IMPROVING HEURISTIC REGRESSION ANALYSIS

Floyd A. Miller
The Glidden Company
P. O. Box 389
Jacksonville, Florida 32201

A B S T R A C T

Heuristic Regression Analysis, a new probabilistic predictor selection concept that allows a computer to automatically "learn" the best regression model, has become a practical and economical tool for profit-oriented industry replacing the usual stepwise regression approach. Wide experience with this computer substitute of human problem solving effort has led to substantial improvements of the technique. Starting with forty variables, simple models having three, four or five predictor terms are rapidly located. An "editor" routine, which has been developed to printout only the best three models generated during the learning iterations, significantly reduces the time required by the user to analyse the final models.

IMPROVING HEURISTIC REGRESSION ANALYSIS

Floyd A. Miller
The Glidden Company
P. O. Box 389
Jacksonville, Florida 32201

INTRODUCTION

Market competition establishes a practical need for process improvement in most industries. Process improvements are technical and economic activities leading to the production of better products. In many commercial processes, unknown complex relationships exist between numerous variables which may affect the product quality or cost.

Not too many years ago, engineers could improve these processes by considering them as simple systems. Now, it is more informative to think in terms of multi-variable systems. These processes are frequently changing - creating information about the performance and structure of the system.

Mathematical modeling of complex systems becomes more acceptable in industry with each economic payoff and as simulation development costs are reduced. Model development forces critical examination of existing information and its quantitative relationships, which verifies knowledge of the process - or emphasizes the lack of it.

AVAILABLE TOOLS

The development of a mathematical model can contribute to major system improvements. A systematic approach for mathematical modeling has been developed. With this established concept of study organization, it has been possible to develop steady-state models of entire plants.

Study teams using modern data collection techniques and computer-based data analysis have been making process improvements, obtaining better scheduling and tightening control in the plants. Initial economic gains from the resulting models are encouraging rapid acceptance of the approach.

Analysis is straight forward if the basic mathematical functional form is known from theoretical considerations or some previous knowledge of the system. Of course, in many system studies, no prior information is available and only an empirical approach is feasible. Consequently, various combinations of the available variables are examined to find a representative mathematical model.

For this type of data analysis, the statistical technique of multiple regression has become increasingly important, although the literature continues to discuss improper use of the method (1-5). Other authors are better utilizing their efforts by making significant contributions to the predictor selection problem of regression analysis (6-9).

THE MAJOR PROBLEM

The major problem in linear multiple regression analysis is that of determining the importance or contributions of the individual predictors used to explain the dependent response variable. This problem has usually been approached by using stepwise regression to resolve how many predictors and which predictors should be included in the final model. Basically, this is a decision problem, in situations of uncertainty, with many choices.

For example, a typical linear multiple regression equation without interaction is shown in Table I. Table II shows, for four predictors, the combination of models that are available to explain the response Y. Disregarding the mean value of Y as a model, each predictor can be used alone to predict Y. This is the combination of four things taken one at a time or four models. Next the predictors can be taken pairwise that is, four things taken two at a time or six more models. Finally, the predictors can be used in triples giving four additional models. These possible models, plus the four term model, give a total of fifteen choices to represent the system and predict the response Y. Various statistics are available to judge the adequacy of each model.

It is this decision problem that has led to widespread use (and mis-use) of stepwise regression - particularly since digital computers have become more available. Many companies have developed some version of stepwise regression (8) and, when properly used, it has proved to be an effective tool for studying complex systems.

THE NEED FOR A NEW APPROACH

As the number of available predictors increase, the selection problem becomes difficult - even with a computer. With stepwise regression, only a few models are tested. Since exhaustive computer search is costly, human selection has been required. Of concern has been the engineer, computer, and calendar time required to do the screening of endless combinations of predictors. Usually, process teams select predictors, make computer studies, evaluate model results and re-select predictors for the next study. This trial-and-evaluation search for the right group of predictors has been harmful to the progress of many process improvement activities.

Research has been done of the man-computer interaction that occurs in a study team's use of a computer to find the right set of predictors. It was found that man can efficiently use computers for data analysis using pre-fixed algorithms and logic, such as in stepwise regression programs. However, it was concluded that teams tended to use the computer less efficiently when attempting to learn something about the system which produced the data. This was particularly true when relying on human trial-and-evaluation search. Such a predictor selection problem is like many combinatorial problems in that a direct solution can be costly. Therefore, special emphasis was placed on a different approach to model development using regression analysis.

HEURISTIC REGRESSION ANALYSIS

Under the name of heuristics, much effort is being made to solve problems for which algorithms are not available (10). Basically, a heuristic is a strategy which drastically limits search for solutions in large problems. Even though heuristic problem solving is potentially powerful, in most computer applications, the heuristic aspects of problem solving are carried out almost wholly separate from the algorithmic aspects (11). The heuristic contributions are made by human problem solvers before their problems get to a computer. Although there are many published heuristic efforts, most heuristic programs, if implemented, fail to solve practical problems. This has limited applications in profit-oriented industry.

Heuristic Regression Analysis is a working digital computer program that solves a real problem. It includes a statistical method that allows a computer to "learn" from available data just which predictors should be included in a good model. The details of the approach have been previously published (12) and will only be reviewed briefly here.

Table III shows how to insert a selection routine as the first step towards applying learning to regression analysis. Next, a goal is added to indicate when a solution is satisfactory. Lastly, some reward-and-penalty procedure is needed to transfer information to the next iteration. This is a heuristic method where the results of the last computer selection contributes to the next choice of predictors. The transition of selection information occurs as a result of knowing at the end of each trial whether a model is satisfactory.

DEVELOPMENT OF THE LEARNING CRITERION

Random selection is straight forward. For example, if there are ten possible predictors, each equal-likely, 1/10 is assigned to each as the probability, (Pi) that it is required in a model. Since these probabilities add up to one, a random number between zero and one is generated to select a predictor. Three, four or five predictors chosen in this random fashion will avoid many of the problems others have experienced with stepwise regression.

Having selected the predictors, the summation matrix is rapidly developed and inverted to provide the usual statistics. Of primary concern are three statistics. First, the coefficient of determination, (Ry^2) gives an overall measure of how all predictors, taken as a group, relate to the response. The second statistic is the t-statistic, which measures the contribution of the individual predictor. And, finally, a measure of the linear dependency which distorts the t-statistic is available in the distortion, (Di^2). To date, these appear to be the only useful statistics available from the inverse of the matrix. As shown in Table IV, the t-prime statistic is developed from the last two statistics using a formula previously published (7). The coefficient of determination is combined with the t-prime statistic to provide a reward-and-penalty criterion. The original probabilities (Pi), are multiplied by this criterion, to generate normalized transition probabilities containing all the selection information. As indicated, this powerful discriminating reward-and-penalty criterion is not a simple arbitrary rule, but is based on the statistics available from the matrix inverse rather than some empirical logical criterion (13). Consequently, the computer can be programmed to efficiently discover and learn about the data.

COMPUTER IMPLEMENTATION

The version of the Heuristic Regression Analysis Program previously published (12) developed only three term models from a set of up to twenty possible predictors as shown in the computer block diagram (Table V). Data are read in and the cumulative probabilities calculated. The predictors are selected randomly by comparing a random number with cumulative probabilities until three different ones are selected. A small matrix, requiring only a few degrees of freedom (observations), is developed and inverted to provide the statistics for the reward-and-penalty criterion. Each of the three transition probabilities is multiplied by this criterion to complete the learning process. All probabilities are normalized for the next iteration. The selection probability of each predictor contains the information as to whether a predictor is desirable or not. Heuristic learning is accomplished as the program iterates twice the number of available predictors.

Table VI shows all the models developed for a test case. The actual model is a four term model with the rest of the predictors being random vectors. In this case, the three term program will end up doing all possible combinations of the four predictors, three at a time. A lengthy study of the computer output allows the user to recognize that four terms are necessary.

DEFICIENCIES OF THE PROGRAM

In using this early three term Heuristic Regression Analysis program, two deficiencies become apparent. First, much time was needed to study all the models to find the ones of interest. And second, if additional terms were needed, subsequent computer runs were required.

The first deficiency was removed by an "editor" routine which stores the top three models during the many iterations. For each model, a performance weight (W) is developed using the same statistics as the learning criterion. This weight is used by the "editor" to suppress the printout of all but the best three models. Table VII shows the new computer output of the test problem previously discussed. Table VIII gives the predicted, actual and residuals values of the three-term model with the largest weight.

The second deficiency was removed by extending the system so that it will calculate four term models as well as five term models. These options allow the user a wider range of models for the "editor" to examine. Also, the number of allowable predictors was increased from twenty to the present forty.

Using the same test problem as above, Table IX shows the computer output for the four term option which includes the original equation. The five term option is given in Table X which shows that a fifth term is not needed.

EXPERIENCES WITH HEURISTIC REGRESSION ANALYSIS

Experience with Heuristic Regression Analysis indicates that it averages only one-fourth the computer time as compared to other types of regression approaches. More importantly, most data are analyzed in a single computer run, giving the user earlier solutions. Table XI shows the overall calculation flow of a computer-based Data Analysis System (14) which includes Heuristic Regression Analysis as a problem solving aid to the engineer and scientist. Although teardown regression is available, it has been virtually replaced by the more powerful and economical Heuristic Regression Analysis program. While developed primarily as a model development tool for physical and economic systems, Heuristic Regression Analysis has been extremely helpful in other areas. Unique non-proprietary applications have included pinpointing the predictors basic to making high salaries in Canada's operations research profession (15) and to isolating the causes of traffic deaths in Jacksonville, Florida (16).

SUMMARY AND EXTENSIONS

Much progress on the predictor selection problem in multiple regression Analysis has been made in recent years. Most authors are using some form of stepwise regression. A new predictor selection criterion using a probabilistic learning technique called Heuristic Regression Analysis has such significant technical and economic advantages that it obsoletes stepwise regression. Experiences with practical problems have led to several improvements; namely, allowing three, four and five term models to be developed from up to forty available predictors. An "editor" routine has also been added to suppress the printout of undesirable models, thereby saving much time of the user.

Future extensions of the program include graphic and verbal analysis (17) of the computer results as further assistance for the user.

TABLE I

TYPICAL LINEAR MULTIPLE REGRESSION EQUATION

FOUR PREDICTORS

(1) $$Y = A_o + A_1X_1 + A_2X_2 + A_3X_3 + A_4X_4$$

WHERE:

Y = THE RESPONSE

A_o = THE INTERCEPT

X_i = THE PREDICTORS

A_i = THE MULTIPLE REGRESSION COEFFICIENTS

TABLE II

ALL POSSIBLE COMBINATIONS OF REGRESSION EQUATIONS

SINGLE PREDICTOR $\quad {}_4C_1 = \frac{4}{1} = 4$ MODELS

(1) $Y = A_o + A_1X_1$

(2) $Y = A_o + A_2X_2$

(3) $Y = A_o + A_3X_3$

(4) $Y = A_o + A_4X_4$

TWO PREDICTORS $\quad {}_4C_2 = \frac{(4)(3)}{(2)(1)} = 6$ MODELS

(5) $Y = A_o + A_1X_1 + A_2X_2$

(6) $Y = A_o + A_1X_1 + A_3X_3$

(7) $Y = A_o + A_1X_1 + A_4X_4$

(8) $Y = A_o + A_2X_2 + A_3X_3$

(9) $Y = A_o + A_2X_2 + A_4X_4$

(10) $Y = A_o + A_3X_3 + A_4X_4$

THREE PREDICTORS $\quad {}_4C_3 = \frac{(4)(3)(2)}{(3)(2)(1)} = 4$ MODELS

(11) $Y = A_o + A_1X_1 + A_2X_2 + A_3X_3$

(12) $Y = A_o + A_1X_1 + A_2X_2 + A_4X_4$

(13) $Y = A_o + A_1X_1 + A_3X_3 + A_4X_4$

(14) $Y = A_o + A_2X_2 + A_3X_3 + A_4X_4$

FOUR PREDICTORS

(15) $Y = A_o + A_1X_1 + A_2X_2 + A_3X_3 + A_4X_4$

WHERE:

Y = THE RESPONSE
A_o = THE INTERCEPT
X_i = THE PREDICTORS
A_i = THE MULTIPLE REGRESSION COEFFICIENTS

TABLE III

APPLYING LEARNING TO REGRESSION ANALYSIS

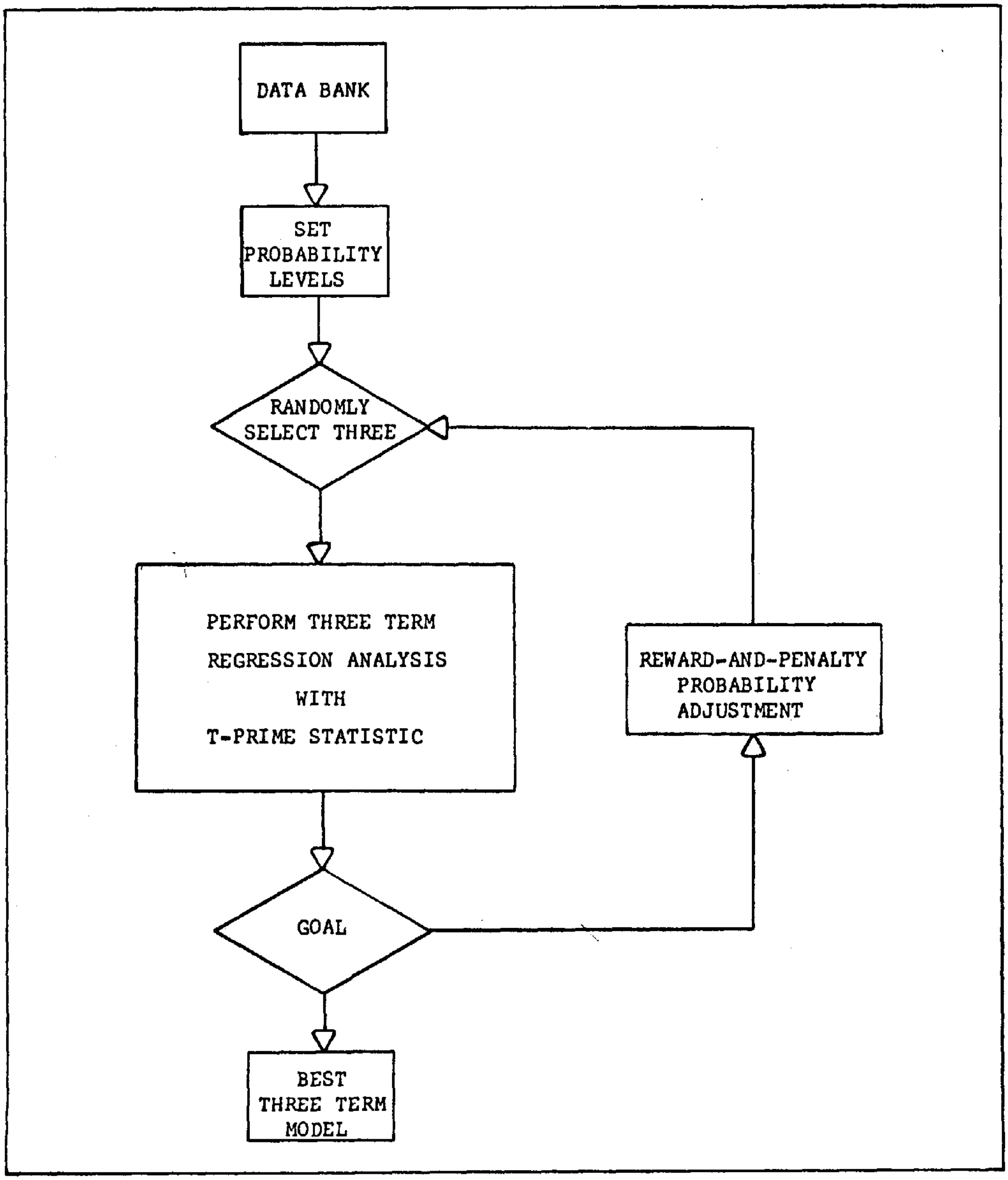


TABLE IV

DEVELOPMENT OF THE LEARNING CRITERION

T-PRIME STATISTIC

$$t'_i = \left[\frac{|t_i|}{1+D_i^2}\right]\left[1 + \left|D_i^2 - \frac{1}{n}\sum_{i=1}^{n} D_i^2\right|\right]$$

WHERE:

t'_i = THE T-PRIME STATISTIC OF THE ITH PREDICTOR IN THE EQUATION.

t_i = THE ABSOLUTE VALUE OF THE T-STATISTIC OF THE ITH PREDICTOR IN THE EQUATION.

D_i^2 = THE DISTORTION FACTOR OF THE ITH PREDICTOR IN THE EQUATION.

TRANSITION PROBABILITY

$$P'_i = (P_i)(Ry^2)(t'_i)$$

WHERE:

P'_i = THE TRANSITION PROBABILITY OF THE ITH PREDICTOR IN THE EQUATION.

P_i = THE ORIGINAL PROBABILITY OF THE ITH PREDICTOR IN THE EQUATION.

Ry^2 = THE COEFFICIENT OF DETERMINATION OF THE EQUATION.

TABLE V

COMPUTER BLOCK DIAGRAM OF HEURISTIC REGRESSION ANALYSIS PROGRAM

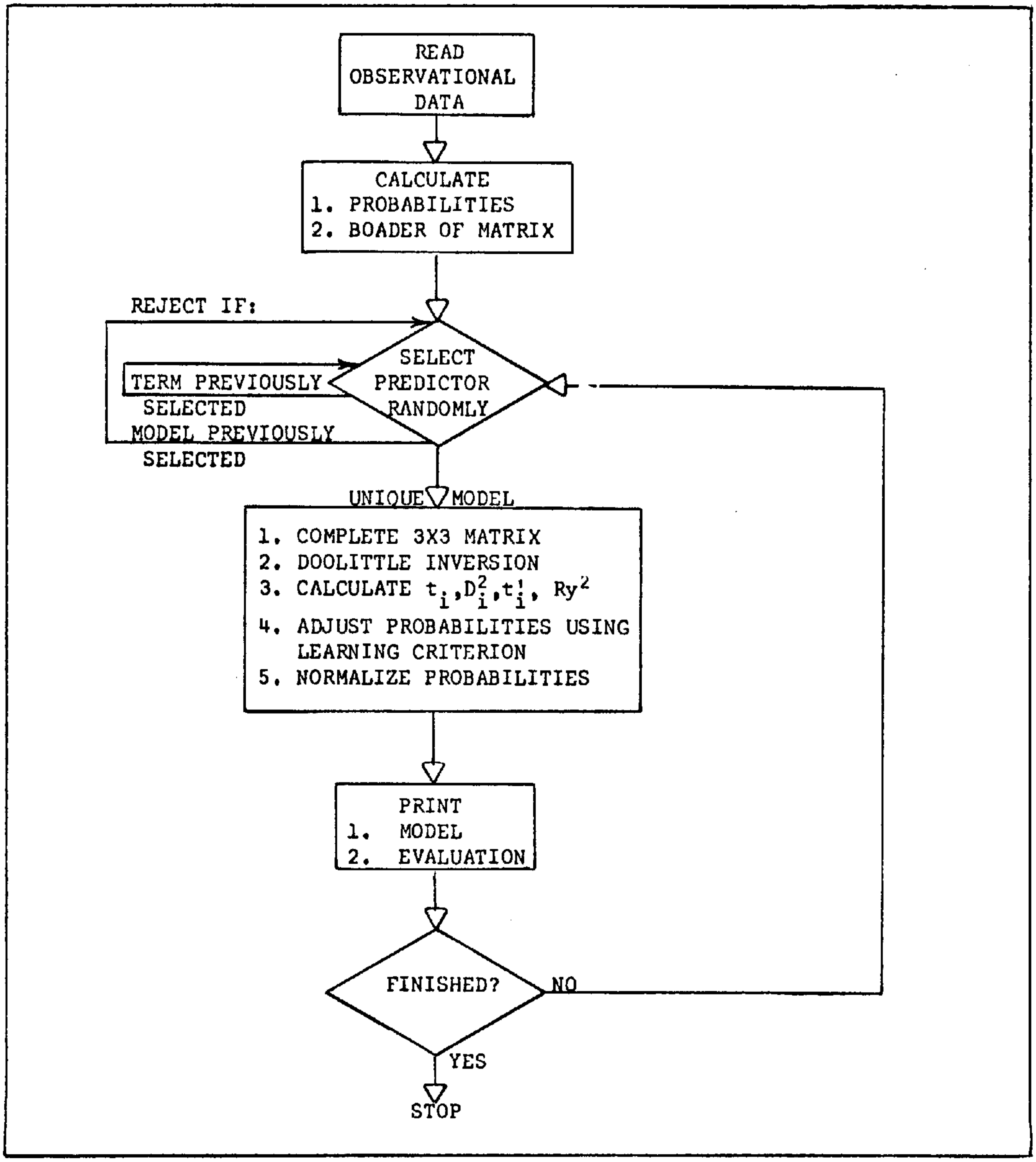


TABLE VI

HEURISTIC REGRESSION ANALYSIS PROGRAM

10 VARIABLES ARE IN HEURISTIC TEST CASE NO. 2 FOR USER F.A.MILLER

$$y = x_1 + x_2 + x_3 + x_4 + \epsilon$$

THREE TERM MODELS

SELECTED VARIABLES	1	2	3	RSQ
COEFFICIENTS	0.974	1.117	1.364	0.72
STUDENT T	5.410	3.190	4.225	
DISTORTION	0.016	0.084	0.079	

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
2	17.	17.	17.	7.	7.	7.	7.	7.	7.	7.										

SELECTED VARIABLES	3	4	7	RSQ
COEFFICIENTS	1.590	0.706	-0.499	0.49
STUDENT T	3.583	2.665	-1.211	
DISTORTION	0.092	0.074	0.043	

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
5	19.	19.	22.	12.	6.	6.	6.	6.	6.	6.										

SELECTED VARIABLES	1	6	8	RSQ
COEFFICIENTS	1.177	-0.065	0.372	0.38
STUDENT T	4.194	-0.242	0.821	
DISTORTION	0.089	0.019	0.106	

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
8	24.	19.	21.	12.	6.	1.	6.	3.	6.	6.										

SELECTED VARIABLES	1	5	9	RSQ
COEFFICIENTS	1.109	0.070	-0.128	0.38
STUDENT T	4.098	0.456	-0.452	
DISTORTION	0.015	0.104	0.101	

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
11	27.	16.	23.	13.	2.	1.	6.	4.	2.	6.										

SELECTED VARIABLES	1	7	10	RSQ
COEFFICIENTS	1.146	-0.077	-0.250	0.41
STUDENT T	4.296	-0.176	-1.427	
DISTORTION	0.035	0.020	0.016	

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
14	27.	17.	24.	14.	2.	1.	1.	4.	2.	8.										

SELECTED VARIABLES	1	3	10	RSQ
COEFFICIENTS	1.039	1.599	-0.049	0.63
STUDENT T	4.934	4.165	-0.335	
DISTORTION	0.032	0.115	0.122	

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
17	27.	18.	27.	14.	2.	1.	1.	4.	2.	3.										

SELECTED VARIABLES	2	3	4	RSQ
COEFFICIENTS	1.533	1.047	0.873	0.64
STUDENT T	3.805	2.735	3.956	
DISTORTION	0.097	0.145	0.079	

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
27	23.	22.	22.	22.	2.	1.	1.	3.	2.	3.										

SELECTED VARIABLES	1	3	5	RSQ
COEFFICIENTS	1.032	1.630	0.028	0.63
STUDENT T	4.940	4.478	0.248	
DISTORTION	0.016	0.023	0.022	

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
30	24.	21.	24.	21.	1.	1.	1.	3.	2.	3.										

SELECTED VARIABLES	1	2	9	RSQ
COEFFICIENTS	1.012	1.499	-0.072	0.56
STUDENT T	4.434	3.476	-0.315	
DISTORTION	0.015	0.027	0.018	

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
33	24.	24.	23.	20.	1.	1.	1.	3.	1.	3.										

SELECTED VARIABLES	1	2	4	RSQ
COEFFICIENTS	1.021	1.651	1.043	0.85
STUDENT T	7.815	6.717	7.722	
DISTORTION	0.013	0.018	0.005	

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
38	23.	23.	21.	23.	1.	1.	1.	3.	1.	2.										

SELECTED VARIABLES	2	3	7	RSQ
COEFFICIENTS	1.194	1.562	-0.444	0.47
STUDENT T	2.413	3.402	-1.045	
DISTORTION	0.111	0.117	0.063	

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
46	22.	24.	24.	22.	1.	1.	1.	3.	1.	2.										

SELECTED VARIABLES	1	3	4	RSQ
COEFFICIENTS	1.053	1.320	0.786	0.79
STUDENT T	6.709	4.701	4.702	
DISTORTION	0.008	0.066	0.059	

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
49	23.	22.	23.	23.	1.	1.	1.	3.	1.	2.										

SELECTED VARIABLES	2	4	10	RSQ
COEFFICIENTS	1.827	1.043	-0.066	0.55
STUDENT T	4.160	4.389	-0.425	
DISTORTION	0.050	0.009	0.050	

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
54	22.	24.	22.	24.	1.	1.	1.	2.	1.	1.										

SELECTED VARIABLES	3	4	6	RSQ
COEFFICIENTS	1.494	0.788	-0.425	0.48
STUDENT T	3.411	2.977	-1.072	
DISTORTION	0.058	0.078	0.023	

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
77	22.	23.	24.	24.	1.	1.	1.	2.	1.	1.										

TABLE VII

HEURISTIC REGRESSION ANALYSIS

10 VARIABLES ARE IN HEURISTIC TEST CASE FOR USER F. A. MILLER

$$Y = X_1 + X_2 + X_3 + X_4 + \epsilon$$

TOP 3 MODELS

SELECTED VARIABLES	1	2	3
COEFFICIENTS	.973	1.117	1.364
T VALUE	5.410	3.190	4.224
DISTORTION	.015	.084	.078
RSQ AND W	.724	2.925	
SELECTED VARIABLES	1	2	4
COEFFICIENTS	1.021	1.651	1.042
T VALUE	7.814	6.717	7.722
DISTORTION	.012	.017	.005
RSQ AND W	.854	6.264	
SELECTED VARIABLES	1	3	4
COEFFICIENTS	1.053	1.319	.786
T VALUE	6.709	4.701	4.701
DISTORTION	.008	.065	.059
RSQ AND W	.789	4.057	

TABLE VIII

HEURISTIC REGRESSION ANALYSIS

10 VARIABLES ARE IN HEURISTIC TEST CASE FOR USER F. A. MILLER

$$Y = X_1 + X_2 + X_3 + X_4 + \epsilon$$

RESIDUALS FOR 40.215 1.021*X 1 1.651*X 2 1.042*X 4

OBS	ACTUAL VALUE	PREDICTED VALUE	DEVIATION	DEVIATION SUM
1	166.00000	142.78458	23.21542	23.21542
2	163.00000	147.95464	15.04536	38.26078
3	187.00000	190.25838	-3.25838	35.00240
4	192.00000	207.87684	-15.87684	19.12556
5	146.00000	167.03064	-21.03064	-1.90508
6	196.00000	177.07129	18.92871	17.02363
7	221.00000	209.46441	11.53559	28.55922
8	182.00000	196.12502	-14.12502	14.43420
9	148.00000	145.97625	2.02375	16.45795
10	169.00000	169.59596	-.59596	15.86199
11	140.00000	134.61396	5.38604	21.24803
12	192.00000	180.43782	11.56218	32.81021
13	109.00000	123.92812	-14.92812	17.88209
14	168.00000	176.55458	-8.55458	9.32751
15	114.00000	110.15437	3.84563	13.17314
16	170.00000	170.61428	-.61428	12.55886
17	162.00000	147.39039	14.60961	27.16847
18	152.00000	156.03804	-4.03804	23.13043
19	117.00000	138.31109	-21.31109	1.81934
20	179.00000	160.38664	18.61336	20.43270
21	90.00000	109.00359	-19.00359	1.42911
22	206.00000	201.92657	4.07343	5.50254
23	191.00000	203.00868	-12.00868	-6.50614
24	209.00000	203.37695	5.62305	-.88309
25	161.00000	167.85453	-6.85453	-7.73762
26	157.00000	162.77081	-5.77081	-13.50843
27	151.00000	144.02657	6.97343	-6.53500
28	162.00000	169.55329	-7.55329	-14.08829
29	163.00000	161.81832	1.18168	-12.90661
30	144.00000	136.31529	7.68471	-5.22190
31	173.00000	174.50367	-1.50367	-6.72557
32	158.00000	154.97448	3.02552	-3.70005
33	229.00000	225.30020	3.69980	-.00025

TABLE IX

HEURISTIC REGRESSION ANALYSIS

10 VARIABLES ARE IN HEURISTIC TEST CASE FOR USER F. A. MILLER

$$Y = X_1 + X_2 + X_3 + X_4 + \epsilon$$

TOP 3 MODELS

SELECTED VARIABLES	1	2	3	4
COEFFICIENTS	.991	1.358	.938	.895
T VALUE	10.634	7.408	5.398	8.950
DISTORTION	.016	.103	.147	.079
RSQ AND W	.928	6.920		
SELECTED VARIABLES	1	2	4	7
COEFFICIENTS	1.038	1.687	1.057	.239
T VALUE	7.917	6.826	7.819	1.095
DISTORTION	.028	.035	.015	.043
RSQ AND W	.860	4.939		
SELECTED VARIABLES	1	2	4	6
COEFFICIENTS	1.019	1.658	1.045	.039
T VALUE	7.680	6.608	7.603	.296
DISTORTION	.013	.026	.009	.012
RSQ AND W	.855	4.671		

TABLE X

HEURISTIC REGRESSION ANALYSIS

10 VARIABLES ARE IN HEURISTIC TEST CASE FOR USER F. A. MILLER

$$Y = X_1 + X_2 + X_3 + X_4 + \epsilon$$

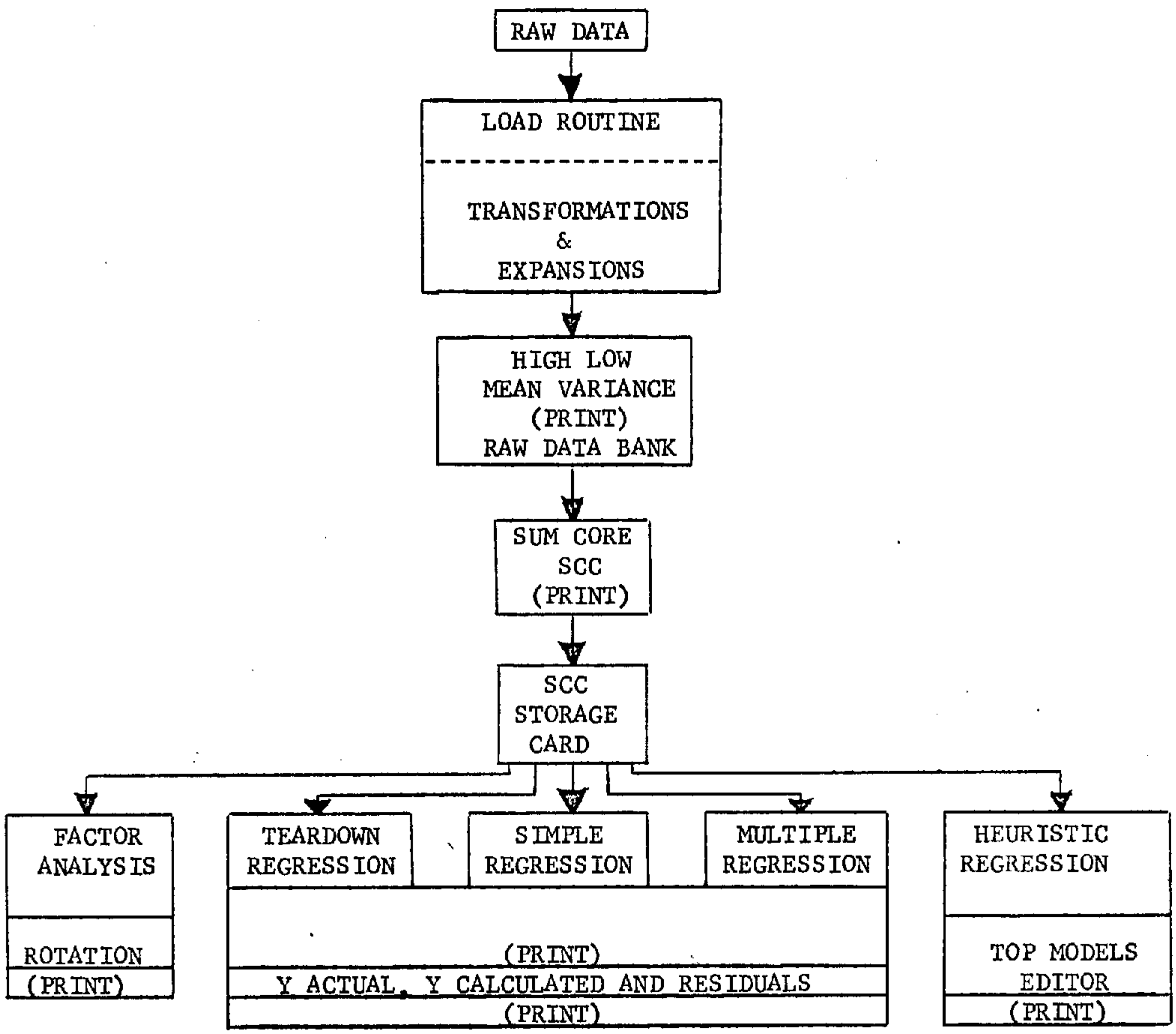
TOP 3 MODELS

SELECTED VARIABLES	1	2	3	4	9
COEFFICIENTS	.994	1.371	.947	.900	.065
T VALUE	10.555	7.371	5.384	8.893	.691
DISTORTION	.017	.113	.152	.084	.033
RSQ AND W	.930	5.664			
SELECTED VARIABLES	1	2	3	4	5
COEFFICIENTS	.985	1.374	.946	.897	-.031
T VALUE	10.392	7.337	5.367	8.865	-.607
DISTORTION	.026	.121	.152	.081	.041
RSQ AND W	.929	5.584			
SELECTED VARIABLES	1	2	3	4	6
COEFFICIENTS	.990	1.365	.938	.897	.038
T VALUE	10.461	7.304	5.317	8.825	.412
DISTORTION	.016	.111	.147	.083	.012
RSQ AND W	.929	5.591			

TABLE XI

DATA ANALYSIS SYSTEM

OVERALL CALCULATION FLOW



REFERENCES

(1) McCune, D. C., "Multiple Regression Analysis" - To Use It Or Not", American Society for Quality Control Conference, Chicago, Illinois, May, 1963.

(2) Kozak, A., "Problems in Multiple Regression Analysis", Biometric Society Conference, June, 1964.

(3) Smillie, K. W., "Remarks on a Recent Paper on Round-Off Errors in Regression Analysis", The American Statistician, October, 1964.

(4) Stout, T. W., Letter to Editor, Instruments and Control Systems, June, 1966.

(5) Hahn, G. J. and Shapiro, S. S., "The Use and Misuse of Multiple Regression", Industrial Quality Control, October, 1966.

(6) Daniel, C., "Factor Screening in Process Development", Industrial and Engineering Chemistry, May, 1963.

(7) Miller, F. A., "Strengthening Stepwise Regression", Association for Computing Machinery Conference, Palm Beach, Florida, June, 1965.

(8) Mallows, C. L., "Some Approaches to Regression Problems", Gordon Research Conference, July, 1965.

(9) Gorman, J. W. and Toman, R. J., "Selection of Variables for Fitting Equations to Data", Technometrics, February, 1966.

(10) Feigenbaum, E. A. and Feldman, J., "Computers and Thought", McGraw-Hill Book Company, 1963.

(11) Licklider, J. C. R., "Man-Computer Partnership", International Science and Technology, May, 1965.

(12) Miller, F. A., "Heuristic Regression Analysis", Operations Research Society of America Conference, Houston, Texas, November, 1965.

(13) Sterling, T., Gleser, M., Haberman, S. and Pollack, S., "Robot Data Screening: A Solution to Multivariate Type Problems in the Biological and Social Sciences", Communications of the Association for Computing Machinery, July, 1966.

(14) Staff, "Users' Manual of the Data Analysis System," Corporate Systems Department, The Glidden Company, October, 1966.

(15) Blackmore, W. R., Cavadies, G., Lach, D., Miller, F. A. and Twery, R., "Annual Salary Survey", Canadian Operations Research Society Bulletin, Fall, 1966.

(16) Miller, F. A., "A Computer Study into the Causes of 1965-1966 Traffic Deaths in Jacksonville, Florida", Police Department Report, August, 1966.

(17) Helm, Carl E., "Simulation Models for Psychometric Theories", Joint Computer Conference, American Federation of Information Processing Societies, Fall, 1965.