THE MISER CRITERION FOR IMBALANCE IN ANOCOVA

by

S. James Press

Technical Report No. 205
December 1984

Prepared under the Auspices of
National Science Foundation
2FPZ459
Ingram Olkin, Project Director

Department of Statistics
Stanford University
Stanford, California
THE MISER CRITERION FOR IMBALANCE IN ANOCOVA

by

S. James Press

Stanford University*

ABSTRACT

This paper describes a new computer intensive statistical technique for comparing unbalanced experimental designs which will be modeled by the univariate analysis of covariance. We propose treating the imbalance in the design by minimizing the Inflation of the Standard Error of a contrast (MISER).

We provide results for both the standard Gauss-Markov model (constant error variance) and the model with heteroscedasticity. We also discuss the problem of attributing the increased variance caused by imbalance in a design to particular covariates.

The effect of implementing the proposed MISER criterion is to generate a design which has great sensitivity to treatment effect differences. The MISER criterion is applied to the Department of Defense’s Enlistment Bonus Test, involving offering cash incentives to induce high quality young men to enlist in the U.S. Army.

* On sabbatical leave from the University of California, Riverside during 1984-1985.
1. Introduction

This research* concerns the problem of pre-experiment balancing in experimental design. In experiments in which there are relevant covariates which cannot be controlled by the experimenter, observed effects might be incorrectly attributed to treatments, when in fact they are generated by there being different average values of the covariates in each of the cells. To correct for concomitant variation we conventionally analyze the experiment by using analysis of covariance (ANOCOVA).

Ideally, we would understand enough about the underlying process to select covariates that will assign most of the variability in the dependent variable to specific causes, so that the residuals in the model are only randomly associated with the treatments (see Rubin, 1974). Usually, we do not have such complete understanding, so we try to compensate for the resulting model misspecification error through a pre-experiment scheme called "balancing," discussed below.

In designing an experiment that we plan to analyze using ANOCOVA we usually try to assign subjects to treatments so that there is balance across cells (there is a sample mean vector of covariates for each of the cells in the experiment, and we try to effect an assignment which will equalize the mean vectors). The reason for this design objective is that while estimators of the unknown parameters (and contrasts) remain unbiased regardless of imbalance (lack of equality of the mean vectors of covariates in each of the cells), so long as the assumptions of the classical "Gauss-Markov model" are satisfied (cf. Haggstrom, 1975, p.5449, p. 37), the variances of all contrasts are inflated when there is imbalance. As a result, a greatly unbalanced design can make it very difficult to detect meaningful differences across test cells (since the difference in effects observed in the experiment might be attributed to sampling variation instead of to differences caused by treatment effects). In experiments in which the covariates may be readily controlled by the experimenter, such as those that typically take place in a physical science context, it is often not too difficult to achieve the objective of a

---

* The author benefited from conversations with Robert M. Bell, Richard L. Fernandez, Jeffrey Garfinkle, Subir Ghosh, Gus W. Haggstrom, and J. Michael Polich. We are also indebted to Robert M. Bell for his excellent suggestions for improving the presentation of the paper. The Rand Corporation sponsored the research. We are also grateful to the National Science Foundation for its additional support.
balanced design. In biological or social experimentation, however it is usually the case that some of the covariates are not controllable very easily, so that in such experiments it is rarely possible to be able to achieve perfect balance.

The problem of balance in such experiments was also studied by Morris, 1973. His context was one of a social experiment designed at The Rand Corporation dealing with health insurance. His procedure, called the Finite Selection Model (FSM), is designed to select an optimal subset of subjects from a finite population available for potential experimentation, on the basis of the values of covariates each subject is likely to have during the experiment. These subjects are selected one at a time by his model using a non-linear integer programming procedure involving a steepest descents algorithm. The criterion for selection of subjects involves minimizing a weighted sum of variances of linear functions of the estimated parameters of interest, subject to constraints. This minimization is carried out using pre-experimental data comparable to the anticipated experimental data. The FSM approach follows the approach used earlier by Conlisk and Watts, 1969, although the latter procedure did not employ pre-experimental data.

The solution proposed here is appropriate for quite a different context. We are proposing a procedure for minimizing imbalance in experimental design, in a context in which we intend to use all subjects potentially available for experimentation, regardless of their anticipated covariate values. Our problem is the one of how to allocate all available subjects to treatments, and of how to set tolerances on the differences of cell means of covariates, so that the design is almost balanced. The MISER criterion function we use is the one of minimum average percent inflation of the standard error of a contrast in the parameters of interest (averaged over all of the contrasts of interest). While both the FSM and the MISER procedures for balancing involve both first and second moments, the MISER procedure compares the various possible configurations of subject-to-treatment allocations of the entire population, while the FSM procedure adds subjects to the design, one by one. So the implementing algorithms are quite different (aside from the fact that one criterion uses inflated standard deviation, while the other uses total variance). The MISER procedure has the advantage over other procedures of providing a “natural scale” for the effects of imbalance. The imbalance scale we use measures the percentage increase of the standard errors of the quantities of interest over
what the standard errors would be in a perfectly balanced situation.

In an experiment where we are attempting to detect an effect attributable to some treatment, we often want to minimize the standard errors of the contrasts of interest (in order to maximize the power of tests of significance for the contrasts). Such a problem involves both selection of appropriate covariates, and also, balancing across cells. Given a set of covariates, the balancing problem reduces to minimizing the percent inflation of such standard errors.

In the sequel we develop statistical models for evaluating the effects of imbalance on the standard errors of the contrasts across the cells. The MISER criterion we propose is computer intensive, and can be used for comparing potential unbalanced designs on the basis of how much a given design inflates the standard errors of the contrasts (and we propose selecting that design which effects the best compromise between close balance on some covariates, and not such close balance on others, so that overall, inflation of the standard error is minimized for the average of all contrasts of interest). We adopt some of the notation used in Haggstrom, 1975. Our main result is given in Theorem (1).

Section 2 develops the theory underlying the MISER criterion for the classical "Gauss-Markov model." It also summarizes the method proposed for implementing MISER. Section 3 provides a result for the more general heteroscedastic model. Section 4 discusses the problems associated with attempting to attribute imbalance to particular covariates. Section 5 presents an application of the MISER criterion to a Department of Defense experiment involving U.S. Army recruiting (the Enlistment Bonus Test). It will be reported that the optimal design found (and actually used in that experiment) had a MISER of just 3%.

2. The MISER Criterion for Classical ANOCOVA

2.1 Theory

We adopt the classical ANOCOVA, one way layout, fixed effects model (for given $z_{ij}$, and $z_i$),

$$ y_i = \sum_{j=1}^{p} \beta_j x_{ij} + \gamma' z_i + e_i, \quad (2.1) $$
\( i = 1, \ldots, n, \) where \( y_i \) denotes the response of subject \( i, \) there are \( p \) groups or cells with \( n_1, n_2, \ldots, n_p \) numbers of subjects assigned to each of the \( p \) cells, \( n = \Sigma_j n_j \) denotes the total number of subjects, \( z_{ij} \) is one, or zero, depending upon whether or not the \( i \)th subject is assigned to the \( j \)th group, \( z_i \) denotes an \((h \times 1)\) vector of covariates for subject \( i, \) \((\beta_j, \gamma)\) are unknown coefficients that must be estimated from the data, and \( \epsilon_i \) denotes mutually uncorrelated disturbance terms with means zero and variances not depending upon \( i. \) Prime denotes the transposed vector.

An alternative formulation involves writing

\[
\beta_j = \tau_j + \alpha, \tag{2.2}
\]

where \( \alpha \) denotes the ambient effect, or the effect in the absence of any treatments, and \( \tau_j \) denotes the effect of treatment \( j \) on a subject in cell \( j. \) In this formulation,

\[
y_i = \sum_{j=1}^{p} (\tau_j + \alpha) x_{ij} + \gamma' z_i + \epsilon_i \\
= \alpha + \sum_{j=1}^{p} \tau_j x_{ij} + \gamma' z_i + \epsilon_i. \tag{2.3}
\]

We now reformulate (2.1) in vector terms, to simplify the algebra. Let \( \beta = (\beta_j) : (p \times 1), \ x_i = (x_{ij}) : p \times 1, \ \delta = (\delta', \gamma')' : (p + h) \times 1, \ w_i = (x_i', z_i')' : (p + h) \times 1. \) Now, (2.1) becomes

\[
y_i = \delta' w_i + \epsilon_i, \tag{2.4}
\]

\( i = 1, \ldots, n. \) If we let \( c = (c_i) : (n \times 1), \) and write the model as

\[
E = c'e = \sum_{i=1}^{n} (y_i - \delta' w_i)^2, \quad E(c) = 0, \quad \text{var} (c) = \sigma^2 I,
\]

it is readily seen that \( E \) is minimized by taking

\[
\delta = A^{-1} g, \tag{2.5}
\]

where, for \( A_{11} : (p \times p), \ A_{22} : (h \times h), \)

\[
A \equiv \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} \Sigma x_i x_i' & \Sigma z_i z_i' \\ \Sigma z_i x_i' & \Sigma z_i z_i' \end{pmatrix}, \tag{2.6}
\]

5
\[ g \equiv \sum_{i=1}^{n} y_i w_i \equiv \left( \sum_i x_i y_i \right)^m \equiv \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}, \quad (2.7) \]

and \( \hat{\delta} \) is the least squares estimator of \( \delta \). We must, of course, have \( A > 0 \); i.e., \( A \) is positive definite (and symmetric). It is also easy to find (from (2.5)) that since

\[ \text{var} (\hat{g}) = \sigma^2 A, \]

\[ \text{var} (\hat{\delta}) = \sigma^2 A^{-1}. \quad (2.8) \]

Define:

\[
A^{-1} = B \equiv \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}.
\]

Then,

\[ \text{var} (\hat{\delta}) = \text{var} \left( \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} \right) = \sigma^2 B. \]

Thus, using, for example, Press, 1982, equations (2.6.3) and (2.6.4),

\[ \text{var} (\hat{\beta}) = \sigma^2 B_{11} = \sigma^2 (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1} \equiv \Sigma_\beta, \quad (2.9) \]

and

\[ \text{var} (\hat{\gamma}) = \sigma^2 B_{22} = \sigma^2 (A_{22} - A_{21} A_{11}^{-1} A_{12})^{-1} \equiv \Sigma_\gamma. \quad (2.10) \]

We also note in passing that \( \text{cov} (\hat{\beta}, \hat{\gamma}) = \sigma^2 B_{12} \).

Let \( \psi \) denote a “contrast” in the effects of the experiment; i.e., \( \psi = c' \beta \), where \( c \equiv (c_j), \sum_{j=1}^{p} c_j = 0 \). For example, for \( p = 3 \), if we take \( c' = (-1, 1, 0), \psi = \beta_2 - \beta_1 \). If \( \hat{\beta} \) denotes the least squares estimator of \( \beta \), \( \hat{\psi} = c' \hat{\beta} \) denotes the least squares estimator of \( \psi \). Thus,

\[ \Delta \equiv \text{var} (\hat{\psi}) = c' [\text{var} (\hat{\beta})] c = c' \Sigma_\beta c. \quad (2.11) \]

We next evaluate \( \Sigma_\beta \).

From (2.5) and (2.7) \( A \hat{\delta} = g \) implies \( A_{11} \hat{\beta} + A_{12} \hat{\gamma} = g_1 \), or

\[ \hat{\beta} = A_{11}^{-1} g_1 - A_{11}^{-1} A_{12} \hat{\gamma}. \quad (1.12) \]

It is well known that the error and estimation spaces in the general linear model (under the Gauss-Markov assumptions that we have made) are orthogonal (see, e.g., Scheffé, 1959, pg. 12). So \( \hat{\gamma} \) and \( g_1 \) (depending only on the \( y_i \)'s, or the errors, for given \( w_i \)'s) are uncorrelated.
Therefore, from (2.12),
\[ \Sigma_\beta = \text{var} (A_{11}^{-1} g_1) + \text{var} (A_{11}^{-1} A_{12} \gamma), \]
or, from (2.11),
\[ \Delta = c'A_{11}^{-1} (\text{var} g_1) A_{11}^{-1} c + c'A_{11}^{-1} A_{12} (\text{var} \gamma) A_{21} A_{11}^{-1} c. \]  
(2.13)

From (2.6) and (2.7), note that
\[ \text{var} (g_1) = \sum_i x_i \text{var} (y_i) x'_i = \sigma^2 \sum_i x_i x'_i = \sigma^2 A_{11}. \]  
(2.14)

Substituting (2.14) into (2.13) gives
\[ \Delta = \sigma^2 c'A_{11}^{-1} c + c'A_{11}^{-1} A_{12} (\text{var} \gamma) A_{21} A_{11}^{-1} c. \]  
(2.15)

Now note that since \( x_{ij} \) denotes an indicator variable, \( x^2_{ij} = x_{ij} \), so that (using (2.6)),
\[ A_{11} = \begin{pmatrix} n_1 & 0 \\ \vdots & \ddots \\ 0 & n_p \end{pmatrix}. \]  
(2.16)

Thus,
\[ c'A_{11}^{-1} c = \sum_{j=1}^{p} \frac{c_j^2}{n_j}. \]  
(2.17)

Substituting (2.17) into (2.15) gives
\[ \Delta = \sigma^2 \sum_{j=1}^{p} \frac{c_j^2}{n_j} + \epsilon, \]  
(2.18)

where:
\[ \epsilon \equiv c'A_{11}^{-1} A_{12} \Sigma \gamma A_{21} A_{11}^{-1} c. \]  
(2.19)

A similar formulation was given by Harville (1975), page 220. The value of \( \epsilon \) is given in the theorem below. First define
\[ Z \equiv (\bar{z}_1, ..., \bar{z}_p)' : (p + h) \times 1, \]
\[ \Phi \equiv \sum_{i=1}^{n} z_i z'_i - \sum_{j=1}^{p} n_j \bar{z}_j \bar{z}'_j, \quad \Phi : (h \times h) \]  
(2.20)

\[ \Phi = \sum_{j=1}^{p} \sum_{i=1}^{n_j} (z_{ij} - \bar{z}_j)(z_{ij} - \bar{z}_j)', \]  
(2.21)
and
\[ z_j \equiv (z_{j1}, ..., z_{jh})' : (h \times 1), \quad z_{jk} \equiv \frac{1}{n_j} \sum_{i=1}^{n_j} z_{ik} z_{ij}. \] (2.22)

**Theorem (1):** If \( \epsilon \) is the portion of the variance of \( \hat{\psi} \) defined in (2.19), \( \epsilon \) is given by the quadratic form
\[ \epsilon = \sigma^2 c' W c \equiv \sigma^2 c'(Z \Phi^{-1} Z') c. \] (2.23)

**Proof:** See Appendix.

**Remark:** Note that \( \Phi \) is the matrix of sums of squares and cross products within cells. Therefore, \( \epsilon \) is a measure of imbalance of the design \( Z \) as measured in its natural metric. \( \epsilon \) is also the inflated portion of \( \text{var} (\hat{\psi}) \).

Referring back to the variance of any contrast, given in (2.18), note that the first term of that expression is strictly positive. Therefore, \( \Delta \) is minimized when \( \epsilon = 0 \) (since, by definition in (2.11), \( \Delta \) must be positive). This case is made specific in the corollary below.

**Corollary (1):** If the sample means of covariate vectors are equal, i.e.,
\[ \bar{z}_1 = \bar{z}_2 = \cdots = \bar{z}_p = \bar{z}, \]
\( \epsilon \), defined in (2.19) vanishes.

**Remark (1):** This result (cf. Haggstrom, 1975, P-5449, p. 37) occurs when the design is balanced. Thus, when the design is balanced, \( W \), the matrix of the positive semidefinite quadratic form in (2.23), is not of full rank.

**Remark (2):** Note that the "balanced design" condition of Corollary (1) is only a sufficient condition for \( \epsilon = 0 \). There are other conditions that would also make \( \epsilon = 0 \) (we merely need \( Z' \epsilon = \sum_1^p c_j \bar{z}_j = 0 \); so for example, with \( p = 3 \), and \( c = (-1, 1, 0)' \), \( \bar{z}_1 = \bar{z}_2 \), and \( \bar{z}_3 = \) anything, would also yield \( \epsilon = 0 \). If, however, we wanted a balanced design regardless of which contrast was of interest, i.e., we wanted \( \epsilon = 0 \) for every possible contrast, we would need \( Z = (\bar{z}, ..., \bar{z})' \).
Proof of Corollary (1): Under the hypothesis of the corollary,

$$Z = (z, ..., z)'$$.

Substituting into (2.23) gives (since $\psi$ is a contrast),

$$Z'c = \bar{z} \sum_{i=1}^{p} c_i = 0. \quad (2.24)$$

So $\varepsilon = 0$, and $\Delta$ is minimized. This result becomes transparent by noting that $c'(Z\Phi^{-1}Z')c$ in (2.23) does not change if we subtract a constant vector $\bar{z}$ from each column of $Z'$, and replace it by $Z'_0 \equiv (\bar{z}_1 - \bar{z}, ..., \bar{z}_p - \bar{z})$.

**INFLATION OF STANDARD ERRORS**

Suppose we wish to design an experiment in which we know we will have to tolerate some degree of imbalance; how do we decide among various unbalanced designs? We propose below a natural criterion for making such a decision, in light of Theorem (1).

Define $\Delta_B$ as the value of $\Delta$ (defined in (2.11)) produced in the case of a balanced design. That is

$$\Delta_B = \sigma^2 \sum_{j=1}^{p} \left( \frac{c_j^2}{n_j} \right). \quad (2.25)$$

Let $\omega$ denote the percent inflation of the standard error of any contrast $\psi = c'\beta$, for any covariate design $Z = (z_1, ..., z_p)'$. Thus,

$$\omega = \left[ \frac{\sqrt{\Delta} - \sqrt{\Delta_B}}{\sqrt{\Delta_B}} \right] (100). \quad (2.26)$$

It should be noted that $\omega$ in equation (2.26) does not depend upon $\sigma^2$.

Suppose we are trying to compare various designs which have the same contrast of interest, but we are free to vary $Z$ so that subjects can be allocated to different treatments in various ways. If we can predetermine $Z$ (at least approximately), based upon external data sources, for each contending design, we can also evaluate $\omega$ for each such design. We might then select that design possessing the minimum value of $\omega$. Alternatively, by careful study of $\omega$ for various competing designs, we might decide that although one design had a higher $\omega$ than another,
the difference was small enough to be tolerable, and the one with the slightly higher \( \omega \) should be selected in light of other biological, economic, political, or social factors; but by examining \( \omega \) we can evaluate the "cost" of such a tradeoff, in terms of effectiveness of the experiment (in terms of loss of probability of finding the effect for which we are designing).

Suppose, alternatively, that there are \( r \) contrasts of interest, and we are free to vary \( Z \) so that subjects can be allocated to different treatments in various ways. We assume that some contrasts may be more important than others. Let \( q_i \) denote the weight to be placed upon the \( i \)th contrast, \( 0 < q_i < 1 \). We might now select a design that possesses the minimum (weighted) average value of \( \omega \); this is the MISER, or minimum average value of the percent inflation of the standard errors of the contrasts, averaged over all contrasts of interest.

\[
\bar{\omega} = \sum_{i=1}^{r} \left( \frac{\sqrt{\Delta_i} - \sqrt{\Delta_B}}{\sqrt{\Delta_B}} \right)(100q_i), \tag{2.27}
\]

where \( \Delta_i \equiv \text{var} (\hat{\psi}_i) \equiv \text{var} (c_i'\hat{\beta}) \), and \( c_i : (p \times 1) \) denotes the weight vector for the \( i \)th contrast. This type of averaging is in the spirit of Cox (1982), page 197.

For example, suppose there are three cells (two test cells) and a control cell, and there are two simple contrasts of interest, namely

\[
\psi_1 = \beta_2 - \beta_1, \quad \psi_2 = \beta_3 - \beta_1.
\]

Then, \( c_1 = (-1, 1, 0)' \), and \( c_2 = (-1, 0, 1)' \). Suppose further that \( \psi_1 \) and \( \psi_2 \) are of equal interest and importance, so that \( q_1 = q_2 = \frac{1}{2} \). Then, from (2.27), for \( r = 2 \),

\[
\bar{\omega} = \frac{(\sqrt{\Delta_1} + \sqrt{\Delta_2} - 2\sqrt{\Delta_B})(100)}{2\sqrt{\Delta_B}}.
\]

2.2 Implementation of MISER

The MISER criterion depends strongly for its implementation upon the use of the computer. We describe below a step by step procedure for applying the criterion in practice.
1. Secure covariate data

It is basic to determine values of the covariate vectors, \( z_1, \ldots, z_n \), all \( (h \times 1) \), for all \( n \) subjects, which are the values likely to obtain at the time the experiment is scheduled to be carried out. This data can be found from earlier, related, studies, historical information, or it can be developed by forecasting methods. In any case, it is fundamental prior information without which we cannot proceed with the balancing of the design of the experiment.

2. Generate a feasible design

A design is a specification of which of the subjects will receive which treatments. A "feasible" design is a design which satisfies the constraints of the problem. In our context we would like to constrain the design to be one for which the means of the covariate vectors in each of the cells are the same. We know we cannot accomplish this in general, however. Alternatively, we introduce "tolerance constraints". A priori, we assign a set of \( h \) numbers (tolerances) \( \mu_k, k = 1, \ldots, h \), and require that differences between all pairs of means of the values of the covariates for all subjects in each of the \( p \) cells be less than or equal to \( \mu_k \), for all \( k \).

For example, suppose there are three cells \( (p = 3) \). Suppose, moreover, a random design allocation yields covariate vector means in each cell, \( \bar{z}_1, \bar{z}_2, \bar{z}_3 \). Let \( z_j = (\bar{z}_{j1}, \ldots, \bar{z}_{jh})' \). The tolerance constraints then require that

\[
| \bar{z}_{1k} - \bar{z}_{2k} | \leq \mu_k, \quad | \bar{z}_{2k} - \bar{z}_{3k} | \leq \mu_k,
\]

\[
| \bar{z}_{1k} - \bar{z}_{3k} | \leq \mu_k, \quad k = 1, \ldots, h,
\]

for preassigned \( \mu_k \).

A design is generated by the computer by assigning all subjects random numbers and allocating treatments randomly, by sampling without replacement. If the tolerances are all permitted to be large, it is likely that the computer will find a feasible design very rapidly (a design which also satisfies the tolerance constraints). As the tolerances are tightened, the computer will take longer to generate a feasible design by cycling through various possible designs. Budgetary considerations may dictate imposing an upper bound on the computer
search time, or the number of iterations permitted. Tolerances that are too demanding may result in no feasible design, in which case, some of the tolerances might have to be relaxed.

Some experiments may require that other types of constraints be imposed as well. For example, external considerations may demand that the ith cell should contain a percentage \( n_i \) of the total number of the available subjects, \( i = 1, ..., p \). In such cases, all constraints could easily be designed into the software created to generate feasible designs.

3. Evaluate covariate moments

For the specific feasible design generated in Step 2, the first two moments should be computed, i.e., compute \( (Z, \Phi) \), defined in eqns. (2.20)-(2.22).

4. Select contrasts

Determine which \( r \) contrasts are of relevance for the experiment, and fix \( c : (p \times 1) \) for each relevant contrast. Also fix the importance weights for the \( i \)th contrast, \( q_i, 0 < q_i < 1, i = 1, ..., r \). They will often be taken as \( q_i = 1/r \).

5. Evaluate \( \omega \)

For each relevant contrast evaluate

(a) \( \frac{\Delta}{\Delta B} = 1 + \frac{c'(Z \Phi^{-1}Z) c}{\sum_{j=1}^p (c_j^2 / n_j)} \)

(b) \( \omega = 100\left(\left(\frac{\Delta}{\Delta B}\right)^{\frac{1}{2}} - 1\right) \)

(c) \( \bar{\omega} = \text{the average } \omega \text{ over all relevant contrasts.} \)

6. Iterate the process

Now repeat Steps 2-5 by finding a new feasible design and evaluating the corresponding \( \omega \). Iterating the procedure many times will yield many \( \bar{\omega} \)'s. The Miser is the minimum \( \bar{\omega} \) obtainable, subject to the budgetary and time constraints on the problem, and the design corresponding to the Miser is the optimal one for pre-experiment balancing.
3. The MISER Criterion for Weighted ANOCOVA

The model treated in Section 2 adopted the Gauss-Markov assumptions \( E(e) = 0, \quad E(e_i e_j) = 0, \ i \neq j, \) and \( \text{var} \ (e_i) = \sigma^2 : i = 1, \ldots, n. \) But suppose alternatively, that

\[
\text{var} \ (e_i) = a_i \sigma^2, \quad (3.1)
\]

while the other assumptions remain the same. That is, we have heteroscedasticity. Suppose the \( a_i \)'s are known constants, however. We see below that the results of Section 2 are readily extended to cover this case as well.

Adopt the model of (2.1), but with the heteroscedasticity assumption of (3.1). Define the transformed variables:

\[
y_i^* = y_i (a_i)^{-\frac{1}{2}}, \quad e_i^* = e_i (a_i)^{-\frac{1}{2}}, \quad z_i^* = z_i (a_i)^{-\frac{1}{2}}, \quad (3.2)
\]

\[
x_{ij}^* = x_{ij} (a_i)^{-\frac{1}{2}}.
\]

The transformed model becomes (for given \( x_{ij}^* \), and \( z_i^* \)),

\[
y_i^* = \sum_{j=1}^{p} \beta_j x_{ij}^* + \gamma' z_i^* + e_i^*, \quad (3.3)
\]

for \( i = 1, \ldots, n \), with

\[
E(e_i^*) = 0, \quad E(e_i^* e_j^*) = 0, \quad i \neq j, \quad \text{var} \ (e_i^*) = \sigma^2. \quad (3.4)
\]

The model now has the general form assumed in Section 2, in terms of transformed variables. So the corresponding results are immediately obtainable. The basic result in Theorem (1) becomes (instead of (2.23)):

\[
\epsilon^* = \sigma^2 e' \Phi^{-1} \epsilon = \sigma^2 e' (Z^* \Phi^{-1} Z^{**}) \epsilon, \quad (3.5)
\]

where:

\[
Z^* \equiv (z_1^*, \ldots, z_p^*)', \quad (3.6)
\]

\[
x_j^* = (z_{j1}^*, \ldots, z_{jk}^*)' : (k \times 1), \quad z_{jk}^* \equiv \frac{1}{n_j^*} \equiv \sum_{i=1}^{n} z_{ik}^* x_{ij}^*, \quad (3.7)
\]

\[
n_j^* = \sum_{i=1}^{n} (z_{ij}^2) = \sum_{i=1}^{n} (\frac{x_{ij}^2}{a_i}) = \sum_{i=1}^{n} (\frac{x_{ij}}{a_i}), \quad (3.8)
\]
and

$$
\Phi^* = \sum_{i=1}^{n} z_i^2 z_i - \sum_{j=1}^{p} n_j^2 z_j^2 z_j.'
$$

(3.9)

Note that (3.7)-(3.9) are weighted averages, weighted sums, or weighted sums of squares, as contrasted with their unweighted analogues in Section 2.

The analogue of (2.18) becomes

$$
\Delta = \sigma^2 \sum_{j=1}^{p} \left( \frac{e_j^2}{n_j} \right) + \epsilon^*,
$$

(3.10)

with $\epsilon^*$ defined in (3.5), and $\Delta = \text{var}(\hat{\psi}) = \text{var}(\epsilon_i \hat{\beta})$. So if

$$
\Delta_B^* = \sigma^2 \sum_{j=1}^{p} \left( \frac{e_j^2}{n_j} \right)
$$

(3.11)

denotes the variance of a contrast for a balanced design (one with $\epsilon^* = 0$), the selection criterion becomes, for the case of heteroscedasticity,

$$
\omega = \frac{\sqrt{\Delta} - \sqrt{\Delta_B^*}}{\sqrt{\Delta_B^*}}
$$

(100),

(3.12)

where $\Delta$ and $\Delta_B^*$ are defined in (3.10) and (3.11). Note again, as in (2.26), $\omega$ does not depend upon $\sigma^2$.

To apply the Miser criterion in the case of heteroscedasticity we first transform the covariate vectors to form $Z^*$ and $\Phi^*$. Then, for a given contrast of interest (choice of $c$), we can evaluate $\omega$ for various competing designs.

$\sigma^2$ is usually not known, but in some situations we may wish to estimate it. For example, suppose our original dependent variables are $Y_i$, and the $Y_i$'s are independently distributed Poisson variates with means $\lambda_i$. If we transform the model so that $y_i = \log(Y_i + 1/2)$ where $y_i$ is given by (2.1), it may be shown (see e.g. Cox, 1955) that approximately,

$$
\text{var}(y_i) = \text{var}[\log(Y_i + 1/2)] = \frac{1}{\lambda_i}.
$$

(3.13)

But in the Poisson distribution, we can use the approximation,

$$
\lambda_i = E(Y_i) \approx Y_i.
$$
Substituting in (3.13) gives:

\[ \text{var} (y_i) \cong \frac{1}{Y_i}. \]

But from (3.1) we can take

\[ \text{var} (y_i) = \text{var} (e_i) \equiv a_i \sigma^2 \cong \frac{1}{Y_i}. \]

So we can take \( \sigma^2 \equiv 1 \) in this instance, but take

\[ a_i = \frac{1}{Y_i}. \] (3.14)

Thus, we adopt the model in (3.3), with \( a_i \) as in (3.14) and \( \sigma^2 = 1 \). We approximate \( Y_i \), prior to an experiment, by using previously obtained data comparable to \( Y_i \). Implementation of the solution is carried out by modifying the procedure of Section 2.2 in an obvious way.

4. Attributing Inflated Standard Errors to a Given Source

Once the degree of inflation of the standard errors of contrasts caused by an unbalanced design has been determined, it is tempting to try to separate the sources of the inflation by attributing these sources to the distinct covariates. It turns out this is very difficult to do, except in unusual circumstances. For simplicity, we examine the nature of this problem below in terms of variances instead of standard deviations, and exhibit the behavior of the inflation explicitly for several sharply defined situations relating to a single contrast. Other cases are more complicated.

The proportion of a contrast variance attributable to imbalance of the design is given in Theorem (1), (2.23), for the classical Gauss-Markov model as

\[ \epsilon = \sigma^2 c' (Z \Phi^{-1} Z') \epsilon. \] (4.1)

That is, \( \epsilon \) denotes the inflation of the variance.

Now suppose that all of the covariates are mutually uncorrelated, so that if \( \sigma^2_n \) denotes
the (sample) variance of the $\alpha$th covariate,

$$
\Phi = n \begin{pmatrix} 
\sigma_1^2 & 0 \\
& \ddots & \ddots \\
0 & \sigma_h^2 
\end{pmatrix}
$$

Then,

$$
\epsilon = \sigma^2 \sum_{\alpha=1}^{h} \sum_{i=1}^{p} \sum_{j=1}^{p} c_i c_j \left( \frac{\bar{z}_{i\alpha} \bar{z}_{j\alpha}}{n\sigma_\alpha^2} \right) = \sum_{\alpha=1}^{h} f_\alpha,
$$

where

$$
f_\alpha = \left( \frac{\sigma^2}{\sigma_\alpha^2} \right) \sum_{i=1}^{p} \sum_{j=1}^{p} c_i c_j \left( \frac{\bar{z}_{i\alpha} \bar{z}_{j\alpha}}{n} \right).
$$

(4.2)

Define

$$
P_\alpha = \frac{f_\alpha}{\epsilon}.
$$

(4.3)

Note that $P_\alpha$ is the proportion of the inflated variance attributable to the $\alpha$th covariate.

Consider a particular (simple) contrast,

$$
c = (-1, 1, 0, \ldots, 0)' : p \times 1.
$$

In this case, from (4.2),

$$
f_\alpha = \left( \frac{\bar{z}_{1\alpha} - \bar{z}_{2\alpha}}{\sigma_\alpha} \right)^2 \left( \frac{\sigma^2}{n} \right).
$$

Let $\theta_\alpha \equiv (\bar{z}_{1\alpha} - \bar{z}_{2\alpha})$ denote the balancing "tolerance" for covariate $\alpha$. That is, when we attempt to balance the design, we attempt to make the sample means for all covariates the same in all of the cells; $\theta_\alpha$ is the amount by which the cell means differ for cells 1 and 2, for covariate $\alpha$. Thus,

$$
f_\alpha = \frac{\theta_\alpha^2}{n\sigma_\alpha^2},
$$

and, from (4.3),

$$
P_\alpha = \frac{(\theta_\alpha^2/\sigma_\alpha^2)}{\sum_{\alpha=1}^{h}(\theta_\alpha^2/\sigma_\alpha^2)}.
$$

(4.4)

Note that for a particular $\alpha$, say $\alpha = 1$, we may express (4.4) in the form

$$
P_1 = \frac{\theta_1^2/\sigma_1^2}{\theta_1^2/\sigma_1^2 + K} = \frac{1}{1 + K\sigma_1^2/\theta_1^2},
$$

(4.5)
where \( K = \sum_{\alpha=2}^{h} (\theta^2_{\alpha}/\sigma^2_{\alpha}) \). Figure 1 shows how the proportion of inflated variance attributable to the first covariate varies with balancing tolerance and covariate standard deviation. We see that \( P_1 \) increases as the square of balancing tolerance, so that the further apart are the covariate cell means, the greater is the variance inflation (increasing by the square of the tolerance). Moreover, \( P_1 \) decreases with increasing covariate standard deviation, so \( P_1 \) is greatest when the first covariate is not free to vary very much. Thus, referring to equation (4.2), when the covariates are uncorrelated, we can minimize the inflated portion of the total variance by choosing covariates which achieve acceptably low values of \( \sigma^2/\sigma^2_{\alpha} \). Accordingly, if there were a choice between two covariates, each of which reduced \( \sigma^2 \) by the same amount, but one had a larger \( \sigma^2_{\alpha} \), we should select the one with the larger \( \sigma^2_{\alpha} \).

We see that because in most ANOCOVA designs the covariates are correlated, we cannot break out the effects due to each covariate separately. The greater the correlations among covariates the more difficult it will be to analyze the individual covariate effects separately, as in Figure 1. In the general case, therefore, we must be content to evaluate \( \omega \) for the design as a whole, and then to make \( \hat{\omega} \) as small as possible.
Figure 1

Balancing Tolerance

Covariate Standard Deviation

Proportion of Inflated Variance Attributable to Balancing Tolerance and Covariate Standard Deviation
5. An Empirical Application of MISER

Since 1973, the U.S. Army has been an all volunteer force. The Army has therefore been faced with the problems of: how to induce eligible people to volunteer for service, how to ensure that an appropriate fraction of those who volunteer will be of "high quality," how to ensure that the volunteers will be agreeable to being assigned to the specialty areas of interest to the Army at the current time, and how to get the volunteers to be agreeable to serving for a reasonable length of time. To determine the effect of a cash incentive system for enlisting, the Department of Defense (DOD) and the Army carried out an experiment during July, 1982-June, 1984 called the Enlistment Bonus Test; MISER was used for pre-experiment balancing. The experiment was mandated by Congress in 1981; it is expected to guide DOD and the Congress in formulating policy on the best configuration of recruiting incentives to attract qualified young people to serve in the armed forces. The experiment was designed at The Rand Corporation, where its results are currently being analyzed.

The experiment involved continuing to offer the same enlistment bonus plan during the experiment as it did prior to the experiment to 70% of the eligible pool of young men, but in addition, offering one of two expanded bonus programs authorized by Congress at the time the test was initiated to each half of the remaining 30%: (1) an increased bonus for a four-year enlistment, raised to $8,000 from the previous maximum of $5,000; and (2) a new $4,000 bonus for a three-year enlistment. The new bonuses were restricted to "high quality" recruits (high school graduates with entrance test scores in the 50th percentile or higher), and to recruits who enlisted in selected Army critical occupational specialties ("skills"). Thus, there were three bonus plans (treatments) in the experiment of which one was a control and two were test bonuses (see Fig. 2). There were three experimental cells corresponding to the treatments. In the test design, local areas were randomly assigned to cells, subject to pre-experiment balancing constraints, which ensured that the cells were well matched on a wide range of factors likely to influence the test outcome. These factors (covariates) are the ten enumerated in Table 1. Variable 11 listed there involved an additional constraint on the design to achieve the (15%, 15%, 70%) figures alluded to above.

There were 66 subjects in the experiment. These were the 66 regions into which the
continental U.S. has been subdivided for purposes of military recruiting. They are called MEPS (Military Enlistment Processing Stations). The design problem was the one of how to allocate the three treatments (bonus plans) to the 66 subjects (MEPS) in such a way that: (1) the percentage constraints of (15%, 15%, 70%) for the two test cells and the control cell were satisfied and (2) the MEPS were geographically located in such a way as to make the mean covariate vectors in the three cells as close to one another as possible. There were two contrasts of particular interest, the main effect in each of the two test cells as compared with the main effect in the control cell.
<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ENLISTMENT RATE</td>
<td>Number of Army high quality enlistments in Oct-Feb FY 1982, as percent of the number of high quality males 17-21 who are “qualified military available” (QMA)</td>
</tr>
<tr>
<td>2. UNEMPLOYMENT RATE</td>
<td>Unemployment rate of workers 16 and older, Jan-Dec, 1982 in percent</td>
</tr>
<tr>
<td>3. WAGE RATE</td>
<td>Ratio of civilian to military wage rate (hourly wage rate for manufacturing production workers divided by hourly rate of basic military pay, Jan-Dec 1981)</td>
</tr>
<tr>
<td>4. HIGH QUALITY CONCENTRATION</td>
<td>Number of high quality QMA males 17-21, as percent of total male population 17-21.</td>
</tr>
<tr>
<td>5. PERCENT NONWHITE</td>
<td>Number of nonwhite males 17-21, as percent of total male population 17-21.</td>
</tr>
<tr>
<td>6. RECRUITER DENSITY</td>
<td>Number of Army production recruiters, Jan-Jun 1981, per thousand QMA males 17-21</td>
</tr>
<tr>
<td>7. ARMY MISSION</td>
<td>Number of Army high quality nonprior service males to be recruited in Oct-Mar FY 1982, as percent of number of high quality QMA males 17-21 in the population</td>
</tr>
<tr>
<td>8. EASTERN REGION</td>
<td>Indicator variable for Census Eastern region</td>
</tr>
<tr>
<td>9. SOUTHERN REGION</td>
<td>Indicator variable for Census Southern region</td>
</tr>
<tr>
<td>10. WESTERN REGION</td>
<td>Indicator variable for Census Western region</td>
</tr>
<tr>
<td>11. SIZE OF AREA*</td>
<td>Number of high quality QMA males 17-21 in the area, as a percentage of the total number in the nation</td>
</tr>
</tbody>
</table>

*This variable is not balanced across cells; rather, the cells are selected so that the total size of each cell approximates percentages desired.
Values of the covariate variables in Table 1 likely to prevail during the experiment were taken to be the corresponding values during the 12 months just prior to the experiment. A computer program was written (by Jeffrey Garfinkle, as a modification of one written by Gus Haggstrom) to generate random designs subject to the appropriate constraints. Tolerance constraints on covariate cell mean differences were first relaxed, and then were gradually tightened, to obtain successively improved designs. We examined as many designs as the budget would permit (upper bounds were placed on computer cycling time—if a feasible design couldn't be found within a given time, we relaxed the tolerance).

Using the MISER criterion, average percent inflation of the standard errors of the two contrasts of interest for the various designs considered for the Enlistment Bonus Test ranged from the MISER of about 3% for the best design, to as much as 25% depending upon which design, and which covariate cell mean inequality tolerances were selected. The final design is shown in Figure 2. Analysis of the results of the experiment is in progress at this time.
Figure 2
Enlistment Bonus Test Design

Legend:
A: $5000 Bonus, 4 Yrs
B: $8000 Bonus, 4 Yrs
C: $10000 Bonus, 1 Yr
REFERENCES


APPENDIX

Proof of Theorem (1): From the definition of \( \epsilon \), in (2.19) and that of \( \Sigma_{\gamma} \), in (2.10), we have

\[
\epsilon = \sigma^2 c' A_{12}^{-1} (A_{22} - A_{21} A_{11}^{-1} A_{12})^{-1} A_{21} A_{11}^{-1} \epsilon. \tag{A.1}
\]

From (2.6),

\[
A_{12} = \sum_{i=1}^{n} z_i x_i' = \begin{pmatrix} n_1 \bar{z}_1' \\ \vdots \\ n_p \bar{z}_p' \end{pmatrix} : (p \times h).
\]

Using (2.16) and (2.20), we find the representation,

\[
A_{12} = A_{11} Z; \quad A_{21} = Z' A_{11}. \tag{A.2}
\]

Substituting (A.2) into (A.1) gives

\[
\epsilon = \sigma^2 c' Z (A_{22} - Z' A_{11} Z)^{-1} Z' \epsilon. \tag{A.3}
\]

From (2.6) and (2.21), it is readily seen that

\[
\Phi = A_{22} - Z' A_{11} Z. \tag{A.4}
\]

Substituting (A.4) into (A.3) gives the result in (2.23).