test, each providing twelve determinations, a total of 96 analyses was reported.

The test was planned having in mind the following sources of variation.

(a) Variation in the test material.

(b) The inability of an analyst to reproduce exactly any determination, using the same materials and working under the same conditions, in so far as they can be controlled.

(c) Differences among analysts working with the same materials under the same conditions.

(d) Differences among analysts working with the same materials under different conditions - i.e. laboratories.

(e) Differences among materials.

(f) Differences between methods.

A statistical analysis of the data should provide information on all these questions.

Lastly, since the true proportions are known, the presence of bias in the methods can be studied.

The results of the analysis, expressed as percent nylon in the mixtures, are given in the following table.

Results of Analyses of Nylon Mixtures - Interlaboratory Trials

METHOD

LAB A

	F		H	
	ANALYST		ANALYST	
MIXTURE	1	2	1	2
1	53.6 54.8	53.1 53.1	53.5 53.0	53.3 53.5
2	50.3 50.9	51.4 51.6	53.2 53.4	53.6 53.1
3	51.8 51.0	51.5 51.8	53.1 53.0	51.7 52.0

LAB B

	ANALYST		ANALYST	
MIXTURE	3	4	3	4
1	53.7 53.6	53.7 53.6	53.6 54.0	53.5 53.4
2	49.3 50.0	47.1 49.4	54.4 54.6	54.2 54.4
3	51.5 51.5	51.7 51.7	52.9 53.0	52.7 52.5

LAB C

	ANALYST		ANALYST	
MIXTURE	5	6	5	6
1	52.7 52.4	52.6 52.3	51.6 55.7	53.8 53.6
2	51.5 52.0	51.8 51.6	54.4 55.6	53.6 53.9
3	51.4 51.1	51.3 51.7	52.9 52.9	53.5 53.5

LAB D

	ANALYST		ANALYST	
MIXTURE	7	8	7	8
1	53.0 53.1	53.3 53.7	53.4 53.6	52.3 52.5
2	51.8 52.0	49.7 49.0	55.3 55.5	55.2 54.8
3	52.1 52.1	51.2 50.2	53.4 53.4	52.7 52.8

21. A rabbit assay of insulin was designed to compare the potencies of two preparations, which may conveniently be called the "standard" and the "unknown". Each preparation was administered at two dosage levels. Each experimental animal was given each of the four doses on four different days. The observations on four animals were arranged in a latin square, "rows" corresponding to days, "columns" to animals, and "treatments" to doses. This latin square was repeated 4 times, using different sets of animals, each set with its own random arrangement of doses, but administering the doses on the same days. The treatments were: treatment 1 - 0.30 units of the unknown preparation, treatment 2 - 0.60 units of the unknown, treatment 3 - 0.30 units of the standard, treatment 4 - 0.60 units of the standard. The table below gives the observations in milligrams of glucose per 100 cc. of blood. The number of the treatment is recorded under each observation. Calculate the appropriate analysis of variance for this experiment and make the required tests of significance.

Milligrams of glucose per 100 cc of blood

Rabbits	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Days																
1	59	56	45	62	42	49	83	56	47	90	79	50	45	52	57	64
	3	2	1	4	1	3	2	4	2	1	3	4	4	3	1	2
2	56	58	41	49	39	61	81	54	46	74	63	69	61	31	30	83
	1	4	3	2	2	1	4	3	4	3	2	1	1	2	4	3
3	41	73	30	63	44	38	101	65	62	61	58	66	45	35	57	74
	2	3	4	1	4	2	3	1	1	2	4	3	3	4	2	1
4	54	69	28	84	61	43	96	58	76	63	87	59	71	81	50	67
	4	1	2	3	3	4	1	2	3	4	1	2	2	1	3	4