8106046

RICHARDS, WINSTON A.

THE RANDOMIZATION ANALYSIS OF COVARIANCE AND CHANGE-
OVER DESIGNS

Iowa State University PH.D. 1980

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark ✓.

1. Glossy photographs ________
2. Colored illustrations ________
3. Photographs with dark background ________
4. Illustrations are poor copy ________
5. Print shows through as there is text on both sides of page ________
6. Indistinct, broken or small print on several pages ________
7. Tightly bound copy with print lost in spine ________
8. Computer printout pages with indistinct print ✓
9. Page(s) ________ lacking when material received, and not available from school or author
10. Page(s) ________ seem to be missing in numbering only as text follows
11. Poor carbon copy ________
12. Not original copy, several pages with blurred type ________
13. Appendix pages are poor copy ________
14. Original copy with light type ________
15. Curling and wrinkled pages ________
16. Other ____

Randomization analysis of covariance

and change-over designs

by

Winston A. Richards

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1980

TABLE OF CONTENTS

Page

1. INTRODUCTION

This thesis is concerned with the analysis of comparative experiments. The general idea of these is that one has a collection of experimental units and one has a number of treatments that one wishes to compare. A comparative experiment consists of assigning each of the treatments to a subset of the collection of experimental units. The outcome of this process in the simplest case is that one will have observations, say $\{y_{jk}\}$, where the subscript j indexes the treatments and the subscript k indexes experimental units within treatments. The task of the experimenter and of the statistician consists of two parts: (a) to decide on an assignment of treatments to experimental units and (b) to decide on modes of interpreting the resultant data.

We assume that the experimental units will have some intrinsic responses under some defined basic conditions. We also assume that the experimenter has some control over the treatments themselves of their application to the experimental units and we wish to determine what differential effects, if any, the treatment exert or appear to exert on the variation of responses of the experimental units. We hope, with some degree of optimism, that the relationship between the responses and the treatment effects can be described in a simple functional algebraic equation, the terms in which represent appropriate variables which are defined over some (limited) specified ranges. By examining the algebraic function we hope to learn more about the underlying true relationships and appreciate the differential affects, if any, on the response variable that are produced by changes in the treatment quantities.

In our developments we shall assume the algebraic functional relationship is linear in unknown parameters which represent the effects of interest. These unknown parameters are estimated under certain other assumptions and the fitted equation is obtained from the data by the method of least squares. The theory and method of least squares is well-known and so will not be discussed her per se. Associated with the least squares procedure is the analysis of variance in which we have a Pythagorean partition of the metric $\Sigma y_1^2 = y'y$ into "orthogonal" components corresponding to the treatment effects and the error or intrinsic variability of the experimental units. The data is then "interpreted" by means of various statistical tests and inferences are made about the treatment effects.

In the situation where the distribution of the underlying population of interest is assumed, e.g., normal distribution say, the validity of the procedures used, or the probabilistic functions based on the resulting sampling distribution will depend, perhaps critically, on the assumptions. Violations of the assumptions may have very serious consequence on the correctness of justification of the inference or decision involved. In the practical world, however, one's experience or prior information about a data situation may be quite inadequate to justify the assumptions. Besides, in many cases the sample size or population of interest may be too small and the appeal to general large sample theory may be without basis. In these situations, distribution free or nonparametric methods may provide suitable procedures for analyzing the data, so let us now try to develop and clarify in a coherent, if somewhat brief way, the logic of the methodology.

Fisher (1926, and in his now classical book The Design of Experiments, 1935) put forward an idea that has dominated comparative experiment methodology for the past 50 years. This idea is that one should examine as well as one can, the experimental material: one should then "arrange" the experimental units in a structure, such as blocks or a two-way array, for example. Then one should use a validated random process, such as a tossing apparatus, to decide the actual assignment of treatments to units. If there were no basis for grouping the experimental units into, say, blocks, one would use the trivial structure that the experimental units are undifferentiated, and one would merely assign treatments completely at random to the units, realizing, of course, that an experimental unit can be assigned only one treatment. This is known, generally, as the completely randomized design (CRD). If there are t treatments, and the experimental units are partitioned into, say, r sets of t units, then the randomization is that the treatments are to be assigned at random to units within each block. This gives the ordinary randomized block design (RBD). Natural extension of the basic idea give incomplete block designs, split-plot designs, Latin square designs and so on.

The randomization process leads, obviously, to a particular realization of the treatment-unit assignment. The performance of the treatment protocols and observation of the outcome leads to the data that the experimenter and the statistician have to interpret. It is at this point that there are strongly different opinions. We shall abbreviate sharply what could be a very long discussion of ideas.

At one pole in the controversy are the statisticians of Bayesian outlook, who take the view that all inference must be done by using three steps. One devises, by some means, e.g., by analogy with previous studies judged to be similar, a parametric probability structure for data sets that can be viewed as representing data that one might obtain in repetitions of the actual study, or that is taken to represent the beliefs of the experimenter on how the data are related to a (belief) probability structure. This probability structure will naturally contain unknown parameters, e.g., treatment effects and parameters that represent the (belief, say) distribution of the contributions to each separate observation of the separate units. The experimenter is to "think out" a representation of his beliefs that is completely specified. Then the task of inference is accomplished by using Bayes Theorem, which gives a (belief) probability structure for the unknowns conditional on the actual data, this being known as the posterior (belief) distribution of the relevant unknown parameters.

The approach of the previous paragraph has had over the past two centuries considerable appeal to some thinkers. It has also been regarded by others as a misleading and useless process. Fisher (1926) was a leader in this outlook. In his 1935 book he rejected the approach and put forward, for the comparative experiment, the idea of randomization. In his simple lady-tasting-tea experiment, Fisher (1935) took the view that the only way to obtain valid evidence was to use randomization in the conduct of the experiment, and then insisted that the value of the statistical test came from the validity brought about by the physical act of randomization. Then in various publications too numerous to cite, Fisher applied the same sort of idea to the ordinary comparative experiment. The essential ideas that ran through the development are that tests of significance for treatment effects would be done by analysis of variance, and a critical aspect of this is that the error mean square which one used as a base for looking at the treatment sum of squares or mean square should be fair, in the sense that the error mean square should measure the variability one would encounter with treatment means and differences in hypothetical repetitions of the randomization. In fact, the expectation under randomization of the one should be a known multiple (such as 2/r, where r is the number of occurrences of each treatment) of the expectation of what is computed as the error mean square. It is curious perhaps, that having insisted on this as a basic requirement, Fisher, clearly, held that one could then proceed with the analysis of the experiment by means of the ordinary additive linear model, with errors that are independently Gaussian with mean zero and the same variance.

The approach that has been followed by some, e.g., Kempthorne (1952, 1955, 1957) and others, is that the use of randomization in choosing an experimental plan should be made the basis for the whole of the ensuing theory of obtaining estimates, variances of estimates, estimated variances, tests of significance and statistical intervals for true treatment comparisons. This general approach has been termed 'Finite Model Theory' or 'Randomization Analysis'.

The general ideas of randomization of the simple widely used randomized design have been worked out largely, and it appears that the ordinary Gauss-Markov normal (GMN) linear model theory provides reasonable approximations to what complete randomization analysis for the common simple designs would give. One would still entertain, however, the possibility that one should, as a final step, rely solely on randomization considerations.

It is the case, however, that the type of validation of normal theory has been accomplished only for a very special class of experimental data configurations: namely, that class in which one has experimental units classified according to a partitional frame and then one has a single arithmetic measurement, the yield, on each experimental unit.

In general, within the domain of frequency statistics and associated randomization analysis, it is relevant to attempt two related targets:

(a) To make a test of significance of the associated null hypothesis of there being no treatment effects; given that we can do this, we can make a test of significance of any set of hypothesized treatment differences, say, $t_k - t_{k'} = \Delta_{k,k'}$, by adjusting the observations to the associated null hypothesis, i.e., forming $\tilde{y}_{k(j)} = y_{k(j)} - t_k$, and then testing for absence of treatment effects in these adjusted observations.

(b) To develop standard errors, which are estimates of standard deviations under randomizations, for treatment difference estimators; if these are obtained, we would like then to obtain statistical intervals, just as we do with the Gauss-Markov normal linear model theory.

From the beginning of the use of linear models for analyzing experiments, the idea of the use of measured, so-called concomitant, variation has been pursued. The idea may be explained by a simple human nutritional experiment. Suppose we are concerned about the physical growth of children from the 6^{th} birthday to the 7^{th} birthday, and we have the idea that supplementation of the diet with, say, 200 units of vitamin D, will improve this growth. Then we shall make a comparative experiment, in which say r children are not given supplementation, while r children are given supplementation. We shall observe achieved growth on the 7^{th} birthday, and we can denote this by $\{y_{jk}\}$, j denoting treatment indexed by j=1 or j=2, and k denoting individual within treatment, k=1,2, ..., r. Now it is obvious that children who were relatively bigger at the 6^{th} birthday will tend to be relatively bigger at the 7^{th} birthday. So a natural idea is to obtain x_{jk}, where x_{jk} is the achieved growth at the 6^{th} birthday of child (jk). It is then natural to consider the model

$$y_{jk} = \mu + t_j + \beta x_{jk} + e_{jk}$$

in which μ is some constant, $\{t_j\}$ represent additive treatments combinations, and βx_{jk} represents a contribution to y by the initial, or concomitant, variable x. The use of linear models of this type was given, first perhaps, by Wishart (1928). The idea is given great prominence in The Design of Experiments by Fisher.

As mentioned above, Fisher believed that having used "proper" randomization, one could then proceed with Gauss-Markov normal linear model theory. There are, however, real obscurities with regard to this step. At this point of the exposition, we merely note that is we envisage unknown unit effects, $\{u_i\}$, then we are to apply randomization ideas to the vectors $Qu_i, x_i\}$ where x_i is the value of the concomitant variable for the i-th unit. So what is involved, then, is the application of randomization analysis ideas to a set of vectors, say, $\{(u_i, x_i)\}$ and not to a set of scalars, say $\{u_i\}$, as is the case in the absence of knowledge of the concomitant variation.

In the analysis of multiple covariance situations, we envisage the vectors $\{(u_i, x_i)\}$, where we have the response variate with unknown unit effects u_i, and a multiple concomitant x_i (a p-tuple). The randomization procedure makes available to the experimenter a random partitioning of the $M = N!/(r!)^t$ partitionings of the $N = rt$ experimental unit vectors (u_i, x_i), (u_2, x_2), ..., (u_t, x_t) into t groups of r. Any statistic of interest, the ratio of adjusted treatment mean square to adjusted error mean square, for example, will be a function of the (u_i, x_i) and the unknown parameter vector of treatments $\tau =)\tau_1, \tau_2, \ldots, \tau_t)$ $\tau \epsilon (R^t)$. We will assume that this function has the simple linear form under additivity and shall investigate it under the conditions of the null hypothesis that there are no differences between the treatments, i.e., $\tau_1 = \tau_2 = \ldots = \tau_t$ = constant.

The consequences of including the covariate is that in the resulting analysis of variance procedure we obtain: (a) an adjusted mean square for residuals, say E^* and (b) an adjusted mean square for treatments, say T^*. However, it is not the case that the expectation over randomizations of T^* is equal to the expectation over randomization of E^*. The failure of this property to hold was stated by Kempthorne (1952, 1977). D.R. Cox (1956) examined this situation. His work is discussed later in this thesis. Kempthorne (1977) cautioned that there may be non-simple problems associated with the concomitant.

As a background for the development of this thesis, in Parts 1 and 2 of Chapter 2, we will give a brief account of randomization analysis in the absence of concomitant variation for the CRD and RBD. The statistics involved in our development include the following:

(a) estimates of treatment differences $(\bar{y}_k - \bar{y}_{k'})$ where $\bar{y}_k$ is the average of the observed responses of the units which receive treatment k;

(b) the ratio of the treatment mean square to the error mean square in the analysis of variance. This is usually denoted as "F" and taken to have an F-distribution under GMN conditions;

(c) the ratio of the treatment sum of squares to the sum of the treatment sum of squares and the error sum of squares. This ratio is denoted as W and is taken to have a beta distribution under GMN conditions.

The behavior of these statistics has been examined by Pitman, Ogawa and others.

In Part 2 of Chapter 2, we will give a brief review of Pitman's work on the moments of W and of Ogawa's work on the behavior of F for the situation where technical errors are included in the observations. We will also discuss briefly the work of H.J. Arnold for the multivariate situation.

One of our overall tasks is to attempt to develop understanding of randomization analysis for the analysis of covariance situation in which the values of one or more concomitant variates are known. In particular we will examine the CRD. We will also give an account of the work of J. Robinson on the RBD. The analysis of covariance tables for the CRD is as follows:

Source	yy	xy	xx
Treatments	T_{yy}	T_{xy}	T_{xx}
Residual	R_{yy}	R_{xy}	R_{xx}
Total	G_{yy}	G_{xy}	G_{xx}

The statistics involved in our investigation of the analysis of covariance include the following:

(a) $\hat{\beta} = R_{xx}^{-1}R_{xy}$;

(b) the adjusted estimates of treatment differences

$$\bar{y}_k - \bar{y}_{k'} - (\bar{x}_k - \bar{x}_{k'})\hat{\beta} \;;$$

(c) the estimate of the variances of the estimates given in (b);

(d) the ratio of the adjusted treatment sum of squares to the sum of the adjusted treatment sum of squares and the adjusted error sum of squares,

$$W^{*} = \frac{\{T_{yy} + R'_{xy}R_{xy}^{-1}R_{xy} - G'_{xy}G_{xx}^{-1}G_{xy}\}}{G_{yy\cdot} - G'_{xy}G_{xx}^{-1}G_{xy}} \ .$$

The examination of these statistics leads us into their randomization distributions and associated randomization tests. These distributions are necessarily discrete and are not available in useful mathematical form. The cardinality of the associated sample spaces increases quite rapidly with the number of available units. For the CRD with t treatments and $N = rt$ units the cardinality is $(rt!)/(r!)^t)$. For the simple RBD with r blocks and t treatments, the cardinality is $(t!)^r$. To get some idea of limiting distributions, we need to make use of the limit theory developed by Wald and Wolfowitz, Noether, Cramer and others. In Part 3 of Chapter 2 we will give some of the results of these workers.

Chapter 3 is presented in 7 parts. In Part 1 we will give a brief account of the analysis of covariance in terms of linear regression under GMN error assumptions. We will then relate the procedure to classificatory models with concomitant variables. In Part 2 we shall attempt to clarify the logic of the randomization procedure and present a formal development of the univariate analysis of covariance in the CRD. In Part 3 we shall obtain the expectations under randomization of various statistics that are important for our purposes. We shall also review the related work of D.R. Cox on weighted randomization. The statistics that we will be considering involve the components of the "reduced normal equations," R_{xx} and R_{xy}, which we have already mentioned above in the analysis of covariance table. The coefficients of variation of these components, assuming, of course, that their expectations are different from zero, each has order of magnitude $O(N^{-1})$, where N is the number of experimental units. This implies that we may consider the components as having asymptotically constant values. In Part 4 we will use ideas of order in probability to approximate some statistics involving these components by Taylor-Mclaurin series. We shall thus obtain the expectations of these statistics to order $O(N^{-2})$. We will then extend the results from the single covariate situation to the analysis of multiple covariation. In Part 5 we shall apply the limit theory ideas previously mentioned to obtain normal law approximations to the limiting distribution under randomization theory of various statistics. In Part 6 we will give an account of the work of J. Robinson on the RBD and a brief discussion of the work of others on the analysis of covariance. In Part 7 we shall present a brief summary of the chapter.

In Chapter 4 of this thesis, we will investigate a different class of problems which have some algebraic features in common with the analysis of covariance. This class of problems involves some designs for treatments applied in sequence to experimental units which we will now discuss in a general way.

A large class of designs called 'change-over' or 'switch-over' designs have been developed by various workers and are widely used in many areas of research, areas such as animal science, agronomy, soil science, medicine and education. Now in these areas of interest, the units may be natural units such as human beings, fields in an agronomic survey, or natural aggregates of units like families, villages, school children of specified ages, victims of a disease of psychological disorder, single plants, rows of plants, or plots of specified size in a field. The treatments may be different drugs, diets, rotation of crops, or sequences of courses in a remedial educational program. To digress a bit from our theme, we note that in planning these experiments, one inevitably has to make contingencies for obvious practical problems. For example, in long term studies involving wide assortments of measurements, problems arise in record keeping and adequate storage of data. Turnover of personnel, change in interests of personnel, change in status of knowledge and techniques over the course of a program may be difficult to assimilate once a project is launched. Moreover, it is sometimes difficult to prepare and maintain a list of experimental units, and in addition to this, patients or subjects may be remiss in complying with or adhering to the regimen described in the protocol of the experiment. Dropout and selective survival may change the composition of the sample and increase heterogeneity. Survival rate may be correlated with measurement variables. In selective sampling, because of repeated participation which is required of subjects, samples may not be representative of the target population. Hence these samples from the outset may run the risk of being selectively biased. For example, the people who volunteer for a certain kind of psychological study may tend to be of higher average intelligence and tend to be of higher socioeconomic status. Such selective sampling may impair generalization. Having made these observations, let us return to our main theme.

In these situations, each subject or experimental unit receives a sequence of different treatments over fixed and usually equal intervals of time. It is quite likely that a treatment applied in one period would interfere with a treatment applied in the subsequent period. Wherever possible, an interval of time is allowed to elapse between treatment periods so that any lingering effects or traces of the treatment in one period would disappear or be minimized before the next treatment in that sequence is applied. In many situations it is difficult to eliminate these residual effects satisfactorily. We usually assume that this interference is simple, but sometimes it persists to more than one subsequent treatment periods. It is used to assume that the actual period in which the treatment is applied does not alter the residual effects of different orders nor do the subsequent treatments, i.e., to assume that the residual effects are additive and that one can use the model

$$y = \mu + \text{subject effect} + \text{period effect} + \text{treatment effect} + \text{residual effects} + \text{error}.$$

A "classical" experiment of the 'change-over' type is that of Cochran, Autrey and Cannon (1941) in which the aim of the experiment was to compare three nutritional treatments for dairy cows. The idea is that there is considerable variation among cows, so a possibility is to apply the several treatments to each cow, necessarily in some sequence, such as treatment 2 in one period, treatment 3 in a following second period, and treatment 1 in a following third period. Here, clearly, all one can do with regard to randomization is to apply at random different treatment sequences to different cows, perhaps after blocking the cows in an attempt to remove some variation from treatment comparisons. It seems natural and appropriate, then, to consider randomization analysis of such experiments. In thinking about the interpretation or logic of such experiments we may envisage the possibility of 'residual' effects, i.e., that a treatment imposed in one period produces an effect in a subsequent period. A simple way to attempt to accommodate such a possibility is to insert an additional term in the linear model to model residual effects. The effect of this is that one is led to a model somewhat like that of covariance which may be written $y_{hiju} = \mu + \theta_g + t_i + r_j + s_u + e_{hiju}$, where y_{hiju} denote the response, under additivity, μ is the grand mean, θ_h is the effect of the h-th period, t_i is the direct effect of treatment i applied in period h, r_j is the residual effect of treatment j, s_u the effect of subject u, and e_{hiju} the error term.

We may be interested in making comparisons among the direct treatment effects as well as the combination of direct effects plus residual effects. In the former case, as well as the latter, the algebraic problems that we have to solve are similar in form to those in the classical analysis of covariance situation.

We assume that residual effects are fixed effects, that is, the residual effects are not affected by the direct treatment effects. However, we face a problem in the fact that the experimenter is powerless to determine how they should be assigned to the experimental units, once he had decided upon the assignment of direct treatment effects to the units. However, formally, from the regression viewpoint, the residual effects generate a column space which we may wish to eliminate in order to obtain information on the direct treatment effects. We will discuss the analysis of covariance from the viewpoint of regression in Chapter 3 and in that setting the previous treatments may, with some license, be loosely considered as covariates. So, formally, from the algebraic viewpoint, the situation may be considered as a case of the analysis of covariance. Logically, however, one may take strong objection to this claim. Since the residual effects are attributes, not of the basic units, but of the treatment plan, they violate two basic requirements of the classical analysis of covariance situation:

1. The covariates are concomitant observations of the basic units with which they may be correlated. An adjustment for the covariates should therefore lead to a reduction in the error term and consequently to a more sensitive analysis. The randomized block design with blocks as the covariate is such a case.
2. The covariates must not influence or be influenced by the treatments.

We will not debate the claim here as the logical differences are quite obvious.

So in Chapter 4 we hope to present a unified account of change-over designs (cods) from the viewpoint of randomization theory. The chapter will be presented in 3 parts. In Part 1 as a background for our development we will include the following: (a) a brief account of Latin square (LS) designs on which the various designs for sequences of treatments in our investigations are based; (b) a brief review of the designs of some workers in the area; (c) a general discussion of the finite population model and the GMN model in LS designs and (d) a theorem of Wilk and Kempthorne which states the expectations, under randomization, of various mean squares in the LS design. In Part 2 we shall investigate the cods in some detail. At first we will specify some of the conditions of balance for the whole class of designs in general which we will need. We will make use of these conditions in our attempt to develop a detailed exposition of each of the following three subclasses; (a) the cod in the absence of residual treatment effects; (b) the cod in the presence of residual treatment effects; and (c) the extra period cod. In Part 3 we summarize the expectations of mean squares for the three subclasses in the various associated analysis of variance tables. We then close the chapter with a brief discussion and conclusion.

Finally, we include 5 Appendices to which we refer in Chapters 3 and 4. In Appendix A we will present the concept of order in probability which we will use to determine approximation to various statistics of interest in Chapter 3. In Appendix B we will present conditions under which a square matrix may be approximated by a convergent series of square matrices. We make use of these conditions to determine approximations to various statistics which involve small powers and inverse powers of the scalar, vector and matrix components $\{R\}$ of the "reduced normal equations," or the residual sum of squares and cross products in the analysis of covariance table for the analysis of multiple covariation. In Appendix C we derive the means, variances and covariances, under randomization, of the "residual" sum of squares and cross products. For these we use the properties of the "design random variables" which determine the randomization procedures completely. In Appendix D we calculate the expected values of various sum of squares for balanced incomplete change-over and extra-period change-over designs. In Appendix E we will give some

This thesis is strongly oriented to the question of whether analyses of variance have the property of unbiasedness. For completeness, we give quotations from Yates (1936).

> "The method of least squares on which the theory of the last section is based involves the assumptions that the experimental values are uncorrelated and are normally distributed. In reality neither assumption is true. In actual experimental work the process of randomization is so arranged that if the treatments produce no effect, then with any one set of experimental values the mean value of the error mean square in all the arrangements from which a random selection is made is equal to the mean value of the treatment mean square."

> "A process of randomization which will equalise the mean values of the treatment and error mean squares does not always exist, and some otherwise admirable experimental arrangements fail on this criterion."

2. REVIEW OF LITERATURE ON RANDOMIZATION THEORY

2.1 Basic Concepts

This chapter is divided into 3 parts. In Part 1 we attempt to clarify some of the underlying concepts which form the formal vocabulary for randomization analysis. This includes a general framework in which our conceptual populations may be defined. In Part 2 we present a brief review of some of the work which has been done by various workers on the analysis of variance procedure in the absence of known concomitant variation from the viewpoint of randomization analysis. In Part 3 we give various conditions under which the limiting distribution of statistics from sequences of populations converge to normal law distributions.

In this thesis it is to be understood that we are attempting to make statistical inference in situations where the probability distribution of the observations is unknown. Incorporation of randomization into the experimental plan gives a strong basis for statistical inference, particularly if additivity, as defined later, holds. In general, one wishes not only to estimate treatment differences but also to obtain a measure of their accuracy. Since the validity of a conclusion depends on the validity of the assumptions on which it is based, the inclusion of randomization allows one to make more accurate inferences than are possible without its use, more accurate in the sense that the probability statements and associated statistical inferences have definitive relations to what is conceptually observable in a situation.

To specify what is conceptually observable as well as to clarify some aspects of our investigations we now present some statistical ideas and definitions that are fundamental to the development of this thesis.

2.1.1 General Ideas for Randomization Analysis

The basic idea of using experimental randomization was given by Fisher in the 1920s and exposited in his book (Fisher 1935). An attempt to formulate a general framework for randomization analysis was made by Kempthorne (1952, 1955, 1957). We modify Kempthorne's procedure for our purposes. In the case of the CRD, say, each experimental unit can receive only one treatment. In the case of the cod, each experimental unit can receive only one sequence of treatments from the set of (sequences of) treatments. However, our conceptual frame must contain the totality of all possible unit by treatment (sequence) combinations. Our randomization procedure determines which treatment (sequence) is applied to the

For the observation on any experimental unit we may assume that the true response depends only on the unit and the treatment (sequence) it receives.

Let $x_{i^*j^*}$ denote the j-th replicate (vector) observation obtained from treatment (sequence) i* . Then i*j* corresponds to some value of k, the index of the experimental units. We regard the set $\{x_{i^*j^*}\}$ as a restricted random sample from the set of random variables $\{y_{ik}\}$ which constitute our population, where y_{ik} is the response of unit i under treatment k.

Ignoring obvious technical errors [measurement error, observer error, etc.], our problem lies in the fact that we can observe only a subset of the $\{y_{ij}\}$ and hence our inference is influenced by sample variation. The restrictions of the experimental design have to be taken account of to obtain a statistical model for the observations $\{x_{i^*j^*}\}$ in terms of the parameters defined on the elements of $\{y_{ij}\}$ and random variables which describe the restrictions of the design. Given the relevant assumptions, one can then obtain and derive algebraically the relevant statistical knowledge and properties of the experiment.

We now present three of the basic statistical concepts by definition which form a basis for the development of this thesis.

2.1.2 Additivity

The term additivity has been used quite often without precise meaning. For example, the model $Y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + e_{ijk}$ is an additive model in the sense that the right hand side is the sum of a number of parameters as opposed to product or a mixture of sums and products. Also, if the symbols i,j are used to denote the levels of factors, it is said that additivity of the factors holds if the interaction term (ab) is unnecessary. The concept of additivity used in randomization is quite different.

Suppose we have N units x_i $(i = 1,2,\ldots,N)$; with x_i as the response or yield for the i-th unit under some basic conditions; then we say that additivity in the strict sense holds if the response of that unit under treatment k is given by

$$y_{ik} = x_i + t_k \qquad \text{for some } \{t_k\}$$

Additivity in the broad sense obtains if we have

$$y_{ik} = x_i + t_k + e_{ik} \quad ,$$

where the e_{ik} are independent random variables over all i and k.

2.1.3 Unbiasedness in Designs

The term "unbiased" occurs with various meanings in the literature. We have, for example, 'unbiased' estimates of estimable functions, 'unbiased' confidence sets, 'unbiased' tests, etc. For the purposes of this thesis, our interest lies in the meaning or use of the word with regard to properties of the design of linear models and related estimable functions. A linear model design was called 'unbiased' by Yates (1936) if the expectation under randomization of the treatment mean square in the analysis of variance is equal to the error mean square when the null hypothesis is true, i.e., in the absence of differential treatment effects. To remove any ambiguities caused by the wide usage of the term, perhaps the qualified "AOV unbiased," i.e., "Analysis of Variance unbiased," suggested by Kempthorne, may be more appropriate. Let E_R denote expectation under randomization. Formally we would like our designs to enable us to calculate from the observations the following:

(1) contrasts among the treatment effects $\{t_k\}$, so that if $\{t_k\}$ are the estimates of t_k, we will have

$$E_R(\Sigma_k c_k t_k) = \Sigma_k c_k t_k, \text{ where } \Sigma_k c_k = 0 ;$$

(2) estimates of the variances of the estimates in 1), so that if the variance of $(\Sigma_k c_k t_k)$ is σ^2, we may obtain an estimate s^2 from the data so that $E_E(s^2) = \sigma^2$;

(3) a mean square between treatments s_b^2 and a mean square for residuals s_r^2 , such that the expectation of s_b^2 is greater than or equal to that of s_r^2 with equality if and only if $t_1 = t_2 \ldots = t$.

We shall examine our designs to see to what extent these criteria are met.

We shall examine our designs to see to what extent these criteria are met in the two analysis of variance situations earlier described.

2.1.4 Randomization Tests

Suppose we have a function of the observations which is a function of the design which we call the criterion C, e.g. treatment sum of squares divided by error sums of squares, say. If the treatments are without effect and all other sources behave as they do in the experiment we would obtain the same set of observations whatever the plan we would have used.

If, however, the presence of treatment effects tends to cause our criterion C to increase then we obtain the probability of having plans with a value of C equal or greater than the observed value C_{obs} under the null hypothesis. This probability is defined to be the significance level under randomization of the null hypothesis that there are no treatment differences as judged by the criterion C.

Tests of significance in the randomized experiment have been presented by way of normal law theory, whereas their validity stems from randomization theory. Fisher says, "The physical act of randomization affords the means, in respect of any particular body of data in examining the wider hypothesis in which no normality of distribution is implied."

2.2 Formal Theory

Let us now look at the work done by various writers on the analysis of variance procedure in the absence of known concomitant variation from the viewpoint of randomization analysis.

We have already described the philosophical ideas of Fisher (1935 and subsequent editions). We shall use the randomized block design as an example to convey the implementation of these ideas. When we do a randomized block experiment, the outcome is a 2-way array of numbers y_{ik}: $i=1,2,\ldots,r$ and $k=1,2,\ldots,t$ in which t is the number of treatments and r is the number of blocks. The standardly used idea of analysis is as follows. We make an analysis of variance:

Source	df	SS	Mean Square
Blocks	$r-1$	x	x
Treatments	$t-1$	T	T^*
Residual	$(r-1)(t-1)$	R	R^*
Total	$(rt-1)$		

We take $\Sigma_k\lambda_k y_{.k}$ $(\Sigma\lambda_k = 0)$ to be the estimate of a treatment contrast. We wish, of course, to have some understanding of the properites of this estimate. If we use a linear model

$$y_{ik} = \mu + b_i + \tau_k + e_{ik}$$

in which the e_{ik} are independent Gaussian random variables with mean 0 and variance σ^2, the estimate is Gaussian with mean $\Sigma\lambda_k\tau_k$ and variance $(\Sigma_k\lambda_k^2)\,\frac{\sigma^2}{r}$. Furthermore if we do the analysis of variance, the expectation of the residual mean square s^2 is σ^2. This leads to the procedures given in essentially all texts. However, Fisher (1926) and Yates (1936) also insists that the estimate of σ^2 be appropriate in a randomization frame. The observation y_{ik} will occur on some unit (ij) of the i-th block, and one may hypothesize that

$$y_{ik} = u_{ij} + \tau_k \quad ,$$

this being a hypothesis that there is no interaction of treatments and experimental units. The requirement is then made that if the expectation of the treatment mean square under the randomization plan, thus involving only $\{u_{ij}\}$, is σ^2, then the variance of the estimate of a treatment comparison shall be $(\Sigma_k \lambda^2_k) \frac{\sigma^2}{r}$, involving the same function σ^2 of the unknown $\{u_{ij}\}$.

This has the consequence that the expectation of the treatment mean square, in the absence of treatment effects, must equal the expectation of the residual mean square, the property of unbiassedness given above. Various randomized designs which at first sight seem to be appropriate for actual use were rejected by Yates because they do not satisfy this property. Both Fisher and Yates, it seems, took the view that if this property of unbiassedness holds, then one may complete the statistical analysis of the experimental data using the simple Gauss-Markov normal distribution procedures.

It was demonstrated e.g. by Eden and Yates (1933) and others, that the significance level that is obtained by this process is "close to" the significance level that is obtained by using the upper tail area of the associated normal distribution theory. Theoretical work on this was done by Pitman (1937), Welch (1937), Ogawa (1974) and Arnold (1964).

This theoretical work was done for the case of the simple designs, randomized block design and Latin square design for the case in which there is not known concomitant variation. In Chapters 3 and 4 of this thesis we will examine the behavior under randomization of the usual Gauss-Markov normal theory statistics in the presence of concomitant variation for these two designs. Pitman (1937) and Welch (1937) and others have examined the standard t and F tests based mainly on univariate situations, when the underlying distribution is non-normal. They have found that these analysis of variance test procedures are remarkably insensitive to deviations or departures from normality. H. J. Arnold (1964) investigated the multivariate generalization of the t test. He considered a statistic which was shown to be a monotonic increasing function of what is commonly known as the Hotelling T^2 statistic. We present here the salient features of Pitman's investigation which give us a general review of the work of the other writers, we will then present a brief review of the work done on randomization theory in linear models by Ogawa (1974) and Arnold (1964).

Pitman (1937) investigated the problem of testing the null hypothesis that the treatments are all equal in the randomized block design for univariate situations. In the usual additive model the observation associated with the j-th treatment in the i-th block is denoted by:

$$y_{ik} = \mu + b_i + \tau_k + e_{ik} \quad , \quad \Sigma b_i = 0 = \Sigma \tau_k, \quad \begin{array}{l} i=1,\ldots,r. \\ k=1,\ldots,t. \end{array}$$

where μ is the average of the whole population, b_i is the average effect assignable to block i, and τ_k the effect corresponding to treatment k. The errors e_{ik} are assumed to be random, independent from normal populations with the same variances. The usual analysis of variance ratio for testing the null hypothesis against the alternative hypothesis that H_1: the effect assignable to treatment k is not zero for all k is

$$F = \frac{r\sum_{k=1}^{t} (y_{.k} - y_{..})^2/(t-1)}{\sum_{i=1} \sum_{k=1} (y_{ik} - y_{i.} - y_{.k} + y_{..})^2/(r-1)(t-1)} = \frac{T/(t-1)}{R/(r-1)(t-1)}$$

Under normality T and R divided by σ^2 , the common variance, are each independently distributed as chi-square variables with (t-1) and (r-1)(t-1) degrees of freedom respectively. Hence the ratio has an F distribution with (t-1) and (r-1)(t-1) degrees of freedom. Instead of the F ratio Pitman (1937) considered the ratio

$$W = \frac{\Sigma(\text{treatments})^2}{\Sigma(\text{treatments})^2 + \Sigma(\text{residuals})^2} = \frac{T}{T + R}$$

It follows that under the above assumptions W has a beta distribution with (t-1) and (r-1)(t-1) degrees of freedom.

Pitman investigated the problem of testing the null hypothesis that the treatments are all equal without making any assumptions about the experimental units y_{ik}. Under the null hypothesis, the treatments are mere labels distributed at random to the various individuals in the blocks. The number of values of W, each value being equally likely for the case of t treatments in r blocks is $(t!)^r$. These $(t!)^r$ values of W contain multipiicities since many of the permutations are equivalent to mere interchanges of treatments. The number of such different interchanges is t!. Thus, the distinct or likely values of W is at most $(t!)^r/t!$ or $(t!)^{r-1}$ An exact test for the null hypothesis would be to compare the actual observed value of W from the experiment with all the other values of W from the $(t!)^{r-1}$ permutations. If not more than some pre-assigned $\alpha\%$ of these $(t!)^{r-1}$ values of W exceed the experimental value one rejects the null hypothesis at the $\alpha\%$ level. Pitman investigated the distribution of W when the underlying distribution deviates from normality. There are two different distributions of W being discussed. One is the distribution of

W which is associated with the underlying distribution of the (infinite) population of which the $\{y\}$ is a sample and hence does not depend on the individual values of the sample. The other is the distribution of W obtained from the permutations and depends directly on the sample. These distributions will be denoted by unconditional and conditional respectively.

Under the assumption that the underlying distribution is normal, the unconditional distribution of W is a beta with (t-1) and (t-1)(r-1) degrees of freedom. If W deviated very little from the beta distribution for any reasonable underlying distribution, then one could "assume normality" and proceed to use the F test without too much concern, since a "α% test" would, in general, reject really true null hypotheses approximately α% of the time.

Pitman (1937) obtained the first four permutation moments of W in terms of the first four moments of the observations. Welch (1937) obtained the first two moments. Let the sum of squares for the basic units effects in the 1-th block be Δ_i , i.e.,

$$\Delta_i = \sum_{j=1}^{t} (u_{ij} - u_{i.})^2 \quad ,$$

where u_{ij} is the basic response of the j-th unit in the i-th block and $u_{i.}$ the mean response in the i-th block. Let $\Delta = \sum_{i=1}^{r} \Delta_i$, and let E_R denote expectation under randomization. Then the first two randomization moments of W are given by

$$E_R(W) = \frac{1}{r}$$

and

$$E_R\{W - E_R(W)\}^2 = \frac{2}{(t-1)} \quad \left(\frac{\Delta^2 - \Sigma\,\Delta_i^{\,2}}{\Delta^2}\right)$$

$$= \frac{2(r-1)}{r(t-1)} \left(1 - \frac{\nu}{b} \right),$$

where $\nu = \dfrac{\Sigma_i(\Delta_i - \Delta)^2}{(r-1)\,\bar{\Delta}^2}$, is the square of the coefficient of variation of the Δ_i.

By looking at the moments of W, Pitman was able to draw certain conclusions about its distribution as the underlying distribution deviates from normality. Of the following moments of W under randomization theory, mean, variance, skewness and kurtosis, only the mean is independent of the particular observations from the experiment.

The mean and variance of a beta distribution with (t-1) and (r-1)(t-1) degrees of freedom are $\frac{1}{r}$ and $\dfrac{2(r-1)}{r^2(rt - r + 2)}$ respectively. The average value of W obtained from randomization is thus the same as the average value for the beta distribution with the proper number of degrees of freedom. When the variance of the beta distribution is equated to the variance of W, one gets the condition

$$\frac{2(r-1)}{r(t-1)} \left(1 - \frac{\nu}{r}\right) = \frac{2(r-1)}{rt - r + 2} .$$

When all the block variances are equal we have $_i =$, $1 \leq i \leq r$ and $\nu = 0$. Then the left hand side of the above equation attains its maximum value which is equal to $\frac{2(r-1)}{r(t-1)}$. The right hand side of the equation can be written as $\frac{2(r-1)}{r(t-1) + 2}$ which is always smaller than $\frac{2(r-1)}{r(t-1)}$ but by an amount which becomes small as r and t increase. The value of the left hand side decreases from $\frac{2(r-1)}{r(t-1)}$ to zero as one block variance becomes much larger than the others. Thus, the variance of W is not greater than $\frac{2(r-1)}{r(t-1)}$ and it takes this value when the block variances are equal. So for large values of r or t when the block variances are approximately equal, the conditional variance of W agrees well with the appropriate beta distribution. If a few of the block variances are very large relative to the others, then the conditional variance of W will be smaller than the variance of the corresponding beta distribution. Pitman, whose ideas were directed to examining by permutation the null hypothesis of no treatment effects, suggested three possibilities when this arises.

(i) Discard the blocks which have large variance relative to the other blocks.

(ii) Fit a beta distribution by the use of the first two moments of W (also investigated by Welch (1937)). This will mean a decrease in the number of degrees of freedom in the beta distribution which applies under normality.

(iii) Make all block variances equal. This of course implies the adjustment of the observations in each block, by scaling the deviations from the block mean, say, to make the necessary changes in the block variances.

We suggest that only the second suggestion is reasonable.

If the block variances σ_1^2 are all equal, or approximately equal, the second moment of W under randomization theory will be too large, but this has an appreciable effect only when r and t are small. The variance of W is fairly insensitive to changes in the values of the block variances when r is large. Pitman recommends that the second moment of W be calculated and compared with the variance of the appropriate beta distribution when t or r is less than 5.

Pitman further remarks that if a beta distribution is fitted by means of the first two moments of W, the third and fourth moments are likely to agree well provided that the second moment is not too small. For equal block variances, fitting a beta distribution in this manner gives a good approximation provided $r(t-1)$ is not too small.

J. Ogawa (1974) examined a class of linear models for finite populations, the true values of which are defined by or derived from the randomization procedure which assigns treatments to units, and the observed values of which differ from the 'true' values by technical errors which are assumed to have GMN properties. The technical error may be due to inaccuracies in measuring the response, say, or due to inaccuracies in the treatment levels actually applied. For example, suppose we have a set of N experimental units which have responses $\{u_i\}$ under some basic conditions and a set of treatments whose effects $\{t_k\}$ we wish to compare. Then under additivity the true response on the i-th unit if treatment k is added to it would be

$$y_{ik} = u_i + t_k \quad .$$

Suppose further that we have a technical error e_{ik}, then our observation will have a value

$$Y_{ik} = y_{ik} + e_{ik} \quad ,$$

where e_{ik} is assumed to have the usual GMN properties. This model is called the Neyman model after Neyman (1935).

Ogawa (1974) considered the Neyman model for the randomized block (RB), the Latin square (LS) and the partially balanced incomplete block (PBIB) designs. Conditional on unit effects being fixed, the ratio of the mean squares for treatment effects and residual error in the analysis of variance has a doubly non-central F-distribution. The distribution of the mean square ratio depends on θ, where θ is the ratio of the treatment sum of squares and the sum of the treatment sum of squares and the residual error sum of squares for the unit effects. So θ is the same as the statistic W which we have just discussed in our review of Pitman's work for the situation in which technical errors are absent. Using Pitman's procedure, Ogawa approximated the randomization distribution of θ by a beta distribution. By integrating the doubly non-central F over the fitted beta distribution of θ, Ogawa showed that the unconditional distribution of the mean square ratio can be approximated by the usual F distribution with the appropriate degrees of freedom.

H. J. Arnold (1964) investigated the multivariate generalization of the t-test. He considered the empirical distribution of the Mahalanobis D^2 statistic, which is a simple monotonic function of Hotelling's "T^2" statistic, for samples from various populations. He found that there was 'mild' disagreement between the nominal significance level of the test based on the assumption that the underlying distribution is normal and the actual randomization significance levels. By equating the parameters of a standard beta distribution to the first two moments of the D^2 statistic, Arnold found that he could obtain 'reasonably' close approximations to the significance levels investigated.

2.3 Conditions for Convergence

We now consider various conditions under which large sample distributions under randomization theory may be approximated by normal law distributions. "Statistical tests based on permutations of the observations" have been proposed and studied by various writers including Fisher, Pitman and Welch whose work we have already discussed. We shall state some of the results of these workers and refer the reader to the citations for proofs. The basic result was obtained by Wald and Wolfowitz (1944) in a paper with the title quoted above. The results of the other workers are variations and refinements of this theme. Of the conditions which follow, the Madow conditions are the strongest and the Noether conditions the weakest. We will prove this statement in a lemma. We shall give equivalent formulations of the Noether conditions and also state some results of Cramer which will be needed for our development in Chapter 3.

2.3.1 Conditions W of Wald and Wolfowitz

Let $H_N = (h_{1N}, h_{2N}, \ldots, h_{NN})$ for $N = 1,2,3, \ldots$ be sequences of real numbers with finite second moments μ_{2N}, and let

$$\mu_r(H_N) = N^{-1} \sum_{i=1}^{N} (h_{iN} - N^{-1} \sum_{j=1}^{N} h_{jN})^r$$

for all integral values of r. For any function $f(N)$ and any positive function $\varphi(N)$ let $f(N) = O(\varphi(N))$ mean that $f(N)/\varphi(N)$ is bounded from above for all N. We say that the sequences H_N $(N = 1,2, \ldots)$ satisfy the conditions W of Wald and Wolfowitz if, for all integral $r > 2$,

$$\mu_r(H_N)/[\mu_2(H_N)]^{r/2} = O(1).$$

We shall omit the second subscript N in the sequences $(h_{1N}, h_{2N}, \ldots, h_{NN})$ where there is no ambiguity. For any value of N let

$$X = (X_1, X_2, \ldots, X_N)$$

be one of the N! permutations of the elements of the sequence $A_N = \{a_i, a_2, \ldots, a_N\}$. Let each permutation of the elements of A_N have the same probability $(N!)^{-1}$. Let $E(y)$ and $\sigma^2(y)$ denote the expectation and variance of a variable y. Then we have the following theorem.

Theorem 2.1 (Wald and Wolfowitz)

Let the sequences $A_N = (a_1, a_2, \ldots, a_N)$ and $D_N = (d_1, d_2, \ldots, d_N)$ $(N = 1,2, \ldots)$ satisfy the condition W. Let the variable L_N be defined as

$$L_N = \sum_{i=1}^{N} d_i x_i$$

Then as $N \to \infty$, the probability of the inequality

$$L_N - E(L_N) < t\,\sigma(L_N)$$

for any real t approaches $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{x^2/2}\,dx$.

Proof: (See Wald and Wolfowitz (1944).)

2.3.2 Madow's Conditions W^*

Madow (1948) defined a sequence $(\varphi_1, \varphi_2, \ldots, \varphi_r, \ldots)$ of finite populations, where for each r, φ_r contains N_r elements u_{ri}, $i = 1, \ldots, N_r$. A simple random sample of size n_r is drawn without replacement from φ_r. The selected elements are denoted by $\{y_{ri}\}$ $i = 1,2, \ldots, n_r$.

The linear function

$$Z_{n_r} = n^{-\frac{1}{2}} \sum_{i=1}^{N_r} (y_{ri} - \bar{U}_{Nr}) \text{ , has the following}$$

distributional properties.

$$\text{Variance} \quad \sigma^2(Z_{n_r}) = \frac{N_r}{N_r - 1} (1 - n_r/N_r)\, \mu_{2N_r}$$

where

$$\mu_{kN_r} = \frac{1}{N_r} \sum_{i=1}^{N_r} (u_{ri} - \bar{U}_{N_r})^k$$

Let $\sigma^2_{y_r}$ denote μ_{2N_r} .

Then the Madow conditions, W^*, are as follows

i) $n_r/N_r < 1 - \epsilon$ where $\epsilon > 0$

and

ii) There exists a finite value λ , such that for all $k \geq 1$

$\lambda_k (N_r) < \lambda$

where

$$\lambda_k (N_r) = \mu_{kN_r} / [\mu_{2N_r}]^{k/2}$$

Theorem 2.2 (Madow (1948).)

If the sequence φ_r satisfies conditions W^*, and a simple random sample of size n_r is selected without replacement from φ_r , then for all t

$$\lim_{\substack{n_r \to \infty \\ N_r \to \infty}} P\{Z^*_{N_r} < t\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2}\, dx$$

where $Z^*_{n_r} = Z_{n_r} / \sigma^2(Z_{n_r})^{\frac{1}{2}}$

Theorem 2.3 (Madow (1948).)

Suppose the elements of φ_r are p-component vectors.

$$u_{ri.} = (u_{ri1}, u_{ri2}, \dots, u_{rip}) \quad i=1,2, \dots, N_r .$$

Assume the conditions W^* are satisfied for each component of this vector.

Let $Z_{n_r j} = n_r^{\frac{1}{2}} \sum_{i=1}^{n_r} (y_{rij} - \overline{Y}_{Nrj}) \qquad j=1,2,\dots,p$

where y_{rij} is the i-th sample element,

and $\qquad \overline{Y}_{N_r j} = \frac{1}{N_r} \sum_{i=1}^{N_r} u_{rij} = \overline{U}_{N_r} \quad .$

Define $Z_{n_r j} = Z_{n_r j} / \sigma^2(Z_{n_r j})$,

where $\qquad \sigma^2(Z_{n_r j}) = \frac{N_r}{N_r - 1} \left[1 - \frac{n_r}{N_r}\right] \mu_{2N_r j}$,

and $\qquad \mu_{kN_r j} = \frac{1}{N_r} \sum_{i=1}^{N_r} (u_{rij} - \overline{U}_{N_r j})^k \qquad k=1,2,\dots$

Suppose that

$$\lim_{r \to \infty} \rho_{rij} = \rho_{ij} \text{ is defined for all i, and j, where}$$

$$\rho_{rij} = N_r^{-1} \sum_{k=1}^{N_r} (u_{rki} - \overline{U}_{N_r})(u_{rkj} - \overline{U}_{N_r})$$

and $\qquad \rho_{ij} > -1+\varepsilon , \qquad \varepsilon > 0.$

Then the limiting distribution of $(Z_{n_r 1}, Z_{n_{r2}}, \dots, Z_{n_r p})$

is multivariate normal, with mean zero and covariance matrix Σ where

$$\Sigma = \begin{pmatrix} 1, & \rho_{12}, & \dots, & \rho_{1p} \\ \rho_{21}, & 1, & \dots, & \rho_{2p} \\ \rho_{p1}, & , & \dots, & 1 \end{pmatrix}$$

2.3.3 Noether Conditions

Noether (1949) introduced a more general set of conditions which we designate as conditions ν.

For each $N = 1, 2, \ldots,$ let $H_N = (h_{1N}, \ldots, h_{NN})$ be an N-tuple of real numbers. Then the sequence H_N satisfies the conditions ν of Noether if for all $r = 3, 4, \ldots,$ we have

$$\sum_{i=1}^{N} (h_{iN} - \bar{h}_N)^r / (\sum_{i=1} (h_{1N} - \bar{h}_N)^2)^{r/2} = o(1),$$

where

$$\bar{h}_N = N^{-1} \sum_{i=1}^{N} h_{iN} \quad .$$

Let $X = (X_1, \ldots, X_N)$ be a random variable which takes each permutation of $A_N = (a_{1N}, \ldots, a_{NN})$ with the same probability $\frac{1}{N!}$. Let the linear expression $L_N = c_{1N}X_1 + c_{2N}X_2 + \ldots \div c_{NN}X_N = \sum_{i=1} c_{iN}X_i$.

Then it is easily proved that

$$E(L_N) = \frac{(\sum_{i=1}^{N} c_{iN})(\sum_{i=1} a_{iN})}{N}$$

and

$$\mathrm{Var}(L_N) = \frac{1}{(N-1)} \sum_{i=1}^{N} (c_{iN} - \bar{c}_N)^2 \sum_{j=1}^{N} (a_{jN} - \bar{a}_N)^2$$

The next theorem is an extension of the Wald-Wolfowitz theorem and was proved by Noether.

<u>Theorem 2.4</u> (Wald and Wolfowitz-Noether)

If the sequence C_N satisfies conditions W of Wald and Wolfowitz and A_N satisfies condition ν of Noether and

Let $$L_N = c_{1N}X_1 + \dots + c_{NN}X_N$$

then $$L_N^{\circ} = (L_N - E(L_N)) / \sigma L_N$$

has a limiting normal distribution with mean 0 and variance 1.

Proof: (See Fraser (1957).)

The proof consists mainly of showing that the moments of L_N° converge to those of $N(0,1)$. The moments of L_N can clearly be expressed in terms of linear combinations of the U statistics. From these we choose the terms which are dominant with respect to the order of magnitude of the product of the U statistics and their coefficients.

Theorem 2.5 (Hoeffding)

The Noether conditions for a sequence H_N is equivalent to either of the following two conditions.

1) $$\lim_{N \to \infty} \frac{\sum_{i=1}^{N} \left| h_{iN} - \bar{h}_i \right|^r}{\left[\sum_{i=1}^{N} (h_{iN} - \bar{h}_N)^2 \right]^{r/2}} = 0 \text{ for some } r > 2. \qquad 2.3.3.1$$

2) $$\max_{N \to \infty}{}_i \frac{(h_{iN} - \bar{h}_N)^2}{\sum_{i=1}^{N} (h_{iN} - \bar{h}_i)^2} \longrightarrow 0 \qquad 2.3.3.2$$

Proof: (See Fraser (1957).)

<u>Theorem 2.6</u> (Fraser)

If A_N satisfies Noether's condition ν, if C_N and D_N satisfy condition W of Wald and Wolfowitz and if the correlation between C_N and D_N

$$\rho_N = \frac{\sum_{i=1}^{N} (c_{iN} - \bar{c}_N)(d_{iN} - \bar{d}_N)}{\left[\sum_{i=1}^{N} (c_{iN} - \bar{c}_N)^2 \sum (d_{iN} - \bar{d}_N)^2\right]^{\frac{1}{2}}}$$

has limit ρ, $\rho \neq 1$, then the limiting distribution of

$$L_N^{\,o} = \frac{L_N - E(L_N)}{\sigma(L_N)} \quad , \qquad L'_N{}^{o} = \frac{L'_N - E(L'_N)}{\sigma(L'_N)} \qquad ,$$

where $L_N = \sum_{i=1}^{N} c_{iN} X_i$, $L'_N = \sum_{i=1}^{N} d_{iN} X_i$, and $(X_1, \ldots X_N)$ is a random permutation of $(a_{1N}, \ldots a_{NN})$ with probability $(N!)^{-1}$, is bivariate normal with means 0, variance 1, and correlation ρ.

<u>Proof</u>: (See Fraser (1957).)

Hájek (1960) has shown that if two double sequences of real numbers satisfy the Noether conditions and a generalized Lindeberg condition, then the limiting distribution under permutations of the inner product is asymptotically normal. We present these conditions in a theorem, then we shall extend the theorem to the multivariate situation where the elements of the sequences are vectors.

Theorem 2.7 (Hájek)

Let $B_r \{b_{ri}, 1 \leq i \leq N_r, r \geq 1\}$ and $A_r = \{a_{ri}, 1 \leq i \leq N_r, r \geq 1\}$ be double sequences of real numbers. Let $X_r = \{X_{r1}, \ldots, X_{rN_r}\}$ be a random vector which takes on the N! permutations of A_r with equal probabilities.

Put $L_r = \sum_{i=1}^{N_r} b_{ri} X_{ri}$

Suppose that the sequences A_r and B_r satisfy the Noether conditions:

$$\lim_{r \to \infty} \frac{\max_{1 \leq i \leq N_r} (a_{ri} - \bar{a}_r)^2}{\sum_{i=1}^{N_r} (a_{ri} - \bar{a}_r)^2} = 0 \qquad 2.3.3.3$$

and

$$\lim_{r \to \infty} \frac{\max_{1 \leq i \leq N_r} (b_{ri} - \bar{b}_r)^2}{\sum_{i=1}^{N_r} (b_{ri} - \bar{b}_r)^2} = 0 \qquad 2.3.3.4$$

Then the statistic $L^* = \dfrac{L_r - EL_r}{\sqrt{\mathrm{var}(L_r)}}$ has an asymptotically normal distribution with mean value 0 and variance 1, if and only if, for any $t > 0$

$$\lim_{r \to \infty} \frac{1}{N_r} \sum_{z_{rij} > t} \sum z_{rij}^2 = 0 \qquad 2.3.3.5$$

where

$$z_{rij} = \frac{(b_{ri} - \bar{b}_r)(a_{rj} - \bar{a}_r)}{[\frac{1}{N_r} \sum_{i=1}^{N_r} (b_{ri} - \bar{b}_r)^2 \sum_{i=1}^{N_r} (a_{ri} - \bar{a}_r)^2]^{\frac{1}{2}}}$$

$(i \leq 1,\ j \leq N,\ r \leq 1)$.

Proof: (See Hājek (1960).)

Suppose the elements of the double sequence $A_r = \{a_{ri},\ 1 \leq i \leq N_r,\ r \geq 1\}$ are p component vectors $a_{ri} = (a_{ri1}, \ldots, a_{rip})'$, and the elements of the double sequence $B_r = \{b_{ri},\ \leq i \leq N,\ r \geq 1\}$ are t component vectors, $b_{ri} = (b_{ri1}, \ldots, b_{rit})'$. Let $X_r = \{X_{ri} \ldots, X_{r_{N_r}}\}$ be a random variable which takes on the $N_r!$ permutations of A_r with equal probabilities. Define the normalized vector sequences, A_r^*, with elements

$$a_{ri}^* = \left(\frac{a_{ri1} - \bar{A}_{N_r 1}}{\sigma_{ar1}}, \frac{a_{ri2} - \bar{A}_{N_r 2}}{\sigma_{ar2}}, \ldots, \frac{a_{rip} - \bar{A}_{N_r p}}{\sigma_{arp}} \right)',$$

and B_r^* with elements

$$b_{ri}^* = \left(\frac{b_{ri1} - \bar{B}_{N_r 1}}{\sigma_{br1}}, \frac{b_{ri2} - \bar{B}_{N_r 2}}{\sigma_{br2}}, \ldots, \frac{b_{rit} - \bar{B}_{N_r t}}{\sigma_{brt}} \right)',$$

where $\bar{A}_{N_r j} = \frac{1}{N_r} \sum_{i=1} a_{rij}$, $\bar{B}_{N_r k} = \frac{1}{N_r} \sum_{i=1} b_{rik}$,

$$\sigma_{arj}^2 = \frac{1}{N_r} \sum_{i=1}^{N_r} (a_{rij} - \bar{A}_{N_r j})^2, \text{ and } \sigma_{brk}^2 = \frac{1}{N_r} \sum_{i=1}^{N_r} (b_{rik} - \bar{B}_{N_r k})^2$$

Put $L_{rkj} = \Sigma_i b_{rik} \cdot X_{rij}$,

and $L_{rkj}^* = \dfrac{L_{rkj} - E\,L_{rkj}}{\sqrt{\operatorname{var}(L_{rkj})}}$

Let $\Sigma_{Br} = \frac{1}{N_r} B_r^* B_r^{*\prime}$ and $\Sigma_{Ar} = \frac{1}{N_r} A_r^* A_r^{*\prime}$. Also let the Euclidian norm, $\|a_{ri}\|$, of a_{ri} be given by $\|a_{ri}\|^2 = \Sigma_{j=1}^{p} a_{rij}^{\ 2}$.

Theorem 2.8

Suppose the vector sequences A_r^* and B_r^* satisfy the following conditions:

$$\lim_{N_r \to \infty} \max_{1 \le i \le N_r} \frac{1}{N_r} \|a_{ri}^*\|^2 = 0 \qquad 2.3.3.6$$

$$\lim_{N_r \to \infty} \max_{1 \le i \le N_r} \frac{1}{N_r} \|b_{ri}^*\|^2 = 0 \qquad 2.3.3.7$$

$$\lim_{N_r \to \infty} \Sigma_{Br} = \Sigma_B \ , \quad \lim_{N_r \to \infty} \Sigma_A = \Sigma_A \qquad 2.3.3.8$$

We assume, without loss of generality, that the Σ's are all positive definite. For any $t > 0$, let $\emptyset_{rt}$ be the set of integers for which $\frac{1}{N_r} \|a_{ri}\| \cdot \|b_{rj}\| > t$

Then we have the following:

1) The limiting joint distribution of the set of t vectors L_{rkj}^* , k=1,2, ..., t is multivariate normal with mean 0 and covariance matrix Σ_B ;

2) The limiting joint distribution of the set of p vectors L_{rkj}^* , j=1, ..., p is multivariate normal with mean 0 and covariance matrix = Σ_A ;

3) The txp matrix $L_r^* = (L_{rkj}^*)$ has limiting joint normal distribution with mean 0 and covariance matrix given by the Kronecker matrix

product $\Sigma_B \circ \Sigma_A$; if and only if for any $t > 0$

$$\lim_{N_r \to \infty} \frac{1}{N_r} \sum_{\phi_{rt}} \|a^*_{ri}\| \, \|b^*_{rj}\| = 0 \qquad 2.3.3.9$$

Proof: It is sufficient to show that any linear combination of the L_{rkj} satisfies the conditions of theorem 2.4 of Hâjek and hence has limiting normal distribution. Since the L^*_{rkj} are invariant for origin and scale changed of A_{rN_r} and of B_{rN_r}, we may assume without loss of generality

$$\Sigma_{i=1}^{N_r} a_{rij} = 0 = \Sigma_{i=1}^{N_r} b_{rik} \qquad j=1, \ldots, p; \quad k=1, \ldots, t \qquad 2.3.3.10$$

and $$\Sigma_{i=1}^{N_r} a^2_{rij} = N_r = \Sigma_{i=1} b^2_{rik} \qquad 2.3.3.11$$

So variance $$L_{rkj} = \frac{1}{N_r} (\Sigma_{i=1}^{N} a^2_{rij})(\Sigma_{i=1}^{N_r} b^2_{rik}) = N \qquad 2.3.3.12$$

For any non-null p-vector $C = (c_1, \ldots, c_p)$, consider the linear combination $W_{rk} = \Sigma_j \; c_j L_{rkj} = L_{rk} C$. Let $U_r = A_r' C$. Then var $(U_r) = C'\Sigma_{Ar} C$. Since Σ_{Ar} converges to the positive definite matrix Σ_A, we can choose N sufficiently large such that if $\epsilon = \frac{1}{2}C'\Sigma_A C$, then $C'\Sigma_{Ar} C > \epsilon$ whenever $W_r > N$.

$\Sigma_{i=1}^p u^2_{ri} = NC_r'\Sigma_{A_r} C$, and $u^2_{ri} = (\Sigma_{j=1}^p c_j a_{rij})^2 \le (\Sigma_{j=1}^p c^2_j)(\Sigma_{j=1}^p a^2_{rij})$

Also $$\Sigma_i^{N_r} b^2_{rik} = N_r = \Sigma_{i=1}^{N_r} a^2_{rij}$$

and $$\Sigma_{k=1}^t b^2_{rik} \ge b^2_{rik'} \qquad 1 \le k' \le t$$

so $$\frac{u^2_{ri}}{\Sigma u^2_{ri}} \le \frac{(\Sigma_{j=1}^p c^2_j)}{\epsilon} \cdot \frac{1}{N_r} \cdot \|a_{ri}\|^2$$

So if the vector sequences A_r^* and B_r^* satisfy conditions 2.3.3.6 - 2.3.3.8 then each component sequences of B_r^*, B_{rk}, k=1,2, ..., t as well as the sequence U_r satisfies the conditions 2.3.3.3 - 2.3.3.5 and so the first and second assertions of theorem 2.8 follows.

The third result follows by considering any linear combination of the t vector $U'_r B_r$ $\{= W_r = (W_{r1}, \ldots, W_{rt})\}$. That is, consider the linear combination $V_r = \Sigma_{k=1}^{t} d_k W_{rk} = W_r'D$ where $D = (d_1, d_2, \ldots, d_t)$ is a non-null t-vector. Then using the same arguments as above we see that the conditions 2.3.3.6 - 2.3.3.8 are sufficient to ensure the limiting normal distribution of V_r. But $V_r = W_r'D = C'A_r B_r'D$

$$= C'LD$$

$$= \Sigma_{kj}\, g_{kj} L_{rkj} \quad ,$$

where $g_{kj} = d_k c_j$. Since d_k and c_j are arbitrary, then we assume that $G = \{g_{11}, \ldots, g_{tp}\}$ is an arbitrary non-null vector and so the result follows.

For the situation where the limiting covariance matrices are semi-definite, we can always make a linear transformation on the sequences to give us a subset of uncorrelated variables which have limiting joint normal distribution and a subset of variables which have constant values with probability one.

Lemma

We shall now show that the conditions W^* of Madow imply that conditions W of Wald and Wolfowitz and that the conditions W of Wald and

Wolfowitz imply the conditions ν of Noether and the conditions ν^* of Hájek. For ease of writing we will drop the subscript r where there is no ambiguity.

If Part ii of Madow's conditions W^* are satisfied, then we have

$$\mu_{kN} \leq \mu_{2N}^{k/2} \lambda \qquad \text{for all } k > 2 \text{ and some constant } \lambda > 0.$$

So

$$\mu_{kN} \leq \mu_{2N}^{k/2} \lambda \qquad \text{for all integral values of } k > 2, \text{ i.e., } k = 3,4,5, \ldots .$$

And

$$\frac{N\mu_{kN}}{(N\mu_{2N})^{k/2}} \leq \frac{N\lambda}{N^{k/2}} = o(1) \text{ for } k = 3,4,5, \ldots .$$

So we have the following result:

conditions W^* imply conditions W imply conditions ν.

Now if the sequences A_r and B_r each satisfy the Madow conditions W^*, by theorem 2.5, they satisfy the conditions given in equations 2.3.3.3 – 2.3.3.4. Let Z_{rij} be defined as in 2.7.3 and for any $t > 0$, let ϕ_{rt} be the set of integer pairs (i,j) for which $Z_{rij} > t$. We may without loss of generality assume that the sequences A_r and B_r have been normalized so that $\Sigma a_{ri} = 0 = \Sigma b_{rj}$, $\Sigma a^2_{ri} = N = \Sigma b^2_{rj}$, and $Z_{rij} = N^{-\frac{1}{2}} \cdot a_{ri}\, b_{rj}$.

So for any $k > 2$ we have

$$\sum_{\phi_{rt}} Z^2_{rij} \leq \Sigma_{\phi_{rt}} |Z_{rij}|^k \; / \; |Z_{rij}|^{k-2} \leq t^{-k+2}\, \Sigma_{\phi_{rt}} |Z_{rij}|^k$$

$$\leq t^{2-k}\, N^{-k/2}\; \Sigma_{\phi_{rt}} |a_{ri}|^k \; |b_{rj}|^k$$

$$\leq t^{2-k} N^{-k/2} \Sigma|a_{ri}|^k \Sigma|b_{ri}|^k .$$

If each of the sequences A_r and B_r satisfies the conditions W^* of Madow, then choosing k to be an even integer we have that $\Sigma |a_{ri}|^k$, $\Sigma |b_{rj}|^k$ are each of order of magnitude $O(N_r)$. So we obtain the required result that:

$$N_r^{-1} \Sigma_{\phi_{rt}} Z^2_{rij} \leq t^{2-k} N^{-k/2} \lambda^* N^2 \quad (N^{2-k/2})$$

$$= O(1) \qquad \text{for } k > 2.$$

2.3.4 Cramer's Conditions

We shall also describe briefly, in Chapter 3, the analysis of covariance in the randomized block design which was examined by J. Robinson (1974). This involves limit theory and the conditions that are required for our discussion are due to Cramer (1970) and are given in the following theorem.

Theorem 2.9 (Cramer)

Let $X_1, X_2, \ldots, X_i = [X_{i1}, X_{i2}, \ldots, X_{ik}]$ be a sequence of random variables in R^k such that every X_N has the probability function $P_N(S)$ with vanishing first order moments and finite second order moments μ_{rsN}.

Suppose that as $N \to \infty$ the following two conditions are satisfied

(1)
$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \mu^{(i)}_{rsN} = \mu_{rs} \qquad (r,s, = 1,2, \ldots, k)$$

where μ_{rs} are not all zero.

(2)
$$\frac{1}{N} \sum_{i=1}^{N} \int_{\|X\| > e\sqrt{N}} \|X\|^2 \, dP_i \to 0$$

for every $e > 0$, where $\|X\|$ denotes $\sqrt{(X_{i1}^2 + \ldots + X_{ik}^2)}$.

Then the probability function of the variable $(X_1 + \ldots + X_N)/\sqrt{N}$ converges to the normal probability function which has first moment zero and the second order moments μ_{rs}.

Proof: (See Cramer (1970).)

In Chapter 3 we shall use Hájek's condition ν^*, which consists of two Noether conditions and a generalized Lindeberg condition, to formulate some limit distributions under randomization for the completely randomized design. We shall also use the conditions of Cramer's theorem in our review of the randomized block design.

3. THE ANALYSIS OF COVARIANCE

This chapter is presented in 8 parts: 1) as a background for the randomization theory, we shall present some of the salient aspects of GMN linear regression in the univariate and the multivariate cases; aspects of the analysis of covariance with the Gauss-Markov error assumption; we then relate the procedure to classificatory models with concomitant variables; 2) here we will try to present a formal development of randomization theory for the CRD. We hope to show that there are serious logico-philosophical differences between the GMN theory and randomization theory with respect to the comparative experiment. We shall derive the expectations under randomization of the first two moments of the various "residual" components of the analysis of covariance table. We shall also discuss the work of D. R. Cox (1956) on weighted randomization; 3) we next use ideas of order in probability (see Appendix A) to approximate some statistics involving the "residual" components by Taylor-McLaurin series to order $O(N^{-2})$; 4) we then extend our results to the situation where we have more than one covariate, i.e., the analysis of multiple covariance; 5) we give various theorems on the limiting distributions under randomization theory of various GMN statistics; 6) we discuss the work of J. Robinson on the RBD; 7) we give a brief discussion of other work related to the analysis of covariance; 8) finally, we will give a brief summary of the chapter.

3.1 GMN Linear Regression

3.1.1 Maximum Likelihood Estimators

Suppose we are considering a model which can be written in the form:

Model 1) $$y = X\beta + e \quad ,$$

where

y is an nxl vector of observation,

X is a known nxp matrix assumed to be of full rank,

β is a pxl vector of parameters,

and

e is a nxl vector of errors with expectation E(e), equal to zero and variance, V(e), equal to $\sigma^2 I$.

The maximum likelihood estimate (MLE) $\hat{\beta}$ of β is given by $\hat{\beta}_1 = (X'X)^{-1}X'Y$ and the MLE $\hat{\sigma}_1^{\,2}$ of σ^2 is $n^{-1}(y'y - \hat{\beta}'X'X\hat{\beta})$. Under model 1 we assume X is fixed and e has GMN distribution. The GMN distribution theory results for the estimates $\hat{\beta}_1$ and $\hat{\sigma}_1^{\,2}$ are given as follows:

i) $\hat{\beta}_1 \sim N(\beta, \sigma^2(X'X)^{-1})$,

ii) $n\hat{\sigma}_1^{\,2}/\sigma^2 \sim X^2_{n-p}$ the Chi-square distribution with n-p degrees of freedom,

iii) $\mathrm{Var}\,(\hat{\beta}\,X) = \sigma^2(1-R^2)(X'X)^{-1}$, where R^2 is the coefficient of determination, i.e., the square of the multiple correlation coefficient of y with the columns of X

and

iv) $\hat{\beta}_1$ and $\hat{\sigma}_1^2$ are independent.

3.1.2 The Analysis of Covariance from the Linear Regression Viewpoint

To get some understanding of the nature of the analysis of covariance in the standard assumed linear model situation with GMN error structure, let us now examine the changes in the estimates of various statistics and the implication of these changes when we fit the following two models :

Model 1) $y = X\beta + e$,

Model 2) $y = X\beta + Z\gamma + e$,

where y is an N-vector of observations from some population of interest, X is an Nxp matrix of values which are assumed to be fixed and known, β is a p-vector of parameters which are unknown and about which we wish to obtain estimate of various contrasts, Z, the covariate, is an Nxq matrix of values concomitant to y. γ is a q vector of parameters which are considered to be nuisance parameters in the experimental context and e is an N vector of errors which are assumed to have Gauss-Markov properties, i.e., $E(e) = 0$, $E(ee') = \sigma^2 I$

The least squares procedure applied to the first model will give us the normal equations.

$$X'X\hat{\beta} = X'Y$$

Let B be such that

$$X'XB = X'$$

Then $$B'X'X = X .$$

So we have

$$XB = B'X'XB = B'X' = P_X$$

$$P_X = P_X^2$$

So $XB = B'X'$ is symmetric and idempotent and so is a projection operator onto $\mathcal{C}(X)$, the column space of X. Hence

$$X\hat{\beta} = B'X'X\hat{\beta} = B'X'y = P_X y.$$

we may without loss of generality assume that X is of full rank and that

$Z'(I - P_X)Z$ is of full rank.

A linear combination of the parameter $\lambda'\beta$ is said to be estimable from model (1) if there exists a vector a, $a' = (a_1\ a_2\ \dots\ a_n) \neq 0$ such that $E(a'y) = \lambda'\beta$.

Also $\lambda'\beta$ is estimable from model (1) if and only if there exists ρ such that

$$X'X\rho = \lambda .$$

The estimate is then $\lambda'\hat{\beta} = \rho'X'X\hat{\beta} = \rho'X'y$ with variance $(\lambda'\hat{\beta}) = \rho'X'X\rho\sigma^2$.

The sum of squares due to the model fit in (1) is $\hat{\beta}X'y = y'P_Xy$.

Also,

$$\text{covariance } (P_Xy, Z\hat{\gamma}) = \sigma^2\ P_X(I - P_X)A' = 0$$

and, consequently,

$$\begin{aligned} \text{Var } (X\tilde{\beta}) &= \text{Var } (P_Xy - P_XZ\hat{\gamma}) \\ &= \text{Var } (P_Xy) + \text{Var } (P_XZ\hat{\gamma}) \\ &= \sigma^2[P_X + P_XZ(Z'(I - P_X)Z)^{-1}Z'P_X] \end{aligned}$$

If $\lambda'\beta$ is estimable from model (1) then we have $\lambda' = \rho'X'X$ and $\lambda'\beta = \rho'X'P_Xy = \rho'X'y$. The estimate of $\lambda'\beta$ in model (2) is then

$$\begin{aligned} \lambda'\tilde{\beta} &= \rho'X'P_X(y - Z\hat{\gamma}) \\ &= \rho'X'P_Xy - \rho'X'P_XZ\hat{\gamma} \end{aligned}$$

with variance equal to

$$\begin{aligned} &\sigma^2[\rho'X'P_XP_XX\rho + \rho'X'P_XZ(Z'(I - P_X)Z)^{-1}ZP_XX\rho] \\ &\quad = \sigma^2[\rho'\lambda + \rho'X'Z\ (Z'(I - P_X)Z)^{-1}Z'X\rho] \end{aligned}$$

We may get some understanding of the second term in the brackets of the right hand side of the last equation by considering the least squares approximation of Z by the column space of X, $\mathcal{C}(X)$.

Consider the model

$$Z = XG + d$$

The normal equations in this situation are

$$X'X\tilde{G} = X'Z \ .$$

The projection of Z onto the column space of X is given by

$$X\tilde{G} = X(X'X)^{-1}X'Z = P_X Z.$$

So we have

$$\rho'X'P_X Z = \rho'X'X\tilde{G} = \lambda'\tilde{G} \quad .$$

and hence

$$\mathrm{Var}\ (\lambda\tilde{\beta}) = \sigma^2[\rho'\lambda + (\lambda'\tilde{G})'[Z'(I - P_X)Z]\tilde{G}'\lambda]$$

What then is the significance of the analysis of covariance in this situation?

If we estimate our contrasts $\{\lambda'\beta\}$ by using model (1) when model (2) is correct, we will have $\lambda'\hat{\beta} = \rho'X'P_X y = \rho'X'P_X(X\beta + Z\gamma + e)$

and $\quad E(\lambda'\hat{\beta}) \quad = \rho'X'X\beta + \rho'X'Z\gamma$

$$= \lambda'\beta + \lambda'\tilde{G}\gamma$$

So the estimate $\lambda'\hat{\beta}$ will have a bias $\lambda'\tilde{G}\gamma$ unless $P_X Z = 0$, that is, unless the column space of Z is disjoint from the column space of X, or

$$\mathcal{C}\ (Z) \wedge \mathcal{C}\ (X) = \{\emptyset\} \text{ , the null-space.}$$

If $\mathcal{C}(Z) \wedge \mathcal{C}(X) \neq \{\emptyset\}$ and if the effects of the nuisance parameters are not negligible, i.e. $Z\gamma \neq \emptyset$, we may reduce σ^2 and gain in precision by using model (2).

If our objective is to obtain information on the set of parameters

β, we may consider the parameters γ, as nuisance parameters included in the algebraic specification of the model to make it more realistic. By this we hope to increase the precision in our estimates. We have seen that the variance of our estimates from model (2) of contrasts $\{\lambda'\beta\}$, estimable from model (1) depend on the regression of X on Z and so does the estimate of σ^2 . Also we have determined the algebraic adjustment needed to estimate $\lambda'\beta$ from model (2). The meaning of the adjustment in this situation seems quite obvious. We surmise that it removes a component of experimental error which may be identified with environmental variation that would inflate the experimental error if it were ignored. We note here that if X is correlated with Z, in the sense that the covariate measurements, A, are influenced by the factors X, the adjustment for Z may distort the nature of the effects of the X factors in which we are interested. If the effects of the X factors disappear after adjusting for Z, this may suggest that the estimated contrasts for model (1) reflect the influence produced by the factors of X on Z; that is, the concomitant Z may be the agent through which the factors of X produce their influence on the response variable y. This situation is not generally considered as a logical case for the analysis of covariance procedure and so we will not pursue this situation any further.

In the completely randomized design, the estimate for treatment differences between treatments k and k' is given by $\widehat{(\tau_k - \tau_{k'})} = y_{k.} - y_{k'.} - b(x_{k.} - x_{k'.})$ where $y_{k.}, x_{k.}$ are the mean values of response variable y and the concomitant x for the experimental units which receive treatment k and $b = R_{xy}/R_{xx}$ where R_{xy} and R_{xx} are the residual sum of products and sum of squares for xy and x respectively. We shall now pursue this situation in more detail.

3.1.3 Classificatory Models: The Univariate Analysis of Covariance in the Comparative Experiment with Gauss-Markoff Assumptions

We shall be concerned with a simple case of this method, where we have one covariate. Suppose we have $N = rt$ experimental units standardized to zero mean and we wish to compare t treatments. We shall assign the treatments to disjoint subsets of r units, and, following the basic idea of randomization, we shall use the completely randomized design. We suppose that we observed a concomitant variable, x, for each of the units before the experiment was performed. The data set available for analysis and interpretation is then

$$\{(y_{kj}, x_{kj}), \quad k=1,2,\ldots,t; \qquad j=1,2,\ldots,r\}$$

The "standard" mode of analysis is well-known, but it needs to be given explicity as the basis for part of what follows. A Gauss-Markov normal linear additive model

Model 3 $$y_{kj} = \mu + \tau_k + \beta x_{kj} + e_{kj}$$

is applied. Routine least squares gives the following results:

a) The estimate of a treatment comparison

$\Sigma_k \lambda_k \tau_k$, $\Sigma_k \lambda_k = 0$ is given by

$\Sigma_k \lambda_k (y_{k.} - \hat{\beta} x_{k.})$

where $\hat{\beta}$ is the estimate of the "regression" on the covariate.

b) The analysis of variance for a univariate observation becomes a multivariate analysis of variance, that is usually given in the following form:

		Sum of Squares		
Source	df	yy	xy	xx
Treatments	t-1	T_{yy}	T_{xy}	T_{xx}
Residual	N-t	R_{yy}	R_{xy}	R_{xx}

c) The estimate of β is taken to be

$$\hat{\beta} = \frac{R_{xy}}{R_{xx}}$$

d) The variance of $\hat{\beta}$ is taken to be

$$\frac{\sigma^2}{R_{xx}}$$

e) The estimate of σ^2 is taken to be

$$\hat{\sigma}^2 = (R_{yy} - R_{xy}{}^2/R_{xx})/(N-t-1) .$$

f) The test of significance of absence of treatment effects is given by the criterion:

$$\frac{\left\{(T_{yy} + R_{yy}) - \dfrac{(T_{xy} + R_{xy})^2}{(T_{xx} + R_{xx})} - R_{yy} + \dfrac{R_{xy}{}^2}{R_{xx}})\right\} /(t-1)}{(R_{yy} - R_{xy}{}^2/R_{xx})/(N-t-1)}$$

which is taken to follow $F_{t-1,N-t-1}$ under the null hypothesis.

g) The variance of $\Sigma\lambda_j \hat{\tau}_j$, $(\Sigma\lambda_j = 0)$, is taken to be

$\sigma^2\{\Sigma\lambda_j{}^2/r + (\Sigma\lambda_j x_{j.})^2/R_{xx}\}$

h) The covariance of $\Sigma\lambda_j\hat{\tau}_j$, and $\Sigma\nu_j\hat{\tau}_j$, $\Sigma\lambda_j=0$, $\Sigma\nu_j=0$ is taken to be $\sigma^2 \{ \Sigma\lambda_j\nu_j/r + (\Sigma\lambda_j x_{j.})(\Sigma\nu_j x_{j.})/R_{xx} \}$.

These facts with normality give all the usual statistical tests and statistical interval procedures.

The estimate of error in the absence of the covariate is $R_{yy}/(N-t)$. So when the covariate is included the estimate of the variance is modified by a factor f where

$$f = \{1 - \frac{R_{xy}^2}{R_{yy} \cdot R_{xx}} \} \cdot \frac{N-t}{N-t-1}$$

$$= (1-r^2)\frac{(N-t)}{(N-t-1)} \quad \text{, say.}$$

If r is small the reduction in estimated variance is negligible. But as r approaches unity we see that precision would be greatly increased.

It is not clear what interpretation should be given to the adjustments in our estimates that are due to the inclusion of the covariate. Consider any contrast $(\widehat{\tau_i - \tau_j})$, say, estimated by

$$y_{i.} - y_{j.} - b(x_{i.} - x_{j.}) \quad .$$

We note that $(x_{i.} - x_{j.})$ measures the difference between the x values for treatment i and treatment j. So if a statistically significant value for a contrast disappears after adjusting for x, we would conclude that the difference indicated by the unadjusted estimate, $y_{i.} - y_{j.}$ was not due to the influence of the treatment factors at all.

We shall now leave the Gauss-Markov situation and try to develop some formal understanding of the logic of the procedure from the randomization model viewpoint. We first examine the incidence matrix or the assignment of treatments to units, and try to make somewhat clear their behavior under the various permutations. We will then examine the behavior as properties of the least squares estimates of statistical parameters under the null hypothesis, that is in the absence of differential treatment effects. To examine the behavior of these estimates for sequences of $N(=rt)$ experimental units, $r=1,2,\ldots$ and t fixed, we shall obtain Taylor-McLaurin expansions of them to order $O(N^{-2})$.

3.2 Randomization Theory

3.2.1 A Randomization Permutation Representation

Suppose we have $N = rt$ experimental units u_i, each with p corresponding concomitant values $X_i = (X_{i1}, X_{i2}, \ldots, X_{ip})$, i.e., we have N $(p+1)$-tuples $(u_i, X_{i1}, X_{i2}, \ldots, X_{ip})$, $i=1,2,\ldots$, $N=rt$.

The usual randomization procedure may be described in two stages (π_1, π_2). In the first stage π_1 we partition the set of N $(p+1)$-tuples (U_i, X_i) into t groups each of size r so that each unit $(U_., X_.)$ has equal probability of falling into any group. There are $N! \,/\, (r!)^t$ ways of doing this. In the second stage π_2, treatments are assigned to groups so that each treatment has equal probability of being assigned to any group. There are $t!$ ways of doing this.

The assignment of units to treatments may be described, alternatively, by a set of sampling random variables $\delta^i_{k(j)}$ $\{0,1\}$, where $\delta^i_{k(j)} = 1$ if unit i is the j-th replicate unit which receives treatment k, $j=1,2,\ldots,r$; and $\delta^i_{k(j)} = 0$ otherwise. The properties of the $\{\delta^i_{k(j)}\}$ are discussed in Appendix C.

If z_{ik} is the response we would have on unit i if treatment k were applied to it, we would have a conceptual population of Nt responses given by z_{ik}, $i=1, 2, \ldots, N$; $k+1,2, \ldots, t$.

Without loss of generality, we may assume that each of the columns of $X = (X_{ij})$, has been normalized so that $X'_{.j} X_{.j} = 1$ and $X'_{.j} \mathcal{J}_N = 0$, where $\mathcal{J}_N$ is the column vector of N ones and $X_{.j}$ denotes the j-th column of X. This is easily done by replacing each element X_{ij} by

$$x_{ij} = \frac{X_{ij} - \bar{X}_{.j}}{\sqrt{\Sigma_i (X_{ij} - \bar{X}_{.j})^2}}$$

where $\bar{X}_{.j} = \frac{1}{N} \Sigma_i X_{ij}$

A potentially useful way of presenting the nature of the completely randomized design is as follows. We have N(=rt) units that are indexed by the numbers 1 to N. There are N! possible permutations and we index these by i. We index the outcome of the i-th permutation by

$$i_{1(1)}, i_{1(2)}, \ldots, i_{1(r)}, i_{2(1)}, \ldots, i_{2(r)}, \ldots, i_{t(r)} \quad .$$

We then assign the k-th treatment to the r units with indices $i_{k(1)}, \ldots, i_{k(r)}$. The ordered set of N units may be written in the form

$$U^* = \{U_{i_{1(1)}}, U_{i_{1(2)}}, \ldots, U_{i_{t(r)}}\}$$

which we may write as

$$U^* = U\pi \quad ,$$

where

$\pi = (\pi_{\ell j})$ is the NxN (elementary) permutation matrix with elements

$\pi_{\ell j} = 0$ or 1. The treatment incidence matrix with respect to the permuted labels is

$$T = \begin{pmatrix} \mathcal{J}_r & \cdot & \cdot & \cdot \\ \cdot & \mathcal{J}_r & \emptyset & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \emptyset & & & \mathcal{J}_r \end{pmatrix}$$

where $\mathcal{J}_r$ is the column of r ones. We want the treatment incidence matrix with respect to the original labels in order to get a general understanding of the results of the randomization procedure. Let $j' = (k-1)r + j$, and let T^* be the treatment incidence matrix with respect to the original labels. If $i_{k(j)} = \ell$, then $T^*_{\ell k} = 1$ and $T^*_{\ell k'} = 0$ for $k \neq k'$. Also $\pi_{\ell j'} = 1$ and $\pi_{\ell' j'} = 0$ for $(\ell', j') \neq (\ell, j')$. So we have $\pi_{\ell j'} T_{j'k} = 1$ if the ℓ-th unit is assigned to treatment group k and equal to zero otherwise. Consequently, we have the algebraic relation $T^*_{\ell k} = \sum_{j=1}^{N} \pi_{\ell j} T_{\ell k}$ or, equivalently, $T^* = \pi T$.

With the original labelling and ordering of units, the model is

$$Y = T^*\tau + X\beta + e$$

in which T^* is the random incidence matrix and $X\beta$ represents the concomitant contribution. We apply least squares to this data-model set up. We assume (T^*, X) to be of full rank equal to t+p. The normal equations are then given by

$$\begin{pmatrix} T^{*\prime}T^{*} & T^{*\prime}X \\ X'T^{*} & X'X \end{pmatrix} \begin{pmatrix} \tau \\ \beta \end{pmatrix} = \begin{pmatrix} T^{*\prime}Y \\ X'Y \end{pmatrix}$$

with $$T^{*\prime}T^{*} = T'\pi'\pi T = T'T \ .$$

The Reduced Normal Equation for $\hat{\beta}$ is

$$X'(I - T^{*}(T^{*\prime}T^{*})^{-1}T^{*\prime})X\hat{\beta} = X'(I - T^{*}(T^{*\prime}T^{*})^{-1}T^{*\prime})Y$$

Let P_τ^{*} denote $T^{*}(T^{*\prime}T^{*})^{-1}T^{*} = \pi T(T'T)^{-1}T'\pi'$.

Then P_τ is (a) Idempotent $P_\tau^{*2} = P_\tau^{*}$

(b) Symmetric $P_\tau^{*\prime} = P_\tau^{*}$

Hence P_τ^{*} is a projection operator.

The reduced normal equations for τ are given by

$$T^{*\prime}(I - X(X'X)^{-1}X')T^{*}\tau = T^{*\prime}(I - X(X'X)^{-1}X')Y$$

$$T'\pi'(I - X(X'X)^{-1} X')\pi T\tau = T'\pi'(I - X(X'X)^{-1}X')\pi \pi'Y$$

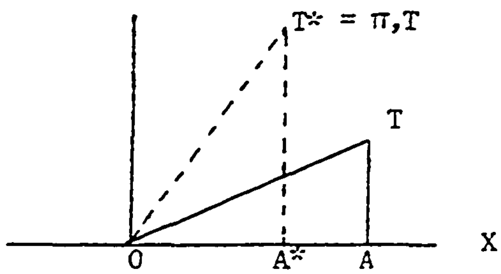
Now let $$P_X^{*} = \pi'X(X'X)^{-1}X'\pi$$

Then P_X^{*} is idempotent, i.e., $P_X^{*2} = P_X^{*}$

and P_X^{*} is symmetric, i.e., $P_X^{*\prime} = P_X^{*}$.

Hence P_X^{*} is a projector operator.

If we could examine the families of projection operators P_X^{*}, P_τ^{*} over the $t!\ N!/(r!)t$ possible outcomes of the set of random matrices $\{\pi\}$ we would see how the estimates $\hat{\tau}$ as well as the reduced sum of squares for τ, $\hat{\tau}'T^{*\prime}(I - X(X'X)^{-1}X')Y$, on the one hand, and $\hat{\beta}$ and $\hat{\beta}'X'(I - P_\tau^{*})Y$, on the other, vary under the randomization scheme. $\pi_1 T$ is obviously a rotation of the columns of T in the N dimensional space of real numbers. The consequence of this is to change, the projection of the column space of T into the column space of X, and, simultaneously, the space orthogonal to X, as suggested by the following diagram.



Symbolically the "vertical" components AT and A^*T^* are orthogonal to column space of X and are identified with, $T'(I - P_X)$ and $T'\pi'(I - P_X)$, while OA and OA^* are identified with $T'P_X$ and $T'\pi'P_X$.

We note that under the null hypothesis of no treatment effects the set of observed responses y, with respect to the initial indexing, is the same for each permutation, i.e. Y = U. Also P_XY and P_XT are unchanged. Hence, the randomization distribution of the various statistics of interest may be considered as being completely determined by the set of permutation $\{\pi\}$. In the following paragraphs we shall show that under additivity of treatment and unit effects, the CRD may have bias in the analysis of variance under randomization theory.

3.2.2 Bias (AOV) in the Analysis of Covariance

Using the notation of Chapter 1 for the completely randomized design (CRD), T, R and G represent partitions of any sum of squares or products into treatments, residuals and total respectively, while the subscripts xx, xy and yy denote the sum of squares for x, the sum of products for x and y and the sum of squares for y respectively. Of course, for each of these, $T + R = G$. Let A_{yy} and E_{yy} denote the adjusted sums of squares for treatments and residual error respectively, which are given as

$$A_{yy} = G_{yy} - G_{xy}G_{xx}^{-1}G_{xy} - (R_{yy} - R_{xy}R_{xx}^{-1}R_{xy})$$

and

$$E_{yy} = R_{yy} - R_{xy}R_{xx}^{-1}R_{xy} \quad .$$

Then with t treatments and $N = rt$ experimental units the corresponding mean squares are given by

$$MS(A) = T_{yy}/(t-1) \text{ and } MS(E) = E_{yy}/(tr - t - 1),$$

so that

$$MS(A) - MS(E) =$$

$$\frac{G_{yy} - G_{xy}G_{xx}^{-1}G_{xy}}{(t-1)} - \frac{(tr - 2)}{(tr-t-1)} \frac{(R_{yy} - R_{xy}R_{xx}^{-1}R_{xy})}{(t-1)}$$

Let E_R denote expectation under randomization. We have

$$E_R(R_{yy}) = G_{yy} \cdot t(r-1)/(tr-1),$$

and the difference in expected mean squares (EMS) is given by

$$\alpha = EMS(A) - EMS(E) =$$

$$(t-1)^{-1} \left[\frac{(tr-2)}{(tr-t-1)} \cdot E_R (R_{xy} R_{xx}^{-1} R_{xy}) - G_{xy} G_{xx}^{-1} G_{xy} - \frac{(t-1)G_{yy}}{(tr-1)(tr-t-1)} \right]$$

We are concerned with 2 questions:

1) in the absence of treatment effects, is $\alpha = 0$

and

2) if there are additive treatment effects, is $\alpha > 0$?

The critical aspect of the criteria above is $E_R \{R_{xy} R_{xx}^{-1} R_{xy}\}$. If R_{xx} is zero for a certain plan, then the x values in each treatment group are the same, so that $R_{xy} = 0$. We then have that $R_{xy}\ R_{xx}^{-1}\ R_{xy}$ is indeterminate. We shall ignore this possibility because it depends on a very special configuration in the concomitant values. In the case of $G_{xy} = 0$, if R_{xx} in a plan is very small it is possible that $E_R\{R_{xy}R_{xx}^{-1}R_{xy}\}$ would be appreciably larger than $(t-1)G_{yy}/(tr-1)(tr-2)$. This suggests that we may have sequences of $\{(U_i,X_i)\}$ in which the AOV bias is appreciable. In general, it is not possible, it seems to establish a general result about whether EMS(A) - EMS(E) is positive or negative under the null hypothesis of no treatment effects. We can see, however, that if we have a sequence of unit sets $\{(y_i,x_i)\}$ such that G_{yy}/N is bounded, then as $r \to \infty$ this tends to the limiting value of $\frac{1}{(t-1)}\{\ E_R(R_{xy}\ R_{xx}^{-1}\ R_{xy}) - G_{xy}\ G_{xx}^{-1}\ G_{xy}\}$. It seems that there is no general reason why this should be positive, zero, or negative. We shall endeavor, inter alia, to get some idea of the magnitude of this. It is of interest, we think, to give a very small and unrealistic example in which we can see exactly what happens.

Example:

Suppose we have a set (4 pairs, say) of observations $\{(x_1,y_1),(x_2,y_2),(x_3,y_3),(x_4,y_4)\}$, subtracted from their mean so that $\Sigma_i y_i = 0 = \Sigma x_i$. We wish to partition the set into 2 groups of two unit pairs each and to assign two dummy treatments to the group at random. There are 2! ways of assigning treatments to groups and there are $\frac{4!}{2!2!} = 6$ ways of choosing the units for the first group. The remaining units form the second group. But there are only $\frac{1}{2} \cdot 6 = 3$ distinct partitions of the set into groups. Given that the dummy treatments have been assigned to groups, the possible partitions and the corresponding analyses of covariance under the null hypothesis of no treatment differences are given below for a simple case.

Suppose $x_1 = 1$, $x_2 = -1$, $x_3 = 0$, $x_4 = 0$, and $\Sigma_i y_i = 0$. Then $G_{xy} = y_1 - y_2$, $G_{xx} = 2$, $G_{yy} = y_1^2 + y_2^2 + y_3^2 + y_4^2$, and we have the following plans.

Partition	Group 1 Y	Group 1 X	Group 2 Y	Group 2 X	Source	Analysis of Covariance xy	xx	yy
1	y_1	1	y_3	0	T	0	0	.
	y_2	-1	y_4	0	R	$y_1 - y_2$	2	.
2	y_1	1	y_2	-1	T	$\frac{1}{2}(y_1 - y_2 + y_3 - y_4)$	1	.
	y_3	0	y_4	0	R	$\frac{1}{2}(y_1 - y_2 - y_3 + y_4)$	1	.
3	y_1	1	y_2	-1	T	$\frac{1}{2}(y_1 - y_2 - y_3 + y_4)$	1	.
	y_4	0	y_3	0	R	$\frac{1}{2}(y_1 - y_2 + y_3 - y_4)$	1	.

$$E \frac{R_{xy}^2}{R_{xx}} = \frac{1}{3}\{\tfrac{1}{2}[(y_1 - y_2)^2 + (y_1 - y_2)^2 + (y_3 - y_4)^2]\}$$

$$= \frac{1}{6}[2(y_1 - y_2)^2 + (y_3 - y_4)^2]$$

$$EMS(A) - EMS(E) = G_{yy} - \frac{G_{xy}^2}{G_{xx}} - 2EMS(E)$$

$$= \quad G_{yy} - 2E(R_{yy}) - \frac{G_{xy}^2}{G_{xx}} + 2E\left(\frac{R_{xy}^2}{R_{xx}}\right)$$

$$= \quad (1 - 2 \cdot \tfrac{2}{3})\, \Sigma y_i^2 - \{\tfrac{1}{2}(y_1 - y_2)^2 - \tfrac{2}{3}(y_1 - y_2)^2 - \tfrac{1}{3}(y_3 - y_4)^2\}$$

$$= \quad -\frac{1}{3}\Sigma y_i^2 + \frac{1}{6}(y_1 - y_2)^2 + \frac{1}{3}(y_3 - y_4)^2$$

$$= \quad - \frac{1}{6}\{2y_1^2 + 2y_2^2 - (y_1 - y_2)^2 + 2(y_3^2 + y_4^2) - 2(y_3 - y_4)^2\}$$

$$= \quad - \frac{1}{6}\{(y_1 + y_2)^2 + 4y_3y_4\}$$

$$= \quad - \frac{1}{6}\{(y_3 + y_4)^2 + 4y_3y_4\}$$

$$= \quad - \frac{1}{6}\{(y_3 + 3y_4)^2 - 8y_4^2\}$$

Bias in this solution is

a) > 0 if $\quad -3 - \sqrt{8} < \dfrac{y_3}{y_4} < -3 + \sqrt{8}$

b) $= 0$ if $\quad \dfrac{y_3}{y_4} = -3 \pm \sqrt{8}$, or if $y_3 = y_4 = 0$

c) < 0, otherwise

The Gauss-Markoff linear regression of y on X is given by $\beta = G_{xy}/G_{xx}$ $= \frac{1}{2}(y_1 - y_2)$. Its estimate from the analysis of covariance procedure is R_{xy}/R_{xx} and its expected value for this example is

$$E(R_{xy}/R_{xx}) = \frac{\frac{1}{3}\{(y_1 - y_2) + (y_1 - y_2 - y_3 + y_4) + (y_1 - y_2 + y_3 - y_4)\}}{2}$$

$$= \frac{(y_1 - y_2)}{2}$$

So in this particular case, the analysis of covariance gives an unbiassed estimate of the product moment regression of y on X.

In general, however, the expected value of $\hat{\beta}$ is not equal to β for

$$E\,\frac{R_{xy}}{R_{xx}} \neq \frac{ER_{xy}}{ER_{xx}} = \frac{G_{xy}}{G_{xx}}$$

3.2.3 Unbiasedness of Treatment Contrasts

In the situation of the completely randomized design we partition the N = rt vectors $\{U_i, X_i\}$, using some simple random mechanism, into t groups of r and assign the t treatments to groups at random.

Let z_{ik} be the response on unit i if treatment k is applied to it. Then we will have a conceptual population of responses

$$z_{ik}\ , \quad i = 1, 2, \ldots, N; \quad k = 1, 2, \ldots, t$$

In actual experiments, however, each unit can receive only one treatment, so we will observe for each i only one z_{ik} , $k=1,2, \ldots, t$.

The randomization procedure is defined by the following sampling random variables $\delta^i_{k(j)}$ where $\delta^i_{k(j)} = 1$ if unit i is the j-th replicate unit which receives treatment k, and zero otherwise.

If we assume additivity we have

$$z_{ik} = u_i + \tau_k = u_. + (u_i - u_.) + \tau_k \quad ,$$

where $u_. = \frac{1}{N} \Sigma_{i=1}^{N} u_i$.

Hence if $y_{k(j)}$ is the observed response of the j-th replicate of treatment k, we will have the observations $y_{k(j)}$, $x_{k(j)}$, where

$$y_{k(j)} = u_. + \tau_k + \sum_i \delta^i_{k(j)} (u_i - u_.)$$

$$x_{k(j)} = x_. + \sum_i \delta^i_{k(j)} (x_i - x_.)$$

We shall then, ordinarily, use the processes of the Gauss-Markov covariance linear model. This will lead to a statistic $\hat{\beta}$, which is intended to reflect the linear association of the unknown $\{u_i\}$ with the known $\{x_i\}$:

$$\hat{\beta} = \frac{\sum_k \sum_j [y_{k(j)} - y_{k(.)}][x_{k(j)} - x_{k(.)}]}{\sum_k \sum_j (x_{k(j)} - x_{k(.)})^2} = \frac{R_{xy}}{R_{xx}}, \quad \text{say.}$$

Then to estimate treatment differences, we use

$$\widehat{\tau_k - \tau_{k'(.)}} = y_{k(.)} - y_{k'(.)} - \hat{\beta}[x_{k(.)} - x_{k'(.)}] ,$$

where $$y_{k(.)} = \frac{1}{r} \sum_{j=1} y_{k(j)} .$$

Part of our task is to develop an understanding of $\hat{\beta}$ and $\widehat{\tau_k - \tau_k}$. We shall look at the latter first. We use E_R to denote expectations over randomizations and E_N, expectations under GMN assumptions. So

$$E_R(\widehat{\tau_k - \tau_{k'}}) = E_R[y_{k(.)} - y_{k'(.)}] - E_R[\hat{\beta}(x_{k(.)} - x_{k'(.)})]$$

$$= \tau_k - \tau_{k'} - E_R\{\hat{\beta}(x_{k(.)} - x_{k'(.)})\}$$

We use here the standard property which holds in the absence of a covariate, that $E_R[y_{k(.)} - y_{k'(.)}] = \tau_k - \tau_{k'}$, assuming that we have additivity of treatment and unit effects. The first property of interest is that

$$E_R \{\hat{\beta} (x_{k(.)} - x_{k'(.)})\} = 0$$

To see this, consider first the partitioning, π, of the N units into t groups of r. This can be done in $\frac{N!}{(r!)^t}$ ways. Let a partitioning give the groups $S_1, S_2, \ldots, S_r$. Then there are t ways of assigning the t treatments to the t groups. Consider a particular partition. Then under additivity of treatments,

$$y_{k(j)} - y_{k(.)} = u_{k(j)} - u_{k(.)} \quad .$$

Hence R_{xy}, R_{xx}, R_{yy} are independent of treatment effects , and so

$$\hat{\beta} = \frac{R_{xy}}{R_{xx}} = \frac{\sum_k \sum_j (u_{k(j)} - u_{k(.)})(x_{k(j)} - x_{k(.)})}{\sum_{k,j} [x_{k(j)} - x_{k(.)}]^2}$$

Clearly, $\hat{\beta}$ does not depend on what the assignment of treatments to the groups of units is. Call this $\hat{\beta}_\pi$. Then

$$E_R\{\hat{\beta}\,[x_{k(.)} - x_{k'(.)}]\} = E_{\tau,\pi}\{\hat{\beta}_\pi\,[x_{k(.)} - x_{k'(.)}]\},$$

where $E_{\tau,\pi}$ is expectation over partitionings and assignment of treatments to the groups given the partitioning. But this is equal to

$$E_\pi\{E_{\tau|\pi}[\hat{\beta}_\pi(x_{k(.)} - x_{k'(.)})]\}$$

$$E_\pi\{\hat{\beta}_\pi[E_{\tau|\pi}(x_{k(.)} - x_{k'(.)})]\}$$

and $\qquad E_{\tau|\pi}[(x_{k(.)} - x_{k'(.)})] = 0 \quad$ so, $\quad E[\hat{\beta}(x_{k(.)} - x_{k'(.)})] = 0$

because each group has the same probability, equal to $\frac{1}{t}$, of receiving any treatment. Hence, writing the estimated difference in treatment effects adjusted for the covariate as

$$(\hat{\tau_k - \tau_{k'}}) = (\tilde{\tau_k - \tau_{k'}}) - \hat{\beta}(x_{k(.)} - x_{k'(.)}),\ 1 \leq k,\ k' \leq t,$$

where $(\tilde{\tau_k - \tau_{k'}}) = (y_{k.} - y_{k'.})$ is the usual estimate of $\tau_k - \tau_{k'}$ in the absence of the covariate and $\hat{\beta}(x_{k(.)} - x_{k'(.)})$ is the adjustment for the covariate, under additivity, we have

$$E(\widehat{\tau_k - \tau_{k'}}) = \tau_k - \tau_{k'} \quad .$$

We may state the above developments as,

Theorem 3.1

With additivity of treatments and units, the estimator of any treatment and units, the estimator of any treatment contrast obtained by the covariance procedure is unbiased under randomization.

This result of course, only the most primitive desired property of randomization. It is worth noting, in passing, that we see already a difference from the case of there being no covariate, because in that case it is elementary that a difference of observed treatment means is an unbiassed estimator of the difference that would be observed if one could place of the treatments k and k' on every unit, and then took the difference of observed mean. That is, without covariance adjustment, we have unbiassedness of treatment comparisons without assumption of additivity of treatments and units. This suggests that role of additivity of treatment and unit effects is much more critical in the covariate case than the non-covariate case. There is a major point with respect to covariance. In section 3.1.2 we showed that under GMN assumptions, if

$$y = X\beta + Z\gamma + e$$

is the correct model, then $\beta = (X'X)^{-1}X'y$ is not unbiassed for β. In the case of the 2 part GM infinite model,

$$y_{kj} = \mu + \tau_k + \beta x_{kj} + e_{kj} \quad ,$$

the least squares (LS) estimator of any treatment contrast, $\Sigma_k \lambda_k \tau_k$, $(\Sigma\lambda_k = 0)$, is given by

$$\Sigma \lambda_k \hat{\tau}_k = \Sigma_k \lambda_k y_{k.} - \beta\Sigma_k \lambda_k x_{k.} \quad .$$

However, if we ignored the covariate the LS estimator would be given by

$$\Sigma_k \lambda_k \tilde{\tau}_k = \Sigma_k \lambda_k y_{k.}$$

Its expectation under GMN assumptions, $E_N(\Sigma_k \lambda_k \tilde{\tau}_k)$ is not equal to $\Sigma_k \lambda_k \tau_k$ unless $\hat{\beta} = 0$ or $\Sigma_k \lambda_k x_{k.} = 0$. But under randomization and additivity we have $E_R(\Sigma_k \lambda_k \tilde{y}_{k.}) = \Sigma_k \lambda_k \tau_k$. So in the randomization experiment with Z as concomitant and with additivity, if we ignored the concomitant we would still obtain unbiassed estimates of treatment contrasts. So there seems to be a logico philosophical problem.

We shall continue with the case in which treatments and units are additive. It then becomes of deep interest to form ideas of properties of subsequent statistical computations. We wish to understand the nature of $\hat{\beta}$. We have seen that it does not depend on the treatment effects. We would surmise that $\hat{\beta}$ is an indicator or an estimator of the slope of the mean square linear regression of the unknown $\{u_i\}$ on the known $\{x_i\}$, with the whole set of doublets $\{(u_i, x_i)\}$. It is appropriate then to attempt to obtain an indicator of $E(\hat{\beta})$, and this we shall do in the next section. We shall also determine the expectation under randomization of the first two moments of various statistics of interest.

3.2.4 Expectations of Various Statistics under Randomization

Our estimate $\hat{\beta}$, of the usual linear regression slope, in the analysis of covariance situation is given by

$$\hat{\beta} = \frac{R_{yx}}{R_{xx}}$$

Under the Gauss-Markov linear model assumption,

$$y_{kj} = \mu + \tau_k + bx_{kj} + e_{kj}$$

with $E_N(e_{kj}) = 0$, for all k, j and x_{kj} known and fixed in repetitions, it is elementary to see that

$$E_N(\hat{\beta}) = \beta$$

where E_N denotes expectation under Gauss-Markov Normal (GMN) conditions.

Our primary task is to form reasonable ideas about the behavior of the GMN Statistics under randomization. We shall proceed in a natural order including here theoretical facts from the literature that are strongly relevant. It is clear that the statistics R_{yy}, R_{yx} and R_{xx} are critical, so it is natural to consider these. In the context of the completely randomized design, the first and second moments of these may be derived from the work of Wilk (1953). We work in detail the variances and covariances of these statistics in Appendix C. To obtain the results, we modify the notation of Wilk (1953) for our purposes. We may write under additivity of treatment and unit effects in the permissible population of all possible responses, i.e. with every treatment being placed on every unit and with y_{ik} = (conceptual) response on unit i with treatment k:

$$\begin{aligned} y_{ik} &= \tau_k + u_i \\ &= u_{.} + \tau_k + (u_i - u_{.}) \end{aligned}$$

Then if we write

$$\sigma_{rw} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - x_{.})^r (u_i - u_{.})^w = \frac{1}{N-1} S_{rw}$$

we will have

$$E_R (R_{xy}) = t(r-1)\, \sigma_{11} = \frac{t(r-1)}{N-1} S_{11}$$

$$V_R\ (R_{xy}) = \frac{(t-1)(r-1)}{r(N-1)(N-2)(N-3)}\ \left[\ (N-2)\cdot S_{20}\ S_{02} + \frac{(N^2-3N+4)}{N-1}\ S_{11}^{\ 2}\ -2NS_{22})\right]$$

$$E_R\ (T_{xy}) = (t-1)\sigma_{11} = \frac{(t-1)}{(N-1)}\ S_{11}$$

$$V_R\ (T_{xy}) = V_R\ (R_{yx})$$

$$Cov_R(R_{xy}, T_{xy}) = -\ V_R\ (R_{xy})$$

We may then write down the same properties for R_{yy}, T_{yy} by replacing x by u, σ_{11} by σ_{02}, and S_{11} by S_{02}, and we get

$$E_R\ (R_{yy}) = t(r-1)\ \sigma_{02} = \frac{t(r-1)}{N-1}\ S_{02}$$

$$V_R\ (R_{yy}) = \frac{(t-1)(r-1)}{r(N-1)(N-2)(N-3)}\ \left(\ \frac{2(N^2-3N+3)}{N-1}\ S_{02}^{\ 2} - 2N\ S_{04}\ \right)$$

etc.

Similarly, we will have

$$E_R\ (R_{xx}) = t(r-1)\ \sigma_{20} = \frac{t(r-1)}{N-1}\ \ S_{40}$$

Also,

$$Cov_R\ (R_{yy}, R_{xx}) = \frac{2(t-1)(r-1)}{r(N-1)(N-2)(N-3)}\ \left(\ (N-2)S_{11}^{\ 2} + \frac{1}{N-1}\ S_{20}S_{02} - NS_{22}\ \right)$$

$$Cov_R\ (R_{xx}, R_{yx}) = \frac{2(t-1)(r-1)}{r(N-1)(N-2)(N-3)}\ \left(\ \frac{(N^2-3N+3)}{N-1}\ \ S_{20}\ S_{11}\ -\ \ NS_{31}\ \ \right)$$

We shall need these results as our thesis develops but now we shall return to our inquiry about the problems of the LS estimates, etc.

Part of our inquiry is directed towards understanding the behavior of the above statistics and statistical tests under the hypothesis only of randomization over a set sequence of N vectors (u_i, x_i) $i = 1,2, \ldots, N$ as N increases.

Kempthorne (1952) noted that the usual randomization basis which is followed in the absence of covariate information does not carry over to the case with a covariate because with T^*, E^* denoting the adjusted treatment mean square and the adjusted residual mean square, respectively, it is not the case that the expectations of these under randomization have the properties

1) $E(T^*) = E(R^*)$ if there are no treatment effects, and

2) If there treatment effects, then necessarily $E(T^*) > E(R^*)$

Cox (1956), whose work we now review briefly, examined these questions to some extent. The bias in the analysis of variance derives from the expectation of the quotient of two statistics, R^2_{xy}/R_{xx}. Cox (1956) has shown that in these simple statistical designs in which adjustments for a concomitant variable are made by the analysis of covariance, an unbiassed between treatments mean square can be produced by weighted randomization.

3.2.5 Weighted Randomization

Suppose that the values $(x_1, \ldots, x_n)$ are available to the experimenter prior to the allocation of treatments to units. Let w be any nonnegative function of $x_1, \ldots, x_n$, defined for each partition of the units into treatment groups. Suppose we have a mechanism which gives each partition of the units in the set a probability of selection proportional to w. This is general weighted randomization. If f is any function of the observations y and x, its expectation under weighted randomization is E(f), equal to $(\Sigma wf)/(\Sigma w)$. By exploiting the symmetrical properties of the expectation under randomization of R_{xx}, R_{xy}, R_{yy}, Cox showed that under the null hypothesis of no treatment effects and weighted randomization with weight proportional to R_{xx}, the expected value of the adjusted treatment mean square and error mean square is proportional to the expected value under randomization of $(R_{xx}R_{yy} - R_{xy}^2)$ and is equal to

$$\frac{1}{N-2}\left(G_{yy} - \frac{G_{xy}^{\ 2}}{G_{xx}}\right) = \sigma_r^{\ 2}\ \text{, say.}$$

This result is true as long as additivity between the treatment effects and unit effects holds, because R_{xx}, R_{xy}, R_{yy} are then independent of treatment effects. Cox called attention to the following:

a) Arrangements with a large value of R_{xx} will have a small value for T_{xx} and conversely. Hence, the weighting proportional to R_{xx} attaches greater chance of selection to those arrangements in which the treatment groups are balanced with respect to the mean value of x.

b) Weighted randomization is, of course, restricted to cases in which the concomitant variable is available prior to the allocation of treatments to units.

c) If weighted randomization is to be done in practice with N not very small, some short-cut method is needed for selecting an arrangement since the enumeration of all arrangements and the calculation of R_{xx} for each would usually be too tedious.

A simple method for weighted randomization given by Professor Tukey is as follows: Let M be the maximum over all arrangements of R_{xx}. Select an arrangement by unweighted randomization and calculate R_{xx} for it. Reject the arrangement with probability $1 - R_{xx}/M$. Continue until an arrangement is accepted.

Conditional on the assignment of units to groups, treatment contrasts are estimable under randomization of treatments to groups. Hence they are estimable under weighted randomization. However, it is not possible to obtain unbiased estimates of their variances.

3.3 Limiting Expectations under Randomization for Large Finite Populations

Suppose we have a sequence of finite populations $\varphi_1, \varphi_2, \ldots, \varphi_r$ containing $N_r(=rt)$ vector units $\{X_{ir}, U_{ir}\}$, $i=1,2, \ldots, N = rt$, where t is a fixed positive integer. Suppose we partition the population φ_r into t groups of r vectors by random permutation. Denote the mean value of the vectors in group k by $(\bar{X}_{kr}, \bar{U}_{kr})$. Dropping the subscript r, we have

Theorem 3.2

Let

$$R_{xu} = \sum_{i=1}^{N} U_i X_i - r \sum_{k=1}^{t} \bar{U}_k \bar{X}_k ,$$

$$R_{xx} = \sum_{i=1}^{N} X_i^2 - r \sum_{k=1}^{t} \bar{X}_k^2 ,$$

$$R_{uu} = \sum_{i=1}^{N} U_i^2 - r \sum_{k=1}^{t} \bar{U}_k^2 ,$$

and

$$S_{pq} = \sum_{i=1}^{N} (X_i - \bar{X})^p (U_i - \bar{U})^q$$

Then the coefficient of variation of R_{xxr} under randomization, i.e. all possible permutations, is given by C.V.(R_{xxr}) =

$$\sqrt{\left\{\frac{2(t-1)(r-1)(N^2-3N+3)}{r(N-1)^2(N-2)(N-3)}\left(1-\frac{N(N-1)}{(N^2-3N+3)}\right)\cdot\frac{S_{40}}{S_{20}^2}\right\}}\cdot S_{20}$$

$$\div\ \frac{(N_r - t)}{(N_r - 1)}\ S_{20}$$

The proof of the theorem is easily derived from the variances and expectations of the $\{R_{\cdot\cdot}\}$ which are derived in Appendix C.

If the sequences φ_{N_r} satisfy the most general of the conditions presented earlier, Noether condition ν, we have the limit

$$\lim_{N_r \to \infty} \frac{S_{40}}{S_{20}^2} = 0$$

The coefficient of variation is then approximately $\sqrt{\frac{2(t-1)}{N_r}}$. The implication of this result is that for $N_r = (rt)$ sufficiently large, and t fixed, in our estimation of the regression coefficients $\hat{\beta}_{N_r} = R_{xxN_r}^{-1} R_{xyN_r}$, and the corresponding sum of squares $\hat{\beta}_{N_r} R_{xyN_r} = R_{xyN_r} R_{xxN_r}^{-1} R_{xyN_r}$, we may consider the ratio $\frac{R_{xxN_r}}{E(R_{xxN_r})}$ to be almost constant for all the possible random assignments of treatments to units without fear of making gross errors. By Tchebychev's inequality

$$P\left\{\frac{R_{xxN_r} - E_R(R_{xxN_r})}{E_R(R_{xxN_r})} \geq \epsilon\right\} \leq \frac{2(t-1)}{N_r^2 \epsilon^2} \longrightarrow 0 \text{ for every } \epsilon > 0.$$ Hence

we may expand small inverse powers of R_{xxN_r} to their second order approximation.

Let $CV_{..N_r}$ denote the coefficient of variation of the relevant residual sum of squares or cross products. Dropping the subscript r we have

$$CV_{20N}^{\;2} = \frac{Var(R_{xxN})}{\{E_R(R_{xxN})\}^2}$$

$$= \frac{2(r-1)(t-1)(N^2-3N+3)}{r(N-1)^2(N-2)(N-3)} \cdot S_{20}^{\;2}$$

$$\cdot \{1- \{ \frac{N(N-1)}{(N^2-3N+3)} \cdot \frac{S_{40N}}{S_{20N}^{\;2}} \} \quad \frac{(N-t)^2}{(N-1)^2} \cdot S_2^2$$

Similarly,

$$CV_{11N}^{\;2} = \frac{(t-1)(r-1)(N^2-3N+2)}{r(N-t)^2(N-2)(N-3)} \cdot \frac{S_{20N}S_{02N}}{S_{11N}^{\;2}} +$$

$$\frac{(N^2 - 3N + 4)}{(N^2 - 3N + 2)} - \frac{N}{(N-2)} \cdot \frac{S_{22N}}{S_{11N}^2}$$

$$CV_{20,11N} = \frac{cov(R_{xxN}, R_{xyN})}{E_R(R_{xxN})\, E_R(R_{xyN})}$$

$$= \frac{2(t-1)(r-1)(N^2 - 3N + 3)}{r(N-t)^2\,(N-2)(N-3)} \{ 1 - \frac{N(N-1)}{(N^2 - 3N + 3)} \cdot \frac{S_{31N}}{S_{20N}S_{11N}} \}$$

$$CV_{20,02N} = \frac{cov(R_{xxN}, R_{yyN})}{E_R(R_{xxN})\, E_R(R_{yyN})}$$

$$= \frac{2(t-1)(r-1)}{r(N-t)^2(N-3)} \cdot \{ \frac{(N-1)S_{11N}^2}{S_{20N}S_{02N}} + \frac{1}{(N-2)}$$

When (x,y) is bivariate normally distributed with correlation coefficient ρ and N is large, we have the following:

$$\lim_{N \to \infty} \frac{(N-1)\, S_{40N}}{S_{20N}^{\;2}} = 3 \quad ,$$

$$\lim_{N\to\infty} r_N^2 = \lim_{N\to\infty} \frac{S_{11N}^2}{S_{20N}S_{02N}} = \rho^2 ,$$ where ρ is the correlation coefficient of the population and r_N is the correlation coefficient of the sample of size N,

$$\lim_{N\to\infty} \frac{(N-1)S_{22N}}{S_{11N}^2} = \frac{1+2\rho^2}{\rho^2}$$

and

$$\lim_{N\to\infty} \frac{(N-1)S_{31N}}{S_{20N}S_{11N}} = 3$$

If the sequences of populations φ_{N_r} are such that the sequences $\{X_1, X_2, \ldots, X_{N_r}\}$ and $\{U_1, U_2, \ldots, U_{N_r}\}$ each satisfy the conditions ν of Noether and if the associated correlation coefficients ρ_N converges to $\rho \neq 0$, then under the randomization model we have the following:

$$\lim_{N\to\infty} CV_{20N}^2 = \frac{2(t-1)}{N^2} = \lim_{N\to\infty} CV_{02N}^2 ,$$

$$\lim_{N\to\infty} CV_{11N}^2 = \frac{(t-1)}{N^2}\left(1 + \frac{1}{\rho^2}\right) , \quad \text{where } \rho_N = \frac{S_{11N}}{\sqrt{(S_{02N}S_{20N})}}$$

$$\lim_{N\to\infty} \frac{Cov(R_{xxN}, R_{xyN})}{E_R(R_{xxN})E_R(R_{xyN})} = \frac{2(t-1)}{N^2} ,$$

and with the mixed product powers of treatment totals denoted by Q_{ij} ,

$$\hat{\beta} = \frac{S_{11N} - \frac{1}{r} Q_{11N}}{S_{20N} - \frac{1}{r} Q_{20N}} = \frac{R_{xyN}}{R_{xxN}}$$

$$= E_R(R_{xyN}) \{1 + \frac{R_{xyN} - E_R(R_{xyN})}{E_R(R_{xyN})} \} .$$

$$\{E_R R_{xxN}(1 + \frac{R_{xxN} - E_R(R_{xxN})}{E_R(R_{xxN})} \}^{-1}$$

Put $$\theta = \frac{R_{xxN} - E_R(R_{xxN})}{E_R(R_{xxN})} .$$

If we use the Taylor expansion of $f(\theta) = (1 + \theta)^{-1}$ around $\theta = 0$, we will have $f(\theta) = 1 - \theta + \theta^2 - \frac{1}{3!} \theta^3 f^{(3)}(\lambda\theta)$, for some $0 < \lambda < 1$. The remainder term is $R(\theta,\lambda) = -\theta^3 (1 + \theta\lambda)^{-4}$ and is bounded by $-\theta^3$ and $-\theta^3(1 + \theta)^{-4}$.

Similarly, for $g(\theta) = (1 + \theta)^{-2}$, we will have

$$g(\theta) = 1 - 2\theta + 3\theta^2 - 4\theta^3(1 - \theta\lambda)^{-4} , \quad 0 < \lambda < 1 .$$

Let $$\beta_N = S_{11N}/S_{20N}$$

and $$\eta = \frac{R_{xyN} - E_R(R_{xyN})}{E_R(R_{xyN})} \quad \text{and } \gamma = \frac{R_{yyN} - E_R(R_{yyN})}{E_R(R_{yyN})} .$$

If the sequences of populations satisfy the Noether conditions we will have

$$\begin{aligned} E_R(\hat{\beta}_N) &= E_R \{\tfrac{N-t}{N-1} S_{11N} (1+\eta)\} \{\tfrac{N-t}{N-1} S_{20N} (1+\theta)\}^{-1} \\ &= \beta_N E_R\{(1+\eta)(1-\theta+\theta^2+R(\theta,\lambda)\} \\ &= \beta_N + \beta[\mathrm{Var}(\theta) - \mathrm{cov}(\eta,\theta) + E_R\{R(\theta,\lambda) + \\ &\quad R(\theta,\lambda) + \theta^2)\eta\}] \\ &= \beta_N + \beta_N[CV_{20N}{}^2 - CV_{20,11N} + o(N^{-2})] \\ &= \beta_N + o(N^{-2}) . \end{aligned}$$

Similarly, $\quad \mathrm{Var}\,(\beta_N) = \beta_N{}^2\{\mathrm{Var}\,(\theta) + \mathrm{Var}\,(\eta) - 2\,\mathrm{Cov}\,(\theta,\eta) + o(N^{-2})\}$

Using the approximations on page 78 we obtain

a) $E_R(\hat{\beta}_N) - \beta_N = o(N^{-2})$

and

b) $$\mathrm{Var}_R(\hat{\beta}_N) = \beta_N{}^2\,\frac{(t-1)}{N^2}\,\frac{(1-\rho_N{}^2)}{\rho_N{}^2} + o(N^{-2})$$

$$= \frac{S_{yyN}(1-\rho_N{}^2)}{S_{xxN}} \cdot \frac{(t-1)}{N^2}$$

$$= S_{20N}{}^{-1}(1-\rho_N{}^2)\,\frac{(t-1)}{N}\,(S_{02N}/N)$$

This result differs from the GMN result by a factor f where $f = \frac{t-1}{N-1}$.

Now $$r\Sigma_i\hat{\tau}_i{}^2 = r\Sigma_k\left[(y_k - y_{..}) - \frac{R_{xy}}{R_{xx}}\,(x_k - x_{..})\right]^2$$

$$= r\Sigma(y_k - y_{..})^2 - 2\,\frac{R_{xy}}{R_{xx}}\,(G_{xy} - R_{xy}) + \frac{R_{xy}{}^2}{R_{xx}{}^2}\,(G_{xx} - R_{xx})$$

$$= T_{yy} + \frac{R_{xy}{}^2}{R_{xx}} - \frac{G_{xy}{}^2}{G_{xx}} + G_{xx}\,\frac{G_{xy}{}^2}{G_{xx}{}^2} - 2\frac{R_{xy}}{R_{xx}} \cdot \frac{G_{xy}}{G_{xx}} + \frac{R_{xy}{}^2}{R_{xx}{}^2}$$

= Adjusted Treatment SS + $S_{20N}(\hat{\beta} - \beta)^2$.

If the sequences $\{X_{N_r}\}$ and $\{Y_{N_r}\}$ $N_r = t, 2t, 3t, \ldots$ each satisfies the conditions' of Noether and $\rho_{N_r} \dashrightarrow \rho$ where ρ is bounded way from 1 then for N_r sufficiently large

$$S_{20N_r} \left\{ \frac{R_{xyN_r}}{R_{xxN_r}} - \frac{S_{11N_r}}{S_{20N_r}} \right\} \simeq S_{20N_r} \operatorname{Var}(\hat{\beta})$$

$$\simeq S_{02N_r} (1-\rho_{N_r}^{\ 2}) \frac{(t-1)}{N_r} \cdot \frac{1}{N_r} \ .$$

And so

$$\Sigma_k r \hat{\tau}_{kN_r}^{\ 2} = r\Sigma_k \{(y_{kN_r} - y_{.N_r}) - \hat{\beta}(x_{kN_r} - x_{.N_r})\}^2$$

$$\approx \text{adj (TSS)}(1 + \frac{1}{N_r}) = A_{yyN_r} (1 + \frac{1}{N_r}) \ .$$

$$\frac{R_{xyN_r}^{\ \ 2}}{R_{xxN_r}} = \frac{(N-t)^2}{(N_r-1)^2} S_{11N_r}^{\ \ 2} (1+2\eta+\eta^2) \Big/ \frac{(N-t)}{(N-1)} S_{20N_r}(1+\theta) \ .$$

So

$$\frac{R_{xyN_r}^{\ \ 2}}{R_{xxN_r}} \simeq \frac{(N_r-t)}{(N_r-1)} \cdot \frac{S_{11N_r}^{\ \ 2}}{S_{20N_r}} \{1+2\eta+\eta^2-\theta-2\eta\theta+\theta^2\} \quad ,$$

and

$$E_{yyN_r} = R_{yyN_r} - \frac{R_{xyN_r}^{\ \ 2}}{R_{xxN_r}}$$

$$= \frac{N_r-t}{N_r-1} S_{02N_r}(1+\gamma) - \frac{N_r-t}{N_r-1} \frac{S_{11N_r}^{\ \ 2}}{S_{20N_r}} (1 + 2\eta + \eta^2 - \theta - 2\eta\theta + \theta^2).$$

But we have shown previously that for large N_r if the sequences $\{X_{N_r}\}$ and $\{Y_{N_r}\}$, $N_r=1,2,\ldots$ each satisfy the Noether conditions N, with $|\rho_{N_r}|$ bounded away from 1 then $\operatorname{Var}(\eta)$, $\operatorname{Var}(\theta)$, $\operatorname{Cov}(\eta,\theta)$ are all

$$O\left(\frac{2(t-1)}{N_r^{\ 2}}\right) \ .$$

So $$E_R(E_{yyN_r}) \simeq S_{02N_r} \frac{S_{11N_r}^{\;2}}{S_{20N_r}} (1 + \mathrm{Var}(\eta) + \mathrm{Var}(\theta) - 2\,\mathrm{Cov}(\eta,\theta)) \frac{N_r - t}{N_r - 1}$$

$$\simeq \frac{(N_r - t)}{(N_r - 1)} S_{02N} \left[1 - \rho_{N_r}^{\;2} - \rho_{N_r}^{\;2} \frac{(1-\rho_{N_r}^{\;2})}{\rho_{N_r}^{\;2}} \frac{(t-1)}{N_r^{\;2}} \right]$$

Hence $$E_R(E_{yyN_r}) \simeq \frac{N_r - t}{N_r - 1} \left(S_{02N_r} - \frac{S_{11N_r}^{\;2}}{S_{20N_r}}\right)$$

and

$$E_R(A_{yyN_r}) = E_R\{T_{yyN_r} + \frac{G_{xyN_r}^2}{G_{xxN_r}} - \frac{R_{xyN}^2}{R_{xxN_r}} \simeq \frac{(t-1)}{N-1} \left(S_{02N_r} - \frac{S_{11N_r}^2}{S_{20N_r}}\right).$$

The number of degrees of freedom for treatments is t-1 and for error is N_r-t-1. Let W_r = adjusted treatment mean square-adjusted error mean square. Then

$$W_r = \frac{(S_{yyN_r} - S_{xyN_r}^2 / S_{xxN_r})}{t-1} - \frac{(R_{yyN_r} - R_{xyN_r}^2 / R_{xxN_r})}{t-1}$$

$$\simeq \frac{(R_{yyN_r} - R_{xyN_r}^2 / R_{xxN_r})}{N_r - t-1} \quad .$$

So under general Noether conditions we have

$$E_R(W) = \frac{S_{yyN_r}}{(t-1)} [1 - \rho_{N_r}^2 - \frac{(N_r - 2)}{(N_r - t-1)} \frac{(N_r - t)}{(N_r - 1)} \{1 - \rho_{N_r}^2 - \rho_{N_r}^2 (\mathrm{Var}(\eta) +$$

$$\mathrm{Var}(\theta) - 2\,\mathrm{cov}(\eta,\theta) + o(N_r^{-2}) \}]$$

$$= - S_{yyN_r}\{(1 - \rho_{N_r}^2)(t-1) - (N_r - 2)(N_r - t)\, \rho_{N_r}^2 \quad (\mathrm{Var}(\theta) + \mathrm{Var}(\ )$$

$$- 2\,\mathrm{cov}(\eta,\theta) + o(N^{-2})\} / (t-1)(N_r-t-1)(N-1).$$

The bias W_r, is approximately equal to

$$\left\{ \frac{2S_{31N_r}}{S_{20N_r}S_{11N_r}} - \frac{S_{22N_r}}{S^2_{11N_r}} - \frac{S_{40N_r}}{S^2_{20N_r}} \right\} \quad \frac{S_{02N_r}}{N_r^2}$$

In order to get some idea of the sign of this bias we would need to know the behavior of the $\{S_{\cdots}\}$ as well as the magnitude of ρ_{N_r}. Also the contents of $O(N_r^{-2})$ could possibly determine the sign of the bias. Nevertheless, if the components of the population sequences satisfy the Noether type conditions, the expectation of W would be at most of order $O(N_r^{-2})$. For samples from a normal population, say, we surmise that when ρ_{N_r} is small in magnitude the expression for the bias is negative and is approximately equal to $S_{yyN_r}(1-\rho^2_{N_r})/N_r^3$, and when ρ_{N_r} is close to 1 the bias is still likely to be negative as it approaches zero from above. Although we cannot make any general statement about the sign of the bias, it is clear that under Noether type conditions the expected mean squares for the adjusted treatment sum of squares and the adjusted error sum of squares have the same asymptotic value.

Also

$$E_R\left(\frac{(A_{yyN_r})}{A_{yyN_r} + E_{yyN_r}}\right) \simeq \frac{(t-1)}{N_r-1} \quad .$$

Denote $\dfrac{S_{11N_r}}{\sqrt{(S_{20N_r}S_{02N_r})}}$ by ρ_{N_r}.

Then $E_{yyN_r} = \frac{N_r-t}{N_r-1} S_{02N_r} (1 + \gamma - \rho_{N_r}^2 (1 + 2\eta + \eta^2 - \theta - 2\eta\theta + \theta^2))$

and since $A_{yyN_r} + E_{yyN_r} = \frac{N_{r-1}}{N_r-1} S_{02N_r}(1 - \rho_{N_r}^2) = \text{constant},$

then $\quad \underset{R}{\mathrm{Var}} (A_{yyN_r}) = \underset{R}{\mathrm{Var}} (E_{yyN_r})$.

So

$$\underset{R}{\mathrm{Var}} \left(\frac{A_{yyN_r}}{A_{yyN_r} + E_{yyN_r}}\right) \simeq \frac{1}{(1 - \rho_{N_r}^2)^2} \mathrm{Var}\,(\gamma - 2\rho_{N_r}^2 \eta + \rho_{N_r}^2 \theta)$$

The numerator of the right hand side is

$$\mathrm{Var}\,(\gamma) + 4\rho_{N_r}^4 \mathrm{var}\,(\eta) + \rho_{N_r}^4 \mathrm{Var}(\theta)$$

$$-4 {}_{N_r}^2 \mathrm{cov}\,(\gamma,\eta) + 2\rho_{N_r}^2 \mathrm{cov}\,(\theta,\gamma) - 4\rho_{N_r}^4 \mathrm{cov}\,(\theta,\eta)$$

Approximations to all the variances and covariances above are given on page 78 in terms of $\frac{2(t-1)}{N_r^2}$ and ρ_{N_r} . Substituting these approximations above we have

$$\underset{R}{\mathrm{Var}}\left(\frac{A_{yyN_r}}{A_{yyN_r} + E_{yyN_r}}\right) \simeq \frac{(1 - 2\rho_{N_r}^2 + \rho_{N_r}^4)}{(1 - \rho_{N_r}^2)^2} \cdot \frac{2(t-1)}{N_r^2} \simeq \frac{2(t-1)}{N_r^2}$$

So we see that for large values of N_r the mean value and variance of the ratio $\frac{A_{yyN_r}}{(A_{yyN_r} + E_{yyN_r})}$ under the randomization model are approximately the same as those obtained under Gauss-Markov Normal assumptions, the bias being of order $O(N_r^{-2})$.

For completeness, we now derive approximations to the mean and variance, under randomization, of the ratio, F_r, of mean squares for treatments and error, adjusted for a single concomitant (p=1). That is, we have

$$F_r = \frac{A_{yyN_r}(N_r-t-p)}{E_{yyN_r}(t-1)} \quad .$$

Write

$$A_{yyN_r} = E_R(A_{yyN_r})\left[1 + \frac{A_{yyN_r} - E_R(_{yyN_r})}{E(A_{yyN_r})}\right] = \frac{(t-1)}{(N-1)} S_{yyN_r}(1-\rho^2_{N_r})(1+e),$$

where e is equal to $[A_{yyN_r} - E_R(A_{yyN_r})]/E_R(A_{yyN_r})$.

Then

$$E_{yyN_r} = S_{yyN_r}(1-\rho^2_{N_r}) - A_{yyN_r} = S_{yyN_r}(1-\rho^2_{N_r})\left[1-\frac{(t-1)}{(N-1)}(1+e)\right].$$

So,

$$F_r = \frac{(N_r-t-p)}{(N_r-t)}(1+e)\left[1-\frac{(t-1)}{(N_r-t)}e\right]^{-1} \simeq 1$$

Assuming Noether type conditions on the basic vector units, we obtain, by the Taylor-MaClaurin expansion, for p = 1, that

$$F_r \simeq \left[1 + \frac{(N_r-1)}{(N_r-t)}e + \frac{(t-1)(N_r-1)}{(N_r-t)(N_r-t)}e^2\right]\frac{(N_r-t-p)}{(N_r-t)} \quad .$$

Now, from the preceding paragraph, we have $E_R(e) = 0$, and for N_r large,

$$E_R(e^2) \simeq \frac{2}{(t-1)} \; .$$

So,

$$E_R(F_r) \simeq \frac{(N_r-t-p)}{(N_r-t)}\left[1 + \frac{2}{(N_r-t)}\right]$$

and

$$\mathrm{Var}_R(F_r) \simeq \frac{2(N_r-t-p)^2(N_r-1)^2}{(N_r-t)^4(t-1)} \simeq \frac{2}{(t-1)} \quad .$$

Also, from the corollary to theorem A1.1 of Appendix A, $(t-1)F_r$ converges in probability to the same limiting distribution as $[A_{yyN_r}(N_r-t-p)/(A_{yyN_r} + E_{yyN_r})]$.

3.4 The Analysis of Multiple Covariance

In this case, omitting the subscript N_r the reduced normal equations for $\hat{\beta}$ may be written in the form

$$R\hat{\beta} = R_y$$

where

$$R = \begin{pmatrix} R_{11}, & R_{12}, & \ldots, & R_{1p} \\ R_{21}, & R_{22}, & \ldots, & R_{2p} \\ \vdots & \vdots & & \vdots \\ R_{p1}, & R_{p2}, & \ldots, & R_{pp} \end{pmatrix} \quad \hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ B_2 \\ \vdots \\ \beta_p \end{pmatrix}, \quad R_y = \begin{pmatrix} R_{1y} \\ R_{2y} \\ \vdots \\ R_{py} \end{pmatrix}$$

The $\{R_{ij}\}$ are the set of Residual sum of squares and cross products obtained from the simple analysis of covariance for the columns i, and j of X, and the $\{R_{iy}\}$ are similarly defined for the columns i of X and the column vector Y. Most of the statistical estimates and tests that we are investigating depend directly on the $\{R_{ij}\}$. So let us now inquire into some of their joint properties.

Now $$E_{\pi_1}(R_{ij}) = \frac{N-t}{N-1}(S_{ij}) = (S_{ij}^{*}) \quad , \quad \text{say}$$

where (S_{ij}) is the square matrix of dimensions pxp and the element $S_{ij} = \sum_{k=1}^{N} x_{ki}x_{kj}$. Let Δ_{ij} denote the difference $R_{ij} - E_{\pi_1}\{R_{ij}\}$ and $\Delta = (\Delta_{ij})$ the pxp matrix with elements Δ_{ij}. Then from theorem B.5 of Appendix B if $\rho(\Delta)$ is the norm of the matrix Δ we have

$$\rho(\Delta) \leq \max_{1 \leq i \leq p} \sum_{j=1}^{p} |\Delta_{ij}|$$

For any $\varepsilon > 0$

$$f \sum_{j=1}^{p} |\Delta_{ij}| > \ , \text{ then } \max_{1 \le j \le p} |\Delta_{ij}| > \frac{\varepsilon}{p} .$$

But from Tchebychev's Inequality,

$$p(|\Delta_{ij}| > \frac{\varepsilon}{p}) \le \text{Var}(\Delta_{ij}) \frac{p^2}{\varepsilon^2} .$$

Hence

$$P((\Delta) > \varepsilon) \le P\{\max_{1 \le i \le p} \sum_{j=1}^{p} \Delta_{ij} > \varepsilon\}$$

$$\le P\{\max_{1 \le i \le p} \max_{1 \le j \le p} \Delta_{ij} > \frac{\varepsilon}{p}\}$$

$$\le \{p^2 \max_{1 \le i \le p} \max_{1 \le i \le p} \text{Var}(\Delta_{ij})\} \cdot \frac{p^2}{\varepsilon^2} .$$

If this last expression is small we may approximate small inverse powers of Δ by a series expansion to order $O(N^{-2})$. (See theorems A.4 and B.6).

Let $S_{x_1^p x_2^q x_3^r x_4^s}$ denote the product sum $\sum_{i=1}^{N} x_{i1}^p x_{i2}^q x_{i3}^r x_{i4}^s$. We previously obtained the following results:

(1) $$\underset{\pi_1}{\text{Var}}(R_{X_1X_1}) = A(S_{X_1^2})^2 + C\, S_{X_1^4} \text{ , where}$$

$$A = \frac{2(t-1)(r-1)(N^2-3N+3)}{r(N-1)^2(N-2)(N-3)} ,$$

and

$$C = \frac{-2(t-1)(r-1)N}{(N-1)(N-2)(N-3)}$$

(2) $$\underset{\pi_1}{\text{Var}}(R_{X_1X_2}) = A\, S_{X_1^2} . S_{X_2^2} + B(S_{X_1X_2})^2 + C\, S_{X_1^2 X_2^2} ,$$

where

$$A = \frac{(t-1)(r-1)(N-2)}{r(N-1)(N-2)(N-3)} ,$$

$$B = \frac{(t-1)(r-1)(N^2-3N+4)}{r(N-1)^2(N-2)(N-3)} ,$$

and

$$C = \frac{-2(t-1)(r-1)(N)}{r(N-1)(N-2)(N-3)}$$

(3) $$\underset{\pi_1}{\mathrm{Cov}}(R_{X_1X_1}, R_{X_2X_2}) = A\ S_{X_1^2}\ S_{X_2^2} + B\ S_{X_1X_2}^{\ 2} + C\ S_{X_1^2X_2^2} ,$$

where

$$A = \frac{2(t-1)(r-1)}{r(N-1)^2(N-2)(N-3)}$$

$$B = \frac{2(t-1)(r-1)}{r(N-1)(N-3)}$$

and

$$C = \frac{-2(t-1)(r-1)N}{r(N-1)(N-2)(N-3)}$$

(4) $$\underset{\pi_1}{\mathrm{Cov}}(R_{X_1X_1}, R_{X_1X_2}) = A\ S_{X_1^2}\ S_{X_1X_2} + C\ S_{X_1^3\ X_2} ,$$

$$A = \frac{2(t-1)(r-1)(N^2-3N+3)}{r(N-1)^2(N-2)(N-3)}$$

and

$$C = \frac{-2(t-1)(r-1)N}{(N-1)(N-2)(N-3)} .$$

We will appeal to the symmetrical nature of the expectations under randomization of products, {variances and covariances}, of the reduced sums of squares and cross products to obtain their values in terms of the {S...}.

(5) $$\underset{\pi_1}{\mathrm{Cov}}\ (R_{X_1X_2}, R_{X_1X_3}) = A\ S_{X_1^2} \cdot S_{X_2X_3} + B\ S_{X_1X_2} \cdot S_{X_1X_3} + C\ S_{X_1^2\ X_2X_3}$$

When $X_2 = X_3$, we have

$$\underset{\pi_1}{\mathrm{Var}}\ (R_{X_1X_2}) = A\ S_{X_1^2} S_{X_2^2} + B\ S_{X_1\ X_2^2} + C\ S_{X_1^2\ X_2^2} .$$

Hence

$$A = \frac{(t-1)(r-1)}{r(N-1)(N-3)} ,$$

$$B = \frac{(t-1)(r-1)(N^2-3N+4)}{r(N-1)^2(N-2)(N-3)} ,$$

and

$$C = \frac{-2(t-1)(r-1)N}{r(N-1)(N-2)(N-3)}$$

(6) $$\underset{\pi_1}{\mathrm{Cov}}\,(R_{X_1X_1},R_{X_2X_3}) = A\, S_{X_1^2} \cdot S_{X_2X_3} + B\, S_{X_1X_2}S_{X_1X_3} + C\, S_{X_1^2\,X_2X_3}$$

When $X_3 = X_2$, we have

$$\underset{\pi_1}{\mathrm{Cov}}\,(R_{X_1X_1},R_{X_2X_2}) = A\, S_{X_1^2} \cdot S_{X_2^2} + B\, S^2_{X_1X_2} + C\, S_{X_1^2\,X_2^2}$$

Hence

$$A = \frac{2(t-1)(r-1)}{r(N-1)^2(N-2)(N-3)} ,$$

$$B = \frac{2(t-1)(r-1)}{r(N-1)(N-3)} ,$$

and

$$C = \frac{-2(t-1)(r-1)N}{(N-1)(N-2)(N-3)}$$

(7) $$\underset{\pi_1}{\mathrm{Cov}}\,(R_{X_1X_2},R_{X_3X_4}) = A\, S_{X_1X_2}S_{X_3X_4} + B\{S_{X_1X_3}S_{X_2X_4} + S_{X_1X_4}S_{X_2X_3}\}$$
$$+ C\, S_{X_1X_2X_3X_4}$$

When $X_4 = X_1$, we have

$$\underset{\pi_1}{\mathrm{Cov}}\,(R_{X_1X_2}R_{X_1X_3}) = B\, S_{X_1^2}\, S_{X_2X_3} + (A+B)\, S_{X_1X_2}S_{X_1X_3} + C\, S_{X_1^2\,X_2X_3}$$

Hence

$$B = \frac{(t-1)(r-1)}{r(N-1)(N-3)} ,$$

$$(A + B) = \frac{(t-1)(r-1)(N^2-3N+4)}{r(N-1)^2(N-2)(N-3)},$$

i.e.
$$A = \frac{2(t-1)(r-1)}{r(N-1)^2(N-2)(N-3)}$$

and
$$C = \frac{-2(t-1)(r-1)N}{r(N-1)(N-2)(N-3)}.$$

Suppose we have a sequence of populations $\varphi_{N_r} = (X_{1N_r}, \ldots, X_{N_rN_r})$ $N_r = rt$, $r=1,2,\ldots$, where the X's are p+1-variates and t is a fixed integer. For ease of writing, where there is no ambiguity we will drop the N_r and X's in the $\{S_{X_{iN_r}X_{jN_r}}\}$ and $\{R_{X_{iN_r}X_{jN_r}}\}$ so that S_{ijN_r} denotes $S_{X_{iN_r}X_{jN_r}}$, We shall use S_{N_r} to denote the matrix of sum of squares and products of the concomitant observations and $S_{.yN_r}$ for the vector of products of the concomitant vectors and the vector of observed responses Y_{N_r}.

R_{N_r} and $R_{.yN_r}$ are similarly defined for residuals.

Suppose each component X_{iN_r} satisfies the Noether conditions and the maximum value of the correlation coefficients between the components is bounded away from one. Let $\rho_{N_r}(\Delta)$ be the norm of $\Delta_{N_r} = (\Delta_{ijN_r})$ where $\Delta_{ijN_r} = R_{ijN_r} - E_R(R_{ijN_r}) = R_{ijN_r} - S_{ijN_r}^{*}$, then from theorem A 2.5

of Appendix B we have for any $\epsilon > 0$ that there exists an N such that for

$N_r > N_\epsilon$

$$P(\rho_{N_r}(\Delta) > \epsilon) \leq p^2 \underset{1\leq i\leq p}{\text{Max}} \; \underset{i\leq 1\leq p}{\text{Max}} \; \text{Var}\,(\Delta_{ijN_r}) \cdot \frac{p^2}{\epsilon^2} \; .$$

$$\simeq \frac{p^4}{\epsilon^2} \frac{2(t-1)}{N_r^{\,2}}$$

$$\to 0 \quad \text{as} \quad N_r \to \infty$$

Hence from theorem A 2.6 we may expand small powers of R_{N_r} in terms of

$S_{N_r}^*$ and Δ_{N_r} .

The reduced normal equations for $\hat{\beta}_{N_r}$ are given by

$$(R_{N_r})\hat{\beta}_{N_r} = R_{.yN_r} \quad .,$$

Hence $\hat{\beta}_{N_r} = R_{N_r}^{-1} R_{.yN_r} = (S_{N_r}^* + \Delta_{N_r})^{-1} (S^*_{.yN_r} + \Delta_{.yN_r})$.

i.e., omitting the subscript N_r we have

$$\hat{\beta} \simeq \{S^{*-1} - S^{*-1}\Delta S^{*-1} + S^{*-1}\Delta S^{*-1}\Delta S^{*-1}\} \{ S^*_{.y} + \Delta_{.y} \}$$

$$\simeq S^{*-1}S^*_{.y} + S^{*-1}\Delta_{.y} - S^{*-1}\Delta S^{*-1}S^*_{.y} - S^{*-1}\Delta S^{*-1}\Delta_{.y}$$

$$+ S^{*-1} S^{*-1} S^{*-1} S^*_{.y} \quad .$$

Now $S^{*-1}S^*_{.y} = S^{-1}S_{.y} = \{\beta_1, \beta_2, \ldots, \beta_p\}' = \beta$, say.

Hence $E(\hat{\beta})$

$$\simeq \beta - E\{S^{*-1}\Delta S^{*-1}\Delta_{.y}\} + E\{S^{*-1}\Delta S^{*-1}\Delta\beta\} \quad .$$

Let (y_k, x_k) denote the mean vector for treatment group k and $(y_. \; x_.)$ denote the mean vector over all groups. Then the estimates of treatment effects are given by the equation $\hat{\tau}_k = (\tau_k \hat{-} \tau_.) = y_k - y_. - \hat{\beta}(x_k - x_.)$ $= y_k^* - \hat{\beta}x_k^*$, say. With T and R as defined for the analysis of multiple covariance i.e., with

$$T + R = S. \quad , \quad T_{.y} + R_{.y} = S_{.y} \quad , \text{ and } T_{yy} + R_{yy} = S_{yy}$$

we have

$$r \sum_{k=1}^{t} \tau_k^2 = T_{yy} - 2\hat{\beta}' T_{.y} + \hat{\beta}' T\hat{\beta}$$

$$= T_{yy} - 2\hat{\beta}'(S_{.y} - R_{.y}) + \hat{\beta}'(S - R)\hat{\beta}$$

$$= \{T_{yy} + \hat{\beta}' R_{.y} - \beta S_{.y}\} + \{\hat{\beta} S \hat{\beta} - 2\hat{\beta}' S\beta + \beta' S\beta\}$$

The first expression in the right hand side of the last inequality is the adjusted sum of squares for treatment effects. The second expression is equal to

$$(\hat{\beta} - \beta)' S(\hat{\beta} - \beta)$$

Now let $W = (W_{ij}) = S^{*-1}$ and $\theta_{.y} = (\theta_{1y}, \theta_{2y}, \ldots, \theta_{py}) = \Delta S^{*-1} \Delta_{.y}$

Then $\underset{R}{E}(\beta_i) \simeq \beta_i - \Sigma_j s^{ij} E(\theta_{jy}) + s^{ii} \beta_i E(W_{ii}) + \sum_{j\, i}^{\neq} s^{ij} \beta_j E(W_{jj})$

$$+ \Sigma_{jk}^{\neq} s^{jj} \beta_k E(W_{jk})$$

$$= \beta_i + o(N^{-2}) \qquad \text{, by theorem 3.3 which follows.}$$

$$\text{Cov}(\hat{\beta}) \simeq \text{Cov}\{S^{-1}\Delta_{.y} - S^{-1}\Delta\beta\}$$

$$= S^{-1}\text{Cov}(\Delta_{.y} - \Delta\beta) \cdot S^{-1}$$

The variances and covariances under randomization theory of the "residual" sum of products $\{R_{ij}\}$ for the various components in the analysis of multiple covariance are derived in Appendix C. We now give the limiting values of the general result in the following theorem.

Theorem 3.3

Suppose we have a sequence of finite population $A_1, A_2, \ldots, A_r \ldots$ such that A_r contains $N_r = rxt$ elements $\{U_{ri}\}, i=1,2,\ldots,N_r$, and suppose the elements are p-component vectors

$$U_{ri} = (U_{ri1}, U_{ri2}, \ldots, U_{rip}) \qquad i=1,2, \ldots, N_r.$$

Assume the general conditions Noether are satisfied by each component sequence $A_{1j}, A_{2j}, \ldots,$ $j=1,\ldots p$ and the maximum coefficient of correlation between the components is bounded away from 1. Then with R_{ij} and S_{ij} as previously defined on A_r, for the analysis of (multiple) covariance, under randomization theory we have the following expectation.

$$\underset{R}{E}\,(R_{ijN_r}\, R_{i'j'N_r}) = (S_{ii'N_r} S_{jj'N_r} + S_{ij'N_r} S_{i'jN_r}) \cdot \left\{ \frac{(t-1)}{N_r^{\,2}} + o(N_r^{-2}) \right\}$$

The conditions ν say , essentially, that each power sum $S_{(\ldots.)}$ of degree $p \geq 3$ has magnitude $o(N^{-1+p/2})$ and for degree $p = 2$ the power sums are of order $O(N)$. The proof of the theorem follows by considering the order of magnitude of each term in the algebraic expressions for the covariances of the R_{ij}'s given in Appendix C.

Let $C_{N_r} = (C_{kk'_{N_r}})$ = the variance matrix of $(\Delta_{.yN_r} - \Delta_{N_r}\beta_{N_r})$.

Dropping the N_r, we have

$$C_{kk'} = \mathrm{cov}(R_{ky} - \Sigma_j R_{kj}\beta_j \, , \, R_{k'y} - \Sigma_j R_{kj}\beta_j)$$

$$= E(R_{ky}R_{k'y}) - \Sigma_j E(R_{ky}R_{kj} + R_{k'y}R_{kj})\beta_j + \Sigma_{jj'} E(R_{kj}R_{kj'})\beta_j\beta_{j'}.$$

Using theorem 3.3, we obtain

$$\frac{N^2}{(t-1)} C_{kk'} = S_{yy}S_{kk'} + S_{ky} \cdot S_{k'y} - \Sigma_j(S_{kk'}S_{jy} + S_{k'y}S_{kj} + S_{kk'}S_{jy}$$
$$+ S_{ky}S_{k'j})\beta_j + \Sigma_{jj'}(S_{kk'}S_{jj'} + S_{kj'}S_{jk'})\beta_j\beta_{j'} \quad .$$

Now writing

$$S = (S_{kk'}) \quad ,$$

we have

$$S\beta = S_{.y} \quad ,$$

or

$$\Sigma_j S_{kj}\beta_j = S_{ky} \quad ,$$

and

$$\beta' S_{.y} = \Sigma_j \beta_j S_{jy} = S_{yy}\rho^2 \quad .$$

where ρ^2 is the square of the multiple correlation coefficient of the regression of the "independent" components on y the dependent component of the population A. Substituting these values in the expression for $C_{kk'}$

we will have the approximate result;

$$\frac{N^2}{(t-1)} C_{kk'} = S_{yy}S_{kk'} + S_{ky} \cdot S_{k'y} - 2(S_{kk'}S_{yy}\rho^2 + 2\, S_{ky}S_{k'y})$$
$$+ S_{kk'}\beta' S_{.y} + S_{ky}.S_{k'y}$$
$$= S_{kk'}S_{yy}(1-\rho^2)$$

Hence $C = 2S(1-\rho^2)(S_{yy/N})(\frac{(t-1)}{N} + o(N^{-1}))$

So formally given the conditions of theorem 3.3 we have the following corollary for the estimated regression coefficient $\hat{\beta}_{N_r} = (R_{N_r}^{-1} R_{yN_r})$

Corollary:

Let $R_{.yN_r}$ and R_{N_r} denote the "residual" sum of cross products with the "dependent" component and among the "independent" components respectively in the conditional analysis of multiple covariance on the sequence A_{N_r}.

Then

$$\mathrm{Cov}(\hat{\beta}_{N_r}) = \mathrm{Cov}(R_{N_r}^{-1} R_{yN_r})$$

$$= S_{N_r}^{-1} (1-\rho_{N_r}^{2}) \frac{S_{yyN_r}}{N_r} \{(t-1)/N + o(N^{-2})\}$$

As in the case of a single covariate, this result differs from its GMN counterpart by a factor f, where $f = \frac{(t-1)}{N}$.

3.5 Normal Law Approximations to the Limiting Distributions under Randomization Theory

In this section we present (normal law) approximations to the limiting distributions of some GMN statistics under randomization theory for various designs. However, we first present two theorems which are directly related,

Theorem 3.4 (Anderson-Cramer)

Let $U_{(N)}$ be an m-component random vector and b a fixed vector. Assume $\sqrt{N}\,(U_{(N)} - b)$ is asymptotically distributed according to $N(0,T)$. Let $W = f(u)$ be a function of a vector U with first and second derivatives existing in a neighborhood of $u = b$.

Let φ_b be the vector of partial derivatives, $\left.\frac{\partial f(u)}{\partial u_i}\right|_{u=b}$.

Then the limiting distribution of $\sqrt{N}[f(U_{(N)}) - f(b)]$ is

$$N(0, \varphi_b' T\varphi_b)$$

Proof: (See Anderson (1958) and Cramer (1946).)

Theorem 3.5 (Aroian)

Suppose x, y are bivariate normal with means μ_x, μ_y, variances σ_x^2, σ_y^2 respectively and coefficient of correlation ρ. Let $c_x = \mu_x/\sigma_x$, $c_y = \mu_y/\sigma_y$ and $z = xy/\sigma_x\sigma_y$. The distribution of z approaches normality with mean μ_z and variance σ_z^2 where $\mu_z = c_x c_y + \rho$, $\sigma_z^2 = c_x^2 + c_y^2 + 2\rho c_x c_y + 1 + \rho^2$ if either of the two conditions are met.

1) $c_x \longrightarrow \infty$, $c_y \longrightarrow \infty$. $-1 + \epsilon < \rho \leq 1$, $\epsilon > 0$.

2) $c_x \longrightarrow \infty$, $c_y \longrightarrow \infty$, $-1 \leq \rho < 1 - \epsilon$, $\epsilon > 0$.

Proof: (See Aroian (1947).)

Aroian uses the convergence of the moment generating function (mgf) of z to obtain the result. We note that mere convergence of a sequence of distribution functions says little about the behavior of the corresponding sequences of mgfs. However, if a mgf exists in some neighborhood of the origin it uniquely determines the corresponding distribution. For a detailed discussion of this see Curtiss (1942) and Rao (1965).

We now apply these theorems to the completely randomized design.

Define sequences of vectors C_r with t components C_{kr} $(k=1,2, \ldots, t)$, i.e. $C_r = (C_{1r}, C_{2r} \ldots C_{tr})$. Each component sequence C_{kr} is comprised of N_r (=rt) elements $\{C_{ikr}, 1 \leq i \leq N_r = rt\}$, $r=1,2, \ldots$ such that

$$C_{ikr} = \begin{cases} c_r, & 1 + (k-1)r \leq i \leq kr \\ 0, & \text{otherwise} \end{cases},$$

where c_r is a constant depending on r.

The correspondence of the vector sequences with the different treatment groups in the CRD is quite obvious. Let the orthogonal t x t matrix $Q' = (Q_0', Q_1', \ldots, Q_{t-1}')$ have elements in the first column Q_0 each equal to $t^{-\frac{1}{2}}$.

Let $C_r^* = C_r Q = (\Lambda_{0r}, \Lambda_{1r}, \ldots, \Lambda_{\overline{t-1},r})$ and

let $\Lambda_r = (\Lambda_{0r}, \Lambda_{1r}, \ldots, \Lambda_{N_r r})$ be such that $(N_r^{-\frac{1}{2}} \Lambda_r)$ is orthogonal, the elements of Λ_0 each being equal to 1.

Let Z_r be a sequence of (p+1)-vectors $Z_r = (Z_{0r}, Z_{1r}, \ldots, Z_{pr})$ with N_r vector elements $\{z_{ir}, 1 \leq i \leq N_r\}$ $r=1,2,\ldots$

where $z_{ir} = (z_{i0r}, z_{i1r}, \ldots, z_{ipr})$.

Let X_r be a random variable which takes on the N! different permutations of Z_r with equal probability.

The following results are easily obtained:

a) $C_r^* C_r^{*'} = C_r C_r'$

b) $X_r' \Lambda_{0r} = \sum_{i=1} z_{ir}'$ = constant with probability 1.

c) $\Lambda_{or}'\Lambda_{kr} = \sum_{i=1}^{rt} \lambda_{ikr} = 0$

d) $E_R(X_r' \; _{kr}) = 0$

e) $\mathrm{Var}\,(N^{-\frac{1}{2}} \; _{kr}'X_r) = \Sigma_{Zr}\;, 1 \leq k \leq N_r - 1$

f) $\mathrm{Cov}\,(\Lambda_{kr}'X_r, \Lambda_{k'r}X_r) = 0 \qquad i \leq k,\; k' \leq N_r - 1,\; k \neq k'$

g) $$(X'X) = (X_r'\Lambda_r\Lambda_r\;'X_r) = (X_r'\Lambda_{or}\Lambda_{or}'X_r) + \sum_{k=1}^{t-1} (X_r'\Lambda_{kr})(\Lambda_{kr}'X_r) + \sum_{k=t}^{rt-1} (X_r'\Lambda_{kr})(\Lambda_{kr}'X_r)$$

h) $$(X_r'C_r)(C_rX) = (X_r'C^*C^{*}{}'X_r) = (X_r'\Lambda_{or}\Lambda_{or}'X_r) + \sum_{k=1}^{t-1}(X_r'\Lambda_{kr})(\Lambda_{kr}'X_r)$$

If the covariance matrix of Σ_{zr} of Z_r converges in the limit to to Σ_z, positive definite, then with C_r, C_r^*, , Λ_r and X_r as defined we have the following results for the analysis of covariance in the CRD.

Theorem 3.6

If the vector sequence Z_r satifies the conditions 2.3.3.6 - 2.3.3.8 then the limiting distribution of $\dfrac{(C_{kr}'\,X_r - EC_{kr}'EX_r)}{c_r t^{\frac{1}{2}}}$, $k=1,2, \ldots, t$, is multivariate normal with mean zero and covariance matrix Σ_Z.

Proof:

Without loss of generality we may assume $\Sigma_i^{rt} z_{ijr} = 0$. $j=1,2,\ldots p$ and $\Sigma_i^{rt}\, z_{ijr}{}^2 = N_r = rt$ etc. by normalizing the vector components of Z_r C_r. So the sequences C_r and Z_r satisfy conditions 2.3.3.6 - 2.3.3.8 Hence the result follows.

Similarly, the limiting distribution of $N^{-\frac{1}{2}}(\Lambda_{kr}'X_r)$ $1 \leq k \leq t-1$, is multivariate normal with mean zero and variance covariance matrix Σ_A.

Theorem 3.7

$\dfrac{C_r^{*'}X_r}{c_r\, t^{\frac{1}{2}}}$ is asymptotically distributed as t-1 independent normal variables with zero mean and variance covariance matrix Σ_Z and a random variable with constant value equal to zero with probability 1.

Corollary 1:

The asymptotic distribution of $A_r = \dfrac{(X_r'C_r)(C_r'X_r)}{c_r^2 t}$ is Wishart $(t-1, \Sigma_z)$.

Corollary 2:

The asymptotic distribution of $R_r = X_r'X_r - A_r$ is Wishart $(N-t, \Sigma_z)$.

Theorem 3.8

$N_r^{-\frac{1}{2}} R_r$ is asymptotically multivariate normally distributed.

Proof:

R_r is the sum of (N_r-t) random variables each having identical limiting independent Wishart distribution. (Independent by virtue of normality, and zero covariance). Hence R_r satisfies the conditions of the Lindeberg central limit theorem.

Let $(Y_r, X_{1r}, \ldots, X_{pr})$ be the components of X_r.

Let

$$R_r = \begin{vmatrix} R_{yyr} & R_{yxr} \\ R_{xyr} & R_{xxr} \end{vmatrix}$$

where R_{yyr} is a scalar, R_{yxr} is a (1xp) vector.

and R_{xx} is p x p matrix etc,

Let $\hat{\beta}_r = R_{xxr}^{-1} R_{xyr}$

Then by theorem 3.4 (Anderson and Cramer) $N^{-\frac{1}{2}}\hat{\beta}_r$ has limiting asymptotic multivariate normal distribution. The expressions A_r and R_r discussed here can be shown to be the same as the expressions for treatments and residuals respectively obtained from the partitioning of the matrix of sums of squares and cross products in the analysis of covariance table previously discussed, i.e. $T_r + R_r = S_r$. Any linear combinations of β_r and and any linear combination of the mean vector for treatment group k can be shown to satisfy the conditions of theorem 3.5 (Aroian). Moreover, the vector $\hat{\beta}_r$ can be shown to have zero correlation in the limit with the set of mean vectors for the different treatment groups. So that the (adjusted) estimates of treatment effects

$\hat{\tau}_{kr} = \bar{y}_{kr} - \bar{X}_{kr}\hat{\beta}_r$ k=1,2, ..., t, will have asymptotic normal distribution. Hence the treatment contrasts, $\Sigma\nu_k \tau_k$, $\Sigma\nu_k = 0$ will have unbiased estimates $\sum_{k=1}^{t} \nu_k\hat{\tau}_{kr}$ with limiting normal distribution having mean value $\sum_{k=1}^{t} \nu_k\tau_k$ and variance equal to $\dfrac{\sigma^2(1-\rho^2)\, t\, \Sigma\nu_k^{\,2}}{r(t-1)}$ where $\sigma_r^{\,2} = \dfrac{S_{yyr}}{N_r-1}$

$\rho_r^{\,2} = S_{xyr}'\, S_{xxr}^{-1}\, S_{xyr}/S_{yyr}$, $\sigma^2 = \lim_{r\to\infty}\sigma_r^{\,2}$ and $\rho = \lim_{r\to\infty}\rho_r$

The total adjusted sum of squares is given by

$$S_{yyr} - S_{xyr}' S_{xxr}^{-1} S_{xyr} = S_{yyr}(1-\rho_r^2) = (N-1)\sigma_r^2 (1-\rho_r^2)$$

The adjusted treatment sum of squares A_{yyr} is given by

$$A_{yyr} = T_{yyr} + R_{xyr}' R_{xyr}^{-1} R_{xyr} - S_{xyr}' S_{xyr}^{-1} S_{xyr} .$$

Let $D_{yyr} = r \sum_{k=1}^{r} (\bar{Y}_{kr} - \bar{X}_{kr} \beta_r)^2$.

We have the following theorem.

Theorem 3.9

$\frac{(N-1-p)}{(N-1)} \frac{D_{yyr}}{\sigma_r^2(1-\rho_r^2)}$ has asymptotic distribution which is chi-square with (t-1) degrees of freedom.

Now from page $D_{yyr} \simeq A_{yyr} (1 + \frac{1}{N_r})$. So A_{yyr} converges to D_{yyr} in probability. We therefore have the following corollary.

Corollary:

$$\chi^2_{(r)} = \frac{(N-1-p)}{(N-1)} \frac{A_{yyr}}{\sigma_r^2(1-\rho_r^2)} = \frac{(N-1-p) A_{yyr}}{(A_{yyr} + E_{yyr})}$$ is asymptotically distributed as a chi-square variate with t-1 degrees of freedom.

Now $\hat{\beta}_r' R_{xyr} = R_{xyr}' R_{xxr}^{-1} R_{xyr} = \hat{\beta}_r' R_{xxr} \hat{\beta}_r$. By theorem 3.4 (Anderson - Cramer) we have the following theorem.

Theorem 3.10

a) $\frac{\sqrt{N_r}}{\sqrt{(t-1) \sigma_r^2 (1-\rho_r^2)}} (\hat{\beta}_r - \beta_r)$ is asymptotically normally distributed with mean 0 and variance Σ_{xx}^{-1}.

b) $\frac{N_r}{(t-1)\sigma_r^2 (1-\rho_r)^2} (\hat{\beta}_n - \beta_n)' R_{xxn}^{-1} (\hat{\beta}_r - \beta_r)$ is asymptotically

distributed as a chi-square variable with p degrees of freedom

c) $\frac{1}{N} (R_{xyr}' R_{xxr}^{-1} R_{xyr} - \frac{(N-t)}{(N-1)} S_{xyr}' S_{xxr}^{-1} S_{xyr})$ converges in

probability to 0.

An approximate test for the null hypothesis for large values of r is to reject the hypothesis if $\chi^2_{(r)} > \chi^2_{t-1(\alpha)}$, the tabulated α significance level of the chi-square distribution with t-1 degrees of freedom. For large values of r this is equivalent to the usual F test, i.e. The usual GMN test seems to be appropriate under the randomization model if the general conditions discussed previously are satisfied. Similar procedures will follow for the other GMN statistics of interest.

If the response variable and the concomitant variates are related such that $\lim_{r \to \infty} \rho_r = \rho, |\rho| > \varepsilon > 0$, then the asymptotic power of the tests are increased by including the covariates in the model. Also more efficient estimates of treatment contrasts are obtained.

Let A_r^*, $R_r^* \chi^2_{(r)}{}^*$ and S_r^* denote their counterparts under the alternative hypothesis. If additivity of treatment and response unit effects hold we have

$$S_{yyr}^* = \Sigma_i (z_{i0r} + \Sigma_k \tau_k \delta_i^k)^2$$

$$= \Sigma_i z_{i0r}^2 + r\Sigma_i \tau_k^2 + 2\Sigma_k \tau_k \Sigma_i z_{i0r} \delta_i^k$$

$$= S_{yyr} + r\Sigma_k \tau_k^2 + 2\Sigma_k \tau_k z_{0r}^{(k)}$$

$$S_{xyr}^* = S_{xyr} + \Sigma_k \tau_k \Sigma_i (z_{i1r}, z_{i2r}, \ldots, z_{ipr})' \delta_i^k$$

$$= S_{xyr} + \Sigma_k \tau_k Z_r^{(k)} \quad , \quad \text{say,}$$

where $Z_r^{(k)}$ is the sum of the covariate vectors in treatment group k and $Z_{0r}^{(k)}$ is the sum of the basal responses for treatment group k.

Now $$S_{xxr}^{*} = S_{xxr} \text{ , and } R_r^{*} = R_r$$

So $$S_{xyr}^{*\prime} S_{xxr}^{-1} S_{xyr}^{*} = S_{xyr}' S_{xxr}^{-1} S_{xyr}$$

$$+ \Sigma_k \tau_k^2 Z_{.r}^{(k)\prime} S_{xxr}^{-1} Z_{.r}^{(k)} + \Sigma_{kk'}^{\neq} \tau_k \tau_{k'} Z_{.r}^{(k)} .$$

$$S_{xxr}^{-1} Z_{.r}^{(k)\prime} \quad ,$$

and $$G^*_{.yyr} = S_{yyr}^{*} - S_{xyr}^{*\prime} S_{xxr}^{*-1} S_{xyr}^{*} = S_{yyr} - S_{xyr}' S_{xxr}^{-1} S_{xyr}$$

$$+ \Sigma \tau_k^2 (r + \frac{t}{N_r-1}) = (N-1)\sigma_r^2 (1-\rho_r^2)(1 + \delta_r^2) \quad ,$$

where $\delta_r^2 = \Sigma \tau_k^2 (r + \frac{t}{N_r-1}) / (N-1)\sigma_r^2 (1-\rho_r^2)$.

Let $\lambda_r = r \Sigma \tau_k^2 / \sigma_r^2 (1-\rho_r^2)$. If $\lambda_r \to \lambda < \infty$, then $\delta_r^2 \to 0$.

If $\lambda_r \to \lambda < \infty$, then $\delta_r^2 \to 0$.

So $\chi_{(r)}^{2*}$ will tend in distribution to a non-central chi-squared variate with t-1 degrees of freedom and non-centrality λ.

3.6 The Analysis of Covariance in the Randomized Block Design

In the case of the randomized block design we have a sequence Z_r of r blocks each with t basal vector units, i.e., $Z_r = \{z_{rij} \ 1 \le i \le r, 1 \le j \le t\}$, $r = 1,2,\ldots$. Each element $z_{rij} = (z_{rij}^{(0)}, z_{rij}^{(1)}, \ldots, z_{rij}^{(p)})$ is a (p + 1) vector with $z_{rij}^{(0)}$ being the response component and $(z_{rij}^{(1)}, \ldots,$

$z_{rij}^{(p)}$, a p-vector concomitant which we assume is not affected by the treatment protocol. The t treatments are applied at random to the units in each block, the randomization in each block being independent. So formally let $Z_r^{(i)} = \{z_{r11}^{(i)}, \ldots, z_{r1t}^{(i)}, \ldots, z_{rr1}^{(i)}, \ldots, z_{rrt}^{(i)}\}$ {i=0,1, ..., p; r = 1,2,...} be p + 1 sequences of real numbers with each component sequence $Z_r^{(i)}$ consisting of r blocks, $z_{rj}^{(i)}$, of t elements each, i.e., $z_{rj}^{(i)} = (z_{rj1}^{(i)}, \ldots, z_{rjt}^{(i)})$. Let $Y_r = \{Y_{r1}, \ldots, Y_{rr}\}$ and $X_r^{(i)} = \{X_{r1}^{(i)}, \ldots, X_{rr}^{(i)}\}$ be random variables such that the matrix$(Y_{rj}, X_{rj}^{(1)}, \ldots, X_{rj}^{(p)})$ takes on with equal probability each of the t! permutations of the block of (p + 1)-vector units of Z_{rj}, j=1,..., r. Tnen using the y_{rjk} as the response variate for the k-th treatment in the j-th block and the $X_{rjk}^{(i)}$ (i=1,..., p) as the corresponding concomitant variates, the analysis of covariance for the RBD is as follows,

Source	yy	xy	xx
Blocks	+	+	+
Treatments	T_{yy}	T_{xy}	T_{xx}
Residuals	R_{yy}	R_{xy}	R_{xx}
Total-Blocks	G_{yy}	G_{xy}	G_{xx}

where, omitting the first subscript r

$$T_{xy} = \Sigma_k \{\Sigma_i (x_{ik} - x_{.k})' \, \Sigma_i (y_{ik} - y_{.k})/r\} \text{ is a p-vector,}$$

$$R_{xy} = \Sigma_{ik}\{ (x_{ik} - x_{i.} - x_{.k} + x_{..})'(y_{ik} - y_{i.} - y_{.k} + y_{..})\}$$

and

$G_{xy} = T_{xy} + R_{xy}$. The dot indicates the average over the missing subscript. The other components of the analysis of covariance table may be obtained by making the obvious substitutions.

In the randomization model, under the null hypothesis of no differential treatment effects, $G_r = S_r$ = constant. Robinson (1973) has investigated the analysis of covariance for the case of a single covariate. There is no added difficulty or difference in procedure for the case of multiple concomitant observations. We shall state some of Robinson's results in the following paragraphs. Then we will give the multiple covariate generalization of the conditions required on the sequences in order that the randomization distribution and tests for various statistics may be approximated fairly accurately by their GMN counterparts. We drop the subscript r where there is no ambiguity.

Write

$$D_{yy} = r\Sigma_j \{y_{.j} - y_{..} - (x_{.j} - x_{..})b_r\}^2$$

$$\rho_r = S_{xy}S_{xx}^{-1}S_{xy} \;/\; S_{yy} \quad ,$$

$$\beta_r = S_{xx}^{-1}S_{xy},\; b_r = R_{xx}^{-1}R_{xy}$$

$$\sigma_{rxy} = \frac{1}{r(t-1)} S_{xyr}$$

$$\sigma_{rxx} = \frac{1}{r(t-1)} S_{xxr}$$

$$\sigma_{ryy} = \frac{1}{r(t-1)} S_{yyr}$$

$$B = \frac{r\,\Sigma_j\{y_{.j} - y_{..} - (x_{.j} - x_{..})\,\beta_r\}^2}{\{G_{yy} - G_{xy}G_{xx}^{-1}G_{xy}\} \;/\; \{r(t-1) - 1\}} \quad .$$

The adjusted treatment sum of squares is given by

$$A_{yy} = T_{yy} + R_{xy}R_{xx}^{-1}R_{xy} - G_{xy}G_{xx}^{-1}G_{xy} \quad .$$

The adjusted error sum of squares is given by

$$E_{yy} = R_{yy} - R_{xy}R_{xx}^{-1}R_{xy} \quad .$$

The ratio of the adjusted means squares for treatment effects and (residual) error is given by

$$F = \frac{A_{yy}/\ (t-1)}{E_{yy}/\ \{(r-1)(t-1) - 1\}} \quad ,$$

and is taken to have an F-distribution with (t-1) and (r-1)(t-1) - 1 degrees of freedom under GMN conditions. The ratio of the adjusted treatment sum of squares and the sum of the adjusted treatment sum of squares and the adjusted error sum of squares is given by

$$W = \frac{A_{yy}}{A_{yy} + E_{yy}}$$

and is taken to have a beta-distribution with parameters $\mu = (t-1)$ and $\nu = (r-1)\ (t-1) - 1$ under GMN conditions.

Let $\alpha_i^2 = \Sigma_k(x_{ik} - x_i)^2$, $\theta_i^2 = \Sigma_k(y_{ik} - y_{i.})^2$ and $\gamma_i^2 = (\alpha_i^2 \,/\, \Sigma\, \alpha_i^2 + \theta_i^2 \,/\, \Sigma\, \theta_i^2)$. Let $Q = (q_{jk})$ be an orthogonal txt matrix with the elements of the t-th row having equal values ($= t^{-\frac{1}{2}}$). Let $U_{ri} = Q\ Y_{ri}$ and $V_{ri} = QX_{ri}$, $i=1,2, \ldots, r$ so that the elements of U_{ri} and V_{ri} are given by

$$u_{rij} = \Sigma_{k=1}^{t} q_{jk}\, y_{rik} \quad \text{and} \quad v_{rij} = \Sigma_{k=1}^{t} q_{jk}\, x_{rik} \,,$$

$$j=1,2,\ldots,\ t$$

$$i=1,2,\ldots,\ r \ .$$

Then we have

$$\sum_{j=1}^{t-1} u_{rij}^{2} = \sum_{k=1}^{t} (y_{rik} - y_{ri.})^{2} = \alpha_{i}^{2}$$

and

$$\sum_{j=1}^{t-1} v_{rij}^{2} = \sum_{k=1}^{t} (x_{rik} - x_{ri.})^{2} = \theta_{i.}^{2} \quad .$$

Let $u_{r.j}^{*} = \Sigma_i u_{rij} \,/\, r^{\frac{1}{2}} \cdot \sigma_{yy}^{(r)}$ and $v_{r.j}^{*} = \Sigma_i\, v_{rij} \,/\, r^{\frac{1}{2}} \cdot \sigma_{xx}^{(r)}$.

We state Robinson's results in the following theorems.

<u>Theorem 3.10</u> (Robinson)

If $\rho_r \dashrightarrow \rho$ as $r \dashrightarrow \infty$ with $|\rho| < 1$, and if for any $\epsilon > 0$

$$\sum_{|\gamma_i| > \epsilon} {}_{i}^{2} \dashrightarrow 0 \ \text{ as } r \dashrightarrow \infty \,, \qquad 3.6.1$$

then the (t-1) pairs $(u_{r,j}^{*} \ , \ v_{r.j}^{*})$ $(j=i, \ldots, t-1)$

are asymptotically distributed as (t-1) independent bivariate normal variates with zero mean, unit variances and correlation.

<u>Proof</u>:

The conditions 3.6.1 imply the conditions of Theorem 2.9 (Cramer) (See Robinson (1974).)

<u>Theorem 3.11</u> (Robinson)

If $\beta_r \to \beta$ and $\rho_r \to \rho$ as $r \to \infty$ with $|\rho| < 1$ and if for any $\epsilon > 0$, 3.6.3 holds, then B is asymptotically distributed as a chi-squared variate with t-1 degrees of freedom.

<u>Theorem 3.12</u> (Robinson)

Under conditions 3.6.1, $r(t-1)\, W - B \to 0$ in probability and the asymptotic distribution of $r(t-1)W^*$ is that of a chi-square variate with (t-1) degrees of freedom.

<u>Proof</u>: (See Robinson (1973).)

By transforming the $N(=rt)$ responses by an NxN orthogonal matrix and proceeding in the fashion outlined in the previous section, we can show that if conditions 3.6.1 hold, the matrix of residual sum of squares and cross products R, may be expressed as the sum of $(r-1)(t-1)$ uncorrelated identically distributed random variables, each with limiting multivariate normal distributions. So by the Lindeberg central limit theorem R has limiting multivariate normal distribution. Let $\sigma_{ryy} = E_{yy} / \{(r-1)(t-1)\}$, then by Theorem 3.4 (Anderson-Cramer) and Theorem 3.5 (Aroian) we have the following theorem.

<u>Theorem 3.13</u>

If conditions 3.6.1 hold, then

1. $$r^{\frac{1}{2}} \sigma_{ryy}^{-\frac{1}{2}} (b_r - \beta_r)$$

has limiting normal distribution with zero mean and unit variance;

2. $$r^{\frac{1}{2}} \sigma_{ryy}^{-\frac{1}{2}} (q^*_{r.j} , V^*_{r.j}) \quad (j=1, \ldots, t-1)$$

are asymptotically distributed as (t-1) independent standard normal variates;

3. $$\sigma_{ryy}^{-1} D_{yy} \text{ , } B, \text{ and } \sigma_{ryy}^{-1} A_{yy}$$

each converge in probability to the same chi-square variate with t-1 degrees of freedom;

4. F is asymptotical F-distribution with (t-1) and (r-1)(t-1) - 1 degrees of freedom.

Suppose $y^*_{rik} = \tau_k + y_{rik}$. Let B^*, W^*, σ_{ry*y*} and σ_{rxy*}, be the values of B, W, σ_{ryy} and σ_{rxy} when y_{rik} is replaced by y^*_{rik} in the corresponding formulae. Let

$$\lambda_r + r\Sigma_k (\tau_k - \tau_.)^2 / (1 - \rho_r^2) \sigma_{yy}^{(r)} .$$

Theorem 3.14 (Robinson)

If $\beta_r \dashrightarrow \beta$ and $\rho_r \dashrightarrow \rho$ with $|\rho| < 1$, as $r \dashrightarrow \infty$, and if for any > 0, the conditions 3.6.1 holds, then

1. if $\lambda_r \dashrightarrow \lambda < \infty$, B^* tends in distribution to a non-central chi-squared variate with t-1 degrees of freedom and non-centrality parameter λ.
2. If $\lambda_r \dashrightarrow \infty$, $B^* \dashrightarrow$ in probability.

Proof: (See Robinson (1973).)

Equivalent results under nonnull values of τ are routine. For the analysis of multiple covariation situation the conditions equivalent to 3.6.1 are as follows.

1. $\lim_{r \to \infty} \Sigma_r = \Sigma$, positive definite, where $\Sigma_r = S_r/r(t-1)$ is the covariance matrix of Z_r .

2. For any $\varepsilon > 0$ $\quad \Sigma_{|\gamma| > \varepsilon} \gamma_i^2 \to 0$ as $r \to \infty$, where

$$\gamma_i^2 = \Sigma_j \alpha_{ij}^2 \quad \text{and} \quad \alpha_{ij}^2 = \Sigma_k (z_{rik}^{(k)} - z_{ri.}^{(k)})^2 \quad .$$

3.7 Other Related Work

In this section we shall give a brief account of some other work related to the analysis of covariance from the viewpoint of randomization theory.

Atiqullah (1964) investigated the effects of departures from normality and linearity and concluded that the requirement of linearity is quite critical for the analysis of covariance. He also claimed that the GMN procedure for the analysis of covariance is much less robust than the corresponding analysis of variance. Alternative approaches to the GMN procedure have been recommended by various workers, in cases where the usual assumptions seem to have been violated. One approach is to search for a suitable transformation. However, its use in parametric analysis may involve obscurities not only with regard to computation but also interpretation. Another approach which is commonly used when there are large samples and only two treatments to be compared is to pick out pairs of observations in which the values of the concomitant variates are approximately equal or "matched" and then to analyse these by one of the standard techniques for paired comparisons. I.J. Bross (1964) developed a test called "covast" which examines a dichotomous response variable (dead-alive) in the presence of a (quantitative) covariate (weight). Bross ordered the data in ascending value of the covariate and considered the distribution of the differences in counts of the number of failures of one treatment which followed a success of the other. The conditions required for validity of covast were (a) a monotonic response between the response and the covariate, and (b) independence of observations. Under the joint null hypothesis of no treatment differential effects and irrelevance of the covariate, each response has probability equal to .5 of success or failure. Bross claims that his procedure guards against spurious effects due to the covariate and also detects differences that may be masked by the covariate. Several workers have investigated the analysis of covariance based on ranks, see for example Puri and Sen (1969) and Quade (1967). The rank tests assume that the concomitant variable X has the same distribution in each (sub) population of interest. Besides this, the behavior for small samples with less than ten observations per group is unknown. These are obvious shortcomings.

3.8 Summary and Conclusions

We have been investigating some statistics from two commonly used GMN statistical designs in the class of comparative experiments under randomization theory. We have been concerned with the differences between the GMN results and the randomization theory results in the completely randomized design and the randomized block design for single response variates y in the presence of a concomitant variate X (possibly multivariate).

The validity of the GMN procedures requires that three parametric assumptions must be satisfied; 1) under the null hypothesis, H_o, of no differences between treatment effects, the conditional distribution of Y given X is normal with expectation $X\beta$, which is linearly dependent on X; 2) Y has variance equal to $\sigma^2 I$; and 3) the variance of y is independent of X. (If the variance is $\sigma^2 V$, where V is not equal to I the Aitkyn procedure is to make a linear transformation $EY = X^* \beta^*$ where $X^* = XT$, $\beta^* = T'\beta$ and $TVT' = I$).

We have shown that under randomization theory, although treatment contrasts are estimable in the analysis of covariance, it is not possible to obtain from the data an unbiased estimate of their variance. Also, in the CRD and RBD, the expectations of mean squares (EMS) for the adjusted treatment effects and adjusted error in the analysis of variance are not the same under the null hypothesis.

The technique of weighted randomization by D.R. Cox (1956) produces equality in the $\{EMS\}$. Also, treatment contrasts are estimable but unbiased estimates of their variances cannot be obtained from the analysis of variance.

We have also attempted to investigate the behavior of various statistics conditional on the sample units for large sample sizes of $N = rt$. Several authors have considered symmetrical functions, "U" statistics, other than the sum of N independent random variables. It is known that there are some type of functions max (X), min (X), for example, the limiting distribution of which, if they exist, may be non-normal.

Also a random variable X_N may converge in probability and hence in distribution to a random variable X, where X posesses finite moments even though $E\{X_i\}$ is not defined. If $\bar{X}_N$ is the mean of N independent $N(\mu,1)$ random variables, with $\mu\neq 0$, it can be shown that $E(\bar{X}_N^{-1})$ does not exist for any finite N. Yet the limiting distribution of $N^{\frac{1}{2}}\mu^{-2}(\bar{X}_N - \mu)$ which is $N(0, \mu^{-4})$. Hence the expectation of $(\bar{X}_N - \mu^{-1})$ is not defined. So, one cannot say that $E\{\bar{X}_N^{-1}\}$ converges to μ^{-1}. So we require that the moments of the sequences of order $r > 2$, should be well behaved in some sense, in order to determine that the sequence of expectations differ from a given sequence by a specified amount. A further problem may be to find a closer asymptotic representation of the distribution functions of these statistics, standardized sums, etc., than that provided by the normal distribution or alternatively to estimate the error committed by replacing them with the limiting distribution when N is finite although quite large. The conditions for convergence discussed in Chapter 2 define different patterns of behavior of various sequences of populations, which are sufficient to ensure asymptotic normality of certain statistics of interest. The weakest of these conditions are given in theorems 2.7 and 2.8. That is, we require Noether type conditions on the elements of each sequence or on the Euclidean norm of each vector element, together with a generalized Lindeberg type condition on the inner product of the (vector) elements. These conditions are sufficient to ensure that the (matrix) inner product of the equiprobable permutations of one (vector) sequence and another (vector) sequence will have asymptotic (possibly multivariate) normal distribution.

The generalized Lindeberg type conditions may be redundant as it appears that they are implied by the two Noether type conditions. However, a proof to this claim has not been resolved.

The conditions given in Chapter 2 are usually met in practice for moderately large samples. However, when the sample size is small, the problem of bias in the CRD or RBD is of paramount importance. One cannot appeal to asymptotic unbiasedness to justify the use of these standard procedures in such situations. The errors may not be small enough or numerous enough to make the difference between the true distribution function of the total error and that of the normal distribution small. Thus, the widely accepted theories based on normal law cannot be relied on with full confidence and may be considered merely as crude approximations.

4. DESIGNS FOR TREATMENTS APPLIED IN SEQUENCE

In this chapter we will attempt to examine the statistical properties of some of the more commonly used designs for comparing treatments that are applied in sequence, from the viewpoint of randomization. As we shall see, the designs are very similar to Latin square designs, so that the symmetric properties of the Latin squares (LS) and orthogonal sets of Latin squares are usually appealed to in order to justify their widespread use in situations where the errors in the observed responses are assumed to have GMN properties. It is standard knowledge that the usual GMN theory for the Latin square design gives, under additivity of treatments and units, estimators of treatment contrasts that are unbiased under randomization properties for designs with sequences of treatments.

An important feature of such designs is that is appropriate to contemplate the possibility of residual effects, that is an effect in a period of the treatment administered in the previous period. This leads to a sort of covariance model, surely a 2 part model in which there are direct and residual treatment effects.

We are prompted to examine this because of a partial resemblance to the covariance type structure considered in Chapter 3. In the case of the completely randomized or randomized block design, we have analysis of variance unbiasedness under additivity, but not, as we have seen, with the introduction of concomitant variation. We see a similar occurrence with change-over designs (cods) with residual effects.

The chapter consists of 3 parts. In Part 1 we will give some particulars of the Latin square design, a brief summary of the literature on change-over designs, and a brief account of the randomization theory on LS designs which includes a theorem due to Wilk and Kempthorne (1957) which illustrates quite clearly that the LS designs are biased for non-additive situations.

In Part 2 we will first give some required conditions for balance in cod's, which exploit the algebraic features and symmetries of orthogonal sets of latin squares, more particularly, incomplete latin squares. Then we will try to develop a coherent and detailed examination of the usual cod's and extra period cod's which satisfy the criteria of balance. We will try to identify the problems and obscurities which are inherent in these designs, from the viewpoint of randomization theory, then we define or specify our models to include residual effects of the first order. We will try to get some understanding of the properties of the usual GMN statistics assuming that possible technical errors are absent. From the GMN viewpoint our investigations may be considered, essentially, as examining the behavior of certain non-centralities in the analysis of variance arising from the variation in the experimental units. Nevertheless, we intend to show that (1), the models are unsatisfactory if (1.a) there are non-additivities between treatment effects and unit effects, (1.b) residual treatment effects of various orders, (of which we shall only consider the first) are present; and (2), fitting constants for these effects does not allow us to obtain from the data that are unbiased with respect to randomization estimators of the variance of the associated estimates of treatment contrasts.

Part 3 consists of a resumé of the chapter in which we give various analysis of variance tables and conclude the chapter with a brief summary and discussion.

4.1 Latin Square Designs

A Latin square design of order k is an incomplete three-way layout in which three factors, usually denoted by rows, columns and letters, are each at k levels. Only k^2 of the k^3 possible combinations occur according to the following pattern. The levels of the first two factors form a square with k rows and k columns. Each level of the third factor (letters) occurs once with each level of the first factor (rows) and each level of the second factor (columns). We may construct a latin square of any order k, by the cyclic permutation of the natural order of the k letters (see square 1 below):

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

Standard Latin squares of order 4

If we permute the rows, columns, letters and their roles of a latin square the result is obviously again a Latin square. The totality of latin squares obtainable in this way from a single square is called a transformation set. A canonical form for any kxk latin square called a standard square is one in which the letters in the first row and first column are arranged in their natural or usual order. From a standard square of order k we may obtain $k!(k-1)!$ squares by permuting the k columns and (k-1) rows which leave the first row in place. Hence, a transformation set contains $k!(k-1)!$ times the number of different standard squares in the transformation set. In order to give all squares of order k the same probability of being selected, we may choose a transformation set at random with probability proportional to the number of standard squares in the set. The conceptual population for the finite model situation using a Latin square of order k is obviously determined by the transformation sets of Latin squares of order k.

Two Latin squares of the same order are said to be orthogonal to each other if all k^2 combinations of letters taking order into account occur exactly once when one square is superimposed on the other. If $k = p^2$ where p is a prime number and m is a positive integer, it can be shown that there are k-1 mutually orthogonal Latin squares, called MOLS, which form a complete set. It has been shown that there is a 1:1 relation between a complete set of MOLS and the finite projective geometry PG(2,s) and conditions for their existence and construction have been given; see W.L. Stevens (1938), R.C. Bose (1938) and Fisher and Yates (1948). Since PG(2,s) does not exist for s = 6,14,21,..., etc., MOLS for these orders of Latin squares do not exist. A completely orthogonalized Latin square of order 4 is given by the following example:

1 2 3 4	1 2 3 4	1 2 3 4
2 1 4 3	3 4 1 2	4 3 2 1
3 4 1 2	4 3 2 1	2 1 4 3
4 3 2 1	2 1 4 3	3 4 1 2

Latin squares are used quite often where the units are suspected of having two major or consistent sources of systematic variation. In change-over or cross-over designs for example, where the experimental units receive different treatments at different time periods, the subjects are considered as one factor and the time periods are considered as another.

4.1.1 Designs for Treatments Applied in Sequence in the Literature

Cochran, Autrey and Cannon (1941), discussed in the introduction, gave the design and analysis of a short-time change-over trial. They used a set of orthogonal Latin squares in the design for comparing three rations. The principal object of the design was to secure accurate comparisions of the effects of the rations and unbiased estimates of experimental errors. The appropriate statistical analyses were illustrated for negligible and non-negligible carry-over effects.

E.J. Williams (1949 and 1950) gave methods of constructing balanced designs for the estimation of direct effects and residual effects or pairs of residual effects when there is no trend with periods. He found that one could obtain balance with one square for an even number of treatments and with two squares for an odd number of treatments.

H.D. Patterson (1950) and H. Lucas (1951, independently) have examined the method of analysis when constants are fitted for residual effects persisting for one period. Their calculations are based on the assumption of the independence of the linear, quadratic, etc. contrasts for each experimental unit. They have shown that bias exists for adjusted effects and permanent (adjusted plus carry over) effects. When $\sigma_1^2 > \sigma_2^2$, where σ_1^2, σ_2^2 are the linear and quadratic, components of error, respectively, the variances and covariances of contrasts of direct effects are underestimated. An average correction can be calculated.

D.R. Cox (1951) gives systematic designs for use when a number of treatments are to be compared, one treatment being applied to each of a number of equally spaced plots. It is assumed that there is a smooth trend between plots and the error is independent in different plots. The designs enable the treatment effects (and the trend) to be estimated simply and accurately.

H.D. Patterson (1951) considered a number of designs and found that confounding in factorial arrangements and incomplete block designs could be introduced by adaptation of the method of single period experiments.

D.R. Cox (1952) considered systematic experimental designs for use when the residual variation is (1) autocorrelated, (2) formed from a trend plus a random error.

H.D. Patterson (1952) constructed balanced designs for experiments involving sequences of treatments, with the use of finite groups, finite fields, abelian groups and combinatorics.

"Extra period Latin square change-over designs" by H.L. Lucas (1957) pointed out several interesting points of contrasts between the extra period and regular Latin square designs.

In the former, each treatment is preceded by every other treatment as in Latin squares, but in addition each treatment is preceded by itself. This renders the one-period residual effects orthogonal, in the least square sense, to direct effects. In contrast to the Latin squares, the residual effects in the extra period designs are orthogonal to sequences. The direct effects are, however, non-orthogonal to experimental units in the extra period patterns, but the degree of non-orthogonality "is not great." Finally the amount of replication of the residual effects is a little greater than in the Latin squares. The reduction in non-orthogonality results in a somewhat simpler analysis for the extra period designs than for Latin squares.

Patterson (1959) considered an extra-period design obtained by repeating the treatment pattern of the last period of any design in a general class of basic change-over designs. The basic designs are derived from Latin squares or incomplete Latin squares and satisfy certain conditions of balance which facilitate the estimation of direct and first residual effects. Patterson and Lucas (1959) also considered designs in incomplete blocks.

The extra period designs have the useful property that the estimates of direct effects are orthogonal, in a least squares sense,to the estimates of first-period residual effects. As with basic designs, the estimates of error given by the method of fitting constants for direct and residual effects are biased.

W.T. Federer (1964) gave methods for the construction and analysis of a class of experimental designs which he called "tied-double-cods." These designs involve t treatments in r rows (periods) and c columns (sequences - experimental units). In some designs he made use of (orthogonal) Latin squares repeating certain rows and columns for special values of r ($r=tq + 1$) and c ($c=ts$). He gave computing formulae for the estimation of direct treatment effects, residual effects, sums of squares in the analysis of variance, etc., when the first period results are omitted and when they are included. Grizzle (1965) gave a method for testing residual effects in the two period cod, and proceeded conditionally on the results. Koch (1972) used a Wilcoxon non-parametric technique for the case in which residual effects are assumed to be absent.

Sharma (1977) also gave methods of constructing designs that permit the estimation of direct effects orthogonal to all other effects when residual effects persists for two consecutive periods. Sharma gave methods of analysis for the situations where the first period observations are omitted from the analysis and where they are included.

Brown (1980) examined the economic feasibility of using the two period cod in clinical trials as compared to the completely randomized design. He concluded it was uneconomical.

Under most favorable conditions where residual effects are absent and Gauss-Markov conditions hold, these designs may be quite useful. However, we face major problems and deep obscurities when first and second order residual effects persist. For the designs in which direct treatment effects are orthogonal, in the least squares sense, to experimental units or subject effects, the residual effects are partially confounded with either direct treatment effects or subject effects or both. If residual treatment effects are orthogonal to the direct effects, the latter will not be orthogonal to subject effects. This is because there are no residual effects in the first period for the first and second order residual treatments, and no effects in the second period for the second and higher order residual treatments. Also it is far from clear how we may be able to justify or determine the conditions under which the assumptions that the linear, quadratic, cubic, ..., etc. contrasts will be independent, which is what Patterson (1950, and Lucas (1951) assumed, as mentioned earlier.

4.1.2 Randomization Theory of the Latin Square Design.

Suppose additivity (previously defined) of treatments and units applies in this situation, that is, there is no interaction between treatments and experimental units. If treatment k is applied to the unit in the i-th row and j-th column, then the response would be given by

$$x_{ijk} = u_{ij} + \tau_k \ , \qquad i,j,k = 1,2,\ldots,t \qquad 4.1.2.1$$

With respect to the u_{ij} we have the algebraic identity

$$u_{ij} = u_{..} + (u_{i.} - u_{..}) + (u_{.j} - u_{..}) + (u_{ij} - u_{i.} - u_{.j} - u_{..}) \ ,$$

where $u_{..}$ is the general mean, etc.,

$$u_{..} = \Sigma_{ij} u_{ij}/k^2 \ ; \quad u_{i.} = \Sigma_j u_{ij}/k \ ; \quad u_{.j} = \Sigma_i u_{ij}/k.$$

And so if additivity holds we will have the Fisher model for the latin square design

$$y_{ijk} = \mu + r_i + c_j + \tau_k + e_{ij} \ , \qquad 4.1.2.2$$

where $\mu = \bar{u}_{..}$

$r_i = u_{1.} - u_{..}$ is the row effect of the i-th row,

$c_j = u_{.j} - u_{..}$ is the column effect of the j-th column,

$e_{ij} = u_{ij} - u_{i.} - u_{.j} + u_{..}$ is the "plot effect" of the plot in the i-th row and j-th column which may be considered as an interaction between row i and column j,

and τ_k is the treatment effect for the k-th treatment.

Since only one treatment occurs on a unit in a given row and column, the set of k^2 observations $\{y_{ijk}\}$ is a random subset of the conceptual set of all k^3 possible treatment unit combinations $\{x_{ijk}\}$.

If additivity does not hold, the model would have to include treatment x row $(r\tau)_{ik}$, treatment x column $(c\tau)_{jk}$, and treatment x plot or treatment x row x column $(rc\tau)_{ijk}$ interaction terms. In addition to this, if we think of the experiment as just one outcome of a sequence of repetitions under the same conditions (which may be impossible) the response x_{ijk} would differ from this conceptual true response X_{ijk} by a technical error η_{ijk}, due to variability in technique, or error caused by the measuring instrument of the observer, so that

$$x_{ijk} = X_{ijk} + \eta_{ijk} \quad ,$$

where the η_{ijk} are regarded as uncorrelated random variables, independent of the X_{ijk}, with zero expectation and variance σ^2 for every i,j,. So we will have the situation

$$x_{ijk} = \mu + r_i + c_j + (rc)_{ij} + \tau_k + (r\tau)_{ik} + (c\tau)_{jk} + (rc\tau)_{ijk} + \eta_{ijk} \tag{4.1.2.3}$$

In the randomization model the expectation of means for treatments, treatment mean squares etc., would be taken over the complete set of latin squares of order t. If we assume that treatments on neighboring plots do not affect each other and we ignore possible technical errors, then our conceptual population would consist of t^3 entities X_{ijk}, i,j,k=1,...t. We may then express the model in terms of the following algebraic identity

$$\begin{aligned} X_{ijk} = {} & X_{...} + (X_{i..} - X_{...}) + (X_{.j.} - X_{...}) + (X_{ij.} - X_{i..} - X_{.j.} + \\ & X_{...}) + (X_{..k} - X_{...}) + (X_{i.k} - X_{i..} - X_{..k} + X_{...}) + (X_{.jk} - \\ & X_{.j.} - X_{..k} + X_{...}) + (X_{ijk} - X_{i.k} - X_{.jk} + X_{..k} - X_{ij.} + X_{i..} \\ & + X_{.j.} - X_{...}) \end{aligned}$$

$$= \mu + r_i + c_j + (rc)_{ij} + \tau_k + (r\tau)_{ik} + (c\tau)_{jk} + e_{ijk} \quad , \qquad 4.1.2.4$$

where we have an obvious correspondence of terms. The usual (summation) side conditions on the parameters are seen to hold by definition, i.e.,

$\Sigma_i r_i = \Sigma_j c_j = \Sigma_k \tau_k = \Sigma_i (r\tau)_{ik} = \Sigma_k (r\tau)_{ik} = \Sigma_j (c\tau)_{jk} = \Sigma_k (c\tau)_{jk} = \Sigma_i (rc)_{ij} = \Sigma_j (rc)_{ij} =$

$\Sigma_i e_{ijk} = \Sigma_j e_{ijk} = \Sigma_k e_{ijk} = 0$.

In any square we will have only t^2 observation $\{y_{ijk}\}$ which form a random subset of these t^3 possible observations $\{y_{ijk}\}$ which comprise our population.

A more general case of the Latin square design consists of choosing at random t levels from each of three factors, rows, columns and letters, say, with R,C,T factors respectively, and combining the levels selected in a Latin square. The effect on the expectations of treatment means, mean squares, etc., of the presence of non-additivities or interaction terms in this situation has been explored by Wilk (1955) and Wilk and Kempthorne (1955 and 1957).

In the usual Gauss-Markov Normal situation, we assume an additive linear model with no interaction terms;

$$y_{ijk} = \mu + r_i + c_j + \tau_k + e_{ijk} \quad , \qquad 4.1.2.5$$

or $\quad Y = \mathcal{J}\mu + Z_1 r + Z_2 c + Z_3 \tau + e \quad ,$

where Y is an $n(=t^2)$ dimensional vector of observations; $r = (r_1, r_2, \ldots, r_t)'$, $c = (c_1, c_2, \ldots, c_t)'$, $\tau = (\tau_1, \tau_2, \ldots, \tau_t)$ are t-dimensional vectors denoting the row, column, and letter (treatment) effects respectively, satisfying $\mathcal{J}'_t r = \Sigma_i r_i = \mathcal{J}_t' c = \mathcal{J}_t' \tau = 0$; and e is an n-dimensional vector of errors which are uncorrelated with mean zero and variance σ^2, Z_1, Z_2, and Z_3 are incidence matrices of dimensions $t^2 \times t$, for the row, column and treatment effects. The column vectors of each of these incidence matrices sum to $\mathcal{J}_n$ and the column subspaces of dimension t-1 orthogonal to $\mathcal{J}_n$ are orthogonal to each other. The totals for rows, columns and treatments are given by

$$R_i = Z_{1i}'Y = \Sigma_{jk}y_{ijk} \quad , \qquad i = i,2,\ldots,t \quad ,$$

$$C_j = Z_{2j}'Y = \Sigma_{ik}y_{ijk} \quad , \qquad j = 1,2,\ldots,t \quad ,$$

$$T_k = Z_{3k}'Y = \Sigma_{ij}y_{ijk} \quad , \qquad k = 1,2,\ldots,t \quad ,$$

where the summation for T, say, is over the t pairs (i,j) where i and j each take on once, each of the values 1,2,...,t for the fixed k according to the Latin square chosen. The sum of squares due to rows, S_R^2, columns, S_C^2, treatments (letters), S_T^2 and error S_E^2 are

$$S_R^2 = \Sigma_i R_i^2/t - t^2y^2_{\ldots}$$

$$S_C^2 = \Sigma_j C_j^2/t - t^2y^2_{\ldots}$$

$$S_T^2 = \Sigma_k T_k^2/t - t^2y^2_{\ldots}$$

$$S_E^2 = \Sigma_{ijk}y_{ijk}^2 - t^2y^2_{\ldots} - S_R^2 - S_C^2 - S_T^2$$

Now under GMN conditions and additivity the ratio $F = S_T^2/(t-1) \div S_E^2/(t^2 - 3t + 2) = (t-2)S_T^2/S_E^2$ will have the F distribution with (t-1) and $(t^2 - 3t +2)$ degrees of freedom. This means that for squares of order 2, 3,4, and 5 we will have 0,2,4 and 12 degrees of freedom for error, respectively. This renders the square of order 2 practically useless. In the situation where "rows" and "columns" are used for blocking, the squares of order 3 and 4 must produce a substantial reduction in error over randomized blocks or the completely randomized design to counterballance the loss of degrees of freedom for error. If the number of treatments to be compared is large, the number of units required (t^2) may be prohibitive. Besides, the experimental error per unit may increase with the size of the square. Sometimes, in latin square designs of small order, more than one square is used in the same experiment. If it is assumed that the treatment effects do not vary from square to square or, in other words, the square x treatment interaction is negligible, it may be pooled with error.

If additivity holds, it is easy to show that $y_{..k} - y_{...}$ is an unbiased estimate of τ_k. If interactions are permitted, with the usual side conditions we will have, ignoring possible technical errors,

$$y_{..k} - y_{...} = \tau_k + \Sigma_{ij}((rc)_{ij} + (rct)_{ijk}) = \tau_k + w_k, \quad \text{say,}$$

where the summation is over the pairs of (i,j) that occur with k. If $v = \lambda'\tau = \Sigma_k\lambda_k\tau_k$ is any contrast in the $\{\tau_k\}$ then its estimate $\hat{v} = \Sigma_k\lambda_k y_{..k}$ will have a bias equal to $\Sigma_k\lambda_k w_k$. This bias will obviously depend on the latin square actually used. Under the randomization model situation, where we randomized in selecting the latin square, the bias, $\Sigma_k\lambda_k w_k$, may be considered as a unit error with zero expectation denoting differences among experimental units from the conceptual population generated by the transformation sets. In this situation $y_{..k} - y_{...}$ would be an unbiased estimate of t_k . The expectations of mean squares were derived by Wilk and Kempthorne (1957) for the ease where the factors may have their levels sampled from a population of levels. The results are presented here in the following theorem which is given without proof.

Suppose we have three factors $\mathcal{R}$, $\mathcal{C}$ and $\mathcal{T}$ with R, C and T levels respectively. And, suppose we have a finite universe of RCT numbers $\{X_{ijk}\}$, $(i=1,\ldots, R_i = 1, \ldots, C_i$ and $k = 1, \ldots, T)$, associated with the RCT possible combinations of the three factors. Let the dot $(\cdot)$ denote the mean value over the missing subscripts in the X_{ijk}'s, so that, for example, $X_{.j.} = \Sigma_{i,k} X_{ijk}/RT$.

Write

$$r_i = (X_{i..} - X_{...}) \; ; \; c_j = (X_{.j.} - X_{...}) \; ; \; \tau_k = (X_{..k} - X_{...}) \; ;$$

$$(rc)_{ij} = (X_{ij.} - X_{i..} - X_{.j.} + X_{...}) \; ;$$

$$(r\tau)_{ik} = (X_{i.k} - X_{i..} - X_{..k} + X_{...}) \; ;$$

$$(c\tau)_{ij} = (X_{.jk} - X_{.j.} - X_{..k} + X_{...}) \; ; \quad \text{and}$$

$$(rc\tau)_{ijk} = (X_{ijk} - X_{i.k} - X_{.jk} + X_{..k} - X_{ij.} + X_{i..} + X_{.j.} - X_{...})$$

and write

$$\sigma_\tau^2 = \Sigma_k(\tau_k)^2/T-1); \quad \sigma_{rc}^2 = \Sigma_{ij}(rc)_{ij}^2/(R-1)(C-1) \quad ;$$

$$\sigma_{rct}^2 = \sum_{ijk} (rc\tau)_{ijk}^2/(R-1)(C-1)(T-1) \quad ; \qquad \text{etc.}$$

Theorem 4.1 (Wilk and Kempthorne (1957).)

If, from the set of levels of each of the factors of $\mathcal{R}$, $\mathcal{C}$ and $\mathcal{T}$ a simple random sample of t levels is chosen, and if, from the subset of t^3 possible numbers $\{X^{**}_{ijk}\}$ associated with the t^3 combinations of the selected levels of $\mathcal{R}$, $\mathcal{C}$, $\mathcal{T}$, a random sample of t^2 numbers $\{X'''_{ijk}\}$ is chosen to satisfy the conditions of the latin square design of order t. Then with $S_T^{\ 2}$ as previously defined and $MS_T = S_T^{\ 2}/(t-1)$, under the randomization procedure, we have the following expectations:

a) $E(S_T^2) = (t-1)[\sigma^2 + (f+t/(RC))\sigma^2_{rc\tau} + \sigma^2_{rc} + (1-t/R)\sigma^2_{r\tau} + (1-t/C)\sigma^2_{c\tau} + \tau\sigma^2_{\tau}]$ 4.1.2.6

b) $E(S_E^2) = (t-1)(t-2)[\sigma^2 + f\sigma^2_{rc\tau} + \sigma^2_{rc} + \sigma^2_{r\tau} + \sigma^2_{c\tau}]$ 4.1.2.7

where $f = (1-1/R - 1/C - 1/T)$. 4.1.2.8

When R = C = T = t we obtain

c) $E(MS_T) = \sigma^2 + (1-2t^{-1})\sigma^2_{rc\tau} + \sigma^2_{rc} + t\sigma^2_{\tau}$ 4.1.2.8

d) $E(MS_E) = \sigma^2 + (1-3t^{-1})\sigma^2_{rc\tau} + \sigma^2_{rc} + \sigma^2_{r\tau} + \sigma^2_{c\tau}$ 4.1.2.9

Proof: See Wilk and Kempthorne (1957).

Using 4.8, 4.9 if there are no interactions of treatments with rows, columns and rows by columns, and if there are no residual effects we have

$$EMS_T = \sigma^2 + \sigma^2_{rc} + t\sigma^2_{\tau} \quad ,$$

and

$$EMS_E = \sigma^2 + \sigma^2_{rc} \quad .$$

This is one way of seeing that the LS design is AOV unbiased.

4.2 The General Balanced Change-Over Design [cod] and Extra Period Design

In this section we shall attempt to give a critical examination of these two types of designs for sequences of treatments. The extra period change-over design may be considered essentially as an ordinary change-over design in which the treatment applied in the final period is repeated in an extra period. As in the work of Patterson (1952), Lucas (1957), Williams (1949) and others, the plans satisfy certain conditions of balance.

In balanced change over designs, designated cod, our plan consists of sequences of treatments arranged to form replicate sets of mutually orthogonal Latin squares or mols, or incomplete Latin squares. Suppose we have N experimental units and t treatments to compare in $p \leq t$ time periods. The sequences are assigned at random to the experimental units so that the scheme ensures the following conditions of balance:

1) treatment k is applied to N/t units in each period;
2) treatment k is applied to each unit at most once;
3) treatment k and k' occur together in $Np(p-1)/t(t-1)$ units;
4) treatment k occurs in $Np(t-p)/t(t-1)$ units in which treatment k' does not occur;
5) treatment k immediately precedes treatment K' in $N(p-1)/t(t-1)$ units;
6) treatment k occurs in $N(p-1)/t(t-1)$ of the N/t units in which treatment k' occurs in any given period;
7) treatments k and k' both occur in $N(p-1)(p-2)/t(t-1)$ units in periods other than the first;
8) treatment k occurs in $N(p-1)(t-p+1)/t(t-1)$ units in which treatment k' does not occur, in periods other than the first.

Conditions 1-4 enable us to derive the expectations under randomization of estimates of treatment effects and sums of squares in the cods in which residual affects are assumed to be absent. Conditions 5 to 8 are included to facilitate the computations of interest in the case where we assume residual effects of the first order.

4.2.1 Randomization Analysis of the General Balanced COD

We now examine the general case of balanced cods in which we wish to compare the effect of t treatments over p periods, $p \leq t$, on N experimental units. The material that follows is related to that of Yates (1936) on incomplete Latin square experiments. We note that since only one of the t treatments can be assigned to an experimental unit in any given period, if there are no residual effects from treatments in preceding periods we will have a random response of the set of t possible responses on that unit at that period. If we assume additivity between treatment, period and experimental unit effects we may write the model for the observed responses in the form

$$Y_{ijk} = \mu + s_i + p_j + t_k + e_{ij} \quad , \qquad 4.2.1.1$$

where Y_{ijk} is the observation on the i-th unit in the j-th period to which treatment k was assigned. μ is the general mean, s_i the unit effect, p_j the period effect, t_k the treatment effect and e_{ij} the error.

In our subsequent investigations of the properties of the balanced cods, we shall not consider directly the particular sequences of treatments but we will make use of the set of Npt design random variables $\{\alpha_{ij}^{k}\}$, $(i=1, \ldots, N;\ j=1, \ldots, p;\ k=1, \ldots, t)$, which define the experimental plan completely and which have the following properties:

$$\alpha_{ij}^{k} = 1 \ , \quad \text{if the k-th treatment is given to the i-th experimental unit in the j-th period,}$$

$$= 0 \quad \text{otherwise.}$$

$\Sigma_k \alpha_{ij}{}^k = 1$, $\Sigma_i \alpha_{ij}{}^k = N/t$.

$P(\alpha_{ij}{}^k = 1)$ $1/t$,

$P(\alpha_{ij}{}^k \alpha_{ij'}{}^k = 1) = 0$ $j \neq j'$,

$P(\alpha_{ij}{}^k \alpha_{i'j}{}^k = 1) = \binom{N/t}{2} / \binom{N}{2} = (N-t)/t^2(N-1)$, $i \neq i'$,

$P(\alpha_{ij}{}^k \alpha_{i'j'}{}^k = 1) = 1/t \cdot N/t(N-1) = N/t^2(N-1)$, $i \neq i'$, $j \neq j'$,

$P(\alpha_{ij}{}^k \alpha_{ij}{}^{k'} = 1) = 0$, $k=k'$

$P(\alpha_{ij}{}^k \alpha_{ij'}{}^{k'} = 1) = 1/t(t-1)$, $j \neq j'$, $k=k'$,

$P(\alpha_{ij}{}^k \alpha_{i'j}{}^{k'} = 1) = N/t^2(N-1)$, $i \neq i'$, $k \neq k'$,

and

$P(\alpha_{ij}{}^k \alpha_{i'j'}{}^{k'} = 1) = (Nt - N - t)/t^2(t-1)(N-1)$, $i \neq i'$, $j \neq j'$, $k \neq k'$

Also we denote the sum $\Sigma_j \alpha_{ij}{}^k$ by α_i^k and we have $\Sigma_i \alpha_i{}^k = Np/t$; $\Sigma_i \alpha_i{}^k \alpha_i{}^{k'} = Np(p-1)/t(t-1)$; and $\Sigma^{\neq}_{ii'} \alpha_i{}^k \alpha_{i'}{}^k = Np(Np-t)/t^2$. The expectations of the products of the α_{ij}'s in the brackets given above are clearly equal to the associated probabilities. We shall use these expectations to examine least square statistics.

The Gauss-Markov normal equations (NE) for balanced cods are as follows

$$Np(\mu + s_{.} + \rho_{.} + \tau_{.}) = Y_{...} = \sum_{ijk} y_{ijk} \qquad 4.2.1.2$$

$$N(\mu + s_{.} + \rho_j + \tau_{.}) = Y_{.j.} = \sum_{ik} y_{ijk} \qquad 4.2.1.3$$

$$p(\mu + s_i + \rho_.) + \sum_k \alpha_i^k \tau_k = Y_{i..} = \sum_{jk} y_{ijk} \qquad 4.2.1.4$$

$$\frac{Np}{t}(\mu + \rho_. + \tau_k) + \Sigma_i \alpha_i^k s_i = Y_{..k} = \sum_{ij} y_{ijk} \qquad 4.2.1.5$$

where the summation is over the occurring combinations of subscripts,

Let $s_. = \frac{1}{N} \Sigma s_i$, $\rho_. = \frac{1}{p} \Sigma \rho_j$ and $\tau_. = \frac{1}{t} \sum_k \tau_k$.

Divide equation 4.2.1.2 by N and subtract the result from equation 4.2.1.4

to obtain $$p(s_i - s_.) + \Sigma_k \alpha_i^k (\tau_k - \tau_.) = Y_{i..} - \frac{Y_{...}}{N} \qquad 4.2.1.6$$

Let $\tau_k^* = \tau_k - \tau_.$, $s_i^* = s_i - s_.$, $P_j^* = P_j - P_.$

Divide equation 4.2.1.2 by t and subtract the result from equation 4.2.1.5

to obtain $$\frac{Np}{t} \tau_k^* + \Sigma_i \alpha_i^k s_i^* = Y_{..k} - \frac{Y_{...}}{t} \qquad 4.2.1.7$$

So eliminating $(s_k - s_.)$ from equation 4.2.1.7 will give us the reduced

normal equations for the τ_k^* as follows:

$$\frac{N(p-1)}{(t-1)} \tau_k^* = Y_{..k} - \frac{1}{p} \Sigma_i \alpha_i^k Y_{i..} = T_k^* \text{ , say} \qquad 4.2.1.8$$

or in matrix notation,

$$N(p-1)/(t-1) \left(I - \frac{1}{t} J\right) \tau = T^* .$$

Imposing the non-estimable constraints

$\Sigma_i s_i = 0 = \Sigma_k \tau_k^* = \Sigma_j \rho_j$, on the solutions of the NE will enable

us to obtain the following estimates of the τ_k^*'s:

$$\hat{\tau}_k = \widehat{(\tau_k - \tau_.)} = \frac{(t-1)}{N(p-1)} \left(Y_{..k} - \frac{1}{p} \Sigma_i \alpha_i^k Y_{i..} \right) , \qquad 4.2.1.9$$

k=1, 2, ..., t_n .

The reduction in sum of squares due to the τ_k ' s is then equal to

$$\frac{(t-1)}{N(p-1)} \Sigma_k T_k^{*2} \quad .$$

Now $T^* = (T_1^*, T_2^*, \ldots, T_k^*)'$ is the right hand side of the reduced NE and so may be written as

$$T^* = Z'(I - P_X)Y \quad ,$$

where Z is the incidence matrix for the treatment effects and P_X is the projection operator for the combined column spaces for the incidence matrices of periods and subject effects. It is easy to show that P_X has rank equal to $1 + N - 1 + p - 1 = N + p - 1$.

If the variance of Y has the form $\sigma^2 V = \sigma^2(I_N + bJ_{NN})$, where I_m is the identity matrix of dimension m x m and $J_{m,n}$ is the m x n matrix of ones, or if we assume GMN error structure, then the covariance matrix for the estimates $\hat{\tau} = (\tau_1, \ldots, \tau_k)'$ would be given by:

$$\text{Cov}\ (\tau) = \frac{(t-1)^2}{N^2(p-1)^2}\ \text{Cov}\ (Z'(I-P_X)Y)$$

$$= \frac{(t-1)^2}{N^2(p-1)^2}\ \sigma^2\ Z'(I - P_X)Z$$

Now $Z'(I - P_X)Z = \frac{N(p-1)}{(t-1)} (I - \frac{1}{t} J)$.

We then have the following results;

$$\text{Var}\ (\tau_k) = \frac{(t-1)^2}{N(p-1)}\ \frac{\sigma^2}{t} \qquad 4.2.1.10$$

$$\text{Cov}\ (\tau_k, \tau_{k'}) = \frac{-(t-1)}{N(p-1)}\ \frac{\sigma^2}{t} \qquad 4.2.1.11$$

so var $(\widehat{\tau_k - \tau_{k'}}) = \frac{2(t-1)}{N(p-1)} \sigma^2$ 4.2.1.12

When p = t we obtain var $(\widehat{\tau_k - \tau_{k'}}) = \frac{2}{N} \sigma^2$. These are the results obtained under GMN assumptions. Also by well known least squares theory the analysis of variance would consist of basically 3 independent sums of squares, $Y'P_XY$, $Y'P_{(Z-X)}Y$ and $Y'(I - P_X - P_{(Z-X)}Y$, where P_X, of rank $(1+p - 1 + N-1) = (N + p - 1)$, is the projection operator determined by the combined column spaces of the incidence matrices for period effects and experimental unit effects; $P_{(Z-X)}$, of rank (t-1), is the projection operator determined by the 'reduced' column space of Z, the incidence matrix for treatment effects, i.e., $P_{(Z-X)} = P_{Z^*}$, where $Z^* = (I - P_X)Z$, and $(I - P_X - P_{Z^*})$, of rank $(Np - N - p - t + 2)$, is the remainder which corresponds to the error space or residual error. The usual GMN tests are then used to make statistical inferences or decisions.

To consider randomization theory, we shall ignore technical error because the contribution of this is the same as under GMN assumptions. Under randomization theory, if additivity holds, since the treatment on the i-th unit in the j-th period is a random selection of t possible treatments, the observed response may be written as $u_{ij} + \Sigma_k \tau_k \alpha_{ij}^k$, where u_{ij} is the response of the i-th experimental unit in the j-th period under some basic conditions. So under randomization theory we have the following:

$$Y_{...} = U_{..} + Np\,\tau_{.} \qquad = Np(u_{..} + \tau_{.}) \qquad = Npy_{...}$$

$$Y_{i..} = U_{i.} + \Sigma_k\, \alpha_i^{\,k} \tau_k \qquad = pu_{i.} + \Sigma_k\, \alpha_i^{\,k} \tau_k \qquad = py_{i..}$$

$$Y_{.j.} = U_{.j} + N\tau_{.} \qquad = N(u_{.j} + \tau_{..}) \qquad = Ny_{.j.}$$

and
$$Y_{..k} = \Sigma_{ij}\, \alpha_{ij}^{\,k}\, u_{ij} + \frac{Np}{t}\tau_k \qquad = \frac{Np}{t} y_{..k}$$

The total sum of squares, TOT, is

$$\Sigma_{ij}\left((u_{ij} - u_{..}) + \Sigma_k\, \tau_k^{*}\, \alpha_{ij}^{\,k}\right)^2$$

$$= \Sigma_{ij}\left[(u_{ij} - u_{..})^2 + \Sigma_k\, \tau_k^{2*}\, \alpha_{ij}^{\,k} + 2\Sigma_k\, (u_{ij} - u_{..})\, \tau_k^{*}\, \alpha_{ij}^{\,k}\right]$$

$$= \Sigma_{ij}(u_{ij} + u_{..})^2 + \frac{Np}{t}\Sigma_k \tau_k^{*2} + 2\Sigma_{ij}\Sigma_k(u_{ij} - u_{..})\, \tau_k^{*}\, \alpha_{ij}^{\,k} \quad .$$

The row sum of squares, RSS, is

$$p\Sigma_i(y_{i..} - y_{..})^2 = p\, \Sigma_i\left[(u_{i.} - u_{..}) + \frac{1}{p}\Sigma_k\, \tau_k^{*}\, \alpha_i^{\,k}\right]^2$$

$$= p\, \Sigma_i(u_{i.} - u_{..})^2 + \frac{1}{p}\Sigma_i\Sigma_k\Sigma_{k'}\tau_k^{*}\tau_{k'}^{*}\alpha_i^k\alpha_i^{k'} + 2\Sigma_i\Sigma_k(u_{i.} + u_{..})\tau_k^{*}\alpha_i^k$$

The column sum of squares, CSS, is

$$N\Sigma_i(y_{.j} - y_{..})^2 = N\Sigma_j(u_{.j} - u_{..})^2$$

The reduced sum of squares for (treatment + error) is TOT - RSS - CSS

$$= \Sigma_{ij}(u_{ij} - u_{..})^2 - p\Sigma_i(u_{i.} - u_{..})^2 - N\Sigma_j(u_{.j} - u_{..})^2$$

$$+ \frac{Np}{t}\Sigma_k\, (\tau_k^{*2} + \frac{1}{p}\Sigma_i\Sigma_k\Sigma_{k'}\, \tau_k^{*}\tau_{k'}^{*}\, \alpha_i^{\,k}\alpha_i^{\,k'}\ N$$

$$+ 2\Sigma_i\Sigma_j\Sigma_k(u_{ij} - u_{..})\, \tau_k^{*}\, \alpha_{ij}^{\,k} - 2\Sigma_i\Sigma_k\, (u_{i.} - u_{..})\, \tau_k^{*}\, \alpha_i^{\,k}$$

$$= \frac{N(p-1)}{t-1}\Sigma_k\tau_k^{*2} + \Sigma_{ij}(u_{ij} - u_{i.} - u_{.j} + u_{..})^2 + G\ , \qquad 4.2.1.13$$

where G is equal to the expression given in the last line above.

Let $u_{ij}^* = u_{ij} - u_{.i} - u_{.j} + u_{..})$

Now $T_k^* = (Y_{..k} - \frac{1}{p}\Sigma_i\alpha_i^k Y_{i..})$

$$= \frac{Np}{t}\tau_k + \Sigma_{ij}u_{ij}\alpha_{ij}^k - \frac{1}{p}\Sigma_i\alpha_i^k u_{i.} - \frac{1}{p}\Sigma_i\alpha_i^k\Sigma_{k'}\alpha_i^{k'}\tau_{k'}$$

$$= \frac{Np}{t}\tau_k(1 - \frac{1}{p}) - \frac{Np(p-1)}{pt(t-1)}(t\tau_{.} - \tau_k) + \Sigma_{ij}(u_{ij} - u_{..})\alpha_{ij}^k$$

$$- \frac{1}{p}\Sigma(u_{i.} - u_{..})\alpha_i^k$$

$$= \frac{N(p-1)}{(t-1)}\tau_i^* + \Sigma_{ij}u_{ij}^*(\alpha_{ij}^k - \frac{1}{p}\alpha_i^k)$$

So, from the expectations of the design variables $\{\alpha_{ij}^k\}$, we have $E(T_k^* - T_{k'}^*) = \frac{N(p-1)}{(t-1)}(\tau_k - \tau_{k'})$, and $E(G) = 0$.

Also,

$$\frac{(t-1)}{N(p-1)}\Sigma_k T_k^* = \Sigma_k\left[\frac{N(p-1)}{(t-1)}\tau_k^{*2} + \frac{(t-1)}{N(p-1)}(\Sigma_{ij}u_{ij}^*(\alpha_{ij}^k - \frac{1}{p}\alpha_i^k))^2\right] + G$$

From Appendix D we have

$$\Sigma_k(\Sigma_{ij}u_{ij}^*(\alpha_{ij}^k - \frac{1}{p}\alpha_i^k))^2 = \frac{N}{(N-1)}\Sigma_{ij}u_{ij}^{*2}$$

So the residual error sum of squares in the analysis of variance is equal to

$$\Sigma_{ij}u_{ij}^{*2}(1 - (t-1)(M-1)(p-1))$$

$$= \Sigma_{ij}u_{ij}^{*2}\left[(N-1)(p-1) - (t-1)\right]/(N-1)(p-1)$$

The expected value of the treatment sum of squares is

$$\frac{N(p-1)}{(t-1)}\Sigma_k\tau_k^* + \Sigma u_{ij}^{*2}(t-1)/(N-1)(p-1)$$

There are (t-1) degrees of freedom for treatment effects and (N-1)(p-1)-(t-1) degrees of freedom for error. Hence the design allows us to obtain estimates of linear contrasts in the treatment effects with the correct expectations. Also the analysis of variance gives the same expected mean square under randomization for treatment effects as the residual error when there are no differential treatment effects.

Since each treatment is assigned at random to one of the t "letters" of the selected plan, the variances of any two T_k^* and $T_{k'}^*$ are the same and the covariance between them is $\frac{-1}{(t-1)}$ V where V is the common variance.

Now we have $\Sigma_k T_k^* = 0$

and

$$E(T_k^{*2}) = (N^2(p-1)^2/(t-1)^2 t)\tau_k^{*2} + (\frac{N}{(N-1)t}) \Sigma_{ij} u_{ij}^{*2} \quad .$$

So we may obtain unbiased estimates of contrasts in the τ_k as well as accurate estimates of their variances. Under GMN assumptions,

$$\mathrm{Var}(T_k^*) = \mathrm{Var}\,(Y_{..k} - \frac{1}{p}\Sigma_i \alpha_i Y_{i..}) = \frac{N(p-1)}{t}\sigma^2$$

and

$$\mathrm{Cov}\,(T_k^*, T_{k'}^*) = \frac{-N(p-1)}{t(t-1)}\sigma^2$$

In the finite model situation if we define σ_r^2 to be equal to $\Sigma_{ij} u_{ij}^{*2}$ / (N-1)(p-1), we will have the results:

$$\mathrm{Var}(T_k^*) = \frac{N(p-1)}{t}\sigma_r^2$$

and

$$\mathrm{Cov}\,(T_k^*, T_{k'}^*) = \frac{-N(p-1)}{t}\frac{\sigma_r^2}{(t-1)}$$

These results are similar in form to the GMN results. For $p \geq 4$, for any four units indexed by (i, i', I, I') and any four periods indexed by (j, j', J, J'), and any two treatments (k,k'), the expectation under randomization of the product α_{ij}^{k}, $\alpha_{i'j'}^{k'}$, $\alpha_{IJ}^{k'}$, $\alpha_{I'J'}^{k'}$ is the same in the balanced incomplete cod as in the corresponding cod. The variance of the treatment sum of squares under randomization theory is given by

$$\lambda^2 N^{-2}[E_R\ \{\Sigma_k(\Sigma_{ij}u_{ij}{}^*\alpha_{ij}^{k})^2\}^2 - \{\Sigma_k E_R(\Sigma_{ij}u_{ij}{}^*\alpha_{ij}^{k})^2\}^2]$$

where $\lambda = \begin{cases} 1 & \text{for the complete cod with } p=t \text{ periods} \\ \frac{t}{p} & \text{for the incomplete cod with } p < t \text{ periods.} \end{cases}$

Hence, the variance of the treatment sum of squares for the cod with $p < t$ periods will have the same form as that for the complete cod with a multiplicative factor f, $f = t^2/p^2$.

Welch (1937) has given an expression for the variance under randomization of the ratio of the treatment sum of squares, to the sum of the treatment sum of squares and the residual error sum of squares. The result is somewhat complicated and is in general cannot be easily related to its GMN counterpart. We give this result in Appendix D.

4.2.2 Change Over Designs in the Presence of First Order Residual Effects

We have seen that with cross-over designs and with any treatments having only an additive effect in the period in which it is applied to the subject, the ordering GMN theory has a strong justification in terms of randomization. Estimated treatment effects are unbiassed and the usual analysis of variance is unbiassed with respect to treatments, where the population of repetitious is merely the randomization set and there are no assumptions about the nature of the subject responses under a uniform treatment.

This argumentation has additional value in that if one were sampling subjects at random, one would expect the subject responses to have a covariance matrix that is not $\sigma^2 I$ for some σ^2. We can easily envisage in this context that there may be serial correlation over the periods and also that variance will change with period.

We now turn to the model in which we assume that treatments have direct effects in the period in which they are applied but also have a residual effect in the following period. The question we address is the validity in randomization terms if the ordinary statistics derived from Gauss-Markoff assumptions.

We now present a development of results already placed in the literature by Cochran et al. (1941), Patterson (1950), Lucas (1951), Sharma (1977) as necessary background for the randomization analysis that follows.

4.2.3 Least Squares Analysis

In the standard assumed model, let y_{hiju} denote the response on the u-th subject to which treatment i is applied in the h-th period and treatment j is applied in the preceding period. Then if we assume additivity between period treatment direct, treatment residual, and subject effects we will have a model

$$y_{hiju} = \mu + \theta_h + \tau_i + r_j + s_u + e_{hiju} \quad ,$$

where μ is the grand mean, θ_h is the effect of the h-th period, τ_i the direct effect of treatment i, r_j the residual effect of treatment j, and

s_u is the effect of the u-th subject.

The matrix form of the model is then given by

$$\underline{y} = \mu + X_\theta \theta + X_\tau \tau + X_r r + X_s s + e \quad ,$$

where $\underline{y}$ is the N vector of observed responses. The matrices $X_\theta, X_\tau, X_r, X_s$ are incidence matrices of N rows; θ, τ, r, s are vectors of effects defined above. The element $x_{(.)ij}$ is the number of times the j-th component of the corresponding effect appears in the i-th observation and $x_{ij} = 0$ or 1. $\mathfrak{J}$ is an N vector of ones, and $\underline{e}$ is an N vector of errors, usually assumed to be $N(0,\sigma^2 I)$.

For example if θ_2 , τ_3 appears in the 5th observation we will have $X_\theta\,(5,2) = 1$; $X_\theta\,(5,j) = 0 \quad j \neq 2$; $X_\tau(5,3) = 1$, $X_\tau(5,j) = 0 \; j \neq 3$, etc.

Designs that are balanced for sequences give the following normal equations

$$Np\,\mu + p\,\Sigma s_i + \Sigma_i \theta_i + \frac{Np}{t}\,\Sigma_k t_k + \frac{N(p-1)}{t}\,\Sigma_k r_k = G \qquad (1)$$

$$p\,\mu + ps_i + \Sigma_j\,\theta_j + \Sigma_k w_{ik} t_k + \Sigma_k z_{ik} r_k = S_i \qquad (2)$$

$$N\,\mu + \Sigma_i s_i + N\theta_1 + \frac{N}{t}\,\Sigma_k t_k = P_1 \qquad (3A)$$

$$N\,\mu + \Sigma_i s_i + N\theta_j + \frac{N}{t}\;{}_k t_k + \frac{N}{t}\,\Sigma_k r_k = P_j, \quad j > 1 \qquad (3B)$$

$$\frac{NP}{t}\,(\mu + \tau_k) + \frac{N(p-1)}{t(t-1)}\,(\Sigma_j r_j - r_k) + \frac{N}{t}\,(\Sigma_i \theta_i + \Sigma_i w_{ik} = T_k \qquad (4)$$

$$\frac{N(p-1)}{t}\,\{(\mu + r_k) + \frac{1}{(t-1)}\,(\Sigma_i \tau_i - \tau_k)\} + \frac{N}{t}\,(\Sigma_i \theta_i - \theta_1) + \Sigma_i z_{ik} s_i = R_k \qquad (5)$$

where

w_{ik} = 1 if direct treatment effect k is applied to subject i ,
= 0 otherwise,

z_{ik} = 1 if residual treatment effect k occurs on subject i ,
= 0 otherwise,

Hence $\Sigma_i w_{ik} = \frac{Np}{t}$; $\Sigma_i z_{ik} = \frac{N(p-1)}{t}$.

The NE in matrix form are given by the following equations.

$$\begin{array}{c}
\begin{array}{ccccc}
\mu & s & \theta & \tau & r
\end{array} \\
\begin{bmatrix}
Np & p\mathcal{J}'_N & N\mathcal{J}'_p & \frac{Np}{t}\mathcal{J}'_t & \frac{N(p-1)}{t}\mathcal{J}'_t \\
p\mathcal{J}_N & pI_N & J_{Np} & W & Z \\
N\mathcal{J}_p & J_{pN} & NI_p & \frac{N}{t}J_{pt} & \frac{N}{t}\begin{bmatrix}\phi \\ J_{p-1,t}\end{bmatrix} \\
\frac{Np}{t}\mathcal{J}_t & W' & \frac{N}{t}J_{tp} & \frac{Np}{t}I_t & \frac{N(p-1)}{t(t-1)}(J-I_t) \\
\frac{N(p-1)}{t}\mathcal{J}_t & Z' & \frac{N}{t}[\phi, J_{t,p-1}] & \frac{N(p-1)}{t(t-1)}(J-I_t) & \frac{N(p-1)}{t}I_t
\end{bmatrix}
\end{array}
\begin{bmatrix} \mu \\ s \\ \theta \\ \tau \\ r \end{bmatrix} = \begin{bmatrix} G \\ S \\ P \\ T \\ R \end{bmatrix}$$

where

N is the number of subjects and p the number of periods,

G is the grand total or sum of all observed responses,

P is the vector of totals for the respective periods, T, R, S are similarly defined for direct treatment effects, residual effects and subject effects, respectively, $\mathcal{J}_n$ is the column vector of n ones, $J_{m,n}$ is the matrix of ones with m rows and n columns, W is a t x N incidence matrix, such that w_{ij} is the number of times the i-th subject receives the j-th direct treatments, i.e. w_{ij} = 0 or 1. Z is similarly defined for residual effects, D = W S is a t-vector of the total responses of all subjects receiving the respective direct treatments, $F = Z\,S$ is a t vector of the total responses of all subjects receiving the respective residual treatment effects, i.e. the sum of the totals of all subjects receiving the respective treatments except in the final period. We note that W and Z, hence D and F are random outcomes with properties depending on the randomization scheme

The columns of each of the submatrices X_θ, X_τ, X_r, X_s sum to $\mathcal{J}_N$ and the sum of the rows of the NE is equal to a constant multiplied by the first row.

It is easily seen that the normal equations

$$(X'X\beta = X'y) \text{ have rank } r \leq 1 + p + t + t + N - 4$$
$$= p + 2t + N - 3 \quad .$$

4.2.3.1 <u>The Reduced Normal Equations for τ and r</u>

If we multiply equation 2 by $\frac{1}{p} w_{ik}$ and sum over i we will obtain

$$\frac{Np}{t}\mu + \Sigma_i w_{ik}\, s_i + \frac{N}{t}\Sigma_i \theta_i =$$
$$\Sigma_i \frac{1}{p} w_{ik} \{ - \Sigma_j w_{ij}\tau_j - \Sigma_j z_{ij} r_j + S_i \} .$$

Subtracting equation (3A) from (3b) will give us

$$N(\theta_i - \theta_1) = P_i - P_1 - \frac{N}{t}\Sigma_j r_j \quad .$$

From equations (5) and (2) we obtain

$$\frac{N(p-1)}{t}\mu + \Sigma_j z_{jk}\, s_j + \frac{N}{t}(\Sigma\,\theta_j - \theta_1) \qquad i=2, \ldots, p$$

$$= \frac{N(p-1)}{t}\mu + \Sigma_i z_{ik}\, s_i + \frac{N(p-1)}{pt}\Sigma_j \theta_j + \frac{N}{pt}[\Sigma_j \theta_j - p\,\theta_1]$$

$$= \frac{1}{p}\Sigma_i z_{ik}(S_i - \Sigma_j w_{ij}\tau_j - \Sigma_j z_{ij} r_j) + \frac{1}{pt}(\Sigma_j P_j - pP_1 - (p-1)\frac{N}{t}\Sigma_j r_j).$$

Also $\Sigma_i w_{ik} = \frac{Np}{t}$; $\Sigma_i w_{ik} w_{ik'} = \frac{Np}{t} \frac{(p-1)}{(t-1)} \qquad k \neq k'$;

$$\Sigma_i w_{ik} z_{ik} = \frac{N(p-1)}{t} \quad ; \quad \Sigma_i w_{ik} z_{ik'} = \frac{Np-1}{t} \frac{(p-1)}{(t-1)} \quad k \neq k' \; ;$$

$$\Sigma_i z_{ik} = \frac{(p-1)}{t} \quad ; \quad \Sigma_i z_{ik} z_{ik'} = \frac{(p-1)}{t} \frac{(p-2)}{t-1}$$

Substituting in equations (4) and (5) we obtain

$$\frac{Np}{t}\tau_k + \frac{N(p-1)}{t(t-1)}(\Sigma r_j - r_k) + \frac{1}{p} D_k - \frac{N}{t}\left(1 - \frac{(p-1)}{t-1}\right)\tau_k - \frac{N}{t}\frac{(p-1)}{(t-1)} \Sigma\tau_j$$

$$- \frac{N(p-1)}{pt}\left(1 - \frac{p-1}{t-1}\right)r_k - \frac{N(p-1)(p-1)}{pt(t-1)} \Sigma r_j = T_k$$

i.e. $\frac{N(p-1)}{t-1} \tau_k - \frac{N(p-1)}{t-1} \frac{1}{t} \Sigma\tau_j - \frac{N(p-1)}{p(t-1)} r_k + \frac{N(p-1)}{p(t-1)} \frac{1}{t} \Sigma r_j$

$$= T_k - \frac{1}{p} D_k$$

Also, $\frac{N(p-1)}{t(t-1)}(\Sigma_j \tau_j - \tau_k) + \frac{N(p-1)}{t} r_i + \frac{1}{p} F_k - \frac{N(p-1)}{pt}\left(1 - \frac{(p-1)}{t-1}\right)\tau_k$

$$- \frac{N(p-1)(p-1)}{pt(t-1)} \Sigma\tau_j - \frac{N(p-1)}{pt(t-1)}(t-p+1) r_k - \frac{N(p-1)(p-2)}{pt(t-1)} \Sigma r_j$$

$$+ \frac{1}{pt}\left(G - pP_1 - (p-1)\frac{N}{t}\Sigma r_j\right) = R_k$$

i.e. $-\frac{N(p-1)}{p(t-1)} \left\{ \left(\tau_k - \frac{1}{t}\Sigma_j \tau_j\right) + \frac{(pt - t - 1)}{t}\left(r_k - \frac{1}{t}\Sigma_j r_j\right) \right\}$

$$= R_k - \frac{G}{pt} + \frac{P_1}{t} - \frac{1}{p} F_k \quad .$$

The reduced normal equations for τ and r are given by the following matrix equations

$$\frac{N(p-1)}{t-1}\begin{bmatrix} I - \frac{1}{t}J & -\frac{1}{p}(I - \frac{1}{t}J) \\ -\frac{1}{p}[I - \frac{1}{t}J] & \frac{(pt-t-1)}{pt}(I - \frac{1}{t}J) \end{bmatrix}\begin{bmatrix} \hat{\tau} \\ \hat{r} \end{bmatrix} = \begin{bmatrix} T - \frac{1}{p}D \\ R - \frac{1}{p}F + (\frac{P_1}{t} - \frac{G}{pt})t \end{bmatrix}$$

$$= \begin{pmatrix} T^* \\ R^* \end{pmatrix}, \text{ say.}$$

The minimum sum of squares for the full model is

$$(y-X\hat{\beta} - Z\hat{\delta})'(y-X\hat{\beta} - Z\hat{\delta}) = y'y - \hat{\beta}'X'y - \hat{\delta}'Z'y$$

$$= y'y - [P_X(y - Z\hat{\delta})]'y - \hat{\delta}'Z'y$$

$= y'y - y'P_Xy - \hat{\delta}'Z'(I-P_X)y$, where P_X is the projection operator defined by X, the matrix of period and subject effects, β, and Z represents the incidence of direct and residual treatment effects, δ.

Hence the reduction in the sum of squares due to the inclusion of τ and π in the least squares fit is given, as in elementary theory, by

$$\delta'Z'(I - P_X)y = (\tau'T^* + r'R^*)$$

It is easy to show that the reduced normal equations have rank

$$r = t-1 + t-1 = 2(t-1)$$

Impose the non-estimable constraints

$$\Sigma\tau_i = 0, \quad \Sigma_i r_i = 0$$

Then $Jr = \phi = Jt$

A set of solutions to the reduced normal equations is then given by the solutions to the equations

$$\begin{bmatrix} aI & bI \\ bI & cI \end{bmatrix} \begin{bmatrix} \tau \\ \pi \end{bmatrix} = \begin{bmatrix} T^* \\ R^* \end{bmatrix}.$$

where $a = \frac{N(p-1)}{(t-1)}$, $b = -\frac{N(p-1)}{p(t-1)}$;

and $c = \frac{N(p-1)}{(t-1)} \frac{(pt-t-1)}{pt}$.

Let

$$A = \begin{bmatrix} aI & bI \\ & \\ bI & cI \end{bmatrix} ,$$

and assume A^{-1} has the form

$$A^{-1} = \begin{bmatrix} xI & wI \\ & \\ wI & qI \end{bmatrix} .$$

Then

$$ax + bw = 1 ,$$

$$aw + bq = 0 ,$$

$$bx + cw = 0 ,$$

and

$$bw + cq = 1 .$$

Hence,

$$x = c/(ac - b^2) ,$$

$$q = a/(ac - b^2) ,$$

and

$$w = -b/(ac - b^2) .$$

The solution to the NE are then given by

$$\hat{\begin{bmatrix} \tau \\ r \end{bmatrix}} \qquad \frac{1}{(ac-b^2)} \cdot \begin{bmatrix} cT^* - bR^* \\ aR^* - bT^* \end{bmatrix}$$

From the reduced normal equations we have

$$a(\widehat{\tau_i - \tau_j}) + b(\widehat{r_i - r_j}) = T_i^* - T_j^* \quad ,$$

and

$$b(\widehat{\tau_i - \tau_j}) + c(\widehat{r_i - r_j}) = R_i^* - R_j^* \quad .$$

Solving, we obtain

$$(\widehat{\tau_i - \tau_j}) = \frac{1}{(ac - b^2)} \{c(T_i^* - T_j^*) - b(R_i^* - R_j^*)\} \quad ,$$

$$(\widehat{r_i - r_j}) = \frac{1}{(ac - b^2)} \{ - b\,(T_i^* - T_j^*) + a(R_i^* - R_j^*) \} \quad .$$

So unbiased estimators of treatment differences and other estimable contrasts may be obtained from the solutions of the NE's given above.

Also from least squares theory we have

$$\binom{T^*}{R^*} = Z'(\,I - P_X\,)y \quad ,$$

So $\mathrm{var}(A^{-1}\binom{T^*}{R^*}) = A^{-1}Z'(I-P_X)V(I-P_X ZA^{-1})$, where V is the variance of y. Suppose $V = \sigma^2(I + J)$. J is clearly in the column space of the incidence matrix for periods and subjects. So $P_X J + J$,

and $\quad \mathrm{var}(A^{-1}\binom{T^*}{R^*})$

$$= \frac{\sigma^2}{(ac-b^2)^2} \left[\begin{array}{c|c} cI & -bI \\ \hline -bI & aI \end{array}\right] \left[\begin{array}{c|c} a(I - 1/t\,J) & b(I - 1/t\,J) \\ \hline b(I - 1/t\,J) & c(I - 1/t\,J) \end{array}\right] \left[\begin{array}{c|c} cI & -bI \\ \hline -bI & aI \end{array}\right]$$

$$= \frac{\sigma^2}{(ac-b^2)^2} \left[\begin{array}{c|c} (ac-b^2)\,(I - 1/t\,J) & \phi \\ \hline \phi & (ac-b^2)(I - 1/t\,J) \end{array}\right] \left[\begin{array}{cc} cI & -bI \\ \hline -bI & aI \end{array}\right]$$

.

$$= \frac{\sigma^2}{(ac-b^2)} \cdot \begin{bmatrix} c(I - 1/t\ J) & -b(I - 1/t\ J) \\ -b(I - 1/t\ J) & a(I - 1/t\ J) \end{bmatrix}$$

Hence, under Gauss-Markov assumptions we have the following results:

$$\mathrm{var}\,(\hat{\tau}_i) = \sigma^2\, c(1-1/t) \quad ,$$

$$\mathrm{var}\,(\hat{r}_i) = \sigma^2\, a(1-1/t) \quad ,$$

$$\mathrm{cov}\,(\hat{\tau}_i, \hat{\tau}_j) = -\,\sigma^2\, c(1/t) \quad ,$$

$$\mathrm{cov}\,(\hat{r}_i, \hat{r}_j) = -\,\sigma^2\, a(1/t) \quad ,$$

$$\mathrm{cov}\,(\hat{\tau}_i, \hat{r}_i) = -\,\sigma^2\, b(1-1/t) \quad ,$$

$$\mathrm{cov}\,(\hat{\tau}_i, \hat{r}_j) = \sigma^2\, b(1/t) \quad , \qquad j \neq i \quad .$$

Hence

$$\mathrm{var}\,(\widehat{\tau_i - \tau_j}) = \sigma^2\, 2c/(ac - b^2)$$

$$\mathrm{var}\,(\widehat{r_i - r_j}) = \sigma^2 2a/(ac - b^2)$$

$$\mathrm{cov}[(\widehat{\tau_i - \tau_j}), (\widehat{r_i - r_j})] = -\,\sigma^2 2b/(ac-b^2)$$

4.2.3.2 Time Trends and Serial Correlations

We shall now apply randomization ideas to situations in which there may be serial correlations, trends with time and different variances among the responses for the different time periods.

It is appropriate first to review quickly the work of Patterson (1950) and Lucas (1951). Their approach was as follows. Consider a vector $u' = (u_1, u_2, \ldots, u_p)$ that represents the responses of a unit under a standard treatment. Then consider the mean of the elements, the linear in time contrast, the quadratic in time contrast and so on. Patterson and Lucas assumed that these are independent in the population of subjects. This can be represented as follows:

Let

$$B_0' = \frac{1}{\sqrt{p}} \mathcal{J}_p'$$

$$B_1' = (b_{11}, \ldots, b_{1p})$$

$$B_2' = (b_{21}, \ldots, b_{2p})$$

$$B_{pi}' = (b_{pi,1}, \ldots, b_{pi,p})$$

represent the vectors representing the mean, the linear trend, the quadratic trend and so on. We may take all of these to be normalized. Then if we write

$$B = (B_0, B_1, \ldots, B_{p-1})$$

Patterson and Lucas assumed that if V is the variance matrix of the p-vector u, then

$$B'VB = \text{diag}\begin{pmatrix} \sigma_0^2 & & & & \\ & \sigma_1^2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \sigma_{p-1}^2 \end{pmatrix}$$

or

$$V = BDB'$$

Patterson and Lucas were then able to obtain variance formulae, and expectations of mean squares of the usual analysis of variance.

A problem is that the assumption on V cannot be justified. We shall consider, first, the conjoining of randomization ideas to the sampling of subjects from a population having responses with a variance matrix V.

The randomization procedure may be defined by the following random variables:

$\alpha_{ij}^{k} = 1,$ if subject i receives treatment k in period j

$= 0,$ otherwise.

$\alpha_{i.}^{k} = 1,$ if subject i receives treatment k in some period

$= 0,$ otherwise $\quad k = 1, \ldots, t; \quad i = 1, \ldots, N$

$\alpha_{i.}^{k} = \Sigma_j \alpha_{ij}^{k}, \quad = w_{ik}$ as previously defined.

Let β_{ij}^{k} and $\beta_{i.}^{k}$ be similarly defined for the residual effects of treatment k. Then $\beta_{i.}^{k} = z_{ik}$ which was previously defined.

The basic properties of these random variables are given in appendix 4. Suppose the p-vector of observations $y_i = [y_{i1}, y_{i2}, \ldots, y_{ip}]$ on the i-th experimental unit has variance equal to $\sigma^2 V$. Let $Q = [Q_1, Q_2, \ldots, Q_p]$ be an orthogonal matrix of eigenvectors of V such that $Q'VQ = D$ and D is a diagonal matrix of eigenvalues of V.

Then we have $y_i = QQ'y_i$. Also with λ, and λ^* fixed px1 vectors, we have var $(\lambda' y_i) = \lambda' Q$ Var $(Q' y_i)\, Q'\lambda$,

$$\text{var} \begin{bmatrix} Q_1' y_i \\ \vdots \\ Q_p' y_i \end{bmatrix} = \text{diag} \qquad \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_p^2 \end{bmatrix} = \sigma_p^2 = \Sigma$$

$$\lambda' QDQ' \lambda = [\, \lambda' Q_1, \lambda' Q_2, \ldots, \lambda' Q_p] \begin{bmatrix} \sigma_2^2\, Q_2' \lambda^* \\ \vdots \\ \sigma_p^2\, Q_p'^{*} \end{bmatrix}$$

$$= \sum_{u=1} \sigma_u^2 \, (\lambda' Q_u)(\lambda^{*\prime} Q_u) \quad .$$

For any two treatments k and k' and any unit, i, say, let

$$\lambda_{ij} = \alpha_{ij}^k - \alpha_{i.}^k \,/\, p - (\alpha_{ij}^{k'} - \alpha_{i.}^{k'} \,/\, p) \;,$$

$$\lambda_{ij}^* = \beta_{ij}^k - \beta_{i.}^k / \, p - (\beta_{ij}^{k'} - \beta_{i.}^{k'} \,/\, p) \;,$$

$$\lambda_i' = (\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{ip}) \;,$$

and

$$\lambda_i^* = (\lambda_{ij}^*, \lambda_{i2}^*, \ldots, \lambda_{ip}^*) \quad .$$

Then in the j-th period the i-th unit has a contribution to $T_k^* - T_{k'}^*$ equal to $\lambda_{ij} y_{ij}$, and a contribution to $R_k^* - R_{k'}^*$ equal to $\lambda_{ij}^* y_{ij}$. So the contribution of that unit to the covariance matrix of $(T_k^* - T_{k'}^*, R_k^* - R_{k'}^*)'$ is

$$\begin{bmatrix} \lambda_i^* \\ \lambda_i^{*\prime} \end{bmatrix} \{Q \Sigma Q' (\lambda_i, \lambda_i^*)\}.$$

Now $E(\lambda_i' Q_u)(\lambda_i^{*\prime} Q_u)$

$$= E \, \Sigma_j \Sigma_{j'} Q_{ju} Q_{j'u} \{[\, \alpha_{ij}^k \beta_{ij'}^k - \alpha_{ij}^k \beta_{ij'}^k - \beta_{ij'}^k \alpha_{ij}^{k'} + \alpha_{ij}^k \beta_{ij'}^{k'}]$$

$$- \frac{1}{p} [\alpha_{i.}^k \beta_{ij'}^k + \beta_{ij'}^{k'} \alpha_{i.}^{k'} - \alpha_{i.}^{k'} \beta_{ij'}^k - \alpha_{i.}^k \beta_{ij'}^{k'} + \beta_{i.}^k \alpha_{ij}^k + \beta_{i.}^{k'} \alpha_{ij}^{k'}$$

$$- \beta_{i.}^{k'} \alpha_{ij}^k - \beta_{i.}^k \alpha_{ij}^{k'}] + \frac{1}{p^2} [\alpha_{i.}^k \beta_{i.}^k + \alpha_{i.}^{k'} \beta_{i.}^{k'} - \alpha_{i.}^k \beta_{i.}^{k'} - \alpha_{i.}^{k'} \beta_{i.}^k]\}$$

$$= \sum_{1 \le j \le p-1} Q_{ju} \Sigma_{j+1,u} \frac{2}{t} - \sum_{1 \ne j' \ne j+1} Q_{ju} Q_{j'u} \cdot \frac{2}{t} \cdot \frac{1}{(t-1)}$$

$$- \frac{1}{p} \Sigma_j Q_{ju} \sum_{j' \ne 1} Q_{j'u} [\, \frac{2}{t} - \frac{2(p-1)}{t(t-1)} \,] - \frac{2}{p} \sum_{j \ne p} Q_{ju} \, \Sigma \, \Sigma_{j'u} \cdot \frac{1}{t}$$

$$+ \frac{2}{p} \sum_{j \ne p} \Sigma_{j'} Q_{ju} Q_{j'u} \frac{1}{t} \frac{(p-2)}{t-1} + \frac{2}{p} \Sigma_{j'} Q_{pu} Q_{j'u} \frac{1}{t} \cdot \frac{(p-1)}{t-1}$$

$$+ \frac{1}{p^2} \Sigma \; Q_{ju} \Sigma \; Q_{j'u} \left[2 \frac{p-1}{t-1} - 2 \frac{(p-1)}{t} \cdot \frac{(p-1)}{(t-1)}\right] .$$

After simplification we obtain

$$E(\lambda' Q_u)(\lambda^{*}' Q_u) = \frac{2}{(t+1)} \{ \Sigma Q_{ju} Q_{j+1,u} + (Q_{1u} + \frac{1}{p} Q_{pu}) \Sigma Q_{ju}$$

$$- \frac{(p+1)}{p^2} (\Sigma Q_{ju})^2 \} .$$

In a similar manner we obtain

$$E(\lambda' Q_u)^2 = \frac{2}{(t-1)} \{ \Sigma Q_{ju}^2 - \frac{1}{p} (\Sigma \, Q_{ju})^2 \} ,$$

and

$$E(\lambda^{*}' Q_u)^2 = \frac{2}{(t-1)} \{ \Sigma Q_{ju}^2 - \frac{(t+1)}{t} Q_{1u}^2 - \frac{(pt+t+1)}{p^2 t} (\Sigma Q_{ju})^2$$

$$+ \frac{(t+1)}{pt} Q_{1u} \Sigma Q_{ju} \} .$$

This line of development can be pursued to the point of examining the nature of expected mean squares. We shall not do this, however, We shall merely note that if Q is such that $Q_1 = p^{-\frac{1}{2}} Q_{\mathcal{J}}$ and Q_2, Q_3, etc. correspond to the linear contrasts, quadratic contrasts and so on, then

$$\Sigma_j Q_{ju} = Q_{\mathcal{J}}' Q_u = 0 \text{ for } u > 1.$$

The results of Patterson (1950) and Lucas (1951) may then be obtained from the expectations given above. However, there seems to be no necessary or even plausible reasons why these conditions should be the case.

4.2.4 Expectations Under Randomization Only

In the finite additive model situation in which we take account of possible technical errors we may write the observation on the i-th subject in the j-th period as

$$u_{ij} + \Sigma_k \alpha_{ij}^k \tau_k + \Sigma_{k'} \beta_{ij}^{k'} r_{k'} + e_{ij}$$

where we assume the e_{ij} are independent of the u_{ij} and are independently and identically distributed with zero mean and common variance equal to σ^2. We will then have

$$T_k^* = \Sigma_i\Sigma_j \{\alpha_{ij}^k (\mu_{ij} + \tau_k + \Sigma_{k'} \beta_{ij}^{k'} r_{k'} + e_{ij})$$

$$- \frac{1}{p}\alpha_i^k \{ u_{ij} + \Sigma_{k'}\alpha_{ij}^{k'} \tau_{k'} + \Sigma_{k'} \beta_{ij}^{k'} r_{k'} + e_{ij}\}$$

$$= \Sigma_i \alpha_i^k \Sigma_j [\alpha_{ij}^k \tau_k - \frac{1}{p}\Sigma_{k'} \alpha_{ij}^{k'} \tau_{k'}]$$

$$+ \Sigma_i\alpha_i^k \Sigma_j(\Sigma_{k'}\alpha_{kj}^k\beta_{ij}^{k'} r_{k'} - \frac{1}{p}\Sigma_{k'}\beta_{ij}^{k'} r_{k'}$$

$$+ \Sigma_i\Sigma_j (\alpha_{ij}^k - \frac{1}{p} \alpha_i^k)(e_{ij} + u_{ij})$$

$$= N p [(\frac{1}{t} - \frac{1}{pt} + \frac{1}{pt} \cdot \frac{(p-1)}{t-1}) \tau_k - \frac{1}{pt} \frac{(p-1)}{t-1} \Sigma_{k'} \tau_{k'}]$$

$$- N(p-1) [(\frac{1}{t(t-1)} + \frac{1}{pt} - \frac{1}{pt}\frac{(p-1)}{(t-1)}) r_k$$

$$- (\frac{1}{t(t-1)} - \frac{(p-1)}{pt(t-1)}) \Sigma_{k'} r_{k'} + \epsilon_{T_k^*}$$

$$= NP [\frac{t(p-1)}{pt(t-1)} \tau_k - \frac{(p-1)}{pt(t-1)} \Sigma_{k'}\tau_{k'}]$$

$$- N(p-1) [t \frac{1}{pt(t-1)} r_k - \frac{1}{pt(t-1)} \Sigma_{k'} r_{k'}] + \epsilon_{T_k^*}$$

where

$$\epsilon_{T_k^*} = \Sigma_i\Sigma_j (\alpha_{ij}^k - \frac{1}{p}\alpha_i^k) (e_{ij} + u_{ij}) .$$

Let $\delta_j^m = 1$, if $j = m$,

$= 0$, otherwise .

Then

$$R_k^* = \Sigma_i\Sigma_j \{(\beta_{ij}^k - \frac{1}{p}\beta_i^k - \frac{1}{pt} + \frac{1}{t}\delta_{ij}) \times$$

$$(u_{ij} + \Sigma_{k'}\alpha_{kj}\,\tau_{k'} + \Sigma_{k'}\beta_{ij}^{k'}\,r_{k'} + e_{ij}\,)\}$$

$$= \Sigma_i\,\beta_i^k\,\Sigma_j\,\{\beta_{ij}^{k}\,r_k - \frac{1}{p}\,\Sigma_{k'}\,\beta_{ij}^{k'}\,r_{k'}\,\}$$

$$+ \Sigma_i\,\beta_i^k\,\Sigma_j\,\{\Sigma_{k'}\,\beta_{ij}^{k}\,\alpha_{ij}^{k'}\,\tau_{k'} - \frac{1}{p}\,\Sigma_{k'}\,\alpha_{ij}^{k'}\,\tau_{k'}\,\}$$

$$- \frac{1}{pt}\,\frac{Np}{t}\,\Sigma_{k'}\,t_{k'} - \frac{1}{pt}\,\frac{N(p-1)}{t}\,\Sigma_{k'}\,r_{k'} - \frac{1}{pt}\,\Sigma_i\Sigma_j\,(e_{ij} + u_{ij})$$

$$+ \frac{N}{t^2}\,\Sigma_{k'}\tau_{k'} + \frac{1}{t}\,\Sigma_i\,(e_{i1} + u_{i1})$$

$$+ \Sigma_i\Sigma_j\,(\beta_{ij}^{k} - \frac{1}{p}\,\beta_i^k\,)\,(e_{ij} + u_{ij})$$

$$= N(p-1)\,(\,\frac{1}{t} + \frac{1}{pt}\,\frac{(p-2)}{t-1} - \frac{1}{pt}\,r_k - \frac{1}{pt}\,(\,\frac{(p-2)}{t-1} + \frac{1}{t}\,)\,\Sigma_{k'}\,r_{k'})$$

$$- \{(\,\frac{1}{t(t-1)} + \frac{1}{pt} - \frac{1}{pt}\,\frac{1}{(t-1)} - \frac{1}{pt}\,\frac{(p-2)}{t-1}\,\tau_k$$

$$- (\,\frac{1}{t(t-1)} - \frac{1}{pt}\,\frac{1}{t-1} - \frac{1}{pt}\,\frac{(p-2)}{t-1}\,)\,\Sigma_{k'}\,\tau'_{k'}\,\} + \epsilon_{R_k}^{*}\,,$$

where

$$\epsilon_{R_k}{}^{*} = \Sigma_i\Sigma_j\,(\beta_{ij}^{k} - \frac{1}{p}\,\beta_i{}^{k} + \frac{1}{t}\,\delta_j{}^{1} - \frac{1}{pt}\,)\,(\,e_{ij} + u_{ij}).$$

Now from the properties of the α_{ij}^{k} 's and β_{ij}^{k} 's we have

$$E(\epsilon_{T_k}{}^{*}) = 0 = E(\epsilon_{R_k}{}^{*})\,.$$

Hence

$$E(T_k{}^{*}) = \frac{N(p-1)}{pt(t-1)}\,[pt\,(\tau_k - \frac{1}{t}\,\Sigma_{k'}\,\tau_{k'}) - t(r_k - \frac{1}{t}\,\Sigma_{k'}\,r_{k'})]\,,$$

$$E(R_k{}^{*}) = \frac{N(p-1)}{pt(t-1)}\,[(pt - t - 1)(r_k - \frac{1}{t}\,\Sigma_{k'}\,r_{k'}) - t(\tau_k - \frac{1}{t}\,\Sigma_{k'}\tau_{k'})\,,$$

$$E\,\{(pt - t - 1)T_k{}^{*} - t\,R_k{}^{*}\} = A(\tau_k - \frac{1}{t}\,\Sigma_{k'}\,\tau_{k'})\,,$$

where $$A = \frac{N(p-1)}{pt(t-1)}\,t(p^2t - pt - t - p)\,,$$

$$E\{ptR_k^* - tT_k^*\} = A\,(r_k - \frac{1}{t}\Sigma_{k'}\, r_{k'})\quad,$$

and

$$E\{(pt-1)\,T_k^* + (pt+t)\,R_k^*)$$

$$= B[\,\tau_k + r_k - \frac{1}{t}\Sigma_{k'}\,(\tau_{k'} + r_{k'})]\;,$$

where $$B = \frac{N(p-1)}{pt(t-1)} \cdot t(p^2t - pt - p - t)\;.$$

So we have the property that the estimators of direct and residual effects given by ordinary least squares are unbiased under randomization.

So $\quad E_R(\widehat{\tau_k - \tau_{k'}}) = \tau_k - \tau_{k'}$,

and $\quad E_R(\widehat{r_k - r_{k'}}) = r_k - r_{k'}$.

Since $\Sigma_k T_k^* = 0 = \Sigma_k R_k^*$,

we have

$$\text{cov}\,(T_k^*, T_{k'}^*) = -\frac{1}{(t-1)}\quad \text{var}\,(T_k^*)\;,$$

$$\text{cov}\,(R_k^*, R_{k'}^*) = -\frac{1}{(t-1)}\quad \text{var}\,(R_k^*)\;,$$

$$\text{cov}\,(T_k^*, R_{k'}^*) = -\frac{1}{(t-1)}\quad \text{cov}\,(T_k^*, R_k^*)\,,$$

and $$\text{var}\; f(T_k^* - T_{k'}^*) + g(R_k^* - R_{k'}^*)$$

$$= \frac{2t}{(t-1)}\{f^2\,\text{var}\,(T_k^*) + g^2\,\text{Var}\,(R_k^*) + 2fg\,\text{cov}\,(T_k^*, R_k^*)\}\;.$$

Given the outcome of the randomization, if we assume that the technical errors e_{ij} have GMN properties, the treatment sum of squares and error sum of squares would contain non-centralities in the numerator as well as the denominator, which are quadratic forms in the $\{u_{ij}\}$. Let us examine these non-centralities.

We will ignore the technical errors in our developments.

Let $U_{..} = \Sigma_{ij}\, u_{ij}/Np$, $u_{i.} = \Sigma_j u_{ij}/p$ and $u_{.j} = \Sigma_i\, u_{ij}/N$.

The average of the observations on subject i under the randomization model is

$$u_{i.} + \frac{1}{p}\Sigma_k\, \tau_k\, \alpha_i^{\ k} + \frac{1}{p}\Sigma_k\, r_k\, \beta_i^{\ k} .$$

The average for period j is

$$u_{.j} + \tau_{.} + r_{.}\,(1 - \delta_j^1) \quad ,$$

and the overall average is

$$u_{..} + \tau_{.} + r_{.}(p-1)/p \qquad .$$

Hence the sum of squares eliminating subjects and periods is

$$\Sigma_{ij}\, \{\, u_{ij} - u_{i.} - u_{.j} + u_{..}\,) + \Sigma_k(\tau_k - \tau_{.})(\alpha_{ij}^{\ k} - \frac{1}{p}\,\alpha_i^{\ k}\,)$$

$$+ \Sigma_k\, (r_k - r_{.})(\beta_{ij}^{\ k} - \frac{1}{p}\,\beta_i^{\ k}) \quad \}^2 .$$

Now we have

$$\Sigma_{ij}\, (\alpha_{ij}^{\ k} - \frac{1}{p}\,\alpha_i^k\,)^2 = \Sigma_{ij}\, (\, \frac{(p-2)}{p}\,\alpha_{ij}^{\ k} + \frac{1}{p^2}\,\alpha_i^{\ k}\,) = N(p-1)/t \quad ,$$

$$\Sigma_{ij}\, (\alpha_{ij}^{\ k} - \frac{1}{p}\,\alpha_i^{\ k})(\alpha_{ij}^{\ k'} - \frac{1}{p}\,\alpha_i^{\ k'}) = -N(p-1)/t(t-1) \quad ,$$

$$\Sigma_{ij}\, (\beta_{ij}^{\ k} - \frac{1}{p}\,\beta_i^{\ k})^2 = \Sigma_{ij}\, (\, \frac{p-2}{p}\beta^k{}_{ij}^{\ k} + \frac{1}{p^2}\, \beta_i^{\ k}) = N(p-1)^2/pt \quad ,$$

$$\Sigma_{ij}\, (\beta_{ij}^{\ k} - \frac{1}{p}\,\beta_i^{\ k})(\beta_{ij}^{\ k'} - \frac{1}{p}\,\beta_i^{\ k'}) \;=\; -\,\frac{N(p-1)(p-2)}{pt(t-1)} \quad ,$$

$$\Sigma_{ij}(\alpha_{ij}^{\ k} - \frac{1}{p}\,\alpha_i^{\ k})(\beta_{ij}^{k} - \frac{1}{p}\,\beta_i^{\ k}) \;=\; -\,\frac{N(p-1)}{pt} \quad ,$$

and

$$\Sigma_{ij}(\alpha_{ij}^{\ k} - \frac{1}{p}\,\alpha_i^{\ k})(\beta_{kj}^{k'} - \frac{1}{p}\,\beta_i^{\ k'}) = \frac{N(p-1)}{pt(t-1)} \qquad .$$

Put $u_{ij}^* = u_{ij} - u_{i.} - u_{.j} + u_{..}$; $\alpha_{ij}^{*k} = \alpha_{ij}^{\ k} - \alpha_i^{\ k}/p$; $\beta_{ij}^{*k} = \beta_{ij}^{k} - \beta_i^{\ k}/p$;

$\tau_k^{\ *} = \tau_k - \tau_{.}$, and $r_k^{\ *} = r_k - r_{.}$.

Another form for the total sum of squares eliminating the sum of squares for subjects and periods is

$$\Sigma_{ij}\, u_{ij}^{*2} + \Sigma_k [a\tau_k^{*2} + 2b\tau_k^{*} r_k^{*} + c r_k^{*2}]$$
$$+ 2\Sigma_{ij}\, u_{ij}^{*}\, \Sigma_k \{ \tau_k^{*} \alpha_{ij}^{*k} + r_k^{*} \beta_{ij}^{*k} \}$$

where $\quad (a,b,c) = \frac{s(p-1)}{(t-1)}[1, -\frac{1}{p}, \frac{pt-t-1}{pt}]$.

Now

$$T_k^{*} = a\tau_k^{*} + b\, r_k^{*} + \Sigma_{ij}\, u_{ij}\alpha_{ij}^{k} - u_i.\alpha_i^{k} \quad .$$

Since $\quad \Sigma_i(\alpha_{ij}^{k} - \frac{1}{p}\alpha_i^{k}) = 0 = \Sigma_j(\alpha_{ij}^{k} - \frac{1}{p}\alpha_i^{k})$,

then we have

$$T_k^{*} = a\tau_k^{*} + b r_k^{*} + \Sigma_{ij}\, u_{ij}^{*}\alpha_{ij}^{*k} \quad ,$$

with expectation under randomization given by

$$E_R T_k^{*} = a\tau_k^{*} + b r_k^{*} \quad .$$

Similarly

$$R_k^{*} = c r_k^{*} + b\tau_k^{*} + \Sigma_{ij}\, u_{ij}^{*}\, \beta_{ij}^{*k}$$

with expectation under randomization given by

$$E_R(R_k^{*}) = c r_k^{*} + b\tau_k^{*} \quad .$$

To obtain expectations of mean squares it is convenient to compute these under the null hypothesis of no direct or residual effects. The actual EMS's are then obtained simply by adjoining the contributions from the direct and residual effects. Therefore, let T_0^{*}, R_0^{*} be the vectors T^{*} and R^{*} under the null hypothesis . Then we have

$$T^{*'}T^{*} = \Sigma_k(a\tau_k^{*} + br_k^{*})^2 + T_o^{*'}T_o + 2\,\Sigma_k\,[\Sigma_{ij}\,u_{ij}^{*}\,\alpha_{ij}^{*k}\,\{a\tau_k^{*} + br_k^{*}\}] \quad ,$$

$$R^{*'}R^{*} = \Sigma_k\{\,cr_k^{*} + b\tau_k^{*}\}^2 + R_o^{*'}R_o^{*} + \Sigma_k\Sigma_{ij}\,u_{ij}^{*}\,\beta_{ij}^{*k}\,\{cr_k^{*} + b\tau_k^{*}\} \quad ,$$

and

$$T^{*'}R^{*} = \Sigma_k\,[\,(a\tau_k^{*} + br_k^{*})(cr_k^{*} + b\tau_k^{*})\,] + T_o^{*'}R_o^{*} + \Sigma_k[\Sigma_{ij}\,(u_{ij}^{*}\alpha_{ij}^{*k})\,\cdot$$

$$(cr_k^{*} + b\tau_K^{*}) + \Sigma_{ij}(u_{ij}^{*}\beta_{ij}^{*k})(a\tau_k^{*} + br_k^{*})] \quad .$$

The sum of squares for direct treatments and residual treatment effects is then equal to

$$(ac - b^2)^{-1}\,\{c\,T^{*'}T^{*} - 2bT^{*}R^{*} + aR^{*'}R^{*}\}$$

$$= \Sigma_k\,\{a\tau_k^{*2} + b\tau_k^{*}r_k^{*} + cr_k^{*2}\,\} + 2\Sigma_{ij}u_{ij}^{*}\Sigma_k\,\{\tau_k^{*}\alpha_{ij}^{*k} + r_k^{*}\beta_{ij}^{*k}\}$$

$$+ \ (ac - b^2)^{-1}\,\{cT_o^{*'}T_o^{*} - 2bT_o^{*'}R_o^{*} + aR_o^{*'}R_o^{*}\}$$

Hence the error sum of squares is

$$\Sigma^{*}u_{ij}^{2} - (ac - b^2)^{-1}\{cT_o^{*'}T_o^{*} - 2bT_o^{*'}R_o^{*} + aR_o^{*'}R_o^{*}\}$$

Now if the direct treatment effects are all the same and the residual treatment effects are all the same, then the sum of the error sum of squares, E_{yy} and the (direct plus residual) treatment sum of squares, T_{yy}, is $\Sigma_{ij}\,u_{ij}^{*2}$, constant. The combined treatment sum of squares, T_{yy}, is then given by $(ac - b^2)^{-1}\,\{cT_o^{*'}T_o^{*} - 2bT_o^{*'}R_o^{*} + aR_o^{*'}R_o^{*}\}$, with expectation equal to

$$\left(\frac{pt^2 - pt - p - t}{p^2t}\right)^{-1} \cdot \frac{(t-1)}{N(p-1)} \cdot \frac{N}{(N-1)}\left\{\frac{2pt - t - 1}{pt} + \frac{2\Sigma_{ij}\,u_{ij}^{*}u_{i,j+1}^{*}}{p\Sigma\,u_{ij}^{*2}}\right.$$

$$\left. - \frac{(t+1)}{t}\,\frac{\Sigma_i u_{i1}^{*2}}{\Sigma_{ij}u_{ij}^{*2}}\right\} \cdot \Sigma_{ij}\,u_{ij}^{*2} \qquad \text{(See Appendix D).}$$

Also, $E_R\,(E_{yy}) = \Sigma u_{ij}^{*2} - E_R\,(T_{yy})$.

Under the null hypothesis, the sum of squares for direct treatment effects eliminating residual treatment effects has expectation equal to

$$E_R \;\left[(ac - b^2)^{-1} \{cT_o^{*\prime}T_o^{*} = 2b\, T_o^{*\prime}R_o^{*} + aR_o^{*\prime}R_o^{*}\} - c^{-1}R_o^{*\prime}R_o^{*}\right]$$

$$= (ac - b^2)^{-1} \frac{N}{(N-1)} \Big\{ \frac{(c^2 + b^2)}{c} \Sigma_{ij} u_{ij}^{*2} + \frac{b^2}{c}\frac{(t+1)}{t} \Sigma_i u_{i1}^{*2}$$

$$- 2b\, \Sigma_{ij}\, u_{ij}^{*}\, u_{i,j+1}^{*} \Big\} \;.$$

Now

$$\frac{N(p-1)}{(t-1)} \cdot \frac{(c^2+b^2)}{c(ac-b^2)} = \frac{(p^2t^2 + t^2 + 1 - 2pt^2 - 2pt + 2t + t^2) \times pt \times p^2t}{p^2t^2(pt - t - 1)(p^2t - pt - p - t)}$$

$$= 1 + \frac{t(2pt - t - 1)}{(pt - t - 1)(p^2t - pt - p - t)} \;,$$

$$\frac{N(p-1)}{(t-1)} \frac{b^2}{c(ac-b^2)} = \frac{pt \cdot p^2t}{P^2(pt - t - 1)(p^2t - pt - p - t)} \;,$$

and

$$\frac{N(p-1)}{(t-1)} \frac{b}{(ac-b^2)} = - \frac{p^2t}{p(p^2t - pt - p - t)} \;.$$

So the sum of squares for direct treatment effects eliminating residual treatment effects has expectation equal to

$$\Big\{\Sigma_{ij}\, u_{ij}^{*2} + t\Big[\frac{(2pt - t - 1)\, \Sigma u_{ij}^{*2} + p(t+1)\Sigma_i\, u_{i1}^{*2}}{(p^{2t} - pt - p - t)(pt - t - 1)}$$

$$+ \frac{2p(pt - t - 1)\Sigma\, u_{ij}^{*}\, u_{i,j+1}}{(p^2t - pt - p - t)(pt - t - 1)}\Big]\Big\} \times \frac{(t-1)}{(N-1)(p-1)} \;.$$

We shall exhibit and discuss these results in section 4.3 where they are laid out in analysis of variance tables.

4.2.5 Balanced Extra Period Change-Over Designs

Balanced extra period change over designs may be considered as balanced change over designs in which the treatment applied in the p-th period is repeated in an extra period the (p+1) period. For example, we may have the following sequences:

	Subjects	1	2	3	4
	1	A	B	C	D
Periods	2	B	C	D	A
	3	C	D	A	B
	4	C	D	A	B

In these extra period designs, however, the properties of the design random variables $\{\alpha_{ij}^{k}$ and $\beta_{ij}^{k}\}$ are somewhat different.

Let $\alpha_{ij}^{k} = 1$ if treatment k is given to the i-th subject in the j-th period

$= 0$ otherwise,

and let $\beta_{ij}^{k} = 1$ if treatment k is given to the i=th subject in the (j-1)-th period, $j > 1$,

$= 0$ otherwise.

Let $\alpha_{i}^{k} = \sum_{j=1}^{p+1} \alpha_{ij}^{k}$, and $\beta_{i}^{k} = \sum_{j=1}^{p+1} \beta_{ij}^{k}$. Then in addition to the obvious properties of the $\{\alpha_{ij}^{k}$' s and the β_{ij}^{k}' s$\}$ the conditions of balance will give us the following:

$$E(\alpha_{i}^{k}) = (p+1)/t \quad ,$$

$$E(\beta_{i}^{k}) = (p)/t \quad ,$$

$$\Sigma_{i}\alpha_{i}^{k}\,\alpha_{i}^{k'} = N(p-1)(p+2)/t(t-1) \quad ,$$

$$\Sigma_{i}\beta_{i}^{k}\,\beta_{i}^{k'} = Np(p-1)/t(t-1) \quad ,$$

$$\Sigma_i \alpha_i^k \beta_i^k = N(p+1)(p-1)/t(t-1) \quad ,$$

$$\Sigma_i \alpha_i^k \beta_{i,p+1}^{\ k} = \frac{2N}{t}$$

$$\Sigma_i \alpha_i^k \beta_{ij} = \frac{N}{t} \qquad 2 \le j \le p$$

$$\Sigma_i \beta_i^k \alpha_{ij}^k = \frac{N}{t}$$

The linear model for the balanced extra period cod with N units, (p+1) periods and t treatments is given by

$$y_{ijkk'} = \mu + s_i + \theta_j + \tau_k + r_{k'} + e_{ijkk'}$$

$i=1, \ldots, N$, $j=1, \ldots, p$, $k=1, \ldots, t$, $k'=1, \ldots, t$,

where μ denotes a general mean and s_i, θ_j, τ_k, $r_{k'}$ corresponds to the subject, period, direct treatment and residual effects respectively and $e_{ijkk'}$ represents the error term. The normal equations are as follows:

$$N(p+1)\ (\mu + s_. + \theta_. + \tau_. + pr/(p+1)) = Y_{....} \quad , \qquad 4.2.5.1$$

$$Np(\mu + s_. + \frac{(p+1)}{p}(\theta_. - \frac{\theta_1}{(p+1)}) + \tau_. + r_. = Y_{....} - Y_{.1..} , \qquad 4.2.5.2$$

$$N(\mu + s_. + \theta_j + \tau_. + (1 - \delta_j^1)r = Y_{.j..} ,\ j > 1 , \qquad 4.2.5.3$$

where $\delta_j^m = 1$ if $j = m$ and $\delta_j^m = 0$ otherwise.

$$(p+1)\ (u + s_i + \theta_.) + \Sigma_k \beta_i^k r_k = Y_{i.} \quad , \qquad 4.2.5.4$$

$$\frac{N(p+1)}{t}(u + \theta_. + \tau_k) + \frac{N(t-p)}{t(t-1)} r_k + \frac{N(p-1)}{(t-1)} r_k + \Sigma_i \alpha_i^k s_i = T_k \ , \qquad 4.2.5.5$$

$$\frac{Np}{t}(u + \frac{(p+1)}{p}(\theta_. - \frac{\theta_1}{(p+1)}) + r_k) + \frac{N(t-p)}{t(t-1)} \tau_k + \Sigma_i \beta_i^k s_i = R_k \ , \qquad 4.2.5.6$$

Let $D_k = \Sigma_i \alpha_i^k Y_{i\ldots}$, that is, D_k is the sum of all subtotals of the experimental units which receive treatment k in some period, and let $F_k = \Sigma_i \beta_i^k Y_{i\ldots}$, the sum of all subtotals of the experimental units which receive treatment k in some period other than the first.

Put $s_i^* = s_i - s_.$, $\tau_k^* = T_k - \tau_.$, and $r_k^* = r_k - r_.$.

Divide equation (1) by N and subtract the result from equation (4) to obtain

$$(p+1)\, s_i^* \, \Sigma_k \left[\tau_k^* \alpha_i^k + r_k^* \beta_i^k\right] = Y_{i\ldots} - Y_{\ldots\ldots}/N \qquad (7)$$

Similarly, we obtain the following:

$$\frac{N(p+1)}{t} \tau_i^* + \Sigma_i \, s_i^* \alpha_i^k + \frac{N(t-p)}{t(t-1)} r_k^* = T_k - Y_{\ldots\ldots}/t \qquad (8)$$

and

$$\frac{NP}{t} r_k^* + \frac{N(t-p)}{t(t-1)} \tau_k^* + \Sigma_i \, s_i^* \beta_i^k = R_k - (Y_{\ldots\ldots} - Y_{.1..})/t \qquad (9)$$

If we eliminate s_i^* from equations (8) and (9) we will obtain

$$\frac{N(p+1)}{t} \tau_k^* - \frac{1}{(p+1)} \Sigma_i \Sigma_{k'} \{\tau_{k'}^* \alpha_i^k \alpha_i^{k'} + r_k^* \alpha_i^* \beta_i^{k'}\}$$

$$+ \frac{N(t-p)}{t(t-1)} r_k^* = T_k - \frac{Y_{\ldots\ldots}}{t} - \frac{1}{(p+1)} \Sigma_i \, \alpha_i^k \left(Y_{i\ldots} - \frac{Y_{\ldots\ldots}}{N}\right)$$

$$= T_k - \frac{1}{(p+1)} \Sigma \alpha_i^k Y_{i\ldots} = T_k^* , \qquad \text{say,}$$

and

$$\frac{Np}{t} r_k^* \frac{1}{(p+1)} \Sigma_i \Sigma_{k'} \{\tau_k^* \beta_i^k \alpha_i^{k'} + r_{k'}^* \beta_i^k \beta_i^{k'}\} + \frac{(t-1)}{t(t-1)} \tau_k$$

$$= R_k - \frac{(Y_{\ldots\ldots} - Y_{.1..})}{t} - \frac{1}{(p+1)} \Sigma_i \, \beta_i^k \left(Y_{i\ldots} - \frac{Y_{\ldots\ldots}}{N}\right)$$

$$= R_k - \frac{1}{(p+1)} \Sigma_i \beta_i^k Y_{i\ldots} - \frac{Y_{\ldots\ldots}}{t(p+1)} + \frac{Y_{.1..}}{t} = R_k^* , \qquad \text{say,}$$

Simplifying the above equations, we obtain the reduced normal equations

$$\begin{bmatrix} f(I - \frac{1}{t} J) & \emptyset \\ \emptyset & g(I - \frac{1}{t} J) \end{bmatrix} \begin{bmatrix} \tau \\ r \end{bmatrix} = \begin{bmatrix} I^* \\ R^* \end{bmatrix},$$

where $(f, g) = \{(p^2t + pt - 2) \ , \ (p^2t - p)\} \frac{N}{t(t-1)(p+1)}$

Impose the non-estimable constraints, $J\underline{\tau} = 0 = J\underline{r}$ on the solutions to obtain

$$\hat{\tau} = T^*/f \ ,$$

and

$$\hat{r} = R^*/g \ .$$

The reduction in sum of squares due to τ and r is equal to

$$T^{*'}T^*/f + R^{*'}R^*/g \ .$$

Now, under randomization theory and additivity, the observation on the i-th subject in the j-th period may be written as

$$u_{ij} + \Sigma_k(\tau_k\alpha_{ij}^k + r_k\,\beta_{ij}^k) \ .$$

So we have $Y_{i...} = U_{i.} + \Sigma_k\,\tau_k\,\alpha_i^k + \Sigma_k r_k\,\beta_i^k$,

$$Y_{.j..} = U_{.j} + N\tau_. + N(1 - \delta_j^1)r_. \ ,$$

$$Y_{....} = U_{..} + N(p+1)\tau_. + Np\,r_.$$

$$T_k = \Sigma_{ij}u_{ij}\alpha_{ij}^k + \frac{N(p+1)}{t}\tau_k + \frac{N(t-p)}{t(t-1)}r_k + \frac{N(p-1)}{(t-1)}r_.$$

and

$$R_k = \Sigma_{ij}u_{ij}\beta_{ij}^k + \frac{Np}{t}r_k + \frac{N(t-p)}{t(t-1)}\tau_k + \frac{N(p-1)}{(t-1)}\tau_.$$

By taking expectations of T^* and R^* under randomization we obtain

$$E(\hat{\tau}_k) = (\tau_k - \tau_.)$$

$$E(\hat{r}_k) = (r_k - r_.)$$

So the design allows us to obtain unbiased estimates of linear contrasts in the parameters, $\underline{\tau}$ and $\underline{r}$.

As in the case of balanced cod,if the vector of observations on each experimental unit has covariance matrix $\sigma^2(aI + bJ)$, or if we assume the usual GMN assumptions, the covariance matrix of the estimate of the parameters may be written in the form

$$\operatorname{cov}\begin{bmatrix}\hat{\tau}\\ \hat{r}\end{bmatrix} = \sigma^2 \begin{bmatrix} q(I - \frac{1}{t}J) & \emptyset \\ \emptyset & w(I - \frac{1}{t}J)\end{bmatrix},$$

where

$$q = 1/f \quad ,$$

and

$$w = 1/g \quad ,$$

So

$$\text{Var}\,(\tau_k \hat{-} \tau_{k'}) = 2\,\sigma^2/f \quad ,$$

$$\text{Var}\,(r_k \hat{-} r_{k'}) = 2\sigma^2/g \quad ,$$

and

$$\text{covar}\,(\tau_k \hat{-} \tau_{k'})(r_k \hat{-} r_{k'}) = 0 \quad .$$

Let $u_{..} = U_{..}/N(p+1) = \Sigma_{ij}\, u_{ij}/N(p+1)$,

$u_{i.} = U_{i.}/(p+1) \quad = \Sigma_j\, u_{ij}/(p+1)$,

and $u_{.j} = U_{.j}/N \quad = \Sigma_j\, u_{ij}/N$.

Then $T_k^* = f(\tau_k - \tau_.) \quad + \quad \Sigma_{ij}\, u_{ij}\, \alpha_{ij}^{\ k} - \Sigma_i \alpha_i^{\ k}\, U_{i.}/(p+1)$

and $\quad R_k^* = g(r_k - r_.) \quad + \quad \Sigma_{ij}\, u_{ij}\, \beta_{ij}^k - \Sigma_i \beta_i^k\, U_{i.}/(p+1)$

$\qquad - U_{..}/t(p+1) + U_{.1}/t$,

where $(f, g) = N(p^2t + pt - 2, p^2t - p)/t(t-1)(p+1)$. From the properties of the α_{ij}^k 's and β_{ij}^k 's, we have $\Sigma_{ij} u_{ij} \alpha_{ij}^k - \Sigma_i U_{i.} \alpha_i^k/(p+1) =$

$\Sigma_{ij}(u_{ij} - u_{i.}) \{\alpha_{ij}^k - \alpha_i^k/(p+1)\}$

$$= \Sigma_{ij}(u_{ij} - u_{i.} - u_{.j} + u_{..})\{\alpha_{ij}^k - \alpha_i^k/(p+1)\}$$

Similarly,

$$\Sigma_{ij} u_{ij} \beta_{ij}^k - \Sigma U_{i.} \beta_i^k/(p+1)$$

$$= \Sigma_{ij} (u_{ij} - u_{i.})(\beta_{ij} - \beta_i^k/(p+1)) ,$$

and $\Sigma_{ij}(u_{.j} - u_{..})(\beta_{ij}^k - \beta_i^k/(p+1))$

$$= \frac{N}{t} (\Sigma u_{.j} - u_{.1}) - \frac{Np}{t(p+1)} \Sigma_j u_{.j} = \frac{U_{..}}{t(p+1)} - \frac{U_{.1}}{t}$$

Hence

$$\Sigma_{ij} u_{ij} \beta_{ij}^k - \Sigma U_{i.} \beta_i^k / (p+1) - U_{..}/(p+1) + U_{.1}/t$$

$$= \Sigma_{ij} (u_{ij} - u_{i.} - u_{.j} + u_{..})(\beta_{ij}^k - \beta_i^k/(p+1))$$

Put $u_{ij}^* = (u_{ij} - u_{i.} - u_{.j} + u_{..})$, $\tau_k^* = (\tau_k - \tau_.)$, $r_k^* = (r_k - r_.)$.

Then $T_k^* = f(\tau_k^*) + h(r_k^*) + \Sigma_{ij} u_{ij}^* \{\alpha_{ij}^k - \alpha_i^k/(p+1)\}$

and $R_k^* = h(\tau_k^*) + g(r_k^*) + \Sigma_{ij} u_{ij}^* \{\beta_{ij}^k - \beta_i^k/(p+1)\}$

The sum of squares eliminating periods and units is

$$\Sigma_{ij} \{ u_{ij} + \Sigma_k \tau_k \alpha_{ij}^k + \Sigma_k r_k \beta_{ij}^k - Y_{i.}/(p+1) - Y_{.j}/N + Y_{..}/N(p+1) \}^2$$

$$= \Sigma_{ij} \{ u_{ij}^* + \Sigma_k \tau_k^* (\alpha_{ij}^k - \alpha_i^k/(p+1)) + \Sigma_k r_k^* (\beta_{ij}^k - \beta_i^k /(p+1))\}^2$$

From the properties of the α_{ij}^{k} 's and β_{ij}^{k} 's, we have the following

$$\Sigma_{ij} (\alpha_{ij}^{k} - \alpha_{i}^{k}/(p+1))^2 = \Sigma_{ij}\{ \frac{p-1}{(p+1)} \alpha_{ij}^{k} + \frac{\alpha_i^k}{(p+1)^2}\} = Np/t ,$$

$$\Sigma_{ij}(\alpha_{ij}^{k} - \alpha_{i}^{k}/(p+1))(\alpha_{ij}^{k'} - \alpha_{i}^{k'}/(p+1)) = - \Sigma_i \alpha_i^{k}\alpha_i^{k'}$$

$$= - N(p+2)(p-1)/t(t-1)(p+1) ,$$

$$\Sigma_{ij} (\beta_{ij}^{k} - \beta_{i}^{k}/(p+1))^2 = Np^2/t(p+1) ,$$

$$\Sigma_{ij} (\beta_{ij}^{k} - \beta_{i}^{k}/(p+1)) (\beta_{ij}^{k'} - \beta_{i}^{k'}/(p+1)) = - Np(p-1)/t(t-1)(p+1) ,$$

$$\Sigma_{ij}(\alpha_{ij}^{k} - \alpha_{i}^{k}/(p+1))(\beta_{ij}^{k} - \beta_{i}^{k}/(p+1) = 0 ,$$

$$\Sigma_{ij}(\alpha_{ij}^{k} - \alpha_{i}^{k} / (p+1))(\beta_{ij}^{k'} - \beta_{i}^{k'}/(p+1)) = 0 .$$ So $h = 0$.

Put $u_{ij}^{*} = (u_{ij} - u_{i.} - u_{.j} + u_{..})$, $\tau_k^{*} = \tau_k - \tau_{.}$, $r_k^{*} = r_k - r_{.}$,

$$\alpha_{ij}^{*k} = \alpha_{ij}^{k} - \alpha_{i}^{k}/(p+1) , \quad \beta_{ij}^{*k} = \beta_{ij}^{k} - \beta_{i}^{k}/(p+1) .$$

Then the sum of squares eliminating subjects and periods is

$$\Sigma_{ij} u_{.j}^{*2} + f \Sigma_k \tau_k^{*2} + g \Sigma_k r_k^{*2} + 2\sum_{kij}\sum u_{ij}^{*} \tau_k^{*} \alpha_{ij}^{*k} + 2\sum_{kij}\sum u_{ij}^{*} r_k^{*} \beta_{ij}^{*k}.$$

Now

$$T^{*'}T^{*} = T_o^{*'}T_o^{*} + f^2 \Sigma_k \tau_k^{*2} + 2f \Sigma_k\Sigma_{ij} u_{ij}^{*}\tau_k^{*}\alpha_{ij}^{*k} ,$$

and

$$R^{*'}R^{*} = R_o^{*'}R_o^{*} + g^2 \Sigma_k r_k^{*2} + 2g \Sigma_k\Sigma_{ij} u_{ij}^{*} r_k^{*} \beta_{ij}^{*k} .$$

So the residual sum of squares is equal to

$$\Sigma_{ij} u_{ij}^{*2} - T_o^{*'}T_o^{*}/f - R_o^{*'}R_o^{*}/g .$$

From Appendix D we have

$$E_R(T_o^{*\prime}T_o^*) = \frac{N}{(N-1)} \cdot (\Sigma u_{ij}^{*2} + 2\Sigma_i u_{ip}^* u_{i,p+1})$$

and

$$E_R(R_o^{*\prime}R_o^*) = \frac{N}{(N-1)} (\Sigma u_{ij}^{*2} - \frac{t+1}{t} \Sigma_i u_{i1}^2) .$$

We give the expectations of the various mean squares under randomization for the extra period cod in Table 3.

4.3 Summary

4.3.1 Expectations of Mean Squares

In this section we summarize the expectations of mean squares under ramdomization for the various models discussed. We shall ignore technical errors except for the case of the Latin square design. In the possible analysis of variance tables we are interested only in the mean squares for a) the direct treatment ignoring residual treatments, b) the residual treatments eliminating direct treatments, c) remainder, and in the mean squares for the alternative ordering. We give the expectations of these means squares in 4 tables.

In Table 1 we give for comparison purposes the expectations of mean squares (EMS's) for a Latin square design with absence of residual effects. In Tables 2a, 2b, and 3, we give results for situations in which residual effects of the first order are present and additivity of treatment effects (direct and residual) and unit effects hold. Tables 2a and 2b give EMS's for the general balanced cod and Table 4 gives EMS's for the extra period cod. The notations used in this section are the same as in the preceding sections unless otherwise specified.

Table 1. The Latin square: non-additivity and technical errors, with factor levels sampled from populations of levels

Source	EMS
Treatments	$\sigma_e^2 + \sigma_{rc}^2 + t\sigma_\tau^2 + (+ \frac{t}{RC})\, \sigma_{rct}^2 + (1 - \frac{t}{R})\, \sigma_{rt}^2 + (1 - \frac{t}{C})\sigma_{ct}^2$
Error	$\sigma_e^2 + \sigma_{rc}^2 + \sigma_{rct}^2 + \sigma_{rt}^2 + \sigma_{ct}^2$

σ_e^2 is the variance of the technical errors.

EMS (treatment) - EMS (error)

$$= \frac{t}{RC} (\sigma_{rct}^2 - C\sigma_{rt}^2 - R\sigma_{ct}^2) + t\sigma_\tau^2 .$$

$$= (1 - \frac{1}{R} - \frac{1}{C} - \frac{1}{T})$$

See theorem 4.1.

Table 2a. The balanced cod with residual effects (ignoring technical error)

Source	EMS
Direct Trt. ign. Res.	$\frac{1}{(N-1)(p-1)} \Sigma\, u_{ij}^{*2} + \frac{N(p-1)}{(t-1)^2} \Sigma_k (\tau_k^* - \frac{1}{p} r_k^*)^2$
Remainder	$\frac{1}{(N-1)(p-1)} \Sigma\, u_{ij}^{*2} + \frac{N(p-1)}{(t-1)} \frac{(p^2t-pt-p-t)}{p^2t(Np-N-p-t+2)} \Sigma_k r_k^{*2}$

Table 2b. The balanced cod with residual effects ignoring technical errors

Source	EMS
Res. after dir. trt.	$\frac{N(p-1)}{(t-1)^2} \cdot \frac{(p^2t-pt-p-t)}{p^2t} \Sigma_k r_k^{*2} + \frac{1}{(N-1)(p-1)} \cdot \{\Sigma_{ij} u_{ij}^{*2} + \frac{p^2t}{(p^2t-pt-p-t)} \cdot [\frac{(pt+p+2t)}{p^2t} \Sigma_{ij} u_{ij}^{*2} + \frac{2}{p} \Sigma_{ij} u_{ij}^{*} u_{i,j+1}^{*} - \frac{(t+1)}{t} \Sigma_i u_{i1}^{*2}]\}$
Res. ign. dir. trt.	$c\Sigma_k (r_k^{*} - \frac{1}{p} \tau_k^{*})^2 + \frac{1}{(N-1)(p-1)} \cdot \frac{pt}{(pt-t-1)} \cdot [\Sigma_{ij} u_{ij}^{*2} - \frac{(t+1)}{t} \Sigma_i u_{i1}^{*2}]$
Dir. trt. after res.	$\frac{N(p-1)}{(t-1)^2} \cdot \frac{(p^2t-pt-p-t)}{(p^2t-pt-p)} \Sigma_k \tau_k^{*2} + \frac{1}{(N-1)(p-1)} \cdot \{\Sigma_{ij} u_{ij}^{*2} + \frac{t}{(p^2t-pt-p-t)(pt-p-1)} \cdot [(2pt-t-1)\Sigma_{ij} u_{ij}^{*2} + p(t+1) \Sigma_i u_{i1}^{*2} + 2p(pt-t-1) \cdot \Sigma_{ij} u_{ij}^{*} u_{i,j+1}^{*})]\}$
Residual error	$\frac{1}{(N-1)(p-1)} \{\Sigma_{ij} u_{ij}^{*2} - \frac{p^2t(t-1)}{(p^2t-pt-p-t)(Np-n-p-2t+3)} \cdot [\frac{(pt+p+2t)}{p^2t} \Sigma_{ij} u_{ij}^{*2} + \frac{2}{p} \Sigma_{ij} u_{ij}^{*} u_{i,j+1}^{*} - \frac{(t+1)}{t} \cdot \Sigma_i u_{i1}^{*2}]\}$

Table 3. The extra period cod (ignoring technical errors)

Source	EMS
Dir. trt.	$f \cdot \Sigma_k t_k^{*2} + \frac{1}{(N-1)p} \{\Sigma_{ij} u_{ij}^{*2} + \frac{2}{(p^2t + pt - 2)} [\Sigma_{ij} u_{ij}^{*2} + pt(p+1)\, \Sigma_i u_{ip} u_{i,p+1}] \}$
Res. trt.	$g\Sigma_k r_k^{*2} + \frac{1}{(N-1)p} [\Sigma_{ij} u_{ij}^{*2} + \frac{(t+1)}{(pt-1)} [\Sigma_{ij} u_{ij}^{*2} - (p+1)\, \Sigma_i u_{i1}^{*2} \}]$
Residual Error	$\frac{1}{(N-1)p} \{ \Sigma_{ij} u_{ij}^{*2} - \frac{(t-1)}{(Np-p-2t+2)} \cdot [\{\frac{(t+1)}{(pt-1)} - \frac{2}{(p^2t+pt-2)} \} \Sigma_{ij} u_{ij}^{*2} - \frac{(t+1)(p+1)}{(pt-1)} \Sigma_i u_{i1}^{*2} + \frac{2pt(p+1)}{p^2t+pt-2)} \Sigma_i u_{ip}^{*} u_{i,p+1}] \}$

From the tables above, it is clear that the designs are biassed. To get some idea of the nature of these results let us make the following simplifying assumptions.

For a given N assume

1) $\Sigma_i u_{ij}^{*2} = N\sigma^2_{.j} = N\sigma^2_{.1}$, that is the variances of the unit responses in each period is the same,

2) $\Sigma_i u^*_{ij} u_{i,j+1}/N\sigma_{.j}\sigma_{.j+1} = \rho_{j,j+1} = \rho$, constant, i.e. the correlation coefficient between observations in successive periods is constant.

Case 1: The ordinary cod.

Under these assumptions we have

$$E(T_{yy}/G_{yy}) = \{ 2 + \frac{(pt + p + 2t - p(t+1) + 2t(p-1)-\rho_{12})}{(p^2 t - pt - p - t)} \} \frac{(t-1)}{(N-1)(p-1)}$$

If p and t are large, (which is not likely in practice), we will have

$$E\left(\frac{T_{yy}}{G_{yy}}\right) = \frac{2(t-1)}{(N-1)(p-1)}$$

which is the result under GMN theory. The relative bias is then equal to

$$\frac{t(2 + (p-1)\rho)}{(p^2 t - pt - p - t)}$$

If ρ is very small or equal to zero the relative bias is positive and approximately equal to $2/p^2$. If ρ is positive, as it approaches 1 in magnitude, the relative bias is positive and increases to

$$\frac{t(p+1)}{(p^2 t - pt - p - t)} = \frac{1}{p} .$$

When ρ is negative the relative bias is negative, for $\rho > 2/(p-1)$.

The bias approaches

$$\frac{-t(p-3)}{(p^2 t - pt - p - t)} = -\frac{1}{p} .$$

Case 2: The extra period cod.

For the extra period design, if we make assumptions that the $\Sigma_i u_{ij}^{*2}$ are the same for all j=1, ..., p+1, we obtain the result

$$E_R \frac{(R_o^{*'} R_o^{*})}{g} = \frac{(t-1)}{(N-1)p} \Sigma_{ij} u_{ij}^{*2}$$

If p and t are assumed to be large, we will obtain the result

$$E_R \frac{(T_o^{*'} T_o^{*})}{f} \simeq \frac{(t-1)}{(N-1)p} \Sigma_{ij} u_{ij}^{*2}$$

and the relative bias is then equal to

$$\frac{2(1 + pt\rho)}{(p^2 t + pt - 2)} \quad .$$

When $\rho = 0$ the relative bias is positive and equal to

$$2/(p^2 t + pt - 2) \quad .$$

When ρ is positive the relative bias is positive and approaches

$$\frac{2(1 + pt)}{(p^2 t + pt - 2)} \approx \frac{2}{p}$$

as ρ approaches 1.

When ρ is negative with magnitude greater than 1/pt, the bias is negative. The bias approaches

$$-\frac{2(pt - 1)}{(p^2 t + pt - 1)} \approx -\frac{2}{p}$$

as ρ approaches -1.

4.3.2 Conclusions

In this chapter of the thesis, we have examined the analysis of designs for sequences of treatments from the viewpoint of randomization analysis. The essential idea was to examine the behavior of the ordinary statistics that come from Gauss-Markoff linear models under randomization.

Some of the literature on treatments applied in sequence has attempted to make experimental inference on the fitting of trend lines, orthogonal polynomials, auto-regressive models as well as treatment parameters. These are essentially the same as covariances which suffer from the defect that an assumption relating the response to the concomitant variable is necessary.

There are two aspects of experimental inference. The first is to form an opinion of the behavior with repetitions of the experiment with the same experimental units. We are facing a problem here since experimental units can be used only once. The second is to extend the inference to some population or experimental material of interest. There are deep obscurities in such an extension based on our experiment. For in assuming that the material of interest is like our own, or that our simple model specifications are adequate for the new material of interest, the unwary experimenter may be tempted to overlook the obvious fact that many new difficulties arise in experimentation with live animals and human beings than, say, in agricultural experimentation. Applying multiple treatments or taking multiple measurements on the same subject will, in general, lead to departures of assumptions of statistical independence of errors residuals and additivity of treatment effects. Psychological interference due to the awareness of the novel or disruptive nature of the experiment, interaction of the history of the subject and treatment, time of measurement, test and treatment carry-over effects, interaction of test carry-over and treatment effects are possible sources of bias. Also when more than one treatment is administered consecutively to the same subject it may be difficult to ascertain the cause of the experimental results or to generalize results to settings in which only one treatment is present.

In the additive model, if the possibility of residual effects can be ignored, accurate comparisons of treatment effects can be obtained from change-over trials. Residual effects are, however, frequently very important. If they are ignored, the resulting comparisons of treatment effects are likely to be erroneous and misleading. In this thesis we have shown that fitting constants for these residual effects does not remove all the difficulties, for we will still be faced with the problems of biasedness, in our estimates of error variance, and reduced accuracy of estimated direct and total effects. The variance of treatment contrasts are not in general estimable. Under certain simplifying assumptions the sign and magnitude of the differences (bias) in the expectations of the mean square for treatment effects (direct and residual) and error mean square depend on the sign and magnitude of the 'average' correlation coefficient between responses in continguous or successive periods. In many experiments of the class, agricultural field experiments, say, the correlation coefficients will usually be positive. Furthermore, it is unlikely that the cod would be used for large values of p, the number of periods and t, the number of treatments, so the fact that biases are $O(p^{-1})$ is not useful.

We have seen that there are major obscurities in the properties of the class of change-over or switch-over designs. Any inference based on their use is at best quite suspect.

5. APPENDIX A

ORDER IN PROBABILITY

In this section we shall present some results in large sample statistical theory which we need to determine the asymptotic behavior of some statistics of interest. We use the concepts of order in magnitude as used in real analysis and order in probability which were introduced by Mann and Wald (1943). Essentially all of this appendix has been abstracted from Fuller (1976).

Let $\{a_n\}$ $n=1,2,\ldots$ be a sequence of real numbers and $\{r_n\}$ $n=1,2,\ldots$ be a sequence of positive real numbers.

<u>Definition</u>: We say a_n is of smaller order than r_n, which we write as

$$a_n = o(r_n)$$

if $\lim\limits_{n\to\infty} \frac{a_n}{r_n} = 0$.

<u>Definition</u>: We say a_n is at most of order r_n

$$a_n = O(r_n)$$

if for some finite real positive number M

$$\frac{a_n}{r_n} \leq M \qquad \text{for all } n.$$

Let $\{a_n\}$, $\{b_n\}$ be sequences of real numbers. Let $\{r_n\}$ and $\{s_n\}$ be sequences of positive real numbers.

<u>Lemma A.1</u>:

1) If $a_n = o(r_n)$ and $b_n = o(s_n)$, then $|a_n|^p = o(r_n^p)$ for $p > 0$,

And $$a_n + b_n = o(\max\{r_n, s_n\}).$$

2) If $a_n = O(r_n)$ and $b_n = O(s_n)$, then $a_n b_n = O(r_n s_n)$

$$|a_n|^p = O(r_n^p) \quad \text{for } p > 0$$

$$a_n + b_n = O(\max\{r_n, s_n\})$$

3) If $a_n = o(r_n)$ and $b_n = O(s_n)$ then $a_n b_n = o(r_n s_n)$.

Definition: The sequence of random variables $\{X_n\}$ converges in probability to the random variable X and we write

$$\text{plim } X_n = X$$

or

$$X_n \xrightarrow{P} X$$

if for every $\epsilon > 0$ and $\delta > 0$ there exists an N such that for $n > N$

$$P\{|X_n - X| > \epsilon\} < \delta$$

or alternatively we write

$$\lim_{n \to \infty} P\{|X_n - X| > \epsilon\} = 0.$$

Let $\{X_n\}$ be a sequence of random variables and $\{r_n\}$ a sequence of positive real numbers.

Definition: We say that X_n is of smaller order in probability than r_n and write

$$X_n = o_p(r_n)$$

if plim $\frac{X_n}{r_n} = 0.$

<u>Definition:</u> We say that X_n is at most of order in probability r_n and write

$$X_n = O_p(r_n)$$

if for every $\epsilon > 0$, there exists a positive real number M_ϵ and an N_ϵ such that

$$P\{|X_n| \geq M_\epsilon r_n\} \leq \epsilon$$

for all $n > N_\epsilon$.

<u>Definition:</u> If X_n is a k-dimensional random variable with elements X_{nj} j=1,2, ..., p and $\{r_n\}$ is a sequence of positive numbers, X_n is said to be at most of order in probability r_n and we write $X_n = O_p(r_n)$. If for every $\epsilon > 0$, there exists a positive real number M and an N such that

$$P\{|X_{jn}| \geq M_\epsilon r_{jn}\} \leq \epsilon, \ j=1,2,\ldots,k$$

for all $n > N_\epsilon$.

We say that X_n is of smaller order in probability than r_n if for every $\epsilon > 0$ and $\delta > 0$ there exists an N such that for $n > N$

$$P\{|X_{jn}| > \epsilon r_n\} < \delta \ , \ j=1,2,\ldots,k$$

Order in probability is similarly defined for a random matrix X_n in terms of the elements X_{ijn} and r_n.

<u>Lemma A.2:</u>

Let X_i, i=1,2,...,n be k dimensional random variables. Then for every $\epsilon > 0$

$$P\left\{\left|\sum_{i=1}^{n} X_i\right| \geq \epsilon\right\} \leq \sum_{i=1} P\left\{|X_i| \geq \frac{\epsilon}{n}\right\}$$

Lemma A .3:

Let X_n be a k dimensional random variable such that

$$\text{plim } X_{jn} = X_j, \quad j=1,2,\ldots,k$$

where X_{jn} is the j-th element of X_n. Then, for k fixed

$$\text{plim} \quad X_n = X$$

Theorem A .1 (Chebychev's inequality)

Let $r > 0$, let X be a random variable such that $E(|X|^r) < \infty$. Then for every $\epsilon > 0$ and $A < \infty$,

$$P\{|X - A| \geq \epsilon\} \leq \frac{E\{|X - A|^r\}}{\epsilon^r}$$

Corollary:

Let $\{X_n\}$ be a sequence of random variables and $\{a_n\}$ a sequence of positive real numbers such that $E\{X_n^2\} = O(a_n^2)$.

Then $$X_n = O_p(a_n).$$

Proof:

Since $E\{X_n^2\} = O(a_n^2)$ there exists an M_1 such that

$$E\{X_n^2\} < M_1^2 a_n^2 \text{ for all n.}$$

By the Tchebychev's inequality for any $M_2 > 0$

$$P\{|X_n| \geq M_2 a_n\} \leq \frac{E\{X_n^2\}}{M_2^2 a_n^2}$$

Hence given $\epsilon > 0$, choose $M_2 \geq M_1 \epsilon^{-\frac{1}{2}}$.

Then $P\{X_n \geq M_2 a_n\} \leq \dfrac{M_1^2 a_n^2}{M_1^2 a_n^2} \epsilon = \epsilon$

Definition: For $r \geq 1$ the sequence of random variables $\{X_n\}$ is said to converge in r-th mean if $E\{|X_n{}^r|\} <$ for all n and $E\{|X_n - X_m|^r\} \to 0$ as $n \to \infty$ and $m \to \infty$. If $\{X_n\}$ converges to X in r-th mean, we denote this by writing $X_n \xrightarrow{r} X$.

Theorem A .2

Let $\{X_n\}$ be a sequence of random variables with finite r-th moments. If there exists a random variable X such that $X_n \xrightarrow{r} X$, then $X_n \xrightarrow{P} X$.

Theorem A .3

Let $\{X_n\}$ be a sequence of real valued k-dimensional random variables such that $\text{plim } X_n = X$. Let g(x) be a function mapping the real k-dimensional vector x into a real p-dimensional space. Let g(x) be continuous. Then

$$\text{plim } g(X_n) = g(X)$$

Mann and Wald (1943) demonstrated that the algebra of the arithmetic order relationship extends to order in probability.

Theorem A .4

Let $\{X_n\}$ be a sequence of k-dimensional random variables with elements $\{X_{jn};\ j=1,2,\ldots,k\}$, and let $\{r_n\}$ be a sequence of k-dimensional vectors with positive real elements $\{r_{jn};\ j-1,2,\ldots,k\}$ such that

$$X_{jn} = O_p(r_{jn}) \qquad j=1,2,\ldots,t$$

$$X_{jn} = o_p(r_{jn}) \qquad j=t+1,\ t+2,\ \ldots,\ k$$

Let $g_n(x_n)$ be a sequence of measurable functions on the k-dimensional Euclidean space and let $\{s_n\}$ be a sequence of positive real numbers. Let $\{a_n\}$ be a non- random sequence of k-dimensional vectors so that

$$g_n(a_n) = O(s_n)$$

whenever $a_{jn} = O(r_{jn}) \qquad j=1,2,\ldots,t$

$$a_{jn} = o(r_{jn}) \qquad j=t+1,\ t+2,\ldots,k.$$

Then $g_n(X_n) = O_p(s_n)$

Proof: (See Fuller (1976).)

If we replace $g_n(a_n) = O(s_n)$

by $g_n(a_n) = o(s_n)$

We may replace

$$g_n(X_n) = O_p(s_n)$$

by $g_n(X_n) = o_p(s_n)$

Corollary:

Let $\{X_n\}$ be a sequence of scalar random variables such that $X_n = a + O_p(r_n)$ where $r_n \to 0$. If g(x) is a function with s continuous derivatives at x = a, then $g(X_n) = g(a) + g^{(1)}(a)(x_n-a) + \ldots + \frac{1}{(s-1)!}$ $g^{(s-1)}(a)(x_n-a)^{s-1} + O_p(r_n{}^s)$, where $g^{(j)}(a)$ is the j-th derivative of g(x) evaluated at x = a. (See Fuller (1976).)

If $X_n = a + O_p(r_n)$ is replaced by $X_n = a + o_p(r_n)$, then the remainder $O_p(r_n{}^s)$ is replaced by $o_p(r_n{}^s)$.

Convergence in Distribution:

Definition: Given $\{X_n\}$ a sequence of random variables with distribution functions $\{F_n\}$, X_n is said to converge in distribution to the random variable X with distribution function F, and we write

$$X_n \xrightarrow{L} X$$

if $\lim_{n \to \infty} F_n = F$ at every continuity point of F.

Theorem A.5

If $\operatorname{plim} |X_n - Y_n| = 0$

and $$X_n \xrightarrow{L} X ,$$

then $$Y_n \xrightarrow{L} X .$$

Corollary:

If $\operatorname{plim} X_n = X$,

then $$S_n \xrightarrow{L} X .$$

Corollary:

If g(X) is a continuous function except on a set D with probability measure 0 and

$$\operatorname{plim} X_n = X ,$$

then $$g(X_n) \xrightarrow{L} g(X) .$$

Proof: (See Fuller (1976).)

Theorem A.6

Let $\{X_n\}$ be a sequence of random variables such that F_n is the distribution function of X_n and let X be a random variable with distribution function F.

If $\qquad F_n \to F, \qquad$ at all continuity points of F,

then $\qquad \int g dF_n \to \int g dF$

for every bounded continuous function g.

Theorem A.7 (Multivariate Central Limit Theorem)

Let F_n denote the joint distribution function of the k-dimensional random variables $(X_{1n}, X_{2n}, \ldots, X_{kn})$. n=1,2,... . Let F be the joint distribution function of the k-dimensional random variable $(X_1, X_2, \ldots, X_k)$. Let $F_{\lambda n}$ be the distribution function of the linear function

$$\sum_{i=1}^{k} \lambda_i X_{in}$$

Then the necessary and sufficient condition that $F_n \to F$ is that

$$F_{\lambda n} \to F_\lambda \text{ for each arbitrary vector } (\lambda_1, \lambda_2, \ldots, \lambda_k).$$

[We obtained this theorem statement from Rao (1965).]

6. APPENDIX B

NORMS AND SPECTRAL RADII OF MATRICES

The concepts of vector norms and spectral radii of matrices play an important role in iterative or approximating methods in numerical analysis. It is often of interest to compare two vectors by their length and likewise it may be convenient to compare two matrices by some measure or norm. Furthermore, the norm of a given matrix quite often gives information on the rate of convergence of functions of that matrix.

The results given below are given in standard texts (See Varga (1962)) and are stated without proof. They are properties of matrices in the N dimensional vector space over the field of complex numbers and hence apply to the N dimensional vector space over the field of real numbers which it contains.

Let C_n be the n dimensional vector space over the field of complex numbers C of column vectors X, where the vector X, its transpose X^T and its conjugate transpose X^* are denoted by

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_n \end{bmatrix} \quad X^T = [x_1, x_2, \ldots, x_n] \text{ , } X^* = [\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_b]$$

where $x_1, x_2, \ldots, x_n$ are complex numbers, and $\bar{x}_i$ is the complex conjugate of x_i . Clearly if X is real we will have $X^T = X^*$.

<u>Definition</u>: Let X be a (column) vector of C_n . Then

$$\|X\| = (X^*X)^{\frac{1}{2}} = (\sum_{i=1}^{n} |x_i|^2)$$

is the Euclidean norm (or length) of X. The following results are well-known.

Theorem B.1

If X and y are vectors of R^n, then

$$\|X\| \geq 0 \text{ with equality when } x = \emptyset;$$

If α is a scalar, then

$$\|\alpha X\| = |\alpha| \cdot \|X\| \quad ,$$

$$\|X + Y\| \leq \|X\| + \|Y\| \quad .$$

If we have an infinite sequence $x^{(0)}$, $x^{(1)}$, $x^{(2)}$ of vectors of R^n. We say that this sequence converges to a vector X of R^n if

$$\lim_{k \to \infty} x_j^{(k)} = x_j, \qquad 1 \leq j \leq n ,$$

where $x_j^{(k)}$ and x_j are respectively, the j-th components of the vectors $x^{(k)}$ and x. Similarly, an infinite series $\sum_{k=0}^{\infty} (y)^k$ of vectors in R^n is said to converge to a vector Y of R^n if

$$\lim_{n \to \infty} \sum_{k=1}^{n} y_j^{(k)} = y_j \qquad 1 \leq j \leq n \quad .$$

Let $A = (a_{ij})$. Then the transpose of A, $A^T = (a_{ji})$

Definition: Let $A = (a_{ij})$ be an nxn real matrix with eigenvalues λ_i, $1 \leq i \leq n$. Then

$$\rho(A) \equiv \max_{1 \leq i \leq n} |\lambda_i|$$

is the spectral radius of the matrix A.

<u>Definition:</u> If A (a_{ij}) is an nxn real matrix, then

$$\|A\| = \sup_{X \neq 0} \frac{\|AX\|}{\|X\|}$$

is the spectral norm of the matrix A. Also, $\|A\| = \sup_{\|X\| = 1} \|AX\|$.

<u>Theorem B.2</u>

If A and B are two nxn matrices, then $\|A\| > 0$ unless A= $\emptyset$, the null matrix. If α is a scalar,

$$\|aA\| = |\alpha| \cdot \|A\|$$

$$\|A + B\| \leq \|A\| + \|B\|$$

$$\|A \cdot B\| \leq \|A\| \cdot \|B\|$$

Moreover,

$$\|AX\| \leq \|A\| \cdot \|X\|$$

for all vectors X, and there exists a non-zero vector Y in R_n for which

$$\|A\,Y\| = \|A\| \cdot \|Y\| \ .$$

<u>Theorem B.3</u>

If $A = (\alpha_{ij})$ is an nxn real matrix, then

$$\|A\| = (\rho(A^T A))^{\frac{1}{2}}$$

<u>Corollary:</u>

If A is an nxn real symmetric matrix, then $\|A\| = \rho(A)$. Moreover, if $g_p(x)$ is any real polynomial of degree p in x, then

$$\|g_p(A)\| = (g_p(A)).$$

<u>Definition:</u> Let A be an nxn real matrix. Then A^r is convergent to zero if the sequence of matrices, A, A^2, A^3, converges to the null matrix 0, and is divergent otherwise.

<u>Theorem B.4</u>

If A is an nxn real matrix, then A^r is convergent if and only if $\rho(A) < 1$.

If A is any real matrix with components a_{11}, a_{12}, ..., a_{nn}, let

$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{ij}| \quad \text{and} \quad \|A\|_\infty \equiv \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}|$$

Then $\rho(A) \leq \|A\|_1$ and $\rho(A) \leq \|A\|_\infty$

<u>Theorem B.5</u>

If M is an arbitrary real nxn matrix with $\rho(M) < 1$, then $I - M$ is non-singular and $(I - M)^{-1} = 1 + M + M^2 + \ldots$, the series on the right converging. Conversely, if the series on the right converges, then $\rho(M) < 1$.

7. APPENDIX C

DESIGN RANDOM VARIABLES

In Chapter 3 of this thesis we introduced the design random variables for the completely randomized design, $\delta'_{k(j)}$ which is 1 if the j-th replicate of treatment k occurs or unit i and zero otherwise. We saw in Chapter 3 that we become involved only in sums over j of $\delta^{i}_{k(j)}$. So it is useful to define

$$\delta^{k}_{2} = \Sigma_j \delta^{i}_{k(j)}$$

and to use the properties of these which we now exposit.

$$\delta^{k}_{i} = \begin{matrix} 1 \text{ if treatment k falls on unit i} \\ 0, \text{ otherwise} \end{matrix} \quad ,$$

$$P(\delta^{k}_{i} = 1) = \frac{r}{N} = \frac{1}{t} \quad ,$$

and for any s+p different units and any two treatments k and k' we have

$$P(\delta^{k}_{(1)} \cdots \delta^{k}_{(s)} \delta^{k'}_{(s+1)} \cdots \delta^{k'}_{(s+p)} = 1) = \binom{r}{s} \quad \binom{r}{p} \qquad \binom{N}{s+p}$$

We shall use the properties of these variables to calculate the first two moments under randomization of the residual sum of squares and cross products in the analysis of variance table.

Throughout this appendix, all expectations are with respect to randomization. Also without loss of generality we may assume $\Sigma x_i = 0 = \Sigma y_j$.

Let
$$Q_{11} = \sum_{k=1}^{t} \left(\sum_{i=1}^{N} \delta_i^k x_i \right)\left(\sum_{j=1}^{N} \delta_j^k y_j \right)$$
$$= \Sigma_k \Sigma_i \delta_i^k x_i y_i + \Sigma_k \Sigma_{ij}^{\neq} x_i y_j \delta_i^k \delta_j^k$$

Then

$$E(Q_{11}) = \Sigma_k \{ \Sigma_i x_i y_i E_k(\delta_i^k) + \Sigma_{ij}^{\neq} x_i y_i E_R(\delta_i^k \delta_j^k) \}$$

$$= t \Sigma x_i y_i \cdot \frac{1}{t} - t \Sigma x_i y_i \frac{1}{t} \cdot \frac{r-1}{N-1}$$

$$= \frac{N-r}{N-1} \cdot \Sigma x_i y_i$$

$$\begin{aligned} Q_{11}^2 = \Sigma_k \{ & \Sigma x_i^2 y_i^2 \delta_i^k + \Sigma_{ij}^{\neq} x_i y_i x_j y_j \delta_i^k \delta_j^k \\ & + 2\{ \Sigma_{ij}^{\neq} x_i y_i (x_i y_j + x_j y_i) \delta_i^k \delta_j^k \\ & + \Sigma_{ij\ell}^{\neq} x_i y_i x_j y_\ell \delta_i^k \delta_j^k \delta_\ell^k \} \\ & + \Sigma_{ij}^{\neq} (x_i^2 y_i^2 + x_i y_i x_j y_j) \delta_i^k \delta_j^k \\ & + \Sigma_{ij\ell}^{\neq} (x_i^2 y_j y_\ell + 2 x_i y_i x_j x_\ell + y_i^2 x_j x_\ell) \delta_i^k \delta_j^k \delta_\ell^k \\ & + \Sigma_{ijm\ell}^{\neq} x_i y_j x_\ell y_m \delta_i^k \delta_j^k \delta_\ell^k \delta_m^k \} \\ & + \Sigma_{kk'}^{\neq} \{ \Sigma_{ij}^{\neq} x_i y_i x_j y_j \delta_i^k \delta_j^{k'} \\ & + 2 \Sigma_{ij\ell}^{\neq} x_i y_i x_j y_\ell \delta_i^k \delta_j^{k'} \delta_\ell^{k'} \\ & + \Sigma_{ij\ell m}^{\neq} x_i y_j x_\ell y_m \delta_i^k \delta_j^k \delta_\ell^{k'} \delta_m^{k'} \} \end{aligned}$$

It is easy to show that

$$\begin{bmatrix} \sum_{i=1}^{N} x_i^2 y_i^2 \\ \Sigma^{\neq} x_i^2 y_i y_j \\ \Sigma^{\neq} x_i^2 y_j^2 \\ \Sigma^{\neq} x_i y_i x_j y_j \\ \Sigma^{\neq} x_i^2 y_j y_\ell \\ \Sigma^{\neq} x_i y_i x_j y_\ell \\ \Sigma^{\neq} x_i x_j y_\ell y_m \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \\ 2 & -1 & 0 \\ 2 & 0 & 1 \\ -6 & 1 & 2 \end{bmatrix} \begin{bmatrix} S_{22} \\ S_{20}S_{02} \\ S_{11}^2 \end{bmatrix}$$

which we write as $\theta = XS$, say.

Let
$$W_1 = \sum_{k=1}^{t} E(\delta_i^k) = t \cdot \frac{1}{t} = 1 ,$$

$$W_2 = 4\sum_{k=1}^{t} E(\delta_i^k \, \delta_j^k) = 4 \cdot t \cdot \frac{r(r-1)}{N(N-1)} = \frac{4(r-1)}{N-1} ,$$

$$W_3 = \sum_{k=1}^{t} E(\delta_i^k \, \delta_j^k) = \frac{(r-1)}{N-1} ,$$

$$W_4 = 2\Sigma_k \, E(\delta_i^k \, \delta_j^k) + \Sigma^{\neq}_{kk'} \, E(\delta_i^k \, \delta_j^{k'}) ,$$

$$= \frac{2(r-1)}{N-1} + t(t-1) \frac{r \cdot r}{N(N-1)} = \frac{2(r-1)}{N-1} + \frac{(N-r)}{N-1} ,$$

$$W_5 = 2\sum_{k=1}^{t} E(\delta_i^k \, \delta_j^k \, \delta_\ell^k) = \frac{t \cdot 2 \cdot r(r-1)(r-2))}{N(N-1)(N-2)} = \frac{2(r-1)(r-2)}{(N-1)(N-2)} ,$$

$$W_6 = 4\Sigma_k \, E(\delta_i^k \, \delta_j^k \, \delta_\ell^k) + 2\, \Sigma^{\neq}_{kk'} \, E(\delta_i^k \, \delta_j^{k'} \, \delta_\ell^{k'}) ,$$

$$= \frac{4(r-1)(r-2)}{(N-1)(M-2)} \; \frac{2t(t-1)r \cdot r(r-1)}{N.(N-1)(N-2)} ,$$

$$= (r-1) \frac{(4(r-2) + 2(N-r))}{(N-1)(N-2)} ,$$

$$W_7 = \sum_{k=1}^{t} E(\delta_i^k \, \delta_j^k \, \delta_k^k \, \delta_\ell^k) + \Sigma^{\neq}{}_{kk'} \, E(\delta_i^k \, \delta_j^k \, \delta_k^{k'} \, \delta_\ell^{k'})$$

$$= t\frac{r(r-1)(r-2)(r-3)}{N(N-1)(N-2)(N-3)} + t(t-1)\,\frac{r(r-1)\;r(r-1)}{N(N-1)(N-2)(N-3)}\;,$$

$$= \frac{(r-1)\;\{(r-2)(r-3) + (N-r)(r-1)\}}{(N-1)(N-2)(N-3)}\;.$$

Then

$$\mathrm{Var}\;(Q_{11}) = E_R(Q_{11}^2) - (E_R Q_{11}^2)$$

$$= W'\theta - \left(\frac{N-r}{N-1}\right)^2 S_{11}^{\;2}\;,$$

where W is obtained from the expectations of the products of $\{\delta_i^{\;k}\}$ given above; and is the following

$$W = \begin{bmatrix} 1 \\ 4\,\dfrac{(r-1)}{(N-1)} \\ (r-1)/(N-1) \\ 2\,\dfrac{(r-1)}{(N-1)} + \dfrac{(N-r)}{(N-1)} \\ 2\,\dfrac{(r-1)(r-2)}{(N-1)(N-2)} \\ (r-1)(4(r-2) + 2(N-r))/(N-1)(N-2) \\ (r-1)\;\{(r-2)(r-3) + (N-r)(r-1)\}/(N-1)(N-2)(N-3) \end{bmatrix}$$

Hence
$$\mathrm{Var}\;(Q_{11}) = AS_{20}\,S_{02} + BS_{11}^2 + CS_{22}\;,$$

where
$$C = \frac{-2(Nr(r-1)(t-1)}{(N-1)(N-2)(N-3)}\;,$$

$$A = \frac{r(t-1)(r-1)(N-2)}{(N-1)(N-2)(N-3)}\;,$$

$$B = \frac{r(t-1)(r-1)(N^2-3N+4)}{(N-1)^2(N-2)(N-3)}\;.$$

If $y_i = x_i$, we will have

$$\mathrm{Var}\;(Q_{20}) = \frac{2r(t-1)(r-1)(N^2-3N+3)}{(N-1)^2(N-2)(N-3)}\;S_{20}^{\;2} - \frac{2Nr(r-1)(t-1)}{(N-1)(N-2)(N-3)}\;S_{40}$$

From symmetrical considerations

$$\text{Cov}\ (Q_{11}, Q_{20}) = A\ S_{20}\ S_{11} + CS_{31}$$

When $y_i = x_i$, we have $\text{Var}\ (Q_{20}) = A\ S_{20}^{\ 2} + CS_{40}$.

$$\text{So Cov}\ (Q_{11}, Q_{20}) = \frac{2r(t-1)(r-1)(N^2-3N+3)}{(N-1)^2(N-2)(N-3)} S_{20}S_{11} - \frac{2Nr(r-1)(t-1)}{(N-1)(N-2)(N-3)} S_{31}\ .$$

$$Q_{20}\ Q_{02} = \Sigma_k \{ \sum_{i=1} x_i^2 y_i^2\ \delta_i^k + \Sigma_{ij}^{\neq} x_i^2 y_j^2 \delta_i^k \delta_j^k$$

$$+ \Sigma^{\neq}\ \{2x_i^2 y_i y_j + 2y_i^2 x_i x_j\}\ \delta_i^k \delta_j^k + \Sigma^{\neq}\ (x_i^2 y_j y_\ell + y_i^2 x_j x_\ell)\ \delta_i^k \delta_j^k \delta_\ell^k$$

$$+ \Sigma^{\neq}\ 2x_i y_i x_j y_j \delta_i^k \delta_j^k + 4\Sigma^{\neq}\ x_i y_i x_j y_\ell \delta_i^k \delta_j^k \delta_\ell^k + \Sigma^{\neq}\ x_i x_j y_\ell y_m \delta_i^k \delta_j^k \delta_\ell^k \delta_m^k \}$$

$$+ \Sigma_{kk'}^{\neq}\ \{\Sigma^{\neq}\ x_i^2 y_j^2 \delta_i^k \delta_j^{k'} + \Sigma^{\neq}\ (x_i^2 y_j y_\ell + y_i^2 x_j x_\ell)\ \delta_i^k \delta_j^{k'} \delta_\ell^{k'}$$

$$+ \Sigma^{\neq}\ x_i x_j y_\ell y_m \delta_i^k \delta_j^k \delta_\ell^{k'} \delta_m^{k'} \} \qquad ,$$

so that

$$\text{Cov}\ (Q_{20}\ Q_{02}) = E_R\ (Q_{20}\ Q_{02}) - \left(\frac{N-r}{N-1} \right)^2 S_{20}\ S_{02}$$

$$= B\ S_{11}^{\ 2} + A\ S_{20}\ S_{02} + CS_{22}$$

When $y_i = x_i$, we have $\text{Var}\ (Q_{20}) = (A + B)\ S_{20}^{\ 2} + CS_{40}$

Here, $C = \dfrac{-2Nr(t-1)(r-1)}{(N-1)(N-2)(N-3)}$,

$$A = \frac{(r-1)}{N-1} + \frac{(N-r)}{N-1} - 2 \left\{ \frac{(r-1)(r-2)}{(N-1)(N-2)} + \frac{(r-1)(N-r)}{(N-1)(N-2)} \right\}$$

$$+ \frac{(r-1)(r-2)(r-3) + (N-r)(r-1)^2}{(N-1)(N-2)(N-3)} - \frac{(N-r)^2}{(N-1)^2} = \frac{2r(t-1)(r-1)}{(N-1)^2(N-2)(N-3)} ,$$

and

$$B = \frac{2r(t-1)(N^2-3N+3)}{(N-1)^2(N-2)(N-3)} - \frac{2r(t-1)(r-1)}{(N-1)^2(N-2)(N-3)} = \frac{2r(t-1)(r-1)}{(N-1)(N-3)}\ .$$

Hence, Cov (Q_{20}, Q_{02})

$$= 2r(t-1)(r-1) \left\{ \frac{S_{11}^{\,2}}{(N-1)(N-3)} + \frac{S_{20}\,S_{02}}{(N-1)^2(N-2)(N-3)} - \frac{NS_{22}}{(N-1)(N-2)(N-3)} \right\}.$$

So we have $$E_R(R_{xx}) = S_{20} - E_R\left(\frac{Q_{20}}{r}\right) = \frac{(N-t)}{N-1} S_{20} \quad ,$$

and $$\mathrm{Var}\,(R_{xx}) = \frac{1}{r^2}\,\mathrm{Var}\,(Q_{20})$$

$$= \frac{2(r-1)(t-1)(N^2-3N+3)}{r(N-2)(N-3)} \cdot \frac{S_{20}}{(N-1)^2} \left\{ 1 - \frac{N(N-1)}{(N^2-3N+3)} \frac{S_{40}}{S_{20}^{\,2}} \right\}.$$

The coefficient of variation, $CV(R_{xx})$, of R_{xx} is

$$\left\{ \frac{2(r-1)(t-1)(N^2-3N+3)}{r(N-2)(N-3)(N-t)^2} \cdot \left(1 - \frac{N(N-1)}{(N^2-3N+3)} \cdot \frac{S_{40}}{S_{20}^{\,2}}\right)\right\}^{\frac{1}{2}}$$

For large N under the general conditions given in Chapter 2, Noether conditions, say

$$\underset{N \to \infty}{L} \frac{S_{40}}{S_{20}^{\,2}} \to 0$$

The coefficient of variation is then approximately $\frac{\sqrt{2(t-1)}}{N}$. The implication of this result is that for N sufficiently large we may consider the ratio, $\frac{R_{xx}}{ER_{xx}}$ to be approximately constant and hence equal to 1.

8. APPENDIX D

CALCULATIONS FOR CODS

The randomization procedure for cods is associated with the design random variables $\{\alpha_{ij}^{k}\}$ (i=1, ..., N; j=1, ..., p; k=1, ..., t) which were introduced in Chapter 4. The α_{ij}^{k}'s were defined as follows:

$$\alpha_{ij}^{k} = \begin{cases} 1 & \text{if the k-th treatment was given to the i-th experimental unit in the j-th period,} \\ 0 & \text{otherwise.} \end{cases}$$

We now introduce a concomitant set of random variables $\{\beta_{ij}^{k}\}$ (i=1, ..., N; j=1, ..., p; k=1, ..., t), such that

$$\beta_{ij}^{k} = \begin{cases} 1 & \text{if the first order residual effect of the k-th treatment occurs in period j for the i-th experimental unit,} \\ 0 & \text{otherwise.} \end{cases}$$

We define $\alpha_{i}^{k} = \Sigma_j \alpha_{ij}^{k}$, $\beta_{i}^{k} = \Sigma_j \beta_{ij}^{k}$

If we assume that k≠k', j≠j' and i≠i' then the basic properties of these random variables are as follows.

$$\Sigma_k \alpha_{ij}^{k} = 1, \quad \beta_{i1}^{k} = 0 \text{ for all } i \text{ and } k,$$

$$\Sigma_k \beta_{ij}^{k} = 1 \quad \text{for } j > 1$$

$$P(\alpha_{ij}^{k} = 1) = \frac{1}{t} = E(\alpha_{ij}^{k})$$

$$P(\alpha_{ij}^{k}\, \alpha_{ij}^{k'} = 1) = 0 = P(\alpha_{ij}^{k}\, , \alpha_{ij'}^{k} = 1)$$

$$P(\alpha_{ij}^{k}\, \alpha_{ij'}^{k'} = 1) = \frac{1}{t(t-1)}$$

$$P(\alpha_{ij}^{k}\, \alpha_{i'j}^{k} = 1) = \frac{N-t}{t^2(N-1)}$$

$$P(\alpha_{ij}^{k}\, \alpha_{i'j'}^{k} = 1) = \frac{N}{t^2(N-1)}$$

and $$P(\alpha_{ij}^{k}\,\alpha_{i'j'}^{k'} = 1) = \frac{(Nt-N-t)}{t^2(N-1)(N-1)}$$

The combinations of the β_{ij}'s have identical probabilities as their counterparts in the α_{ij}'s except for j=1 in which case the probabilities are all zero. The expectations of the combinations are clearly equal to their probabilities.

Also we have

$$P(\alpha_{ij}^{k}\,\beta_{ij'}^{k} = 1) = \begin{cases} \frac{1}{t}, & \text{for } j' = j+1 \\ 0, & \text{otherwise} \end{cases}$$

$$P(\alpha_{ij}^{k}\,\beta_{i'j'}^{k} = 1) = \frac{N-t}{t^2(N-1)}, \quad j' = j + 1 \leq p.$$

$$P(\alpha_{ij}^{k}\,\beta_{ij'}^{k'} = 1) = 0, \quad \text{for } j' = j + 1$$

$$P(\alpha_{ij}^{k}\,\beta_{i'j'}^{k'} = 1) = \frac{Nt-N-t}{t^2(t-1)(N-1)}, \quad i \neq j' \neq j+1, \text{ and } j'=j$$

$$P(\alpha_{ij}^{k}\,\beta_{i'j'}^{k} = 1) = \frac{N}{t^2(N-1)}, \quad j' = j+1 \quad .$$

The other products of the α_{ij}^{k}'s and $\beta_{i'j'}^{k}$'s have identical properties as the corresponding products obtained by substituting $\alpha_{i'j'}^{k'}$ for $\beta_{i'j'}^{k'}$.

In the change over design plan, let u_{ij} be the response of the i-th experimental unit in the j-th period under some basal conditions.

Let $$u_{..} = \Sigma_{ij}\, u_{ij} / Np, \quad i=1,2,\ldots,N; \quad j=1,2,\ldots,p;$$

$$u_{i.} = \Sigma_j\, u_{ij} / p,$$

$$u_{.j} = \Sigma_i\, u_{ij} / N,$$

and let $\Sigma_{ii'}^{\neq}$ denote $\Sigma_i \Sigma_{i' \neq i}$.

Then if we write u_{ij}^{*} for $(u_{ij} - u_{i.} - u_{.j} + u_{..})$

we will have the following algebraic results.

$$\Sigma_i u^*_{ij} = 0 = \Sigma_j u^*_{ij}$$

$$\Sigma_{ij} u^{*2}_{ij} = -\Sigma_i \Sigma^{\neq}_{jj'} u^*_{ij} u^*_{ij'} = -\Sigma_j \Sigma^{\neq}_{ii'} u^*_{ij} u^*_{i'j} = \Sigma^{\neq}_{ii'} \Sigma^{\neq}_{jj'} u^*_{ij} u^*_{i'j'}$$

We omit the * from u^*_{ij} in what follows.

Let

$$T_o{}'R_o = \Sigma_k\{ \Sigma_{ij} u_{ij} (\alpha^k_{ij} - \alpha^k_i /p) \Sigma_{ij} u_{ij} (\beta^k_{i'j'} - \beta^k_{i'}/p)\}$$

$$= \Sigma_{ij} u^2_{ij} \Sigma_k \alpha^k_{ij} \beta^k_{ij}$$

$$+ \Sigma_i \Sigma^{\neq}_{jj'} u_{ij} u_{ij'} \Sigma_k \alpha^k_{ij} \beta^k_{ij'}$$

$$+ \Sigma_j \Sigma^{\neq}_{ii'} u_{ij} u_{i'j} \Sigma_k \alpha^k_{ij} \beta^k_{i'j}$$

$$+ \Sigma^{\neq}_{ii'} \Sigma^{\neq}_{jj'} u_{ij} u_{i'j'} \Sigma_k \alpha^k_{ij} \beta^k_{i'j'}$$

Taking note of the special properties of the $\{\alpha^k_{ij}\}$, $\{\beta^k_{ij}\}$ and the u_{ij} we need only consider the cases where $j' = j+1$ for the first expressions in the summands. The other expression will cancel each other.

So $T_o{}'R_o = \Sigma_i \Sigma_{j<p} u_{i,j} u_{i,j+1} \Sigma_k (\alpha^k_{i,j} \beta^k_{i,j+1})$

$$+ \Sigma_{ii'}\{\Sigma_j u_{ij} u_{i'j} \Sigma_k (\alpha^k_{ij} \beta^k_{i'j}) + \Sigma_j \Sigma_{j'} u_{ij} u_{i'j'} \Sigma_k (\alpha^k_{ij} \beta^k_{i'j'})\}$$

$$E(T_o{}'R_o) = \Sigma_i \Sigma_{j<p} u_{i,j} u_{i,j+1} \{1 + \{N - (N-t)\}/t(N-1)\}$$

$$= \frac{N}{(N-1)} \Sigma_{ij} u_{i,j} u_{i,j+1} \quad .$$

Replacing the β_{ij}'s by the corresponding α_{ij}'s in the summands for $T_o{}'R_o$ we obtain

$$T_o^*{}'T_o^* = \Sigma_k\{\Sigma_{ij} u_{ij} (\alpha^k_{ij} - \alpha^k_i/p)\}^2$$

$$= \Sigma_{ij}\, u_{ij}^2\, \Sigma_k \alpha_{ij}^k + \Sigma_j \Sigma_{ii'}^{\neq} u_{ij}\, u_{i'j}\, \Sigma_k\, (\alpha_{ij}^k\, \alpha_{i'j}^k)$$

$$+ \Sigma_{ii'}^{\neq}\, \Sigma_{jj'}^{\neq} u_{ij}\, u_{i'j'}\, \Sigma_k (\alpha_{ij}^k\, \alpha_{i'j'}^k)$$

So $E(T_o'T_o) = \Sigma_{ij}\, u_{ij}^2\, \{1 -(N - t - N)/t(N - 1)\}$

$$= \frac{N}{(N-1)}\, \Sigma_{ij}\, u_{ij}^2 \;.$$

We note that with the summation $\Sigma_{jj'}^{\neq}$, when $\beta_{ij}^k \beta_{i'j'}^k$ occur in its summand neither j nor j' can take the value 1 replacing the α_{ij}^k's by the corresponding β_{ij}^k's in the summand above for $T_o{}'R_o$ we obtain

$$E(R_o{}'R_o) = \Sigma_k\, E\{\, \Sigma_{ij} u_{ij}\, \beta_{ij} \quad \}^2$$

$$= \frac{N}{(N-1)}\, \Sigma_{ij}\, u_{ij}^2 - \Sigma_i u_{i1}\{1 - (N - t -2N)/t(N-1)\}$$

$$= \frac{1}{(N-1)}\, \{N\, \Sigma_{ij} u_{ij}^2 - \frac{N(t+1)}{t}\, \Sigma_i\, u_{i1}^2\}$$

The variance for $T_o{}'T_o$ is rather complicated. An expression for it was obtained by Welch (1937) for complete cod's under additivity.

Let $D = \Sigma_{ij}\, u_{ij}^4$, $F = (\Sigma_{ij}\, u_{ij}^2)^2$,

$$G = \{\Sigma_i (\Sigma_j\, u_{ij}^2)^2 + \Sigma_j (\Sigma_i\, u_{ij}^2)^2\} ,$$

and $H = \Sigma_{im}^{\neq}\, (\Sigma_j\, u_{ij}\, u_{mj})^2$.

Then Var $(T_o{}'T_o)$

$$= \{\frac{1}{(N-1)(N-2)^2(N-3)}\; [\, 2N^2(N-1)D + (N^4 -4N^3 + 2N^2 + 6N - 6)F$$

$$- 2N(N^2 - 3N + 3)G - 2(N^2 -6N + 6)H]$$

$$+ \frac{\Sigma^{\neq}\, \theta_{kk'}}{\{N(N-1)(N-2)(N-3)\}^2}\; [2N^2(N-1)^2D + 2(2N^2 - 6N +3)F$$

$$- 2N(N-1)(N^2 - 3N + 3)G + 2(N^4 - 6N^3 + 13N^2 - 12N + 6)H]$$

where for any square $\theta_{kk'}$ is the number of rows in which treatment k and k' occur in the same columns in reverse sequence. (See Welch (1937).)

8.1 Calculations for Extra Period cods

In the extra period designs we have p+1 periods instead of p and the treatments in the p-th period is repeated.

We use design random variables $\{\alpha_{ij}^k\}$ and $\{\beta_{ij}^k\}$ as in the ordinary cods. The properties of these random variables are identical in certain combinations for the two designs, but quite different in others. For our purposes we note the following changes.

$$\begin{aligned}
\Sigma_k(\alpha_{i,p}^k\ \alpha_{i,p+1}^k) &= 1 \quad , \\
p(\alpha_{i,p}^k\ \alpha_{i,p+1}^k = 1) &= 1/t \quad , \\
p(\alpha_{i,p}^k\ \alpha_{i',p+1}^k = 1) &= (N-t)/t^2(N-1) \ , \\
p(\alpha_{i,p+1}^k\ \beta_{i,p+1}^k = 1) &= 1/t \quad , \\
p(\alpha_{i,p+1}^k\ \beta_{i',\,p+1}^k = 1) &= (N-t)/t^2(N-1) \ .
\end{aligned}$$

For the extra period cod let $T_o^{*}{}'R_o^{*}$ denote the expression corresponding to $T_o'R_o$ in the ordinary cod just described.

Then the only changes in the symmetry to affect the form of the results come from the first expression in the first and third summands when j=j'=p+1.

So $T_o^{*}{}'R_o^{*} = \frac{N}{(N-1)} \Sigma_{ij}\ u_{i,j}\ u_{i,j+1} + \Sigma_i\ u_{i,p+1}^2 + \Sigma_{ii'}\ u_{i,p+1}\ u_{i',p+1}\ \frac{t}{t(N-1)}$

$$= \frac{N}{(N-1)} (\Sigma_{ij}\ u_{i,j}\ u_{i,j+1} + \Sigma_i\ u_{i,p+1}^2)$$

Similarly, the first expressions in the second and fourth summands when

$j=p$ and $j'=p+1$, etc. give us the relevant changes for $T_o^{*}{'}T_o^{*}$

So $$E(T_o^{*}{'}T_o^{*}) = \frac{N}{(N-1)}\ \Sigma_{ij}\ u_{ij}^2 + 2\Sigma_i\ u_{i,p}\ u_{i,p+1} - 2\Sigma_{ii'}\ u_{i,p}\ u_{i',p+1}\ \frac{t}{t(N-1)}$$

$$= \frac{N}{(N-1)}\ (\Sigma_{ij}\ u_{ij}^2 + 2\Sigma_i\ u_{i,p}\ u_{i,p+1}) \quad .$$

Also, $R_o^{*}{'}R_o^{*}$ has the same form as $R_o{'}R_o$ and so

$$E(R_o^{*}{'}R_o^{*}) = \frac{1}{(N-1)}\ (N\Sigma_{ij}\ u_{ij}^2 - N(\frac{t+1}{t})\ \Sigma_i\ u_{i1}^2\ ) \quad .$$

9. APPENDIX E

A MONTE CARLO SIMULATION OF THE ANALYSIS OF COVARIANCE

In this section we give the results of a small Monte Carlo study in which the usual analysis of covariance model with GMN assumptions,

$$y_{ik} = \mu + u_i + t_k + \beta x_{ik} + \epsilon_{ik}$$

is fitted to the following two situations:

1) $u_i = ax_i + e_i$

and

2) $u_i = ax_i^2$,

where $\{u_i\}$ is the set of basal responses, $\{x_i\}$ is the set of concomitant observations and $\{e_i\}$ is a set of random errors.

We investigated 2 examples of 10 different data sets, for each situation. Each data set consisted of 10 units, $\{(x_i,u_i)\ ,\ i = 1, 2, \ldots, 10\}$, which were arranged into 2(treatment) groups of five units each. For each data set we obtained the mean values and variances, under randomization and the null hypothesis of no treatment differences, of the following statistics:

a) the slope

$$\hat{\beta} = R_{xx}^{-1} R_{xy} \quad ,$$

b) the adjusted treatment difference

$$(\widehat{t_1 - t_2}) = (y_{.1} - y_{.2}) - (x_{.1} - x_{.2})\,\hat{\beta} \quad ,$$

c) the ratio (F) of mean squares of the adjusted sum of squares for treatments and the adjusted sum of squares for error

$$F = \frac{A_{yy}}{E_{yy}} \frac{(N-t-1)}{(t-1)} \quad ,$$

and

d) the ratio (beta) of the adjusted sum of squares for treatments and the sum of the adjusted sum of squares for treatments and the adjusted sum of squares for error

$$W = \frac{A_{yy}}{A_{yy} + E_{yy}} .$$

The mean values and variances of these four statistics under the usual GMN assumptions are as follows:

$$E(\hat{\beta}) = \beta ; \qquad Var(\hat{\beta}) = \sigma^2(1-\rho^2)R_{xx}^{-1} ;$$

$$E(t_k - t_{k'}) = t_k - t_{k'} ; \qquad Var(t_k - t_{k'}) = \sigma^2(1-\rho^2)\left[\frac{2}{r} + (\bar{x}_k - \bar{x}_{k'})' R_{xx}^{-1} (\bar{x}_k - \bar{x}_{k'})\right] ;$$

and, under the null hypothesis

$$E(F_{m,n}) = \frac{n}{n-2} ; \qquad Var(F_{m,n}) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} ;$$

$$E(W) = E[beta(m,n)] = \frac{m}{m+n}$$

and

$$Var(W) = Var[beta(m,n)] = \frac{2mn}{(m+n)^2(m+n+2)}$$

where $m = t-1$ and $n = N-t-p$.

We derived approximations to the means and variances, under randomization, of these statistics in section 3.3 of this thesis. These approximations, except for the variance of $\hat{\beta}$, were shown to have the same limiting values as the corresponding means and variances under GMN assumptions and the null hypothesis. The approximation to the variance of $\hat{\beta}$ under randomization was shown to be much smaller than the

GMN variance of $\hat{\beta}$, their ratio being approximately equal to $\frac{(t-1)}{(N-1)}$ for large values of N.

We obtained the percentage frequencies, over the different randomization plan for each set, with which the GMN critical values at the 10%, 5%, 2.5% and 1% levels for the last three statistics were exceeded.

In examples 2 and 4 the concomitant values $\{x_i, i = 1, \ldots, 10\}$ were $\{\pm 4.5, \pm 3.5, \pm 2.5, \pm 1.5, \pm 0.5\}$. In examples 1 and 3 we altered the x values given above by adding pseudo random N(0,1) random variables $\{e_i, i = 1, \ldots, 10\}$ to obtain $x_i^* = x_i + e_i$. The $\{e_i\}$ were generated by the IMSL package GGNPM with initial seed value equal to 47361. Also for each of examples 3 and 4 in which $u_i = 5x_i + e_i$ an additional 10 pseudo random N(0,1) "errors" $\{e_i, i = 1, \ldots, 10\}$ were generated by the IMSL package.

The mean and variance, under GMN assumptions, of the mean squares ratio (F-ratio) and also the ratio, W (beta) of the adjusted sums of squares are denoted by E and Var respectively. They have the same values in each of the four examples. The mean, under the null hypothesis, of the adjusted treatment difference is zero for both the GMN and randomization models. We give the mean value over all plans in a set, of the absolute value of the difference, $|t_1-t_2|$. In each set the usual GMN estimates of the variances of the estimators of the slope and the adjusted treatment difference vary from plan to plan. We indicate the average of these, under all randomization plans, by GMVAR, and the actual variances, under randomization, by RDVAR.

The results of the Monte Carlo simulations show that the mean and variance of the ratio W (beta) and the mean of the F-ratio are about the same for both models, but the variance of the F-ratio, under randomization, varies by a factor between .46 and 3.42 relative to its variance under GMN assumptions. The average value of the estimator of the slope, over all plans, is given for each data set. The actual value is given with the corresponding data set. The variance of the estimator of the slope varies by a factor between .074 and .150 relative to its variance under GMN assumptions. This agrees rather well with the theoretical factor of 1/9 for the intermediate value of N(=10). The results also show that the GMN 10% critical values for the beta ratio, F-ratio and treatment difference are exceeded with frequencies between 7.9% and 12.7%. But the lower percentage points may be exceeded with frequencies having factors quite different than the postulated GMN values. So our conclusion is that, for small or intermediate values of N, the use of the usual GMN assumptions in the analysis of covariance situation, is at best valid only as a very crude approximation and may lead to quite misleading and incorrect results.

MONTE CARLO SIMULATION OF ANOCOVA I

10 UNITS 2 TREATMENTS U = X*X

GMN RESULTS: E(BETA) =0.1250 VAR(BETA) =0.0219 E(F) =1.4000 VAR(F) =7.8400

MEANS AND VARIANCES UNDER RANDOMIZATION

BETA RATIO		F-RATIO		SLOPE	RDVAR	GMVAR	|T1-T2|	RDVAR	GMVAR
0.121	0.023	1.415	9.764	1.658	0.107	1.194	3.557	24.252	40.399
0.122	0.023	1.457	11.384	-0.368	0.110	1.043	3.448	26.453	41.346
0.119	0.020	1.239	3.595	-1.404	0.098	1.262	3.844	26.094	46.359
0.125	0.025	1.483	8.767	-0.052	0.107	0.786	2.356	27.212	33.458
0.122	0.024	1.620	26.818	-1.063	0.071	0.755	1.894	14.741	19.912
0.124	0.024	1.509	11.775	0.491	0.110	0.713	2.212	19.927	24.922
0.120	0.021	1.300	5.156	0.302	0.059	0.753	1.765	11.848	16.888
0.120	0.021	1.516	20.701	0.124	0.132	1.101	3.307	19.294	32.215
0.121	0.022	1.375	7.775	0.572	0.103	1.200	1.915	39.324	47.338
0.120	0.021	1.306	5.443	-0.589	0.109	1.469	0.904	51.910	59.312

PERCENTAGE EXCEEDING GMN CRITICAL VALUES AT 10%, 5%, 2.5%, AND 1%

BETA RATIO				F RATIO				TRT. DIFF.			
10.3	5.6	2.4	0.8	10.3	5.6	2.4	0.8	10.3	5.6	2.4	0.8
11.1	4.0	2.4	1.6	11.1	4.0	2.4	1.6	11.1	4.0	2.4	1.6
8.7	4.0	1.6	0.0	8.7	4.0	1.6	0.0	8.7	4.0	1.6	0.0
7.9	6.3	5.6	1.6	7.9	6.3	5.6	1.6	7.9	6.3	5.6	1.6
12.7	4.8	2.4	0.8	12.7	4.8	2.4	0.8	12.7	4.8	2.4	0.8
10.3	3.2	3.2	3.2	10.3	3.2	3.2	3.2	10.3	3.2	3.2	3.2
7.9	5.6	2.4	1.6	7.9	5.6	2.4	1.6	7.9	5.6	2.4	1.6
7.9	4.8	2.4	0.8	7.9	4.8	2.4	0.8	7.9	4.8	2.4	0.8
9.5	4.0	2.4	0.8	9.5	4.0	2.4	0.8	9.5	4.0	2.4	0.8
11.1	4.0	1.6	0.8	11.1	4.0	1.6	0.8	11.1	4.0	1.6	0.8

MONTE CARLO SIMULATION OF ANOCOVA 1 U = X*X 10 DATA SETS

X	-4.007	-2.762	-2.171	-1.096	-0.029	0.174	1.217	1.983	5.203	4.862
Y	16.059	7.629	4.715	1.201	0.001	0.030	1.480	3.933	27.069	23.640
SL	1.648									
X	-5.177	-3.868	-2.188	-1.552	0.708	-0.609	2.188	3.328	4.387	3.772
Y	26.800	14.965	4.789	2.409	0.501	0.371	4.789	11.074	19.248	14.224
SL	-0.354									
X	-5.820	-3.528	-2.980	-1.771	-0.671	1.033	1.009	2.469	2.812	4.245
Y	33.875	12.445	8.881	3.136	0.450	1.067	1.019	6.095	7.905	18.018
SL	-1.391									
X	-4.510	-3.649	-3.728	-0.347	-3.298	-0.091	2.780	1.518	4.322	4.455
Y	20.344	13.316	13.898	0.121	10.878	0.008	7.728	2.305	18.683	19.844
SL	-0.062									
X	-4.332	-3.611	-3.433	-1.189	-0.077	1.115	-0.801	2.830	3.471	1.153
Y	18.768	13.038	11.786	1.413	0.006	1.244	0.641	8.008	12.065	1.329
SL	-1.066									
X	-3.964	-3.680	-2.100	-2.074	-2.717	0.096	0.953	2.319	4.207	4.327
Y	15.717	13.545	4.412	4.303	7.381	0.009	0.907	5.376	17.698	18.722
SL	0.476									
X	-3.427	-2.477	-2.823	-1.119	0.900	0.981	-0.381	2.195	2.106	4.219
Y	11.747	6.136	7.969	1.251	0.809	0.962	0.145	4.816	4.437	17.804
SL	0.291									
X	-4.723	-1.687	-1.120	-0.728	0.298	1.496	4.066	3.526	2.718	3.625
Y	22.303	2.846	1.253	0.530	0.089	2.239	16.529	12.432	7.389	13.139
SL	0.155									
X	-3.929	-4.128	-2.812	0.165	1.453	0.447	3.902	2.807	2.295	5.554
Y	15.434	17.042	7.909	0.027	2.112	0.199	15.224	7.877	5.267	30.844
SL	0.570									
X	-3.961	-5.374	-0.891	-3.553	-0.048	-0.116	1.241	2.133	3.387	5.083
Y	15.686	28.876	0.793	12.624	0.002	0.014	1.541	4.550	11.473	25.832
SL	-0.593									

MONTE CARLO SIMULATION OF ANOCOVA 2

10 UNITS 2 TREATMENTS U = X*X

GMN RESULTS: E(BETA) =0.1250 VAR(BETA) =0.0219 E(F) =1.4000 VAR(F) =7.8400

MEANS AND VARIANCES UNDER RANDOMIZATION

BETA RATIO		F-RATIO		SLOPE	RDVAR	GMVAR	|T1-T2|	RDVAR	GMVAR
0.121	0.022	1.354	6.286	-0.000	0.081	0.955	2.854	20.216	31.507
0.121	0.022	1.354	6.286	-0.000	0.081	0.955	2.854	20.216	31.507
0.121	0.022	1.354	6.286	-0.000	0.081	0.955	2.854	20.216	31.507
0.121	0.022	1.354	6.286	-0.000	0.081	0.955	2.854	20.216	31.507
0.121	0.022	1.354	6.286	-0.000	0.081	0.955	2.854	20.216	31.507
0.121	0.022	1.354	6.286	-0.000	0.081	0.955	2.854	20.216	31.507
0.121	0.022	1.354	6.286	-0.000	0.081	0.955	2.854	20.216	31.507
0.121	0.022	1.354	6.286	-0.000	0.081	0.955	2.854	20.216	31.507
0.121	0.022	1.354	6.286	-0.000	0.081	0.955	2.854	20.216	31.507
0.121	0.022	1.354	6.286	-0.000	0.081	0.955	2.854	20.216	31.507

PERCENTAGE EXCEEDING GMN CRITICAL VALUES AT 10%, 5%, 2.5%, AND 1%

BETA RATIO				F RATIO				TRT. DIFF.			
7.9	4.8	1.6	1.6	7.9	4.8	1.6	1.6	7.9	4.8	1.6	1.6
7.9	4.8	1.6	1.6	7.9	4.8	1.6	1.6	7.9	4.8	1.6	1.6
7.9	4.8	1.6	1.6	7.9	4.8	1.6	1.6	7.9	4.8	1.6	1.6
7.9	4.8	1.6	1.6	7.9	4.8	1.6	1.6	7.9	4.8	1.6	1.6
7.9	4.8	1.6	1.6	7.9	4.8	1.6	1.6	7.9	4.8	1.6	1.6
7.9	4.8	1.6	1.6	7.9	4.8	1.6	1.6	7.9	4.8	1.6	1.6
7.9	4.8	1.6	1.6	7.9	4.8	1.6	1.6	7.9	4.8	1.6	1.6
7.9	4.8	1.6	1.6	7.9	4.8	1.6	1.6	7.9	4.8	1.6	1.6
7.9	4.8	1.6	1.6	7.9	4.8	1.6	1.6	7.9	4.8	1.6	1.6
7.9	4.8	1.6	1.6	7.9	4.8	1.6	1.6	7.9	4.8	1.6	1.6

MONTE CARLO SIMULATION OF ANOCOVA 2 U = X*X   10 DATA SETS

X	-4.500	-3.500	-2.500	-1.500	-0.500	0.500	1.500	2.500	3.500	4.500
Y	20.250	12.250	6.250	2.250	0.250	0.250	2.250	6.250	12.250	20.250
SL	0.0									
X	-4.500	-3.500	-2.500	-1.500	-0.500	0.500	1.500	2.500	3.500	4.500
Y	20.250	12.250	6.250	2.250	0.250	0.250	2.250	6.250	12.250	20.250
SL	0.0									
X	-4.500	-3.500	-2.500	-1.500	-0.500	0.500	1.500	2.500	3.500	4.500
Y	20.250	12.250	6.250	2.250	0.250	0.250	2.250	6.250	12.250	20.250
SL	0.0									
X	-4.500	-3.500	-2.500	-1.500	-0.500	0.500	1.500	2.500	3.500	4.500
Y	20.250	12.250	6.250	2.250	0.250	0.250	2.250	6.250	12.250	20.250
SL	0.0									
X	-4.500	-3.500	-2.500	-1.500	-0.500	0.500	1.500	2.500	3.500	4.500
Y	20.250	12.250	6.250	2.250	0.250	0.250	2.250	6.250	12.250	20.250
SL	0.0									
X	-4.500	-3.500	-2.500	-1.500	-0.500	0.500	1.500	2.500	3.500	4.500
Y	20.250	12.250	6.250	2.250	0.250	0.250	2.250	6.250	12.250	20.250
SL	0.0									
X	-4.500	-3.500	-2.500	-1.500	-0.500	0.500	1.500	2.500	3.500	4.500
Y	20.250	12.250	6.250	2.250	0.250	0.250	2.250	6.250	12.250	20.250
SL	0.0									
X	-4.500	-3.500	-2.500	-1.500	-0.500	0.500	1.500	2.500	3.500	4.500
Y	20.250	12.250	6.250	2.250	0.250	0.250	2.250	6.250	12.250	20.250
SL	0.0									
X	-4.500	-3.500	-2.500	-1.500	-0.500	0.500	1.500	2.500	3.500	4.500
Y	20.250	12.250	6.250	2.250	0.250	0.250	2.250	6.250	12.250	20.250
SL	0.0									
X	-4.500	-3.500	-2.500	-1.500	-0.500	0.500	1.500	2.500	3.500	4.500
Y	20.250	12.250	6.250	2.250	0.250	0.250	2.250	6.250	12.250	20.250
SL	0.0									

MONTE CARLO SIMULATION OF ANOCOVA 3

10 UNITS 2 TREATMENTS U = 5X + E

GMN RESULTS: E(BETA) =0.1250 VAR(BETA) =0.0219 E(F) =1.4000 VAR(F) =7.8400

MEANS AND VARIANCES UNDER RANDOMIZATION

BETA RATIO		F-RATIO		SLOPE	RDVAR	GMVAR	|T1-T2|	RDVAR	GMVAR
0.127	0.025	1.610	16.763	4.917	0.003	0.015	0.062	0.522	0.491
0.123	0.022	1.347	5.176	5.044	0.001	0.012	0.091	0.459	0.492
0.126	0.026	1.721	25.007	4.982	0.003	0.021	-0.020	0.806	0.782
0.123	0.022	1.381	6.820	5.100	0.001	0.009	0.323	0.262	0.381
0.127	0.022	1.403	5.609	4.942	0.003	0.019	0.111	0.513	0.493
0.126	0.025	1.549	10.965	5.139	0.005	0.027	-0.103	1.001	0.950
0.123	0.019	1.257	3.300	4.856	0.002	0.016	-0.257	0.282	0.364
0.127	0.026	1.532	9.469	5.184	0.002	0.007	-0.099	0.225	0.215
0.127	0.025	1.570	13.915	5.045	0.002	0.010	0.041	0.406	0.387
0.125	0.019	1.275	3.387	4.923	0.002	0.010	-0.186	0.380	0.406

PERCENTAGE EXCEEDING GMN CRITICAL VALUES AT 10%, 5%, 2.5%, AND 1%

BETA RATIO				F RATIO				TRT. DIFF.			
8.7	7.9	2.4	0.8	8.7	7.9	2.4	0.8	8.7	7.9	2.4	0.8
10.3	6.3	3.2	0.8	10.3	6.3	3.2	0.8	10.3	6.3	3.2	0.8
10.3	4.8	3.2	2.4	10.3	4.8	3.2	2.4	10.3	4.8	3.2	2.4
11.1	3.2	3.2	0.8	11.1	3.2	3.2	0.8	11.1	3.2	3.2	0.8
10.3	5.6	3.2	0.8	10.3	5.6	3.2	0.8	10.3	5.6	3.2	0.8
11.1	4.8	4.0	2.4	11.1	4.8	4.0	2.4	11.1	4.8	4.0	2.4
9.5	5.6	1.6	0.0	9.5	5.6	1.6	0.0	9.5	5.6	1.6	0.0
9.5	7.1	4.0	1.6	9.5	7.1	4.0	1.6	9.5	7.1	4.0	1.6
10.3	7.1	2.4	0.8	10.3	7.1	2.4	0.8	10.3	7.1	2.4	0.8
10.3	3.2	0.8	0.0	10.3	3.2	0.8	0.0	10.3	3.2	0.8	0.0

MONTE CARLC SIMULATION OF ANOCOVA 3 U = 5X + E 10 DATA SETS

X	-4.007	-2.762	-2.171	-1.096	-0.029	0.174	1.217	1.983	5.203	4.862
Y	-18.954	-14.575	-10.077	-3.842	1.634	-0.194	6.273	11.178	25.781	24.104
SL	4.918									
X	-5.177	-3.868	-2.188	-1.552	0.708	-0.609	2.188	3.328	4.387	3.772
Y	-25.777	-21.141	-10.406	-7.583	3.905	-3.303	12.513	15.890	22.500	17.672
SL	5.044									
X	-5.820	-3.528	-2.980	-1.771	-0.671	1.033	1.009	2.469	2.812	4.245
Y	-28.908	-16.226	-16.209	-7.739	-3.963	6.829	3.342	13.908	13.901	20.806
SL	4.983									
X	-4.510	-3.649	-3.728	-0.347	-3.298	-0.091	2.780	1.518	4.322	4.455
Y	-21.225	-18.861	-19.967	-0.841	-16.271	-0.548	14.842	9.003	22.762	22.230
SL	5.101									
X	-4.332	-3.611	-3.433	-1.189	-0.077	1.115	-0.801	2.830	3.474	1.153
Y	-20.875	-18.618	-16.746	-5.024	-1.860	7.478	-3.880	14.286	16.817	4.953
SL	4.942									
X	-3.964	-3.680	-2.100	-2.074	-2.717	0.096	0.953	2.319	4.207	4.327
Y	-20.928	-17.784	-10.088	-12.421	-12.173	-1.439	3.595	11.622	23.146	22.298
SL	5.136									
X	-3.427	-2.477	-2.823	-1.119	0.900	0.981	-0.381	2.195	2.106	4.219
Y	-17.548	-10.120	-14.128	-5.004	4.765	3.896	-1.362	11.005	10.874	20.369
SL	4.856									
X	-4.723	-1.687	-1.120	-0.728	0.298	1.496	4.066	3.526	2.718	3.625
Y	-25.171	-9.445	-5.278	-3.358	1.947	6.436	20.918	17.378	14.339	18.687
SL	5.182									
X	-3.929	-4.128	-2.812	0.165	1.453	0.447	3.902	2.807	2.295	5.554
Y	-20.299	-20.965	-15.410	-0.985	7.905	1.872	20.330	13.081	9.884	27.004
SL	5.045									
X	-3.961	-5.374	-0.891	-3.553	-0.048	-0.116	1.241	2.133	3.387	5.083
Y	-20.302	-25.081	-6.214	-17.947	-0.421	0.032	6.473	10.707	17.168	24.763
SL	4.924									

MONTE CARLO SIMULATION OF ANOCOVA 4

10 UNITS 2 TREATMENTS U = 5X + E

GMN RESULTS: E(BETA) =0.1250 VAR(BETA) =0.0219 E(F) =1.4000 VAR(F) =7.8400

MEANS AND VARIANCES UNDER RANDOMIZATION

BETA RATIO		F-RATIO		SLOPE	RDVAR	GMVAR	|T1-T2|	RDVAR	GMVAR
0.127	0.025	1.587	15.775	4.924	0.003	0.015	0.070	0.535	0.500
0.123	0.021	1.321	4.769	5.012	0.002	0.015	0.048	0.475	0.502
0.127	0.026	1.681	19.930	4.949	0.004	0.023	-0.052	0.804	0.772
0.122	0.022	1.379	7.825	5.080	0.001	0.012	0.299	0.299	0.411
0.129	0.022	1.383	4.402	4.910	0.004	0.014	0.062	0.531	0.464
0.127	0.026	1.722	21.900	5.145	0.006	0.029	-0.075	1.003	0.941
0.122	0.019	1.273	3.840	4.894	0.001	0.011	-0.250	0.297	0.379
0.127	0.026	1.561	11.638	5.171	0.001	0.007	-0.160	0.215	0.220
0.127	0.025	1.499	8.735	4.993	0.002	0.012	-0.018	0.424	0.400
0.123	0.017	1.206	2.454	4.992	0.002	0.014	-0.121	0.411	0.446

PERCENTAGE EXCEEDING GMN CRITICAL VALUES AT 10%, 5%, 2.5%, AND 1%

BETA RATIO				F RATIO				TRT. DIFF.			
8.7	7.9	1.6	0.8	8.7	7.9	1.6	0.8	8.7	7.9	1.6	0.8
11.1	4.8	3.2	0.8	11.1	4.8	3.2	0.8	11.1	4.8	3.2	0.8
7.9	4.8	4.0	2.4	7.9	4.8	4.0	2.4	7.9	4.8	4.0	2.4
10.3	3.2	2.4	0.8	10.3	3.2	2.4	0.8	10.3	3.2	2.4	0.8
11.9	5.6	3.2	0.0	11.9	5.6	3.2	0.0	11.9	5.6	3.2	0.0
10.3	4.8	4.0	2.4	10.3	4.8	4.0	2.4	10.3	4.8	4.0	2.4
7.9	4.0	3.2	0.0	7.9	4.0	3.2	0.0	7.9	4.0	3.2	0.0
10.3	7.9	4.0	1.6	10.3	7.9	4.0	1.6	10.3	7.9	4.0	1.6
10.3	7.1	2.4	0.8	10.3	7.1	2.4	0.8	10.3	7.1	2.4	0.8
7.9	1.6	0.8	0.0	7.9	1.6	0.8	0.0	7.9	1.6	0.8	0.0

MONTE CARLO SIMULATION OF ANOCOVA 4 U = 5X + E   10 DATA SETS

X	-4.500	-3.500	-2.500	-1.500	-0.500	0.500	1.500	2.500	3.500	4.500
Y	-21.417	-18.265	-11.720	-5.862	-0.722	1.448	7.690	13.762	17.267	22.293
SL	4.923									
X	-4.500	-3.500	-2.500	-1.500	-0.500	0.500	1.500	2.500	3.500	4.500
Y	-22.392	-19.299	-11.965	-7.322	-2.133	2.242	9.072	11.751	18.063	21.315
SL	5.012									
X	-4.500	-3.500	-2.500	-1.500	-0.500	0.500	1.500	2.500	3.500	4.500
Y	-22.307	-16.087	-13.808	-6.385	-3.108	4.165	5.795	14.063	17.342	22.082
SL	4.950									
X	-4.500	-3.500	-2.500	-1.500	-0.500	0.500	1.500	2.500	3.500	4.500
Y	-21.172	-18.116	-13.827	-6.605	-2.280	2.407	8.442	13.912	18.650	22.457
SL	5.082									
X	-4.500	-3.500	-2.500	-1.500	-0.500	0.500	1.500	2.500	3.500	4.500
Y	-21.714	-13.064	-12.080	-6.580	-3.975	4.402	7.624	12.637	16.949	21.689
SL	4.911									
X	-4.500	-3.500	-2.500	-1.500	-0.500	0.500	1.500	2.500	3.500	4.500
Y	-23.606	-16.882	-12.086	-9.549	-1.089	0.581	6.332	12.529	19.611	23.164
SL	5.144									
X	-4.500	-3.500	-2.500	-1.500	-0.500	0.500	1.500	2.500	3.500	4.500
Y	-22.911	-15.235	-12.513	-6.911	-2.233	1.493	8.043	12.532	17.842	21.772
SL	4.894									
X	-4.500	-3.500	-2.500	-1.500	-0.500	0.500	1.500	2.500	3.500	4.500
Y	-24.058	-18.510	-12.181	-7.216	-2.041	1.455	8.090	12.249	18.248	23.064
SL	5.169									
X	-4.500	-3.500	-2.500	-1.500	-0.500	0.500	1.500	2.500	3.500	4.500
Y	-23.156	-17.824	-13.849	-9.310	-1.861	2.139	8.322	11.548	15.909	21.736
SL	4.994									
X	-4.500	-3.500	-2.500	-1.500	-0.500	0.500	1.500	2.500	3.500	4.500
Y	-22.999	-15.713	-14.260	-7.682	-2.682	3.114	7.766	12.542	17.732	21.850
SL	4.993									

10. BIBLIOGRAPHY

Anderson, T. W. 1958. An introduction to multivariate statistical analysis. John Wiley and Sons, Inc., New York.

Arnold, H. J. 1964. Permutation support for multivariate techniques. Biometrika 51:65-70.

Aroian, L. A. 1947. The probability function of the product of two normally distributed random variables. AMS 18:265-271.

Atiqullah, M. 1964. The robustness of the covariance analysis of a one-way classification. Biometrika 51:365-372.

Bose, R. C. 1938. On the application of Gabor's fields to the problem of the construction of Hyper-Graeco Latin squares. Sankhya 3: 323-338.

Bross, I. J. 1964. Taking a covariable into account. J. Amer. Stat. Assoc. 59:725-736.

Brown, Jr. B. W. 1980. The crossover experiment for clinical trials. Biometrics 36:69-79.

Chung, K. L. 1974. A course in probability theory. 2nd ed. Academic Press, New York.

Cochran, W. G., K. M. Autrey and C. Y. Cannon. 1941. A double change-over design for dairy cattle feeding experiments. J. Dairy Sci. 24:937-951.

Cox, D. R. 1951. Some systematic experimental designs. Biometrika 38:312-323.

Cox, D. R. 1952. Some recent work on systematic experimental designs. J. Statist. Soc. B 22:172-176.

Cox, D. R. 1956. A note on weighted randomization. Ann. Math. Stat. 27:1144-1151.

Cramer, H. 1946. Mathematical method of statistics. Princeton U. Press, Princeton, New Jersey.

Cramer, H. 1970. Random variables and probability distributions. 3rd ed. Cambridge U. Press, London, England.

Curtiss, J. H. 1942. A note on the theory of moment generating functions. AMS 13:430-434.

Eden, T. and F. Yates. 1933. On the validity of Fisher's Z-test when applied to an example of non-normal data. J. Agric. Sci. 23:6-16.

Federer, W. T. 1964. Tied-double-change-over designs. Biometrics 20:168-181.

Fisher, R. A. 1926. The arrangement of field experiments. J. Min. Agric. Eng. 33:503-513.

Fisher, R. A. 1935. The design of experiments. Oliver and Boyd, London, England.

Fisher, R. A. and F. Yates. 1948. Statistical tables. Oliver and Boyd, Edinbourgh, England.

Fraser, D. 1957. Nonparametric methods in statistics. John Wiley and Sons, Inc., New York.

Fuller, W. A. 1976. Introduction to statistical time series. John Wiley and Sons, Inc., New York.

Grizzle, J. E. 1965. The two-period change-over design and its use in clinical trials. Biometrics 21:467-480.

Hâjek, J. 1960. Some extensions of the Wald-Wofowitz-Noether theorem. Pub. Math. Inst. Hung. Acad. Sci. 5:506-523.

Hiduogton, M. A. 1974. Estimation of regression parameters for finite population. Unpublished Ph.D. thesis, Iowa State University, Ames, Iowa.

Hotelling, H. 1931. The generalization of students ratio. Annals of Math. Stat. 2:360-378.

Kempthorne, O. 1952. Design and analysis of experiments. 2nd ed. John Wiley and Sons, Inc., New York.

Kempthorne, O. 1955. The randomization theory of experimental inference. J. Ann. Stat. Assn. 50:946-967.

Kempthorne, O. 1977. Why randomize? J. of Stat. Plan. and Inf. 1:1-25.

Koch, G. G. 1972. The use of non-parametric methods in the statistical analysis of the two-period change-over design. Biometrics 28:577-584.

Lucas, H. L. 1951. Bias in estimation of error in change-over trials with dairy cattle. J. Agric. Sci. 41:146-148.

Lucas, H. L. 1957. Extra-period Latin square change-over design. J. Dairy Sci. 40:225.

Madow, W. G. 1948. On the limiting distribution of estimates based on samples from finite universes. Inst. of Math. Stat. 19:535-545.

Neyman, J., K. Iwaszkiewiez and St. Koludziejczyk. 1935. Statistical problems in agricultural experimentation. J. Roy. Stat. Soc. Suppl. 2:107-154.

Noether, G. E. 1949. On a theorem by Wald and Wofowitz. Ann. Math. Stat. 20:455-458.

Ogawa, J. 1974. Statistical theory of the analysis of experimental designs. Marcel Dekker, Inc., New York.

Patterson, H. D. 1950. The analysis of change-over trials. J. Agric. Sci. 40:375-80.

Patterson, H. D. 1951. Change-over trials. J. Roy. Stat. Soc. B. 13:256-271.

Patterson, H. D. 1952. The construction of balanced designs for experiments involving sequences of treatments. Biometrika 39:32-48.

Patterson, H. D. and H. L. Lucas. 1959. Extra period change-over designs Biometrics 15:116-132.

Pitman, E. J. G. 1937. Significance tests which may be applied to samples from any population II. J. Roy. Stat. Soc., Supp. 4: 225-240.

Pitman, E. J. G. 1938. Significance tests which may be applied to samples from any population III. Biometrika 29:322-335.

Puri, M. L. and P. K. Sen. 1969. Analysis of covariance on general rank scores. Ann. Math. Stat. 40:610-618.

Quade, D. 1967. Rank analysis of covariance. J. Amer. Stat. Assoc. 62:1187-1200.

Rao, C. R. 1965. Linear statistical inference and its applications. John Wiley and Sons, Inc., New York.

Robinson, J. 1973. The analysis of covariance under a randomization model. J. Roy. Stat. Soc., B. 35:368-376.

Sharma, V. K. 1977. Change-over designs with complete balance for first and second residual effects. The Can. J. of Sta., C. 5(1):121-123.

Snedecor G. W. and W. G. Cochran. 1967. Statistical methods. Iowa State U. Press, Ames, Iowa.

Stevens, W. L. 1938. The completely orthogonalized Latin square. Ana. Eug. 9:82-93.

Varga, R. S. 1962. Matrix iterative analysis. Prentice-Hall, Englewood Cliffs, New Jersey.

Wald, A. and Wolfowitz, J. 1944. Statistical tests based on the permutations of the observations. Ann. Math. Stat. 15:358-372.

Welch, B. L. 1937. On the Z-test in randomized blocks and Latin squares. Biometrika 29:21-52.

Welch, B. L. 1951. On the comparison of several mean values: an alternative approach. Biometrika 38:330-336.

Wilk, M. B. 1953. The randomization analysis of a generalized randomized block design. Unpublished M.S. Thesis. Iowa State U., Ames, Iowa.

Wilk, M. B. 1955. Linear models and randomized experiments. Unpublished Ph.D. thesis. Iowa State U., Ames, Iowa.

Wilk, M. B. and O. Kempthorne. 1955. Fixed, mixed, and random models. J. Amer. Stat. Assoc. 50:1144-1167.

Wilk, M. B. and O. Kempthorne. 1957. Nonadditivities in a Latin square design. J. Amer. Stat. Assoc. 52:218-236.

Yates, F. 1936. Incomplete Latin squares. J. Agric. Sci. 26:301-315.

Williams, E. J. 1949. Experimental designs balanced for the estimation of residual effects of treatments. Aust. J. Sci. Res. A, 2:149-168.

Williams, E. J. 1950. Experimental designs balanced for pairs of residual effects. Aust. J. Sci. Res. A, 3:351-363.

Wishart, J. 1928. The generalized product moment distribution in samples from a normal multivariate population. Biometrika 20:32-52.

11. ACKNOWLEDGEMENTS

I wish to acknowledge and thank Professor Oscar Kempthorne, under whose supervision this thesis was done, for his patience, time and advice; the members of my committee: Professors David Harville, Colby Kegley, William Kennedy, and Leroy Wolins; Rebecca Heymann and Phyllis Carr for many hours of typing and retyping; my parents and various friends whose support has been invaluable; and the faculty and staff at Iowa State University, particularly at the Statistical Laboratory, for their general assistance.