BC 2-M

LECTURES ON THE DESIGN OF EXPERIMENTS AND STATISTICAL METHODOLOGY

by

W.T. Federer
Cornell University

presented on
Dec.13, 1948, Jan.17, 1949, Jan.31, 1949
and Feb.7, 1949,

at the
Geneva Experiment Station
Geneva, N.Y.,

and sponsored by the
Geneva Chapter of Sigma Xi.

The four three-hour lectures consisted of a formal lecture of approximately one hour's duration with the remainder of the time being devoted to discussion. A number of the questions raised, with their answers, are presented herein.

The article written by Professor G.W. Snedecor was not discussed in the lectures but is included because it is considered to be an excellent introduction to a discussion on the design of experiments. Also, the procedure for testing ranked means by Tukey's method was not presented in the lectures but is included in the present write-up because several of the questions discussed were related to this subject. Additional material on the latin square design and the factorial experiment was given but these discussions were not included in the present manuscript.

DESIGNS FOR RESEARCH AND OBSERVATIONAL EXPERIMENTS

Chapter I.

INTRODUCTION

by

Walter T. Federer

Facts concerning the design of experiments are scattered throughout statistical literature. No single text suffices entirely for a course in experimental design. Leonard and Clark (1939), Goulden (1939), and Love (1936, 1943), have written texts which are suitable for students of agronomy. The above three references are excellent for their discussions on experimental techniques and procedures. Fisher's (1942) "The Design of Experiments" is suitable as a reference for research workers who have mastered the elementary statistical methods and is more advanced than desired for students learning the elements of experimental design. Cochran and Cox (unpublished) have planned to write a book on experimental designs but as far as known little or nothing will be included on the estimation and use of variance components. Therefore, in an attempt to fill this need and to give the elementary mathematical theory on experimental designs, the present manuscript has been written.

The student will need to understand basic statistical methods as illustrated and explained by Professors G.W. Snedecor, "Statistical Methods", 1946, and R.A. Fisher, "Statistical Methods for Research Workers", 1944. The latter reference is somewhat more advanced in some sections than is required for an understanding of the present dissertation on experimental designs. For the full comprehension of the mathematical theory underlying the analysis of various designs, the student would need to have mastered the principles of calculus and matrix algebra. However, an attempt was made to write

the text in such a fashion that the student may be able to grasp the essential ideas concerning the mathematical theory with only a good basic understanding of college algebra. Occasionally, some concepts of calculus and matrix algebra may creep in but in these cases the explanation will be given. The less advanced students should concentrate on the first part of the chapters, confining themselves to learning the "how" rather than the "why" of designing experiments for research or observational work.

At this point an explanation should be given concerning the terms "Research Experiments" and "Observational Experiments". But first of all it is necessary to define what is meant by the term "research". In the broad sense, research is the collection and analysis of data. As is well-known, there are many degrees of research. The scale runs all the way from doing something absolutely new down to the "refinding" of wellknown and well established facts. The upper part of the scale might be termed research while the lower end, the refinding of wellknown facts, might be termed "re-search". The scale in between has no clear-cut division but perhaps some experimental facts are desired and may be obtained by people who have not had the training required for carrying out the more technical phases of research problems. "Reflective thinking" and historical studies should not (unfortunately they are by some groups) be termed research <u>unless</u> the results are used to bring out new ideas and facts. In addition, to be of most value the results of a research study need to be put in a form that is available to others. Experimentation for personal satisfaction alone is useless to the advance of science and as one able administrator once put it: "One could not imagine a greater fools' paradise than to merely experiment in

whatever direction the mind might wander, with little heed paid to and no deductions made concerning the results obtained."

The term "research experiment" may be applied to a study investigating new ideas or facts. The exercisable degree of control on experimental conditions by the experimenter is quite adequate and is almost complete in some cases. The external factors affecting the character under observation may either be controlled or eliminated, i.e. they are held constant while the character under observation is allowed to vary. In some more complex research experiments two or more characters are allowed to vary while other factors that may affect the experimental results are kept constant. Of course, it is impossible to control all external factors and in practice the main ones are controlled and the factors with lesser effects are allowed to vary.

On the other hand, the experimenter may have little or no control over the influential external factors and he may have to choose a range of the character that appears in the population under surveillance. Also he may know, in general, the expected results of the experiment but wishes to obtain some measure of the character under observation. In this case the experimenter may choose some experimental layout or design from the following chapters or some other source which may be the same design as that for the research experiment. Since the degree of control over experimental conditions is limited and since the range of the character may be fixed in the population, it is suggested that this type of study be called an observational experiment in contradistinction to the research experiment which may include a range of the character larger than present in the population and is subject to a high degree of

control. Of course the argument over the division line between the two types of experiments boils down to an argument similar to the pros and cons over "What is truth?" Regardless of the terms used to designate the various experiments, the experimental designs or the field or laboratory layout will be the same. Therefore, this manuscript will be confined to a discussion of the various experimental designs with regard to the choice of the experimental material size, shape, and number of the individual units; the number of repetitions; construction, randomization, field or laboratory lay-out, and statistical analysis; the choice of appropriate experimental errors; the choice of an experimental design; and the elementary theory involved in tests of significance and in estimation of variance components.

In all cases except for the chapter on systematic designs, the discussion will be limited to designs which are subject to statistical analysis. The allotment of a particular treatment (variety, feeding ration, size of farm, level of fertilizer, baking condition, etc.) to a specified plot or "area" in the experimental area (the site at which an experiment may be conducted, e.g. a greenhouse bench, a set of ovens, a period of time, etc.) must be at random. The element of chance in the random allotment of the treatment to the experimental plot permits the experimenter to obtain unbiased estimates of the means and variance and to make probability statements concerning them. Fisher, "The Design of Experiments," 1942, gives an excellent discussion on validity and randomization and tests of hypotheses.

The following is a general classification of experimental designs

Systematic
Completely Randomized
Randomized Complete Blocks

Latin Squares and Variations
- Latin Square
- Graeco-latin Squares
- Other Latin Squares (plaid, half-plaid, and quasi-)
- Cross-over
- Switch-back or Reversal

Incomplete Blocks
- Split-plot
- Lattice
- k^n or n-dimensional Lattices
- p x q Lattices
- Incomplete Lattice Squares
- Youden Squares

As with all statistical manuscripts, the present one is subject to the criticism that the author did not use "standard" statistical symbolism. Since the last three words of the preceding sentence imply something different to writers in the statistical field, it is perhaps best to list and define the symbols used in the majority of places in the text.

X_i or Y_i, i= 1,2,..., n= a specified number, is the individual measurement or count on the observed character. X_1= measurement on the first individual, X_2=measurement on the second individual.

X_{ij} or Y_{ij}, i=1,2,...,n and j=1,2,...,m, is the record of the i th individual of the j th classification.

$\frac{\sum_{i=1}^{n} X_i}{n} = \frac{X.}{n} = \bar{x}$, i.e., the mean of n observations $X_1, X_2, \ldots, X_n$, is the sum of the observations, $X_1 + X_2 + \ldots X_n$, divided by the number n. $\bar{y}$ is obtained similarly. Some texts use m for a mean with m_x denoting the mean of the X_i and m_y the mean of the Y_i. $\bar{x}$ in normal populations is an unbiased estimate of the population mean μ, a known or unknown parameter.

$\frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} X_{ij} = \frac{X_{..}}{mn} = \bar{x}$ = overall mean of a 2-way classification.

$\frac{1}{m} \sum_{j=1}^{m} X_{ij} = \frac{X_{i.}}{m} = \bar{x}_{i.}$ = the mean over all $j = 1,2,\ldots, m$ for the i th classification. Likewise,

$\frac{1}{n} \sum_{i=1}^{n} X_{ij} = \frac{X_{.j}}{n} = \bar{x}_{.j}$.

$(X_i - \bar{x}) = x_i$ = deviation from the mean.

$\sum_{i=1}^{n} (X_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2$ = sum of squares of deviations from the mean.

$\frac{\sum_{i=1}^{n} X_i^2}{n-1} = s^2$ = the variance or sometimes mean square of a single observation and is an unbiased estimate of the population parameter σ^2 in normal populations.

$\frac{\sum_{i=1}^{n} x_i^2}{n(n-1)} = s^2_{\bar{x}}$ = the variance of a mean of n individuals and is an unbiased estimate of the population parameter $\sigma^2_{\bar{x}}$ in normal populations.

$\sqrt{s^2} = s$ = the standard error of a single observation or the standard deviation.

$\sqrt{s^2_{\bar{x}}} = s_{\bar{x}}$ = the standard error of a mean.

$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = s_{\bar{x}_1 - \bar{x}_2} = s_d$ = the standard error of a difference between 2 means based on n_1 and n_2 individuals, respectively. In the event that $n_1 = n_2$, $s\sqrt{\frac{2}{n}} = \sqrt{2}\, s_{\bar{x}} = s_d$.

b_{12} = regression coefficient of the variate X_{1i} on the variate X_{2i} and is an unbiased estimate of the population parameter β_{12}. Other symbols will be used to designate regression coefficients also but these will be explained when used.

$b_{12.3}$ = partial regression coefficient of X_{1i} on X_{2i} independent of X_{3i}.

r_{12} = total or zero order (Snedecor, 1946) correlation coefficient of X_{1i} with X_{2i} and is an estimate (though biased) of the population parameter, ρ.

v = coefficient of variation or variation coefficient, which has often been designated as c.v. or just c.

d.f. = degrees of freedom.

s.s. = sums of squares.

m.s. = mean square.

A. of V. = analysis of variance.

EMS = error mean square.

Eff = efficiency expressed as a percentage.

$F = \frac{\text{greater m.s.}}{\text{lesser m.s.}}$ = Snedecor's F = a test of significance.

t = "Student's" t = a test of significance.

X^2 = chi-square = a test of significance.

$z = \frac{1}{2} \log_e F$ = Fisher's z = a test of significance.

A <u>significance level</u> is a predetermined probability of obtaining deviations as large as or larger than a specified number. Values from tests of significance of approximately the specified constant or larger are deemed significant.

A <u>variable</u> is the characteristic or trait under observation and of interest at the moment. The variable may be continuous (measurement data) or discrete (enumeration data). The individual measurements or counts of the variable are called the <u>variates</u>. Thus heights of people represent a variable, say X, and the height 67 inches represents the variate.

A <u>population</u> represents all the individuals of interest or of a given characteristic (or characteristics). The number of individuals in the population may be finite or infinite. For example a discrete population might represent the plants on a greenhouse bench. Usually the population about which inferences are made is infinitely large and a fraction is all that is ever observed.

A <u>sample</u> represents a fraction of the population. A <u>representative sample</u> is as its name specifies, representative of the population from which drawn. A <u>random sample</u> is a sample drawn in such manner that every possible sample will have an equal chance of being drawn. For example there are 20 possible samples of 3 taken from 6 objects and a random sample of 3 items would be one that was drawn in such a manner that any of the other 19 samples have equal chances of being drawn.

A <u>parameter</u> is a fixed value, e.g. 2, and is the value of the variable which specifies the population. The population parameter may be known or unknown.

A <u>statistic</u> is an estimate of the population parameter. If

the average of a large number of estimates does not approach the population parameter, then the statistic is said to be biased, and conversely for unbiased statistics.

A <u>distribution</u> of the variates for the population may be represented by some frequency curve or by a mathematical formula, for example the formula for a random sample of n individuals from the normal, binomial, and Poisson distributions, respectively are of the form

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} e^{-\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2\sigma^2}} dX_i ,$$

$$\sum_{X=0}^{n} {}_nC_X p^X q^{n-X} ,$$

and

$$\sum_{X=0}^{n} e^{-m} \frac{m^X}{X!} .$$

An <u>individual unit</u> or <u>unit of observation</u> represents the smallest unit for which an observation is made.

The <u>treatments</u> in an experiment represent items which are being tested. The term "treatments" is very general and is often used in place of varieties, fertilizers, baking treatments, feeding treatments, method of dying fabrics, dosages of drugs, etc.

The <u>experimental site</u> or <u>area</u> represents the place or time over which an experiment is conducted. In field trials, it may represent the land which the experiment occupies; in home economics experiments, it may represent the time and pieces of equipment used, etc.

If all treatments have been included a proportional (usually equal) number of times and are together in one part of the experimental area, this fraction of the total site is called the <u>replicate</u>,

If the replicate is repeated, this is known as a replication or repetition of the treatments. The treatments may be replicated without having replicates (see Chapter III).

Uniformity or blank trial data represent the measurements taken for a variable on several (usually many) units of observation of the same treatment. Various experimental designs may be laid out on the uniformity trial data and the "treatments" may be assigned to the plot yields at random. The comparisons among these "treatments" would represent comparisons for the same treatment or "dummy" comparisons.

The term deviation has been explained previously to denote the discrepancy between the variate, X_i, and the estimate of the parameter, say $\hat{\mu}$. The term difference refers to the discrepancy between two variates, say X_1 and X_2, or between two treatment means (Snedecor, 1946, Examples 2.6 and 2.7).

The following material is reprinted from the Proceedings of the Auburn Conference on Statistics applied to research in the Social Sciences, Plant Sciences, and Animal Sciences, held September 7-9, 1948 at Auburn, Alabama, and it is the lecture given by Professor G.W. Snedecor, Statistical Laboratory, Iowa State College, Ames, Iowa, at the conference.

SOME PRINCIPLES OF EXPERIMENTAL DESIGN

by

George W. Snedecor, M.A.
Research Professor of Statistics
Iowa State College

Our courses in statistical methods and experimental design tend to emphasize the mechanical features of the subject--the lay-out of the experiment and the calculation of statistics. It is my purpose to discuss and illustrate two of the more fundamental principles that govern design. These are simple and obvious; yet they seem to include the important practical requirements of good experimentation. The philosophical aspects of the subject I do not intend to mention.

The two principles of design which I have in mind are as follows:

First: To provide unambiguous answers to the questions asked, and

Second: To get the answers with a minimum expenditure of resources.

As a preliminary, let us consider for a moment the questions that may be asked. From a long experience in reading project outlines, I conclude that these questions are often not very definite. Frequently I have found the proposed experiment wholly inadequate to realize the stated objectives. For instance, the objective may read, "To determine the adaptability of some varieties to Iowa conditions,"

followed by a proposal to try the varieties during a single season on some one set of plots. Of course, I realize that Iowa is supposed to exhibit a uniform climate and soil, but no Iowa agronomist would admit this.

At times, it would appear that the experimenter is concerned merely in demonstrating his own superior knowledge. I have in mind the case of a man who had designed a split-plot experiment to answer this question: When is the best time in the fall to make the last cutting of a certain perennial forage crop? While I was examining the first year's results, he explained that one average was unusual because variety A should have been superior to variety B in the earlier cuttings. He even described the physiological characteristics that made this so. The other averages he considered normal. When I asked for the results in succeeding years in order to get an answer to the question presumably asked--which treatment maximized the yield over the effective life of the plots?--I was blandly informed that other experiments demanded his attention after one season so that this one was abandoned. I gathered that he already knew all the answers to the experiment as conducted and that he really wasn't much interested in the specified objective of his project. The experiment seemed to be a filler which warranted the payment of his salary and expenses until something more promising turned up.

If you say that my examples are unusual, I fear you are wrong. My experience leads me to suspect that some large fraction of our so-called experimentation, perhaps even a majority of it, is conducted in an equally vague and unproductive fashion. Experimenters often remind me of small boys investigating the mechanism of a

clock. They are eager to see what makes it tick, but not much interested in devising a superior time-piece, much less in writing a scientific report of their investigation.

Some good experimenters do not require very definite answers to questions. This is notably true in breeding tests with large numbers of varieties. The answer desired is given by an array of the varieties according to their performance in one or more respects. An over-all test of significance is available for one character--tests of other characters would not be independent of the first, I suppose. No way is provided for combing the various measurements into a single one determining excellence. No test is furnished for distinguishing between individual varieties. The only question asked, usually, is about the yield of those varieties that prove to have acceptable stalk strength, for example; or ear height. Even nutritive value is not often used as a criterion of selection. Yet the notable achievements of the breeders is convincing evidence that this simple answer is the only one needed.

Contrasted with these experiments in which vague or few questions are asked are those in which many individual comparisons are planned, each with an appropriate test of significance. The factorial experiment is a shining example of this type. But I wish to avoid any implication that an experimenter should choose a design merely because it answers many questions. The efficient design is the one that answers the questions which he is asking. If he has a single question, then answers to other questions may be superfluous and the more complicated designs may be inefficient.

Assuming then that the experimenter has one or more definite questions, he should choose a design that will furnish unambiguous

answers to them. This is the first of the principles which I wish to discuss. A single, clear-cut answer from an experiment is not easy to attain. Some type of replication is always necessary. Appropriate controls must be set up to segregate the desired effects from those that confuse the issue. Correlated variables difficult to regulate may be measured and controlled by covariance. Tests of significance must be provided to distinguish between treatment effects and sampling variation. These are some of the familiar devices for achieving unambiguous answers.

Ambiguous answers mean little or no information from an experiment. Let me illustrate. Not long ago I was asked by an experimenter to give an opinion as to which of four methods was superior for testing significance. Six protein supplements had been tried on chicks and all the 15 pairs of treatments were to be tested, ignoring the fact that only five degrees of freedom were available. This aroused my suspicion, so I made detailed inquiries about the design. I found that all the chicks receiving Treatment A were kept in a single separate enclosure, as were the lots getting the other treatments. I found that the sexes were mixed in the lots in unknown ratios. I found that one supplement had been substituted for another in equal weights irrespective of protein content. Of course, no record was kept of individual food consumption.

Practically any question asked of this experiment would have half a dozen different answers. If the lot having Treatment A gained significantly more than that with Treatment B the answer might be (i) that the concentration of protein in A was greater than that in B; or (ii) that the biological value of the first protein was superior to that of the second; or (iii) that there was a larger

proportion of males in lot A than in lot B; or (iv) that environmental conditions (including incidence of disease) were more favorable in pen A than in pen B; or (v) that there was correlation among the gains of the birds within the lots resulting in a downward bias in the estimate of error; or (vi) that ration A was more appetizing than B. So far as the gain is concerned, the last answer could be eliminated by using some ratio such as gain per unit of food eaten; but the ambiguity would not be removed because there is no test of significance provided for the ratio.

Remember, now, that this man was asking about the merits of various tests of significance, seemingly unaware that the differences he was testing were completely meaningless. Such is the effect of the manner in which we teach statistics!

I have cited an extreme example, but not one that is unique. I remember some data that were sent around to several statisticians only a few years ago. The data were from a ten-year experiment on cultural treatments of pecan trees. All the statisticians agreed that the investigator had been particularly ingenious in arranging the experiment so that every comparison was ambiguous. But the U.S.D.A. paid money to this man for 10 years, money from your pocket and mine, with never a competent check on the design or progress of the experiment. I suspect that there are many of this type of experiment now in progress.

Another common kind of ambiguity is one already mentioned, the trial of three or more treatments with no specified degrees of freedom to be tested. The means are estimated in an unbiased manner and can be arrayed according to the order in which they turned up in the particular experiment. But there is no way to differentiate the populations. Even though the F-test indicates population differences,

the answer as to which population differs from which has no single answer.

For illustration, I return to the chick-feeding experiment. There were five protein supplements tried in addition to the standard. Of course, each of the five could be compared to the standard, but this is not an efficient set of comparisons. It turned out that there were three vegetables and two animal sources, so that four of the possible five independent comparisons could be specified. The only remaining ambiguity would be in the two degrees of freedom for vegetable sources.

I believe that careful forethought would eliminate this type of ambiguity in a great many experiments. Among the experimental treatments, there are nearly always relations which would lead to at least a partial set of orthogonal comparisons. Questions based on these comparisons would be answered unambiguously, with less definite information about the remaining ones.

A somewhat more subtle type of ambiguity is that which springs from samples which are too small. An experimenter may know with reasonable certainty that a difference cannot be larger than 20 percent, yet he may use only enough replications to detect a difference of 50 percent. At the end, he is unable to distinguish between a real difference and sampling error.

In reviewing project outlines, I have often had to call attention to the fact that the proposed sample size was too small to detect any but the most obvious differences. Experimenters often take a peculiar view of this situation. They say that they don't expect significant differences, but they just wish to observe how the treatments behave under trial. They are the small boys of whom

I spoke before--they like to play around with the experimental material, hoping, no doubt, that something interesting may turn up.

It is hard for many experimenters to realize that, before the experiment is performed, definite statements can be made about the prospect of an unambiguous answer. Foreknowledge of the probability that an experiment will detect a specified difference is confused with fore-knowledge that there is such a difference. If you tell these people that, under specified conditions, the probability is 0.8 that their experimental questions will be answered, they scornfully say, "Well, why do the experiment? You can tell us now what the answer is."

I think this is one of the most spectacular feats of statistics. Experimenters have always begged us to tell them how large a sample to take, but when we tell them, the language seems unrealistic. Also, the necessary number of replications is ofttimes so great that it will not be believed. The new tables of Mood, et al., now in press, enable one to make decent estimates of the number of replications necessary to yield unambiguous answers to specified questions. The experimenter himself must decide on the size of the population difference he wishes to detect and must help to estimate the standard deviation of the population. The latter is the weakest link in the chain, but it can often be made quite strong. Under the specified assumptions, a determinate sample size will, with stated probability, yield a significant difference. If it does not, the answer is that any difference there may be in the population is less than the one which the experiment was designed to detect.

The experimenter who insists on using too small a sample often feels justified by his results. Even if the population difference

is zero, he has a five-in-a-hundred chance of significance; while if there is a real difference, his chance of detecting it is greater. But the canny investigator will not easily be misled by such an outcome. He will be wary of a large sample difference, because he knows that a large population difference would have been discovered long ago. He will be equally wary of a small experimental error--he knows the usual sizes in such experimentation. But with an inadequate sample size, about the only chance of attaining significance, in case the population difference is as small as suspected, is an unusually large sample difference or an unusually small sample error. So the competent experimenter, knowing the size of sample necessary to detect a reasonable population difference if it exists, is not deceived by significance in an experiment of inadequate size. But, of course, this competent investigator would not have wasted his time and my money in carrying out such a too-small experiment. It is the incompetent who braves the inadequate sample and proudly publishes his accidental findings.

I had quite a lengthy correspondence, a couple of years ago, with a botanist in one of our greater universities who asked me about the adequacy of the size of his experiment. He was seeking to detect a small difference with large variation. When I told him the required size he was frankly scornful. I cited formulas and suggested that he have them checked by a girl in his employ who had completed a course in statistics given by an uncontaminated, pure mathematician. I never heard from him again. I suspect his work has been fruitful of publications and that he is held in great esteem by his brethren.

Having assured ourselves of an experiment that will yield

unambiguous results, we set up the second principle that it shall be done at a reasonable cost. "At minimum cost" is the phrase which I used in the beginning, but that is an unattainable ideal. I think we shall not be censured even if we habitually waste a dollar or two on our experiments. In general, economy means low experimental error.

For a selected set of answers, to be attained with specified accuracy, economy is achieved in three familiar ways. First, there is the selection of homogeneous experimental material. When this is possible, the experimenter must beware lest he make the base of his structure too narrow. He might get answers about a small segment of his population, and these might not hold in other related parts. If this is a reasonable conjecture, then he should first investigate such possible interactions. If he finds them absent, he can proceed with his more detailed experiment with the selected portion of the population. Otherwise, he must enlarge his plans so as to answer the questions, at least ultimately, for each of the homogeneous parts of the population concerned. I suspect that the existence of these interactions is unusual in many fields of investigation. Nonetheless, their possible presence must not be ignored.

The second method of achieving economy is by choice of an appropriate plan for the experiment. Randomized blocks, the latin square, and various incomplete block devices are familiar to you. This is the method of economizing which most taxes the ingenuity of the investigator. In this conference, you will hear one of the speakers describe his device of splitting plants in order to have more homogeneous material for his experiments. He also had to

decide how to use the pairs of plants to test three fertilizers--calcium, phosphorous and potassium. He could have chosen a balanced incomplete block plan with two combinations per block; this would have given him equal amounts of information on main effects and interactions. But he decided that the interactions were not pertinent to his inquiry, so he chose a combination of paired experiments in which the three main effects were evaluated with maximum efficiency, leaving the interactions to be tested with the large error due to the variation among half-plants from different plants.

This is a nice decision which experimenters tend to shirk. Over and over again I have tried to get men to decide which set of treatments to put on the split plots and which on the main. They are avaricious of information and are reluctant to sacrifice any from either set. The same kind of decision often has to be made in order to get an experiment large enough to promise unambiguous answers--some treatments must be sacrificed, and giving them up is like having eyeteeth pulled. The ideal experimenter is he who can choose that plan which will give unambiguous answers to necessary questions without undue expenditure.

The third method of achieving economy, or reducing experimental error, is by statistical control of extraneous variables. This is the appropriate method if homogeneous material is either undesirable or not available. It has the virtue of broadening the basis of inference. If related variables can be measured rather than experimentally controlled, their effects can be eliminated from the estimate of error; and in the end one has not only the answers desired about the experimental variable but also measures of its relations to the others. I think this method of covariance is growing in popularity. It is effective not only in reducing

experimental error but also in broadening the scope of information at small cost.

In conclusion, I wish to discuss one of the causes of ambiguity, with consequent high cost, in much of our experiment station work. A friend of mine, shrewd but inclined to be caustic, placed the fraction of useful expenditure of station funds at 3 percent. My own estimate is somewhere between 20 percent and 40 percent. Why is this?

One reason is that experimenters, eager for information, are too impatient to allow the necessary time to get it. They want it all the first year. The result is that even a big experiment is far too small to yield unambiguous answers. So, the inadequate experiment must be tried again next year. This goes on indefinitely. After ten or twenty years of this kind of thing, the fundamental questions are still unanswered, but the leader is ever hopeful that next year he will have the longed-for results. If he is told that the experiment is too small to yield the answers he wants, he complacently replies that the director will not provide him with sufficient funds; so, he continues to fritter away the resources he has.

Every experimenter should have some kind of twenty-year plan. The early years should be spent in getting reasonably certain answers to subsidiary questions about techniques, the effects of environment, etc. Each experiment should be large enough to get an unambiguous answer to at least one question. As the climax approaches, the main question emerges with increasing clarity, and at the end is answered without ambiguity.

I am not so naive as to suppose that this ideal program will

move along smoothly and without interruption. At the end of twenty years the question may seem as far from solution as it was at the start. But every preliminary answer is available information and denotes progress toward ultimate success.

I hope I have succeeded in convincing you that my main th[illegible] is true: that the investigator may, with specifiable certainty, know in advance that his experiment will, without excessive cost, yield unambiguous answers to the questions asked.

Chapter II.

SYSTEMATIC DESIGNS

by

W.T. Federer.

Prior to the development of modern experimental designs, experimenters had tried various arrangements which are not subject to the laws of chance. Various systematic schemes of ordering or arranging the treatments in the various repetitions have been devised. One such scheme might be to arrange all duplicates, triplicates or etc. of the treatment together. Suppose the experimenter wished to test three treatments A, B, and C and that he decided to have 4 repetitions of each treatment. With the above scheme the arrangement of the 3 treatments over the experimental area could be one of the following:

A	A	A	A	B	B	B	B	C	C	C	C

,

A	A	A	A
B	B	B	B
C	C	C	C

,

A	B	C
A	B	C
A	B	C
A	B	C

,

A	A	B	B	C	C
A	A	B	B	C	C

From fertilizer, yield, and other trials or experiments it was evident that it might be better to test treatments A, B, and C together in a compact block and then to repeat these blocks with a systematic ordering of the treatments in each block or repetition. One of the more common types of systematic arrangements in which the treatments are repeated several times is the following:

Replicate I	Replicate II	Replicate III
A B C	A B C	A B C

or

Replicate I	A B C
" II	A B C
" III	A B C

In this case the ordering of the treatments is exactly the same in every replicate (a unit which contains all the treatments). Another systematic arrangement is the following:

Replicate I	Replicate II	Replicate III
A B C	C A B	B C A

In this case each treatment occupies each order in the replicate.

From the last systematic arrangement experimenters may have felt it necessary to place the treatments so as to eliminate soil heterogeneity in two directions and proposed the "diagonal square" (see Fisher, 1942); for three treatments the design would be:

A	B	C
C	A	B
B	C	A

and for 5 treatments the design would be:

A	B	C	D	E
E	A	B	C	D
D	E	A	B	C
C	D	E	A	B
B	C	D	E	A

In order to eliminate the effect of A appearing on one diagonal, a systematic arrangement involving the Knight's Move was used, i.e. one down and two over. This arrangement for 5 treatments in 3 replicates gives the following design:

Replicate I	A	B	C	D	E
" II	D	E	A	B	C
" III	B	C	D	E	A

and involving 5 replicates the arrangement is:

A	B	C	D	E
D	E	A	B	C
B	C	D	E	A
E	A	B	C	D
C	D	E	A	B

Fisher (1942, The Design of Experiments) states that the above design has been known in Denmark since about 1872 but that it is usually ascribed to the Norwegian, Knut Vik, and called the Knut Vik Square. A fairly good account of this design is given by Fisher (1942), but the statistical analysis for it has only recently been worked out (Øvind Nissen, 1949, unpublished paper.)

Numerous other systematic arrangements have been devised, the various experimenters attempting to outguess natural variation. Regardless of the type of systematic design they all have relatively the same advantages and disadvantages. The advantages are often given as (the quotation marks are those of the author):

(i) Simplicity. Many experimenters feel that planting, note-taking, and harvesting in agronomic trials are facilitated by using systematic arrangements. In judging or scoring experiments it is sometimes felt that the judge will be better able to "discriminate" between the treatments if he (the judge) knows the order in which the treatments occur in the different repetitions.

(ii) The systematic design provides "adequate" sampling of the experimental area. That is, it allows for "intelligent placement" of the various treatments.

(iii) Varieties may be arranged in order of maturity or fertilizer treatments in order of increasing fertility.

(iv) It may be desirable to alternate dissimilar varieties (say, bearded versus beardless barley) so that natural crossing or mechanical mixtures can be detected in subsequent years.

(v) There is no need to randomize since the heterogeneity of the experimental site is such as to randomize the effects on the treatments. (This does not lessen the effect of one treatment on another or of a single arrangement; these facts should not be ignored.)

The disadvantages of the systematic designs are that there is no correct measure of the variation and the correlation between adjacent plots may lead to systematic errors in assessing treatment differences. The latter point is easily illustrated by the following systematic arrangement:

Replicate I	Replicate II	Replicate III
A B C	A B C	A B C

,

where the yield gradient is assumed to exist from left to right. Even though treatments A, B, and C may be the same thing, the experiment would show A to be better than B, and B better than C. In the event that the treatments were different, their differences may be exaggerated or underestimated depending upon the arrangement of the treatments.

Fisher (1942, The Design of Experiments, sections 27 and 34) discusses the effect of systematic arrangements on tests of significance and in the estimation of an error variance. Suppose that systematic arrangements are tried on uniformity trial data, (Plot data on the same treatment over the whole of the experimental

site or area).[1] Then, the treatments in the experimental area would all be the same thing. The total sum of squares would be a constant regardless of what arrangement was chosen. If the experimenter was able to "intelligently" place the treatments so that all were subjected to about the same heterogeneity, then the sum of squares due to the differences between dummy or pseudo-varieties would be decreased. The decrease must be counterbalanced by an increase in the error or remainder sum of squares since the total is a constant, or

Total s.s. = s.s. among dummy var.+ s.s.within var.

If, on the other hand, the experimenter does a lousy job of "intelligently" placing the dummy varieties, the estimate of the error sum of squares will be smaller than it really should be and the differences between the varieties will be exaggerated. Some arrangements may consistently underestimate the error variance. The amount of underestimation is unknown and any attempt to obtain an estimate of the error variance from systematic arrangements is pretty much a matter of guesswork.

It is suggested that students in experimental design read this chapter for its historical value. They should never design an experiment in a systematic manner but rather should choose an experimental design that is subject to statistical analysis. The remainder of the manuscript will be confined to a discussion of such designs.

---

1. For work on use of systematic designs on uniformity trial data see Odland and Garber, 1928, Journ. Amer.Soc.Agron.,20:93-108; Tedin,1931, Jour.Agr.Sci.21:191-208; and Pan,1935, Jour.Amer. Soc.Agron. 27:279-285. The first reference states that the standard deviations obtained from systematic arrangements were somewhat lower in all cases, than those obtained from random arrangements on soybean uniformity trial data. Tedin(1931) found that the variation within 6x5 blocks was uninfluenced by arrangement, diagonal or random, for estimating the error. He suggested random arrangements for the highest degree of scientific accuracy. On rice uniformity yield data Pan (1935) found that the deviations among varieties (in reality they are dummy or pseudo-varieties) were much larger than might be explained on the basis of random sampling.

Chapter III.

COMPLETELY RANDOMIZED DESIGNS.

by

W. T Federer.

The simplest of all designs having a random arrangement is the completely randomized design. The design may be defined as one in which the treatments are randomly arranged over the whole of the experimental site. No effort is made to confine treatments to any portion of the whole area. The number of repetitions of any one treatment may vary. The completely randomized design is usually chosen when the variation over the whole experimental unit is relatively small. An example of the lay-out of a completely randomized design would be the following, in which the five treatments A, B, C, D, and E are repeated four times each on the twenty units representing the whole of the experimental area. -

(E) (1)	(E) (8)	(C) (9)	(B) (16)	(E) (17)
(A) (2)	(D) (7)	(D) (10)	(B) (15)	(A) (18)
(B) (3)	(C) (6)	(A) (11)	(C) (14)	(B) (19)
(E) (4)	(D) (5)	(A) (12)	(D) (13)	(C) (20)

Such an experiment might have been designed for 20 pots on a greenhouse bench, a series of 20 soil analyses, the 20 animals in a feeding trial, 20 cake-pans in an oven or the 20 successive bakings of single cakes in an oven, records on five litters of four pigs each, or some other type of experimental material.

Example III-1

In order to best illustrate the statistical analysis a numerical example was chosen from a guayule experiment on the dry weight of the shrub (leaves not included) on plants that had completed one year's growth in the field. The plants of variety 109 (a 54 ± chromsome strain of guayule) were classified with regard to trueness of type. The characteristic plants of 109 were listed as normals = N. The remainder of plants differed considerably in appearance from the normals and were divided into two categories, offtypes = O and aberrants = A. It was desired to know if the three types of plants varied with regard to dry weight of shrub. A random selection was made of 5 plants of each type and the dry weights obtained. The 15 selected plants were scattered over the experimental area in the following manner (the type of plant, A, N, or O, is listed first followed by number of plant and dry weight of shrub in grams):

A-1-34

O-3-84 N-2-87

A-4-12

A-5-20 N-6-167

N-7-112

N-9-104 O-8-134

A-10-5

O-11-86

N-12-106

O-13-120

A-14-48 O-15-108

The table of yields and sums of squares is given below:

	Normals =N	Offtypes =O	Aberrants =A	Total
	87	84	34	
	167	134	12	
	112	86	20	
	104	120	5	
	106	108	48	
Totals = ΣX	576	532	119	1227
Means	115.2	106.4	23.8	81.8
ΣX_i^2	70054	58472	4029	132555
$(\Sigma X)^2/n$	66355.2	56604.8	2832.2	100368.6
Σx_i^2	3698.8	1867.2	1196.8	32186.4

Since such comparisons as normals versus offtypes and aberrant versus the mean of the normals and offtypes are logical comparisons to make and since they formed a part of the hypothesis in designing the experiment, these contrasts are given in the following analysis of variance table:

Source of variation.	Degrees of freedom.		Sum of squares.		Mean squares.	
Among types	2		25,423.6		12,712.8	
Nvs O		1				193.6
N+Ovs2A		1				25,230.0
Within types	12		6,762.8		563.6	
Within N		4		3,698.8		924.7
" O		4		1,867.4		466.8
" A		4		1,196.8		299.2
Total	14		32,186.4			

The total sum of squares is

$$\sum_{i=1}^{n}\sum_{j=1}^{k} X_{ij}^2 - \frac{X^2_{..}}{nk}$$

$$= 87^2 + 167^2 + \ldots + 5^2 + 48^2 - \frac{(1227)^2}{15}$$

$$= 132555.0 - 100,368.6 = 32,186.4 \text{ with } 14 \text{ d.f.}$$

The sum of squares among types is

$$\sum_{i=1}^{n} \frac{X_{i.}^2}{k} - \frac{X^2_{..}}{nk}$$

$$= 125,792.2 - 100,368.6 = 25,423.6 \text{ with 2 d.f.}$$

The sum of squares among plants within normals is

$$\sum_{j=1}^{k} X_{1j}^2 - \frac{X_{1.}^2}{k} = 87^2 + \ldots + 106^2 - \frac{576^2}{5}$$

$$= 70,054 - 66,355.2 = 3,698.8 \text{ with 4 d.f.}$$

The sums of squares among plants within offtypes and aberrants is 1,867.2 and 1,196.8 respectively. The pooled within type sum of squares is 3,698.8 + 1,867.2 + 1,196.8 = 32,186.4 with 12 degrees of freedom.

An orthogonal set of comparisons among the three types would be:

Comparison	N	O	A
NvsO	+	-	0
N+O vs A	+	+	-2

The sum of squares for the comparison normals versus aberrants, is

$$\frac{(576 - 532)^2}{5(1 + 1)} = 193.6$$

and for the comparison, normals and offtypes versus aberrants, is

$$\frac{\{576 + 532 - 2(119)\}^2}{5\ (1 + 1 + 4)} = 25,230.0$$

The pooled error mean square, 563.6, with 12 degrees of freedom is used to test the comparisons of the normals with offtypes and the normals and offtypes with the aberrants, with the respective values of

$$F = \frac{193.6}{563.6} = 0.34 \quad \text{and}$$

$$F = \frac{25230.0}{563.6} = 44.77$$

There is a strong hint that the within type variances for the three types of plants are different. However, one should not expect to detect differences in variances unless the degrees of freedom are fairly numerous or unless the variances are extremely divergent. Bartlett's test for homogeneity of variances as illustrated by Snedecor (1946,p.250) for equal numbers of individuals per lot follows:

$$\chi^2 = 2.3026\ (5-1)\ \left(3 \log_{10} 563.6 - \log_{10} 924.7 - \log_{10} 466.8 - \log_{10} 299.2 \right)$$

$$= 1.306 \text{ with 2 degrees of freedom.}$$

The corrected χ^2 value is

$$\chi^2_e = \frac{1.306}{1.111} = 1.18$$

A chi-square value of 1.18 or larger with 2 degrees may be expected to be exceeded in random sampling from a homogeneous population in about 75% of the times. There is little evidence that within type variances are different, hence the pooled error mean square 563.6 was used to test the comparisons made.

The conclusion would be reached that the difference between the means of normal and offtype plants is less than ordinarily expected in random sampling from the same population. On the other hand the difference between the means of the offtype and normal plants and of the aberrants is much larger than can be attributed to chance sampling fluctuations. There is little doubt that the aberrants are much lower in dry weight of shrub than are the other types of plants.

Various other statistics may be computed from the data given. For example the standard error of a mean is

$$s_{\bar{x}} = \sqrt{\frac{563.6}{5}} = 10.62,$$

the standard error of a mean difference is,

$$s_{\bar{d}} = \sqrt{2}\ s_{\bar{x}} = 1.414(10.62) = 15.02,$$

the coefficient of variation is

$$v = \frac{s}{\bar{x}} = \frac{\sqrt{563.6}}{81.8} = 29 \text{ percent},$$

the intraclass correlation is (see Snedecor, P.245)

$$\frac{12712.8 - 563.6}{12712.8 + 4(563.6)} = .81,$$

and so forth.

In addition the experimenter may wish to compute such statistics as the t values for various comparisons. The contrast of the means of normals and offtypes is tested by

$$t = \frac{115.2 - 106.4}{\sqrt{\frac{2}{5}\left(\frac{3698.8+1867.4}{4+4}\right)}} = \frac{8.8}{16.7} = 0.53$$

The standard error of a mean difference, 16.7 is the appropriate one (see Fisher, 1942, The Design of Experiments) for comparing these means. The pooled error variance, 695.8, has 8 degrees of freedom. Therefore, a t value of 0.53 or larger with 8 degrees of freedom has a probability of occurrence greater than 50 per cent in sampling from homogeneous populations. In the event that a pooled error variance with 8 d.f. was not considered appropriate, but that each within type variance was an estimate of a different parameter the resulting t value would correspond to the tabled t values for <u>four</u> degrees of freedom (see Snedecor, P.83).

The remainder of the contrasts may be made in a similar manner. A single least or minimum significant different (lsd or msd) would be appropriate only if the variances were considered to be from the same population. The experimental evidences against heterogeneity is insignificant ($.7 < p < .8$ for [illegible] value) and one may be justified in computing

a single lsd or msd equal to

$$t_{.05}s_{\bar{d}} = 2.179\ (15.02) = 32.7,$$

since the comparisons of interest do not represent a grouping of the data after the results have been scrutinized but were made prior to the selection of the plants for dry weights. Cochran (Emp.Jour.Agric, 6:157,1938) gives an excellent discussion of the various tests of significance among a group of treatments. Fisher, (1942. Section 24, The Design of Experiments) Love (1943,P.34), and Leonard and Clark (1939,Chapter 11) are among other writers who have discussed this problem and there should be no need for repetition except that experimenters have consistently misused least significant differences, especially with regard to the highest versus the lowest. For the comparisons of fortuitous groupings of the data, the present tables of probability values have little value.

The above illustration was given to illustrate the computational procedure for completely randomized design with equal numbers of repetitions of each of the treatments. For the case of n classes with k individuals per class the following breakdown of the degrees of freedom are appropriate:

Source of variation.	Degrees of freedom.
Among n classes	n - 1
Among individuals within classes	n(k-1)
Total	nk -1

Some experimenters are not so lucky as to always obtain equal numbers for each class. If the experimenter is working with animals, some of the experimental animals may become sick or die, leaving the experimenter with unequal numbers. Likewise, in the laboratory, an assistant may unwittingly bulk items, may forget to record the data, or

may inadvertently lose some results in one way or another and the experimenter is left with unequal numbers of individuals.

The analysis for unequal numbers in a completely randomized design is little affected; the only real effect is that comparisons among treatments with fewer numbers is less precise than among treatments with larger numbers.

Example III-2.

Example III-1 represents only a part of the plants of 109 for which dry weight of shrub was obtained. From the entire area planted to variety 109, 54 plants were selected at random. Of these plants 27 were normals, 15 offtypes and 12 aberrants. The dry weight of shrub for the plants of the three types are given in Table III.1 along with the means, sums of squares and standard errors of a mean.

The analysis of variance for example III.2 is

Source of variation.	Degrees of freedom.		Sums of squares.		Mean squares.
Among types	2		67,566.7		33,783.4
N vs.O		1		2,436.8	
N +O vs 2A		1		65,129.9	
Within types	51		45,750.1		897.1
Within N		26		29,348.3	1,128.8
" O		14		13,430.9	959.4
" A		11		2,970.9	270.1
Total	53		113,316.8		

Thetotal sum of squares is obtained by squaring all individual weights and subtracting the overall sum squared divided by the total number,

$$\sum_{i=1}^{n} \sum_{j=1}^{k_i} X_{ij}^2 - \frac{X_{..}^2}{\sum_{i=1}^{n} k_i}$$

$$= 58^2 + 109^2 + \ldots + 17^2 + 65^2 - \frac{4935^2}{54}$$

$$= 564,321.0 - 451,004.2 = 113,316.8 \text{ with 53 d.f.}$$

TABLE III-1. Dry weight of shrub (without leaves) in grams.

Number	Normals	Offtypes	Aberrants	Total
1	58	103	34	
2	109	84	12	
3	87	88	20	
4	101	109	5	
5	105	134	48	
6	94	106	32	
7	167	86	21	
8	141	149	19	
9	112	64	24	
10	104	120	20	
11	58	108	17	
12	98	82	65	
13	106	112	-	
14	120	129	-	
15	65	22	-	
16	100	-	-	
17	117	-	-	
18	82	-	-	
19	133	-	-	
20	172	-	-	
21	133	-	-	
22	165	-	-	
23	150	-	-	
24	116	-	-	
25	120	-	-	
26	192	-	-	
27	117	-	-	
Total	3122	1496	317	4935
Mean	115.6	99.7	26.4	91.4
$s.s.=\Sigma X_i^2$	390344	162632	11345	564321
$(\Sigma X_i)^2/k$	360995.7	149201.1	8374.1	451004.2
Σx_i^2	29348.3	13430.9	2970.9	113316.8
$s_{\bar{x}}$	6.5	8.0	4.7	

The sum of squares among types with 2 d.f. is

$$\sum_{i=1}^{3} \frac{X_{i.}^2}{k_i} - \frac{X_{..}^2}{\sum_{i=1}^{3} k_i} = \frac{3122^2}{27} + \frac{1496^2}{15} + \frac{317^2}{12} - \frac{4935^2}{54}$$

$$= 518{,}570.9 - 451{,}004.2 = 67{,}566.7$$

The sums of squares among plant weights within normals, offtypes and aberrants are obtained from the formula

$$\sum_{j=1}^{k_i} X_{ij} - \frac{X_{i.}^2}{k_i}$$

and are given in Table III-1.

The sums of squares for the 2 orthogonal comparisons given in the analysis of variance table are

$$\frac{3122^2}{27} + \frac{1496^2}{15} - \frac{4618^2}{42} = 2436.8 \qquad \text{and}$$

$$\frac{4618^2}{42} + \frac{317^2}{12} - \frac{4935^2}{54} = 65{,}129.9$$

The standards errors of a mean (Table III-1) are computed from the formula

$$s_{\bar{x}} = \sqrt{\frac{\sum_{j=1}^{k_i} X_{ij}^2 - \frac{X_{i.}^2}{k_i}}{k_i\,(k_i - 1)}} \; .$$

The individual within type variances appear to be quite different. Bartlett's test for homogeneity results in the following chi-square value,

$$X^2 = 2.3026\,(51 \log 897.1 - 26 \log 1128.8 - 14 \log 959.4 - 11 \log 270.1)$$
$$= 6.29$$

$$X^2_c = \frac{6.29}{1 + \frac{1}{6}\left(\frac{1}{26} + \frac{1}{14} + \frac{1}{11} - \frac{1}{51}\right)} = \frac{6.29}{1.028} = 6.12,$$

with 2 d.f. The probability of obtaining a chi-square value as large or larger than 6.12 would occur about 2-5 per cent of the times in random

sampling. Hence it is concluded that the variances differ. The X^2_c value = 6.12 with 2 d.f. may be partitioned into two single d.f. with corrected chi-square values of 0.12 and 6.00 for comparisons among plant variances for normals versus offtypes and of the pooled among plant variances for normals and offtypes versus aberrants. Apparently the variation in individual plant weights is much smaller for aberrants than for the other plant types, but the variation among plant weights for normals and offtypes is approximately equal. One could then use the pooled within plant variance, 1069.5 with 26 + 14 = 40 degrees of freedom for testing the difference between the means of the normals and offtypes thus,

$$F = \frac{2436.8}{1069.5} = 2.28$$

The corresponding F value for 1 and 40 degrees of freedom at the 5 per cent level is 4.08 and we would conclude that the difference in means of normals and offtypes could be obtained fairly frequently in random sampling. The experimenter may wish to be more conservative and also may not wish to assume that the error variances in normals and offtypes are estimates of the same parameter. He may have observed that the last weight for the offtype plants, 22, was unusually low and that for the last aberrant plant, 65, was unusually high. These values tend to increase the variances of both types. A copying error was suspected but a check showed this not to be the case. A misclassification was suspected but could not be verified. In view of this then, the experimenter may wish to make a more conservative test. Such a test would be to use an F with less than 40 degrees of freedom, say 14 or mid-way between 14 and 26 degrees of freedom. An approximate significance level of t may be computed from the formula given by Cochran and Cox (1944,Experimental Designs, mimeo), and

illustrated by Snedecor (P.84, 1946),

$$t_{o5} = \frac{t_{o5}(k_1\text{-ld.f.})\ \frac{s_1^2}{k_1} + t_{o5}\ (k_2\text{-ld.f})\ \frac{s_2^2}{k_2}}{\frac{s_1^2}{k_1} + \frac{s_2^2}{k_2}}$$

$$= \frac{2.056\left(\frac{1128.8}{27}\right) + 2.145\left(\frac{959.4}{15}\right)}{\frac{1128.8}{27} + \frac{959.4}{15}}$$

= 2.110, which is equivalent to about 17 degrees of freedom. The experimental t value is

$$t = \frac{115.6 - 99.7}{\sqrt{\frac{1128.8}{27} + \frac{959.4}{15}}} = \frac{15.9}{10.28} = 1.55,$$

which is somewhat smaller than the calculated five per cent value, 2.110. In this instance the means agree sufficiently well so that the same conclusion is reached regardless of the test used. An illustration of the opposite situation is givem by Snedecor (section 4.6, 1946)

Similarly the mean difference of aberrants and offtypes may be tested by the statistic

$$t = \frac{99.7 - 26.4}{\sqrt{\frac{959.4}{15} + \frac{270.1}{12}}} = \frac{73.3}{9.30} = 7.88$$

The 5 per cent level of t for this comparison is

$$t_{o5} = \frac{2.145\left(\frac{959.4}{15}\right) + 2.201\left(\frac{270.1}{12}\right)}{\frac{959.4}{15} + \frac{270.1}{12}} = 2.160$$

and the 1 per cent level of t is

$$t_{o1} = \frac{2.977\left(\frac{959.4}{15}\right) + 3.106\left(\frac{270.1}{12}\right)}{\frac{959.4}{15} + \frac{270.1}{12}} = 3.011$$

The mean weight difference for the 2 types of plants, offtypes and aberrants, is much larger than could be logically attributed to chance sampling fluctuations.

Other comparisons forming a part of the hypothesis may be tested similarly.

The coefficient of variation for the whole experiment has little meaning, but for illustrative purposes it is,

$$v = \frac{s}{\bar{x}} = \frac{\sqrt{897.1}}{91.4} = 33 \text{ per cent.}$$

The coefficients of variations for the three types of plants, normals, offtypes, and aberrants, are

$$\frac{\sqrt{1128.8}}{115.6} = 29 \text{ per cent,}$$

$$\frac{\sqrt{959.4}}{99.7} = 31 \text{ per cent, and}$$

$$\frac{\sqrt{270.1}}{26.4} = 62 \text{ per cent.}$$

One might suspect that the means and standard deviations were related in a linear manner and then the coefficients of variations should have been approximately equal (this still may be true if there was a misclassification of the last individual, for both offtypes and aberrants, i.e. the values 22 and 65). Despite this, there appears to be a relationship between the means and variances and in order to use the generalized error with 51 degrees of freedom some transformation of the data is necessary (M.S. Bartlett has discussed this subject to some extent, see Biometrics 3: 39,1947)

The above examples illustrate the procedures and complexities that may be encountered in experimental work. Rubber percentage data were taken for the 3 types of plants (see problem III-1) on the same 54 plants. The error variances or within type variances for normals, offtypes, and aberrants are the reverse in order of magnitude. The normals seemed to be less variable than are the other types with the aberrants being the most variable. On the other hand, the variation in grams of rubber per plant (see

Problem III-2) and resin percentage is approximately equal for the three plant types.

The chief advantages of the completely randomized design are:

a) The ease of laying out the design.

b) The design allows for the maximum number of degrees of freedom for the error sum of squares.

c) Ease of analysis. A completely randomized design has the simplest analysis of all experimental designs subject to statistical analysis.

d) Unequal numbers of repetitions for the various treatments may be included without complicating the analysis in most cases.

The chief disadvantage of the design is that it is usually suited only for small numbers of treatments. When large numbers of treatments are included the material must necessarily be spread over a relatively large experimental area. This generally increases the variation among the treatment responses. For the case in which the variation over the whole of the experimental area is relatively large, it is possible to select more efficient designs than the completely randomized one. Quite frequently the treatment means are measured more precisely in the more efficient designs with fewer replicates. Completely randomized blocks are seldom, if ever, used for field lay-out of experiments, the reason being that experience has shown that other designs are much more suitable.

Problem III-1.

Fifty-four plants were selected at random from the area planted to variety 109. These are the same plants on which dry weight of shrub was obtained in Table III-1. The character rubber percentage was obtained on the individual plants. The data are:

Plant No.	Normals.	Offtypes.	Aberrants.	Total.
1	6.97	6.21	4.28	
2	7.11	5.70	7.71	
3	7.26	6.04	6.48	
4	6.80	4.47	7.71	
5	7.01	5.22	7.37	
6	7.00	5.55	7.20	
7	6.35	4.45	7.06	
8	6.37	4.84	6.40	
9	7.29	5.88	8.93	
10	7.31	5.82	5.91	
11	6.86	6.09	5.51	
12	6.81	5.59	6.36	
13	6.43	6.06		
14	7.43	5.59		
15	6.68	6.74		
16	7.29			
17	7.12			
18	6.68			
19	7.34			
20	5.15			
21	6.41			
22	6.45			
23	6.32			
24	6.82			
25	6.86			
26	6.48			
27	7.28			
Total				
Mean				

(i) Test the mean differences of normals and offtypes and of offtypes and aberrants by t-test.

(ii) Are the variances homogeneous?

(iii) Run covariance analysis of rubber-percentage (Y) on dry weight of shrub.

(iv) Does the regression of the means differ from the average within regression?

(v) Do the within-type regressions differ from the average within regression?

(vi) Is the variation among dry weight of shrubs significantly greater than that among rubber percentages for aberrants?

Problem III-2.

The following data on estimated grams of rubber per plant were obtained on the same plants of variety 109 as given in problem III-1.

Plant No.	Normals.	Offtypes.	Aberrants.	Total.
1	4.07	6.39	1.46	
2	7.73	4.77	0.89	
3	6.29	5.33	1.30	
4	6.84	4.88	0.41	
5	7.35	7.00	3.57	
6	6.57	5.90	2.34	
7	10.60	3.82	1.51	
8	8.99	7.23	1.20	
9	8.16	3.77	2.19	
10	7.58	7.00	1.17	
11	4.00	6.61	0.91	
12	6.67	4.61	4.11	
13	6.78	6.77		
14	8.91	7.23		
15	4.32	1.51		
16	7.30			
17	8.30			
18	5.47			
19	9.74			
20	8.86			
21	8.52			
22	10.62			
23	9.46			
24	7.93			
25	8.20			
26	12.47			
27	8.52			
Total				

(i) Are the within type variances homogeneous?

(ii) Under the assumption of homogeneity of variances test the mean of highest versus lowest and give level of significance at 5 and 1 per cent levels.

(iii) Compute coefficients of variation for each type and for the experiment.

Problem III-3.

The 54 plants variety 109 were analyzed for resin percentages. The data follow:

Plant No.	Normals.	Offtypes.	Aberrants.	Totals.
1	5.71	6.17	3.97	
2	6.15	6.04	6.65	
3	6.05	5.89	5.44	
4	5.64	5.91	7.20	
5	3.85	5.22	6.52	
6	5.62	5.75	6.51	
7	5.60	5.38	5.92	
8	5.00	5.99	6.81	
9	6.06	5.44	7.34	
10	6.05	5.88	5.55	
11	5.24	6.13	5.22	
12	5.66	5.83	5.95	
13	5.53	5.88		
14	6.25	6.34		
15	6.06	5.83		
16	6.10			
17	6.07			
18	7.13			
19	6.53			
20	5.83			
21	5.85			
22	5.67			
23	6.01			
24	5.64			
25	5.88			
26	5.63			
27	6.35			
Total				

(i) Do the types differ with regard to resin percentages at the end of one year's growth?

(ii) Compute a least or minimum significant difference. Does it have any meaning for these data?

(iii) Compute the coefficient of variation. Do you believe that the variation among plants was so great as to obscure differences among the types for resin percentages?

Problem III-4.

For the students' interest it is suggested that the following list of problems or examples be scrutinized for whatever value they might have in connection with the design and analysis for a completely randomized design.

G.W. Snedecor, Statistical Methods, 1946

Page	Example	
226	Example	10.5
227	"	10.8
232	Table	10.12
235	Example	10.15
235	"	10.16
236	Table	10.15
242	Example	10.19
244	Table	10.19
247	"	10.20
318	"	12.1
341	"	13.1

C.H. Goulden, Methods of Statistical Analysis, 1939

Page 125 - Example 29

W.H.Leonard and A. Clark, Field Plot Technique, 1939

Chapter 12, Table 1.

Chapter IV.

RANDOMIZED COMPLETE BLOCKS DESIGN.

by

W.T. Federer.

A randomized complete blocks design is one in which the site of the experiment is divided into a number of compact blocks, each block containing as many plots as there are treatments. The treatments are assigned at random to the plots in each block. There are slight variations in the various randomized blocks designs. In some instances, a check variety or treatment may be included more than once in each block.

Using the same example as in the design above, the five treatments A, B, C, D, and E may be included in each of the four blocks once and only once. The following diagram illustrates the experimental lay-out for the field, laboratory, or greenhouse.

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

The breakdown of the total degrees of freedom is:

Source of variation	Degrees of freedom	Mean square
Among 5 treatments	4	T
" 4 blocks	3	R
Remainder or error	12	E
Total	19	

As is apparent from the analysis, three of the degrees of freedom are segregated from the error degrees of freedom, for a completely randomized design. These three degrees of freedom are associated with the sum of squares attributable to the differences among the means of the four blocks.

If an experiment had been conducted as a randomized complete blocks design, it is possible to determine what the efficiency would have been had the experiment been conducted as a completely randomized design. The calculated variance for the latter design is obtained from the sum of squares for blocks plus the sum of squares obtained by multiplying the error mean square by the treatment plus error degrees of freedom and dividing by the total degrees of freedom. Symbollically, this is:

$$E' = \frac{(\text{treatment plus error degrees of freedom})E + (\text{Block degrees of freedom})R}{\text{treatment + error + block degrees of freedom}}$$

The efficiency of the randomized complete blocks design relative to the completely randomized design is the ratio of the amount of informations on the designs. The amount of information (Fisher, Design of Experiments 1942, and Snedecor, Statistical Methods, 1946) is defined as the reciprocal of the error variance. The efficiency of the randomized complete blocks design relative to what it would have been had a completely randomized design been used is

$$\frac{1}{E} \Big/ \frac{1}{E'} = \frac{E'}{E} \quad \text{in percent.}$$

The increase in efficiency due to the use of randomized complete blocks is equal to

$$1 - \frac{E'}{E} \quad \text{in percent.}$$

The shape of the complete block is usually as nearly square as possible, the reason for this being that the experimenter has little or no knowledge regarding the variation in perpendicular directions. Therefore, he usually selects a square or nearly square complete block and hopes that the variation is about equal in both directions. In some instances, it may be extremely undesirable to use square blocks. An example of this would be the lay-out of a randomized complete blocks experiment on contours. A single block should probably be confined to one contour, which would result in a long narrow block. The variation down one contour would probably be more nearly equal to the variation along the oontour than if the complete block were designed to include several contours. Another example of this would be in the design of a greenhouse experiment for the case where the heat source might be at one side of the experiment. Here again, the experimenter might profitably choose a long narrow block rather than a square one. Although it had been commonly advocated that only square blocks be used, the experimenter may be better off on the average if he follows the general rule to select a replicate shape that would make the variation in both directions approximately equal, and, consequently, making the variation within the whole block as small as possible. The variation among the replicates possible for the experimental site should be maximized.

The size and shape of the plots within a complete block have been discussed by a number of workers. (Love, Hutchinson and Panse, Cochran, etc.) In general it may be advocated that long narrow plots are preferable to square ones. The object in this case is to select plot sizes and shapes so that the variation among them is as small as possible.

The amount of replication required will depend upon the precision with which the experimenter wishes to measure the treatment means. He usually has some idea regarding the co-efficient of variation in the material under observation. Also he has some idea of the size of the difference between two treatments which is of practical significance. With these facts then, he may decide the approximate number of replicates to use, by choosing the number of replicates giving the desired degree of precision.

The chief advantages of the randomized complete blocks design are:

(i) Accuracy. This design has been shown to be more accurate than the previous design for most types of experimental work. The elimination of the blocks sum of squares from the error sum of squares usually results in a decrease in the error mean square.

(ii) Flexibility. The design places no restrictions on the number of treatments or on the number of replicates. In general, however, at least two replicates are required to obtain tests of significance,(see later chapters for exceptions). In addition, the standard or check treatments may be included more than once with little complication to the analysis.

(iii) Ease of analysis. The statistical analysis is simple and rapid. Moreover, the error of any treatment comparison can be isolated and any number of treatments may be omitted from the analysis without complicating it. These facilities may be useful when certain treatment differences turn out to be very large, when some treatments produce crop failures or when the experimental material is heterogeneous.

The chief disadvantage of the randomized blocks design is that it is not too suitable for large numbers of treatments, or for cases in which the complete block contains considerable variability.

Because of its advantages regarding accuracy, flexibility, and ease of analysis, the randomized complete blocks design is probably the most widely used of any design.

The advantages, computational procedure, and efficiency of randomized complete block designs are illustrated in the following examples.

Example IV-1.

Uniformity trial data on corn (Zuber, 1940) were used to construct the first numerical example illustrating statistical computations involved for data obtained from randomized complete blocks experiments. The 12 plot yields in Table IV-1 represent the yield in pounds of ear corn per 2x10 hill plot. The spacing between hills was 3.5 feet. Thus, the dimensions of an individual plot are 7x35 feet and of the complete block, 35x21 feet. The plot and block shape agree fairly well with the principles enunciated in the foregoing section. Before proceeding with the computations, it should be remembered that the 3 so-called varieties, A, B, and C, are the same thing. The comparisons among the "varieties" may be called dummy comparisons, but these are made only to illustrate the numerical procedure and some interpretations involved in the course of experimentation.

All computations necessary to obtain the analysis of variance table are given in Tables IV-1 and 2. The F values were obtained as

TABLE IV-1. Field arrangement of 3 varieties of corn in 4 randomized complete blocks. Plot sizes of 2x10 hills of corn with 3.5 feet between hills.

I			II		
1 - C	2 - B	3 - A	4 - A	5 - B	6 - C
30.6	32.0	30.3	33.0	34.4	33.1
12 - C	11 - A	10 - B	9 - A	8 - C	7 - B
29.9	31.6	32.5	31.0	29.2	29.7
IV			III		

70 ft.

42 feet

	Totals	Means
Replicate I	92.9	-
" II	100.5	-
" III	89.9	-
" IV	94.0	-
Variety A	125.9	31.48
" B	128.6	32.15
" C	122.8	30.70
Total	377.3	31.44

TABLE IV-2. Sums of squares and analysis of variance table for data in Table IV-1.

Correction term = $\frac{(377.3)^2}{12}$ = 11,862.94

Total sum of squares =

$30.6^2 + \ldots + 29.9^2 - \frac{(377.3)^2}{12}$ = 11,890.97 - 11,862.94
= 28.03 with 11 d.f.

Replicate sum of squares =

$\frac{92.9^2 + \ldots + 94.0^2}{3} - \frac{(377.3)^2}{12}$ = 11,882.89 - 11,862.94
= 19.95 with 3 d.f.

Variety sum of squares =

$\frac{125.9^2 + 128.6^2 + 122.8^2}{4} - \frac{(377.3)^2}{12}$ = 11,867.15 - 11,862.94
= 4.21 with 2 d.f.

Interaction sum of squares by subtraction =

28.03 - 19.95 - 4.21 = 3.87 with 6 d.f.

Analysis of variance of yields:

Source of variation	d.f.	s.s.	m.s.	F
Replicates	3	19.95	6.650	10.31
Varieties	2	4.21	2.105	3.26
Error	6	3.87	0.645	
Total	11			

$$F = \frac{2.105}{0.645} = 3.26 \quad (F_{05} = 5.14)$$

and

$$F = \frac{6.650}{0.645} = 10.31 \quad (F_{01} = 9.78)$$

The latter F value exceeds the tabulated F value at the one percent level of probability, for 3 degrees of freedom in the greater mean square and 6 in the lesser mean square. A more appropriate F test of the differences among whole blocks (for the case of uniformity trial data <u>only</u>) would be the replicate mean square divided by the within replicate mean square,

$$F = \frac{6.650}{\frac{4.21 + 3.87}{8}} = \frac{6.650}{1.01} = 6.58$$

where F_{05} (3 and 8 d.f.) = 4.07 and F_{01} = 7.59.
In either case, the particular layout was effective in removing variation.

The F test of the "variety" differences indicates that the variation among the 3 means are not to be considered unusually large. The probability of obtaining an F of 3.26 or larger may be <u>approximated</u> from the formula given by E. Paulson (Annals.Math. Stat. 1942). This formula is applicable for 3 or more d.f. in the error variance. Using Paulson's formula an F of 3.26 or larger occurs in 20 to 30 percent of the cases, thus

$$U = \frac{\left(1 - \frac{2}{9n_e}\right) F^{\frac{1}{3}} - \left(1 - \frac{2}{9n_g}\right)}{\sqrt{\frac{2}{9n_e} F^{2/3} + \frac{2}{9n_g}}}$$

(where n_e = error degrees of freedom and n_g = d.f. for mean square in the numerator of F)

$$= \frac{(1 - \frac{2}{9(6)})(3.26)^{1/3} - (1 - 2/9(2))}{\sqrt{\frac{2}{9(6)}(3.26)^{2/3} + \frac{2}{9(2)}}}$$

= 1.23, which may be compared with the tabulated values of t for 6 d.f. (see Fisher, Statistical Methods, Table IV, p.169, 1944).

A more accurate probability value for an experimental F may be obtained from the formula set forward by Bancroft (Annals of Math.Stat. 194?)

In general practice, non-significance among the 3 means would be the end of computations. However, the experimenter may still want to know the coefficient of variability and the magnitude of differences necessary for significance. The coefficient of variation is

$$v = \frac{s}{\bar{x}} = \frac{\sqrt{0.645}}{31.44} = \frac{0.803}{31.44} = 2.6 \text{ percent,}$$

which is low for most experimental work on corn yields with 2x10 hill plots (the average coefficient of variation for corn yield trials in Iowa is about 9-12 percent for randomized complete blocks designs. Federer, 1948).

The average least significant difference between 2 variety means is

$$\text{lsd} = t_{05}(6 \text{ d.f.})s_d = 2.447\sqrt{\frac{2(0.645)}{4}} = 1.39.$$

The least significant difference for the comparison of the highest with the lowest yielding variety in a sample of 3 (Snedecor, 1946 Table 5.5) is

$$\text{lsd} = 3.34(.568) = 1.90.$$

The standard error of a mean is

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{0.803}{\sqrt{4}} = 0.40 \quad .$$

The efficiency of this design compared to what it would have been had a completely randomized design been used is the ratio of the two variances,

$$\frac{\dfrac{(0.645)(2 + 6) + 6.650(3)}{2 + 6 + 3}}{0.645} = \frac{2.283}{0.645} = 354 \text{ percent.}$$

This means that (3.5)(4) = 14 replicates of a completely randomized design would have been required to estimate the means with the same precision as the present design, with only 4 replicates. It might be pointed out that such large gains in efficiency, 254 percent, are not generally expected.

The z test (Fisher, Statistical Methods) may be used in lieu of the F test if desired. For this example

$$Z = \frac{1}{2}\left(\log_e 2.105 - \log_e 0.645\right)$$

$$= \frac{1}{2}\left(.7444 - 9.561 + 10 \right) = .592 \ ,$$

and the z value (Table VI, Fisher, Statistical Methods) at the 5 percent point for $n_2 = 6$ and $n_1 = 2$ degrees of freedom is 0.8188. The z test agrees with the F as it should since

$$z = \frac{1}{2} \log_e F.$$

<u>Example IV-2.</u>

The preceding example illustrates the analysis for one unit per plot. In some instances several units per plot may be measured. In a feeding trial of weight groups and rations several steers

could be included in each cell or lot of the two-way classification or for a randomized blocks field experiment several observations of the same variable could be taken on each plot.

The data presented in Table IV-3 represent the grams of rubber obtained from 2 randomly selected plants in a plot for each of the 7 varieties of guayule planted in the 5 replicates. The allocation of the varieties to the seven plots in each replicate was random. The plot size was 28 plants long by 12 rows wide with 20" between plants within a row and 24" between rows resulting in a plot of $\frac{1}{12}$ [(28x20) x (12x24)] = $46^{2}/3'$ x 24'. The replicate size was 7x24' by $46^{2}/3'$ = 168' x $46^{2}/3'$. The shape of the replicate may not have been the most desirable except for the fact that the irrigation was perpendicular to the length of the replicate. In such an experiment, the best guess was to have a rectangular shaped replicate in preference to a square one, since the plots in each replicate would be treated similarly with regard to time and amount of irrigation. However, plots 6 rows wide by 56 plants long may have been the best shape in relation to replicate shape for this experiment. In instances where the plot shape is fixed the rectangular shaped replicate may prove to be the most efficient for irrigation experiments.

The sums, means, and sums of squares for the data in Table IV-3 are presented in Table IV-4. The results are summarized in Table IV-5. The mean squares are obtained from the division of the sums of squares by the appropriate degrees of freedom.

TABLE IV-3. Field arrangement of 7 varieties of guayule in 5 randomized complete blocks and weight (grams) of rubber for 2 randomly selected plants.

≠ 7-130-4.06 -3.75	6-406-6.65 -6.17	5-593-6.85 -4.94	4-109-1.46 -6.39	3-416-2.96 -2.71	2-405-2.53 -6.93	1-407-2.06 -6.12
8-109-4.07 -7.73	9-593-5.92 -5.00	10-405-1.85 -6.44	11-406-4.06 -6.65	12-416-4.35 -5.85	13-130-9.27 -6.64	14-407-5.00 -5.12
21-593-3.88 -6.22	20-407-2.59 -4.79	19-406-7.77 -6.91	18-416-2.03 -5.08	17-130-6.42 4.72	16-405-5.20 0.90	15-109-6.29 -4.77
22-130-4.43 7.31	23-109-6.84 -0.89	24-405-6.49 8.55	25-416-5.41 -0.87	26-593-6.71 6.67	27-407-6.46 -10.66	28-406-6.12 -8.21
35-593-5.82 -5.08	34-130-6.64 -5.92	33-416-0.48 1.97	32-405-7.30 -4.19	31-406-8.11 -5.95	30-109-7.35 -5.33	29-407-7.66 -5.00

5x28= 140 ft.

12x7 =84 ft.

≠ First no.= plot no.; second no.= varietal designation, and last two numbers equal weight of rubber in grams from the two plants.

TABLE IV-4. Totals of plots yields and sums of squares.

Variety	Replicate Number I	II	III	IV	V	Total.	Mean.
109	7.85	11.80	11.06	7.73	12.68	51.12	5.112
130	7.81	15.91	11.14	11.74	12.56	59.16	5.916
405	9.46	8.29	6.10	15.04	11.49	50.38	5.038
406	12.82	10.71	14.68	14.33	14.06	66.60	6.660
407	8.18	10.12	7.38	17.12	12.66	55.46	5.546
416	5.67	10.20	7.11	6.28	2.45	31.71	3.171
593	11.79	10.92	10.10	13.38	10.90	57.09	5.709
Total	63.58	77.95	67.57	85.62	76.80	371.52	5.307

Total sum of squares:

$$2.06^2 + 6.12^2 + 2.53^2 + \ldots + 5.82^2 + 5.08^2 - \frac{(371.52)^2}{70} = 2287.489 - 1971.816$$
$$= 315.673$$

Sum of squares for replicates

$$\frac{63.58^2 + \ldots + 76.80^2}{14} - \frac{(371.52)^2}{70} = 1993.811 - 1971.816 = 21.995$$

Sum of squares for varieties:

$$\frac{51.12^2 + \ldots + 57.09^2}{10} - \frac{(371.52)^2}{70} = 2042.747 - 1971.816 = 70.931$$

Sum of squares of plot totals:

$$\frac{7.85^2 + 11.80^2 + \ldots + 10.90^2}{2} - \frac{(371.52)^2}{70} = 2148.102 - 1971.816 = 176.286$$

Sum of squares for interaction of replicates and varieties by subtraction =

$176.286 - 70.931 - 21.995 = 83.360$ with 24 d.f.

Within plot sum of squares either by subtraction or by

$$\frac{(6.12 - 2.06)^2}{2} + \frac{(6.93 - 2.53)^2}{2} + \ldots + \frac{(5.82 - 5.08)^2}{2}$$

$$= 6.12^2 + 2.06^2 - \frac{(6.12 + 2.06)^2}{2} + 6.93^2 + 2.53^2 - \frac{(6.93 + 2.53)^2}{2}$$

$$+ \ldots + 5.82^2 + 5.08^2 - \frac{(10.90)^2}{2}$$

$$= 2287.489 - 2148.102 = 139.387.$$

Table IV-5. Analysis of Variance for the Data of Table IV-3.

Source of variation	d.f.	s.s.	m.s.	F
Replicates	4	21.995	5.4988	1.58
Varieties	6	70.931	11.8218	3.40
Experimental error	24	83.360	3.4733	
Sampling error	35	139.387	3.9825	
Total	69	315.673	-	

In the preceding table, two errors are listed, experimental and sampling. The experimenter may often be in a quandary as to which one to use. The answer depends upon which hypothesis he desires to test. If the worker wishes to confine his remarks to the particular 5 replicates used above, then the sampling error should be used for testing the variation among variety means. If on the other hand, the experimenter is not so conservative and wishes to make an inference about the true differences among the 7 varieties from the sample of 5 replicates, then the experimental error should be used. The last cited instance is the one of practical importance in most cases.

The sampling error for the data in Table IV-3 is larger than the experimental error but not significantly so. If the variation of plot means from plot to plot after removing replicate and variety effect is zero in the population then it would be expected that the experimental error would be smaller in about 50 percent and larger in 50 percent of the samples. If the latter error is significantly smaller than the sampling error, it would be concluded that a significant negative intraclass correlation (Snedecor and Fisher) existed. The explanation would depend upon the particular type of biological material involved.

Even though the experimental error is smaller, it is the best estimate of the error term for testing the significance of the difference among treatment means. The experimenter may wish to be more conservative and use the sampling error and the degrees of freedom associated with the experimental error. Other schemes could be followed, but the most logical one is to use the experimental error as the estimate of error variation in making various tests of hypotheses.

The F test of the difference among the 7 treatment means is

$$F = \frac{11.8218}{3.4733} = 3.40.$$

For 24 and 6 degrees of freedom the F values at the 5 and 1 percent points are 2.51 and 3.67, respectively. An F value as large as or larger than the experimental F value, has a probability of occurrence equal 2 to 3 percent.

The next question of importance would be to determine which, if any, of the 7 presumably unrelated varieties are significantly different with respect to yield of rubber at the end of one growing season. The answer to this question has only recently been supplied. Duncan (1947, Ph.D.Thesis, Iowa State College) obtained the test of significance for any pair among the 3 or 4 items in the test. Tukey (Biometrics, In press), using a different approach, considered the problem by segregating the k means into homogeneous subgroups. This test probably is more conservative than the significance level indicates, but it is easy to use and makes uses of published probability tables.

In making use of Tukey's ingenious method to determine which varieties differ significantly, the various steps are illustrated with the data of the present example.

Step 1. Choose a significance level.

The 5 percent level is chosen for this example.

Step 2. Calculate the difference which would have been significant if there were but two varieties.

This is equivalent to computing a least or minimum significant difference. The standard error of a variety mean is

$$s_{\bar{x}} = \sqrt{\frac{3.4733}{10}} = 0.589.$$

Therefore,

$$\text{lsd} = t_{05}\ (24\text{d.f.})\ \sqrt{2}\ s_{\bar{x}} = 2.064(1.414)(0.589)$$
$$= 1.719.$$

Step 3. Arrange the means in order and if any two adjacent means deviate by more than the lsd, consider them as subgroup endpoints.

The seven means arranged in order and the difference between 2 adjacent means are:

Variety	Mean Yield	Difference
406	6.660	
130	5.916	0.744
593	5.709	0.207
407	5.546	0.163
109	5.112	0.434
405	5.038	0.074
416	3.171	1.867
Average	5.307	

The mean of variety 416 is more than one lsd lower than the next adjacent mean, variety 405. Therefore at the end of step 3, the 7 varieties are divided into 2 groups, one containing 6 varieties and the other one variety. If no group contains more than 2 means, the process terminates.

Step 4. In each group of 3 or more means find the grand mean, the most divergent mean, and the difference of these 2 divided by $s_{\bar{x}}$. Convert these ratios into approximate unit normal deviates by finding

$$\frac{\frac{\bar{x}_t-\bar{x}_d}{s_{\bar{x}}} - \frac{6}{5}\log_{10} k}{3\left(\frac{1}{4}+\frac{1}{n_e}\right)} \quad (k > 3 \text{ means in a group}).$$

or

$$\frac{\frac{\bar{x}_t-\bar{x}_d}{s_{\bar{x}}} - \frac{1}{2}}{3\left(\frac{1}{4}+\frac{1}{n_e}\right)} \quad (3 \text{ means in a group}).$$

Separate off any straggling or divergent mean for which this is significant at the chosen two-sided significance level for the normal.

For the group of 6=k means the average is $\bar{x}_t$=5.664 and the most divergent mean from the general mean is that for variety 406, $\bar{x}_d$= 6.660. The error degrees of freedom (Table IV-5) is n_e=24. Substituting these values in the formula,

$$\frac{\frac{|\bar{x}_t-\bar{x}_d|}{s_{\bar{x}}} - \frac{6}{5}\log k}{3\left(\frac{1}{4}+\frac{1}{n_e}\right)} = \frac{\frac{6.660-5.664}{0.589} - \frac{6}{5}\log 6}{3\left(\frac{1}{4}+\frac{1}{24}\right)}$$

$$= \frac{8}{7}(1.691 - 0.934) = 0.865.$$

A normal deviate of 0.86 or larger is expected to occur about 39 percent of the time in random sampling (Table I, Fisher, Statistical Methods).

Thus this step has produced no further subdivision of the group of 6 means. If such had been the case the

above process would have been continued until no further subdivision into subgroups was possible.

Step 5. Calculate the sum of squares of deviations from the group mean, and the corresponding mean square for each group or subgroup of 3 or more resulting from step 4. Using $s_{\bar{x}}^2$ as the denominator, calculate the variance ratio and apply the F test.

The sum of squares of the deviations of the 6 means from their average mean is

$$6.660^2 + 5.916^2 + \ldots + 5.038^2 - 6(5.6635)^2$$

$$= 194.2194 - 192.4514 = 1.7680.$$

The mean square is 0.3536 and the F is

$$F = \frac{0.3536}{3.4733/10} = 1.02.$$

The nonsignificant F indicates no overall evidence of difference in yield for the 6 varieties.

Thus, the method of Tukey indicates that variety 416 was significantly lower than the others in yield of rubber and that the variation in yield among the remaining 6 varieties was no larger than might logically be ascribed to chance.

The above method is applicable to a group of unrelated varieties or treatments and was carried through to illustrate the procedure. However, considerable information concerning the relationships of the 7 varieties was available. Variety 109 was the only $54^{\pm}$ chromosome variety in the group, the remaining were in the $72^{\pm}$ category. A logical comparison would be the mean of the 72's versus the mean of the 54 chromosome variety,

$$\frac{[6(51.12) - 59.16 - 50.38 - \ldots - 57.09]^2}{10[36 + 1 + 1 + 1 + 1 + 1 + 1]} =$$

$$\frac{51.12^2}{10} + \frac{[59.16 + \ldots + 57.09]^2}{60} - \frac{371.52^2}{70} = 0.4456.$$

Also, it was known that varieties 406 and 130 were selections from 593. The 2 degrees of freedom among these three means could logically be partitioned into 2 single degrees of freedom representing the comparison of the two selections with the parent variety and the comparison between the selections.

<u>Among 130, 406, and 593:</u>

$$\frac{59.16^2 + 66.60^2 + 57.09^2}{10} - \frac{182.85^2}{30} = 5.0026.$$

<u>130 + 406 versus 593:</u>

$$\frac{[59.16 + 66.60 - 2(57.09)]^2}{10[1+1+4]} = \frac{57.09^2}{10} + \frac{(59.16 + 66.60)^2}{20}$$

$$- \frac{[59.16 + 66.60 + 57.09]^2}{30} = 2.2349.$$

<u>130 versus 406:</u>

$$\frac{(59.16 - 66.60)^2}{10(1+1)} = 2.7677.$$

Furthermore varieties 130, 406, and 593 are phenotypically different from the remaining 3 varieties, 405, 407, and 416. The former have round greenish leaves and short branching habit while the latter group have long serrated grayish green leaves and longer branches. A logical comparison would be between the 2 groups of means,

$$\frac{[59.16 + 66.60 + 57.09 - 50.38 - 55.46 - 31.71]^2}{10(1 + 1 + 1 + 1 + 1 + 1)} = 34.2015.$$

The remaining 2 degrees of freedom make up the comparisons among the 3 varieties 405, 407, and 416, with the following sum of squares,

$$\frac{50.38^2 + 55.46^2 + 31.71^2}{10} - \frac{137.55^2}{30} = 31.2813.$$

It was not known what relationship existed among the 3 varieties and without further information, the partitioning of the variety sum of squares is finished. Tukey's test may be applied to these 3 means and it was found that 2 subgroups are formed consisting of 405 and 407 in one group and 416 in the other.

The sums of squares are summarized in Table IV-6 and as a partial check they should add up to the total, 70.931.

TABLE IV-6. Partitioning of the treatment sum of squares from Table IV-5.

Source of variation.	d.f.	Sum of squares.	Mean square.	F
Varieties	6	70.931	11.8218	3.40
109 vs.others	1	0.4456	0.4456	-
130+406 vs. 593	1	2.2349	2.2349	-
130 vs.406	1	2.7677	2.7677	-
130,406,593 vs. 405,407,416	1	34.2015	34.2015	9.85
Among 405,407,416	2	31.2813	15.6406	4.50
Exp. error	24	83.360	3.4733	

F= 9.85 exceeds the tabulated F at the one percent point and F=4.50 exceeds the F value at the 5 percent point. The mean of the 3 varieties, 130, 406, and 593 and of the 3, 405,407,and 416 cannot be considered as coming from the same general population. However, upon examination of the latter 3 varieties, it was found that they did not represent a homogeneous group and that the very low yield of variety 416 accounts for the large F values in both instances.

The amount of variability relative to the mean in this experiment was much higher than desired. The coefficient of variation is

$$\frac{\sqrt{3.4733/2}}{5.307} = \frac{1.318}{5.307} = 25 \text{ percent.}$$

The standard deviation <u>per plant mean yield</u> is $\sqrt{3.4733/2}$ or it is the standard deviation resulting from an analysis of the plot means. The verification that division of the error mean square by 2 (equals number of items from each plot) results in the same value as that obtained from using the <u>plot means</u> in the analysis, is left as an exercise for the student.

The efficiency of this design relative to what it would have been had a completely randomized design been used is

$$\text{Eff.} = \frac{\dfrac{21.995 + 3.4733(6+24)}{4 + 6 + 24}}{3.4733} = \frac{3.7116}{3.4733} = 107 \text{ percent,}$$

or a gain in efficiency of 7 percent.

Problem IV-1.

For example IV-1, make the assumption that varieties B and C are two entries of the same variety, say D. Complete the following analysis of variance table:

Source of variation	d.f.
Replicates	3
D vs. A	1
B vs. C	1)
Remainder	6)
Total	11

Under what conditions would the pooled error with 7 d.f. be used? Compute the standard error for testing the difference between the means of varieties A and D and the least or minimum significant difference.

Problem IV-2.

Compute the coefficient of variation, the standard error of a mean, and standard error of a difference for a mean on a plot total basis for example IV-2. If the means were on a basis of the plot total of 2 plants, would the efficiency of the design be changed? Explain.

Problem IV-3.

Use Tukey's method for testing the significance of the means in Tables 10.3 and 11.9 and example 11.28, Snedecor, Statistical Methods, 1946. Assume no relationship among the entries in the test. Why?

Problem IV-4.

Obtain the expected values for the example IV-1 from the formula

$$\hat{X}_{ij} = \text{experimental mean} + \text{variety effect} + \text{replicate effect}$$

$$= \bar{x} + v_i + r_j$$

$$= \text{variety mean} + \text{replicate mean} - \text{experiment mean.}$$

and compute the following sums of squares

$$\sum_{j=1}^{4} (X_{1j}-\hat{X}_{1j})^2 = \Sigma e_{1j}^2$$

$$\sum_{j=1}^{4} (X_{2j}-\hat{X}_{2j})^2 = \Sigma e_{2j}^2$$

and

$$\sum_{j=1}^{4} (X_{3j}-\hat{X}_{3j})^2 = \Sigma e_{3j}^2 .$$

What conclusions would you draw from the computations made?

Problem IV-4.

The following data on resin percentage and shrub weight (grams) of plants were obtained for the varieties and replicates of example IV-2. Do the varieties differ with regard to resin percentage after being corrected to a constant shrub weight? (See De Lury, Biometrics, Sept. 1948). Would you suggest a transformation for these percentages? Why or why not?

TABLE 15.10

Yields of Wheat in a 2^3 Factorial Experiment laid down in a Latin Square. Nitrogen (n), Phosphorous (p), and Potassium (k) were tried at Two Levels, None and Some. (1) Indicates No Fertilizer.

p	n	np	k	nk	(1)	npk	pk
18.8	12.2	18.3	15.8	11.4	11.5	19.4	18.9
n	nk	pk	npk	p	k	np	(1)
12.9	7.3	17.4	17.2	19.7	12.0	19.0	15.6
nk	np	n	p	(1)	npk	pk	k
10.7	17.5	10.4	18.0	9.8	16.6	17.5	14.3
pk	k	npk	(1)	n	np	p	nk
18.3	12.6	14.2	12.2	11.4	14.5	16.9	16.1
np	(1)	nk	n	pk	p	k	npk
17.9	12.8	13.3	11.3	16.5	15.6	10.9	16.7
k	pk	(1)	np	npk	n	nk	p
14.9	18.2	12.8	17.1	15.8	9.5	8.9	20.6
npk	p	k	pk	np	nk	(1)	n
19.0	18.9	11.2	17.1	17.9	8.6	10.2	14.5
(1)	npk	p	nk	k	pk	n	np
17.5	20.4	20.8	16.4	16.8	18.5	13.6	23.0

Treatment Sums.

(1)	102.4	p	149.3	np	145.2	pk	142.4
n	95.8	k	108.5	nk	92.7	npk	139.3

Preliminary Analysis of Variance.

Source of Variation	Degrees of Freedom.	Sum of Squares.	Mean Square.
Treatments	7	513.79	73.4
Error	42		2.190

TABLE 15.10b.

	d.f.	s.s.	m.s.	F	Effects
Treatments	7	513.789			
N	1	13.690		6.25	-29.6
P	1	488.410		223.02	176.8
K	1	1.501)			-9.8
NP	1	3.610)			15.2
NK	1	1.051)	2.338	1.07	-8.2
PK	1	3.901)			-15.8
NPK	1	1.626)			10.2
Error	42		2.190		

F_{05} (42, 1 d.f) = 4.07

F_{01} (42, 1 d.f.) = 7.27

TABLE 15.11a

Effects of Nitrogen in Wheat Experiment

Treatment	Sum of Yields	N	NP	NPK
(1)	102.4	-6.6	2.5	10.2
n	95.8			
p	149.3	-4.1		
np	145.2			
k	108.5	-15.8	12.7	
nk	92.7			
pk	142.4	-3.1		
npk	139.3			
Total		-29.6	15.2	10.2

TABLE 15.11b

		K	NK	Treatments		P	Kp
(1)	102.4	6.1	-9.2	(1)	102.4	46.9	
k	108.5			p	149.3		-13.0
n	95.8	-3.1		k	108.5	33.9	
nk	92.7			pk	142.4		
p	149.3	-6.9	1.0	n	95.8	49.4	
pk	142.4			np	145.2		- 2.8
np	145.2	-5.9		nk	92.7	46.6	
npk	139.3			npk	139.3		
		-9.8	-8.2			176.8	-15.8

QUERY 1.

The following design of 3 treatments, check, A and B was used.

A	ch	A	ch
ch	B	ch	B
A	ch	A	ch
ch	B	ch	B
A	ch	A	ch
ch	B	ch	B
A	ch	A	ch
ch	B	ch	B

The treatments were two salt sprays to control weeds in beets and the check was ordinary cultivation. It is desired to determine if the salt sprays result in a decrease in yield.

ANSWER.

The experimenter used the above design. It was suspected that treatments A and B would give a lower yield than the check. The opposite result was obtained, that is, the yields were increased. Since the arrangement is entirely systematic, no correct measure of variation is possible. The experimenter is at a loss then to determine if the yields for treatments A and B were significantly higher than the yield of the check. The only way he can find this out is to re-run the experiment in a design that is subject to statistical analysis.

The experimenter who designed this experiment was fortunate in another way. The experiment was conducted at another location. The interaction of treatments and locations may be used as an estimate of the error for testing the differences between treatments. The experimenter could have gained considerably more information by using several latin squares or randomized blocks designs rather than the systematic arrangement used. The soil gradient in this experiment was such that the check yields on one end of the experiment were different from those on the other end, the yield being progressively smaller.

QUERY 2.

In a storage experiment with 1-year-old cherry trees, there are eight treatments of 80 trees each. Because of the size and shape of the plot of ground where they are to be planted, they will be planted in 32 rows of 20 trees each.

Measurements and observations to be made are the number surviving, the length of shoot growth the first year, and the diameter increase the first and second year. The trees will probably be removed after the second year.

What kind of design should be used in planting?

ANSWER.

In any experiment of this nature the storage treatments should also be replicated in the place of storage. If four replicates are to be used, then it would be wise to divide the 80 trees for each treatment into four lots of 20 and place them in four positions in the storage place. If it is possible all eight treatments should be put into one part of the storage place and the second eight treatments in another part of the storage place, and so forth. In this way, the effects of different places of storage will be confounded with the differences among the replicates in the field. If more replicates in the field are to be used then it will probably be advisable to increase the number of replicates in the storage place accordingly.

Of course, the most efficient design would be to use one tree per plot and thus replicate the treatments 80 times. Usually this is considered impractical. Therefore it is suggested that the experimenter use four or eight replicates, depending upon the precision with which he wishes to measure the treatment differences, and to use plots of either one row by 20 trees or one row by 10 trees. The randomized complete blocks design of either four or eight replicates will probably be quite suitable, since the area of the experimental site is small. The complete block should be such that the variation in both directions is approximately equal. Since the area is relatively small, the best shape for the complete block would probably be approximately a square.

QUERY 3.

In evaluating the results of an experiment by the analysis of variance, is it permissible to exclude from the averages of each replicate of each treatment, the data from individual count plants which are obviously extremes because of extraneous influencing factors?

For example: Experiment designed to measure the effect of certain sprays on yield per vine.

Yield data (pounds of fruit per vine) from grape test in which all count vines were "control-pruned." There were several treatments included in the experiment which was planted in 4 randomized complete blocks. Yield data were obtained on 8 vines per treatment per replicate. The yield data for treatment one follows:

Treatment I Bordeaux 8-8-100 (3 applications)

Count vine No.	1	2	3	4	5	6	7	8	1-8	Averages 9/25
Rep. A	15	17	28	15	14	16	2	17	15.5	15.7
" B	14	18	12	27	16	15	30	13	18.1	14,7
" C	15	15	16	18	18	14	12	10	14.8	14.8
" D	18	17	2	15	1	14	13	14	11.8	15.2
									15.1	15.1

The averages in the last column were obtained by excluding the yields of the vines which were less than 9 or more than 25 pounds.

The extremely high yields from A3, B4, and B7 and the extremely low yields from A7, D3, and D5 are obviously not the effect of the treatment which is being evaluated. These extremes are probably caused by incorrect pruning, mortality, or mechanical severance of a portion of the vine or some other incidental factor.

In evaluating the data, is it permissible to arbitrarily exclude from the replicate averages of all treatments, the data from those count vines which yielded more than 25 pounds and those which yielded less than 10 pounds?

If not, is there a statistical device by which the extremes to be excluded could be determined?

ANSWER:

In discarding data from an experiment one should proceed with caution. If the experimenter is able to explain the excessively low or high yields as due to a specific cause, then the data may be omitted from the averages. If he truly knows that a particular mishap made one or more vines deviate considerably from the remainder, then it would be only a matter of common sense to exclude these.

Frequently, the reason for extreme deviations is not known with any degree of certainty. In this event, it is suggested that the so-called extreme deviates be retained in computing averages and the resulting analysis of variance. The researcher may have to contend with the fact that grape vine yields are apt to be quite variable and that unless the differences in treatment yields are extreme it would be necessary to use more replicates and perhaps more vines per replicate.

It may be possible to obtain a measure of the amount of mechanical injury and the severity of pruning. By the statistical technique known as covariance, it would be possible to adjust vine yields to a common intensity of pruning and mechanical injury. Notes on the last two characters could be taken prior to harvest. Also, it may be possible to take notes on pruning intensity prior to the random selection of vines for yield. The vines which were too severely

QUERY 3 cont'd.

pruned or which were not pruned at all could be excluded from the experiment prior to the application of the treatments. This would allow the experimenter to use more homogeneous material and hence obtain less variable results. Any exclusion should be made prior to harvest. After the yields have been obtained, or even after the grapes have set on the vines, the experimenter should proceed with caution in excluding certain grape vines from the experiment.

For further information on this problem the writer would have to know more about the grape vines under consideration. However it is suggested that the questioner write Professor G.W. Snedecor, Iowa State College, Ames, Iowa, for further information on the general subject and with regard to the last question.

QUERY 4.

What is the rule for rounding figures in a table to be presented in a journal paper, and for the calculation of various statistics?

ANSWER.

There is no set rule for the rounding of figures to a specified number of significant digits. However, one may consult such references as G.W. Snedecor, Statistical Methods, section 5.5, 1946, and F. Yates, The Design and Analysis of Factorial Experiments, page 91, Imperial Bureau of Soil Sc., Tech.Comm. No.35, 1937, in formulating rules of procedure to follow.

In general two or three significant figures for the original observations are all that need to be retained if the coefficient of variation (c.v.) is over say seven per cent, three if the c.v. is not less than 3-4 per cent, and more than 3 figures if the observations are measured so accurately that the c.v. is less than 3 per cent. Also, the mean should usually be carried to one more figure than the unit of observation, although for large samples more significant figures should be included. The standard error of a mean should be of the same order of magnitude as the mean, i.e. if the mean is carried to 2 decimal places then so should the standard error of a mean.

Snedecor states that although rules could be devised for carrying out certain computations, "they would have to be discarded promptly when a whole series of calculations" are made.

The usual rule for rounding figures is to round to the next even number. For example, 1.35 would be rounded to 1.4 and 1.45 to the same figure, 1.4.

QUERY 5.

In trueness to type studies made by certain organizations, it is desirable to test seed stocks for genetic variation and mixtures. Often the samples vary from 50 to 200 or more seeds.

Is there any basis for determining how many seeds (assuming almost complete germination) to include of any one seed source or stock, in order to obtain a representative sampling of the population involved?

ANSWER.

The answer to such a query depends upon several things. First of all, what percentage of mixtures or plants of different types is allowable, .1%, 1%, 5%, 10%? Also would there be interest in ascertaining genetic variation in a character affected by 1, 2, or 3 factors and segregating in 3:1, 15:1 or 63:1 ratios?

In the most practical aspects of this query, the organization will presumably be interested only in knowing if the seed stock might be considered relatively true to type with less than 5 to 10 per cent of offtypes, regardless of nature. In this case set p = fraction of defectives allowable and q =1-p = fraction of non-defectives. Then the term, q^n from the binomial expansion

$$(p + q)^n,$$

gives the probability of selecting a sample with only non-defectives and n may be determined large enough so that one might be relatively certain of including some of the mixtures if such exists in the sample.

For example, if 10 percent of defectives or offtypes are allowable, p =0.1 and q = 1-p = 0.9. In a random sample of 50 seeds the probability of obtaining no defectives is $(0.9)^{50}$ = .0052 and in a sample of 40 seeds the probability of all true or non-defectives is $(0.9)^{40}$ = .0148. Thus in either a sample of 40 or 50 seeds, it is quite unlikely that a random sample of seeds would contain all nondefectives if there actually were 10% defectives in the population.

The decision as to whether or not the sample is representative will be determined by the method of selecting the original lot of 50 to 200 seeds or more. For example, suppose that ten 100-pound sacks of seed represent the seed source. Five of the sacks may contain mixtures and the other five may be relatively pure. If the method of taking the original sample is to select a 300 seed sample from one of the sacks at random, then the probability of choosing a relatively pure sack is one-half. Thus the method of choosing the seed sample needs to be such as to give a representative seed sample.

QUERY 6.

In the conduct of germination experiments on 4 lots of 100 seeds each, it has been found that in some instances 3 of the 4 lots are within the usual range or latitude but the fourth is outside. Should one repeat the series or discard the low one?

ANSWER.

In germination experiments the experimenter will usually have some idea of the variation expected for a series of 4 germination tests on a lot of seed. His conceived idea of range is something of an average, probably the mode. The experimental material will yield ranges that vary around this average latitude. It is quite doubtful that the questioner would wish to discard a series that was too uniform, so why should he wish to discard a series that is a little more variable than average unless he has definite knowledge that some external condition or event caused one germination test to be lower? In the event that some external event has caused the low germination then it would be wise to exclude the observation. In case of doubt repeat the series and average the results.

The idea of discarding "divergent" samples probably dates back to the techniques used in some college laboratory courses. The so-called divergent observation may be the correct one or it may be the one that contributes the correct amount to obtain a very close estimate of the true germination percentage in the seed stock being tested. In any event the average of the 4 seed germination tests (under comparable conditions) yields an unbiased estimate of the true germination percentage.

In some cases the 4 seed lots are germinated at the same time and location. It may be well to point out that as far as germination conditions are concerned there is only one replication. Also, if the seed sample has been taken as one single sample then there is also one replication on seed sampling regardless of how many subsamples are taken from the single sample.

QUERY 7.

In view of the apparent differences in growth and biochemical processes between varieties and between species of plants, what one variety could be chosen to make studies on photosynthesis, hormone action, etc? Is there any solution to the problem always encountered that results obtained with one variety or one species cannot be applied to other varieties and species? Do you think that there is any possibility of plant researchers concentrating on one type, such as has been done in animal work with the white rat?

ANSWER:

The answer to the various questions may be partially answered by considering an even more extreme case than that cited in the last question. The literature on plant reaction contains results obtained on a single plant or on plants from a single variety; then these results were used by the author to generalize to the species or even to the whole plant kingdom. Naturally there would be many contradictions in the literature on a particular subject, due to variation among plants within a variety and/or environment. Some experimenters apparently did not realize that even though the selected variety might be relatively homozygous for agronomic or other characters, it may be extremely variable for the character being studied and that the results obtained on one plant may be quite different from those obtained on another plant. If the variety or species were uniform with respect to the studied character, which also did not react with the environment, then the results from a single plant may be sufficient for generalizing to the species. This case would be extremely rare. The experimenter almost always must contend with the phenomenon of variation.

In studying respiration or the like for a species it may be well to select a few (as many as practical) phenotypically different varieties or subspecies and to select a relatively large (50 to 200) number of plants per variety. One could study then the relative within and between variety variation. The study could be extended to several species. Then, if a general law exists for all species the experimenter would have indications of such. The results of the experiment would indicate if the species or even varieties varied with regard to the phenomenon under consideration. The sample of plants would be more diverse in nature resulting in a considerably more substantial basis for discussing such items as respiration, photosynthesis, etc.

In designing an experiment for a study like the one outlined above, it would be well to consult a statistician relative to the proper experimental design.

QUERY 8.

A randomized blocks design of 24 treatments in 3 replicates is to be used. Since the experiment (with the same treatments) is to be conducted at different locations up to several miles away, or in different years at the same location, but on different parts of the experimental area,

(i) is it necessary to rerandomize for each location or year?
(ii) is it more advisable?
(iii) why?

ANSWER.

It may not be absolutely necessary, (but it is advisable) to rerandomize the experiment each time if the heterogeneity at the different places is such as to produce a random effect on the treatment yield and if the treatment yield is not adversely or beneficially affected by neighboring treatments. Since the experimenter usually has little or no knowledge concerning the two conditions of the preceding sentence, he would be better off on the average to use a new randomization of the design for each repetition of the experiment. The results of a group of experiments on the same set of treatments are more likely to be unbiased if a new randomization is used each time. Alternatively, the experimenter could use the same randomization (unless the experiment was to be repeated a large number of times) <u>if</u> the treatment numbers 1 to 24 were reassigned a new treatment number at random each time the experiment was repeated; this might prove more troublesome than to rerandomize the original design each time.

Furthermore, it may be desirable to pool the results of several experiments (whether over several years or for different locations) in which case different randomizations should be used for the experiment at the different places in order to satisfy the theoretical conditions required for combining the results of random samples (see G. W. Snedecor, Statistical Methods, 1946).

QUERY 9.

Would you comment on the factorial experiment for fertilizer treatments using individual tree plots and on the analysis of data with regard to

(i) flaws in such experiments
(ii) likelihood of errors in the field
(iii) selection of experimental trees in a mature orchard for an experiment to be conducted over a four year period.

ANSWER:

The advantages and disadvantages of factorial experiments have been considered in detail by F.Yates, The Design and Analysis of Factorial Experiments, Imperial Bureau of Soil Science, Technical Communication No.35, 1937.

The particular point on individual tree plots has not been covered in the reference cited and therefore will be discussed herein. The discussion refers to all types of experiments including the factorial.

The genetic variation among trees in an orchard may be large despite the fact that they may be asexually propagated. In some cases the roots may be from a different strain than the tops and although the tops represent cuttings from the same tree the root stocks may have been obtained from seedlings. The point is that the various trees in an orchard may be variable due to genetic causes. The effect of the genetic variation (and also environmental variation) on a treatment mean is inversely proportional to the number of experimental trees used. Thus, if only one tree per plot were grown, the whole of the genetic variation would be included with the differences among plots (see Powers, L., Genetics 27: 561, 1942). The experimenter on oats, corn, etc. seldom if ever uses individual plant plots in assessing treatment differences. Even though these crops are much less expensive per plant, the more plants are used, in order to lessen the effects of genetic diversity on treatment responses.

Single tree plots are undesirable for another reason , i.e., trees are apt to die, resulting in missing plots. Also, variation may be increased due to severe weather or disease conditions. Considering these effects, which are beyond the control of the experimenter, it is suggested that single treeplots be avoided if possible.

Single tree plots are undesirable for another reason. In fertilizer trials, guard or border trees are usually necessary. The ratio of experimental to guard trees is the smallest for single tree plots; this is usually undesirable.

In a factorial experiment all trees within a replicate are utilized in obtaining a treatment effect. The net result is that several trees enter into each effect even though there are single tree plots. For example, suppose that $16 = 2^4$ factorial treatments are included in 6 randomized blocks with single tree plots. In each replicate, 16 trees are used in obtaining a treatment effect and in the experiment 96 trees enter into the mean treatment effect.

For a factorial experiment the effect of genetic or environmental diversity is minimized and single tree plots are not so undesirable because of this fact. However, the chance of losses or injury and the unfavorable ratio of experimental to guard trees are features making single tree plots undesirable even in factorial experiments.

QUERY 9 cont'd.

In the selection of trees from a mature orchard one should obviously not include trees which have been injured in one way or another unless it is desired to make recommendations for all trees both injured and not injured. The experimenter should have a representative sample of the population about which he wishes to obtain estimates. The sample may be homogeneous or heterogeneous depending upon the parent population postulated by the experimenter. Usually it is desired to make recommendations regarding certain fertilizer treatments on, say, apple trees. Recommendations may be made for a variety or for a species depending upon the scope of the investigations and the magnitude of the variety by fertilizer interaction. The experimental trees should be representative of the apple tree population. If the material is made too homogeneous, the sample may not be representative of the population. The experimental material should be made as homogeneous as possible without sacrificing representativeness.

QUERY 10.

The following analyses of variance are from 3 locations. The treatments were the same at all 3 places.

	Degrees of freedom.		
	Location I	Location II	Location III
Treatments	9	9	9
Replicates	3	4	5
Error	27	36	45
Total	39	49	59

What is the method of computation and breakdown of degrees of freedom for the combined analysis? How does one compute a least significant difference for place or location means, i.e. what is the n in the formula

$$\sqrt{\frac{2}{n}\text{ error mean square}}?$$

ANSWER.

The following is the breakdown of the total degrees of freedom for the combined analysis of variance:

Source of variation	Degrees of freedom
Total	149
Locations or places	2
Replicates within locations	12
Treatments	9
Treatments by locations	18
Treatments by replicates within locations	108

The total sum of squares is obtained by squaring all the items for the 3 places and subtracting off the grand total squared divided by 150 = the number of items, or symbolically,

$$\sum_{j=1}^{10}\sum_{i=1}^{3}\sum_{k=1}^{r_i} X_{ijk}^2 - \frac{(\Sigma\Sigma\Sigma X_{ijk} = X_{...})^2}{150}$$

Where j =1,2,...,10 = treatment subscript, i = 1,2,or 3 = the place subscript, and k_1= 1,2,3,4 at place I, k_2= 1,2,3,4,5 at place II and k_3= 1,2,3,4,5,6 at place III.

The location sum of squares is obtained as the sum of the correction terms for the individual experiments or the location totals squared divided by the number of items minus the overall correction term,

$$\frac{\left(\sum_{j=1}^{10}\sum_{k=1}^{4} X_{1jk}\right)^2}{40} + \frac{\left(\sum_{j=1}^{10}\sum_{k=1}^{5} X_{2jk}\right)^2}{50} + \frac{\left(\sum_{j=1}^{10}\sum_{k=1}^{6} X_{3jk}\right)^2}{60} - \frac{X_{...}^2}{150}.$$

The replicates within location sum of squares is obtained by adding the replicate sum of squares from the 3 individual analyses. This sum of squares will have 3 + 4 + 5 = 12 degrees of freedom.

The treatment sum of squares is obtained from the sum of the squares from the 10 treatment totals divided by 15 minus the overall correction term,

QUERY 10 contd.

$$\frac{\left(\sum_{i=1}^{3}\sum_{k=1}^{r_i} X_{i1k}\right)^2 + \left(\sum_{i=1}^{3}\sum_{k=1}^{r_i} X_{i2k}\right)^2 + \ldots + \left(\sum_{i=1}^{3}\sum_{k=1}^{r_i} X_{i10k}\right)^2}{15} - \frac{X_{\ldots}^2}{150}$$

The treatment by location sum of squares (with 18 degrees of freedom) is obtained by summing the treatment sum of squares from the individual analyses to obtain a treatment within location sum of squares (with 27 degrees of freedom) and then subtracting out the treatment sum of squares (with 9 degrees of freedom) obtained in the preceding paragraph.

The treatment by replicate within locations sum of squares may be obtained by summing the error sum of squares for the 3 places with 27 + 36 + 45 = 108 degrees of freedom.

As a partial check the sum of the above sums of squares should equal the total sum of squares.

The second question: "How does one compute a least significant difference for place or location means"? might be answered with the question: "Why would one ever want to do such a thing?"

However it is possible to compute a least significant difference (l.s.d.) relative to certain hypotheses, for place means. Under the hypothesis that the treatments in the experiment represent the sole treatment means of interest, then the appropriate error for testing the differences among location means would be the replicates within locations mean square with 12 degrees of freedom. The l.s.d. under this hypothesis for the difference between the means of locations I and II would be

$$t_{.05}\sqrt{\text{Rep. within locations mean square}\left(\frac{1}{40}+\frac{1}{50}\right)}$$

The other l.s.d's should be apparent from the above expression.

Under the hypothesis that the replicates at the locations were not a sample of the replicates but were the only replicates of interest, then the appropriate error mean square would be the treatments by location mean square. However, one may be only interested (this would certainly be a unique situation) in the particular replicates and treatments included in the test, i.e., these entities represent the whole of the population. In this case, the error mean square for testing the differences among location means would be the treatments by replicates within place mean square.

For excellent discussions of this topic consult G.W. Snedecor's "Statistical Methods", chapter 11, 1946, or R.A. Fisher's "The Design of Experiments", section 65, 1942.

The experimenter should not become confused by the difference in numbers at the different locations since in this case they are all proportional. Thus there is no need for computing the analysis of variance by the "method of fitting constants" or the "method of weighted squares of means" (see Snedecor, G.W. Statistical Methods, ch.11, 1946), since the same results will be obtained by these analyses as was obtained above.

QUERY 11.

Ten fertilizer treatments, A, B, C, D, E, F, G, H, I, and J, were tested in 4 randomized complete blocks on cabbage in the following arrangement of plots:

H	G	I	J	A	D	E	B	C	F

The replicate was divided into two parts across all fertilizer plots. Then, at random, two applications of DDT dusting (none and dusting) were made, at random, to the two halves of the plots, thus

No dust
Dust

Is there any way to calculate the interaction of fertilizers treatments and applications of DDT dust?

ANSWER.

In an experiment such as the above there are 4 replicates each for fertilizer treatments and dust applications. The design differs from a split plot in that the application of DDT dust is made in a systematic manner over all whole plots of fertilizer treatment plots. Such an arrangement does not affect the estimation of the interaction of fertilizer treatments and dusting applications. The interaction is estimated with the greatest precision.

The breakdown of the total degrees of freedom, with arrows showing the pairs of mean squares to use in making F tests, follows:

Source of variation	Degrees of freedom
Replicates	3
Fertilizer treatments	9
Reps x Fert.tr.	27
DDT applications	1
DDT applications x reps.	3
Fert.treatments x DDT appl.	9
Fert.tr.x DDT appl.X reps.	27
Total	79

QUERY 12.

An experiment was conducted on 4 fertility levels, A,B,C,and D and 4 levels of plant populations, a,b,c,and d. The crop was sweet corn. The following field arrangement of a 4 x 4 latin square with the split plots ordered in each column was used.

A c b d a	C	B	D
B b c a d	A	D	C
D a d b c	B	C	A
C d a c b	D	A	B

The remaining columns are likewise subdivided but with a different randomization.

What is the analysis, taking into consideration the latin square arrangement of the split or sub plots?

ANSWER. The analysis for this design is more easily understood if the columns are considered singly. The analysis of variance degrees of freedom for one column is:

Source of variation	d.f.
Whole plots	3
Order in whole plots	3
Among population levels	3
Error	6
Total	15

The breakdown of the degrees of freedom for the other columns are obtained similarly.

The combined analysis of variances would be the following:

Sources of variation	d.f.
Columns	3
Among whole plots in columns	12(= 4x3)
Rows	3
Fert.levels	3
Error	6
Among population levels in cols.	12(= 4x3)
Population levels	3
Pop.levels x fert.levels	9
Among orders within columns	12(= 4x3)
Error in columns	24(= 4x6)
Total	63

QUERY 13.

Four fertilizer treatments on peas were tested in 3 randomized complete blocks in 1946 and in 4 blocks in 1948. The following breakdown of the total degrees of freedom and sums of squares was obtained for the individual years:

Source of variation	1946 d.f.	1946 s.s.	1948 d.f.	1948 s.s.
Replicates	2	A	3	G
Treatments	3	B	3	H
Error = Reps.x tr.	6	C	9	I
Total	11	D	15	J
Correction for mean	1	E	1	K
Total uncorrected s.s.	12	F	16	L

What is the combined analysis of variance with regard to breakdown of degrees of freedom? How are the sums of squares for years and treatment by years interaction calculated?

ANSWER.

Despite the fact that different numbers of replicates for treatments are used in the two years, there is little difficulty in obtaining the combined analysis. The number of treatment plots are proportional in the two years, i.e.there are 3 plots of each treatment in 1946 and 4 in 1948.

The correction for the mean for year 1946 is

$$E = \frac{(\text{total yield of all 12 plots})^2}{12 = \text{total no.of plots}}$$

The uncorrected sum of squares (F) is the sum of squares of the 12 plot yields. The corrected sum of squares is

$$D = F - E$$

In computing the sum of squares for the combined analysis a new correction for the mean is required. It is the total of all 28 plot yields squared divided by the total number of plots or

$$M = \frac{(\text{total of 28 plot yields})^2}{28}$$

The sum of squares for years is

$$E + K - M$$

In obtaining the sum of squares for the year by treatment interaction, the treatment sum of squares over the two years is required. This is obtained by summing the squares of the treatment totals and dividing by 3 + 4 = 7 or the number of plot yields making up each total and then subtracting the correction for the overall mean, M. Denoting the treatment sum of squares by N, the treatment by year interaction sum of squares is the difference between the treatment within years (B+H)and the treatment (N) sums of squares,

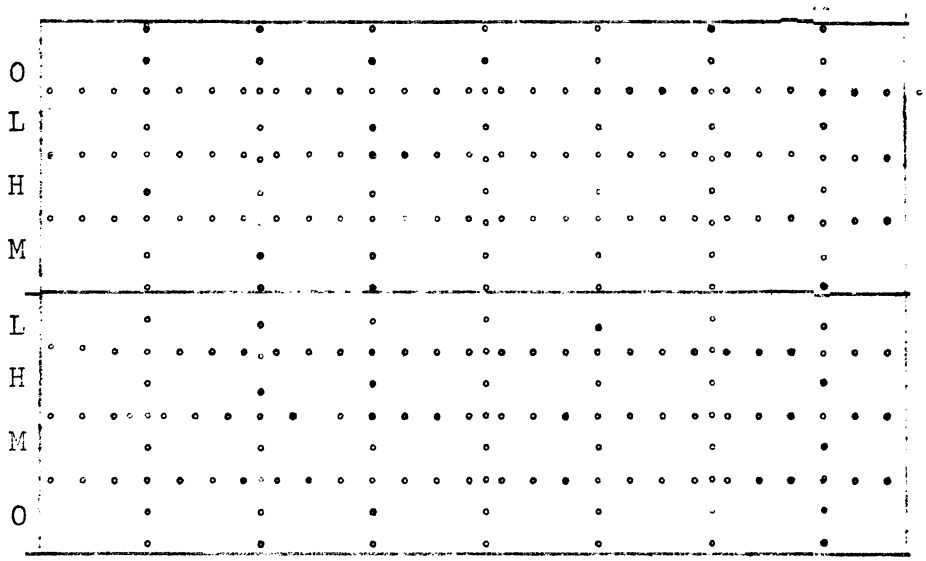
$$B + H - N = \text{treat. x yr. sum of squares},$$

The combined analysis of variance is

Source of variation	d.f.	Sum of squares
Years	1	E + K -M
Reps.within years	5	A + G
Treatments	3	N
Treats.x yrs.	3	B + H -N
Rep.x tr.in years	15	C + I
Total	27	F + L -M
Correction for mean	1	M
Total uncorrected	28	F + L

QUERY 14.

Eight varieties of cabbage were arranged in two 8 x 8 latin squares giving 16 replicates on varieties. The first 4 rows and the last 4 rows of a square were treated with 4 levels of fertilizer (none = 0, 500 lbs.=L, 1000 lbs.=M, 2000 lbs.=H). The fertilizer levels were allotted to the rows at random. The following is the field arrangement of the first 8 x 8 latin square. The first row received no fertilizer, the second 500 pounds, the third 2000 pounds and the fourth row received 1000 pounds of fertilizer.



The last 4 rows of the 8 x 8 latin square are a second replicate of the 4 levels of fertilizer. The second 8 x 8 latin square is similarly divided into 2 replicates. How are the data to be analyzed to obtain the most information about the variety by fertilizer interaction?

ANSWER.

A complete answer cannot be given until more information is available concerning the disposition of varieties to the various fertilizer levels. A partial answer will be given by quoting the analysis given by Miss Gertrude M. Cox (unpublished class notes) for an 8 x 8 half plaid square of 4 varieties and 2^3 treatments:

Systematic arrangement of an 8 x 8 half plaid square for four varieties and $2^3 = 8$ treatments

v_1	(1)	abc	ab	c	bc	a	ac	b
v_2	c	ab	abc	(1)	b	ac	a	bc
v_4	a	bc	b	ac	abc	(1)	c	ab
v_4	b	ac	a	bc	c	ab	abc	(1)
v_1	abc	(1)	c	ab	a	bc	b	ac
v_2	ab	c	(1)	abc	ac	b	bc	a
v_3	bc	a	ac	b	(1)	abc	ab	c
v_4	ac	b	bc	a	ab	c	(1)	abc

The effects confounded are V'AB, V''AC, and V'''BC where

$$V' = -v_1 - v_2 + v_3 + v_4,$$

$$V'' = v_1 - v_2 - v_3 + v_4,$$

$$V''' = v_1 - v_2 + v_3 - v_4,$$

AB = interaction of factors a and b,

AC = " " " a and c,

BC = " " " b and c,

and where the main effects and interactions of a 2^3 factorial are the same as given by Yates (The Design and Analysis of Factorial Experiments, Tech.Comm.No. 35, 1937).

The analysis of variance for one 8 x 8 half plaid square is

Source of variation		Degrees of freedom	
Rows	7		
Replicates		1	
Varieties		3	
Rep. x var.		3	
Columns	7		
Replicates		1	
Confounded effects (V'AB,V''AC,V'''BC)		3	
Reps. x confounded effects		3	
Subplots	49		
Treatments		7	
A			1
B			1
C			1
AB			1
AC			1
BC			1
ABC			1
Variety x treatment		18	
Var.x A			3
" x B			3
" x C			3
" x AB			2'
" x AC			2'
" x BC			2'
" x ABC			3
Error		24	
Total	63		

The analysis of two 8 x 8 half plaid squares with the same effects confounded in the second squares as in the first is

QUERY 14 cont'd. (3)

Source of variation	Degrees of freedom	
Squares	1	
Rows in squares	14	
Reps. in squares		2
Varieties		3
Var. x square		3
Var. x reps. in square		6
Columns in squares	14	
Reps. in squares		2
Effects confounded		3
Effects confounded x squares		3
Rep. x effects confounded in squares		6
Subplots in squares	98	
Treatments		7
Treatments x squares		7
Treatments x varieties		18
Treat. x var. x squares		18
Error in squares		48
Total	127	

QUERY 15.

The following field design of a 4 x 4 latin square superimposed on a 4 x 4 latin square was made (the arrangement is systematic here but was random in the field):

	Col. I				Col. II				Col. III				Col. IV			
	a	b	c	d	b	c	d	a	c	d	a	b	d	a	b	c
Row I		A				B				C				D		
Row II		D				A				B				C		
Row III		C				D				A				B		
Row IV		B				C				D				A		

The treatments a,b,c,and d in column I are laid out in a systematic manner over the plots containing varieties A,B,C,and D in column I, i.e., treatment a continues across the 4 rows of the latin square containing the 4 varieties, A, B, C, and D. Treatments a,b,c,and d are arranged so that each one occupies each of the 4 orders within the columns.

What is the breakdown of the degrees of freedom in the analysis of variance for such a design?

ANSWER.

The analysis of variance breakdown of the total degrees of freedom follows:

Source of variation	Degrees of freedom
Rows	3
Columns	3
Varieties (A,B,C,D)	3
Error for varieties	6
Treatments (a,b,c,d)	3
Orders (position in column)	3
Error for treatments	6
Interaction of varieties and treatments	9
Remainder = error for interaction	27
Total	63
