DOCUMENT RESUME

ED 113 377 TM 004 874

| | |
|---|---|
| AUTHOR | Simon, Charles W. |
| TITLE | Methods for Improving Information from "Undesigned" Human Factors Experiments. Technical Report No. p75-287. |
| INSTITUTION | Hughes Aircraft Co., Culver City, Calif. |
| SPONS AGENCY | Air Force Office of Scientific Research, Arlington, Va. |
| REPORT NO | TR-P75-287 |
| PUB DATE | Jul 75 |
| NOTE | 91p. |
| EDRS PRICE | MF-$0.76 HC-$4.43 Plus Postage |
| DESCRIPTORS | Bias; Computer Programs; Correlation; Data Analysis; *Experiments; Mathematical Applications; Matrices; Military Training; *Multiple Regression Analysis; Orthogonal Rotation; Performance; *Predictor Variables; *Research Design; *Research Methodology; Research Problems |
| IDENTIFIERS | Least Squares Analysis; Ridge Regression Analysis; Variance (Statistical) |

ABSTRACT

An "undesigned" experiment is one in which the predictor variables are correlated, either due to a failure to complete a design or because the investigator was unable to select or control relevant experimental conditions. The traditional method of analyzing this class of experiment--multiple regression analysis based on a least squares criterion--gives rise to a number of interpretation problems when the effects of individual predictors are to be assessed. Some difficulties and their effects on the quality of information are discussed. Two methods are described in this report for improving the information obtained from the undesigned human factors experiment. One is to collect more information at a few data points selected at locations that improve the orthogonality of this non-orthogonal design. The other is to use a ridge regression analysis in place of the conventional least squares analysis, in which a slight bias is introduced into the data in such a way that the combined bias and variance error is smaller than the variance error of unbiased estimates from the least squares analysis. The ridge analysis produces more stable and meaningful regression coefficients. Computational aids--both references and complete computer programs--are supplied. (Author)

TECHNICAL REPORT NO. P75-287

Prepared under contract for the
Air Force Office of Scientific Research,
Air Force Systems Command,
United States Air Force.



# METHODS FOR IMPROVING INFORMATION FROM "UNDESIGNED" HUMAN FACTORS EXPERIMENTS

Charles W. Simon



AEROSPACE GROUPS

HUGHES

DISPLAY SYSTEMS AND
HUMAN FACTORS DEPT.
HUGHES AIRCRAFT COMPANY
CULVER CITY, CALIFORNIA

JULY 1975

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|
| 1. REPORT NUMBER: P75-287 | 2. GOVT ACCESSION NO. / 3. RECIPIENT'S CATALOG NUMBER: |
| 4. TITLE (and Subtitle): Methods for Improving Information from "Undesigned" Human Factors Experiments | 5. TYPE OF REPORT & PERIOD COVERED: Technical Report July 1974 to July 1975 |
| | 6. PERFORMING ORG. REPORT NUMBER: P75-287 |
| 7. AUTHOR(s): Charles W. Simon | 8. CONTRACT OR GRANT NUMBER(s): F44620-72-C-0086 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS: Hughes Aircraft Company, Centinela and Teale, Culver City, California 90230 | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research, 1400 Wilson Blvd., Arlington, Va. 22209 | 12. REPORT DATE: July 1975 |
| | 13. NUMBER OF PAGES: 82 |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) | 15. SECURITY CLASS. (of this report): Unclassified |
| | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Distribution list included at end of report.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

"Undesigned" experiment
Non-orthogonal designs
Correlated predictors
Multiple regression
Ridge regression analysis
Repairing ill-conditioned matrices
Least squares analysis

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

An "undesigned" experiment is one in which the predictor variables are correlated, either due to a failure to complete a design or because the investigator was unable to select or control relevant experimental conditions. The traditional method of analyzing this class of experiment -- multiple regression analysis based on a least squares criterion -- gives rise to a number of interpretation problems when the effects of individual predictors are to be assessed. Some difficulties and their effects on the quality of information are discussed.

(Continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (Continued)

Two methods are described in this report for improving the information obtained from the undesigned human factors experiment. One is to collect more information at a few data points selected at locations that improve the orthogonality of this non-orthogonal design. The other is to use a ridge regression analysis in place of the conventional least squares analysis, in which a slight bias is introduced into the data in such a way that the combined bias and variance error is smaller than the variance error of unbiased estimates from the least squares analysis. The ridge analysis produces more stable and meaningful regression coefficients. Computational aids -- both references and complete computer programs -- are supplied.

# METHODS FOR IMPROVING INFORMATION FROM "UNDESIGNED" HUMAN FACTORS EXPERIMENTS

Charles W. Simon
Hughes Aircraft Company

Technical Report No. P75-287

Prepared under contract with the Air Force Office of Scientific Research, (AFSC), United States Air Force.

July 1975

AEROSPACE GROUPS

Hughes Aircraft Company, Culver City, California

## ABSTRACT

An "undesigned" experiment is one in which the predictor variables are correlated, either due to a failure to complete a design or because the investigator was unable to select or control relevant experimental conditions. The traditional method of analyzing this class of experiment -- multiple regression analysis based on a least squares criterion -- gives rise to a number of interpretation problems when the effects of individual predictors are to be assessed. Some difficulties and their effects on the quality of information are discussed.

Two methods are described in this report for improving the information obtained from the undesigned human factors experiment. One is to collect more information at a few data points selected at locations that improve the orthogonality of this non-orthogonal design. The other is to use a ridge regression analysis in place of the conventional least squares analysis, in which a slight bias is introduced into the data in such a way that the combined bias and variance error is smaller than the variance error of the unbiased estimates from the least squares analysis. The ridge analysis produces more stable and meaningful regression coefficients. Computational aids -- both references and complete computer programs -- are supplied.

## FOREWORD

While the "undesigned" experiment is used extensively in personnel selection research, it has been virtually ignored as a viable approach in equipment design and training research. Traditionally, in these latter problem areas, systematic designs have been used in which the primary experimental variables are all controlled. As a result, variables that are difficult or impossible to control are often excluded from the experimental plan even when they are relevant and have an important effect on performance. Consequently, much of the performance variability in the experiment remains unexplained and the data is of limited value when applied to real-world problems.

Unmanageable sources of variance, however, can be accounted for if they are treated as variables of an "undesigned" experiment. Thus a most effective use of the methods described in this report to enhance "undesigned" experiments is to combine them with the "advanced methodologies" described in previous reports (e.g., "Economical Multifactor Designs" and "Methods of Handling Sequence Effects..." (Simon, 1973; 1974)). By properly using these methods in combination, we become capable of doing experiments that will account for most of the variance associated with the performance of a real-world task and to eliminate major sources of irrelevant variance.

I would be interested in hearing about applications of these techniques by behavioral scientists and am willing to discuss efforts in this regard. Comments and criticisms are always welcomed.

Charles W. Simon
1975

## ACKNOWLEDGMENTS

This paper was prepared as a part of a research project conducted at Hughes Aircraft Company, Culver City, California, under Contract No. F44620-72-C-0086 with the Air Force Office of Scientific Research, Air Force Systems Command, United States Air Force. Dr. Charles E. Hutchinson was technical monitor for AFOSR.

The special help given by the following people in the preparation of this report is gratefully acknowledged:

- Edward J. Dragavon, who spent considerable time interpreting and explaining the mathematics involved in the different techniques.
- Rose E. Konrath and Masse Bloomfield of the Hughes Aircraft Company library, who facilitated the collection of the original papers reviewed in this report.
- Mary G. Gallegos, who prepared the complete ridge regression program.
- Maryann Hill of the Health Sciences Computing Facility at the University of California, Los Angeles, who gave permission to use her article on ridge regression analysis.
- Dr. Robert G. Demaree of Texas Christian University, Ft. Worth, Texas, who helped clarify questions regarding ridge regression analysis.
- Paula Lintz, who made valuable suggestions in the editing of this report.

# TABLE OF CONTENTS

## SYMBOLOGY

| | |
|---|---|
| $A^{-1}$ | Inversion of matrix A where $A = X'X$ |
| $b_i$ | Regression coefficient of variable i for raw score data |
| $\beta$ | True beta (standard regression) coefficient |
| $\hat{\beta}$ | Estimated beta (standard regression) coefficient from least squares analysis |
| $\hat{\beta}^*$ | Biased estimator of ridge coefficient from ridge regression analysis |
| $E$ | Statistical expectation; weighted integral of --- |
| I | Identity matrix: In a correlation matrix, all diagonal values equal one and off-diagonal values equal zero |
| k | Constant used to distort correlation matrix in ridge regression analysis |
| $\lambda$ | Eigenvalue |
| $L_i^2$ | Squared distance between true and estimated coefficients |
| $\Pi$ | Product of --- |
| $\sigma^2$ | Residual error variance; sigma squared, mean square error |
| $\sum_1^p$ | Sum of --- (from 1 to p items) |
| $V(\hat{y})$ | Variance of an estimated response |
| $x_o'$ | Transpose of vector $x_o$ (experimental condition) |
| $X_i$ | Predictor variable i |
| X'X or (X'X) | Sum of squares and cross-product matrix (of value of every predictor multiplied by one another yielding $X_i^2$ and $X_iX_j$ values); x vector multiplied by its transpose; square root of elements of X'X divided by N equals correlation matrix |
| $\lvert X'X \rvert$ | Determinant of the (X'X) matrix |
| X'Y | Cross-product between predictor and performance |
| $\hat{Y}$ or $\overline{Y}$ | Estimated performance (from regression equation) |

# SECTION I
# INTRODUCTION

This report describes two methods of improving the information obtained from the "undesigned" experiment. In the first approach, additional data is collected in order to facilitate the interpretation of data already collected. The second approach is a relatively new technique of data analysis that provides better solutions than does the traditional least squares analysis.

## DEFINITION

An "undesigned" experiment is one in which some experimental variables cannot be or are not controlled by the experimenter. To be included in an experiment, therefore, the level of each variable must be known or measured at the time each performance measurement is made. Under these circumstances, variables in an undesigned experiment are correlated mathematically to some degree, a condition which markedly complicates the interpretation of the results.

## EXAMPLES OF UNDESIGNED EXPERIMENTS

The following fictitious situations are examples of undesigned experiments in human factors engineering research:

1. The Army has rewritten its maintenance manuals in a style that will enable the ordinary technician to understand and use the information better. They are interested in measuring the impact of this revision on system performance. Old and new manuals are made available at a number of maintenance depots where the technicians differ in training levels and experience with the particular equipment. At the depots, differences also exist in the availability of critical parts, the maintenance philosophy and schedules, the unit morale levels, and other factors that could conceivably affect the quality of maintenance. Since it is impossible to control these associated factors to any degree, a daily record is kept on each of them along with several

criteria of maintenance performance over a six-month period. This data taken as a whole can be treated as an undesigned experiment.

2. The Air Force wishes to determine the optimum parameters for the manual control configuration of a missile-delivery system. They wish to reach a solution derived from empirical data collected under operational conditions. A flight test is planned in which the strike accuracy of a dummy air-to-ground missile is to be studied as a function of changes in control parameters. There is little opportunity to make a great many flights to offset the effects of such uncontrolled but critical factors as visibility, turbulence, and variations in the target itself. However, these variables can be measured at the time each missile is fired. While the control parameters can be systematically varied, the existence of the other uncontrolled but presumably critical factors make this a partially undesigned experiment.

3. The Navy has built a research-oriented pilot-training simulator. A study is conducted to determine the least expensive simulator configuration that will result in the greatest transfer in pilot performance from simulator to aircraft. Two groups of pilots are selected for the study -- those with less than 2000 flying hours and those with more than 5000 flying hours. It is recognized that flying time per se is not sufficient to characterize pilot skill and that such things as the type of aircraft, the nature of the flying experience (military or civilian; war-time or peace-time), and recency of this experience also should be taken into consideration. Since it is necessary to use all available pilots as subjects without an opportunity to control these other factors, pilot characteristics must be included in the analysis and handled as variables of an undesigned experiment.

4. Over a twelve year period, a research organization has conducted experiments relating equipment parameters to success in acquiring ground targets on an airborne display. During this time the effects of over fifteen variables associated with the sensor, the display,

and the briefing information have been examined, but in a series of small experiments of two and three variables each. Since no overall research strategy was ever planned, the frequency with which certain variables and levels of variables occur in this data varies considerably. The resulting lack of a balanced design leaves predictor variables correlated. Thus this belated effort to combine the results of several experiments to develop a single prediction equation takes on the characteristics and problems of an undesigned experiment.

5. The levels of a factorial design are used as the data collection plan in a drug-therapy experiment. While the study is being run, it becomes apparent that two of the extreme conditions cannot be measured at all because they exceed physiological safety limits. This destroys the orthogonality of the design. The data that remains to be analyzed takes on the characteristics of an undesigned experiment.

## DESIGNED VERSUS UNDESIGNED EXPERIMENTS

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 independently 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 -- alone or in conjunction with a balanced design -- offers a viable alternative.

## Characteristics of Designed Experiments

The value of a designed experiment rests on the fact that the experimental conditions are selected in such a way that critical effects can be isolated and the interpretation of the results simplified. However, there is a price to be paid for these advantages, for the rigidity of the design forces the experimenter to:

- anticipate in advance the variables he will include in his study;
- be able to control the exact levels of any variables that will be included in the study;
- include conditions that may be unrealistic or otherwise undesirable.

## Positive Features of Undesigned Experiments

The undesigned experiment, because it generally accepts as the experimental conditions those which exist at the moment a performance measurement is made, does not face the same problems. The very lack of control of the conditions under which performance data must be acquired yields the following advantages for the undesigned experiment:*

1. The costs of collecting performance data are no longer as rigidly related to the number of factors being investigated. As many variables as desired can be considered as long as the level of each can be ascertained at the time performance data is being collected.
2. It is not always necessary to anticipate critical variables in advance of the data collection phase. If appropriate records are available, these may be used later to introduce more variables into the analysis.

*It is by definition that these advantages fall to the undesigned rather than the designed experiment. Obviously a number of these advantages could exist for experiments that are planned by an experimenter who intends to use some analysis of covariance design. However, to identify the class of problems that will benefit from the techniques discussed in this report, any situation in which variables are included in which the level selection is not under the investigator's complete control is considered an undesigned experiment.

3. Particularly in field studies, there is a greater likelihood that the majority of critical variables will be present (although not necessarily identified) and that the values that are used will be more representative of reality.

4. The regression approach permits an iteration in the search for the more critical variables. If the proportion of variance accounted for by the regression is low, other variables may be tried to see if they fit the data better, provided the necessary measures are available.

Thus the undesigned experiment has the advantage of allowing (or forcing) the experimenter to study the world as it really is. If the levels of experimental variables are not selected artificially but are allowed to vary naturally, the chances are higher that performance will be measured under more representative circumstances with the relevant and critical variables operating.

Measurement Sources. The only alternative to controlling the levels of variables to be included in an experiment is to measure their levels as they exist at the time performance is measured. The following are the most common sources from which these measurements can be obtained:

- Concurrent measurements. As an event unfolds and performance is measured, concomittant variables of importance are also measured. (Example: Measuring air turbulence in a flight test.)
- Historical measurements. The data is obtained from past records that can in some way be associated with the conditions occurring at the time performance is being measured. (Example: Using subject aptitude scores from tests taken prior to his entering the pilot training course.)
- Incomplete measurements. The levels of each variable are already known, having been assigned as levels of a designed experiment which became degraded when certain conditions were omitted by choice or by accident. (Example: A factorial design is planned and data is collected at all but two corner points when a data recorder failed to operate.)

In practice, all or some of these sources may be used in a single experiment.

## Difficulties with Undesigned Experiments

There are penalties, however, associated with the freedom of data collection for the undesigned experiment. The imbalance among combinations of variables that is bound to occur when no systematic experimental design is used leaves predictor variables correlated. As a result, the derived equations are subject to greater error and information becomes scrambled and difficult to isolate.

In the next section, some problems of interpreting the results of the undesigned experiment are described along with general concepts and terminology used in regression analysis that are useful when reading this report.

## SECTION II.

## ANALYSIS AND INTERPRETATION -- PROBLEMS, CONCEPTS, AND TERMINOLOGY

In this section, basic concepts and terminology relevant to multiple regression analysis will be reviewed and problems in interpreting the results from undesigned experiments identified. The discussion is simplistic and intended only to supply the minimum detail required for a reader to appreciate the value of the alternate techniques described in subsequent sections. For an in depth explanation of multiple regression analysis, the reader is encouraged to read the excellent books and papers that are available on this topic (e.g., Draper and Smith, 1968; Darlington, 1968; Kerlinger and Pedhazur, 1973).

### RAW DATA MATRICES

The experimental conditions and related performance in both designed and undesigned experiments can be organized into a matrix format such as shown in table T.1.

| Observation # | Subj # | Predictor Variables ($X_i$) X1 | X2 | X3 | X4 | X5 | XN | Obtained (Y) Performance |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0.23 | 1 | 25.6 | 0 | Hi | 1 | 27% |
| 2 | 14 | 0.11 | 1 | 7.8 | 0 | Med | 5 | 18% |
| 3 | 3 | 0.07 | 2 | 14.1 | 1 | Med | 3 | 87% |
| 4 | 8 | 0.54 | 1 | 3.0 | 0 | Hi | 4 | 52% |
| 5 | 11 | 0.27 | 2 | 22.9 | 0 | Lo | 1 | 15% |
| 6 | 4 | 0.33 | 1 | 15.6 | 0 | Med | 2 | 38% |
| 7 | 5 | 0.19 | 1 | 1 | 1 | Hi | 5 | 77% |

[T.1]

Each line represents one observation, i. e., the conditions, $X_i$, under which performance was measured and the performance, Y, that was obtained. There could of course be more than one performance measure for any observation*, e. g., speed and accuracy, and subject characteristics could be included as factors among the predictor variables.

A primary difference between raw data matrices of designed and undesigned experiments lies in the arrangement of the levels of the predictor variables. In the designed experiment, these levels, being systematically controlled by the experimenter, are generally selected in a balanced fashion so that the main effects of the predictor variables are orthogonal (i. e., uncorrelated). The factorial design is one of the more familiar examples in which the levels of each variable are combined equally often with every other variable to achieve this orthogonality. As a result, the analysis and interpretation of the results are simplified. In the undesigned experiment, this balance is not achieved because the experimenter is unable (or fails to) select or control the levels of the experimental conditions. As a result, main effects of predictor variables are correlated with one another, a condition that makes the analysis and interpretation of the results more difficult. This correlation is a mathematical dependence, a happenstance of the levels that occurred at the time the measures were taken, and does not necessarily imply a causal relationship between the pair of variables.

## CORRELATION MATRICES

The distinction between a designed and undesigned experiment is easier to illustrate if the raw data matrix is transformed into a correlation matrix composed of the linear (Pearson product moment) correlations among all variables.

---

*Conventional regression analysis handles only a single performance variable per analysis. Kerlinger and Pedhazur (1973, 376-381) describe a method of doing multivariate regression analysis with two dependent variables at the same time, illustrating how the combined analysis provides a clearer interpretation of the data than two analyses each with single and different dependent variables.

For a designed experiment, a fictitious correlation matrix for three predictor variables and a performance variable might look like table T.2.

| Linear Correlations | | Predictor Variables X1 | X2 | X3 | Performance (Y) |
|---|---|---|---|---|---|
| Predictor Variables | X1 | 1. | 0 | 0 | 0.342 |
| | X2 | 0 | 1. | 0 | -0.167 |
| | X3 | 0 | 0 | 1. | 0.523 |

[T.2]

The table of intercorrelations can be broken into two parts: one, the predictor matrix of correlations among each predictor variable and every predictor variable including itself, and two, the performance column vector of correlations between each predictor variable and performance.

Note that since each predictor variable correlates perfectly and positively with itself, the diagonal values are all one. Note further that with the designed experiment, all off-diagonal values are zero, showing that the linear components of the predictors are all orthogonal to one another. A matrix with only zeros off the diagonal is referred to as a diagonal matrix. When the numbers on the diagonal are all ones, the matrix is called a unit matrix.

In the undesigned experiment, the intercorrelation matrix for the predictor variables is not likely to have zero correlations in the off-diagonal positions. Instead, for the undesigned experiment, the correlation table might look like table T.3.

| Linear Correlations | | 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 |

[T.3]

When the off-diagonal elements are non-zero, the predictor variables are correlated. In that case, the matrix is said to be ill-conditioned and the original experimental design said to be non-orthogonal.

Note that in both tables T.2 and T.3, the predictor matrix is symmetrical about the diagonal. In some texts, only half of the matrix (above or below the diagonal) will be written out.

To be able to analyze this data by regression analysis, the matrix must be non-singular. This means that each row (or column) of the matrix must be linearly independent of every other row (or column). No row (or column) is produced from any linear combination of others in the matrix.

## MULTIPLE REGRESSION ANALYSIS

Given the information in raw data or correlational form, the investigator ordinarily subjects it to an analysis that reduces it to a linear polynomial equation that will provide the "best" estimate of performance under specific conditions of the predictor variables.

Each line of the raw data or correlation matrix represents an equation. Performing a multiple regression analysis on the data is the same as finding the common solution to a set of simultaneous equations.

The equation derived from an analysis of the raw data will be written in the following form:

$$b_0X_0 + b_1X_1 + b_2X_2 + \ldots b_NX_N = \overline{Y} \qquad [E.1]$$

where $b_0X_0$ is a constant and $b_i$ ($i$ = 1 through N) are regression coefficients, for the N independent variables, $X_i$ ($i$ = 1 through N), respectively. In practice, the $X_i$ terms can represent main effects or transgenerations of main effects, such as cross-products ($X_iX_j$) or higher order terms ($X_i^2$), each treated in the analysis as if it were another variable. A regression

coefficient, $b_i$, is the average change in performance that will occur for each unit change in the particular variable; this change may be positive or negative. The value, $\overline{Y}$, is the performance estimated by the equation for particular values of the predictor variables, $X_i$.

Least Squares Fit

The coefficients derived by multiple regression analysis are the ones used in the polynomial to provide the "best" fit of the data. The criterion for a "best" fit is met when the sum of the squares of the differences between the observed and the estimated performance values is at a minimum.* The difference between the observed and the estimated performance values is called the residual; thus the "best" fit is obtained from the equation that minimizes the residual sum of squares (RSS).

Standard Regression Equation

Variables are commonly measured in different units and on different scales. In order to compare the coefficients of these variables, the values in the raw data table can be converted to standard measures, or Z scores. This is done for each variable as follows:

$$\frac{X - M}{\sigma} = Z$$

where X is the raw score to be converted, M is the variable mean value, and $\sigma$ is the standard deviation. If these standard scores are subjected to a multiple regression analysis, then the resulting polynomial is referred to as a standard regression equation of the following form:

$$\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \ldots \beta_N Z_N = \overline{Z}_Y \qquad [E.2]$$

*There are other criteria for judging the merits of an equation. Kiefer (1959) discusses a number of these in detail. Later on in this paper, some weaknesses of a least squares solution of data obtained from non-orthogonal experimental designs will be discussed and alternative criteria proposed.

wherein the regression coefficients, $b_i$, of Equation E.1 are replaced by beta coefficients, $\beta_i$, and there is no longer a constant term. A regression analysis of the data in a correlation matrix results in a standard regression equation.

If either the ordinary regression equation or the standard regression has been calculated, the other can be derived according to the following relationships:

$$b_i = \beta_i \frac{\sigma Y}{\sigma X_i} \quad \text{or} \quad \beta_i = b_i \frac{\sigma X_i}{\sigma Y}$$

The constant, $b_0 X_0$, for the multiple regression equation is found as follows:

$$b_0 X_0 = Y_{Mean} - (b_1 X_{1Mean} + b_2 X_{2Mean} \ldots b_N X_{NMean})$$

Interpreting Multiple Regression Analysis

From his regression analysis, an investigator ordinarily is interested in obtaining the following information:

- An equation to be used to estimate performance at specific coordinates of the experimental space.
- Measures of the relative importance of the experimental variables.

This information is generally easy to obtain when orthogonal designs are used. However, this is not the case when results from undesigned (non-orthogonal) experiments are to be interpreted. Let us examine both cases.

Orthogonal Designs. The equation derived from a multiple regression analysis might appear as in the case of this fictitious example:

$$\overline{Y} = 0.45 + 1.53 X_1 - 8.49 X_2 + 0.67 X_3$$

where $\overline{Y}$ is the estimated number of targets found as a function of $X_1$, dynamic range of the display in log foot lamberts; $X_2$, sensor resolution in 10-foot units; and $X_3$, display size (diameter) in inches. With an orthogonal design, the coefficients in the equation can be interpreted as follows. Each time the dynamic range on the display increased one log foot-lambert, 1.53 more targets were found on the average. Each time, the resolution of the sensor was increased by 10-feet, 8.49 fewer targets on the average were found. Each time the diameter of the display was increased an inch, 0.67 more targets were found on the average. It is understood, as in all regression analyses, that these relationships hold only within the boundaries set by the data collection points in the original experiment.

In equipment design problems, this information may be enough to compare the relative importance of the different predictor variables. Since the levels of all three variables can be converted to a common scale of engineering costs (to achieve a particular resolution, dynamic range, or size), no refinement of the equation is actually required to decide their relative importance in the applied situation. This is not often the case in other fields of psychology where no common base among variables exists. Thus it would be quite difficult to know which contributed more to success in school by studying the raw score coefficients in an equation that relates school success to scores on a reading test and a math test. There is no common base to work with. In that case, to compare the relative importance of the individual variables on the performance, the predictor variables must be changed to standard scores and the equation written in a standard regression form.

For example, a multiple regression analysis of the correlation table, T.2, would yield the following standard regression equation:

$$Y_z = 0.342 Z_1 - 0.167 Z_2 + 0.523 Z_3$$

The effect of Variable 1 on performance is twice as large as the effect of Variable 2 (in terms of their standard scores, Z). However, it is questionable whether or not the use of standard coefficients is as meaningful for an applied

problem that can be related to a common cost scale as the coefficients of the ordinary regression equation would be.

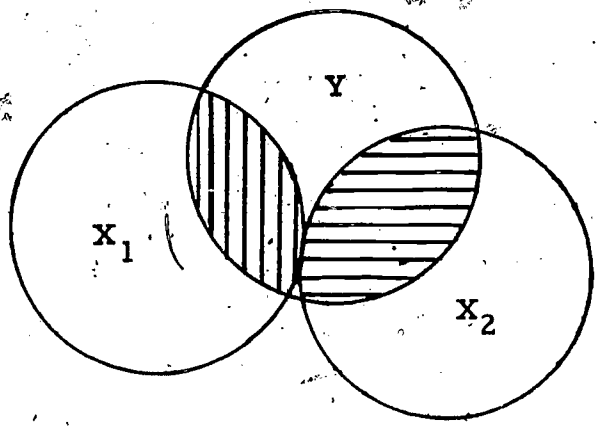
Calculations, however, can be made from the coefficients of the standard regression equation that might add to the understanding of the data. When derived from fully orthogonal designs,

1. These coefficients are the same as the linear correlation between each predictor and performance, as seen in the XY column of the table of intercorrelation.
2. The square of each coefficient shows the proportion of the total variability in performance that each predictor accounts for.
3. The sum of the squared coefficients shows the proportion of the total performance variability that can be explained by the total standard regression equation, and one minus that value shows the proportion that is not explained.

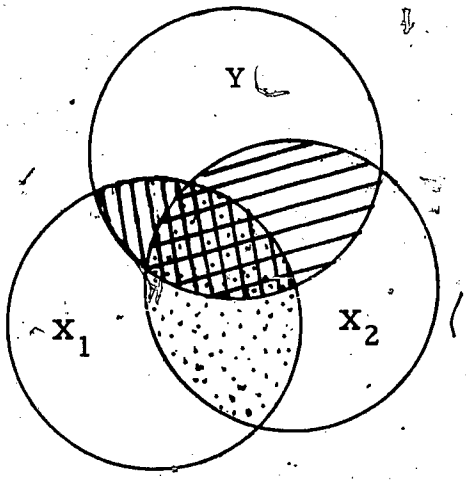
Non-Orthogonal Designs. While regression equations from undesigned (predictor-correlated) experiments are mathematically the same as those from orthogonally designed experiments, pragmatically they are not. Although in both cases the overall equation does represent the best fit of the data (according to the least squares criterion), in the case of data from undesigned experiments, the beta coefficients of individual terms should not be considered independently. However tempting it may be to do so, when predictor variables are markedly correlated, the beta coefficients should not be individually interpreted to show the relative importance of the variables. The relative magnitude of these coefficients are the result in part of arbitrary decisions made by the investigator during the analysis. This can best be explained by example.

In figure F.1, two factors, $X_1$ and $X_2$ account for 25 and 36 percent, respectively, of the total variability in performance (Y). A standard regression equation based on these two factors alone would be:

$$\overline{Z}_Y = 0.5\, Z_{X1} + 0.6\, Z_{X2}$$




PROPORTION OF PERFORMANCE VARIANCE, Y, ACCOUNTED FOR BY TWO INDEPENDENT PREDICTOR VARIABLES, $X_1$ AND $X_2$ [F.1]




PROPORTION OF PERFORMANCE VARIANCE, Y, ACCOUNTED FOR BY TWO CORRELATED PREDICTOR VARIABLES, $X_1$ AND $X_2$ [F.2]

Interpretation in this case is straight-forward. The relative contributions of each variable can be estimated; thirty-nine percent of the performance variability is still left unexplained.

In figure F.2, the two factors, $X_1$ and $X_2$ again overlap Y by 25 and 36 percent, respectively. This time, however, they are also correlated 0.60 with one another. It is no longer a simple matter to decide how much

of an effect each variable has on performance. Where $X_1$ overlaps $X_2$ and Y, how can one determine whether the effect on Y is due to $X_1$ or $X_2$ If the effect on Y in the overlap portion is due to $X_1$, then $X_2$ does not have as much of an effect as the simple correlation between $X_2$ and Y suggests. If the effect on Y in the overlap portion is actually due to $X_2$, then $X_1$ does not have as much effect as its correlation with Y suggests. Because the data itself does not directly suggest which alternative is correct, using regression analysis on data with correlated predictors can give a number of solutions, depending on the order in which variables are introduced into the analysis.

In the above example, if the effect of $X_1$ (including the $X_1X_2$ overlap) were removed first, only 14 percent of $X_2$ would be left (excluding the X1X2 overlap) to affect Y. In that case, the equation would be written:

$$\overline{Z}_Y = 0.5\, Z_{X1} + 0.375\, Z_{X2}$$

On the other hand, had the full effect of $X_2$ been removed first, then the effect of $X_1$ that remained after taking into consideration the X1X2 overlap would have been reduced and the equation would have been:

$$\overline{Z}_Y = 0.6\, Z_{X1} + 0.175\, Z_{X2}$$

Both equations would estimate $\overline{Z}_Y$ equally well, each accounting for 0.39 of the total variance. In both equations, the first beta coefficient corresponds to the full correlation between that variable and performance; the second beta coefficient, however, corresponds to a semi-partial correlation after the effect of all prior variables has been removed from the predictor under consideration. As the number (N) of correlated variables increases, the number of ways in which they can be ordered into the equation (N!) illustrates the numerous solutions that are possible and why interpreting the individual coefficients is a meaningless exercise.

Because of this, there are some like Darlington (1968, p. 169), who after reviewing the problem at some length, concludes: "It would be better to simply concede that the notion of 'independent contribution to variance' has no meaning when predictor variables are intercorrelated."

Eigenvalues

Given a correlation matrix A, such as table T.3 (or any real symmetric matrix), there exists a set of eigenvalues, $\lambda$, such that:

$$\left|\lambda I - A\right| = 0$$

For a four-variable study, the determinant of the correlation ($a_{ij}$) matrix in the above expression could be written as follows:

$$\begin{vmatrix} \lambda - a_{11} & -a_{12} & -a_{13} & -a_{14} \\ -a_{12} & \lambda - a_{22} & -a_{23} & -a_{24} \\ -a_{13} & -a_{23} & \lambda - a_{33} & -a_{34} \\ -a_{14} & -a_{24} & -a_{34} & \lambda - a_{44} \end{vmatrix} = 0$$

The expansion of this determinant yields a polynomial $\phi(\lambda)$ of a degree n in $\lambda$ which is known as the characteristic polynomial of the matrix A. The equation, $\phi(\lambda) = 0$, is called the characteristic equation of A and its roots $\lambda_1, \lambda_2, \ldots \lambda_{N=4}$ are called the characteristic roots (or eigenvalues) of A. For the purposes of this report it is not necessary for the reader to understand the mathematics required to calculate eigenvalues since even for a matrix of modest size, a computer would be required to perform the calculations. It is important though that the reader be aware of some of the ways they can be used to facilitate the interpretation of data from the undesigned experiment.

The set of eigenvalues, $\lambda_i$, for an orthogonal matrix (e. g., T.2) would all be equal to one. This should be obvious from the above explanation, since an orthogonal correlation matrix is a unit matrix with all ones in the diagonal which would yield a determinant equal to zero only if the ($\lambda$I) were also all ones.

For a non-orthogonal matrix, however, the eigenvalues are no longer either equal or necessarily one. Instead, some of them are larger than one and some smaller than one. The more non-orthogonal the design matrix, the greater the range of values. For example, the eigenvalues for a fictitious moderately non-orthogonal design of eight variables might be as follows:

$$\begin{aligned} \lambda_1 &= 1.55 \\ \lambda_2 &= 1.36 \\ \lambda_3 &= 1.15 \\ \lambda_4 &= 1.03 \\ \lambda_5 &= 0.97 \\ \lambda_6 &= 0.85 \\ \lambda_7 &= 0.64 \\ \lambda_8 &= 0.45 \end{aligned} \qquad [T.4]$$

while the eigenvalues for a fictitious more severely non-orthogonal design of eight variables might be as follows:

$$\begin{aligned} \lambda_1 &= 3.22 \\ \lambda_2 &= 2.18 \\ \lambda_3 &= 1.30 \\ \lambda_4 &= 0.74 \\ \lambda_5 &= 0.31 \\ \lambda_6 &= 0.18 \\ \lambda_7 &= 0.05 \\ \lambda_8 &= 0.02 \end{aligned} \qquad [T.5]$$

Note how the range has increased in the second case, T.5, and how small some of the eigenvalues are. Both sets sum to 8.00.

Given the set of eigenvalues for a matrix, however, an investigator can use them as a means of better understanding his data. The following applications can be made.

The sum of the eigenvalues will always equal N, the number of predictor variables in the experiment, whether the matrix be from a designed or undesigned experiment. However, with undesigned experiments, since some of the eigenvalues can be very small fractions, less than the total number of eigenvalues may be needed to almost sum to N. This can provide a clue as to how many critical variables are actually influencing performance. For example, in the set of eigenvalues for a severely non-orthogonal design, T.5, 99 percent of the variation is explained by the first six eigenvalues. Although the fact that no eigenvalue is zero indicates that all eight variables have some effect on performance, but for all practical purposes, only six are probably really critical. This could be important to know if an investigator wished to eliminate some of the terms in an equation. (Note: There is no one-to-one relationship between the numerical ordering of eigenvalues and variables.)

The sum of the reciprocals of the eigenvalues is an indication of the degree of matrix non-orthogonality. This value for a completely orthogonal design, of course, equals to the number of predictors, N. The more correlated the predictor variables, however, the smaller some eigenvalues will become and therefore the larger the sum of their reciprocals. This sum divided by N shows how many times greater the squared distance is between sample (estimated) and population (true) beta coefficients for the non-orthogonal design than it would have been for an orthogonal design.

The product of the eigenvalues equals the determinant of the matrix. The larger the determinant (up to N for an orthogonal matrix), the more orthogonal the design. Later in this report, the determinant will be used as a criterion for selecting the coordinates of data collection points, which when added to the conditions of an undesigned experiment, will make it more orthogonal.

## IMPROVED METHODS OF HANDLING DATA FROM NON-ORTHOGONAL DESIGNS

To extract the most information from the undesigned experiment, full advantage must be taken of any technique that can offset the problems associated with this class of experiment. In the last two sections of this report, two approaches will be described that are superior to the more conventional techniques in popular use today. These approaches involve:

- Collecting additional data at specific coordinates of the experimental space to improve the orthogonality of the design.
- Using "ridge regression" analysis to provide more stable and meaningful regression coefficients with which to fit the data from non-orthogonal experimental designs.

Conceptually, these techniques are relatively easy to understand. Implementing them, however, will require the talents of the investigator, a computer programmer, and possibly a statistician. In all cases the only practical way in which these techniques will be employed is with the aid of a high-speed computer. In the body of this report, no detailed discussion of the computations required for the analyses will be given. However, in the appendices both general and specific references regarding the computational efforts are supplied along with listings of complete programs. When these are not sufficient, the reader is encouraged to refer to the original papers.

## SECTION III

## COLLECTING ADDITIONAL DATA TO ORTHOGONALIZE THE UNDESIGNED EXPERIMENT

The non-orthogonality of the undesigned experiment complicates the interpretation of results. In this section, methods of collecting additional data that will alleviate this situation are proposed. Specifically, information will be provided here to tell the reader:

- How to select the coordinates of new data points that will improve the orthogonality of the original design.
- How to handle irrelevant shifts in performance that may occur between the time when data is collected on original and subsequent runs.

### PRACTICAL CONSIDERATIONS

Since most undesigned experiments are those in which the experimenter has little or no control over the levels of his variables, it may appear presumptuous to suggest an approach that requires just such control. The point in fact is that there are circumstances when this approach can be used and an investigator should be aware that such an approach exists and be prepared to use it should the occasion arise. Sometimes, if only a few additional points are needed, an investigator can make a concerted effort to set up the required conditions in a way that would not be justified for an entire experiment. At other times, once the principles involved in adding points are understood, experimental conditions that are not located optimally can be considered which will still improve the orthogonality of the design and the interpretability of the data. All in all, the knowledge of how to properly add data points is a useful experimental tool that has applications beyond the immediate problem.*

*Other useful applications of these techniques for the design of experiments are cited in Appendix E.

Other factors that must be taken into consideration before this technique is employed include:

- Computer facilities must be available because of the amount and complexity of the computations required.
- Variables should be measurable on quantitative and continuous scales.
- The added costs of data collection must be weighed against any anticipated improvement in data interpretability.

Little effort is made in this report to help an investigator select which alternative method he should use for his particular problem. Nor is more than a superficial effort made to identify and handle special problems that might arise uniquely in behavioral research.

## SELECTING NEW DATA POINTS TO IMPROVE DESIGN ORTHOGONALITY

Adding additional experimental conditions at the proper coordinates within the experimental space can reduce the non-orthogonality of an undesigned experiment. When an ill-conditioned design can be repaired this way sufficiently, the data may be interpreted with the finesse ordinarily reserved for data obtained from orthogonal designs. Improved orthogonality depends solely on the location of the experimental conditions and is independent of the responses obtained under those conditions.

Two methods of selecting these additional data points have been proposed. These are:

- Search the entire region of interest in the experimental space to find one or more points that satisfy the selection criterion.
- Examine a group of "candidate" points to see which one best meets the selection criterion.

The first will be called the "random search approach" and the second, the "candidate selection approach".

## The Selection Criterion

With either approach, given an initial set of non-orthogonal conditions,

> the orthogonality of an experimental design will be improved if the next condition is chosen at the point in the region of interest where the variance of the fitted response, V(y), is largest.

When data has been collected within the experimental space in some non-systematic fashion, the precision of the data throughout the continuous response surface will be irregular, with greater precision naturally lying in the vicinity of where the greatest amount of data was collected and vice versa. The selection criterion says that to improve orthogonality additional data should be collected at the point or points in the response surface where the precision is poorest, i.e., the variance is highest. In Appendix A, methods of discovering and measuring this point of maximum variance will be discussed.

When a data point is added to the non-orthogonal design at the point on the response surface where variance is highest, the following occur:

- The non-orthogonal design becomes more orthogonal.
- The variance at that point is reduced.
- The design becomes more "rotatable" over a spherical region of interest. (A rotatable design is one in which the variances of estimated values equidistant from the center of the design will be equal. See Box and Hunter, 1958, 1, 167.)
- The overall variance of the polynomial is reduced.
- The confidence regions about the regression coefficients are reduced.

Mathematically, adding a new data point in the region of interest where the variance is largest also maximally increases the determinant of the revised — old plus new points — experimental design matrix. All of the effects cited above will also occur as a consequence of maximizing the determinant.

Therefore, if it is practical to do so, selecting the point that will maximize the determinant of the revised matrix could be substituted for the criterion of selecting the point on the response surface where variance is highest. Computer programs for calculating determinants are cited in Appendix A.

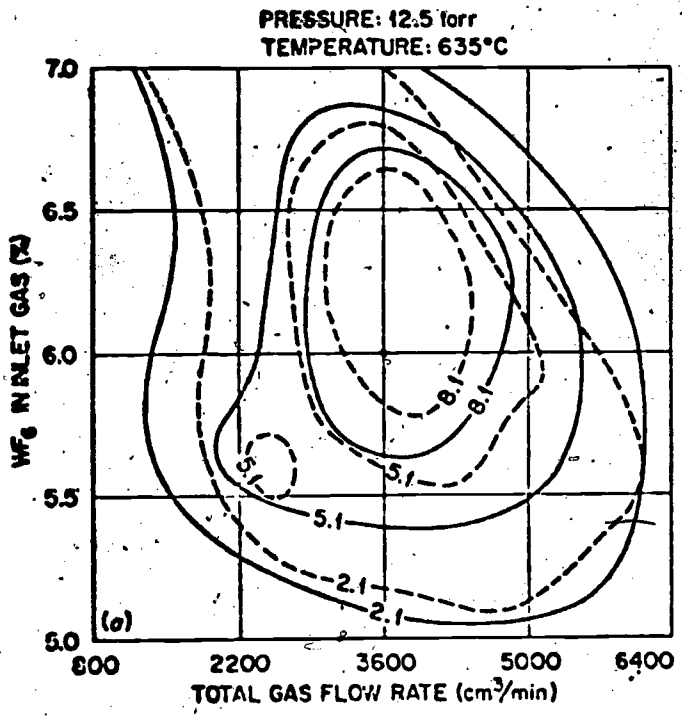
Random Search Approach

Hebble and Mitchell (1972) propose that a computer be used to randomly search the existing experimental space (of the original undesigned experiment) to find where the variance of the estimated reponse is maximum. When found, that point would be the next condition to add to the experimental design. The process is then repeated, seeking the point where the variance is maximum within the space now defined by the original plan plus the first additional point. A third point will be added where the variance is maximum within the space defined by the original and two additional points. This process continues until the investigator is satisfied with the degree of correction obtained. Once a sufficient number of data points have been selected, the performance data can be collected.

Hebble and Mitchell (1972, p. 768) state: "When there are not more than two independent variables, ... we use a grid search procedure. When the factor space is of higher dimension, ... we favor a random search technique. We chose random search in preference to more sophisticated optimization procedures for the following reasons: (i) The random search technique is easier to use, especially when the region of interest R is constrained in strange ways. (ii) We feel that the random search technique can be most easily extended to the simultaneous considerations of several criteria."

Example. Hebble and Mitchell (1972, p. 776) show how their random search approach can be applied to repair the non-orthogonal design used in a chemical problem. Four predictor variables were involved. They had planned to use a third-order rotatable design requiring 81 runs, but during the experiment, some combinations were never run because of equipment limitations. As a result, the orthogonality of the design was destroyed.

To repair the design, they added five new design points using the random search approach. The change in the "information contour" of two dimensions of the response surface before and after the extra data points were added is shown in figure F.3. This contour of constant information, I, is inversely related to the variance contour, i.e., $I = \sigma^2 / V(\hat{y})$. The improvement in rotatability after the points had been added is visually obvious. There was a corresponding improvement in the other qualities affected by adding points at the maximum $V(\hat{y})$ which also maximizes the determinant of the augmented design.




[F.3]

ILLUSTRATING THE IMPROVEMENT IN ROTATABILITY IN VARIANCE CONTOURS AFTER FIVE DATA POINTS HAD BEEN ADDED WITHOUT BLOCKING

- - - Original contours

—— Improved contours

[Adapted from Figure 6 in paper by Hebble and Mitchell (1972)]

Precaution. The original purpose for adding more data points to the original undesigned experiment was to improve the orthogonality of the design, which in turn could facilitate the interpretation of the results. It should be noted however that although using the maximum variance criterion does improve the design for that purpose, it does not necessarily provide an optimum design. Hebbie and Mitchell (1972, p. 778) recognized this when they wrote: "...in many cases, 'bias' caused by fitting an inadequate model will be a more important source of error in the fitted response than will variance." Bias error can be present, for example, if a higher order relationship exists in fact between variables and performance but these effects cannot be isolated by the existing experimental design. Because the bias criterion would be a more difficult one to meet, Hebbie and Mitchell ignore the problem. In the next section, Dykstra suggests some ways of meeting it.*

Candidate Selection Approach

Dykstra (1971) proposes that instead of searching randomly through the region of interest for the point where the estimated response variance, $V(\hat{y})$, is maximum, a group of candidate points should be selected on some rational basis. Then the $V(\hat{y})$ of these candidate points would be calculated for the existing design and the one with the largest $V(\hat{y})$ would be used for the next run. Candidates would continue to be evaluated this way for each successive run.

Of course, none of the candidate points will necessarily be located precisely at the point on the response surface where the $V(\hat{y})$ is maximum. This makes the results somewhat less accurate initially than the random search approach. However when a series of runs is made, the approach becomes self-correcting. One advantage of this approach over the random search approach is the reduction in computer time.

---

*Designs that satisfy both bias and random error criteria have been proposed by Box and Hunter (1958). Tests of the goodness of fit of a specific model are applied. If the fit is found inadequate, data points that will enable a higher-order to be fit are added to the original design. (See Simon, 1970 and 1973.)

Selecting a Set of Candidate Points. There are a number of practical considerations affecting the rational selection of a set of candidate points. For example, the investigator would avoid selecting points:

- That are not feasible to run.
- Where no response is likely to occur.

Thus unlike the random search technique, the use of rationally selected candidates permits the experimenter to impose his judgment onto the mathematical criteria by selecting points of practical interest as well. This enables the number of different levels of each factor to be kept reasonably low, an important consideration when changing experimental conditions is difficult or time consuming.

Dykstra (1971), by choosing candidate points, is more able to attack the problem of equation bias that Hebbie and Mitchell ignored. The candidate points should be selected in a way that not only improves orthogonality and the associated reduction in variance but also develops into a design of a model that will adequately fit the data. He suggests the following:

> "In choosing specific combinations, however, one should be guided by the model. For a first-order model the procedure will select points at the extremes of the experimental space, so that only corner points need be specified as candidates. For a second-order model the list of candidates should include the axial points and a center point, in addition to corner points. A cubic model should have the candidates at four levels of the controllable variables, and so on." (p. 684).

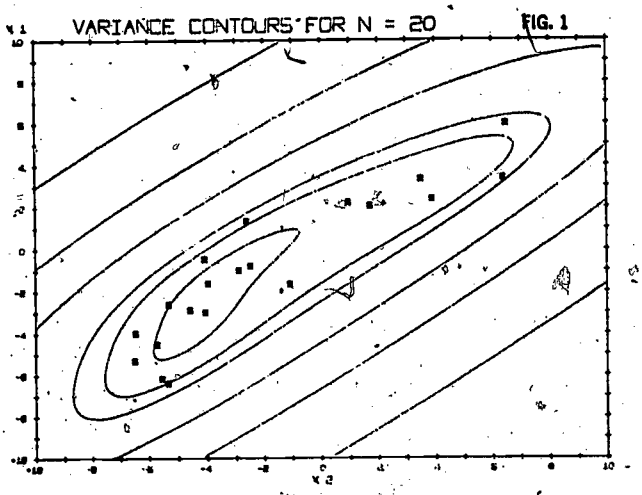
Selecting the Point for the Next Run. Given a set of candidate experimental conditions, the one selected to run next is the one that gives the highest variance for the estimated response at that point, i.e., where the value of $V(\hat{y})$ is greatest, or when added to the existing design, maximizes the determinant for the augmented matrix. Each additional candidate point is selected sequentially in the same way until there is a decision to stop. The number of candidate points to be used, although at the discretion of the experimenter, may be based partially on the number required to meet the characteristics of the model and partially on the improvement needed in the precision of the equation.

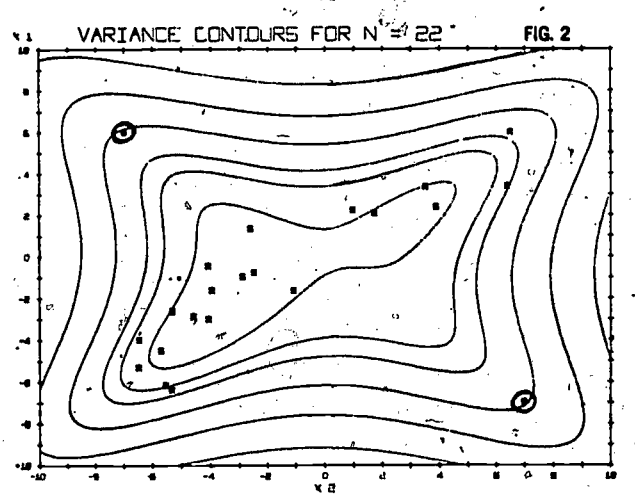
Mitchell (1974) proposes his own algorithm "DETMAX" for design augmentation which searches for complete subsets of candidate points that will optimize (almost) the determinant of the X'X matrix. He states that this method will give higher values of the $|X'X|$ than Dykstra's one-point-at-a-time approach, but admits that the latter "is seldom far off, and takes much less time on the computer. In many practical situations, when the object is to find a good (not necessarily 'optimal') design quickly, the sequential procedure will be quite satisfactory." (p. 206).

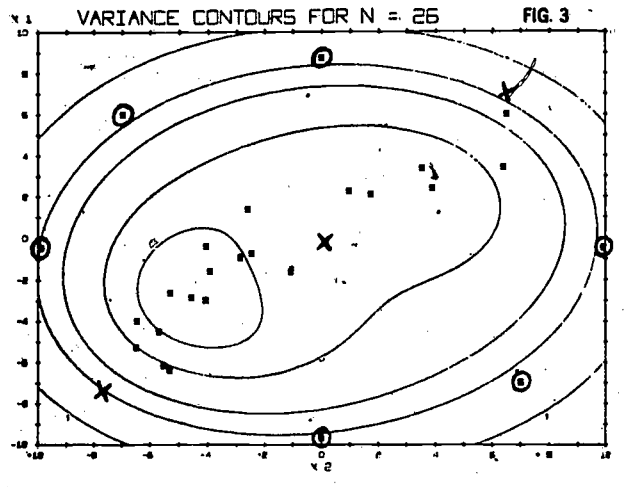
Example. Dykstra (1971) improved the orthogonality of a 20-run "undesigned" experiment with two correlated predictors by sequentially adding six out of nine candidates needed to improve a second-order design. The nine candidate points were the four corners of a square, the four extremes of the axes, and one center point, with the non-center points placed equidistant from the center of the space. The changes in the variance contour before and after several degrees of augmentation are shown in figure F. 4. The shift toward a more rotatable design is visibly obvious. After 26 points had been added, the determinant of the augmented design for the 26 points is $3.39 \times 10^5$ times larger than it had been for the original 20 points.

## HANDLING PERFORMANCE SHIFTS BETWEEN ORIGINAL AND ADDED POINTS

Characteristically in human performance research, if experimental conditions are measured sequentially, changes in performance may be observed that are not due to the experimental variables. If there is a considerable interval between the time the performance data are collected from the original undesigned experiment and from the additional points, unexplained and undesired performance shifts may occur. This can be due to changes in the subject, in the environment, in the equipment, or any number of unknown factors. In any case, unless this shift in performance between blocks of data is dealt with properly, it will distort the information of interest.




[F.4]

Original points

Added points

X Candidate points never selected

CHANGE IN VARIANCE CONTOURS FROM ORIGINAL 20 EXPERIMENTAL CONDITIONS (FIG. 1) AS TWO (FIG. 2) AND SIX (FIG. 3) DATA POINTS ARE ADDED

[Adapted from Figures 1, 2, and 3 in the paper by Dykstra (1971).]

This blocking problem can be handled in two ways,

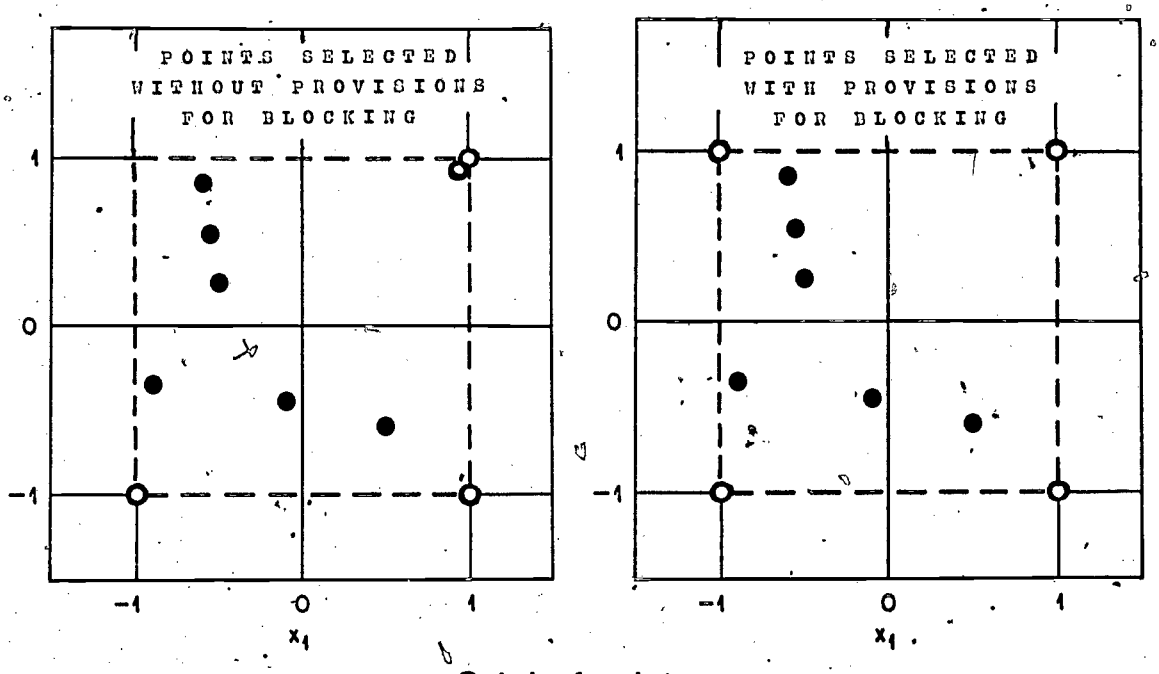
- By including a blocking term in the regression model.
- By adding data points in balanced pairs.

## Adding a Blocking Term

Hebble and Mitchell (1972, p. 771) suggest that a blocking term be included in the regression model to account for a possible difference in overall response level between the initial design and the runs that are chosen to augment it. They say (p. 771): "....when a constant $\beta_o$ is already in the model, we can account for a possible block effect simply by introducing a 'dummy variable', which takes the value of 0 for each run in the initial design and a value of 1 for each additional run. When this is done, the model for the original design is unchanged by the introduction of the blocking variable. Thus, the first new point in the design can be selected without introducing a blocking variable. To select further new points, the blocking variable should be included, and a 1 should appear in the blocking column of the X matrix in every row which is part of the additional block of runs." They note that different data points will be selected when a blocking term is and is not included (See figure F.5). Generally, it is wiser to include a blocking term. Procedural precautions against sequence effects (Simon, 1974) also should be employed whenever possible.

## Adding Data Points in Pairs

Dykstra (1966, p. 279) suggests that another way of handling this problem is to augment the design with pairs of data points. Orthogonal blocking will be obtained if pairs of data points are selected so that the averages of the coordinates of these new pairs equal the average of the corresponding coordinates of all conditions in the original design. For example, if the original




[F.5]

• Original points

o Selected points

ILLUSTRATING HOW FOUR CANDIDATE POINTS ARE SELECTED DIFFERENTLY DEPENDING ON WHETHER A BLOCKING TERM IS (B) OR IS NOT (A) INCLUDED

[Adapted from Figure 5 in the paper by Hebble and Mitchell (1972).]

---

design of three variables had been made up of data points at the following coordinates:

| | Variable | | | |
|---|---|---|---|---|
| Data Points | I | II | III | Variable levels (or coordinates) |
| No. 1 | 3 | 2 | 7 | |
| No. 2 | 1 | 3 | 2 | |
| No. 3 | 4 | 5 | 2 | Original design composed of four data points |
| No. 4 | 2 | 4 | 1 | |
| Average: | 2.5 | 3.5 | 3.0 | |

to be orthogonal, the two new data points would have to be selected at points

where their mean coordinates equal the averages of the points already in the design, for example:

| | | | |
|---|---|---|---|
| No. 5 | 4 | 5 | 2 |
| No. 6 | 1 | 2 | 4 |
| Average: | 2.5 | 3.5 | 3.0 |

Dykstra in his 1971 article did not discuss this blocking method when he used the maximum $V(\hat{y})$ criterion to find the coordinates where the next data point is to be added. However, it could still be used if candidate points were designated in pairs and the criterion for selecting the proper pair would be that which maximizes the determinant of the augmented design. Mitchell's (1974) DETMAX, for example, might be used for this application.

## SECTION IV

## RIDGE REGRESSION ANALYSIS

The purpose of regression analysis is to obtain a set of coefficients for an equation that will fit the existing data without bias and with a minimum amount of variable error. The conventional criterion of a best equation is one in which the sum of the errors squared between estimated and observed responses will be at a minimum.

When an orthogonal design has been employed (i.e., the predictor variables are mathematically independent), the estimated beta coefficients are reasonable representations of the true beta coefficients, within the limits set by the error estimate. When a non-orthogonal design has been employed, the individual betas calculated on the basis of the least squares criterion are often unsatisfactory. While the overall equation may be adequate for prediction, the relative effects of individual terms cannot be evaluated. With non-orthogonal designs, beta coefficients derived from a least squares fit may not make sense in the real world.

Hoerl and Kennard (1970a, b) cite the following characteristics of coefficients estimated from ill-conditioned experimental designs:

- The coefficients become too large in absolute value.
- Some coefficients may have the wrong sign.
- Collectively the coefficients are unstable; another set of performance data would be unlikely to give the same beta values.
- Individual coefficients may be over or under estimates of the strength of a particular factor.

The more non-orthogonal the original design, the poorer the equation is likely to be.

All of these conditions stem from the correlations among the predictor variables. In the past, in order to untangle the relationship among the factors, it has either been necessary to drop those predictors that correlate the

highest with the others or to treat the total equation as a black box. However some of the power of the regression model is lost by either of these approaches. As an alternative to conventional multiple regression (least squares) analysis with non-orthogonal data, Hoerl and Kennard propose "ridge regression". This analysis, they suggest, will obtain a better prediction equation in which:

- The estimated coefficients will be closer to the true coefficients on the average;
- The signs will be more meaningful;
- A point estimate of a response can be made with a smaller mean square error;
- The coefficients will be more stable and likely to be repeated if new data is taken.

## MATHEMATICAL BASIS FOR RIDGE REGRESSION

Hoerl and Kennard (1970a, b) supply the mathematical basis for ridge regression analysis. Only the rudiments of their explanation will be supplied here. The reader should refer to the original papers if more details are desired. Marquardt (1970) also deals with the mathematics of ridge regression as part of a broader class of biased linear estimators employing generalized inverses.

Essentially, Hoerl and Kennard (1971a) show that in conventional multiple regression analysis, the average value of the squared distance, $E(L^2)$, between the estimated, $\hat{\beta}$, and the true, $\beta$, beta coefficients is equal to the error variance, $\sigma^2$, of the data multiplied by the sum of the reciprocals of the eigenvalues, i.e.,

$$E(L^2) = (\hat{\beta} - \beta)' (\hat{\beta} - \beta) = \sigma^2 \sum_{1}^{\rho} (1/\lambda_i) . \qquad [E.3]$$

When the predictor variables are uncorrelated, the eigenvalues, $\lambda_i$, are each equal to one. In that case, the average squared distance between estimated and true beta coefficients will be equal to the error variance of the data

multiplied by the number of variables, p, involved. However when the predictors are correlated, as in the case of the undesigned experiment, some of the eigenvalues become very small and their reciprocals very large. This increases the average squared distance between the estimated and true beta coefficients. A least squares fit of data from a non-orthogonal experimental design also produces coefficients that are too large in their absolute value. To compensate for these large positive coefficients, other coefficients are estimated that are too negative which often may be the incorrect sign. The more ill-conditioned the design matrix, the worse these conditions are likely to be.

To correct for this, Hoerl and Kennard propose to add a small positive quantity, k, to the unit diagonal of the intercorrelation matrix of the predictor variables. For example, if the original intercorrelation matrix were:

| Variables | $X_1$ | $X_2$ | $X_3$ | $X_4$ | Y |
|---|---|---|---|---|---|
| $X_1$ | 1.0 | 0.23 | 0.45 | 0.67 | 0.14 |
| $X_2$ | 0.23 | 1.0 | 0.15 | 0.36 | 0.26 |
| $X_3$ | 0.45 | 0.15 | 1.0 | 0.89 | 0.54 |
| $X_4$ | 0.67 | 0.36 | 0.89 | 1.0 | 0.22 |

Correlations between $X_i X_j$ or $X_i Y$

then the new matrix would be, for example, if k = 0.2, would be:

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | Y |
|---|---|---|---|---|---|
| $X_1$ | 1.2 | 0.23 | 0.45 | 0.67 | 0.14 |
| $X_2$ | 0.23 | 1.2 | 0.15 | 0.36 | 0.26 |
| $X_3$ | 0.45 | 0.15 | 1.2 | 0.89 | 0.54 |
| $X_4$ | 0.67 | 0.36 | 0.89 | 1.2 | 0.22 |

Note that the k = 0.2 has been added to the 1's in the diagonal. Next, a conventional least squares fit is done using the perturbed matrix. The results produce what Hoerl and Kennard call "ridge coefficients," $\hat{\beta}^*$. The distinction

between the conventional beta, $\hat{\beta}$, coefficients and the ridge coefficients, expressed in matrix algebra, is:

$$\hat{\beta} = (X'X)^{-1} X'Y,$$

and

$$\hat{\beta}^* = (X'X + kI)^{-1} X'Y.$$

Values of k between 0 to 0.9 may be substituted with finer increments being used at the lower end of the scale below 0.1 where changes in the estimated ridge coefficients are greater. Whereas the betas estimated from the conventional least squares are unbiased with minimum variance, the ridge coefficients contain both a bias and a variable error. These two error components are present in the equation (written in matrix algebra) for the average squared distance between values of the ridge coefficients and the true coefficients thus:

$$E[L_i^2(k)] = \underline{\sigma^2 \sum_i^p \lambda_i / (\lambda_i + k)^2} + \underline{k^2 \beta' (X'X + kI)^{-2} \beta} \qquad [E.4]$$

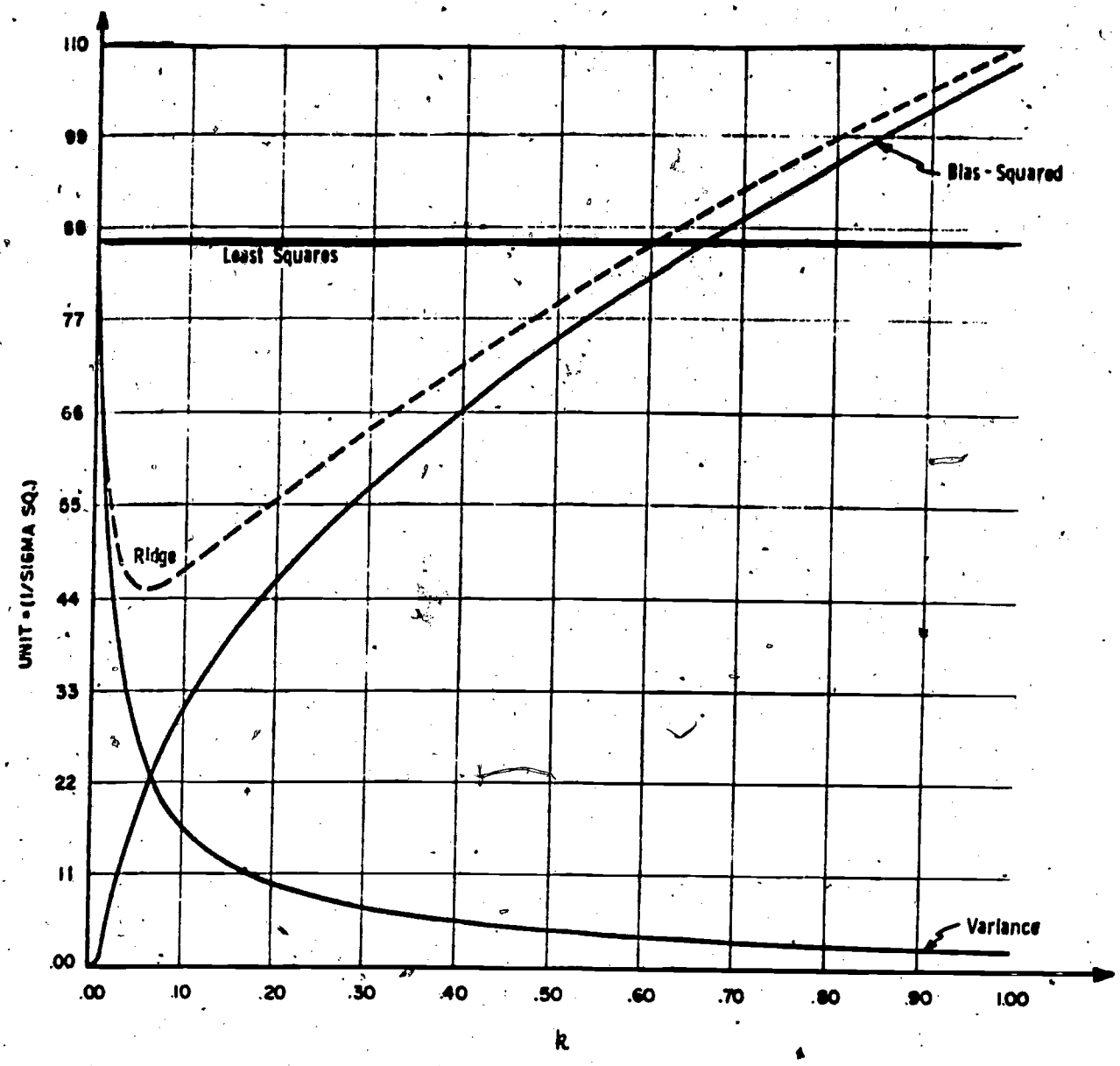
The first component represents the variance and the second, the bias. Note that when k = 0, the second component disappears, leaving the unbiased estimates of the coefficients found by a conventional least squares fit. As k increases, so does the bias error.

However, Hoerl and Kennard demonstrate that as k increases, the variance error decreases more rapidly than the bias error increases. This means that at some value of k, the mean square error — the combination of bias and error variance — for the ridge coefficients will be smaller than it would be for the conventional coefficients.

Exactly what has happened in this process is simple to understand if equations E.3 and E.4 are referred to. In equation E.3, it can be seen that the small eigenvalues have the greatest impact on the estimations. The smaller some of the eigenvalues get (as a result of a non-orthogonal design), the larger their reciprocals and the greater the squared distance between estimated and true beta coefficients becomes. In equation E.4, it can be seen that adding a constant k to the correlation matrix diagonal has the effect of adding k to the eigenvalues of the variance component. For the very small eigenvalues, the addition of even a small k can do much to decrease the size of the reciprocals of the eigenvalues and to decrease the squared distance between estimated and true beta coefficients.

This phenomenon is illustrated in figure F.6. In this figure, both the bias squared and the variance of the ridge regression coefficients have been standardized by dividing each by the residual error variance of the response data. The least squares variance (normalized) of the estimated beta coefficients is represented by the horizontal line (a constant) across the top of the graph. When k equals 0, of course, the variance of the ridge coefficients is identical to the variance of the estimated beta coefficients, and the bias squared (normalized) is zero. As k increases, however, it can be seen that the variance decreases and the bias squared increases, each in a monotonic function. The sum of these two effects, the mean square error (as represented by the dashed "ridge" line), drops initially only to rise later on. There will always be for some value of k a portion of the ridge trace where the mean square error is smaller than it would be had no distortion been introduced. In this example, the mean square error is at a minimum for k = 0.05, nearly half the magnitude of the original variable error. While there are other criteria than the minimum ridge value for selecting the k where the ridge coefficients would be found, this figure does illustrate how adding the bias can actually reduce the mean square error, and thereby improve the estimates of the coefficients.

Two computer programs for performing ridge regression analysis are listed in Appendices B and C, a print-out of the latter is given in Appendix D, and some discussion on both programs is held in Appendix A.



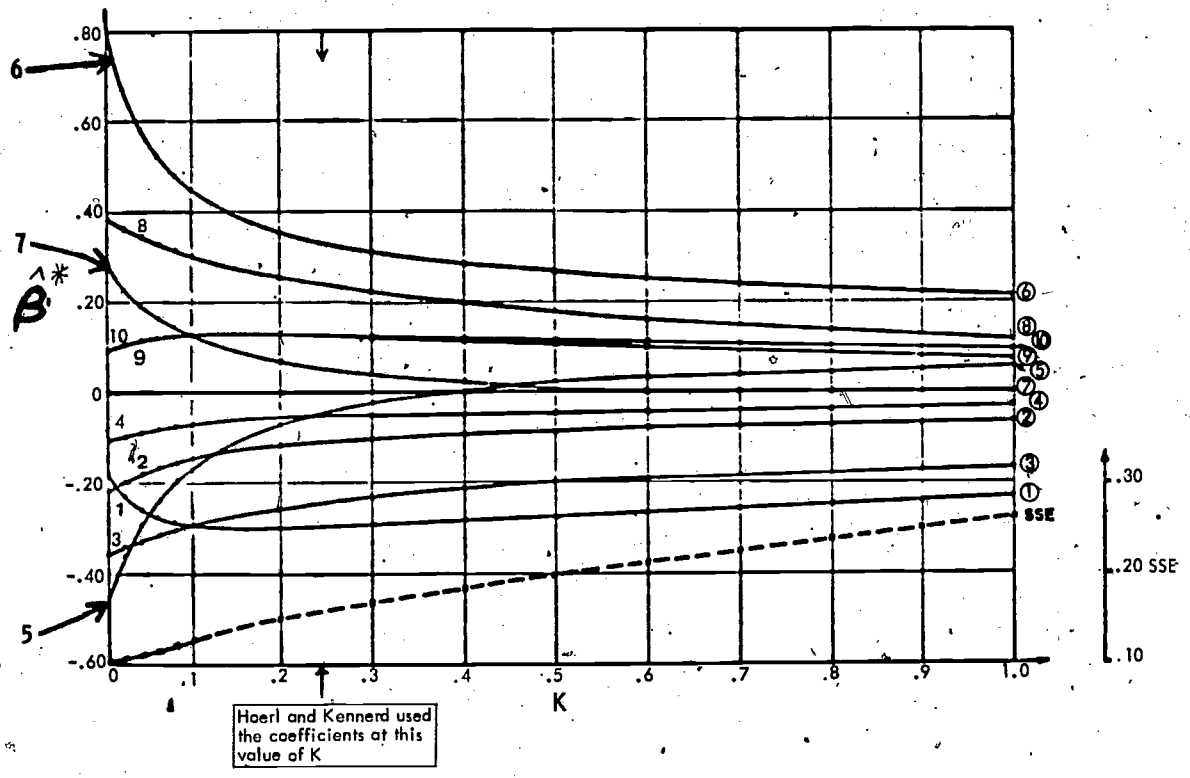

[F.6]

RIDGE REGRESSION MEAN SQUARE ERROR FUNCTIONS

[From Figure 1 in Hoerl and Kennard's (1970a) paper.]

## INTERPRETING RIDGE TRACES

One of the advantages of ridge regression analysis over conventional least squares is the ability to portray the sensitivity of the beta estimates graphically. A two-dimensional ridge trace of the ridge coefficients is obtained by plotting the estimated ridge coefficients against the values of k. This is illustrated in figure F.7.




[F.7]

RIDGE TRACE: TEN FACTOR EXAMPLE

[From Figure 1 in Hoerl and Kennard's (1970b) paper.]

The plot of solid lines illustrates how, as k increases, the ridge coefficients ($\hat{\beta}^*$) diminish in absolute magnitude and begin to stabilize. If k were to go to infinity (ad absurdum), of course, the above processes would be complete, for all coefficients would be equal to zero. From figure F.7, it is apparent that long before that point is reached the distance between the estimated and true coefficients would be too large to be of practical value. Quite obviously, therefore, it is necessary to select a minimum value of k that will adequately provide an improved set of coefficients, ones that are more meaningful and will result in more accurate predictions.

The dashed line at the bottom of figure F.7 is a plot of the residual sum of squares as a function of k. It is normal that as bias is introduced into the design matrix, the lack of fit of the original data would become poorer

(i.e., SSE become larger). It is only when the equation with the ridge coefficients is used to estimate performance on new data that the estimation has been improved.

Hoerl and Kennard did not feel that an automatic — mathematical — solution for selecting the best k was justified. They stated (1970a, p. 64):

> "The inherent boundedness assumptions in using $\hat{\beta}^*$ make it clear that it will not be possible to construct a clear-cut, automatic estimation procedure to produce a point estimate (a single value of k or a specific value for each k), as can be constructed to produce $\hat{\beta}$. However, this is no drawback to its use because with any given set of data it is not difficult to select a $\hat{\beta}^*$ that is better than $\hat{\beta}$."

They propose that instead of seeking a mathematical solution for k, the ridge regression chart be examined visually. The following conditions should be looked for when selecting the value of k:

1. The beta values and particularly their orders of magnitude 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 square variance.

In the analysis illustrated by figure F.7, Hoerl and Kennard (1970b) selected a k between 0.2 and 0.3. Note how the coefficients have begun to stabilize, how variables 6 and 7 have reduced considerably in magnitude (with 7 losing its effectiveness almost completely), and how variable 5 has begun (but not completed) a shift from a large negative to a low positive coefficient. At k = 0.25, the residual sum of squares (SSE) has increased approximately 60 percent, from 0.10 to 0.16 while the expected squared distance between

of the coefficient estimate from true coefficient values has reduced to 26 percent of its original value.

## ALTERNATE METHODS OF SELECTING K

Techniques other than Hoerl and Kennard's have been proposed for selecting the desired k value. Several investigators reanalyzed the data from Gorman and Toman's (1966) 10-variable study that Hoerl and Kennard had used for the analysis shown in figure F.7 in this report. For the same data, using different criteria, the individuals cited below selected the following k values:

| Source | Value of k | Basis of Selection |
|---|---|---|
| Hoerl and Kennard (1970b, p. 72) | 0.2500 | Inspection |
| Lindley and Smith (1972, p. 17) | 0.0390 | Bayesian |
| Mallows (1973, p. 672) | 0.0200 | $C_p$ plot |
| Farebrother (1975, p. 128) | 0.0029 | Min. MSE |

Lindley and Smith (1972) argue that since there is usually prior information about the parameters relating predictors to performance, this information should be exploited to find improved estimates of the parameters. They apply Bayesian methods to linear regression analysis arguing that in the case of non-orthogonal data, the Bayesian method reaches the same conclusion as the ridge method but has the added advantage of dispensing with the rather arbitrary choice of k and allows the data to estimate it. Using Gorman and Toman's (1966) data, they compare the coefficients obtained by the three methods — least-squares, Bayes, and ridge. They note that the Bayes approach like ridge gets rid of the three major complaints against betas obtained from least-squares — large absolute values, incorrect signs, and instability. Comparing the results of the Bayesian versus the ridge approach, they note that all the estimated coefficients are pulled towards zero with those from the ridge being smaller since "a considerably larger value of k than the data suggest" (p. 17) was used.

Farebrother (1975, p. 128) concludes from his own data and from a reexamination of Mallow's k value — which he thought should have been smaller — that "Hoerl and Kennard's quest for stability has led them too far from the unbiased estimator."

It is appropriate to remind the reader that each of the above investigators was applying a different criterion when he selected the optimum k value, and what may be best for one purpose may not be best for another. For example, Mallow's (1972) $C_p$ criterion ("standardized total squared error") is a measure that combines both bias and variable error and he selects the k value where $C_p$ is minimum. This could conceivably correspond to the minimum ridge value considered by Hoerl and Kennard in figure F.6 but which was not used in selecting the k in the analysis shown in figure F.7. Which is better? The difference might be in whether one is more interested in a good prediction without too great an increase in RSS, in which case the mean square error or $C_p$ should be minimized or if one is more interested in comparing individual terms, in which case the stability of the individual coefficients becomes more important. Only experience is going to decide how the numerous criteria must be traded off against one another.

McDonald and Schwing (1973) used ridge regression analysis on a problem relating air pollution to mortality. They selected their value of k (which was not necessarily optimal in so far as the mean square error was concerned) according to three criteria, i.e., at the point where:

- The order of magnitude of the coefficients had stabilized;
- The residual sum of squares and coefficient of determination had values consistent with problems of that type;
- The ridge coefficients are within the 95 percent confidence ellipsoid for the unknown true coefficients, assuming normally distributed errors.

Newhouse and Oman (1971) propose several methods of choosing a k value to use in ridge regression and investigate their properties using Monte Carlo experiments with two predictors. It appears that an optimal choice of k

(or interval of k values) is an open question at this time unless one has prior knowledge about the length and/or direction of the unknown coefficient vector.

Although the problem of how and where to select the best value of k has not yet been resolved, the overall superiority of ridge regression over least squares regression in the analysis of non-orthogonal data has not been seriously questioned.

Theobald (1972) independently demonstrated that provided

$$k < 2\sigma^2/\beta'\beta$$

the mean square error of the ridge coefficients will always be smaller than the least squares value of the conventional regression coefficients. For this to be true, the values of the ridge coefficients must be bounded, a realistic condition. Theobald does not attempt to precisely locate the optimum value of k within the limits set by the equation.

Banerjee and Carr (1971) suggest a different and "more meaningful" criterion against which to assess the accuracy of the biased estimator, $\hat{B}^*$, that Hoerl and Kennard used. Hoerl and Kennard compared the size of the mean square error of the bias estimators (where $k > 0$) against the variance of the unbiased coefficients from the conventional least square fit ($k = 0$) to show there always exists a k at which the new mean square error would be less than the original variance. Banerjee and Carr however argue that it would be more meaningful to compare the mean square error of the biased estimators against a modified variance,

$$E\ (\hat{\beta} - \beta)^2 \text{ squared bias} = \sigma^2 \sum_{1}^{p} \left[1/(\lambda_i + k)\right]$$

rather than the one Hoerl and Kennard used,

$$\sigma^2 \sum_{1}^{p} \left[1/\lambda_i\right].$$

However, they also show that even against this modified criterion, there still exists a k where the biased estimators have a smaller mean square error, although the effect is less pronounced. Banerjee and Carr suggest that the "gain in accuracy may better be exhibited in relative terms, that is in terms of percentages (or fractions) of the variance $\hat{\beta}_A$ or $\hat{\beta}$, rather than in absolute terms" (p. 898). $\hat{\beta}_A$ as used here refers to the ridge coefficients.

Goldstein and Smith (1974, p. 288) propose a modification of the ridge approach "which might be appropriate if one were especially interested in some particular $\beta_i$'s, or were worried that the Ridge estimate might distort the estimation of those $\beta_i$'s which could be estimated accurately anyway." They suggest the possibility of choosing different values of k for different predictor components. They disagree with Hoerl and Kennard that this procedure would offer little improvement over the use of a constant k. It would depend, they claim, on what the optimum k would be for each component; if it differed widely, then an improvement in the mean square error could be expected.

## IDENTIFYING CRITICAL VARIABLES

When an equation is exceptionally long and if many of its terms are found to be inconsequential, some investigators will want to drop the terms. In the case of the designed experiment in which variables are orthogonal to one another, dropping terms of insignificant effects is a straight-forward process. In the case of the undesigned experiment, traditionally (because of the intercorrelation among the variables) dropping a term just because a coefficient is small would be unwise. However, since a shortened equation is simpler and more convenient and economical to use, a variety of algorithms have been devised to find the "best" subset of variables out of the total considered in the undesigned experiment that will fit the data about as well as the complete equation.

Techniques for selecting the "best" subset regression equations have been primarily of two types: one, those that literally compare all possible (or all reasonable) subsets of regression equations against some criterion of

goodness, or two, those with no exact criterion of goodness but which depend upon a heuristic algorithm that will supply a group of potentially good candidates from which the investigator will select the "best".* Mathematical criteria for comparing subset regressions have traditionally been either the minimum error variance (which is the least squares fit criterion) or a minimum total (bias and variance) error, with minor variations in the exact form involved. Hocking (1972), Allen (1971), Helms (1974) and Beale (1970) critically review these criteria. Since there can be $(k! - 1)$ possible subsets of an original equation with k variables, the main emphasis in developing selection techniques that compare many subset regressions has been to reduce the computation time required for the analysis. Some recent efforts in this regard are those of Furnival and Wilson (1974) and LaMotte and Hocking (1970). Among the techniques employing the less exact criterion for selection, the stepwise regression algorithm has been perhaps the one used most frequently by behavioral scientists analyzing nonorthogonal data. This and related techniques are discussed by Draper and Smith (1966), Chapter 6, by Beale (1970), and Kerlinger and Pedhazur (1973, pp 289-295).

Ridge regression provides an alternative solution to the subset selection problem. By stabilizing the coefficients, it enables the relative importance of predictor variables to be assessed more directly. However, in

---

*Other techniques, such as conical equations, factor analysis, and so forth have been proposed for isolating subsets of variables. While these are undoubtedly useful for certain purposes, they may be of limited value for certain equipment design problems. The reason for this is that solutions from these techniques result in composite variables. That is, a solution will provide a set of mathematically independent variables which are mathematical mixtures of the original variables. Such solutions, while probably useful in test construction or personality assessment, will ordinarily not be adequate for problems of equipment design. While any technique that aids in interpreting data should be considered, beware of relying on techniques that don't fit the particular problem under investigation.

spite of this, Hoerl and Kennard (1970b) specifically recommend that factors with small coefficients not be dropped from the equation.* Instead they recommend the following procedure when some variables have small coefficients and are believed to have small effects:

> "To 'discard' a factor, set it at its average value for all predictions, which is the equivalent of setting the coefficient equal to zero. But do not delete and reestimate . . . ." (p. 75).

The average value for any predictor is the mean of all levels of that predictor used in the experiment. They demonstrate how eliminating low effect predictors completely can result in an even more unstable solution than when all predictors are retained.

## APPLYING RIDGE REGRESSION ANALYSIS TO A TARGET ACQUISITION PROBLEM

Zaitzeff (1971) at the Boeing Company, Seattle, performed an undesigned experiment to discover the function relating fifteen selected target and background characteristics (table T.6) to the probability of acquiring targets. Observers were required to find a variety of targets visually in a dynamically changing scene. The empirical data thus obtained was subjected to regression analysis.

First a stepwise regression was carried out on the data without stopping until all of the variables had been entered into the equation. The order in which they were entered into (or deleted from) the equation corresponds to the order in which they are listed in table T.6. The first variable, "probability of finding a static target", accounted for more than 80 percent of the

---

*This recommendation must assume that the variables in the complete equation are there because of some rational variable selection and not merely on the whim of an investigator who'd "just like to see what would happen" if they were included. As Hays (1963, p. 577) says: "Tracing relationships among variables is the legitimate business of the scientist, but simply asking if anything relates linearly to anything else in a large set of variables is a pretty crude way to do business." This point is discussed further in the paragraphs on "Data Selection" in Appendix A.

## VARIABLES IN TARGET ACQUISITION STUDY IN ORDER OF THEIR APPEARANCE IN STEP-WISE REGRESSION EQUATION

| Order No. | | Variable ID No. |
|---|---|---|
| 1. | Probability of finding a static target (PSTAT) | (11) |
| 2. | Number of filtered brightness elements in scene (NAVG) | (15) |
| 3. | Number of confusion areas in scene (AMBIG) | (10) |
| 4. | Small dimension (LIT DM) | (3) |
| 5. | Target width (LIT DC) | (9) |
| 6. | Detail contrast (DCONTR) | (7) |
| 7. | Target length (BIG DC) | (8) [T.6] |
| 8. | Area 1 variance (VARAR1) | (14) |
| 9. | Target area (TGAREA) | (4) |
| 10. | Target contrast* (TCONTR) | (6) |
| 11. | Heterogeneity (HETERO) | (12) |
| 12. | Scan variance (VARAVG) | (16) |
| 13. | Large dimension (BIG DM) | (2) |
| 14. | Detail size (DETSIZ) | (5) |
| 15. | Area 1 count (NAREA1) | (13) |

*In the step-wise process, "Target Contrast" originally entered the equation in the step following the entry of "Number of confusion areas in scene", was later deleted in step following entry of "Target length", and finally reentered in position indicated. (From Zaitzeff, 1971)

total performance variance. All 15 variables accounted for 93 percent of the variance. Seven variables would have accounted for 90 percent of the variance.

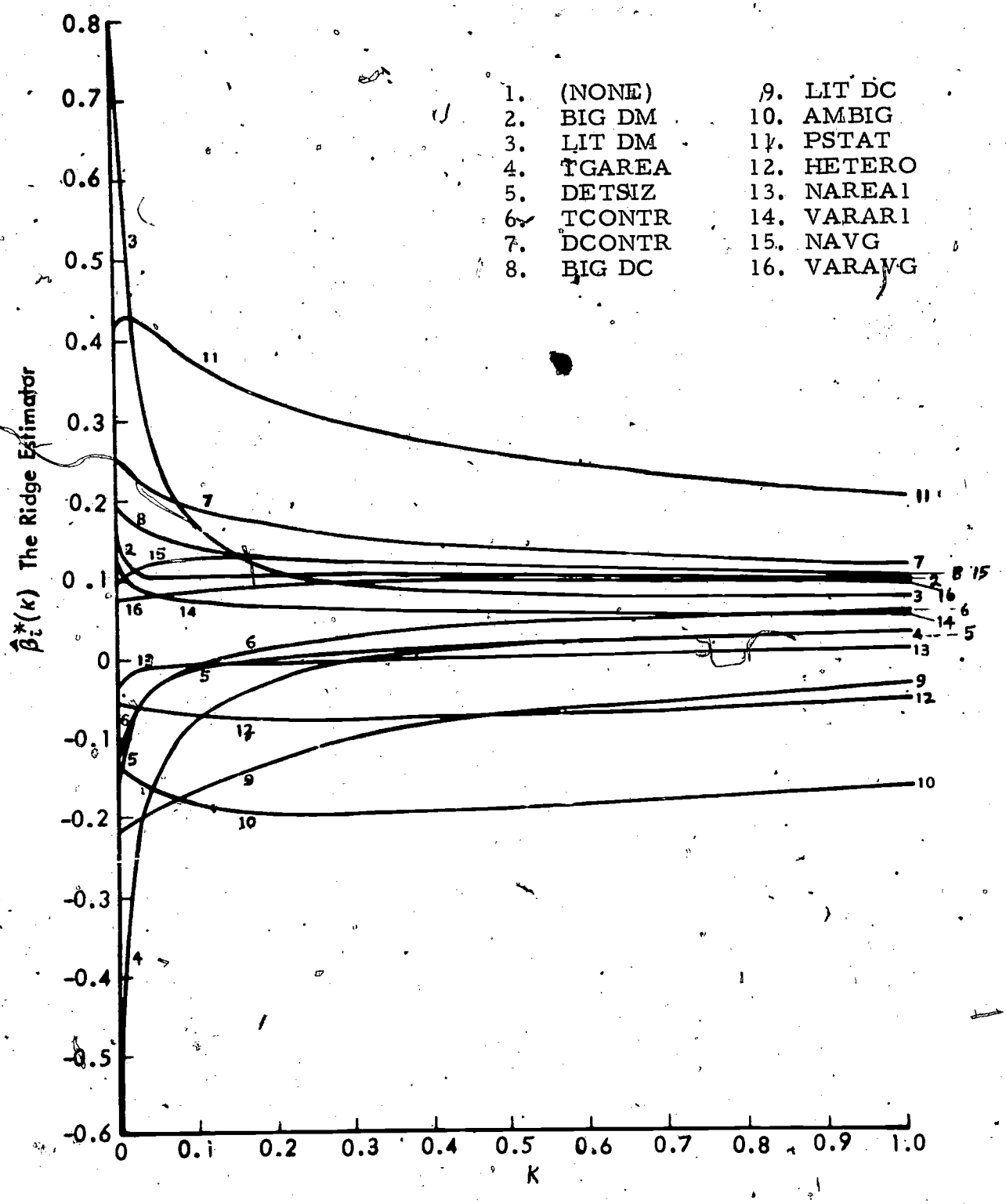
In examining the coefficients from the stepwise regressions, Zaitzeff commented on the unsatisfactory results:

> "Thus, it is disconcerting to see relatively large negative coefficients assigned to such variables as Target Area, Detail Size, Target Contrast, and Target Width, when the factor analysis has shown them to be positively correlated with dynamic acquisition probability" (p. 51).

He also cites other limitations of equations developed with the least squares criterion.

Next a ridge regression analysis was carried out on the same data. The ridge trace of the 15 target-background variables are shown in figure F.8. The instability of a number of the variables is immediately evident. For example, "Little Dimension" (#3) changed from having the largest positive coefficient to one that ranks sixth, and "Target Area" (#4) changed from having the largest negative coefficient to a slight positive one. Zaitzeff decided on the basis of visual inspection that the coefficients were reasonably stable at $k = 0.7$. In addition to "Target Area", #4, both "Detail Size" (#5) and "Target Contrast" (#6) show a sign change that appears more meaningful in the light of what is known about visual perception.

Variable Elimination. Zaitzeff eliminated variables that had stable coefficients but low predicting power (a coefficient less than 0.05) and those with unstable coefficients that failed to hold their predictive value. He also eliminated two other variables, "Big Dimension" (#2) and "Area 1-Count" (#13), which correlated highly positive with two other variables and were considered redundant. In addition, although "Static Acquisition Probability" (#11), is shown to be the single best overall predictor of dynamic acquisition probability, it was eliminated because it was an unwieldy value to acquire and because Zaitzeff felt that it was actually a function of the other physical and psychophysical variables rather than a distinct target-background variable in and of itself.




[F.8]

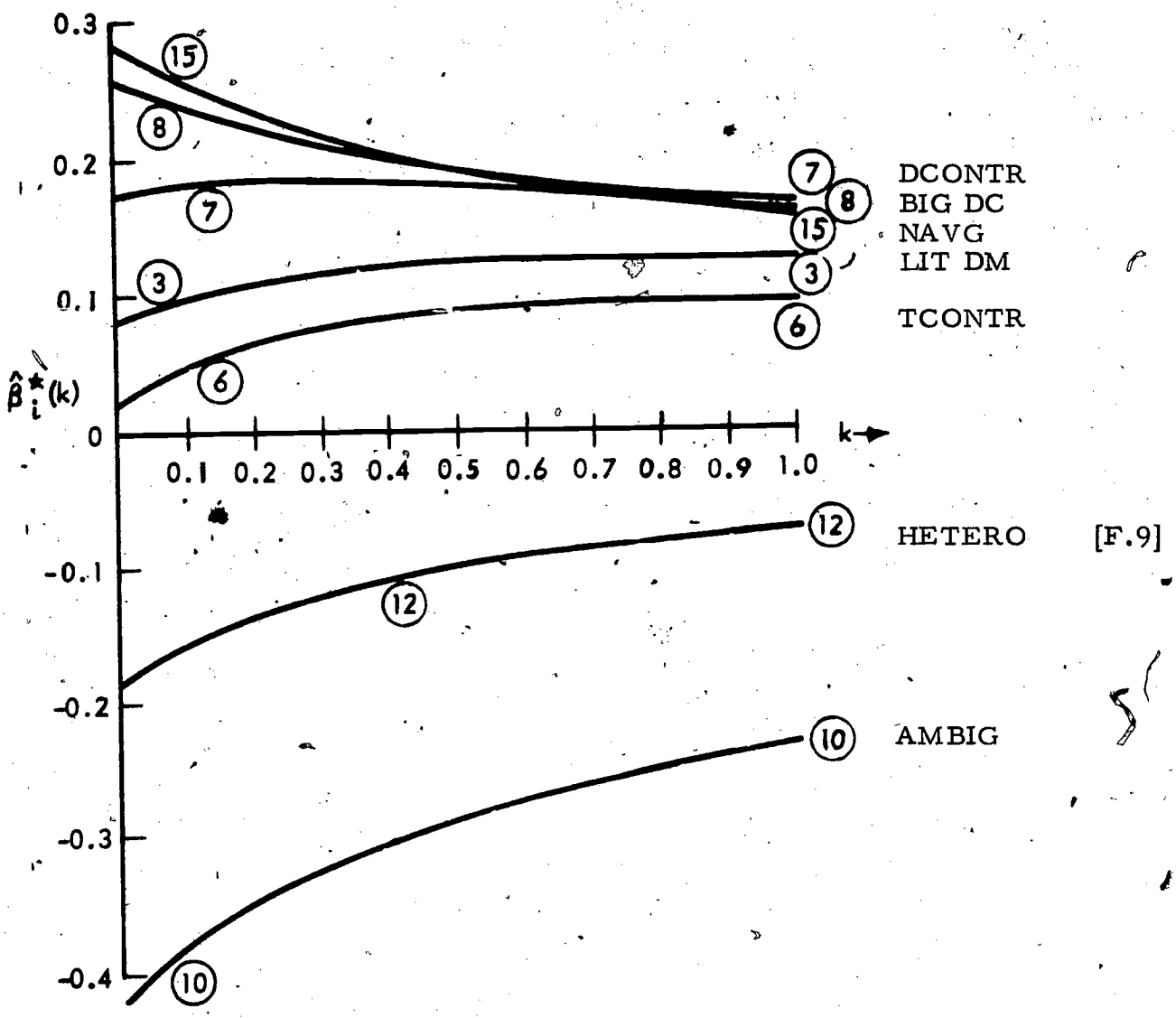
RIDGE TRACE OF FIFTEEN TARGET-BACKGROUND VARIABLES

[From Figure 5-13 in Zaitzeff's (1971) paper]

A ridge regression was run on the remaining seven variables resulting in the new ridge regression pattern shown in figure F.9.* Zaitzeff selected k = 0.4 as the place where the coefficients appear to stabilize. This reduced set of variables accounted for 79 percent of the observed variance (as opposed to 96 percent when all variables are used). However, in the light of what we know about shrinkage and the instability of the original coefficients, the drop is not disturbing. Zaitzeff also noted that by reducing the number to only four easily attainable physical measures: "Detail Contrast", "Target Contrast", and "Target Length" and "Width", 66 percent of the observed variance could still be accounted for. "Target Length" and "Target Contrast" alone accounted for 48 percent of the variance. Several attempts to include interaction or transgenerated terms proved less effective and were aborted. Zaitzeff did not follow the procedure for eliminating variables recommended by Hoerl and Kennard.

This study is one of the better ones attempting to relate target and background characteristics to target acquisition performance and illustrates the advantages of ridge regression over stepwise regression analysis.

*The ridge pattern in figure F.9, showing primarily a relatively orderly compression of coefficient values, is similar to the pattern found when the bias is introduced in an orthogonal design.




RIDGE TRACE IN TARGET-BACKGROUND STUDY FOR ONLY SEVEN VARIABLES [F.9]

[From Figure 5-14 in Zaitzeff's (1971) paper]

## APPENDIX A

## SOURCES OF COMPUTER PROGRAMS NEEDED TO IMPLEMENT THE TECHNIQUES DESCRIBED IN THIS REPORT

Computations used in the techniques in this report - to be implemented - will require the aid of a high-speed computer. Presumably the talents of the experimenter, a computer programmer, and possibly a statistician must be combined to provide the software required for the computations. In this appendix, references are given to sources of computer programs and subroutines needed to support the techniques, along with some general references on the mathematics involved. The original papers are an excellent place to begin to understand the computational requirements of these techniques.

### GENERAL REFERENCES

Some general references* on statistical techniques, mathematics, and computer programs relevant to this report are:

Regression analysis

- Darlington (1968). Multiple regression in psychological research and practice
- Draper and Smith (1966). Applied regression analysis
- Hader and Grandage (1958). Simple and multiple regression analysis
- Kerlinger and Pedhazur (1973). Multiple regression in behavioral research

Matrix mathematics

- Ayres, Jr., (1962). Theory and problems of matrices
- Draper and Smith, Applied regression analysis
- Kerlinger and Pedhazur (1973). Multiple regression in behavioral research. Appendix A.

*Complete references can be found in the Reference list at the end of the complete report.

Computer programs

- Dixon, (1970). BMD: Biomedical computer programs.
- JUG Computer programs directory**
- Kerlinger and Pedhazur (1973). Multiple regression in behavioral research. Appendices B and C.
- Kuo (1972). Computer applications of numerical methods.
- Nie, Bent, and Hull (1970). Statistical package for the social sciences.
- NASA Computer program abstracts*.

## COMPUTER PROGRAMS FOR ADDING DATA POINTS

Computer routines will be needed to calculate the variance of the estimated response at a point or the determinant of the X'X matrix (e.g., Table T.3). Random search or optimization routines are also required for that method of adding data points to improve orthogonality of the undesigned experiment.

### Variance Criterion

If the variance of the estimated response, $V(\hat{y})$ at point $x_o$ on a response surface, is to be used as the criterion, it can be calculated using the equation

---

*Computer Program Abstracts is an indexed abstract journal listing documented computer programs developed by or for the National Aeronautics and Space Administration and the Department of Defense, which are offered for sale through NASA-sponsored Industrial Applications Centers and the Computer Software Management and Information Center (COSMIC).

Computer Program Abstracts is available to the public on subscription or by individual issues from the Superintendent of Documents, United States Government Printing Office, Washington, D.C. 20402, USA. Rates as of August 1975 for an annual subscription were: $3.30 domestic; $4.15 foreign.

**Joint User Group (JUG) of the Association for Computer Machinery Computer Programs Directory was begun in 1971 and updated several times since then. Its purpose is to exchange program documentation among computer user groups. It is published by CCM Information Corp., subsidiary of Crowell Collier and Macmillan, Inc., 909 Third Ave., NY, NY 10022.

employed by Dykstra (1971, p. 683), Draper and Smith (1968, p. 56) and others:

$$V(\hat{y}) \Big| x_o = \sigma^2 x_o' A^{-1} x_o .$$

This expression, in matrix algebra, requires the following operations and computer subroutines to perform the operations:

| | |
|---|---|
| $A^{-1'} = X'X^{-1}$ | Matrix X Multiplication, $X'X = A$* |
| | Matrix A inversion, $A^{-1}$ |
| $x_o$ to $x_o'$ | Vector $x_o$ transposition, horizontal to vertical** |
| $x_o' A^{-1} x_o$ | Matrix/vector Multiplication |

Since the $\sigma^2$ in the above equation (i.e., the error variance) is a constant, it need not be included if the equation is to be used only to compare various data points.

Determinant Criterion

Computer programs for calculating determinants can be found in the general references cited above. Also most computer manufacturers supplying subroutine packages with their systems include programs for calculating the determinant of a matrix and eigenvalues. It must be remembered that the product of the eigenvalues of a matrix equals the determinant. The main problem of selecting a program is not whether it calculates the desired values but does it do it most efficiently.

*X as used here is the matrix formed by the elements of the independent variables, such as in tables T.2 and T.3 in the text.

**$x_o$ is a vector of values, e.g. 4, 6, 3, 7, which represents the levels or coordinates of four variables, A, B, C, D, which thus describes the data point (experimental condition).

The determinant is calculated and printed out in the ridge regression analysis program provided in Appendix C. However, it ordinarily would not be economical to use it to calculate the determinant as a criterion for selecting data points to repair an undesigned experiment.

Random search routines

No general purpose search program is recommended here. However, Mitchell and Miller (1970) employ the same principles to construct D-optimal experimental designs as would be needed to add data points to the matrix of an undesigned experiment using the determinant criterion. D-optimal designs are those for which the determinant of the $X'X$ matrix is maximum, where X is the matrix of independent variables in the usual linear regression model. Mitchell (1974) describes application of the algorithm, DETMAX, to construct D-optimal designs. In his paper, Mitchell states (p. 209): "A FORTRAN listing of DETMAX is available on request to the Computer Sciences Div., Math. and Stat. Research Staff, Union Carbide Corp. Nuclear Division, P. O. Box Y, Building 9704-1, Oak Ridge, Tennessee."

Box and Draper (1971) mention an optimization routine due to Powell (1964) of the direct search type, that maximizes the determinant. However in their application it was only suitable for relatively small designs (np less than 30). Hebble and Mitchell (1972, p. 768) refer to a paper by Spang (1962) for a general discussion of random search procedures.

If the candidate approach is used to add data points, then there is no need for a random search program. Instead the variance at the candidate points or the determinant of the new matrix when each point is added to the original design can be determined and compared — which is the largest?

## COMPUTER PROGRAMS FOR RIDGE REGRESSION ANALYSIS

A convenient method of developing a computer program for ridge regression analysis can be obtained by modifying a conventional multiple regression program. The bias, k, is introduced by adding a constant, k, to the unit diagonal of the correlation matrix and doing a least squares fit on the modified matrix. This process is iterated using different k values until enough ridge coefficients are obtained to plot the data and select the location of stability, meaningfulness, and so forth.

If the University of California Biomedical Data Processing (BMDP) Manual* is available, then a modification of the BMDR2R (Stepwise regression) program, developed by Maryann Hill (1975) of the UCLA Health Sciences Computing Facility, can be used for a ridge regression analysis. The complete article describing this modification has been reproduced in Appendix B.

A more complete computer program for ridge regression was prepared by Mary G. Gallegos of the Display Systems and Human Factors Department, Hughes Aircraft Company, Culver City, California. This is reproduced in Appendix C along with a sample problem in Appendix D. The program, however, was written for a particular problem and, as listed, has dimension statements and other features specific to that problem. With relatively little effort, a competent programmer could use it as a guide to fit other parameters and other computers.

## PRELIMINARY PREPARATIONS FOR THE DATA ANALYSIS

There is no such thing as a completely automated data analysis, only automated aids to data analysis. The computers are available to handle the routine manipulation of numbers, but they are not intended to decide what manipulations are required, what assumptions are to be made, what data is to be fed into them, nor how to interpret the output. These are the responsibilities of the investigator.

*The BMDP Manual of statistical computer programs is available from University of California Press, 2223 Fulton Street, Berkeley, California 94720 at $10.00 per copy.

Without sufficient background himself, the investigator — to employ the techniques proposed in this report — will need the help of a competent computer specialist and statistician. However, while their technical aid can be of considerable value, the investigator must understand exactly what is being done by a computer performing an analysis and why and must not allow critical decisions to be made for him. He is the only one who knows the intended use of the data, its sources, and other critical factors. Any employment of outside talent should be a part of a team effort, with the experimenter in complete control.

## Data Selection

Before any computations are begun, a preliminary analysis of all of the available data should be made to be assured that all should be included in the formal analysis. Particularly with the undesigned experiment where variable upon variable can be added by simply making more measurements (sometimes after the fact), the dangers of a superabundance of inconsequential variables included by an overcurious investigator should be avoided. A preanalysis ought to consider seriously the relevance of the variables under consideration, and even an examination of a table of intercorrelations could suggest which variables are mathematically identical and should not be included twice.

Anscombe (1967, p. 38) has this to say about this matter:

> "In considering multiple regression with large numbers of potential 'explanatory' variables, I would like to echo and extend Dr. Yates's remark on the value of understanding the x-variables first, before seeking to relate any of them to the y-variable. Put very briefly, I have never come across an occasion where one wanted to construct a multivariate relationship without already knowing enough about the x's not to have to do a formal 'search' operation of the multiple regression form. One must be extraordinarily uninformed about one's subject-matter simply to wish to put all 'possible' variables into a multiple-regression black-box and trust to least squares to sort them out. Modern technology may now facilitate almost incomprehensibly vast multiple regression analyses at almost

incredible speeds, but this can only serve to verify the the dictum that the computer shows things to be unnecessary which were previously impossible."

Gorman and Toman (1966, p. 27) discuss the idea of a preliminary examination of the data in a slightly different way, thus:

> "Before variables are selected the data must be examined carefully for statistical difficulties such as split plotting, serious departures from normal distribution of residuals, serial correlation of residuals, and outliers, and for functional difficulties such as the improper choice of X's in the complete equation. Statistical defects are usually spotted by a careful examination of residuals after the equation has been fitted with all k variables present (1, 2, 3). The choice of the X's and their functional forms (i.e., $X_1 = 1/T$, $X_2 = \log (SV)$, etc.) is really a matter of technical judgment by experts in the field from which the data are drawn. Here again, careful examination of residuals can expose improper choices of functional forms."

There is just no substitute for the early application of intuitive judgment by an investigator who knows his business.

## Input Accuracy

Before data is fed to the computer, it should be carefully checked for accuracy. When a great deal of data must be key-punched, it is all too easy for mistakes to be made. Much frustration can be avoided if the investigator takes the extra time at the beginning to inspect a print-out of the input cards himself. It is amazing how easy it is for a person who knows how the information should appear to spot errors that would never be evident to a key-punch operator nor a less-informed technician.

## Program Precision

For handling large multiple regression analyses, it is wise to request the computer be programmed to handle double precision arithmetic. The problem of rounding errors, with serious consequences to the results, in analyses of this type has been discussed by Draper and Smith (1968, pp 144-145)

and by Freund (1963). Neither of the ridge analysis programs listed in Appendices B and C is written in double precision. It has been pointed out that although the consequences of imprecision increase with ill-conditioned matrices, the very process of ridge analysis corrects the sensitive condition.

Then, too, before the consequences of single and double precision can be estimated, it is necessary to know how many bits per word are involved, and this depends on the particular computer. Single precision for one computer may be more precise than double precision for a smaller one.

Draper and Smith (1968, p. 148) point out the value of working from the correlation matrix. They say: "Transforming the regression problem into a form in which it involves correlations is good in general because it makes all of the numbers in the calculations lie between -1 and 1. When numbers are all of this order the adverse effects of roundoff error are minimized." Certainly avoiding sources of imprecision is a matter of prudence. However, the value of double precision for the particular set of data must be weighed against the requirement for a larger computer memory and a possible limitation on the amount of data that could be analyzed.

# APPENDIX B

## RIDGE REGRESSION USING BMDP2R

Maryann Hill

in BMD COMMUNICATIONS, Health Sciences Computing Facility, University of California Los Angeles, February 1975, No. 3

In a regression analysis when the independent variables are highly correlated, the data are often said to be ill conditioned. The resulting regression coefficients may be quite unstable and not useful for future predictive purposes on a new sample. Ridge regression is a technique that is used to "tame" the estimates of regression coefficients, to portray sensitivity of the estimates to the particular set of data being used, and to obtain point estimates with smaller mean square error (although the estimates will be biased).

In the regression model $Y = Z\beta + e$, the ridge estimate of the coefficient vector $\beta$ is

$$\hat{\beta}^* = (Z'Z + \lambda I)^{-1} Z'Y$$

where Z is the matrix (n cases by p variables) of the standardized independent variables and Y is the vector of the standardized dependent variable. The usual least squares estimate is obtained when $\lambda=0$.

Plotting the resulting coefficients for a number of values of $\lambda$ gives an indication of the stability of the coefficients. You hope to find the value of $\lambda$ where the coefficients begin to smooth out and no longer make sudden changes (e.g., switching signs). The estimates of the coefficients eventually approach zero as $\lambda$ goes to infinity.

By adding "dummy" cases to the end of the standardized data file and using the zero intercept option (TYPE=0), you can try this technique with your own data using BMDP2R. The "dummy" cases determine the amount added to the diagonal of the Z'Z matrix. Add one "dummy" case for each of the p independent variables with $\sqrt{(n-1)\lambda}$ as the value of the corresponding variable and zeros for the remaining variables. Note that the Z'Z matrix is (n-1) times the correlation matrix. It is useful to think of ridge regression in terms of the correlation matrix: the size of the value added to the diagonal elements of the correlation matrix is then comparable from problem to problem. In this context values of $\lambda$ less than one are of most interest.

Example: Hoerl (1962) discussed a ridge technique in an article dealing with the measurement of the performance of a chemical process. He specified a relationship between three highly correlated process variables and a response variable, added random noise to the response variable and then analyzed the data. Although the specified relationship had all positive coefficients, the usual least squares solution produced inflated coefficients – one of which was negative. He then applied a ridge technique to these data showing the taming effect on the coefficients and producing solutions closer to the "true" values.

To see the effect of $\lambda = .16$ on the regression coefficients for the Hoerl data, we compute $\sqrt{(n-1)\lambda} = \sqrt{9 \times .16} = 1.2$ and submit the following cards for the HSCF system:

```
// EXEC BIMEDT,PROG=BMDP2R
//TRANSF DD *
      IF(KASE.GT.10)GO TO 1
      X(1)=(X(1)-1.82)/.4022
      X(2)=(X(2)-1.86)/.4088
      X(3)=(X(3)-1.88)/.4492
      X(4)=(X(4)-28.9)/4.0213
   1  CONTINUE
/*
//GO.SYSIN DD *
PROBLEM TITLE IS RIDGE./
INPUT VARIABLES ARE 4.
      FORMAT IS '(4F4.1)'./
REGRESSION DEPENDENT IS 4.
      ENTER IS .001. REMOVE IS 0.
      TYPE IS ZERO./
END/
   11 11 11 223
   14 15 11 223
   17 18 20 292
   17 17 18 270
   18 19 18 285
   18 18 19 304
   19 18 20 311
   20 21 21 314
   23 24 25 328
   25 25 24 340
   12  0  0   0
    0 12  0   0
    0  0 12   0
/*
//
```

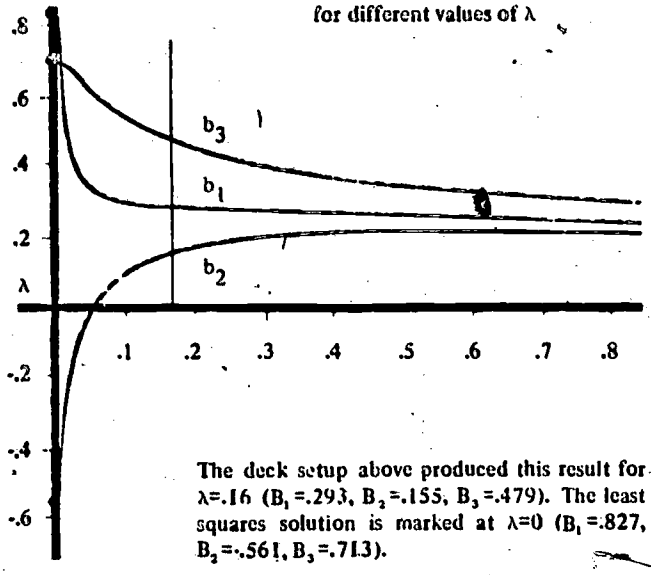
Using the sample $\bar{x}$ and s to standardize the independent variables, X(1),X(2),X(3) and the dependent variable, X(4) for 10 cases

10 cases of raw data

3 dummy cases

Inserting different values of $\sqrt{(n-1)\lambda}$ and rerunning the program, we obtain and plot ridge estimates of the coefficient for each $\lambda$ (see figure). (Note: A number of problems can be run together and the BIMEDT procedure can be used to change the values of the dummy cases in each problem.)

**PLOT OF RIDGE COEFFICIENTS**
**for different values of λ**




The deck setup above produced this result for λ=.16 ($B_1$=.293, $B_2$=.155, $B_3$=.479). The least squares solution is marked at λ=0 ($B_1$=.827, $B_2$=-.561, $B_3$=.713).

You will also want to plot the residual sum of squares versus λ. The most desirable coefficients hopefully will correspond to that value of λ where the residuals have not started to increase rapidly, but yet the values of the coefficients have settled down.

## APPENDIX C

## A SAMPLE PROGRAM FOR RIDGE REGRESSION ANALYSIS

Programmer: Mary G. Gallegos

(Original program was written by Charles Bahun for a GE-635 machine in Basic. The Basic program was converted to FORTRAN IV for a Xerox Sigma V machine, with subsequent modifications. Final program size is 11.8 thousand words, 32 bits per word.)

```
C**********RIDGE REGRESSION PROGRAM********
C ONE MAIN - ONE SUBROUTINE - MAXIMUM MATRIX DIMENSIONS ARE IN THE
C MAIN.  VARIABILITY IN THE DIMENSIONS IS ACCOMPLISHED THROUGH THE
C SUBROUTINE CALL.  SUBROUTINE DONT DOES ALL THE WORK.  SUBROUTINES
C USED FOR MATRIX OPERATIONS AND EIGENVALUES ARE CALLED FROM THE
C USER LIBRARY.  THESE ROUTINES ARE PART OF THE XEROX NUMERICAL
C SUBROUTINE PACKAGE.
C
C
C FUNCTION -- TO DO A RIDGE REGRESSION ANALYSIS ON A GIVEN STUDY.
C COMPUTES:
C         1 MEAN VALUE FOR EACH COLUMN IN MATRIX X
C         2 MEAN VALUE FOR EACH COLUMN IN MATRIX Y
C         3 STANDARD DEVIATION FOR EACH COLUMN IN MATRIX X
C         4 STANDARD DEVIATION FOR EACH COLUMN IN MATRIX Y
C         5 TABLE OF INTERCORRELATIONS
C         6 EIGEN VALUES FOR TABLE OF INTERCORRELATIONS
C         7 DETERMINANTS FOR TABLE OF INTERCORRELATIONS
C         8 LEAST SQUARES AND RIDGE COEFFICIENTS FOR TABLE OF
C           INTERCORRELATIONS
C
C
C
C THE RIDGE REGRESSION PROGRAM CAN HANDLE TWO KINDS OF DATA FOR THE
C TABLE OF INTERCORRELATIONS:
C         1 RAW DATA -- THE X AND Y MATRICES ARE READ IN AND THE TABLE
C           OF INTERCORRELATIONS (MATRIX R) IS COMPUTED
C         2 MATRIX R IS READ IN STRAIGHT FROM CARDS -- THE TABLE OF
C           INTERCORRELATIONS ALREADY COMPUTED.
C
C
C IF DESIRED, CERTAIN COLUMNS READ INTO MATRIX X CAN BE SINGLED
C OUT OF THE COMPUTATION FOR THE TABLE OF INTERCORRELATIONS.  THIS IS
C ONLY AVAILABLE IF THE DATA FOR THE TABLE OF INTERCORRELATIONS IS TO
C BE COMPUTED WITHIN THE PROGRAM.
C
C
C VARIABLES
C         X - INDEPENDENT VARIABLE IN FUNCTION
C         Y - DEPENDENT VARIABLE
C         R - TABLE OF INTERCORRELATIONS (X)
C         RL - EIGEN VALUES
C         DETER - DETERMINANTS
C         Z - TABLE OF INTERCORRELATIONS (Y)
C         E - STANDARD DEVIATIONS FOR Y
C         G - STANDARD DEVIATION FOR X
C         MEAN - MEAN VALUE (USED FOR BOTH X AND Y)
```

```
C        K9 = ERROR FACTORS
C        N = NUMBER OF ROWS IN MATRIX X (OR NUMBER OF DATA SETS)
C        IX = NUMBER OF COLUMNS TO REMAIN IN X MATRIX FOR COMPUTATION
C             IN TABLE OF INTERCORRELATIONS.
C        IS = ACTUAL NUMBER OF COLUMNS READ IN FOR X MATRIX
C        IY = ACTUAL SIZE OF Y MATRIX (NUMBER OF ACTUAL COLUMNS--MAX=9)
C        IN = NUMBER OF K9'S TO BE INTRODUCED(READ IN)
C
C
C*****************************************************************
C
      DIMENSION A(15,60)
      DIMENSION BX(15,3)
      DIMENSION B(15,3)
      DIMENSION BB(15,3)
      DIMENSION C(15,3)
      DIMENSION D(15,15)
      DIMENSION E(3)
      DIMENSION F(3,15)
      DIMENSION FB(15,15)
      DIMENSION FF(15,15)
      DIMENSION G(15,1)
      DIMENSION P(1,15)
      DIMENSION Q(1,3)
      DIMENSION R(15,15)
      DIMENSION RL(15)
      DIMENSION RLL(15)
      DIMENSION RM(3,3)
      DIMENSION T(15,3)
      DIMENSION U(15,1)
      DIMENSION W(15,15)
      DIMENSION X(60,15)
      DIMENSION Y(60,3)
      DIMENSION Z(15,3)
C
C VARIABLES IX AND IY DETERMINE DIMENSION WITHIN SUBROUTINE
C
      READ (105,961) N,IX,IS,IY,IN
      CALL DENT      (A,B,BB,C,D,E,F,FB,FF,G,P,Q,R,RM,T,U,W,X,Y,Z,IX,IY,
     *N,IN,BX,IS)
  961 FORMAT (5(I2,2X))
      END
```

```
      SUBROUTINE DENT(A,B,BB,C,D,E,F,FB,FF,G,P,Q,R,RM,T,U,W,X,Y,Z,IX,IY,
     *N,IN,BX,IS)
      DIMENSION A(IX,N)
      DIMENSION B(IX,IY)
      DIMENSION BB(IX,IY)
      DIMENSION BX(IX,IY)
      DIMENSION C(IX,IY)
      DIMENSION D(IX,IX)
      DIMENSION DETER(15)
      DIMENSION E(IY)
      DIMENSION F(IY,IX)
      DIMENSION FB(IX,IX)
      DIMENSION FF(IX,IX)
      DIMENSION G(IX,1)
```

```
      DIMENSION P(1,IX)
      DIMENSION Q(1,IY)
      DIMENSION R(IX,IX)
      DIMENSION RM(IY,IY)
      DIMENSION T(IX,IY)
      DIMENSION U(IX,1)
      DIMENSION X(N,IX)
      DIMENSION W(IX,IX)
      DIMENSION Y(N,IY)
      DIMENSION Z(IX,IY)
C
      DIMENSION BXX(15)
      DIMENSION RL(15)
      DIMENSION RLL(15)
      DIMENSION WK1(15)
      DIMENSION WK2(15)
      DIMENSION LAB(18)
      DIMENSION IT(2)
      DIMENSION ICH(15)
      DIMENSION S9(3)
      DIMENSION VAL(15)
      DIMENSION IFORM(12)
      DIMENSION TEMP(15)
C
      DIMENSION PEIG(15)
      REAL K9
      REAL MEAN
      INTEGER ANS
      INTEGER ANSR
      INTEGER YES
C
      DATA YES /4HYES /
      DATA IT /4HXXXX,4HYYYY/
      DATA BLANK /4H    /
      DATA IFORM /4H(X,A,4H4,1X,4H,  (,4HF6.3,4H,1X),4H,  (,4HA4,3,
     * 4HX),',4H)  ',3,4H(F6.,4H3,1X,4H))  /
C
C IANS ..EITHER A YES OR A NO.  THIS DETERMINES WHETHER THE TABLE OF
C INTERCORRELATIONS IS TO BE COMPUTED WITHIN THE PROGRAM.
C IF YES, GO TO 1111.
C
      READ(105,957) IANS
      IF(IANS.EQ.YES) GO TO 1111
      DO 830 I=1,IX
      DO 880 J=1,IY
  880 READ(105,980) Z(I,J)
      DO 890 K=1,IX
  890 READ(105,950) R(I,K)
  830 CONTINUE
      GO TO 9988
C
C THIS SECTION LETS ONE PULL OUT CERTAIN COLUMNS IN X MATRIX FOR
C COMPUTING THE TABLE OF INTERCORRELATIONS
C
 1111 NUM=IS=IX
      IF(NUM.EQ.0) NUM=1
      READ (105,965) (ICH(J),J=1,NUM)
      DO 50 I=1,N
      READ (105,900) (TEMP(J),J=1,IS)
       READ (105,950) (Y(I,K),K=1,IY)
```

```
      IX=1
      DO 45 J=1,IS
      DO 40 K=1,NUM
      IF(J.EQ.ICH(K)) GO TO 45
   40 CONTINUE
      X(I,IX)=TEMP(J)
      IX=IX+1
   45 CONTINUE
C
C COMPUTES MEAN VALUES AND STANDARD DEVIATION FOR EACH COLUMN IN THE
C X AND Y MATRICES
C
      IX=IX-1
      DO 30 J=1,IX
   30 P(1,J)=P(1,J)+X(I,J)
      DO 31 K=1,IY
      E(K)=E(K)+ (Y(I,K))**2
   31 Q(1,K)=Q(1,K)+Y(I,K)
      WRITE (108,902) IT(1), (X(I,K),K=1,IX)
      WRITE (108,902) IT(2), (Y(I,K),K=1,IY)
   50 CONTINUE
      CALL S004 (X,W,N,IX)
      CALL S003 (A,X,W,IX,N,IX)
      CALL S003 (A,Y,T,IX,N,IY)
      WRITE (108,954)
      WRITE (108,964) IT(1)
      DO331 J=1,IX
      MEAN=P(1,J)/N
  331 WRITE (108,950) MEAN
      WRITE (108,901)
      WRITE (108,964) IT(2)
      DO 32 J=1,IY
      MEAN=Q(1,J)/N
   32 WRITE (108,950) MEAN
      WRITE (108,901)
      WRITE (108,963) IT(1)
      DO 60 I=1,IX
      G(I,1)=SQRT(    (W(I,I)*(P(1,I)**2)/N)/N)
      WRITE (108,950) G(I,1)
   60 CONTINUE
      WRITE (108,901)
      WRITE (108,963) IT(2)
      DO 70 I=1,IY
      E(I)=SQRT(    (E(I)-(Q(1,I)**2)/N)/N)
      WRITE (108,950) E(I)
   70 CONTINUE
C
C THE TABLE OF INTERCORRELATIONS.  R FOR X MATRIX, Z FOR Y MATRIX.
C
      DO 100 I=1,IX
      DO 80 J=1,IY
   80 Z(I,J)=(T(I,J)-P(1,I)*Q(1,J)/N)/N/G(I,1)/E(J)
      DO 90 K=1,IX
   90 R(I,K)=(W(I,K)-P(1,I)*P(1,K)/N)/N/G(I,1)/G(K,1)
  100 CONTINUE
 9988 CONTINUE
      WRITE (108,954)
      WRITE (108,969)
      WRITE (108,954)
      READ(105,952) (LAB(K),K=1,18)
```

```
      WRITE(108,903) (LAB(K),K=1,18)
      IH=15-IX
      DO 111 I=1,IX
      IF(IH.EQ.0) WRITE (108,909) LAB(I),(R(I,K),K=1,IX),
     * (Z(I,K),K=1,IY); GO TO 111
      ENCODE(4,968,IFORM(3),IDUM),IX
      ENCODE(4,968,IFORM(6),IDUM),IH
      WRITE (108,IFORM)LAB(I),(R(I,K),K=1,IX),(BLANK,K=1,IH),
     * (Z(I,K),K=1,IY)
  111 CONTINUE
      WRITE (108,901)
C
C EIGEN VALUES AND DETERMINANTS FOR THE TABLE OF INTERCORRELATIONS.
C
      CALL CHANGESS (R,FF,IX,0,1)
 2112 CALL EIGEN (FF,FB,IX,1)
 2113 CALL CHANGESS(FF,RL,IX,1,2)
      RLL(1)=1/RL(1)
      DETER(1)=RL(1)
      DO 113 I=2,IX
      DETER(I)=DETER(I-1)*RL(I)
  113 RLL(I)=RLL(I-1)*1/RL(I)
      DO 3113 I=1,IX
      PEIGT=PEIGT+RL(I)
      PEIG(I)=PEIGT/IX
 3113 CONTINUE
      WRITE (108,910)
      DO 114 J=1,IX
  114 WRITE (108,911)       RL(J),RLL(J),PEIG(J),DETER(J)
      WRITE (108,954)
      DO 7115 KK=1,IX
 7115 IF(RL(KK).LT.0) GO TO 999
C
C OPTION TO PAUSE AFTER EACH BETA PRINT
      READ (105,957)  ANS
C
C RIDGE COEFFICIENTS FOR THE TABLE OF INTERCORRELATIONS.
C * 'PREVIOUS' IN PRINT OUT = MATRIX BB
C *'CURRENT' IN PRINT OUT = MATRIX B
C * 'CURRENT B' IN PRINT OUT = MATRIX BX
C * TOTAL UNDER 'CURRENT B' = BXX
C
      DO 200 I=1,IN
      DO 129 J=1,IY
  129 BXX(J)=0.
      WRITE (108,901)
      READ (105,951) K9
      WRITE (108,904) K9
      DO 130 J=1,IX
  130 R(J,J)=1+K9
      CALL MO19 (R,D,IX,IX,0)
      CALL INVERT(D,IX ,DET,WK1,WK2)
      CALL SO03 (D,Z,B,IX,IX,IY)
      IF(I.GT.1) GO TO 140
      CALL MO19 (B,C,IX,IY,0)
      CALL SO04 (C,F,IX,IY)
  140 CALL SO03 (D,D,W,IX,IX,IX)
      CALL SO03 (W,C,T,IX,IX,IY)
      CALL SO03 (F,T,RM,IY,IX,IY)
      WRITE (108,906)
```

```
195.          DO 141 K=1,IY
196.          DO 141 J=1,IX
197.          BX(J,K)=B(J,K)*(E(K)/G(J,1))
198.          BXTEMP=(B (J,K))**2
199.      141 BXX(K)=BXX(K)+BXTEMP
200.          DO 150 J=1,IX
201.      150 WRITE (108,905) (BB(J,K),B(J,K),BX(J,K),K=1,IY)
202.          WRITE (108,966) (BXX(K),K=1,IY)
203.          WRITE (108,901)
204.      C
205.          IF(ANS.NE.YES) GO TO 155
206.          WRITE (102,958)
207.          READ (102,953)  ANSR
208.          IF(ANSR.NE.YES) GO TO 999
209.      C
210.      C LEAST SQUARES FOR THE TABLE OF INTERCORRELATIONS.
211.      C * 'R SQUARED' = H
212.      C * 'ERROR SQUARED' = S
213.      C * 'VARIANCE' = VAR
214.      C * 'BIAS SQUARED' = B9
215.      C * 'RIDGE' = RIDGE
216.      C
217.      155 DO 170 J=1,IY
218.          H=0.
219.          V1=0.
220.          DO 160 K=1,IX
221.          H=H+R(K,J)*Z(K,J)
222.          V1=V1+RL(K)/((RL(K)+K9)**2)
223.      160 CONTINUE
224.          IF(I.EQ.1) S9(J)=1-H
225.      165 S=1-H
226.          VAR=S9(J)*V1
227.          B9=(K9**2)*RM(J,J)
228.          RIDGE=(S9(J)*V1+B9)
229.          WRITE (108,908) J,H,S,VAR,(VAR/S9(J)),B9,(B9/S9(J)),RIDGE,
230.         * (RIDGE/S9(J))
231.      170 CONTINUE
232.          WRITE (108,954)
233.          CALL M019 ( B,BB,IX,IY,0)
234.      200 CONTINUE
235.      C
236.      900 FORMAT (2(5(F10.4,2X),/),5(F10.4,2X))
237.      901 FORMAT (5(1H ))
238.      902 FORMAT (1X,'MATRIX ',A1,/,3(5(1X,F12.6),/))
239.      903 FORMAT (43X,'X-X INTERCORRELATIONS',47X,'X-Y INTERCORRELATIONS',
240.         2 //,6X,15(A4,3X),1X,3(A4,3X))
241.      904 FORMAT (1X,'K VALUE= ',F8.4,///)
242.      905 FORMAT (3(2(F12.3)))
243.      906 FORMAT (1X,'BETA COEFFICIENTS ARE: ',/,1X,
244.         2 3('PREVIOUS',4X,'CURRENT',5X,'CURRENT B',7X))
245.      908 FORMAT (1X,'FOR DEPENDENT VARIABLE ',I2,',R SQUARED IS ',F7.3,/,
246.         1 1X,' ERROR SQUARED = ' F10.4,/,1X,'VARIANCE = ',F10.4,2X,F10.4,
247.         2 /,1X,'BIAS-SQUARED = ',F10.4,2X,F10.4,/,1X,'RIDGE = ',F10.4,
248.         3 2X,F10.4,//)
249.      909 FORMAT (1X,A4,1X,15(F6.3,1X),'|',3(F6.3,1X))
250.      910 FORMAT (1X,'EIGEN VALUES FOR MATRIX R: ',/,12X,'EIGEN VALUE',
251.         2 05X,'RECIPROCAL SUM',5X,'PROPORTION EIGEN',3X,'DETERMINANT',//)
252.      911 FORMAT (9X,15(   F12.4,3X,F12.4,9X,F12.4,9X,F12.4,/) )
253.      950 FORMAT (F12.6)
254.      951 FORMAT(F10.4)
```

```
255.      953 FORMAT(A4)
256.      952 FORMAT (15(A4,1X),/,3(A4,1X))
257.      954 FORMAT (1H1)
258.      957 FORMAT (49X,A4)
259.      958 FORMAT (1X,'CONTINUE ')
260.      960 FORMAT (5X,'EIGENVECTORS FOR MATRIX R: ',/)
261.      962 FORMAT (49X,I2)
262.      963 FORMAT (1X,'STANDARD DEVIATION FOR EACH COLUMN -- MATRIX ',A1)
263.      964 FORMAT (1X,'MEAN VALUE FOR EACH COLUMN -- MATRIX ',A1)
264.      965 FORMAT (15(2X,I2))
265.      966 FORMAT (/,3(12X,F12.6,12X))
266.      968 FORMAT (1,1,I2,'(')
267.      969 FORMAT (30X,'RIDGE REGRESSION ANALYSIS',//,
268.         230X,'DETERMINANT',//,
269.         330X,'TABLE OF INTERCORRELATIONS',//,
270.         430X,'EIGENVALUES',//,
271.         530X,'LEAST SQUARES AND RIDGE COEFFICIENTS',////)
272.      999 END
```

(Questions regarding this program should be referred to the Real Time Simulation Section, Displays and Human Factors Department, Hughes Aircraft Company, Culver City, California 90230.)

## DATA SET-UP

FIRST CARD:

| | |
|---|---|
| Col. 1-2 | Number of observations (N) |
| Col. 5-6 | Number of predictor (X) variables to be analyzed (maximum = 15) (see THIRD CARD, below) |
| Col. 9-10 | Number of predictor (X) variables in the total data set (maximum = 15) |
| Col. 13-14 | Number of dependent (Y) variables (maximum = 3) |
| Col. 17-18 | Number of k values [k's are the constants used to bias the diagonal of the correlation matrix and is referred to as "k9-error factors" in the program.] (see SEVENTH CARD below) |

SECOND CARD:

| | |
|---|---|
| Col. 50-52 | If you are starting the analysis by inputting the raw data values of the predictors and associated performance, write YES |
| or | otherwise |
| Col. 50-51 | If you are going to start by inputting a previously calculated correlation matrix, write NO. |

THIRD CARD: If all of the predictor (X) variables indicated in Col. 9-10 of the first card are to be included in the analysis, leave this card BLANK.

If some of the predictor variables are not to be included in the analysis, then these must be identified by their identification number, i.e., the number of their position on the DATA INPUT cards below (which also could be used as the Label ID required by the FOURTH card).

The predictor variables to be excluded from the analysis are entered:

Cols. 1-2 BLANK

Cols. 3-4 ID number of first predictor to be excluded.

Cols. 5-6 BLANK

Cols. 7-8 (After two blank spaces, the next two spaces are used to enter the ID numbers, 01 to 15, of each variable to be omitted from the analysis, until all are indicated.)

DATA INPUT CARDS -- N SETS -- FOLLOW AT THIS POINT:

For each observation, a set of data input cards for the values of the predictors and the performance is required. The order in which the variables are listed on the cards is fixed and their position can be used as their identification number (see THIRD and FOURTH cards).

For predictor (X) variables, a maximum of three cards can be used with five inputs on each card in a decimal-number format. Ten columns per input, with four decimal places, right-justified. There are two spaces between each input.

For dependent (Y) variables, one per card, maximum of three cards. Twelve columns per input, with six decimal places, right-justified. Y variable cards follow each corresponding set of X variable cards.

FOURTH CARD: Label identification of X predictor variables (maximum = 15) on Table of Data Correlations

Col. 1-4 Label ID for first X variable, right justified

Col. 6-9 Label ID for second X variable, right justified

Col. 11-14 (Continue with four characters per ID and one space between until all are labelled. Leave remainder BLANK.)

FIFTH CARD: Label identification of Y dependent variables (maximum = 3) on Table of intercorrelations.

Col. 1-4 Label ID for first Y variable, right justified

Col. 6-9 Label ID for second Y variable, right justified

Col 11-14 Label ID for third Y variable, right justified

(If fewer than three Y variables, leave extra columns BLANK.)

SIXTH CARD: Do you want the printer to pause after each beta printout?

Col. 50-51 If you do not wish the printer to pause after printing the beta coefficients for each k, write NO. In this case, it will analyze and print out betas for all k values indicated on SEVENTH card.

or Otherwise

Col. 50-52 If you do wish the printer to pause after printing the beta coefficients for each k in order to inspect the values and possibly decide to abort the program from that point on, write YES.

SEVENTH CARD (SET) Use one card of this set for each k factor to be added to the matrix. (Maximum = 14)

Col. 5-10 Use decimal-number format, left justified, beginning with 0, then the decimal, and then the numbers of the k, e.g.,
0.0
0.022
0.06
0.1
0.5

EIGHTH CARD:

Col. 1-4 RFIN

# APPENDIX D

## SAMPLE PRINT-OUT OF RIDGE REGRESSION PROGRAM

These are sample print-outs of critical information in the ridge regression program listed in Appendix C. Included are: raw score data matrix, means and standard deviations of all variables, correlation matrix, eigenvalues, sum of eigenvalue reciprocals, cumulative proportion accounted for by eigenvalues, determinant of the matrix, and for each value of k: $R^2$, error squared, normalized variance, normalized bias squared, normalized ridge, and ridge coefficients for standardized and raw score measures.

```
RUN
MATRIX X
     .000000     50.000000      1.000000      5.000000      8.000000
    1.000000      1.000000      1.000000      1.000000       .000000
    2.000000      1.000000   2500.000000      1.000000      5.000000

MATRIX Y
    2.701360
MATRIX X
     .000000     50.000000      1.000000      5.000000      8.000000
    1.000000      1.000000      1.000000      3.000000       .000000
    2.000000      1.000000   2500.000000

MATRIX Y
    3.335770
MATRIX X
     .000000     50.000000      1.000000
    1.000000      1.000000      1.000000      4.000000       .000000
    2.000000      1.000000   2500.000000      1.000000     20.000000

MATRIX Y
    3.850150
MATRIX X
     .000000     50.000000      1.000000      2.000000      6.000000
```

Partial print out of raw score data for fifteen predictor (X) variables and one response (Y) variable.

```
MEAN VALUE FOR EACH COLUMN -- MATRIX X
     .622642
   31.509430
    2.867924
    2.849056
    6.254717
     .863208
     .207547
     .226415
    2.000000
     .113208
    1.867924
     .754717
 1426.415039
   88.452820
    5.622641

MEAN VALUE FOR EACH COLUMN -- MATRIX Y
    2.963115

STANDARD DEVIATION FOR EACH COLUMN -- MATRIX X
     .484726
   20.822357
    8.956996
    1.323178
```

Means and standard deviations for each predictor (X) and response (Y) variable.

# CORRELATION MATRIX AND EIGENVALUE CALCULATIONS

X-X INTERCORRELATIONS

| | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 01 | 1.000 | -.803 | .162 | -.618 | -.228 | -.209 | -.067 | -.124 | -.314 | .278 | -.263 | -.444 | -.750 | .122 | -.549 |
| 02 | -.803 | 1.000 | .040 | .460 | .158 | .284 | -.068 | -.036 | .235 | -.369 | .268 | .347 | .975 | .054 | .417 |
| 03 | .162 | .040 | 1.000 | .024 | -.037 | .124 | -.088 | -.102 | .000 | -.075 | .071 | -.366 | -.013 | .971 | .020 |
| 04 | -.618 | .460 | .024 | 1.000 | .289 | .427 | .165 | .241 | -.066 | -.229 | .217 | .034 | .379 | .018 | .658 |
| 05 | -.228 | .158 | -.037 | .289 | 1.000 | .105 | -.035 | -.109 | -.190 | -.064 | .060 | .101 | .131 | -.028 | -.011 |
| 06 | -.209 | .284 | .124 | .427 | .105 | 1.000 | -.250 | -.241 | -.212 | -.563 | -.096 | -.148 | .233 | .093 | .194 |
| 07 | -.067 | -.068 | -.088 | .165 | -.035 | -.250 | 1.000 | -.794 | .177 | .578 | .143 | .242 | -.069 | -.066 | .309 |
| 08 | -.124 | -.036 | -.102 | .241 | -.109 | -.241 | .794 | 1.000 | .235 | .599 | .166 | .280 | -.043 | -.077 | .423 |
| 09 | -.314 | .235 | .000 | -.066 | -.190 | -.212 | .177 | .235 | 1.000 | .000 | .389 | .151 | -.213 | .000 | .628 |
| 10 | .278 | -.369 | -.075 | -.229 | -.064 | -.563 | -.578 | -.599 | .000 | 1.000 | .121 | .204 | -.325 | -.056 | -.145 |
| 11 | -.263 | .268 | .071 | .217 | .060 | -.096 | .143 | .166 | .389 | .121 | 1.000 | .144 | .251 | .053 | .306 |
| 12 | -.444 | .347 | -.366 | .034 | .101 | -.148 | .242 | .280 | .151 | .204 | .144 | 1.000 | .359 | -.275 | .078 |
| 13 | -.750 | .975 | -.013 | .379 | .131 | .233 | -.069 | -.043 | .213 | -.325 | .251 | .359 | 1.000 | .011 | .357 |
| 14 | .122 | .054 | .971 | .018 | -.028 | .093 | -.066 | -.077 | .000 | -.056 | .053 | -.275 | .011 | 1.000 | .015 |
| 15 | -.549 | .417 | .020 | .658 | -.011 | .194 | .309 | .423 | .628 | -.145 | .306 | .078 | .357 | .015 | 1.000 |

X-Y INTERCORRELATIONS

| | 01 | 02 | 03 |
|---|---|---|---|
| 01 | .361 | | |
| 02 | -.098 | | |
| 03 | .077 | | |
| 04 | -.446 | | |
| 05 | -.338 | | |
| 06 | -.127 | | |
| 07 | .032 | | |
| 08 | .058 | | |
| 09 | .210 | | |
| 10 | -.064 | | |
| 11 | -.504 | | |
| 12 | -.187 | | |
| 13 | -.052 | | |
| 14 | .084 | | |
| 15 | .021 | | |

EIGEN VALUES FOR MATRIX R:

| EIGEN VALUE | RECIPROCAL SUM | PROPORTION EIGEN | DETERMINANT |
|---|---|---|---|
| 4.1082 | .2434 | .2739 | 4.1082 |
| 3.0127 | .5753 | .4747 | 12.3768 |
| 2.1108 | 1.0491 | .6154 | 26.1248 |
| 1.4312 | 1.7478 | .7109 | 37.3885 |
| 1.2439 | 2.5518 | .7938 | 46.5067 |
| .9243 | 3.6337 | .8554 | 42.9861 |
| .6477 | 5.1775 | .8986 | 27.8436 |
| .5004 | 7.1761 | .9319 | 13.9317 |
| .3865 | 9.7631 | .9577 | 5.3852 |
| .2464 | 13.8208 | .9741 | 1.3272 |
| .1748 | 19.5436 | .9858 | .2319 |
| .1499 | 26.2150 | .9958 | .0348 |
| .0292 | 60.4496 | .9977 | .0010 |
| .0207 | 108.7842 | .9991 | .0000 |
| .0131 | 185.0477 | 1.0000 | .0000 |

99%

Table of intercorrelation among 15 predictor variables and 1 response variable.

Eigenvalues for the above matrix;

Cumulative proportion of variance accounted for by the eigenvalues;

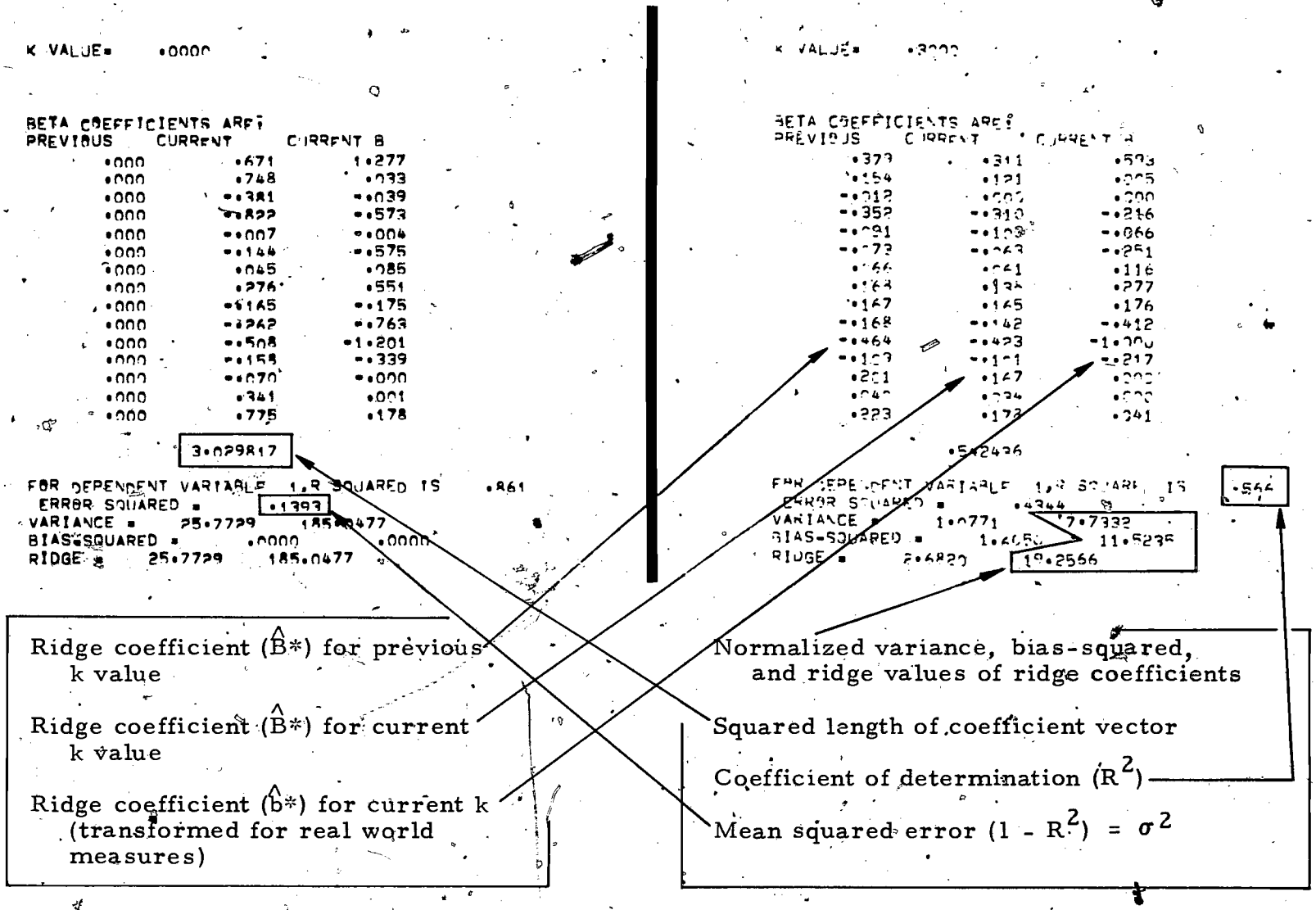
Sum of the reciprocals of the eigenvalues;

Determinant of the X'X matrix (insufficient number of spaces to show a significant figure).

74

81

# ORIGINAL (k = 0, unbiased least squares) AND SUBSEQUENT (k = 0.3) RIDGE ANALYSES



## APPENDIX E

## OTHER APPLICATIONS IN THE DESIGN OF EXPERIMENTS FOR WHICH THE TECHNIQUES DESCRIBED IN THIS REPORT MIGHT BE USED

Hebble and Mitchell (1972) illustrate how the maximum Var $(\hat{y})$ or maximum $|X'X|$ criterion can be used for other important purposes, such as adding data points to:

1. Expand a square region of interest in a second-order model.
2. Alter the model to fit the space.
3. Shift the region of interest.

In these situations they employ candidate points much in the manner proposed by Dykstra (1971).

Mitchell (1974) uses the maximized $|X'X|$ criterion (with a specified linear model and a value of n) to:

1. Exchange data points to improve a design.
2. Determine whether more data points might improve the design.
3. Select a best design made up of a subset of candidate points when limits are placed on the value of n.
4. Supplement "screening" designs (see Simon, 1973) to isolate two-factor interactions.

# REFERENCES


Allen, D. M. Mean square error of prediction as a criterion for selecting variables. Technometrics, 1971, 13, 469-481.

Anscombe, F. J. Topics in the investigation of linear relations fitted by the method of least squares. Journal of the Royal Statistical Society, Series B. 1967, 29, 1-52.

Ayres, Jr., F. Outline of Theory and Problems of Matrices. Schaum's Outline Series. NY: McGraw-Hill, 1962.

Banerjee, K.S., and Carr, R.N. A comment on ridge regression. Biased estimation for non-orthogonal problems. Technometrics, 1971, 13, 895-898.

Beale, E. M. Note on procedures for variable selection in multiple regression. Technometrics, 1970, 12, 909-914.

Box, M. J., and Draper, N.R. Factorial designs, the /X'X/ criterion, and some related matters. Technometrics, 1971, 13, 731-742.

Box. G. E. P., and Hunter, J. S. Experimental designs for the exploration and exploitation of response surfaces. In Chew, V. (ed.) Experimental design in industry. New York: Wiley, 1956, pp 138-190.

Darlington, R. B. Multiple regression in psychological research and practice. Psychological Bulletin, 1968, 69, 161-182.

Davies, O. L. Design and analysis of industrial experiments. N. Y.: Hafner, 1967.

Dixon, W. J. (ed.) BMD: Biomedical computer programs. Berkeley: University of California Press, 1970.

Draper, N.R., and Smith, H. Applied regression analysis. N. Y.: Wiley, 1968.

Dykstra, O., Jr. The orthogonalization of undesigned experiments. Technometrics, 1966, 8, 279-290.

Dykstra, O., Jr. The augmentation of experimental data to maximize /X'X/. Technometrics, 1971, 13, 682-688.

Farebrother, R. W. The minimum mean square error linear estimator and ridge regression. Technometrics, 1975, 17, 127-128.

Freund, R. J. A warning of round-off errors in regression. American Statistician, 1963, 17, 13-15.

Furnival, G.M., and Wilson, R.W., Jr., Regressions by leaps and bounds. Technometrics, 1974, 16, 499-511.

Goldstein, M., and Smith, A.F.M. Ridge-type estimators for regression analysis. Journal of the Royal Statistical Society, Series B, 1974, 36, 287-291.

Gorman, J.W. and Toman, R.J. Selection of variables for fitting equations to data. Technometrics, 1966, 8, 27-51.

Hader, R.J. and Grandage, A.H.E. Simple and multiple regression analysis. In Chew V. (ed.) Experimental designs in industry. NY: Wiley, 1956, pp. 108-137.

Hays, W.L. Statistics. NY: Holt, Rinehart, and Winston. 1963.

Helms, R.W. The average estimated variance criterion for the selection-of-variables problem in general linear models. Technometrics, 1974, 16, 261-272.

Hebble, T.L., and Mitchell, T.J. "Repairing" response surface designs. Technometrics, 1972, 14, 767-779.

Hill, Maryann. Ridge regression using BMDP2R. BMD Communications, No. 3, (UCLA Health Sciences Computing Facility), February, 1975, 1-2.

Hocking, R.R. Criteria for selection of a subset regression: which one should be used? Technometrics, 1972, 14, 967-970.

Hoerl, A.E. Applications of ridge analysis to regression problems. Chemical Engineering Progress, 1962, 58, 54-59.

Hoerl, A.E., and Kennard, R.W. Ridge regression: biased estimation for nonorthogonal problems. Technometrics, 1970a, 12, 55-67.

Hoerl, A.E., and Kennard, R.W. Ridge regression: applications to non-orthogonal problems. Technometrics, 1970b, 12, 69-82.

Joint User Group Computer Programs Directory. NY: CCM Information Corporation, 1971 (First edition) (Later editions have been published).

Kerlinger, F.N., and Pedhazur, E.J. Multiple regression in behavioral research. N.Y.: Holt, Rinehart, and Winston, 1973.

Kiefer, J. Optimum experimental designs. Journal of the Royal Statistical Society, Series B, 1959, 21, 272-319.

Kuo, S.S. Computer applications of numerical methods. Reading, Mass.: Addison-Wesley, 1972.

LaMotte, L. R., and Hocking, R. R. Computational efficiency in the selection of regression variables. Technometrics, 1970, 12, 83-93.

Lindley, D. V., and Smith, A. F. M. Bayes estimates for the linear model. Journal of the Royal Statistical Society, Series B, 1972, 34, 1-41.

Mallows, C. L. Some comments on $C_p$. Technometrics, 1973, 15, 661-675.

Marquardt, D. W. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. Technometrics, 1970, 12, 591-612.

McDonald, G. C., and Schwing, R. C. Instabilities of regression estimates relating air pollution to mortality. Technometrics, 1973, 15, 463-481.

Mitchell, T. J. An algorithm for the construction of "D-optimal" experimental designs. Technometrics, 1974, 16, 203-210.

Mitchell, T. J. and Miller, F. L., Jr. Use of "design repair" to construct designs for special linear models. Math. Div. Ann. Prog. Rept. (ONRL-4661), Oak Ridge National Laboratory, 1970.

National Aeronautics and Space Administration, Computer Program Abstracts, Washington: Supt. Doc., Govt. Print. Office.

Newhouse, J. P. and Oman, S. D. An evaluation of ridge estimators. Santa Monica, Calif: Rand Corp., Report No. R-716-PR, 1971.

Nie, N. H., Bent, D. H., and Hull, C. H. Statistical package for the social sciences. NY: McGraw-Hill, 1970.

Powell, M. J. D. An efficient method of finding the minimum of a function of several variables without calculating derivatives. The Computer Journal, 1964, 7, 155-162.

Simon, C. W. The use of central-composite designs in human factors engineering experiments. Hughes Aircraft Co., Technical Report No. AFOSR-70-6, December 1970. (AD 748 277)

Simon, C. W. Economical multifactor designs for human factors engineering experiments. Hughes Aircraft Co., Technical Report No. P73-326, June 1973 (AD 767 739).

Simon, C. W. Methods of handling sequence effects in human factors engineering experiments. Hughes Aircraft Co., Technical Report No. P74-541, December 1974. (AD A006 240)

Spang, H. A., III. A review of minimization techniques for non-linear functions. SIAM Review, 1962, 4, 343-365.

Theobald, C. M. Generalizations of mean square error applied to ridge regressions. Journal of the Royal Statistical Society, Series B, 1974, 36, 103-106.

Zaitzeff, L. P. Target background scaling and its impact on the prediction of aircrew target acquisition performance. Seattle: Boeing Company, Aerospace Group, Attack Subsystems. Technical Report No. D180-14156-1, December 1971.

## DISTRIBUTION LIST

HQ USAF/RDPS
Washington, D.C. 20330

HQ AFSC/DLSE
Andrews AFB, D.C., 20334

HQ AFHRL
Brooks AFB, Texas 78235

Lt. Colonel A.W. Kibler, Director
Human Resources Research Office
Advanced Research Projects Agency (ARPA)
1400 Wilson Boulevard
Arlington, Virginia 22209

Dr. Tom Gray
AFHRL/FT,
Williams AFB, Arizona 85224

Mr. Robert G. Mills
6570th AMRL/HEB
Wright-Patterson AFB,
Ohio 45433

Dr. W.G. Matheny
President & Technical Director
Life Sciences, Inc.
227 Loop 820 Northeast
Hurst, Texas 76053

Dr. Gordon A. Eckstrand
AFHRL/TR
Wright-Patterson AFB, Ohio 45433

Dr. Marty Rockway
Technical Director
AFHRL, Technical Training Division
Lowry AFB, Colorado 80230

Dr. Roy A. DeGaugh
AFHRL/PE
Lackland AFB, Texas 78236

Dr. H. L. Parris
Chief Scientist
AFHRL/CCS
Brooks AFB, Texas 78235

Dr. Edgar M. Johnson
U.S. Army Research Institute for the Behavioral and Social Sciences
1300 Wilson Blvd,
Arlington, VA 22209

Mr. Vernon E. Carter
Pilot Training Systems
Orgn 3750/62
Northrop Corp./Aircraft Div.,
3901 W. Broadway
Hawthorne, Calif. 90250

Dr. Frederick A. Muckler
Code 311
Navy Personnel Research and Development Center
San Dieto, California 92152

Dr. Stanley N. Roscoe
Bldg. 6, MS D-120
Hughes Aircraft Company
Culver City, California 90230

ERIC
Processing and Reference Facility
4833 Rugby Ave., Suite 303
Bethesda, MD 20014