This is to certify that the

thesis entitled

Correction for Dependence

in Two Level Nested Designs

presented by

Suwatana Sookpokakit

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Counseling,

Personnel Services and

Educational Psychology

A C Porter
Major professor

Date 2/23/81

O-7639

CORRECTION FOR DEPENDENCE

IN TWO LEVEL NESTED DESIGNS

By

Suwatana Sookpokakit

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling, Personnel Services
and Educational Psychology

1981

ABSTRACT

CORRECTION FOR DEPENDENCE
IN TWO LEVEL NESTED DESIGNS

By

Suwatana Sookpokakit

A commonly used design in educational research involves hierarchically nested data. Classrooms of students are randomly assigned to receive one of two or more alternative educational treatments. Since dependent variables in educational research are typically defined on students, however, the design results in students nested within classrooms and classrooms nested with treatments.

A fully specified model for the design includes sources of variation for treatments, classrooms and students. Given the fully specified model, the null hypothesis about treatments can be tested using $F_{C:T} = MS_T/MS_{C:T}$. There has been resistance on the part of educational researchers to use the $F_{C:T}$ test statistic because for these studies the test has few degrees of freedom for error and so limited statistical power. As a result, researchers have sometimes turned to a pooled model which ignores classroom variance. By ignoring classroom variance the sources of variation become treatments and students. The apparent test statistic for the treatment null hypothesis is then $F_{S:T} = MS_T/MS_{S:T}$.

The test statistic, $F_{S:T}$, for the pooled model requires that observations on students be independent of each other. Violation of the independence assumption when using $F_{S:T}$ has been shown to yield a test which can be either too liberal or too conservative (Glendening, 1977; Paull, 1950). What is needed, then, are analysis strategies which have greater degrees of freedom error than the $F_{C:T}$ test statistic and which are valid when there is dependence at level of individuals among observations on the dependent variable.

Glendening and Porter (1974) suggested the possibility of using ANCOVA to adjust for positive dependence. They pointed out that the effects of positive dependence could be conceptualized as similar in form to the problem created by confounding in quasi experiments. Index of response is another adjustment strategy which is closely tied to ANCOVA. Thus, index of response was also considered in this investigation.

Four possible situations of dependence were classified for an experimental study that involves two level hierarchically nested data. Dependence could arise because students were not randomly assigned to classrooms (initial dependence) and/or class effects which occur during the study (during-experiment dependence). Crossing these two dichotomous possibilities defined the four situations, one of which was independence.

Investigation of the utility of index of response and ANCOVA was restricted to use of a pre test to adjust for initial dependence.

Further, classroom populations were assumed to be normally distributed on the dependent variable with a common variance but different means.

Results from the investigation indicated that index of response is an appropriate analysis strategy when class effects on the covariate are perfectly correlated with class effects on the dependent variable. Under this condition, the pooled index of response model provides a valid test statistic for the treatment hypothesis with higher power than the F-test from the full index of response model. With an additional stipulation of equality of functional regression slopes at the class level and at the individual level, the pooled ANCOVA model also provides correct adjustment for dependence and a valid test with higher power than the F-test of the full ANCOVA model. The gain in the power through the use of index of response and ANCOVA models was primarily a function of larger degrees of freedom error. Thus, both analysis of variance of index of response and analysis of covariance can be used with a pooled model to provide a more powerful test of the null hypothesis about treatments even when initial dependence is present in post test results. For designs having few classrooms per treatment condition, the increase in power is substantial.

DEDICATED TO

the memory of my father,

Yuan Sookpokakit,

and my brother,

Teeranit Sookpokakit.

ACKNOWLEDGEMENTS

I would like to acknowledge the members of my dissertation committee for their assistance. I am deeply greatful to Professor Andrew C. Porter, my committee chairman, advisor, teacher and friend for his generous and endless advice, encouragement, support and patience. Working with him has greatly increased my knowledge and professional development. Special thanks go to Professors Richard Houang, Dennis Gilliland, Robert Floden and William H. Schmidt for their help and valuable comments.

Working in the Office of Research Consultation (ORC) and with the Institute for Research on Teaching provided me with invaluable experience. Many thanks to my friends and colleagues in the ORC and in the Institute who were very kind to me, especially Professors Joe L. Byers and John H. Schweitzer, who hired me in the ORC, and Professor Andrew C. Porter, Professor Jere Brophy and Mary Rohrkemper, who hired me at the Institute.

I acknowledge with appreciation the support of the American Association of University Women who gave me a fellowship during the 1977-78 academic year.

Finally, I extend heartfelt gratitude to my mother, Sugunya Sookpokakit, to my husband, Rungsit Suwanketnikom, to my sisters, Puangpet Gunyabarn, Benjalug and Raywadee Sookpokakit, and to Mrs. Martha Ward for their sympathy, encouragement and patient support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

In experimental studies concerned with classroom learning and classroom teaching, the sampling frame typically involves at least two levels of nested data. This is a common characteristic of experimental design in education. For example, in many studies individual students are nested within classrooms, and classrooms in turn are nested within treatments. One question often raised when dealing with data from such hierarchically nested designs is what should be the appropriate analysis procedure (Glendening & Porter, 1976; Hannan & Young, 1976). This question is sometimes viewed as the problem of selecting the appropriate unit of analysis (Cronbach, 1976; Peckham, Glass, & Hopkins, 1969; Porter & Chibucos, 1975).

Consider a two-level balanced nested design with fixed treatment effects and one dependent variable, as shown in Figure 1.1. In this design, an equal number of students are nested within each classroom and an equal number of classrooms are nested within each treatment. One concern of the researcher is to test the hypothesis of no treatment effects.

Within the context of analysis of variance, the appropriate linear model for this two-level nested data is:

	T_1			T_2	
c_1	c_2	c_3	c_4	c_5	c_6
s_{11}	s_{21}	s_{31}	s_{41}	s_{51}	s_{61}
s_{12}	s_{22}	s_{32}	s_{42}	s_{52}	s_{62}
s_{13}	s_{23}	s_{33}	s_{43}	s_{53}	s_{63}
s_{14}	s_{24}	s_{34}	s_{44}	s_{54}	s_{64}
s_{15}	s_{25}	s_{35}	s_{45}	s_{55}	s_{65}

T = treatment

c = class

s = student

Figure 1.1:
A Data Matrix of a Two Level Balanced Nested Design

$$Y_{ijk} = \mu + \alpha_i + A^*_{ij} + E_{ijk} \qquad \begin{array}{l} i = 1,2,\ldots,t \\ j = 1,2,\ldots,c \\ k = 1,2,\ldots,s \end{array}$$

Where

Y_{ijk} is an observation of the outcome variable Y on student k in class j receiving treatment i,

μ is the grand mean,

α_i is the effect of being in treatment i,

A^*_{ij} is the effect of being in class j which is nested within treatment i, and

E_{ijk} is an individual error of student k.

In this model, μ and α_i are unknown constants, and A^*_{ij} and E_{ijk} are random variables which are assumed to be independently, identically and normally distributed with zero means and variances $\sigma^2_{A^*}$ and σ^2_E , respectively.

This model, which will be referred to as the "full" model, is considered fully specified because it accounts for both classroom and student sources of variation in the two-level nested data. The analysis of variance table under the "full" model is shown in Figure 1.2. The hypothesis of no treatment effects can be stated as:

$$\text{Ho:} \quad \sum_{i=1}^{t} \alpha_i^2 = 0 \; .$$

Under the hypothesis of no treatment effects, the expected mean square of treatments is equal to the expected mean square of classes nested within treatment or

$$E(MS_T) = E(MS_{C:T}) = s\sigma^2_{A^*} + \sigma^2_E \; .$$

Thus, from Figure 1.2, the ratio $MS_T/MS_{C:T}$ (i.e., the ratio of mean square for treatments over mean square for classrooms nested within treatments) will be the appropriate test statistic for the no treatment hypothesis given the assumptions of the model.

Although the full model is the appropriate model for the nested data, in practice, researchers often assume a model that ignores the classroom grouping variable. That kind of model is under specified and gives misleading results. An example is Anderson's large scale study (1941) of aptitude by treatment interaction. Two teaching methods were compared: drills and meaningful emphasis. Though the

Model: $Y_{ijk} = \mu + \alpha_i + A^*_{ij} + E_{ijk}$

Assumptions: $A^*_{ij} \sim \text{NID}\ (0, \sigma^2_{A^*})$

and $E_{ijk} \sim \text{NID}\ (0, \sigma^2_E)$

Source of Variation	d.f.	E(MS)
Treatments (T)	(t-1)	$\sigma^2_E + s\sigma^2_{A^*} + cs \sum_{i=1}^{t} \frac{\alpha^2_i}{t-1}$
Classes (C:T)	t(C-1)	$\sigma^2_E + s\sigma^2_{A^*}$
Students (S:CT)	tc(S-1)	σ^2_E

Treatments are fixed;

Classrooms are random; and

Students are random.

Figure 1.2: ANOVA Table of the Full Model

methods were delivered in classrooms, Anderson's analysis disregarded class membership and pooled all students within a treatment. Having done so, Anderson found a significant interaction of the investigated teaching methods with the aptitude variable. The interaction was interpreted at the individual level. Cronbach and Webb (1978), reanalyzed Anderson's data by adding the class membership variable. They found that after controlling for the aptitude by treatment interaction at the class level, the interaction at the individual level was no longer significant. This inconsistency of conclusions when analyzing the same set of data by different models or units of analysis is a well known issue in educational evaluation. Discussion of this issue can also be found in publications related to the Follow Through Project (Porter, 1972; Porter & Chibucos, 1975).

Even when the appropriate unit of analysis has been recognized there can be circumstances which prevent use of that unit. Porter (1973) evaluated two teaching strategies, TABA and BASICS, that were delivered in two different schools. Recognizing that the school should be the unit of analysis, Porter chose to use students as the unit of analysis. Porter explained that,

> When identifying the unit of analysis for a study, a crucial consideration is one of independence. This is because all tests of significance are based on the assumption that the units of analysis are independent of each other. Since TABA and BASICS teacher training took place in groups, and since group discussion is one of the most important aspects of the two programs, it follows directly that children in a school were not exposed to the programs in a way such that their exposure represented independent replications of the programs. It could, however, be argued that multiple program schools would have represented independent replications of the program. Unfortunately, with only one program school and one control school, using school as the unit of analysis would have resulted in no tests of significance, i.e., there would have been zero degrees of freedom. (p. 25)

In both the Anderson and Porter examples, students and classrooms were grouped together within treatment levels. By ignoring the natural grouping of students the model assumes that the treatment effects are the only source of systematic variation among the data. This model referred to as the "pooled" model is,

$$Y_{ijk} = \mu + \alpha_i + E^*_{ijk} \qquad \begin{array}{l} i = 1,\ldots t \\ j = 1,\ldots s \\ k = 1,\ldots n \end{array}$$

Where

Y_{ijk} , μ and α_i are defined as before, and

E^*_{ijk} is an individual error on the outcome variable Y for student k.

In this "pooled" model, E^*_{ijk} is a random variable which is assumed to be independently, identically and normally distributed with zero mean and variance $\sigma^2_{E^*}$.

The analysis of variance table for the "pooled" model is shown in Figure 1.3. If the "pooled" model is assumed for the data, the null hypothesis of no treatment effects is tested using $F_{S:T} = MS_T/MS_{S:T}$. Given that the distributional assumptions for the "pooled" model hold, that is, the observations on the students are independent, identically and normally distributed the test statistic $F_{S:T}$ will have a central F distribution with t-1 and t(cs-1) degrees of freedom under the null hypothesis.

To test the hypothesis of no treatment effects, the choice of the "full" model is analogous to choosing classrooms as the units of analysis while the choice of the "pooled" model is analogous to choosing students as the units of analysis (Glendening & Porter, 1976). For either conceptualization, the assumption of independence of observations for the students is the central issue. If this assumption can be met, the choice of the "pooled" model or students as the units of the analysis will be the appropriate one. Otherwise, Model: $Y_{ijk} = \mu + \alpha_i + E^*_{ijk}$

Assumption: $E^*_{ijk} \sim NID\ (0,\ \sigma^2_{E^*})$

Source of Variation	d.f.	E(MS)
Treatments (T)	(t-1)	$\sigma^2_{E^*} + cs \sum_i \frac{\alpha_i^2}{t-1}$
Students (S:T)	t(cs-1)	$\sigma^2_{E^*}$

Treatments are fixed, and

Students are random.

Figure 1.3: ANOVA Table of the Pooled Model

the "full" model using classrooms as the units of analysis will be the appropriate choice.

Independence of observations is one of three assumptions for analysis of variance models. The other two assumptions are normality and homogeneity of variances. Since assumptions are rarely met exactly in real world situations, researchers must be aware of the consequences of violating assumptions. Violations of the assumptions of a model may affect both the significance level and the sensitivity of a test (Cochran & Cox, 1957, p. 91). Though the F tests underlying analysis of variance may be robust with respect to violations of the assumptions of normality and homogeneity of variances under certain circumstances (Glass & Stanley, 1970), they are not robust with respect to violation of the independence assumption (Glendening, 1977; Paull, 1950).

For the "pooled" model, the assumption of independence at the individual level is equivalent to the assumption of no grouping effects (i.e., $\sigma^2_{A^*} = 0$). That is, the individuals (E^*_{ijk}) will be independent if $\sigma^2_{A^*}$ is zero. When class effects are present, however, disturbances for each individual are correlated within a class. The degree of correlation among individual units within a class can be measured by an intraclass correlation coefficient. Similar to $\sigma^2_{A^*}$, when the intraclass correlation coefficient is zero, the condition of independence is met. Analytic and empirical results from Glendening's (1977) work show that dependence affects the validity of the $F_{S:T}$ test leaving the researchers no choice but the full model or the $F_{C:T}$ test.

In comparing the test statistics from the full and pooled models, a distinction in degrees of freedom can be made. While the degrees of freedom for the numerators of both statstics are the same, (i.e., t-1) the degrees of freedom for the denominators are different. For the $F_{C:T}$ ratio, the degrees of freedom for the denominator are t(c-1), which is dependent on the number of treatment levels and the number of classrooms nested within treatments. For the $F_{S:T}$ ratio, the degrees of freedom for the denominator are t(cs-1) which also depends on the number of students within each classroom. Except for the extreme situation where there is only one student in each classroom, t(cs-1) will always be greater than t(c-1).

Peckham et al. (1969) computed power estimates of the $F_{C:T} = MS_T/MS_{C:T}$ test and the $F_{S:T} = MS_T/MS_{S:T}$ test under the condition of individual independence. Their results, reproduced in Table 1.1, showed that the power of the $F_{S:T}$ test is higher than the $F_{C:T}$ test. The gain in power by using the $F_{S:T}$ test is largest when the treatment effect is small and/or the number of levels of the grouping variable (e.g., classrooms) is small. Under independence, the $F_{S:T}$ test has higher power than the $F_{C:T}$ test, so the pooled model or using individual as the unit is the better choice.

Table 1.1: Power Computations Using Groups and Individuals as Unit of Analysis, Given Individuals are Independent

			Power ($\alpha = .05$) Treatment Effect ($\mu_1 - \mu_2$ in sigma unit)		
Analysis Unit	F-Test	d.f. of Denominator	.25	.50	.75
Individuals	$F_{S:T}$	198	.42	.94	.991
Groups	$F_{C:T}$	6	.25	.82	.987

Since many experimental studies can afford only a few classrooms, the difference in power between the full and pooled models is typically substantial. Further, small treatment effects, if any, are also not uncommon in educational research. Peckham et al. (1969) called researchers' attentions to the importance of power when testing educational effects:

> Most studies in education assume that the individual is the unit of statistical analysis. If this assumption is seriously in error, one would find an abundance of significant effects. (Indeed, there would be many instances of contradictory significant effects if random differences were being made "significant" through the utilization of illegitimate power.) History suggests that this is not the typical case. Significant effects related to different methods of instruction are relatively rare, even when we utilize the power afforded by treating the pupil as the unit. All too frequently, rigor in statistical analysis is defined as the avoidance of inaccurate probability statements concerning a Type I error. A more comprehensive notion of rigor would include a similar concern for the avoidance of Type II errors and for the maximum use of available data. (p. 345)

In conclusion, the two available test statistics, the $F_{C:T}$ test and the $F_{S:T}$ test, which underlie the full and the pooled models are not always satisfactory. When treatment effects are small and there are few classrooms, $F_{C:T}$ lacks power. While $F_{S:T}$ has greater degrees of freedom and so greater power, the test requires independence at the student level.

The purpose of this study, then is to consider alternative models which use individuals as the units of analysis and at the same time account for the dependence that may exist in the data. For an alternative model to have utility, the model must yield a test statistic for the null hypothesis about treatments that has a known sampling distribution and greater power than the $F_{C:T}$ test from the full model. The investigation will proceed within the context of a two-level balanced nested design where the available number of classrooms is small. Independence among classrooms will be assumed throughout. For example, the kinds of experimental situations that are considered here are those where intact classrooms are assigned randomly to treatment levels.

Before examining alternative models, however, one needs to understand the conditions under which independence and dependence may occur in an experimental study. Chapter II provides (1) an operational definition of independence among individuals in the two-level balanced nested design, (2) a discussion of an approach for empirically testing the independence assumption, (3) a discussion of an unsuccessful procedure for taking dependence into account, and (4) discussions of index of response and analysis of covariance as alternative models which might allow individuals to be the units of analysis.

Chapter III presents a classification of independence and types of dependence in experimental studies. A discussion of ways in which dependence can occur is also provided. Finally, a particular type of dependence is selected for investigating the utility of index of response and analysis of covariance as alternative models which allow individuals to be the units of analysis.

Chapter IV provides an examination of the index of response and the analysis of covariance strategies. Chapter V presents summary and conclusions.

CHAPTER II

REVIEW OF LITERATURE

An Operational Definition of Independence

Glendening (1977) investigated the independence assumption at the individual level for two-level nested designs using an analysis of variance model. Given independence at the group (classroom) level, Glendening defined independence at the individual level to be a condition where the variance of the group units could be predicted from group size and the variance of the individual units. Consequently, within the context of analysis of variance Glendening stated the operational definition of independence as a condition where the expected mean square of classrooms nested within treatments ($E(MS_{C:T})$) equals the expected mean square of students nested within classrooms and treatments ($E(MS_{S:CT})$). Given classrooms as a random factor, this operational definition applies regardless of whether individuals are a random factor (as shown in Figure 1.2) or a fixed factor (as shown in Appendix A). Only when the individual effects are random, however, is the condition of independence as defined by Glendening equivalent to σ_A^2* being zero. It is this latter case, random individual (student) effects, that is the focus of the present study. Thus, a non-zero value of σ_A^2* implies dependence and a zero value of σ_A^2* implies independence.

<u>Positive Dependence and Negative Dependence</u>

Using the operational definition of independence that she defined, Glendening (1977) classified two types of dependence: positive and negative dependence. She defined positive dependence to be a condition where the expected mean square of classrooms nested within treatments is greater than the expected mean square of students nested within classrooms and treatments (i.e., $E(MS_{C:T}) > E(MS_{S:CT})$). Negative dependence, on the other hand, was defined as a condition where the expected mean square of classrooms nested within treatments is smaller than the expected mean square of students nested within classrooms and treatments (i.e., $E(MS_{C:T}) < E(MS_{S:CT})$). Glendening pointed out that positive dependence was possible whenever individual effects were random. This can be seen from Figure 1.2 where $E(MS_{C:T}) = s\sigma_A^{2}* + \sigma_E^2$ which is equal to or larger than $E(MS_{S:CT}) = \sigma_E^2$. Negative dependence, however, can only occur when the individual effects are fixed. That is from Appendix A, $E(MS_{C:T}) = s\sigma_A^{2}*$ which may be lesser than $E(MS_{S:CT}) = \sigma_E^2$.

Glendening studied analytically and empirically, the effects of dependence on the sampling distributions of $F_{C:T}$ and $F_{S:T}$ test statistics for the full and the pooled models respectively. Her results showed that neither type of dependence among individuals affected the validity of the $F_{C:T}$ test. However, both kinds of dependence affected the validity of the $F_{S:T}$ test. Specifically, within the context of analysis of variance, Glendening found that positive dependence made the $F_{S:T}$ test liberal and resulted in spuriously high power. Negative dependence, on the other hand, made the $F_{S:T}$ test conservative and yielded spuriously low power. Since the sampling distribution of the $F_{S:T}$ test was affected by degrees of dependence, Glendening recommended the use of the full model when dependence was suspected.

The definition of independence in terms of an equality of expected mean squares is helpful in two respects. First, it is an operational definition for the assumption of independence in analysis of variance models. The degrees of dependence in a study can be easily estimated from an analysis of variance table, using a $MS_{C:T}/MS_{S:T}$ ratio. Second, the definition allows for the possibility of either positive or negative dependence. Negative dependence is rarely discussed in the literature.

Preliminary Testing on the Full Model and Conditional Pooling

As has been said, the test statistic $F_{S:T}$ is appropriate only when the assumption of independence is met at the level of individuals. The test statistic $F_{C:T}$ is correct even when individuals are dependent, but the test suffers from low power. If a researcher could decide when individuals are independent, for those situations the best strategy would be to use $F_{S:T}$. When dependence is present, however, the $F_{C:T}$ test must be used. Glendening's operational definition of independence provides a test for dependence and so might be used to guide the researcher in deciding between the pooled and full models.

Preliminary testing for dependence and conditional pooling results in a two-stage testing procedure. For a two-level nested design, as shown in Figure 1.1, the procedure starts with the preliminary test (i.e., $F = MS_{C:T}/MS_{S:CT}$) for independence at the individual level. The null hypothesis of the preliminary test is Ho: $E(MS_{C:T}) = E(MS_{S:CT})$ which is the same as the condition of independence defined by Glendening (1977). If the preliminary test results in rejecting the independence hypothesis, the treatment hypothesis is tested as if the independence assumption were not valid using $F_{C:T}$. If the preliminary test fails to reject the independence hypothesis, the treatment hypothesis is tested as if the independence assumption is valid using $F_{S:T}$ (i.e., the pooled model). Because the choice of test statistic for the null hypothesis about treatments depends on the decision made at the preliminary stage, the consequent test of this two-stage procedure is conditional.

The preliminary test represents an attempt to avoid using a pooled model when dependence is present. Peckham et al. (1969) warned researchers that the preliminary test is not an infallible indication of whether or not independence exists (i.e., either Type I or Type II errors are possible). Thus, they recommend the testing and pooling procedure to be used only when a researcher has an a priori notion that independence among individual observations exists.

To be successful, the preliminary testing and conditional pooling procedure must keep the actual alpha level of the conditional test close to the nominal alpha level. Further, the power of the conditional test must be greater than the power of the unconditional, always correct, $F_{C:T}$ test.

Paull (1950) studied factors that affect the distributional properties of the conditional test when individuals are a random factor. He showed that the distribution of the conditional test involved three main contingencies. The contingencies were (1) magnitude of the dependence (which he defined as $E(MS_{C:T})/E(MS_{S:CT})$), (2) probability of Type I error at the preliminary test and at the consequent test, and (3) number of classes per treatment and number of students per class.

Paull found that when $\sigma^2_A* = 0$, the preliminary test was effective in making the power of the conditional test greater than the power of the unconditional test. However, as σ^2_A* increased from zero, Paull found that the observed alpha level of the conditional test increased to a maximum and then decreased slowly to being equal to the nominal alpha level. Thus, given positive dependence, Paull found that for a fixed probability of a Type I error, the conditional test was generally more liberal than the unconditional test, $F_{C:T}$.

Paull also found that, given dependence, the number of classes per treatment and the number of students per class affected the discrepancy between the distributions of the conditional test and its reference distribution. Paull found that a large number of classes per treatment was desirable in two respects. First, as the number of classes per treatment increased, the power of the preliminary test increased, and pooling inappropriately happened less often. Second, when pooling was prescribed, the pooled mean square was weighted in favor of the correct mean square error, $MS_{C:T}$. As the number of students per class increased, again the power of the preliminary test increased. But when pooling was prescribed under the dependence condition, the wrong mean square, $MS_{S:CT}$, received greater weight. Thus, the effect of class size to the distribution of the consequent test was not simple and largely dependent on the value of positive dependence.

Lastly, Paull examined the effect of increasing the nominal alpha of the preliminary test. Increasing the nominal alpha level of the preliminary test will increase its power and so decrease the frequency of pooling. Less frequent pooling should in turn result in less liberalness of the conditional test under positive dependence. Paull, however, found that increasing the alpha level of the preliminary test did not always result in decreasing the liberalness of the conditional test. From his finding, there was a critical alpha level above which increasing the alpha level of the preliminary test resulted in increasing the liberalness of the consequent test.

To stabilize the disturbances between the distributions of the conditional and its reference distribution for a given amount of positive dependence, Paull finally recommended $2F_{50}$ be used as the critical value for the preliminary test. (F_{50} is the 50th percentile in the central F distribution with (c-1) and tc(s-1) degrees of freedom). However, it is not clear how $2F_{50}$ stabilizes the disturbances between the distribution of the conditional test and its reference distribution under positive dependence. If the goal is to increase the power of the preliminary test, Glendening (1977) pointed out that "taking twice the critical value given a large alpha of .50

had the same effect as selecting a small alpha level in the first place."

Results from Paull's study indicated that positive dependence is an important threat to the validity of the conditional test. Only under the condition of independence and under extreme positive dependence was the conditional test valid. Unfortunately, given an intermediate value of positive dependence, the preliminary test can mistakenly prescribe pooling and make the validity of the conditional test questionable.

Glendening (1977) examined the utility of preliminary testing and conditional pooling under both positive and negative dependence. Analytically and empirically, Glendening's findings opposed the use of the procedure. Similar to Paull (1950), Glendening concluded that given an intermediate value of dependence, the preliminary F test was not sensitive enough to help a researcher guard against having an undesirably distorted probability of Type I error for the conditional F test.

The preliminary testing and conditional pooling procedure is not likely to be useful in experimental studies in education since effectiveness of the procedure is limited by its insensitivity to moderate degrees of dependence. Degrees of dependence that occur within educational research studies usually range from small to moderate. Glendening (1977) investigated two research studies in elementary schools. She found that on achievement scores, the degree to which classroom variation accounted for total variation among students ranged consistently from 20 to 50 percent.

Analysis strategies that can account for moderate degrees of dependence would be useful for analyzing nested data. The following sections present discussions of such strategies. First the use of a quasi-F ratio to correct for dependence is considered. The chapter concludes by considering the possibility of using adjustment strategies which rely on information provided by a covariate.

A Quasi-F Statistic

There is a class of factorial designs for which analysis of variance does not provide direct tests of certain hypotheses even when all assumptions of the model have been met (e.g., Kirk, 1968; Winer, 1972). For example, the fixed main effect is not directly testable in a completely crossed factorial design having one fixed and two random factors with random replication in each cell of the design. In such situations a quasi-F statistic is sometimes constructed to provide an appropriate test of the fixed main effect. As will be seen, there was some reason to believe that a quasi-F test might hold potential for providing a valid test statistic with greater power than $F_{C:T}$ in situations of positive dependence.

A linear combination of independent chi-square statistics is distributed approximately as a chi-square distribution with degrees of freedom estimated from a function of mean squares and degrees of freedom (Satterthwaite, 1941, 1946). For example, let $\chi^2_{\nu_1}$, $\chi^2_{\nu_2}$ and $\chi^2_{\nu_3}$ be chi-square statistics which are independently distributed as central chi-square distributions with ν_1, ν_2 and ν_3 degrees of freedom respectively. Also, let MS_1, MS_2 and MS_3 be mean squares associated with $\chi^2_{\nu_1}$, $\chi^2_{\nu_2}$ and $\chi^2_{\nu_3}$. Then, $\chi^2_{\nu_1} + \chi^2_{\nu_2} - \chi^2_{\nu_3} \sim \chi^2_{\nu_4}$ where ν_4 is estimated by

$$\nu_4 = \frac{= MS_1 + MS_2 + (-MS_3)}{\frac{MS_1^2}{\nu_1} + \frac{MS_2^2}{\nu_2} + \frac{(-MS_3)^2}{\nu_3}}$$

A quasi-F statistic is simply the ratio of two estimated variances, at least one of which has been formed through a linear combination of independent mean squares (Hudson & Krutchkoff, 1968; Galor & Hopper, 1969). The potential utility of the quasi-F statistic in situations of positive dependence can be seen by returning to the $F_{S:T}$ statistic and its inadequacies. Given positive dependence and the null hypothesis for treatment,

$$E(MS_T) = \sigma^2_E + s\sigma^2_{A^*} ,$$

while

$$E(MS_{S:T}) = \sigma^2_E + \frac{(c-1)}{(cs-1)} \sigma^2_{A^*} ,$$

(Glendening, 1977). Applying the strategy of constructing a quasi-F test the following ratio can be formed

$$F' = \frac{MS_T - MS_{C:T} + MS_{S:CT}}{MS_{S:CT}} .$$

The expected values of the numerator and the denominator of F' are equal under the null hypothesis of no treatment effects. This equality can be seen by recalling that

$$E(MS_T) = \sigma^2_E + s\sigma^2_{A^*} + sc \sum_{i=1}^{t} \frac{\alpha^2_i}{t-1} ,$$

$$E(MS_{C:T}) = \sigma^2_E + s\sigma^2_{A^*} ,$$

and $$E(MS_{S:CT}) = \sigma^2_E .$$

Thus,

$$E(MS_T - MS_{C:T} + MS_{S:CT}) = \sigma_E^2 + sc \sum_{i=1}^{t} \frac{\alpha_i^2}{t-1}$$

and so, under the null hypothesis of no treatment effects, $E(MS_T - MS_{C:T} + MS_{S:CT})$ and $E(MS_{S:CT})$ estimate the same parameter, σ_E^2 .

The apparent reference distribution for F' is a central F distribution with first and second degrees of freedom, ν_1 and ν_2 , provided by

$$\hat{\nu}_1 = \frac{MS_T - MS_{C:T} + MS_{S:CT}}{\frac{(MS_T)^2}{(t-1)} + \frac{(-MS_{C:T})^2}{t(c-1)} + \frac{(MS_{S:CT})^2}{tc(s-1)}}$$

and $\nu_2 = tc(s-1)$.

Unfortunately, F' is not a legitimate quasi-F ratio. As can be seen the complex variance of the numerator and the simple variance of the denominator are not independent; $MS_{S:CT}$ is used in both places. The F' statistic which appeared to hold promise as a test with greater power than $F_{C:T}$ in situations of dependence and with no cost of additional information, has been found to lack a known distribution.

In the search for a more powerful test of treatment effects in situations of dependence, two approaches have been considered. Both the procedure of preliminary testing for dependence and conditional pooling and the building of a quasi-F statistic have been seen to yield test statistics with unknown distributions. As an alternative to these two approaches, perhaps there exists ways to adjust the individual observations for the dependence they reflect. This possibility is considered in the following section. For the approach to be successful in correcting dependence, the approach must meet two criteria: (1) provide adjustment that leads to the condition of independence on the adjusted observations; and (2) provide a statistical test that has a known sampling distribution and has higher power than the $F_{C:T}$ test.

Using a Covariate to Adjust for Positive Dependence

Glendening and Porter (1974) suggested the possibility of using analysis of covariance (ANCOVA) to adjust for positive dependence. The effects of positive dependence, as they pointed out, can be conceptualized as similar in form to the problem of confounding in quasi experiments. Given positive dependence, the problematic variance of the class effects (σ^2_A*) exists in the expected mean square for treatments (i.e., $E(MS_T) = sc \sum_{i=1}^{t} \frac{\alpha_i^2}{t-1} + s\sigma^2_A* + \sigma^2_E$. Thus σ^2_A* might be removed from $E(MS_T)$ by ANCOVA procedures conceptually leaving the adjusted observations free from positive dependence.

Glendening and Porter conjectured that removing positive dependence from the individual observations was possible if (1) the covariate, X, and the dependent variable, Y, have equal degrees of dependence (i.e., $E(MS^X_{C:T})/E(MS^X_{S:CT}) = E(MS^Y_{C:T}/E(MS^Y_{S:CT}))$, and (2) the correlation of X and Y within classrooms was equal to one.

Following their lead, the present study investigated the possibility of ANCOVA as an analysis strategy in situations of positive dependence. Their rationale for ANCOVA, however, applies equally well to an index of response strategy (i.e., Z = Y - KX where Z is the index of response). Conceptually, ANCOVA and index of response are closely tied, the main difference being that ANCOVA estimates the value of K while index of response requires that K be set a priori. The investigation in this study started with an examination of the potential of an index of response model. Only given a reasonable solution using the index of response strategy would the utility of ANCOVA be investigated. If no solutions for the dependence problem is found using a model of index of response, it is unlikely that a solution exists in the corresponded but more complex ANCOVA model.

Unlike the preliminary testing and conditional pooling strategy or the quasi-F approach, the index of response and the ANCOVA strategies require information in addition to observations on the dependent variable. To examine the index of response and ANCOVA strategies, it is important to distinguish between when and how dependency may arise in a study. These distinctions facilitate understanding the causes of dependence and so may help to inform the selection of an appropriate covariable.

In summary, there were two main related tasks that this study intended to accomplish:

(1) to classify situations of independence and dependence in experimental studies that assume independence at the group level, and

(2) to investigate the possibility of using an index of response and its corresponded ANCOVA models to correct for positive dependence among the individual units.

CHAPTER III

SITUATIONS OF INDEPENDENCE AND DEPENDENCE

As has been stated, the problems created by positive dependence are analogous to the problem created by confounding variables in quasi experiments. If there are no classroom effects in the two level nested design under consideration, there is independence at the level of individuals. Thus, an attempt to remove dependence from the data can be viewed as an attempt to adjust classroom effects to zero. If classrooms were confounded with treatments in a quasi experiment, the same adjustment procedure which removed dependence would also remove the confounding effect of classrooms in that quasi experiment.

Two essential assumptions to the success of an "adjustment" strategy in removing confounding effects in a quasi-experiment are correct specification of the covariable and proper specification of the analytic model (Olejnik, 1977). Therefore, correct specification of the covariate and appropriate specification of the analytic model are also required in this study. The problems inherent in the general assessment of these two assumptions are formidable (Cronbach, Pagosa, Floden, & Price, 1977; Olejnik, 1977).

Importantly, to be successful in the adjustment, the two assumptions place the responsibility on the researcher. Beside being knowledgeable in the substantive aspect of his experiment, a researcher must understand design problems that lead to situations of dependence. When dealing with hierarchically nested data, it is helpful for a researcher to understand "when" and "how" dependence problems can arise in an experiment. Given this knowledge, the researcher may be able to avoid problems of dependence through careful design of his study. Further, when experimental control is not possible for a certain type of dependence, the researcher will be in a better position to specify and measure the dependence (i.e., correctly specify the covariate and analysis model).

The intention of this chapter is to provide better understanding of correct specification of a covariate. The discussion starts with a classification of independence and dependence situations in an experimental study that involves two level hierarchically nested data. As stated previously, independence at the class level is assumed. At the outset of the study, two conditions are identified: with and with<u>out</u> random assignment of students to classrooms. During the experimental period, an additional two conditions are identified: <u>no</u> class effects and class effects. Together, these two dichotomous dimensions classify four possible situations in an experimental study. Each situation will be discussed to generate potential sources of dependence (which in turn will be used to inform selection of covariates).

<u>Initial Dependence and During-Experiment Dependence</u>

The following classification of situations of independence and dependence in an experimental study is similar to how Porter (1972) classifies situations of confounding variables in a quasi-experimental study. Two important questions are "when" and "how" the dependence arises. Answers to these two questions can serve to inform the design of an experiment involving hierarchically nested data.

When measurement of an outcome variable is taken immediately after the intervention, there are two places where dependence can arise, 1) at the outset of the experiment and 2) during the experiment. The first kind of dependence, initial dependence, is likely to occur from the lack of random assignment of analysis units to classrooms. To protect against initial dependence in a nested design, random assignment to classrooms is essential.

The second type of dependence, occurring during the experiment, may arise from interactions among analysis units while they receive treatments (Cox, 1958). For example, when a treatment is delivered to intact classrooms, common class experiences may reduce the variability of students within the same class. Cronbach (1976) suggested that unless a researcher is prepared to assume that students within an intact classroom are treated independently by a treatment and respond independently from each other, students are not independent. While Cronbach's definition of dependence is in terms of process, the process he identifies may also result in dependence as it has been defined here in terms of observations on the dependent variable. Since during-experiment dependence occurs after the start of an experiment, it can exist even in a completely randomized experiment.

Four Possible Situations of Dependence and Independence

In summary, dependence can arise because students were not randomly assigned to classrooms (initial dependence) and/or because of classroom effects which occur during the study (during-experiment dependence). Crossing these two dichotomous possibilities defines four situations (Figure 3.1). Situation I is the only situation that does not violate the assumption of independence. Situations II and III suffer from initial and during-experiment dependence respectively. In Situation IV, both types of dependence exist.

Before Experiment	During-Experiment	
	No Class Effects	Has Class Effects
With random assignment of students to classes	I	III
With<u>out</u> random assignment of students to classes	II	IV

I = independence situation

II = initial dependence situation

III = during-experiment dependence situation

IV = initial plus during-experiment dependence situation

Figure 3.1: Four Possible Situations of Independence and Dependence in a Two Level Hierarchically Nested Design

Situation I: Independence. Situation I, the independence situation, is possible when at the outset of an experiment students are randomly assigned to classes and during the course of intervention treatments are delivered independently to individual students. An example of this kind of treatment might be a study of different types of individualized instruction. While an experiment dealing with individualized instruction might be conducted in classroom settings, students might still be required to react individually to the instructional packets. Under Situation I, $E(MS_{C:T})$ and $E(MS_{S:CT})$ are equal. An analysis of variance using the pooled model ($F_{S:T}$) is the best strategy for testing the no treatment effect hypothesis (Glendening, 1977; Cronbach, 1976).

Situation II: Initial Dependence. Situation II includes initial dependence only. In this situation, common experience effects (i.e., class effects) during the experiment are controlled but students are not randomly assigned to classrooms. When students are not randomly assigned to classrooms, the samples of students within classes are best thought of as coming from distinct populations (Cronbach et al., 1977). In general, the class populations will have different distributions on the dependent variable regardless of treatment effects. Consequently, $E(MS_{C:T})$ is greater than $E(MS_{S:CT})$, and one has the problem of positive dependence.

For Situation II, the value of σ^2_{A*} is solely a function of initial dependence. If dependence is to be removed from the data through index of response or ANCOVA strategies the covariable must reflect initial class differences that are predictive of class differences in the dependent variable. For experiments which fit Situation II, a pre test seems potentially the best covariable for removing dependence. Nevertheless, the utility of a pre test for removing dependence will be a function of the relationships between the type of natural growth on the dimension measured by the pre and post tests and the relationship between pre and post tests reflected in the analytic model (Bryk & Wiseburg, 1977).

Situation III: During-Experiment Dependence. Situation III suffers from dependence which occurs during the experiment. In this situation, a researcher is able to randomly assign students to classes, creating initial independence. However, the researcher may not be able to eliminate, through design, effects of common class experiences (i.e., class effects) that occur during the course of the intervention.

Webb (1977) perceived class effects as group process effects that cause dependence on the outcome variable dimension. He explained that knowledge of group processes in a particular class is crucial for understanding and estimating the degrees of dependence in a classroom. Since the knowledge of group processes would guide a researcher as to where to look for potential covariate, Webb concluded that studying group process may be the only way to get at this dependence.

Group process effects, however, are a function of complex and global variables including effects due to subject matter, teacher effectiveness, teaching strategies, student interactions and classroom milieu which are not part of the treatment effects. For example, differences in teacher effectiveness may result in classroom effects and so dependence at the level of individuals. But predicting differential teacher effectiveness is a substantive problem in its own right (Dunkin & Buddle, 1974). In the literature of research in classroom teaching and classroom learning, other group process variables identified above have been placed in the black box of classroom setting about which little is known (e.g., Bloom, 1976; Duncan & Biddle, 1974).

Confusion in specifying the class effects to be removed to create independence increases when one considers the possibility that some class effects may also be part of treatment effects. For example, a researcher investigating effects of teaching methods on achievement of elementary students may include differences in instructional skills (one kind of class effects) of teachers as a part of the treatment effects. However, differences in management skills (another kind of class effects) of which he may not be aware could cause dependence among students within the same classrooms. Further, since during-experiment dependence occurs within the same period as the treatments, potential interactions between dependence and treatments must be considered. Thus, in Situation III, it is important for a researcher to understand and carefully describe what constitutes the treatment effects and what may be nuisance variables that would induce dependence among units of analysis. If class distinctions are not made between treatment effects and nuisance variables, an adjustment to create independence might at the same time remove part of the treatment effects from the data as well.

Clearly, the criteria for selecting covariates to remove dependence in Situation III are quite different from the criteria for Situation II. In Situation II, the covariate must predict initial differences while, in Situation III, the covariate must predict effects of classrooms that are not part of treatment effects. The use of a pre test as an adjustment variable in Situation III is unlikely to be helpful. As Porter (1972) stated:

> Although I believe pre tests to be the best predictors of initial differences it does not necessarily follow that they are also the best predictors of differences that occur in the dependent variable dimension during program participation which are not a function of program participation. My reasoning is that initial differences are a function of all that has preceded the study in the life of the child, while differences that occur during the study other than due to program most likely are primarily a function of the child's environment at that time. (p. 19)

<u>Situation IV: Initial and During-Experiment Dependence</u>. The last situation, Situation IV, is the most complicated. This situation suffers from initial dependence and during-experiment dependence. An experimental study that falls into this category has two design problems. First, it lacks random assignment of students to classes, and, second, it deals with treatments that are delivered in group settings. Thus, both the magnitude of initial dependence and the magnitude of during-experiment dependence are contained in σ_A^{2*}. Specification of an appropriate covariable or covariables to adjust σ_A^{2*} to zero (independence) is extremely complicated. Application of structural equation strategies to classify important causal variables in longitudinal data (Schmidt, 1975) may be helpful in identifying

appropriate covariables in Situation IV. In addition to all the difficulties identified for Situations II and III, a researcher in Situation IV must not ignore the possibility of interactions between initial differences and class process differences.

Having described three types of situations in which there is dependence at the level of individuals the present study limits its focus to the use of the index of response and ANCOVA models in Situation II. There are two reasons for focusing on Situation II. First, the problem of modeling dependence in a design is not well understood in the literature. To be able to understand this modeling problem, one should start with a simple case. Second, the initial dependence problem, underlying Situation II, is comparable to the initial confounding problem in a quasi-experimental study. The literature contains a great deal of discussion about the utility of index of response and ANCOVA for controlling the effects of confounding in quasi-experiments.

CHAPTER IV

CORRECTIONS FOR INITIAL DEPENDENCE USING INDEX OF RESPONSE AND ANALYSIS OF COVARIANCE STRATEGIES

Given dependence among individuals the $F_{C:T}$ statistic provides a valid test of the null hypothesis about treatments but for most educational research its power is low. The goal of this study is to explore alternative tests for treatment effects when dependence is present. These alternatives will be evaluated against the criteria of an actual Type I error in agreement with the nominal value and power that exceeds the $F_{C:T}$ test.

The review of literature in Chapter II provided two helpful conceptions for the investigation. The first conception was an operational definition of independence and dependence. Given a two level balanced nested design where both the group effect (A^*_{ij}) and the individual effect (E_{ijk}) are random, independence is defined as the condition when $\sigma^2_{A^*}$ is equal to zero. Dependence is defined as the condition when $\sigma^2_{A^*}$ is greater than zero. The second conception was to recast dependence as equivalent to the effect of confounding at the class level. The reconceptualization suggests approaches to analysis that use adjustment strategies comparable to those used to remove the effects of confounding in quasi-experimental studies. Two such adjustment strategies that are investigated here are index of response and analysis of covariance.

In dealing with adjustment strategies, both correct specification of adjustment variables (covariables) and proper specification of the analytic model are necessary. Chapter III provided a classification of independence and dependence situations which could facilitate proper specification of covariates. In this chapter the parameters of the two analytic models are specified for the situation of initial dependence.

Modeling Situation II Dependence

Situation II represents designs in which students were not randomly assigned to classrooms. Thus as has been noted, each classroom must be considered a separate population. Given normality, these populations may differ in both means and variances. The set of Situation II designs considered here is restricted, however, to classroom populations which differ only in terms of means. A linear system of structural equations consistent with the above restriction can be specified as follows:

Model:

$$Y_{ijk} = \mu + \alpha_i + A^*_{ij} + E_{ijk} \qquad \begin{array}{l} i = 1,2,\ldots,t \\ j = 1,2,\ldots,c \\ k = 1,2,\ldots,s \end{array}$$

$$X_{ijk} = \mu^X + A_{ij} + V_{ijk}$$

$$A^*_{ij} = B_1 A_{ij} + H_{ij}$$

and $E_{ijk} = B_2 V_{ijk} + G_{ijk}$

where

Y_{ijk} is a post test score of individual k in class j receiving treatment i;

μ is the grand mean of Y;

α_i is the effect of treatment i;

A^*_{ij} is the effect (as measured by post test) of being in class j which is nested in treatment i;

E_{ijk} is the specification error at post test;

X_{ijk} is a pre test score of individual k in class j receiving treatment i;

μ^X is the grand mean of X;

A_{ij} is the initial effect (as measured by pre test) of being in class j which is nested in treatment i;

V_{ijk} is the specification error at pre test;

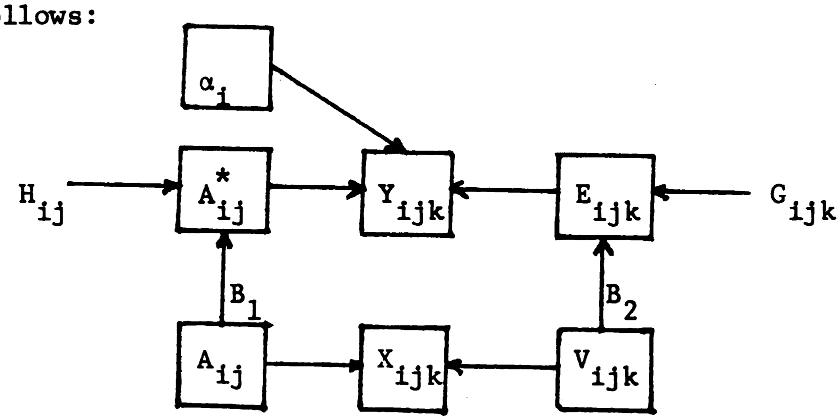
H_{ij} is the residual of A^*_{ij} given A_{ij};

G_{ijk} is the residual of E_{ijk} given V_{ijk};

B_1 is the structural regression coefficient that predicts A^*_{ij} from A_{ij};

and B_2 is the structural regression coefficient that predicts V_{ijk} from E_{ijk}.

This structural model can be represented by a causal diagram as follows:



Additional distributional assumptions of the model are

$$A_{ij} \sim \text{NID}(0, \sigma_A^2)$$

$$H_{ij} \sim \text{NID}(0, \sigma_H^2)$$

$$V_{ijk} \sim \text{NID}(0, \sigma_V^2)$$

$$G_{ijk} \sim \text{NID}(0, \sigma_G^2)$$

and $\sigma_{AH} = \sigma_{VG} = \sigma_{AV} = \sigma_{AG} = \sigma_{HV} = \sigma_{HG} = 0$

where σ_p^2 denotes the variance of variable p and σ_{pq} denotes the covariance of variables p and q.

From the above assumptions of the model, the covariance structures of Y, X and XY are in the form of super diagonal matrices. Specifically,

$$\Sigma^Y = \begin{bmatrix} M^Y & \phi & \cdots & \phi \\ \phi & M^Y & & \vdots \\ \vdots & & \ddots & \phi \\ \phi & \cdots & \phi & M^Y \end{bmatrix}$$

where

Σ^Y is the covariance matrix of Y with dimentions of tcs x tcs;

M^Y is the covariance matrix of students within a classroom only with dimensions of s x s;

$$M^Y = \sigma_A^2{}^* \underline{1}\,\underline{1}' + \sigma_E^2 I = \begin{bmatrix} (\sigma_E^2 + \sigma_A^2{}^*) & \sigma_A^2{}^* & \cdots & \sigma_A^2{}^* \\ \sigma_A^2{}^* & (\sigma_E^2 + \sigma_A^2{}^*) & & \vdots \\ \vdots & & \ddots & \sigma_A^2{}^* \\ \sigma_A^2{}^* & \cdots & \sigma_A^2{}^* & (\sigma_E^2 + \sigma_A^2{}^*) \end{bmatrix}$$

and ϕ is a null matrix of s x s dimensions.

$$\Sigma^X = \begin{bmatrix} M^X & \phi & \cdots & \phi \\ \phi & M^Y & & \vdots \\ \vdots & & \ddots & \phi \\ \phi & \cdots & \phi & M^X \end{bmatrix}$$

where

Σ^X is the covariance matrix of X with dimensions of tcs x tcs;

M^X is the covariance matrix of students within a class on X with s x s dimensions.

$$M^X = \sigma_A^2 \underline{1}\,\underline{1}' + \sigma_V^2 I = \begin{bmatrix} (\sigma_A^2 + \sigma_V^2) & \sigma_A^2 & \cdots & \sigma_A^2 \\ \sigma_A^2 & (\sigma_A^2 + \sigma_V^2) & & \vdots \\ & & & \sigma_A^2 \\ \vdots & & \ddots & \\ \sigma_A^2 & \cdots & \sigma_A^2 & (\sigma_A^2 + \sigma_V^2) \end{bmatrix}$$

and

$$\Sigma^{XY} = \begin{bmatrix} M^{XY} & \phi & \cdots & \phi \\ \phi & M^{XY} & & \vdots \\ \vdots & & \ddots & \phi \\ \phi & \cdots & \phi & M^{XY} \end{bmatrix}$$

where

Σ^{XY} is the covariance matrix of the cross product XY with dimensions tcs x tcs;

M^{XY} is the covariance matrix of students within a classroom on XY with dimensions s x s;

$$M^{XY} = \sigma_{AA}{}^* \underline{1}\,\underline{1}' + \sigma_{VE} I = \begin{bmatrix} (\sigma_{AA}{}^* + \sigma_{VE}) & \sigma_{AA}{}^* & \cdots & \sigma_{AA}{}^* \\ \sigma_{AA}{}^* & (\sigma_{AA}{}^* + \sigma_{VE}) & & \vdots \\ & & & \sigma_{AA}{}^* \\ \vdots & & \ddots & \\ \sigma_{AA}{}^* \cdots & \sigma_{AA}{}^* & & (\sigma_{AA}{}^* + \sigma_{VE}) \end{bmatrix}$$

Dependence is reflected in non zero covariances among students within a class. Given the structural models under consideration, these non zero covariances are seen to equal σ_A^2* which is the class effect. For simplicity, the structural model made the restriction that every class be characterized by the same covariance structure. Since classrooms are assumed to be independent, the covariance of any two individuals from different classrooms is zero. Consequently, the covariance matrix of Y, Σ^Y , is a super diagonal matrix. Further, given initial dependence, where X is a pre test, the covariance matrices of X and XY (Σ^X and Σ^{XY}) are also super diagonal matrices.

Given independence, σ_A^2* is zero and M^Y is a diagonal matrix. Consequently, the Σ^Y matrix becomes diagonal. Thus, to appropriately adjust for dependence an analysis strategy must have a linear model with a residual term that has a subject by subject diagonal covariance matrix.

<u>Index of Response Strategy</u>

One alternative to using the full model for analysis of variance (ANOVA) of the two level nested design, is analysis of variance of index of response using the pooled model. An index of response is defined by:

$$Z = Y = KX$$

where

Z is the index of response;

Y is the post test observation;

X is the pre test observation;

and K is a known constant.

Using Z as the dependent variable in ANOVA the linear model is

$$Z_{ijk} = \mu^Z + \alpha_i^Z + E_{ijk}^Z$$

where

μ^Z is the grand mean of index of response,

α_i^Z is the treatment i effect,

and E_{ijk}^Z is the specification error.

Since for the set of designs under investigation classrooms are randomly assigned to treatments, a treatment effect on Z is equal to a treatment effect on Y. A treatment effect on Z is defined

$$\alpha_i^Z = \mu_i^Z - \mu^Z .$$

Since Z is a linear composite of X and Y,

$$\mu_i^Z = \mu_i - K\mu_i^X$$

and

$$\mu^Z = \mu - K\mu^X$$

so that by substitution

$$\alpha_i^Z = \mu_i - \mu - K(\mu_i^X - \mu^X) .$$

But, given random assignment of classrooms to treatments

$$\mu_1^X = \mu_2^X = \ldots = \mu_t^X = \mu^X$$

and

$$\alpha_i^Z = \mu_i - \mu = \alpha_i$$

regardless of the value of K.

Thus, the null hypothesis for treatment effects can be stated

$$\sum_{i=1}^{t} \alpha_i^2 = 0 .$$

Given the assumptions of the model $F = MS_T^Z / MS_{S:T}^Z$ can be used to test the null hypothesis. Of particular concern here is the assumption of independence which can be stated

$$\underset{tcsx1}{E^Z_{ijk}} \sim NID\ (\underset{tcsx1}{0}\ ,\ \sigma^2_{E^Z} \underset{tcsxtcs}{I}) \ .$$

Returning to the linear model for Z and restating in terms of parameters of X and Y

$$Z_{ijk} = \mu - K\mu^X + \alpha_i + A^*_{ij} - KA_{ij} + E_{ijk} - KV_{ijk}$$

Using the relationships in the structural model under consideration this becomes

$$Z_{ijk} = \mu - K\mu^X + \alpha_i + (B_1 - K)\ A_{ij} + H_{ij} + (B_2 - K)\ V_{ijk} + G_{ijk}$$

<u>The Covariance Matrix of Z</u>

The covariance structure of Z has the form of a super diagonal matrix

$$\underset{tcsxtcs}{\Sigma^Z} = \begin{bmatrix} M^Z & \phi & \cdots & \phi \\ \phi & M^Z & & \vdots \\ \vdots & & \ddots & \phi \\ \phi & \cdots & \phi & M^Z \end{bmatrix}$$

where the within-class covariance matrix M^Z is

$$\underset{sxs}{M^Z} = \underset{sxs}{M^Y} + K^2 \underset{sxs}{M^X} - 2K\underset{sxs}{M^{XY}} \ .$$

In terms of the parameters of M^Y , M^X and M^{XY} , M^Z can be expressed as

$$M^Z = (\sigma_A^2{}^* + K^2\sigma_A^2 - 2K\sigma_{AA}{}^*)\ \underline{1}\,\underline{1}' + (\sigma_E^2 + K^2\sigma_V^2 - 2K\sigma_{VE})I$$

The diagonal element of M^Z is

$$(\sigma_A^2{}^* + K^2\sigma_A^2 - 2K\sigma_{AA}{}^*) + (\sigma_E^2 + K^2\sigma_V^2 - 2K\sigma_{EV}) ,$$

and the off diagonal element is

$$(\sigma_A^2{}^* + K^2\sigma_A^2 - 2K\sigma_{AA}{}^*) .$$

For independence to exist at the level of individuals for Z, Σ^Z must be diagonal. The Σ^Z matrix will be a diagonal if M^Z is diagonal. To make M^Z diagonal,

$$K^2\sigma_A^2 - 2K\sigma_{AA}{}^* + \sigma_A^2{}^* = 0 ,$$

(a gradratic equation in K).

To solve for K,

$$K = \frac{2\sigma_{AA}{}^* \pm \sqrt{(2\sigma_{AA}{}^*)^2 - 4\sigma_A^2\sigma_A^2{}^*}}{2\sigma_A^2}$$

Define

$$\rho_{A^*A} = \frac{\sigma_{AA}{}^*}{\sigma_A\,\sigma_A{}^*}$$

Then,

$$\sigma_{AA}{}^* = \rho_{AA}{}^*\ \sigma_A\sigma_A{}^* ,$$

$$K = \frac{2\rho_{AA}{}^*\ \sigma_A\sigma_A{}^*\ \sqrt{4\rho_{AA}^2{}^*\ \sigma_A^2\sigma_A^2{}^* - 4\sigma_A^2\sigma_A^2{}^*}}{2\sigma_A^2}$$

$$= \rho_{AA}{}^*\ \frac{\sigma_A{}^*}{\sigma_A} \pm \sqrt{\rho_{AA}^2{}^*\ \frac{\sigma_A^2{}^*}{\sigma_A^2} - \frac{\sigma_A^2{}^*}{\sigma_A^2}}$$

Thus,

$$K = \frac{\sigma_A{}^*}{\sigma_A} \rho_{AA}{}^* \pm \sqrt{\rho^2_{AA}{}^* - 1} \quad .$$

The absolute value of the correlation coefficient, $|\rho_{AA}{}^*|$, ranges from zero to one. If $|\rho_{AA}{}^*|$ is less than one, K will be an imaginary number. Thus, the only real solution of K is when $|\rho_{AA}{}^*|$ is one and $K = \sigma_A{}^*/\sigma_A$. But, $|\rho_{AA}{}^*| = 1$ only when the class effects on X, A_{ij}, are perfectly correlated with the class effects on Y, A^*_{ij} . This perfect relationship between A^*_{ij} and A_{ij} dictates the specification errors at the class level in the structural model, i.e., H_{ij} , be zero. Thus, the revised structural model that is appropriate for the index of response strategy must be

$$Y_{ijk} = \mu + \alpha_i + A^*_{ij} + E_{ijk}$$

$$X_{ijk} = \mu + A_{ij} + V_{ijk}$$

$$A^*_{ij} = B_1 A_{ij}$$

$$E_{ijk} = B_2 V_{ijk} + G_{ijk} \ .$$

It is this structural model that is used through the rest of this study.

In conclusion, when there is perfect correlation between class effects for X and class effects for Y and $K = \sigma_{A_*}/\sigma_A$, ANOVA of index of response correctly adjusts for dependence among individuals. Since K must be known a priori, it is useful to explore different ways of thinking about the ratio σ_{A_*}/σ_A . Since from the revised structural model,

$$A^*_{ij} = B_1 A_{ij}$$

B_1 is the regression coefficient of A^*_{ij} on A_{ij} .

It can also be shown that under certain conditions the common within treatment regression coefficient, $\beta_{S:T}$, for Y on X is equal to σ_A*/σ_A .

$$\beta_{S:T} = \frac{\text{Cov } (X_{ijk} - \mu_i^X)\ (Y_{ijk} - \mu_i)}{\text{Var } (X_{ijk} - \mu_i^X)}$$

Under the structural model $\beta_{S:T}$ can be expressed as

$$\beta_{S:T} = \frac{B_1\sigma_A^2 + B_2\sigma_V^2}{\sigma_A^2 + \sigma_V^2}$$ (See Appendix B).

If $B_1 = B_2$, then

$$\beta_{S:T} = B_1 \ .$$

Since

$$B_1 = \sigma_A*/\sigma_A \ ,$$

$$\beta_{S:T} = \sigma_A*/\sigma_A \ .$$

Thus, in addition to perfect correlation of the class effects when the structural regression coefficients at the class level, B_1, and at the individual level, B_2, are equal, the common within treatment regression coefficient, $\beta_{S:T}$ is a correct adjustment coefficient to create independence.

Further, it should be noted that for two level nested data, if $B_1 = B_2$ then

$$\beta_{S:CT} = \beta_{C:T} = \beta_{S:T} = \beta_{Tot} = B_1 \; .$$

Define

$$\beta_{S:CT} = \frac{\text{Cov } (X_{ijk} - \mu_{ij}^X)\ (Y_{ijk} - \mu_{ij})}{\text{Var } (X_{ijk} - \mu_{ij}^X)}$$

Then,

$$\beta_{S:CT} = \frac{B_2\ \sigma_V^2}{\sigma_V^2}$$

Given $B_1 = B_2$, $\beta_{S:CT} = B_1$.

Define

$$\beta_{C:T} = \frac{\text{Cov } (\overline{X}_{ij} - \mu_i^X)\ (\overline{Y}_{ij} - \mu_i)}{\text{Var } (\overline{X}_{ij} - \mu_i^X)}$$

Then

$$\beta_{C:T} = \frac{B_1\sigma_A^2 + B_2\sigma_V^2/s}{\sigma_A^2 + \sigma_V^2/s}$$

Given $B_1 = B_2$, $\beta_{C:T} = B_1$.

Define

$$\beta_{Tot} = \frac{\text{Cov } (X_{ijk} - \mu^X)\ (Y_{ijk} - \mu)}{\text{Var } (X_{ijk} - \mu^X)}$$

Then,

$$\beta_{Tot} = \frac{B_1\sigma_A^2 + B_2\sigma_V^2}{\sigma_A^2 + \sigma_V^2}$$

Given $B_1 = B_2$, $\beta_{Tot} = B_1$.

(The above illustrations use information from Appendices C and D.)

Test Statistic

The test statistic for testing the treatment hypothesis can be developed from the analysis of variance table of Z as shown in Figure 4.1. From Figure 4.1, the expected mean square of treatments, $E(MS_T^Z)$, and the expected mean square of students pooled within treatments, $E(MS_{S:T}^Z)$, estimate the same parameter, $(B_2-B_1)^2 \sigma_V^2 + \sigma_G^2$, under the null hypothesis of no treatment effects. Given the assumptions about V_{ijk} and G_{ijk}, the $F_{S:T}^Z = MS_T^Z / MS_{S:T}^Z$ ratio is distributed as a central F distribution with t-1 and t (cs-1) degrees of freedom under the null hypothesis of no treatment effects.

The power of the $F_{S:T}^Z$ test is higher than the $F_{C:T}$ test. The power of the $F_{S:T}^Z$ test is a function of the probability of Type I error, the size of the treatment effects (α_i), $E(MS_{S:T}^Z)$ and degrees of freedom (t-1, and t(cs-1)). Since treatment effects are identical for the two tests and the probability of Type I error can be held constant, differences in power are only a function of the last two factors.

The degrees of freedom for the error terms of $F_{S:T}^Z$ and $F_{C:T}$ are t(cs-1) and t(c-1) respectively. Thus, everything else held constant, $F_{S:T}^Z$ has higher power than $F_{C:T}$ because of higher degrees of freedom.

The effects on power of the two error terms is not straightforward. Recall that

$$E(MS_{S:T}^Z) = (B_2 - B_1)^2 \sigma_V^2 + \sigma_G^2$$

and

$$E(MS_{C:T}) = s\sigma_A^{2*} + \sigma_E^2 .$$

Model

$$Z_{ijk} = Y_{ijk} - KX_{ijk} \qquad \begin{array}{l} i = 1,2,\ldots,t \\ j = 1,2,\ldots,c \\ k = 1,2,\ldots,s \end{array}$$

or

$$Z_{ijk} = (\mu - B_1\mu^X) + \alpha_i + (B_2 - B_1)\, V_{ijk} + G_{ijk}$$

Assumptions

$$V_{ijk} \sim \text{NID}\ (0,\ \sigma_V^2)$$

$$G_{ijk} \sim \text{NID}\ (0,\ \sigma_G^2)$$

$$\sigma_{VG} = 0$$

Source of Variation	d.f.	Sum Square	Mean Square	E(MS)
Treatment (T)	t-1	$sc\sum_i(\bar{Z}_{i..} - \bar{Z}..)^2$	MS_T^Z	$sc\frac{\Sigma\alpha_i^2}{t-1} + (B_2-B_1)^2\ \sigma_V^2 + \sigma_G^2$
Student (S:T)	t(cs-1)	$\sum_{ijk}(Z_{ijk} - Z_{i..})^2$	$MS_{S:T}^Z$	$(B_2 - B_1)^2\ \sigma_V^2 + \sigma_G^2$

Given Ho: $\alpha_i = 0\ ,\ \frac{MS_T^Z}{MS_{S:T}^Z} \sim F_{t-1,\ t(cs-1)}$

Figure 4.1: ANOVA Table for the Index of Response Model Using Individual Students as the Units of Analysis

$E(MS_{C:T})$ can be expressed as

$$E(MS_{C:T}) = s\sigma^2_{A^*} + B^2_2\sigma^2_V + \sigma^2_G \ .$$

When $B_2 = B_1$, $E(MS^Z_{S:T})$ is clearly smaller than $E(MS_{C:T})$. Also, if B_1 and B_2, have the same sign, $E(MS^Z_{S:T})$ is smaller than $E(MS_{C:T})$. But, when B_1 and B_2 have different signs, the magnitude of dependence and class size must be taken into consideration. Thus, generally the $F^Z_{S:T}$ test has smaller error variance than the $F_{C:T}$ test.

Given that data on a covariate exist for building an index of response, the power of the $F^Z_{S:T}$ test might more appropriately be compared to the power of $F^Z_{C:T}$. Given that $\rho_{A^*A} = 1.0$ and $K = \sigma_{A^*}/\sigma_A$,

$$E(MS^Z_{C:T}) = E(MS^Z_{S:T}) \ .$$

(The analysis of variance table for index of response using the full model is reported in Appendix E.) Since the two test statistics, $F^Z_{S:T}$ and $F^Z_{C:T}$, have equal expected mean square errors, their difference in power is solely a function of their difference in degrees of freedom. Thus, because of greater degrees of freedom $F^Z_{S:T}$ has greater power than $F_{C:T}$.

In conclusion, index of response is a useful method of analysis for two level nested data when the class effects on the covariate are perfectly correlated with the class effects on the dependent variable: i.e., $A^*_{ij} = B_1A_{ij}$. Given that $A^*_{ij} = B_1A_{ij}$, the adjustment coefficient K which must be known a priori is equal to a ratio of σ_{A^*} over σ_A . Under this condition, another test can be developed by using classroom as the unit of analysis, i.e., $F^Z_{C:T}$. The $F^Z_{S:T}$ and the $F^Z_{C:T}$ statistics test the correct treatment hypothesis. The power of $F^Z_{S:T}$ is higher than $F_{C:T}$ because of larger degrees of freedom and a generally smaller error term. When comparing $F^Z_{S:T}$ to $F^Z_{C:T}$, the gain in the power by using $F^Z_{S:T}$ is a function of the gain in the degrees of freedom alone.

Analysis of Covariance

When the adjustment coefficient $K = \sigma_A{*}/\sigma_A$ is not know a priori, an alternative strategy that allows the coefficient to be estimated from data would be helpful. One such strategy is analysis of covariance (ANCOVA).

Consider a one way ANCOVA. The linear model involving a random covariable is

$$Y_{ijk} = \mu + \alpha^W_i + \beta_{S:T} (X_{ijk} - \mu^X) + E^W_{ijk}$$

where

Y is the post test,

$\beta_{S:T}$ is the common bivariate regression slope of Y on X,

μ is the grand mean of Y,

W is used to denote adjustment by covariate X,

α^W_i is the adjusted treatment i effect,

X is the pre test,

μ^X is the grand mean on post test, and

E^W_{ijk} is the error term.

The null hypothesis is stated

$$Ho: \sum_{i=1}^{t} (\alpha_i^W)^2 = 0 \ .$$

For the model, a treatment effect is defined

$$\alpha_i^W = \mu_i - \mu - \beta_{S:T} (\mu_i^X - \mu^X) \ .$$

But given random assignment

$$\mu_i^X = \mu_i^X = \ldots = \mu_t^X = \mu^X \ ,$$

and

$$\alpha_i^W = \mu_i - \mu = \alpha_i$$

regardless of the value of $\beta_{S:T}$. Therefore, when the pooled model ANCOVA is applied to data considered in this study, treatment effects are unbiased and the correct treatment hypothesis is tested.

An F test of the null hypothesis about treatments can be provided given the assumptions that

$$\underset{tcs \times 1}{\underline{E}^W_{ijk}} \sim NID\,(\underset{tcs \times 1}{\underline{0}} \ , \ \sigma^2_{E^W} \underset{tcs \times tcs}{I}) \ .$$

In addition, ANCOVA assumes that the within treatment slopes are equal across all values of i and that Y and X are bivariate normal. These assumptions are the same as those for classical ANCOVA except that the covariable is random. In using ANCOVA when X is random, one can still obtain unbiased estimaters and valid confidence intervals and tests from the usual analysis. The only difference from the classical result is that the variances of the estimators are larger. Discussion of a random covariate in ANCOVA can be found in DeGracie (1968), Winer (1971), and Huitema (1980).

The Covariance Matrix of E^W

The pooled model ANCOVA can be expressed in terms of parameters in the structural model as follows:

$$Y_{ijk} = \mu - \beta_{S:T}\,\mu^X + \alpha_i + (B_1 - \beta_{S:T})\,A_{ij} + (B_2 - B_{S:T})\,V_{ijk} + G_{ijk}$$

where

$$E^W_{ijk} = (B_1 - \beta_{S:T})\,A_{ij} + (B_2 - \beta_{S:T})\,V_{ijk} + G_{ijk}\ .$$

Under the independence assumption, the covariance matrix of E^W_{ijk} must be diagonal. From the assumptions of the structural model

$$\sigma_{AV} = \sigma_{AG} = \sigma_{VG} = 0\ .$$

Therefore,

$$\underset{tcsxtcs}{\Sigma^{E^W}} = (B_1 - \beta_{S:T})^2 \underset{tcsxtcs}{\Sigma^A} + (B_2 - \beta_{S:T})^2 \underset{tcsxtcs}{\Sigma^V} + \underset{tcsxtcs}{\Sigma^G}\ .$$

$$\underset{tcsxtcs}{\Sigma^A} = \begin{bmatrix} M^A & \phi & \cdots & \phi \\ \phi & M^A & & \vdots \\ \vdots & & \ddots & \phi \\ \phi & \cdots & \phi & M^A \end{bmatrix}$$

where

$$\underset{sxs}{M^A} = \sigma_A^2\, \underset{sx1}{\underline{1}}\ \underset{1xs}{\underline{1}'}$$

But Σ^V and Σ^G are diagonal, i.e.,

$$\Sigma^V = \sigma_V^2 I$$

and

$$\Sigma^G = \sigma_G^2 I\ .$$

Thus,

$$\underset{tcsxtcs}{\Sigma^{A}} = \begin{bmatrix} M^{E^W} & \phi & \cdots & \phi \\ \phi & M^{E^W} & & \vdots \\ \vdots & & \ddots & \phi \\ \phi & \cdots & \phi & M^{E^W} \end{bmatrix}$$

where the diagonal matrix of Σ^{E^W} is

$$\underset{sxs}{M^{E^W}} = (B_1 - \beta_{S:T})^2 \underset{sxs}{M^{A}} + (B_2 - \beta_{S:T})^2 \underset{sxs}{M^{V}} + \underset{sxs}{M^{G}}$$

$$= (B_1 - \beta_{S:T})^2 \sigma_A^2 \underline{1}\,\underline{1}' + (B_2 - B_{S:T})^2 \sigma_V^2 I + \sigma_G^2 I \ .$$

The diagonal element of M^{E^W} is

$$(B_1 - \beta_{S:T})^2 \sigma_A^2 + (B_2 - \beta_{S:T})^2 \sigma_V^2 + \sigma_G^2 \ ,$$

and the off diagonal element is

$$(B_1 - \beta_{S:T})^2 \sigma_A^2 \ .$$

For independence to exist at the level of individuals, M^{E^W} must be diagonal.

Given the structural model considered in this study and given $B_1 = B_2$, it was shown that

$$\beta_{S:T} = \frac{\sigma_A}{\sigma_A*} = B_1 \ .$$

When $B_1 = \beta_{S:T}$, the off diagonal elements of M^{E^W} are zero, and Σ^{E^W} becomes a diagonal matrix.

Test Statistic

Given the structural model and the further restrictions that $\rho_{A^*A} = 1$ and $B_1 = B_2$, the pooled model ANCOVA provides a valid test of the null hypothesis about treatments. The test statistic is

$$F^W_{S:T} = MS^W_T / MS^W_{S:T} \sim F_{t-1,\ t(cs-1)} \ .$$

To be a useful analysis strategy, however, $F^W_{S:T}$ must also have greater power than the test statistic for the full model. The most appropriate comparison is to the test statistic for the full model ANCOVA, $F^W_{C:T}$, since the use of a covariate will in general provide a more powerful test statistic than the full model ANOVA.

$$F^W_{C:T} = \frac{MS^{W'}_T}{MS^W_{C:T}} \sim F_{t-1,\ t(c-1)}$$

where the prime is used to distinguish the treatment mean square of the full model ANCOVA from the treatment mean square of the pooled model ANCOVA.

A treatment effect defined by the pooled ANCOVA model is

$$\alpha^W_i = \mu_i - \mu - \beta_{S:T}\ (\mu^X_i - \mu^X) \ .$$

A treatment effect defined by the full ANCOVA model is

$$\alpha_i^{W'} = \mu_i - \mu - \beta_{C:T}\ (\mu^X_i - \mu^X) \ .$$

Treatment effects for the pooled and the full ANCOVA models are identical under the condition that $B_1 = B_2$ since $\beta_{S:T} = \beta_{C:T}$. The estimates of treatment effects for the two models differ only in the estimates of slopes. Under the null hypothesis for treatments,

$$E(MS_T^W) = E(MS_{S:T}^W) = \sigma_G^2 \left(1 + \frac{1}{t(cs-1) - 2}\right)$$

and

$$E(MS_T^{W'}) = E(MS_{C:T}^{W'}) = s \frac{\sigma_G^2}{s} \left(1 + \frac{1}{t(c-1) - 2}\right)$$

$$= \sigma_G^2 \left(1 + \frac{1}{t(c-1) -2}\right) .$$

Then,

$$E(MS_T^{W'}) = \frac{\left(1 + \frac{1}{t(c-1) - 2}\right)}{\left(1 + \frac{1}{t(cs-1) - 2}\right)} E(MS_T^W) .$$

Therefore, $E(MS_T^{W'}) > E(MS_T^W)$. Similarly, $E(MS_{C:T}^W) > E(MS_{S:T}^W)$. The gain in power by using the pooled ANCOVA model instead of the full ANCOVA model is a function of increased degrees of freedom and a slightly smaller mean square error.

CHAPTER V

SUMMARY AND CONCLUSIONS

A commonly used design in educational research involves hierarchically nested data. Classrooms of students are randomly assigned to receive one of two or more alternative educational treatments. Since dependent variables in educational research are typically defined on students, however, the design results in students nested within classrooms and classrooms nested with treatments.

A fully specified model for the design includes sources of variation for treatments, classrooms, and students. Given the fully specified model, the null hypothesis about treatments can be tested using $F_{C:T} = MS_T/MS_{C:T}$. There has been resistance on the part of educational researchers to use the $F_{C:T}$ test statistic because for these studies the test has few degrees of freedom for error and so limited statistical power. For example, a design comparing two treatments and having three classrooms per treatment might include over 100 students and still have only four degrees of freedom error for the $F_{C:T}$ test statistic.

As a result, researchers have sometimes turned to a pooled model which ignores classroom variance. By ignoring classroom variance the sources of variation become treatments and students. The apparent test statistic for the treatment null hypothesis is then

$F_{S:T} = MS_T/MS_{S:T}$. The motivation for the pooled model can be illustrated with the previous example. What had been four degrees of freedom error for $F_{C:T}$, becomes more than 100 degrees of freedom error for $F_{S:T}$.

The test statistic, $F_{S:T}$, for the pooled model requires that observations on students be independent of each other. Violation of the independence assumption when using $F_{S:T}$ has been shown to yield a test which can be either too liberal or too conservative (Glendening, 1977; Paull, 1950). What is needed, then, are analysis strategies which have greater degrees of freedom error than the $F_{C:T}$ test statistic and which are valid when there is dependence at level of individuals among observations on the dependent variable.

The search for more powerful tests of treatment effects in hierarchically nested data began with an operational definition of independence. Given balanced designs with two levels of nested data, Glendening (1977) defined independence as equivalent to when the expected values of classrooms and students mean squares are equal. If students are considered a random factor in the design, as they were in the present investigation, dependence occurs when there are classroom effects which make the expected value of the mean square for classroom exceed the expected value of the mean square for students. Glendening labeled this situation as positive dependence. Glendening also considered the possibility of negative dependence which can result when students are a fixed factor in the design. Under positive dependence, the $F_{S:T}$ test is too liberal and so has spurious power (Glendening, 1977; Paull, 1950). The first analysis strategy posed as a solution to the problem which dependence creates for the $F_{S:T}$ statistic, involves a preliminary test for independence. Based on the outcome of this preliminary test, either the full model or the pooled model is used to test the null hypothesis about treatments (Glendening, 1977; Paull, 1950; Peckham, et al., 1969).

To be successful, the preliminary test and conditional pooling procedure must keep the actual alpha level of the consequent test (the conditional test) close to the nominal alpha level. Further, in order that the strategy be useful, the power of the conditional test must be greater than the power of the unconditional $F_{C:T}$ test.

Glendening (1977) examined the validity of the preliminary test and conditional pooling strategy. Analytically and empirically, Glendening's findings opposed the use of the procedure. Glendening concluded that given a moderate value of dependence, which is often the case, the preliminary F test is not sensitive enough to help a researcher guard against having a distorted probability of Type I error for the conditional F test. This conclusion was consistent with Paull(1950).

Since a preliminary test for independence and conditional pooling does not result in a valid test of the treatment hypothesis, the present investigation considered additional alternatives. First, the use of a quasi-F ratio to correct for dependence was considered (Satterthwaite, 1941, 1946). The potential utility of the quasi-F statistic in situations of positive dependence can be seen in the following ratio.

$$F' = \frac{MS_T - MS_{C:T} + MS_{S:CT}}{MS_{S:CT}}$$

Under the null hypothesis of treatment effects $E(MS_T - MS_{C:T} + MS_{S:CT})$ and $E(MS_{S:CT})$ estimate the same parameter, σ_E^2 . The apparent reference distribution for F' is F_{ν_1,ν_2} where ν_2 was t(cs-1) and ν_1 , could be computed from a formula provided by Satterthwaite. But, the complex variance of the numerator and the simple variance of the demoninator are not independent. Thus, the F' statistic which appeared to hold promise as a test with greater power than $F_{C:T}$ in situations of dependence and with no cost of additional information lacks a known distribution.

Glendening and Porter (1974) suggested the possibility of using ANCOVA to adjust for positive dependence. They pointed out that the effects of positive dependence could be conceptualized as similar in form to the problem created by confounding in quasi experiments. They speculated that ANCOVA would remove the problematic variance of the class effects (σ_A^2*) from both $E(MS_{C:T})$ and $E(MS_T)$. Conceptually, this would be equivalent to the ANCOVA procedures creating adjusted observations that are free from positive dependence.

Index of response is another adjustment strategy which is closely tied to ANCOVA. Thus, index of response was also considered in this investigation.

To consider the utility of index of response and ANCOVA strategies, it became important to understand the possible causes of dependence. Distinctions among when and how dependence can arise in an experimental study could help to inform the selection of an appropriate covariable.

Two main related tasks, then, were set for the investigation of this study: (1) to classify situations of independence and dependence in experimental studies; and (2) to investigate the possibility of using an index of response and ANCOVA models to correct for positive dependence among individual units.

As a result of the first task, four possible situations were classified for an experimental study that involves two level hierarchically nested data. Dependence could arise because students were not randomly assigned for classrooms (initial dependence) and/or class effects which occur during the study (during-experiment dependence). Crossing these two dichotomous possibilities defined the four situations (Figure 3.1).

Situation I, independence, occurs when students are randomly assigned to classrooms and there are no class effects. Analysis of variance using the pooled model is the best analysis strategies for testing the no treatment effect hypothesis (Glendening, 1977; Cronbach, 1976).

Situation II includes dependence due solely to students not being randomly assigned to classrooms. When students are not randomly assigned to classrooms, each classroom represents a population that could have a different distribution on the dependent variable. To the extent that classroom populations have different means, there will be a classroom effect and so dependence among observations taken on individual students. If index of response or ANCOVA are to have potential for creating an adjustment which eliminates dependence, the covariate must reflect the dependence present at the outset of the experiment. A pretest would seem to hold the greatest potential.

Situation III includes dependence which occurs during the experiment. Common class experinces or class effects are the causes of this dependence. Possible class effects are complex and global: they might include effects due to subject matter, teacher effectiveness, teaching strategies, student interactions and classroom milieu which are not part of the definitions of treatments. Defining covariables which reflect during-experimental dependence and which might be used in index of response or ANCOVA is a substantive issue worthy of study in its own right.

In Situation IV, both initial and during-experiment dependence occur together. Under the simplest case when initial dependence and during-experiment dependence do not interact with each other or with treatments, the resulting classroom effects could be thought of as having two additive parts, one for each type of dependence. To adjust for Situation IV dependence the researcher would need to identify a set of covariables reflecting initial dependence and a second set of covariables reflecting during-experiment dependence.

In this investigation the utility of index of response and ANCOVA for adjusting for dependence was restricted to use of a pre test in Situation II conditions. Further, classroom populations were assumed to be normally distributed on the dependent variable with a common variance but different means. In order for either approach to be judged as having utility, their resulting test statistics of the null hypothesis about treatments must have: 1) a known sampling distribution, and 2) greater power than the test statistics which results from the full model.

The structural model, assumed for the dependent variable, Y, the covariable, X, and their interrelationship is given

$$Y_{ijk} = \mu + \alpha_i + A^*_{ij} + E_{ijk}$$

$$X_{ijk} = \mu^X + A_{ij} + V_{ijk}$$

$$A^*_{ij} = B_1 A_{ij} + H_{ij}$$

$$E_{ijk} = B_2 V_{ijk} + G_{ijk}$$

Assumptions

$$A_{ij} \sim NID\ (0, \sigma^2_A)$$

$$H_{ij} \sim NID\ (0, \sigma^2_H)$$

$$V_{ijk} \sim NID\ (0, \sigma^2_V)$$

$$G_{ijk} \sim NID\ (0, \sigma^2_G)$$

and

$$\sigma_{AH} = \sigma_{VG} = \sigma_{AV} = \sigma_{AG} = \sigma_{HV} = \sigma_{HG} = 0$$

Under dependence, the covariance structure of Y, X and XY (i.e., Σ^X , Σ^Y and Σ^{XY}) are the form of super diagonal matrices. The diagonal matrices of Y, X, and XY (i.e., M^Y, M^X and M^{XY}) are within class covariance matrices and are expressed as

$$M^Y = \sigma^2_{A^*} \underline{1}\,\underline{1}' + \sigma^2_E I$$

$$M^X = \sigma^2_A \underline{1}\,\underline{1}' + \sigma^2_V I$$

$$M^{XY} = \sigma_{AA^*} \underline{1}\,\underline{1}' + \sigma_{EV} I$$

Under independence, Σ^X, Σ^Y and Σ^{XY} are diagonal matrices. Thus, an analysis strategy that correctly adjusts for dependence will yield a linear model with specifications errors that have a diagnonal covariance matrix.

Given the structural model, it was discovered that an index of response would only adjust for dependence when class effects on X and Y are perfectly correlated (i.e., $|\rho_{AA^*}| = 1.0$) . Given this condition, an index of response

$$Z = Y - \frac{\sigma_{A^*}}{\sigma_A} X ,$$

was seen to yield observations that are independent at the level of individuals. The covariance matrix of Z was found to be

$$\Sigma^Z = ((B_2 - K)^2 \sigma^2_V + \sigma^2_G)I$$

Thus, a pooled model which uses Z as the dependent variable provides a valid test statistic for the treatment null hypothesis, $F^Z_{S:T} = MS^Z_T/MS^Z_{S:T}$. The $F^Z_{S:T}$ test has greater power than the $F^Z_{C:T}$ test statistic from the full model using the same index of response because of greater degrees of freedom. The $F^{Z}_{S:T}$ test provides greater power than the $F_{C:T}$ test from the full model using the dependent variable both because of greater degrees of freedom and because the error term for the $F_{S:T}$ test will generally be smaller than the error term for the $F_{C:T}$ test.

Investigation of the pooled model ANCOVA revealed a valid test for the treatment null hypothesis given the structural model for X, Y and their interrelationship only when $\rho_{A^*A} = 1.0$ and $B_1 = B_2$. When $B_1 = B_2$, it was seen that $\beta_{S:T}$, the pooled within slope of Y on X, was equal to the constant used to define the correct index of response (i.e., $\beta_{S:T} = \sigma_{A^*}/\sigma_A$). Under this condition the ANCOVA pooled model yielded specification errors which have a diagonal covariance matrix

$$\Sigma^W = \sigma^2_G I ,$$

and the test statistic $F^W_{S:T} = MS^W_T/MS^W_{S:T}$. $F^W_{S:T}$ was seen to have greater power than the full model ANCOVA test statistic, $F^W_{C:T}$, because of larger degrees of freedom and a smaller mean square error.

The key to the success of index of response and ANCOVA strategies was the structural model and the further condition that class effects on Y be completely accounted for by the class effects on X; i.e., $A^*_{ij} = B_1 A_{ij}$. Index of response did not require any correlation between X and Y for individuals within classrooms. ANCOVA, however, required that the constant specifying the relationship between X and Y at the level of individuals be equal to the constant that specified the relation between class effects. Nevertheless, it was seen that the correlation between X and Y for individuals within classrooms need not be perfect. ANCOVA was seen to adjust for dependence when there was specification error in Y that was not accounted for by specification error in X.

In general, then the results of this investigation supported and expanded upon the conjecture by Glendening and Porter (1974) about the utility of ANCOVA for adjusting for dependence. However, their suggestion that X and Y must be perfectly correlated can be relaxed to a requirement that the classroom effects be perfectly correlated.

APPENDIX A

ANOVA TABLE OF THE FULL MODEL

GIVEN STUDENTS ARE FIXED

APPENDIX A

ANOVA TABLE OF THE FULL MODEL

GIVEN STUDENTS ARE FIXED

Model: $Y_{ijk} = \mu + \alpha_i + A^*_{ij} + E_{ijk}$

Assumptions: $A^*_{ij} \sim \text{NID}\ (0, \sigma^2_A{}^*)$

Source of Variation	d.f.	E(MS)
Treatments (T)	(t-1)	$s\sigma^2_A{}^* + \frac{cs\sum_i \alpha^2_i}{t-1}$
Classes (C:T)	t(c-1)	$s\sigma^2_A{}^*$
Students (S:CT)	tc(s-1)	$\frac{\sum\sum\sum_{ijk} E^2_{ijk}}{tc(s-1)}$

Treatments are fixed;

Classrooms are random; and

Students are fixed.

APPENDIX B

THE REGRESSION SLOPE OF THE POOLED ANCOVA MODEL

APPENDIX B

THE REGRESSION SLOPE OF THE POOLED ANCOVA MODEL

Define $\beta_{S:T}$ to be:

$$\beta_{S:T} = \frac{\text{Cov } (X_{ijk} - \mu_i^X)\ (Y_{ijk} - \mu_i)}{\text{Var } (X_{ijk} - \mu_i^X)}$$

$$= \frac{\text{Cov } (A_{ij} + V_{ijk})\ (A_{ij}^* + E_{ijk})}{\text{Var } (A_{ij} + V_{ijk}}$$

$$= \frac{\sigma_{AA^*} + \sigma_{VE}}{\sigma_A^2 + \sigma_V^2}$$

From

$$A_{ij}^* = B_1 A_{ij}$$

$$E_{ijk} = B_2 V_{ijk} + G_{ijk}$$

and $\sigma_{VG} = 0$,

$$\sigma_{AA^*} = B_1 \sigma_A^2$$

and $\sigma_{VE} = B_2 \sigma_V^2$.

Then,

$$\beta_{S:T} = \frac{B_1 \sigma_A^2 + B_2 \sigma_V^2}{\sigma_A^2 + \sigma_V^2}$$

If $B_1 = B_2$, $\beta_{S:T} = B_1$.

APPENDIX C

VARIANCE COMPONENTS OF DEVIATED SCORES

APPENDIX C

VARIANCE COMPONENTS OF DEVIATED SCORES

Model:

$$Y_{ijk} = \mu + \alpha_i + A^*_{ij} + E_{ijk}$$

$$X_{ijk} = \mu^X + A_{ij} + V_{ijk}$$

Assumptions:

$$A^*_{ij} \sim NID\ (0,\ \sigma^2_{A^*})\ ,$$

$$E_{ijk} \sim NID\ (0,\ \sigma^2_E)\ ,$$

$$A_{ij} \sim NID\ (0,\ \sigma^2_A)\ ,$$

$$V_{ijk} \sim NID\ (0,\ \sigma^2_V)\ ,$$

and $\sigma_{A^*E} = \sigma_{AV} = \sigma_{A^*V} = \sigma_{AE} = 0$

Define:

$$\mu_i = \mu + \alpha_i$$

$$\mu_{ij} = \mu + \alpha_i + A^*_{ij}$$

$$\mu^X_i = \mu^X$$

$$\mu^X_{ij} = \mu^X + A_{ij}$$

Variable	Model	Expected Value	Variance
$Y_{ijk} - \mu_i$	$A^*_{ij} + E_{ijk}$	0	$\sigma^2_{A^*} + \sigma^2_E$
$Y_{ijk} - \mu_{ij}$	E_{ijk}	0	σ^2_E
$\overline{Y}_{ij.} - \mu_i$	$A^*_{ij} + \overline{E}_{ij.}$	0	$\sigma^2_{A^*} + \frac{\sigma^2_E}{s}$
$Y_{ijk} - \mu$	$\alpha_i + A^*_{ij} + E_{ijk}$	α_i	$\sigma^2_{A^*} + \sigma^2_E$
$X_{ijk} - \mu^X_i$	$A_{ij} + V_{ijk}$	0	$\sigma^2_A + \sigma^2_V$
$X_{ijk} - \mu^X_{ij}$	V_{ijk}	0	σ^2_V
$\overline{X}_{ij.} - \mu^X_i$	$A_{ij} + \overline{V}_{ij.}$	0	$\sigma^2_A + \frac{\sigma^2_V}{s}$
$X_{ijk} - \mu^X$	$A_{ij} + V_{ijk}$	0	$\sigma^2_A + \sigma^2_V$

APPENDIX D

COVARIANCE COMPONENTS OF DEVIATED SCORES

APPENDIX D

COVARIANCE COMPONENTS OF DEVIATED SCORES

Model:

$$Y_{ijk} = \mu + \alpha_i + A^*_{ij} + E_{ijk}$$

$$X_{ijk} = \mu^X + A_{ij} + V_{ijk}$$

Assumptions:

$$A^*_{ijk} \sim NID\ (0,\ \sigma^2_{A^*})\ ,$$

$$E_{ijk} \sim NID\ (0,\ \sigma^2_E)\ ,$$

$$A_{ij} \sim NID\ (0,\ \sigma^2_A)\ ,$$

$$V_{ijk} \sim NID\ (0,\ \sigma^2_V)\ ,$$

and $$\sigma_{A^*E} = \sigma_{AV} = \sigma_{A^*V} = \sigma_{EV} = 0$$

Define: $$\mu_i = \mu + \alpha_i$$

$$\mu_{ij} = \mu + \alpha_i + A^*_{ij}$$

$$\mu^X_i = \mu^X$$

and $$\mu^X_{ij} = \mu^X + A_{ij}\ .$$

$\text{Cov}\ (X_{ijk} - \mu_i^X)\ (Y_{ijk} - \mu_i)$	$\text{Cov}\ (A_{ij} + V_{ijk})\ (A_{ij}^* + E_{ijk})$	$\sigma_{AA^*} + \sigma_{VE}$
$\text{Cov}\ (X_{ijk} - \mu_{ij}^X)\ (Y_{ijk} - \mu_{ij})$	$\text{Cov}\ (V_{ijk}\ ,\ E_{ijk})$	σ_{VE}
$\text{Cov}\ (\overline{X}_{ij.} - \mu_i^X)\ (\overline{Y}_{ij.} - \mu_i)$	$\text{Cov}\ (A_{ij} + \overline{V}_{ij.})\ (A_{ij}^* + \overline{E}_{ij.})$	$\sigma_{AA^*} + \frac{\sigma_{VE}}{s}$
$\text{Cov}\ (X_{ijk} - \mu^X)\ (Y_{ijk} - \mu)$	$\text{Cov}\ (A_{ij} + V_{ijk})\ (A_{ij}^* + E_{ijk})$	$\sigma_{AA^*} + \sigma_{VE}$

APPENDIX E

ANOVA TABLE FOR THE INDEX OF RESPONSE MODEL

USING CLASSROOMS AS THE UNITS OF ANALYSIS

APPENDIX E

ANOVA TABLE FOR THE INDEX OF RESPONSE MODEL USING CLASSROOMS AS THE UNITS OF ANALYSIS

Model:

$$\bar{Z}_{ij.} = \bar{Y}_{ij.} - K\bar{X}_{ij.} \qquad i = 1,2,\ldots,t \quad j = 1,2,\ldots,c \quad k = 1,2,\ldots,s$$

or

$$\bar{Z}_{ij.} = (\mu - B_1\mu^X) + \alpha_i + (B_2 - B_1)\,\bar{V}_{ij.} + \bar{G}_{ij.}$$

Assumptions:

$$\bar{V}_{ij.} \sim NID\ (0,\ \sigma_V^2/s)$$

$$\bar{G}_{ij.} \sim NID\ (0,\ \sigma_G^2/s)$$

$$\sigma_{VG} = 0$$

Source of Variation	d.f.	Sum of Square	Mean Square	E(MS)
Treatment (T)	t-1	$sc\sum_i(\bar{Z}_{i..}-\bar{Z}_{...})^2$	MS_T^Z	$\frac{sc\sum_{i=1}^{t}\alpha_i^2}{t-1} + (B_2-B_1)^2\,\sigma_V^2 + \sigma_G^2$
Classroom (C:T)	t(c-1)	$s\sum_{j,k}(\bar{Z}_{ij.}-\bar{Z}_{i..})^2$	$MS_{C:T}^Z$	$(B_2-B_1)^2\,\sigma_V^2 + \sigma_G^2$

Given Ho: $\alpha_i = 0$, $\frac{MS_T}{MS_{C:T}} \sim F_{t-1,\ t(c-1)}$

REFERENCES

REFERENCES

Anderson, G.L. A comparison of the outcomes of instruction under two theories of learning. Unpublished doctoral dissertation, University of Minnesota, 1941.

Bloom, B.S., Human Characteristics and School Learning, New York: McGraw-Hill, 1976.

Bryk, A.S. and Weisberg, H.I. Use of the Nonequivalent Control Group Design When Subjects Are Growing, Psychological Bulletin, 1977, 84, 950-962.

Cronbach, L.J. Research on classrooms and schools: formulation of questions, design, and analysis. Occasional papers of the Stanford Evaluation Consortium, Stanford University, California, July, 1976.

Cronbach, L.J. & Webb, N. Between-class and within-class effects in a reported aptitude X treatment interaction: Reanalysis of a study by G.L. Anderson. Journal of Educational Psychology, 1975, 67, 717-724.

Cronbach, L.J., Pagosa, D.R., Floden, R.E., & Price, G.G. Analysis of covariance in non randomized experiments: parameters affecting bias. Occasional paper, the Stanford Evaluation Consortium, Stanford, California, 1977.

Cochran, W.G., & Cox, G.M. Experimental Designs. New York: Wiley & Sons, 1957.

Cox, D.R. Planning of Experiments. New York: Wiley & Sons, 1958.

DeGracie, J.S., "Analysis of covariance when the concomitant variable is measured with error." Unpublished Ph.D. thesis, Library, Iowa State University of Science and Technology, Ames, Iowa, 1968.

Dunkin, M.J. & Biddle, B.J. The Study of Teaching, New York: Holt, Rinehart, and Winston, Inc., 1974.

Gaylor, D.W. & Hopper, F.N., Estimating the degrees of freedom for linear combinations of mean squares by Satterthwaite's Formula. Technometrics, 1969, 11, 691-706.

Glass, G.V., & Stanley, L.C. Statistical Methods in Education and Psychology. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

Kirk, R.E. Experimental Design: Procedure for the Behavioral Sciences. Belmont, California: Brooks/Cole, 1968.

Hannan, M.T., & Young, A.A. On certain similarities in the estimation of multi-wave panels and multi-level cross-sections. Paper presented at the Conference on Methodology for Aggregating Data in Educational Research, Stanford University, October 23-24, 1976.

Hudson, J.D., & Krutchkoff, R.G. A monte earlo investigation of tests employing Satterthwaite's synthetic mean squares. Biometrika, 1968, 55, 431-433.

Huitema, B.E. The Analysis of Covariance and Alternatives. New York: Wiley & Sons, 1980.

Olejnik, S.F. Data analysis strategies for quasi-experimental studies where differential group and individual growth rates are assumed. Unpublished Ph.D. dissertation, East Lansing: Michigan State University, 1977.

Paull, A.E. On a preliminary test for pooling mean squares in the analysis of variances. Annals of Mathematical Statistics, 1950, 21, 539-556.

Peckham, P.D., Glass, G.V., & Hopkins, K.D. The experimental unit in statistical analysis. The Journal of Special Education, 1969, 3, 337-349.

Porter, A.C. Some design and analysis concerns for quasi-experiments such as Follow Through. Paper presented at the meeting of the American Psychological Association, Hawaii, August, 1972.

Porter, A.C. An Evaluation of the TABA and BASICS Teaching Strategies Programs in the Lansing Public Schools. Paper presented for the Michigan Department of Education, College of Education, Michigan State University, November, 1973.

Porter, A.C., & Chibucos, T.R. Selecting an analysis strategy. In G. Borich (ed.), Evaluating Educational Programs and Products. Educational Technology Press, 1974.

Porter, A.C., & Chibucos, T.R. Common problems of design and analysis in evaluative research. Sociological Methods & Research, 1975, 3, 235-257.

Satterthwaite, F.W. Synthesis of variance. Psychometrika, 1941, 6, 309-16.

Satterthwaite, F.W. An approximate distribution of estimates of variance components. Biometrics Bulletin, 1946, 2, 110-2.

Schmidt, W.H. Structural equation models and their application for longitudinal data. Paper prepared for the Conference on Longitudinal Statistical Analysis, Boston, 1975.

Winer, B.J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.