DESIGN OF EXPERIMENTS

GROUP TRAINING COURSE IN
INDUSTRIAL STANDARDIZATION AND QUALITY CONTROL
1973

Presented jointly
By
Tadakazu Okuno

Chief, Division of Statistical Research,
National Institute of Agricultural Sciences,
Ministry of Agriculture & Forestry,

and

Tosiro Haga

Manager,
Electronic Data Processing Division,
Sanyo & Kokusaku Pulp Co., Ltd.

JAPANESE STANDARDS ASSOCIATION

1. SIMPLE COMPARATIVE EXPERIMENTS

1.1 Description of the Experiment

The experiment was required to test whether or not treatment with a certain chlorinating agent increased the abrasion resistance of a particular type of rubber. The experimenter took five test-pieces of the material and divided each piece into two. One half was treated and the other half was left untreated, the choice which half of the specimen should receive the treatment being made by tossing a coin. The abrasion resistance of the five pairs of specimens were then tested by a machine, the specimens being taken in random order. The five differences, abrasion resistance of treated specimen minus abrasion resistance of untreated specimen, are shown in Table 1.1.

Table 1.1

Test-piece	Treated	Untreated	Difference	Test-piece mean
1	14.7	12.1	2.6	13.40
2	14.0	10.9	3.1	12.45
3	16.2	14.5	1.7	15.35
4	10.2	9.3	0.9	9.75
5	12.4	11.2	1.2	11.80
Mean	13.5	11.6	1.9	

The mean difference ($\bar{x} = 1.9$) was positive, suggesting that the treatment had been favourable, but much variation occurred in the individual results. In order to decide how much confidence could be placed in the apparent improvement a test of significance was made. The appropriate test is the t-test, which will be shown in 1.3.

1.2 Test of Significance

In this place, we notice only the signs of the differences in Table 1.1 and find that all the five differences are positive. Based on this information much, how could we draw inferences? The statistical test follows as shown below:

(1) The hypothesis it is desired to test is decided upon. The observer wants to find out whether there is probably a real difference, and to do this he considers whether, in the light of the experimental results, the hypothesis that there is really no difference is tenable. In statistical terminology this hypothesis of no real difference is called the Null Hypothesis H_0. Under this Null Hypothesis it is recognized that the probability p of getting the positive sign of the individual difference is,

$$H_0: \quad p = 1/2 \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1.1)$$

(2) The information of the separate measurements is replaced by a single criterion, and is calculated the probability of drawing a more discrepant sample from H_0 than the actually obtained data, provided that the Null Hypothesis is true. In this particular instance it is seen that the chance of obtaining these five positive signs is,

$$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = 1/32 \doteq 0.03 \ (3\%) \qquad \ldots\ldots\ldots \quad (1.2)$$

(3) A decision is made whether the departure from H_0 indicated by the criterion is such that this hypothesis must be abandoned. Working intuitively, the observer would have to decide this from common sense and past experience. Except in extreme instances he would find it very difficult to do this.

In the precise statistical test this phase corresponds to the reference of the above-obtained probability.

(i) If the probability thus obtained is negligibly small, say less than 5% or 1% the observer would likely reject H_0, and the alternative Hypothesis that a real increase has occured will be accepted. In doing so, he faces the possibility that he is wrong. Such is the kind of risk always run by those who test hypotheses and rest decisions on the tests.

(ii) If the probability thus obtained does not show an unusually small value, the observer would not likely reject H_0. Of course, he may be in error, provided that H_0 is not true. But the discrepancy, if any, is so small that the data have given no convincing evidence of rejecting H_0.

In this example, the experimenter could reject the Null Hypothesis and state that the treatment had increased the abrasion resistance, with the risk of about 3% that this statement might be wrong, because this sample might be one of those 3 per 100 samples which were expected to come to hand even when H_0 is true. This probability is usually called "level of significance" because it signifies rejection. Of course, we might not reject H_0 and put off decision till obtaining the further data.

1.3 Test of Difference in a Paired Experiment (t -test)

We are now taking into consideration the differences themselves given in Table 1.1 and performing the t -test, which involves a comparison of the mean difference with its standard error. The procedures are shown in Table 1.2 and below.

Table 1.2

Difference x	Code $u=(X-2.0)\times 10$	u^2
2.6	6	36
3.1	11	121
1.7	-3	9
0.9	-11	121
1.2	-8	64
	17 -22	351
		- 5 = CT
	$T_u = -5$	$S = 346$
$\bar{x} = 2.0+(-1)/10$	$\bar{U} = -1$	$s_u^2 = 346/4 = 86.5$
$= 1.9$	$CT = (-5)^2/5 = 5$	s^2 86.5 $/10^2 = 0.865$

(1) The Null Hypothesis H_0 is $\mu = 0$, where μ denotes the population mean difference. It is assumed that the observations may be regarded as representing independent drawings from a normal distributed universe having mean μ.

$$H_0: \mu = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1.3)$$

(2) The data reported in Table 1.2 are summarized by the following two values:

Mean difference: $\bar{x} = 1.9$

Variance: $s^2 = 0.865$

From these the standard error of the mean $\bar{x}$ is obtained as,

$$\text{Standard error of the mean} = \sqrt{s^2/n} = \sqrt{0.865/5} = 0.416$$

Then we calculate the quantity t, known as 'student's' t, the distribution of which was discovered by W.S. Gosset in 1908 and perfected by R.A. Fisher in 1924 as,

$$t = \frac{\text{deviation of the estimated mean } \bar{x} \text{ from } \mu}{\text{standard error of mean } \bar{x}}$$

$$= (\bar{x} - \mu)/\sqrt{s^2/n}$$

when we can replace 0 for μ under H_0.

$$= (1.9 - 0)/0.416 = 4.57 \quad \ldots\ldots\ldots\ldots\ldots \quad (1.4)$$

(3) Is this an ordinary value of t in sampling from a normal population with $\mu = 0$, or is it so unusual as to cause rejection of the Null Hypothesis?

The distribution of t is laid down in t-table. In the case of large sample, it is practically normal with $\mu = 0$ and the variance $\sigma^2 = 1$. Like the normal distribution, the t-distribution is symmetrical about the mean, which allows the probability in the table to be stated as that of a larger absolute value, sign ignored. Entering the t-table with the degrees of freedom equal to four (or in general one less than the number of differences), it is found that the appropriate 1% significance point is 4.604.

The value of $t = 4.57$ is therefore significant at the approximately 1% level and the experimenter could conclude that the treatment probably yielded a real increase in abrasion resistance.

1.4 Interval Estimation

In the above example the hypothesis that the true mean difference in abrasion resistance between treated and untreated specimens was zero was rejected by the t-test. The argument was that if μ denoted the true mean difference, then in repeated experiments the statistic,

$$(\bar{x} - \mu)/(s/\sqrt{n}) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1.5)$$

would be distributed to a sufficiently close approximation as the tabled quantity t. When μ was put equal to zero, the value required by the Null Hypothesis, the ratio equalled 4.57, a value which would be expected to be equalled or exceeded nearly once in a hundred times: the hypothesis that $\mu = 0$ was therefore rejected. This same form of test would be equally appropriate to test the plausibility of any other postulated value for μ. Suppose it was decided to make the test at the 5% level, then since from the t-table, t has a 5% chance of lying outside the limits ± 2.776, those hypothetical values of μ which made $(1.9 - \mu)/0.416$ fall outside the limits ± 2.776 would be rejected, that is to say, those values of μ falling outside the limits 1.90 ± 1.15. It can then be said that at the 5% level of significance the observations are consistent with any value for the true mean difference μ lying between 0.75 and 3.05, but that values for μ outside these limits are contradicted by the data. This

argument is due to R.A. FISHER, who called these limits the Fiducial Limits. In general, the fiducial limits for μ would be,

$$\bar{x} \pm t(f;\alpha)\ s/\sqrt{n} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1.6)$$

where $t(f;\alpha)$ is the value exceeded with some small probability the degrees of freedom being f. A different justification was used by J. NEYMAN to arrive at these limits, which he called Confidence Limits.

NEYMAN showed that if these limits were adopted, that is to say, if on completion of an experiment of this kind it was always said that lay within the limits $\bar{x} \pm t(f;\alpha)\ s/\sqrt{n}$, then in the long run this μ statement would be right in a proportion $1-\alpha$ of the time. $1-\alpha$ is called the Confidence Coefficient, and the limits are called the $100(1-\alpha)\%$ Confidence Limits. For example, if α is taken equal to 0.05, they are the 95% Confidence Limits.

Note:

(i) If we wish to increase the Confidence Coefficient, we would have a wider interval.

(ii) If, on the contrary, we were contented with the low Confidence Coefficient, we could obtain a narrow interval.

(iii) In order to get a reasonably narrow interval with the considerably high confidence, we would have to either increase the precision of experiments themselves or increase the number of replications.

1.5 The Reason for Taking the Specimens in Pairs

In order that comparison should be made between specimens as like as possible, each test-piece was cut into halves, one half being treated and the other half not. The results analysed were the differences in abrasion resistance between the treated and untreated specimens cut from the same test-piece. The comparisons were thus kept entirely within a test-piece and the large variation between test-pieces was eliminated both in the design and in the subsequent analysis.

The essential plan is to limit comparisons to within aggregates of material which are made homogeneous than the whole; these aggregates are called Blocks. In the example given the blocks are the test-pieces from each of which a treated and an untreated specimen is obtained. The error appropriate for making the test is then that due to the variation between a pair of specimens from the same test-piece if no treatment were applied, the additional variation from one test-piece to another being eliminated. This will be seen more clearly by referring to the individual results given at the right-hand column of Table 1.1.

1.6 Comparison of Two Randomized Groups

Sometimes there may not be enough knowledge of behavior in the experimental material to warrant pairing or blocking. The alternative is merely to assign the individual materials (specimens, test-pieces etc.) at random to the two groups, and then apply one of the treatments to each group. It will be only the two means that are compared, not the individual measurements. Experimental error will be determined by the average variation among the individuals within each of the groups.

In order to illustrate the analysis of this type of comparative experiments, suppose that the data in Table 1.1 were obtained from the experiment, which utilized ten test-pieces, one specimen being taken from each test-piece.

Two groups of five test-pieces each are assumed to be drawn at random from populations in which different treatments may have differentiated the means but not the variances. Denote the population means of the treated and the untreated group by μ_1 and μ_0, respectively. For evidence we have the group means $\bar{x}_1 = 13.5$ and $\bar{x}_0 = 11.6$.

Then, the familiar question is this: Is the difference, $\bar{x}_1 - \bar{x}_0 = 1.9$, attributable to a population difference $\mu_1 - \mu_0$, or may it be random variation from a single population mean, (that is to say, $\mu_1 = \mu_0$)?

To answer this question the following steps are taken:

(1) We set up the null hypothesis,

$$H_0: \ \mu_1 = \mu_0 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1.7)$$

The t-distribution furnishes a test:

$$t = \left[(\bar{x}_1 - \bar{x}_0) - (\mu_1 - \mu_0)\right] / s(\bar{x}_1 - \bar{x}_0) \quad \ldots\ldots \quad (1.8)$$

which becomes, with H_o

$$t = (\bar{x}_1 - \bar{x}_0) / s(\bar{x}_1 - \bar{x}_0) \quad \ldots\ldots\ldots\ldots\ldots \quad (1.9)$$

where $s(\bar{x}_1 - \bar{x}_0)$ denotes the standard error of the difference of the two means $x_1 - x_0$. It remains only to calculate the denominator.

(2) Each group provides an estimate of σ^2, the variance common to the two experimental populations. The estimates are obtained as follows:

Treated group		Untreated group	
$u_i = (x_i - 10.0) \times 10$	u_i^2	$u_i = (x_i - 10.0) \times 10$	u_i^2
47	2,209	21	441
40	1,600	9	81
62	3,844	45	2,025
2	4	-7	49
24	576	12	144
175	8,233	80	2,740
$\bar{u}_1 = 35$	-6,125	$\bar{u}_2 = 16$	-1,280
	$S_1 = 2,108$		$S_2 = 1,460$
	$s_1^2 = 527$		$s_2^2 = 365$

The average of these two estimates is used as the best estimates of σ^2.

$$s^2 = \frac{1}{2}(s_1^2 + s_2^2) = \frac{1}{2}(5.27 + 3.65) = 4.46 \quad \ldots\ldots \quad (1.10)$$

Now, as we know, the variance of each of the two means is,

$$s_{\bar{x}}^2 = s^2 / n = 4.46 / 5 = 0.892 \quad \ldots\ldots\ldots\ldots\ldots \quad (1.11)$$

By referring to the mathematical theory or by drawing random samples experimentally we will be able to know that the variance of random differences is double that of the original population. We have, therefore,

$$s_{(\bar{x}_1 - \bar{x}_0)}^2 = s_{\bar{x}_1}^2 + s_{\bar{x}_0}^2 = 2s^2/n = 2 \times 0.892 = 1.784 \quad \ldots \quad (1.12)$$

Substituting the difference between the group means $\bar{x}_1 - \bar{x}_0 = 1.9$, along with its standard error, $s(\bar{x}_1 - \bar{x}_2) = \sqrt{1.784} = 1.34$, we have from (1.9),

$$t = (\bar{x}_1 - \bar{x}_0) / s_{(\bar{x}_1 - \bar{x}_0)} = 1.9/1.34 = 1.42$$

There are $n - 1 = 4$ degress of freedom associated both with s_1^2, and with s_2^2, making $2(n - 1) = 8$ in total.

Comparison of $t = 1.42$ with the figures corresponding to d.f. = 8 in t - table shows that the probability for this t -value to be equalled or exceeded is approximately equal to 0.20.

(3) Since the probability 0.20 is of modest size, the null hypothesis can not be rejected.

Note:

(i) The general formula for obtaining the standard error of the difference of two means, one from n_1 observations and another n_2 observations, is

$$s_{(\bar{x}_1 - \bar{x}_2)}^2 = s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2 = (1/n_1 + 1/n_2)\, s^2$$

$$= (n_1 + n_2)/(n_1 n_2 s^2)$$

(ii) In section 1.3 and 1.2 the same null hypothesis was rejected with the risk of about 1% and 3%, respectively, whereas, here we cannot reject it. This comes from having ignored the pair in the testing procedure of the latter case.

(iii) With increasingly precise experimentation in any field, resulting from more exact knowledge of the behavior of the experimental material, group comparisons are likely to be replaced by those of individuals (paired comparison).

(iv) If the investigator is interested more in estimates than in tests, he may choose to use the Confidence Limits rather than t -test.

He may report that $\bar{x}_1 - \bar{x}_0 = 1.9$ with the 95% Fiducial Limits:

$$x_1 - x_0 \pm t\,(0.05;\ 8)\, s_{(\bar{x}_1 - \bar{x}_0)}$$

$$= 1.9 \pm 2.306 \times 1.34$$

$$= 1.9 \pm 3.1$$

$$= (-1.2 \sim +5.0)$$

2. TEST OF EQUALITY OF MORE THAN TWO POPULATION MEANS

This is a natural extension of comparison of two randomized groups described in section 1.6. These populations are thought of as having means characterized by the treatments but with a common variance unaffected by treatment.

Theoretically, they are normal populations all having the same variance, σ^2, but each with its peculiar mean μ. As examples, there may be several lots of animals, every lot receiving a different ration; or several classes of children in the 6th grade on which different methods of instruction are being tried. Table 2.1 contains the data from such an experiment.

Table 2.1. Percent Loss of Product in Manufacture of an Organic Chemical

Batch No.	A_1	A_2	A_3	A_4	
1	22.4	15.4	6.0	12.4	
2	25.6	20.4	14.9	10.6	
3	22.5	20.0	15.6	8.9	Grand
4	19.7	15.3	14.4	7.5	total
Total	90.2	71.1	50.9	39.4	251.6
Mean	22.6	17.8	12.7	9.9	15.725

Table 2.2. Symbolical Representation of an Experiment

Level of Treatment	A_1	A_2 ----- A_3 ----- A_4	
Individual observations Rows have no physical meaning because the X_{ij} can be rearranged within each of the columns ($n = \Sigma_{ni}$)	X_{11} X_{12} ⋮ X_{1n_1}	A_{21} ---- X_{i1} ---- X_{a1} X_{22} ---- X_{i2} ---- X_{a2} ⋮ ⋮ ⋮ X_{2n_2} X_{ij} ⋮ X_{in_i} X_{an_a}	Grand Total
Total	T_1	T_2 ----- T_i ----- T_a	G
Mean	$\bar{X}_1$	$\bar{X}_2$ ----- $\bar{X}_i$ ----- $\bar{X}_a$	

The experimenter wished to learn how the percent loss was influenced by the four different blends. The usual method of calculation is shown below:

(1) The correction term:

$$CT = G^2/n$$
$$= (251.6)2/16 - 3,956.41$$

(2) The total sum of squares:

$$S_T = \sum_{i=1}^{a} \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2 = \sum_{i=1}^{a} \sum_{j=1}^{n_i} X_{ij}^2 - CT$$

$$= (22.4^2 + 25.6^2 + \cdots\cdots + 7.5^2) - CT$$

$$= 4,449.18 - 3,956.41 = 492.77$$

(3) The sum of squares for blends (between groups):

$$S_A = \sum_{i=1}^{a} n_i(\bar{X}_i - \bar{X})^2 = \sum_{i=1}^{a} T_i^2/n_i - CT$$

$$= (90.2^2 + 71.1^2 + 50.9^2 + 39.4^2)/4 - CT$$

$$= 4,333.605 - 3,956.41 = 377.195$$

(4) The sum of squares for batches (within groups):

$$S_e = \sum_{i=1}^{a} \sum_{j=1}^{n_i} (\bar{X}_{ij} - \bar{X}_i)^2 = \sum_{i=1}^{a} \sum_{j=1}^{n_i} X_{ij}^2 - \sum_{i=1}^{a} T_i^2/n_i$$

$$= S_T - S_A = 492.77 - 377.195 = 115.575$$

(5) The results from the last 3 steps, with corresponding degrees of freedom, are entered in Table 2.3.

Table 2.3. Analysis of Variance of Percent Loss

Source of variation	Degrees of freedom	Sum of squares	Mean square	F	Expected value of mean square
Total T	16-1=15	S_T=492.77			
Treatment A (between groups)	4-1= 3	S_A=377.195	125.732	12**	2_{+4} A^2
Error e (within groups)	4(4-1)=12	S_e=115.575	9.631		2

Comments:

(i) The degrees of freedom and sum of squares in the last line are got by subtraction, taking advantage of the addition theorem characterizing this analysis.

(ii) The mean square or variance is obtained from dividing the sum of squares by the corresponding degrees of freedom.

This partition of degrees of freedom and corresponding sums of squares is called analysis of variance. Under the assumptions outlined above, sampling from normal populations with common σ^2, 9,631 is an estimate of this σ^2. But the mean square for treatment, 125.732 seems to have an additional component due to the different behaviors of the blends in the chlorsulphonation.

As to the constancy of the variance, the ranges in the samples are evidence. For A_1, the range is 25.6 - 19.7 = 5.9; for the others, 5.1, 9.6, 4.9.

(6) Now, these data introduce a familiar question: is ordinary random sampling accountable for the large discrepancy between the mean square for treatment and error, or shall we conclude that the treatment means are differentiated by causes other than sampling fluctuations?

The appropriate null hypothesis to be tested is H_0: $\mu_1 = \mu_2 \ldots\ldots = \mu_a$, which specifies the population sampled by the levels of treatment. In order to test H_0, a new test criterion is calculated, the ratio

$$F = \frac{\text{Mean square for treatment means}}{\text{Mean square for error}}$$

To find the 5% and 1% points in F-table, look in the column headed by $f_1 = 3$ and down to the rows $f_2 = 12$. The required points are these: F (0.05; 3, 12) = 3.49 and F (0.01; 3, 12) = 5.95. On the other hand, from the analysis of variance in table 2.3, we get

$$F = 125.932/9.631 > 12$$

Thus, from the distribution specified in the hypothesis, there is less than one chance in 100 of drawing a sample having a larger value of this F. Evidently the samples come from populations with different μ. The conclusion is that the four blends have different effects on the percent loss in the chlorsulphonation.

(7) Least significant difference

The estimated standard error of the difference between any two treatment means is

$$s_d = \sqrt{\frac{2s^2}{n}} = \sqrt{\frac{2 \times 9.631}{4}} = 2.19$$

Hence the least significant difference is obtained as follows:

$$\text{l.s.d.} = t(0.05;\ 12)\ s_d$$

$$= 2.179 \times 2.19 = 4.77,$$

from which we can state that the blends A_4 and A_3 have the smaller loss than those of A_1 and A_2 and that the loss of A_2 is smaller than that of A_1.

3. COMPLETE BLOCK DESIGN

3.1 Basic Principles for Design of Experiments

When several experimental treatments are to be compared it is clearly desirable that all other conditions shall be kept as nearly constant as is practicable. Random variations will occur and appear as experimental error, and some "Replication" under similar conditions will be required to compare the treatments with sufficient reliability; such replication also supplies the information to estimate the experimental error, and this is required to assess the reliability.

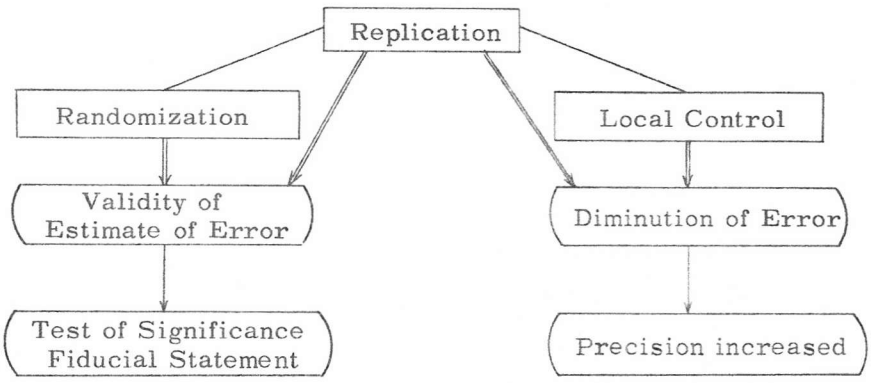
In practice, particularly when the number of trials required is large, it is often difficult to ensure such similar conditions owing to the natural variability of the materials and processes involved. (Even if the variability could be eliminated it is doubtful whether it would be wise to try, since such variability broadens the basis of comparison and so renders the results more generally applicable.) It is often possible, however, to split up a set of trials into smaller groups within which such variations are likely to be less than in the set as a whole. Thus pieces cut from the same sheet of rubber are expected to be more alike in their properties than pieces cut from different sheets, and samples taken from a plant over a short period vary less than those taken over a longer period. Where these conditions hold, the precision of experiment can be increased by dividing it into "Blocks," within each of which the random variations are likely to be smaller than in the experiment as a whole. In Randomized Block Design each block contains all the treatments concerned (this --→ being called Complete Block), while the use of several blocks ensures that the number of observations is sufficient to give the required precision to the experiment as a whole. Where the block is too small to accommodate all the treatments, resource may be had to the more complex Incomplete Block Designs discussed in Chapter 5. This device 'Blocking' is called "Local Control" and may be considered as an extension of paired comparison stated in Section 1.3. Sometimes this blocking can be effected in more than one way; for instance, in a multiple plant the various units may differ in performance, and in addition there may exist a trend in time, as in certain electrolytic and catalytic processes. Under such circumstances the most efficient design for the comparison of different experimental treatments is known as Latin Square.

Although the variability within any block is likely to less than that in the experiment as a whole, there may be a systematic variation within the individual blocks. Thus pieces from the centre of a sheet of rubber are likely to differ systematically from those cut from the edge of the sheet. If therefore the treatments are introduced in the same relative positions in successive blocks, spurious effects due to the systematic variations associated with position within a block are likely to be introduced into the results. To overcome this the arrangement of the treatments must be different in each block, the actual positions of the treatments within any block being chosen by an adequate "Random" process.

R.A. FISHER gave the following diagram to show the above three basic principles for design of experiments.

Fig. 3.1

Basic Principles for Design of Experiments



3.2 Randomized (Complete) Block Design

(1) Introduction

Suppose it is required to compare the effects of five treatments, say five lots of material prepared in different ways, or five temperatures of reaction, and in order to reduce the uncertainties caused by experimental error it is decided to test each treatment four times, making twenty trials in all, then the ideal design will provide for all twenty trials to be carried out under uniform conditions, apart from deliberate variation of the treatments. In practice it may be impossible to do this because, for example, sufficient raw material of uniform quality for twenty trial cannot be made. But it may be that a homogeneous batch of raw material sufficient for five trials can be made, and if so the experiment may be arranged so that all five treatments are tested on each of four batches which are homogeneous but not necessarily identical, with the consequence that any variation from batch to batch does not affect comparisons between the treatments. A typical example arises in the testing of rubber or other material in sheet form. Suppose five methods of treating the rubber are to be tested and large sheets are available. Adjacent samples cut from a rubber sheet are usually more alike than non-adjacent samples, and this property suggests that comparisons between treatments should be made between adjacent samples of rubber. To compare five treatments, replicating the experiment four times, four pieces should be cut from different parts of the sheet and each piece cut into five, for in this way the variation from one set of five to another does not affect comparisons between the five treatments, which are made entirely within the sets. Had twenty pieces of rubber been cut from the sheet and the five treatments applied at random the experiment would have been less sensitive, because the heterogeneity of the material would have inflated the experimental error.

In this context the set of five pieces taken from the same part of the sheet is called a 'Block'. As a precaution against systematic variation from one trial to another within a block, it is desirable to arrange the treatments within each block in random order, and when this has been done the results is a Randomized Block Design. The terminology originated with agricultural field trials. In order to minimize the effects of differential fertility the experimental area is divided into compact blocks, each supposed more homogeneous than the whole, and the blocks are subdivided into plots, one treatment being assigned to each plot. If the plots were arranged systematically within blocks the experiment would be vitiated by any fertility gradient occurring in the same direction; consequently the treatments must be allotted to the plots in random order.

(2) Layout --- Manufacture of an organic chemical

This investigation concerned the chlorsulphonation of acetanilide. The yield is somewhat below the theoretical, mainly because of loss of product in the mother liquor (i.e. the liquor removed when the product is filtered). It was required to test whether different blends of acetanilide gave different losses, and to do this five different blends were made and four batches of products were prepared from each blend. If the blends had been tested one at a time, i.e. four batches from blend 1, then four batches from blend 2, and so on, any variation from blend to blend might have been due to a time trend in the process and not to any real differences between the blends. To eliminate this effect the experiment was designed in randomized blocks. There were twenty batches in all, divided into four blocks of five, the five in one block being prepared from the five blends of acetanilide in random order. The following table gives the order of preparation and the blend used.

Block	I					II					III					IV				
Batch	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Blend	B	A	C	E	D	A	E	B	C	D	C	A	B	D	E	B	D	C	E	A

(3) Data obtained

Table 3.1 Percentage Loss (X_{ij})

Block	Blend of acetanilide A	B	C	D	E	Total $T_{.j}$	Mean $\bar{x}_{.j}$	Effect $\bar{x}_{.j}-\bar{x}$
I	18.3	17.1	17.3	15.1	16.7	84.5	16.9	-0.6
II	18.8	18.3	18.1	15.9	16.9	88.0	17.6	0.1
III	19.8	19.2	17.2	17.8	16.5	90.5	18.1	0.6
IV	18.3	18.2	17.0	16.0	17.5	87.0	17.4	-0.1
Total $T_{i.}$	75.2	72.8	69.6	64.8	67.6	350.0 = G		
Mean $\bar{x}_{i.}$	18.8	18.2	17.4	16.2	16.9		17.5 = $\bar{x}$	
Effect $\bar{x}_{i.} - \bar{x}$	+1.3	+0.7	-0.1	-1.3	-0.6			0.0

(4) Expected values and residual

Table 3.2 Expected Value

$$\hat{X}_{ij} = \bar{X} + (\bar{X}_{i.} - \bar{X}) + (\bar{X}_{.j} - \bar{X})$$

Block \ Blend	A	B	C	D	E	Total
I	18.2	17.6	16.8	15.6	16.3	84.5
II	18.9	18.3	17.5	16.3	17.0	88.0
III	19.4	18.8	18.0	16.8	17.5	90.5
IV	18.7	18.1	17.3	16.1	16.8	87.0
Total	75.2	72.8	69.6	64.8	67.6	350.0

Table 3.3 Residual

$$X_{ij} - \hat{X}_{ij} = X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}$$

Block \ Blend	A	B	C	D	E	Total
I	0.1	-0.5	0.5	-0.5	0.4	0.0
II	-0.1	0.0	0.6	-0.4	-0.1	0.0
III	0.4	0.4	-0.8	1.0	-1.0	0.0
IV	-0.4	0.1	-0.3	-0.1	0.7	0.0
Total	0.0	0.0	0.0	0.0	0.0	0.0

(5) Computation of sums of squares

1) $X_{ij} - \bar{X} = (\bar{X}_{i.} - \bar{X}) + (\bar{X}_{.j} - \bar{X}) + (X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X})$

(Deviation from grand mean) (treatment effect) (block effect) (residual)

2) $\sum_i \sum_j (Xij-\bar{X})^2 = r \sum_i (\bar{X}i.-\bar{X})^2 + a \sum_j (\bar{X}.j-\bar{X})^2 + \sum\sum (Xij-\bar{X}i.-\bar{X}.j+\bar{X})^2$

S_T S_A S_R S_e

3) $(ar - 1) = (a - 1) + (r - 1) + (a - 1)(r - 1)$

f_T f_A f_R f_e

Total sum of squares:

$$S_T = \sum_i \sum_j (Xij - X)^2$$

$$= (18.3-17.5)^2 + (18.8-17.5)^2 + \cdots + (16.5-17.5)^2 + (17.5-17.5)^2$$

$$= 25.84$$

Sum of squares between treatments (blends):

$$S_A = r \sum_i (\bar{X}i - \bar{X})^2$$

$$= 4 \left\{ 1.3^2 + 0.7^2 + (-0.1)^2 + (-1.3)^2 + (-0.6)^2 \right\} = 16.96$$

Sum of squares between blocks:

$$S_R = a \sum_j (\bar{X}.j - \bar{X})^2$$

$$= 5 \left\{ (-0.6)^2 + 0.1^2 + 0.6^2 + (-0.1)^2 \right\} = 3.70$$

Sum of squares due to error (from the residuals in Table 3.3)

$$S_e = \sum_i \sum_j (Xij - \bar{X}i. - \bar{X}.j + \bar{X})^2$$

$$= 0.1^2 + (-0.1)^2 + \cdots + (-1.0)^2 + 0.7^2 - 5.18$$

(6) Computational Procedure

Table 3.4 Code Number

$Uij = (Xij - 17.0) \times 10$

Block \ Blend	A	B	C	D	E	$T_{.j}$	$T_{.j}^2$
I	13	1	3	-19	-3	-5	25
II	18	13	11	-11	-1	30	900
III	28	22	2	8	-5	55	3025
IV	13	12	0	-10	5	20	400
$T_{i.}$	72	48	16	-32	-4	100	$4350/5 = 870 = (I)_R$
$T_{i.}^2$	5184	2304	256	1024	16	$8784/4 = 2196 = (I)_A$	
						$100^2/20 = 500 = CT$	

Table 3.5 U_{ij}^2

Block \ Blend	A	B	C	D	E	
I	169	1	9	361	9	
II	324	169	121	121	1	
III	784	484	4	64	25	
IV	169	144	0	100	25	
Total	1446	798	134	646	60	3084 = (II)

$S_T = (II) - CT = 3084 - 500 = 2584$

$S_A = (I)_A - CT = 2196 - 500 = 1696$

$S_R = (I)_R - CT = 870 - 500 = 370$

$S_e = S_T - S_A - S_R = 518$

(7) Analysis of Variance Table

Table 3.6 Analysis of Variance (decoded) (% loss)

Source of Variation	Degrees of freedom	Sum of squares	Mean square (Variance)	Variance ratio F	Expected value of Mean square
Total T	$f_T = 19$	$S_T = 25.84$			
Treatment A (Blends)	$f_A = 4$	$S_A = 16.96$	$V_A = 4.24$	9.81**	$\sigma 2 + 4k_A^2$
Block R	$f_R = 3$	$S_R = 3.70$	$V_R = 1.23$	2.85	$\sigma 2 + 5k_R^2$
Error e	$f_e = 12$	$S_e = 5.18$	$V_e = 0.432$		$\sigma 2$

The variation among blends is significant at the 1% level:

$$F = 9.81 \quad > \quad F(4, 12; 0.01) = 5.41$$

The test of significance is appropriate in this example because we were interested in testing the existence of blend-to-blend variation which, on chemical grounds, did not seem likely. Since the mean square between blends is highly significant at the 1% level, we conclude that there is clear evidence of blend-to-blend variation. If the existence of this variation was not in question but only its magnitude, the problem would be one of estimation and the result would require to be interpreted in terms either of errors of the first and second kinds or of confidence limits.

Although the apparent variation among blocks is not confirmed (i.e. it might well be ascribed to experimental error), future experiments should still be carried out in the same way. There is no clear evidence of a trend in this set of trials, but it might well appear in another set, and no complication in experimental arrangement is involved.

(8) Least significant difference

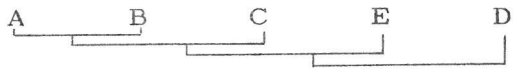
The estimated standard error of the difference between any two blend means is

$$s_d = \sqrt{2V_e / r} = \sqrt{2 \times 0.432 / 4} = 0.465$$

from which, we can calculate the least significant difference

$$l.s.d. = t(12; 0.05)\, S_d = 1.01.$$

Referring to the treatment means given at the bottom of Table 3.1, we can conclude that:



(9) Missing plot technique

It has been assumed hitherto that every treatment is equally represented in every block, so that the effects of block and treatment variation can be assessed by simple averaging. It may happen that in one of the blocks one treatment is not tested, or more generally the results of one or more treatments in one or more blocks are not available. We cannot then use the simple analysis given above.

Suppose that for some reason the result of one trial, say batch 14 (block III, blend D) is not available. The procedure is to calculate from the remaining trials the most probable value for batch 14, using the estimates of the effects of block III and of blend D.

To do this we insert symbol y in place of the missing values, carry out the analysis of variance including this symbol, and then calculate value for it such that the residual or error variance is a minimum. This is the best estimate of the missing value in the sense that it minimises the error variance.

The estimate of the missing value is obtained by the following formula:

$$y = \frac{a\,A + r\,R - G}{(a-1)\,(r-1)}$$

where,

A = total of known observations of the treatment containing y

R = total of known observations of the block containing y

G = total of all known results

omitting y from these total.

From Table 3.4 it is found that

$$A = -40, \qquad R = 47, \qquad G = 92$$

Therefore substituting these values will yield the estimate:

$$y = \frac{5 \times (-40) + 4 \times 47 - 92}{(5-1)\,(4-1)} = \frac{-104}{12} = -9$$

An appropriate analysis of variance can now be carried out in the usual way, using the estimate for batch 14, with the difference that the error variance has only 11 degrees of freedom instead of 12, because one of the results is estimated. The resulting analysis is shown in Table 3.7.

Table 3.7 Analysis of variance with one value missing

Source of Variation	Degrees of freedom	Sum of squares	Mean square (variance)	F
Total T	$f_T = 18$	$S_T = 27.57$		
Treatment A (Blends)	$f_A = 4$	$S_A = 21.96$	$V_A = 5.49$	17.2**
Block R	$f_R = 3$	$S_R = 2.10$	$V_R = 0.70$	2.2
Error e	$f_e = 11$	$S_e = 3.51$	$V_e = 0.319$	

3.3 Latin Square

(1) Introduction

In an experiment which is not divided into blocks the replicated treatments are distributed at random over the whole of the experimental material, the heterogeneity of the material together with the testing errors giving the residual (or error) variation. In a randomized block experiment the heterogeneity is controlled by applying the treatments over compact blocks of relatively homogeneous material, thus reducing the residual variation and making the comparisons more sensitive. Under suitable conditions the experiment may be subdivided in more than one way, giving in each case greater homogeneity.

Some examples are shown below.

(a) An experiment on the weaving of cotton cloth. The purpose of this experiment is to investigate the effect of sizing treatment applied to the warp. The criterion is the number of breaks in the warp during weaving. Four warps, each with a different sizing treatment, are assumed to be woven simultaneously on 4 different looms, which can be supervised by a single weaver. Then each warp is moved to a different loom of the set so that after 4 periods every sizing treatment have been tested on all 4 looms. If A, B, C, D represent the 4 warps, the Latin square used is as follows:

Periods	Looms			
	1	2	3	4
I	A	D	B	C
II	D	C	A	B
III	C	B	D	A
IV	B	A	C	D

This arrangement can eliminate constant differences among the looms, and also differences among the 4 periods of weaving.

(b) In the preparation of an explosive mixture used in primers, variation may occur either in the mixing of the ingredients of the explosive or in the process of charging. Suppose that one experiment of this type includes 4 mixing-blending teams and 4 charging operators. On each day, the product of each team is sent to a different charging operator, the arrangement being changed daily according to the following Latin square.

	Charging operators			
	1	2	3	4
Monday	A	D	B	C
Tuesday	B	A	C	D
Wednesday	C	B	D	A
Thursday	D	C	A	B

A, B, C and D represent the mixing-blending teams.

The Latin square analysis of variance enables us to isolate consistent differences amongst the teams and consistent differences amongst the charges, as well as day-to-day variation.

A rough rule for randomization is this: Having written down any systematic arrangement of the letters, rearrange at random the rows and colums; then assign the treatments at random to the letters.

For refinements, see the Fisher and Yates's Table.

(2) Layout ----- wear testing of textile fabrics

In a wear-testing machine with four positions the results obtained in the four positions may vary apart from testing error, and comparisons between different materials will be more precise if all are tested in the same position. This means, however, that they can not be tested in the same run of the machine. There may also be variation from run to run, and comparisons between different materials will be more precise if all are tested in the same run, in which case they cannot all be tested in the same position. If the variation between runs or between positions did not exist or could be ignored a randomized block design could be used, but in order to eliminate both sources of variation from the comparison of treatments a more elaborate design would be required. It is impossible to make every comparison, say treatment 1 with treatment 2, in both the same run and the same position, but it is possible to ensure that each treatment is tested the same number of times in every position and also in every run, provided the number of runs is equal to the number of positions. The average response for each treatment is then an average over all runs and over all positions, and provided the position effect is independent of the run effect all the treatment averages are equally affected by the run and position variations. The relative values of these averages are thus unaffected and they are estimated as precisely as if the run and position variations did not exist.

The arrangement of the trials to ensure that each treatment is tested once in every run and once in every position makes use of the Latin Square. This is a square containing $\underline{m}$ rows and $\underline{m}$ columns, and consequently $\underline{m}^2$ cells. Each cell contains one of $\underline{m}$ letters, corresponding to $\underline{m}$ treatments, and each letter occurs once and once only in each row and each column.

The following is an example of an experiment relating to the testing of rubber-covered fabric in the MARTINDALE wear tester carried out as a 4 x 4 Latin Square. The machine consists of four rectangular brass plates on each of which is fastened an abrading surface consisting of special-quality emery paper. Four weighted bushes, into which test samples of fabric are fixed, rest on the emery surfaces, and a mechanical device moves the bushes over the surface of the emery, thus abrading the test specimens. The loss in weight after a given number of cycles is used as a criterion of resistance to abrasion. There are slight differences between the four positions on the machine, and it is known from past experience that if a run is repeated under apparently identical conditions and using apparently identical samples, slightly different results are obtained, ascribable partly to variations in the emery paper and partly to variations in temperature, humidity, etc.

If samples of fabric are to be compared it is clearly desirable to eliminate the effect of variation from one position to another, and from one run to another, as far as possible. It could be assumed that the three factors, positions, runs, and materials acted independently, so that if one position gave a higher rate of wear than another it would do so in every run and on any material, and so on.

(3) Data obtained

The experiment described involved four materials tested together in each of four runs on the machine. These numbers were chosen because the machine has four positions, so that a Latin Square design could be used. The entries of Table 3.8 denote the loss in weight (units of mg.) in a run of standard length, and the letters A to D refer to the four materials.

Table 3.8 Results of wear-testing experiment: Latin Square

Run	Position in machine 1	2	3	4	Total	Average
1	B 21.2	A 27.0	D 22.7	C 22.9	93.8	23.5
2	C 23.5	D 23.9	A 26.3	B 22.6	96.3	24.1
3	D 23.4	B 24.0	C 23.0	A 26.7	97.1	24.3
4	A 25.2	C 24.2	B 21.8	D 24.0	95.2	23.8
Total	93.3	99.1	93.8	96.2	382.4	
Average	23.3	24.8	23.5	24.1		23.9
Total	A 105.2	B 89.6	C 93.6	D 94.0	382.4	
Average	26.3	22.4	23.4	23.5		23.9

From a simple inspection of the averages it appears that:

(i) Material B is best (lowest loss)

(ii) Run 3 gives higher losses on the average than the others.

(iii) Position 2 is more severe than the others.

However, a statistical analysis is required to test the significance of these apparent differences.

(4) Analysis of variance

Table 3.9 Code number $u = (x - 23.0) \times 10$

Run	Position in machine C_1	C_2	C_3	C_4	Total	$(\text{Total})^2$
R_1	B -18	A 40	D -3	C -1	18	324
R_2	C 5	D 9	A 33	B -4	43	1849
R_3	D 4	B 10	C 0	A 37	51	2601
R_4	A 22	C 12	B -12	D 10	32	1024
Total	13	71	18	42	144	$5798/4 = 1449.5 = (I)_R$
Total squared	169	5041	324	1764	$7298/4 = 1824.5 = (I)_C$	
Total	A 132	B -24	C 16	D 20	$144 - - - (144)^2/16 = 1296 = CT$	
Total squared	17424	576	256	400	$18656/4 = 4664\ (I)_A$	

sum of square of all entires: $\sum_i \sum_j u^2_{ij.} = 5502 = (II)$

Total sum of squares: $S_T = (II) - CT = 5502-1296 = 4206$

Sum of squares between runs: $S_R = (I)_R - CT = 1449.5-1296 = 153.5$

s.s. between positions: $S_C = (I)_C - CT = 1824.5-1296 = 528.5$

s.s. between materials: $S_A = (I)_A - CT = 4664 - 1296 = 3368$

Error sum of squares: $S_e = S_T - S_R - S_C - S_A = 156.0$

Table 3.10 Analysis of variance

Source of variation	Degrees of freedom	Sum of squares	Mean square	F
Total T	$f_T = m^2 - 1 = 15$	$S_T = 42.060$		
Runs R	$f_R = m - 1 = 3$	$S_R = 1.535$	$V_R = 0.512$	1.97
Positions C	$f_C = m - 1 = 3$	$S_C = 5.285$	$V_C = 1.762$	6.78*
Materials A	$f_A = m - 1 = 3$	$S_A = 33.680$	$V_A = 11.227$	> 40**
Error e	$f_e = (m-1)(m-2) = 6$	$S_e = 1.560$	$V_e = 0.260$	

The value of F corresponding to materials greatly exceeds the 1% level. The design has been successful in eliminating a considerable amount of variation associated with positions which would otherwise have decreased markedly the sensitivity of the comparisons between materials.

(5) Least significant difference

$$l.s.d. = t(6; 0.05)\, S_d$$

$$= 2.447 \times \sqrt{2 \times 0.260/4} = 0.88$$

As for the loss in weight we can conclude that

$$A > (C, D) > B$$

(6) Estimate of missing value

The method of § 3.2 (9) may be used to estimate missing values.

For one missing value y and with the notation

R = total of known values in row (run) containing y

C = total of known values in column (position) containing y

A = total of known values in treatment containing y

G = total of all available values

m = number of treatments, columns or rows

then

$$y = \frac{m(R + C + A) - 2G}{(m-1)(m-2)}$$

4. FACTORIAL EXPERIMENTS AND SPLIT-PLOT DESIGN

4.1 Factorial designs

Frequently in scientific investigations, particularly where an empirical approach has to be adopted, problems arise in which the effects of a number of different factors on some property or process require to be evaluated. Such problems can usually be most economically investigated by arranging the experiments according to an ordered plan in which all the factors are varied in a regular way. Provided the plan has been correctly chosen, it is then possible to determine not only the effect of each individual factor but also the way in which each effect depends on the other factors (i.e. the Interactions). This makes it possible to obtain a more complete picture of what is happening than would be obtained by varying each of the factors one at a time while keeping the others constant. Designs of this sort, moreover, lend themselves well to statistical analysis and can if required provide their own estimates of experimental error. In this field the design and the analysis of the results are closely linked, and unplanned experimental work is liable to confuse the effects sought in such a way that much of the information which would otherwise be available is sacrificed.

4.2 Example --- wear resistance of vulcanized rubber

The data given in Table 4.1 are taken from the results of an investigation into the effects on the physical properties of vulcanized rubber of varying a number of factors, the property recorded being the wear resistance of the samples, and the factors being:

A : five qualities of filler

B : three methods of pretreatment of the rubber

C : four qualities of the raw rubber

Table 4.1 Data of a 5x3x4 factorial experiment
(Wear resistance of vulcanized rubber)

Level of factor A	Level of factor C											
	C_1			C_2			C_3			C_4		
	B_1	B_2	B_3	B_1	B_2	B_3	B_1	B_2	B_3	B_1	B_2	B_3
A_1	404	478	530	381	429	528	316	376	390	423	482	550
A_2	392	418	431	239	251	249	186	207	194	410	416	452
A_3	348	381	460	327	372	482	290	313	358	383	376	496
A_4	296	291	333	165	232	242	158	279	220	301	306	330
A_5	186	198	225	129	157	197	107	163	190	213	209	255

Note that, as always in a complete factorial design, all combinations of the levels of the different factors are tested.

Table 4.2 Two-way table of sums for factors A and B

	B_1	B_2	B_3	Sum	Mean
A_1	1.524	1.765	1.998	5.287	440.6
A_2	1.227	1.292	1.326	3.845	320.4
A_3	1.348	1.444	1.788	4.580	381.7
A_4	920	1.108	1.125	3.153	262.7
A_5	633	718	867	2.218	184.8
Sum	5.652	6.327	7.104	19.083	
Mean	282.6	316.4	355.2		318.1

Table 4.3 Analysis of variance

Source of variation	d.f.	s.s.	m.s.	F
Total	59	766,297		
Main effects A	4	478,463	119,616	374 **
B	2	52,794	26,397	82.5 **
C	3	150,239	50,080	156 **
Interactions AB	8	16,807	2,101	6.57**
AC	12	53,890	4,491	14.0 **
BC	6	6,416	1,069	3.34 *
Remainder=interaction ABC	24	7,688	320	

4.3 Split-Plot Designs

In factorial experiments, it is often desirable to lay out the levels of one factor in relatively large units (or whole plots) designed as a randomized block, Latin square, or other design, on account of the following circumstances:

(1) By necessity: The very nature of the levels of this factor may be such as to exclude the use of small plots or units.

(2) By design: The experimenter may know that the levels of this factor usually differ in yield, and he can forego precise information on this factor.

Since the 'whole plots' are large by necessity or by design, it may be desirable to compare levels of another factor on each (smaller) plot, the several levels being allotted to the split-plots or sub-units of each whole plot at random. Such an arrangement is called 'split-plot design', and is a type of factorial arrangement, in which the main effect of one factor is completely confounded with whole plot differences.

The whole plot treatments as well as the split-plot treatments may themselves represent a factorial arrangement. The terminology 'whole' or 'split-plot' treatment does not necessarily refer to the levels of a single factor.

Numerous examples of factors which require large experimental units are available in all fields of research. A few of these are listed below to illustrate the diversity of material for which split-plot designs are suitable.

(i) In field experiments on varieties or fertilizers on small plots, some cultural practices with large machines may be tried on whole groups of the smaller plots, each group containing all the varieties. Irrigation is one such practice that demands fairly large areas per treatment.

(ii) In greenhouse temperature studies, it may be necessary to keep the entire greenhouse at a constant temperature. Several treatments may be conducted in the greenhouse, but the greenhouse is used as a unit. Heat chambers, storage cellars, freezing units, baking ovens etc., must also be utilized as a single experimental unit.

(iii) The smallest unit in some plant-response studies is a single plant, but the plant may be subdivided into subsamples to study the methods of chemical analysis to determine plant composition.

(iv) In the preparation of metal alloys a smelting or blast furnace requires large amount of material, whereas some treatments, such as types of mould, require relatively small amounts.

(v) In experimental education a movie film may be used by several teachers and on several sets of students; the film is a single experimental unit as far as replication on the film is concerned. Perhaps replication on films could be obtained by filming the material on different types of film under different conditions, by different operators and cameras, etc.

Randomization

(1) The randomization procedure for the units (or whole plots) is determined by the particular design chosen, such as a randomized block, Latin square or others.

(2) Randomization of the sub-unit (or split-plot) treatments is newly done in each whole plot.

A consequence is that the experimental error for sub-plot treatments is different (characteristically smaller) than that for whole-plot treatments.

5. INCOMPLETE BLOCK DESIGN

5.1 Introduction

The principles of designing experiments for comparing a number of treatments when uniform conditions can be maintained within blocks of observations, each accommodating one replicate (or the same number of replicates) of each treatment, were discussed in Chapter 3. Sometimes experimental conditions will not permit blocks large enough to include every treatment, so that all treatments cannot be tested under uniform conditions; nevertheless, by suitably designed experiments, valid and efficient comparisons between the treatments can be made without being disturbed by the differences between the blocks.

5.2 Balanced Incomplete Block Designs (BIB)

(1) Simple examples:

(i) $v = 4,\ k = 3,\ b = 4,\ r = 3,\ \lambda = 2,\ e = 89\%$

Block				
	(1)	1	2	3
	(2)	1	2	4
	(3)	1	3	4
	(4)	2	3	4

(ii) $v = 7,\ k = 3,\ b = 7,\ r = 3,\ \lambda = 1,\ e = 78\%$

Block				
	(1)	1	2	4
	(2)	2	3	5
	(3)	3	4	6
	(4)	4	5	7
	(5)	5	6	1
	(6)	6	7	2
	(7)	7	1	3

(2) Fundamental relations among parameters:

(i) $vr = bk$ where

(ii) $\lambda = \dfrac{r(k-1)}{v-1}$

v: number of treatments
k: size of (incomplete) block, ($\leqq v$)
b: number of incomplete blocks
r: number of repetitions of each treatment
λ: number of times (an integer) a treatment occurs with each of the other treatments within same incomplete blocks.

Note: This relation (ii) demands that λ should be equal for all pairs of treatments and be equal to an integer.

The efficiency factor, e, is defined as the fraction of total information contained in intra-block comparisons when inter-block and intra-block contrasts are of equal accuracy, and is equal to

(iii) $e = \dfrac{1 - 1/k}{1 - 1/v} = \dfrac{v(k-1)}{k(v-1)} = \dfrac{v\,\lambda}{k\,r}$

6. FRACTIONAL FACTORIAL DESIGN BY USING ORTHOGONAL ARRAYS

6.1 Introduction

A complete factorial experiment, in which all possible combinations of all the levels of the different factors are investigated, will involve a large number of tests when the number of factors is large. Thus an investigation of five factors each at two levels will entail $2^5 = 32$ observations, each under a different set of experimental conditions. An experimenter might well consider 32 observations excessive, even after consideration of the advantages of the factorial design given in §4. In any case the experimenter may not require the high degree of accuracy in the estimates of the effects given by the complete factorial design, and he may be satisfied from prior knowledge of the process or similar processes that many of the interactions, particularly the higher-order ones, are not appreciable and moreover he may already have a sufficiently reliable estimate of the experimental error. Even when an experiment of this magnitude may eventually be required he would prefer to carry out the work as a series of smaller experiments.

The object of Fractional Factorial Designs, sometimes called Fractional Replicates, is to obtain information on the main effects and as many of the interactions as seems necessary with a smaller number of observations than is required by the complete design. Careful consideration of the best combinations of experimental treatments is needed, and the theory brings out in detail what becomes of the interactions neglected in any particular design, and what are the results if, unexpectedly, they are not negligible in reality.

There are two approaches to the problem of obtaining a suitable design in any particular case. We may begin with the full factorial design appropriate to the number of factors to be investigated, and by equating (or confounding) interactions which are likely to be small or unimportant with one another and with other interactions considered worth measuring we can arrive at a design in which all (or almost all) of comparisons are made between effects likely to be important. The number of observations needed will clearly depend on the number of such comparisons. If, for instance, an experimenter wants to estimate all the main-effects and all the first-order interactions (that is: two-factor interactions) of ten factors each at two levels, the number of main-effects and of first-order interactions being 10 and 45, respectively, it follows that 1/16 fractional replicates of 2^{10} factorial design could be suitable, which requires only 64 tests in all.

Alternatively, we may begin with the full factorial design corresponding to a number of factors smaller than is actually under investigation and substitute the remaining factors for those comparisons which measure effects considered unlikely to be appreciable. This procedure can be easily followed by using specially devised "Table of Orthogonal Arrays" given below.

The types of investigations considered in this chapter are those in which the experimenter requires to know the behaviour of several factors over a defined range for each, beginning with two levels only. There are many investigations in which we wish to compare two conditions only of each of a number of factors, e.g. two different qualities (crude and purified) of one or more of the materials used in a process; high and low agitation rates, these rates being restricted by the design of the plant; two different designs of filter presses or filter cloths; two different units of a plant; slow and vigorous reflux conditions; two different solvents; and so on. Fractional replication of 3^n factorial designs is also considered.

The use of fractional factorial designs has now become widely accepted as an efficient way to carry out experiments involving a large number of factors. However, one of the difficulties in adopting fractional factorial experiments is that it is neither easy nor feasible to compute the estimates of all the factorial effects and to obtain the indispensable analysis of variance table on desk calculating machines. But, now, with electronic computers, it takes only a few minutes to complete the necessary computations for each character obtained from such experiments of moderate size. Thus the specially devised "Table of Orthogonal Arrays" and the data processing on electronic computers have made it quite easy for the industrial research workers in Japan to use various types of fractional replicates of 2^n and 3^n factorial designs.

6.2 Design for 3 factors in 2^2 observations

The basic principles and methods of construction of fractional factorial designs can be explained very simply by means of a 2^2 design. Denoting the factors by A and B, the lower level of each factor by - and the upper level by +, the four combinations of the factor levels constituting the complete design may be represented as in Table 6.1.

Table 6.1

Notation for 2^2 factorial designs

Observation	Level of factor A	Level of factor B	(A B)	Treatment combination
Y_1	−	−	+	(1)
Y_2	+	−	−	a
Y_3	−	+	−	b
Y_4	+	+	+	ab

The fourth column is added for a purpose shown later. The first observation, denoted by − −, involves the lower levels of both A and B; the second observation, denoted by + −, involves the higher level of A and the lower level of B; and so on for the other two observations. In the other notation, where the levels of factor A and B are denoted by (1), a; and (1), b respectively, the 4 (= 2^2) possible treatment combinations, i.e. the combinations of the factor levels, are shown respectively by (1), a, b, and ab. The sign notation for indicating levels of factors and treatment combinations is a convenient one, and is also appropriate to apply to the corresponding observations to calculate the main effects.

Thus

the effect of A at lower level of B: $Y_2 - Y_1$ or a − (1)

the effect of A at upper level of B: $Y_4 - Y_3$ ab − b

Main effect of A (average): $1/2\,(Y_2 + Y_4 - Y_1 - Y_3)$

or $1/2\,\{\, a + ab - (1) - b \,\}$

In the similar way,

Main effect of B: $1/2\,(Y_3 + Y_4 - Y_1 - Y_2)$

or $1/2\,\{\, b + ab - (1) - a \,\}$

These represent two of the comparisons between the four observations; the third comparison is obtained as the difference of the effect of A at lower level of B and that at upper level of B, yielding the interaction AB; and the signs given in the fourth column of Table 6.1 are simply the products of the signs for A and B, so that

Interaction AB: $1/2\,\{\,(Y_4 - Y_3) - (Y_2 - Y_1)\,\} = 1/2\,(Y_1 - Y_2 - Y_3 + Y_4)$

or $1/2\,\{\,(ab - b) - (a - (1))\,\} = 1/2\,\{(1) - a - b + ab\}$

Assuming the two factors A and B do not interact, the comparison represented by AB will be zero, apart from experimental error. It is then possible to utilize the AB comparison to measure the effect of a third factor (provided, as shown later, that the third factor does not interact with A or B). Denoting the additional factor by C, we "equate" C to AB in Table 6.1, and the signs then give the levels of factor C. This results in the design of Table 6.2 for measuring three factors in four observations.

Table 6.2 Design for three factors in four observations

Observation	A	B	C (=AB)	Treatment combination
Y_1	−	−	+	c
Y_2	+	−	−	a
Y_3	−	+	−	b
Y_4	+	+	+	abc

The main effect of C is

$$1/2\,(Y_1 - Y_2 - Y_3 + Y_4) \text{ or } 1/2 \left\{ c - a - b + abc \right\}$$

Using the comparison representing interaction AB to measure a main effect C is referred to as "equating" C to AB.

Another design exists for three factors in four observations, obtained by equating C to -AB. The design is similar to that of Table 6.2, but with the signs of column C reversed. Both are half-factorial designs and the two together comprise the complete factorial design of three factors at two levels (2^3). The treatment combinations for the two designs are:

Design 1 (C = AB): c, a, b, abc

Design 2 (C = -AB): (1) ac bc ab

It is interesting to note that the difference between the two designs represents the three-factor interaction ABC, which is given by

$$\text{Interaction ABC} = 1/4 \left\{ a + b + c + abc - (1) - ab - ac - bc \right\}$$

Important relationships are apparent upon inspection of Table 6.2: the fact that C has been equated to AB is shown by the signs of column C being the products of the signs of A and B, so that the comparison measures C when AB is zero. We see also that the signs of A are the products of B and C, while the signs of B are the products of those of A and C. It follows, therefore, that:

> The four treatment combinations of Table 6.2 may be used to estimate three main effects provided all interactions between them are zero, or may be assumed negligible.

In general, the factors A, B, and C may or may not interact, and therefore it follows that:

> After equating C to AB in a 2^2 design, the comparison measures C + AB, and the other comparisons measure A + BC and B + AC respectively. Similarly, after equating C to -AB, the comparison measures C - AB, and the other comparisons measure A - BC and B - AC respectively.

6.3 Construction of L_4 and L_8 Tables of Orthogonal Arrays

Our L_4 - Table of Orthogonal Arrays is only a version of Table 6.1, where the signs + and - are replaced by 1 and 2, respectively, turn it upside down, and the first two columns headed by Level of factor A and by that of B are interchanged, the result being shown in Table 6.3.

Table 6.3 L_4 - Orthogonal Arrays

No. of observation	Array Number (1)	(2)	(3)	
1	1	1	1	
2	1	2	2	
3	2	1	2	
4	2	2	1	
Array label	a	b	a b	
Group Number	1		2	

The original factorial effects of 2^2 design shown in Table 6.1 is referred to the "array label" given by the corresponding 'small' letters at the bottom of each of arrays. 'Capital' letters are left to indicate the factorial effects of a particular fractional factorial experiment, which are to be assigned to the arrays of the L_4-Table.

In the similar way, we can easily construct L_8-Orthogonal Arrays from the 2^3 factorial design, as shown in Table 6.4 and 6.5.

Table 6.4 Notation for 2^3 factorial designs

No. of observation	Factorial effects A	B	A B	C	A C	B C	A B C	Treatment combination
1	−	−	+	−	+	+	−	(1)
2	+	−	−	−	−	+	+	a
3	−	+	−	−	+	−	+	b
4	+	+	+	−	−	−	−	ab
5	−	−	+	+	−	−	+	c
6	+	−	−	+	+	−	−	ac
7	−	+	−	+	−	+	−	bc
8	+	+	+	+	+	+	+	abc

Table 6.5 L_8-Orthogonal Arrays

No. of observation	Array Number (1)	(2)	(3)	(4)	(5)	(6)	(7)
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2
Array lavel	a	b	a b	c	a c	b c	a b c
Group number	1		2			3	

The orthogonal arrays given in the above tables are conveniently used for getting the necessary treatment combinations in any fractional factorial designs in 4 and 8 observations, if the factorial effects required are assigned to the appropriate arrays of the table. The 'group number' at the bottom of the table is useful in finding arrays, to which factors in split-plot designs are assigned, because the figure 1 or 2 in the arrays with 'group number' 1 to 3 appears in groups of 4, 2 and 1.

The notation of the Table L_8 (2^7) shows that:

(1) L is the capital letter of Latin Square, indicating that this orthogonal property comes from Latin Square.

(2) this table is used for any experiment of 8 observations

(3) it has 7 orthogonal arrays, the number corresponding to the total degrees of freedom for 8 observations

(4) all factors applied to this experiment are those at 2 levels.

The 'orthogonal' property means that:

(1) every array has the same number of 1 and 2. (4 for L_8)

(2) every pair of arrays has the same number of the combinations (1, 1), (1, 2), (2, 1), and (2, 2). (2 for L_8)

6.4 Design of 8 observations (L_8)

Designs of 4 observations cannot cope with more than 3 factors because three factors exhaust all the independent comparisons between the four observations. The next larger fractional factorial design of the kind discussed above is one of eight observations. A design of this size can usually be carried out without serious practical difficulties, even with complex industrial processes.

Eight observations are sufficient to supply estimates of all main effects and all interactions for a complete factorial design with three factors each at two levels. Three main effects denoted by a, b and c, three-two-factor interaction by ab, ac and bc, and one three-factor interaction by abc are given as "array labels" at the bottom of the arrays, respectively.

If the three-factor interaction ABC may be assumed zero or negligible, the relevant comparison may be used to examine another factor D by 'equating' D to the array labelled by abc of Table 6.5.

The design is as Table 6.6.

Table 6.6 Design for four factors in 8 observations - $L_8(2^7)$

No. of test	Array Number (1)	(2)	(3)	(4)	(5)	(6)	(7)	Treatment combination
1	1	1	1	1	1	1	1	abcd
2	1	1	1	2	2	2	2	ab
3	1	2	2	1	1	2	2	ac
4	1	2	2	2	2	1	1	ad
5	2	1	2	1	2	1	2	bc
6	2	1	2	2	1	2	1	bd
7	2	2	1	1	2	2	1	cd
8	2	2	1	2	1	1	2	(1)
Array label	a	b	a b	c	a c	b c	a b c	
Factors assigned	A	B		C			D	

This table represents one half of a 2^4 factorial design, and it is therefore a half-replicate. In an investigation of the conditions of polymerization process of synthetic resin the object was to improve the strength of product. Four factors were examined:

A: percentage of additive agent: $A_1 = 6\%$, $A_2 = 5\%$

B: speed of agitation: $B_1 = 20$ r.p.m., $B_2 = 15$ r.p.m.

(r.p.m. = rounds per minute)

C: percentage of catalyser: $C_1 = 1.0\%$, $C_2 = 1.2\%$

D: supplier of raw material: D_1 D_2

It was considered unlikely that there would be large interactions between these factors, and as a first step it was decided to examine them in eight observations, i.e. by means of $L_8(2^7)$-Table. The comparisons obtained from the arrays No.(3), (5) and (6) in Table 6.6 were considered to measure the effects caused by experimental error. Randomization of the order of testing these eight treatment combinations was indispensable. The resulting experimental plan and the data are shown in Table 6.7.

Table 6.7 Experimental Plan

Order	No. of test	Level of factors A	B	C	D	data x (coded value)
1	4	6%	15 r.p.m.	1.2%	1	17
2	7	5%	15 r.p.m.	1.0%	1	14
3	2	6%	20 r.p.m.	1.2%	2	5
4	5	5%	20 r.p.m.	1.0%	2	0
5	1	6%	20 r.p.m.	1.0%	1	20
6	8	5%	15 r.p.m.	1.2%	2	1
7	6	5%	20 r.p.m.	1.2%	1	1
8	3	6%	15 r.p.m.	1.0%	2	26

6.5 Statistical analysis

There are several ways of calculating the effects and variances. The most straightforward way is to write the data in standard order and calculate the difference of the two sums; for the first level and for the second level, as shown in Table 6.8.

Table 6.8 Calculation of effects and variances

No. of test	Array Number (1)	(2)	(3)	(4)	(5)	(6)	(7)	data x	x^2
1	1	1	1	1	1	1	1	20	400
2	1	1	1	2	2	2	2	5	25
3	1	2	2	1	1	2	2	26	676
4	1	2	2	2	2	1	1	17	289
5	2	1	2	1	2	1	2	0	0
6	2	1	2	2	1	2	1	1	1
7	2	2	1	1	2	2	1	14	196
8	2	2	1	2	1	1	2	1	1
sum for the first level	68	26	40	60	48	38	52	84	1588
sum for the second level	16	58	44	24	36	46	32		
difference (d)	52	-32	-4	36	12	-8	20		
effect (d/4)	13	-8	-1	9	3	-2	5		
variance ($d^2/8$)	338	128	2	162	18	8	50		

Yates has developed a systematic tabular method which is particularly convenient when there are four or more factors. Write down the treatment combinations and the observations in standard order, as in the first two columns of Table 6.9.

Table 6.9 Calculation of effects and variances by Yates's method

	No. of test	Data (0)	(1)	(2)	(3)	effect (3)/4	S.S. = variance $(3)^2/8$
1	abc	20	25	68	84 = total	(10.5)*	882 **
2	ab	5	43	16	36 = 4 C	9.0	162
3	ac	26	1	24	-32 = 4 B	-8.0	128
4	a	17	15	12	-8 = 4 BC = e	-2.0	8
5	bc	0	15	-18	52 = 4 A	13.0	338
6	b	1	9	-14	12 = 4 AC = e	3.0	18
7	c	14	-1	6	-4 = 4 AB = e	-1.0	2
8	(1)	1	13	-14	20 = 4 ABC = 4 D	5.0	50
					160	Total	1588

* (3)/8, average, ** correction factor = $(\text{total})^2/8$

The column marked (1) is derived from the data column marked (0) as follows. The first entry in column (1) is the sum of the first two data (20 + 5); the second entry is the sum of the second pair of data (26 + 17); the third and fourth entries are the sums of the succeeding pairs (0 + 1), (14 + 1) respectively. This completes the top half of column (1). The lower half is derived from the data column by taking the differences of the same pairs, the second from the first in every case. Thus the first entry in the lower half is (20 - 5), the second (26 - 17), and the other two (0 - 1), (14 - 1). Column (2) is derived from column (1), and column (3) is derived from column (2), in exactly the same way as column (1) was derived from column (0), by summing and differencing the pairs of values. These operations yield exactly the same figures as have been calculated in Table 6.8. The actual effects are then obtained by dividing the entries in column (3) by 4. The number of operations (i.e. columns) of summing and differencing is equal to the number of factors.

The seven effects, A, B, AB, ..., ABC from 2^3 factorial design involve seven independent comparisons between the eight observations, and the capital letters showing these effects are given in the column (3) of Table 6.9, corresponding to the second level of arrays a, b and c, and according to the assignment of the actual factors to the array labels (given in Table 6.6) we can easily identify the factorial effects concerned. Since each of the quantities in column (3) of Table 6.9 is the algebraic sum of eight observations (in fact a difference between two sums of four), the sums of squares (equal in this case to the respective variances, there being only one degree of freedom for each effect) are obtained by squaring the corresponding quantities in column (3) and dividing by 8, yielding identically the same results as those obtained by the method usually given for analysis of variance. These quantities are given in the final column of table 6.9. Their total including correction term is 1588, which should agree with the raw sum of squares for the column (0), viz.:

$$\text{Raw sum of squares} = 20^2 + 5^2 + 26^2 + 17^2 + 0^2 + 1^2 + 14^2 + 1^2$$
$$= 1588$$

This check is usually sufficient, since compensating errors are unlikely to occur. A more detailed check on the computation is to use the fact that the sum of all the figures in the column (3) is equal to 8 times of the first data (8 x 20 = 160).

It is convenient to present the statistical analysis in the form of Table 6.10.

Table 6.10 Analysis of variance of Table 6.9

Source of variation	Degress of freedom	Sum of squares	Variance	F
Total	7	706		
A	1	338		36.3 **
B	1	128		13.8 *
C	1	162		17.4 *
D	1	50		5.4
error	3	28	9.3	

Combining three sums of squares for the arrays (3), (5) and (6), we obtain an estimate of error variance of 9.3, based on three degrees of freedom. The 5% and 1% values of F for $f_1 = 1$, $f_2 = 3$ are 10.1 and 34.1 respectively, so that the effect of percentage of additive agent A is highly significant; those of speed of agitation and of percentage of catalyser are significant at 5% level, and the difference of suppliers of raw material is not significant at 5% level, and in further experimental work on this process this factor may be disregarded.

The actual magnitudes of the significant effects are (see Table 6.9):

Effect of increase in percentage of additive agent A = 13.0

Effect of decrease in speed of agitation B = -8.0

Effect of decrease in percentage of catalyser C = 9.0

The condition for maximising the strength of product is then recognized as $A_1\ B_2\ C_1$ (i.e. 6% additive agent, 15 r.p.m. and 1.0% catalyser) and the expected strength under this condition is estimated by

(average) (effect of A_1) (effect of B_2) (effect of C_1)

$$10.5 + \frac{1}{2} \times 13.0 + \frac{1}{2} \times (+8.0) + \frac{1}{2} \times 9.0$$

$$= 25.5,$$

which is quite close to the value of the No. 3 test.

6.6 Confusion of effects in a fractional design. Aliases

Assuming that interactions of all orders are real, it follows that in a half-replicate each effect is confused with another effect, that is, the effects occur in pairs; in a quarter-replicate the effects occur in sets of four and each effect is confused with three others; in a eighth-replicate the effects occur in sets of eight, and so on. The effects which are confused in this way are termed Aliases. It is clearly imperative to determine the aliases for any proposed fractional design in order to avoid confusion of important effects. These aliases may be found from the assignment of the effects to the orthogonal arrays by multiplying the "array labels" in all possible ways and grouping the effects measured by the same comparison, that is, which are assigned to the same array. Applying these procedures to Table 6.6 gives the results shown in Table 6.11.

Table 6.11 Factorial effects of 2^4 assigned to orthogonal arrays

Array Number	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Defining contrast
Array label	a	b	a b	c	a c	b c	a b c	
Effects assigned (Table 6.6)	A = B C D	B = A C D	A B = C D	C = A B D	A C = B D	B C = A D	D = A B C	I = ABCD
Effects assigned (alternative design)	A = A B C D	B = C D	A B = A C D	C = B D	A C = A B D	D = B C	A D = A B C	I = BCD

It should be noted that in the multiplication of array labels the power is obtained in binary system (that is, mod. (2)); for example

$$a \times abc = a^2bc \equiv a^0bc = bc$$

$$ab \times bc = ab^2c \equiv ab^0c = ac$$

It is also possible to systematize the method of obtaining aliases because all the sets may be obtained from the defining contrasts. For this example, the defining contrast is

$$I = ABCD$$

Multiplying successively by A, B, C, etc., yields the above aliases given in Table 6.11. This result is perfectly general for all fractional designs, and leads to the rule:

> The effects confused with any given effect in a fractional design are derived by multiplying the defining contrasts by the given effect. The whole set of comparisons is derived by multiplying the defining contrasts successively by the main-effects, two-factor interactions, etc. until all the effects have been accounted for.

In this design we also note that all the main-effects are aliased with only three-factor interactions, and any two-factor interactions with another two-factor interaction. Therefore, if all three-factor interactions are negligible, this design can measure all main-effects A, B, C and D. And moreover if all two-factor interactions of D with A, B and C were zero, the two-factor interactions AB, AC and BC could be evaluated by this design.

At the bottom of Table 6.11, the alternative design is given, where the array to which the factor D is assigned is No. (6), being different from the original design. In this design, all main-effects except A are confused with some of two-factor interactions, and hence this alternative design is not better than the original one.

The following assumptions and principles are usually kept in mind for constructing fractional factorial designs:

(1) All three-factor interactions (and also interactions involving more than three factors) are considered negligible.

(2) All main-effects should be estimable (that means: they should not be aliased with other main-effects nor with any two-factor interactions).

(3) To maximise the number of estimable two-factor interactions. (That is: two-factor interactions which are not aliased with any other two-factor interactions).

References:
1 O. L. Davis: Design and Analysis of Industrial Experiment, Oliver & Boyd, 1956.
2 G. W. Snedecor & W. G. Cochran: Statistical Methods, 6th edition, The Iowa State University Press, 1967.
3 W. G. Cochran & G. M. Cox: Experimental Designs, John Wiley, 1957.
4 V. Chew: Experimental Design in Industry, John Wiley, 1958.
5 D. J. Finney: The Theory of Experimental Design, The Chicago Press, 1960.
6 T. Okuno & T. Haga: Experimental Design (in Japanese) Baifukan, 1969.