ON REDUCING BIAS IN
OBSERVATIONAL
STUDIES

By

Oscar Chi-Hsuing Lin

Bachelor of Arts
National Taiwan University
Taipei, Taiwan, Republic of China
1965

Master of Science
Western Illinois University
Macomb, Illinois
1973

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
July, 1978

ON REDUCING BIAS IN

OBSERVATIONAL

STUDIES

Thesis Approved:

Lyle Broemeling
Thesis Adviser

Don Holbert

David L. Weeks

M. Palmer Terrell

Norman N. Durham
Dean of the Graduate College

ACKNOWLEDGMENTS

I am deeply indebted to my principal adviser, Dr. Lyle D. Broemeling, for his guidance, concern, and encouragement through this study. A special thanks is extended to Dr. J. Leroy Folks for his guidance, suggestions, and patience at the initial stage of this study before he went on sabbatical leave in July, 1977.

I wish to thank all the members of my advisory committee for their assistance and constructive suggestions.

Special thanks are also extended to Dr. David L. Weeks for his interest in my topic and for being a constant source of encouragement.

I would like to thank Dr. Don Holbert for serving on my advisory committee and for his assistance in correcting the final copy of my thesis.

I would like to thank Drs. Ronald W. McNew, Marvin P. Terrell, Kenneth E. Case for serving on my advisory committee.

Appreciation is extended to my friend, Peter Nelson, for helping me to improve my English writing, especially in this thesis.

I would like to extend my thanks to faculty members and graduate students for the friendship and help they have given me.

Finally, I wish to express my appreciation to Sara Veitl, who served as my typist.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Background and Need for Study

In recent years the number of studies which attempt comparisons of treatments effects without adequate randomization has increased rapidly, particularly in disciplines and areas of research involving human populations. The term "observational" has been employed to denote this type of investigation which can be somewhat vaguely described as a nonexperiment. Perhaps one of the clearest discussions of this type of investigation was that presented by Cochran (1965) who suggested two main distinguishing characteristics:

> 1. The objective is the investigation of possible cause-effect relationships.
> 2. This is implemented by the comparison of groups subject to different "treatments" which were preassigned in a nonrandom manner (p. 234).

Following Cochran's suggested characteristics (1965), the term "observational study", as employed in this thesis, will denote that type of study which is concerned with investigating relationships among characteristics of human populations, after the manner of an experiment, but comparing groups among which the "treatments" are not randomly assigned.

Without random assignment to insure homogeneity of groups, statistical tools employed in the design and analysis of experiments to reduce variation were somewhat naturally adapted to the removal of bias. Two principal questions to be considered in designing observational studies, as noted by McKinlay (1975), are:

> (i) What criteria should be used in determining the most important sources of bias in a comparison of two variables?
> (ii) Which method, or combination of methods, will most effectively remove such bias from the comparison? This question of effectiveness contains two related considerations--the number of covariables included and the proportion of bias eliminated for any given set of covariables (p. 512).

Cochran (1953) did briefly consider the first question, on the selection of covariables in relation to the multiple correlation coefficient. In comparing pair-matching with independent samples, the reduction in variation of a response variable due to matching is ($1 - R^2$) when several independent covariables are considered (here R is the multiple correlation coefficient). As this reduction is not substantial for $R < 0.5$, Cochran suggested that selection of covariables be based on the size of the individual correlation coefficients, matching on those with $r_{yx} \geq 0.3$. This simple rule of thumb leads to the obvious violation of the independence among the covariables which is assumed in his paper.

In considering the second question, the outstanding tool among available techniques which have been so employed is pair-matching. McKinlay (1975) noted two important reasons for the almost universal adoption of this tool:

> (i) as a technique for eliminating unwanted effects it is conceptually easy for the statistically unsophisticated researcher to comprehend; and (ii) pair-matching is applicable as a method regardless of the types, or distributions of variables being considered (p. 504).

However, the disadvantages of pair-matching and the inability of pair-matching to remove all bias have been pointed out by Thorndike

(1942), McNemar (1940), Billewicz (1965), Cochran (1953), and many others. McKinlay (1975) also concluded that pair-matching will not be the optimal choice, as the loss of potential information will not be offset by any commensurate increase in either efficiency or effectiveness in removing bias.

A growing awareness of the problems associated with pair-matching has led to a search for other methods to remove bias from two study groups. Various methods have been developed and discussed. Most of these studies have been concerned with the effects of a single independent variable acting on a single study variable.

A review of the literature shows that while a variety of statistical tools are applicable to the collection and analysis of observational data, the emphasis on the comparisons of the statistical methods has been on the efficiency of the methods, in terms of precision. Only very recently has attention been focused on the effectiveness of these methods in reducing bias (the unwanted effects of concomitant variation) which is a principal concern in observational studies.

Within the rapidly expanding field of observational research, the need to develop new methods for design and analysis of observational studies in various situations would seem paramount. With this goal in mind, this thesis is intended to develop a method, or combination of methods, which will more effectively remove the bias from the estimate of the treatment comparison of two groups. We will consider the case where the response variable has a quadratic relation with a single continuous covariable. The study of the quadratic relation may indicate which method will be more effective in reducing bias in situations where the response variable and the continuous covariable are nonlinearly (or nearly quadratically) related. This thesis also presents two methods of reducing the bias in observational studies with two covariables. These methods consist of the combination of transformation and stratification. For example, these methods may be used to reduce bias for two normally distributed covariables. We consider the situation where the covariables have the same covariance matrix, but have different means.

Organization of This Thesis

The organization of this thesis is as follows. The literature pertaining to the methods to remove bias is reviewed in Chapter II. In Chapter III, randomization analysis for a single covariable is studied. In Chapter IV, we use the combination of stratification and covariance adjustment for the treatment effect. A Monte Carlo study is presented. Extension of the stratification to two covariables, and different methods of transformations combining with stratification is discussed in Chapter V. The thesis is then briefly summarized in Chapter VI.

CHAPTER II

LITERATURE REVIEW

Methods to Remove Bias

Most of the following discussion will be confined to studies in which we compare two populations, which will be called the experimental population and the control population. We shall suppose that we cannot create the experimental population, but must take it as we find it. In the comparison of the two populations, pair-matching is perhaps the most popular technique to remove bias in an observational study. Each member of the experimental sample is taken in turn, and a partner is sought from the control population which has the same values as the experimental member (within defined limits) for each of the covariables. Emphasis in the observational studies has increasingly been given to the investigation of multivariate sources of variation rather than simply being restricted to the removal of bias from the comparison in groups for a single covariable. Consequently, various problems have been encountered by using the pair-matching technique in the field of observational studies.

McKinlay (1975) found that, for samples of equal size and an equivalent number of matching categories of a qualitative response, only 50 percent of the maximum matches could be expected in pair-matching and that even with a reduction in matching categories, the number of pairs could never be expected to reach the maximum number of pairs as he concluded. For unequal samples, it was noted that a ratio of at least 1:5 would be needed in most instances to obtain a near maximum number of pairs, provided that the smaller sample size exceeded the number of matching categories.

Rubin (1973) compared directly the effectiveness of pair-matching and covariance analysis in removing bias from a quantitative response. He concluded that pair-matching was not preferable to the use of independent samples. His conclusion is consistent with the findings of McKinlay, although the latter is concerned with a qualitative response. This awareness of the problems with pair-matching has led to a search for other alternatives which will effectively reduce the bias in the observational studies and will not suffer as many problems as pair-matching does.

An early suggestion for the use of regression analysis rather than pair-matching was made by Peters (1941), whose primary aim was to avoid the loss of "unmatchables". He calculated an expected value, using regression coefficients estimated from a control group. This method was in essence a covariance analysis employing regression coefficients estimated from the control group only.

When dealing with the problems associated with observational work in epidemiological research, Greenberg (1953) found that the combination of balancing (equating covariable means) and covariance analysis yielded the most precise estimate among pair-matching and analysis of covariance.

Belson (1956) also used covariance analysis with estimates of the regression coefficients from the control group as a possible solution to the problem of non-parallel regression.

Another popular alternative to pair-matching is the method of stratification (stratified matching). In this method, the distribution of the covariables is divided into c subclasses. For each group of subjects, the mean value of the response is calculated separately within each subclass. Then a weighted mean of these subclass means is calculated for each group, using the same weights for every group. The actual weights employed depend on the judgment of the investigator.

Cox (1957) considered the optimum grouping of a population on a continuous variable. For a normal distribution, the percentage of information retained by dividing the distribution into three groups was about 80%, increasing to between 90% and 95% for six groupings. Moreover, there was little difference in the corresponding percentages for optimal and equal strata. The implication for stratifying on quantitative covariates is that between three and six divisions should be sufficient in most cases, at least for those distributions which were near normal.

Stratified matching, which is more expeditious than pair-matching, is superior to pair-matching in removing bias and maximizing precision as the initial bias is increased. Although the covariance analysis gives greater gains in removing the bias than stratification when the response variable is linearly related to the covariable, as noted by Cochran (1968) and Billewicz (1965), when the response variable has a curvilinear relation with the covariable the stratified matching should be preferred.

From these findings, it would be worthwhile to further explore and study the method of stratification, especially when the means of the covariables are different. Also if pair-matching is difficult to accomplish, we should use the stratification method.

Comparison of Different Methods

In comparing two populations, where one is the experimental population and the other the control population, matching of the experimental and control samples with respect to the covariables can be accomplished in a number of ways. Conceptually, the simplest method is the method of pairing. It is difficult to discuss the effectiveness of pairing in realistic terms. The advantages of pairing and of covariance analysis are usually demonstrated by means of a linear regression model.

Let y denote the variable by which the effects of the experimental factor are measured, and x denote the covariable. The model assumes y has a linear regression on x with the same slope in each population. The equations are as follows:

Experimental population $y_1 = t_1 + \beta x_1 + e_1$ (1.1)

Control population $y_2 = t_2 + \beta x_2 + e_2$ (1.2)

The variables x and e are independently distributed; the deviations e_1, e_2 have zero means in both populations and constant variance σ_e^2. Further, it is assumed that the means u_1, u_2 of x in the two populations are equal and that $t_1 - t_2$ represents the true effect of the experimental factor, i.e., we do not have unsuspected biases.

With this model, the precision given by paired samples can be compared with that given by independent random samples drawn from the two populations. The effect of the experimental factor will be estimated by the difference $(\bar{y}_1 - \bar{y}_2)$ between the means of the two samples in either method. For independent samples, each of size n, the variance V_i of $(\bar{y}_1 - \bar{y}_2)$ is

$$V_i = \frac{2}{n}\sigma_y^2 . \tag{1.3}$$

Here we assume for simplicity that σ_y is the same in both populations.

On the other hand, with samples paired on x the variance V_p of $(\bar{y}_1 - \bar{y}_2)$ is

$$V_p = \frac{2}{n}\sigma_e^2 . \tag{1.4}$$

From (1.1), assuming that ρ is the correlation coefficient between y and x, we obtain

$$V_p = \frac{2}{n}\sigma_y^2 (1 - \rho^2) . \tag{1.5}$$

Comparison of (1.5) with (1.3) shows that pairing has higher precision. If pairing is accomplished for several x-variables, all linearly related to y, the variance of V_p is $\frac{2}{n}\sigma_y^2 (1-R^2)$, where R is the multiple correlation coefficient between y and x.

If, instead of pairing, we draw random samples of size n from each population and adjust the sample mean by covariance, then, on the average, the variance V_a of $(\bar{y}_{1a} - \bar{y}_{2a})$, the adjusted mean difference for the covariables, given by Cochran (1953), is

$$V_a = \frac{2}{n}\sigma_y^2 (1 - \rho^2)\left\{ 1 + \frac{1}{2(n-2)} \right\} , \tag{1.6}$$

when the means u_1, u_2 of the covariables are the same.

Pairing requires that data on the values of the covariables in the control population be readily accessible; this may not be the case. One disadvantage is the time spent in constructing the pairs. When the means u_1, u_2 of x in the two populations are different, some difficulty may be experienced in finding control partners for the experimental sample. With the covariance method the corresponding variance, when the means u_1, u_2 of x in the two populations are different, may be shown to be approximately

$$V_a = \frac{2}{n} \sigma_y^2 (1 - \rho^2) \left\{ 1 + \frac{1}{2(n-2)} + \frac{n(u_1 - u_2)^2}{4(n-2) \sigma_x^2} \right\} \tag{1.7}$$

assuming x is normally distributed with the same variance in each population.

In comparing pair-matching and covariance analysis, Billewicz (1965) noted that the effectiveness of pair-matching appeared to decrease considerably in comparison to covariance adjustments for a quantitative response, as (i) the correlation between the covariate and response variable increased, and (ii) sample size decreased, given that regressions were parallel and linear.

Cochran (1968) analytically derived the bias removal and variance reduction for stratified matching by assuming the distribution of the covariable to be normal. The proportion of the initial bias that is removed is approximately

$$\sum_{i=1}^{c} M_i(f_{i-1} - f_i). \tag{1.8}$$

In this expression, f_{i-1} and f_i are the ordinates of the density f(x) at the boundaries x_{i-1} and x_i of the ith subclass, and M_i is the mean value of x in the ith subclass, i = 1, 2, . . ., c, assuming that the initial bias is small.

When x has the same distribution f(x) in the two groups and the regression of y on x is linear, the variance V_s of $(\bar{y}_1 - \bar{y}_2)$ after the adjustment by stratification is

$$V_s = \frac{2}{n} \sigma_y^2 \left\{ 1 - (1-g)\rho^2 \right\} . \tag{1.9}$$

Here $g = \sum_{i=1}^{c} p_i \frac{\sigma_i^2}{\sigma_x^2}$, p_i is the proportion of the population and σ_i^2 is the variance of x within the ith subclass.

For the normal distribution, N(0,1), $1 - g = \Sigma M_i (f_{i-1} - f_i)$. This is the proportional reduction in the variance of $\bar{x}_1 - \bar{x}_2$ due to adjustment by stratification and is also the proportional reduction in bias.

CHAPTER III

RANDOMIZATION ANALYSIS

Univariate with Parallel Slope

To realize the function of pair-matching, we first perform the pairing of the units so that within each pair the covariables have the same value. After pairing is completed, treatments are applied at random to units within each pair. Following the derivation of the analysis of randomized experiments by Kempthorne (1973), we postulate the existence of a real (unknown) number Y_{ijk} which represents the true response if the ith unit in the jth pair is subjected to treatment k, where i, k = 1,2, and j = 1,2, . . ., N. Further, we assume that Y is linearly related to a covariable X with the true relation being

$$Y_{ijk} = t_k + \beta X_{ij}.$$

In general, we are able to observe only a subset of the Y_{ijk} and hence our inferences will be influenced by additional variabilities. The function of randomization is to control, in a statistical sense, these additional variabilities, and to enable us to obtain valid estimates of the treatment effects.

In order to write an explicit model for the Y_{ijk} in terms of the parameters of interest it is useful to introduce some additional definitions and notations. Let d_{ij}^{k} be a random variable with the following properties:

$d^k_{ij} = 1$ if treatment k is applied to the ith unit in the jth pair

$= 0$ otherwise

$P\left\{ d^k_{ij} = 1 \right\} = \frac{1}{2}$ for any i, j, k

Given that

$d^k_{ij} = 1$, then $d^k_{i'j} = 0$ for all $i \neq i'$ and $d^{k'}_{ij} = 0$ for all $k \neq k'$

d^k_{ij} and $d^{k'}_{i'j'}$ are independent if $j \neq j'$ for any i, i', k, k'

Given that

$d^k_{ij} = 1$, $P\left\{ d^{k'}_{i'j} = 1 \right\} = 1$ for $i \neq i'$, $k \neq k'$.

These properties are an expression of the fact that we randomize the positions of the treatments in each pair separately and, of course, that a treatment occurs on only one subject within a pair and that any subject receives only one treatment.

Now let us examine the estimates of treatment effects. Random assignment of treatments implies that

$$\sum_i d^k_{ij} Y_{ijk} = \sum_i d^k_{ij} (t_k + \beta x_{ij}).$$

Thus

$$y_{jk} = t_k + \sum_i d^k_{ij} \beta x_{ij}, \text{ since } d^k_{ij} \text{ is one when } i = k, \text{ and}$$

$$E(\bar{y}_{.1} - \bar{y}_{.2}) = t_1 - t_2.$$

Therefore $\bar{y}_{.1} - \bar{y}_{.2}$ is an unbiased estimate of the treatment effect.

The function of pair-matching can be understood by considering the variance of the estimate of the treatment effect which is

$$V_{rp} = \frac{\beta^2}{N^2} \sum_j (x_{1j} - x_{2j})^2. \qquad (3.1)$$

If we have successfully matched the covariable for each pair, the variance V_{rp} should be very small.

The analysis of variance is given in Table I.

TABLE I

ANALYSIS OF VARIANCE
PAIR-MATCHING AND
PARALLEL

Due to	Sum of Squares	Expectation of Mean Square
Pairs	$\sum_{jk} (\bar{y}_{j.} - \bar{y}_{..})^2$	PMS*
Treatments	$\sum_{jk} (\bar{y}_{.k} - \bar{y}_{..})^2$	TMS*
Remainder	$\sum_{jk} (y_{jk} - \bar{y}_{j.} - \bar{y}_{.k} + \bar{y}_{..})^2$	RMS*
Total	$\sum_{jk} (y_{jk} - \bar{y}_{..})^2$	

$$PMS^* = \frac{2\beta^2}{N-1} \sum_j (\bar{x}_{.j} - \bar{x}_{..})^2$$

$$TMS^* = \frac{N}{2} (t_1 - t_2)^2 + \frac{\beta^2}{N} \sum_{jk} (x_{kj} - \bar{x}_{.j})^2$$

$$RMS^* = \frac{\beta^2}{N} \sum_{jk} (x_{kj} - \bar{x}_{.j})^2.$$

From Table I, the treatment effect can be tested by using TMS/RMS; the distribution of the covariable may be ignored. However, if one is interested in finding the power of the test, distributional properties of the covariable x should be assumed.

Instead of using the pair-matching procedure, a balancing ("mean" matching) may be performed prior to the random assignment of the treatments applied to the groups which are well matched on the basis of their covariable means.

The random variable d_{ij}^{k} has the following properties:

d_{ij}^{k} = 1 if treatment k occurs on the jth unit in the ith group

= 0 otherwise

$P\{ d_{ij}^{k} = 1 \} = \frac{1}{2}$ for every i, j, k

$P\{ d_{ij}^{k} = 1 \mid d_{i'j'}^{k'} = 1 \} = 1$ if $i = i'$, $k = k'$ and $j \neq j'$

= 0 otherwise.

The results of mean-matching are:

$$E(\bar{y}_{.1} - \bar{y}_{.2}) = t_1 - t_2.$$

This is the same as in pair-matching. The variance of $\bar{y}_{.1} - \bar{y}_{.2}$ is

$$V_{rm} = \frac{\beta^2}{N^2}\left(\sum_j x_{1j} - \sum_j x_{2j} \right)^2. \tag{3.2}$$

To decide whether pair-matching or mean-matching should be used, we may compare the equations (3.1) and (3.2), given that the relation between y and x is linear and parallel in the two groups. In general, it is easier to apply the mean-matching procedure than pair-matching. After ordering the data $x_{(1)}, x_{(2)}, \ldots, x_{(2N)}$, one may choose $x_{(1)}$ and $x_{(2N)}$ for one group and choose $x_{(2)}$ and $x_{(2N-1)}$ for the other group, continuing this procedure until all the data are used. This procedure guarantees that one always can make the value in equation (3.2) smaller than that in equation (3.1). One disadvantage is that mean-matching depends on the model assumed. If the model assumed is a true model then mean-matching is preferable.

Univariate with Unequal Slopes

We now formulate a model in which additivity does not hold. We may write

$$Y_{ijk} = t_k + \beta_k x_{ij}$$

and examine the effect of pair-matching for this model. The random procedure used for pair-matching is identical to that when we assume that the model is additive. Random assignment of treatment k to unit i in the jth pair gives the results

$$\sum_i d_{ij}^k \, Y_{ijk} = t_k + \sum_i d_{ij}^k \, \beta_k \, X_{ij}.$$

It is instructive to examine that expectation of the usual estimates if this is the true model. We have

$$E\,(\bar{y}_{.1} - \bar{y}_{.2}) = t_1 - t_2 + (\beta_1 - \beta_2)\bar{x}_{..} \, . \tag{3.3}$$

The variance of the mean difference $(\bar{y}_{.1} - \bar{y}_{.2})$ is

$$V'_{rp} = \frac{(\beta_1 + \beta_2)^2}{4N^2} \sum_j (x_{1j} - x_{2j})^2. \tag{3.4}$$

In this analysis, the observed mean of a treatment estimates the mean response we would obtain had all the experimental units been subjected to that treatment. One obvious function in pair-matching is to increase the precision as we can see from equation (3.4)

Now let us examine the usual analysis of variance which is given in Table II. Under the null hypothesis that the treatments have identical effects on all units, we may use TMS/RMS to test the treatment effect. It should be noted that we are considering the estimation problem. The analysis of variance given in Table II is entirely irrelevant from the point of view of the testing of the hypothesis that there are no treatment effects, for we have obtained the expectations over the population of possible experiments that we could have obtained. As regards the testing of the hypothesis, we shall obtain one experiment only, and we shall apply the randomization test procedure to that one experiment. This test procedure would consist of superimposing all the possible randomizations on the set of yields we would obtain in the particular experiment and evaluating some criterion for each randomization. If this criterion is in the critical region, we reject the hypothesis that there are no treatment effects.

TABLE II

ANALYSIS OF VARIANCE
PAIR-MATCHING AND
NONPARALLEL

Due to	Sum of Squares	Expectation of Mean Square
Pairs	$\sum_{jk} (\bar{y}_{j.} - \bar{y}_{..})^2$	PMS*
Treatments	$\sum_{jk} (\bar{y}_{.k} - \bar{y}_{..})^2$	TMS*
Remainder	$\sum_{jk} (y_{jk} - \bar{y}_{j.} - \bar{y}_{.k} + \bar{y}_{..})^2$	RMS*

$$PMS^* = \frac{(\beta_1 + \beta_2)^2}{2(N-1)} \sum_j (\bar{x}_{.j} - \bar{x}_{..})^2 + \frac{(\beta_1 - \beta_2)^2}{8N} \sum_j (x_{1j} - x_{2j})^2$$

$$TMS^* = \frac{N}{2} \{t_1 - t_2 + (\beta_1 - \beta_2)\bar{x}_{..}\}^2 + \frac{(\beta_1 + \beta_2)2}{8N} \sum_j (x_{1j} - x_{2j})^2$$

$$RMS^* = \frac{(\beta_1 - \beta_2)^2}{2(N-1)} \sum_j (\bar{x}_{.j} - \bar{x}_{..})^2 + \frac{(\beta_1 + \beta_2)^2}{8N} \sum_j (x_{1j} - x_{2j})^2.$$

If the treatment effects are not additive, comparisons of the observed means will be of value to the experimenter, because they give estimates of treatment differences over a well-defined population. In observational studies, we are interested in estimating the mean response we would have obtained had all the experimental units been subjected to that treatment. Consequently, the quantity to be estimated should be

$$t_1' - t_2' = t_1 - t_2 + (\beta_1 - \beta_2)\bar{x}_{..}$$

when the treatment is non-additive for a fixed population.

To estimate the treatment effect in this case, we first estimate the slope from either group one or group two, depending on which group more nearly represents the whole population on the basis of the covariables. Let β be estimated from group one as follows:

$$\hat{\beta}_1 = \frac{\sum_j (x_{1j} - \bar{x}_{1.})y_{1j}}{\sum_j (x_{1j} - \bar{x}_{1.})^2} \quad . \tag{3.5}$$

The treatment effect is estimated by

$$t_1' - t_2' = (\bar{y}_{.1} - \hat{\beta}_1\bar{x}_{1.}) - (\bar{y}_{.2} - \hat{\beta}_1\bar{x}_{2.}). \tag{3.6}$$

The expectation of $t_1' - t_2'$ is

$$E\,(t_1' - t_2') = t_1 - t_2 + (\beta_1 - \beta_2)\bar{x}_{2.} \tag{3.7}$$

The quantity $t_1' - t_2'$ is an unbiased estimate of the treatment effect if $\bar{x}_{2.}$ is the population mean of the covariable x.

The purpose of this chapter is to shed some light on the estimation problem when the response surfaces are nonparallel in the observational studies. For illustrative purposes, we use the example given by Wang, Novick, Isaacs, and Ozenne (1977). In that example, they examined the effectiveness of compensatory education as compared to a standard treatment.

The random assignment of individual students cannot be accomplished because of the disruption of school routine and burden of cost involved. Thus for all practical purposes, it is necessary to work with two intact classes, one of which is thought to be well below some desired level of functioning, possibly because of disadvantages homes, neighborhoods, or school environment, the other is a normal class. Then the special treatment is assigned to the disadvantaged class and the other to control. In this case, we would expect to observe a larger difference in the post-test scores (y) from the two treatments for those students with higher pretest scores (x) than for those students with lower pretest scores if the compensatory education is effective in a fixed period. This means that the students with higher pretest scores should benefit more from the compensatory education than those with lower scores. Consequently, the effectiveness of compensatory education for these fixed set of students is estimated by

$$(\bar{y}_{.1} - \hat{\beta}_1\bar{x}_{1.}) - (\bar{y}_{.2} - \hat{\beta}_1\bar{x}_{2.}).$$

This seems to be very reasonable if the means of the initial scores are identical. In order to provide a sound basis for explanation in this study, the best approach is to carefully match groups with their respective pretest means.

The above analyses suggest a reasonable method to estimate the treatment effect when the treatment effects are nonadditive in the comparison of two groups in the observational studies.

CHAPTER IV

STRATIFICATION AND COVARIANCE ADJUSTMENT IN REMOVING BIAS

Univariate Stratification

We will present the results obtained by Cochran (1968) using stratification on a single covariable at the beginning of this chapter; these results will be referred to throughout this chapter and the later chapters.

Let $u(x)$ represent the population regression of y on x. If y_{1j}, y_{2j} are random responses from the two populations, the model is

$$y_{1j} = t_1 + u(x_{1j}) + e_{1j},$$
$$y_{2j} = t_2 + u(x_{2j}) + e_{2j}, \tag{4.1}$$

where e_{1j}, e_{2j} are random residuals with zero means in the respective populations. The quantity to be estimated is $(t_1 - t_2)$. For the unadjusted means of y in the two groups, it follows that

$$E(\bar{y}_{1.}) = t_1 + \bar{u}_1, \quad E(\bar{y}_{2.}) = t_2 + \bar{u}_2,$$

where

$\bar{u}_1 = \int u(x) f_1(x)\, dx$, $\bar{u}_2 = \int u(x) f_2(x)\, dx$, and $f_1(s)$, $f_2(x)$ are the marginal density functions of x of the two populations. Hence if no adjustment is made, the initial bias due to x is $\bar{u}_{1.} - \bar{u}_{2.}$.

In the stratification, the distribution of x in group one is divided into two, three, or more subclasses. For each group of subjects, the mean value of y is calculated separately within each subclass. Then a weighted mean of these subclass means is calculated for each group, using the same weights for every group.

In the ith subclass, let the boundaries of x be x_{i-1} and x_i and let the sample means of y be $\bar{y}^i_{1.}$ and $\bar{y}^i_{2.}$ The expectation of $\bar{y}^i_1 - \bar{y}^i_2$ in the ith subclass is

$$E\,(\bar{y}^i_{1.} - \bar{y}^i_{2.}) = t_1 - t_2 + \bar{u}_{1i} - \bar{u}_{2i}$$

where

$$\bar{u}_{ki} = \int_{x_{i-1}}^{x_i} u(x) \quad f_k(x)dx \;\Big/\; \int_{x_{i-1}}^{x_i} f_k(x)dx,\ k=1,2.$$

After adjustment, the remaining bias due to x is

$$\sum_i w_i\ (\bar{u}_{1i} - \bar{u}_{2i}),$$

where w_i is the weight assigned to subclass i.

The percent reductions in the bias of $\bar{x}_{1.} - \bar{x}_{2.}$ were calculated by Cochran using simulation for the case in which $f_1(x)$ is the normal distribution $N(u,\sigma^2)$, $f_2(x)$ is $N(0,\sigma^2)$, and the regression is linear. In this simulation, the boundary points for population one were chosen such that the proportions of the population subclasses were the same. Thus equal weights were used. For $u/\sigma = 1$, 0.5, 0.25 and for two, three, four, five, and six subclasses, the percent reductions in the bias of $\bar{x}_{1.} - \bar{x}_{2.}$ are shown in Table III.

Table III indicates that for initial biases which are not too large ($u/\sigma \leq 0.5$), the percent bias removed may be almost independent of the value of u/σ.

Based on the above observations, Cochran obtained the results by an analytical approach in which u/σ is assumed small.

TABLE III

PERCENT REDUCTIONS IN BIAS

LINEAR REGRESSION,

X NORMAL

	Number of Subclasses				
u/σ	2	3	4	5	6
1	61.8	78.2	85.3	89.1	91.5
0.5	63.2	79.1	85.9	89.6	91.8
0.25	63.6	79.3	86.0	89.7	91.9

Let $f(x)$ depend on a parameter u that has a nonzero value in population one and is zero in population two. For the adjustments, the range of x is divided into c subclasses by division points x_0, $x_1, \ldots, x_c$. In the ith subclass let $P_i(u)$ denote the proportion of the population one and $M_i(u)$ the mean value of x. The weights used may be the $P_i(0)$, the $P_i(u)$ or a combination of the two. Since u tends to zero in his approach, these different choices of weights become identical. Here the $P_i(0)$ are used.

If $M(u)$ denotes the overall mean of x, the initial bias, $M(u) - M(0)$ may be written as

$$\sum_{i=1}^{c} \{P_i(u)M_i(u) - P_i(0)M_i(0)\}$$

$$\doteq u \sum_{i=1}^{c} \{P_i \frac{dM_i}{du} + \frac{dP_i}{du}\} = u \frac{dM}{du} , \tag{4.2}$$

assuming u small, where the derivatives are taken at $u = 0$. After adjustment, the bias remaining is

$$\sum_{i=1}^{c} P_i(0) \{M_i(u) - M_i(0)\} \doteq u \sum_{i=1}^{c} P_i \frac{dM_i}{du} . \tag{4.3}$$

Consequently, the proportion of the initial bias that is removed is approximately

$$\sum_{i=1}^{c} M_i \frac{dP_i}{du} / \frac{dM}{du} . \tag{4.4}$$

The utility of this expression depends, of course, on whether the functions that enter into (4.4) are easily found analytically. If $f_1(x) = f(x-u)$, $f_2(x) = f(x)$, the denominator of (4.4) becomes one, since the initial bias in (4.2) is u. Further,

$$P_i(u) = \int_{x_{i-1}}^{x_i} f(x-u)dx = \int_{x_{i-1}-u}^{x_i-u} f(x)dx,$$

so that at $u = 0$,

$$\frac{dP_i}{du} = f(x_{i-1}) - f(x_i).$$

Finally, the proportional reduction in bias becomes

$$\sum_{i=1}^{c} M_i \{f(x_{i-1}) - f(x_i)\} \tag{4.5}$$

Further, Cochran showed that when the covariable has a normal distribution N(0,1), the proportional reduction in the variance of $\bar{x}_{1.} - \bar{x}_{2.}$ due to adjustment by stratification is equal to the proportional reduction in bias.

Estimate of the Treatment Effect

Cochran (1953) concluded that if the regression is nonlinear the precision of the covariance analysis will be decreased unless the presence of nonlinearity is recognized in the covariance analysis and we go to the trouble of fitting the appropriate type of regression curve. Stratification analysis may be considered as an alternative to covariance analysis. In the stratification, we may encounter the difficulty of not enough observations falling in certain subclasses, especially when there are too many subclasses. For example, if there are n_i observations in the ith subclass from group one, one may not be able to obtain the same number n_i of observations in the ith subclass from group two to reduce the bias in the means of the co-variables in the two groups.

Based on the consideration of removing the bias effectively, without too many subclasses, as well as the consideration of detecting the nonlinear trend in the response curve, we combine the stratification and covariance analysis to reduce the bias. This method requires fewer subclasses and should be able to remove nearly all the bias when the response curves in the two groups are parallel. If there are not too many subclasses, it should be easier to obtain the same number of observations in each subclass to reduce the bias in the means of the covariables in the two groups, especially when the means of the covariables are different.

For a random sample of size N from group one, the range of the covariable x is partitioned into c subclasses such that there are n_j observations in the jth subclass. Group two is constrained to have

n_j observations in this subclass also. For any observation from the ith group, we assume the model in (4.1).

Based on the assumption of a linear model $u(x_{ij}) = \beta x_{ij}$, the most common estimate of β comes from fitting the parallel linear response surface model by least squares. The $\hat{\beta}_p$ is calculated as

$$\hat{\beta}_p = \frac{\sum_{ij} (x_{ij} - \bar{x}_{i.})\, y_{ij}}{\sum_{ij} (x_{ij} - \bar{x}_{i.})^2} .$$

The treatment effect is estimated by

$$t_1 - t_2 = \bar{y}_{1.} - \bar{y}_{2.} - \hat{\beta}_p (\bar{x}_{1.} - \bar{x}_{2.}).$$

This method is the standard approach of the analysis of covariance for two groups, when we have random samples. The estimate is unbiased under the assumption $u(x_{ij}) = \beta x_{ij}$. The variance of this estimate is the same as that in (1.7). In general, the expectation of $t_1 - t_2$ given the x_{ij} is

$$\begin{aligned} E(t_1 - t_2) &= t_1 - t_2 + u(\bar{x}_{1.}) - u(\bar{x}_{2.}) - (\bar{x}_{1.} - \bar{x}_{2.}) \\ &\quad \cdot \frac{\sum_{ij} (x_{ij} - \bar{x}_{i.}) u(x_{ij})}{\sum_{ij} (x_{ij} - \bar{x}_{i.})^2} \\ &= t_1 - t_2 + \text{bias}. \end{aligned} \tag{4.6}$$

With this combination of stratification and analysis of covariance, we calculate β_h within each subclass. For x_{hij} and y_{hij} in the hth subclass from the ith group, the estimate of the linear coefficient β_h is

$$\hat{\beta}_h = \frac{\sum_{ij} (x_{hij} - \bar{x}_{hi.}) y_{hij}}{\sum_{ij} (x_{hij} - \bar{x}_{hi.})^2} . \tag{4.7}$$

The treatment effect is calculated by using a weighted average of the adjusted mean differences, namely

$$\sum_h w_h \{\bar{y}_{h1.} - \bar{y}_{h2.} - \hat{\beta}_h(\bar{x}_{h1.} - \bar{x}_{h2.})\}, \tag{4.8}$$

where $w_h = n_h/n$.

The expectation of the treatment effect (TES) from (4.8) for given x is

$$E(TES) = t_1 - t_2 + B, \tag{4.9}$$

where

$$B = \sum_h \frac{n_h}{N} \{\bar{u}_{h1.} - \bar{u}_{h2.} - (\bar{x}_{h1.} - \bar{x}_{h2.}) \frac{\sum_{ij}(x_{hij} - \bar{x}_{hi.})u_{hij}}{\sum_{ij}(x_{hij} - \bar{x}_{hi.})^2}\} .$$

The bias B in (4.9) should be expected to be smaller than the bias in (4.6) if $u(x_{ij}) \neq \beta x_{ij}$. When $u(x_{ij}) = \beta x_{ij}$, TES is an unbiased estimate of the treatment effect. The variance of TES is approximately

$$V(TES) \doteq \frac{\sigma_y^2(1-\rho^2)}{N} \left\{ 1 + \frac{h}{2(N-h)} + \frac{1}{h} \sum_i \frac{N(u_{i1}-u_{i2})^2}{4(N-h)\ \sigma_i^2} \right\} \tag{4.10}$$

where σ_i^2 is the variance of the covariable x in the ith subclass for $n_i = N/h$ and u_{ik} is the mean of x in ith class from kth group. When N is large (4.10) is nearly equal to (1.7), the variance of the adjusted mean difference. The precision in (4.10) is slightly lower than the usual covariance analysis if the model is linear. However, if the regression is nonlinear (4.10) should be smaller than (1.7).

Monte Carlo Investigation

When dealing with finite matched samples, the expectations required to calculate the reductions in bias are analytically intractable. Hence, we will turn to Monte Carlo methods in order to obtain numerical values for reduction in bias of the different estimates in nonlinear situations.

Cochran (1968) investigated three non-normal distributions (χ^2, t, and beta) in stratification; he concluded that the percent reductions in bias differ only trivially from those for the normal distribution. We will assume that in group i, $x \sim N(u_i, \sigma^2)$, $i = 1,2$. A quadratic relation between y and x may be representative of the non-linear situation. The true relation between y and x is $y_{ij} = t_i + (x_{ij} - 0.1)^2$ where $t_1 = 4$ and $t_2 = 2$ in this investigation. We choose $u_2 = u_1 - 0.5$ under the consideration that when u/σ is less than 0.5 the percent bias removed may be almost independent of the value u/σ. The values of u_1 used in this investigation are 0.5 (0.3) 2.6; the sample sizes are 10, 20, 50, and 100 for each pair of u_i's. With respect to stratification, we use two subclasses. The sample of size n from group one is generated by the subroutine GAUSS of the normal generator from the IMSL package of subroutines. The division points are $-\infty$, $\bar{x}_{1.}$, $+\infty$. In each subclass, we have n_j observations in the jth subclass for group one after stratification. A random sample of n_j is generated for the second group accordingly. The y values are generated according to the true relation $y_{ij} = t_i + (x_{ij} - 0.1)^2$ plus the standard normal deviate generated from the same subroutine. For the combination of stratification and covariance adjustment (SCA), we calculate the pooled estimate of the slope within each subclass; the treatment effect is adjusted accordingly. The overall treatment effect is the weighted average of the adjusted treatment effects from each subclass. Table IV gives Monte Carlo values of the percent reductions in bias after two types of adjustment; covariance adjustment

(CA) and combination of stratification and covariance adjustment (SCA).

TABLE IV

PERCENT REDUCTIONS IN BIAS AFTER ADJUSTMENT BY COVARIANCE AND COMBINATION OF STRATIFICATION AND COVARIANCE EQUAL SAMPLE SIZE IN SUBCLASS

u_1 ($u_2 = u_1 - 0.5$)	n=10		n=20		n=50		n=100	
	CA	SCA	CA	SCA	CA	SCA	CA	SCA
0.5	74.0	88.7	62.0	91.3	76.0	91.3	66.0	97.3
0.8	97.3	91.8	91.8	96.7	81.8	99.6	99.6	99.8
1.1	84.0	91.9	87.2	98.3	92.0	99.2	92.3	99.7
1.4	99.0	97.1	89.8	98.8	95.6	99.5	94.5	99.8
1.7	94.1	98.0	93.9	97.7	99.5	99.9	99.8	99.8
2.0	98.5	99.9	98.7	99.1	95.0	99.6	99.8	99.8
2.3	96.3	97.1	96.9	99.4	98.7	99.7	99.2	99.7
2.6	97.0	98.3	97.0	99.3	96.0	99.9	98.4	99.6

CA - Covariance Adjustment

SCA - Combination of Stratification and Covariance Adjustment

$x_{ij} \sim N(u_i, 1)$

Table V gives Monte Carlo values of the expected variance of the treatment effects.

TABLE V

VARIANCES OF TREATMENT EFFECTS WITH EQUAL SAMPLE SIZE IN SUBCLASS

u_1	n=10		n=20		n=50		n=100	
	CA	SCA	CA	SCA	CA	SCA	CA	SCA
0.5	.278	.132	.168	.025	.082	.009	.037	.006
0.8	.356	.082	.138	.034	.086	.010	.033	.006
1.1	.216	.086	.203	.031	.091	.101	.036	.006
1.4	.327	.067	.196	.032	.089	.011	.046	.005
1.7	.301	.097	.141	.035	.071	.010	.034	.006
2.0	.274	.083	.138	.029	.089	.011	.038	.005
2.3	.286	.073	.140	.031	.105	.010	.033	.004
2.6	.376	.085	.160	.033	.063	.007	.042	.005

In the previous experiment, the subclass sample sizes n_j within the first group determined the corresponding subclass sizes within the second group. In the second experiment, samples of size n are generated independently for each group and the combination of stratification and covariance adjustment is applied. The results for this experiment are given in Tables VI and VII.

TABLE VI

PERCENT REDUCTIONS IN BIAS AFTER ADJUSTMENT BY COVARIANCE AND COMBINATION OF STRATIFICATION AND COVARIANCE WITH INDEPENDENT SAMPLES

u_1 ($u_2 = u_1 - 0.5$)	n=10		n=20		n=50		n=100	
	CA	SCA	CA	SCA	CA	SCA	CA	SCA
0.5	72.7	-10.0	72.7	-132.0	63.3	-126.7	84.0	-153.3
0.8	94.7	57.6	85.6	27.8	92.0	24.4	97.8	15.1
1.1	81.2	69.5	93.6	56.0	93.9	51.2	97.9	50.7
1.4	95.2	81.9	98.9	65.0	96.1	63.4	98.1	61.0
1.7	93.2	87.6	93.6	75.4	98.9	74.1	97.8	71.2
2.0	98.6	9o.9	98.6	79.2	98.3	78.9	99.1	77.0
2.3	98.2	86.3	97.1	82.0	99.4	80.5	99.0	80.7
2.6	94.8	89.6	96.8	82.8	98.2	84.3	99.2	82.8

CA - Covariance Adjustment.

SCA - Combination of Stratification and Covariance Adjustment.

$x_{ij} \sim N(u_i, 1)$

TABLE VII

VARIANCES OF TREATMENT EFFECTS WITH INDEPENDENT SAMPLES

u_1	n=10		n=20		n=50		n=100	
	CA	SCA	CA	SCA	CA	SCA	CA	SCA
0.5	.260	.233	.181	.131	.056	.039	.034	.018
0.8	.424	.141	.241	.111	.084	.043	.040	.022
1.1	.377	.168	.141	.107	.074	.049	.039	.022
1.4	.312	.273	.225	.107	.063	.038	.038	.029
1.7	.310	.212	.183	.084	.066	.041	.039	.022
2.0	.363	.217	.116	.097	.071	.039	.029	.021
2.3	.358	.158	.185	.088	.074	.041	.034	.024
2.6	.313	.199	.228	.091	.111	.040	.030	.021

Table IV indicates that SCA technique removes nearly all the bias when the sample size is larger than 20 and when the subclass sample sizes are equal in each group. The precision of this method (SCA) is higher than the corresponding precision of the covariance analysis (CA) method in both experiments; the experiment when the subclass has equal sample size for each group and the experiment with random sample for each group. The SCA method is superior to CA method in removing the bias even when the sample size is less than 20. Thus it is recommended that when y has a quadratic relation with the covariable x the SCA method should be used. As the mean of the covariable increases, the percent reductions in bias become nearly equal for the two techniques. This is due to the fact that when the mean of the covariable is large, the relation between y and x is nearly linear. The result also indicates that when y is linearly related to x, the SCA technique with equal subclass sample sizes in each group is more effective than the CA method in removing the bias due to the covariable. When the mean of the covariable is close to 0.1, the CA method is less effective in reducing the bias than when the mean of the covariable is far from 0.1. This is due to the fact that the relation between y and x is more nonlinearly related when u_i is close to 0.1. The negative numbers in Table VI indicate that the bias is increased after adjustment. It should be noted that the SCA method is not recommended when the samples are randomly selected unless there are equal subclass sample sizes in each group. The conclusion from the above discussion is that the SCA method with equal subclass sample sizes in each group is more effective in removing the bias than the CA method if the relation between y and x is linear or quadratic.

When the slope is the same in each subclass, other estimates of the slope may be preferable. For instance, if the true regression coefficients β_h are the same in all subclasses, we may use the combined estimate

$$\hat{\beta}_c = \frac{\sum_h \sum_{ij} (x_{hij} - \bar{x}_{hi.})\, y_{hij}}{\sum_h \sum_{ij} (x_{hij} - \bar{x}_{hi.})^2}$$

and use this common estimate to adjust the treatment effect. However, this type of estimate needs further study. As far as we can judge, if the regressions are linear and if β_h appears to be the same in all subclasses, the combined estimate seems to be preferred.

CHAPTER V

EXTENSION OF THE STRATIFICATION

Most of the literature in the past has been concerned with the effect of a single stratifying variable. However, the multivariate situation is more common, with several variables both available and desirable for stratification.

Let $\underline{x}_i$ have a bivariate normal distribution with mean $\underline{u}_i$ and covariance matrix Σ for group i, i = 1,2. Here we assume that the covariables have different mean vectors in the two study groups but have the same covariance matrix. Notice that the matching on the covariables is intended to reduce the bias of some linear combination of the covariables $\underline{\beta}'\underline{x}$. For any given matching method, the method that reduces the bias in the means of the covariables does not necessarily reduce the bias of the linear combination of the means of the covariables. For example, let $\underline{\beta}' = (1, 1)$, and $\underline{u}_1' = (1,0)$, $\underline{u}_2' = (0,1)$. After some method of matching, say, we observe $\underline{u}_{1*}' = (.1, .1)$ and $\underline{u}_{2*}' = (-.1, -.1)$. The initial bias is zero for the linear combination of the covariables but the final bias is 0.4. Hence, if the covariables are thought to be linearly related to the response variable, a method that will guarantee bias reduction is desirable.

Bivariate Normal with $\rho = 0$

Let the model be

$$y_{ij} = t_k + \beta_1 x_{1ij} + \beta_2 x_{2ij} + e_{ij} \text{ for } i = 1,2;\ j = 1,2, \ldots, N. \tag{5.1}$$

Here the e_{ij} have mean zero, constant variance, and are independent of the covariables. In this model we assume that the response variable y in the two groups has the same linear relation with the covariables.

Without loss of generality, we assume that the mean values of the covariables x_1, x_2 in group one are u_1, u_2 and the mean values of the covariables x_1, x_2 in group two are zeros. A random sample of size n is obtained from each group. Without adjustment, the expected difference in the sample means of y from equation (5.1) is

$$E(\bar{y}_{1.} - \bar{y}_{2.}) = t_1 - t_2 + \beta_1 u_1 + \beta_2 u_2. \tag{5.2}$$

So the initial bias is $\beta_1 u_1 + \beta_2 u_2$.

In order to effectively reduce the initial bias we may consider an extension of the stratification method. We begin our study with a simple case; the covariance matrix is diagonal and $\underline{x}_i$ has a normal distribution in each group. The results for the stratification on a single variable can be easily extended.

We denote $f_1(x_1,x_2)$, $f_2(x_1,x_2)$ as the bivariate normal with mean $\underline{u}_i = (u_1,u_2)$ and $\underline{u}_2' = (0,0)$ and assume the covariance matrix in group one and group two are the same. The range of x_1, the first covariable in group one, is partitioned into c subclasses with the division points $x_{10}, x_{11}, \ldots, x_{1c}$ and range of x_2, the second covariable in group one, is divided into h subclasses in the same way with the division points $x_{20}, x_{21}, \ldots, x_{2h}$.

We shall need a set of notations which we define below:

$$P_{ij}(\underline{u}) = \int_i \int_j f_1(x_1,x_2)dx_1dx_2 = p_i(u_1)p_j(u_2)$$

$$P_{ij}(\underline{0}) = \int_i \int_j f_2(x_1,x_2)dx_1dx_2 = p_i(0_1)p_j(0_2). \tag{5.3}$$

The equations in (5.3) give the proportion in the (i,j)th cell for group one and group two.

$$f_i(x_1) = \int f_i(x_1, x_2)dx_2$$

$$f_i(x_2) = \int f_i(x_1, x_2)dx_1 \tag{5.4}$$

The above are the marginal densities of x_1, x_2 in the ith group.

$$M_{ij}(u_1) = M_i(u_1) = \frac{1}{P_i(u_1)} \int_i x_1 f_1(x_1)dx_1$$

$$M_{ij}(u_2) = M_j(u_2) = \frac{1}{P_j(u_2)} \int_j x_2 f_2(x_2)dx_2$$

$$M_{ij}(0_1) = M_1(0_1) = \frac{1}{P_i(0_1)} \int_i x_1 f_2(x_1)dx_1$$

$$M_{ij}(0_2) = M_i(0_2) = \frac{1}{P_j(0_2)} \int_j x_2 f_2(x_2)dx_2 \tag{5.5}$$

The equations in (5.5) are the means of the covariables x_1,x_2 in the (i,j)th cell in each group.

$$\sigma^2_{ij}(u_1) = \frac{1}{p_i(u_1)} \int_i x_1^2 \ f(x_1)dx_1 - \{M_i(u_1)\}^2 = \sigma^2_{1i}$$

$$\sigma^2_{ij}(u_2) = \frac{1}{P_j(u_2)} \int_j x_2^2 \ f(x_2)dx_2 - \{M_j(u_2)\}^2 = \sigma^2_{2j} \tag{5.6}$$

where σ^2_{1i} and σ^2_{2j} are the variances for the first and second covariables in the (i,j)th cell for each group.

In the study of a single covariable, we know that the precent bias removed due to stratification is almost independent of the value

u/σ when it is less than 0.5. So we may assume that $u_i \le 0.5$ for every i. For the adjustment, let $p_{ij}(\underline{0})$ be the weights used. The initial bias may be written as

$$\begin{aligned}
\beta_1 u_1 + \beta_2 u_2 &= \beta_1 \sum_{i=1}^{c} \{p_i(u_1)M_i(u_1) = p_i(0)M_i(0)\} \\
&\quad + \beta_2 \sum_{j=1}^{c} \{p_j(u_2)M_j(u_2) - p_j(0)M_j(0)\} \\
&\doteq \beta_1 u_1 \sum_{i=1}^{h} \{p_i \frac{dM_i}{du_1} + M_i \frac{dp_i}{du_1}\} \\
&\quad + \beta_2 u_2 \sum_{j=1}^{h} \{p_j \frac{dM_j}{du_2} + M_j \frac{dp_j}{du_2}\}
\end{aligned} \tag{5.7}$$

assuming u_1, u_2 small, where the derivatives are taken at $u_1 = 0$, $u_2 = 0$. After adjustment, the bias remaining is approximately

$$\beta_1 u_1 \sum_{i=1}^{c} p_i \frac{dM_i}{du_1} + \beta_2 u_2 \sum_{j=1}^{h} p_j \frac{dM_j}{du_2} \tag{5.8}$$

which is the same result as in (4.3). The generalization of the result in (4.5), the approximate proportional reduction in bias, becomes

$$\frac{\beta_1 u_1 R_1 + \beta_2 u_2 R_2}{\beta_1 u_1 + \beta_2 u_2} \tag{5.9}$$

where

$$R_1 = \sum_{i=1}^{c} M_i \{f(x_{1i-1}) - f(x_{1i})\} ,$$

$$R_2 = \sum_{j=1}^{h} M_j \{f(x_{2j-1}) - f(x_{2j})\} .$$

Expression (5.9) is a weighted average of proportional reduction in bias. The reduction in bias depends on the weights $\beta_1 u_1$ and $\beta_2 u_2$ as well as the sign of the weights. The adjustment may increase the bias. For example, let $\beta_1 u_1 = 1$, $\beta_2 u_2 = -2$ and $R_1 = 0.4$, $R_2 = 0.9$; then the approximate proportional reduction in bias in (5.9) is -1.4, that is, the bias is increased by 40% due to the adjustment, although we have reduced the proportion of the bias in the second covariable by 90%. In order to ensure that the adjustment will not increase the bias, expression (5.9) may give us some indication as how we should do the stratification. One obvious solution to this is to make $R_1 = R_2$. Then the proportion of reduction in bias is equal to the common proportion reduced in bias in each covariable. This suggests that, whenever possible, we should make the proportional reduction in bias for each covariable as nearly equal as possible. In some situations, other methods of stratification may be preferred, e.g., when the $\beta_i u_i$'s are known. In practice, we do not have the knowledge about $\beta_i u_i$. The conclusions from (5.9) are: (1) the proportional reduction in bias for each covariable should be the same, and (2) to satisfy the condition in (1), one way to accomplish it is to have the same number of subclasses for each covariable as well as to have the division points the same standardized distance from their respective means.

Effect of the Adjustment

If independent samples are drawn from the two groups, with no adjustment, the variance of $\bar{y}_{1.} - \bar{y}_{2.}$ is $2\,\sigma_y^2/n$. Of this, a part, $2(\rho_1^2 + \rho_2^2)\sigma_y^2/n$, is due to variations in $\underline{x}$ and a part, $2(1-\rho_1^2 - \rho_2^2)\sigma_y^2/n$, is due to other sources of variability. Here ρ_1 and ρ_2 are the correlation coefficients between y and x_1 and y and x_2 respectively. With stratified matching on x_i, the average value of the variance of x_i is $2\sum_j p_j \sigma_{ij}^2/n$.

The effect of stratified matching is therefore that the contribution of variations in $\underline{x}$ to $V(\bar{y}_{1.} - \bar{y}_{2.})$ is reduced from

$2(\rho_1^2 + \rho_2^2)\sigma_y^2/n = 2(\beta_1^2\sigma_1^2 + \beta_2^2\sigma_2^2)/n$ to

$$\frac{2\sigma_y^2}{n} \{\rho_1^2 \sum_{i=1}^{c} \frac{p_i\sigma_{1i}^2}{\sigma_1^2} + \rho_2^2 \sum_{j=1}^{h} \frac{p_j\sigma_{2j}^2}{\sigma_2^2}\} \tag{5.10}$$

Here σ_{1j}^2 and σ_{2j}^2 are the variances of x_1 and x_2 in the (i,j)th cell, and σ_1^2 and σ_2^2 are the variances of x_1, x_2.

If we let $g_1 = \Sigma p_i\sigma_{1i}^2/\sigma_1^2$ and $g_2 = \Sigma p_j\sigma_{2j}^2/\sigma_2^2$, the net result is to reduce $V(\bar{y}_{1.} - \bar{y}_{2.})$ from $2\sigma_y^2/n$ to

$$\frac{2\sigma_y^2}{n} \{1-\rho_1^2(1-g_1) - \rho_2^2(1-g_2)\}. \tag{5.11}$$

When $R_1 = R_2$, the proportional reduction in variance is equal to R_i. Though this equivalence of the proportional reduction in variance and bias appears to hold only for the normal distribution, it gives some indication about the reduction in variance due to stratification for other distributions which do not differ too much from the normal distribution.

Bivariate Normal with $\rho \neq 0$

As mentioned earlier, the selection of covariables should be based on the size of the individual correlation coefficients between the response variable y and the covariable x. Cochran suggested to match on those covariables where $r_{yx} > 0.3$. Thus we should study the situation when the covariables are correlated and have a linear relation with y. The results derived in the previous section are based on the assumption that the covariables are uncorrelated. A natural way to consider the stratification on the covariables which are correlated is to transform the covariables to be uncorrelated. This type of transformation requires the knowledge of the structure of the covariance matrix. However, this knowledge will not, in general, be available. A more practical assumption on the knowledge of r, the ratio of the variances of the covariables, seems not unrealistic in many observational studies. For example, the ratio of the variances of weight and height of the human beings is less variable than the variances of weight or height. The sources to obtain the information about the ratio of the variances are: (1) past surveys of similar variables, (2) a pilot study may be conducted prior to the study, and (3) use the ratio of the ranges as an estimate. Thus we propose a method that transforms the correlated covariables into uncorrelated ones and, hence, independent under normal assumption, assuming that we know r, the ratio of the variances of the covariables. Stratification is performed on these transformed independent covariables (STI).

For a given r, the ratio of the variances of the covariables x_1, x_2, we may consider a matrix

$$P = \begin{bmatrix} \frac{1}{\sqrt{2r}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2r}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

then

$$P^{-1} = \begin{bmatrix} \frac{\sqrt{r}}{\sqrt{2}} & -\frac{\sqrt{r}}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} .$$

where P^{-1} is the inverse matrix of P.

Let $\underline{x}_i'$ have a bivariate normal distribution with mean $\underline{u}_i'$ and covariance matrix Σ for each i. Here $\underline{u}_1' = (u_1, u_2)$, $\underline{u}_2' = (0,0)$, and Σ has the form

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} .$$

Since r is known, Σ will have the form

$$\Sigma = \sigma_2^2 \begin{bmatrix} r & \sqrt{r}\,\rho \\ \sqrt{r}\,\rho & 1 \end{bmatrix} .$$

The model given in (5.1) may be written as

$$\begin{aligned} y_{ij} &= t_i + \underline{\beta}'\underline{x}_{ij} + e_{ij} \\ &= t_i + \underline{\beta}'P^{-1}P\underline{x}_{ij} + e_{ij} \\ &= t_i + \alpha_1 z_{1ij} + \alpha_2 z_{2ij} + e_{ij} \end{aligned} \tag{5.12}$$

where

$$\alpha_1 = (\sqrt{r}\,\beta_1 + \beta_2)/\sqrt{2}$$

$$\alpha_2 = (-\sqrt{r}\,\beta_1 + \beta_2)/\sqrt{2} ,$$

and

$$z_{1ij} = x_{1ij} / \sqrt{2r} + x_{2ij}/\sqrt{2}$$

$$z_{2ij} = -x_{1ij} /\sqrt{2r} + x_{2ij} /\sqrt{2} .$$

Consequently, $\underline{z}_i'$ will have a bivariate normal distribution

$$\underline{z}_1 \sim MVN_2 \left[\begin{pmatrix} u_1/\sqrt{2r} + u_2/\sqrt{2} \\ -u_1/\sqrt{2r} + u_2/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{pmatrix} \sigma_2^2 \right]$$

$$\underline{z}_2 \sim MVN_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{pmatrix} \sigma_2^2 \right] . \tag{5.13}$$

After the transformation, we stratify on the z_{1ij}, z_{2ij} such that we have the same number of subclasses for each covariable. The division points used in the stratifications of z_{1ij} and z_{2ij} are chosen so that they are nearly same standardized distance from their respective means. The results derived from the previous section when $\rho = 0$ follow except that the sample sizes in each cell from each group may not be equal. However, the difference of the sample size from each group in the (i,j)th cell should be very small when ρ is less than 0.2. When ρ is large, stratification on the transformed covariables may encounter empty cells more often than the original covariables. In this case, we may reduce the number of cells in the stratification. If r, the ratio of the variances of the covariables, is unknown and is estimated from the sample variances with sample size n, the expected value of $\hat{r}$ is approximately

$$E\left(\frac{s_1^2}{s_2^2}\right) \doteq \frac{E(s_1^2)}{E(s_2^2)} - \frac{\mathrm{Cov}(s_1^2, s_2^2)}{\{E(s_2^2)\}^2} + \frac{E(s_1^2)}{\{E(s_2^2)\}^3} \cdot V(s_2^2)$$

$$= \frac{\sigma_1^2}{\sigma_2^2} - \frac{2\rho^2\sigma_1^2\sigma_2^2}{n(\frac{n-1}{n}\sigma_2^2)^2} + \frac{n-1}{n}\sigma_1^2 \times$$

$$\left\{\frac{n}{(n-1)}\right\}^3 \sigma_1^2 \left\{\frac{n}{(n-1)\sigma_2^2}\right\}^3 \frac{2}{n}\sigma_2^4$$

$$= \sigma_1^2/\sigma_2^2 \left\{1 - \frac{2n(1-\rho^2)}{(n-1)^2}\right. .$$

When the sample size is large the estimate $\hat{r}$ should be very close to σ_1^2/σ_2^2.

One of the major factors which affects the bias reduction in the model (5.1) is the size of the linear coefficients β_1, β_2. Transformation of the covariables to be uncorrelated is intended to make the two-way stratification easier to handle in bias reduction. In particular, we may use the results obtained in the adjustment of a single covariable when the $\beta_i u_i$'s are known. For example, the percent reduction in bias for a univariate x having normal distribution ranges over 63, 79, 86, 89, 92 for two to six subclasses when the original bias u/σ is less than 0.5. If $\beta_1 u_1 = 2$, $\beta_2 u_2 = -1$ and we use 5 or 6 subclasses for the first covariable, then we should use 3 or 4 subclasses for the second covariable whenever it is possible. In this way, we can reduce nearly all the bias. Without knowing the $\beta_i u_i$'s, transforming the covariables to be independent still does not guarantee that the resulting stratification will reduce the bias. Thus other types of transformation should be considered.

The second proposed transformation is based on large sample sizes. With large sample sizes, one should be able to estimate u_1, u_2 with high precision. For given u_1, u_2 (or estimate of the means), let $d = \sqrt{u_1^2 + u_2^2}$, and the transformation matrix Q be

$$Q = \frac{1}{d}\begin{bmatrix} u_1 & u_2 \\ -u_2 & u_1 \end{bmatrix}$$

We have

$Q'Q = QQ' = I$ the identity matrix.

The model in (5.1) can be written as

$$y_{ij} = t_i + \underline{\beta}'Q'Q\underline{x}_{ij} + e_{ij}$$

$$= t_i + \gamma_1 z_{1ij} + {}_2 z_{2ij} + e_{ij}$$

with

$$\gamma_1 = (\beta_1 u_1 + \beta_2 u_2)/d, \quad \gamma_2 = (-\beta_1 u_2 + \beta_2 u_1)/d,$$

and

$$z_1 = (x_1 u_1 + x_2 u_2)/d, \quad z_2 = (-u_2 x_1 + u_1 x_2)/d.$$

The transformed variables z_1 and z_2 have means d and zero in group one and zeros in group two. The main characteristic of this type of transformation is to change the contribution to the original bias from two variables to a single variate. After the transformation, the covariance matrix $Q\Sigma Q'$ is

$$\frac{1}{d^2}\begin{bmatrix} u_1^2\sigma_1^2 + u_2^2\sigma_2^2 + 2u_1u_2\sigma_{12} & \sigma_{12}(u_1^2 - u_2^2)+(\sigma_2^2 - \sigma_1^2)u_1u_2 \\ & u_1^2\sigma_1^2 + u_2^2\sigma_2^2 - 2u_1u_2\sigma_{12} \end{bmatrix}$$

Since all the bias comes from the first variable after transformation, the way to reduce the bias is simplified. The covariance matrix $Q\Sigma Q'$ indicates that if the magnitudes of u_1, u_2 and σ_1^2, σ_2^2 are the same, the two-way stratification after transformation can be simplified to the case of one-way stratification, ignorning the fact that the original σ_{12} is not zero. The implication from the above observation is that whenever the standardized means from the original variables have equal or nearly equal magnitudes we should transform the covariables. When $u_i/\sigma < 0.5$, the transformed covariables are nearly uncorrelated. We should expect that when $u_i/\sigma < 0.5$ the two-way stratification on the transformed variables will reduce nearly all the bias for 3x3 subclasses.

Monte Carlo Investigation

Monte Carlo methods are employed in this study. We assume that y, x_1, x_2 have a joint normal distribution in both groups. The following set of parameters are used in the model (5.1):

t_i = treatment effect in the ith group,

u_i = mean of the covariable x_i, $u_i = 0$ for the covariable x_i in the second group,

β_i = coefficient of x_i in model (5.1),

σ_y = standard deviation of y,

σ_i = standard deviation of the covariable x_i,

ρ = correlation coefficient between x_1 and x_2,

ρ_i = correlation coefficient between y and x_i,

R^2 = multiple correlation coefficient of y and x_i,

and $e_{ij} \sim N(0,1)$

The conditional distribution of y has a normal distribution with the following restrictions:

$$R^2 = (\rho_1^2 + \rho_2^2 - 2\rho\rho_1\rho_2) / (1 - \rho^2) < 1$$

$$\sigma_y^2 = 1 / (1 - R^2)$$

$$\beta_1 = \sigma_y(\rho_1 - \rho\rho_2) / \sigma_1(1 - \rho^2)$$

$$\beta_2 = \sigma_y(\rho_2 - \rho\rho_1) / \sigma_2(1 - \rho^2)$$

$$1 - \rho \geq -2\rho_1\rho_2 \text{ and } 1 + \rho \geq 2\rho_1\rho_2.$$

For values of ρ, we choose -0.2, -0.3, and -0.5, and σ_1, σ_2 are chosen to be one. With $\beta_1 = -1.4$ and $\beta_2 = -1.8$, we have

$$\sigma_y^2 = \beta_1^2\sigma_1^2 + \beta_2^2\sigma_2^2 + 2\beta_1\beta_2\rho\sigma_1\sigma_2 + 1,$$
$$= 6.2 + 5.04\rho.$$

The Table VIII will summarize the values of the parameters for given ρ.

All the values selected are based on the consideration of the previous discussions except $\rho_1 = -0.26$ (ρ_i should be greater than 0.3) which was selected so that we can see how it would affect the stratification on the transformed covariables. The data are generated from the IBM subroutine GGNRM and GGNRM1 with sample sizes n = 50, 100, 200. We investigate three types of two-way stratification on the same

TABLE VIII

VALUES OF THE PARAMETERS
FOR GIVEN ρ

ρ	σ_y^2	ρ_1	ρ_2	R^2
-0.2	5.19	-0.457	-0.667	0.807
-0.3	4.69	-0.397	-0.637	0.787
-0.5	3.68	-0.26	-0.573	0.728

set of data: (1) Regular two-way stratification (RTS), (2) Transformation to independence followed by two-way stratification (TIS), and (3) Transformation of the bias to one variable followed by two-way stratification (TOS). In the case of one-way stratification, the proportional bias reduction is approximately 80% for three subclasses. Cochran (1968) also suggested that it will be sufficient to use three to six subclasses in most cases. Here we use three subclasses on each covariable for a total of nine (3x3) cells. The boundary points are chosen $\bar{x}_{i.} \pm 0.44$ s.d. (sample standard deviation). In this investigation, the samples are generated independently with equal sample sizes for a given pair of u_1, u_2. Thus the sample size in the (i,j)th cell may be different in two groups.

Tables IX and X indicate that: (1) The percent reduction in bias is usually between 60% and 80% if we use the regular two-way stratification (RTS). The percent reductions in bias decreases as

TABLE IX

PERCENT REDUCTIONS IN BIAS ON BIVARIATE NORMAL DUE TO STRATIFICATION

	$u_1 = 0.25$			$u_2 = -0.25$					
	$\rho = -0.2$			$\rho = -0.3$			$\rho = -0.5$		
	RTS	TIS	TOS	RTS	TIS	TOS	RTS	TIS	TOS
n = 50	64.4	89.4	86.2	72.0	81.3	67.8	59.2	64.5	74.3
n = 100	70.4	88.8	64.5	74.2	71.3	67.1	60.1	78.6	64.5
n = 200	84.6	91.4	77.3	84.4	72.8	72.6	53.2	80.1	79.8
	$u_1 = -0.5$			$u_2 = 0.25$					
n = 50	85.5	80.9	83.1	57.6	82.6	75.4	64.1	85.1	76.2
n = 100	77.0	66.4	68.7	68.0	81.0	71.1	66.3	79.5	78.9
n = 200	74.3	78.5	75.6	68.5	80.6	72.4	62.8	77.2	75.3
	$u_1 = 0.25$			$u_2 = 0.25$					
n = 50	74.3	82.3	79.6	73.4	79.0	77.2	61.7	80.7	75.9
n = 100	74.9	81.6	79.3	73.5	78.1	78.8	60.9	79.7	73.4
n = 200	75.9	80.0	79.8	74.2	78.5	79.3	70.6	78.2	75.4

TABLE X

EXPECTED VARIANCES IN BIAS DUE TO STRATIFICATION ON BIVARIATE NORMAL

	$\rho = -0.2$			$\rho = -0.3$			$\rho = -0.5$		
	RTS	TIS	TOS	RTS	TIS	TOS	RTS	TIS	TOS
$u_1 = 0.25$				$u_2 = -0.25$					
n = 50	.072	.045	.043	.048	.041	.039	.023	.009	.015
n = 100	.021	.018	.022	.018	.019	.017	.013	.005	.008
n = 200	.011	.009	.010	.010	.009	.008	.006	.003	.007
$u_1 = -0.5$				$u_2 = 0.25$					
n = 50	.055	.044	.061	.054	.038	.032	.023	.016	.019
n = 100	.030	.016	.024	.026	.015	.017	.011	.007	.009
n = 200	.010	.010	.010	.009	.009	.007	.005	.003	.004
$u_1 = 0.25$				$u_2 = 0.25$					
n = 50	.040	.046	.063	.037	.031	.043	.018	.014	.017
n = 100	.023	.020	.015	.013	.018	.015	.010	.007	.008
n = 200	.013	.010	.010	.010	.008	.010	.012	.003	.005

ρ's increase. In some cases the proportional bias reductions decrease as the sample sizes increase. This is due to the fact that $\beta_1 \neq \beta_2$. As mentioned earlier, the reduction of the bias in the means of the covariables does not necessarily imply the reduction of the bias in the linear combination of the means. We also observe that the precision is lower than those of TIS and TOS techniques in most cases. In particular, the RTS is not preferred as far as bias reduction is concerned when ρ is larger than 0.3 and the sample size is larger than 50. This implies that we should use TIS or TOS technique when we estimate the ratio of the variances of the covariables with high precision or when we can estimate the biases precisely. (2) The transformation to independence (TIS) technique will remove the bias nearly equal to 80% in most cases. This is the expected result as we concluded from equation (5.9) and Table III. In the univariate case, the percent reductions in bias due to stratification for three subclasses is approximately 79.3 percent as shown in Table III. If the proportional reduction in bias for each covariable is nearly the same when the covariates are independent, then the proportion of reduction in bias is equal to the common proportion reduced in bias in each covariable. Also the precision is slightly higher than the other two techniques when we use TIS technique. (3) The equivalence in bias reduction between TOS and TIS techniques is due to the fact that $Q\Sigma Q'$ is a diagonal matrix after transformation. That is, the transformation of the bias to a single variate results the independence between the transformed covariables when the magnitudes of the bias from each covariable are the same (or nearly the same). From the above observation, the conclusion is that the two-way stratification can be simplified to a one-way stratification by using TOS technique when $u_1/\sigma_1 = u_2/\sigma_2$ and the covariables have a bivariate normal distribution. (4) The boundary points used in this study are not the optimum boundary points. This choice of the boundary points is based on the univariate case with approximately equal sample size in each subclass. A discussion of optimum stratification point is given by Ghosh (1963). However, the method given in that reference requires several stages of iteration. (5) In this study the pairs of means were chosen to represent those cases one is likely to encounter. Other possible pairs of the means for the covariables should produce results similar to one of the three types discussed.

From the above results, we may conclude that if it is possible the TOS or TIS technique should be applied to the two-way stratification whenever the response variable y is thought to be linearly related to the covariables.

If the number of cells is increased, the percent bias reduction should be expected to be higher and should be similar to Table III when we use TIS or TOS technique; however one may encounter the empty cells often.

Generalization of Multivariate Stratification

When y is linearly related to the covariables, in the bivariate normal case the TOS method is reduced to the case of a single variate when the sample size is not too small and when the magnitudes of the biases are nearly equal. This will imply that the transformed variables are nearly independent under normal assumption. Since in the observational studies an investigator generally has some knowledge of the biases of the covariables, it seems useful to generalize the TOS technique to multivariate normal. In this section we just discuss how to choose the Q matrix which will transform the bias from three covariables to univariate. The choice of Q matrix is not unique.

Let y be linearly related to three covariables x_1, x_2, x_3 with bias from each covariable being u_1, u_2, u_3 respectively. One way to choose the matrix Q is as follows: first transform the bias from x_1 and x_2 to a single variable, say x_1', with a matrix Q_1. Then the bias due to x_1' and x_3 is transformed to a single variable with the matrix Q_2. Finally $Q = Q_2Q_1$. For example, let the original bias vector be $\underline{u}' = (u_1, u_2, u_3)$ and $d_1 = (u_1^2 + u_2^2)^{\frac{1}{2}}$, $d_2 = (u_1^2 + u_2^2 + u_3^2)^{\frac{1}{2}}$. Then

$$Q_1 = \frac{1}{d_1}\begin{bmatrix} \begin{bmatrix} u_1 & u_2 \\ -u_2 & u_1 \end{bmatrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & 1 \end{bmatrix}$$

for the first transformation.

$$Q_2 = \begin{bmatrix} d_1/d_2 & 0 & u_3/d_2 \\ 0 & 1 & 0 \\ -u_3/d_2 & 0 & d_1/d_2 \end{bmatrix}$$

for the second transformation. Then

$$Q = Q_2Q_1 = \begin{bmatrix} u_1/d_1 & u_2/d_1 & u_3/d_1 \\ -u_2/d_1 & u_1/d_1 & 0 \\ -u_1u_3/d_1d_2 & -u_2u_3/d_1d_2 & d_1/d_2 \end{bmatrix}$$

is the final transformation matrix and $Q\underline{x}$ has the expected value $E(Q\underline{x}) = (d_2, 0, 0)'$. Consequently, $Q\underline{x}$ transforms the bias from three sources to that of a single variate. A three-way stratification may follow after this transformation. The generalization to the case of more than three covariable should present no difficulty. It may be fruitful to investigate the percent reduction in bias for more than three covariables. However, in practice, three covariables should explain a high percentage of the variation of the response variable. If more than three covariables are needed, the techniques of principal components analysis may be used for stratification. This technique seems particularly attractive with the advent of modern computing technology. Little work has been done, however, to investigate the efficiency of this technique relative to the standard approaches used with stratification variables.

For large sample sizes, we may consider the transformation of the covariables to be independent in the multivariate normal distribution. After this transformation the proposed TOS method may be applied. This procedure needs further investigation.

CHAPTER VI

SUMMARY

The primary objective of this study was to answer the question as to which method, or combination of methods, will be most effective in removing the bias from estimates of comparison of treatments in observational studies. The randomization analysis presented in Chapter III provides a method to estimate the treatment effect when there is an interaction between the treatments and the covariables. The role played by matching covariables in randomized experiments was also discussed. When a completely randomized experiment is feasible and interaction between the treatments and the covariables exists, we may estimate the linear coefficient from either the control group or the treatment group and the treatment effect is adjusted accordingly. The group from which the linear coefficient should be estimated may be judged on the basis of which group most nearly represents the whol population on the basis of the covariable. Whether the interaction exists or not, matching on the covariables is an important and essential step in a randomized experiment, as well as in observational studies.

The combination of stratification and covariance adjustment (SCA) method presented in Chapter IV is more effective in removing the bias than the covariance adjustment (CA) technique when the subclass sample sizes are equal in each group. The simulation shows that the

SCA technique removes nearly all the bias when y is linearly (or nearly linearly) related to the covariable x as well as when y has a quadratic relation with the covariable. As far as we can judge, this technique is a safer method to use if there is a nonlinear relationship between the response variable and the covariable; however, the investigator may fail to detect the existence of the nonlinearity. The only disadvantage of this technique is the complexity of the calaculations. However, if a high percentage of bias reduction is required, the gains in the bias reduction should more than compensate for the increased complexity of the calculations.

In Chapter V we proposed the TOS procedure which appears to be a very promising technique. It is effective in removing bias even when the correlation coefficients between the covariables are large provided that the standardized bias in each covariable is nearly equal. A simple one way stratification rather than a complex multi-way stratification may be used with the TOS method; we assume that the biases of the covariables are known or can be estimated with high precision. The TIS technique should be preferred if we have the knowledge of the structure of the covariance matrix in the covariables. This technique will guarantee the proportional reduction in bias to be nearly equal to the common reduction in bias for each covariable. In the simulation study, we estimated r, the ratio of the variances of the covariables, and applied the TIS procedure by transforming the covariables to be independent followed by stratification. The simulation study showed that TIS is the most effective technique in reducing the bias among the three techniques considered. Thus generalization of the TIS technique to the multivariate normal needs to be further studied. As noted earlier, the β_i's will affect the bias reductions in the stratification technique. Thus different combinations of the regression coefficients should be investigated. Also, it appears worthwhile to investigate the statistical properties of these transformation.

A SELECTED BIBLIOGRAPHY

Belson, W. A. "A Technique for Studying the Effects of a Television Broadcast." Applied Statistics, Vol. 5, No. 3 (1956), 195-202.

Billewicz, W. A. "The Efficiency of Matched Samples: An Empirical Investigation." Biometrics, Vol. 21 (1965), 623-644.

Campbell, D. T. and Erlebacher, A. "How Regression Artifacts in Quasi-Experimental Evaluations Can Mistakenly Make Compensatory Education Look Harmful." Compensatory Education: The Disadvantaged Child, Vol. III, New York: Brunner-Mazel, 1970.

Cochran, W. G. "Matching in Analytical Studies." American Journal of Public Health, Vol. 43 (1953), 684-691.

Cochran, W. G. Sampling Techniques (2nd Ed.). New York: Wiley, 1963.

Cochran, W.G. "The Planning of Observational Studies of Human Populations." Journal of the Royal Statistical Society, Ser. A, 128, Part 2 (1965), 234-266.

Cochran, W. G. "The Effectiveness of Adjustment by Sub-Classification in Removing Bias in Observational Studies." Biometrics, Vol. 24 (1968), 295-313.

Cox, D. R. "Note on Grouping." Journal of the American Statistical Association, Vol. 52 (1957), 543-547.

Ghosh, S. P. "Optimum Stratification with Two Characters." Annals of Mathematical Statistics, Vol. 34 (1963), 866-872.

Greenberg, B. G. "The Use of Covariance and Balancing in Analytical Surveys." American Journal of Public Health, Vol. 43 (1953), 692-699.

Kempthorne, O. The Design and Analysis of Experiments. New York: Robert E. Krieger Publishing Co., 1973.

McKinlay, S. M. "The Design and Analysis of the Observational Study - A Review." Journal of the American Statistical Association, Vol. 70 (1975), 503-520.

McKinlay, S. M. "An Investigation of the Effectiveness of Pair-Matching and the Stratification of Independent Samples in Eliminating Bias from a Comparison of Qualitative Variables." (Ph.D. thesis, Aberdeen, Scotland, 1973.)

McNemar, Q. "A Critical Examination of the University of Iowa Studies of Environmental Influences upon the I. Q." Psychological Bulletin, Vol. 37 (1940), 63-92.

Peters, C. C. "A Method of Matching Groups for Experiment with No Loss of Population." Journal of Educational Research, Vol. 34, No. 8 (1941), 606-612.

Rubin, D. B. "Matching to Remove Bias in Observational Studies." Biometrics, Vol 29 (1973), 159-183.

Rubin, D. B. "The Use of Matched Sampling and Regression Adjustment to Remove Bias in Observational Studies." Biometrics, Vol. 29 (1973), 185-203.

Thorndike, R. L. "Regression Fallacies in the Matched Groups Experiment." Psychometrika, Vol. 7 (1942), 85-102.

Wang, M., Novick, M. R., Isaacs, G. L., and Ozenne, D. "A Bayesian Data Analysis System for the Evaluation of Social Program." Journal of the American Statistical Association, Vol. 72 (1977), 711-722.

VITA

Oscar Chi-Hsuing Lin

Candidate for the Degree of

Doctor of Philosophy

Thesis: ON REDUCING BIAS IN OBSERVATIONAL STUDIES

Major Field: Statistics

Biographical:

Personal Data: Born in Taipei, Taiwan, Republic of China, November 26, 1940, the third son of Mr. and Mrs. Ah-Chung Lin.

Education: Graduated from Wen Shan High School, Taiwan, China, in 1960; received Bachelor of Arts degree in English from National Taiwan University, Taiwan, China, in June, 1965; received Master of Science degree in Mathematics from Western Illinois University in 1973; completed requirements for the Doctor of Philosophy degree at Oklahoma State University in July, 1978.

Professional Experience: Junior College instructor of English at Yen Pei Institute and Technology, Taiwan, China, 1966-1967. Junior High School teacher of math, English at Kin Shan Junior High School, Taiwan, China, 1967-1969. Graduate assistant, Mathematics Department, Western Illinois University, 1971-1973. Graduate teaching associate, Statistics Department, Oklahoma State University, 1973-1978.

Professional Organizations: American Statistical Association, Mu Sigma Rho.