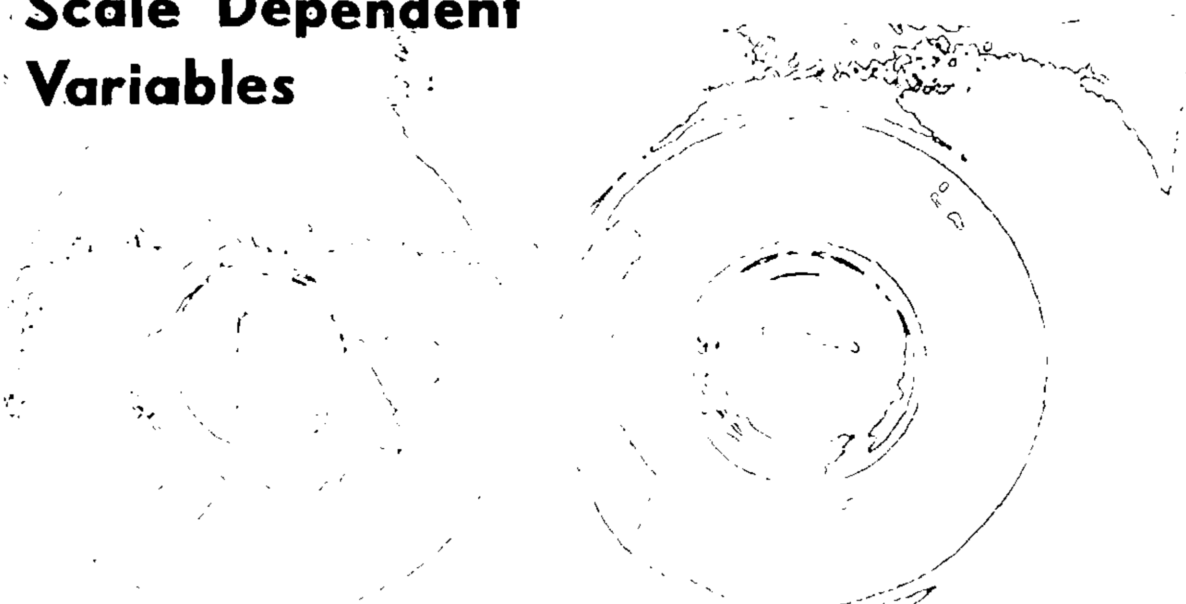
3461

THAID

A Sequential Analysis Program for the Analysis of Nominal Scale Dependent Variables



JAMES N. MORGAN
ROBERT C. MESSENGER

THAID

A Sequential Analysis Program for the Analysis of Nominal Scale Dependent Variables

by

JAMES N. MORGAN

ROBERT C. MESSENGER

Survey Research Center • Institute for Social Research
The University of Michigan
Ann Arbor, Michigan
1973

ISR CODE No. 3461

Library of Congress Card No. 72-619720
ISBN: 0-87944-136-4 Paperbound
ISBN: 0-87944-137-2 Clothbound

Printed by Thomson-Shore
Ann Arbor, Michigan

Manufactured in the United States of America

Acknowledgements

The authors are grateful to the National Science Foundation whose capability-building grant NSF(GI-29904) to the Institute for Social Research allowed us to develop and test this program. Useful criticisms and suggestions were received from Frank Andrews, Charles Cowan, Terry Davidson, Ray Faith, Laura Klem, Lewis Mandell, Patrick O'Malley, Judith Rattenbury and Roger Wright. Patrick also provided us with a useful comparison of THAID with other techniques. Peter Solenberger improved and extended the programming, and Linda Donehue and Carolynn Crandall did most of the typing. The realization that categorical dependent variables need attention was greatly stimulated by Frank Andrews, whose parallel developments of a computer program to handle additive regression solutions appear in a companion monograph <u>Multivariate Nominal Analysis</u>. The realization that neither change nor level can be studied separately from the other in panel studies, so that combinations of them may have to be treated as a categorical dependent variable, was stimulated by the staff working on the OEO Panel Study of Family Income Dynamics, particularly Jonathan Dickinson.

TABLE OF CONTENTS

APPENDIX

CHAPTER I INTRODUCTION

1.1 METHODOLOGICAL APPROACH

Data analysis may be divided into two basic approaches. The first is a searching procedure sometimes referred to as "ransacking" in which an appropriate model is sought to explain the phenomena of interest. In arriving at an appropriate model the analyst decides such things as whether the model is: predictive/causal, linear/non-linear, additive/non-additive, etc., as well as which variables are dependent/independent or nominal/ordinal.

The second approach is that of fitting a specific model to estimate a set of population parameters in an attempt to simplify and summarize the raw data in the context of the model being used. Here the analyst concerns himself with such notions as confidence intervals, significance tests, etc., as well as the calculation of parameter estimates. For a discussion of the philosophical differences see (Blalock, 1970, Chapter IV).

Where the amount of data is severely limited and/or theories to be investigated are relatively specific, then the procedures of model fitting, testing and parameter estimation apply. In much of social science, however, we have a multitude of competing theories not only about the functional form of the model, but even about which predictor variables matter, and why. In addition, with modern surveys we have substantial bodies of data, often in a micro or non-aggregate form. In this situation, the investigator often ends up fitting a whole series of alternative models to his data and selecting among them somehow.

The role of THAID and other model searching approaches is to provide an efficient and effective means for examining a large set of these alternative models. In the THAID program this search is conducted using a sequential binary split algorithm based on one of two possible "loss" or "criterion" statistics, Theta and Delta. (See Binary Split Algorithm, Chapter II.) The functional form of the model is predictive and non-additive* in THAID and what the user extracts from the results is a subset of predictors and possible interactions (non-additivities) which provide an explanatory model for the particular dependent variable chosen.

The authors view this procedure as being useful in two basic ways. First, if one is interested purely in a simple subgrouping of data to provide strong explanatory power without probing for a sophisticated theory of underlying causes and effects, then the THAID approach stands alone as an end in itself. One of the most important advantages of the program for this usage is the simplicity and direct interpretability of the results. The output is simply a set of subgroups, characterized by a few (one to four) independent variable attributes, whose dependent variable distributions are maximally different. An example of this is the identification of market segments for automobile size/price classes and is illustrated in a recent article describing the THAID algorithm (Messenger and Mandell, 1972). The second usage would be as a prior procedure for the identification of predictors and any associated interactions that provide strong explanatory power. Results of THAID ransacking can be used, by fitting the resulting model to fresh data, converting the implied predictors and

---

*In its simplest form this means that the relationship between one of the predictors and the dependent variable is different when one sub-groups the data by (controls for) a second predictor variable.

interactions into a linear model that can be fitted by regression. One such program, MNA, is an appropriate second analysis technique using a series of dummy variable regressions successively on each dummy dependent variable formed by dichotomizing each dependent variable code against all others in turn (Andrews and Messenger, 1973). (See also Chapter II, Comparison With Other Techniques.) Using this approach the body of theory surrounding classical least squares regression is made available to the user after having acquired a workable model for a detailed analysis.

1.2 APPLICABLE DATA

The data for THAID must be categorical (nominal or ordinal). The dependent variable is intended to be nominal but there are instances where ordinal scales may be used. When the dependent variable is ordinal one has a choice of assuming interval properties and proceeding with conventional product-moment correlation/regression techniques (Labovitz, 1970) or forgoing the ordinal properties and focusing only on distribution shape changes. The decision reflects the user's orientation towards either an analysis that reflects "closeness" of prediction or one based on a right/wrong prediction scheme. The independent variables in THAID must be categorical only in the form used by the program. This means that ordinal or nominal predictors may be used directly and that even interval predictors may be given discrete values (such as decile break points) and subsequently used. A given THAID analysis could therefore contain predictors from each of the types of scales. If a predictor is ordinal and one wishes to preserve this, then the provided option will insure ordinality for any splits involving this predictor (See

Chapter II and Appendix A). The data in THAID may be weighted. Weights are considered purely as a device for sampling correction and have no analytical usage. Missing data are not eliminated automatically and are therefore treated as valid categories unless filtered or recoded. (See Appendix A).

1.3 MAJOR RESTRICTIONS

The major program restrictions in THAID are as follows:

A. Number of cases: unlimited

B. 40 predictors or less

C. All variable codes (except weights) in the range (0-9).

Details and other minor restrictions can be found in Chapter IV and Appendix A.

1.4 EXAMPLES OF NOMINAL DEPENDENT VARIABLES

Many phenomena of interest involve what could be thought of immediately as a nominal dependent variable. For example, analyses involving a single-time choice of political party, brand of product, housing status, etc., are candidates. In addition, if one has "panel data" (that is, data at two or more times on the same individuals) then changes in political party, brands of products, etc. may be of interest. The example found in Chapter IV with corresponding output in Appendix C is based on a change in a family composition variable constructed using data on family status in 1968 and 1971.

A second major source for nominal variables is the use of interval scaled variables whose behavior is best explained by a nominal scale treatment. As an example, consider the following set of employment categories for families:

Head works only, regular hours (1800-2200/year).

Head works only, long hours (>2200/year).

Head works only, short hours (<1800/year).

Head and wife work regular hours

Head and wife work long hours

Head and wife work short hours

Other

This variable is a seven category combination of two interval variables whose joint behavior is not appropriately explained by interval scale techniques. Numerous other examples exist in the study of the joint relationship of two or more interval scaled variables using panel data.

1.5 DICHOTOMOUS DEPENDENT VARIABLES

Many interesting dependent variables are dichotomous, such as political party choice (DEMOCRAT/REPUBLICAN), purchasing decisions (BUY/DID NOT BUY), disease status (AFFLICTED/HEALTHY), etc. In this situation THAID is applicable along with other programs* designed to handle interval dependent variables. The authors feel, however, that THAID is more robust in giving results less sensitive to sampling effects in all instances and particularly when the dichotomy is extremely unimodal, i.e., distributed $\frac{2}{3}$, $\frac{1}{3}$ or more unevenly. (See Chapter II Monte Carlo Simulation and Chapter IV Theta or Delta Criterion for more discussion). In addition, the option of expanding the dichotomy to three or more codes such as (DEMOCRAT/REPUBLICAN/OTHER) adds versatility.

*Closest in basic structure and intent is the AID program (Sonquist, Baker and Morgan, 1971)

CHAPTER II - Mathematical Description

The intent of this chapter is to discuss the THAID algorithm, to define all statistics and to discuss important properties. First is a discussion of the binary split algorithm; then a detailed description of the statistics Theta and Delta. Also included are comparisons with other programs and techniques, results of Monte Carlo simulations and statement of an unproven theorem related to the Delta statistic.

Keep in mind that the program partitions a set of data by a sequence of binary divisions, each of which produces two subgroups from a parent group in such a way as to maximize some criterion for that split. Two criteria are allowed, one maximizing the number of correct predictions if each case must be predicted, using the mode of its group, the other maximizing the weighted sum of the differences in proportions in the two subgroups, assuming that one is predicting distributions rather than one case at a time.

2.1 BASIC TERMS*

C_i ≡ Modal code in i^{th} split group

d_r ≡ Number of codes in r^{th} predictor

$\delta_{y/x} = \sum_i N_i \sum_j |P_j - p_{ij}| = \sum_{ij}\sum |fe_{ij} - f_{ij}|$ = Split group Delta

$\delta \doteq \sum_\ell N_\ell \sum_j |P_j - p_{\ell j}| = \sum_{\ell j}\sum |fe_{\ell j} - f_{\ell j}|$ = THAID model Delta

δ_{min} ≡ Minimum delta increase for a split

$\mathcal{F}_j$ ≡ Frequency for j^{th} dependent variable code of sample

F_j ≡ Frequency for j^{th} dependent variable code of parent group

f_{ij} ≡ Frequency for j^{th} dependent variable code of i^{th} split group

$f_{eij} = \frac{(\sum_\alpha f_{i\alpha})\ (\sum_\beta f_{\beta j})}{N_1+N_2}$ ≡ Split group expected frequency

G ≡ Number of dependent variable codes

$H_y = N_s \sum_j P_j \ln P_j$ ≡ Entropy for parent group

$H_{y/x} = N_1 \sum_j p_{1j} \ln p_{1j} + N_2 \sum_j p_{2j} \ln p_{2j}$ ≡ Entropy for split

i ≡ Split group subscript (1 or 2).

j ≡ Dependent variable code subscript ($j=1,2,\ldots,G$).

k ≡ Unique subgroup number

ℓ ≡ Set of terminal groups forming THAID model

M ≡ Modal frequency in sample

M_i ≡ Modal frequency in i^{th} split group

N_i ≡ Frequency in i^{th} split group

---

*The basic terms defined here are used in the entire monograph. Minor points with local usage of notation will have definitions contained in the context of the discussion. For notational simplification, all terms will appear unweighted. The generalization to weighted data is straightforward (mainly replacing frequencies by weights) and is left to the reader. The analytical use of weights is not considered in THAID.

$N\ min \equiv$ Minimum split group frequency

$N_s \quad = N_1 + N_2 \equiv$ Total frequency for split

$\mathcal{P}_j \quad = \mathcal{F}_j/N \equiv$ Proportion of total sample in j^{th} dependent variable code

$P_j \quad = \frac{F_j}{N_s} \equiv$ Proportion of unsplit (parent) group frequency in j^{th} dependent variable code

$P_{ij} \quad = \frac{f_{ij}}{N_i} \equiv$ Proportion of i^{th} split group frequency in j^{th} dependent variable code

$Q \quad \equiv$ Set of terminal group subscripts

$r \quad \equiv$ Predictor subscript

$R \quad \equiv$ Number of terminal subgroups

$\theta_{y/x} \quad = \frac{M_1 + M_2}{N_1 + N_2} \equiv$ Theta (proportion classed correct) for split

$\Theta \quad = \frac{\sum_{\ell} M_{\ell}}{N} \equiv$ Theta (proportion classed correct) for model

$\theta\ min \equiv$ Minimum Theta increase for a split

$V_y \quad = \frac{N_s}{2} - \frac{1}{2N_s} \sum_j F_j^{\ 2} \equiv$ Gini variance for parent group

$V_{y/x} \quad = \frac{N_1}{2} - \frac{1}{2N_1} \sum_j f_{1j}^{\ 2} + \frac{N_2}{2} - \frac{1}{2N_2} \sum_j f_{2j}^{\ 2} \equiv$ Gini variance for split

2.2 BINARY SPLIT ALGORITHM

The binary split algorithm is the basis for the THAID program, which makes it natural for notation to be defined in the context of this algorithm. The generalization of statistics to other models, such as bivariate association in two-way tables, is straightforward and most details are left to the interested reader.

The primary idea behind the binary split algorithm is that data be sequentially partitioned into two parts, determined by an independent variable's codes, so as to optimize locally a criterion function for the dependent variable. Since THAID is designed for nominally scaled dependent variables, criterion functions were chosen to be appropriate only for this scale. This is the major difference between AID and THAID. (See details under Relationship to AID.)

The algorithm begins by performing "permissible" splits on the entire data set, each time evaluating the criterion function (either Theta or Delta - discussed below). A permissible split is a binary partition of data based on independent variable (predictor) codes satisfying the following constraints:

A. Predictor ordinality (optional): If a predictor is felt to be ordinally scaled, and for this or other reasons the user wishes to preserve ordinality within groups, the user may choose to limit permissible splits in this manner. For example, if a predictor has codes {1,2,3} and this option is chosen, the splits {1}, {2,3} and {1,2}, {3} are permissible while the split {1,3},{2} is not.

Both split criteria make no use of the nominal or ordinal scale designation of predictors other than in deciding the permissibility of a split. Thus even strictly ordinal predictors need not be specified as such and strictly nominal predictors may be treated ordinally if a specific grouping is desired. It is much more economical to treat predictors ordinally because of the large reduction in permissible splits, however, so that users should specify nominal predictors only when needed.

B. Minimum Group Frequency: The user can specify a minimum group frequency, such that a split is not permissible if one or both split groups has frequency less than the stated value. THAID compares the given minimum, Nmin, against the truncated half of the group frequency, $[\frac{N_s}{2}]$, and if

$$[\frac{N_s}{2}] < N_{min} \tag{2-1}$$

no split on this group is permissible and it becomes a "terminal group." (Terminal groups remain unsplit and form the final model in the THAID tree.) If (2-1) does not hold, splits are examined for

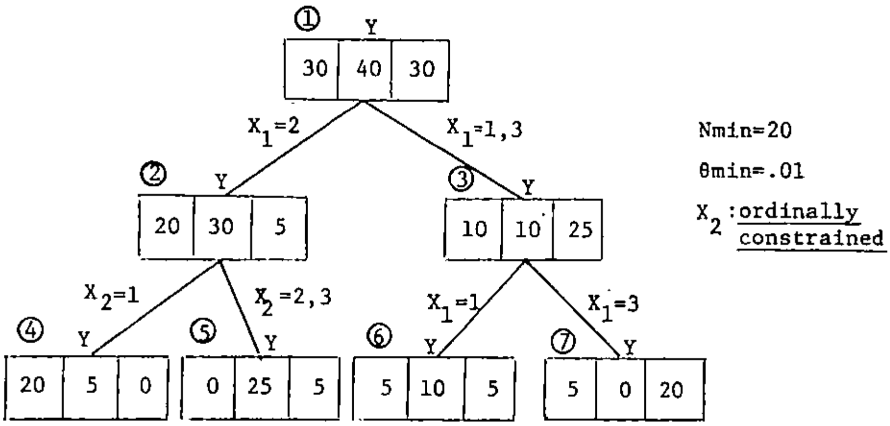
$$N_1, N_2 > N_{min} \tag{2-2}$$

and a split may still be not permissible.

C. Minimum criterion improvement: Here the user specifies a minimum criterion improvement for the split, θmin or δmin, and a split is declared permissible only if these values are exceeded. (See Appendix A - Set-up Instructions for details.)

After all permissible splits have been tried the one having the largest criterion improvement is used as the initial split. Then the search of permissible splits for maximum criterion improvement is repeated on each of the two new groups and assuming permissible splits continue to occur the number of new groups grows as the sequence {2,4,8,...}. The procedure terminates when all groups unsplit are terminal groups, or in other words, when no permissible splits remain. The terminal groups then represent the THAID model and the entire sequential array of groups is called the THAID "tree." Figure II-1 shows a simplified THAID tree for two predictors X_1, X_2 and a dependent variable Y which is assumed trichotomous and has marginal frequency distributions shown within each subgroup rectangle.

Figure II-1 THAID TREE



Notice here that the terminal groups 4-7 do not have permissible splits because of constraint (2-1); X_1 is used to split both the original data, group 1, and group 3. For convenience of identification, the groups are numbered (both in Figure II-1 and in the output) sequentially first from left to right and then top to bottom in the THAID tree. Discussion of the split criteria is left for the following sections and the THAID tree shows modal prediction using Theta, θ. (Delta, δ, could also have been used.)

It is worth mentioning that an early version of the THAID program used a splitting algorithm that sequentially partitioned the data into as many new groups as codes of the independent variable. In effect, this is the same as forming a multi-dimensional contingency table by locally optimizing the criterion at each dimension. The criterion is much more rapidly increased using the procedure but the results become unstable after only a few dimensions. More specifically, Monte Carlo simulation of the null hypothesis shows that sampling errors have sharp prohibitive increases for the number of groups greater than 8% of the sample size. This is quickly reached even for large data sets because the number of sub-groups, S, is equal to the product of the number of predictor codes, d_r. That is,

$$S = \prod_r d_r \tag{2-3}$$

For this reason, the algorithm was abandoned in favor of the binary split.

There are two basic statistics used in THAID, Theta, θ, and Delta, δ. At present either can be used as split criteria and the user must decide which is more appropriate for his analysis.*

2.3 THETA

The Theta statistic is defined as the proportion of the sample classed correctly when using the optimal-prediction-to-the-mode strategy. In univariate prediction this is simply the ratio of modal frequency to total frequency and is denoted θ_y. In bivariate prediction, one predicts to the mode of the marginal frequency distributions associated with the predictor and the success probability is denoted $\theta_{y/x}$. For a binary split we have

$$\theta_{y/x} \equiv \sum_{i=1}^{2} \left(\frac{N_i}{N_s}\right)\left(\frac{M_i}{N_i}\right) = \frac{1}{N_s}\,(M_1+M_2) \tag{2-4}$$

*For a discussion of considerations determining the choice between Theta and Delta, see Chapter IV, Theta or Delta Criterion.

or the sum of products of the probability of being in the i^{th} group and the probability of correct placement given the i^{th} group.

Note that in order to have a non-zero Theta improvement (Constraint C) it is necessary that the modal <u>codes</u> C_1, C_2 be different. Otherwise $M_1+M_2=M$. This is illustrated in Table II-1.

<u>TABLE II-1</u>

		Y: 1	Y: 2	Total
X	1	40	20	60
	2	30	10	40
	Total	70	30	100

Because $C_1=C_2=1$ we have both θ_y and $\theta_{y/x}$ for the only split {1}, {2} equal to .70.

To illustrate the difference between a binary split and the complete bivariate table (equivalent to a split with d_r groups) consider TABLE II-2.

<u>TABLE II-2</u>

		Y: 1	Y: 2	Y: 3	Total
X	1	20	10	10	40
	2	10	20	10	40
	3	5	5	10	20
	Total	35	35	30	100

Here predictor X and dependent variable Y both have 3 codes and the total frequency is 100 displayed as a 2-way table. We have

$$\theta_y = \frac{35}{100} = .35 \tag{2-5}$$

$$\theta_{y/x} = \frac{20+20+10}{100} = .50 \tag{2-6}$$

Also there are 3 binary split Thetas, namely,

$$\theta_{y/x}(\{1\}, \{2,3\}) = \frac{20+25}{100} = .45 \tag{2-7}$$

$$\theta_{y/x}(\{1,2\}, \{3\}) = \frac{30+10}{100} = .40 \tag{2-8}$$

$$\theta_{y/x}(\{1,3\}, \{2\}) = \frac{25+20}{100} = .45 \tag{2-9}$$

Equation (2-6) gives a higher value than either (2-7), (2-8) or (2-9). In practice, however, the binary split algorithm captures the majority of the bivariate relationship and if a significant portion remains a second split may still occur on the same variable as in Figure II-1. From here on $\theta_{y/x}$ will refer to a binary split of a group with univariate θ_y.

This statistic is based on the same prediction scheme as Goodman and Kruskal's λ_y statistic, (Goodman and Kruskal, 1954). In fact,

$$\lambda_y = \frac{\theta_{y/x} - \theta_y}{1-\theta_y} \tag{2-10}$$

In addition, since λ_y is a linear transform of $\theta_{y/x}$ and $(\frac{1}{1-\theta_y}) > 0$ both rank order predictors identically and are hence <u>equivalent</u> criteria. Furthermore, maximizing (2-4) for each split maximizes $\theta_{y/x}, \lambda_y$ and the overall model prediction since the sum of modes determines the maximum of each. The choice of θ over λ is one of convenience of direct predictive interpretation over a proportional reduction of error interpretation.

2.4 DELTA

The Delta criterion is based on the simple notion that one should find split groups whose probability distributions differ maximally from the original group and hence from each other. To account for differing split group frequencies the probability distribution "distances" are frequency-weighted. Also, because of sampling instabilities and accentuation of large departures from expected value (often due to non-normally distributed errors such as coding, key punching, etc.) inherent in chi-square and related statistics, it was decided to use the absolute value distance metric. This led to the formula

$$\delta_{y/x} = N_1 \sum_{j=1}^{G} |P_j - p_{1j}| + N_2 \sum_{j=1}^{G} |P_j - p_{2j}| \qquad (2\text{-}11)$$

A geometric interpretation of Delta can be gained by visualizing the probability distributions $[P_i]$, $[p_{1i}]$ and $[p_{2i}]$ as points in G-dimensional space. For G=3, these points are contained within the unit cube and $\delta_{y/x}$ is the frequency weighted sum of absolute value distances between $[P_i]$, $[p_{1i}]$ and $[P_i]$, $[p_{2i}]$.*

In a form more analogous to chi-square, one can first think of the split group frequency data as a 2XG bivariate table (rows indicate the split group

*The absolute value metric can be thought of as the minimum distance when following a path always parallel to an axis of the space and is sometimes referred to as the "city block" metric.

and columns the dependent variable code). If the expected frequency is then defined by

$$fe_{ij} = \frac{(\sum_{\alpha} f_{i\alpha})\ (\sum_{\beta} f_{\beta j})}{N_s} \tag{2-12}$$

as in chi-square, then equation (2-11) can be represented as

$$\delta_{y/x} = \sum_{i=1}^{2} \sum_{j=1}^{G} |f_{e_{ij}} - f_{ij}| \tag{2-13}$$

since

$$N_i p_{ij} = f_{ij} \quad \text{and} \tag{2-14}$$

$$N_i P_j = f_{e_{ij}} \tag{2-15}$$

A convenient computational property of Delta is seen by noting that for any dependent variable code j

$$F_j = f_{1j} + f_{2j} \quad \text{or} \tag{2-16}$$

$$N_s P_j = N_1 p_{1j} + N_2 p_{2j} \tag{2-17}$$

substituting $(N_1 + N_2)$ for N_s and rearranging terms gives

$$N_1 (P_j - p_{1j}) = N_2 (p_{2j} - P_j) \tag{2-18}$$

taking absolute values of both sides and summing on j gives

$$N_1 \sum_{j=1}^{G} |P_j - p_{1j}| = N_2 \sum_{j=1}^{G} |P_j - p_{2j}| \tag{2-19}$$

Therefore (2-13) can be rewritten as

$$\delta_{y/x} = 2N_1 \sum_{j=1}^{G} |p_j - p_{1j}| = 2N_2 \sum_{j=1}^{G} |p_j - p_{2j}| \tag{2-20}$$

Computationally, and for analysis of distributional properties, this means one need be concerned only with the term corresponding to one split group. This saves nearly 50% of the THAID algorithm binary split execution time.

To illustrate the calculation of δ consider TABLE II-3.

TABLE II-3

Y

		1	2	3	Total
	{1}	20	12	18	50
X	{2,3}	16	14	20	50
	Total	36	26	38	100

By the original formula (2-11)

$$\delta_{y/x} = 50[|.36-.40|+|.26-.24|+|.38-.36|] +$$
$$50[|.36-.32|+|.26-.28|+|.38-.40|] = 8$$

and by (2-13)

$$\delta_{y/x} = |18-20|+|13-12|+|19-18| +$$
$$|18-16|+|13-14|+|19-20| = 8$$

Note the term by term equality between corresponding dependent variable codes of split groups.

In order to provide a statistic in the (0-1) range that is comparable across differing sample sizes, the printed version of $\delta_{y/x}$ is divided by its maximum value achieved when each marginal frequency distribution has one non-zero frequency. In this situation the dependent variable is perfectly predicted and this statistic, δ', is given by

$$\delta'_{y/x} = \frac{\delta_{y/x}}{2N(1-\frac{\sum N_i^2}{N})} \qquad (2\text{-}21)$$

2.5 RELATIONSHIP TO AID

The THAID program was written with the idea of generalizing the successful AID procedure to nominal dependent variables. The present version of THAID could be viewed as a simplified version of the present AID (Sonquist, Baker and Morgan, 1971) with the Theta/Delta criteria replacing the explained sum of squares criterion given by

$$E_{y/x} = N_1(\overline{Y}_1-\overline{Y}_S)^2 + N_2(\overline{Y}_2-\overline{Y}_S)^2 \qquad (2\text{-}22)$$

where

$\overline{Y}_S \equiv$ Mean for parent group

$\overline{Y}_i \equiv$ Mean of subgroup i

For dichotomous dependent variable analysis, both THAID and AID are applicable as mentioned earlier. However, the trees will in general be different since the THAID criteria and explained sums of squares criterion do not rank order predictors identically. Roughly speaking, AID tends to split giving outlying subgroups of small size much more than does THAID.

2.6 MONTE CARLO SIMULATIONS

In order to compare the Theta and Delta criteria with other potential split criteria the following null hypothesis simulation test was devised. A A data set was created containing 1000 cases of a variable generated uniformly along a ten-point (1-10) scale. Next, the data were partitioned into two groups having equal probability of receiving each case. The following statistics were then computed:

A. Chi-square: χ^2

B. Delta: $\delta_{y/x}$

C. Entropy: H (Attneave, 1959)

D. Gini Variance: = V (Light and Margolin, 1971)

E. Theta: $\theta_{y/x}$ where

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{10} \frac{(f_{e_i} - f_{ij})^2}{f_e} \tag{2-23}$$

$$\delta_{x/y} = \sum_{i=1}^{2} \sum_{j=1}^{10} |f_{e_{ij}} - f_{ij}| \tag{2-24}$$

$$H_{y/x} = \frac{N_s \sum_j P_j \ln P_j - [N_1 \sum_j \ln p_{1j} + N_2 \sum P_{2j} \ln p_{2j}]}{N_s \sum_j P_j \ln P_j} \tag{2-25}$$

$$V_{y/x} = \frac{N_s}{2} - \frac{1}{2N_s} \sum_j F_j^2 - [\frac{N_1}{2} - \frac{1}{2N_1} \sum_j f_{1j}^2 +$$

$$\frac{N_2}{2} - \frac{1}{2N_2} \sum_j f_{2j}^2] \tag{2-26}$$

The procedure was replicated 10 times giving a total of 10 values for each statistic which were replaced by ranks and the Kendall's tau b rank correlations computed. Table II-4 gives these correlations.

TABLE II-4

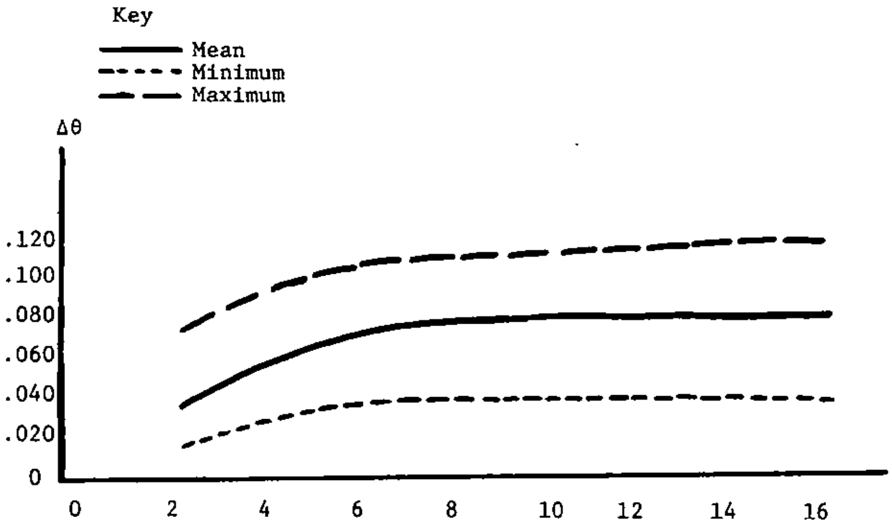
	δ	H	V	χ^2
H	0.82			
V	0.87	0.96		
χ^2	0.82	1.00	0.96	
θ	0.24	0.24	0.29	0.24

The basic finding appears to be that the triad of χ^2, H and V are essentially the same in split selection with δ moderately related to all three, while θ is only weakly related to the other four.

The weak .24 correlation between θ and δ indicates that two quite different trees may well be produced on the same data using the two different criteria. Thus a careful decision weighing the considerations mentioned in Chapter IV Theta or Delta Criterion is warranted.

A second Monte Carlo simulation was conducted to indicate the sampling properties of the Theta criterion under the null hypothesis. The test involved generating a statistically independent data set and observing the incremental Theta growth with the number of subgroups. The dependent variable was uniformly dichotomous and eight nominal predictors were used, each uniformly distributed on a five point scale. The sample size was 500 and the test was replicated 10 times. The results appear in Figure II-2.

Figure II-2



SPLIT GROUPS

The choice of distribution type, nominal predictors, number of codes and number of predictors was made to make the results conservative and hence an upper bound on the growth of $\Delta\theta$. Further simulations have indicated that the curves of Figure II-2 are closely proportional to the square root of the reciprocal sample size. Hence users can rescale by $\sqrt{\frac{500}{N}}$ to have curves appropriate for their sample.

2.7 DELTA THEOREM

An unproven theorem of moderate importance can be stated as follows. If one maximizes the Delta statistic for a given split then this implies stepwise maximization of the Delta statistic for the current model.* Mathematically

*In other words, maximizing the split Delta simultaneously gives an optimal Delta value on the THAID tree including that split.

this would mean that maximizing (2-13) simultaneously maximizes

$$\delta = \sum_{\ell j}\sum |fe_{\ell j} - f_{\ell j}| \tag{2-27}$$

where

$$fe_{\ell j} = N_\ell \mathcal{P}_j \tag{2-28}$$

Parallel theorems for the AID criterion and for the THAID Theta criterion can be easily proven. For example, one can show that the explained sum of squares for the AID model at any stage

$$E_M = \sum_\ell N_\ell (\overline{Y}_\ell - \overline{Y})^2 \tag{2-29}$$

can be written as

$$E_M = E_{\overline{M}} + E_{y/x} \tag{2-30}$$

where $E_{\overline{M}}$ is the E prior to the latest split (whose explained sum of squares is $E_{y/x}$, equation (2-22)).

Similarly, for the Theta criterion maximizing equation (2-4) maximizes the sum of modal frequencies ($M_1 + M_2$) which simultaneously maximizes the model statistic,

$$\Theta = \frac{\sum_\ell M_\ell}{N} \tag{2-31}$$

Because of these two facts, the lack of a counterexample, and the economy (equation 2-20) of maximizing the split Delta rather than the model Delta at each split, the program is coded to maximize Delta at each split.

2.8 COMPARISON WITH OTHER TECHNIQUES

The non-additive model and single code prediction/maximum distribution shape change criteria of THAID make it difficult to compare with other techniques. We will, however, mention the similarities of three techniques which appear closest in basic intent.

Multiple discriminant function analysis using dummy independent variables and "centour" ellipse prediction (Nunnally, 1967)is first. The data are not grouped by independent variable codes but a prediction scheme is used to generate categorical predictions. Each case is assigned a point in G-dimensional discriminant space and a set of distances to the G dependent variable centroids (d_1, d_2, ..., d_G) are computed. These distances are then converted into a set of percentile scores (P_1, P_2, ..., P_G) where a percentile score P means P percent of the sample have d's < d'. Minimum P_j determines the predicted category.

A second approach is a set of G least-squares regressions using dummy independent variables successively on the set of G dummy dependent variables associated with each dependent variable code (Theil, 1970). The MNA program (Andrews and Messenger, 1972) does this under the assumptions of additivity and homogeneity of variance. It further produces a categorical prediction $\hat{z}$ for each case, chosen as the code associated with the maximum of the prediction or "forecast" vector ($\hat{p}_1$, $\hat{p}_2$, ..., $\hat{p}_G$).

A final approach is the formation of a multi-dimensional contingency table with each predictor defining a dimension and predicting the mode of each marginal frequency distribution. A stepwise version of this approach was discussed above under BINARY SPLIT ALGORITHM. This method gives the closest comparison with THAID but suffers heavily from sampling instability as mentioned.* An empirical comparison of these three approaches has been conducted (O'Malley, 1972). The

*A variant of this technique using a within dependent variable code independence of predictors assumption is much less sensitive to sampling instabilities (Overall and Kleet, 1972).

test involved three independent variables and 1371 cases.

The prediction results summarized as a proportion of the sample correctly classed are as follows:

THAID $\frac{788}{1371} = .575$

MDF $\frac{788}{1371} = .575$

MNA $\frac{799}{1371} = .583$

CONTINGENCY TABLE $\frac{803}{1371} = .586$

Note that the contingency table value of .586 represents a maximum for any method on the sample and that for these data all techniques were very near optimal. The pattern of hits and misses for THAID was most like that of the contingency table (not significantly different) and least like MDF (significantly different) as measured by the chi-square statistic of differences in the classification matrices. The classification matrix is formed by rows giving predicted scores and columns giving actual scores for individuals.

CHAPTER III PROGRAM DESIGN

The intent of this chapter is to give the user a basic idea of the operation of THAID, of programming decisions made during its writing and of possible future improvements. Further details may be found in Appendices A, B and C.

3.1 DESIGN CHARACTERISTICS

THAID is written in FORTRAN G and uses standard OSIRIS subroutines (University of Michigan, 1970a) written in IBM Basic Assembly Language and Fortran G. The program has an overlayed total length of 14900_{16} bytes and needs the following configuration:

TWO TAPE/DISK/CARD FILES:	DICTIN DATAIN
FIVE DISK FILES*:	ISR01 FT03F001 FT04F001 FT05F001 FT07F001
CARD READER	
PRINTER	

THAID begins by reading the user global set-up cards**. It interprets them, sets the appropriate default values and checks for errors. If errors exist the appropriate message is printed and the program terminates. If no errors exist the interpreted global set-up cards are printed and the local

* DCB and SPACE requirements as annotated in sample output.

**THAID operates in an "analysis packet" mode. This requires a set of "global" control cards input initially to recode, filter and write the data onto a faster direct access file. Following this are sets of "local" control card packets to process a series of independent analyses using the same original data.

control cards (analysis packets) are processed. An error disqualifies the packet and causes a message to be printed. All usable packets are written onto work file FT07F001 in decoded form.

After all packets have been processed, the program checks for continuation if errors were found (Global keyword BADPKTS). If continuing, the input dictionary is then used and printed (University of Michigan, 1970b, Appendix D). The input data are then read and optionally filtered (subset), recoded (University of Michigan, 1970b, Appendix K) and written onto the work file ISR01.

Actual processing begins next by reading the first good analysis packet from FT07F001, local filtering, weighting and writing the processed data onto work file FT03F001 (if needed). A programming decision was made here to allow the use of two different processing modes. If the storage requirements are low enough all storage and processing is done in-core to increase speed. If storage requirements are high then the data and possibly also the pointer array indicating group assignment of cases will be written on the disk files FT03F001 and FT04F001, FT05F001, respectively.* The exact test for in-core/disk mode choice is:

$m \cdot n > 12{,}000$ (disk required for data)

$n \quad > 10{,}000$ (disk required for pointer)

where

$m \equiv$ number of predictors +1 (plus 1 if weights)

$n \equiv$ number of cases

Next comes the actual binary splitting beginning with generation of trial

*FT04F001, FT05F001 are used simultaneously for pointer arrays.

splits. If trial distributions are requested (local keyword PRINT) then the details for the best permissible split for each predictor and the bivariate table are printed. After the best split is found the pointer array is updated and summary statistics for the split stored, the algorithm continues on the new groups until no further permissible splits are found. To reiterate, the permissible split constraints are:

A. Predictor ordinality (optional)

B. Minimum group frequency

C. Minimum criterion improvement

D. A maximum of four splits (iterations) defining a group.

After exhausting all permissible splits the analysis packet is completed and the summary statistics on the THAID tree terminal groups are printed.

It is worth mention that two considerations will greatly affect run times. First the analysis packet feature is designed to provide economy in minimizing the relatively slow tape reading and character to binary conversion of data. Therefore, as many packets as possible should be run in a single job step. In addition, the global keyword BADPKTS should be used to guard against the processing of only a few of the originally intended analysis packets. A second consideration is that significant time differences exist between the processing of trial splits for nominal and ordinal predictors. This is because the number of trial splits for the r^{th} predictor treated ordinally is d_r-1 whereas the same predictor treated nominally has $2^{d_r-1}-1$ trial splits. For example, if $d_r=10$ then the number of ordinal splits is 9 whereas the number of nominal splits is 511. Thus users should make judicious use of nominal treatments. An ordinal set of k categories can be recoded into k dichotomous predictors and used more economically.

3.2 RESTRICTIONS

Due to the dual processing mode (in-core/disk) for analysis packets, the restrictions in the THAID program are minimal. These restrictions are:

A. Number of predictors $\leq$ 40 (each packet).

B. Total number of variables $\leq$ 200.

C. Number of cases: unlimited.

D. All variables in range 0-9 (except weight).

E. Weight variable in range 0-32,767.

F. Weighted frequencies $\leq 2^{31} - 1$.

3.3 FURTHER WORK

There are a number of areas of investigation for possible improvements in THAID. The first is an analytical and simulation study of the Delta statistic. The object would be to produce a confidence band under the null hypothesis using an analysis parallel to that done for Theta (See Figure II-2) supported by appropriate analytical work. In addition, an empirical search for a counter-example to the DELTA THEOREM of Chapter II is planned.

A second category of potential improvements is the evaluation of techniques found in the AID programs (Sonquist and Morgan, 1970) for possible applicability in THAID. The most important of these would be a two-stage (Draper and Smith, pp. 173-177) residuals option in which a set of predictors is used to generate an initial THAID tree and the "residuals" from this tree analyzed by a second set of predictors. Residuals are defined as follows. Consider the data of TABLE III-1.

TABLE III-1 RESIDUAL SCORES

DEP. VAR. VALUE		DUMMY DEPENDENT VARIABLES			TERMINAL GROUP PROPORTION			RESIDUALS		
	Codes	1	2	3	1	2	3	1	2	3
1		1	0	0	.80	.10	.10	.20	-.10	-.10

The dummy dependent variable set for the trichotomous (codes 1, 2 and 3) dependent variable are (1,0,0) and the corresponding terminal group proportions are(.80, .10,.10). The residuals are then defined as the set of differences between the dummy dependent variables and the terminal group proportions leading to (.20,-.10,-.10). A second technique would be the use of a "look-ahead" algorithm in which the improvement in either Theta or Delta is done on a two or three sequential split basis rather than the current single split. This would eliminate most of the loss of optimality associated with the stepwise maximization portion of the THAID algorithm. Another technique, an alternative to the "two-step" residuals approach, is a predictor ranking scheme. Here the predictors are assigned a rank and if no permissible splits are found at a certain rank a search of trial splits is then made over predictors at the next highest rank.

Finally, an investigation of faster ways for finding the best permissible split of a nominally treated predictor is planned. It is hoped that a theorem analogous to the one used to great advantage in the AID series of programs (Sonquist and Morgan, 1964, Appendix C) holds allowing a substantial reduction in the number of trial splits required. In brief, the AID theorem states that rank ordering nominal predictor codes by mean value of the dependent variable data associated with that code and trial splitting the rank ordered codes as if they belonged to an ordinal variable (a total of d_r-1 splits) is sufficient to find the optimal split. If no such theorem can be found, approximate techniques will be explored. One approximation scheme would be to perform a stepwise addition of nominal codes to a split group code set maximizing Theta/Delta at each step. In other words, one searches for the best one-subclass-versus-all-the-others to maximize Delta, then searches for the best subclass to add to the one so that those two versus the rest would maximize Delta, and so on.

CHAPTER IV USER STRATEGY

Any searching program requires a large data set, and careful attention to the possibility of random results--capitalizing on chance. We discuss in this chapter the strategy decisions, roughly in the order in which they must be made, with this in mind:

The dependent classification can have as many as 10 subclasses, but the number should be kept small wherever possible. The problems of small probabilities are not so serious as with MNA (or any other regression using one or a series of dichotomous dependent variables), but the random isolation of groups which differ on some minor pair of subclasses which mostly move together, remains a possible problem.

4.1 PREDICTOR SPECIFICATION

The selection of predictors must face the problem of intercorrelations, as in any other multivariate analysis. To put in the age of the family head and the age of his wife is to ask for trouble, since once one of them is used to divide the sample, the other will disappear from view. Where highly correlated predictors occur one can omit one, combine them into an index, or even develop a new set of uncorrelated predictors such as the average age of head and wife, and the difference in their ages.

One must always ask whether the explanatory characteristics are at the same level in the causal process or whether some of them could be the cause of others of them, but not the reverse. In this case, to include them all in a single multivariate analysis is inappropriate. One must make some choice of the kind of explanatory model involved, even if the particular variables and the patterns of their influence are still open. One might work in stages--using background factors like age, education, race and sex, --and then analyze the residual distributions against more current environmental or individual behavioral variables, though the present program does not allow this. Or, if it is meaningful to think of the current forces affecting some groups and not others, one can rank predictors to hold off the current factors until the effects of the background factors had been "removed" by dividing the sample on them. It is important to remember a general principle which carries over from variance analysis with a numerical dependent variable, that a relatively small number of subclasses on a predictor will exhaust most of its explanatory power. This is not intuitively obvious, and many people think that only the full numerical detail of predictors, or a full dummy variable regression with many subclasses of each predictor, really removes the variance accounted for by those predictors. Indeed, when the effects of a predictor are not linear, a few subclasses may account for more than a linear relationship using all the detail.

The number of classes in each predictor is also a point of strategy, and should again be kept small to minimize chance results. Where the predictor has no ordinal ranking, it is often true that it is also not one-dimensional, and should be divided into several predictors each of which is uni-dimensional. On the other hand, some dichotomous predictors like sex, and marital status, should be combined, as in the case of family heads who fall into only three classes: single men, single women, and married men (given our male chauvinist definition of "head of family").

Having selected and defined a "dependent" classification, and a set of explanatory characteristics and their details, one must then decide on:

1. Whether to maintain the order of each predictor.
2. How far to push the sequence of divisions.

In general we urge maintaining the order of all predictors except a few, such as Region, and restricting those to only 3 or 4 classes in order to avoid idiosyncratic results. If some other ordering of such a predictor would have been superior it will show up in the print-out, yet erratic orderings will not actually be used to split the sample.

4.2 MINIMUM FREQUENCY AND CRITERION CONTROLS

There are two rules for stopping the process, one based on the added explanatory power of the marginal split and the other on the number of cases left in the group to be split--a proxy for the reliability or potential significance of the next split.

We suggest using both of them, with say, a minimum group size of 50 (or even more), and a minimum increase in Theta of 0.5% (.005). In fact, of course, one can always truncate the results, but there is no point wasting computer time and paper and attention on random noise.

The double criterion is particularly important where weighted data are being used, since one might have increases in explanation based on widely different numbers of cases, and with widely different reliability.

4.3 THETA OR DELTA CRITERION

When one faces the decision of criterion choice, two basic considerations need be made. First is whether or not the researcher desires an individual prediction model rather than one giving groups whose overall distribution varies but whose ability to predict individuals may not be enhanced. The first thought requires the Theta criterion and the latter the Delta criterion. The second consideration is one meant to indicate the feasibility of the Theta criterion if it is desired. Since modal prediction requires a change in modal location from one split group to the other it must be possible to form binary splits having this property to be successful. As a rule of thumb, when the dependent variable sample distribution has a ratio of modal to second largest frequency exceeding 2:1 (termed "extremely unimodal") the Theta criterion is not workable and Delta is advised.

At the present time the authors are leaning towards the Delta criterion as it has shown to be more versatile and often leads to groups at the 2nd to 3rd iteration having different modal locations even when the sample distribution is extremely unimodal. The greatest advantage appears to be the more probabilistic orientation of Delta, making it sensitive to a statistical association confined to the <u>non-modal</u> codes and thereby not detected by Theta. In an attempt to exploit Delta to the fullest the program computes and prints an a posteriori set of Theta statistics on the THAID tree giving predictive model statistics even though the non-predictive Delta criterion was used.

4.4 STRATEGY IN LOOKING AT THE RESULTS

The first question to be answered by looking at the results is whether the decisions made at each split were clear, unambiguous and with no close competitors. Clearly if there were alternative splits which would have done nearly as well, another sample of data might have produced a different split decision. But there are two possible interpretations:

If the competitors were used in subsequent splits, then it is quite possible that the final groups are reliable, and other samples would have given similar results, getting to them by different routes. One could split off old people, then men, while in another sample one could segregate first all men, then old ones.

On the other hand, if the competitors disappeared from view and never showed any power in the subgroups after the split, then clearly there were correlated, or at least competive, predictors and with another sample we might well have selected the alternative. Where they are correlated there is some justification for eliminating one, or combining them into a combined or index predictor. Where they are uncorrelated but competitive, it may well be necessary to impose theoretical or substantive considerations.

The data in Figure IV-1 come from a panel study of family income dynamics conducted by the Survey Research Center for the Federal Office of Economic Opportunity. Results are given for only the THAID tree terminal groups and key subgroup proportion differences are indicated by inequality signs,(<,>).

FIGURE IV-1 Change in Family Composition Between 1968 and 1971

	18 - 34 (2)								35 or older (3)							
	Married (4)				Not Married (5)				35 - 54 (6)				55 or older (7)			
	18-24 (8)		25-34 (9)		Women* (10)		Men (11)		35-44 (12)		45-54 (13)		55-64 (14)		65 or older (15)	
	Unemployment rate of area in 1968 was:		Family money income/needs in 1967 was:		Age in 1968 was:		Money income/needs in 1967 was:		Actual-required rooms in 1968 was:		Actual-required rooms in 1968 was:		Actual-required rooms in 1968 was:		Marital Status in 1968 was:	
	less than 4%	4% or more	less than 2	2 or more	18-24	25-34	less than .80	.80 or more	1 or less	2 or more	1 or less	2 or more	1 or less	2 or more	married	single
0. No change	22%	37%	46%	50%	39%	46%	73%	40%	42%	56%	24%	45%	47%	69%	71%	85%
1. Single splitoff	0	2	3	1	3	1	0	5	7	6	14 >	6	9 >	4	2	2
2. Married splitoff	2	0	3	1	2	6	14 >	6	16 >	7	22 >	17	14 >	8	3	1
3. Female head got married	4	5	2	2	52 >	29	1	0	2	1	2	1	3	1	0	1
4. Wife became head	8	7	10 >	3	0	0	0	0	3	4	1	2	4	3	12	0
5. Head got married	2	5	1	1	0	0	11 <	41	2	1	2	1	0	1	0	0
6. Head single, was married	6	4	6	4	0	0	0	0	1	2	2	1	0	1	4	0
7. More people more adults	1	0	3	2	0	5	0	3	4	3	4	5	2	4	1	4
8. More people, no more adults	55 >	41	24 <	33	2	8	0	2	4	4	0	2	1	1	0	0
9. Fewer people in family	1	0	3	2	0	6	0	3	20 >	15	26 >	22	19 >	8	.6	.6
% of Sample	3.3	1.8	4.9	8.5	1.7	2.4	0.4	1.6	10.2	12.7	8.8	13.1	4.2	11.8	17.9%	6.1%
N	155	94	361	288	109	219	18	75	719	491	642	507	268	445	254	195

It turned out that one could not study changes in family income without first looking at changes in family composition. For the families in the panel as of 1971, including new families formed by splitting off from the original families, we can develop a measure of changes since the first interview in 1968. Since for most people the most common thing is no change at all,* modal prediction would not be useful and the Delta criterion was used. Details of the input and output appear with annotation in Appendix C.

It is ofren useful to start with the array of final groups, as in the Figure IV-1 presented here, rather than complicate things with all the intermediate groups. However, when a particular effect is of interest, it may be useful to see whether it actually appears farther back for larger subgroups, or whether it never showed up except in some subgroups. If the latter is true, as it was with the first two pairs of columns in Figure IV-1, then these effects are the kind most likely to have been missed by methods which insist on additivity--that an effect should work for everyone--as in multiple regression, or MNA. (The explanatory factors were measured as of the beginning of the period, including the age of the then head, who is no longer head in groups 1-4).

Unless there were important effects working on several branches, where the pooled effect for larger subgroups is of interest, it is often sufficient to present the final subgroup distributions only, as in Figure IV-1. The reader could always re-pool to get combined group distributions, particularly with unweighted data.

It is also clear from the Figure that substantial interaction effects are present. Environmental forces like employment and earning opportunities

---

*In the sample 51.3% showed no change with code 9 (fewer people in family) being second most frequent with 12.4%.

affect younger families, and individual housing condition (overcrowding) affects the undoubling of older families. And young women get married when they are young, but young men when they can afford it!

4.5 STRATEGY IN THE LARGE

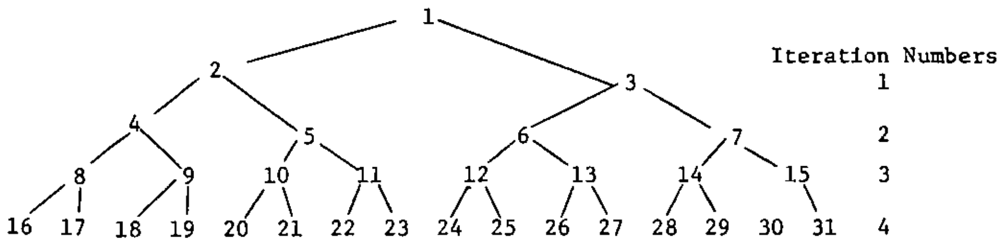
It is always useful to work out a path analysis diagram in which the variables are related to one another and advantage taken of any known one-directional relations. One can then decide what analyses need to be done, and which explanatory characteristics are appropriate. Programs like this one were developed to avoid unsystematic ransacking and trial and error procedures, and to encourage a pre-stated strategy which would at least be reproducible by another researcher, or applicable to another set of data. Hence, if many runs are made, it should not be because one cannot decide what the relevant predictors are, but because there are several parts of a complex theoretical model which call for separate assessment and specification. The process is one of looking for the best specified model, but not from among all possible models, rather from a limited subset. In fact, one is really selecting a subset of k explanatory characteristics from a limited, defined and justified list, and selecting a pattern of interactions among them. It is analogous to selecting the best functional form, in a rather flexible sense.

Where some predictors are prior in a causal sense to others, one might be tempted to use a two-stage approach taking residuals from the effects of the first set and pooling them to assess the effects of the second set. This is particularly urgent, since categorical dependent variables require more data (because distributions require more than means). What is the meaning of a residual or deviation of an individual's actual position from an "expectation" which is a distribution? Suppose there are three categories, and the end-group distribution is .30, .40, .30. Then an individual in that group who is in the first class would have a set of residuals .70, -.40, -.30, and if one pooled all the residuals from the individuals in that group they would add to 0, 0, 0.

It should then be possible to pool the residuals across the end groups of a first-stage analysis (they would sum to zero over the sample for each subclass), and do an analysis of them, provided the algorithm is changed to allow for negative numbers and to handle up to ten such residuals per case. Where there may be specific interactions with one of the first-stage predictors, they can be handled by reintroducing that predictor in the second stage analysis. What is also possible is ranking of predictors, so that some are only allowed in after a prior set has had its chance. (See Chapter III, Future Work). Remember that this, too, is not yet programmed.

The choice of strategy between two-stage analysis and ranked predictors, where there are stages in the causal process, depends on theoretical and practical considerations. Ranking requires more data since one looks for further splits on subgroups some of which may already be small. But if one is searching for effects of second-stage variables which may exist only for sub-groups of the population, then ranking is more likely to find them.

Rather than number the groups as they are developed, as in AID III, this program numbers them systematically as though a symmetric tree were to be developed in every case, and each row were to be produced by splits starting with the left-most group. The resulting group numbers are thus:



And the rows are called "iterations" and numbered starting with the second row. Figure IV-1 thus gives the data for groups 16 through 31, in that order.

For the first iteration, an option, local keyword PRINT, allows printing the bivariate table on each predictor, that is, the distribution of the criterion classification for each subclass of each predictor.

For each group, the best binary division on each predictor is shown together with the two subgroup distributions of the criterion classification, absolutely and as difference from the sample distribution proportions. The codes which are used for each subgroup definition are given with the Delta measure of the power of the split.

Then the best predictor is specified, and its Delta repeated.

In the case of a weighted run a list of the unweighted frequencies in each of the "child" groups is then given. Finally a summary of the splits appears, giving the variable and codes used, the weighted frequency and modal frequency and the ratio (Theta) of mode to total for each child, and the Delta for each split. (See Appendix C: SAMPLE OUTPUT)

The final summary is useful in forming the branching diagram, but one wants, at least for the final groups, the actual distributions, which are available in the earlier part, for children numbers 16 through 31. And while copying off those distributions, one can look at the Delta statistics for that predictor and all its competitors to see how close the competition was. If other predictors would have done nearly as well, then the probability that another sample of data would give a different branching diagram is high. Indeed, this scanning of the extent to which the chosen split was unambiguous or the result of close comparisons, should be carried back through all the splits. Table IV-1, from the example given in Figure IV-1, shows one useful format for assembling these Deltas for examination.

TABLE IV-1 Delta Measure of Explanatory Power of Alternative Predictors of Change in Family Composition - 1968 to 1971

	Predictor (Initial Condition)					
Initial Condition	Age	Sex-Marital Status	Race	Income/ Needs	Unemployment in County	Actual Required Rooms
1. All	.168	.094	.028	.029	.017	.104
2. Under 35	.122	.217	.039	.087	.037	.114
4. Married	.152	--	.040	.075	.044	.067
8. 18-24	--	--	.041	.109[b]	.138	.081
9. 25-34	--	--	.044	.106	.046	.086[b]
5. Not married	.125	.268	.106	.074	.056	.099
10. Women	.184	--	.184[c]	.091	.060	.120
11. Men	.104	--	.087	.189	.092	.163[d]
3. 35 or older	.140	.065	.030	.033	.012	.105
6. 35-54	.128	.058	.038	.063	.036	.113[a]
12. 35-44	--	.079	.041	.111[e]	.039	.114
13. 45-54	--	.082	.043	.083	.054	.138
7. 55 or older	.190	.156	.086	.078	.059	.160[f]
14. 55-64	--	.102	.053	.111	.040	.174
15. 65 or older	.059	.224	.192	.146	.119	.178[b]

a) A competitor which lost out here, but made the next splits on both children.

b) Competitors which could have produced more symmetry, but at a loss of predictive power of about a fifth.

c) Extremely close - a slightly different classification of change in family composition led to a split on race here, and a different sample easily could, too.

d) A competitor that would have made this split more like the right half of the diagram.

e) This alternative would have made this split more like the left half of the diagram.

f) Like a), lost out but retained its power with the children.

For the final splits, it is possible to use an inferior split, as we almost did in one case in Figure IV-1, if it produces a more symmetric figure, or is theoretically more interesting, provided one makes clear what was done.

The most interesting result in Table IV-1 is that the most serious competitive predictors were not competitive because of intercorrelations, hence did not lose their power when the alternative split was made, but were able to dominate subsequent splits in most cases. This means that with a fresh sample, the end-groups would very likely have been the same, but the order of the splits that produced them might have been different.

The one real tied decision was for the young single women where one could either report that the youngest ones were more likely to get married, or that the Whites were more likely to get married, and the Blacks more likely to have children.

There will be occasions with other data where intercorrelations are serious and where predictors which look important disappear once the sample is divided on some related predictor. And there may be occasions where there are a number of predictors of roughly equal power and also correlated, so that the resulting divisions of the sample are clearly unstable and unlikely to be repeated exactly if a fresh sample of similar data were used. In such cases one might want to compress the predictors into a combination index or scale or factor, or to eliminate some of them.

It may be useful, after presenting the final subgroup distributions, to show subgroup distributions farther back along the "tree", either to show how some effects were hidden in more aggregated groups, or that they appeared early and were able to effect a split only after a more powerful factor has made a prior split. Table IV-2 shows an example of each type, from the earlier distributions not shown in Figure IV-1.

TABLE IV-2

Examples of Hidden and of Persistent Effects

	18-34 Married (Group 4)				35-54 (Group 6)	
	Unemployment rate in area in 1968		Family Income/Needs		Actual Required Rooms	
Change in family Structure	less than 4%	4% or more	less than 2	2 or more	1 or less	2 or more
0. No change	42	44	40	45	34	50
1. Single splitoff	1	2	2	1	10	6
2. Married splitoff	2	0	2	1	19	12
3. Female head got married	2	4	3	2	2	1
4. Wife became head	6	7	8	5	3	3
5. Head got married	1	2	2	1	2	1
6. Head single, was married	5	5	6	4	2	1
7. More people, more adults	2	1	2	1	4	4
8. More people, no more adults	36	34	32	38	2	2
9. Fewer people in family	2	1	2	1	23	19
	99	100	99	99	101	99

When the married couples 18-34 are examined, neither local unemployment nor income adequacy appear to affect change in family structure, but once we divide into two finer age groups, (group 4 into groups 8 and 9), the 18-24 year olds do appear to be affected by local unemployment, and the 25-34 year olds by their initial income adequacy, in decisions to have children! This makes sense, and is a classic example of the power of this algorithm to uncover relationships hidden by interactions (that only operate within a subgroup).

On the other hand, while the initial housing adequacy affected the amount of undoubling (splitoffs, and empty-nest couples with fewer people in the family) for all those 35 or older, three divisions by age came first because the total amount of undoubling varied more by age, and was greatest where the head of the original family was 45-54. The table shows that for two of the age groups combined (ages 35-54) the main pattern appeared that was in the two separate age groups (35-44 and 45-54) in Figure IV-1.

Finally, if one is concerned about "capitalizing on chance," the proper question is not whether a fresh sample would give the same "tree," but whether the subgroups developed, when forced on a fresh sample, would give evidence that such a set could increase our ability substantially to predict population distributions. With complex samples, it is necessary to take account of sample design in forming independent subsamples, and with such a subsample, it is difficult to estimate sampling errors, but relatively easy to estimate population parameters, and explanatory power.

The authors would be happy to hear from users of the program about their experiences and whether they discovered anything about the real world that more orthodox procedures might not have revealed.

Bibliography

1. Andrews, Frank M. and Messenger, Robert C., Multivariate Nominal Analysis, Ann Arbor: University of Michigan, Institute for Social Research, 1973

2. Attneave, F., Applications of Information Theory to Psychology, New York: Holt, Rinehart, 1959

3. Blalock, Hubert M. Jr., An Introduction to Social Research, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970

4. Draper, N. R. and Smith, H., Applied Regression Analysis, New York: Wiley, 1966

5. Goodman, L. A. and Kruskal, W. H., "Measures of Association for Cross Classification," Journal of the American Statistical Association, 49 (December 1954), 732-64

6. Labovitz, Sanford, "The Assignment of Numbers to Rank Order Categories," American Sociological Review, 35 (June 1970), 515-24

7. Light, R. J. and Margolin, B. H., "An Analysis of Variance for Categorical Data," Journal of the American Statistical Association, 66 (September 1971), 534-544

8. Messenger, Robert C. and Mandell, Lewis M., "A Modal Search Technique for Predictive Nominal Scale Multivariate Analysis," Journal of the American Statistical Association, 67 (December 1972)

9. Nunnally, Jum C., Psychometric Theory, New York: McGraw-Hill, Inc., 1967

10. O'Malley, P. An Empirical Comparison of MNA and MDF. Unpublished paper, Department of Psychology, University of Michigan, 1972

11. Overall, J. E. and Klett, C. J., Applied Multivariate Analysis, New York: McGraw-Hill, 1972

12. Sonquist, John A. and Morgan, James N., The Detection of Interaction Effects, Ann Arbor: University of Michigan, Institute for Social Research, 1964

13. Sonquist, John A. and Morgan, James N., The Detection of Interaction Effects, Ann Arbor: University of Michigan, Institute for Social Research, 1970 (Superseded by #14 below), Rev. Ed.

14. Sonquist, John A., Baker, Elizabeth L. and Morgan, James N., Searching for Structure, Ann Arbor: University of Michigan, Institute for Social Research, 1971

15. Theil, Henri, "On the Estimation of Relationships Involving Qualitative Variables," American Journal of Sociology, 76 (July 1970), 103,154

16. University of Michigan, Institute for Social Research, OSIRIS/40 User's Manual, Ann Arbor: University of Michigan, Institute for Social Research, 1

17. University of Michigan, Institute for Social Research, OSIRIS/40 Subroutine Manual, Ann Arbor: University of Michigan, Institute for Social Research, 1970b

APPENDIX A: SET-UP INSTRUCTIONS

These details are taken from the OSIRIS 40 User's Manual (University of Michigan, 1970a) and are intended to give the reader a more complete understanding of some of the features of THAID. THAID is part of ISR's OSIRIS package of computer programs. The control cards therefore conform to OSIRIS format.

INPUT

1. Dictionary and data, on cards, tape, or disk.
2. Control cards.

OUTPUT

1. Printout: listing of dictionary for accessed variables.
2. Printout: number of cases passing global filter.
3. Printout: number of cases passing local filter
4. Printout: univariate distribution of dependent variables.
5. Printout: parameter card interpretation
6. Printout: (optional) trial distributions for each predictor at each split.
   - A. Best split on each predictor with:
     - (1) split codes
     - (2) frequencies
     - (3) proportions
     - (4) proportion deviations from sample
     - (5) Theta/Delta
   - B. Bivariate-frequency table
7. Printout: summary of split groups; data for each group:
   - A. Independent-variable codes
   - B. Frequency
   - C. Mode
   - D. Modal frequency

E. Theta for the group

8. Printout: Theta/Delta for each split

9. Printout: Theta for each iteration (weighted average for all split groups at end of iteration).

10. Printout: normal termination message.

(NOTE: 3-9 printed for each analysis packet)

RESTRICTIONS

1. Maximum number of cases: no limit.
2. Maximum number of predictor variables per analysis packet: 40.
3. Maximum predictor or dependent-variable code: 9.
4. Maximum number of iterations: 4.
5. Total number of dependent, predictor, and weight variables for all analysis packets: 200.

MISSING-DATA TREATMENT

Missing data is not checked by the program. Users may exclude missing data in the global filter or by use of maximum-code checking (recoding if necessary). All cases having at least one variable exceeding the maximum code are eliminated.

EVALUATION SUMMARY

1. Users should attempt to make both independent and dependent variables as uniformly distributed as possible. This can be done by using a recoding scheme to "build up" lower frequencies or by filtering the modal codes.

2. A minimum group frequency should be chosen, after considering probable sampling errors in the data and the number of probable dependent-variable codes, to produce groups in which the mode of the group has a high probability of being the population sample group mode.

3. Theta vs. Delta: Users must choose between a categorical prediction analysis (implying the choice of <u>Theta</u>) and a distributional shape-change analysis (implying the <u>Delta</u> criterion). When the univariate distribution for the top two codes has a frequency ratio exceeding 2:1, Theta is generally not useful. Otherwise, either may be used.

SETUP SUMMARY

```
JOB card
//  EXEC   ISRSYS
//DICTx  DD  parameters for input dictionary
             (omit if on cards)
//DATAx  DD  parameters for input data file
             (omit if on cards)
//SETUP  DD  *
$RUN  THAID
$RECODE card                     } optional
  Recode-statements cards        }
$SETUP card
  1.  Global-filter card (optional).
  2.  Global label card.
  3.  Global parameter card.
  4.  Local filter card (optional).  }
  5.  Local label card.              } analysis
  6.  Local parameter card .         } packet
  7.  Predictor cards.
$DICT card                          }
  Dictionary-descriptor card        } if dictionary on cards
  T-cards                           }
```

$DATA card

  Data cards   } if data on cards

/* card

DESCRIPTION OF CONTROL CARDS

1. Global-filter card (optional). Filter cards select a subset of cases by specifying certain values of certain variables. A "global" filter applies to the entire program run. A "local" filter applies only to one computation (e.g., one analysis packet in a THAID run).

   Examples: (a) INCLUDE V2=1-5 AND V7=23,27,35, AND V8=1,3,4,6*

   (b) EXCLUDE V10=2-3,6,8-9 AND V30=001-004 OR V91=025*

   1. Punch INCLUDE or EXCLUDE, beginning anywhere on the first card; continue punching on the first card; if necessary extend by ending the first card with a comma or conjunction (AND or OR) and continuing on the next card.
   2. Maximum number of expressions per run: 15 (an expression includes V, the variable number, an equals sign, and a list of values).
   3. An asterisk must terminate the list of expressions.
   4. Alphabetic variables may be specified for data-management programs; values are enclosed by primes (e.g., 'DETROIT').
   5. Variables may appear in any order and in more than one expression.
   6. Expressions are connected by AND or OR:
      a. AND indicates that the values in all connected expressions must be found in order to select the case.
      b. OR indicates the case will be selected if any or all

of the specified values are found.

c. AND expressions are executed before OR expressions.

d. Example: expression 1 AND expression 2 OR expression 3,A case would be selected if any expression 3 values were found <u>or</u> if all expression 1 and 2 values were found, but not if only expression 1 values were found or if only expression 2 values were found.

7. Values specified must have the field width of the pertinent variable; lead zeros must be punched.

8. Values may be specified singly, separated by commas, or in a range (e.g., 001-004 using a dash).

9. Values may be a positive or negative, but a value <u>range</u> may not vary from a negative value to a positive value (separate into two ranges).

10. Negative ranges should be expressed as in example (a) for global filter, as in (a) or (b) for local filter:

(a) V1=-10--1,0-9*

(b) V1=01--10,0-9*

11. Error messages are self-explanatory; e.g., EQUAL SIGN NOT PRECEDED BY A V or INCONSISTENT FIELD WIDTH.

2. <u>Global label card</u>: 1-80 columns to title the output, punched free-form.

3. <u>Global parameter card</u>: Keywords (defaults underlined) are separated by commas or blanks; list is terminated by an asterisk. If all defaults are chosen, <u>punch a single asterisk on parameter</u> card.

INFILE=<u>IN</u>/xxxx	Allows the user to specify a 1 to 4 character acter input ddname suffix.

PRINT=DICT/NODICT — Print the input dictionary?

BADDATA=SKIP/MD1/MD2/TERMINATE/STOP

When non-numeric characters (including embedded blanks and all-blank fields) are found in numeric variables:

SKIP: Skip the case.

MD1 : Convert value to MD1 code.

MD2 : Convert value to MD2 code.

TERM: Terminate the run.

STOP: Terminate the run.

For SKIP, MD1, MD2: a message indicates the number of cases so treated.

BADPKTS=STOP/GO — Allows the user to specify whether to continue when there is an error in one analysis packet if other analysis packets are still good.

4. Local filter card (optional): selects a subset of cases for one analysis.
5. Local label card: 1-80 columns to title the individual analysis.
6. Local parameter card: Keywords (defaults underlined) are separated by commas or blanks; list is terminated by an asterisk. If all defaults are chosen, punch a single asterisk on parameter card.

MINF=n — Specifies minimum group frequency (default = 10).

TMIN=n — Specifies minimum Theta/Delta increase for split (default = 0).

NOWEIGHT/WEIGHT=(n,m)	Specifies weight variable number (n) and maximum value (m). (Default for m if n is specified = 10,000)
DEPVAR=(n,m)	Specifies dependent variable number and maximum value. (Default for m = 9)
PRINT=n	Print trial distributions? (Default: PRINT = 0) n = 0: None n > 0: Trial distributions for all iterations; bivariate tables for first n iterations.
CRIT=DELTA/THETA	Specifies the split criterion

7. Local predictor cards: the general format for the local predictor cards (one for each predictor) is:

PRED=(LIST), M/F, MAXC=9/I* or

PRED=(LIST), M/F, MAXC=9/I, END*

where PRED indicates a predictor card

LIST ≡ String of variable numbers (e.g., V1-V3,V5)

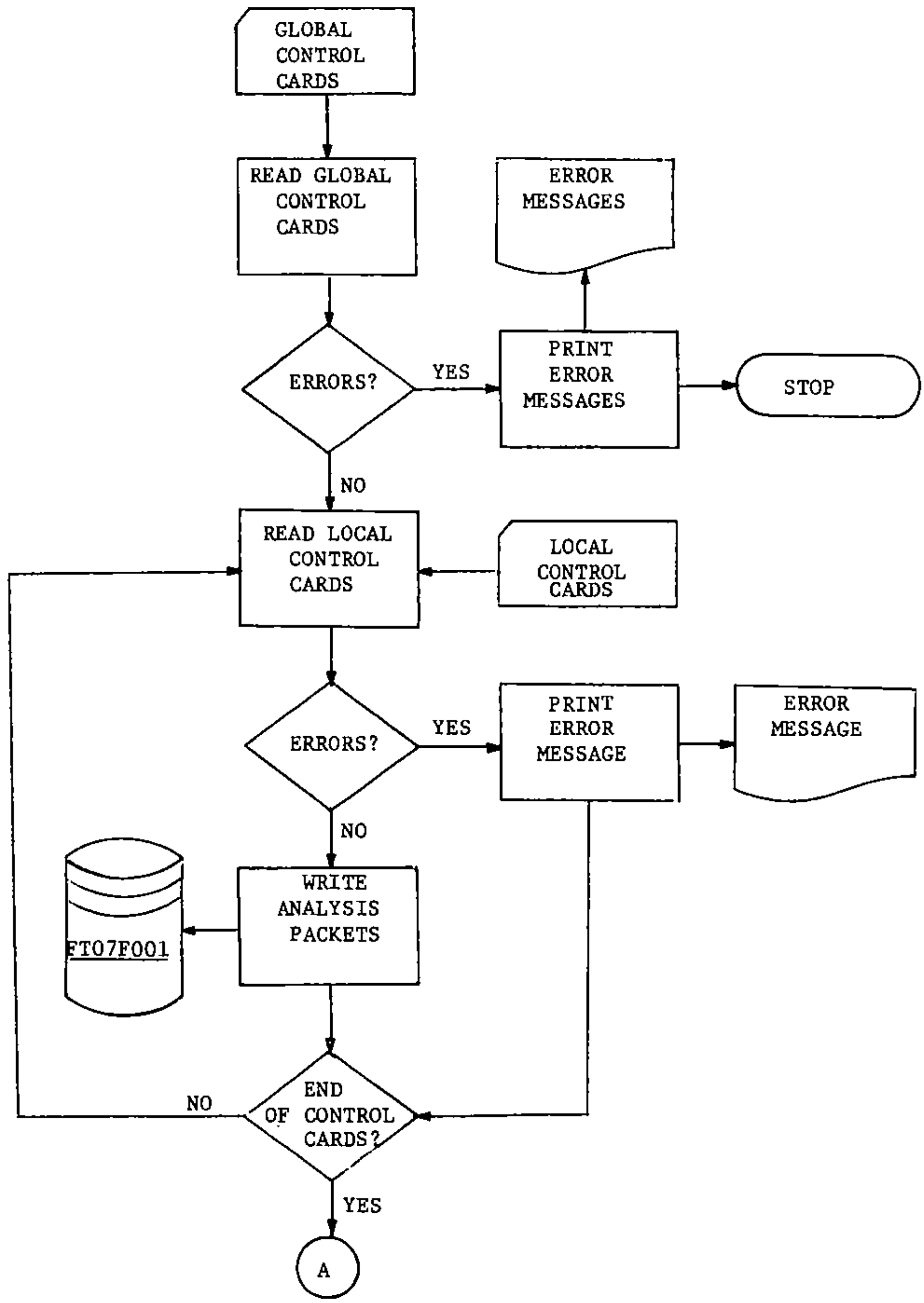
M/F ≡ Monotonic ("ordinally treated") or
Free ("nominally treated")

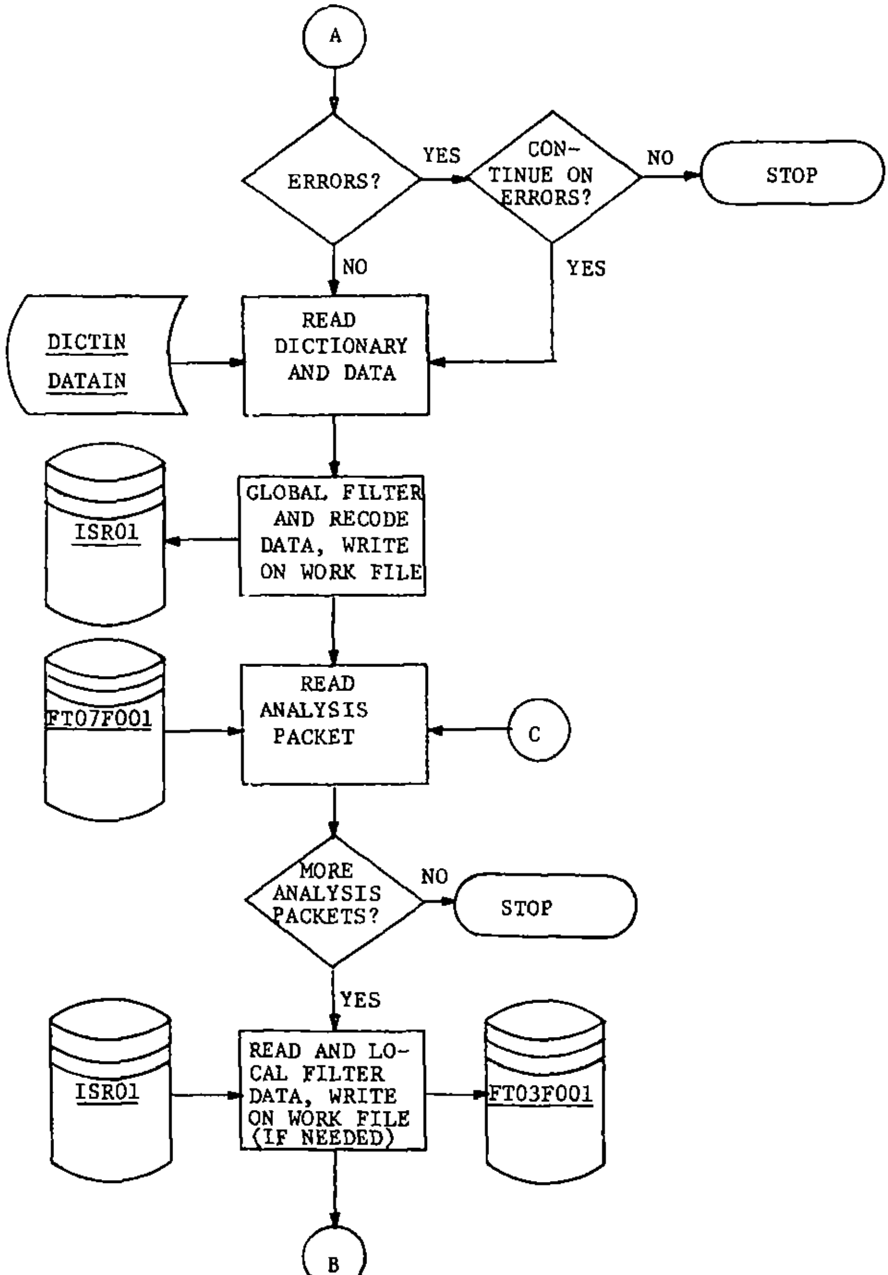
MAXC ≡ Maximum code

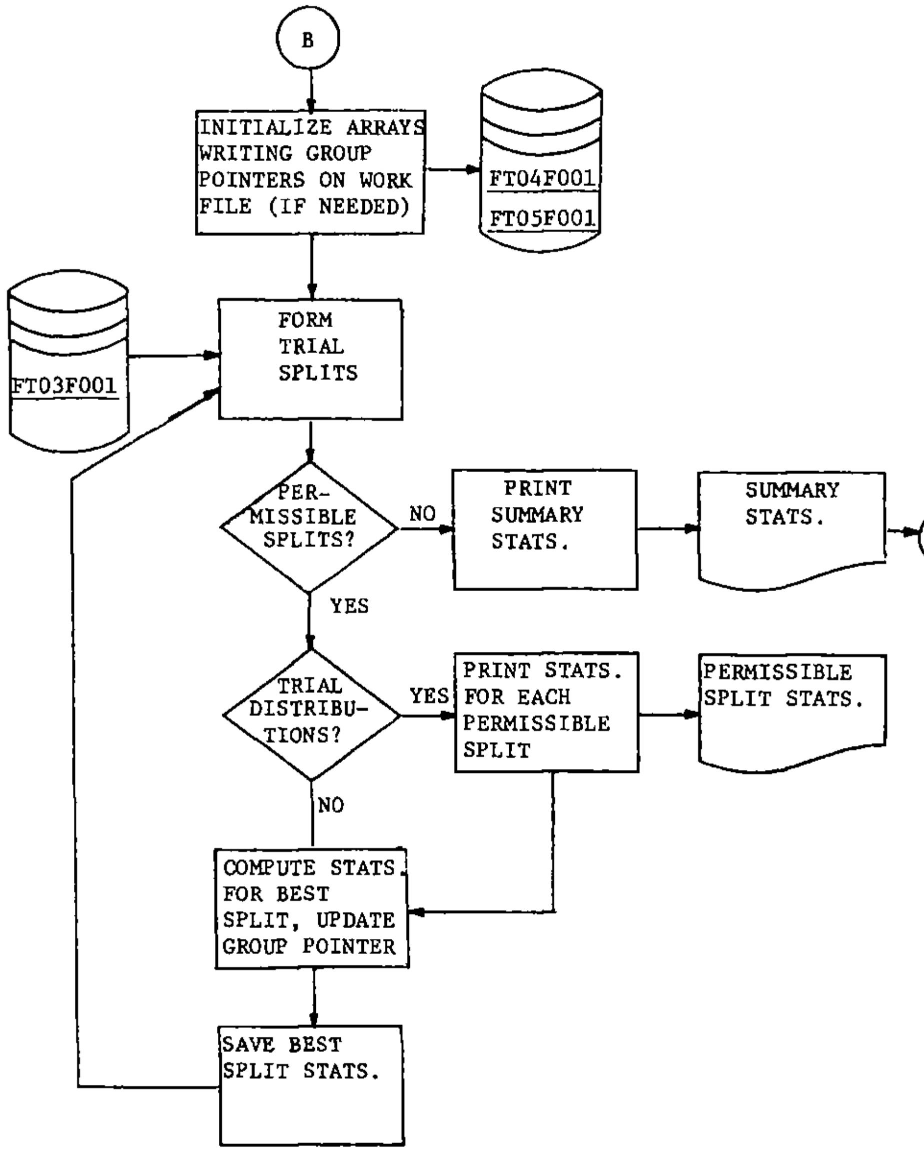
END indicates the last predictor card for a packet

* indicates end of a single predictor card.

APPENDIX B: MACRO FLOW CHART




A
ERRORS?
YES
CON-
TINUE ON
ERRORS?
NO
STOP
NO
YES
DICTIN
DATAIN
READ
DICTIONARY
AND DATA
ISR01
GLOBAL FILTER
AND RECODE
DATA, WRITE
ON WORK FILE
FT07F001
READ
ANALYSIS
PACKET
C
MORE
ANALYSIS
PACKETS?
NO
STOP
YES
ISR01
READ AND LO-
CAL FILTER
DATA, WRITE
ON WORK FILE
(IF NEEDED)
FT03F001
B


B
INITIALIZE ARRAYS WRITING GROUP POINTERS ON WORK FILE (IF NEEDED)
FT04F001
FT05F001
FT03F001
FORM TRIAL SPLITS
PER-MISSIBLE SPLITS?
NO
PRINT SUMMARY STATS.
SUMMARY STATS.
YES
TRIAL DISTRIBU-TIONS?
YES
PRINT STATS. FOR EACH PERMISSIBLE SPLIT
PERMISSIBLE SPLIT STATS.
NO
COMPUTE STATS. FOR BEST SPLIT, UPDATE GROUP POINTER
SAVE BEST SPLIT STATS.

APPENDIX C: SAMPLE OUTPUT

```
//M009307  JOB  (,                                                         JOB 522
// 468473,THAI,MONO,15,,,,5),MESSENGER,MSGLEVEL=(1,0)
// EXEC ISRSYS
XXOSIRIS PROC LIB=OSIRPGM,LIB1=SRCLIB,LIB2=CPSLIB,LIB3=ISRLIB,      00000020
XX             OI='*.FT48F001,VOL=REF=*.FT48F001',                    00000040
XX             OA='*.FT47F001,VOL=REF=*.FT47F001',                    00000060
XX             P=ISRSYS,SP1=1000,SP2=600                              00000080
XXGO   EXEC  PGM=&P                                                   00000100
IEF653I SUBSTITUTION JCL - PGM=ISRSYS
XXSTEPLIB  DD  DSN=&LIB,DISP=SHR                                      00000120
IEF653I SUBSTITUTION JCL - DSN=OSIRPGM,DISP=SHR
XX         DD  DSN=&LIB1,DISP=SHR                                     00000140
IEF653I SUBSTITUTION JCL - DSN=SRCLIB,DISP=SHR
XX         DD  DSN=&LIB2,DISP=SHR                                     00000160
IEF653I SUBSTITUTION JCL - DSN=CPSLIB,DISP=SHR
XX         DD  DSN=&LIB3,DISP=SHR                                     00000180
IEF653I SUBSTITUTION JCL - DSN=ISRLIB,DISP=SHR
XXSYSPUNCH  DD  SYSOUT=B                                              00000200
XXSYSPRINT  DD  SYSOUT=A                                              00000220
XXSYSOUT  DD  SYSOUT=A                                                00000240
XXFT02F001  DD  SYSOUT=B                                              00000260
XXFT03F001  DD  UNIT=SYSDA,SPACE=(TRK,(100,50)),                      00000280
XX   DCB=(RECFM=VBS,LRECL=200,BLKSIZE=204,BUFNO=1)                    00000300
XXFT04F001  DD  UNIT=SYSDA,SPACE=(TRK,(200,50)),                      00000320
XX   DCB=(RECFM=VBS,LRECL=200,BLKSIZE=204,BUFNO=1)                    00000340
XXFT05F001 DD UNIT=SYSDA,SPACE=(TRK,(50,50)),                         00000360
XX   DCB=(RECFM=VBS,LRECL=200,BLKSIZE=204,BUFNO=1)                    00000380
XXFT06F001  DD  SYSOUT=A                                              00000400
XXFT07F001  DD  UNIT=SYSDA,SPACE=(TRK,(50,50)),                       00000420
XX   DCB=(RECFM=VBS,LRECL=200,BLKSIZE=204,BUFNO=1)                    00000440
XXFT08F001  DD  UNIT=SYSDA,SPACE=(TRK,(50,10)),                       00000460
XX  DCB=(RECFM=FB,LRECL=80,BLKSIZE=80)                                00000480
XXFT46F001 DD UNIT=SYSDA,SPACE=(TRK,(5,5)),                           00000500
XX  DCB=(RECFM=FB,LRECL=80,BLKSIZE=800)                               00000520
XXFT47F001 DD UNIT=SYSDA,SPACE=(TRK,(100,20)),                        00000540
XX  DCB=(RECFM=FB,LRECL=80,BLKSIZE=3520)                              00000560
XXFT48F001 DD UNIT=SYSDA,SPACE=(TRK,(100,20)),                        00000580
XX  DCB=(RECFM=FB,LRECL=80,BLKSIZE=3520)                              00000600
XXFT49F001  DD  UNIT=SYSDA,SPACE=(TRK,(200,50)),                      00000620
XX  DCB=(RECFM=FB,LRECL=80,BLKSIZE=3520)                              00000640
XXFT50F001  DD  DSN=ISRNEWS,DISP=SHR,LABEL=(,,,IN)                    00000660
XXFT09F001 DD  VOL=REF=*.FT47F001,DSN=*.FT47F001,DISP=(OLD,DELETE)   00000680
XXISR01  DD  UNIT=SYSDA,SPACE=(TRK,(&SP1),,CONTIG)                    00000700
IEF653I SUBSTITUTION JCL - UNIT=SYSDA,SPACE=(TRK,(1000),,CONTIG)
XXISR02  DD  UNIT=SYSDA,SPACE=(TRK,(&SP2),,CONTIG)                    00000720
IEF653I SUBSTITUTION JCL - UNIT=SYSDA,SPACE=(TRK,(600),,CONTIG)
XXISR03  DD  UNIT=SYSDA,SPACE=(TRK,(50),,CONTIG)                      00000740
XXISR09  DD  UNIT=SYSDA,SPACE=(TRK,(200),,CONTIG)                     00000760
XXISR10  DD  UNIT=SYSDA,SPACE=(TRK,(200),,CONTIG)                     00000780
XXUCLOAD  DD  DSN=*.FT07F001,VOL=REF=*.FT07F001,DISP=(OLD,DELETE)    00000800
XXSORTWK01  DD  DSN=*.ISR01,DISP=(OLD,DELETE),VOL=REF=*.ISR01         00000820
XXSORTWK02  DD  DSN=*.ISR02,DISP=(OLD,DELETE),VOL=REF=*.ISR02         00000840
XXSORTWK03  DD  DSN=*.ISR09,DISP=(OLD,DELETE),VOL=REF=*.ISR09         00000860
XXSORTWK04  DD  DSN=*.ISR10,DISP=(OLD,DELETE),VOL=REF=*.ISR10         00000880
XXSORTWK05  DD  DSN=*.ISR03,DISP=(OLD,DELETE),VOL=REF=*.ISR03         00000900
XXSORTWK06  DD  DSN=*.FT04F001,DISP=(OLD,DFLETE),VOL=REF=*.FT04F001   00000920
XXSORTLIB  DD  DSN=SYS1.SORTLIB,DISP=SHR                              00000940
XXSORTIN   DD  VOL=REF=*.FT47F001,DSN=*.FT47F001,DISP=(OLD,PASS)      00000960
XXSORTOUT  DD  VOL=REF=*.FT47F001,DSN=*.FT47F001,DISP=(OLD,PASS),     00000980
XX   DCB=(RECFM=FB,LRECL=80,BLKSIZE=800)                              00001000
```

THIS IS A GENERAL PURPOSE PROCEDURE FOR OSIRIS PROGRAMS. (// EXEC ISRSYS)

DISK FILE FOR LOCAL DATA. REQUIRES 2 BYTES PER VARIABLE PER CASE. (FT03F001)

DISK FILE FOR GROUP POINTER. REQUIRES 2 BYTES PER CASE. (FT04F001)

SAME AS FT04F001 (FT05F001)

CONTAINS ANALYSIS PACKET DECODED VALUES. REQUIRES LESS THAN ONE TRACK PER 10 PACKETS. (FT07F001)

DISK FILE FOR GLOBAL DATA. REQUIRES 4 BYTES PER VARIABLE PER CASE. (ISR01)

```
X/DICTIN  DD  DSN=&DI,DISP=(OLD,PASS),                                  00001020
IEF653I SUBSTITUTION JCL - DSN=*.FT48F001,VOL=REF=*.FT48F001,DISP=(OLD,PASS),
XX  DCB=BUFNO=1                                                         00001040
//DATAIN DD UNIT=TAPE,VOL=SER=(1750,1751),LABEL=2,DSN=DA8901CH,DISP=OLD-----------INPUT DATA
X/DATAIN  DD  DSN=&DA,DISP=(OLD,PASS),                                  00001060
IEF653I SUBSTITUTION JCL - DSN=*.FT47F001,VOL=REF=*.FT47F001,DISP=(OLD,PASS),
XX  DCB=BUFNO=1                                                         00001080
XXDICTOUT  DD  VOL=REF=*.FT48F001,DSN=*.FT48F001,DISP=(OLD,PASS),       00001100
XX  DCB=BUFNO=1                                                         00001120
XXDATAOUT   DD  VOL=REF=*.FT47F001,DSN=*.FT47F001,DISP=(OLD,PASS),      00001140
XX  DCB=BUFNO=1                                                         00001160
XXFT01F001 DD UNIT=SYSDA,SPACE=(TRK,(50,10)),DCB=(RECFM=F,BLKSIZE=80)   00001180
XXSYSIN DD DSN=*.FT01F001,DISP=(OLD,DELETE),VOL=REF=*.FT01F001          00001200
//SYSUDUMP DD SYSOUT=A
//SETUP DD *
//
IEC209I M009307  1750   281  TR=001,TW=000,EG=000,CL=000,N=000,SIO=00122
IEC209I M009307  1750   281  TR=000,TW=000,EG=000,CL=000,N=000,SIO=04571
IEC209I M009307  1751   282  TR=000,TW=000,EG=000,CL=000,N=000,SIO=00297
IEF283I   SYS72356.T024354.RF000.M009307.R0000122      NOT DELETED 8
IEF283I   VOL SER NOS= MFT1    1.
IEF283I   SYS72356.T024354.RF000.M009307.R0000119      NOT DELETED 8
IEF283I   VOL SER NOS= MFT1    1.
IEF283I   SYS72356.T024354.RF000.M009307.R0000125      NOT DELETED 8
IEF283I   VOL SER NOS= ISRB    1.
IEF283I   SYS72356.T024354.RF000.M009307.R0000126      NOT DELETED 8
IEF283I   VOL SER NOS= ISRA    1.
IEF283I   SYS72356.T024354.RF0CC.M009307.R0000128      NOT DELETED 8
IEF283I   VOL SER NOS= ISRB    1.
IEF283I   SYS72356.T024354.RF000.M009307.R0000129      NOT DELETED 8
IEF283I   VOL SER NOS= ISRA    1.
IEF283I   SYS72356.T024354.RF000.M009307.R0000127      NOT DELETED 8
IEF283I   VOL SER NOS= MFT1    1.
IEF283I   SYS72356.T024354.RF000.M009307.R0000116      NOT DELETED 8
IEF283I   VOL SER NOS= ISRB    1.
IEF283I   SYS72356.T024354.RF000.M009307.R0000130      NOT DELETED 8
IEF283I   VOL SER NOS= ISRA    1.
ISR011I STEP GO       EXECUTION TIME =   470.42 SEC.
ISR013I STEP GO       CORE USAGE =  102K HSC
ISR012I TOT. M009307  EXECUTION TIME =   470.42 SEC.
ISR016I TIME OF DAY =  4.19.18. DATE = 72.356
```

```
          INSTITUTE FOR SOCIAL RESEARCH MONITOR SYSTEM       7/1/72
     *****    M F T / H A S P  ***  M F T / H A S P     M F T / H A S P  ***  M F T / H A S P     *****
     *****                              OCTOBER 16, 1972                                          *****
     *****    ************************************************************************            *****
     *****                                                                                        *****
     *****    PLEASE REMEMBER TO BRING PROBLEMS DIRECTLY TO THE CSF COUNSELLORS IN ROOM 180       *****

*****TIME IS    3:59:55

*****LISTING OF SET-UP FOLLOWS:

   CARD                    1         2         3         4         5         6         7         8
    NO.           12345678901234567890123456789012345678901234567890123456789012345678901234567890
     1            $RUN THAID
     2            $RECODE
     3                 NAME R1'CHANGE FAMILY STRUCTURE'
     4                 IF(V1809 IN(5,6)OR V1109 IN(5,6)ORV542 IN(5,6))ANDV2072 IN(2-5)THEN
     5                *R1=1 AND GO TO END
     6                 IF(V1809 IN(5,6)OR V1109 IN(5,6)OR V542IN(5,6))AND V2072 EQ 1 THEN
     7                *R1=2 AND GO TO END
     8                 IF V1809 IN(4,7,8)OR V1109 IN(4,7,8)OR V542 IN(4,7,8)THEN R1=3 AND GOTOEND
     9                 IF V1809 EQ 3     OR V1109 EQ 3     OR V542 EQ 3     THEN R1=4 AND GOTOEND
    10                 IF(V1809 EQ 2     OR V1109 EQ 2     OR V542 EQ 2)AND V2072 EQ 1 THEN
    11                *R1=5 AND GO TO END
    12                 IF(V1809 EQ 2     OR V1109 EQ 2     OR V542 EQ 2)AND V2072 NE 1 THEN
    13                *R1=6 AND GO TO END
    14                 IF V1809 EQ 0 AND V1109 EQ 0 AND V542 EQ 0 THEN R1=0 AND GO TO END
    15                 IF V1941 GT V115 AND V2303 GT V116 THEN R1=7 AND GO TO END
    16                 IF V1941 GT V115 AND V2303 LE V116 THEN R1=8 AND GO TO END
    17                 IF V1941 LT V115 THEN R1=9 ELSE R1=0
    18            END  CONTINUE
    19            $SETUP
    20            THE IMPACT OF CHANGE IN FAMILY STRUCTURE FROM YEAR 1 TO 4
    21            *
    22            THE IMPACT OF CHANGE IN FAMILY STRUCTURE FROM YEAR 1 TO 4
    23            MINF=50,WEIGHT=(V2321,256),PRINT=1,DEPVAR=(R1,9)*
    24            PRED=(V368,V407,V414,V381)*
    25            PRED=(V181,V360),F,END*
    26            $END
```

CALLS FOR THE THAID PROGRAM

THIS IS A SEPARATE RECODE PROGRAM CREATING THE DEPENDENT "VARIABLE"

```
RECODE SCANNER BEGINS-    4    0  11   96

RECODE STATEMENTS:

    1            NAME R1'CHANGE FAMILY STRUCTURE'
    2            IF(V1809 IN(5,6)OR V1109 IN(5,6)ORV542 IN(5,6))ANDV2072 IN(2-5)THEN
                *R1=1 AND GO TO END
    3            IF(V1809 IN(5,6)OR V1109 IN(5,6)OR V542IN(5,6))AND V2072 EQ 1 THEN
                *R1=2 AND GO TO END
    4            IF V1809 IN(4,7,8)OR V1109 IN(4,7,8)OR V542 IN(4,7,8)THEN R1=3 AND GOTOEND
    5            IF V1809 EQ 3     OR V1109 EQ 3     OR V542 EQ 3     THEN R1=4 AND GOTOEND
    6            IF(V1809 EQ 2     OR V1109 EQ 2     OR V542 EQ 2)AND V2072 EQ 1 THEN
                *R1=5 AND GO TO END
    7            IF(V1809 EQ 2     OR V1109 EQ 2     OR V542 EQ 2)AND V2072 NE 1 THEN
                *R1=6 AND GO TO END
    8            IF V1809 EQ 0 AND V1109 EQ 0 AND V542 EQ 0 THEN R1=0 AND GO TO END
    9            IF V1941 GT V115 AND V2303 GT V116 THEN R1=7 AND GO TO END
   10            IF V1941 GT V115 AND V2303 LE V116 THEN R1=8 AND GO TO END
   11            IF V1941 LT V115 THEN R1=9 ELSE R1=0
   12      END   CONTINUE

RECODE SCANNER ENDS -    4    0  26   31
```

```
THAID    OSIRIS    NOMINAL SCALE SEQUENTIAL ANALYSIS    DECEMBER 19, 1972

                   THE IMPACT OF CHANGE IN FAMILY STRUCTURE FROM YEAR 1 TO 4

*

MINF=50,WEIGHT=(V2321,256),PRINT=1,DEPVAR=(R1,9)*

PRED=(V368,V407,V414,V381)*

PRED=(V181,V360),F,END*

THE VARIABLE LIST IS:

  -1  368  407  414  381  181  360 2321
```

	VAR.	TYPE	VARIABLE NAME	TLOC	WIDTH	NODEC	RESP.	MDCODE1	MDCODE2	REFNO	ID	TSEQNO
R	1	0	RECODED VARIABLE	0	0	0	1					
T	115	0	NUMBER IN FAMILY 5:51-52	285	2	0	1		0000099			00000
T	116	0	NO ADULTS IN FAM 5:53	287	1	0	1		0000009			00000
T	181	0	RACE 6:59	367	1	0	1		0000009			00000
T	360	0	SEX-MARITAL STATUS 8V360	631	1	0	1	0000009				00000
T	368	0	BKT AGE OF HEAD 8V368	639	1	0	1	0000009				00000
T	381	0	ACTUAL-REQ'D ROOMS 8V381	657	1	0	1	0000009				00000
T	407	0	BKT MONEY Y/NEEDS 8V407	709	1	0	1					00000
T	414	0	UNEMPLOY RATE 8V414	716	1	0	1					00000
T	542	0	CHANGE IN FU COMP 15:25	1070	1	0	1					00000
T	1109	0	CHANGE IN FU COMP 21:18	2021	1	0	1					00000
T	1809	0	CHANGE IN FU COMP 33:22	3521	1	0	1					00000
T	1941	0	# IN FU 38:40-41	3853	2	0	1					00000
T	2072	0	MARITAL STATUS 41:34	4027	1	0	1					00000
T	2303	0	NUMBER OF ADULTS 1V503	4384	2	0	1					00000
T	2321	0	WEIGHT 1V521	4412	2	0	1				COR	00000

TOTAL N (AFTER APPLICATION OF THE TOTAL FILTER,IF REQUESTED) IS 4840

THIS LISTS THE VARIABLES WITH THEIR SPECIFICATIONS FROM THE DICTIONARY WHICH IS WITH THE DATA SOURCE

THE IMPACT OF CHANGE IN FAMILY STRUCTURE FROM YEAR 1 TO 4

*****TIME 4 10 50 35

NUMBER OF CASES = 4840

*** UNIVARIATE DISTRIBUTION ***

CODES :	0	1	2	3	4	5	6	7	8	9
FREQUENCIES :	78327	7817	14122	4676	5972	2747	3007	4758	12426	18980
PERCENTAGES :	0.513	0.051	0.092	0.031	0.039	0.018	0.020	0.031	0.081	0.124

THIS IS THE OVERALL WEIGHTED DISTRIBUTION OF THE DEPENDENT (CRITERION) CLASSIFICATION

```
MODAL CODE    =      0
MODAL FREQ    =  78327
THETA         =  0.513
UNWEIGHTED N  =   4840
WEIGHTED   N  = 152832
```

PREDICTOR NAME	VARIABLE NO.	MAXIMUM CODE	SCALE
BKT AGE OF HEAD 8V368	368	9	ORDINAL
BKT MONEY Y/NEEDS 8V407	407	9	ORDINAL
UNEMPLOY RATE 8V414	414	9	ORDINAL
ACTUAL-REQ'D ROOMS 8V381	381	9	ORDINAL
RACE 6:59	181	9	NOMINAL
SEX-MARITAL STATUS 8V360	360	9	NOMINAL

THIS REPEATS WHAT THE SET-UP INSTRUCTIONS WERE, AS TO PREDICTORS, AND WHETHER THEIR ORDER IS TO BE MAINTAINED

DEP. VAR. NAME	VARIABLE NO.	MAXIMUM CODE
CHANGE FAMILY STRUCTURE	-1	9

AND THE DEPENDENT VARIABLE

MINIMUM THETA/DELTA INCREASE = 0.0
MINIMUM GROUP FREQUENCY = 50

AND THE STRATEGY

*** WEIGHT VAR. NO.2321 MAX. CODE = 256

AND THE VARIABLE LOCATION

*** TRIAL DISTRIBUTIONS REQUESTED , BIVARIATE TABLES THROUGH ITERATION 1

AND WHETHER, FOR THE WHOLE SAMPLE, ALL THE TWO-WAY TABLES ARE TO BE PRINTED OUT

```
***TRIAL UNIVARIATE DISTRIBUTIONS FOR ITERATION NO. 1  SPLIT NO. 1

                              *** CHILD NO.  2                                              *** CHILD NO.  3

VARIABLE NO.      CODES :    0    1    2    3    4    5    6    7    8    9          0    1    2    3    4    5    6    7    8    9


BKT AGE OF HEAD 8V368
VAR. NO. 368              16379  659  979 3232 1808 1536 1374  71310568  775        61948 715813143 1444 4164 1211 1633 4045 185818205
PERCENT                    0.43 0.02 0.03 0.09 0.05 0.04 0.04 0.02 0.28 0.02         0.54 0.06 0.11 0.01 0.04 0.01 0.01 0.04 0.02 0.16
PERCENT DIFF.             -0.08-0.03-0.07 0.05 0.01 0.02 0.02-0.01 0.20-0.10         0.03 0.01 0.02-0.02-0.00-0.01-0.01 0.00-0.07 0.03
SPLIT CODES :              0 1 2**************                                           3 4 5 6 7 8 9******
THETA/DELTA = 0.168

*** BIVARIATE TABLE *** IND. VAR. (ROWS) X DEP. VAR. (COLS.)

               0              0    0    0    0    0    0    0    0    0    0

               1           4170  195  334 1716  606  981  415   94 4039   48

               2          12209  464  645 1516 1202  555  959  619 6529  727

               3          17543 2289 3849  569 1244  549  594 1181 1395 5989

               4          12334 3211 6380  482  613  475  449 1443  202 8083

               5          15642 1294 2400  305  839  143   93  913  211 2775

               6          10974  208  288   88  872   44  352  330   50 1078

               7           5391   52  226    0  596    0  145  178    0  228

               8              0    0    0    0    0    0    0    0    0    0

               9             64   64    0    0    0    0    0    0    0   52

BKT MONEY Y/NEEDS 8V407
VAR. NO. 407              35115 3568 6751 2723 3081 1285 1482 2485 4813 9162        43212 4249 7371 1953 2891 1462 1525 2273 7613 9818
PERCENT                    0.50 0.05 0.10 0.04 0.04 0.02 0.02 0.04 0.07 0.13         0.52 0.05 0.09 0.02 0.04 0.02 0.02 0.03 0.09 0.12
PERCENT DIFF.             -0.01-0.00 0.00 0.01 0.00 0.00 0.00 0.00-0.01 0.01         0.01 0.00-0.00-0.01-0.00-0.00-0.00-0.00 0.01-0.00
SPLIT CODES :              0 1 2 3 4**********                                           5 6 7 8 9**********
THETA/DELTA = 0.029

*** BIVARIATE TABLE *** IND. VAR. (ROWS) X DEP. VAR. (COLS.)

               0           1451  357  427  180  208   21  189  185   83  375

               1           7178  737 1647  544  486   98  272  498  683 1946

               2           8654  740 1261  876  904  444  369  615 1111 2159

               3           9457 1083 1201  394  852  325  190  665 1190 2069

               4           8375  651 2215  729  631  397  462  522 1746 2613

               5           7772  923 1167  491  639  346  175  431 1736 2172

               6           9694  891 2030  533  626  150  487  433 1892 2723

               7          11609 1288 1883  554  580  497  262  706 2325 2578
```

THIS SPECIFIES THE BEST DIVISION OF THIS PREDICTOR AND GIVES THE TWO NEW DISTRIBUTIONS

THIS GIVES THE EXPLANATORY POWER OF THIS SPLIT, RANGING FROM 0 TO 1,000

8	9731	1057	1642	312	889	356	542	458	1359	1802
9	4406	90	649	63	157	113	59	245	301	543

```
UNEMPLOY RATE      8V414
VAR. NO. 414        49476 4674 9185 3101 4104 1865 1963 2754 832711920     28851 3143 4937 1575 1868  882 1044 2004 4099 7060
PERCENT              0.51 0.05 0.09 0.03 0.04 0.02 0.02 0.03 0.09 0.12      0.52 0.06 0.09 0.03 0.03 0.02 0.02 0.04 0.07 0.13
PERCENT DIFF.       -0.00-0.00 0.00 0.00 0.00 0.00 0.00-0.00 0.00-0.00      0.01 0.01-0.00-0.00-0.01-0.00-0.00 0.00-0.01 0.00
SPLIT CODES :         0 1 2**************                                      3 4 5 6 7 8 9*******
THETA/DELTA = 0.017
```

*** BIVARIATE TABLE *** IND. VAR. (ROWS) X DEP. VAR. (COLS.)

0	0	0	0	0	0	0	0	0	0	0
1	5640	222	757	151	277	51	183	478	1013	1092
2	43836	4452	8428	2950	3627	1814	1780	2276	731410828	
3	20856	2304	3378	1273	1215	636	786	1462	3042	5089
4	6282	588	1123	268	391	246	214	274	707	1377
5	1713	251	436	34	262	0	44	268	350	594
6	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0

```
ACTUAL-REQ'D ROOMS 8V381
VAR. NO. 381        24291 4216 7266 3239 2053 1717 1135 1874 5413 9174     54036 3601 6856 1437 3919 1030 1872 2884 7013 9806
PERCENT              0.40 0.07 0.12 0.05 0.03 0.03 0.02 0.03 0.09 0.15      0.58 0.04 0.07 0.02 0.04 0.01 0.02 0.03 0.08 0.11
PERCENT DIFF.       -0.11 0.02 0.03 0.02-0.01 0.01-0.00-0.00 0.01 0.03      0.07-0.01-0.02-0.02 0.00-0.01 0.00 0.00-0.01-0.02
SPLIT CODES :         0 1 2 3 4**********                                      5 6 7 8 9**********
THETA/DELTA = 0.104
```

*** BIVARIATE TABLE *** IND. VAR. (ROWS) X DEP. VAR. (COLS.)

0	253	91	210	232	99	0	37	74	1	203
1	610	299	591	71	5	0	44	61	18	540
2	2361	588	790	497	199	574	104	144	135	1005
3	5654	1249	2153	957	484	294	267	605	1356	2651
4	15413	1989	3522	1482	1266	849	683	990	3903	4775
5	19107	1567	2491	888	1704	487	717	1184	2944	4308
6	17555	1134	2330	443	1162	445	729	829	1994	2670
7	10359	548	1097	61	497	48	219	344	1188	1661
8	5725	137	802	45	506	50	207	455	733	866
9	1290	215	136	0	50	0	0	72	154	301

```
VAR. NO. 181          69961 631912523 4207 5147 2346 2410 36361083516062      8366 1498 1599  469  825  401  597 1122 1591 2918
PERCENT                0.52 0.05 0.09 0.03 0.04 0.02 0.02 0.03 0.08 0.12      0.43 0.08 0.08 0.02 0.04 0.02 0.03 0.06 0.08 0.15
PERCENT DIFF.          0.01-0.00 0.00 0.00-0.00-0.00-0.00-0.00-0.00-0.00     -0.08 0.03-0.01-0.01 0.00 0.00 0.01 0.03 0.00 0.03
SPLIT CODES :          1 9****************                                      0 2 3 4 5 6 7 8****
THETA/DELTA = 0.028

*** BIVARIATE TABLE *** IND. VAR. (ROWS) X DEP. VAR. (COLS.)

              0            0    0    0    0    0    0    0    0    0    0

              1        69772 631512521 4206 5147 2344 2410 36361079716010

              2         7059 1422 1356  391  707  271  589  780 1336 2516

              3          939   31  199   74   98   85    8  195  194  257

              4            0    0    0    0    0    0    0    0    0    0

              5            0    0    0    0    0    0    0    0    0    0

              6            0    0    0    0    0    0    0    0    0    0

              7          368   45   44    4   20   45    0  147   61  145

              8            0    0    0    0    0    0    0    0    0    0

              9          189    4    2    1    0    2    0    0   38   52

SEX-MARITAL STATUS 8V360
VAR. NO. 360          55438 542210951 1005 5925 1194 2936 30691192914792     22889 2395 3171 3671   47 1553   71 1689  497 4188
PERCENT                0.49 0.05 0.10 0.01 0.05 0.01 0.03 0.03 0.11 0.13      0.57 0.06 0.08 0.09 0.00 0.04 0.00 0.04 0.01 0.10
PERCENT DIFF.         -0.02-0.00 0.00-0.02 0.01-0.01 0.01-0.00 0.02 0.01      0.06 0.01-0.01 0.06-0.04 0.02-0.02 0.01-0.07-0.02
SPLIT CODES :          3*****************                                       0 1 2 4 5 6 7 8 9**
THETA/DELTA = 0.094

*** BIVARIATE TABLE *** IND. VAR. (ROWS) X DEP. VAR. (COLS.)

              0            0    0    0    0    0    0    0    0    0    0

              1         6331  629  686   51    0 1550   70  378   66  604

              2        16383 1766 2458 3620   47    3    1 1311  431 3584

              3        55438 542210951 1005 5925 1194 2936 30691192914792

              4            0    0    0    0    0    0    0    0    0    0

              5            0    0    0    0    0    0    0    0    0    0

              6            0    0    0    0    0    0    0    0    0    0

              7            0    0    0    0    0    0    0    0    0    0

              8            0    0    0    0    0    0    0    0    0    0

              9          175    0   27    0    0    0    0    0    0    0

******************************
BEST PREDICTOR IS:
BKT AGE OF HEAD 8V368
```

THIS TELLS (TO SAVE SEARCHING) WHICH PREDICTOR'S BEST SPLIT WAS BEST OVER ALL THE PREDICTORS, AND HENCE WAS USED TO FORM THE TWO NEW GROUPS - CHILD NO. 2 AND CHILD

VAR. NO.   368
THETA/DELTA = 0.168
****

NO. 3 - NAMELY AGE RETURN TO THE BEGINNING OF THIS ITERATION TO FIND WHICH GROUPS

GROUP 2 IS NOW SEARCHED
FOR THE BEST SPLIT INTO
GROUPS 4 AND 5.

```
***TRIAL UNIVARIATE DISTRIBUTIONS FOR ITERATION NO. 2  SPLIT NO. 1

                           *** CHILD NO.  4                                            *** CHILD NO.  5

VARIABLE NO.     CODES :    0    1    2    3    4    5    6    7    8    9          0    1    2    3    4    5    6    7    8    9


BKT AGE OF HEAD 8V368
VAR. NO. 368             4170  195  334 1716  606  981  415   94 4039   48       12209  464  645 1516 1202  555  959  619 6529  727
PERCENT                  0.33 0.02 0.03 0.14 0.05 0.08 0.03 0.01 0.32 0.00        0.48 0.02 0.03 0.06 0.05 0.02 0.04 0.02 0.26 0.03
PERCENT DIFF.           -0.18-0.04-0.07 0.11 0.01 0.06 0.01-0.02 0.24-0.12       -0.03-0.03-0.07 0.03 0.01 0.00 0.02-0.01 0.18-0.10
SPLIT CODES :            0 1****************                                          2 3 4 5 6 7 8 9****
THETA/DELTA = 0.122

BKT MONEY Y/NEEDS 8V407
VAR. NO. 407             7302  372  624 1976  985  918  679  418 4171  433        9077  287  355 1256  823  618  695  295 6397  342
PERCENT                  0.41 0.02 0.03 0.11 0.06 0.05 0.04 0.02 0.23 0.02        0.45 0.01 0.02 0.06 0.04 0.03 0.03 0.01 0.32 0.02
PERCENT DIFF.           -0.10-0.03-0.06 0.08 0.02 0.03 0.02-0.01 0.15-0.10       -0.06-0.04-0.07 0.03 0.00 0.01 0.01-0.02 0.24-0.11
SPLIT CODES :            0 1 2 3 4**********                                          5 6 7 8 9**********
THETA/DELTA = 0.087

UNEMPLOY RATE      8V414
VAR. NO. 414             1184    3   37    2   40    5   36  148  897  138       15195  656  942 3230 1768 1531 1338  565 9671  637
PERCENT                  0.48 0.00 0.01 0.00 0.02 0.00 0.01 0.06 0.36 0.06        0.43 0.02 0.03 0.09 0.05 0.04 0.04 0.02 0.27 0.02
PERCENT DIFF.           -0.04-0.05-0.08-0.03-0.02-0.02-0.01 0.03 0.28-0.07       -0.08-0.03-0.07 0.06 0.01 0.03 0.02-0.02 0.19-0.11
SPLIT CODES :            0 1****************                                          2 3 4 5 6 7 8 9****
THETA/DELTA = 0.037

ACTUAL-REQ'D ROOMS 8V381
VAR. NO. 381             7851  468  531 2394  828 1123  640  547 4712  595        8528  191  448  838  980  413  734  166 5856  180
PERCENT                  0.40 0.02 0.03 0.12 0.04 0.06 0.03 0.03 0.24 0.03        0.47 0.01 0.02 0.05 0.05 0.02 0.04 0.01 0.32 0.01
PERCENT DIFF.           -0.11-0.03-0.07 0.09 0.00 0.04 0.01-0.00 0.16-0.09       -0.05-0.04-0.07 0.02 0.01 0.00 0.02-0.02 0.24-0.11
SPLIT CODES :            0 1 2 3 4**********                                          5 6 7 8 9**********
THETA/DELTA = 0.114

RACE 6:59
VAR. NO. 181            14358  556  764 2877 1528 1406 1115  437 9407  528        2021  103  215  355  280  130  259  276 1161  247
PERCENT                  0.44 0.02 0.02 0.09 0.05 0.04 0.03 0.01 0.29 0.02        0.40 0.02 0.04 0.07 0.06 0.03 0.05 0.05 0.23 0.05
PERCENT DIFF.           -0.08-0.03-0.07 0.06 0.01 0.02 0.01-0.02 0.20-0.11       -0.11-0.03-0.05 0.04 0.02 0.01 0.03 0.02 0.15-0.08
SPLIT CODES :            1 9****************                                          0 2 3 4 5 6 7 8****
THETA/DELTA = 0.039

SEX-MARITAL STATUS 8V360
VAR. NO. 360            12266  405  485  762 1807  446 1369  45910164  485        4113  254  494 2470    1 1090    5  254  404  290
PERCENT                  0.43 0.01 0.02 0.03 0.06 0.02 0.05 0.02 0.35 0.02        0.44 0.03 0.05 0.26 0.00 0.12 0.00 0.03 0.04 0.03
PERCENT CIFF.           -0.08-0.04-0.08-0.00 0.02-0.00 0.03-0.02 0.27-0.11       -0.07-0.02-0.04 0.23-0.04 0.10-0.02-0.00-0.04-0.09
SPLIT CODES :            3*****************                                           0 1 2 4 5 6 7 8 9**
THETA/DELTA = 0.217

*******************************
BEST PREDICTOR IS:
SEX-MARITAL STATUS 8V360
VAR. NO.   360
THETA/DELTA = 0.217
*******************************
```

WHICH TURNS OUT TO BE A SPLIT
ON SEX-MARITAL STATUS

GROUP 3 IS NOW SEARCHED FOR THE BEST SPLIT INTO GROUPS 6 AND 7.

```
***TRIAL UNIVARIATE DISTRIBUTIONS FOR ITERATION NO. 2  SPLIT NO. 2

                             *** CHILD NO.  6                                          *** CHILD NO.  7

VARIABLE NO.     CODES :    0    1    2    3    4    5    6    7    8    9          0    1    2    3    4    5    6    7    8    9


BKT AGE OF HEAD 8V368
VAR. NO. 368             29877 550010229 1051 1857 1024 1043 2624 159714072     32071 1658 2914  393 2307  187  590 1421  261 4133
PERCENT                   0.43 0.08 0.15 0.02 0.03 0.01 0.02 0.04 0.02 0.20      0.70 0.04 0.06 0.01 0.05 0.00 0.01 0.03 0.01 0.09
PERCENT DIFF.            -0.08 0.03 0.06-0.02-0.01-0.00-0.00 0.01-0.06 0.08      0.19-0.02-0.03-0.02 0.01-0.01-0.01-0.00-0.08-0.03
SPLIT CODES :             0 1 2 3 4**********                                       5 6 7 8 9**********
THETA/DELTA = 0.140

BKT MONEY Y/NEEDS 8V407
VAR. NO. 407             40452 4871 9222 1124 2989  732 1226 2833 136113481     21496 2287 3921  320 1175  479  407 1212  497 4724
PERCENT                   0.52 0.06 0.12 0.01 0.04 0.01 0.02 0.04 0.02 0.17      0.59 0.06 0.11 0.01 0.03 0.01 0.01 0.03 0.01 0.13
PERCENT DIFF.             0.00 0.01 0.03-0.02-0.00-0.01-0.00 0.01-0.06 0.05      0.08 0.01 0.01-0.02-0.01-0.00-0.01 0.00-0.07 0.01
SPLIT CODES :             0 1 2 3 4 5 6******                                       7 8 9**************
THETA/DELTA = 0.033

UNEMPLOY RATE      8V414
VAR. NO. 414             38639 4277 8320  911 2955  937 1033 2274 119911254     23309 2881 4823  533 1209  274  600 1771  659 6951
PERCENT                   0.54 0.06 0.12 0.01 0.04 0.01 0.01 0.03 0.02 0.16      0.54 0.07 0.11 0.01 0.03 0.01 0.01 0.04 0.02 0.16
PERCENT DIFF.             0.03 0.01 0.02-0.02 0.00-0.00-0.01 0.00-0.06 0.03      0.03 0.02 0.02-0.02-0.01-0.01-0.01 0.01-0.07 0.04
SPLIT CODES :             0 1 2**************                                       3 4 5 6 7 8 9******
THETA/DELTA = 0.012

ACTUAL-REQ'D ROOMS 8V381
VAR. NO. 381             16440 3748 6735  845 1225  594  495 1327  701 8579     45508 3410 6408  599 2939  617 1138 2718 1157 9626
PERCENT                   0.40 0.09 0.17 0.02 0.03 0.01 0.01 0.03 0.02 0.21      0.61 0.05 0.09 0.01 0.04 0.01 0.02 0.04 0.02 0.13
PERCENT DIFF.            -0.11 0.04 0.07-0.01-0.01-0.00-0.01 0.00-0.06 0.09      0.10-0.01-0.01-0.02 0.00-0.01-0.00 0.01-0.07 0.01
SPLIT CODES :             0 1 2 3 4**********                                       5 6 7 8 9**********
THETA/DELTA = 0.105

RACE 6:59
VAR. NO. 181             55603 576311759 1330 3619  940 1295 3199 142815534      6345 1395 1384  114  545  271  338  846  430 2671
PERCENT                   0.55 0.06 0.12 0.01 0.04 0.01 0.01 0.03 0.01 0.15      0.44 0.10 0.10 0.01 0.04 0.02 0.02 0.06 0.03 0.19
PERCENT DIFF.             0.04 0.01 0.02-0.02-0.00-0.01-0.01 0.00-0.07 0.03     -0.07 0.05 0.00-0.02-0.00 0.00 0.00 0.03-0.05 0.06
SPLIT CODES :             1 9****************                                       0 2 3 4 5 6 7 8****
THETA/DELTA = 0.030

SEX-MARITAL STATUS 8V360
VAR. NO. 360             43172 501710466  243 4118  748 1567 2610 176514307     18776 2141 2677 1201   46  463   66 1435   93 3898
PERCENT                   0.51 0.06 0.12 0.00 0.05 0.01 0.02 0.03 0.02 0.17      0.61 0.07 0.09 0.04 0.00 0.02 0.00 0.05 0.00 0.13
PERCENT DIFF.             0.00 0.01 0.03-0.03 0.01-0.01-0.00-0.00-0.06 0.05      0.10 0.02-0.01 0.01-0.04-0.00-0.02 0.02-0.08 0.00
SPLIT CODES :             3******************                                       0 1 2 4 5 6 7 8 9**
THETA/DELTA = 0.065

******************************
BEST PREDICTOR IS:
BKT AGE OF HEAD 8V368
VAR. NO.   368
THETA/DELTA = 0.140
******************************
```

WHICH TURNS OUT TO BE AGE AGAIN

WE NOW START SEARCHING THE GROUPS IN THE THIRD ROW TO FORM GROUPS 8-15 OF THE FOURTH

```
***TRIAL UNIVARIATE DISTRIBUTIONS FOR ITERATION NO. 3  SPLIT NO. 1

                         *** CHILD NO.  8                                          *** CHILD NO.  9

VARIABLE NO.     CODES :   0    1    2    3    4    5    6    7    8    9          0    1    2    3    4    5    6    7    8    9


BKT AGE OF HEAD 8V368
VAR. NO. 368             2205   45  104  340  605  210  414   50 3975   37       10061  360  381  422 1202  236  955  409 6189  448
PERCENT                  0.28 0.01 0.01 0.04 0.08 0.03 0.05 0.01 0.50 0.00        0.49 0.02 0.02 0.02 0.06 0.01 0.05 0.02 0.30 0.02
PERCENT DIFF.           -0.24-0.05-0.08 0.01 0.04 0.01 0.03-0.02 0.42-0.12       -0.03-0.03-0.07-0.01 0.02-0.01 0.03-0.01 0.22-0.10
SPLIT CODES :            0 1****************                                         2 3 4 5 6 7 8 9****
THETA/DELTA = 0.152
```

NOTE THAT THIS GROUP HAS ONLY ONE AGE GROUP SO NO SPLITS CAN BE TRIED ON AGE

```
BKT MONEY Y/NEEDS 8V407
VAR. NO. 407             4734  264  290  379  984  272  674  209 3876  249        7532  141  195  383  823  174  695  250 6288  236
PERCENT                  0.40 0.02 0.02 0.03 0.08 0.02 0.06 0.02 0.32 0.02        0.45 0.01 0.01 0.02 0.05 0.01 0.04 0.01 0.38 0.01
PERCENT DIFF.           -0.12-0.03-0.07 0.00 0.04 0.00 0.04-0.01 0.24-0.10       -0.06-0.04-0.08-0.01 0.01-0.01 0.02-0.02 0.29-0.11
SPLIT CODES :            0 1 2 3 4**********                                         5 6 7 8 9**********
THETA/DELTA = 0.075

UNEMPLOY RATE      8V414
VAR. NO. 414             8019  219  465  374 1148  231  929  337 6889  430        4247  186   20  388  659  215  440  122 3275   55
PERCENT                  0.42 0.01 0.02 0.02 0.06 0.01 0.05 0.02 0.36 0.02        0.44 0.02 0.00 0.04 0.07 0.02 0.05 0.01 0.34 0.01
PERCENT DIFF.           -0.09-0.04-0.07-0.01 0.02-0.01 0.03-0.01 0.28-0.10       -0.07-0.03-0.09 0.01 0.03 0.00 0.03-0.02 0.26-0.12
SPLIT CODES :            0 1 2**************                                         3 4 5 6 7 8 9******
THETA/DELTA = 0.044

ACTUAL-REQ'D ROOMS 8V381
VAR. NO. 381             4972  278  260  487  827  286  635  349 4445  324        7294  127  225  275  980  160  734  110 5719  161
PERCENT                  0.39 0.02 0.02 0.04 0.06 0.02 0.05 0.03 0.35 0.03        0.46 0.01 0.01 0.02 0.06 0.01 0.05 0.01 0.36 0.01
PERCENT DIFF.           -0.13-0.03-0.07 0.01 0.03 0.00 0.03-0.00 0.26-0.10       -0.05-0.04-0.08-0.01 0.02-0.01 0.03-0.02 0.28-0.11
SPLIT CODES :            0 1 2 3 4**********                                         5 6 7 8 9**********
THETA/DELTA = 0.067

RACE 6:59
VAR. NO. 181            10995  352  387  710 1528  436 1115  316 9257  366        1271   53   98   52  279   10  254  143  907  119
PERCENT                  0.43 0.01 0.02 0.03 0.06 0.02 0.04 0.01 0.36 0.01        0.40 0.02 0.03 0.02 0.09 0.00 0.08 0.04 0.28 0.04
PERCENT DIFF.           -0.08-0.04-0.08-0.00 0.02-0.00 0.02-0.02 0.28-0.11       -0.11-0.03-0.06-0.01 0.05-0.01 0.06 0.01 0.20-0.09
SPLIT CODES :            1 9****************                                         0 2 3 4 5 6 7 8****
THETA/DELTA = 0.040

******************************
BEST PREDICTOR IS:
BKT AGE OF HEAD 8V368
VAR. NO.   368
THETA/DELTA = 0.152
******************************
```

```
***TRIAL UNIVARIATE DISTRIBUTIONS FOR ITERATION NO. 3  SPLIT NO. 2

                         *** CHILD NO. 10                                    *** CHILD NO. 11

VARIABLE NO.      CODES :    0    1    2    3    4    5    6    7    8    9        0    1    2    3    4    5    6    7    8    9


BKT AGE OF HEAD 8V368
VAR. NO. 368             1965  150  230 1376    1  771    1   44   64   11       2148  104  264 1094    0  319    4  210  340  279
PERCENT                  0.43 0.03 0.05 0.30 0.00 0.17 0.00 0.01 0.01 0.00       0.45 0.02 0.06 0.23 0.0  0.07 0.00 0.04 0.07 0.06
PERCENT DIFF.           -0.09-0.02-0.04 0.27-0.04 0.15-0.02-0.02-0.07-0.12      -0.06-0.03-0.04 0.20-0.04 0.05-0.02 0.01-0.01-0.07
SPLIT CODES :            0 1****************                                       2 3 4 5 6 7 8 9****
THETA/DELTA = 0.125

BKT MONEY Y/NEEDS 8V407
VAR. NO. 407             2236  108  167 1190    1  524    5  110  295  106       1877  146  327 1280    0  566    0  144  109  184
PERCENT                  0.47 0.02 0.04 0.25 0.00 0.11 0.00 0.02 0.06 0.02       0.41 0.03 0.07 0.28 0.0  0.12 0.0  0.03 0.02 0.04
PERCENT DIFF.           -0.04-0.03-0.06 0.22-0.04 0.09-0.02-0.01-0.02-0.10      -0.11-0.02-0.02 0.25-0.04 0.10-0.02-0.00-0.06-0.08
SPLIT CODES :            0 1 2 3************                                       4 5 6 7 8 9********
THETA/DELTA = 0.074

UNEMPLOY RATE      8V414
VAR. NO. 414             2818  178  400 1816    1  697    1  143  239  236       1295   76   94  654    0  393    4  111  165   54
PERCENT                  0.43 0.03 0.06 0.28 0.00 0.11 0.00 0.02 0.04 0.04       0.46 0.03 0.03 0.23 0.0  0.14 0.00 0.04 0.06 0.02
PERCENT DIFF.           -0.08-0.02-0.03 0.25-0.04 0.09-0.02-0.01-0.04-0.09      -0.06-0.02-0.06 0.20-0.04 0.12-0.02 0.01-0.02-0.11
SPLIT CODES :            0 1 2**************                                       3 4 5 6 7 8 9******
THETA/DELTA = 0.056

ACTUAL-REQ'D ROOMS 8V381
VAR. NO. 381              481   38  127  533    0  374    0    9   23   64       3632  216  367 1937    1  716    5  245  381  226
PERCENT                  0.29 0.02 0.08 0.32 0.0  0.23 0.0  0.01 0.01 0.04       0.47 0.03 0.05 0.25 0.00 0.09 0.00 0.03 0.05 0.03
PERCENT DIFF.           -0.22-0.03-0.02 0.29-0.04 0.21-0.02-0.03-0.07-0.09      -0.04-0.02-0.04 0.22-0.04 0.07-0.02 0.00-0.03-0.09
SPLIT CODES :            0 1 2**************                                       3 4 5 6 7 8 9******
THETA/DELTA = 0.099

RACE 6:59
VAR. NO. 181             3367  200  406 2218    0 1023    0  121  150  158        746   54   88  252    1   67    5  133  254  132
PERCENT                  0.44 0.03 0.05 0.29 0.0  0.13 0.0  0.02 0.02 0.02       0.43 0.03 0.05 0.15 0.00 0.04 0.00 0.08 0.15 0.08
PERCENT DIFF.           -0.07-0.02-0.04 0.26-0.04 0.12-0.02-0.02-0.06-0.10      -0.08-0.02-0.04 0.11-0.04 0.02-0.02 0.05 0.07-0.05
SPLIT CODES :            1 3****************                                       0 2 4 5 6 7 8 9****
THETA/DELTA = 0.106

SEX-MARITAL STATUS 8V360
VAR. NO. 360             2723  122  264 2455    1    3    1  176  354  218       1390  132  230   15    0 1087    4   78   50   72
PERCENT                  0.43 0.02 0.04 0.39 0.00 0.00 0.00 0.03 0.06 0.03       0.45 0.04 0.08 0.00 0.0  0.36 0.00 0.03 0.02 0.02
PERCENT DIFF.           -0.08-0.03-0.05 0.36-0.04-0.02-0.02-0.00-0.03-0.09      -0.06-0.01-0.02-0.03-0.04 0.34-0.02-0.01-0.06-0.10
SPLIT CODES :            2*****************                                        0 1 3 4 5 6 7 8 9**
THETA/DELTA = 0.268

******************************
BEST PREDICTOR IS:
SEX-MARITAL STATUS 8V360
VAR. NO.   360
THETA/DELTA = 0.268
******************************
```

```
***TRIAL UNIVARIATE DISTRIBUTIONS FOR ITERATION NO. 3  SPLIT NO. 3

                              *** CHILD NO. 12                                                  *** CHILD NO. 13

VARIABLE NO.      CODES :    0    1    2    3    4    5    6    7    8    9            0    1    2    3    4    5    6    7    8    9


BKT AGE OF HEAD 8V368
VAR. NO. 368             17543 2289 3849  569 1244  549  594 1181 1395 5989       12334 3211 6380  482  613  475  449 1443  202 8083
PERCENT                   0.50 0.07 0.11 0.02 0.04 0.02 0.02 0.03 0.04 0.17        0.37 0.10 0.19 0.01 0.02 0.01 0.01 0.04 0.01 0.24
PERCENT DIFF.            -0.01 0.01 0.02-0.01-0.00-0.00-0.00 0.00-0.04 0.05       -0.15 0.04 0.10-0.02-0.02-0.00-0.01 0.01-0.08 0.12
SPLIT CODES :             0 1 2 3************                                        4 5 6 7 8 9********
THETA/DELTA = 0.128

BKT MONEY Y/NEEDS 8V407
VAR. NO. 407             18843 3694 7482  779 1144  593  692 1885 119810233       11034 1806 2747  272  713  431  351  739  399 3839
PERCENT                   0.40 0.08 0.16 0.02 0.02 0.01 0.01 0.04 0.03 0.22        0.49 0.08 0.12 0.01 0.03 0.02 0.02 0.03 0.02 0.17
PERCENT DIFF.            -0.11 0.03 0.07-0.01-0.01-0.01-0.00 0.01-0.06 0.10       -0.02 0.03 0.03-0.02-0.01 0.00-0.00 0.00-0.06 0.05
SPLIT CODES :             0 1 2 3 4 5 6******                                        7 8 9**************
THETA/DELTA = 0.063

UNEMPLOY RATE      8V414
VAR. NO. 414             17823 3317 6465  563 1354  797  737 1502  993 8981       12054 2183 3764  488  503  227  306 1122  604 5091
PERCENT                   0.42 0.08 0.15 0.01 0.03 0.02 0.02 0.04 0.02 0.21        0.46 0.08 0.14 0.02 0.02 0.01 0.01 0.04 0.02 0.19
PERCENT DIFF.            -0.09 0.03 0.06-0.02-0.01 0.00-0.00 0.00-0.06 0.09       -0.05 0.03 0.05-0.01-0.02-0.01-0.01 0.01-0.06 0.07
SPLIT CODES :             0 1 2**************                                        3 4 5 6 7 8 9******
THETA/DELTA = 0.036

ACTUAL-REQ'D ROOMS 8V381
VAR. NO. 381              9904 2998 5543  643  742  589  450 1086  643 6634       19973 2502 4686  408 1115  435  593 1538  954 7438
PERCENT                   0.34 0.10 0.19 0.02 0.03 0.02 0.02 0.04 0.02 0.23        0.50 0.06 0.12 0.01 0.03 0.01 0.01 0.04 0.02 0.19
PERCENT DIFF.            -0.17 0.05 0.10-0.01-0.01 0.00-0.00 0.01-0.06 0.10       -0.01 0.01 0.03-0.02-0.01-0.01-0.00 0.01-0.06 0.06
SPLIT CODES :             0 1 2 3 4**********                                        5 6 7 8 9**********
THETA/DELTA = 0.113

RACE 6:59
VAR. NO. 181              3587 1069  958  106  280  266  233  571  317 1591       26290 4431 9271  945 1577  758  810 2053 128012481
PERCENT                   0.40 0.12 0.11 0.01 0.03 0.03 0.03 0.06 0.04 0.18        0.44 0.07 0.15 0.02 0.03 0.01 0.01 0.03 0.02 0.21
PERCENT DIFF.            -0.11 0.07 0.01-0.02-0.01 0.01 0.01 0.03-0.05 0.05       -0.07 0.02 0.06-0.01-0.01-0.01-0.01 0.00-0.06 0.08
SPLIT CODES :             2 3****************                                        0 1 4 5 6 7 8 9****
THETA/DELTA = 0.038

SEX-MARITAL STATUS 8V360
VAR. NO. 360              3917 1172 1706  772    0    0    0  673   72 2329       25960 4328 8523  279 1857 1024 1043 1951 152511743
PERCENT                   0.37 0.11 0.16 0.07 0.0  0.0  0.0  0.06 0.01 0.22        0.45 0.07 0.15 0.00 0.03 0.02 0.02 0.03 0.03 0.20
PERCENT DIFF.            -0.14 0.06 0.07 0.04-0.04-0.02-0.02 0.03-0.07 0.09       -0.07 0.02 0.05-0.03-0.01-0.00-0.00 0.00-0.06 0.08
SPLIT CODES :             2*****************                                         0 1 3 4 5 6 7 8 9**
THETA/DELTA = 0.058

******************************
BEST PREDICTOR IS:
BKT AGE OF HEAD 8V368
VAR. NO.    368
THETA/DELTA = 0.128
******************************
```

```
***TRIAL UNIVARIATE DISTRIBUTIONS FOR ITERATION NO. 3  SPLIT NO. 4

                        *** CHILD NO. 14                                          *** CHILD NO. 15

VARIABLE NO.     CODES :    0    1    2    3    4    5    6    7    8    9         0    1    2    3    4    5    6    7    8    9


BKT AGE OF HEAD 8V368
VAR. NO. 368             15642 1294 2400  305  839  143   93  913  211 2775     16429  364  514   88 1468   44  497  508   50 1358
PERCENT                   0.64 0.05 0.10 0.01 0.03 0.01 0.00 0.04 0.01 0.11      0.77 0.02 0.02 0.00 0.07 0.00 0.02 0.02 0.00 0.06
PERCENT DIFF.             0.12 0.00 0.01-0.02-0.00-0.01-0.02 0.01-0.07-0.01      0.26-0.03-0.07-0.03 0.03-0.02 0.00-0.01-0.08-0.06
SPLIT CODES :             0 1 2 3 4 5********                                       6 7 8 9************
THETA/DELTA = 0.190

BKT MONEY Y/NEEDS 8V407
VAR. NO. 407             21609 1177 1740  345 1845  139  534  948  163 3248     10462  481 1174   48  462   48   56  473   98  885
PERCENT                   0.68 0.04 0.05 0.01 0.06 0.00 0.02 0.03 0.01 0.10      0.74 0.03 0.08 0.00 0.03 0.00 0.00 0.03 0.01 0.06
PERCENT DIFF.             0.17-0.01-0.04-0.02 0.02-0.01-0.00-0.00-0.08-0.02      0.22-0.02-0.01-0.03-0.01-0.01-0.02 0.00-0.07-0.06
SPLIT CODES :             0 1 2 3 4 5 6******                                       7 8 9**************
THETA/DELTA = 0.078

UNEMPLOY RATE       8V414
VAR. NO. 414             20816  960 1855  348 1601  140  296  772  206 2273     11255  698 1059   45  706   47  294  649   55 1860
PERCENT                   0.71 0.03 0.06 0.01 0.05 0.00 0.01 0.03 0.01 0.08      0.68 0.04 0.06 0.00 0.04 0.00 0.02 0.04 0.00 0.11
PERCENT DIFF.             0.20-0.02-0.03-0.02 0.02-0.01-0.01-0.00-0.07-0.05      0.16-0.01-0.03-0.03 0.00-0.02-0.00 0.01-0.08-0.01
SPLIT CODES :             0 1 2**************                                       3 4 5 6 7 8 9******
THETA/DELTA = 0.059

ACTUAL-REQ'D ROOMS 8V381
VAR. NO. 381              6536  750 1192  202  483    5   45  241   58 1945     25535  908 1722  191 1824  182  545 1180  203 2188
PERCENT                   0.57 0.07 0.10 0.02 0.04 0.00 0.00 0.02 0.01 0.17      0.74 0.03 0.05 0.01 0.05 0.01 0.02 0.03 0.01 0.06
PERCENT DIFF.             0.06 0.01 0.01-0.01 0.00-0.02-0.02-0.01-0.08 0.05      0.23-0.02-0.04-0.03 0.01-0.01-0.00 0.00-0.08-0.06
SPLIT CODES :             0 1 2 3 4**********                                       5 6 7 8 9**********
THETA/DELTA = 0.160

RACE 6:59
VAR. NO. 181              2288  326  382    8  245    5  105  106  113  985     29783 1332 2532  385 2062  182  485 1315  148 3148
PERCENT                   0.50 0.07 0.08 0.00 0.05 0.00 0.02 0.02 0.02 0.22      0.72 0.03 0.06 0.01 0.05 0.00 0.01 0.03 0.00 0.08
PERCENT DIFF.            -0.01 0.02-0.01-0.03 0.01-0.02 0.00-0.01-0.06 0.09      0.21-0.02-0.03-0.02 0.01-0.01-0.01 0.00-0.08-0.05
SPLIT CODES :             2 7****************                                       0 1 3 4 5 6 8 9****
THETA/DELTA = 0.086

SEX-MARITAL STATUS 8V360
VAR. NO. 360             19061 1034 2276    0 2261  143  590  766  254 2869     13010  624  638  393   46   44    0  655    7 1264
PERCENT                   0.65 0.04 0.08 0.0  0.08 0.00 0.02 0.03 0.01 0.10      0.78 0.04 0.04 0.02 0.00 0.00 0.0  0.04 0.00 0.08
PERCENT DIFF.             0.14-0.02-0.01-0.03 0.04-0.01 0.00-0.00-0.07-0.03      0.27-0.01-0.05-0.01-0.04-0.02-0.02 0.01-0.08-0.05
SPLIT CODES :             3******************                                       0 1 2 4 5 6 7 8 9**
THETA/DELTA = 0.156

******************************
BEST PREDICTOR IS:
BKT AGE OF HEAD 8V368
VAR. NO.   368
THETA/DELTA = 0.190
******************************
```

WE NOW START SEARCHING THE GROUPS OF THE 4TH ROW TO FORM THE FINAL GROUPS 16-31

```
***TRIAL UNIVARIATE DISTRIBUTIONS FOR ITERATION NO. 4  SPLIT NO. 1

                          *** CHILD NO. 16                                     *** CHILD NO. 17

VARIABLE NO.     CODES :    0    1    2    3    4    5    6    7    8    9       0    1    2    3    4    5    6    7    8    9

BKT MONEY Y/NEEDS BV407
VAR. NO. 407              374   38   58  107  205  113  129    0  653    0    1831    7   46  233  400   97  285   50 3322   37
PERCENT                  0.22 0.02 0.03 0.06 0.12 0.07 0.08 0.0  0.39 0.0    0.29 0.00 0.01 0.04 0.06 0.02 0.05 0.01 0.53 0.01
PERCENT DIFF.           -0.29-0.03-0.06 0.03 0.08 0.05 0.06-0.03 0.31-0.12  -0.22-0.05-0.09 0.01 0.02-0.00 0.03-0.02 0.45-0.12
SPLIT CODES :            0 1 2***************                                  3 4 5 6 7 8 9******
THETA/DELTA = 0.109

UNEMPLOY RATE      BV414
VAR. NO. 414             1140    1   97  207  397   79  313   50 2802   30    1065   44    7  133  208  131  101    0 1173    7
PERCENT                  0.22 0.00 0.02 0.04 0.08 0.02 0.06 0.01 0.55 0.01   0.37 0.02 0.00 0.05 0.07 0.05 0.04 0.0  0.41 0.00
PERCENT DIFF.           -0.29-0.05-0.07 0.01 0.04-0.00 0.04-0.02 0.47-0.12  -0.14-0.04-0.09 0.02 0.03 0.03 0.02-0.03 0.33-0.12
SPLIT CODES :            0 1 2***************                                  3 4 5 6 7 8 9******
THETA/DELTA = 0.138

ACTUAL-REQ'D ROOMS BV381
VAR. NO. 381              966    8   58  226  286  132  240    2 2103   37    1239   37   46  114  319   78  174   48 1872    0
PERCENT                  0.24 0.00 0.01 0.06 0.07 0.03 0.06 0.00 0.52 0.01   0.32 0.01 0.01 0.03 0.08 0.02 0.04 0.01 0.48 0.0
PERCENT DIFF.           -0.27-0.05-0.08 0.03 0.03 0.01 0.04-0.03 0.44-0.12  -0.20-0.04-0.08-0.00 0.04 0.00 0.02-0.02 0.40-0.12
SPLIT CODES :            0 1 2 3 4***********                                  5 6 7 8 9**********
THETA/DELTA = 0.081

RACE 6:59
VAR. NO. 181              192    1    5    5   98    0  107    2  335    0    2013   44   99  335  507  210  307   48 3640   37
PERCENT                  0.26 0.00 0.01 0.01 0.13 0.0  0.14 0.00 0.45 0.0    0.28 0.01 0.01 0.05 0.07 0.03 0.04 0.01 0.50 0.01
PERCENT DIFF.           -0.25-0.05-0.09-0.02 0.09-0.02 0.12-0.03 0.37-0.12  -0.23-0.05-0.08 0.02 0.03 0.01 0.02-0.02 0.42-0.12
SPLIT CODES :            2*******************                                  0 1 3 4 5 6 7 8 9**
THETA/DELTA = 0.041

******************************
BEST PREDICTOR IS:
UNEMPLOY RATE      BV414
VAR. NO.   414
THETA/DELTA = 0.138
******************************
```

NOTE THAT THERE IS ONLY ONE AGE GROUP AND ONLY ONE SEX-MARITAL STATUS GROUP

```
***TRIAL UNIVARIATE DISTRIBUTIONS FOR ITERATION NO. 4  SPLIT NO. 2

                           *** CHILD NO. 18                                        *** CHILD NO. 19

VARIABLE NO.     CODES :    0    1    2    3    4    5    6    7    8    9          0    1    2    3    4    5    6    7    8    9


BKT MONEY Y/NEEDS 8V407
VAR. NO. 407             3462  219  232  125  746   62  492  209 1832  212       6599  141  149  297  456  174  463  200 4357  236
PERCENT                  0.46 0.03 0.03 0.02 0.10 0.01 0.06 0.03 0.24 0.03       0.50 0.01 0.01 0.02 0.03 0.01 0.04 0.02 0.33 0.02
PERCENT DIFF.           -0.06-0.02-0.06-0.01 0.06-0.01 0.05-0.00 0.16-0.10      -0.01-0.04-0.08-0.01-0.00-0.00 0.02-0.02 0.25-0.11
SPLIT CODES :            0 1 2 3 4**********                                         5 6 7 8 9**********
THETA/DELTA = 0.106

UNEMPLOY RATE      8V414
VAR. NO. 414             6879  218  368  167  751  152  616  287 4087  400       3182  142   13  255  451   84  339  122 2102   48
PERCENT                  0.49 0.02 0.03 0.01 0.05 0.01 0.04 0.02 0.29 0.03       0.47 0.02 0.00 0.04 0.07 0.01 0.05 0.02 0.31 0.01
PERCENT DIFF.           -0.02-0.04-0.07-0.02 0.01-0.01 0.02-0.01 0.21-0.10      -0.04-0.03-0.09 0.01 0.03-0.01 0.03-0.01 0.23-0.12
SPLIT CODES :            0 1 2**************                                         3 4 5 6 7 8 9******
THETA/DELTA = 0.046

ACTUAL-REQ'D ROOMS 8V381
VAR. NO. 381             4006  270  202  261  541  154  395  347 2342  287       6055   90  179  161  661   82  560   62 3847  161
PERCENT                  0.45 0.03 0.02 0.03 0.06 0.02 0.04 0.04 0.27 0.03       0.51 0.01 0.02 0.01 0.06 0.01 0.05 0.01 0.32 0.01
PERCENT DIFF.           -0.06-0.02-0.07-0.00 0.02-0.00 0.03 0.01 0.18-0.09      -0.00-0.04-0.08-0.02 0.02-0.01 0.03-0.03 0.24-0.11
SPLIT CODES :            0 1 2 3 4**********                                         5 6 7 8 9**********
THETA/DELTA = 0.086

RACE 6:59
VAR. NO. 181             8982  308  288  397 1021  226  808  268 5665  329       1079   52   93   25  181   10  147  141  524  119
PERCENT                  0.49 0.02 0.02 0.02 0.06 0.01 0.04 0.01 0.31 0.02       0.46 0.02 0.04 0.01 0.08 0.00 0.06 0.06 0.22 0.05
PERCENT DIFF.           -0.02-0.03-0.08-0.01 0.02-0.01 0.02-0.02 0.23-0.11      -0.06-0.03-0.05-0.02 0.04-0.01 0.04 0.03 0.14-0.07
SPLIT CODES :            1 9****************                                         0 2 3 4 5 6 7 8****
THETA/DELTA = 0.044

*******************************
BEST PREDICTOR IS:
BKT MONEY Y/NEEDS 8V407
VAR. NO.   407
THETA/DELTA = 0.106
*******************************
```

```
***TRIAL UNIVARIATE DISTRIBUTIONS FOR ITERATION NO. 4  SPLIT NO. 3

                       *** CHILD NO. 20                                                *** CHILD NO. 21

VARIABLE NO.     CODES :   0    1    2    3    4    5    6    7    8    9          0    1    2    3    4    5    6    7    8    9


BKT AGE OF HEAD BV368
VAR. NO. 368              1029   76   59 1369    1    3    1    0   64   11        1694   46  205 1086    0    0    0  176  290  207
PERCENT                   0.39 0.03 0.02 0.52 0.00 0.00 0.00 0.0  0.02 0.00        0.46 0.01 0.06 0.29 0.0  0.0  0.0  0.05 0.08 0.06
PERCENT DIFF.            -0.12-0.02-0.07 0.49-0.04-0.02-0.02-0.03-0.06-0.12       -0.06-0.04-0.04 0.26-0.04-0.02-0.02 0.02-0.00-0.07
SPLIT CODES :             0 1****************                                          2 3 4 5 6 7 8 9****
THETA/DELTA = 0.184

BKT MONEY Y/NEEDS BV407
VAR. NO. 407              1499   74   90 1175    1    3    1   76  295  106        1224   48  174 1280    0    0    0  100   59  112
PERCENT                   0.45 0.02 0.03 0.35 0.00 0.00 0.00 0.02 0.09 0.03        0.41 0.02 0.06 0.43 0.0  0.0  0.0  0.03 0.02 0.04
PERCENT DIFF.            -0.06-0.03-0.07 0.32-0.04-0.02-0.02-0.01 0.01-0.09       -0.10-0.04-0.03 0.40-0.04-0.02-0.02 0.00-0.06-0.09
SPLIT CODES :             0 1 2 3************                                          4 5 6 7 8 9********
THETA/DELTA = 0.091

UNEMPLOY RATE      BV414
VAR. NO. 414              1966   88  214 1801    1    3    1   65  189  186         757   34   50  654    0    0    0  111  165   32
PERCENT                   0.44 0.02 0.05 0.40 0.00 0.00 0.00 0.01 0.04 0.04        0.42 0.02 0.03 0.36 0.0  0.0  0.0  0.06 0.09 0.02
PERCENT DIFF.            -0.08-0.03-0.04 0.37-0.04-0.02-0.02-0.02-0.04-0.08       -0.09-0.03-0.06 0.33-0.04-0.02-0.02 0.03 0.01-0.11
SPLIT CODES :             0 1 2**************                                          3 4 5 6 7 8 9******
THETA/DELTA = 0.060

ACTUAL-REQ'D ROOMS BV381
VAR. NO. 381               267   38  127  533    0    0    0    7   23   64        2456   84  137 1922    1    3    1  169  331  154
PERCENT                   0.25 0.04 0.12 0.50 0.0  0.0  0.0  0.01 0.02 0.06        0.47 0.02 0.03 0.37 0.00 0.00 0.00 0.03 0.06 0.03
PERCENT DIFF.            -0.26-0.02 0.03 0.47-0.04-0.02-0.02-0.02-0.06-0.06       -0.05-0.04-0.07 0.33-0.04-0.02-0.02 0.00-0.02-0.09
SPLIT CODES :             0 1 2**************                                          3 4 5 6 7 8 9******
THETA/DELTA = 0.120

RACE 6:59
VAR. NO. 181              2128   68  147 2159    0    0    0   77  100   86         595   54  117  296    1    3    1   99  254  132
PERCENT                   0.45 0.01 0.03 0.45 0.0  0.0  0.0  0.02 0.02 0.02        0.38 0.03 0.08 0.19 0.00 0.00 0.00 0.06 0.16 0.09
PERCENT DIFF.            -0.07-0.04-0.06 0.42-0.04-0.02-0.02-0.01-0.06-0.11       -0.13-0.02-0.02 0.16-0.04-0.02-0.02 0.03 0.08-0.04
SPLIT CODES :             1******************                                          0 2 3 4 5 6 7 8 9**
THETA/DELTA = 0.184

******************************
BEST PREDICTOR IS:
BKT AGE OF HEAD BV368
VAR. NO.   368
THETA/DELTA = 0.184
******************************
```

```
***TRIAL UNIVARIATE DISTRIBUTIONS FOR ITERATION NO. 4  SPLIT NO. 4

                       *** CHILD NO. 22                                          *** CHILD NO. 23

VARIABLE NO.     CODES :   0    1    2    3    4    5    6    7    8    9        0    1    2    3    4    5    6    7    8    9


BKT AGE OF HEAD 8V368
VAR. NO. 368              936   74  171    7    0  768    0   44    0    0       454   58   59    8    0  319    4   34   50   72
PERCENT                  0.47 0.04 0.09 0.00 0.0  0.38 0.0  0.02 0.0  0.0       0.43 0.05 0.06 0.01 0.0  0.30 0.00 0.03 0.05 0.07
PERCENT DIFF.           -0.04-0.01-0.01-0.03-0.04 0.37-0.02-0.01-0.08-0.12     -0.08 0.00-0.04-0.02-0.04 0.28-0.02 0.00-0.03-0.06
SPLIT CODES :            0 1****************                                        2 3 4 5 6 7 8 9****
THETA/DELTA = 0.104

BKT MONEY Y/NEEDS 8V407
VAR. NO. 407              393    0   77    8    0   61    0    0    0    0       997  132  153    7    0 1026    4   78   50   72
PERCENT                  0.73 0.0  0.14 0.01 0.0  0.11 0.0  0.0  0.0  0.0       0.40 0.05 0.06 0.00 0.0  0.41 0.00 0.03 0.02 0.03
PERCENT DIFF.            0.22-0.05 0.05-0.02-0.04 0.10-0.02-0.03-0.08-0.12     -0.12 0.00-0.03-0.03-0.04 0.39-0.02-0.00-0.06-0.10
SPLIT CODES :            0 1****************                                        2 3 4 5 6 7 8 9****
THETA/DELTA = 0.189

UNEMPLOY RATE      8V414
VAR. NO. 414             1298  132  186   15    0  899    4   78   50   72        92    0   44    0    0  188    0    0    0    0
PERCENT                  0.47 0.05 0.07 0.01 0.0  0.33 0.00 0.03 0.02 0.03       0.28 0.0  0.14 0.0  0.0  0.58 0.0  0.0  0.0  0.0
PERCENT DIFF.           -0.04-0.00-0.02-0.03-0.04 0.31-0.02-0.00-0.06-0.10     -0.23-0.05 0.04-0.03-0.04 0.56-0.02-0.03-0.08-0.12
SPLIT CODES :            0 1 2 3************                                        4 5 6 7 8 9********
THETA/DELTA = 0.092

ACTUAL-REQ'D ROOMS 8V381
VAR. NO. 381              214    0    0    0    0  374    0    2    0    0      1176  132  230   15    0  713    4   76   50   72
PERCENT                  0.36 0.0  0.0  0.0  0.0  0.63 0.0  0.00 0.0  0.0       0.48 0.05 0.09 0.01 0.0  0.29 0.00 0.03 0.02 0.03
PERCENT DIFF.           -0.15-0.05-0.09-0.03-0.04 0.62-0.02-0.03-0.08-0.12     -0.04 0.00 0.00-0.02-0.04 0.27-0.02-0.00-0.06-0.10
SPLIT CODES :            0 1 2**************                                        3 4 5 6 7 8 9******
THETA/DELTA = 0.163

RACE 6:59
VAR. NO. 181              143    0    0    8    0   19    4   34    0    0      1247  132  230    7    0 1068    0   44   50   72
PERCENT                  0.69 0.0  0.0  0.04 0.0  0.09 0.02 0.16 0.0  0.0       0.44 0.05 0.08 0.00 0.0  0.37 0.0  0.02 0.02 0.03
PERCENT DIFF.            0.17-0.05-0.09 0.01-0.04 0.07-0.00 0.13-0.08-0.12     -0.07-0.00-0.01-0.03-0.04 0.36-0.02-0.02-0.06-0.10
SPLIT CODES :            2 9****************                                        0 1 3 4 5 6 7 8****
THETA/DELTA = 0.087

******************************
BEST PREDICTOR IS:
BKT MONEY Y/NEEDS 8V407
VAR. NO.   407
THETA/DELTA = 0.189
******************************
```

```
***TRIAL UNIVARIATE DISTRIBUTIONS FOR ITERATION NO. 4  SPLIT NO. 5

                              *** CHILD NO. 24                                          *** CHILD NO. 25

VARIABLE NO.      CODES :    0    1    2    3    4    5    6    7    8    9          0    1    2    3    4    5    6    7    8    9


BKT MONEY Y/NEEDS BV407
VAR. NO. 407               4727  913 1784  203  315  118  102  582  306 2411      12816 1376 2065  366  929  431  492  599 1089 3578
PERCENT                    0.41 0.08 0.16 0.02 0.03 0.01 0.01 0.05 0.03 0.21       0.54 0.06 0.09 0.02 0.04 0.02 0.02 0.03 0.05 0.15
PERCENT DIFF.             -0.10 0.03 0.06-0.01-0.01-0.01-0.01 0.02-0.05 0.09       0.03 0.01-0.01-0.02 0.00 0.00 0.00-0.01-0.04 0.03
SPLIT CODES :              0 1 2 3************                                       4 5 6 7 8 9********
THETA/DELTA = 0.111

UNEMPLOY RATE       BV414
VAR. NO. 414              15787 2013 3144  561 1164  549  568 1110 1255 5133       1756  276  705    8   80    0   26   71  140  856
PERCENT                    0.50 0.06 0.10 0.02 0.04 0.02 0.02 0.04 0.04 0.16       0.45 0.07 0.18 0.00 0.02 0.0  0.01 0.02 0.04 0.22
PERCENT DIFF.             -0.01 0.01 0.01-0.01-0.00-0.00-0.00 0.00-0.04 0.04      -0.06 0.02 0.09-0.03-0.02-0.02-0.01-0.01-0.05 0.09
SPLIT CODES :              0 1 2 3************                                       4 5 6 7 8 9********
THETA/DELTA = 0.039

ACTUAL-REQ'D ROOMS BV381
VAR. NO. 381               6595 1070 2519  309  544  303  148  571  581 3085      10948 1219 1330  260  700  246  446  610  814 2904
PERCENT                    0.42 0.07 0.16 0.02 0.03 0.02 0.01 0.04 0.04 0.20       0.56 0.06 0.07 0.01 0.04 0.01 0.02 0.03 0.04 0.15
PERCENT DIFF.             -0.09 0.02 0.07-0.01-0.00 0.00-0.01 0.01-0.04 0.07       0.05 0.01-0.02-0.02-0.00-0.01 0.00 0.00-0.04 0.02
SPLIT CODES :              0 1 2 3 4**********                                       5 6 7 8 9**********
THETA/DELTA = 0.114

RACE 6:59
VAR. NO. 181              15543 1818 3438  491 1052  449  493  829 1133 5154       2000  471  411   78  192  100  101  352  262  835
PERCENT                    0.51 0.06 0.11 0.02 0.03 0.01 0.02 0.03 0.04 0.17       0.42 0.10 0.09 0.02 0.04 0.02 0.02 0.07 0.05 0.17
PERCENT DIFF.             -0.00 0.01 0.02-0.01-0.00-0.00-0.00-0.00-0.04 0.05      -0.10 0.05-0.01-0.01 0.00 0.00 0.00 0.04-0.03 0.05
SPLIT CODES :              1 7****************                                       0 2 3 4 5 6 8 9****
THETA/DELTA = 0.041

SEX-MARITAL STATUS BV360
VAR. NO. 360               1833  330  799  492    0    0    0  235   54  976      15710 1959 3050   77 1244  549  594  946 1341 5013
PERCENT                    0.39 0.07 0.17 0.10 0.0  0.0  0.0  0.05 0.01 0.21       0.52 0.06 0.10 0.00 0.04 0.02 0.02 0.03 0.04 0.16
PERCENT DIFF.             -0.12 0.02 0.08 0.07-0.04-0.02-0.02 0.02-0.07 0.08       0.00 0.01 0.01-0.03 0.00 0.00-0.00-0.00-0.04 0.04
SPLIT CODES :              2*****************                                        0 1 3 4 5 6 7 8 9**
THETA/DELTA = 0.079

******************************
BEST PREDICTOR IS:
ACTUAL-REQ'D ROOMS BV381
VAR. NO.   381
THETA/DELTA = 0.114
******************************
```

```
***TRIAL UNIVARIATE DISTRIBUTIONS FOR ITERATION NO. 4  SPLIT NO. 6

                             *** CHILD NO. 26                                                      *** CHILD NO. 27

VARIABLE NO.      CODES :     0    1    2    3    4    5    6    7    8    9          0    1    2    3    4    5    6    7    8    9


BKT MONEY Y/NEEDS 8V407
VAR. NO. 407                6715 1688 4261  358  360  205  394  896  157 5276       5619 1523 2119  124  253  270   55  547   45 2807
PERCENT                     0.33 0.08 0.21 0.02 0.02 0.01 0.02 0.04 0.01 0.26       0.42 0.11 0.16 0.01 0.02 0.02 0.00 0.04 0.00 0.21
PERCENT DIFF.              -0.18 0.03 0.12-0.01-0.02-0.01-0.00 0.01-0.07 0.14      -0.09 0.06 0.07-0.02-0.02 0.00-0.02 0.01-0.08 0.09
SPLIT CODES :               0 1 2 3 4 5 6******                                         7 8 9***************
THETA/DELTA = 0.083

UNEMPLOY RATE      8V414
VAR. NO. 414                7027 18C7 4039  270  401  347  299  819  197 5250       5307 1404 2341  212  212  128  150  624    5 2833
PERCENT                     0.34 0.09 0.20 0.01 0.02 0.02 0.01 0.04 0.01 0.26       0.40 0.11 0.18 0.02 0.02 0.01 0.01 0.05 0.00 0.21
PERCENT DIFF.              -0.17 0.04 0.11-0.02-0.02-0.00-0.01 0.01-0.07 0.13      -0.11 0.06 0.08-0.01-0.02-0.01-0.01 0.02-0.08 0.09
SPLIT CODES :               0 1 2***************                                        3 4 5 6 7 8 9******
THETA/DELTA = 0.054

ACTUAL-REQ'D ROOMS 8V381
VAR. NO. 381                3309 1928 3024  334  198  286  302  515   62 3549       9025 1283 3356  148  415  189  147  928  140 4534
PERCENT                     0.24 0.14 0.22 0.02 0.01 0.02 0.02 0.04 0.00 0.26       0.45 0.06 0.17 0.01 0.02 0.01 0.01 0.05 0.01 0.22
PERCENT DIFF.              -0.27 0.09 0.13-0.01-0.02 0.00 0.00 0.01-0.08 0.14      -0.06 0.01 0.07-0.02-0.02-0.01-0.01 0.01-0.07 0.10
SPLIT CODES :               0 1 2 3 4**********                                         5 6 7 8 9**********
THETA/DELTA = 0.138

RACE 6:59
VAR. NO. 181               10620 2613 5831  454  525  307  317 1155  147 7231       1714  598  549   28   88  168  132  288   55  852
PERCENT                     0.36 0.09 0.20 0.02 0.02 0.01 0.01 0.04 0.01 0.25       0.38 0.13 0.12 0.01 0.02 0.04 0.03 0.06 0.01 0.19
PERCENT DIFF.              -0.15 0.04 0.11-0.02-0.02-0.01-0.01 0.01-0.08 0.12      -0.13 0.08 0.03-0.02-0.02 0.02 0.01 0.03-0.07 0.07
SPLIT CODES :               1******************                                         0 2 3 4 5 6 7 8 9**
THETA/DELTA = 0.043

SEX-MARITAL STATUS 8V360
VAR. NO. 360                9189 2103 5194  166  613  123  385  957  184 6573       3145 1108 1186  316    0  352   64  486   18 1510
PERCENT                     0.36 0.08 0.2C 0.01 0.02 0.00 0.02 0.04 0.01 0.26       0.38 0.14 0.14 0.04 0.0  0.04 0.01 0.06 0.00 0.18
PERCENT DIFF.              -0.15 0.03 0.11-0.02-0.02-0.01-0.00 0.01-0.07 0.13      -0.13 0.08 0.05 0.01-0.04 0.03-0.01 0.03-0.08 0.06
SPLIT CODES :               3******************                                         0 1 2 4 5 6 7 8 9**
THETA/DELTA = 0.082

******************************
BEST PREDICTOR IS:
ACTUAL-REQ'D ROOMS 8V381
VAR. NO.   381
THETA/DELTA = 0.138
******************************
```

```
***TRIAL UNIVARIATE DISTRIBUTIONS FOR ITERATION NO. 4  SPLIT NO. 7

                          *** CHILD NO. 28                                              *** CHILD NO. 29

VARIABLE NO.     CODES :    0    1    2    3    4    5    6    7    8    9        0    1    2    3    4    5    6    7    8    9


BKT MONEY Y/NEEDS 8V407
VAR. NO. 407             8063  813 1226  257  567   95   93  488  113 1988      7579  481 1174   48  272   48    0  425   98  787
PERCENT                  0.59 0.06 0.09 0.02 0.04 0.01 0.01 0.04 0.01 0.15      0.69 0.04 0.11 0.00 0.02 0.00 0.0  0.04 0.01 0.07
PERCENT DIFF.            0.08 0.01-0.00-0.01 0.00-0.01-0.01 0.00-0.07 0.02      0.18-0.01 0.02-0.03-0.01-0.01-0.02 0.01-0.07-0.05
SPLIT CODES :            0 1 2 3 4 5 6******                                       7 8 9**************
THETA/DELTA = 0.111

UNEMPLOY RATE      8V414
VAR. NO. 414            14222 1099 2193  305  738   99   49  690  211 2581      1420  195  207    0  101   44   44  223    0  194
PERCENT                  0.64 0.05 0.10 0.01 0.03 0.00 0.00 0.03 0.01 0.12      0.58 0.08 0.09 0.0  0.04 0.02 0.02 0.09 0.0  0.08
PERCENT DIFF.            0.13-0.00 0.01-0.02-0.01-0.01-0.02-0.00-0.07-0.01      0.07 0.03-0.01-0.03 0.00 0.00-0.00 0.06-0.08-0.04
SPLIT CODES :            0 1 2 3************                                       4 5 6 7 8 9*********
THETA/DELTA = 0.040

ACTUAL-REQ'D ROOMS 8V381
VAR. NO. 381             3092  596  926  202  238    5    1  147   58 1273     12550  698 1474  103  601  138   92  766  153 1502
PERCENT                  0.47 0.C9 0.14 0.03 0.04 0.00 0.00 0.02 0.01 0.19      0.69 0.04 0.08 0.01 0.03 0.01 0.01 0.04 0.01 0.08
PERCENT DIFF.           -0.04 0.04 0.05 0.00-0.00-0.02-0.02-0.01-0.07 0.07      0.18-0.01-0.01-0.02-0.01-0.01-0.01 0.01-0.07-0.04
SPLIT CODES :            0 1 2 3 4**********                                       5 6 7 8 9**********
THETA/DELTA = 0.174

RACE 6:59
VAR. NO. 181              973  147  190    8    1    5    5   58   63  496     14669 1147 2210  297  838  138   88  855  148 2279
PERCENT                  0.50 0.08 0.10 0.00 0.00 0.00 0.00 0.03 0.03 0.25      0.65 0.05 0.10 0.01 0.04 0.01 0.00 0.04 0.01 0.10
PERCENT DIFF.           -0.01 0.02 0.01-0.03-0.04-0.02-0.02-0.00-0.05 0.13    . 0.13-0.00 0.01-0.02-0.00-0.01-0.02 0.01-0.07-0.02
SPLIT CODES :            2 7****************                                       0 1 3 4 5 6 8 9****
THETA/DELTA = 0.053

SEX-MARITAL STATUS 8V360
VAR. NO. 360            10516  845 1896    0  839   99   93  627  204 2100      5126  449  504  305    0   44    0  286    7  675
PERCENT                  0.61 0.05 0.11 0.0  0.05 0.01 0.01 0.04 0.01 0.12      0.69 0.06 0.07 0.04 0.0  0.01 0.0  0.04 0.00 0.09
PERCENT DIFF.            0.10-0.00 0.02-0.03 0.01-0.01-0.01 0.01-0.07-0.00      0.18 0.01-0.02 0.01-0.04-0.01-0.02 0.01-0.08-0.03
SPLIT CODES :            3******************                                       0 1 2 4 5 6 7 8 9**
THETA/DELTA = 0.102

******************************
BEST PREDICTOR IS:
ACTUAL-REQ'D ROOMS 8V381
VAR. NO.   381
THETA/DELTA = 0.174
******************************
```

```
***TRIAL UNIVARIATE DISTRIBUTIONS FOR ITERATION NO. 4  SPLIT NO. 8

                         *** CHILD NO. 30                                              *** CHILD NO. 31

VARIABLE NO.     CODES :    0    1    2    3    4    5    6    7    8    9        0    1    2    3    4    5    6    7    8    9



BKT AGE OF HEAD 8V368
VAR. NO. 368            10974  208  288   88  872   44  352  330   50 1078       5455  156  226    0  596    0  145  178    0  280
PERCENT                  0.77 0.01 0.02 0.01 0.06 0.00 0.02 0.02 0.00 0.08       0.78 0.02 0.03 0.0  0.08 0.0  0.02 0.03 0.0  0.04
PERCENT DIFF.            0.26-0.04-0.07-0.02 0.02-0.01 0.00-0.01-0.08-0.05       0.26-0.03-0.06-0.03 0.05-0.02 0.00-0.01-0.08-0.08
SPLIT CODES :            0 1 2 3 4 5 6******                                        7 8 9**************
THETA/DELTA = 0.059

BKT MONEY Y/NEEDS 8V407
VAR. NO. 407            11170  364  470   88 1120    0  382  415   50 1196       5259    0   44    0  348   44  115   93    0  162
PERCENT                  0.73 0.02 0.03 0.01 0.07 0.0  0.03 0.03 0.00 0.08       0.87 0.0  0.01 0.0  0.06 0.01 0.02 0.02 0.0  0.03
PERCENT DIFF.            0.22-0.03-0.06-0.02 0.03-0.02 0.01-0.00-0.08-0.05       0.35-0.05-0.09-0.03 0.02-0.01-0.00-0.02-0.08-0.10
SPLIT CODES :            0 1 2 3 4**********                                        5 6 7 8 9**********
THETA/DELTA = 0.146

UNEMPLOY RATE      8V414
VAR. NO. 414            10631  203  276   88  991   44  247  196   50  603       5798  161  238    0  477    0  250  312    0  755
PERCENT                  0.80 0.02 0.02 0.01 0.07 0.00 0.02 0.01 0.00 0.05       0.73 0.02 0.03 0.0  0.06 0.0  0.03 0.04 0.0  0.09
PERCENT DIFF.            0.29-0.04-0.07-0.02 0.04-0.01-0.00-0.02-0.08-0.08       0.21-0.03-0.06-0.03 0.02-0.02 0.01 0.01-0.08-0.03
SPLIT CODES :            0 1 2**************                                        3 4 5 6 7 8 9******
THETA/DELTA = 0.119

ACTUAL-REQ'D ROOMS 8V381
VAR. NO. 381             6788  271  377    0  792    0  193  230   50  964       9641   93  137   88  676   44  304  278    0  394
PERCENT                  0.70 0.03 0.04 0.0  0.08 0.0  0.02 0.02 0.01 0.10       0.83 0.01 0.01 0.01 0.06 0.00 0.03 0.02 0.0  0.03
PERCENT DIFF.            0.19-0.02-0.05-0.03 0.04-0.02 0.00-0.01-0.08-0.02       0.31-0.04-0.08-0.02 0.02-0.01 0.01-0.01-0.08-0.09
SPLIT CODES :            0 1 2 3 4 5********                                        6 7 8 9************
THETA/DELTA = 0.178

RACE 6:59
VAR. NO. 181             1259  179  192    0  244    0  100  148   50  489      15170  185  322   88 1224   44  397  360    0  869
PERCENT                  0.47 0.07 0.07 0.0  0.09 0.0  0.04 0.06 0.02 0.18       0.81 0.01 0.02 0.00 0.07 0.00 0.02 0.02 0.0  0.05
PERCENT DIFF.           -0.04 0.02-0.02-0.03 0.05-0.02 0.02 0.02-0.06 0.06       0.30-0.04-0.08-0.03 0.03-0.02 0.00-0.01-0.08-0.08
SPLIT CODES :            2 3****************                                        0 1 4 5 6 7 8 9****
THETA/DELTA = 0.192

SEX-MARITAL STATUS 8V360
VAR. NO. 360             8545  189  380    0 1422   44  497  139   50  769       7884  175  134   88   46    0    0  369    0  589
PERCENT                  0.71 0.02 0.03 0.0  0.12 0.00 0.04 0.01 0.00 0.06       0.85 0.02 0.01 0.01 0.00 0.0  0.0  0.04 0.0  0.06
PERCENT DIFF.            0.20-0.04-0.06-0.03 0.08-0.01 0.02-0.02-0.08-0.06       0.34-0.03-0.08-0.02-0.03-0.02-0.02 0.01-0.08-0.06
SPLIT CODES :            3*****************                                         0 1 2 4 5 6 7 8 9**
THETA/DELTA = 0.224

******************************
BEST PREDICTOR IS:
SEX-MARITAL STATUS 8V360
VAR. NO.   360
THETA/DELTA = 0.224
******************************
```

*** UNWEIGHTED FREQUENCIES ***

CHILD NO.	UNWEIGHTED N
1	4840
2	1319
3	3521
4	898
5	421
6	2359
7	1162
8	249
9	649
10	328
11	93
12	1210
13	1149
14	713
15	449
16	155
17	94
18	361
19	288
20	109
21	219
22	18
23	75
24	719
25	491
26	642
27	507
28	268
29	445
30	254
31	195

WITH WEIGHTED DATA, YOU NEED THE UNWEIGHTED SUBGROUP SIZES TO ASSESS THE STABILITY (RELIABILITY) OF THE RESULTS

*****RESULTS BASED ON WEIGHTED DATA*****

WEIGHT VARIABLE NUMBER = 8
VARIABLE NAME : WEIGHT 1V521

```
**** ITERATION NUMBER = 1  ****

          SPLIT NUMBER = 1        VARIABLE NUMBER =     2        VARIABLE NAME :BKT AGE OF HEAD 8V368

                     LEFT SIDE RESULTS                                RIGHT SIDE RESULTS     THIS RECAPITULATES THE SPLITTING
                                                                                             SEQUENCE
          CODES       :   0 1 2**************                         3 4 5 6 7 8 9******
          FREQUENCY   :   55536                                        114809
          MODAL FREQ.:    16379                                         61948
          MODAL LOC. :       0                                          0
          THETA       :  0.295                                        0.540
          *** CHILD NO.  2                                            *** CHILD NO.   3      THIS GIVES THE POWER OF THE SPLIT
                                                                                             USING THE DISTANCE MEASURE
          THETA/DELTA FOR THIS SPLIT =  0.168

    ****************************************************************************************************************
    OVERALL THETA AT COMPLETION OF THIS ITERATION = 0.513                                    THIS GIVES THE POWER OF MODAL
                                                                                             PREDICTION WITH 2 GROUPS, NO
                                                                                             BETTER THAN ONE SINCE THEY HAVE
                                                                                             THE SAME MODE
```

```
**** ITERATION NUMBER = 2  ****

          SPLIT NUMBER = 1        VARIABLE NUMBER =     7        VARIABLE NAME :SEX-MARITAL STATUS 8V360

                          LEFT SIDE RESULTS                                    RIGHT SIDE RESULTS

          CODES       :   3******************                           0 1 2 4 5 6 7 8 9**
          FREQUENCY   :   28648                                            9375
          MODAL FREQ.:    12266                                            4113
          MODAL LOC. :      0                                             0
          THETA       :  0.428                                          0.439
          *** CHILD NO.  4                                              *** CHILD NO.  5

          THETA/DELTA FOR THIS SPLIT =  0.217


          SPLIT NUMBER = 2        VARIABLE NUMBER =     2        VARIABLE NAME :BKT AGE OF HEAD 8V368

                          LEFT SIDE RESULTS                                    RIGHT SIDE RESULTS

          CODES       :   0 1 2 3 4**********                           5 6 7 8 9**********
          FREQUENCY   :   68874                                           45935
          MODAL FREQ.:    29877                                           32071
          MODAL LOC. :      0                                             0
          THETA       :  0.434                                          0.698
          *** CHILD NO.  6                                              *** CHILD NO.  7

          THETA/DELTA FOR THIS SPLIT =  0.140

    *************************************************************************************************************
    OVERALL THETA AT COMPLETION OF THIS ITERATION = 0.513
```

NOTE THAT SINCE THE FOUR GROUPS HAVE THE SAME MODE, NO GAIN IN THETA (MODAL PREDICTION) POWER OCCURS

```
SPLIT NUMBER = 1        VARIABLE NUMBER =    2        VARIABLE NAME :BKT AGE OF HEAD 8V368

                 LEFT SIDE RESULTS                             RIGHT SIDE RESULTS

CODES       :   0 1****************                      2 3 4 5 6 7 8 9****
FREQUENCY  :    7985                                       20663
MODAL FREQ.:    3975                                       10061
MODAL LOC. :     8                                         0
THETA       :  0.498                                     0.487
*** CHILD NO.  8                                         *** CHILD NO.  9

THETA/DELTA FOR THIS SPLIT =  0.152

SPLIT NUMBER = 2        VARIABLE NUMBER =    7        VARIABLE NAME :SEX-MARITAL STATUS 8V360

                 LEFT SIDE RESULTS                             RIGHT SIDE RESULTS

CODES       :   2*****************                       0 1 3 4 5 6 7 8 9**
FREQUENCY  :    6317                                        3058
MODAL FREQ.:    2723                                        1390
MODAL LOC. :     0                                         0
THETA       :  0.431                                     0.455
*** CHILD NO. 10                                         *** CHILD NO. 11

THETA/DELTA FOR THIS SPLIT =  0.268

SPLIT NUMBER = 3        VARIABLE NUMBER =    2        VARIABLE NAME :BKT AGE OF HEAD 8V368

                 LEFT SIDE RESULTS                             RIGHT SIDE RESULTS

CODES       :   0 1 2 3************                      4 5 6 7 8 9********
FREQUENCY  :   35202                                       33672
MODAL FREQ.:   17543                                       12334
MODAL LOC. :     0                                         0
THETA       :  0.498                                     0.366
*** CHILD NO. 12                                         *** CHILD NO. 13

THETA/DELTA FOR THIS SPLIT =  0.128

SPLIT NUMBER = 4        VARIABLE NUMBER =    2        VARIABLE NAME :BKT AGE OF HEAD 8V368

                 LEFT SIDE RESULTS                             RIGHT SIDE RESULTS

CODES       :   0 1 2 3 4 5********                      6 7 8 9***********
FREQUENCY  :   24615                                       21320
MODAL FREQ.:   15642                                       16429
MODAL LOC. :     0                                         0
THETA       :  0.635                                     0.771
*** CHILD NO. 14                                         *** CHILD NO. 15

THETA/DELTA FOR THIS SPLIT =  0.190

***********************************************************************************************************
OVERALL THETA AT COMPLETION OF THIS ITERATION = 0.524
```

NOTE THAT WITH MODAL PREDICTION THERE IS ALMOST NO GAIN OVERALL SINCE ALMOST ALL GROUPS HAVE THE SAME MODE

```
**** ITERATION NUMBER = 4  ****

     SPLIT NUMBER = 1        VARIABLE NUMBER =    4        VARIABLE NAME :UNEMPLOY RATE      8V414

                   LEFT SIDE RESULTS                                    RIGHT SIDE RESULTS

     CODES       :   0 1 2**************                          3 4 5 6 7 8 9******
     FREQUENCY   :    5116                                           2869
     MODAL FREQ.:     2802                                           1173
     MODAL LOC. :       8                                              8
     THETA       :  0.548                                          0.409
     *** CHILD NO. 16                                              *** CHILD NO. 17

     THETA/DELTA FOR THIS SPLIT =  0.138


     SPLIT NUMBER = 2        VARIABLE NUMBER =    3        VARIABLE NAME :BKT MONEY Y/NEEDS 8V407

                   LEFT SIDE RESULTS                                    RIGHT SIDE RESULTS

     CODES       :   0 1 2 3 4**********                          5 6 7 8 9**********
     FREQUENCY   :    7591                                          13072
     MODAL FREQ.:     3462                                           6599
     MODAL LOC. :       0                                              0
     THETA       :  0.456                                          0.505
     *** CHILD NO. 18                                              *** CHILD NO. 19

     THETA/DELTA FOR THIS SPLIT =  0.106


     SPLIT NUMBER = 3        VARIABLE NUMBER =    2        VARIABLE NAME :BKT AGE OF HEAD 8V368

                   LEFT SIDE RESULTS                                    RIGHT SIDE RESULTS

     CODES       :   0 1****************                          2 3 4 5 6 7 8 9****
     FREQUENCY   :    2613                                           3704
     MODAL FREQ.:     1369                                           1694
     MODAL LOC. :       3                                              0
     THETA       :  0.524                                          0.457
     *** CHILD NO. 20                                              *** CHILD NO. 21

     THETA/DELTA FOR THIS SPLIT =  0.184


     SPLIT NUMBER = 4        VARIABLE NUMBER =    3        VARIABLE NAME :BKT MONEY Y/NEEDS 8V407

                   LEFT SIDE RESULTS                                    RIGHT SIDE RESULTS

     CODES       :   0 1****************                          2 3 4 5 6 7 8 9****
     FREQUENCY   :     539                                           2519
     MODAL FREQ.:      393                                           1026
     MODAL LOC. :       0                                              5
     THETA       :  0.729                                          0.407
     *** CHILD NO. 22                                              *** CHILD NO. 23

     THETA/DELTA FOR THIS SPLIT =  0.189
```

```
                  LEFT SIDE RESULTS                                          RIGHT SIDE RESULTS

CODES       :   0 1 2 3 4***********                                      5 6 7 8 9**********
FREQUENCY   :   15725                                                       19477
MODAL FREQ.:     6595                                                       10948
MODAL LOC. :       0                                                        0
THETA       :  0.419                                                      0.562
*** CHILD NO. 24                                                          *** CHILD NO. 25

THETA/DELTA FOR THIS SPLIT =  0.114


SPLIT NUMBER = 6          VARIABLE NUMBER =     5         VARIABLE NAME :ACTUAL-REQ'D ROOMS 8V381

                  LEFT SIDE RESULTS                                          RIGHT SIDE RESULTS

CODES       :   0 1 2 3 4***********                                      5 6 7 8 9**********
FREQUENCY   :   13507                                                       20165
MODAL FREQ.:     3549                                                        9025
MODAL LOC. :       9                                                        0
THETA       :  0.263                                                      0.448
*** CHILD NO. 26                                                          *** CHILD NO. 27

THETA/DELTA FOR THIS SPLIT =  0.138


SPLIT NUMBER = 7          VARIABLE NUMBER =     5         VARIABLE NAME :ACTUAL-REQ'D ROOMS 8V381

                  LEFT SIDE RESULTS                                          RIGHT SIDE RESULTS

CODES       :   0 1 2 3 4***********                                      5 6 7 8 9**********
FREQUENCY   :    6538                                                       18077
MODAL FREQ.:     3092                                                       12550
MODAL LOC. :       0                                                        0
THETA       :  0.473                                                      0.694
*** CHILD NO. 28                                                          *** CHILD NO. 29

THETA/DELTA FOR THIS SPLIT =  0.174


SPLIT NUMBER = 8          VARIABLE NUMBER =     7         VARIABLE NAME :SEX-MARITAL STATUS 8V360

                  LEFT SIDE RESULTS                                          RIGHT SIDE RESULTS

CODES       :   3*******************                                      0 1 2 4 5 6 7 8 9**
FREQUENCY   :   12035                                                         9285
MODAL FREQ.:     8545                                                         7884
MODAL LOC. :       0                                                        0
THETA       :  0.710                                                      0.849
*** CHILD NO. 30                                                          *** CHILD NO. 31

THETA/DELTA FOR THIS SPLIT =  0.224

**************************************************************************************************************************
OVERALL THETA AT COMPLETION OF THIS ITERATION = 0.528


*****TIME  4 18 50 53  NORMAL TERMINATION OF THAID
```

```
*****TIME IS   4:18:52
 NO MORE RUN CARDS IN SETUP. STEP TERMINATED
```

APPENDIX D: OBTAINING THE PROGRAM AND ADAPTING IT TO OTHER SYSTEMS

OBTAINING PROGRAM

Information on obtaining THAID, as part of the OSIRIS/40 package, may be gotten from the Institute for Social Research program librarian by writing to:

Program Librarian
Insitute for Social Research
Room 106
University of Michigan
Ann Arbor, Michigan 48104

ADAPTATION

THAID is written to run on an IBM 360/40 under HASP/MFT with a Fortran G level compiler, IBM basic ASSEMBLER and a 104K byte partition. If the adaptation is to another IBM machine with the same language software, at least 104K bytes of available core and sufficient peripheral devices (tape drive and 2314 disk unit or equivalent); no problems should be encountered if the instructions given with the package are followed carefully. Listed below are common adaptation problems and solutions. If the user's adaptation does not fall into one or more of these categories, it is suggested that a systems programmer be consulted.

LESS THAN 104K

The simplest change is to reduce the dimensionality of the array DATA (12000) which is declared INTEGER*2. If DATA is made to have length L < 12000 then the tests for in-core/disk mode choice (Chapter III, Design Characteristics) become

$m*n > L$ (disk required for data)

$n \quad > 10{,}000$ (disk required for pointers)

where

$m \equiv$ number of predictors + 1 (plus 1 if weighted).

$n \equiv$ number of cases.

Thus more frequent use of disk will be made increasing I/O time.

FORTRAN E or EARLIER COMPILER

The basic problem here is non-recognition of the FORTRAN G LEVEL INTEGER*2 type statements, which may be converted to INTEGER*4. In addition, in the OSIRIS s routines the GETDIC argument LIST must be changed to INTEGER*4 and GETDIC reprogrammed. All "logical ifs" can be easily converted to their "arithmetic if" counterparts.

NON-IBM MACHINE

The major problem here is that almost without exception the Assembler coded subroutines GPIN and DIRECT cannot be assembled. This would necessitate either rewriting these subroutines in the available Assembler language or reprogramming the I/O and data manipulation portions of THAID.

THAID: A SEQUENTIAL ANALYSIS PROGRAM FOR THE ANALYSIS OF NOMINAL SCALE DEPENDENT VARIABLES
by James N. Morgan and Robert C. Messenger; 1973

a computer program using the same sequential searching algorithm as the AID program in Searching for Structure, but with a categorical dependent classification; the program searches for subgroups that differ maximally as to their distributions; it assumes neither additivity nor linearity, so requires substantial samples of 1,000 or more; 92 pp.

LC 72-619720
ISBN 0-87944-136-4 *paperbound $4*
ISBN 0-87944-137-2 *clothbound $6*

MULTIVARIATE NOMINAL SCALE ANALYSIS
by Frank M. Andrews and Robert C. Messenger; 1973

a new multivariate computer technique for analyzing data with categorical independent and dependent variables, 135 pp.

LC 72-619721
ISBN 0-87944-134-8 *paperbound $4*
ISBN 0-87944-135-6 *clothbound $6*

MULTIVARIATE MODEL BUILDING:
The Validation of a Search Strategy
by John A. Sonquist; 1970, reprinted 1971

undertakes the validation of the Automation Interaction Detection (AID) technique; 264 pp.

LC 76-630039
ISBN 0-87944-081-3 *paperbound $5*
ISBN 0-87944-082-1 *clothbound $7*

SEARCHING FOR STRUCTURE
by John A. Sonquist, Elizabeth Lauh Baker, and James N. Morgan; 1971

presents an approach to analysis of substantial bodies of micro-data and documentation for the AID III computer program (supersedes The Detection of Interaction Effects); 287 pp.

LC 72-168543
ISBN 0-87944-109-7 *paperbound $3*
ISBN 0-87944-110-0 *clothbound $5*

MULTIPLE CLASSIFICATION ANALYSIS:
A Report on a Computer Program for Multiple Regression Using Categorical Predictors
by Frank M. Andrews, James N. Morgan, and John A. Sonquist; 1967, reprinted 1971

a technique for examining the interrelationship between several predictor variables and a dependent variable within the context of an additive model; 221 pp.

LC 67-64745
ISBN 0-87944-055-4 *paperbound $3*

INFERENCE FROM SURVEY SAMPLES:
An Empirical Investigation
by Martin R. Frankel; 1971, reprinted 1972

an investigation of the sampling errors of multivariate statistics, as well as simpler means and differences of means for complex samples; 173 pp.

LC 72-161550
ISBN 0-87944-012-0 *paperbound $3*
ISBN 0-87944-013-9 *clothbound $5*

SEPP: SAMPLING ERROR PROGRAM PACKAGE
by Leslie Kish, Martin R. Frankel, and Neal Van Eck; 1972

the user's manual for a set of three programs—Balanced Repeated Replications Package, Paired Selection Algorithm for Multiple Subclasses, and ABSERD—designed to compute sampling errors from complex (clustered and stratified) samples; includes continuous updates; 184 pp.

LC 72-87231
ISBN 0-87944-131-3 manual in 3-ring binder $25

MEASURES OF SOCIAL PSYCHOLOGICAL ATTITUDES
by John P. Robinson and Phillip R. Shaver; 1969, reprinted 1973

review and evaluation of the major empirical measures of social psychological attitudes: life satisfaction and happiness, self-esteem, alienation and anomia, authoritarianism, values, general attitudes toward people, religious attitudes, and methodological scales; 650 pp.

LC 79-627967
ISBN 0-87944-130-5 *paperbound $8*
ISBN 0-87944-069-4 *clothbound $12*

MEASURES OF POLITICAL ATTITUDES
by John P. Robinson, Jerrold G. Rusk, and Kendra B. Head; 1968, reprinted 1972

brings together the major measuring instruments of political attitudes from the first scientific survey work in the 1930s to the present; 710 pp.

LC 68-65537
ISBN 0-87944-132-1 *paperbound $8*
ISBN 0-87944-060-0 *clothbound $12*

ECONOMIC SURVEY METHODS
by John B. Lansing and James N. Morgan; 1971, reprinted 1973

deals with the survey tool and its uses, the design of surveys, sampling problems in economic surveys, methods of data collection, getting data ready for analysis, analysis, and the financing, organization, and utilization of survey research; 448 pp.

LC 71-633672
ISBN 0-87944-008-2 *paperbound $5*
ISBN 0-87944-009-0 *clothbound $10*

SURVEY RESEARCH CENTER • INSTITUTE FOR SOCIAL RESEARCH
THE UNIVERSITY OF MICHIGAN