3-1-1976

Research for the Development of Guidelines for Conducting and Analyzing an Environmental Water Quality Study to Determine Statistically Meaningful Results

Melvin D. Springer
University of Arkansas, Fayetteville

PUB37

#37

Do not take

Last copy

Charge $9.00

PROJECT COMPLETION REPORT

PROJECT NO.: A-033-ARK AGREEMENT NO.: 14-31-0001-5057

Starting Date: January 1975 Completion Date: June 1975

RESEARCH FOR THE DEVELOPMENT OF GUIDELINES FOR CONDUCTING AND ANALYZING AN ENVIRONMENTAL WATER QUALITY STUDY TO DETERMINE STATISTICALLY MEANINGFUL RESULTS

Melvin D. Springer

Principal Investigator

UNIVERSITY OF ARKANSAS

Arkansas Water Resources Research Center

Engineering Experiment Station

Fayetteville, Arkansas

March 1976

ACKNOWLEDGEMENT

The work upon which this report is based was supported in part with funds provided by the Office of Water Research and Technology, U.S. Department of the Interior under project A-033-ARK as authorized under the Water Resources Research Act of 1964, P.L. 88-379 as amended by P.L. 89-404 and P.L. 92-375.

Abstract

This report presents and discusses the basic statistical models and methods which are useful to researchers in the field of water resources research, as well as in other fields. These models and methods are presented from the standpoint of type (parametric and nonparametric - or distribution free) and purpose (e.g., simultaneous comparison of several means, comparison of two or more variances, establishment of a difference between two means with a specified confidence, etc.). The material is presented with emphasis primarily upon methodology, including the necessary assumptions upon which each model is based. No derivations or proofs are given, since these are found in numerous textbooks on statistics readily accessible to the reader. Emphasis is also placed upon the need for the researcher to determine before obtaining data the type of statistical model and analysis required, so that he can use that model or method which is most powerful, and so that he will have the proper data to permit the most efficient analysis. Failure to carry out such preliminary planning relevant to the selection and application of a statistical model will almost always result in either a lack of sufficient relevant data or in the gathering of extraneous data, either of which is unnecessarily costly. Each method is illustrated by an example, together with an interpretation of the result.

TABLE OF CONTENTS

RESEARCH FOR THE DEVELOPMENT OF GUIDELINES FOR CONDUCTING AND ANALYZING AN ENVIRONMENTAL WATER QUALITY STUDY TO DETERMINE STATISTICALLY MEANINGFUL RESULTS

I. IDENTIFYING THE STATISTICAL PROBLEM AND SELECTING THE APPROPRIATE MODEL

Almost all water research projects require the collection and analysis of data. The type and amount of data required depend upon the objectives of the research, the method of analysis, the desired conclusions, and the confidence with which one wishes to assert that said conclusions are correct.

The mere collection, classification, and inspection of data in itself is at best suggestive, and may lead to erroneous conclusions unless subjected to a valid statistical analysis. An important capability which should be present in any water research project involving data collection and analysis is the ability to extract as much as possible--ideally, all--of the information which is inherent in the data, thereby promoting efficiency and cost reduction. The researcher needs a certain amount of information--no more and no less--to reach a specific conclusion with a stated confidence. That is, when he states a conclusion, he must--to be convincing--be able to state what odds he is willing to give that this conclusion is correct. If he wishes to wager with nine-to-one odds (90 percent confidence) that his conclusion is correct, he needs a certain amount of information--more data than if he is content to wager with eight to-two odds (80 percent confidence). Also, the amount of data ("sample size") depends upon both the type of conclusion and the confidence to be attached to the conclusion. For example, if one wishes to establish, with 95 percent confidence, that the average concentration (density) of algae differs at two depth strata in a particular river during a given time period, the amount of data needed differs from the amount required to show that the variation in concentration differs in the two strata.

When, as is frequently the case, there is a choice among several statistical procedures ("statistical models") for data analysis in a particular water research project, it is important to choose the one which will provide the most powerful results. The choice depends upon whether the data satisfy the necessary assumptions underlying the model. An intelligent choice requires both a knowledge of available statistical models and their associated assumptions. Selection of the most appropriate statistical model will result in improved efficiency in the required analysis and a corresponding reduction in cost.

In short, there are many factors which must be taken into consideration if a water research project involving data analysis is to yield maximum results per dollar expended. Moreover, it is imperative to take these factors into consideration during the planning stage of the research study before data collection begins, if conclusions are to be established at a desired confidence level. Necessarily included in the planning stage are:

(1) a statement of specific objectives of the research project

(2) identification of the relevant variables

(3) selection of the appropriate statistical model

(4) specification of the confidence to be attached to each conclusion

(5) determination of the amount of data (sample size) required to achieve this confidence.

Too often the data are collected with little or no thought given to these factors. The result is that frequently a considerable amount of the data gathered is not statistically useful, while at the same time some of the data required to obtain statistically meaningful results are not obtained. Most scientists engaged in water research need guidelines for obtaining and analyzing data so that the conclusions reached are statistically meaningful and defensible. Without such guidelines, a research study may reduce to nothing more than a large-scale data collection project leading to general comparisons and statistically indefensible--and often erroneous--conclusions.

The main function of such guidelines is that of guiding the researcher in the proper planning of the research project prior to obtaining the necessary data and the subsequent methods of analysis after the data have been obtained. The various statistical models (tests, procedure, etc.) which are available, as well as their purpose, must be known to the researcher so that he can choose that statistical model which will best serve his objectives. He should know (before any data are actually collected) what data and how much data (sample size) are needed in order that a conclusion can be reached at a specified confidence level when the appropriate model is used. When there is a choice among several possible models, the most appropriate one should be chosen. The selection of the model involves trade-offs between the use of a "parametric" statistical model based on rather strong assumptions and the use of the weaker "nonparametric" statistical models which make virtually no assumptions. The nature, use, advantages, and disadvantages of various types of sampling (e.g., simple random, stratified, systematic) should be understood.

Accordingly, this project is one of information dissemination, rather than research per se. It presents the various statistical models which are frequently required in research, and explains their use in the analysis of data. Specifically, the various types of models and their variations are presented and discussed.

First, the well-known t-test for comparing two means is presented and discussed. There are two forms of this test: the parametric and nonparametric forms. The first requires rather strong assumptions (which nevertheless are quite often met); the second, a very weak assumption, which in reality is practically always satisfied. There are, of course, trade-offs involved, which are also mentioned. Similarly, one must often compare two sample variances or two sample proportions to determine whether they are significantly different. Appropriate methods for implementing this comparison--either from the standpoint of confidence intervals or hypothesis testing--are presented.

The extension of this model to one which permits the <u>simultaneous</u> comparison of a set of means (more than two) is also presented and discussed; namely, the Analysis of Variance

Model which may have many forms (randomized block, Latin square, factorial designs, etc.). Each is adapted to particular situations as explained. Parametric and nonparametric versions of these models are discussed.

Modification of some of these models, when some data are lost, invalidated, or otherwise unavailable, is also included. This situation sometimes arises in the analysis of real world problems. For instance, in a project involving algae at various water depths, the data obtained at a given time or location may be invalidated by malfunctioning of the equipment at that time or location.

Somewhat less traumatic--but nonetheless important--consequences ensue when complete data are obtained but under differing conditions. For example, the experiment may involve several different laboratories; or several different locations; or several different brands of equipment; etc. This introduces an element of confounding, which can bias the results unless the usual designs are properly modified. A method for accomplishing this modification is discussed under the term "confounding." Related to the problem of confounding is that of "fractional replication" which enables one to reduce the size (and, hence, the cost) of the experiment. This is particularly pertinent in the case of complex experiments involving many combinations of factors. It is not uncommon, for example, to have ten factors at two levels entering into a research experiment, giving rise to a total of $2^{10} = 1024$ possible cases to be examined. There are ways of reducing the size of such an experiment without hindering the analyst in reaching the correct decision. Specifically, in the case of a $2^{10} = 1024$ factorial experiment, it is not unusual to reduce the size of the experiment to $2^6 = 64$ cases or even less, by properly selecting the 64 cases. The procedure for making the proper choice of 64 out of 1024 cases is discussed under the heading "fractional factorial designs."

In problems of estimation and prediction, the method of regression and correlation analysis is often useful, which method is discussed in this document.

The problem of bracketing the central 95 percent (or more generally, the $100(1-\alpha)\%$ portion of a population is at times an important one. Methods for obtaining such limits (frequently referred to as "tolerance limits") are discussed.

Finally, it is possible that a particular problem may be more efficiently solved by the use of stratified or systematic sampling techniques rather than by random sampling. Some comments are made relative to the effect of such a change in sampling methods. However, for the most part, random sampling techniques are appropriate. This is particularly fortuitous, since much of the theory of statistical inference appropriate when random sampling obtains is not directly applicable to situations in which stratified, systematic, or other types of sampling methods are used. The methods of analysis for these situations are considerably more complicated and are beyond the scope of this document.

II. ESTIMATING A POPULATION MEAN: POINT AND INTERVAL ESTIMATES

One of the most common problems confronting researchers, engineers, and businessmen is that of estimating the true value of a "population" mean. Thus, in water research, the researcher may have to decide what is the value of the concentration of algae at a particular water depth in a stream or lake. Or an engineer may have to determine the length of time which the guidance unit of a space system will perform before failure. Again, a canner must determine whether the correct amount of fruit juice is being packed in his 20 ounce cans. Since the label reads "20 ounces", the canner cannot afford to pack much less than 20 ounces for fear of losing customer acceptance or running afoul of the law, while at the same time he cannot afford to pack much more than 20 ounces for fear of losing a substantial part of his profit. In all such problems, one cannot expect a specific value to obtain in each of the above situations. Thus, the concentration of algae will not be the same in samples taken at a specified depth from the same general area because of differences due to random factors beyond human control, such as environmental factors. Similarly, because of imperfect quality control, it would be unrealistic to expect several guidance units to have their first failure occurring at identical times, even though the units were produced by the same manufacturer, installed by the same mechanic, and operated by the same operator. All that one can reasonably require is that the true mean (average) lie within a satisfactory range.

A. Point Estimate

Consider, then, the problem of estimating the true mean μ of a given population with a single estimate; i.e., a point estimate. Several types of point estimates could be used (e.g., the sample mean, median, geometric mean, harmonic mean, etc.) but whatever point estimate is used, it must be based on sample data. It can be shown that the best point estimate is given by the sample mean (rather than, say, the sample median or some other sample statistic). It is "best" from the following standpoints:

(a) It "zeros in" on the true but unknown population mean μ as the sample size increases

(b) It is less variable than other estimates

(c) It extracts all the relevant information from the sample

Thus, for a single estimate (point estimate) of a population mean μ, one uses the sample estimate

$$\overline{X} = \frac{1}{n} \sum_{1}^{n} X_i \quad ,$$

where n denotes the sample size.

Usually, one is more interested in obtaining an interval estimate than a point estimate of μ, since he realizes that the sample mean $\overline{X}$, though the best estimate of μ, will never coincide with μ. It is preferable, therefore, to determine an interval estimate $\overline{X} \pm C$, such that one can assert with a specified confidence that the true mean μ lies within this interval. Thus, if $\overline{X} \pm C$ is a 95% confidence interval for μ, one can wager 95 to 5 odds that the true mean is within this range; and if he did so many times, he would be correct in 95% of the cases.

B. <u>Confidence Interval for the Population Mean μ.</u>

In the following discussion concerning the determination of an interval estimate for the population mean μ it will be assumed that the population is normal; i.e., that the distribution of items in the population is characterized by the probability density function

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}, \quad -\infty < X < \infty$$

where μ, σ, and σ^2 denote, respectively, the population mean, standard deviation, and variance. The method for estimating the constant C for the confidence interval $\overline{X} \pm C$ depends upon whether one knows (at least approximately) the value of σ, perhaps from past experience. If he does, then he can utilize the fact that the variable

$$Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$$

has a standardized normal distribution with mean zero and standard deviation one. Hence, if the sample has actually been drawn from a normal population with mean μ and standard deviation σ, the probability that

$$-1.96 < Z < 1.96$$

is 0.95. Or more generally, the probability that the variable Z will be between the values $\pm Z_{\alpha/2}$ is $1-\alpha$, where $Z_{\alpha/2}$ satisfies the relationship

$$\frac{1}{\sqrt{2\pi}} \int_{-Z_{\frac{\alpha}{2}}}^{Z_{\frac{\alpha}{2}}} e^{\frac{-Z^2}{2}} \, dZ = 1 - \alpha .$$

For this reason, the interval $\left(\overline{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}},\ X + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$ covers the unknown population mean μ with a probability of $1-\alpha$. Or equivalently,

$$X - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < X + Z_{\frac{}{2}} \frac{\sigma}{\sqrt{n}}$$

is a $100\,(1-\alpha)\%$ confidence interval for μ.

Example 1. In a laboratory experiment, 50 engineering students separately measured the specific heat of aluminum, obtaining a mean of 0.2210 calories per centigrade degree per gram. It is known from past experience that this type of measurement has a normal distribution with $\sigma = 0.0240$. Find a 95% confidence interval for the true specific heat (population mean μ) of aluminum.

From the foregoing discussion, it is clear that $Z_{\frac{\alpha}{2}} = Z_{.025} =$ 1.96, so that

$$0.2210 - \frac{1.96}{50}(0.0240) < \mu < 0.2210 + \frac{1.96}{50}(0.0240)$$

or

$$0.2144 < \mu < 0.2276$$

which is the desired 95% confidence interval.

Now suppose that it is known that a population is normal, but one has no knowledge of the population standard deviation σ. Then the best that one can do is to estimate σ from sample data. An unbiased estimate of σ is obtainable from the sample standard deviation, provided it is calculated from the formula

$$s = \sqrt{\frac{1}{n-1}\sum_{1}^{n-1}(x_i - \bar{X})^2}$$

$$= \sqrt{\frac{n\sum_{1}^{n} x_i^2 - (\sum_{1}^{n} x_i)^2}{n(n-1)}} .$$

Then one knows that

$$t = \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}$$

has the so-called t distribution

In other words, the variable $Z = (\bar{X}-\mu)/\frac{\sigma}{\sqrt{n}}$ no longer obtains, but is replaced by the variable t defined above. The reason that the variable t does not have a standardized normal distribution as did Z, is because the constant σ is now replaced with the sample statistic s, which varies from sample to sample. Now a $100(1-\alpha)\%$ confidence interval for μ is given by

$$\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}$$

where $t_{\alpha/2}$ satisfies the relationship

$$\int_{-t_{\frac{\alpha}{2}}}^{t_{\frac{\alpha}{2}}} f(t)dt = 1-\alpha.$$

Unfortunately, this integral cannot be evaluated in closed form. For this reason, values of t_α have been evaluated and tabulated [1] the specific values $\alpha = 0.005, 0.01, 0.025, 0.05, 0.10$ and for the degrees of freedom $n - 1 \leq 29$. (When $n - 1 \geq 29$, the t distribution may for all practical purposes be considered as identical with the normal distribution, so that one can use the Z variable instead of t.) These values, together with a knowledge of the sample size n and sample standard deviation s, enable one to determine the corresponding confidence limits, as the following example shows.

Example 2. A random sample of 25 measurements of the coefficient of thermal expansion of nickel have a mean of 12.81 and a standard deviation of 0.04. Construct a 95% confidence interval for the actual coefficient of expansion. Assume that the 25 measurements constitute a random sample from a normal population.

Using the t distribution, since the population variance is unknown, one has

$$\overline{X} - t_{.025}\frac{s}{\sqrt{n}} < \mu < \overline{X} + t_{.025}\frac{s}{\sqrt{n}}$$

where $t_{.025} = 2.064$ corresponds to $n - 1 = 24$ degrees of freedom. Thus,

$$12.81 - 2.064\ (\frac{0.04}{5}) < \mu < 12.81 + 2.064\ (\frac{0.04}{5})$$

or

$$12.793 < \mu < 12.827$$

That is, one is 95% confident that the true value of the coefficient of expansion of nickel lies between 12.793 and 12.827. Or, equivalently, he can wager with 95 to 5 odds that such is the case.

C. <u>Determination of Sample Size Required for a Specified Confidence</u>

When one uses a sample mean to estimate the mean of a population, he knows that although he is using a method of estimation which has certain desirable properties, the probability is essentially zero that the estimate is equal to the mean μ. Hence, it would seem desirable to accompany such a point estimate of μ with some statement as to how close one might reasonably expect the estimate to be. To examine this error, one makes use of the fact that for large n

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

is a value of a random variable having approximately the standard normal distribution. (If the sample is taken from a normal population, Z is normally distributed regardless of sample size.) Consequently, as previously pointed out, one can assert with probability $1 - \alpha$ that

$$- Z_{\alpha/2} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < Z_{\alpha/2};$$

or equivalently,

$$\frac{|\overline{X} - \mu|}{\sigma/\sqrt{n}} < Z_{\alpha/2},$$

where $Z_{\alpha/2}$ is such that the normal curve area to its right equals $\alpha/2$

(which is equivalent to the integral definition of $\alpha/2$ given previously). Since $|\overline{X} - \mu|$ is the numerical error E which results from estimating μ with the sample mean $\overline{X}$, one can write the above inequality as

$$E < Z_{\alpha/2}\, \sigma/\sqrt{n}$$

with probability $1 - \alpha$. That is, if one estimates μ by means of a random sample of size n, he can assert with a probability of $1 - \alpha$ that the error $E = |\overline{X} - \mu|$ is less than $Z_{\alpha/2}\, \sigma/\sqrt{n}$, at least if n is not too small (say at least 30, as a rule of thumb). Conversely, one can solve the last inequality for n, obtaining

$$n < Z^2_{\alpha/2}\, \sigma^2/E^2.$$

This inequality states that if one selects a sample of size n such that

$$n = Z^2_{\alpha/2}\, \sigma^2/E^2,$$

one can assert with probability $1 - \alpha$ that the error of estimating μ by means of $\overline{X}$ will be less than E. To be able to use this formula for computing the sample size needed to estimate μ in a given situation, it is necessary to specify α, σ, and E. Thus, one must give not only the maximum tolerable error E and the population standard deviation σ, but also the probability $1 - \alpha$ with which one wishes to assert that the maximum error will be less than E. (Note that one cannot determine $Z_{\alpha/2}$ until α - or equivalently $1 - \alpha$ - is specified.). The population standard deviation is usually estimated with prior data of a similar kind, and sometimes a good guess will have to do. The following examples are illustrative.

Example 3. Suppose a utilities company estimates the mean amount of its past-due accounts by taking a random sample of 81 bills. If the mean is \$9.87 and the standard deviation is \$5.14, what is the probability that an error of not more than \$1.00 is made when estimating the mean delinquent account to be \$9.87?

Solution: Use the inequality

$$E < z_{\alpha/2}\ \sigma/\sqrt{n}\ ,$$

and note that $E = 1.00$, $\sigma = 5.14$, $n = 81$.

This inequality will be satisfied if

$$z_{\alpha/2} < \frac{1.00}{0.57} = 1.75\ .$$

From a standardized normal table one notes that for $z_{\alpha/2} = 1.75$, the right tail of the normal curve is $\frac{\alpha}{2} = 0.0401$, so that $1 - \alpha = 0.92$. That is, the probability that an error of no more than 1.00 results is 0.92.

Example 4. The mean of a sample of n "0 gauge" wires is used to check the mean diameter of an incoming shipment of a large number of such wires. How large a sample is necessary if one wishes to be 95% confident that the error in estimating the mean of the shipment is to be less than 0.006? Assume that it is known that the diameters of "0 guage" wires are normally distributed with standard deviation of 0.006 inch.

Solution: Use the aforementioned formula

$$n = z^2_{\alpha/2}\ \sigma^2/E^2$$

where $E = 0.006$, $\sigma = 0.012$, $\alpha/2 = 0.025$. Then

$$n = \frac{(1.96)^2\ (0.012)^2}{(0.006)^2}$$

$$= (3.8416)(4)$$

$$= 15.37$$

$$= 16.$$

Thus, one can be 95% confident that the mean of a sample of 16 measurements will differ from the shipment mean by not more than 0.006 inch.

D. Hypothesis Testing

Closely related to the problem of interval estimation is that of hypothesis testing. Thus, one may wish to know whether a sample of size n and with variance s^2 came from a normal population whose mean is μ and whose variance is unknown. This question can be answered by utilizing the fact that

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

has a t distribution with n-1 degrees of freedom. The following example is illustrative.

Example 5. A random sample of boots worn by 50 soldiers in a desert region showed an average life of 1.24 years with a standard deviation of 0.55 years. Under standard conditions, such boots are known to have an average life of 1.40 years. Is there reason to assert at the 5% level of significance that use in the desert causes the average life of such boots to decrease? In other words, does this sample come from a normal population with known mean but unknown variance?

The answer to this question involves testing a "null" hypothesis H_0 (i.e., a hypothesis to be either accepted or rejected) against an alternative hypothesis H_1. Specifically:

$$H_0: \mu = 1.24$$

$$H_1: \mu < 1.24$$

where μ denotes the mean life of boots worn in the desert.

To reach a decision, one must compute $t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$

where $\mu = 1.24$, $s = 0.55$, $n = 50$. If $t > t_{\alpha/2}$, where $t_{\alpha/2}$ is the critical value of t obtained from a t table, corresponding to a confidence level α and n-1 degrees of freedom, one rejects H_0 In this particular case, let $\alpha = 0.05$. Then

$$t = \frac{1.24 - 1.50}{\frac{0.55}{\sqrt{50}}} = \frac{0.16}{0.0778} = 2.057$$

From the t table, $t_{.025} = 1.96$ for 49 degrees of freedom. Hence, $t = 2.057 > t_{.025} = 1.96$, so one rejects H_0. That is, one concludes that the average life of boots worn in desert regions is different from the average life under standard conditions.

III. ESTIMATING A POPULATION VARIANCE: POINT AND INTERVAL ESTIMATES

Of somewhat less importance than the estimation of the mean - but nevertheless often necessitated - is the estimation of the variance. For while the mean is an estimate of central tendency, the variance about the mean indicates the variation of the items about the mean. In problems of estimation, the accuracy of a predicted value of a population statistic, such as the population mean depends heavily upon the population variance. Estimation of that variance is, therefore, of considerable importance in such a situation - and many others.

A. Point Estimate of the Population Variance σ^2

Just as the sample mean can be used to estimate the population mean, so the sample variance can be used to estimate the population variance. However, unlike the sample mean, the sample variance

$$s^2 = \frac{1}{n} \sum_{i=1} (X_i - \bar{X})^2$$

is not an unbiased estimate of the population variance σ^2. This means that if many sample variances σ^2 (as just defined) were averaged, their average (mean) would not approach σ^2 as the number of samples increased.

In fact,

$$\overline{S^2} = \lim_{N \to \infty} \frac{1}{N} \sum_{j=1}^{N} S_j^2$$

$$= \frac{n-1}{n} \sigma^2$$

However, $\frac{n}{n-1} S^2$ is an unbiased estimate of σ^2. For this reason the sample variance is sometimes defined as

$$S'^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

$$= \frac{n}{n-1} S^2;$$

it is an unbiased estimate of σ^2.

B. Interval Estimate of the Population Variance σ^2

In order to determine an interval estimate (σ_L^2, σ_U^2) which covers the unknown value σ^2 with a specified confidence, one must know the sampling distribution of the variance; or equivalently, the sampling distribution of $\frac{NS^2}{\sigma^2}$

which is merely the sample variance S^2 multiplied by a constant n/σ^2. The reason for utilizing the statistic nS^2/σ^2 is the fact that its sampling distribution is the well-known chi-square (χ^2) distribution with (n-1) degress of freedom, namely,

$$f(X^2) = \frac{1}{2^{\frac{n-1}{2}} \Gamma(\frac{n-1}{2})} (\chi^2)^{\frac{n-1}{2} - 1} e^{-\chi^2/2}, \quad 0 \leq \chi^2 < \infty$$

It is assumed, of course, that the sample of size n, from which χ^2 is computed, was drawn from a normal population. Then

$$\chi^2 = \frac{nS^2}{\sigma^2}$$

is a chi-square variable with n-1 degrees of freedom. Solving this formula for σ^2 in terms of S^2 and χ^2, one obtains

$$\sigma^2 = \frac{nS^2}{\chi_\alpha^2}$$

Then a 100 (1 - α)% confidence interval for σ^2 is given by

$$\frac{n^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{n^2}{\chi^2_{1-\frac{\alpha}{2}}}$$

where $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ are read from a χ^2- table, but satisfy the relations

$$\int_0^{\chi^2_{\alpha/2}} f(\chi^2)\, d\chi^2 = 1 - \frac{\alpha}{2}$$

$$\int_0^{\chi^2_{1-\alpha/2}} f(\chi^2)\, d\chi^2 = \alpha/2.$$

A lower 100 (1-α)% confidence limit on σ^2 is $\frac{nS^2}{\chi^2_\alpha}$, where χ^2_α is determined from the relation

$$\int_0^{\chi_\alpha^2} f(\chi^2)\, d\chi^2 = 1 - \alpha$$

Similarly, an upper 100 (1 - α)% confidence limit on σ^2 is $\chi^2_{1-\alpha}$ determined from the equation

$$\int_0^{\chi^2_{1-\alpha}} f(\chi^2)\, d\chi^2 = \alpha$$

Example - The diameters of random sample of 12 bolts have a variance of $S^2 = 0.000050$. Assuming that the diameters of such bolts constitute a normal population with variance σ^2, find 95% confidence limits for σ^2.

As was previously stated, a 100 (1-α)% confidence interval for σ^2 is given by

$$\frac{n\sigma^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{n\sigma^2}{\chi^2_{1-\frac{\alpha}{2}}} .$$

In the present problem, $\alpha = 0.05$ and $\chi^2_{\alpha/2}$, as well as $\chi^2_{1-\frac{\alpha}{2}}$, has eleven degrees of freedom. Using a χ^2 table, one finds that $\chi^2_{.975}$ = 3.816 and $\chi^2_{.025}$ = 21.920, so that the desired confidence interval is

$$\frac{12(0.00005)}{21.920} < \sigma^2 < \frac{12(0.00005)}{2.816}$$

or $-.000027 < \sigma^2 < 0.000157$

Or equivalently,

$$0.0052 < \sigma < 0.0125$$

C. Hypothesis Testing

Instead of bracketing the unknown population variance, one might choose to view the problem as one of hypothesis testing. That is, he may wish to test the null hypothesis

$$H_0: \sigma^2 = \sigma_0^2$$

against the alternative hypothesis

$$H_0: \sigma^2 \neq \sigma_0^2$$

particularly if he wants to know whether σ^2 differs significantly - in either direction - from some desired value σ_0^2. Or, if he is primarily concerned that the (normal) population variance not exceed some critical value σ_0^2, he would test the null hypothesis

$$H_0: \sigma^2 = \sigma_0^2$$

against the alternative hypothesis

$$H_1 \ \sigma^2 > \sigma_0^2,$$

Similarly, he might test the null hypothesis

$$H_0: \sigma^2 = \sigma_0^2$$

against the alternative hypothesis

$$H_1. \ \sigma^2 < \sigma_0^2$$

The following examples illustrate the procedure. It should be remembered that one necessarily assumes that the sample has come from a normal population.

Example 6. Past data indicate that the variance of measurements made on sheet metal stampings by experienced quality control inspectors is 0.16 square inches. Such measurements made by an inexperienced inspector could have too large a variance (perhaps because of inability to read the instruments properly) or too small a variance (perhaps because unusually high or low measurements are discarded). If a new inspector measures 15 stampings with a variance of 0.11 square inches, test at the $\alpha = 0.05$ level of significance whether the inspector is making satisfactory measurements.

Solution. The null and alternative hypothesis are, respectively,

$$H_0: \ \sigma^2 = 0.11$$

$$H_1: \ \sigma^2 \neq 0.11.$$

To test H_0 one computes nS^2/σ^2, which, as stated previously, has a chi-square distribution with n-1 = 14 degrees of freedom. If the value of nS^2/σ^2 is less than $\chi^2_{0.975} = 5.629$ or greater than $\chi^2_{.025} = 26.119$, one rejects H_0. In this particular case,

$$nS^2/\sigma^2 = \frac{15\,(0.16)}{0.11}$$

$$= 25.818 > 26.119$$

and one concludes, therefore, that the new inspector is not making satisfactory measurements. Specifically, they are in general too large.

If one were only concerned with the measurements being too large, he could sharpen the procedure by carrying out a one-tail test at the α level of significance:

$$H_0: \sigma^2 = 0.11$$

$$H_1: \sigma^2 > 0.11 \quad , \quad \alpha = 0.05.$$

Since $\alpha = 0.05$, one notes that the critical value of χ^2 is $\chi^2_{.05} = 23.685$

and that $\frac{nS^2}{\sigma^2} = \frac{15(0.16)}{(0.11)}$

$$= 25.818 > 23.685$$

Thus, the hypothesis H_0 is rejected as before. Note, however, that the value of $\frac{nS^2}{\sigma^2}$ does not have to be quite as large as before in order to reject H_0; specifically, the critical value is now 23.685 as compared with 26.119. This results in a sharper test.

IV COMPARISON OF TWO SAMPLE MEANS, VARIANCES, OR PROPORTIONS

In many real world problems, one needs to determine whether two sample means are significantly different. That is, one must frequently decide whether two particular samples come from two populations with identical means, or with different means. For example, it may be necessary for the army to determine whether boots worn by soldiers in a desert region have an average life which is less than the average life of boots worn in Arctic regions. The conclusion in such a case is based on comparing the mean life of a sample of boots worn in the desert region with that of a sample of boots worn in Arctic regions.

Similarly, it may be necessary to determine whether the length of life of boots worn in a desert region is more variable than that of boots worn in Arctic regions. That is, are the variances of the two populations (life under desert wear and life under Arctic conditions) significantly different? Again, such a conclusion must be reached by comparing the sample variances of two samples, one from each population.

Actually, the procedure utilized depends upon whether one knows what mathematical function (probability density function) characterizes the distribution of the items in a population. For example, can one assume that the life of boots both in a desert and in an Arctic environment is characterized by a normal distribution? If so, one has a parametric problem which is solved by using the t-test. If one can only assume that the life distribution is characterized by a continuous mathematical function (probability density function), then a nonparametric analysis is used; namely, a nonparametric t-test. For a comparison of variances, the same situation exists. That is, if the populations are normal, a (parametric) F-test is used; if not, a nonparametric or distribution-free F test is used.

A. Parametric Analysis

1. Parametric t-test

This test is designed to determine whether two samples came from two normal populations with identical means; or equivalently, to determine whether the means of two samples from normal populations are significantly different. This test requires that the two normal populations have identical variances - a rather strong assumption. The use of the test is best explained by an example.

Example 7. Members of an army evaluation team are attempting to evaluate the relative merits of two designs of antitank projectiles. A sample of 10 projectiles of Type A are fired at maximum range, with a mean target error of 24 feet and a variance of 16 feet. A sample of 8 projectiles of Type B are fired, with a mean target error of 30 feet and a variance of 25 feet. Is there a significant difference between the mean target errors of the two kinds of projectiles at the $\alpha = 0.01$ level of significance?

Solution.

It is well-known that the statistic

$$t = \frac{\bar{X}_1 - \bar{X}_2 - \delta}{\sqrt{(n_1-1)S_1^2 + (n_2-1)S_2^2}} \sqrt{\frac{n_1 n_2 \; (n_1 + n_2 - 2)}{n_1 + n_2}}$$

where $\delta = \mu_1 - \mu_2$ is the difference between the two population means, has a t distribution. Hence, the problem reduces to testing the null hypothesis H_0 against the alternative hypothesis H_1:

$$H_0: \quad \mu_1 - \mu_2 = 0$$

$$H_1: \quad \mu_1 - \mu_2 \neq 0$$

on the basis of the t value obtained by substituting the relevant sample data into the above t formula. If $t > t_{\alpha/2}$, where $t_{\alpha/2}$ is obtained from the t-table [1, p. 399] with $n_1 + n_2 - 2$ degrees of freedom the hypothesis H_0 is rejected; otherwise, it is accepted. In the problem at hand, $n_1 + n_2 - 2 = 16$, $\delta = 0$, $\alpha/2 = 0.005$, $t_{\alpha/2} = 2.120$,

and $t = \dfrac{24-30}{\sqrt{9(16) + 7(25)}} \sqrt{\dfrac{10(8)\ (16)}{18}}$

$= 2.833 > 2.120$

Thus, one rejects H_0.

If one wished to test the hypothesis H_0: $\mu_1 - \mu_2 = 0$ against the alternative H_1: $\mu_1 - \mu_2 < 0$, one would compare the computed value $t = 2.833$ with the critical value $t_{0.05} = 1.746$. Since $t = 2.833 > 1.746$, H_0 would be rejected. It is assumed, of course, that both samples came from normal populations.

2. Parametric F test for comparing two variances.

The parametric F test is designed to determine whether two samples came from two normal populations with identical variances. Thus, for the two samples in Example 6, one might wish to test the hypothesis

$$H_0: \sigma_1^2 = \sigma_2^2$$

against the alternative hypothesis

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

where σ_1^2 and σ_2^2 denote the variances of the normal populations from which the two samples were taken; namely, the variances of all projectiles of Types A and B which would ever be produced. In words, one wishes to know whether the variability in the magnitude of the error differs for the two types of rounds. To answer this question, one utilizes the fact that the ratio S_1^2/S_2^2 of two sample variances has an F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom, provided that the samples were drawn from normal populations. Critical values of the F distribution corresponding to $\alpha = 0.01$ and $\alpha = 0.05$ levels of significance have been tabulated for various degrees of freedom[1,p.401] Since the entries in the F tables are all equal to or greater than 1, it is necessary to place the larger variance in the numerator of the F ratio. Then, in determining the critical value of F, the degrees of freedom corresponding to the numerator are located at the top of the F table and those corresponding to the denominator are located at the side of the F table. If the F ratio exceeds this critical value, the hypothesis

H_0: $\sigma_1^2 = \sigma_2^2$ is rejected

To illustrate the procedure, consider testing the hypothesis

$$H_0: \sigma_1^2 = \sigma_2^2$$

against the alternative

$$H_1: \sigma_1^2 < \sigma_2^2$$

with $\alpha = 0.05$ for the data of Example 7. Then

$$F = S_2^2/S_1^2$$
$$= 25/16$$
$$= 1.562 .$$

Entering the $F_{0.05}$ table with $n_2 - 1 = 7$ degrees of freedom at the top and $n_1 - 1 = 9$ degrees of freedom at the side, one finds the critical value F* to be F* = 3.29. Since $F = 1.562 < 3.29$, one accepts H_0.

Now suppose one wishes to know whether σ_1^2 was different from σ_2^2 either larger or smaller. That is, one wishes to test the hypothesis

$$H_0: \sigma_1^2 = \sigma_2^2$$

against the alternative hypothesis

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

using the $F_{.05}$ table. Again one calculates the value

$$F = S_2^2/S_1^2$$
$$= 25/16$$
$$= 1.562$$

as before, by placing the larger variance in the numerator of the F ratio. Also, as before, the critical value $F^* = 3.29$ is obtained and one accepts H_0. But now the level of significance is no longer $\alpha = 0.05$ but $\alpha = 0.10$; i.e., the level has doubled. This is due to the fact that the critical values in the F table make no allowance for F ratios which are less than one, and there is a probability of ½ that if one constructs the F ratio as

$$F = \frac{\text{variance of first sample}}{\text{variance of second sample}}$$

the value of F will be less than one.

B. Nonparametric Analysis

1. Nonparametric t test.

Nonparametric t tests are designed to determine whether two samples came from populations with different means, in which it is not required that the populations be normal. In fact, no assumption is placed on the distribution of the items in the population, other than the weak condition that the distribution be described by a continuous mathematical function. While several such tests are available, only the Mann-Whitney U test will be considered here. It is based on rank-order statistics, as is indicated by the following example.

Example 8.

An experiment designed to compare the tensile strength of two kinds of yarn produced the following results (in pounds):

Yarn A: 143.6, 144.8, 145.2, 144.8, 145.6, 146.0,
143.0, 147.4, 144.0, 145.6, 145.5, 144.8

Yarn B: 146.6, 147.8, 144.4, 140.8, 143.0, 148.8,
153.0, 142.4, 146.8, 143.2, 140.9, 150.6

Use the U-test at the $\alpha = 0.05$ level of significance to test the null hypothesis

$$H_0: \mu_1 = \mu_2$$

against the alternative hypothesis

$$H_1: \mu_1 \neq \mu_2$$

To reach a decision, one first jointly arranges the 24 observations according to size, retaining the sample identity of each observation. Then one assigns the ranks 1, 2, 3, . . ., 24 as shown in the following table:

Yarn	B	B	B	B	A	B	A
Observation	140.8	140.9	142.4	143.0	143.0	143.2	143.6
Rank	1	2	3	4.5	4.5	6	7
Yarn	A	B	A	A	A	A	A
Observation	144.0	144.4	144.8	144.8	144.8	145.2	145.5
Rank	8	9	11	11	11	13	14
Yarn	A	A	A	B	B	A	B
Observation	145.6	145.6	146.0	146.6	146.8	147.4	147.8
Rank	15.5	15.5	17	18	19	20	21
Yarn	B	B	B				
Observation	148.8	150.6	153.0				
Rank	22	23	24				

Note that if two or more observations are tied in rank, one assigns to each of the observations the mean of the ranks they jointly occupy.

Construct the statistic

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

where n_1 = size of sample #1

n_2 = size of sample #2

R_1 = the sum of the ranks occupied by the first sample.

It can be shown that if both n_1 and n_2 are greater than 8, the distribution of the U statistic is approximately normal with mean

$$\mu_u = \frac{n_1 n_2}{2}$$

and variance

$$\sigma_u^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}.$$

Thus,

$$z = \frac{U - \mu_u}{\sigma_u}$$

has approximately the standardized normal distribution. In this example

$$\mu_u = 72$$

$$\sigma_u = 300$$

$$= 17.32,$$

$$R_1 = 4.5 + 7 + 8 + 11 + 11 + 11 + 13 + 14$$

$$+ 15.5 + 15.5 + 17 + 20$$

$$= 147.5$$

$$U = 144 + 78 - 147.5$$

$$= 74.5$$

Thus,

$$z = \frac{U - \mu_u}{\sigma_u}$$

$$= \frac{74.5 - 72}{17.32}$$

$$= 0.14$$

Since z is (approximately) normally distributed, $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$. Thus, $z = 0.14 < 1.96$, so H_0 is accepted and it is concluded that the population means are identical.

If n_1 and n_2 are not both greater than 8, the critical value of U is no longer Z , but can be obtained from a special table [13] . If the sample value of U is greater than this critical value, Ho is rejected. Otherwise, H_0 is accepted.

2. Nonparametric F test for comparing two variances

If the ranks are assigned in a somewhat different manner than that used in Example 8, the U statistic can also be used to test the null hypothesis of identical populations against the alternative hypothesis that the populations have unequal variances. The ranks are assigned "from both ends toward the middle" by assigning Rank 1 to the smallest observation, Ranks 2 and 3 to the largest and second largest observations, Ranks 4 and 5 to the second and third smallest, ranks 6 and 7 to the third and fourth largest, and so on. All other aspects are identical with those of the Mann-Whitney U test discussed in Example 8. The example below illustrates the use of the U-test as a nonparametric F test.

Example 9.

Use the U-test to determine whether the two populations from which the samples in Example 8 were taken have equal variances.

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

Yarn	B	B	B	B	A	B	A	A	B
Observation	140.8	140.9	142.4	143.0	143.0	143.2	143.6	144.0	144.4
Rank	1	4	5	8.5	8.5	12	13	16	17

Yarn	A	A	A	A	A	A	A	A	B
Observation	144.8	144.8	144.8	145.2	145.5	145.6	145.6	146.0	146.6
Rank	21.67	21.67	21.67	23	22	18.5	18.5	15	14

Yarn	B	A	B	B	B	B
Observation	146.8	147.4	147.8	148.8	150.6	153.0
Rank	11	10	7	6	3	2

The ranking was carried out as explained above - differently from the ranking method used in the nonparametric t test. Then, as in Example 8,

$$n_1 = n_2 = 12$$

$$n_u = 72$$

$$^{2}_{u} = 300$$

$$\mu_u = 17.32$$

$$\alpha = 0.05$$

$$R_1 = 8.5 + 13 + 16 + 21.67 + 21.67 + 21.67$$
$$+ 23 + 22 + 18.5 + 18.5 + 15 + 10$$
$$= 209.51$$

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$
$$= 144 + 6(13) - 209.51$$
$$= 12.49$$

$$z = \frac{U - \mu_u}{\sigma_u}$$
$$= \frac{12.49 - 72}{17.32}$$
$$= -3.44.$$

Since $|z| = 3.44 > z_{0.025} = 1.96$, the hypothesis H_0 is rejected. That is, one concludes that the population variances are different.

3. <u>A nonparametric test for comparing two proportions</u>

There are frequently situations in which one wishes to know if two proportions are significantly different. For example, it may be desired to know if the proportion π_1 of voters in the state who favor a piece of legislation is significantly different from the proportion π_2 who are opposed. Or it may be necessary to determine whether the population proportion π_1 of patients responding to drug A is different from the population proportion π_2

of patients responding to drug B. As before, the decision must be based upon a comparison of two sample proportions $p_1 = \frac{X_1}{M_1}$ and $p_2 = \frac{X_2}{M_2}$. The question is whether these two sample proportions are significantly different.

If the samples utilized are large - say 50 or greater - then if the two population proportions are identical, the standardized difference of the two corresponding sample proportions is normally distributed. More specifically, if two samples of sizes n_1 and n_2 yield proportions $p_1 = X_1/n_1$ and $p_2 = X_2/n_2$, and if the corresponding population proportions π_1 and π_2 are identical, then

$$z = \frac{\frac{X_1}{n_1} - \frac{X_2}{n_2}}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

has, approximately, a standardized normal distribution with mean zero and standard deviation one, where

$$p = \frac{X_1 + X_2}{n_1 + n_2}$$

is a pooled estimate of the (assumed) common proportion $\pi_1 = \pi_2$.

To illustrate the procedure for determining whether two sample proportions are significantly different, consider the following problem. A manufacturer of electronic equipment wishes to subject two competing brands of transistors to an accelerated environmental test. Of the 80 transistors from the first manufacturer, 25 failed the test, whereas of the 50 transistors from the second manufacturer, 21 failed the test. Using the level of significance $\alpha = 0.05$, test whether there is a difference between the two products.

In essence, one must test the null hypothesis

$$H_0: \pi_1 = \pi_2$$

against the alternative hypothesis

$$H_1: \pi_1 \neq \pi_2$$

where π_1 and π_2 denote, respectively, the true proportions of failures for the two brands of equipment. Thus,

$$p = \frac{25 + 21}{80 + 50}$$

$$= \frac{46}{130}$$

$$= 0.3538$$

$$p_1 = \frac{X_1}{n_1} = \frac{25}{80} = 0.3125$$

$$p_2 = \frac{X_2}{n_2} = \frac{21}{50} = 0.42$$

$$z = \frac{\frac{21}{50} - \frac{25}{80}}{\sqrt{0.3538(0.6462)\ (\frac{1}{80} + \frac{1}{50})}}$$

$$= 1.247$$

Since $z = 1.247 < z_{.025} = 1.96$, one accepts H_0; i.e., he concludes that there is no difference between the two products.

There are some instances in which we are interested in determining whether it is reasonable to conclude that a population proportion has a specified value π. For example, in acceptance sampling one is concerned with the proportion of defectives in a lot. Or in life testing one is concerned with the percentage of certain components which will perform satisfactorily during a stated period of time. This type of problem can again be treated as a problem in hypothesis testing:

$$H_0: \quad p = \pi_0$$

$$H_1: \quad p \neq \pi_0$$

For example, suppose a medical research worker wants to know whether a new muscle relaxant will produce beneficial results in a higher proportion of patients suffering from a neurological disorder than the 0.70 receiving beneficial results from standard treatment. How should he interpret an experiment (at the $\alpha = 0.05$ level of significance) if 156 of 200 patients obtained beneficial results with the new relaxant?

To answer this question, we test the hypothesis

$$H_0: \; p = \pi_0 = 0.70$$

against the alternative hypothesis

$$H_1: \; p > 0.70$$

by computing

$$z = \frac{x - n\pi_0}{\sqrt{n\,\pi_0(1-\pi_0)}}$$

$$= \frac{156 - 200\,(0.70)}{\sqrt{200\,(0.70)\,(0.30)}}$$

$$= 2.468$$

and comparing the value with $z_{.025} = 1.96$. Since $z = 2.468 > 1.96$, one rejects H_0 and concludes that the new muscle relaxant is more beneficial than the old.

One could also obtain a 100 $(1 - \alpha)$% confidence interval for a population proportion π, namely,

$$\frac{X}{n} - z_{\alpha/2}\sqrt{\frac{\frac{X}{n}(1 - \frac{X}{n})}{n}} < \pi < \frac{X}{n} + z_{\alpha/2}\sqrt{\frac{\frac{X}{n}(1 - \frac{X}{n})}{n}} \, .$$

For example, in the muscle relaxant problem just considered, a 95% confidence interval for the new muscle relaxant would be

$$\frac{156}{200} - 1.96\sqrt{\frac{\frac{156}{200}(1 - \frac{156}{200})}{200}} < \pi < \frac{156}{200} + 1.96\sqrt{\frac{\frac{156}{200}(1 - \frac{156}{200})}{200}}$$

or

$$0.7226 < \pi < 0.8374 \, .$$

Such confidence intervals on can also be obtained graphically from specially constructed charts [14] , which do not require that n be large.

One final point regarding proportions. The magnitude of the error one incurs by using $\frac{X}{n}$ as an estimate of π depends upon n. For n large (say 50 or greater), the maximum error which one risks with a probability of $1 - \alpha$ is

$$E = z_{\alpha/2}\sqrt{\frac{\pi(1 - \pi)}{n}} .$$

However, since one does not know the value of π the value of E cannot really be obtained from this formula unless he has some approximate estimate of π. In the absence of such an estimate,

$$E \leq z_{\alpha/2}\sqrt{\frac{1}{4n}}$$

since $\pi(1 - \pi) \leq \frac{1}{4}$. Conversely, if one wishes the estimate $\frac{X}{n}$ to differ from π by not more than E, with a probability of $1 - \alpha$, he must utilize a sample of size

$$n \doteq \pi(1 - \pi)\left[\frac{z_{\alpha/2}}{E}\right]^2 ,$$

or if no estimate of π is available,

$$n = \sqrt{\frac{1}{4}\,\frac{z_{\alpha/2}}{E}}^{\,2} .$$

V. THE PROBLEM OF ESTIMATION AND COMPARISON OF SEVERAL MEANS

A. Introduction

Analysts, researchers, and engineers dealing with real world problems are frequently faced with the necessity of deciding which means (averages) in a set of means are different and which are identical. Thus, in determining whether the mean (average) concentration of algae differs for five depth strata, one must decide whether all five means in the set of five depth strata are identical. If they are not all identical, which ones differ--some, or all five? Similarly, in evaluating the effectiveness of, say, three different drugs in the treatment of a disease, one wishes to decide whether all three are or are not equally effective by comparing the mean response of the patients to each drug.

B. Selection of the Model

There are basically two types of models from which to choose the test: parametric and distribution-free (nonparametric). Parametric models require a knowledge of the distribution of the observed values of the relevant variable, whereas distribution-free models only require that the distribution of the relevant variable be (mathematically) continuous, without the necessity of specifying its mathematical form. In many cases the parametric model requires that the variable be normally distributed, which is often approximately true.

One might well wonder, then, why the distribution-free test is not always used, since one would then never run the risk of violating assumptions which, when not satisfied, could cause wrong conclusions with serious resultant consequences. The reason is that the distribution-free test is, in statistical jargon, less powerful than its parametric counterpart. This means, for example, that the differences between means in a set of population means must be large in order to be detected by a distribution-free F test. Thus, with a distribution-free F test, we are more apt to conclude that the mean density of algae is the same for n different depth strata, when in fact it is not. What this means is that a parametric test should be used when the underlying assumptions are not seriously violated. If there is a serious violation of assumptions, then the ideal approach is to apply both the parametric and distribution-free tests. If they lead to the same conclusion at the desired significance level, then the analyst knows that he has a bona fide conclusion at the stated confidence level. If the two tests lead to different conclusions, one would abide by the conclusion from the distribution-free test, if he felt the assumptions of the parametric test were seriously violated. If he had no indication as to such violations, he would probably strike a compromise. Thus, if the parametric F test indicated that the algae concentration means were significantly different at the 5% level of significance, but the distribution-free F test indicated a significant difference at the 15% level, the researcher or analyst would probably conclude that the difference was significant at the 10% level. What this means is that the researcher would conclude that the true mean algae concentrations are not the same for all depth strata; the probability that this conclusion is incorrect is, in the compromise approach, 0.10. An example along this line will be given later.

Basically there is one parametric model for testing a set of means: the conventional analysis of variance or F test. Actually, the purpose of this model is threefold:

(1) To determine whether treatment differences that are of interest exist, and if so, to estimate these treatment differences. In this statement both the words "treatment" and "difference" are used in a rather loose sense; e.g., a treatment difference might be the difference between the mean yields of any two of five varieties in a plant-breeding trial, or the relative toxicity of an unknown to a standard poison in a dosage mortality experiment. One wants such estimates to be efficient. That is, roughly speaking, one wants the difference between the estimate and the true value to have as small a variance as can be attained from the data that are being analyzed.

(2) To obtain some idea of the accuracy of our estimates, e.g., by attaching to them confidence limits.

(3) To perform tests of significance with the parametric model. This consists of carrying out the conventional F test (which will be explained later) to test the hypothesis that a set of population means all have the same value. We should like this test to have

the property that if one attaches a confidence coefficient of 0.05 to the conclusion that the means are not all identical, the probaility of getting the observed result (F value) or a more discordant one (larger F value) when in fact the population means are identical, is equal to or less than 0.05.

C. Elements Involved In the Conventional Parametric Analysis of Variance Model

When using an analysis of variance model, one generally recognizes three types of effects:

(1) treatment effects--the effects deliberately introduced by the experimenter--e.g., five strata depths at which the algae concentration will be measured.

(2) environmental effects--these are certain features of the environment which the analysis enables one to measure--e.g., the effect of speed, driver, make of car, etc. on gasoline mileage when one is determining whether some brands of gasoline are superior to others insofar as average mileage is concerned.

(3) experimental errors--this term includes all elements of variation not taken into account in (1) or (2).

D. Underlying Assumptions in Parametric Analysis of Variance Model

The assumptions required in the analysis of variance model in order for the foregoing properties to hold are as follows:

1. The numbers in each category are random variables which are distributed about the true population mean for that category. For example, the algae concentration values for a given depth stratum are random variables which are distributed about the true mean for that depth stratum.
2. Additivity. For example, in a two way classification involving rows and columns,

$$m_{ij} = m_{..} + (m_{i.} - m_{..}) + (m_{.j} - m_{..})$$

where

m_{ij} = element in the ithe row and j column of the population

$m_{..}$ = mean of the entire population

$m_{i.}$ = mean of i<u>th</u> row of population

$m_{.j}$ = mean of j<u>th</u> column of population

3. Homogeneity of variances. The variance of the items is assumed to be the same for each class or category. The items are also assumed to be mutually uncorrelated.
4. The items in the population are normally distributed -- or for the most general case, have a multivariate normal distribution. In the algae example, the concentration values for a given depth stratum are assumed to be normally distributed about the mean concentration for that depth stratum.

It has been found that the (conventional) parametric analysis of variance model is quite robust with regard to all of the above assumptions except that of homogeneity of variances. What this means is that primarily one needs to be most concerned about the homogeneity of variance assumption--the others are relatively less important. Lack of homogeneity of variances has a serious effect on the efficiency of the test, and will tend to indicate significant differences among population category means when there is no difference. The analysis of variance is not much affected by a moderate lack of normality. The same is true of nonadditivity effects. The assumption of mutual independence among the items is important, but is usuallv reasonably well satisfied.

VI. VARIATIONS IN THE CONVENTIONAL PARAMETRIC ANALYSIS OF VARIANCE MODEL

There are many variations in the parametric analysis of variance model, all of which are based on the conventional parametric F test. Each one has its particular place and enables one to reach certain decisions (or in statistical vernacular, to test certain hypotheses) with a particular degree of efficiency. The basic variations in this model will now be illustrated.

A. Completely Randomized Design (One-Way Classification)

In a completely randomized design, or one-way classification, a set of sample or treatment means is compared, with no attempt to remove the effects of extraneous sources of variability. The effects of these extraneous sources are randomized over the entire experiment so that it will not bias the conclusions. It will, however, inflate the unit of measure (namely, the error variance) for detecting differences between treatment population means. For example, one may wish to test four hull designs of motorboats to determine whether one is superior to another insofar as top speed is concerned. To make such a determination, suppose that four boats, each with a different hull design, were run on a marked course in random sequence, and the time required (in minutes) to cover the course was observed. The sequence was chosen in a random manner so as to "average out" the effects from extraneous sources such as condition of the water (e.g., calm, moderate, or choppy). The results are shown in Table 1.

	Day 1	2	3	Totals
Design A	45	46	51	142
Design B	42	44	50	136
Design C	36	41	48	125
Design D	49	47	54	150
	172	178	203	553

Table 1. Specific Results in the Completely Randomized Design

The means for designs A, B, C, and D, are, respectively, 47.333, 45.333, 41.667, and 50. The question is whether these four samples came from normal populations whose means μ_i, $i = 1, 2, 3, 4$ are all identical, or whether some are different.

To answer this question, we set up the so-called null hypothesis

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

to be tested against the alternative hypothesis

$$H_1: \text{ the } \mu_i \text{ are not all identical.}$$

(The term "null hypothesis" merely means that a hypothesis is to be tested for acceptance or rejection.) The evaluation is carried out with the usual analysis of variance computations, which are readily available in text books of experimental design and will not be elaborated upon here.

$$C = \frac{(\sum_i \sum_j x_{ij})^2}{rc} = \frac{(45 + 42 + \ldots + 48 + 54)^2}{(4)\ (3)}$$

$$= \frac{(553)^2}{12} = 25{,}484.08$$

r = number of rows
c = number of columns

SST = Total sum of squares

$$= (45)^2 + (42)^2 + \ldots + (54)^2 - 25{,}484.08$$

$$= 264.92$$

SS(Tr) = Sum of squares for treatments

= SSR (Sum of squares between row means)

$$= \frac{(142)^2 + (136)^2 + (125)^2 + (150)^2}{3} - 25{,}484.08$$

$$= \frac{17161 + 21025 + 18496 + 19881}{3} - 25{,}484.08$$

$$= \frac{76785}{3} - 25{,}484.08$$

$$= 25595 - 25484.08$$

$$= 110.92$$

SSE = Error sum of squares = 264.92 - 110.92 = 154

The results are summarized below in the usual analysis of variance format.

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Square	F	Critical Value of $F_{.05}$
Hull designs	3	110.92	36.97	1.921	4.07
Error	8	154	19.25		
Total	11	264.92			

Table 2.

Since the computed value $F = 1.921$ is less than the critical value of $F_{.05} = 4.07$ for 3 and 8 degrees of freedom, one accepts the null hypothesis H_0. That is, one concludes that the three hull designs are not significantly different insofar as their effort on top speed of the boat is concerned. The probability that one is wrong in this conclusion and that at least two of the hull designs are significantly different in their effect upon average speed is 0.05.

In the above design, one assumes that the items in the ith row come from a normal population with mean μ_1 and variance σ^2, i = 1, 2, 3. One further assumes that the population element for Design i and Day j is expressible in the form

$$\mu_{ij} = \mu + (\mu_i - \mu) + \epsilon_{ij}$$

i = 1, 2, 3, 4, j = 1, 2, 3; i.e., as a linear additive function of the population grand mean $m_{..}$ and the population hull design mean m_1. The ϵ_{ij} are standardized normal random variables.

Ninety-five percent confidence limits on the mean speed for each hull design are:

$$\bar{X}_i - t_{.025}\sqrt{\frac{s^2}{3}} < \mu_i < \bar{X}_i + t_{.025}\sqrt{\frac{s^2}{3}}$$

where $s = \sqrt{\frac{19.25}{3}} = 2.533$, $t_{.025} = 2.306$, for 8 degrees of freedom. Specifically

$$41.492 < \mu_1 < 53.174$$

$$39.492 < \mu_2 < 51.174$$

$$35.826 < \mu_3 < 47.508$$

$$44.159 < \mu_4 < 55.841$$

B. Randomized Block Design--Two-Way Classification

In the example of Table 1, no attempt was made to remove the effect of extraneous sources upon boat speed, such as the condition of the water (e.g., calm, moderate, chppy). Now the effect of these water conditions was randomized over the experiment (since the sequence of experiments was randomized) and did not bias or otherwise invalidate the experiment--it merely made it more difficult to pick up an actual difference among hull designs by inflating the unit of measure for any such difference (namely, the SSE or error variance). One can, however, deflate the unit of mearusre or error variance by removing the effect of extraneous factors by using other designs. In particular, the effect of one such factor can be removed by using a randomized block design.

To illustrate the removal of an extraneous factor, suppose the four hull designs were tested under water conditions characterized as calm, moderate, or choppy, and which occurred, respectively, on days 1, 2, and 3. Thus, columns 1, 2, and 3 are now labeled calm, moderate, and choppy.

The additional element in the two-way analysis of variance is the removal of the effect of observed differences in column (water condition) means upon the unit of measure (error variance). In this particular example, it is of sufficient magnitude to change the conclusion, namely, to reach the decision that the population means (speeds for the three different hull designs) are significantly different. This conclusion was not reached with the completely randomized design because the differences among treatment means were masked by the effect of water condition differences on the error variance (which was increased). The results are shown in Table 3.

SSC = Sum of squares between (water condition) means

$$= \frac{(172)^2 + (178)^2 + (203)^2}{4} - 25{,}484 = 135$$

$$SS(Tr) = \frac{(142)^2 + (136)^2 + (125)^2 + (150)^2}{3} - 25{,}484 = 111$$

SSE = SST - SS(Tr) - SSC

= 265 - 111 - 135 = 19

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F	Critical Value $F_{.05}$
Hull designs	3	111	37.0	11.6	4.76
Water Conditions	2	135	67.5	21.1	9.55
Error	6	19	3.2		
Total	11	265			

Table 3. Analysis of Variance Summary

Since $F > F_{.05}$ for both hull designs and days, one concludes that:

1. The condition of the water has a significant effect on the top speed for the three hull designs.
2. The hull design means (average boat speeds for the hull designs) are not all identical.

In other words, one rejects the aforementioned null hypothesis

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

in favor of the alternative hypothesis

$$H_1: \text{ the } \mu_i \text{ are not all identical.}$$

For the factorial, two-way table, one has the following confidence limits on the three means:

$$\overline{X}_i - t_{.025} \frac{s}{\sqrt{n}} < \mu_i < \overline{X}_i + t_{.025} \frac{s_{\overline{x}}}{\sqrt{n}}, \quad i = 1, 2, 3, 4$$

where $s = 1.789$, $n = 3$, $t_{.025} = 2.447$ (for six degrees of freedom), and $t_{.05}\ s/\sqrt{n} = 2.528$. Specifically,

$$44.805 < \mu_1 < 49.861$$
$$42.805 < \mu_2 < 47.861$$
$$39.139 < \mu_3 < 44.195$$
$$47.472 < \mu_4 < 52.528$$

Note that the widths of the confidence intervals are considerably smaller for this factorial design than for the randomized block, simply because the effect of the water conditions (days) on the error variance has been removed.

At this point, however, one does not know whether all three hull design means are different, or whether two of the three means are identical but different from the third. This decision is made on the basis of what is called, in statistical jargon, a multiple comparisons test which is discussed in the next section.

C. Multiple Comparisons Tests

Once a significant F-test has been obtained, there are several methods available to determine which of the means are significantly different. One of the more commonly used of such tests is the Duncan Multiple Comparisons Test, which will be applied here.

To apply this test, one proceeds through the following steps:

1. Arrange the four means in order from low to high
2. Enter the analysis of variance table (Table 3) and take the error mean square with its degrees of freedom.
3. Obtain the standard error of the mean for each treatment

$$s_{\bar{x}_{i.}} = \sqrt{\frac{\text{error mean square}}{\text{number of observations upon which } \bar{x}_{i.} \text{ is based}}}$$

where the error mean square is the one used as the denominator in the F test.

4. Enter a Studentized range table (e.g., [1] Appendix, Table X) of significant ranges at the α level desired (in our problem, $\alpha = 0.05$), using d.f. = degrees of freedom for error mean square and $p = 2, 3, \ldots, k$, and list these k ranges (k = number of means; in our problem, $k = 4$).
5. Multiply these ranges by the various $s_{\overline{X}_i}$ to form a group of $k - 1$ least significant ranges.
6. Test the observed ranges between means, beginning with the largest versus smallest, which is compared with the least significant range for $p = K$; then test largest versus second smallest with the least significant range for $p = k - 1$; and so on. Continue this for second largest versus smallest, and so forth, until all $k(k-1)/2$ possible pairs have been tested. This procedure results in various subsets of means, such that no two means in the same subset can be declared significantly different.

Following the steps given above, one has, for the foregoing example, the following multiple comparison analysis:

	(4)	(3)	(2)	(1)
1. k = 4 means are	41.667	44	47.333	50
for treatments	C	B	A	D

2. From Table 3, error mean square is 3.2 with six degrees of freedom.

3. Standard error of mean is

$$s_{\bar{x}_{i.}} = \sqrt{\frac{M.S.E.}{n}} = \sqrt{\frac{3.2}{3}} = \frac{1.789}{1.732} = 1.033$$

4. From Appendix Table X [1] , at the 5% level, the significant ranges are, for 6 degrees of freedom,

p =	2	3	4
ranges =	3.46	3.59	3.65

5. Multiplying by the standard error 0.894, the least significant ranges are

p =	2	3	4
LSR =	3.574	3.708	3.770

6. Largest versus smallest: 1 versus 4 = 8.333 > 3.770

   Largest versus next smallest: 1 versus 3 = 6 > 3.708 .

   Largest versus next largest: 1 versus 2 = 2,667 < 3.574.

   Second largest versus smallest: 2 versus 4 = 5.666 > 3.708 .

Second largest versus next smallest: 2 versus 3 = 4.333 > 3.574.

Third largest versus smallest: 3 versus 4 = 2.333 < 3.574.

41.667	44	47.333	50
C	B	A	D
(4)	(3)	(2)	(1)

Here there is a significant difference between (1) and (4), (1) and (3), (2) and (4), and (2) and (3) but not between (1) and (2) or (3) and (4). Therefore, either design A or D is preferable over C and B insofar as top speed is concerned, but A is not significantly different from D, and B is not significantly different from C.

D. Latin Square Design

Suppose one wishes to remove the effect of two extraneous factors on the treatment means: what experimental design can he use? One possible design, which is quite efficient, is the Latin square design, which is discussed in any basic textbook on experimental design, and is illustrated by the following example.

Three experimental fuels are tested to determine if there is any difference in the length of time an engine will operate on one gallon of the fuel. The number of minutes three engines E_1, E_3, and E_3, tuned up by mechanics M_1, M_2, and M_3, operated with one gallon of fuel, A, B, or C, is observed and the results are recorded. The experimental design used is a so-called Latin square in which each fuel (A, B, C) occurs once and only once in each row and each column of a 3 x 3 Latin square (Figure 1). The experiment was replicated three times, in order to obtain a smaller error variance (based on more degrees of freedom) which will make the F test more powerful. An n x n Latin square with no replication has only (n - 1)(n - 2) degrees of freedom. Thus, a Latin square of order 3, such as the one in this example, has only two degrees of freedom, which does not provide a very sensitive F test.

	E_1	E_2	E_3	
M_1	A 16	B 21	C 14	51
M_2	B 25	C 18	A 23	66
M_3	C 16	A 21	B 26	63
	57	60	63	180

(a)

	E_1	E_2	E_3	
M_1	A 17	B 23	C 13	53
M_2	B 28	C 19	A 21	68
M_3	C 12	A 20	B 25	57
	57	62	59	178

(b)

Figure 1(a) (b) (c). Three Replications of a Third Order Latin Square Design.

	E_1	E_2	E_3	
M_1	A 15	B 20	C 14	49
M_2	B 26	C 16	A 24	66
M_3	C 19	A 24	B 28	71
	60	60	66	186

(c)

The basis for the analytical computations are given in any basic experimental design textbook, and therefore will not be discussed in detail here. Conceptually, the effect of differences in row (mechanics) means and column (engine) means on the error variance has been removed, thereby decreasing the error variance and increasing the power of the F-test. The necessary calculations are shown below.

$$C = \frac{T^2_{...}}{rn^2} = \frac{(180 + 178 + 186)^2}{3(9)} = \frac{(544)^2}{27} = 10{,}960.5926$$

$$\underset{\substack{i \\ \text{(rows)}}}{\Sigma} \; \underset{\substack{j \\ \text{(cols)}}}{\Sigma} \; \underset{\substack{k \\ \text{(Latin letters)}}}{\Sigma} Y^2_{ijk} = 16^2 + \ldots + 26^2 + 17^2 + \ldots + 25^2 + 15^2 + \ldots + 28^2 = 11536$$

$$SST = \sum_i \sum_j \sum_k Y_{ijk}^2 - C = 11{,}536 - 10{,}960.593 = 575.407$$

$$SSR = \frac{1}{rn} \sum_{i=1}^{n} T_{i.}^2 - C = \frac{1}{9}((153)^2 + (200)^2 + (191)^2)$$

$$- 10{,}960.593$$

$$= 138.296 \text{ (for row effects)}$$

(r = number of replicates)

$$SSC = \frac{1}{rn} \sum_{j=1}^{n} T_{.j}^2 - C = \frac{1}{9}((174)^2 + (182)^2 + (188)^2)$$

$$- 10{,}960.593$$

$$= 10.963 \text{ (for column effects)}$$

SS(Tr) = SS (Latin letters)

$$= \frac{1}{rn} \sum_{k=1}^{n} T_{(k)}^2 - C$$

$$= \frac{1}{9}[(181)^2 + (222)^2 + (141)^2] - 10{,}960.593$$

$$= 11{,}325.111 - 10{,}960.593$$

$$= 364.518 \text{ (for treatment, i.e., fuel effects).}$$

(The figure 181 is the sum of the nine values for A:

16 + 23 + 21 + 17 + 21 + 20 + 15 + 24 + 24 = 181.

Similarly, 222 is the sum of the nine values for B and 141 the sum of the nine values for C.)

$$SS(Rep) = \frac{1}{n^2} \sum_{\ell=1}^{r} T_{..\ell} - C$$

$$= \frac{1}{9}((180)^2 + (178)^2 + (186)^2) - 10{,}960.593$$

$$= 10{,}965.444 - 10{,}960.593 = 4.851 \text{ (for replicates)}$$

$$SSE = SST - SS(Tr) - SSR - SSC - SS(Rep)$$

$$= 575.407 - 138.296 - 10{,}963 - 364.518 - 4.851$$

$$= 56.779$$

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F	$F_{.05}$	$F_{.01}$
Fuel (Treatments, i.e., Latin letters)	2	364.518	182.259	57.787	3.55	6.01
Mechanics (Rows)	2	138.296	69.148	21.924		
Engines (Cols)	2	10.963	5.482	1.738		
Replication	2	4.851	2.425	0.769		
Error	18	56.779	3.154			
Total	26	575.407				

Table 4. Analysis of Variance Summary For Latin Square.

From an $F_{.05}$ table (see, for example, [1], Table 6), the F value for 2 and 18 degrees of freedom is 3.55. The $F_{.01}$ value for 2 and 18 degrees of freedom is 6.01. One concludes, therefore, that the fuel means are significantly different at both the 5% and 1% levels of significance. However, the fuel means may or may not all be different. To reach a decision on this point, the Duncan test is applied corresponding to $\alpha = .01$ (i.e., the 1% level of significance).

$$s_{\bar{x}} = \sqrt{\frac{M.S.E.}{n}} = \sqrt{\frac{3.154}{9}} = \frac{1.7760}{3} = 0.5920$$

C	A	B
(3)	(2)	(1)
15.67	20.11	24.67

$\alpha = 0.01$

p	2	3
q	4.07	4.25

p	2	3
q	2.409	2.516

Largest vs. smallest: (1) vs. (3) = 9 > 2.516

Largest vs. next smallest: (1) vs. (2) = 4.56 > 2.409

Second largest vs. smallest: (2) vs. (3) = 4.34 > 2.409

<u>C</u> <u>A</u> <u>B</u>

One concludes, therefore, that all three means are significantly different from each other.

It might also be pointed out that the "mechanics" means are not all identical, so that it was worthwhile to remove the effects of this variable from the error variance. Since one is not particularly interested in which "mechanics means" are significantly different (or whether they all differ from each other) the Duncan test is not applied. The "engines means" are not significantly different, so that one could have used a randomized block design and still reached the correct decision. In other words, in hindsight one now knows that the extraneous variable "engines" does not affect the results, but it was necessary to use a Latin square design to determine this fact.

Again, concerning the assumptions of normality and homogeneity of variances, one assumes that the items in each row and in each column have normal distributions with the same variance σ^2. Similarly, the response (operation time) for each fuel (A, B, C) has a normal distribution with variance σ^2. In actuality, the null hypothesis

$$H_0: \mu_1 = \mu_2 = \mu_3$$

was tested against the alternate hypothesis

$$H_1: \text{ the } \mu_i,\ i = 1, 2, 3 \text{ are not all identical,}$$

with the result that H_0 was rejected in favor of H_1 at the 1% level of significance.

E. Graeco-Latin Squares

The use of a Latin square enables one to remove the effect of two extraneous variables (rows and columns) upon the error mean square. By using a Graeco-Latin square, the effect of yet another extraneous variable on the error variance can be removed. The method by which this is done will now be illustrated by an example. A discussion of the Graeco-Latin square can be found in any standard text on experimental design. It will merely be pointed out here that in a Graeco-Latin square, each Latin-letter (treatment) occurs once and only once in each row and in each column; each Greek letter occurs once and only once in each row and in each column; and each Greek letter occurs in combination with each Latin letter once and only once. The only additional calculation (beyond those for the Latin square) is for the sum of squares between "Greek letter means".

Consider, then, the following example of the application of a Graeco-Latin square. A processor of breakfast foods wishes to study the effectiveness of different kinds of packaging upon scales. He used a fifth order (n = 5) Graeco-Latin square design, with the results shown in Table 5, where α, β, γ, δ, and ε represent (in increasing order of magnitude) the amount of money spent on newspaper ads for the product on the day before the experiment, and the rows represent different locations within identically designed supermarkets, which are represented in turn by the five columns. The figures are the number of sales of breakfast food from 9 A.M. to 11 A.M.

Aα 50	Bβ 51	Cγ 53	Dδ 55	Eε 56	265
Bγ 51	Cδ 50	Dε 50	Eα 45	Aβ 49	245
Cε 45	Dα 37	Eβ 39	Aγ 40	Bδ 41	202
Dβ 39	Eγ 40	Aδ 41	Bε 44	Cα 37	201
Eδ 43	Aε 47	Bα 41	Cβ 42	Dγ 42	215
					1128

Table 5. Graeco-Latin Square

$$C = \frac{T_{\cdot\cdot\cdot}^2}{n^2} = (1128)^2/25 = 50895.36$$

$$\text{Latin} \quad SSL = \frac{1}{n}\sum_{k=1}^{n} T_{(k)}^2 - C = \frac{254,500}{5} - 50,895.36 = 4.64$$

$$\text{Rows} \quad SSR = \frac{1}{n}\sum_{k=1}^{n} T_{i\cdot\cdot}^2 - C = \frac{257,680}{5} - 50,895.36 = 640.64$$

Columns $SSC = \frac{1}{n} \sum_{j=1}^{n} T^2_{.j.} - C = \frac{254,486}{5} - 50,895.36 = 1.84$

Greek (treatments) $SSG = \frac{1}{n} \sum_{g=1}^{n} T^2_{g} - C = \frac{255,040}{5} - 50,895.36 = 112.64$

$$SST = \sum_{j=1}^{n} \sum_{i=1}^{n} Y^2_{ij(k)(g)} - C = 51,008 - 50,895.36 = 772.64$$

$$SSE = SST - SS(Tr) - SSR - SSC - SS6 = 22.88$$

	Source	d.f.	Sum of Squares	Mean Square	F	$F_{0.01}(4,8)$
(Latin)	Treatments (kinds of packaging)	4	4.64	1.16	0.72	7.01
(Rows)	Locations	4	640.64	160.16	99.5**	7.01
(Columns)	Supermarkets	4	1.84	0.46	0.28	7.01
(Greek)	Money spent on advertising	4	112.64	28.16	17.5**	7.01
	Error	8	12.88	1.61		
	Total	24	572.64			

Table 6. Analysis of Variance for Graeco-Latin Square

One concludes that the amount of money spent on advertising is significant at the 1% level, as is the location within identically designed supermarkets. The differences between identically designed supermarkets is not significant; neither are the differences in packaging.

F. Factorial Designs

In the previous sections, one was primarily concerned with the effects of one variable, whose values were referred to as "treatments". Extraneous values were accommodated by means of blocks, replicates, or the rows and columns of Latin and Graeco-Latin square designs, so that the effects of such extraneous variables on the error variance were removed. This section, however, will deal with the individual and joint effects of several variables, and combinations of the values, or levels, of these variables will play the roles of the different treatments. Extraneous variables, if any, will be handled as before; i.e., their effect on the error variance will be removed by randomization , replication, etc.

The additional element of interest in a factorial design is the interaction between the various factors. For example, in a simple two-factor (two-variable) experiment, it might be desired to determine the effects of the flue temperature and oven width on the time required to make coke. There may, however, be an interaction between oven-width and flue temperature, and the result for one combination, say T_1W_2, may be different than that for another combination T_2W_3. In a factorial design, it is possible to determine, at a stated significance level α, if such interactions exist.

It should be pointed out that the two factors whose interaction is of interest here are "design variables" that are "built in" to the experimental design structure, rather than extraneous variables such as mild, moderate, or choppy water conditions in the hull design problem. Of course, factorial designs could also be used to determine whether there are interactions between extraneous variables, but this is not usually necessary. Usually the F-test is sufficiently powerful when the effect of differences in factor (row, column, Latin letter) means is removed from the error variance, without removing the effect of interaction.

Of course, higher order factorial designs permit the analysis of higher order interactions. For example, a three-factor factorial with replication permits one to test the significance of two-factor and three factor interactions. Usually, however, one is not concerned--from a practical standpoint--with interactions beyond three-factor interactions; and in the great majority of real-world problems, one is primarily interested in one-factor (main effects) and two-factor interactions.

1. Two-Factor Factorial Design

To illustrate the analysis of a two-factor factorial design, consider a specific example.* Suppose it is desired to learn the effects of two kinds of soil treatments on the yield of wheat. One kind of soil treatment is chemical, while the other is a "humus and fertilizer" type of treatment. There are four variations (levels) of chemical treatments: None, N + 0 (nitrogen plus oxygen), CO_2 gas, and carbonic acid. There are also four variations (levels) in the humus and fertilizer treatments: None, straw, straw + PO_4, straw + PO_4 + lime. Thus, one has a two-factor factorial design in which each factor consists of four levels. It is desired to determine whether:

(1) the chemical treatment means are significantly different

(2) the humus and fertilizer treatment means are significantly different

---

*See Ref. [15], p. 276.

(3) whether there is interaction between chemical treatment and humus and fertilizer treatment.

The effects of chemical levels or humus and fertilizer levels are called "main effects"; i.e., main effect for chemicals and main effects for fertilizer. The interaction between the two factors (chemical and fertilizer) is called a two-factor interaction or a first-order interaction.

This was a greenhouse experiment involving three replications. That is, three pots of wheat were treated with the same combination of chemical and fertilizer levels. For example, "soil plus straw" with "carbon dioxide gas" was laid down in three pots. To eliminate the possibility that position might affect the pots differentially, the 48 pots were placed at random in the greenhouse. This randomization element is an important precaution. If one places together three pots containing one combination of treatments, and in a second place those with another combination, the effects of position and treatment will be confounded (mixed or intertwined) and the three pots may be worth no more than a single determination.

The results of the experiment are shown below in Table 7. The formulas used to determine the various sums of squares are derived in any basic text on the design of experiments, and will not be derived here.

$$C = \frac{T_{...}^{2}}{N} = \frac{(711.7)^2}{48} = 10,552.44$$

$$SST = (21.4)^2 + (21.2)^2 + (20.1)^2 + (12.0)^2 + \ldots + (14.0)^2$$

$$= 367.15$$

$$SSH \text{ (humus)} = \frac{(230.2)^2 + \ldots + (164.6)^2}{12} - 10,552.44 = 306.24$$

$$SSC \text{ (chemical)} = \frac{(180.1)^2 + \ldots + (172.6)^2}{12} - 10,552.44 = 9.17$$

$$SS \text{ (Subclasses)} = \frac{(62.7)^2 + (38.3)^2 + \ldots + (40.6)^2}{3} - 10,552.44$$

$$= 340.87$$

$$SS \text{ (HC)} = \text{Interaction Sum of Squares}$$

$$= 340.87 - (306.24 - 9.17) = 25.46$$

$$SSE = 367.15 - 340.87 = 26.28$$

From the results summarized in Table 7a, one notes that the difference in humus treatment means, chemical treatment means, and interaction means are significant at the 5% level of significance* under the aforementioned assumptions upon which the F test is based. The most critical of these assumptions, the homogeneity or equality of variances, will now be subjected to an appropriate test.

*Actually, the humus and interaction treatment means are significant at the 1% level also.

Table 7.

YIELD OF WHEAT IN 48 POTS. GREENHOUSE EXPERIMENT WITH TWO SERIES OF SOIL TREATMENTS, THREE POTS FOR EACH COMBINATION

Grams

Humus and Fertilizer Treatment	Pot	Chemical Treatment				Sum, 12 Pots	Yield per Pot
		None	N + O	CO_2 Gas	Carbonic Acid		
None	1	21.4	20.9	19.6	17.6		
	2	21.2	20.3	18.8	16.6		
	3	20.1	19.8	16.4	17.5		
	Sum	62.7	61.0	54.8	51.7	230.2	19.2
Straw	1	12.0	13.6	13.0	13.3		
	2	14.2	13.3	13.7	14.0		
	3	12.1	11.6	12.0	13.9		
	Sum	38.3	38.5	38.7	41.2	156.7	13.1
Straw + PO_4	1	13.5	14.0	12.9	12.4		
	2	11.9	15.6	12.9	13.7		
	3	13.4	13.8	13.1	13.0		
	Sum	38.8	43.4	38.9	39.1	160.2	13.4
Straw + PO_4 + lime	1	12.8	14.1	14.2	12.0		
	2	13.8	13.2	13.6	14.6		
	3	13.7	15.3	13.3	14.0		
	Sum	40.3	42.6	41.1	40.6	164.6	13.7
Sum, 12 pots		180.1	185.5	173.5	172.6	711.7	
Yield per pot		15.0	15.5	14.5	14.4		

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F	$F_{.05}$	$F_{.01}$
Humus Treatments	3	306.24	102.08	124.49**	2.90	4.47
Chemical Treatments	3	9.17	3.06	3.73*	2.90	4.47
Interaction	9	25.46	2.83	3.452**	2.19	3.03
Error	32	26.28	0.82			
Total	47	367.15				

Table 7 a.

Analysis of Variance For A Two-Factor Factorial Design

2. Test For Homogeneity of Variances

One of the assumptions— in fact, by far the most important one - underlying the parametric (conventional) analysis of variance model is that of homogeneity of variance - i.e., that the distributions of the items in all classes (rows, columns, interaction cells, etc.) are normal distributions with identical variances σ^2. There are several tests for checking this assumption; e.g., tests by Hartley, Cochran, and Bartlett. Bartlett's test has the advantage that it does not require the sample size n_j in each of the treatment classes to be the same, but it is more complex than Hartley's or Cochran's tests and does require that no n_j be smaller than 3 and preferably not smaller than 5. Cochran's test is somewhat stronger (more sensitive than Hartley's), and will be used in the present analysis. The requirement in Cochran's test that the n_j be equal is not really a serious limitation, since in most analysis of variance problems the n_j values are the same. In the event that they are not the same, Bartlett's test (which is readily available [3, p. 95] can be used.

It will now be shown that the row variances are significantly different at the $\alpha = 0.05$ level, which is sufficient to (theoretically) invalidate the conventional analysis of variance (F) test, since the latter assumes that the variances are identical for all classes. The application of Cochran's C test to the row variances is basically very simple. Specifically, one evaluates the ratio of the largest of the row variances to the sum of the row variances:

For Table 7

$$\bar{X}_1 = 230.2/12 = 19.2 \qquad s_1^2 = 2.87$$

$$\bar{X}_2 = 156.7/12 = 13.1 \qquad s_2^2 = 0.72$$

$$\bar{X}_3 = 160.2/12 = 13.4 \qquad s_3^2 = 0.75$$

$$\bar{X}_4 = 164.6/12 = 13.7 \qquad s_4^2 = 0.63$$

$$\text{largest variance} = s_1^2 = 2.87$$

Thus,

$$C = \frac{2.87}{2.87 + 0.72 + 0.75 + 0.63}$$

$$= \frac{2.87}{4.97}$$

$$= 0.58.$$

Since the critical value of C [3, Table B.8, p. 654] for $k = 4$ variances and $n-1 = 12-1 = 11$ degrees of freedom for each of the row variances is $C^* = 0.4831$, one concludes at the $\alpha = 0.05$ level that the row variances are not all identical. This being the case, one would tend to rely on the distribution-free analysis of variance test rather than the conventional analysis of variance (F) test. However, in this example, both methods (see Section VII) indicate that the row (humus-fertilizer) means and column (chemical) means are significantly different at the 0.1% level of significance, so that one would not question these conclusions. However, as can also be seen from the results of Section VII, the two methods do not lead to the same conclusion relative to the two factor interaction RC. In this case, one would tend to rely on the distribution-free analysis (see Section VII), which indicates that the two-factor interaction is significant at about the $\alpha = 0.15$ level.

3. Three-Factor Factorial Design

Frequently one encounters real world problems whose analysis requires the use of factorial designs with more than two factors. While the analysis of factorial designs with more than three factors is in general quite complex (unless each factor involves only two levels), the implementation of the analysis in the general multifactor case can be deduced from a discussion of the three-factor factorial design, which is the subject of this section.

To illustrate the analysis of a multifactor factorial design, consider the following problem. To study the performance of three detergents at different temperatures, a laboratory technician performed a 2 x 2 x 3* factorial experiment with three replicates. The results are shown below; the entries are "whiteness" readings obtained with specially designed equipment. Analyze the data to determine whether:

a. The whiteness effect differs among the three detergents, when the effects of washing time and water temperature have been removed.

b. The whiteness effect differs for the various washing times, when the effects of detergent and water temperature have been removed.

c. The whiteness effect differs for the various water temperatures, when the effects of detergents and washing time have been removed.

*This means that factor A involves 2 levels; factor B, 2 levels; and factor C, three levels.

d. An interaction exists between detergent and washing time, detergent and water temperature, or washing time and temperature.

e. A three-factor interaction exists.

First one needs to apply the Cochran C test (or a similar test) to check the assumption that the variances within the various classes are identical. Thus, one must determine whether the three variances corresponding to the three classes A_1, A_2, A_3 are significantly different. These variances are obtained from samples of size 12, i.e., the 12 entries in Table 8 corresponding to A_1, the 12 entries corresponding to A_2, and the 12 corresponding to A_3. Likewise, the variances corresponding to the two classes B_1 and B_2 must be compared to see if they are significantly different. This entails comparing the variances of two random samples of sizes 18. The variances for C_1 and C_2 must be similarly compared, each variance again being based upon 18 observations. For the interaction AB, one must compare six variances (Table 8 a) each based on six observations. For the interactions AC and BC, respectively, one must compare six variances each based on six observations and four variances based on nine observations. Finally, to determine whether the homogeneity of variance condition is satisfied relative to the three-factor interaction ABC, one compares the twelve variances obtained from the twelve rows in Table 8 . Since an example illustrating this test as applied to row variances is carried out in section G, no calculations will be presented here. It will merely be stated that the homogeneity of variance assumption is satisfied for all main effects and interactions at the $\alpha = 0.05$ level.

One can now proceed with the regular analysis. First, the preliminary calculations shown below are carried out. The usual notation is employed.

A (Detergent)	B (Washing Time)	C (Water Temperature)	Rep. 1	Rep. 2	Rep. 3	Total
A_1	10	hot	76	72	73	221
A_1	10	warm	51	48	50	149
A_1	20	hot	77	74	79	230
A_1	20	warm	61	62	62	185
A_2	10	hot	63	62	60	185
A_2	10	warm	45	48	43	136
A_2	20	hot	63	64	59	186
A_2	20	warm	55	53	58	166
A_3	10	hot	64	60	63	187
A_3	10	warm	47	42	49	138
A_3	20	hot	65	66	62	193
A_3	20	warm	56	54	54	164
		Total	723	705	712	2,140

Table 8. Data for the Analysis of Detergents, Washing Time, and Water Temperature.

$T... = 76 + 51 + \ldots + 62 + 54 = 2,140$

$C = T...^2/N = (2,140)^2/36 - 127,211.11$

$SST = (76)^2 + (51)^2 + \ldots + (54)^2 - 127,211.11 = 3,305$

$SS(Rep) = 1/12 \quad (723)^2 + (705)^2 + (712)^2 \quad - 127,211.11 = 13.72$

$SS(Tr, A, B, C) = 1/3 \quad (221)^2 + (149)^2 + \ldots + (164)^2 \quad - 127,211.11$

$= 3188.22$

$SSE = SST - SS(Tr, A, B, C) - SSR$

$= 3305 - 3188.22 - 13.72 = 103.06$ (1)

The quantity T... denotes the sum of all the whiteness entries in Table 8, while C is a so-called correction factor used throughout the analysis. The symbol SST refers to a sum of squares for the entire table which is broken down into its various component sums according to the Fundamental Theorem of Analysis of Variance, namely, SST = SS(Tr, A, B, C) + SS(Rep) + SSE

= (Treatment Sum of Squares) + (Replication Sum of Squares) + (Error Sum of Squares)

The most important element in this sum is SSE, which is the denominator of each F test yet to be performed. The term SS(Rep) removes the effects of differences in the replicates, while

SS(Tr, A, B, C) = SSA + SSB + SSC + SS(AB) + SS(AC) + SS(BC) + SS(ABC) (2)

That is, SS(Tr, A, B, C) consists of the combined sums of squares for all main effects and interactions. These sums of squares will now be determined.

Tables 8 a and the calculations following constitute the basis for the analysis of the main effects A, B and the interaction AB. Similarly, the basis for analyzing C and AC is Table 8b and the subsequent calculations. The remaining two-factor interaction, AB, requires the calculations following Table 8c. Finally the quantity SS(ABC) can now be determined from equation (2).

The data necessary for carrying out the requisite F tests are shown in Table 9. The table simply converts the various sums of squares into the variances (Mean Square Values) required by the relevant F tests. Each variance or mean square value is readily determined by dividing the corresponding sum of squares by the proper degrees of freedom. The degrees of freedom value for a main effect is one less than the number of levels for that effect, whereas the number of degrees of freedom for an interaction is the product of the degrees of freedom for each component or factor entering into the interaction. Also, the mean square for replicates is one less than the number of replicates. Finally, the degrees of freedom for the

		B: $B_1 = 10$	B: $B_2 = 20$	
A	A_1	370	415	785
	A_2	321	352	673
	A_3	325	357	682
		2016	1124	2140

Table 8a. Data For Analysis of AB Interaction

$$SS(Tr, A, B) = 1/6 \ ((370)^2 + (321)^2 + \ldots + (357)^2) - 127{,}211.11$$

$$= 979.56$$

$$SSA = 1/12 \ ((785)^2 + (673)^2 + (682)^2) - 127{,}211.11$$

$$= 645.39$$

$$SSB = 1/18 \ ((1016)^2 + (1124)^2) - 127{,}211.11$$

$$= 324.00$$

$$SS(AB) = SS(Tr, A, B) - SSA - SSB$$

$$= 979.56 - 645 - 324.00$$

$$= 10.17$$

		C: C_1	C: C_2	
A	A_1	451	334	785
	A_2	371	302	673
	A_3	380	302	682
		1202	938	2140

Table 8b. Data For Analysis of AC Interaction

$$SS(Tr, A, C) = 1/6 \ ((451)^2 + (371)^2 + \ldots + (302)^2) - 127{,}211.11$$

$$SSA = 645.39$$

$$SSC = 1/12 \ ((1202)^2 + (938)^2) - 127{,}211.11$$

$$= 1936.00$$

SS(AC) = SS(Tr, A, C) - SSA - SSC

= 2689.89 - 645.39 - 1936.00

= 108.5

		C		
		C_1	C_2	
B	B_1	593	423	1016
	B_2	609	515	1124
		1202	938	2140

Table 8c. Data for Analysis of BC Interaction

$SS(Tr, B, C) = 1/9\ ((593)^2 + (609)^2 + (423)^2 + (515)^2) - 127{,}211.11$

$= 2420.45$

SSB = 324

SSC = 1936

SS(BC) = SS(Tr, B, C) - SSB - SSC

= 160.45

SS(ABC) = SS(Tr, A, B, C) - SSA - SSB - SSC - SS(AB) - SS(AC) - SS(BC)

= 3.71

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F	F.05	F.01
Replicates	2	13.72	6.86	1.47	3.44	5.72
Main Effects						
A	2	645.39	327.70	68.90**	3.44	5.72
B	1	324.00	324.00	69.23**	4.30	7.95
C	1	1936.00	1936.00	413.70**	4.30	7.95
Two Factor Interactions						
AB	2	10.17	5.08	1.09	3.44	5.72
AC	2	108.5	54.25	11.59**	3.44	5.72
BC	1	160.45	160.45	34.28**	4.30	7.45
Three Factor Interaction						
ABC	2	3.71	1.85	0.39	3.44	5.72
Error	22	103.06	4.68			
Total	35	3305				

Table 9. Summary of Analysis of Variance Data

(All main effects as well as the AC and BC interactions are significant at the $\alpha = 0.01$ level. Results which are significant at the $\alpha = 0.01$ level are indicated by a double asterisk; those significant at the $\alpha = 0.05$ level, by a single asterisk.)

total effect is N - 1 = 36 - 1 = 35, from which it follows that the degrees of freedom for the "error" or random effect is 35 - 2 - 2 - 1 - 1 - 2 - 2 - 1 - 2 = 22. This latter fact follows from the fact that the effects are "additive," as are also their degrees of freedom.

To determine the significance of any element, one evaluates the F ratio obtained by dividing the mean square for the element by the error mean square (MSE: in this case, MSE = 4.68). If the ratio so computed is greater than the proper entry in the F table, the element is significant. Any such element is, of course, a main effect, interaction, or replicate.

Thus, to test the main effect A, one notes that

$$F_A = \frac{327.70}{4.68} = 68.90 \qquad (3)$$

Since $F_A > F_{.01}$, where $F_{.01}$ is the entry in the F-table found in the cell determined by the column headed 2 and the row headed 22, the main effect A is significant at the 1% level. The column heading is always equal to the degrees of freedom associated with the variance in the numerator of the F ratio - i.e., the variance being tested. The row heading corresponds to the variance in the denominator of the F ratio, namely, the "error" or random variance.

The F test is based on a comparison of the ratio of the variance between class means to a strictly random (error) variance. If the ratio is significantly larger than one, it is because the variance between class means is significantly larger than the random variance in the denominator. This will be so only when the class means are not all identical, since then the variance between class means contains not only a random variance component but also a variance component due to the separation of the class means. A class mean could, of course, be a "main effect" mean (e.g., a column mean) or an interaction mean.

Clearly, from an examination of the results in Table 9, one concludes that the two-factor interaction AB and the three-factor interaction ABC are not significant at the 5% level but that all the main effects and the interactions AC and BC are highly significant (i.e., at better than the 0.1% level). In layman's language, it means that the detergents differ in the degree of whiteness which they impart to clothing. Furthermore, the degree of whiteness imparted on the average by these detergents depends upon what temperature of water is used. It also depends upon the washing time utilized. Furthermore, the average degree of whiteness, averaged over all detergents for a given combination (BC) of washing time and water temperature, varies over the different combinations of B and C.

Finally, the question remains as to whether <u>all</u> <u>three</u> detergents are significantly different. The multiple comparisons test below indicates that A_1 is different from both A_2 and A_3, but A_2 and A_3 are equivalent.

Detergent	A_2	A_3	A_1
Mean ($\bar{a}_i$)	56.08	56.83	65.42

Thus, from the table "Critical Values for Duncan's New Multiple Range Test" corresponding to $\alpha = 0.05$ and 22 degrees of freedom (see Reference [4], one has (by linear interpolation)

p	2	3
r_p	2.935	3.085

Multiplying each value of r_p by $\sigma_{\bar{x}} = 0.6245$ gives

p	2	3
R_p	1.833	1.927

Hence, since

$$\bar{a}_1 - \bar{a}_2 = 65.42 - 56.08 = 9.34 > 1.927,$$

the detergents are not all equivalent, as has already been seen from the F test. Furthermore, since

$$\bar{a}_1 - \bar{a}_3 = 65.42 - 56.83 = 8.59 > 1.833,$$

detergent A_1 is significantly different from A_3 insofar as whiteness effect is concerned. But since

$$\bar{a}_3 - \bar{a}_2 = 56.83 - 56.08 = 0.75 < 1.833$$

detergents A_2 and A_3 are equivalent. Consequently, detergent A_1 is preferable over A_2 and A_3 insofar as whiteness effect is concerned; i.e., if whiteness effect is the only criterion. Insofar as a second choice is concerned, it makes no difference whether A_2 or A_3 is chosen when whiteness effect is the only criterion.

4. 2^N Factorial Designs

A special case of the general multi-factor factorial design is the particular case in which each factor occurs at two levels: the so-called 2^N factorial design. In the first place, the number of experimental conditions in a factorial experiment increases multiplicatively with the number of levels of each factor. Hence, if many factors are to be investigated simultaneously, it may not be economically feasible to include more than two levels of each factor. A second reason for treating 2^N factorial experiments separately is that there exist computational short-cuts which apply only to this case. This design is particularly useful in exploratory or screening experiments, in which one considers what are intuitively (or otherwise) believed to be the "best" and "worst" levels of each factor. Fewer factors (ordinarily those which are found to be significant) may then be examined at more than two levels.

A 2^N factorial experiment requires 2^N experimental conditions; since their number can be fairly large, it is convenient to represent these experimental conditions by a special notation and to list them in a so-called special order. With this notation, each experimental condition is represented by the product of lower case letters corresponding to the factors which are taken at level 1, called the higher level. If a lower-case letter corresponding to a factor is missing, this indicates that the factor is taken at level 0, referred to as the lower level. For example, in a three-factor experiment, ab represents the experimental condition in which factors A and B are taken at the higher levels and factor C is taken at the lower level. Similarly c denotes the experimental condition in which factor C is at the higher level and factors A and B are at the lower levels. Finally, the symbol "1" represents the experimental condition in which all factors are at the lower level.

The model is, of course, still a linear model; i.e., each experimental condition is expressed as a linear function of the main effects, interactions, replication effect, and random errors. For example, the model for a 3-factor experiment is represented by the linear function

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \rho_\ell + \epsilon_{ijk\ell}$$

$i = 0, 1$; $j = 0, 1$; $k = 0, 1$; $\ell = 1, 2, \ldots, r$; where the Greek letters denote the main effects, $(\alpha\beta)_{ij}, \ldots, (\beta\gamma)_{jk}$ denote the two-factor interactions, $(\alpha\beta\gamma)_{ijk}$ represents the three-factor interaction, and ρ_ℓ denotes the replication effect. The term ϵ_{ijk} represents the effect due to random, uncontrollable factors, and is assumed to be normally distributed with mean zero. Thus, Y_{ijkl} represents the experimental condition corresponding to $A_i B_j C_k$ in a given replicate, $i = 0, 1$; $j = 0, 1$; $k = 0, 1$; and which, when expressed in terms of the new symbolism, becomes $a^i b^j c^k$, where the i, j, and k may be regarded as mathematical exponents with the usual interpretation $a^o = 1$, $a^1 = a$, etc.

A 2^N factorial experiment can be analyzed as a multi-factor factorial design in the manner previously explained. However, Yates [5] has devised a method which is considerably simpler and easier to expedite, as will presently become evident.

To illustrate the analysis of a 2^N factorial design, consider the following example.

Example

A taste-testing experiment was performed to discover what effect, if any, the physical properties of a certain food have on its taste. The results, expressed as ratings by a judge on a scale from 1 to 10, are given in the following table:

	A Color	B Consistency	C Texture	Ratings Rep. 1	Rep. 2	Total
(1)	light	light	fine	8	6	14
a	dark	light	fine	7	6	13
b	light	heavy	fine	9	9	18
ab	dark	heavy	fine	8	9	17
c	light	light	coarse	7	8	15
ac	dark	light	coarse	8	6	14
bc	light	heavy	coarse	2	1	3
abc	dark	heavy	coarse	3	2	5
			Total	52	47	99

Table 10. Treatment Combinations in Standard Form

Determine whether color, consistency, or texture has a significant effect on taste, and if so, if there are any interactions between these factors.

In order to use Yates' method, one must first arrange the data in so-called standard order, as in Table 10. This order begins with 1, followed by a; the next two entries are obtained by multiplying 1 and a by b in that order. These are then multiplied, respectively, by c which gives the next four entries, which completes the standard order arrangement.

Yates' method begins by operating on the values in the "Treatment Totals" column, adding and subtracting pairwise. Thus, the first four entries in Column (1) are, respectively, the pair sums 14 + 13 = 27 , 18 + 17 = 35, 15 + 14 = 29, and 3 + 5 = 8. The remaining entries in Column (1) are the corresponding pair differences, namely, 13 - 14 = -1, 17 - 18 = 1, 14 - 15 = -1, and 5 - 3 = 2. (Table 10 a)

	Total	(1)	(2)	(3)	(4)=(3) /16 (S.S.)	(5) (M.S.)
1	14	27	62	99	612.56	
a	13	35	37	-1	0.06	0.06
b	18	29	-2	-13	10.56	10.56
ab	17	8	1	3	0.57	0.57
c	15	-1	8	-25	39.06	39.06
ac	14	-1	-21	3	0.56	0.56
bc	3	-1	0	-29	52.56	52.56
abc	5	2	3	3	0.57	0.57

Table 10 a. Calculation of M.S. Values by Yates' Method

For a 2^N factorial design, this procedure is continued until N columns have been generated (in this case, 3 columns). The mean square column values are obtained by squaring column (3) and dividing by $r \cdot 2^N$, where r denotes the number of replicates. Thus, the mean square for A is $(-1)^2/2(8) = \frac{1}{16} = 0.06$. The first entry in the mean square column, corresponding to I (which stands for the grand total) is $T...^2/r2^N = \frac{99^2}{16} = 612.56$. Actually, the first entry in column (3) provides on the computation to that point, inasmuch as it must equal the grand total (99). The sum of squares for replication and the error sum of squares cannot be obtained by Yates' method, but are necessarily determined from the preliminary analysis shown below.

$$T... = 99;\ C = \frac{T...^2}{16} = \frac{9801}{16} = 612.56$$

$$SST = (8)^2 + (7)^2 + \ldots + (9)^2 + (2)^2 - 612.56 = 110.44$$

$$SS(Tr) = 1/2\ \ (14)^2 + (15)^2 + \ldots + (17)^2 + (5)^2\ \ - 612.56 = 103.94$$

$$SS(Rep) = 1/8\ \ (52)^2 + (47)^2\ \ = 612.56 = 1.56$$

$$SSE = SST - SS(Tr) - SS(Rep)$$

$$= 110.44 - 103.92 - 1.56$$

$$= 4.94$$

These results, coupled with Yates' analyses, lead to the summary and conclusions shown in Table (11).

Source of Variation	Degrees of Freedom	S.S.	M.S.	F	$F_{.05}$	$F_{.01}$
Replicates	1	1.56	1.56	2.23	5.59	12.20
Main Effects						
A	1	0.06	0.06	0.09	"	"
B	1	10.56	10.56	15.09**	"	"
C	1	39.06	39.06	55.80**	"	"
Interactions						
AB	1	0.57	0.57	0.81	"	"
AC	1	0.56	0.56	0.80	"	"
BC	1	52.56	52.56	75.09**	"	"
ABC	1	0.57	0.57	0.81	"	"
Error	7	4.94	0.70			
Total	15	110.44				

Table 11. Analysis of Variance Summary

Thus, the main effects B, C and the interaction BC are significant at both the $\alpha = 0.05$ and $\alpha = 0.01$ levels of significance. In other words the consistency and texture of food has a significant effect on its taste, but the color does not. The replication factor is not really of interest insofar as its significance or nonsignificance is concerned. In this case it means that one could consider the two replicates as identical, and hence one could pool the SS(Rep) with the SSE to obtain a new SSE value with more degrees of freedom, which would strengthen the test somewhat (very little in this particular problem). This would not ordinarily be done unless certain main effects and/or interactions were "borderline"; i.e., just short (mathematically) of being significant.

G. Confounding in a 2^N Factorial Experiment

In some experiments, it is impossible to run all the required treatment combinations in one block. For instance, if a 2^3 factorial experiment involves eight combinations of paint pigments that are to be applied to a surface and baked in an oven which can accommodate only four specimens, the eight treatments must be divided into two blocks (ovens) in each replicate. Or, if six brands of tires are to be tested, only four could be tried on a given car. Again, such a block (car) would be incomplete, having only four of the six "treatments" in it.

Returning to the 2^3 factorial experiment mentioned above, involving eight combinations of paint pigments and two ovens, the question arises as to which treatments (combinations of paint pigments) should be included in the first oven and which in the second. The question is not merely academic, since when the experimental conditions are distributed over several blocks, one or more of the effects may become confounded (inseparable) with possible block effects, i.e., between-block differences. For example, if the treatment combinations a, ab, ac, and abc are included in one oven (Block 1) and (1), b, c, and bc are included in a second oven (Block 2), then the block effect, i.e., the difference between the two block totals, is given by

$$(a) + (ab) + (ac) + (abc) - ((1) + (b) + (c) + (bc)). \qquad (4)$$

But (4) also measures the main effect of factor A, so that the main effect of factor A is confounded with blocks. That is, for each treatment combination in Block 1 with A at the high level there is in Block 2 the corresponding treatment combination with A at the low level; namely, a vs (1); ab vs b; ac vs c; and abc vs bc. Thus, the difference in (4) is a measure of the main effect A if there is no difference in the two ovens. If there is a difference in the two ovens, it will be reflected in the difference (4). Hence, if the usual analysis of variance for a randomized block is carried out, the difference in (4) will probably show up as significant. However, one does not really know whether or not this is really the case, since (4) measures both the "block difference," and the main effect of factor A. The way to avoid "confounding" the main effect of A with blocks is to redistribute the treatment combinations among the two ovens in such a way (and there are several possible ways) that the "high levels" and "low levels" of factor A are "balanced" in the two blocks. One possibility is to place a, ab, c, bc in Block 1 and ac, abc, (1), and b in Block 2. Then the effect of any difference between blocks will affect the high level of factor A in the same manner as the low level, and the block effect on A will be separated out. To see why this is so, it will be helpful to note how each main effect and interaction is measured in terms of treatment effect totals. This is most easily seen by constructing a table (Table 12) of signs for a 2^3 factorial design which can be implemented in only one block and in which, therefore, there is no confounding. The sign in a given column, say column A, is plus or minus, according as the treatment combination in the corresponding row does or does not contain the letter a. The signs in the various interaction columns, say AB, are obtained by multiplying the signs in columns A and B; or more generally, by multiplying the signs in the columns corresponding to the component factors. Thus, the signs in the interaction column ABC are obtained by multiplying the corresponding signs in columns A, B, and C.

Treatment Combinations	I	A	B	AB	C	AC	BC	ABC
(1)	1	-1	-1	1	-1	1	1	-1
a	1	1	-1	-1	-1	-1	1	1
b	1	-1	1	-1	-1	1	-1	1
ab	1	1	1	1	-1	-1	-1	-1
c	1	-1	-1	1	1	-1	-1	1
ac	1	1	-1	-1	1	1	-1	-1
bc	1	-1	1	-1	1	-1	1	-1
abc	1	1	1	1	1	1	1	1

Table 12. Calculation of Effect Totals for a Complete Factorial

To determine the value of the total response for the main effects of factor A, one simply multiplies the response value for each treatment combination by the corresponding entry in column A, takes the algebraic sum, and divides the value of this sum by -8r, where r denotes the number of replicates. Thus,

$$A = \frac{-(1) + a - b + ab - c + ac - bc + abc}{8r} . \quad (5)$$

Similarly,

$$B = \frac{-(1) -a + b + ab - c - ac + bc + abc}{8r}$$

For the interactions AB and ABC, one has, respectively, from Table 12 :

$$AB = \frac{(1) - a - b + ab + c - ac - bc + abc}{8r}$$

$$ABC = \frac{-(1) + a + b - ab + c - ac - bc + abc}{8r}$$

Returning to the above example involving two ovens (blocks), one notes that if the treatment combinations a, ab, ac, and abc are placed in Oven #1 and (1), b, c, and bc in Oven #2, then any bias effect from Oven #1 will always be added and any bias effect from Oven #2 will always be subtracted in the analysis of the main effects of factor A, as is seen from (5). If, however, the treatment combinations (1), ab, ac, bc were placed in one block, say Block #1, and a, b, c, abc in the other, then any bias effect on the main effects of A due to either block would be removed, since they are both added and subtracted and hence would cancel out.

The end result is always that something must be sacrificed when all treatment combinations cannot be placed in the same block. If the experiment involves two blocks, then precisely one main effect or interaction must be sacrificed. If four blocks are involved, precisely two or more effects must be sacrificed. The following illustration indicates how one determines what distribution of treatment combinations will sacrifice specified effects, and what additional effects are also sacrificed.

Suppose in a 2^4 factorial design it is necessary to carry out the experiment in four blocks. That is, the 16 treatment combinations (1), a, b, abc, ac, bc, abc, d, ad, bd, abd, cd, acd, bcd, abcd must be placed in four blocks, each containing four treatment combinations. Suppose further that one is willing to sacrifice the four-factor interaction ABCD and the two-factor interaction AD. How must the 16 treatment combinations be distributed among the four blocks?

One procedure for determining this distribution is the so-called even-odd rule. Applying this rule to accomplish the confounding of ABCD with blocks, one merely places each treatment combination which has an even number of letters in common with ABCD into one block, and those having an odd number in common with ABCD into the other block. This procedure then results in the following distribution of the 16 treatment combinations among two blocks.

Block 1: (1), ab, ac, bc, ad, bd, cd, abcd

Block 2: a, b, c, abc, d, abd, acd, bcd,

If now one applies the even-odd rule to each of these two blocks relative to the interaction AD, both Block 1 and Block 2 are subdivided into two blocks, resulting in the four blocks shown below.

Block 1: (1), bc, ad, abcd

Block 2: ab, ac, bd, cd

Block 3: b, c, abd, acd

Block 4: a, abc, d, bcd

The effect (ABCD) (AD) = A^2BCD^2 = BC is also confounded with blocks. That is, if two effects, such as ABCD and AD, are confounded with blocks, then their product, reduced mod 2, is also confounded. (When one is working on the mod 2 basis, only the numbers 0 and 1 are involved, so that any number (e.g., an exponent) reduces to either 0 or 1. That is, 0 = 0, 1 = 1, 2 = 0, 3 = 1, 4 = 0, 5 = 1, etc. Thus, if one carries out a 2^4 factorial analysis using the above four blocks, he can test all main effects and interactions except the interactions AD, BC, and ABCD; and he can do this if each treatment combination is used twice or more. It should be pointed out that replicating the entire experiment will in no way retrieve the information lost by confounding; it will only enable one to measure and test the significance of the unconfounded effects more accurately.

To illustrate how the need may arise to analyze a problem in which confounding occurs, consider the previous example involving a taste-testing experiment. In this experiment, suppose that the judge rates foods in sets of four, with a rest period in between, and that the experiment was actually performed so that each replicate consisted of two blocks with ABC confounded. The two blocks so obtained (i.e., by confounding ABC with blocks) are readily found, by application of the even-odd rule, to be (1), ab, bc, ac and a, b, c, abc. The problem now is to test the main effects and two-factor interactions, the three-factor interaction being unmeasurable because it is confounded with blocks. This kind of analysis is called an intrablock analysis of the data.

The main effects and two-factor interactions may be determined by a Yates' analysis just as in the original situation when there was no confounding with blocks. However, instead of a sum of squares for replicates, one now has a sum of squares for blocks. That is, one can no longer separate out the effects of replication, since ABC is confounded with blocks (and hence also with replicates). To find the sum of squares for blocks (SS (Bl)), one first constructs the tables below and then calculates the SS(Bl) as shown. It has already been shown that ABC will be confounded with blocks if a, b, c, abc are placed in one block and (1), ab, bc, ac are placed in the other block. The complete experiment for both replicates is again shown below for the convenience of the reader.

Treatment Combination	Rep. 1	Rep. 2	Total
(1)	8	6	14
a	7	6	13
b	9	9	18
ab	8	9	17
c	7	8	15
ac	8	6	14
bc	2	1	3
abc	3	2	5
Totals	52	47	99

	Block 1				Block 2		
Treatment Combination	Rep. 1	Rep. 2	Total	Treatment Combination	Rep. 1	Rep. 2	Total
a	7	6	13	(1)	8	6	14
b	9	9	18	ab	8	9	17
c	7	8	15	bc	2	1	3
abc	3	2	5	ac	8	6	14
	26	25	51		26	22	48

	Block 1	Block 2
Rep. 1	26	26
Rep. 2	25	22

$$SS(B1) = \frac{26^2 + 25^2 + 26^2 + 22^2}{4} - 612.56 = 2.69$$

Combining this result with those in the Yates' table previously presented, one obtains the Analysis of Variance table below.

Source of Variation	Degrees of Freedom	S.S.	M.S.	F	$F_{.05}$	$F_{.01}$
Blocks	3	2.69	0.90	1.23	4.76	9.78
Main Effects						
A	1	0.06	0.06	0.08	5.99	13.70
B	1	10.55	10.55	14.45**	"	"
C	1	39.06	39.06	53.51**	"	"
Unconfounded Interactions						
AB	1	0.56	0.56	0.77	"	"
AC	1	0.56	0.56	0.77	"	"
BC	1	52.56	52.56	72**	"	"
Intrablock Error	6	4.39	0.73			
Total	15	110.43				

Clearly, the main effects of factors B and C are significant at the $\alpha = 0.01$ level, as is also the two-factor interaction BC.

VII. DISTRIBUTION-FREE ANALYSIS OF VARIANCE

As was previously mentioned, a distribution-free analysis of variance model removes essentially all of the restrictive assumptions of the parametric or conventional analysis of variance tests. There are several such distribution-free tests, but the simplest and best of these, from an overall point of view for testing main effects and interactions of any order, is Wilson's Distribution-Free Test of Analysis of Variance Hypotheses [6]. Wilson's method essentially tests the significance of differences among a set of class means (say column means) on the basis of a Chi-square test involving the number of items in each class which are above the overall median for the entire experiment. A stepwise description of the test is given below.

Description of the Test

1. The median value, Md, for the entire set of n observations is determined. This median should not be interpolated but should be determined only as a "boundary" which divides the entire set of observations, as nearly as possible, into two groups of equal size. In the notation used here, n_a represents the number of observations greater than or equal to Md and n_b represents the number of observations less than the Md.

2. A 2 x r x c contingency table is set up where r is the number of row conditions in the experimental design and c is the number of column conditions. The third dimension in this table corresponds to the division of the scores by Md. Thus, the frequency ${}_b f_{ij}$ represents the number of observations less than Md for the cell in row i and column j, and ${}_a f_{ij}$ represents the number of

observations in this cell which are greater than or equal to Md. Obviously,

$$n_a = \sum_i \sum_j {}_a f_{ij}, \qquad n_b = \sum_i \sum_j {}_b f_{ij}.$$

3. The total Chi-square, χ^2_T, can be computed from formula (6) below without restrictions on n_a, n_b, and n_{ij}:

$$\chi^2_T = \sum_i \sum_j \left[\frac{({}_a f_{ij} - n_{ij} n_a/n)^2}{(n_{ij}\, n_a)/n} + \frac{({}_b f_{ij} - n_{ij} n_b/n)^2}{(n_{ij} n_b)/n} \right]^2 \qquad (6)$$

In this formula for χ^2_T, the expected frequencies (i.e., the terms $n_{ij}n_a/n$ and $n_{ij}n_b/n$) are predicated on the null hypothesis that the main effects and interactions effects produce no change in the distribution of scores. According to this hypothesis, one should expect that for each cell the proportion of the n_{ij} scores below Md would be $n_{ij}n_b/n$, and the proportion above Md would be $n_{ij}n_a/n$. In all cases, χ^2_T has (rc-1) degrees of freedom.

4. The Chi-square values of the row effects, χ^2_R, and the column effects, χ^2_C, are computed using the marginal totals of the 2 x n x c contingency table. Formulas (7a) and 7b) given below provide these Chi-square values.

$$\chi^2_R = \sum_i \left[\frac{({}_a f_{i.} - n_{i.} n_a/n)^2}{n_{i.} n_a/n} + \frac{({}_b f_{i.} - n_{i.} n_b/n)^2}{n_{i.} n_b/n} \right] \qquad (7a)$$

where $n_{i.} = \sum_j n_{ij}$.

$$\chi^2_C = \sum_j \left[\frac{({}_a f_{i.} - n_{.j} n_a/n)^2}{n_{.j} n_a/n} + \frac{{}_b f_{.j} - n_{.j} n_b/n)^2}{n_{.j} n_b/n} \right] \qquad (7b)$$

where $n_{.j} = \sum_i n_{ij}$.

In both of the formulas for χ^2_R and χ^2_C, the expected frequencies for the main effects are obtained for the null hypothesis that the distributions of scores are identical for all levels of the raw or column conditions. Thus, for χ^2_R, one should expect the proportion n_b/n of each $n_{i.}$ to be below Md and the proportion n_a/n to be above it. In all cases, χ^2_R and χ^2_C have (r-1) and (C-1) degrees of freedom, respectively.

5. The Chi-square value for the interaction effect, χ^2_I, is most easily computed by subtraction. That is,

$$\chi^2_I = \chi^2_T - \chi^2_R - \chi^2_c , \tag{8}$$

where χ^2_I has (r-1) (c-1) degrees of freedom.

6. The tests for the main effects and interaction are made by comparing the obtained values of χ^2_R, χ^2_C, and χ^2_I with values from the cumulative Chi-square distribution for the appropriate degrees of freedom and significance level .

If there are three factors, say R, B, C involved in the experiment the three-factor interaction may also be tested for significance. The 2 x r x b, 2 x r x c, and 2 x b x c tables may be analyzed as previously indicated to determine the Chi-square values χ^2_R, χ^2_B, χ^2_C, ${}_{RB}\chi^2_T$, ${}_{RC}\chi^2_T$, ${}_{BC}\chi^2_T$, ${}_{RB}\chi^2_I$, ${}_{RC}\chi^2_I$, and ${}_{BC}\chi^2_I$, from which the various main effects and two-factor interactions may be tested for significance. (The notation ${}_{RB}\chi^2_T$ and ${}_{RB}\chi^2_I$ denotes the total χ^2 and interaction χ^2 values for the 2 x r x b table. The symbols ${}_{RC}\chi^2_T$, ${}_{RC}\chi^2_I$, ${}_{BC}\chi^2_T$, and ${}_{BC}\chi^2_I$ have similar interpretations and are obtained from the 2 x r x c and 2 x b x c tables. Specifically,

$${}_{RB}\chi^2_I = {}_{RB}\chi^2_T - \chi^2_R - \chi^2_B$$

$${}_{RC}\chi^2_I = {}_{RC}\chi^2_T - \chi^2_R - \chi^2_C$$

$${}_{BC}\chi^2_I = {}_{BC}\chi^2_T - \chi^2_B - \chi^2_C$$

$$\chi^2_T = \sum_i \sum_j \sum_k \left[\frac{({}_a f_{ijk} - n_{ijk} n_a/n)^2}{n_{ijk} n_a/n} + \frac{({}_b f_{ijk} - n_{ijk} n_b/n)^2}{n_{ijk} n_b/n} \right]$$

$$\chi^2_I = \chi^2_T - \chi^2_R - \chi^2_C - \chi^2_B$$

$${}_{RBC}\chi_I = \chi^2_I - {}_{BC}\chi^2_I - {}_{RB}\chi^2_I - {}_{RC}\chi^2_I$$

The interaction ${}_{RBC}\chi^2_I$ has (r-1) (b-1) (c-1) degrees of freedom.

A word of caution is in order. For when the expected values are small (less than three), the Chi-square test may be unreliable unless the contingency table has 30 or more degrees of freedom. In the following example the expected values involved in testing both fertilizers (rows) and chemicals (columns) are 5.5 and 6.5, which is a satisfactory situation. Also in this example, the expected values relevant to the interaction test are 11/8 and 3/8, which are a bit small and call for caution in interpreting the interaction as highly significant, since the contingency table has less than 30 degrees of freedom.

Example

To illustrate the implementation of the methods, a distribution-free analysis of variance will be carried out for the data relative to the two-factor factorial design in Table 7. Arranging the items in increasing order of magnitude (Table 13), one finds the median to be Md=13.8. Tables 15 and 16 constitute, respectively, the 2 x c and the 2 x r tables from which the χ^2 values for the main effects for chemicals and humus fertilizer, respectively, are calculated. The total Chi-square value, χ^2_T, is calculated from Table 14 using equation (6).

SAMPLE ASCENDING ORDER	
11.6	13.8
11.9	13.8
12.0	13.9
12.0	14.0
12.0	14.0
12.1	14.0
12.4	14.1
12.8	14.2
12.9	14.2
12.9	14.6
13.0	15.3
13.0	15.6
13.1	16.4
13.2	16.6
13.3	17.5
13.3	17.6
13.3	18.8
13.4	19.6
13.5	19.8
13.6	20.1
13.6	20.8
13.7	20.9
13.7	21.2
13.7	21.4

Md = 13.8

n_a = number of observations equal to or greater than the median

n_b = number of observations less than the median

= 22

Table 13. Determination of the Median of the Data in Table 3.

Humus and Fertilizer Treatment		Chemical Treatment: None	N+O	CO_2 Gas	Carbonic Acid
NONE	${}_af_{ij}$	3	3	3	3
	${}_bf_{ij}$	0	0	0	0
STRAW	${}_af_{ij}$	1	0	1	2
	${}_bf_{ij}$	2	3	2	1
STRAW + PO_4	${}_af_{ij}$	0	3	1	1
	${}_bf_{ij}$	3	0	2	2
STRAW + PO_4 + LIME	${}_af_{ij}$	1	2	1	2
	${}_bf_{ij}$	2	1	2	1

Table 14. Data for Interaction Analysis.

	None	N+ O	CO_2 Gas	Carbonic Acid	Total
$n_{a.j}$	5	8	6	8	27
$n_{b.j}$	7	4	6	4	21
$n_{.j}$	12	12	12	12	48

Table 15. Analysis of Chemical Means (Main Effect)

Using formula (7b) and the data in Table 15, one has

$$\chi^2_c = \frac{(5-6.5)^2}{6.5} + \frac{(8-6.5)^2}{6.5} + \frac{(4-6.5)^2}{6.5} + \frac{(7-6.5)^2}{6.5}$$

$$+ \frac{(7-5.5)^2}{5.5} + \frac{(4-5.5)^2}{5.5} + \frac{(8-5.5)^2}{5.5} + \frac{(5-5.5)^2}{5.5}$$

$$= \frac{11}{6.5} + \frac{11}{5.5} = 1.69 + 2 = 3.69$$

From the Chi-square table, one obtains (for 3 degrees of freedom)

$$\chi^2_\alpha = 7.815, \ \alpha = 0.05$$

Since 3.69 < 7.815, the chemical means are not significantly different (or equivalently, the chemical main effect is not significant). One makes this assertion with a confidence (probability of being correct) of 0.95.

Similarly, from the data in Table 16, one has, from formula (7a),

	NONE	STRAW	STRAW + PO_4	STRAW PO_4, LIME	TOTAL
N_a	12	4	5	6	27
N_b	0	8	7	6	21
$n_{i\bullet}$	12	12	12	12	48

Table 16. Analysis of Humus-Fertilizer Main Effect

$$\chi^2_R = \frac{(12-6.5)^2}{6.5} + \frac{(3-6.5)^2}{6.5} + \frac{(3-6.5)^2}{6.5} + \frac{(6-6.5)^2}{6.5}$$

$$+ \frac{(0-5.5)^2}{5.5} + \frac{(9-5.5)^2}{5.5} + \frac{(9-5.5)^2}{5.5} + \frac{(6-5.5)^2}{5.5}$$

$$= \frac{55}{6.5} + \frac{55}{6.5} = \frac{110}{6.5} = 16.92$$

Again, from the Chi-square table (with 3 degrees of freedom) one notes that $\chi^2_\alpha = 7.815.$, $\alpha = 0.05$.

Since 16.92 > 7.815, one concludes with a confidence of 0.95, that there is a significant difference among the row (humus-fertilizer) means.

Finally, by applying equation (6) to Table14, where the reader is reminded that

$_af_{ij}$ = number of items in cell (i, j) - i.e., in $i^{\underline{th}}$ row and $j^{\underline{th}}$ column - whose value is equal to or greater than Md = 13.8

$_bf_{ij}$ = number of items in cell (i, j) whose values are less than Md = 13.8,

one obtains

$$\chi_T^2 = \frac{(3-\frac{13}{8})^2}{\frac{13}{8}} + \frac{(3-\frac{13}{8})^2}{\frac{13}{8}} + \cdots + \frac{(1-\frac{13}{8})^2}{\frac{13}{8}} + \frac{(2-\frac{13}{8})^2}{\frac{13}{8}}$$

$$+ \frac{(0-\frac{11}{8})^2}{\frac{11}{8}} + \frac{(0-\frac{11}{8})^2}{\frac{11}{8}} + \cdots \frac{(2-\frac{11}{8})^2}{\frac{11}{8}} + \frac{(1-\frac{11}{8})^2}{\frac{11}{8}}$$

$$= 14.77 + 11.48$$

$$= 26.25$$

Then

$$_{RC}\chi_I^2 = \chi_T^2 - \chi_R^2 - \chi_B^2$$

$$= 26.25 - 16.92 - 3.69$$

$$= 26.25 - 20.61$$

$$= 5.64$$

Note that, for four degrees of freedom,

$$\chi^2_{.05} = 9.488.$$

Since $5.64 < 9.488$, the interaction between humus-fertilizer and chemical treatments is not significant at the $\alpha = 0.05$ level. Or equivalently, one asserts with a confidence of 0.95, that this interaction does not exist.

As can be seen, the chemical main effect is not significant at the 5% level but the humus-fertilizer main effect is. From a rather detailed χ^2 table [16], one notes that the value $\chi_c^2 = 3.69$ for the chemical main effect is significant at about the 30% level of significance, whereas with the parametric (conventional) factorial model (Table 7a) it is significant at about the 5% level. Similarly the humus-fertilizer, main effect (R) is significant [16] at slightly better than the 0.1% level of significance, as indicated by a calculated value of $\chi_R^2 = 16.92$; but with the parametric (conventional) analysis of variance model, it is significant at a considerably lower level. As for the interaction, Wilson's distribution-free model yields a value of $\chi_{RC}^2 = 5.648$ which is significant at about the 25% level (4); this compares with significance at the 1% level on the basis of the parametric analysis of variance test.

Thus, both the parametric and distribution-free analysis of variance tests agree that the humus-fertilizer main effect is significant at the $\alpha = 0.001$ level; so that one is highly confident that the humus-fertilizer means are not all identical. If one had some doubts as to the validity of the parametric analysis of variance assumptions, he would probably conclude that the chemical main effect and the interaction were significant at about the $\alpha = 0.15$ (15%) level, since this is intermediate between the significance levels indicated by the parametric and distribution-free models.

In summary, if one had doubts concerning the assumptions underlying the parametric model and wished to make decisions at the 5% level, he would reject the hypothesis that the humus-fertilizer means are identical, but accept the hypothesis that the chemical means are identical. Similarly, he would accept the hypothesis that the interaction means are identical. In a layman's language, he would conclude that the kind of humus-fertilizer used (among the four considered) is important but the kind of chemical treatment (among the four considered) is not, while there is no interaction between the four chemicals and the four humus types considered.

Finally, the important point to emphasize--as is illustrated in this example--is that if one is confident that the assumptions underlying the parametric model are satisfied (particularly the homogeneity of variance assumption), he should reach his conclusions on the basis of the parametric analysis of variance model, which in such a case usually yields considerably stronger results. If he has no such confidence, he should abide with the conclusions from the distribution-free analysis of variance test. If the assumptions are somewhat questionable, he should probably compromise between the two results when they differ; when they agree, there is no problem.

Actually, as stated in the preceding section, the row variances are significantly different, as are the column variances. Consequently, one should abide by the conclusions reached by the distribution-free analysis of variance test, namely, that there is a significant difference among the humus-fertilizer treatments but not among the chemical treatments, and that there is no interaction between the two (at the $\alpha = 0.05$ level of significance).

VIII. INCOMPLETE DESIGNS

There are some instances in which one cannot run the complete experimental design which he originally planned. One may find, for example, that in a randomized block it may not be possible to apply all treatments in every block. Thus, if one wishes to test six brands of tires, only four can be tried on a given car (not using the spare), and such a block would be incomplete, having only four out of the six treatments in it. Similarly, one might wish to use a 6 x 4 Latin square design but find that only three treatments are possible (e.g., in one block because only three positions are available) and where there are four blocks altogether, the design is an incomplete Latin square. Such a square is called a Youden square. Again, in a 2^{10} factorial design, one may not be able to consider all 2^{10} possible treatment combinations, perhaps for cost reasons. He can usually reach completely valid and accurate conclusions by analyzing a fraction of the complete design, perhaps by considering only 2^5 of the possible 2^{10} complete factorial designs. This would be called a $\frac{1}{32}$nd fractional factorial design. In all of these types of incomplete designs, the analysis differs from that of the corresponding complete designs. The proper analyses for these three types of incomplete designs will now be discussed.

A. Balanced Incomplete Randomized Block

An incomplete block design is one in which there are more treatments than can be put in a single block. A balanced incomplete block design is an incomplete block design in which every pair of treatments occurs the same number of times in the experiment. The number of blocks necessary for balancing will depend upon the number of treatments that can be run in a single block. The following example illustrates the method of analysis for a balanced incomplete block design.

Example Data on screen color difference on a television tube measured in degrees K are to be compared for four operators. On a given day only three operators can be used in the experiment. A balanced incomplete block design gave results as follows:

	Operator				
Day (Block)	A	B	C	D	Totals
Monday	780	820	800	-	2,400
Tuesday	950	-	920	940	2,810
Wednesday	-	880	880	820	2,580
Thursday	840	780	-	820	2,440
Totals:	2,870	2,480	2,600	2,580	10,230

Carry out a complete analysis of these data and discuss the results with regard to differences between operators.

In this design, only treatments A, B, and C are run on the first day (Monday); A, C, and D on the Tuesday; B, C, and D on Wednesday; and A, B, and D on Thursday. Note that each pair of treatments, such as AC, occurs twice in the experiment. As in the case of randomized complete block designs, the order in which the three treatments (operators) are run on a given day is completely randomized.

The notation used is explained below.

b = number of blocks in the experiment (4)

t = number of treatments in the experiment (4)

k = number of treatments per block (3)

r = number of replications of a given treatment throughout the experiment (3)

N = total number of observations

$= bk$

$= tr\ (12)$

λ = number of times each pair of treatments appears together throughout the experiment

$= r\ (k-1)/(t-1)\ (\lambda = 2).$

The steps for analyzing a balanced incomplete block design are delineated below.

1. Calculate the total sum of squares as usual.

$$SS_{total} = \sum_i \sum_j X_{ij}^2 - \frac{T_{..}^2}{N}$$

$$= 66{,}900 - 27{,}075$$

$$= 39{,}825$$

2. Calculate the block sum of squares, ignoring treatments.

$$SS_{block} = \sum_{i=1}^{b} \frac{T_{i.}^2}{k} - \frac{T_{..}^2}{N}$$

$$= \frac{(2400)^2 + (2810)^2 + (2580)^2 + (2440)^2}{} - \frac{(10{,}230)^2}{12}$$

$$= \frac{26{,}266{,}100}{3} - \frac{104{,}652{,}900}{12}$$

$$= 8{,}755{,}366.667 - 8{,}721{,}075$$

$$= 34{,}291.667$$

3. Calculate treatment effects, adjusting for blocks.

$$SS_{treatments} = \frac{\sum_{j=1}^{t} Q_j^2}{K\lambda t}$$

where $Q_j = k\ t._j - \sum_i n_{ij} T_{i}.$

and where $n_{ij} = 1$ if treatment j appears in block i, and $n_{ij} = 0$ if treatment j does not appear in block i. Note that $\sum_i n_{ij}\ T_{i.}$ is merely the sum of all block totals which contain treatment j. Specifically,

$Q_1 = 3(2570) - (2400 + 2810 + 2440) = 60$

$Q_2 = 3(2480) - (2400 + 2580 + 2440) = 20$

$Q_3 = 3(2600) - (2400 + 2810 + 2580) = 10$

$Q_4 = 3(2580) - (2810 + 2580 + 2440) = -90$

$$\frac{SS_{treatments}}{K\lambda t} = \frac{60^2 + 20^2 + 10^2 + 90^2}{3(2)(4)}$$

$$= \frac{12,200}{24}$$

$$= 508.333$$

Note that $\sum_{j=1}^{t} Q_j = 0.$

4. Calculate the error sum of squares by subtraction

$$SS_{error} = SS_{total} - SS_{blocks} - SS_{treatments}$$

$$= 39825 - 34,291.667 - 508.333$$

$$= 5024.997$$

The ANOVA* results are summarized in the table below.

Source	df	SS	MS	F
Blocks (Days)	3	34,291.667	11,430.556	
Treatments (Adjusted)	3	508.333	169.444	0.1686
Error	5	5,024.997	1,005.00	
Totals	11	29,824.997		

Table 17. ANOVA Treatment Analysis for Balanced Incomplete Randomized Block

The value F = 0.1686 with 3 and 5 degrees of freedom is not significant at the 5% level, the critical value being F*= 5.41. The error degrees of freedom are determined in the usual way - by subtraction - rather than as the product of the block and treatment degrees of freedom. However, this error degrees of freedom can be regarded

*Standard abbreviation for Analysis of Variance as the product of treatment and block degrees of freedom ($3 \times 3 = 9$) if 4 is subtracted for the four missing values in the design.

In some instances it may be desirable to test for block effects as well as treatment effects. In Table 17, the mean square for blocks was not computed, since it had not been adjusted for treatments. When one has a symmetrical balanced incomplete randomized block design, where $b = t$, the block sum of squares may be adjusted in the same manner as the treatment sums of squares

$$Q_i' = rT_{i.} - \sum_j n_{ij}\, T_{.j} \quad .$$

Specifically, for the problem at hand:

$$Q_1' = 3(2400) - 7650 = -450$$

$$Q_2' = 3(2810) - 7750 = 680$$

$$Q_3' = 3(2580) - 7660 = 80$$

$$Q_4' = 3(2440) - 7630 = \underline{-310}$$

$$\sum_i Q_i' = 0$$

$$SS_{blocks} = \sum_{i=1}^{b} (Q_i')^2 / r\lambda b$$

$$= \frac{(-450)^2 + (680)^2 + (80)^2 + (-310)^2}{3(2)(4)}$$

$$= \frac{767{,}400}{24}$$

$$= 31{,}975$$

$$SS_{treatments\ (unadj.)} = \frac{(2570)^2 + (2480)^2 + (2600)^2 + (2580)^2}{3} - \frac{(10{,}230)^2}{12}$$

$$= \frac{26{,}171{,}700}{3} - 8{,}721{,}075$$

$$= 8{,}723{,}900 - 8{,}721{,}075$$

$$= 2825$$

The results of this adjustment, together with the previous results for treatments, are summarized in Table 18.

Source	d.f.	SS	MS	F
Blocks (adjusted)	3	31,975	10,658.333	10.605
Blocks	(3)	(34,291.667)		
Treatments (adjusted)	3	508.333	169.444	0.1686
Treatments	(3)	(2,825)		
Errors	5	5,024.997	1,005	
Totals	11	39,824.997		

Table 18. ANOVA Treatment and Block Analysis for Balanced Incomplete Randomized Block

The terms in parentheses are so designated merely to show how the error term was computed for one adjusted effect and one unadjusted effect:

$$SS_{error} = SS_{total} - SS_{treatment\ (adj.)} - SS_{blocks}$$
$$= 39,824.997 - 508.333 - 34,291.667$$
$$= 5024.997$$

or equivalently,

$$SS_{error} = SS_{total} - SS_{treatments} - SS_{blocks\ (adj.)}$$
$$= 39,824.997 - 2825 - 31,975$$
$$= 5024.997$$

It bears stating that the final sum of squares values used in Table 18 to get the mean square values do not add up to the total sum of squares, as is the case with any nonorthogonal design.

B. <u>Youden Squares</u>

A special case of a balanced incomplete randomized block is a Latin square with one column missing. For example, when the conditions for a Latin square of order 4 are met except for the fact only three treatments are possible (e.g., in one block because only three positions are available) and where there are four blocks altogether, the design is an incomplete Latin square, called a Youden square. The nature and analysis of a Youden square is illustrated in the following example.

Example. In a research study at Purdue University [7] on a metal-removal rate, five electrode shapes were studied, shapes A, B, C, D, and E. The removal was accomplished by an electric discharge between the electrode and the material being cut. For this experiment four holes were cut in five work pieces, and the order of electrodes was arranged so that only one electrode shape was used in the same position on each of the five workpieces. The times in hours necessary to cut the holes are recorded in the table below. Analyze the data for this design, which constitutes a Youden square.

(It might be pointed out that it would be desirable to cut five holes in each workpiece, thereby achieving a (complete) Latin square. However, if there is only enough room on each strip (workpiece) for four holes, it becomes necessary to use a Youden square design.)

POSITIONS ON STRIP

		I	II	III	IV	Totals
	1	A(1.5)	B(0.1)	C(0.5)	D(1.5)	3.6
	2	E(0.6)	A(1.3)	B(0.1)	C(0.5)	2.5
STRIPS	3	D(0.9)	E(0.6)	A(1.5)	B(0.7)	3.7
	4	C(0.5)	D(0.9)	E(1.0)	A(1.3)	3.7
	5	B(0.1)	C(0.3)	D(1.7)	E(1.2)	3.3
Totals		3.6	3.2	4.8	5.2	16.8

The analysis of a Youden square is completely similar to that of an incomplete randomized block. Specifically,

$t = b = 5$

$r = k = 4$

$\lambda = 3$

Treatment totals of A, B, C, D, and E are, respectively, 5.6, 1.0, 1.8, 5.0, and 3.4. Then

$$SS_{total} = \sum_{ijkl}\Sigma\Sigma\Sigma X^2_{ijkl} - \frac{T^2_{....}}{N}$$

$$= 19.16 - \frac{(16.8)^2}{20}$$

$$= 5.048$$

Position effect may first be ignored, since every position occurs once and only once with each treatment, so that positions are orthogonal to blocks and treatments. Thus,

$$SS_{block\ (ignoring\ treatments)} = \frac{3.6^2 + 2.5^2 + 3.7^2 + 3.7^2 + 3.3^2}{4} - \frac{(16.8)^2}{20}$$

$$= \frac{57.48}{4} - \frac{282.24}{20}$$

$$= 14.37 - 14.112$$

$$= 0.258$$

For treatment sum of squares adjusted for blocks one obtains

$$Q_A = 4(5.6) - (3.6 + 2.5 + 3.7 + 3.7) = 8.9$$

$$Q_B = 4(1.0) - (3.6 + 2.5 + 3.7 + 3.3) = -9.1$$

$$Q_C = 4(1.8) - (3.6 + 2.5 + 3.7 + 3.3) = -5.9$$

$$Q_D = 4(5.0) - (3.6 + 3.7 + 3.7 + 3.3) = 5.7$$

$$Q_E = 4(3.4) - (2.5 + 3.7 + 3.7 + 3.3) = 0.4$$

$$\sum_i Q_i = 0$$

$$SS_{treatments} = \frac{8.9^2 + (09.1)^2 + (05.9)^2 + 5.7^2 + 0.4^2}{4(3 \times 5)}$$

$$= \frac{229.48}{60}$$

$$= 3.8247$$

$$SS_{positions} = \frac{(3.6)^2 + (3.2)^2 + (4.8)^2 + (5.2)^2}{5} - \frac{(16.8)^2}{20}$$

$$= \frac{73.28}{5} - 14.112$$

$$= 14.656 - 14.112$$

$$= 0.544$$

$$SS_{error} = SS_{total} - SS_{block} - SS_{treatment\ (adj.)} - SS_{position}$$

$$= 5.048 - 0.258 - 3.8247 - 0.544$$

$$= 0.4213$$

The results of the analysis are summarized in Table 19.

Source	d.f.	SS	MS	F
Treatments or Electrodes (adjusted)	4	3.8247	0.956	18.038
Blocks (strips)	4	0.258	-	
Positions	3	0.544	0.181	3.415
Error	8	0.4213	0.053	
Totals	19	5.048		

Table 19. ANOVA Results for Youden Square Example.

The treatment or electrode effect is highly significant, since the critical value $F_{.01}$ (4, 8) = 7.01 for α = 0.01. Since $F_{.05}$ (3,8) = 4.07 for α = 0.05, the position effect is not significant at the 5% level. The block mean square not given as blocks must be adjusted by treatments if one is to determine whether block effects are significant. The procedure is identical with that previously discussed for randomized balanced incomplete block designs.

C. Fractional Factorial Designs

There are some situations in which one cannot afford (timewise or costwise) to test all possible treatment combinations, even though distributed over many blocks. For example, the design of complex equipment (as, for example, the "main battle tank" at one time proposed for use by the NATO countries) would involve at least ten basic factors each at two levels. This means that there are $2^{10} = 1024$ possible designs ("treatment combinations"). However, it is not feasible - nor in fact necessary - to examine each of these possible designs. In fact, in this case it was necessary to examine at most $2^6 = 64$ of the possible 1024 designs. These 64 designs cannot, however, be arbitrarily chosen. In fact, the key to the solution of this problem by examining only 64 of the designs lies in the proper selection of these 64 designs ("treatment combinations").

The basis for the validity of fractional factorial designs is the fact that high order interactions are almost always negligible; i.e., they usually do not exist. If, for example, the interaction ABCD does not really exist (or in statistical jargon, is not significant) then ABCD may be confounded with blocks and no information is lost if the experiment is carried out in both blocks. As has already been shown, these two blocks are:

Block 1: (1), ab, ac, bc, ad, bd, cd, abcd

Block 2: a, b, c, abc, d, abd, acd, bcd

If, however, only the treatment combinations in either one of these blocks are tested and analyzed, such an analysis would involve only half of the possible 16 treatment combinations. That is, one has used only a fraction (1/2) of the total number of possible combinations.

This design (i.e., Block 1 or Block 2) is called a fractional factorial design consisting of a 1/2 replicate of the complete factorial design.

The question now arises as to what conclusions can be reached concerning the four factors A, B, C, D and their interactions. Clearly, one would not expect to accomplish the same results from using just one of the two blocks and hence just half as much data as if the entire factorial experiment were carried out. There is a difference. For even if only those treatment combinations in the chosen block were significant, the use of the second block would increase the power of the analysis of variance test. However, the increased cost of the testing for the additional block is seldom if ever worth the cost; otherwise, one would not have used a fractional factorial design to begin with.

The main difference between the confounded complete factorial design and the fractional factorial design is that not only is the effect confounded with blocks unmeasurable, but certain other effects are "aliased" with each other; i.e., are inseparable from each other. Usually this causes no problem, since often these effects are negligible anyway.

In order not to engulf the reader in generalities which are often hard to follow, an example will be given in which both the concepts and the analysis associated with a fractional factorial design will be illustrated. Among other things, Yates' method for the analysis of a 2^N factorial design must be modified.

Example

The following factors are to be studied in a half-replicate of a 2^6 factorial experiment, determined by selecting one of the blocks obtained by confounding ABCDEF with blocks, and designed to evaluate several chemicals as insecticides.

	Factor	Level 0	Level 1
A	BMC	0%	5%
B	Malathion	3%	6%
C	Tedion	1%	2%
D	Chlordane	2%	5%
E	Lindane	1%	4%
F	Pyrethrum	2%	4%

Each experimental unit consists of 10 insects, and the average lifetimes (in seconds) after application of the respective insecticides are as follows, in the random order in which they are obtained:

ce	181	acdf	162	bd	135	abdf	131
ae	172	(1)	182	df	171	ab	136
abef	140	bf	171	acef	159	bcde	105
bcdf	165	cf	176	bc	179	abcdef	109
acde	139	be	187	ac	165	af	176
ef	186	abce	131	bcef	181	ad	150
de	164	abcf	125	cdef	163	abde	115
abcd	112	adef	158	bdef	128	cd	166

Table 20 Data Obtained for Treatment Combinations in a Half-Replicate of a 2^6 Factorial Design With Defining Contrast ABCDEF.

The effect confounded with blocks is called a "defining contrast." Also, while either block may be chosen, it will simplify the discussion if one always chooses the block containing the treatment combination

(1), which procedure will be followed whenever fractional factorial analysis is involved in this document.

The first step is to determine the two blocks resulting from confounding ABCDEF with blocks, and then selecting the block which contains (1). Using the even-odd rule previously explained, one obtains the following two blocks:

Block 1		Block 2	
(1)	af	a	f
ab	bf	b	abf
ac	cf	c	acf
bc	abcf	abc	bcf
ad	db	d	adf
bd	abdf	abd	bdf
cd	acdf	acd	cdf
abcd	bcdf	bcd	abcdf
ae	ef	e	aef
be	abef	abe	bef
ce	acef	ace	cef
abce	bcef	bce	abcef
de	adef	ade	def
abde	bdef	bde	abdef
acde	cdef	cde	acdef
bcde	abcdef	abcde	bcdef

As stated before, either block may be used as a half replicate, but Block 1 will be chosen since the policy will always be to choose that block which contains (1). The question now arises as to what effect are aliased (inseparable) from others. For a half-replicate, the aliases occur in pairs, which are obtained in this Example by multiplying each effect by ABCDEF and reducing mod 2. Thus,

$$A\ (ABCDEF) = A^2BCDEF = BCDEF,$$

from which it follows that A is aliased with BCDEF.

Alias Pairs	
A, BCDEF	CD, ABEF
B, ACDEF	CE, ABDF
C, ABDEF	CF, ABDE
D, ABCEF	DE, ABCF
E, ABCDF	DF, ABCE
F, ABCDE	EF, ABCD
AB, CDEF	ABC, DEF
AC, BDEF	ABD, CEF
AD, BCEF	ACD, BEF
AE, BCDF	BCD, AEF
AF, BCDE	ABE, CDF
BC, ADEF	ACE, BDF
BD, ACEF	BCE, ADF
BE, ACDF	ADE, BCF
BF, ACDE	BDE, ACF
	CDE, ABF

Table 21. Alias Pairs for Half Replicate of 2^6 With Defining Contrast ABCDEF.

What this means is that when one tests the main effect A in the analysis of variance test, he is measuring either A or BCDEF, but he cannot tell which one. However, if BCDEF is negligible, then the main effect A is measurable and can be tested for significance. From Table 21, one can see that if the 5-factor, 4-factor, and 3-factor interactions are negligible, then all the main effects and 2-factor interactions are measurable and can be tested for significance, as will be shown presently.

Before one proceeds further with this example, it would probably be helpful to the reader to consider the aliases in a half replicate of a 2^3 factorial design, the structure of which is simpler and easier to understand. As previously shown, when ABC is confounded with blocks, the resulting two blocks are (1), ab, bc, ac and a, b, c, abc. For a half replicate fractional factorial analysis using the first of these blocks, one has available only the responses (1), ab, bc, ac. Furthermore, the alias pairs are

A, BC
B, AC
C, AB.

As has already been stated, when two effects are aliased, one does not know which effect he is measuring. This becomes clear from an examination of Table 12. Note that in the A column (using only the treatment combinations (1), ab, bc, ac) the main effect of factor A is measured as

$$A = (-1) + ab + ac - bc.$$

Similarly, from Table 12,

$$BC = (1) - ab - ac + bc = -A.$$

That is, one doesn't know whether he is measuring A or BC, since they are based on the same data and are numerically identical. Similarly,

$$B = -AC$$
$$C = -AB$$

which is the significance of the alias pairs listed above. A similar interpretation applies to the alias pairs in Table 21.

Returning to the insecticide problem involving a 2^5 half replicate fractional factorial design, one has the information on the treatment combinations in Block 1 shown in Table 20; they are listed again for the convenience of the reader in column 3 of Table 22. In order to carry out a Yates analysis, the 32 treatment combinations must be placed in a modified standard order, which is constructed in the following manner: Place the complete 2^5 factorial design corresponding to the five factors A, B, C, D, E in standard form, as shown in the first column of Table 22. Then append the sixth factor F by multiplying each treatment combination in column 1 by f, and place the treatment combinations so generated in Column 2 of Table 22. To determine the desired modified standard order in which the 32 treatment combinations must be placed, one places them in the order determined by columns 1 and 2 of Table 22. Thus, one and only one of the entries in Row 1 of Columns 1 and 2 of Table 22 must be found in Block 1; it is (1). Hence, (1) is the first treatment combination in the modified standard order, listed in Column 4 of Table 22. Similarly af (and not a) is contained in the half-replicate and is found in Row 2 of Table 22; it then becomes the second treatment combination in the modified standard order. Again, bf - found in Row 3 of Table 22 and also in the half-replicate - becomes the third item in the modified standard form. Continuing in this fashion, one obtains the modified standard order shown in Column 4 of Table 22 for a half replicate of a 2^6 factorial design for which the defining contrast is ABCDEF.

Finally, Column 5 shows what effect is actually being measured and tested for significance in the analysis of variance. As in the Yates analysis of a complete factorial design- I is the grand total for the 32 items, namely, 4920. (This constitutes a good check at this point in the construction of the Yates Table.) The next item is labeled af, but it actually measures the total response for the main effect A. Similarly, bf is used to test the significance of the main effect B. The treatment combination that one is testing in

Column 1	Column 2	Column 3 Half Replicate	Column 4 Half Replicate in Modified Standard Form	Column 5 Effect Measured and Tested
(1)	f	(1)	(1)	I
a	af	ab	af	A
b	bf	ac	bf	B
ab	abf	bc	ab	AB
c	cf	ad	cf	C
ac	acf	bd	ac	AC
bc	bcf	cd	bc	BC
abc	abcf	abcd	abcf	ABC
d	df	ae	df	D
ad	adf	be	ad	AD
bd	bdf	ce	bd	BD
abd	abdf	abce	abdf	ABD
cd	cdf	de	cd	CD
acd	acdf	abde	acdf	ACD
bcd	bcdf	acde	bcdf	BCD
abcd	abcdf	bcde	abcd	ABCD = EF
e	ef			
ae	aef	af	ef	E
be	bef			
abe	abef	bf	ae	AE
ce	cef	cf	be	BE
ace	acef	abcf	abef	ABE
bce	bcef	df	ce	CE
abce	abcef	abdf	acef	ACE
de	def			
ade	adef	acdf	bcef	BCE
bde	bdef	bcdf	abce	ABCE = DF
abde	abdef	ef	de	DE
cde	cdef	abef	adef	ADE
acde	acdef	acef	bdef	BDE
bcde	bcdef	adef	abde	ABDE = CF
		bcef	cdef	CDE
		bdef	acde	ACDE = BF
abcde	abcdef	cdef	bcde	BCDE = AF
		abcdef	abcdef	ABCDE = F

Table 22. Determination of Modified Standard Order For the Example

the j^{th} row of the modified standard order is the treatment combination in Column 1 and Row j of Table 22, whether or not this treatment combination occurs in Column 4. Using this criterion, one identifies the effects that are being tested by the entries in Column 4 of Table 22. Note that ABCD is aliased with EF; ABCE with DF; ABDE with CF; ACDE with BF; BCDE with AF; and ABCDE with F.

Table 23 consists of a regular Yates analysis when the treatment combinations are placed in the modified standard order in the "Treatment Combination" column and the corresponding numerical responses are given in the next column. The error variance (M.S.E.) used in the demonstrator of each F test consists of the three-factor interactions designated as "Error" in Table 23, namely, ABC, ABD, ACD, BCD, ABE, ACE, BCE, ADE, BDE, and CDE, inasmuch as they are assumed to be negligible. Each of these three-factor interactions has one degree of freedom, so that

$$\text{M.S.E.} = (50 + 465.125 + 0.125 + 15.125 + 128 + 60.5 + 18 + 153.125 + 253.125 + 91.125) / 10 = 123.425$$

and has ten degrees of freedom. The F value for each effect is then obtained by dividing the mean square for that effect by 123.425, which result appears in the F column of Table 23. The critical values of F for 1 and 10 degrees of freedom are, respectively, $F^*_{.05}(1,10) = 4.96$ and $F^*_{.01}(1,10) = 10.00$. It is seen from Table 23 that the main effects A, B, and D (marked with **) are significant at the 1% level, while the main effect F and the interaction AB (marked with *) are significant at the 5% level.

Treatment Combination	Total Response	(1)	(2)	(3)	(4)	(5)		SS= $(5)^2/32$	MS= SS/d.f.	F
(1)	182	358	665	1310	2502	4920	I			
af	176	307	645	1192	2418	-360	A	4050.	4050.	32.8**
bf	171	341	587	1337	-188	-420	B	5512.5	5512.5	44.66**
ab	136	304	605	1081	-172	-144	AB	657.	657.	5.32*
cf	176	321	685	-106	-144	-84	C	220.5	220.5	1.79
ac	165	266	652	-82	-226	-68	AC	144.5	144.5	1.17
bc	179	328	565	-133	-104	12	BC	4.5	4.5	0.04
abcf	125	277	516	-39	-40	-40	ABC	50.	Error	
df	171	358	-41	-88	-2	-374	D	4371.125	4371.125	35.41**
ad	150	327	-65	-106	-82	118	AD	435.12	435.12	3.52
bd	135	340	-25	-59	-56	-126	BD	496.125	496.125	4.02
abdf	131	312	-57	-167	-12	122	ABD	465.125	Error	
cd	166	322	-61	-72	18	22	CD	15.125	15.125	0.12
acdf	162	243	-72	-32	-6	2	ACD	0.125	Error	
bcdf	165	302	-19	-61	80	-22	BCD	15.125	Error	
abcd	112	214	-20	21	40	-22	EF	15.125	15.125	0.12
ef	186	-6	-51	-20	-118	-84	E	220.5	220.5	1.79
ae	172	-35	-37	18	-256	16	AE	8.	8.	0.06
be	187	-11	-55	-33	24	-32	BE	32.	32.	0.26
abef	140	-54	-51	-49	94	64	ABE	128.	Error	1.04
ce	181	-21	-31	-24	-18	-80	CE			
acef	159	-4	-28	-32	-108	44	ACE	60.5	Error	
bcef	181	-4	-79	-11	40	-24	BCE	18.	Error	
abce	131	-53	-88	-1	82	120	DF	450.	450.	3.65
de	164	-14	-29	14	38	-138	DE	595.125	595.125	4.82
adef	158	-47	-43	4	-16	70	ADE	153.125	Error	
bdef	128	-22	17	3	-8	-90	BDE	253.125	Error	
abde	115	-50	-49	-9	10	42	CF	55.125	55.125	0.45
cdef	163	-6	-33	-14	-10	-54	CDE	91.125	Error	
acde	139	-13	-28	-66	-12	18	BF	10.125	10.125	0.08
bcde	105	-24	-7	5	-52	-2	AF	0.125	0.125	0.001
abcde	109	4	28	35	30	82	F	210.125	210.125	1.702

Table 23 Yates Analysis of Half Replicate of 2^6 Fractional Factorial Design for the Example.

IX. CORRELATION AND REGRESSION

A. Simple Linear Correlation and Regression

It is not uncommon for a situation to arise in which one needs to know if there is any relationship between a pair of variables with which he is concerned. Currently, for example, there is the important question of whether there is any relationship between smoking and cancer. Again, in a less serious vein, one might ask if there is a relationship between music appreciation and scientific aptitude, between radio reception and sunspot activity, between beauty and brains.

In the case of simple correlation, the problem consists basically of determining to what extent two variables x and y are related. The determination of the answer to this question centers around the analysis of a set of n pairs of measurements (x_1, y_1), (x_2, y_2), ..., (x_n, y_n). Usually one's first attempt to discover the approximate form of the relationship consists of graphing the data as n points in the x, y plane, the graph being referred to as a scatter diagram. From the scatter diagram one can usually determine the nature of the relationship, i.e., whether it is linear, quadratic, exponential, etc. Once the nature of the relationship is determined, there is still the problem of determining the best-fitting function (curve) of the relevant type, which is used to express mathematically the functional relationship between the two variables. This function is called the regression function, and is used to estimate or predict the value of a dependent variable y from a knowledge of the related independent variable x. This best-fitting function is obtained by the method of least squares [9]. An indication of how good one can expect the estimate or prediction to be is given by the correlation coefficient (usually called the correlation ratio for non-linear correlation). The subject is perhaps best introduced by the following Example.

Example. On the basis of the data below, calculate the correlation coefficient relative to the correlation between carbon content and permeability for sinter mixtures. Also determine the regression function for estimating the permeability index (y) from a knowledge of carbon content (x).

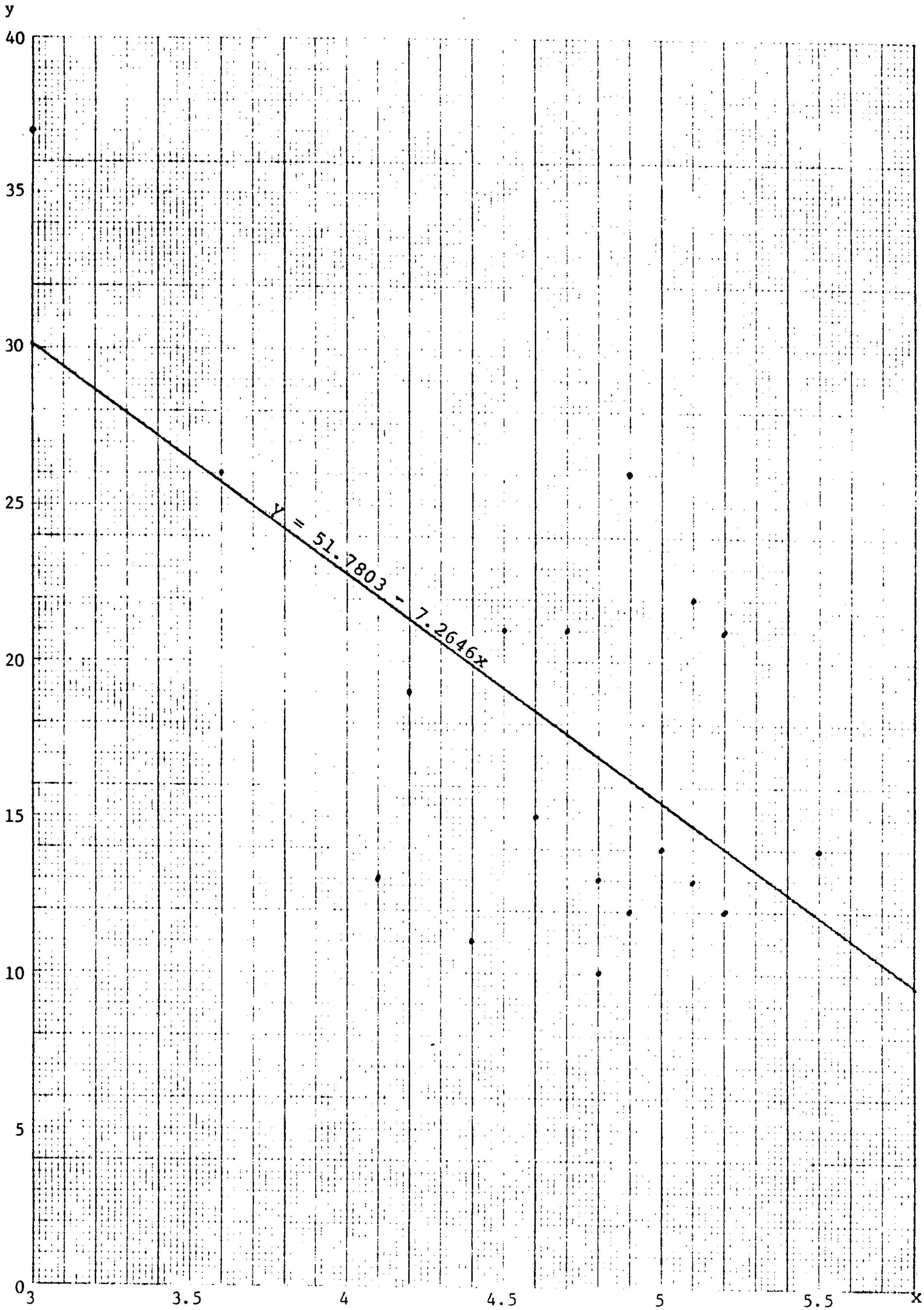
Carbon Content (%) (x_i)	Permeability Index y_i	Carbon Content (%) (x_i)	Permeability Index (y_i)
4.1	13	5.1	22
4.9	12	4.5	21
4.4	11	5.1	13
4.7	10	3.0	37
5.1	13	4.8	13
5.0	14	4.2	19
4.7	21	5.2	12
4.6	14	5.5	14
3.6	26	5.2	21
4.9	25	4.4	29

Table 24. Data for Analyzing Correlation Between Carbon Content and Permeability of Sinter Mixtures.

As is seen from a graph of these data points, there is a general tendency for small values of x to be associated with small values of y, and for large values of x to be associated with large values of y. Also, the trend of the data appears to be linear, although the width of the linear band is appreciable. This suggests that the trend is linear - i.e., the appropriate estimating function is a linear function - but that the correlation is not too pronounced.

Consider first the linear correlation coefficient, defined as

$$r = \sum_{i=1}^{n} \frac{(x_1-\bar{x})(y_1-\bar{y})}{ns_x s_y}$$



GRAPH OF DATA POINTS FOR THE EXAMPLE PROBLEM

It should be pointed out that r is independent of the units and the origin of the variables x and y. This is a particularly convenient characteristic of r, since when large values of x and y are involved, the computation can be considerably simplified by changing the units and the origin of the variables involved. Also, the regression line always passes through the point $(\bar{x}, \bar{y})$. The symbols s_x and s_y denote the standard deviation of x and of y, respectively. Actually, to obtain a value of r which is an unbiased estimate of the population correlation r, one should calculate s_x and s_y from the formulas

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

$$= \left\{ \frac{n\sum_{i=1}^{n} x_i^2 - (\sum_{i-1}^{n} x_i)^2}{n\,(n-1)} \right\}^{\frac{1}{2}}$$

$$s_y = \left\{ \frac{1}{n-1} \sum_{i-1}^{n} (y_i - \bar{y})^2 \right\}^{\frac{1}{2}}$$

$$= \left\{ \frac{n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2}{n\,(n-1)} \right\}^{\frac{1}{2}}$$

The value of r will always be between -1 and 1.

It is particularly convenient from an interpretation standpoint to shift the origin to the point $(\bar{x}, \bar{y})$. Then, among other things, one notes that when the points of the scatter diagram are concentrated in the first and third quadrants, the slope of the regression line will be positive. If, on the other hand, the points are concentrated in the second and fourth quadrants, the slope of the regression line is negative.

The formula for the correlation coefficient is also expressible in another form more convenient for computation; namely,

$$r = \frac{S_{xy}}{[S_{xx} S_{yy}]^{1/2}}$$

where

$$S_{xx} = n\sum_i x_i^2 - (\sum_i x_i)^2$$

$$S_{yy} = n\sum_i y_i^2 - (\sum_i y_i)^2$$

$$S_{xy} = n\sum_i x_i y_i - (\sum_i x_i)\ (\sum_i y_i).$$

The required sums as obtained from the data in Table 24 are

$\sum_i x_i = 93.0$ $\qquad$ $\sum_i y_i = 360$

$\sum_i x_i^2 = 439.14$ $\qquad$ $(\sum_i x_i)^2 = 8649$

$\sum_i x_i y_i = 1625.4$ $\qquad$ $(\sum_i y_i)^2 = 129{,}600$

$\sum_i y_i^2 = 7452.0$

Then

$$S_{xx} = 20(439.14) - (93.0)^2 = 133.8$$

$$S_{yy} = 20(7452.0) - (360)^2 = 19440$$

$$S_{xy} = 20(1625.4) = (93.0)\ (360) = -972$$

and

$$r = \frac{-972}{(133.8)\ (19{,}440)}$$

$$= \frac{-972}{1612.78}$$

$$= -0.603$$

The negative value of the correlation coefficient indicates that y varies inversely with x; i.e., as the carbon content increases, the permeability decreases.

The question now arises as to whether this value of r ($^{-}0.603$) is significantly different from zero, or whether this value arose strictly by chance. To answer this question, one utilizes the fact that

$$z = \frac{\sqrt{n-3}}{2} \ln \left(\frac{1 + r}{1 - r}\right)$$

is approximately distributed as a standardized normal variable with mean zero and standard deviation one, when the population correlation coefficient ρ is zero. Specifically, in the problem at hand,

$$z = \frac{\sqrt{17}}{2} \ln \left(\frac{0.397}{1.60}\right)$$

$$= -2.873$$

Since z > 1.96, one concludes that the true (population) correlation between carbon content and permeability in sinter mixtures is different from zero.

Since the true correlation between carbon content and permeability is different from zero, one should be able to predict permeability (y) from a knowledge of carbon content (x) for a given sinter mixture. The linear predicting or estimating function is the equation of the best-fitting line to the scatter diagram of data points. It can be shown, by using the method of least squares, that the best-fitting straight line (regression line) has the equation

$$y = a + b\,x$$

where

$$a = \left(\sum_i x_i^2 \sum_i y_i - \sum_i x_i \sum_i x_i\, y_i\right) / \left(n \sum_i x_i^2 - \left(\sum_i x_i\right)^2\right)$$

$$b = \left(n \sum_i x_i\, y_i - \sum x_i \sum y_i\right) / \left(n \sum_i x_i^2 - \left(\sum_i x_i\right)^2\right)$$

In this particular problem, the regression line is

$$y = 51.7803 - 7.2646\, x.$$

In predicitng or estimating y for a given value of x, one does not expect to predict an individual value of y, but rather the average value of y corresponding to a given value of x. Thus, if one wishes to predict the weight (y) when an individual's height (x) is known, one does not really hope to predict the weight y of a specific individual whose height is x, but rather the average weight y of all individuals having the same height x.

The question naturally arises as to whether one can determine a confidence interval for a particular estimate or prediction y. Actually, such a confidence interval can be constructed if one assumes that the variables x and y have a bivariate normal distribution:

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\left\{\left(\frac{x-m_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-m_1}{\sigma_1}\right)\left(\frac{y-m_2}{\sigma_2}\right) + \left(\frac{y-m_2}{\sigma_2}\right)^2\right\}/2(1-\rho^2)\right]^*.$$

Up to this point, all that has been said about correlation and regression is valid regardless of what form the joint distribution f (x, y) takes. However, in order to construct a confidence interval for an estimated value y, one must assume that f (x, y) has the bivariate normal distribution. Under this assumption, if $y' = a + b\, x_o$ is the predicted value of y for a given value x_o, then the 100 (1 - α)% confidence interval for the true value of y is given by

$$a + b\, x_o \pm t_{\alpha/2}\, s_e \sqrt{1 + \frac{1}{n} + \frac{n\,(x_o - \bar{x})^2}{S_{xx}}}\ ,$$

where $t_{\alpha/2}$ is the t-variable with n-2 degrees of freedom and corresponding to a confidence level α, and where

$$s_e = \left\{(S_{xx}S_{yy} - S_{xy}^2)/n(n-2)\, S_{xx}\right\}^{\frac{1}{2}}\ .$$

* ρ denotes the true but unknown population correlation coefficient.

For example, if one wished to obtain a 95% confidence interval for the true permeability of a sinter mixture having a carbon content of 4%, the predicted value is

$$y' = a + b\,x_o$$

$$= 51.7803 - 7.2646\ (4)$$

$$= 22.722$$

and the 95% confidence interval is

$$22.722 \pm 2.101\ s_e \sqrt{1 + \frac{1}{20} + \frac{20(4-4.65)}{133.8}}\ .$$

Since

$$s_e = \left\{\frac{(133.8)(19{,}440) - (-972)^2}{20(18)(133.8)}\right\}^{\frac{1}{2}}$$

$$= \left\{\frac{1{,}656{,}288}{48168}\right\}^{\frac{1}{2}}$$

$$= \left\{34.386\right\}^{\frac{1}{2}}$$

$$= 5.864$$

one obtains the interval

$$22.72 \pm 13.196 = \left\{\begin{matrix}35.916\\ 9.524\end{matrix}\right\}$$

The aforementioned definition of the correlation coefficient r is only valid when there is a linear relationship between x and y. It is equivalent (in the case of a linear relationship between x and y) to a more general definition which is valid regardless of the functional relationship between x and y - whether this relationship be linear, quadratic, exponential, or whatever. This definition expresses the correlation coefficient in terms of the reduction in the original variance of y achieved by using the regression function. Specifically,

$$R = \left\{1 - \frac{S_y^{\,2}}{s_y^{\,2}}\right\}^{\frac{1}{2}}$$

$$= \left\{\frac{s_y^2 - S_y^2}{s_y^2}\right\}^{\frac{1}{2}} ,$$

where S_y^2 is the variance about the regression function (or curve):

$$S_y^2 = \frac{1}{n} \sum_i \left[y_i - (a + bx_i)\right]^2$$

$$s_y^2 = \frac{1}{n} \sum_i (y_i - \bar{y})^2 .$$

This definition of ρ, it should be emphasized, is valid regardless of what type of regression function or curve is involved. To differentiate between r and R, the term "correlation ratio" is often used when referring to R.

In the above example, if one calculates , he obtains

$$R = \left\{1 - \frac{S_y^2}{s_y^2}\right\}^{\frac{1}{2}}$$

$$= \left\{1 - \frac{618.9405}{972}\right\}^{\frac{1}{2}}$$

$$= \left[1 - 0.63677\right]^{\frac{1}{2}}$$

$$= \sqrt{0.36323}$$

$$= 0.603$$

The definition of R does not, however, indicate whether the sign is positive or negative. This information must be obtained from the scatter diagram, from which it is seen whether y varies directly with x (in which case the sign of R is positive) or whether y varies inversely with x, in which case the sign of R is negative.

B. Spearman's Rank Correlation Coefficient

It is sometimes necessary to determine the correlation or degree of association between two nonquantitative variables. Thus, one may wish to determine whether there is a correlation between the degree of food coloring used on a food and the consumer's rating of the taste of the food, or whether there is a correlation between a salesman's merit rank by his employer and his number of years of service. One way to measure the correlation between such variables is by using Spearman's rank correlation coefficient, which is based on the difference d in the two ranks for each individual or item. If several items are tied for rank, each item is given the average rank. The formula for computing this correlation coefficient is

$$r = 1 - \frac{6 \Sigma d^2}{(N-1)N(N+1)} \quad .$$

The derivation of this formula can be found in most textbooks in basic statistics, and will not be presented here. An example will illustrate its application.

Example. Twelve salesmen are ranked in order of merit for efficiency by their manager. They are also ranked in accordance with their length of service. Would you conclude that there is a correlation between length of service and efficiency?

Salesman	Years of Service	Rank According to Service (x)	Rank According to Efficiency (y)	d=x-y	d^2
A	5	7.5	6	1.5	2.25
B	2	11.5	12	-0.5	0.25
C	10	2	1	1	1
D	8	4	9	-5	25
E	6	6	8	-2	4
F	4	9	5	4	16
G	12	1	2	-1	1
H	2	11.5	10	1.5	2.25
I	7	5	3	2	4
J	5	7.5	7	0.5	0.25
K	9	3	4	-1	1
L	3	10	11	-1	1
					58

The value of r as computed from the above formula is

$$r = 1 - \frac{6(58)}{11(12)(13)}$$

$$= 0.80$$

It can be shown [8] that for N > 10, Spearman's rank correlation coefficient r is approximately normally distributed with mean zero and standard deviation $\sigma = \left[\frac{1}{18} N(N-1)(2N+5)\right]^{\frac{1}{2}}$. Thus, $z = \frac{r}{\sigma}$ is approximately a standardized normal variable. In the problem at hand, $\sigma = 14.58$ and

$$z = \frac{58}{14.58} = 3.978.$$

Since $z > z_{.01} = 2.576$, one concludes that r is significantly different from zero at the $\alpha = 0.01$ level of significance.

C. Multiple Linear Correlation and Regression

It could well be that the method of linear regression for estimating y from x, discussed in the previous section, yields poor results not because the relationship is far removed from the linear one assumed, but because there is no single variable related closely enough to the variable being estimated to yield good results. However, it may be that there are several variables that, when taken jointly, will serve as a satisfactory basis for estimating the particular variable of interest. Thus, the strength of a ceramic coating is approximately a linear function of kiln temperature, humidity and iron content; the price of a product can be approximately estimated as a linear function of factory production, consumption level, and stocks in storage. As in the case of simple linear regression, the accuracy of the estimate depends upon the multiple linear correlation coefficient $R_{y.x_1x_2...x_n}$. Geometrically, the regression function represents the best-fitting hyperplane to a set of points in a hyperspace, and is again determined by the method of least squares. A discussion of multiple linear regression is, of course, beyond the scope of this document. However, an example will be given to indicate the manner in which multiple correlation and regression is used.

Example. Twelve samples of cold-reduced sheet steel, having different copper contents and annealing temperatures, are measured for hardness with the following results.

Hardness (Rockwell 30-T) (y)	Copper Content (per cent) (X_1)	Annealing Temperature (degrees F) (X_2)
78.8	0.02	1000
65.1	0.02	1100
55.4	0.02	1200
56.2	0.02	1300
80.9	0.10	1000
69.5	0.10	1100
57.4	0.10	1200
55.2	0.10	1300
85.6	0.18	1000
71.8	0.18	1100
60.2	0.18	1200
58.7	0.18	1300

Table 25. Data for Multiple Linear Regression Example

Determine the regression function $y' = b_o + b_1 X_1 + b_2 X_2$, where X_1 denotes copper content, X_2 denotes annealing temperature, and y denotes hardness.

Using the method of least squares, one can show that the values of b_o, b_1, and b_2 which yield the desired regression function are obtained by solving the following system of three linear equations in b_o, b_1, b_2:

$$\sum_{i=1}^{n} y_i = n\, b_o + \left(\sum_{i=1}^{n} X_{1i}\right) b_1 + \left(\sum_{i=1}^{n} X_{2i}\right) b_2$$

$$\sum_{i=1}^{n} X_{1i} y_i = \left(\sum_{i=1}^{n} X_{1i}\right) b_o + \left(\sum_{i=1}^{n} X_{1i}^2\right) b_1 + \left(\sum_{i=1}^{n} X_{1i} X_{2i}\right) b_2$$

$$\sum_{i=1}^{n} X_{2i}\, y_i = \left(\sum_{i=1}^{n} X_{2i}\right) b_o + \left(\sum_{i=1}^{n} X_{1i} X_{2i}\right) b_1 + \left(\sum_{i=1}^{n} X_{2i}^2\right) b_2 .$$

From the data in Table 29, one obtains

$$\sum_{i=1}^{2} y_i = 794.80, \quad \sum_{i=1}^{12} X_{1i} = 1.20, \quad \sum_{i=1}^{12} X_{2i} = 13.800$$

$$\sum_{i=1}^{12} X_{1i}\, y_i = 81.144 \quad \sum_{i=1}^{12} X_{1i}^2 = 0.1712, \quad \sum_{i=1}^{12} X_{1i}\, X_{2i} = 1380$$

$$\sum_{i=1}^{12} X_{2i}\, y_i = 901.070 \quad \sum_{i=1}^{12} X_{2i}^2 = 16,020,000,$$

so that the resulting system of equations is

$$794.80 = 12\, b_o + 1.20\, b_1 + 13,800\, b_2$$

$$81.144 = 1.2\, b_o + 0.1712\, b_1 + 1380\, b_2$$

$$901,070 = 13,800\, b_o + 1380\, b_1 + 16,020,000\, b_2.$$

Solving this system of equations simultaneously for b_o, b_1, and b_2 yields

$$b_o = 162.2, \; b_1 = 32.5, \; b_2 = -0.08633.$$

The regression equation is, therefore,

$$y' = 162.2 + 32.5\ X_1 - 0.086333\ X_2.$$

From this equation, one sees that if X_2 (annealing te ature) is held constant, the hardness (y') increases by 32.5 s for each per cent increase in copper content. Similarly, i (per cent copper content) is held constant, the hardness (y') eases by 0.0863 units for each degree increase in annealing temp ure.

To get some indication of whether the y' estimates are reasonably good or not, one needs to calculate the multiple ar correlation coefficient $R_{y.X_1X_2}$:

$$R_{y.X_1X_2} = \left\{1 - \frac{S_y^2}{s_y^2}\right\}^{\frac{1}{2}}$$

where

$$S_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_i')^2$$

$$s_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

y_i = observed value of y

y_i' = estimated value of y obtained from the multiple linear regression function

The values of y_i and y_i', i=1, 2, ..., 12 are given in Table 26.

i	Observed Value y_i	Estimated Value y'_i	$(y_i - y'_i)^2$
1	78.8	76.517	5.212
2	65.1	67.884	7.751
3	55.4	59.250	14.822
4	56.2	50.617	31.170
5	80.9	79.117	3.179
6	69.5	70.484	0.968
7	57.4	61.850	19.802
8	55.2	53.217	3.932
9	85.6	81.717	15.078
10	71.8	73.084	1.649
11	60.2	64.450	18.063
12	58.7	55.817	8.312

Table 26. Observed and Estimated Values of y for Multiple Regression Example

$$S_y^2 = \frac{1}{12}(129.938) = 10.828$$

$$s_y^2 = 108.503$$

$$R_{y.X_1X_2} = \left[1 - \frac{10.828}{108.503}\right]^{\frac{1}{2}}$$

$$= \left[0.9002055\right]^{\frac{1}{2}}$$

$$= 0.949$$

To test this value of $R_{y.X_1X_2}$ to see if it is significantly different from zero, one computes the F ratio

$$F = \frac{\frac{1}{2}\left\{\sum_1^{12}(y_i - \bar{y})^2 - \sum_1^{12}(y_i - y'_i)^2\right\}}{\frac{1}{9}\sum_1^{12}(y_i - y'_i)^2}$$

$$= \frac{\frac{1}{2}\left[1302.036 - 129.938\right]}{\frac{1}{9}(129.938)}$$

$$= 40.591$$

The critical value $F^*_{.01}$ corresponding to 2 and 9 degrees of freedom and the $\alpha = 0.01$ level of significance is $F^* = 8.02$ Hence, $R_{y.X_1X_2}$ is significant at better than the 1% level.

The above F ratio is appropriate for testing the significance of the multiple linear correlation coefficient because it is the ratio of two independent variances. The numerator of the ratio is the "explained variance" or the reduction in the original variance achieved by using the regression function to estimate y. The denominator is the original variance - or equivalently, the variance of the observed values of y about the estimated values when using the estimator $y = \bar{y}$ instead of the regression function. If this reduction in variance (numerator) is significantly greater than the random variance (denominator), then the regression function (estimator) does a better job of estimating y than does the simple estimating function $y = \bar{y}$. This means that utilizing information of X_1 and X_2 in the regression function yields a better estimate of y than one obtains when he uses the estimating function $y = \bar{y}$ which does not make use of the relevant information about X_1 and X_2. This indicates, therefore, that there is correlation between y and X_1, X_2 in the parent population.

It should perhaps be stated that in the general case, the denominator of the F ratio has N-1-K degrees of freedom, where N is the sample size and K is the number of constants in the multiple regression equation, while the numerator has K-1 degrees of freedom.

X. TOLERANCE LIMITS

Frequently one wishes to know what proportion of the population from which samples are drawn lies between two specified limits. Thus, one would sometimes like to be able to assert with a degree of confidence $1 - \alpha$ that the proportion of the population contained between $\bar{X} - Ks$ and $\bar{X} + Ks$ is at least P, where $\bar{X}$ and s are the mean and standard deviation of a random sample drawn from that population. In other cases, one may wish to determine with a confidence $1 - \alpha$ what proportion P of the population lies between X_1 and X_n, where X_1 and X_n denote the smallest and largest items in a random sample taken from this population.

Consider first the matter of bracketing the proportion p of the population contained between $\bar{X} \pm K\,s$, with confidence $1 - \alpha$. Tables [10] have been constructed giving such values of K for $P = 0.90, 0.95, 0.99$, $1 - \alpha = 0.95$ and 0.99, and selected values of n from 2 to 1000. The following Example illustrates how tolerance limits can be applied to real world problems.

<u>Example</u>. In a study designed to determine the time required to assemble a given piece of machinery, 50 workers averaged 42.5 minutes with a standard deviation of 3.8 minutes. Establish tolerance limits for which one can assert with a degree of confidence 0.95 that at least 90 per cent of the workers (in the population of workers from which the sample was selected) can assemble the piece of machinery within these limits. Assume that assembly time is a normally distributed variable.

From the aforementioned Table [10], one finds, for $n = 50$, $1 - \alpha = 0.95$, $P = 0.90$, that $k = 1.996$, so that

$$\bar{X} - ks = 34.92$$

$$\bar{X} + ks = 50.08.$$

That is, one has a confidence of 0.95 that at least 90% of the workers require between 34.92 and 50.08 minutes to assemble the designated piece of machinery.

Nonparametric tolerance limits can be obtained on the basis of X_1, and X_n, the smallest and largest observations in a sample. Wilks [11] has shown that the equation

$$nP^{n-1} - (n-1)P^n = \alpha$$

relates the quantities n, P, and α, where P is the minimum proportion of the population contained between X_1 and X_n. An approximate solution for n, which is usually quite satisfactory, is given by

$$n \cong \frac{1}{2} + \frac{1 + P}{1 - P} \cdot \frac{\chi_\alpha^2}{4},$$

where χ_α^2 is the value of chi-square for 4 degrees of freedom that corresponds to a right-hand tail of area α.

<u>Example</u>. Suppose that in the preceding example the population probability distribution function was unknown, and that $X_1 = 36$, $X_{50} = 48$. Find the value of P corresponding to $1 - \alpha = 0.95$.

Solving the above equation for P in terms of n and χ_α^2, one has

$$P = \frac{4n - 2 - \chi_\alpha^2}{\chi_\alpha^2 + 4n - 2}$$

which, for this Example, becomes

$$P = \frac{200 - 2 - 9.488}{9.488 + 200 - 2}$$

$$= \frac{188.512}{207.488}$$

$$= 0.908 .$$

That is, based on a sample of 50 workers for which the shortest assembly time was 36 minutes and the longest 48 minutes, one can assert with a confidence of 0.95 that 90.8 per cent of all the individuals in the population from which this sample was selected require between 36 and 48 minutes to assemble the given piece of machinery.

The formula

$$n \cong \frac{1}{2} + \frac{1 + P}{1 - P} \cdot \frac{X_{\alpha}^{2}}{4}$$

shows approximately the sample size required to assert with confidence $1 - \alpha$ that at least 100 P% of the population lies between the smallest (X_1) and largest (X_n) items in the sample, where the values α and P specified in advance. Thus, if one wished to have a confidence of 0.95 that at least 90% of the population was contained between the values X_1 and X_n, he must use a sample size

$$n = \frac{1}{2} + \frac{1.90}{0.10} \left(\frac{9.488}{4}\right)$$

$$= \frac{1}{2} + 19(2.372)$$

$$= 45.568;$$

i.e., a sample size $n \doteq 46$.

XI. SPECIAL SAMPLING TECHNIQUES

The methods of analysis discussed up to this point have been based on data which has been obtained by simple random sampling. That is, the sampling is carried out in such a way that each item in the population from which the sample was drawn has an equally likely chance of being chosen. In the great majority of situations requiring statistical analysis, simple random sampling is appropriate. This is most fortunate, since most of the methods of statistical inference are not applicable when other sampling methods are used. For the sake of completion, a few of these methods are mentioned here. A discussion of these methods is beyond the scope of this document, but can be found in standard texts on sampling techniques [12].

Stratified Random Sampling

In stratified random sampling, the population of N items is divided into subpopulations of N_1, N_2, ..., N_L items, respectively. These subpopulations are called strata; when they have been determined, a sample is drawn from each, the drawings being carried out independently in the different strata. The sample sizes are labeled n_1, n_2, ..., n_L. If a simple random sample is drawn from each stratum, the procedure is referred to as stratified random sampling.

Stratified random sampling is used in the following situations:

1. When data of known precision are wanted for certain subdivisions of the population.

2. When administrative convenience dictates the use of stratification, as, for example, when an agency conducting a survey has field offices, each of which can supervise the survey for a part of the population.
3. When sampling problems differ greatly in various parts of the population.

The formulas and relationships involved in carrying out the relevant analyses are considerably different from those utilizing simple random sampling, and will not be discussed here.

Systematic Random Sampling

This method of sampling is considerably different from simple random sampling. Consider a population of N units which are numbered 1 to N in some order, and suppose one selects a sample of n units by taking a unit at random from the first k units and every k^{th} unit thereafter. Thus, if k is 20 and the first unit drawn is the number 16, then the subsequent units are numbers 36, 56, 76, and so on. Note that the selection of the first unit determines the whole sample. Such a sample is called an "every k^{th} systematic sample."

Systematic sampling has a few advantages over simple random sampling. In the first place, it is easier to draw a sample and often easier to carry out without mistakes. Also, it appears to be somewhat more precise than simple random sampling. Actually, it stratifies the population into n strata, consisting of the first k units, the second k units, etc. The difference is that with a systematic sample, the units are taken at the same relative position in the stratum, while with a stratified random sample, the position in the sample is determined separately by randomization within each stratum. The fact that the systematic sample is spread more evenly over the population may sometimes make systematic sampling somewhat more precise than stratified random sampling.

It should be pointed out, however, that the performance of systematic sampling is greatly dependent on the properties of the population. There are some populations for which systematic sampling is more precise and others for which it is less precise than simple random sampling. In fact, as Cochran points out [12, p 214], for some populations and some values of n, the variance of the mean of systematic samples may even increase when a larger sample is taken - quite a contrast for the good behavior of the mean of simple random samples.

As in the case of stratified random sampling, the formulas and relationship involved in carrying out the relevant analyses are quite different from those utilizing simple random sampling, and will not be discussed here.

XII. COMPUTER PROGRAMS

Since computer programs are available for the statistical models presented in this report, no discussion was included relevant to computer programs. (One excellent source is the IBM Scientific Computer Subroutine Package.) The one possible exception is Wilson's Distribution Free Analysis of Variance model, which to the author's knowledge has not previously been programmed. Since it is a rather useful nonparametric model, a computer program has been written (in PL-1 language) and is included in the Appendix, together with an example. The example is strictly illustrative, and serves primarily to indicate the format used in presenting the data output. Three factors, one with three levels and the others with two, are analyzed. An asterisk after the D. F. number of a given factor indicates significance at the 5% level. For example, if there were an asterisk after the degree number "2" of the first factor, it would indicate that factor 1 is significant at the 5% level. (The significance level is an input and can be varied.) In fact, none of the factors or interactions are significant at the 5% level in this example. (D. F. denotes degrees of freedom.)

XIII. APPENDIX: COMPUTER PROGRAM FOR WILSON'S DISTRIBUTION-FREE ANOVA MODEL.

```
STMT LEVEL NEST
  1                 PGM:     PROC OPTIONS(MAIN);
                    /*************************************************************************/
                    /*SOURCE FILE : ALPHA, 30 NUMBERS OF THE CRITICAL V.L__ _.               */
                    /*               CHI-SQUARE.                                             */
                    /*SYSIN FILE:      FIRST CARD---NAME OF THE PROBLEM                      */
                    /*                 SECOND CARD : NUMBER OF FACTORS,                      */
                    /*                               NUMBER OF LEVELS ... EACH FACTOR,      */
                    /*                               MAXIMUN NUMBER OF REPLICATIONS ...     */
                    /*                               THE PROBLEM.                            */
                    /*                 3+ CARDS :NUMBER OF REPLICATIONS IN THIS CELL         */
                    /*                           INPUT DATA FOR THIS CELL                    */
                    /*************************************************************************/
  2     1                    DCL(F(5),N_WAY,MAX#REP) FIXED,(TITLE)CHAR(..),
  3     1                    DCL(#) FIXED;
  4     1                    DCL SOURCE FILE;
  5     1                    DCL SET(30);
  6     1                    OPEN FILE(SOURCE);
  7     1                    GET FILE(SOURCE)LIST(ALPHA);
  8     1                    GET FILE(SOURCE)LIST((SET(JL) DO JL=1 TO ...),
  9     1                    F=1;
 10     1                    ON ENDFILE(SYSIN) GO TO EOJ;
 12     1           START:   GET EDIT(TITLE)(COL(1),A(80));
 13     1                    GET LIST(N_WAY);
 14     1                    GET LIST((F(I) DO I=1 TO N_WAY));
 15     1                    GET LIST(MAX#REP);
 16     1                    #=MAX#REP*F(1)*F(2)*F(3)*F(4)*F(5);
                    /*************************************************************************/
 17     1                    BEGIN;
 18     2                    DCL  (#REP(F(1),F(2),F(3),F(4),F(5)));
 19     2                    DCL(DATA(F(1),F(2),F(3),F(4),F(5),MAX#REP),LIST(#));
 20     2                    DCL(A(F(1),F(2),F(3),F(4),F(5)),B(F(1),F(2),F(3),F(4),F(5)),
                             SQU(32),DEG(32));
 21     2                    DCL(#P,#A,#B);
 22     2                    DCL CH(32) CHAR(10);
 23     2                    DCL(MJ(5))FIXED;
 24     2                    DCL(ONE(5),TWO(10),THR(10),FOU(5));
 25     2                    M=0;
 26     2                    SQU=0;   DEG=0;
 28     2                    ONE=0;   TWO=0;   THR=0;   FOU=0;
 32     2           Z11:     DO I1=1 TO F(1);
 33     2      1    Z12:     DO I2=1 TO F(2);
 34     2      2    Z13:     DO I3=1 TO F(3);
 35     2      3    Z14:     DO I4=1 TO F(4);
 36     2      4    Z15:     DO I5=1 TO F(5);
 37     2      5             GET LIST(#REP(I1,I2,I3,I4,I5));
 38     2      5    Z16:     DO I6=1 TO #REP(I1,I2,I3,I4,I5);
 39     2      6             M=M+1;
 40     2      6             GET LIST(LIST(M));
 41     2      6             DATA(I1,I2,I3,I4,I5,I6)=LIST(M);
 42     2      6             END Z16; END Z15; END Z14; END Z13; END Z12; END Z11;
                    /*************************************************************************/
 48     2           SORT:    DO LL=1 TO M-1;
 49     2      1    COMPAR:  IF LIST(LL)>=LIST(LL+1) THEN GO TO OK;
 51     2      1             STORE=LIST(LL);   LIST(LL)=LIST(LL+1);   LIST(LL+1)=STORE;
 54     2      1             LL=LL-1;   IF LL=0 THEN LL=2;
 57     2      1             GO TO COMPAR;
```

```
        PGM:     PROC OPTIONS(MAIN);

STMT LEVEL NEST
 58      2    1    OK:      END SORT;
 59      2                  LLL=M/2.+0.5;    MD=LIST(LLL);
                   /*************************************************************************/
 61      2         CLASS:   A=0;   B=0;
 63      2         Z21:     DO I1=1  TO  F(1);
 64      2    1    Z22:     DO I2=1  TO  F(2);
 65      2    2    Z23:     DO I3=1  TO  F(3);
 66      2    3    Z24:     DO I4=1  TO  F(4);
 67      2    4    Z25:     DO I5=1  TO  F(5);
 68      2    5    Z26:     DO I6=1  TO  #REP(I1,I2,I3,I4,I5);
 69      2    6             IF DATA(I1,I2,I3,I4,I5,I6)<MD THEN B(I1,I2,I3,I4,I5)=
                                                               B(I1,I2,I3,I4,I5)+1;
 71      2    6                   ELSE A(I1,I2,I3,I4,I5)=A(I1,I2,I3,I4,I5)+1;
 72      2    6             END Z26; END Z25; END Z24; END Z23; END Z22; END Z21;
 78      2                  TT_A=SUM(A);   TT_B=SUM(B);   F_A=TT_A/M;   F_B=TT_B/M;
                   /*************************************************************************/
 82      2                  PUT PAGE EDIT(TITLE)(COL(20),A);
 83      2                  PUT SKIP(2)EDIT(N_WAY,' WAY EXPERIENCE')(COL(45),F(2),A);
 84      2                  PUT SKIP EDIT('MEDIAN POINT IS ',MD)(COL(45),A,F(4));
 85      2                  PUT SKIP(3);
 86      2                  DO N=1  TO  N_WAY;
 87      2    1             PUT EDIT('FACTOR',N)(X(5),A,F(2));
 88      2    1             END;
 89      2                  PUT EDIT('TOTAL #','# OF ABOVE','# OF BELOW')(X(10),
                              3 (X(5),A));
 90      2                  DO MJ(1)=1  TO  F(1);
 91      2    1             DO MJ(2)=1  TO  F(2);
 92      2    2             DO MJ(3)=1  TO  F(3);
 93      2    3             DO MJ(4)=1  TO  F(4);
 94      2    4             DO MJ(5)=1  TO  F(5);
 95      2    5             PUT SKIP(2);
 96      2    5             PUT EDIT((MJ(L)  DO L=1  TO  N_WAY))(X(9),F(2));
 97      2    5             PUT EDIT(#REP(MJ(1),MJ(2),MJ(3),MJ(4),MJ(5)),
                                      A(MJ(1),MJ(2),MJ(3),MJ(4),MJ(5)),
                                      B(MJ(1),MJ(2),MJ(3),MJ(4),MJ(5)))(X(12), 3(X(10),
                                F(3)));
 98      2    5             END;  END;   END;   END;   END;
                   /*************************************************************************/
103      2         COMPUT:  PROC;
104      3                  #P=#A+#B;
105      3                  SGU( I)=SGU(I)+((#A-#P*F_A)**2)/(#P*F_A)
                                          +((#B-#P*F_B)**2)/(#P*F_B);
106      3                  END COMPUT;
                   /*************************************************************************/
107      2         TOTAL:   I=1;
108      2         W1:      DO I1=1  TO  F(1);
109      2    1    W2:      DO I2=1  TO  F(2);
110      2    2    W3:      DO I3=1  TO  F(3);
111      2    3    W4:      DO I4=1  TO  F(4);
112      2    4    W5:      DO I5=1  TO  F(5);
113      2    5             #A=A(I1,I2,I3,I4,I5);
114      2    5             #B=B(I1,I2,I3,I4,I5);
115      2    5             CALL COMPUT;
116      2    5             END W5;   END W4;   END W3;   END W2;   END W1;
121      2                  DEG(I)=F(1)*F(2)*F(3)*F(4)*F(5)-1;
```

```
STMT LEVEL NEST
                        /*******************************************************************/
 122     2              ONE1:   I=I+1;
 123     2                      DO J=1 TO F(1);
 124     2     1                #A=SUM(A(J,*,*,*,*));   #B=SUM(B(J,*,*,*,*));
 126     2     1                CALL COMPUT;
 127     2     1                END;
 128     2                      ONE(1)=SQU(I);
 129     2                      DEG(I)=F(1)-1;
 130     2                      CH(I)='1';
                        /*******************************************************************/
 131     2              ONE2:   I=I+1;
 132     2                      DO K=1 TO F(2);
 133     2     1                #A=SUM(A(*,K,*,*,*));   #B=SUM(B(*,K,*,*,*));
 135     2     1                CALL COMPUT;
 136     2     1                END;
 137     2                      ONE(2)=SQU(I);
 138     2                      DEG(I)=F(2)-1;
 139     2                      CH(I)='2';
 140     2                      IF N_WAY=2 THEN DO;
 142     2     1                I=I+1;       SQU(I)=SQU(1)-SQU(2)-SQU(3);
 144     2     1                CH(I)='1*2';
 145     2     1                      DEG(I)=(F(1)-1)*(F(2)-1);
 146     2     1                   GO TO OUT;   END;
                        /*******************************************************************/
 148     2              ONE3:   I=I+1;
 149     2                      DO L=1 TO F(3);
 150     2     1                #A=SUM(A(*,*,L,*,*));   #B=SUM(B(*,*,L,*,*));
 152     2     1                CALL COMPUT;
 153     2     1                END;
 154     2                      ONE(3)=SQU(I);
 155     2                      DEG(I)=F(3)-1;
 156     2                      CH(I)='3';
 157     2                      IF N_WAY=3 THEN GO TO TWO12;
                        /*******************************************************************/
 159     2              ONE4:   I=I+1;
 160     2                      DO M=1 TO F(4);
 161     2     1                #A=SUM(A(*,*,*,M,*));   #B=SUM(B(*,*,*,M,*));
 163     2     1                CALL COMPUT;
 164     2     1                END;
 165     2                      ONE(4)=SQU(I);
 166     2                      DEG(I)=F(4)-1;
 167     2                      CH(I)='4';
 168     2                      IF N_WAY=4 THEN GO TO TWO12;
                        /*******************************************************************/
 170     2              ONE5:   I=I+1;
 171     2                      DO N=1 TO F(5);
 172     2     1                #A=SUM(A(*,*,*,*,N));   #B=SUM(B(*,*,*,*,N));
 174     2     1                CALL COMPUT;
 175     2     1                END;
 176     2                      ONE(5)=SQU(I);
 177     2                      DEG(I)=F(5)-1;
 178     2                      CH(I)='5';
                        /*******************************************************************/
 179     2              TWO12:  I=I+1;
 180     2                      SQRINT=SQU(1)-ONE(1)-ONE(2)-ONE(3)-ONE(4)-ONE(5);
```

```
      PGM:     PROC OPTIONS(MAIN);

 STMT LEVEL NEST

 181      2                     DO J=1 TO F(1);
 182      2       1             DO K=1 TO F(2);
 183      2       2             #A=SUM(A(J,K,*,*,*));   #B=SUM(B(J,K,*,*,*));
 185      2       2             CALL COMPUT;
 186      2       2             END;  END;
 188      2                     DEG(I)=(F(1)-1)*(F(2)-1);
 189      2                     TWO(1)=SQU(I)-ONE(1)-ONE(2);
 190      2                     CH(I)='1*2';
 191      2                     SQU(I)=TWO(1);
                   /*************************************************************************/
 192      2          TWO13:     I=I+1;
 193      2                     DO J=1 TO F(1);
 194      2       1             DO L=1 TO F(3);
 195      2       2             #A=SUM(A(J,*,L,*,*));   #B=SUM(B(J,*,L,*,*));
 197      2       2             CALL COMPUT;
 198      2       2             END;  END;
 200      2                     DEG(I)=(F(1)-1)*(F(3)-1);
 201      2                     TWO(2)=SQU(I)-ONE(1)-ONE(3);
 202      2                     CH(I)='1*3';
 203      2                     SQU(I)=TWO(2);
                   /*************************************************************************/
 204      2          TWO23:     I=I+1;
 205      2                     DO K=1 TO F(2);
 206      2       1             DO L=1 TO F(3);
 207      2       2             #A=SUM(A(*,K,L,*,*));   #B=SUM(B(*,K,L,*,*));
 209      2       2             CALL COMPUT;
 210      2       2             END;  END;
 212      2                     DEG(I)=(F(2)-1)*(F(3)-1);
 213      2                     TWO(5)=SQU(I)-ONE(2)-ONE(3);
 214      2                     SQU(I)=TWO(5);
 215      2                     CH(I)='2*3';
 216      2                     IF N_WAY=3 THEN DO;
 218      2       1                       I=I+1;
 219      2       1                       SQU(I)=SQRINT-TWO(1)-TWO(2)-TWO(5);
 220      2       1             CH(I)='1*2*3';
 221      2       1                       DEG(I)=(F(1)-1)*(F(2)-1)*(F(3)-1);   GO TO OUT; END;
                   /*************************************************************************/
 224      2          TWO14:     I=I+1;
 225      2                     DO J=1 TO F(1);
 226      2       1             DO M=1 TO F(4);
 227      2       2             #A=SUM(A(J,*,*,M,*));   #B=SUM(B(J,*,*,M,*));
 229      2       2             CALL COMPUT;
 230      2       2             END;  END;
 232      2                     DEG(I)=(F(1)-1)*(F(4)-1);
 233      2                     TWO(3)=SQU(I)-ONE(1)-ONE(4);
 234      2                     SQU(I)=TWO(3);
 235      2                     CH(I)='1*4';
                   /*************************************************************************/
 236      2          TWO24:     I=I+1;
 237      2                     DO K=1 TO F(2);
 238      2       1             DO M=1 TO F(4);
 239      2       2             #A=SUM(A(*,K,*,M,*));   #B=SUM(B(*,K,*,M,*));
 241      2       2             CALL COMPUT;
 242      2       2             END;  END;
 244      2                     DEG(I)=(F(2)-1)*(F(4)-1);
```

```
PGM:     PROC OPTIONS(MAIN);

STMT LEVEL NEST

245     2                  TWO(6)=SQU(I)-ONE(2)-ONE(4);
246     2                  SQU(I)=TWO(6);
247     2                  CH(I)='2*4';
             /**********************************************************************/
248     2        TW034:    I=I+1;
249     2                  DO L=1 TO F(3);
250     2     1            DO M=1 TO F(4);
251     2     2            #A=SUM(A(*,*,L,M,*));   #B=SUM(B(*,*,L,M,*));
253     2     2            CALL COMPUT;
254     2     2            END;  END;
256     2                  DEG(I)=(F(3)-1)*(F(4)-1);
257     2                  TWO(8)=SQU(I)-ONE(3)-ONE(4);
258     2                  SQU(I)=TWO(8);
259     2                  CH(I)='3*4';
260     2                  IF N_WAY=4 THEN GO TO THR123;
             /**********************************************************************/
262     2        TW015:    I=I+1;
263     2                  DO J=1 TO F(1);
264     2     1            DO N=1 TO F(5);
265     2     2            #A=SUM(A(J,*,*,*,N));   #B=SUM(B(J,*,*,*,N));
267     2     2            CALL COMPUT;
268     2     2            END;  END;
270     2                  DEG(I)=(F(1)-1)*(F(5)-1);
271     2                  TWO(4)=SQU(I)-ONE(1)-ONE(5);
272     2                  CH(I)='1*5';
273     2                  SQU(I)=TWO(4);
             /**********************************************************************/
274     2        TW025:    I=I+1;
275     2                  DO K=1 TO F(2);
276     2     1            DO N=1 TO F(5);
277     2     2            #A=SUM(A(*,K,*,*,N));   #B=SUM(B(*,K,*,*,N));
279     2     2            CALL COMPUT;
280     2     2            END;  END;
282     2                  DEG(I)=(F(2)-1)*(F(5)-1);
283     2                  TWO(7)=SQU(I)-ONE(2)-ONE(5);
284     2                  CH(I)='2*5';
285     2                  SQU(I)=TWO(7);
             /**********************************************************************/
286     2        TW035:    I=I+1;
287     2                  DO L=1 TO F(3);
288     2     1            DO N=1 TO F(5);
289     2     2            #A=SUM(A(*,*,L,*,N));   #B=SUM(B(*,*,L,*,N));
291     2     2            CALL COMPUT;
292     2     2            END;  END;
294     2                  DEG(I)=(F(3)-1)*(F(5)-1);
295     2                  TWO(9)=SQU(I)-ONE(3)-ONE(5);
296     2                  CH(I)='3*5';
297     2                  SQU(I)=TWO(9);
             /**********************************************************************/
298     2        TW045:    I=I+1;
299     2                  DO M=1 TO F(4);
300     2     1            DO N=1 TO F(5);
301     2     2            #A=SUM(A(*,*,*,M,N));   #B=SUM(B(*,*,*,M,N));
303     2     2            CALL COMPUT;
304     2     2            END;  END;
```

```
      PGM:     PROC OPTIONS(MAIN);

STMT LEVEL NEST

306      2                       DEG(I)=(F(4)-1)*(F(5)-1);
307      2                       TWO(10)=SQU(I)-ONE(4)-ONE(5);
308      2                       CH(I)='4*5';
309      2                       SQU(I)=TWO(10);
                     /******************************************************************/
310      2           THR123:  I=I+1;
311      2                    DO J=1 TO F(1);
312      2       1            DO K=1 TO F(2);
313      2       2            DO L=1 TO F(3);
314      2       3            #A=SUM(A(J,K,L,*,*));   #B=SUM(B(J,K,L,*,*));
316      2       3            CALL COMPUT;
317      2       3            END;  END;  END;
320      2                    DEG(I)=(F(1)-1)*(F(2)-1)*(F(3)-1);
321      2                    THR(1)=SQU(I)-ONE(1)-ONE(2)-ONE(3)-TWO(1)-TWO(2)-TWO(5);
322      2                    CH(I)='1*2*3';
323      2                    SQU(I)=THR( 1);
                     /******************************************************************/
324      2           THR124:  I=I+1;
325      2                    DO J=1 TO F(1);
326      2       1            DO K=1 TO F(2);
327      2       2            DO M=1 TO F(4);
328      2       3            #A=SUM(A(J,K,*,M,*));   #B=SUM(B(J,K,*,M,*));
330      2       3            CALL COMPUT;
331      2       3            END;  END;  END;
334      2                    DEG(I)=(F(1)-1)*(F(2)-1)*(F(4)-1);
335      2                    THR(2)=SQU(I)-ONE(1)-ONE(2)-ONE(4)-TWO(1)-TWO(3)-TWO(6);
336      2                    CH(I)='1*2*4';
337      2                    SQU(I)=THR( 2);
                     /******************************************************************/
338      2           THR134:  I=I+1;
339      2                    DO J=1 TO F(1);
340      2       1            DO L=1 TO F(3);
341      2       2            DO M=1 TO F(4);
342      2       3            #A=SUM(A(J,*,L,M,*));   #B=SUM(B(J,*,L,M,*));
344      2       3            CALL COMPUT;
345      2       3            END;  END;  END;
348      2                    DEG(I)=(F(1)-1)*(F(3)-1)*(F(4)-1);
349      2                    THR(4)=SQU(I)-ONE(1)-ONE(3)-ONE(4)-TWO(2)-TWO(3)-TWO(8);
350      2                    CH(I)='1*3*4';
351      2                    SQU(I)=THR( 4);
                     /******************************************************************/
352      2           THR234:  I=I+1;
353      2                    DO K=1 TO F(2);
354      2       1            DO L=1 TO F(3);
355      2       2            DO M=1 TO F(4);
356      2       3            #A=SUM(A(*,K,L,M,*));   #B=SUM(B(*,K,L,M,*));
358      2       3            CALL COMPUT;
359      2       3            END;  END;  END;
362      2                    DEG(I)=(F(2)-1)*(F(3)-1)*(F(4)-1);
363      2                    THR(7)=SQU(I)-ONE(2)-ONE(3)-ONE(4)-TWO(5)-TWO(6)-TWO( 8);
364      2                    CH(I)='2*3*4';
365      2                    SQU(I)=THR( 7);
366      2                    IF N_WAY=4 THEN DO;
368      2       1            SQU(I)=SQTOT-TWO(1)-TWO(2)-TWO(3)-TWO(5)-TWO(6)-TWO(8)
                                      -THR(1)-THR(2)-THR(4)-THR(7);
```

```
         PGM:     PROC OPTIONS(MAIN);

STMT LEVEL NEST
 369     2    1             DEG(I)=(F(1)-1)*(F(2)-1)*(F(3)-1)*(F(4)-1);
 370     2    1         CH(I)='1*2*3*4';
 371     2    1            GO TO OUT;  END;
                /******************************************************************/
 373     2      THR125: I=I+1;
 374     2              DO J=1 TO F(1);
 375     2    1         DO K=1 TO F(2);
 376     2    2         DO N=1 TO F(5);
 377     2    3         #A=SUM(A(J,K,*,*,N));   #B=SUM(B(J,K,*,*,N));
 379     2    3         CALL COMPUT;
 380     2    3         END;  END;  END;
 383     2              DEG(I)=(F(1)-1)*(F(2)-1)*(F(5)-1);
 384     2              THR(3)=SQU(I)-ONE(1)-ONE(2)-ONE(5)-TWO(1)-TWO(4)-TWO(7);
 385     2              CH(I)='1*2*5';
 386     2              SQU(I)=THR( 3);
                /******************************************************************/
 387     2      THR135: I=I+1;
 388     2              DO J=1 TO F(1);
 389     2    1         DO L=1 TO F(3);
 390     2    2         DO N=1 TO F(5);
 391     2    3         #A=SUM(A(J,*,L,*,N));   #B=SUM(B(J,*,L,*,N));
 393     2    3         CALL COMPUT;
 394     2    3         END;  END;  END;
 397     2              DEG(I)=(F(1)-1)*(F(3)-1)*(F(5)-1);
 398     2              THR(5)=SQU(I)-ONE(1)-ONE(3)-ONE(5)-TWO(2)-TWO(4)-TWO(9);
 399     2              CH(I)='1*3*5';
 400     2              SQU(I)=THR( 5);
                /******************************************************************/
 401     2      THR145: I=I+1;
 402     2              DO J=1 TO F(1);
 403     2    1         DO M=1 TO F(4);
 404     2    2         DO N=1 TO F(5);
 405     2    3         #A=SUM(A(J,*,*,M,N));   #B=SUM(B(J,*,*,M,N));
 407     2    3         CALL COMPUT;
 408     2    3         END;  END;  END;
 411     2              DEG(I)=(F(1)-1)*(F(4)-1)*(F(5)-1);
 412     2              THR(6)=SQU(I)-ONE(1)-ONE(4)-ONE(5)-TWO(3)-TWO(4)-TWO(10);
 413     2              CH(I)='1*4*5';
 414     2              SQU(I)=THR( 6);
                /******************************************************************/
 415     2      THR235: I=I+1;
 416     2              DO K=1 TO F(2);
 417     2    1         DO L=1 TO F(3);
 418     2    2         DO N=1 TO F(5);
 419     2    3         #A=SUM(A(*,K,L,*,N));   #B=SUM(B(*,K,L,*,N));
 421     2    3         CALL COMPUT;
 422     2    3         END;  END;  END;
 425     2              DEG(I)=(F(2)-1)*(F(3)-1)*(F(5)-1);
 426     2              THR(8)=SQU(I)-ONE(2)-ONE(3)-ONE(5)-TWO(5)-TWO(7)-TWO( 9);
 427     2              CH(I)='2*3*5';
 428     2              SQU(I)=THR( 8);
                /******************************************************************/
 429     2      THR245: I=I+1;
 430     2              DO K=1 TO F(2);
 431     2    1         DO M=1 TO F(4);
```

PGM: PROC OPTIONS(MAIN);

```
STMT LEVEL NEST

 432     2     2              DO N=1 TO F(5);
 433     2     3              #A=SUM(A(*,K,*,M,N));   #B=SUM(B(*,K,*,M,N));
 435     2     3              CALL COMPUT;
 436     2     3              END;  END;  END;
 439     2                    DEG(I)=(F(2)-1)*(F(4)-1)*(F(5)-1);
 440     2                    THR(9)=SQU(I)-ONE(2)-ONE(4)-ONE(5)-TWO(6)-TWO(7)-TWO(10);
 441     2                    CH(I)='2*4*5';
 442     2                    SQU(I)=THR(9);
                    /*****************************************************************/
 443     2          THR345:   I=I+1;
 444     2                    DO L=1 TO F(3);
 445     2     1              DO M=1 TO F(4);
 446     2     2              DO N=1 TO F(5);
 447     2     3              #A=SUM(A(*,*,L,M,N));   #B=SUM(B(*,*,L,M,N));
 449     2     3              CALL COMPUT;
 450     2     3              END;  END;  END;
 453     2                    DEG(I)=(F(3)-1)*(F(4)-1)*(F(5)-1);
 454     2                    THR(10)=SQU(I)-ONE(3)-ONE(4)-ONE(5)-TWO(8)-TWO(9)-TWO(10);
 455     2                    CH(I)='3*4*5';
 456     2                    SQU(I)=THR(10);
                    /*****************************************************************/
 457     2          FO1234:   I=I+1;
 458     2                    DO J=1 TO F(1);
 459     2     1              DO K=1 TO F(2);
 460     2     2              DO L=1 TO F(3);
 461     2     3              DO M=1 TO F(4);
 462     2     4              #A=SUM(A(J,K,L,M,*));   #B=SUM(B(J,K,L,M,*));
 464     2     4              CALL COMPUT;
 465     2     4              END;  END;  END;  END;
 469     2                    DEG(I)=(F(1)-1)*(F(2)-1)*(F(3)-1)*(F(4)-1);
 470     2                    FOU(1)=SQU(I)-ONE(1)-ONE(2)-ONE(3)-ONE(4)-TWO(1)-TWO(2)
                                -TWO(3)-TWO(5)-TWO(6)-TWO( 8)-THR(1)-THR(2)-THR(4)-THR( 7);
 471     2                    CH(I)='1*2*3*4';
 472     2                    SQU(I)=FOU(1);
                    /*****************************************************************/
 473     2          FO1235:   I=I+1;
 474     2                    DO J=1 TO F(1);
 475     2     1              DO K=1 TO F(2);
 476     2     2              DO L=1 TO F(3);
 477     2     3              DO N=1 TO F(5);
 478     2     4              #A=SUM(A(J,K,L,*,N));   #B=SUM(B(J,K,L,*,N));
 480     2     4              CALL COMPUT;
 481     2     4              END;  END;  END;  END;
 485     2                    DEG(I)=(F(1)-1)*(F(2)-1)*(F(3)-1)*(F(5)-1);
 486     2                    FOU(2)=SQU(I)-ONE(1)-ONE(2)-ONE(3)-ONE(5)-TWO(1)-TWO(2)
                                -TWO(4)-TWO(5)-TWO(7)-TWO( 9)-THR(1)-THR(3)-THR(5)-THR( 8);
 487     2                    CH(I)='1*2*3*5';
 488     2                    SQU(I)=FOU(2);
                    /*****************************************************************/
 489     2          FO1245:   I=I+1;
 490     2                    DO J=1 TO F(1);
 491     2     1              DO K=1 TO F(2);
 492     2     2              DO M=1 TO F(4);
 493     2     3              DO N=1 TO F(5);
 494     2     4              #A=SUM(A(J,K,*,M,N));   #B=SUM(B(J,K,*,M,N));
```

```
STMT LEVEL NEST

 496     2     4            CALL COMPUT;
 497     2     4            END;  END;  END;  END;
 501     2                  DEG(I)=(F(1)-1)*(F(2)-1)*(F(4)-1)*(F(5)-1);
 502     2                  FOU(3)=SQU(I)-ONE(1)-ONE(2)-ONE(4)-ONE(5)-TWO(1)-TWO(3)
                              -TWO(4)-TWO(6)-TWO(7)-TWO(10)-THR(2)-THR(3)-THR(6)-THR( 9);
 503     2                  CH(I)='1*2*4*5';
 504     2                  SQU(I)=FOU(3);
                  /*****************************************************************/
 505     2        F01345:   I=I+1;
 506     2                  DO J=1  TO F(1);
 507     2     1            DO L=1  TO F(3);
 508     2     2            DO M=1  TO F(4);
 509     2     3            DO N=1  TO F(5);
 510     2     4            #A=SUM(A(J,*,L,M,N));   #B=SUM(B(J,*,L,M,N));
 512     2     4            CALL COMPUT;
 513     2     4            END;  END;  END;  END;
 517     2                  DEG(I)=(F(1)-1)*(F(3)-1)*(F(4)-1)*(F(5)-1);
 518     2                  FOU(4)=SQU(I)-ONE(1)-ONE(3)-ONE(4)-ONE(5)-TWO(2)-TWO(3)
                              -TWO(4)-TWO(8)-TWO(9)-TWO(10)-THR(4)-THR(5)-THR(6)-THR(10);
 519     2                  CH(I)='1*3*4*5';
 520     2                  SQU(I)=FOU(4);
                  /*****************************************************************/
 521     2        F02345:   I=I+1;
 522     2                  DO K=1  TO F(2);
 523     2     1            DO L=1  TO F(3);
 524     2     2            DO M=1  TO F(4);
 525     2     3            DO N=1  TO F(5);
 526     2     4            #A=SUM(A(*,K,L,M,N));   #B=SUM(B(*,K,L,M,N));
 528     2     4            CALL COMPUT;
 529     2     4            END;  END;  END;  END;
 533     2                  DEG(I)=(F(2)-1)*(F(3)-1)*(F(4)-1)*(F(5)-1);
 534     2                  FOU(5)=SQU(I)-ONE(2)-ONE(3)-ONE(4)-ONE(5)-TWO(5)-TWO(6)
                              -TWO(7)-TWO(8)-TWO(9)-TWO(10)-THR(7)-THR(8)-THR(9)-THR(10);
 535     2                  CH(I)='2*3*4*5';
 536     2                  SQU(I)=FOU(5);
 537     2                  I=I+1;
 538     2                  DEG(I)=(F(1)-1)*(F(2)-1)*(F(3)-1)*(F(4)-1)*(F(5)-1);
 539     2                  SQU(I)=SQRINT-TWO(1)-TWO(2)-TWO(3)-TWO(4)-TWO(5)-TWO(6)
                           -TWO(7)-TWO(8)-TWO(9)-TWO(10)-THR(1)-THR(2)-THR(3)-THR(4)
                           -THR(5)-THR(6)-THR(7)-THR(8)-THR(9)-THR(10)-FOU(1)-FOU(2)
                           -FOU(3)-FOU(4)-FOU(5);
 540     2                  CH(I)='1*2*3*4*5';
                  /*****************************************************************/
 541     2        OUT:
                            PUT PAGE EDIT(TITLE)(COL(20),A);
 542     2                  PUT SKIP(2)EDIT(N_WAY,' WAY EXPERIENCE')(COL(45),F(2),A);
 543     2                  PUT SKIP(3);
 544     2                  PUT SKIP EDIT('FACTOR','CHI-SQUARE','DEGREE')(COL( 2),A,
                                 COL(24),A,COL(45),A);
 545     2        MAIN:     DO L=2 TO I;
 546     2     1            PUT SKIP(2) EDIT(CH(L),SQU(L),DEG(L))(COL(5),A,COL(25),F(7,3)
                                      ,COL(45),F(3));
 547     2     1            IF SQU(L)>=SET(DEG(L))  THEN PUT EDIT('*')(X(5),A);
 549     2     1            END MAIN;
 550     2                  PUT SKIP (3) EDIT('(AN * MEANS SIGNIFICANT AT ',
```

```
STMT LEVEL NEST

                                        ALPHA,' LEVEL.)')(COL(20),A,F(5,2),A);
551       2                     END;
                         /**********************************************************************/
552       1                     GO TO START;
553       1              EOJ:   END PGM;
```

CROSS-REFERENCE TABLE

DCL NO.	IDENTIFIER	REFERENCES
20	A	61,71,71,76,97,113,124,133,150,161,172,183,195,207,227,239,251,265 277,289,301,314,328,342,356,377,391,405,419,433,447,462,47[illegible],494,510 526
	ALPHA	7,550
20	B	52,70,70,79,97,114,125,134,151,162,173,184,196,208,226,24[illegible],252,266 278,290,302,315,329,343,357,378,392,406,42[illegible],434,448,463,47[illegible],495,511 527
22	CH	130,139,144,156,167,178,190,202,215,22[illegible],235,[illegible]47,259,272,[illegible],[illegible],30[illegible] 322,336,350,364,376,385,399,413,427,441,455,471,487,503,51[illegible],[illegible]3[illegible],54[illegible] 546
61	CLASS	
49	COMPAR	57
103	COMPUT	115,126,135,152,163,174,185,197,209,229,241,255,267,279,291,303,31[illegible] 330,344,358,379,393,407,421,435,449,464,480,496,512,52[illegible]
19	DATA	41,69
20	DEG	27,121,129,138,145,155,166,177,188,200,212,221,232,244,256,270,282 294,306,320,334,348,362,369,383,397,411,425,439,453,464,485,501,517 533,538,546,547
553	EOJ	11
2	F	9,14,16,16,16,16,16,20,20,20,20,20,20,20,20,20,20,19,1[illegible],1[illegible],1[illegible],1[illegible],1[illegible] 1[illegible],1[illegible],18,18,32,33,34,35,36,63,64,65,66,67,90,91,92,93,[illegible],[illegible],1[illegible],11[illegible] 1[illegible],1[illegible]2,121,121,121,121,121,125,129,132,136,145,145,14[illegible],[illegible],1[illegible],1[illegible] 171,177,181,182,18[illegible],188,193,194,200,200,205,206,212,212,2[illegible],[illegible],[illegible] 225,226,232,232,237,238,244,244,249,250,256,256,263,264,2[illegible],[illegible],27[illegible] 276,282,282,287,288,294,294,299,300,306,30[illegible],311,312,313,320,320,320 325,326,327,334,334,334,339,340,341,348,348,348,353,354,355,362,362 362,369,369,369,369,374,375,376,383,383,383,385,389,390,397,397,397 402,403,404,411,411,411,416,417,418,425,425,425,430,431,432,439,439 439,444,445,446,453,453,453,456,459,460,461,469,469,469,469,474,475 476,477,485,485,485,485,490,491,492,493,501,501,501,501,506,507,508 509,517,517,517,517,522,523,524,525,533,533,533,533,538,538,538,538 538
	F_A	80,105,105
	F_B	81,105,105
457	F01234	
473	F01235	

```
     PGM:     PROC OPTIONS(MAIN);

DCL NO.          IDENTIFIER         REFERENCES
489              F01245
505              F01345
521              F02345
24               FOU                31,470,472,486,488,502,504,518,520,534,536,539,539,539,539,539
     *********   I                  14,14,105,105,107,121,122,122,128,129,130,131,131,137,138,139,142
                                    142,143,144,145,148,148,154,155,156,159,159,165,166,167,170,170,176
                                    177,178,179,179,188,189,190,191,192,192,200,201,202,203,204,204,212
                                    213,214,215,216,218,219,220,221,224,224,232,233,234,235,236,236,244
                                    245,246,247,248,248,256,257,258,259,262,262,270,271,272,273,274,274
                                    282,283,284,285,286,286,294,295,296,297,298,298,306,307,308,309,310
                                    310,320,321,322,323,324,324,334,335,336,337,338,338,348,349,350,351
                                    352,352,362,363,364,365,368,369,370,373,373,383,384,385,386,387,387
                                    397,398,399,400,401,401,411,412,413,414,415,415,425,426,427,428,429
                                    429,439,440,441,442,443,443,453,454,455,456,457,457,469,470,471,472
                                    473,473,485,486,487,488,489,489,501,502,503,504,504,505,517,518,519
                                    520,521,521,533,534,535,536,537,537,538,539,540,545
     *********   I1                 32,37,38,41,63,68,69,70,70,71,71,108,113,114
     *********   I2                 33,37,38,41,64,68,69,70,70,71,71,109,113,114
     *********   I3                 34,37,38,41,65,68,69,70,70,71,71,110,113,114
     *********   I4                 35,37,38,41,66,68,69,70,70,71,71,111,113,114
     *********   I5                 36,37,38,41,67,68,69,70,70,71,71,112,113,114
     *********   I6                 38,41,68,69
3                #                  15,19
21               #A                 104,105,113,124,133,150,161,172,183,195,207,227,239,251,265,277,289
                                    301,314,328,342,356,377,391,405,419,433,447,462,478,494,510,526
21               #B                 104,105,114,125,134,151,162,173,184,196,208,228,240,252,266,278,290
                                    302,315,329,343,357,378,392,406,420,434,448,463,479,495,511,527
21               #P                 104,105,105,105,105
18               #REP               37,38,68,97
     *********   J                  123,124,125,181,183,184,193,195,196,225,227,228,253,255,256,311,314
                                    315,325,328,329,339,342,343,374,377,378,388,391,392,402,405,406,458
                                    462,463,474,476,479,490,494,495,506,510,511
     *********   JL                 8,8
     *********   K                  132,133,134,182,183,184,205,207,208,237,239,240,275,277,278,312,314
                                    315,326,328,329,353,355,357,375,377,378,416,419,420,430,433,434,459
                                    462,463,475,475,479,491,494,495,522,526,527
```

```
        PGM:     PROC OPTIONS(MAIN);

DCL NO.          IDENTIFIER              REFERENCES

     ********* L                         96,96,149,150,151,194,195,196,206,207,208,249,251,252,287,289,290
                                         313,314,315,340,342,343,354,356,357,389,391,392,417,419,420,444,447
                                         448,460,462,463,476,478,479,507,510,511,523,526,527,545,545,546,546
                                         547,547

19   ********* LIST                      40,41,49,49,51,52,52,53,60

     ********* LL                        48,49,49,51,52,52,53,54,54,55,56

     ********* LLL                       59,60

     ********* M                         25,39,39,40,41,48,59,80,81,160,161,162,226,227,228,238,239,240,250
                                         251,252,299,301,302,327,328,329,341,342,343,355,356,357,403,405,406
                                         431,433,434,445,447,448,461,462,463,492,494,495,508,510,511,524,526
                                         527

545              MAIN

2                MAX#REP                 15,16,19

     ********* MD                        60,69,84

23               MJ                      90,91,92,93,94,96,97,97,97,97,97,97,97,97,97,97,97,97,97,97,97

     ********* N                         85,87,171,172,173,264,265,266,276,277,278,288,289,290,300,301,302
                                         376,377,378,390,391,392,404,405,406,418,419,420,432,433,434,446,447
                                         448,477,478,479,493,494,495,509,510,511,525,526,527

2                N_WAY                   13,14,83,86,96,140,157,168,216,260,366,542

54               OK                      50

24               ONE                     28,128,137,154,165,176,180,180,180,180,189,189,189,201,[illegible],215
                                         233,233,245,245,257,257,271,271,283,283,295,295,307,307,321,321,321
                                         335,335,335,349,349,349,363,363,363,384,384,384,398,398,[illegible]
                                         412,426,426,426,440,440,440,454,454,454,470,470,470,470,[illegible]
                                         486,502,502,502,502,518,518,518,518,534,534,534,534

122              ONE1

131              ONE2

148              ONE3

159              ONE4

170              ONE5

541              OUT                     146,222,371

1                PGM

5                SET                     8,547

48               SORT
```

PGM: PROC OPTIONS(MAIN);

DCL NO.	IDENTIFIER	REFERENCES
4	SOURCE	6,7,8
	SQRINT	180,219,368,539
20	SQU	25,105,105,128,137,143,143,143,143,154,165,176,180,189,191,201,203 213,214,219,233,234,245,246,257,258,271,273,283,285,295,297,307,309 321,323,335,337,349,351,363,365,366,384,386,398,400,412,414,426,428 440,442,454,456,470,472,486,488,502,504,518,520,534,536,539,546,547
12	START	552
	STORE	51,53
	SUM	78,79,124,125,133,134,150,151,161,162,172,173,183,184,195,196,207 208,227,228,239,240,251,252,265,266,277,278,289,290,301,302,314,315 328,329,342,343,356,357,377,378,391,392,405,406,419,420,433,434,447 448,462,463,478,479,494,495,510,511,526,527
	SYSIN	10,12,13,14,15,37,40
	SYSPRINT	82,83,84,85,87,89,95,96,97,541,542,543,544,546,548,550
24	THR	30,321,323,335,337,349,351,363,365,366,366,368,368,384,386,398,400 412,414,426,428,440,442,454,456,470,470,470,470,486,486,486,486,502 502,502,502,518,518,518,518,534,534,534,534,539,539,539,539,539,539 539,539,539,539
310	THR123	261
324	THR124	
373	THR125	
338	THR134	
387	THR135	
401	THR145	
352	THR234	
415	THR235	
429	THR245	
443	THR345	
2	TITLE	12,82,541
107	TOTAL	
	TT_A	78,80
	TT_B	79,81

DCL NO.	IDENTIFIER	REFERENCES
24	TWO	29,189,191,201,203,213,214,219,219,219,233,234,245,245,257,259,271 273,283,285,295,297,307,309,321,321,321,335,335,335,349,349,349,363 363,363,366,366,368,368,368,368,384,384,384,398,398,398,412,412,412 426,426,426,440,440,440,454,454,454,470,470,470,470,470,470,486,486 486,486,486,486,502,502,502,502,502,502,518,518,518,518,518,518,534 534,534,534,534,534,539,539,539,539,539,539,539,539,539,539
179	TW012	156,169
192	TW013	
224	TW014	
262	TW015	
204	TW023	
236	TW024	
274	TW025	
248	TW034	
286	TW035	
298	TW045	
108	W1	
109	W2	
110	W3	
111	W4	
112	W5	
32	Z11	
33	Z12	
34	Z13	
35	Z14	
36	Z15	
38	Z16	
63	Z21	
64	Z22	
65	Z23	

PGM: PROC OPTIONS(MAIN);

DCL NO.	IDENTIFIER	REFERENCES
66	Z24	
67	Z25	
68	Z26	

```
PGM:    PROC OPTIONS(MAIN);

STORAGE REQUIREMENTS.
---------------------

THE STORAGE AREA FOR THE PROCEDURE LABELLED PGM IS 536 BYTES LONG.
THE STORAGE AREA FOR THE ON UNIT AT STATEMENT NO. 10 IS 184 BYTES LONG.
THE STORAGE AREA FOR THE BEGIN BLOCK AT STATEMENT NO. 17 IS 1960 BYTES LONG.
THE STORAGE AREA (IN STATIC) FOR THE PROCEDURE LABELLED COMPUT IS 184 BYTES LONG.
THE PROGRAM CSECT IS NAMED PGM       AND IS 41374 BYTES LONG.
THE STATIC  CSECT IS NAMED ****PGMA AND IS 4224 BYTES LONG.

*STATISTICS*     SOURCE RECORDS =      501.PROG TEXT STMNTS =      553.OBJECT BYTES =    41374
```

EXAMPLE OF THE APPLICATION OF WILSON'S DISTRIBUTION FREE ANOVA TEST

3 WAY EXPERIENCE
MEDIAN POINT IS 30

FACTOR 1	FACTOR 2	FACTOR 3	TOTAL #	# OF ABOVE	# OF BELO
1	1	1	6	2	4
1	1	2	6	3	3
1	2	1	6	3	3
1	2	2	6	3	3
2	1	1	6	4	2
2	1	2	6	3	3
2	2	1	6	5	1
2	2	2	6	2	4
3	1	1	6	4	2
3	1	2	6	2	4
3	2	1	6	3	3
3	2	2	6	3	3

FACTOR	CHI-SQUARE	D.F.
1	0.778	2
2	0.056	1
3	1.390	1
1*2	0.111	2
1*3	2.113	2
2*3	0.056	1
1*2*3	1.446	2

AN ASTERISK * AFTER THE D.F. NUMBER INDICATES SIGNIFICANCE AT THE α=0.05 LEVEL.

REFERENCES

[1] Miller, I. and Freund, J. Probability and Statistics for Engineers, Prentice-Hall, 1965.

[2] Ibid, p. 279.

[3] Winer, B. J. Statistical Principles in Experimental Design, McGraw-Hill, 1962; p. 95.

[4] Miller, I. and Freund, J. Probability and Statistics for Engineers, Prentice-Hall, 1962; p. 410.

[5] Ibid., p. 316.

[6] Wilson, Kellog V. "A Distribution-Free Test of Analysis of Variance Hypotheses," Psychological Bulletin, Vol. 53.

[7] Hicks, C.R. Fundamental Concepts in the Design of Experiments, (Second Edition), Holt, Rinehart, and Winston, 1972; Problem 5.5, p.84.

[8] Kendall, M. G. Rank Correlation Methods, Griffin, 1948. Chapters 3 and 4.

[9] Miller, I. and Freund, J. Probability and Statistics for Engineers, Prentice-Hall, 1965; Chapter 12.

[10] Ibid, p. 413.

[11] Wilks, S.S. "Determination of Sample Sizes for Setting Tolerance Limits," Annals of Mathematical Statistics, Vol. 12, 1941; pp. 91-96.

[12] Cochran, W. G. Sampling Techniques (Second Edition), Wiley, 1963.

[13] Owen, D. B. Handbook of Statistical Tables, Addison -Wesley, 1962; p. 331.

[14] Miller, I. and Freund, J. Probability and Statistics For Engineers, Prentice-Hall, 1965; pp. 407-408.

[15] Snedecor, G. W. Statistical Methods, (4th Edition) Iowa State College Press, Ames, Iowa; 1948.

[16] Yates, F. and Fisher, R. A. Statistical Tables for Biological, Agricultural, and Medical Research; Hafner, 1949; p. 33.