Winter 1980

A Monte Carlo Approach to Ridge Discriminant Analysis

Lee Wooldridge
University of Central Florida

A MONTE CARLO APPROACH TO RIDGE DISCRIMINANT ANALYSIS

BY

LEE WOOLDRIDGE
B.S.E., Florida Technological University, 1973
B.A., Florida Technological University, 1973

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Engineering
Industrial Engineering and Management Systems
in the Graduate Studies Program of the
College of Engineering at the
University of Central Florida; Orlando, Florida

Winter Quarter
1980

ACKNOWLEDGMENTS

This report represents the culmination of over five years work during which many kind and dedicated people contributed invaluable assistance.

The author would like to gratefully acknowledge the helpful guidance, candid criticism and long-suffering tolerance of my graduate advisor, Dr. Christian Bauer, and my thesis committee, Dr. D. G. Linton and Dr. H. I. Klee. June Stillman provided crucial and exacting editorial services and the interpretation of format regulations needed to mold this report into a professional product.

Don Norman, of the Human Factors Laboratory, NAVAL TRAINING EQUIPMENT CENTER, Orlando, Florida, gave the author permission to use the performance measurement development data for further analysis. Robert Coward, of the Flight Training Division, Air Force Human Resources Laboratory, Luke Air Force Base, Phoenix, Arizona, also gave his permission to use the Air Combat Maneuvering data.

A. C. Bittner, Jr., of the Naval Missile Center, Point Mugu, California, cheerfully relinquished several hours of his time to provide timely information, critical to the development of the statistical simulation necessary to further the author's research.

Clay Brice, of Appli-Mation, Inc., contrived insightful and encouraging mathematical reality checks which conforted the author as we tread on unexplored territory.

Ira Goldstein, of the Human Factors Laboratory, NAVAL TRAINING EQUIPMENT CENTER, Orlando, Florida, as mentor, nurtured that little spark of creativity that drove this research.

A special acknowledgment is required for a very distinctive individual and friend, Don Vreuls, who insulated the author from interfering elements by providing a corporate structure and orchestrating the conditions under which this single thread of research could mature. Don is a well from which ideas and enthusiasm endlessly spring.

Of preeminent and direct importance to the ultimate completion of this work, was the unerring and patient organizational and proofreading services, typing and retyping performed by Helen Foard under the most trying of conditions.

Finally, Joyce, who stood behind the author, soothed his nerves, and anchored his life during all his late hours, frustration, fretting, wringing of hands, and gnashing of teeth that is the sheer joy of discovery.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

SECTION I

INTRODUCTION

Systems comprised of one or more human elements many times evade reliable mathematical description. In complex tasks, such as aircraft control, human performance is deeply embedded in many subsystems. Extracting information about crew performance requires direct and indirect measurement of many variables and poses extreme methodological challenges. Therefore, it is no surprise that policy makers, training managers, instructors and scientists make many decisions without the aid of valid human performance information.

The search for human performance measurement necessitated the development of 'ridge' discriminant analysis. This new adaptation of discriminant analysis has been instrumental in the successful implementation of an automated performance measurement system. So far, utilized solely on its face value, more work is needed to adequately describe and validate the properties of this new statistical tool.

In this paper, the author will attempt to highlight the relevant background information that led to the generation of 'ridge' discriminant analysis. A review of discriminant analysis theory and evidence of the merits of the 'ridge' version will also be presented.

Background

The underlying concept of the performance measurement development method, mentioned previously, was to have a combined analytical and empirical technique to define performance measures for automated training of instrument flight maneuvers (IFM) in a simulator. The final measure set was to represent a comprehensive, yet minimum set of measures which were sensitive to the skill changes that occurred during training.

The criteria for measurement selection and the fundamental techniques and algorithms for selecting measures were developed by Vreuls, Obermayer, Goldstein and Lauber (1973) and by Vreuls, Obermayer and Goldstein (1974). After careful search of the literature and due consideration, it was felt by these researchers that a form of multivariate discriminant analysis was the best statistical technique to incorporate into the measurement selection and weighting scheme. The measures, once selected, could be weighted and combined in a simple first order equation upon which an automated training system could track a student's comparative skill level.

Vreuls and Obermayer implemented programs directly from Cooley and Lohnes' book entitled <u>Multivariate Data Analysis</u> (1971). Their analysis algorithm centered primarily around the MANOVA and DISCRIM programs. A set of candidate performance measures was iteratively reduced on the basis of low communality until all members of the set retained a communality above some minimum value. The last discriminant analysis in the cycle provided the coefficients for the

performance measurement model. Figure 1 is a simplified flow chart of this portion of their analysis.

Analyses of empirical data were performed. The measures were selected, weighted and summarily combined into linear equations describing student learning on four basic IFM exercises. The automated training device using these equations was then to be implemented and evaluated by comparing its relative efficiency to other scoring algorithms. It was during debugging for the evaluation phase of the study that certain scoring system insufficiencies became apparent.

When introduced to simulator training, many novice, private pilots performed quite erratically at first. This initial performance in the simulator was unlike anything observed in the student population used to develop the performance measurement models. The discriminant linear models were very sensitive to these pilot behaviors and tended to misclassify some obvious novice pilots as experts.

A combination of incorrect sign and inappropriate magnitude of the discriminant coefficients was felt to contribute most to the instability of the measurement models. Other researchers have also observed contradictory coefficients when using multivariate regression analysis. Ridge regression has shown great success in damping these instabilities by biasing the model in a controlled and rational fashion. An analog to ridge regression was developed by the author for discriminant analysis and implemented for reanalysis of the performance measurement data. The resulting performance models were found to have merit when compared to existing measurement methods.

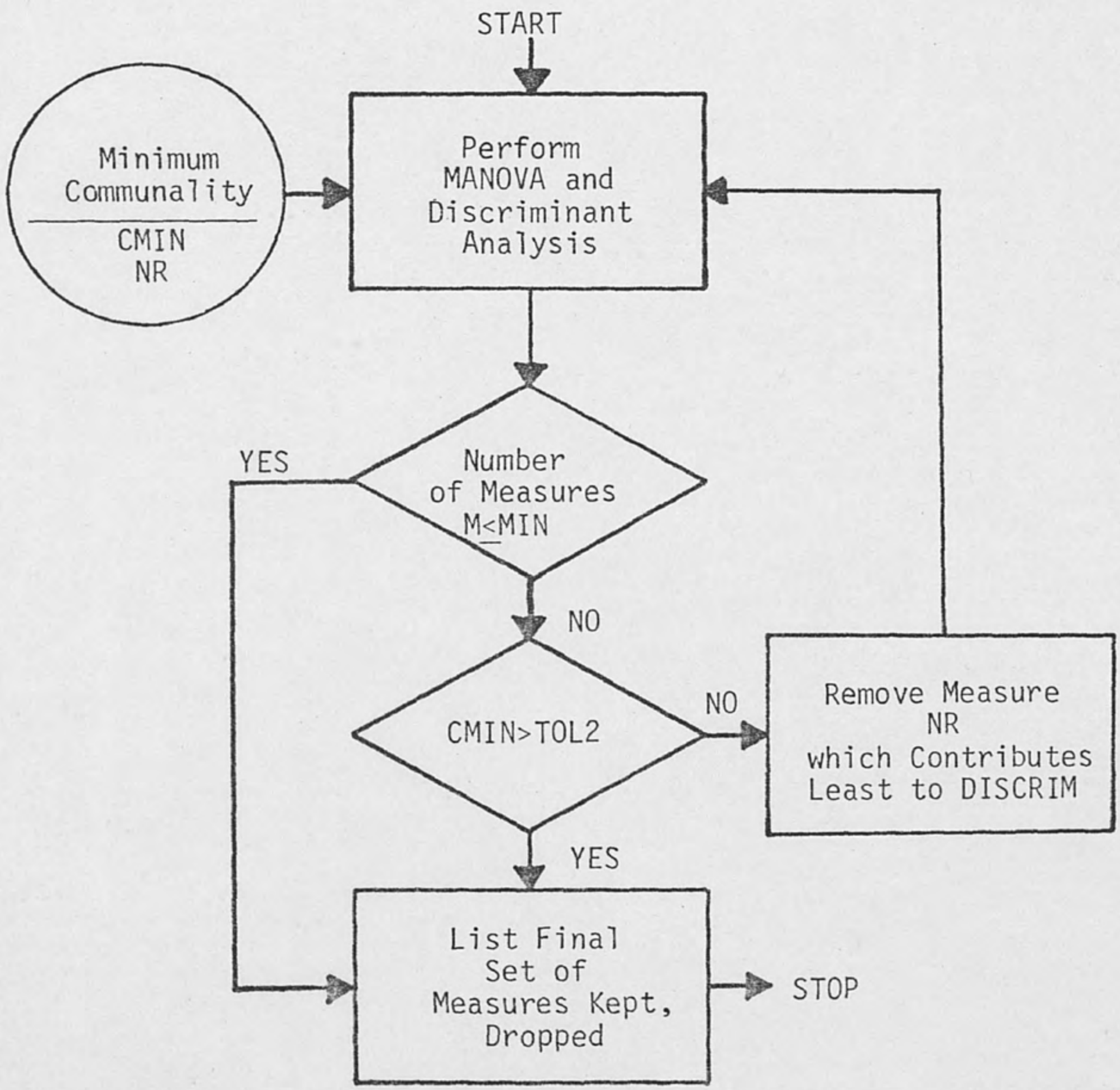

Figure 1. Discriminant Analysis Flow Chart

The performance measurement study, described above, is anecdotal evidence for the necessity and possible validity of multivariate ridge discriminant analysis. More detail on performance measure selection is contained in later sections of this report, along with a complete description of 'ridge' discriminant analysis. A simulation technique, using actual performance data, will be described and used to estimate the relative validities of different performance measurement analyses.

SECTION II

METHODS

Discriminant Analysis

Decision making requires intelligent integration of information. A linear model is usually the most convenient and optimal method for combining component variables. Merely testing for significance or performing analysis of variance is not often adequate for optimally determining group differentiation with respect to more than one variable. The discriminant model is useful for reducing this multivariate problem by the determination of a linear function of the variates which maximizes the difference between populations:

$$\pi = b_1 x_1 + b_2 x_2 + b_3 x_3 + \ldots + b_n x_n.$$

In review, multiple discriminant analysis projects data points from their initial measurement space into a suitable subspace. This subspace is univariate and usually referred to as discriminant space. The discriminant model determines those components which best separate the groups in measurement space and weights them to maximize this difference in discriminant space. The geometric interpretation of discriminant analysis can be seen, for the case of two groups and two variates, in figure 2.

In figure 2 can be seen two partially overlapping bivariate-normal scatter diagrams projected onto a new axis Π. The two new

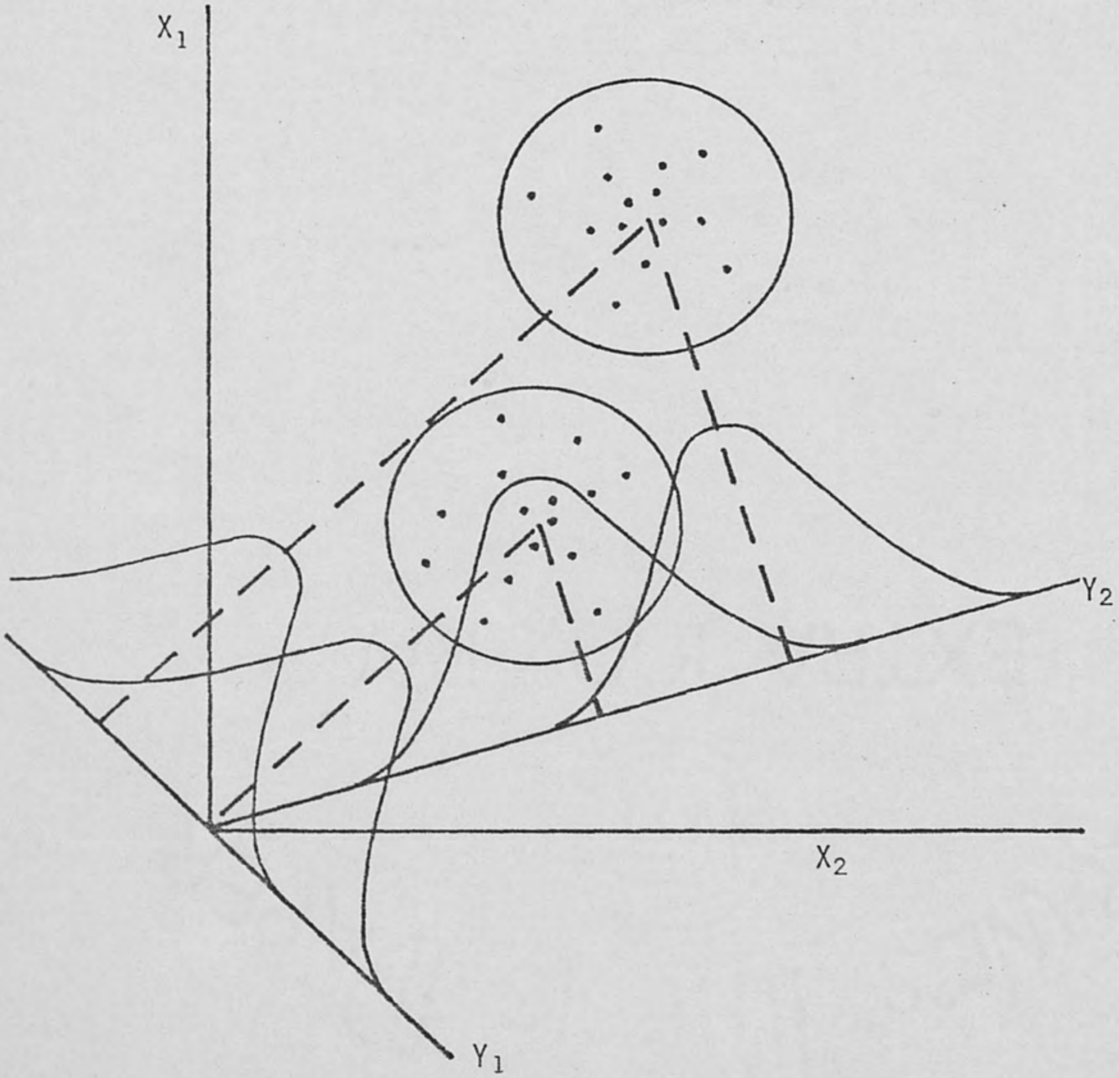

Figure 2. Different Linear Combinations for Two Variables

overlapping distributions represent the two groups projected onto an arbitrary discriminant axis. The degree of overlap can be manipulated by varying the discriminant coefficients used to transform the multivariate points onto the resulting discriminant axis. The objective of discriminant analysis is to find a set of coefficients which minimize this projected overlap for two or more groups consisting of many normally distributed variables.

Via Maurice Tatsuoka (1970) comes a simplified mathematical interpretation of the multivariate discriminant solution starting with a measure of the overlap between the two distributions π_1 and π_2. This difference could be calculated by dividing the difference between the two group means, $\overline{\Pi}_1$ and $\overline{\Pi}_2$, by the standard deviation of one of the groups. For example,

$$\frac{\overline{\Pi}_1-\overline{\Pi}_2}{s_{\pi,2}}$$

would describe the distance between the group 1 Π-mean and the group 2 Π-mean in units of standard deviation of Π in the second group.

This unit of distance favors one group over the other, depending on the choice of standard deviation. The pooled within-groups standard diviation, elemental in the t-test of significance of the difference between two means, would help provide a less biased estimate of the distance. Using

$$s_{\pi,W} = \frac{(n_1-1)s^2_{\pi,1}+(n_2-1)s^2_{\pi,2}}{n_1+n_2-2}$$

a "neutral" measure of degree of overlap becomes

$$f = \frac{\overline{\Pi}_1 - \overline{\Pi}_2}{s_{\pi,w}} .$$

f becomes smaller as the degree of overlap becomes greater. As a convenience, to avoid negative values, the final measure of the separation between the distributions of two groups projected in discriminant space becomes

$$f^2 = \left(\frac{\overline{\Pi}_1 - \overline{\Pi}_2}{s^2_{\pi,w}} \right)^2$$

Thus, by computing f^2 for systematic variation in the linear combinations, the optimum combination could be found when f^2 is largest. To avoid numerous computations, an expression is needed of f^2 that is a function of the coefficients b_1, b_2 through b_n. Differential calculus can be used to determine the set of b_i values which will maximize f^2 $(b_1, b_2, \ldots, b_n)$.

Before this can be done, the index, f^2, must be generalized to be used for cases involving two or more groups. It is a natural extension of the two group case that the square of the difference between the means is also provided by the variance of three or more quantities.

When three or more groups ($l \geq 3$) are being compared

$$f^2_l = \frac{\mathrm{Var}\ (\overline{\Pi})}{s^2_{\pi,w}}$$

where
$$\mathrm{Var}(\overline{\Pi}) = \frac{\sum_{g=1}^{l} (\overline{\Pi}_1-\overline{\Pi}_2)^2}{l-1}$$

with
$$\overline{\Pi} = \frac{\overline{\Pi}_1 + \overline{\Pi}_2 + \ldots + \overline{\Pi}_l}{l}$$

and
$$s^2_{\pi,w} = \frac{\sum_{g=1}^{l} (n_g-1)s^2_{\pi,g}}{N-l}.$$

$s^2_{\pi,w}$ is now the within-groups mean-square, MS_w; used in the analysis of variance. The total sample size is

$$N = \sum_{g=1}^{l} n_g.$$

In the f^2 the $\mathrm{Var}(\overline{\Pi})$ does not account for unequal groups. The size of each group can be accounted for by replacing $\mathrm{Var}(\overline{\Pi})$ by the between-groups mean-square

$$MS_b = \frac{\sum_{g=1}^{l} n_g\ (\overline{\Pi}_g-\overline{\Pi}')^2}{l-1}$$

where
$$\overline{\Pi}' = \frac{\Sigma n_g\ \overline{\Pi}_g}{N}$$

is the grand mean of Π in the total sample comprised of all l groups.

Using the notation of analysis of variance,

$$f^2_l = \frac{MS_b}{MS_w}.$$

To further simplify

$$f_l^2 = \frac{SS_b/(l-1)}{SS_w/(N-l)} = \frac{SS_b}{SS_w} \frac{N-l}{l-1}$$

and discarding $(N-l)/(l-1)$, since it is constant for any given problem,

$$f_l^2 = \frac{SS_b}{SS_w} = \lambda$$

where

$$SS_b = \sum_{g=1}^{l} n_g \, (\overline{\Pi}_g - \overline{\Pi}')^2$$

and

$$SS_w = \sum_{g=1}^{l} \sum_{i=1}^{n_g} (\overline{\Pi}_{g,i} - \overline{\Pi}_g)^2 .$$

The quantity λ is called the discriminant criterion. The next step of defining SS_b, SS_a and the linear combination Π in terms of the unknown coefficients $b_1, b_2, \ldots, b_n$ will require matrix notation.

If we define the vectors

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{bmatrix}, \; X = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix}, \; \overline{X}_g = \begin{bmatrix} \overline{X}_{1,g} \\ \overline{X}_{1,g} \\ \cdot \\ \cdot \\ \cdot \\ \overline{X}_{1,g} \end{bmatrix}, \text{ and } \overline{X} = \begin{bmatrix} \overline{X}_1 \\ \overline{X}_2 \\ \cdot \\ \cdot \\ \cdot \\ \overline{X}_n \end{bmatrix}$$

the foregoing expressions become

$$\Pi_{g,j} = b' X_{g,i}$$

$$\overline{\Pi}_g = b' \overline{X}_g$$

and $$\overline{\Pi} = b'\overline{X}$$

where b' is the transpose of b, the row vector $\{b_1, b_2, \ldots, b_n\}$. The Π-scores for all N individuals can be expressed as

$$\Pi' = b'X$$

and it follows that

$$\overline{\Pi}'_g = b'\overline{X}_g J$$

and $$\overline{\Pi}' = v'\overline{X}J$$

where $$\overline{X}J = \left[\overline{X}\ \overline{X}\ \overline{X}\ \ldots\ \overline{X}\right]$$

and $$\overline{\Pi}' = \left[\overline{\Pi}\ \overline{\Pi}\ \overline{\Pi}\ \ldots\ \overline{\Pi}\right]$$

for N repeated columns and

$$\overline{X}_g J = \overline{X}_{1,1}\ \overline{X}_{2,1}\ \overline{X}_{3,1}\ \ldots\ \overline{X}_{n1,1}\ \overline{X}_{1,2}\ \overline{X}_{2,2}\ \overline{X}_{3,2}\ \overline{X}_{n2,2}\ \ldots$$

$$\overline{X}_{1,g}\ \overline{X}_{2g}\ \ldots\ \overline{X}_{ng,g}\ .$$

Since $$\pi_{1,i} = \Pi_{1,i} - \overline{\Pi}_1 \quad (i = 1, 2, \ldots, n_1)$$

and $$\pi_{2,i} = \Pi_{2,i} - \overline{\Pi}_2 \quad (i = 1, 2, \ldots, n_2)$$

and so on, $$SS_w = (\Pi - \overline{\Pi}_g)'(\Pi - \overline{\Pi}_g).$$

In order to express the right-hand side of the previous formula in terms of b, substitute

$$\Pi' - \overline{\Pi}'_g = b'X - b'\overline{X}j_g$$

$$= b'(X - \overline{X}J_g)$$

and find $$SS_w = b'(X - \overline{X}J_g)\ \ b'(X - \overline{X}J_g)'$$

$$= b'\ \ (X - \overline{X}J_g)\ \ \ (X - \overline{X}J_g)'\ \ b.$$

Note for future reference that the within-groups sum-of-squares of X form the diagonal elements and the within-groups sum-of-cross-

products between the elements of X form the off-diagonal elements of $(X-\overline{X}J_g)(X-\overline{X}J_g)'$, which is usually denoted by W. This is called the within-groups sums-of-squares-and-cross-products (SSCP) matrix. Thus, $SS_w = b'Wb$.

Following the previous example

$$SS_b = (\overline{\Pi}_g - \Pi)' \ (\overline{\Pi}_g - \Pi)$$

$$= b'(\overline{X}J_g-\overline{X}J) \quad b'\overline{X}J_g-\overline{X}J)\ '$$

$$= b' \ (\overline{X}J_g-\overline{X}J)(\overline{X}J_g-\overline{X}J) \ b$$

or

$$SS_b = b'Bb.$$

B is called the between-groups SSCP matrix where the between-groups sum-of-squares of X make up the diagonal elements and the between-groups sum-of-cross-products are the off-diagonal elements.

Now the discriminant criterion

$$\lambda = \frac{SS_b}{SS_w}$$

can be expressed as

$$\lambda = \frac{b'Bb}{b'Wb} .$$

The coefficients in discriminant analysis are the result of maximizing the ratio of the among-groups variance over the between-groups variance:

$$\lambda = \left. \frac{b'Ab}{b'Wb} \right|_{\text{maximum}}$$

subject to the restriction

$$b'b = 1.$$

The maximum value λ and the associated vector of weights b are found using the largest eigen value and its eigenvector of the equation

$$(W^{-1}A-\lambda I)b=0 \text{ (Tatsuoka 1970).}$$

The purpose for the inclusion of this extensive derivation of the discriminant criterion is to give the reader an intuitive feel for the meaning of the among- and between-groups SSCP matrices. It is this understanding that was critical to the development of 'ridge' discriminant analysis.

'Ridge' Discriminant Analysis

It makes no difference to the formal logic of the discriminant model whether the variates in measurement space are the dependent variables and the discriminant function is the independent predictor vector or the groups consist of independent treatment variables and group membership is the dependent variable vectors. In the case of the performance measurement development, previously discussed, the groups represented relative skill levels of pilots composed of dependent and uncontrolled performance measures and the group membership was the only controlled independent variable.

Undesigned Experiments

Charles Simon (1975) would refer to this kind of method for developing a performance model as an "undesigned" experiment. An "undesigned" experiment, according to Simon, is one in which some

experimental variables cannot or are not controlled by the experimenter. Each variable is known or measured only at the time of measurement during the experiment. Therefore, variables in an undesigned experiment may often be correlated mathematically to some degree. This correlation can dramatically effect the results of a multivariate analysis.

For example, in the performance measurement study, mentioned above, many of the aircraft parameters that were transformed to provide the measures were dynamically related. The aircraft attitude is usually directly linked to the pilots' control actions. Also, various measures are sometimes just different transforms of the same parameter. Even so, the interrelationships between the resulting measures are complex and the aspects of performance they mutually describe (i.e., their simple correlation) are unpredictable for any new maneuver. In other words, even though a relationship between basic aircraft and aircraft control parameters can be demonstrated, transformed measures of these parameters may not always be perfectly correlated. Elimination from the analysis is suggested for one measure from a near-perfectly correlated pair, since they obviously describe the same aspect or variance of the performance.

A different kind of mathematical dependence can occur, which is an artifact of the uncontrolled experimental conditions at the time of measurement. These accidental or artifical correlations are not a result of any casual relationship between the predictor variables.

Regardless of the source, even seemingly small correlations have the potential for producing less than satisfactory models for some applications, and it is not intended to argue here the merits of "designed" versus "undesigned" experiments other than to reiterate Charles Simon's viewpoint:

> "The goals of a good experiment should be to obtain new, relevant, important, and lasting information which is capable of explaining most of the performance variability associated with a particular real-world task. In the behavioral sciences, unlike the physical sciences, performance cannot be examined or evaluated independly of the context in which it occurs and can only be described or predicted as a function of this context. The more generalizable data therefore will be derived from experiments in which critical context factors are varied rather than held constant.
>
> If, however, an investigator decides to study behavior in a realistic context, he may find himself in circumstances where his ability to control and adjust the levels of critical parameters is sorely limited. This means that he can no longer plan and carry out a totally designed experiment and must either limit the questions he can ask or resort to another approach. The undesigned experiment--along or in conjunction with a balanced design--offers a viable alternative."

Interpretation of "Undesigned" Experiments

When the predictor variables are correlated with each other, the intercorrelation matrix will have non-zero correlations in the off-diagonal positions. A hypothetical intercorrelation matrix for the discriminant situation appears in table 1. This table can be broken into three parts: 1) the predictor matrix of correlations among each predictor variable and every other predictor variable; 2) the diagnonal of the matrix which is the correlation of each predictor variable to itself; and, 3) the group membership row vector

TABLE 1

EXAMPLE INTERCORRELATION MATRIX FROM
CHARLES SIMON'S (1975) REPORT

		Predictor Variables X1	X2	X3	Performance (Y)
Predictor Variables	X1	1.00	0.145	0.352	0.674
	X2	0.145	1.00	0.022	0.532
	X3	0.352	0.022	1.00	0.348

of correlations between each predictor variable and group membership.

Note that since each predictor variable correlates perfectly and positively with itself, the diagonal values are all one. Since the off-diagonal elements are not zero, the matrix is said to be ill-conditioned and the original experimental design is classified as non-orthogonal.

A multivariate least-squares regression as well as a multivariate discriminant analysis will both suffer from similar failings when applied to data characterized by an ill-conditioned intercorrelations matrix. Indeed, there is a popular contention that the regression and discriminant models should always find the same solution for any given data. It has been suggested that any notable differences are due to computational problems. The author feels that the regression criterion

$$(Y-Xb)'(Y-Xb)\Big|_{Min}$$

and the discriminant criterion

$$\frac{b'Ab}{b'Wb}\Bigg|_{Max}$$

behave somewhat differently under adverse conditions. Should the groups have significantly different shapes, be non-normally distributed or from "undesigned" experiments, the two criteria may result in quite different discriminant models. Until this relationship is mathematically or practically proven, the author reserves the right to discuss the two methods as different, yet analogous,

procedures. For the time being, it is this analogy that is of critical importance.

Users of either analysis technique, McDonald and Schwing (1973) for instance, have noted certain instabilities in the resulting ordinary coefficients when analyzing non-orthogonal systems. Some coefficients have extreme magnitudes or incorrect signs resulting in linear functions that respond unsatisfactorily when supplied with new data. This erratic behavior was also noted in the Vreuls and Wooldridge (1976) performance measurement study when new pilots were obviously misclassified by the normal discriminant functions.

Hoerl and Kennard (1970a,b) suggested adding a small positive quantity, k, to the unit diagonal of the intercorrelation matrix, $X'X$, of the predictor variables in regression analysis. The conventional least-squares fit is done using this new matrix to produce what are called 'ridge' coefficients. The standard regression

$$b = (X'X)^{-1}X'Y$$

becomes

$$b = (X'X + kI)^{-1}X'Y$$

where I is the unit diagonal matrix.

The 'ridge' comes from the fact, that as k increases, the variance error decreases more rapidly than the bias error increases. As can be seen in figure 3, for some value of k, the sum of the bias error and variance error (the mean-square error) is minimized and smaller than it would be for the conventional coefficients. Although the 'ridge' can be mathematically demonstrated to exist, little success has been made in calculating a specific value of k

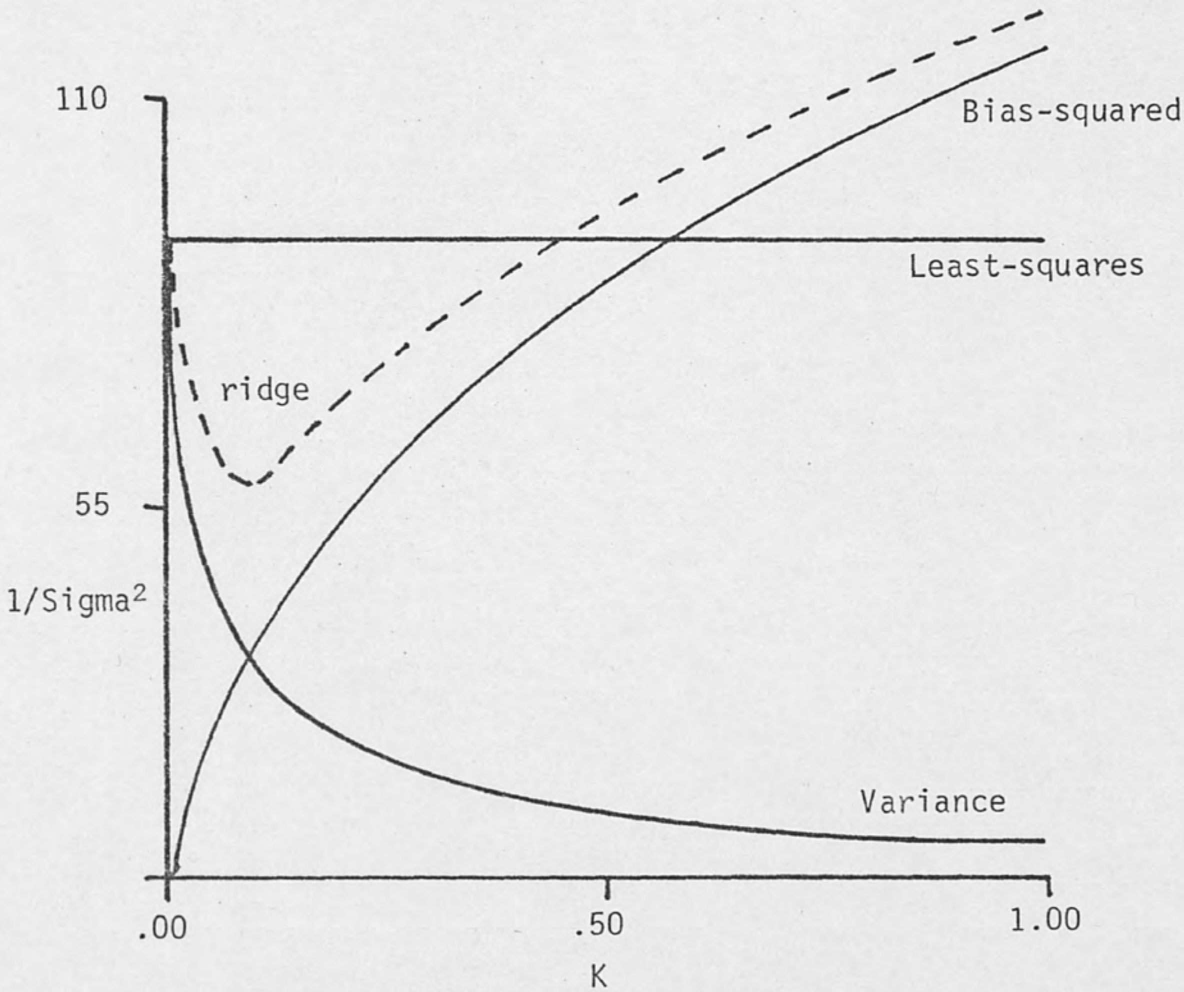

Figure 3. Example Ridge Regression
Mean-Square-Error Functions

that minimizes the mean-square error. Lindley and Smith (1972); Mallows (1973); and Farebrother (1975), to name a few, all suggest various mathematical criteria for selecting a value of k which would improve the set of coefficients without unduly biasing the estimate.

Hoerl and Kennard did not feel that a mathematical solution for selecting the best k was justified. They proposed visual inspection of the 'ridge' trace. Figure 4 is an example 'ridge' trace. This plots the change in regression coefficients various values of k between 0 and 1.

The following conditions should be looked for when selecting the value of k (in lieu of a mathematical formula):

1. The beta values and particularly their orders of magnitudes have begun to stabilize.
2. The coefficients no longer have unrealistically large absolute values.
3. The coefficients with logically incorrect signs are approaching or have reached the proper sign.
4. The residual sum-of-squares is not unreasonably inflated.
5. The ridge trace (representing the mean-square error) is smaller than the unbiased least-squares variance.

Mathematical Rationale for 'Ridge' Discriminant Analysis

Hoerl and Kennard (1970a) were able to mathematically demonstrate the existence of the ridge estimator for regression by calculating the expected value of the squared distance between $\hat{\beta}^*$ and

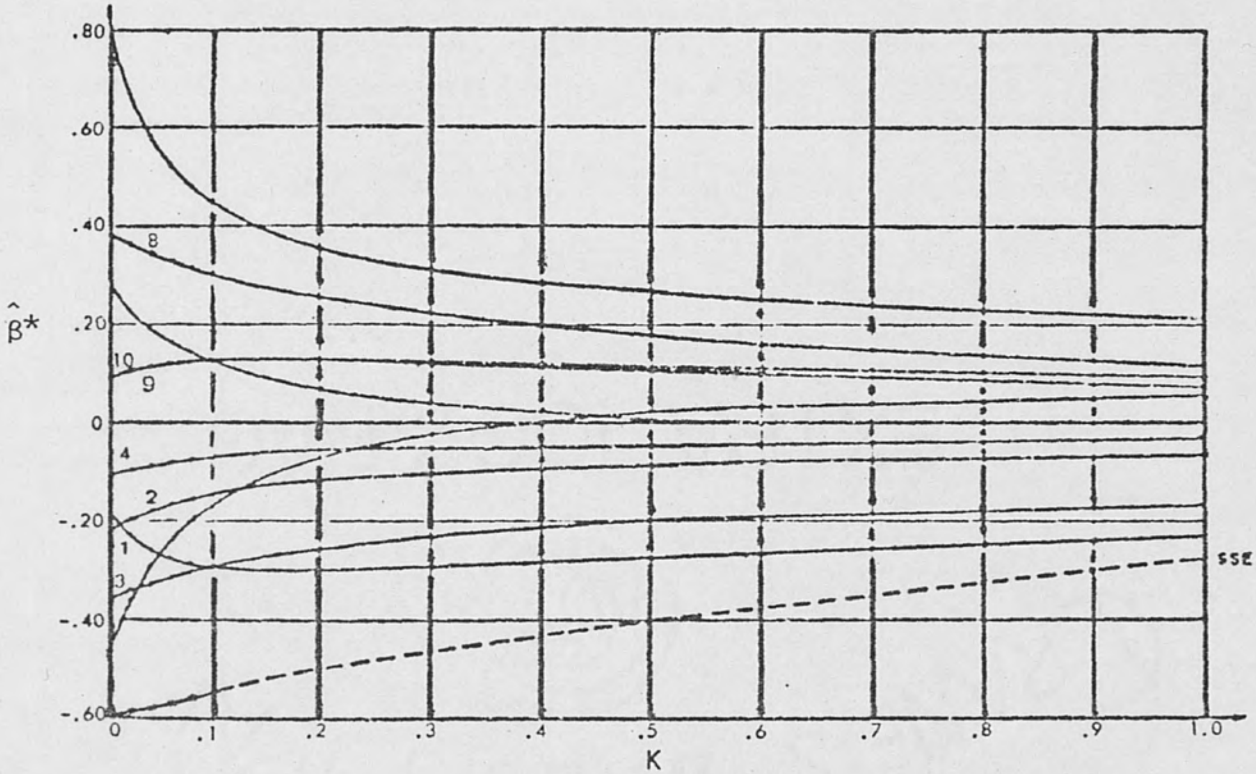

Figure 4. Example Coefficient Trace (Hoerl and Kennard 1970)

β. β is the vector of the true regression coefficients and $\hat{\beta}^*$ is the 'ridge' estimates of β. This function

$$E\left[L^2(K)\right] = (\hat{\beta}^*-\beta)'(\hat{\beta}^*-\beta)$$

becomes

$$E\left[L^2(K)\right] = \sigma^2\sum_{1}^{p} \lambda_i/(\lambda_i+K)^2 + K^2\beta'(X'X+KI)^{-2}\beta$$

or

$$E\left[L^2(K)\right] = \gamma_1(K)+\gamma_2(K).$$

The first term, $\gamma_1(K)$, can be shown to be the sum of the variances, or total variance, of the estimate. $\gamma_2(K)$ can be considered the square of the bias resulting when $\hat{\beta}^*$ is used rather than $\hat{\beta}$. The function $\gamma_1(K)$ is a monotonic decreasing function of K, while $\gamma_2(K)$ is monotonic increasing. As previously discussed, the sum of the two provide the 'ridge'.

An existence theorem can demonstrate that, although the derivative of $E\left[L^2(K)\right]$ is positive as K approaches ∞, there always exists a $K>0$ such that $dE\left[L^2(K)\right]/dK<0$. This describes the property of $E\left[L^2(K)\right]$ of always going through a minimum as K goes from 0 to ∞. Appropriate values of K have been looked for by various solutions to the first derivative of $E\left[L^2(K)\right]$.

Development of a similar expectation function for a 'ridge' discriminant analog poses a challenge which this author may not have to overcome. Although the author advances demonstration by simulation, it is useful to examine the discriminant function to grasp an intuitive feel for the existence of the 'ridge' adjustment.

In review, discriminant analysis attempts to maximize the criterion,

$$\lambda = \frac{b'Ab}{b'Wb}.$$

Finding the smallest eigen value of W to maximize λ is reasonably comparable to finding the largest eigen value of $W^{-1}A$ used to solve $(W^{-1}A-\lambda I)b=0$. The smallest eigen value of W, λ_1, would be very 'small' if two or more experimental variables were correlated. Then if b were chosen to correspond to λ_1 of W then,

$$b'Wb=b'(Wb)=b'(\lambda_1 b)=\lambda_1 b'b=\lambda_1,$$

since $b'b\underline{\Delta}1$. Since λ_1 is 'small', $\frac{b'Ab}{\lambda_1}$ is very 'large'. Undesirably, the solution would disregard or be insensitive to the values found in A. Somewhat larger eigen values of W would be more desirable. One way to force the eigen values to be larger is to replace W with W+IK, where K is a small scalar. Then,

$$b'(R+IK)b=\lambda_1+K.$$

Now, regardless of how small the smallest eigenvector of W is, the size of λ_1+K can be no smaller than K. K being of reasonable size, $\frac{b'Ab}{\lambda_1+K}$ would not be as large as $\frac{b'Ab}{\lambda_1}$. In maximizing $\frac{b'Ab}{b'(W+IK)b}$, A would determine more of b's direction than before. Thus, a very small bias, K, would have the beneficial result of improving the sensitivity of λ to values found in A. Again, as in 'ridge' regression there is a trade-off between bias error and variance. It is hypothesized that the adjusted discriminant function has similar

minimizable properties as in the case of least-squares regression. A simulation would also be expected to demonstrate these properties.

Predictably, one characteristic of the 'ridge' discriminant analysis will be quite different from that of the regression version. Usually, the regression case requires only very small values of K, much smaller than 1, usually less than 0.1, to minimize $E\left[L^2(K)\right]$. The mathematics of the situation suggests that the discriminant analysis may require adjustments much larger than those typical for regression.

To explain further, remember that W contains the within-groups sum-of--squares of X on the diagonal elements and the within-groups sum-of-cross--products of X form the off-diagonal elements. $X'X$ relates to W when XX' is expressed as the correlation matrix, r. When $X'X=r$, it follows that

$$r_{ij} = \frac{W_{ij}}{\sigma_i \sigma_j}$$

Thus, the elements of W can be much larger in magnitude than those of $X'X$; to be precise, $\sigma_i \sigma_j$ times greater. As a result, it can take larger values of K to affect the eigen structure of W^{-1}.

Monte Carlo Validation of 'Ridge" Discriminant Analysis

Given the lack of 'ridge' discriminant analysis application experience and the absence of a rigorous mathematical proof, a Monte Carlo simulation (Gordon 1969) would be a practical way to demonstrate the effectiveness of this new multivariate technique

(Bittner 1974). Using simulated data, the percentage of classification error could be calculated for various values of k. Thus providing an immediate indication of the relative improvement in the discriminant function associated with each value of k.

As was previously discussed, several researchers when using 'ridge' regression have chosen values for k based on a wide variety of mathematical or pragmatic criteria. This simulation technique, on the other hand, can provide pseudo-empirical evidence for the selection of a near-optimal discriminant function for any sample simply by finding the minimum classification error as the value of k increases from zero. Rather than base a decision on a highly contestable mathematical assumption, perhaps a data simulation should be incorporated as the criteria whenever using a 'ridge' analysis.

Mathematical Basis for Simulation

As a first consideration, the simulated data must have a multivariate distribution much like the original sample. For the most part, human performance data are normally distributed with error scores having highly skewed normal or translated poisson distributions. This mixture of distributions could be satisfactorily simulated with random vectors generated from a multivariate normal population. Since the interrelationships existing in the performance measures was the undesirable characteristic responsible for the development of 'ridge' analyses, it is crucial that the same variance-covariance structure of the data be represented in the simulation. Thus it was important not only to simulate a

multivariate normal population, but to also be able to specify the variance-covariance matrix of the population.

Scheuer and Stoller (1960) suggest a method for generating random vectors from a multivariate normal population with a specified variance covariance matrix based on matrix equations. To simplify description of the technique, it will be assumed at first that the mean of the random vectors is zero. The result is no loss in generality, for a vector x with a mean of zero and a variance-covariance matrix Σ, the vector $x+\mu$ has the same variance-covariance matrix Σ and mean vector μ. It is then possible to concentrate on generating a random vector $x=(x_1, x_2, \ldots x_n)$ from $N(0,\Sigma)$, the multivariate normal distribution with mean vector zero and variance-covariance matrix:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \sigma_{n1} & \ldots & \sigma_{nn} \end{bmatrix}.$$

Let y be distributed $N(0,I_n)$, where I_n is the unit matrix of size n, and let $x=Cy$. Then x is distributed $N(0,CC')$. It is required that CC' be equal to Σ in this case. The matrix C is unique and readily determined if C is lower triangular. The elements of C are determined recursively as follows:

$$C_{i1} = \sigma_{i1} / \sqrt{\sigma_{i1}}, \quad 1 \leq i \leq n,$$

$$C_{ij} = \sqrt{\sigma_{ii} - \sum_{k=1}^{i-1} C_{ik}^2}, \quad 1 \leq i \leq n,$$

$$C_{ij} = \left[\sigma_{ij} - \sum_{k=1}^{j-1} C_{ik}\, C_{jk}\right] \Big/ C_{jj}, \; 1 < j < i \leq n$$

$$C_{ij} = 0, \qquad\qquad i < j \leq n.$$

This technique is referred to as the "square root" method and C is the "square root" of Σ.

Once C has been determined, x is obtained by

$$x_i = \sum_{j=1}^{i} C_{ij} y_j, \; i = 1, \ldots, n$$

where y_1, ... y_n are independent standard normal variables, N(0,1).

Box and Miller (1958) suggest a method for computation of random normal deviates. This approach has been shown to be more accurate than other known methods for generating normal deviates from independent ramdom numbers; 1) the inverse Gaussian function of the uniform deviates, 2) Teichroew's approach, 3) a rational approximation such as that developed by Hastings, 4) the sum of a fixed number of uniform deviates, and 5) rejection type approach.

The method may be used to generate a pair of random deviates from the same normal distribution starting from a pair of random numbers. Letting U_1 and U_2 be independent variables from the same rectangular density function on the interval (0,1),

$$y_1 = (-2 \text{ Log}_e \text{ U}_1)^{\frac{1}{2}} \text{ COS } 2\pi\text{U}_2 \text{ and}$$

$$y_2 = (-2 \text{ Log}_e \text{ U}_1)^{\frac{1}{2}} \text{ SIN } 2\pi\text{U}_2$$

provides a pair of independent random variables, (y_1, y_2), from the same normal distribution with mean zero, and unit variance.

The new random vector x can now be computed given the original μ and Σ as

$$x_i = \left[\sum_{j=1}^{i} C_{ij}\, y_i \right] + \mu_i,\ i = 1, \ldots, n.$$

A program was then written which would compute any number of new independent random vectors with the same means and variance-covariance matrix. The simulation would require a predetermined number of generated vectors from each population to then be classified as to their respective populations using various discriminant functions. The resulting misclassifications can then be used to compare the discriminant functions by calculating their respective percent error of classification. An acceptable criterion must be developed to optimally classify vectors for each discriminant function.

The score s will be calculated by:

$$s = b_1x_1+b_2x_2+b_3x_3+\ldots+b_nx_n,$$

where the b's are the discriminant coefficients for n number of performance measures or x's. In other words, the b's make up the discriminant function for the sample vector composed of n number of x's. s is now a new variate in discriminant space belonging

to one of two distributions in this case. The cost function can be derived using the score distributions in discriminant space of the sample used to empirically develop the discriminant function, s. There will be two of these distributions: π_1 and π_2. Any single score, s, will then fall somewhere in these distributions, with probabilities of $P_1(s)$ and $P_2(s)$.

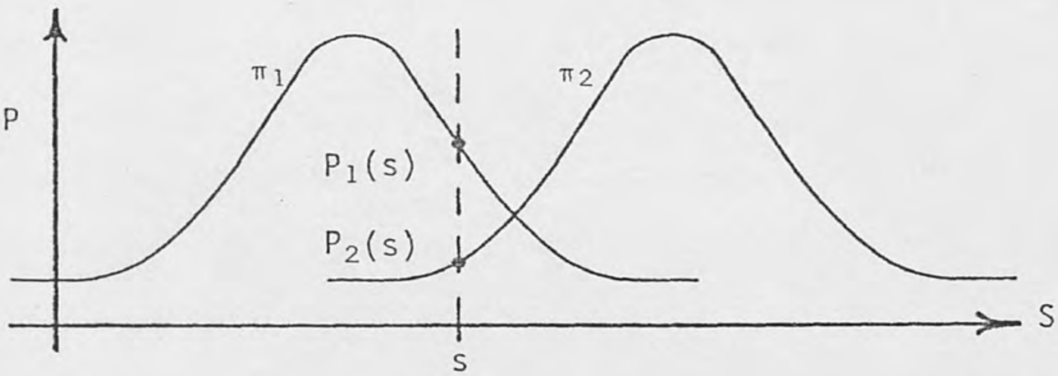

Figure 5. Two Distributions in Discriminant Space

In classifying s as belonging to distribution π_1 or π_2, two errors can be made. If s actually belonged to distribution π_1, an error would be made if s was classified as belong to π_2 (see figure 5).

On the other hand, s could belong to π_2 and be classified as belonging to π_1. There is a cost associated with each type of error. Let $C(2|1)$ be the cost of the first type of error and $C(1|2)$ be the cost of the second. Table 2 is a logic table of the costs of correct and incorrect classification. It follows, that an effective classification scheme should minimize the cost of classification.

TABLE 2

CLASSIFICATION

		π_1	π_2
Actual	π_1	0	C(2|1)
Membership	π_2	C(1|2)	0

If we select a score, s, in discriminant space, the potential cost of using that point for classification can be estimated. The probability that s will be classified as belonging to π_2 even though it belongs to π_1 is

$$P(2|1,s) = \int_{s}^{\infty} P_1(s)ds.$$

Given that we already have an actual sample distribution where the number of observations in π_1, n_1, is known and the number of scores less than s in π_1 can be summed as m_1,

$$P(2|1,s) = \frac{n_1-m_1}{n_1}.$$

The probability of misclassification of an observation from π_2 is then

$$P(1|2,s) = \int_{-\infty}^{s} P_2(s)ds = \left(1 - \frac{n_2-m_2}{n_2}\right),$$

where n_2 is the number of observations in π_2 and m_2 is the number of scores in π_2 less than s.

The probability of a π_1 pilot achieving a particular score, s, is $P_1(s)$. This probability can be calculated using the number of observations in π_2 falling in the period of integration bounded by s, $m_{s,1}$, divided by n_1 or simply

$$P_1(s) = \frac{m_{s,1}}{n_1}.$$

Thus, the probability associated with misclassifying a score from π_1 is

$$P_1(s)P(2|1,s) \text{ or } \frac{m_{s,1}}{n_1} \; \frac{n_1-m_1}{n_1}$$

and the probability of misclassifying a score from π_2 is

$$P_2(s)P(1|2,s) \text{ or } \frac{m_{s,2}}{n_2}\left(1 - \frac{n_2-m_2}{n_2}\right).$$

The average or expected loss from costs of misclassification is the sum-of-the-products costs of each misclassification multiplied by the probability of its occurrence;

$$C(2|1)P(2|1,s)P_1(s) + C(1|2)P(1|2,s)P_2(s), \text{ or}$$

$$C(2|1) \; \frac{n_1-m_1}{n_1} \; \frac{m_{s,1}}{n_1} \quad + \quad C(1|2)\left(1- \frac{n_2-m_2}{n_2}\right)\frac{m_{s,2}}{n_2}.$$

If $C(1|2) = C(2|1) = 1$, the expected loss is

$$\frac{n_1-m_1}{n_1}\,\frac{m_{1,s}}{n_1} + \left(1- \frac{n_2-m_2}{n_2}\right)\frac{m_{s,2}}{n_2}.$$

Assuming π_1 and π_2 are normally distributed, for a given score s, the probability of misclassification is minimized by assigning s to the sample that has the higher conditional probability. Thus, the rule is:

π_1 is chosen if

$$\frac{n_1-m_1}{n_1}\,\frac{m_{s,1}}{n_1} \geq \left(1- \frac{n_2-m_2}{n_2}\right)\frac{m_{s,2}}{n_2}$$

and π_2 is chosen if

$$\frac{n_1-m_1}{n_1}\,\frac{m_{s,1}}{n_1} < \left(1- \frac{n_2-m_2}{n_2}\right)\frac{m_{s,2}}{n_2}.$$

This line of reasoning follows closely the derivation put forth by Anderson (1958) for discriminant classification criteria. For the purposes of this research, it suffers from two fatal problems. First, the computation of $P_1(s)$ and $P_2(s)$ relies on the arbitrary determination of a period of integration. Since these values can change directly with the length of the integration period selected, the cost function itself becomes a problem to define. Secondly, the distributions to be analyzed are known not to be normally distributed. This upsets the decision rule described above.

A slight departure from these results may provide a practical decision algorithm for realistic data analysis conditions. Assuming that the intrinsic cost associated with misclassifications of both kinds is still 1, the total probability of misclassification for any s provides a relative metric for comparison. The total probability of misclassification is

$$P(1|2,s) + P(2|1,s) \text{ or}$$

$$\frac{n_1 - m_1}{n_1} + \left(1 - \frac{n_2 - m_2}{n_2}\right).$$

For an existing sample, $P(1|2,s)$ and $P(2|1,s)$ are simply the percentage of misclassifications of π_1 and π_2 respectively. Their sum is the total percentage of misclassifications for s in discriminant space. Given a sample, this total percentage can be evaluated along the entire discriminant space to locate the minimum error that will occur somewhere between the means of π_1 and π_2. The value in discriminant space where the minimum error occurs will be called the break even point.

Several potential discriminant functions can be derived for each sample of data. A classification criteria for each function can be determined by finding the break even point using the sample discriminant scores, and the discriminant functions can then be compared using their respective percentage of classification errors on generated independent variates resembling the original sample. The simulation concept is now complete.

SECTION III

ANALYSIS SOFTWARE ARCHITECTURE

All of the simulation elements were then coded and assembled into a single analysis algorithm. The heart of the program was the existing discriminant analysis shown in figure 1 of the Introduction. As can be seen in figure 6, the generation of random variables, the 'ridge' adjustment, and simulated test of the discriminant model have been added to the original program.

First, a thousand simulated measurement vectors are generated for each of the two groups analyzed. These 'measures' are saved on a disc and reused to test each set of discriminant coefficients resulting from the various stages of analysis. An economy of computational time is realized by only generating one large set of simulated data. The statistical reliability that would be gained by recalculating the test sample for each discriminant model would be insignificant. Figure 7 shows the functional step involved in the generation of the test data and follows closely the conceptual material presented in the previous sections.

Discriminant analysis is then iteratively performed, removing measures which contribute least to the model, until all remaining measures retain a communality higher than some arbitrary minimal value. Communality is therefore the criterion for retention in the model. Control of this process is done by varying the minimum

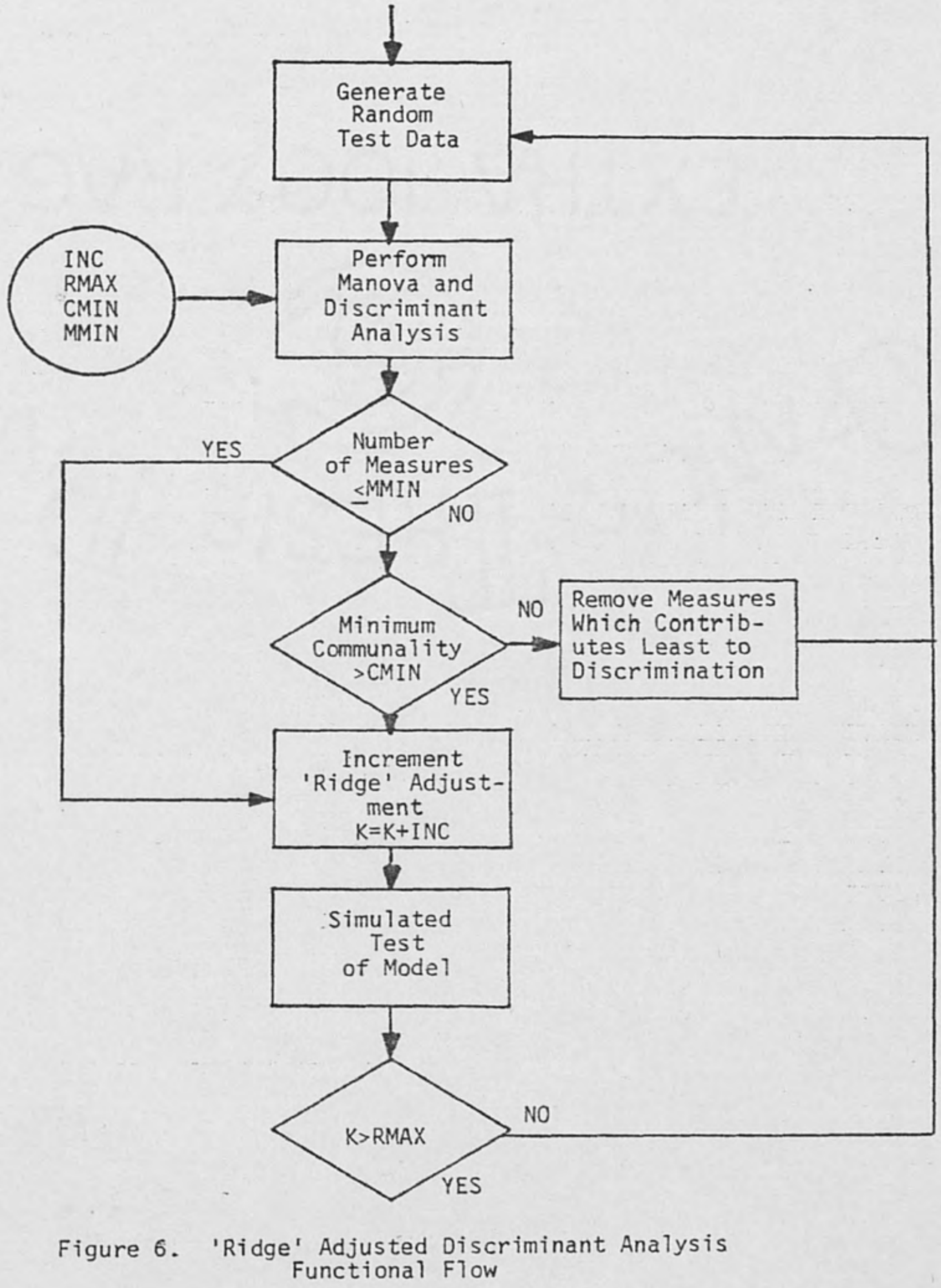

Figure 6. 'Ridge' Adjusted Discriminant Analysis Functional Flow

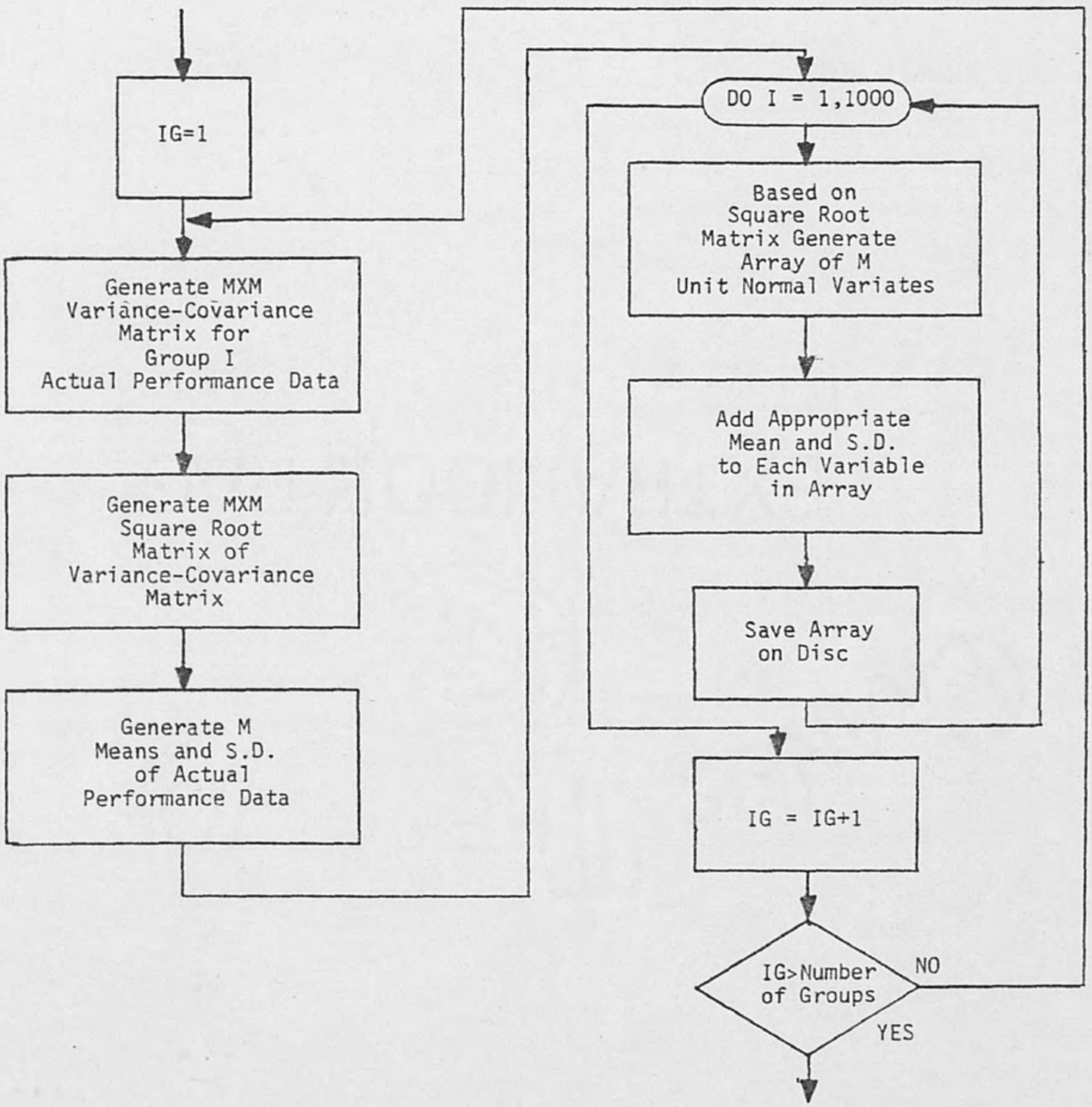

Figure 7. Function Flow of Random Test Data Generation Program for M Variates

communality allowed or by initial inclusion of measures at the start of the process.

Once the final set of measures has been determined, the 'ridge' adjustment to the model is made. A predetermined increment is added to K, which in turn is added to the diagonal of the W matrix, and the discriminant analysis is performed again. This cycle is repeated until the 'ridge' adjustment, K, is greater than the limit set by the operator.

In between each discriminant analysis, during the 'ridge' adjustment cycles, each set of discriminant coefficients undergoes the simulation test. As shown in figure 8, the resulting coefficients are first applied to the actual data, using the cost function, to find a break-even point. The predicted error of classification can also be calculated by summing the misclassifications and dividing by the total number of observations. The simulated data are then weighted by the discriminant coefficients and the same break point is used as the classification criterion. The percentage of misclassifications then becomes a gauge of the relative reliability of the resulting models.

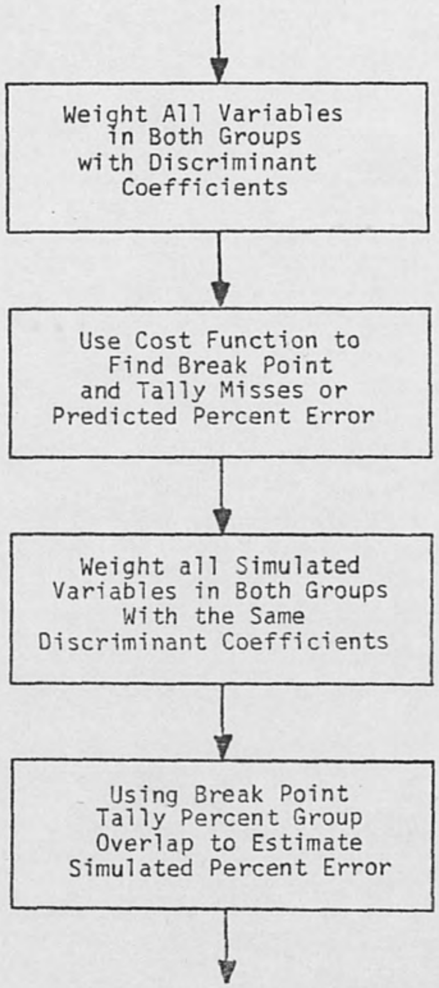

Figure 8. Simulated Test of Discriminant Model Functional Flow

SECTION IV

APPLICATION OF THE 'RIDGE' DISCRIMINANT ANALYSIS

The 'ridge' adjusted discriminant analysis was used, without the benefit of the simulation feature, to develop performance measurement models (Vreuls, Wooldridge 1976). These models were capable of differentiating between trained and untrained pilots in an instrument flight simulator. The implementation of these empirically derived models resulted in a 40 percent reduction in the time required to train pilots to a fixed level of proficiency. In the context of that study, the discriminant model represented an improvement in control criteria for an automated training system.

Fortunately, the successful results of the first empirical measurement study lent a degree of validity to an untested, new statistical procedure. During the analysis stage of that study, the 'ridge' adjustment seemed to produce the desired effects on the coefficients without unduly disturbing the significance of the discrimination. Otherwise, very little else was known about the effects of the adjustment on the discriminant models. The level of adjustment was determined purely by visual inspection of the 'ridge' trace. Values of K were selected when the unrealistically large coefficients began to stabilize at lower levels and before any noticeable, undesirable changes in the R^2 or χ^2 resulted.

To review in more detail, the data (used in the Vreuls and Wooldridge study) were gathered during an experiment, sponsored by the Naval Training Equipment Center (NTEC) to develop automated performance measurement for an instrument flight maneuvers simulator. This simulator, the Training Device Computer System (TRADEC) located at NTEC, was configured as a fixed-wing aircraft (F-4E). The TRADEC primary hardware included an XDS Sigma-7 computer and associated peripherals and an aircraft cockpit mounted on top of a motion platform. The cockpit contained all of the necessary flight controls and instrument displays found in a jet fighter front seat. A digital computer program provided the basic flight simulation.

Twelve relatively low-time students and private pilots were used as trainees. All participants had some familiarity with instrument flight, but were unskilled as jet fighter pilots. Each participant was trained to fly four basic instrument flight maneuvers, straight and level flight, standard rate climbs and descents, level turns and climbing and descending turns. Aircraft weight and resultant center-of-gravity (C.G.) shift, and turbulence were varied according to a predetermined schedule during training. Each pilot was trained on each experimental condition, for example, straight and level flight with fore C.G. and no turbulence, 12 times during a training day. For any particular condition, 12 trainees would provide 144 observations on any single training day.

It was assumed that after seven days of training, the pilots would become relatively proficient at the basic maneuvers. Therefore, a comparison of performance differences between day 1 and

day 7 would reveal those measures which were sensitive to the skill change that occurred during training. Discriminant analysis was selected as a method to highlight these performance differences.

During the execution of each exercise, performance data were collected on several continuous variables. The transforms or summary statistics performed on these system variables constitute the performance measures to be analyzed. For example, the 15 candidate measures for the level turn exercises are contained in table 3. Level turns required the trainee to hold a constant bank angle (30 degrees) while holding the aircraft altitude and airspeed constant.

The level turn exercise with no turbulence and fore C.G. was arbitrarily chosen from the instrument flight maneuvers syllabus to provide the example data used to develop a working version of the Monte Carlo simulation. After a complete standard analysis only six measures were found to have any significant contribution to the discriminant function. An exhaustive simulation of the 'ridge' adjusted discriminant models was then performed on these remaining measures. This was done by iteratively performing a discriminant analysis and incrementing the value of K between each iteration as previously described. Monte Carlo classification tests were performed on the discriminant functions resulting from 64 discrete values of K between 0 and 400. A cost function minimizing the sum of Type I and Type II error was recalculated for every case. Appendix A contains tables of a few of these discriminant analyses.

TABLE 3

CANDIDATE MEASURES FOR LEVEL TURNS

Meas. No.	Vari-ables	Desired Value	Trans-form	Abbrevia-tion in Analysis	Glossary	
1	ELVS	0	FLTR	ELF1	Elevator Stick	Crossover Power
2			AAE	ELF2		Average Displacement
3	ALPH	0	RNG	ALRG	Angle of Attack	Range
4			SDEV	ALSD		Standard Deviation
5	PTCH	0	SDEV	PTSD	Pitch Attitude	Standard Deviation
6	AILS		FLTR	AIF1	Aileron Stick	Crossover Power
7			AAE	AIF2		Average Displacement
8	ROLL	30[1]	AAE	ROAA	Roll Attitude	Average Absolute Error
9	PED	0	FLTR	PDF1	Rudder Pedal	Crossover Power
10			AAE	PDF2		Average Displacement
11	BETA	0	PNG	BERG	Side Slip	Range
12			RMS	BERM		Root-Mean-Squared Error
13	A/S	350	AAE	ASAA	Airspeed	Average Absolute Error
14	ALT	25000	AAE	HAA	Altitude	Average Absolute Error
15	THRR	0	RNG	THRG	Right Throttle	Range

Segmentation Rules: Start Meas. at Beginning of Trial (Speed Brake In).
Stop Meas. at End of Trial (45^{0} Heading Change).

[1] One-half of the trials were right (+) 30 degree bank turns, the other half were left (-) 30 degree bank turns.

The trace of the coefficients for values of K between 0 and 3 is shown in figure 9. Most of the change in the magnitude of the discriminant coefficients was found where K was less than 3 in this case. The increments in K used for the analysis in this region were .05 in order to maximize the resolution where the most activity occurs. Larger increments, ranging between 0.1, 0.5, 1.0, 5.0, 10.0, and 50.0 were used as K became larger and the differential impact on the discriminant analyses became more miniscule. Inspection of the coefficient trace shows a definite effect brought about by the addition of a small amount of bias, K.

It is not until one observes the change in the simulated percent of group misclassifications caused by differing values of K, that the benefit of the 'ridge' adjustment becomes apparent. Figure 10 shows the average total percent of Monte Carlo group membership misclassifications for values of K between 0 and 300. With the exception of minor 1 to 3 percent fluctuations, an obvious trend appears. The misclassification error for the standard discriminant model is 35.2 percent, but decreases to a minimum 15.8 percent when K is 1.8. It is evident that a global minimum occurs between the extreme values of K used in this series of simulations producing a classical 'ridge' trace. As predicted, though, the effectiveness of the discriminant model is improved with values of K of far greater magnitude than those usual for 'ridge' regression.

This was only one opportunity to use the simulation feature coupled with the 'ridge' discriminant analysis. Thus it constitutes only a limited demonstration of a developing performance measurement

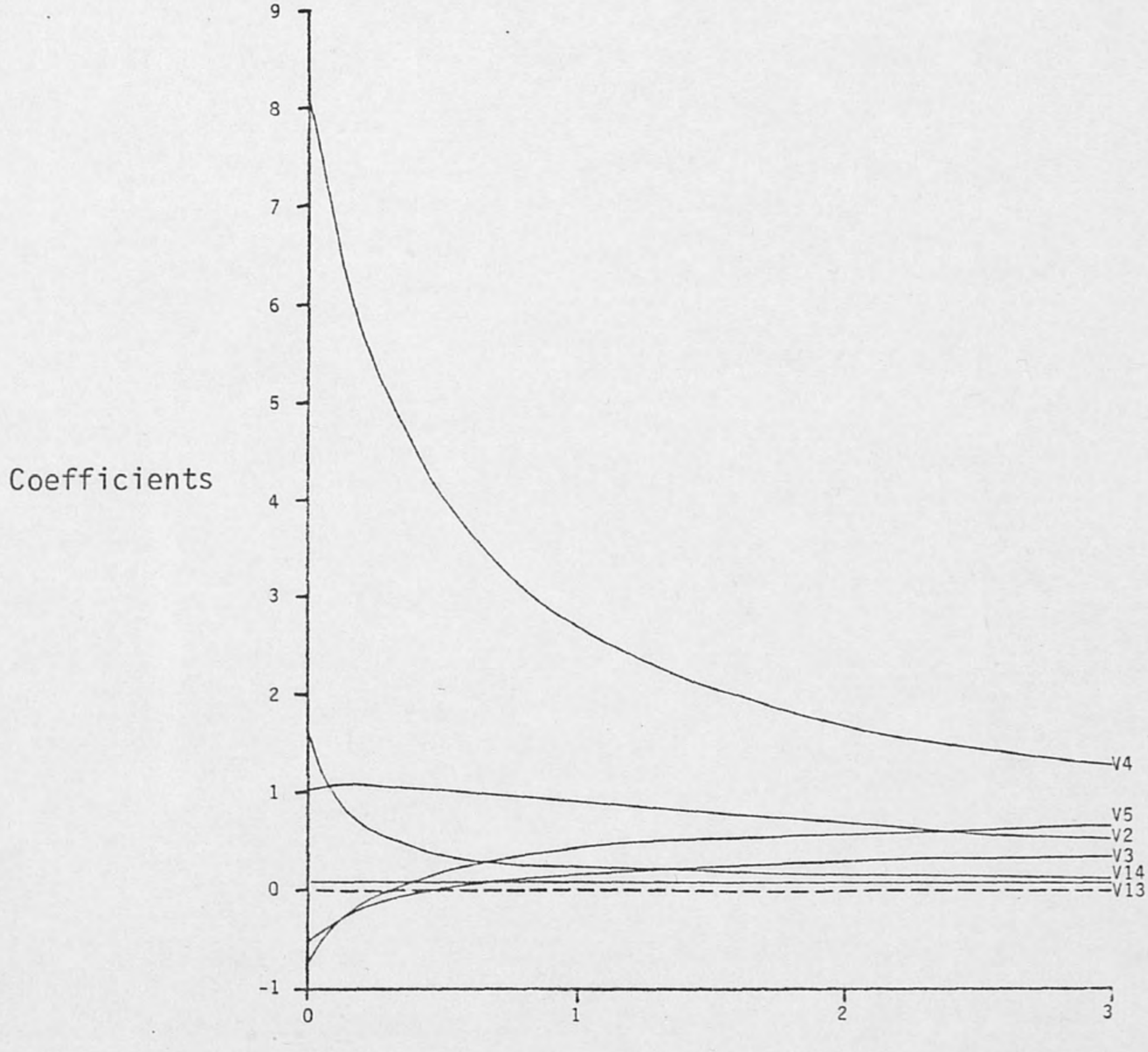

Figure 9. Coefficient Trace for Instrument Flight Maneuver Example

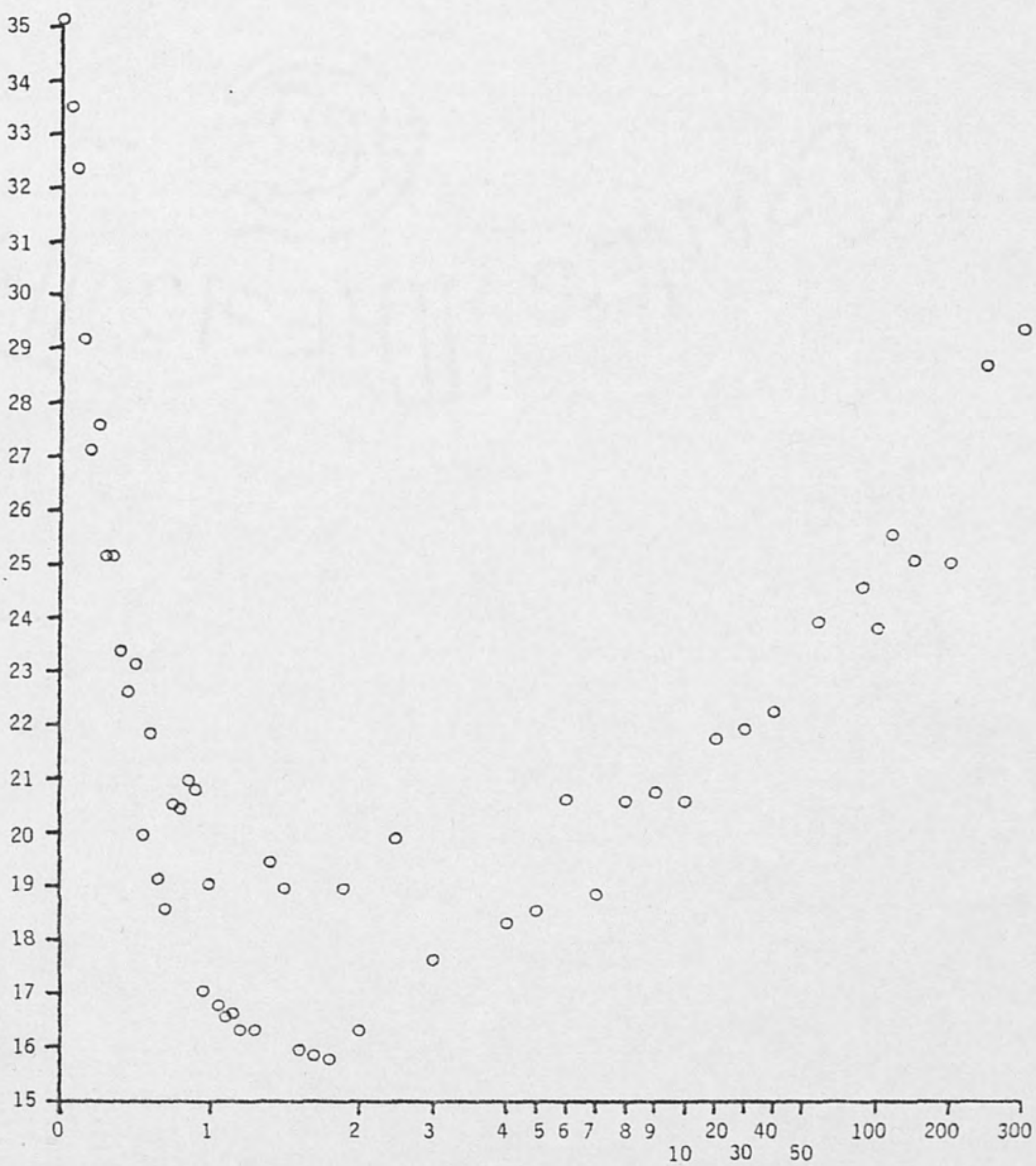

Figure 10. Total Percent Misclassifications for Increasing Values of K for the Instrument Flight Maneuvers Example

modeling philosophy. All of the measures had relatively high communalities and there was plenty of room for improvement on the existing discriminant function. Hence, inadvertently, there were "ideal" conditions under which to affect an improvement. Another data set was needed to further demonstrate the procedure and to replicate its successful application under entirely different conditions.

A more recent study provided these data with an entirely different complexion. A preliminary analysis showed that the standard discriminant model with 13 variables resulted in a very small classification error of 9.6 percent. The set of measures was composed of a mixture of those with very high communalities and a few with low communalities and questionable contribution to the discriminant model. This was the extreme opposite from the previous situation. Would the addition bias to the analysis immediately detract from the models discriminatory ability or is there still room for improvement?

A Naval Training Equipment Center and Air Force Human Resources Laboratory jointly sponsored research study, entitled 'Air Combat Maneuvering Performance Measurement' provided this opportunity to analyze 'fresh' data. This study was to develop an approach to Air Combat Maneuvering Performance Measurement for the assessment of one-on-one free engagements on the Simulator for Air-to-Air Combat (SAAC) at Luke AFB, Arizona (Kelly, Wooldridge, et al, 1979).

The SAAC is composed of two F-4 cockpits controlled by a common computer system. Each cockpit has a 296° horizonal and 150° vertical field-of-view system. For this study the visual system displayed a

computer generated scene around each cockpit which included a sun, sky, horizon, and 'checker-board' ground. The opponent's aircraft was a computer controlled aircraft model electronically superimposed into each pilot's visual scene which realistically presented the opponent's flight activity. Acceleration cues were only provided by an inflatable G-seat and G-suit.

Thirty F-4 pilots were divided into three groups of ten pilots each, representing three experience levels ranging from novice to expert. Six pilots (two from each experience group) 'flew' each week for five consecutive weeks. Each of the pilots flew against the other five pilots three times that week. Three initial starting positions were used for each pair of pilots which gave neither pilot an advantage. Various other engagements were flown to comprise the total, but they will not be discussed, as they were not used in any formal analysis to date.

Twenty-eight candidate performance measures were developed that could be collected automatically during ACM engagements on the SAAC. These were measures of aircraft maneuvering, control activity, energy management, engagement outcome, and relative aircraft position. Three variables were added for analysis purposes to account for any learning effect and any difference between the three set-up conditions. The final set of 31 variables are listed in table 4.

A two-group analysis was desired; one that would simply delineate between the lesser skilled pilots and those more competent. As was mentioned, each pilot in this study fell into one of three categories, each representing a different level of skill or experience.

TABLE 4

WHOLE ENGAGEMENT CANDIDATE MEASURES FOR MULTIVARIATE ANALYSES

Meas. No.	Variable	Trans-form	Units	Abbre-viation
1	Altitude Rate	AAV[1]	Ft./Sec.	ARAA
2	Current Out of View	%[2]	--	OOVP
3	IAS Range (Maximum-Minimum)	AVG[3]	Knots	ASRA
4	Speed Brake Position	AVG	--	SBPA
5	Fuel Flow	AVG	Lbs. Hr.	FFA
6	Altitude Rate S.D. Ratio (p/o)[5]	---	--	ARSR
7	Energy Management Index	RMS[4]	--	EMRM
8	Offense Time	%	--	OFFP
9	Offensive with Advantage[6]	%	--	SADP
10	Throt 1[7]	AVG[7]	--	TH1A
11	Throt 2	AVG	--	TH2A
12	Throt 3	AVG	--	TH3A
13	Throt 4	AVG	--	TH4A
14	Hdg Rate	RMS	Deg./Sec.	HDGRM
15	Hdg Rate	AAV	Deg./Sec.	HDGRAA
16	Time in a Lead Pursuit Position	%	--	LEADP
17	In Range	%	--	IRP
18	Roll Rate	AAV	Deg./Sec.	RRAA
19	Roll Rate X Alt Rate	AAV	Deg. Ft./Sec.	RRARA
20	Plane of Action[8]	AAV[8]	Deg.	PLOAAA
21	Defensive with Disadvantage[9]	%	--	DDP
22	AOA>28^{o}	%	--	A28P
23	AIM-9 Success	---	--	AIM9P
24	Initialization 2 (Negative State for Initialization 1)	---	--	INIT2

TABLE 4-Continued

Meas. No.	Variable	Trans-form	Units	Abbre-viation
25	Gun Kill Success	---	--	GUN
26	Ground Kill Success	---	--	GND
27	g's	---	--	G
28	Bingo Kill Success	---	--	BING
29	Leading/Offense[10]	---	--	L/O
30	Sequence Number[11]	---	--	SEQN
31	Initialization 3 (Negative State for Initialization 1)	---	--	INIT3

[1]AAV = Absolute Average
[2]% = Percent of engagement time
[3]AVG = Average
[4]RMS = Root-Mean-Square
[5]The ratio of proponent standard deviation to opponent standard deviation
[6]When the line of sight to the target is $<60^{\circ}$ and aspect is $<90^{\circ}$
[7]The number of transitions into each throttle zone divided by the number of data samples in each engagement
[8]The average absolute maneuvering plane of each cockpit
[9]When the line of sight $>120^{\circ}$, aspect $>90^{\circ}$, and range <4000 ft.
[10]Lead pursuit time divided by offensive time
[11]A count of successive free engagements for each pilot throughout Day 2 and Day 3

A rationale was determined for parsing the pilots into two distinct and justifiable groups. It was felt that the least experienced pilots, those that fell into category one, would always be doing something during any engagement that would be characteristic of their capability, regardless of the engagement outcome. Category three pilots, the most experienced in the study, would likewise exhibit their more advanced skills, regardless of the opponent or engagement outcome. Preliminary analyses also showed that the category two group of pilots formed a distinct intermediate group somewhere in between the two other groups. Therefore, all category one engagements, and all category three engagements, regardless of the opponents category, represented the extreme skill levels with the most promising capability to highlight differing ACM performance characteristics.

All of the data from category one and three pilots were then sorted into one of two respective groups for analysis. To keep things simple, the category two proponent data were excluded from these two groups, since it was not known precisely what part of the skill spectrum they represented. Of course, their effect would show up as an opponent's influence whenever another pilot from the other two categories engaged them. The results of this sort were 130 group one, or novice observations and 130 group two, expert observations of 31 performance measures each.

Several performance measures were eliminated on the basis of their low communalities, and, therefore, lack of contribution to the discriminant model. Some measures, such as 'Bingo Kill Success' or

'Ground Kill Success', were removed for being ultimate engagement outcomes, as the final model was to measure intermediate control or strategy skill devoid of ultimate criteria.

The same iterative 'ridge' adjusted discriminant analysis was performed on the remaining 13 measures. Appendix B contains tables of a few of these discriminant analyses.

The trace of the coefficients for values of K between 0 and 3 is shown in figure 11. As before, most of the change in the coefficients occurred in this range. There were 31 analyses performed with K ranging from 0 to 10. Examination of the trace indicates a somewhat less orderly change in the coefficients than in the previous example. There exist discontinuities and/or several points of inflection on each trace.

Figure 12 shows that these traits are reflected in the total percent misclassifications for increasing values of K. Most of the points plotted between 0.0 and 0.4 can be seen to be slight improvements over the original 9.6 percent. After 0.35 there is general trend upwards with a few exceptions in either direction. The entire trend is hardly as clear-cut as in the previous example. An improvement of over 1.5 percent was gained with a very small (0.3) 'ridge' adjustment, whereas a slightly higher bias produced a less satisfactory result.

In this example, great care would be needed to select an appropriate value for K. There was no indication in the usual discriminant statistics of the rapid breakdown in discriminability evidenced by the simulation.

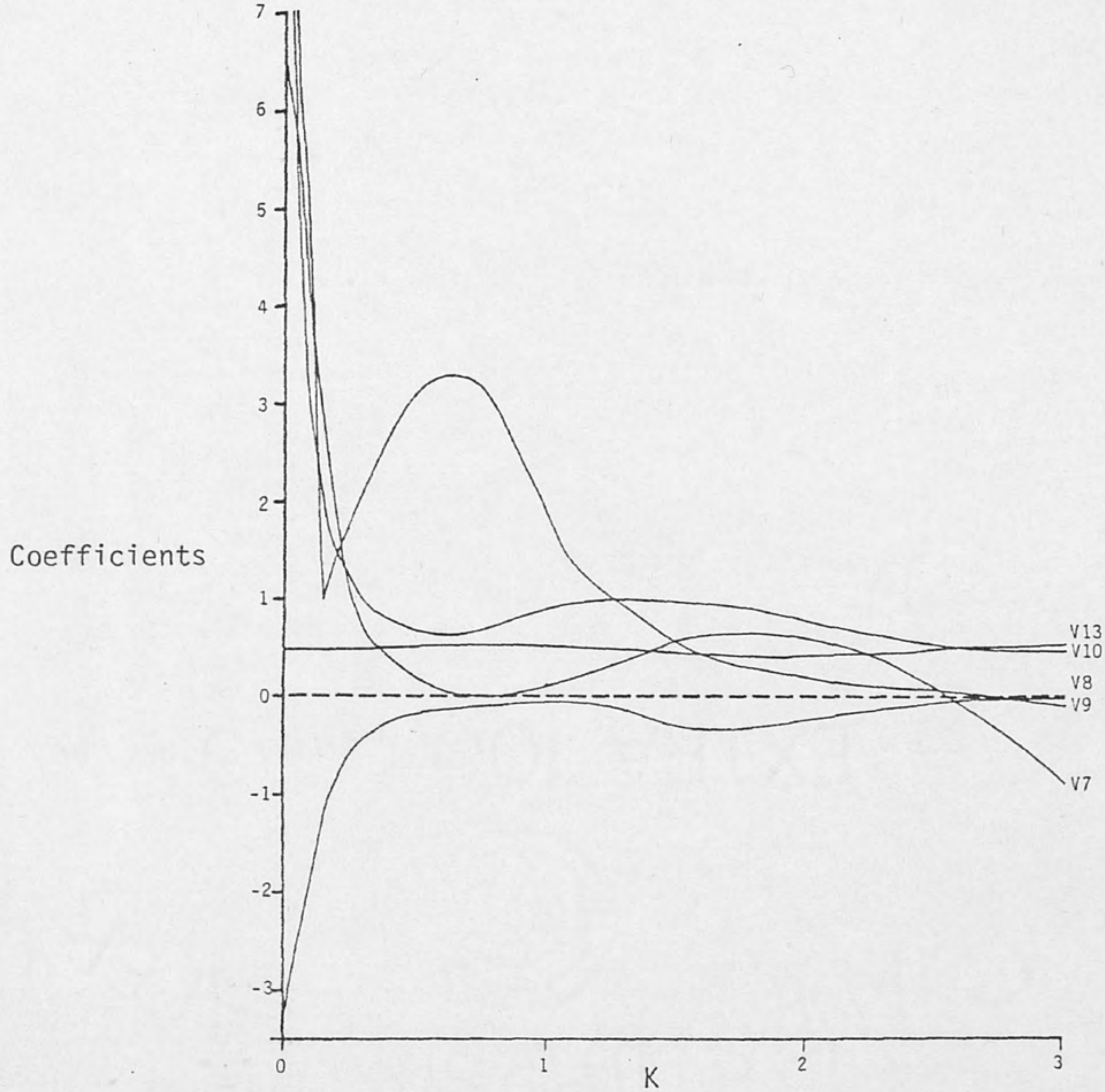

Figure 11. Coefficient Trace for Selected Measures for ACM Example

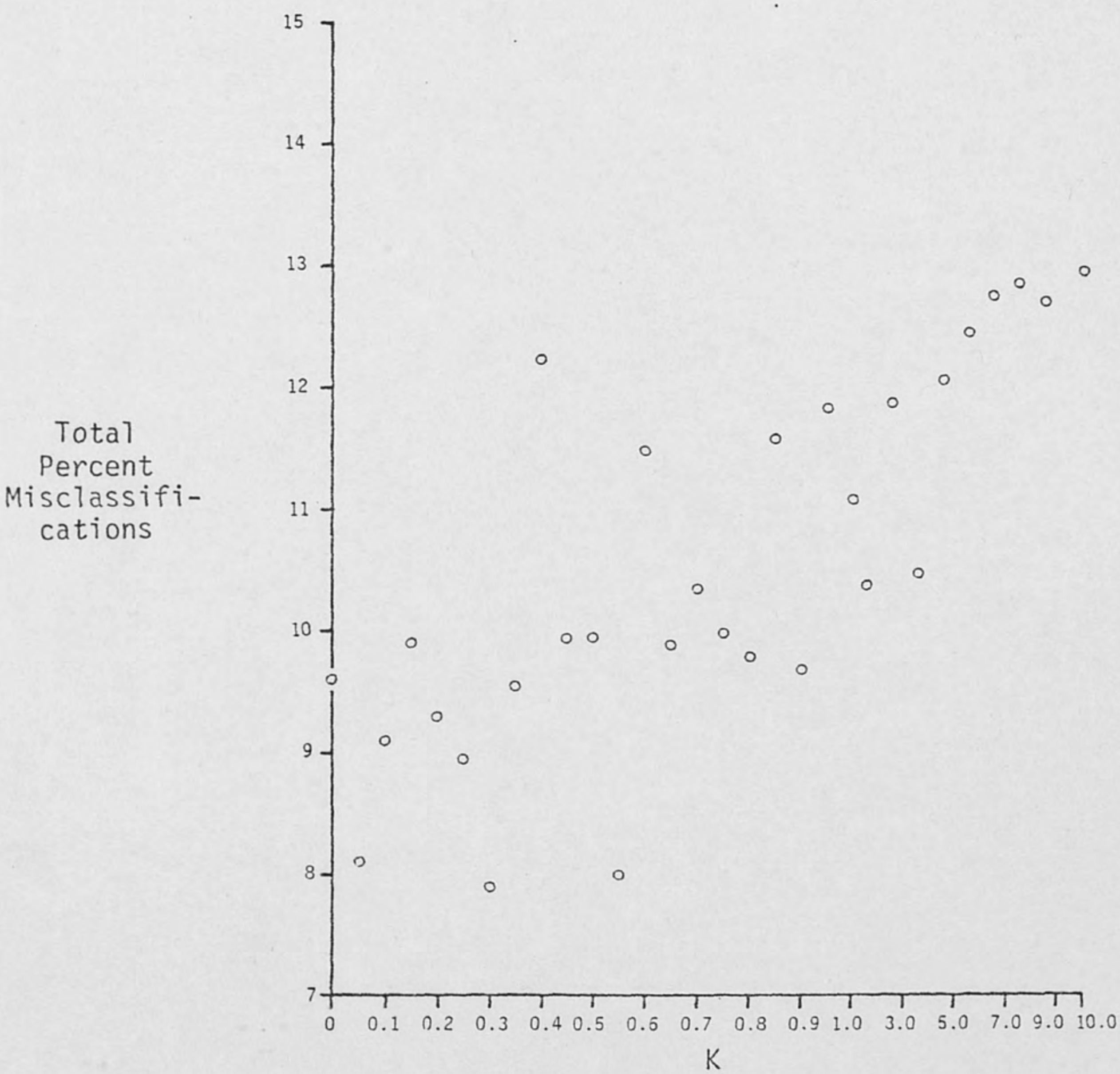

Figure 12. Total Percent Misclassifications for Increasing Values of K for the ACM Example

SECTION V

CONCLUSIONS

Circumstances surrounding performance measurement research underway at the NTEC Human Factors Laboratory necessitated the invention of 'ridge' adjusted discriminant analysis. For lack of rigorous mathematical proof, this new technique was applied to research only on the basis of educated supposition and positive empirical evidence. A Monte Carlo simulation was devised to provide a more thorough demonstration of the relative effectiveness of these biased discriminant models.

Two characteristically different sets of data have been subjected to a comprehensive 'ridge' discriminant analysis. Monte Carlo simulations were used to estimate the percentage of expected misclassifications for each level of adjustment. Both series of analyses exhibited fundamentally similar results. In both cases there were some coefficients which shrunk in magnitude and even a few that changed sign. Thus, indicating a reduction of 'overfit'. The simulations provided further evidence of improvements in the discriminant functions for certain levels of bias (K).

There were notable dissimilarities in the results between the two examples. The first example, the IFM data, exhibited a smoothly changing set of coefficients as a function of K. The simulation demonstrated a highly robust and unmistakable trend with a minimum

error that has approximately 17 percent better than the unbiased discriminant model.

On the other hand, the ACM data, produced a coefficient trace with discontinuities and greater variability. The outcome of the simulation reflected these less desirable traits with noticeable variability and only spurious improvements in the percentage of misclassifications. Several features of the ACM data may explain these differences. Examination of Appendix B shows that a few of the variables included to satisfy practical criteria added little to the discriminant model, as evidenced by their relatively small communalities. These measures may have contributed undesirable elements of variability. Two of the measures, INIT2 and GUN, were discrete, coded dummy variables which may have negatively effected the discriminant analysis as well as the simulation. There was no facility for the generation of binomial distributions in the simulation program. Such a departure from the multivariate normal distribution assumed by discriminant analysis would also more than likely have negative effects on the resulting discriminant functions. Lastly, since there was less than a 10 percent overlap of the two distributions in discriminant space, any simulation would have been dramatically effected by the ability to simulate the 'tails' of the two distributions. Even though generating a thousand data points, using the extremes of the distributions may have taxed the capability of the simulation. Combinations of any or all of the above explanations would be enough to cause the 1 to 3 percent variability in the simulation results with small changes in K.

In general, the simulation had only the capability to generate multivariate normal distributions. The data being simulated was not so restricted. Thus, the Monte Carlo misclassifications are only proportioned to the actual expected misclassifications had the model been validated with empirical data. It is also not clear which direction these estimates are biased. The generalizability of the simulation would be enhanced by the ability to simulate distributions of various characteristics at will. Inclusion of a Pearson distribution technique (Thomas 1966) which generates distributions, with the first four moments specifiable, might provide this added flexibility.

The simulation, as implemented, did provide a relative metric for choosing one bias level over another. This capability is advantageous over educated guess work and should probably be incorporated as a standard feature in future discriminant analyses. The 'ridge' discriminant analysis, through the use of the simulation feature, demonstrated a controllable improvement in discrimination over the standard analysis.

APPENDIX A

DISCRIMINANT MODELS

AND

ASSOCIATED STATISTICS

FOR

INSTRUMENT FLIGHT

MANEUVERS DATA

(First Example)

TABLE 5

DISCRIMINANT MODEL COEFFICIENTS, COMMUNALITIES, R^2, χ^2, Λ, AND PERCENT MISCLASSIFICATIONS FOR INCREASING VALUES OF K USING INSTRUMENT FLIGHT MANEUVERS DATA

Variable Number	Variable Name	Key Name	K=0.0		K=0.05		K=0.10		K=0.20		K=0.50	
			COEF(I)	COMM(I)	COEF(I)	COMM(I)	COEF(I)	COMM(I)	COEF(I)	COMM(I)	COEF(I)	COMM(I)
2	ELF2	V2	1.026	.622	1.056	.394	1.070	.399	1.073	.404	1.014	.405
3	ALRG	V3	-.495	.615	-.402	.630	-.325	.642	-.203	.659	.014	.683
4	ALSD	V4	8.234	.854	7.470	.856	6.836	.856	5.847	.854	4.099	.842
5	PTSD	V5	-.680	.706	-.499	.724	-.357	.737	-.148	.756	.199	.783
13	ASAA	V13	.107	.485	.107	.497	.107	.506	.106	.520	.104	.539
14	HAA	V14	1.617	.574	1.220	.582	.973	.589	.688	.600	.376	.617
	R^2		.460		.454		.449		.442		.429	
	χ^2		149		146		144		.141		136	
	Λ		.54		.55		.55		.56		.57	
Percent Misclassifications		Group 1	34.7		32.0		30.3		24.3		19.6	
		Group 2	35.7		35.0		34.3		30.2		26.8	
		Total	35.2		33.5		32.3		27.2		23.2	

TABLE 5-Continued

Variable Number	Variable Name	Key Name	K=1.0		K=1.55		K=2.0		K=3.0		K=5.0	
			COEF(I)	COMM(I)	COEF(I)	COMM(I)	COEF(I)	COMM(I)	COEF(I)	COMM(I)	COEF(I)	COMM(I)
2	ELF2	V2	.887	.396	.774	.386	.701	.380	.584	.369	.447	.354
3	ALRG	V3	.184	.698	.277	.705	.324	.708	.388	.712	.454	.715
4	ALSD	V4	2.774	.826	2.077	.815	1.739	.809	1.301	.799	.905	.786
5	PTSD	V5	.448	.796	.572	.799	.628	.799	.693	.795	.732	.786
13	ASAA	V13	.103	.552	.103	.559	.103	.563	.104	.570	.107	.581
14	HAA	V14	.239	.626	.189	.629	.168	.628	.143	.626	.120	.620
	R^2		.418		.412		.409		.404		.397	
	χ^2		131		129		127		125		122	
	Λ		.58		.59		.59		.60		.60	
Percent Misclassifications		Group 1	16.7		12.2		12.8		14.3		16.5	
		Group 2	23.1		19.6		20.5		21.1		20.8	
		Total	19.9		15.9		16.6		17.7		18.6	

TABLE 5-Continued

Variable Number	Variable Name	Key Name	K=10.0		K=20.0		K=40.0		K=80.0		K=150	
			COEF(I)	COMM(I)	COEF(I)	COMM(I)	COEF(I)	COMM(I)	COEF(I)	COMM(I)	COEF(I)	COMM(I)
2	ELF2	V2	.301	.336	.201	.317	.137	.294	.094	.265	.065	.232
3	ALRG	V3	.522	.716	.563	.704	.564	.669	.511	.597	.420	.506
4	ALSD	V4	.570	.765	.379	.733	.265	.681	.188	.601	.135	.508
5	PTSD	V5	.715	.762	.638	.726	.529	.677	.410	.611	.306	.539
13	ASAA	V13	.114	.604	.126	.642	.146	.701	.172	.781	.199	.859
14	HAA	V14	.097	.605	.077	.585	.060	.559	.044	.527	.032	.490
	R^2		.386		.371		.349		.321		.291	
	χ^2		118		112		104		94		83	
	Λ		.61		.63		.65		.68		.71	
Percent Misclassifications		Group 1	21.8		20.8		22.7		24.4		21.7	
		Group 2	19.6		22.9		21.9		25.1		28.7	
		Total	20.7		21.8		22.3		24.7		25.2	

TABLE 5-Continued

Variable Number	Variable Name	Key Name	K=300 COEF(I)	K=300 COMM(I)	K=400 COEF(I)	K=400 COMM(I)
2	ELF2	V2	.042	.195	.035	.181
3	ALRG	V3	.304	.401	.259	.364
4	ALSD	V4	.089	.406	.074	.369
5	PTSD	V5	.208	.457	.174	.427
13	ASAA	V13	.225	.925	.233	.944
14	HAA	V14	.022	.443	.018	.425
	R^2		.257		.243	
	χ^2		72		67	
	Λ		.74		.76	
Percent Misclassifications		Group 1	29.0		28.3	
		Group 2	30.1		35.3	
		Total	29.5		31.8	

APPENDIX B

DISCRIMINANT MODELS
AND
ASSOCIATED STATISTICS
FOR
AIR COMBAT
MANEUVERS DATA
(Second Example)

TABLE 6

DISCRIMINANT MODEL COEFFICIENTS, COMMUNALITIES, R^2, χ^2, Λ, AND PERCENT MISCLASSIFICATIONS FOR INCREASING VALUES OF K USING AIR COMBAT MANEUVERS DATA

Variable Number	Variable Name	Key Name	K=0.0		K=0.05		K=0.10		K=0.20		K=0.30	
			COEF(I)	COMM(I)	COEF(I)	COMM(I)	COEF(I)	COMM(I)	COEF(I)	COMM(I)	COEF(I)	COMM(I)
1	ARAA	V1	- .00448	.219	- .00493	.235	- .00518	.235	-.00539	.242	- .00529	.244
2	OOVP	V2	- .36752	.399	- .39692	.437	- .38777	.421	-.41900	.460	- .41236	.454
4	SBPA	V3	.06642	.157	.06209	.168	.05900	.168	.05555	.172	.05923	.174
5	FFA	V4	- .00006	.241	- .00006	.260	- .00006	.260	-.00006	.267	- .00006	.268
7	EMRM	V5	.00985	.416	.01091	.448	.01126	.447	.01223	.461	.01213	.465
8	OFFP	V6	.76394	.502	1.16190	.543	1.22550	.509	1.61332	.562	1.42345	.559
10	TH1A	V7	6.57811	.370	5.71979	.386	3.94562	.376	1.71096	.374	2.26921	.382
12	TH3A	V8	-3.33729	.186	-3.16944	.194	-2.35618	.198	-.93464	.198	-1.21061	.197
13	TH4A	V9	7.93816	.396	6.64023	.409	4.54950	.392	1.86919	.390	2.68429	.399
17	IRP	V10	7.71285	.486	5.04694	.387	2.99058	.301	1.37972	.258	3.05413	.314
18	RRAA	V11	.00305	.011	.00787	.011	.01079	.011	.01389	.011	.01131	.012
24	INIT2	V12	.64878	.067	.58708	.066	.67910	.076	.53560	.065	.55259	.065
25	GUN	V13	.48438	.064	.52789	.064	1.17136	.138	.88203	.096	.62540	.071
	R^2		.533		.523		.524		.515		.514	
	χ^2		167		162		162		158		158	
	Λ		.47		.48		.48		.48		.49	
Percent Misclassifications		Group 1	12.9		8.8		10.8		6.2		10.4	
		Group 2	6.3		7.4		7.4		12.4		5.4	
		Total	9.6		8.1		9.1		9.3		7.9	

TABLE 6-Continued

Variable Number	Variable Name	Key Name	K=0.4 COEF(I)	K=0.4 COMM(I)	K=0.5 COEF(I)	K=0.5 COMM(I)	K=1.0 COEF(I)	K=1.0 COMM(I)	K=1.5 COEF(I)	K=1.5 COMM(I)	K=2.5 COEF(I)	K=2.5 COMM(I)
1	ARAA	V1	-.00534	.235	-.00557	.246	-.00566	.250	-.00580	.260	-.00594	.263
2	OOVP	V2	-.39288	.424	-.42231	.461	-.41444	.447	-.42442	.449	-.41318	.436
4	SBPA	V3	.05630	.167	.05450	.175	.05515	.178	.05410	.184	.05438	.187
5	FFA	V4	-.00006	.259	-.00006	.272	-.00006	.277	-.00006	.287	-.00006	.290
7	EMRM	V5	.01187	.447	.01283	.470	.01305	.477	.01360	.496	.01386	.501
8	OFFP	V6	1.4021	.509	1.68472	.559	1.58624	.538	1.58550	.541	1.54639	.518
10	TH1A	V7	1.28620	.364	.36438	.374	.12633	.380	.51686	.395	.06730	.397
12	TH3A	V8	-.58346	.198	-.27356	.198	-.01843	.201	-.30167	.202	-.04919	.204
13	TH4A	V9	1.92499	.378	.30995	.386	.19050	.389	.59169	.400	.06450	.398
17	IRP	V10	1.43987	.248	.67867	.232	.80047	.229	.98149	.234	.51714	.212
18	RRAA	V11	.01330	.011	.01511	.012	.01490	.012	.01519	.012	.01517	.012
24	INIT2	V12	.64030	.072	.52097	.064	.55478	.067	.46795	.057	.52350	.063
25	GUN	V13	1.34647	.161	.84108	.089	.89153	.095	.62314	.066	.73840	.076
	R^2		.521		.509		.505		.495		.491	
	χ^2		161		155		154		149		148	
	Λ		.48		.49		.49		.51		.66	
Percent Misclassifications		Group 1	5.6		6.6		5.8		6.3		6.2	
		Group 2	18.9		13.3		16.4		14.5		17.6	
		Total	12.1		9.9		11.1		10.4		11.9	

TABLE 6-Continued

Variable Number	Variable Name	Key Name	K=5.5 COEF(I)	K=5.5 COMM(I)	K=10.0 COEF(I)	K=10.0 COMM(I)
1	ARAA	V1	-.00636	.283	-.00677	.301
2	OOVP	V2	-.40128	.405	-.39331	.367
4	SBPA	V3	.05344	.201	.05219	.213
5	FFA	V4	-.00007	.312	-.00008	.331
7	EMRM	V5	.01495	.539	.01616	.573
8	OFFP	V6	1.34798	.467	1.14908	.407
10	TH1A	V7	.00068	.426	.07591	.453
12	TH3A	V8	-.01054	.210	-.04624	.215
13	TH4A	V9	-.00789	.412	.08613	.422
17	IRP	V10	.25281	.190	.26147	.167
18	RRAA	V11	.01663	.013	.01732	.014
24	INIT2	V12	.52428	.056	.46207	.049
25	GUN	V13	.71399	.059	.53424	.052
	R^2		.473		.452	
	χ^2		140		131	
	Λ		.53		.55	
Percent Misclassifications		Group 1	7.0		8.5	
		Group 2	18.0		17.5	
		Total	12.5		13.0	

BIBLIOGRAPHY

Anderson, T. W. An Introduction to Multivariate Statistical Analysis. New York: John Wiley & Sons, 1958.

Anderson, D. A., and Carney, E. J. Ridge Regression Estimation Procedures Applied to Canonical Correlation Analysis. BU-509-M. Ithica, New York: Biometrics Unit, Cornell University, 1974.

Banerjee, K. S., and Carr, R. N. "A Comment on Ridge Regression." Technometrics 12 (1970): 895-898.

Bittner, A. C., Jr. Computerized Accommodated Percentage Evaluation (CAPE) Model for Cockpit Analysis and Other Exclusion Studies. TP-75-49. Point Mugu, California: Naval Missile Test Center, 1975.

Bittner, A. C., Jr. Reduction in Potential User Population as the Result of Imposed Anthropometric Limits: Monte Carlo Estimation. TP-74-6. Point Mugu, California: Naval Missile Test Center, 1974.

Box, G. E. P., and Miller, M. E. "A Note on Generation of Random Normal Deviates." Annals of Mathematical Statistics 29 (1958): 610.

Cooley, W. W., and Lohnes, P. R. Multivariate Data Analysis. New York: John Wiley & Sons, 1971.

Farebrother, R. W. "The Minimum Mean-Square-Error Linear Estimator and Ridge Regression." Technometrics 17 (1975): 127-128.

Gordon, G. System Simulation. Englewood Cliffs, N. J.: Prentice Hall, 1969.

Hoerl, A. E., and Kennard, R. W. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." Technometrics 12 (1970): 55-67.

Hoerl, A. E., and Kennard, R. W. "Ridge Regression: Applications to Nonorthogonal Problems." Technometrics 12 (1970): 69-82.

Kelly, M. J., and Wooldridge, A. L. "Air Combat Maneuvering Performance Measurement." In Proceedings of the Human Factors Society 23rd Annual Meeting, Boston, Massachusetts, October 29-November 1, 1979, pp. 324-328. Edited by Carolyn K. Bensel. Santa Monica, California, 1979.

Kelly, M. J.; Wooldridge, A. L.; Hennessy, R. T.; Vreuls, D.; Barneby, S. F.; Cotton, J. C.; and Reed, J. C. Air Combat Maneuvering Performance Measurement. AFHRL-TR-79-3. Williams AFB, AZ: Flying Training Division, Air Force Human Resources Laboratory, September 1979.

Lane, N. E. The Influence of Selected Factors on Shrinkage and Overfit in Multiple Correlation. Monograph 17. Pensacola, FL: Naval Aerospace Medical Research Laboratory, Naval Aerospace Medical Institute, 1971.

Lindley, D. V., and Smith, A. F. M. "Bayse Estimates for the Linear Model." Journal of Royal Statistical Society, B 34 (1972): 1-18.

McDonald, G. C., and Schwing, R. C. "Instabilities of Regression Estimates Relating Air Pollution to Mortality." Technometrics 15 (1973): 463-481.

Mallows, C. L. "Some Comments on Cp." Technometrics 16 (1973): 225-236.

Marquardt, D. W. "Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation." Technometrics 12 (1970): 591-613.

Obenchain, R. L. "Ridge Analysis Following a Preliminary Test of the Shrunken Hypothesis." Technometrics 17 (1975): 431-441.

Rozeboom, W. W. "Ridge Regression: Bonanza or Beguilment?" Psychological Bulletin 86 (1979): 242-249.

Simon, C. W. Methods for Improving Information from 'Undesigned' Human Factors Experiments. TRP75-287. Culver City, CA: Hughes Aircraft Company, 1975.

Scheuer, E. M., and Stoller, D. S. On the Generation of Normal Random Vectors. Journal of the American Statistical Association (1960): 278-281.

Tatsuoka, M. M. Discriminant Analysis-The Study of Group Differences. Champaign, IL: The Institute for Personality and Ability Testing, 1970.

Thomas, D. G. Computer Methods for Generating Pseudo-Random Numbers from Pearson Distributions. Blacksburg, VA: Virginia Polytechnic, May 1966.

Vreuls, D.; Obermayer, R.; Goldstein, I.; and Lauber, J. Measurement of Trainee Performance in a Captive Rotary-Wing Device. 71-C-0194-1. Orlando, FL: U. S. Naval Training Equipment Center, July 1973.

Vreuls, D., and Obermayer, R. W. Study of Crew Performance Measurement for High-Performance Aircraft Weapon System Training: Air-to-Air Intercept. NAVTRADEVCEN 70-C-0059-1. Orlando, FL: U. S. Naval Training Device Center, 1971.

Vreuls, D.; Obermayer, R.; and Goldstein, I. Trainee Performance Measurement Development Using Multivariate Techniques. NAVTRAEQUIPCEN 73-C-0066-1. Orlando, FL: U. S. Naval Training Equipment Center, September 1974.

Vreuls, D.; Wooldridge, L.; Obermayer, R.; Johnson, R.; Norman, D.; and Goldstein, I. Development and Evaluation of Trainee Performance Measures in an Automated Instrument Flight Maneuvers Trainer. NAVTRAEQUIPCEN 74-C-0063-1. Orlando, FL: U. S. Naval Training Equipment Center, October 1976.

Vreuls, D., and Wooldridge, A. L. "Aircrew Performance Measurement." In Proceedings of the Symposium on Productivity Enhancement: Personnel Performance Assessment in Navy Systems, San Diego, CA, October 1977, pp. 5-31. Edited by Louis T. Pope, and David Meister. San Diego, CA: Navy Personnel Research and Development Center, 1977.

Wichern, D. W., and Churchill, D. W. "A Comparison of Ridge Estimates." Technometrics 20 (1978): 301-311.

Wooldridge, A. L. Air Combat Maneuvering: Whole Mission Success Factors. Orlando, FL: Canyon Research Group, Inc., In Press.

Wooldridge, A. L.; Breaux, R.; and Weinman, D. Statistical Issues in the Use of Multivariate Methods for Selection of Flight Simulator Performance Measures. NAVTRAEQUIPCEN 75-C-0091-1. Orlando, FL: U. S. Naval Training Equipment Center, In Press.

LIST OF REFERENCES

Anderson, T. W. An Introduction to Multivariate Statistical Analysis. New York: John Wiley & Sons, 1958.

Bittner, A. C., Jr. Reduction in Potential User Population as the Result of Imposed Anthropometric Limits: Monte Carlo Estimation. TP-74-6. Point Mugu, California: Naval Missile Test Center, 1974.

Box, G. E. P., and Miller, M. E. "A Note on Generation of Random Normal Deviates." Annals of Mathematical Statistics 29 (1958): 610.

Cooley, W. W., and Lohnes, P. R. Multivariate Data Analysis. New York: John Wiley & Sons, 1971.

Farebrother, R. W. "The Minimum Mean-Square-Error Linear Estimator and Ridge Regression." Technometrics 17 (1975): 127-128.

Gordon, G. System Simulation. Englewood Cliffs, N. J.: Prentice Hall, 1969.

Hoerl, A. E., and Kennard, R. W. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." Technometrics 12 (1970): 55-67.

Hoerl, A. E., and Kennard, R. W. "Ridge Regression: Applications to Nonorthogonal Problems." Technometrics 12 (1970): 69-82.

Kelly, M. J.; Wooldridge, A. L.; Hennessy, R. T.; Vreuls, D.; Barneby, S. F.; Cotton, J. C.; and Reed, J. C. Air Combat Maneuvering Performance Measurement. AFHRL-TR-79-3. Williams AFB, AZ: Flying Training Division, Air Force Human Resources Laboratory, September 1979.

Lindley, D. V., and Smith, A. F. M. "Bayse Estimates for the Linear Model." Journal of Royal Statistical Society, B 34 (1972): 1-18.

McDonald, G. C., and Schwing, R. C. "Instabilities of Regression Estimates Relating Air Pollution to Mortality." Technometrics 15 (1973): 463-481.

Mallows, C. L. "Some Comments on Cp." Technometrics 16 (1973): 225-236.

Simon, C. W. Methods for Improving Information from 'Undesigned' Human Factors Experiments. TRP75-287. Culver City, CA: Hughes Aircraft Company, 1975.

Scheuer, E. M., and Stoller, D. S. On the Generation of Normal Random Vectors. Journal of the American Statistical Association (1960): 278-281.

Tatsuoka, M. M. Discriminant Analysis-The Study of Group Differences. Champaign, IL: The Institute for Personality and Ability Testing, 1970.

Thomas, D. G. Computer Methods for Generating Pseudo-Random Numbers from Pearson Distributions. Blacksburg, VA: Virginia Polytechnic, May 1966.

Vreuls, D.; Obermayer, R.; Goldstein, I.; and Lauber, J. Measurement of Trainee Performance in a Captive Rotary-Wing Device. NAVTRAEQUIPCEN 71-C-0194-1. Orlando, FL: U. S. Naval Training Equipment Center, July 1973.

Vreuls, D.; Obermayer, R.; and Goldstein, I. Trainee Performance Measurement Development Using Multivariate Techniques. NAVTRAEQUIPCEN 73-C-0066-1. Orlando, FL: U. S. Naval Training Equipment Center, September 1974.

Vreuls, D.; Wooldridge, L.; Obermayer, R.; Johnson, R.; Norman, D.; and Goldstein, I. Development and Evaluation of Trainee Performance Measures in an Automated Instrument Flight Maneuvers Trainer. NAVTRAEQUIPCEN 74-C-0063-1. Orlando, FL: U. S. Naval Training Equipment Center, October 1976.

Vreuls, D., and Wooldridge, A. L. "Aircrew Performance Measurement." In Proceedings of the Symposium on Productivity Enhancement: Personnel Performance Assessment in Navy Systems, San Diego, CA, October 1977, pp. 5-31. Edited by Louis T. Pope, and David Meister. San Diego, CA: Navy Personnel Research and Development Center, 1977.