DOCUMENT RESUME

ED 137 340 TM 006 141

AUTHOR Magidson, Jay
TITLE Partitioning a Control Group to Achieve Appropriate Comparison Subgroups for Assessing Program Impacts on Completers and Noncompleters.
PUB DATE [Apr 77]
NOTE 40p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Attrition (Research Studies); Career Education; *Control Groups; Discriminant Analysis; Experimental Groups; Mathematical Models; Nonparametric Statistics; Predictor Variables; Probability; Program Effectiveness; *Program Evaluation; *Research Problems; *Statistical Analysis

ABSTRACT

In evaluation research studies, it often occurs that several program participants (experimentals) drop out of the program prior to completion. Since noncompleters generally differ substantially from completers in many respects, a control group which originally was representative of the participant group will most likely not be representative of either the completers or noncompleters considered separately. This paper presents a maximum likelihood procedure for partitioning a control group in such a situation into separate comparison subgroups for assessing program impacts on completers and noncompleters. The approach was used in evaluating the Mountain Plains Career Education Program. (Author)

PARTITIONING A CONTROL GROUP TO ACHIEVE APPROPRIATE COMPARISON SUBGROUPS FOR ASSESSING PROGRAM IMPACTS ON COMPLETERS AND NONCOMPLETERS

by

Jay Magidson

Abt Associates Inc.

Paper presented at the annual meetings of the American Educational Research Association, April 1977.

Abstract

In evaluation research studies, it often occurs that several program participants (experimentals) drop out of the program prior to completion. Since noncompleters generally differ substantially from completers in many respects, a control group which originally was representative of the participant group will most likely not be representative of either the completers or noncompleters considered separately.

This paper presents a maximum likelihood procedure for partitioning a control group in such a situation into separate comparison subgroups for assessing program impacts on completers and noncompleters. The approach was used in evaluating the Mountain Plains Career Education Program.

1. Scope of the Problem

The problem of the nonequivalent control group is probably one of the most urgent but yet one of the most troublesome dilemmas facing evaluators of social programs today. Even in the case that participants are randomly assigned to the treatment and control conditions, differential rates of (and causes of) attrition will often result in the emergence of some important differences between the treated and untreated.

For example, the evaluation design for the Mountain Plains Career Education Program called for eligible participant families to be randomly assigned to the program or a control group. As it turned out, approximately 30% of the program participants failed to complete the program for a variety of reasons. As might be expected, the noncompleters differed substantially from completers in many important respects. As a result, while the control families were a good match to the participants prior to program entry, they could not reasonably be viewed as being representative of the 70% of the participants who completed the program.

It is well known (Campbell and Stanley, 1966) that if a control group differs from an experimental group on any of the relevant factors affecting the outcome measures of performance, the observed differences between the groups on these outcome measures (or lack of observed differences, for that matter) may in fact be artifacts of the pre-existing differences rather than representing the effect of the program. The problem of interpretation presented by this confounded setting quickly becomes devastating when we realize that there is often no way to be sure that a given method of dealing with the situation has accomplished the intended result. Lord (1967) has asserted:

> . . . no logical or statistical procedure can be counted upon to make proper allowances for uncontrolled pre-existing differences between groups.

Nonetheless, reality constraints force evaluators to explicitly deal with these recurring problems. Evaluators must make certain assumptions, choose one or more approaches and exclude others. Moreover, they often must do so within more or less rigid time constraints.

In the case of the Mountain Plains program, follow-up data was obtained not only for program completers and the control families, but also for the participant families who exited from the program prior to completion. This allowed us, the evaluators, to identify some major differences between the completers and noncompleters and to attempt to partition the control group into two comparison subgroups for assessing the effect of the program on completers and noncompleters separately.

This paper addresses the particular instance of the nonequivalent control group problem which we confronted in evaluating the Mountain Plains Career Education Program. We discuss the approach which we used and the rationale behind it.

In the next section we provide some descriptive and historical information on the Mountain Plains program. The extent of the problem posed by the substantial noncompletion rate is discussed in section 3. Section 4 describes our general approach to partitioning the control group and in section 5 we give our rationale for choosing the maximum likelihood logit method over the classical discriminant analysis approach to the problem. The results are summarized in section 6. We conclude with some suggestions and implications for future research on the nonequivalent control group situation.

(Readers interested in the results of the evaluation of the Mountain Plains program are referred to Bale and Molitor, 1970.)

2. The Mountain Plains Program

In July, 1974 Abt Associates, Inc. was awarded a 30-month contract by the National Institute of Education (Contract NIE-C-74-1047) to conduct a longitudinal follow-up study of participants and a control group in the Mountain Plains Career Education and Employment Program. The Mountain Plains Program is a residential, family-centered career education, employment and training program oriented to the needs of multi-problem disadvantaged families caught in the cycle of poverty. It offers an integrated program of services to the entire family based on its fundamental assumption that in order to break the cycle of poverty it is necessary to address problems and needs simultaneously at the family level. The program generally takes from 6 to 12 months to complete depending upon the particular type(s) of courses in which the family enrolls.

The experimental design, planned and implemented by the Research Services Division of the Mountain Plains Education and Economic Development Program, Inc., called for randomized assignment to the participant and control groups. Although it is believed that the randomized nature of the experiment was not strictly adhered to, the control families were in fact found to be very similar to the participant families in most respects (Bale and Park, 1976). Control families however were earning somewhat less money and consisted of approximately 10% more female headed families and unmarried parents.[1]

Since its conception in February of 1972, the Mountain Plains program has served over 1200 disadvantaged families from the rural areas of Idaho, Montana, Nebraska, North Dakota, South Dakota, and Wyoming, a sparsely populated region that encompasses nearly one-fifth of the continental United States. Eighty percent of the families served by the program are white and the majority of the nonwhite families are Indian. The 1200 families served represents over 4500 individuals.

The population for this paper consists of all 160 control families together with the 914 participant families who had exited the program by February, 1976, the cutoff date for purposes of the evaluation.

3. The Problem Posed by the Noncompleters

Of the 914 participant families who had exited the program as of the cutoff date, 287 or 31% left prior to completion. A preliminary discriminant analysis on 23 salient characteristics indicated that those families failing to complete the program differed significantly (at the .05 level) from the completers with respect to (X_1) Education, (X_2) Housing Status, (X_3) Number of Children, (X_4) Race, and (X_5) Income. Table 1 presents the standardized coefficients associated with these variables together with the average scores for the completers and noncompleters.

The completion rate was much higher for whites than nonwhites. The drop out rate was especially high for the 130 Indian families, 60% of them failing to complete the program compared to a noncompletion rate of only 27% for the white families. Noncompleters also tended to be families living with others prior to entering the program as opposed to renting or owning their own housing. The completers were generally more educated and had slightly more children that the noncompleters.

These results might lead one to speculate on possible reasons for resigning from the program. First, it is possible that the program was geared more to the higher educated families who had more employable skills

Table 1

A Comparison of the 754 Completer and 287 Noncompleter Families on Statistically Significant Discriminators and the Associated Standardized Beta Weights

Discriminators	BETA	Average Value: Completers	Average Value: Noncompleters
(X_1) Years of Education (Head of Household)	.50	10.9	10.1
(X_2) Housing Status[a]	.35	84%[c]	68%[c]
(X_3) Number of Children	.30	1.8	1.6
(X_4) Race[b] (Head of Household)	.27	85%[d]	67%[d]
(X_5) Gross Family Employment Income	.20	$3600	$2800

[a] Housing status was coded as (1) own or rent and (0) live with others

[b] Race was coded as (1) White and (0) Other

[c] This is the percentage of owners and renters

[d] This is the percentage of White families

at the time of program entry. Also, the presentation may not have communicated well with the Indian population. Regardless of the reasons for not remaining in the program it is reasonably clear that the completers as a group were substantially different with respect to certain potentially important criteria used to evaluate the program (post program employment income, occupational status, self concept, etc.)

If the control families were representative of the group of all participants, they would definitely not be representative of the two very different classes of participants, the completers and the noncompleters. Moreover, in assessing the impact of the program on the completers (those families receiving the full "treatment") ideally the completers should be compared with only that subgroup of control families who would have completed the program had they entered. But of course no control family had entered the program so that it was impossible to distinguish with certainty the potential completer from the potential noncompleter families.

Bale and Park (1976) questioned whether the control group was comparable to the participants. On the basis of comparisons across 58 characteristics they concluded that the controls were comparable to the total participant group but not to either the completers or noncompleters considered separately. This left us with the problem of finding two control groups--one comparable to program completers in terms of pre-program characteristics, the other comparable to the resignees.

4. The General Approach

We decided to partition the control families into two subgroups, the completer-controls and the resignee-controls. Since these subgroups would be used in separate comparisons there was no need to restrict the subgroups to be disjoint or nonoverlapping. (In this respect the present application differs from the typical application of discriminant analysis.)

It is the probability of completing the program that we were interested in estimating for each of the control families. We could then choose two cutoff probabilities. The control families whose estimated probability of completing the program was greater than the lower cutoff would be the completer-controls. Those families below the upper cutoff would comprise the resignee-controls. Those between the two cutoff probabilities would be included in both comparison subgroups.

The approach we decided upon was as follows:

1. Estimate the probability of completing the program for a family with given values on the discriminators $X = (X_1, X_2, X_3, X_4, X_5)$. Use only the data for the 914 participants to accomplish this. Formally, estimate the probability function f where

$$P(X) = f\ (X_1, X_2, X_3, X_4, X_5) \qquad (1)$$

The implicit assumption here is that families with the same values on these 5 variables have the same completion probabilities. In other words, it assumes that the probability of completing the program is determined (causally) by an exact function of the discriminators.

2. Assume that this model (estimated using only data from the participants) also holds for the control families. The implicit assumption here is that the control families are exchangeable with the participants with respect to a) the latent factors determining whether a family will complete the program and b) the relationship of these latent factors to the discriminators.

3. Rank the control families on the estimated completion probabilities from high to low.

4. Choose the two cutoff points discussed above so as to maximize the goodness of match between the completer-participants and completer-controls and between the resignee-participants and resignee-controls by minimizing the number of significant differences on pre-program characteristics. In the case that some significant differences are inevitable choose the cutoff points that yield significant differences for those same characteristics that were found to yield significant differences between all participants and all controls. Only those cutoff points utilizing at least half of the 160 control families would be considered in order to assure that the comparison group would be large enough for the evaluation of the programs to be considered reasonably reliable.

To implement this five-step approach we needed to:

- choose the functional form of the probability f

and
- choose the method of estimation.

5. Estimation of the Completion Probability

Classical discriminant analysis (Anderson, 1958) is based on the assumption that the joint distribution of the discriminators is multivariate normal with the same covariance matrix within the two groups (completers and resignees). If these conditions were met the completion probability would satisfy the multivariate logistic distribution (Truett, Cornfield and Kannel, 1967)

$$P(X) = 1 | \{1 + \exp(-g(X))\} \quad (2)$$

where $g(X) = \alpha+\beta_1X_1 +\beta_2X_2 +\beta_3X_3 +\beta_4X_4 +\beta_5X_5 \quad (3)$

is the linear discriminant function.

The probabilities could thus be estimated by substituting the estimated discriminant function in place of g(X) in equation 2. These estimates would be unconditional maximum likelihood estimates under the normality assumptions.

Alternatively, the probabilities could be estimated directly by the conditional maximum likelihood approach employed by logit analysis in which case the discriminant function estimate, if desired, can be obtained by computing the log odds, i.e. $g(X) = \ln P(X) | (1-P(X))$. Halperin, Blackwelder and Verter (1970) show that extreme biases can result such as estimating a probability of .9 when the true value is only .2 if the classical discriminant function estimation procedure is employed when one or more of the discriminators is nonnormal but dichotomous and equation (2) holds. Since equation (2) holds under a family of exponential distributions, they recommend the logit estimation procedure for cases involving qualitative/categorical variables.

However, there is no reason to suppose that g(X) will be linear. In general, interaction effects will be present and g(X) will not be linear in X. Without information regarding the joint distribution of the discriminators it was not clear how to formulate a functional form which included interaction terms.

In the context of all categorical variables, Goodman (1972) employs the nonparametric concept of a saturated model which includes all possible interaction terms in the form of a generalized logistic function. Maintaining equation (2), if all of the X's were dichotomous taking on the values 1 and -1, Goodman's representation could be expressed as follows (Magidson, 1976):

$$
\begin{aligned}
g(X) = {} & \text{constant} + \text{five main effects} + \text{10 bivariate interaction effects} \\
& + \text{10 trivariate interaction effects} \\
& + \text{5 4th-order interaction effects} \\
& + \text{1 5 way interaction effect}
\end{aligned}
$$

Regardless of the true distribution of the X's, the estimated probabilities using the saturated model (2), (4) and dichotomized X's will always equal the observed proportions. Various unsaturated models formed by omitting some of the interaction effects can be tested using a chi square test with degrees of freedom equal to the number of effects omitted. While the general approach can be used with any quantal response model, the logistic function equation (2) has computational advantages and can be derived using the classical ANOVA formulation for the effects as applied to the natural logs of the cell counts (Goodman, 1972; Bishop, Fienberg and Holland, 1975).

We decided to use equations (2), (4) as our model in order to gain the use of interaction terms. We dichotomized (X_1) Education, (X_3) Number of Children and (X_5) Income at the medians in an attempt to minimize the number of empty cells in the five way contingency table. The resulting discriminators were (E) Education (High, Low), (R) Race (White, Other), (H) Housing Status (Own/rent, Other), (C) Number of Children (2 or more, less than 2) and (I) Income (High, Low).

For ease in interpretation we included only main effects and first order interaction effects in our initial model. This also served to reduce the scope of our model selection problem. A chi square goodness of fit statistic was available for testing whether we needed to hypothesize higher order interactions. Estimates were calculated using Goodman's ECTA (Everyone's Contingency Table Analyzer) computer program which maximizes the likelihood function of the multinomial distribution.

Regarding the possible loss of information due to dichotomization, after finding an initial model using the dichotomies, one could always test for loss of information by expanding the dichotomies to trichotomies and using the chi square statistic to test if such expansion significantly improves discrimination. The variables can continue to be expanded in this manner so long as the number and pattern of empty cells does not cause the model to be underidentified. We felt that dichotomies would probably be sufficient for our purposes and the results could be summarized by the thirty-two classes of profiles formed from the five way classification scheme. This would provide us with a convenient method for presenting the results in tabular form which could be easily scanned to determine whether the results appeared reasonable.

6. Partitioning the Control Group

In Table 2, the 32 profiles of participant families are ranked from high to low on the estimated probability of completing the Mountain Plains Program. The higher educated White families who owned or rented their housing were estimated as being most likely to complete the program. A comparison of the estimated probabilities with the actual completion rates observed show a reasonably good match. Notice that the few large discrepancies which do appear occur for profiles containing fewer than three participant families.

Table 3 displays summary statistics for the 9 models estimated. These models were selected by the process of backwards elimination as discussed by Goodman (1971). For our purposes, model H_2, the model with the highest probability value was chosen. For a given degree of parsimony as measured by the degrees of freedom, the higher the probability value the closer the fit to the observed proportions.

A comparison of model H_2 with the main-effects-only model, model H_8, indicates that the interaction effects included in model H_2 significantly improve the fit to the observed proportions. The difference between the chi-square statistics associated with these models has a chi-square distribution with 5 degrees of freedom (the difference between the degrees of freedom associated with the two models) under the null hypothesis that the main-effects-only model is correct. The difference between the chi-square values

Table 2.

Probability of Participants Completing the Program for Each of 32 Profiles of Characteristics Based on Education and Race of Family Head, Housing Status, Number of Children and Gross Family Employment Income

	PROFILE OF CHARACTERISTICS						PROBABILITY OF COMPLETING	
	(E) EDUCATION (1 = high) (0 = low)	(R) RACE (1 = white) (0 = other)	(H) HOUSING (1=own/rent) (0 = other)	(C) # CHILDREN 1 = ≥ 2 0 = < 2	(I) INCOME (1 = high) (0 = low)	n	ESTIMATED	OBSERVED
1.	1	1	1	1	1	82	.89	.89
2.	1	1	1	1	0	54	.86	.85
3.	1	1	1	0	1	83	.86	.87
4.	1	0	1	1	0	10	.85	.90
5.	1	0	1	1	1	9	.82	.78
6.	1	1	1	0	0	69	.82	.87
7.	0	1	1	0	1	74	.78	.78
8.	1	1	0	1	1	9	.74	.89
9.	0	1	1	1	1	109	.73	.70
10.	1	1	0	0	1	11	.69	.82
11.	0	1	0	0	1	27	.69	.62
12.	1	1	0	1	0	6	.68	.67
13.	1	0	0	1	0	7	.67	.57
14.	1	0	1	0	0	11	.64	.64
15.	0	1	0	1	1	3	.63	1.00
16.	1	1	0	0	0	24	.62	.58
17.	1	0	0	1	1	1	.62	.00
18.	0	0	1	1	1	8	.61	.75
19.	0	1	1	0	0	55	.61	.58
20.	1	0	1	0	1	13	.59	.31
21.	0	1	1	1	0	72	.55	.56
22.	0	0	1	1	0	24	.53	.54
23.	0	0	0	1	1	7	.50	.43
24.	0	1	0	0	0	34	.49	.44
25.	0	0	1	0	1	16	.45	.62
26.	0	1	0	1	0	15	.43	.47
27.	0	0	0	1	0	11	.42	.45
28.	1	0	0	0	0	17	.39	.29
29.	0	0	1	0	0	25	.37	.36
30.	0	0	0	0	1	3	.35	.33
31.	1	0	0	0	1	2	.34	1.00
32.	0	0	0	0	0	23	[illegible]	.35

COMPLETERS

RESIGNEES

Table 3

The Main Effects and Interaction Effects Models and Associated Goodness of Fit Chi Square Values for Models Fit Using Data on 914 Known Completers and Resignees

Model	Explanatory Variables	Degrees of Freedom	Chi Square X_G^2	Chi Square X_L^2	Probability Value
H_0	Saturated Model	0	0	0	--
H_1	I,H,C,E,R,CR,IE,IR,HE, CE,IH	20	20.5	22.4	.32
H_2	I,H,C,E,R,CR,IE,IR,HE, CE	21	21.2	23.3	.33
H_3	I,H,C,E,R,CR,IE,IR,HE	22	23.5	25.5	.27
H_4	I,H,C,E,R,CR,IE,HE,CE	22	23.3	25.0	.30
H_5	I,H,C,E,R,CR,IE,HE	23	25.8	27.4	.24
H_6	I,H,C,E,R,CR,HE	24	28.7	29.6	.20
H_7	I,H,C,E,R,CR	25	30.6	31.6	.17
H_8	I,H,C,E,R	26	37.9	38.3	.06

is 15 (or 17 if the likelihood ratio chi-square X_L^2 is used instead of the Peason chi-square X_G^2) which is sufficiently large so as to reject the main-effects-only model in favor of the interaction effects model, H_2. (It was also found that the use of the main-effects-only model would result in substantially different control groups than those defined in Table 2. The definition of the control groups is discussed later in this section.)

Table 4 presents the estimated effects associated with the variables. The most important variables for predicting program completion are Education, Race and Housing Status. The relatively large negative interaction effect associated with Children and Race indicates that the effect of the number of children upon completing the program depends upon race. Specifically, since the magnitude of this interaction effect equals the magnitude of the main effect associated with the number of children, the model states that the number of children does not influence the decision of white families to complete or resign from the program, but it has a strong effect upon other families.

In order to see if the interaction effect was reflected in the data we cross classified completers and noncompleters by Race and Number of Children. Sure enough, there was a strong interaction effect. White families completed the program at the same rate (73%) regardless of whether they had few (0 or 1) or many (more than 2) children. Non-white families however, were much less likely to resign from the program if they had many children. While 58% of non-white families with 0 or 1 child left the program prior to completion, only 39% of non-white families with 2 or more children did so. These results are displayed in Table 5. Similar checks also supported the reasonability of the selected model.

The partition of Table 2 was then applied to the control population. All control families having an estimated probability of completing at .61 or higher comprise the completer-control group. This is designated in Table 2 by an upward arrow beginning at the estimated probability of .61. These control families consist of the 19 profiles most likely to complete the program as estimated by the maximum likelihood procedure described earlier. An alternative cutoff probabilitiy of .62 was also considered but the resulting match was somewhat better for the .61 cutoff.

Table 4

The Main Effects and Interaction Effects Estimated by Model H_2, as Fit Using 914 Known Completers and Resignees

Variable	Main Effects*
Education (E)	+.202
Race (R)	+.187
Housing (H)	+.186
Children (C)	+.112
Income (I)	+.075

Variable	Interaction Effects
Children/Race (CR)	-.113
Housing/Education (HE)	+.070
Income/Education (IE)	-.065
Income/Race (IR)	+.062
Children/Education (CE)	+.062

*Positive values indicate propensity to complete program.

Table 5

Completers and Noncompleter Participants Classified by Race and Number of Children

	White: 0 or 1 child	White: 2 or more	Nonwhites: 0 or 1 child	Nonwhites: 2 or more
Completers	277 (73%)	257 (73%)	46 (42%)	47 (61%)
Resignees	100 (27%)	93 (27%)	64 (58%)	30 (39%)
	377 (100%)	350 (100%)	110 (100%)	77 (100%)

A similar procedure was used to define the resignee-controls. This is designated by the downward pointing arrow below .85. Other probability cutoffs considered for the resignee-controls which led to some worse matches were .80, .73 and .70.

The goal to be accomplished by the partitioning of the control group was to arrive at comparison groups which are comparable (exchangeable) with the associated participant groups (completers and noncompleters) with respect to causal factors determining the outcome measures upon which the program would be evaluated. Although these factors are unknown and/or unmeasured one might make inferences about these comparisons from the goodness of match between the groups on pre-program characteristics. The better the match on pre-program characteristics the more likely that the comparison groups are similar on related unmeasured variables which effect the various outcome measures.

Table 6-1 through 6-4 display the pre-program comparisons for all participants, as well as separate comparisons for completers and resignees with their corresponding control groups. The probability values associated with the completers and with the resignees are mostly substantial in size indicating quite good matches given by the control subgroups as defined in this section. The match for the completers is somewhat better than that for the resignees. By selecting appropriate subsets of the control families by the methods described earlier we have obtained what appears to be reasonably good comparison groups for the completers and resignees.

Let us now examine a few of these comparisons in detail. Notice first that there are good matches (nonsignificant differences) between the completer-controls and resignee-participants with respect to the five variables used in defining these control subgroups (i.e., Education and Race of Head, Housing Status, Number of Children and Total Family Income). For each of these five variables the use of the control subgroups is a substantial improvement over the use of all control families as a comparison group for both completers and resignees.

For example, consider "Race." While 85% of the completers are white and 67% of the resignees are white, the total control population consists of 81% white families. By restricting the set of control families to the completer-control subset and the resignee-control subset of families, this latter percentage becomes 91% and 74% white families respectively, which

Table 6-1

Demographic Characteristics: Pre-Center Comparisons for the Actual Population by All Participants/All Controls, Completer-Participants/Completer-Controls, Resignee-Participants/Resignee Controls

VARIABLE DESCRIPTION	ALL PARTICIPANTS n	ALL PARTICIPANTS %	ALL CONTROLS n	ALL CONTROLS %	χ^2 (signif)	COMPLETER PARTICIPANTS n	COMPLETER PARTICIPANTS %	COMPLETER CONTROLS n	COMPLETER CONTROLS %	χ^2 (signif)	RESIGNEE PARTICIPANTS n	RESIGNEE PARTICIPANTS %	RESIGNEE CONTROLS n	RESIGNEE CONTROLS %	χ^2 (signif)
SEX, HEAD OF FAMILY (P002)															
Male	732	80.1	114	71.2		494	78.8	82	69.5		238	82.9	85	74.6	
Female	182	19.9	46	28.7	.0157*	133	21.2	36	30.5	.0364*	49	17.1	29	25.4	.0769
SEX, SPOUSE (P013)															
Male	4	0.6	0	0		4	0.8	0	0		0	0	0	0	
Female	722	99.4	111	100.0	.9641	487	99.2	79	100.0	.9371	235	100.0	84	100.0	NA
RACE OF HEAD OF FAMILY (P012)															
White	727	79.5	130	81.3		534	85.2	108	91.5		193	67.2	84	73.7	
Indian	130	14.2	26	16.2		56	8.9	8	6.8		74	25.8	26	22.8	
Other	57	6.2	4	2.5	.1509	37	5.9	2	1.7	.1132	20	7.0	4	3.5	.2895
RACE OF SPOUSE (P018)															
White	582	80.2	83	74.8		450	85.5	69	87.3		162	68.9	59	70.2	
Indian	97	13.4	23	20.7		39	7.9	8	10.1		58	24.7	21	25.0	
Other	47	6.5	5	4.5	.1007	32	6.5	2	2.5	.3277	15	6.4	4	4.8	.8647
MARITAL STATUS (P006)															
Married	731	80.0	111	69.4		495	78.9	79	66.9		236	82.2	84	73.7	
Not Married	183	20.0	49	30.6	.0037*	132	21.1	39	33.1	.0065*	51	17.8	30	26.3	.0743
STATE OF ORIGIN (P010)															
Idaho	157	17.2	24	15.0		107	17.1	14	11.9		50	17.4	22	19.3	
Montana	152	16.6	27	16.9		102	16.3	20	16.9		50	17.4	20	17.5	
Nebraska	150	16.4	27	16.9		104	16.6	23	19.5		46	16.0	18	15.8	
North Dakota	137	15.0	29	18.1		109	17.4	26	22.0		28	9.8	15	3.2	
South Dakota	143	15.6	25	15.6		95	15.2	13	11.0		48	16.7	21	18.4	
Wyoming	175	19.1	28	17.5	.9168	110	17.5	22	18.6	.4670	65	22.6	18	15.8	.6946
HOUSING STATUS (P211)															
Own	78	8.5	21	13.1		60	9.6	13	11.0		18	6.3	14	12.3	
Rent	644	70.5	108	67.5		467	74.5	92	78.0		177	61.7	59	60.5	
Live With Others	192	21.0	31	19.4	.1774	100	15.9	13	11.0	.3742	92	32.1	31	27.2	.1143
FLUSH TOILET (P217)															
Yes	870	95.2	155	96.9		606	96.7	117	99.2		264	92.0	109	95.6	
No	44	4.8	5	3.1	.4598	21	3.3	1	0.8	.2394	23	8.0	5	4.4	.2852
TELEPHONE (P218)															
Yes	470	51.4	77	48.1		352	56.1	62	52.5		118	41.1	50	43.9	
No	444	48.6	83	51.9	.4940	275	43.9	56	47.5	.5348	169	58.9	64	56.1	.6963

Table 6-1 (continued)

Demographic Characteristics: Pre-Center Comparisons for the Actual Population by
All Participants/All Controls, Completer-Participants/Completer-Controls, Resignee-Participants/Resignee Controls

VARIABLE DESCRIPTION	ALL PARTICIPANTS		ALL CONTROLS		F RATIO (signif)	COMPLETER PARTICIPANTS		COMPLETER CONTROLS		F RATIO (signif)	RESIGNEE PARTICIPANTS		RESIGNEE CONTROLS		F RATIO (signif)
	MEAN	SD	MEAN	SD		MEAN	SD	MEAN	SD		MEAN	SD	MEAN	SD	
EDUCATION, HEAD OF FAMILY (highest grade completed) (P005)	10.7	1.7	10.8	1.6	.4947	10.9	1.6	11.2	1.5	.1369	10.1	1.8	10.2	1.6	.7291
EDUCATION, SPOUSE (highest grade completed) (P016)	10.7	1.8	10.5	1.5	.3219	10.9	1.7	10.7	1.5	.5345	10.4	1.8	10.4	1.5	.9714
AGE, HEAD OF FAMILY (HAGE)	26.6	7.0	27.2	7.0	.2912	27.1	7.0	27.6	7.4	.3247	25.5	6.9	26.7	7.1	.1182
AGE, SPOUSE (SAGE)	23.5	5.8	23.9	6.0	.5059	24.0	6.0	24.4	6.4	.6283	22.3	5.2	23.8	6.0	.0372*
NUMBER OF CHILDREN (NUMKIDS)	1.8	1.4	1.8	1.2	.6735	1.8	1.4	1.8	1.3	.8792	1.6	1.4	1.7	1.2	.5403
NUMBER IN HOUSEHOLD (P007)	3.6	1.4	3.5	1.3	.6472	3.6	1.5	3.5	1.4	.3937	3.5	1.4	3.4	1.3	.8718
NUMBER OF ROOMS (P216)	4.3	1.6	4.3	1.6	.7590	4.4	1.5	4.3	1.5	.5903	4.1	1.6	4.3	1.4	.2444
PEOPLE PER ROOM (PPROOM)	0.9	0.5	0.9	0.5	.6157	0.9	0.4	0.9	0.5	.9250	1.0	0.6	0.9	0.4	.0730
NUMBER OF CITIES LIVED IN IN LAST FIVE YEARS (P221)	2.3	1.3	2.1	1.1	.1178	2.3	1.3	2.1	1.1	.0630	2.3	1.3	2.2	1.1	.1789
NUMBER OF STATES LIVED IN IN LAST FIVE YEARS (P222)	1.6	0.8	1.6	0.7	.6572	1.6	0.8	1.6	0.7	.7712	1.7	0.9	1.6	0.7	.5023
GROSS RENT (rent plus utilities) (GROSRENT)	82.0	53.0	80.0	52.0	.6109	88.1	50.0	89.6	50.0	.7622	$70	$58	$73	$54	.6602
GROSS RENT AS % OF TOTAL INCOME (RENTOINC)	0.3	0.5	0.4	0.8	.0658	0.3	0.5	0.4	0.7	.2655	0.3	0.4	0.4	0.9	.0380*

Table 6-2

Employment Characteristics of Head and Spouse: Pre-Center Comparisons for the Actual Population by All Participants/All Controls, Completer-Participants/Completer-Controls, Resignee-Participants/Resignee-Controls

VARIABLE DESCRIPTION	ALL PARTICIPANTS		ALL CONTROLS		χ^2	COMPLETER PARTICIPANTS		COMPLETER CONTROLS		χ^2	RESIGNEE PARTICIPANTS		RESIGNEE CONTROLS		χ^2
	n	%	n	%	(signif)	n	%	n	%	(signif)	n	%	n	%	(signif)
HOH PRESENTLY EMPLOYED (P149)															
Yes	387	42.3	72	45.0		284	45.3	56	47.5		103	35.9	47	41.2	
No	527	57.7	88	55.0	.5889	343	54.7	62	52.5	.7399	184	64.1	67	58.8	.3776
HOH LOOKING FOR WORK (P150)															
Yes	452	49.5	73	45.6		284	45.3	53	44.9		168	58.5	58	50.9	
No	462	50.5	87	54.4	.4192	343	54.7	65	55.1	.9802	119	41.5	56	49.1	.1993
HOH TYPE OF INDUSTRY most recent job (P024)															
Agriculture, Forestry and Fisheries	86	9.4	13	8.1		66	10.5	10	8.5		20	7.0	10	8.8	
Mining	18	2.0	3	1.9		12	1.9	1	0.8		6	2.1	2	1.8	
Construction	176	19.3	31	19.4		113	18.0	22	18.6		63	22.0	22	19.3	
Manufacturing	160	17.5	26	16.2		106	16.9	17	14.4		54	18.8	22	19.3	
Transportation, Communications and Public Utilities	44	4.8	7	4.4		29	4.6	6	5.1		15	5.2	4	3.5	
Wholesale, Retail Trade	143	15.6	31	19.4		108	17.2	28	23.7		35	12.2	21	18.4	
Finance, Insurance and Real Estate	7	0.8	1	0.6		7	1.1	1	0.8		0	0	1	0.9	
Business and Repair Services	45	4.9	8	5.0		35	5.6	7	5.9		10	3.5	6	5.3	
Personal Services	33	3.6	7	4.4		21	3.3	2	1.7		12	4.2	6	5.3	
Entertainment, Recreation	7	0.8	1	0.6		3	0.5	0	0		4	1.4	1	0.9	
Professional Services	61	6.7	6	3.7		43	6.9	5	4.2		18	6.3	1	0.9	
Public Administration	37	4.0	8	5.0		21	3.3	5	4.2		16	5.6	6	5.3	
Military	7	0.8	1	0.6		4	0.6	1	0.8		3	1.0	0	0	
Other	17	1.9	1	0.6		9	1.4	1	0.8		8	2.8	0	0	
Student in School	8	0.9	1	0.6		5	0.8	1	0.8		3	1.0	1	0.9	
Not Employed in Last Year	65	7.1	15	9.4	.9730	45	7.2	11	9.3	.9406	20	7.0	11	9.6	.3080

Table 6-2 (continued)

Employment Characteristics of Head and Spouse: Pre-Center Comparisons for the Actual Population by All Participants/All Controls, Completer-Participants/Completer-Controls, Resignee-Participants/Resignee Controls

	ALL PARTICIPANTS		ALL CONTROLS		χ^2	COMPLETER PARTICIPANTS		COMPLETER CONTROLS		χ^2	RESIGNEE PARTICIPANTS		RESIGNEE CONTROLS		χ^2
VARIABLE DESCRIPTION	n	%	n	%	(signif)	n	%	n	%	(signif)	n	%	n	%	(signif)
SPOUSE PRESENTLY EMPLOYED (P152)															
Yes	149	20.5	17	15.3		112	22.8	12	15.2		37	15.7	11	13.1	
No	577	79.5	94	87.7		379	77.2	67	84.8		198	84.3	73	86.9	
No Spouse	237		49		.2496	175				.1686	82				.6854
SPOUSE PRESENTLY LOOKING FOR WORK (153)															
Yes	118	16.3	18	16.2		73	14.9	13	16.5		45	19.1	12	14.3	
No	608	83.7	93	83.8		418	85.1	66	83.5		190	80.9	72	85.7	.4050
No Spouse	237				.8980	175				.8441	82				
SPOUSE TYPE OF INDUSTRY most recent job (P104)															
Agriculture, Forestry and Fisheries	9	1.2	0	0		8	1.6	0	0		1	0.4	0	0	
Mining	3	0.4	0	0		3	0.6	0	0		0	0	0	0	
Construction	25	3.4	9	8.1		22	4.5	5	6.3		3	1.3	7	8.3	
Manufacturing	35	4.8	2	1.8		23	4.7	1	1.3		12	5.1	0	0	
Transportation, Communications and Public Utilities	5	0.7	1	0.9		3	0.6	1	1.3		2	0.9	22	26.2	
Wholesale, Retail Trade	176	24.2	32	28.8		113	23.0	27	34.2		63	26.8	0	0	
Finance, Insurance and Real Estate	5	0.7	0	0		1	0.2	0	0		4	1.7	0	0	
Business and Repair Services	7	1.0	0	0		6	1.2	0	0		1	0.4	3	3.6	
Personal Services	53	7.3	6	5.4		29	5.9	4	5.1		24	10.2	0	0	
Entertainment, Recreation	11	1.5	0	0		4	0.8	0	0		7	3.0	0	0	
Professional Services	58	8.0	3	2.7		41	8.4	3	3.8		17	7.2	2	2.4	
Public Administration	19	2.6	3	2.7		12	2.4	1	1.3		7	3.0	3	3.6	
Military	2	0.3	0	0		2	0.4	0	0		0	0	0	0	
Other	9	1.2	0	0		5	1.0	0	0		4	1.7	0	0	
Student in School	0	0	0	0		0	0	0	0		0	0	0	0	
Not Employed in Last Year	309	42.6	55	49.5	.1205	219	44.6	37	46.8	.5019	90	38.3	45	53.6	.0082*

Table 6-2 (continued)

Employment Characteristics of Head and Spouse: Pre-Center Comparisons for the Actual Population by All Participants/All Controls, Completer-Participants/Completer-Controls, Resignee-Participants/Resignee Controls

VARIABLE DESCRIPTION	ALL PARTICIPANTS		ALL CONTROLS			COMPLETER PARTICIPANTS		COMPLETER CONTROLS			RESIGNEE PARTICIPANTS		RESIGNEE CONTROLS		
	MEAN	SD	MEAN	SD	F RATIO (signif)	MEAN	SD	MEAN	SD	F RATIO (signif)	MEAN	SD	MEAN	SD	F RATIO (signif)
HEAD OF HOUSEHOLD															
NUMBER OF WEEKS UNEMPLOYED LAST YEAR (WKUNEMH)	20.1	16.9	22.1	17.0	.1836	18.3	16.9	18.8	16.6	.7456	24.2	16.2	24.0	16.4	.9061
NUMBER OF JOBS IN LAST YEAR (NJOBH)	2.1	1.3	2.1	1.3	.8631	2.1	1.3	2.1	1.4	.7416	2.2	1.3	2.1	1.3	.7057
EMPLOYMENT INCOME IN LAST YEAR (EINCH)	$3078	2016	$2746	1813	.0517*	$3301	$2113	$3070	$1895	.2669	$2588	$1961	$2518	$1710	.7096
AVERAGE DURATION IN DAYS OF JOBS HELD LAST YEAR (ADURH)	299	642.0	306	514.0	.9022	340.3	735.4	345.2	574.5	.9449	208.6	346.2	279.0	479.4	.1025
OCCUPATIONAL STATUS most recent job (NP030)	24.5	19.3	24.5	20.1	.9877	25.7	20.1	25.4	20.0	.8788	21.8	17.2	23.0	19.5	.5438
MONTHLY SALARY most recent job (P031)	$382	206	$334	186	.0066*	$ 389	$ 215	$ 344	$ 196	.0352*	$ 367	$ 187	$ 334	$ 184	.1187
SPOUSE															
NUMBER OF WEEKS UNEMPLOYED LAST YEAR (WKUNEMS)	39.6	16.9	43.3	14.3	.0315*	38.7	18.0	42.2	15.3	.0993	41.6	14.4	45.0	12.3	.0617
NUMBER OF JOBS IN LAST YEAR (NJOBS)	0.8	0.9	0.7	0.8	.1522	0.8	0.9	0.7	0.8	.7597	0.8	0.9	0.6	0.8	.0593
EMPLOYMENT INCOME IN LAST YEAR (EINCS)	$762	1143	$566	952	.0857	$ 807	$1210	$ 654	$1054	.2888	$ 669	$ 907	$ 475	$ 855	.1119
AVERAGE DURATION IN DAYS OF JOBS HELD LAST YEAR (ADURS)	136	421	98	304	.3530	121.5	276.9	116.4	353.3	.8838	166.7	622.4	57.6	102.9	.1114
OCCUPATIONAL STATUS most recent job (NP110)	16.7	19.3	13.0	16.0	.0506*	15.9	18.9	14.3	16.9	.4623	18.4	20.0	10.7	13.5	.0011*
MONTHLY SALARY most recent job (P111)	$159	155	$126	138	.0357*	$ 154	$ 156	$ 133	$ 140	.2768	$ 170	$ 153	$ 116	$ 140	.0048*

Table 6-3

Family Income Characteristics: Pre-Center Comparisons for the Actual Population by All Participants/All Controls, Completer-Participants/Completer-Controls, Resignee-Participants/Resignee Controls

VARIABLE DESCRIPTION	ALL PARTICIPANTS		ALL CONTROLS		F RATIO (signif)	COMPLETER PARTICIPANTS		COMPLETER CONTROLS		F RATIO (signif)	RESIGNEE PARTICIPANTS		RESIGNEE CONTROLS		F RATIO (signif)
	MEAN	SD	MEAN	SD		MEAN	SD	MEAN	SD		MEAN	SD	MEAN	SD	
TOTAL FAMILY INCOME FROM EMPLOYMENT IN LAST YEAR (FAMEINC)	$3377	2035	$2994	1879	.0266*	$3629	2089	$3367	1925	.2055	$2826	$1793	$2714	$1793	.5759
TOTAL FAMILY INCOME FROM WELFARE IN LAST YEAR (WELINC)	$ 441	846	$ 471	765	.6688	$ 477	903	$ 421	691	.4248	$ 361	$ 699	$ 473	$ 783	.1653
TOTAL FAMILY INCOME FROM "OTHER" SOURCES IN LAST YEAR (P163)	$ 186	691	$ 188	639	.9751	$ 187	720	$ 206	689	.7918	$ 183	$ 624	$ 130	$ 398	.3905
TOTAL FAMILY INCOME FROM ALL SOURCES IN LAST YEAR (TOTINC)	$4003	2006	$3652	1801	.0388*	$4293	2041	$3994	1768	.1363	$3370	$1774	$3316	$1748	.7830
TOTAL FAMILY INCOME FROM ALL SOURCES PER PERSON (INCPP)															
GROSS DEBTS PAYABLE IN NEXT YEAR (P166)	$ 615	713	$ 618	943	.9577	$ 689	775	$ 680	996	.9050	$ 452	$ 521	$ 546	$ 902	.1910
FAMILY EMPLOYMENT INCOME PER PERSON (EINCPP)	$1071	765	$ 919	644	.0181*	$ 1146	792	$1034	667	.1522	$ 908	$ 674	$ 845	$ 627	.3909

Table 6-4

Financial and Social Characteristics: Pre-Center Comparisons for the Actual Population by
All Participant 'All Controls, Completer-Participants/Completer-Controls, Resignee-Participants/Resignee Controls

VARIABLE DESCRIPTION	ALL PARTICIPANTS n	ALL PARTICIPANTS %	ALL CONTROLS n	ALL CONTROLS %	χ^2 (signif)	COMPLETER PARTICIPANTS n	COMPLETER PARTICIPANTS %	COMPLETER CONTROLS n	COMPLETER CONTROLS %	χ^2 (signif)	RESIGNEE PARTICIPANTS n	RESIGNEE PARTICIPANTS %	RESIGNEE CONTROLS n	RESIGNEE CONTROLS %	χ^2 (signif)
DOES FAMILY USE FOOD STAMPS (FOODSTAM)															
Yes	401	43.9	64	40.0		292	46.6	45	38.1		109	38.0	43	37.7	
No	513	56.1	96	60.0	.4090	335	53.4	73	61.9	.1123	178	62.0	71	62.3	.9476
DOES FAMILY USE A BUDGET (P173)															
Yes	410	44.9	59	36.9		302	48.2	51	43.2		108	37.6	33	28.9	
No	504	55.1	101	63.1	.0732	325	51.8	67	56.8	.3753	179	62.4	81	71.1	.1268
DOES FAMILY HAVE A SAVINGS ACCOUNT (P167)															
Yes	205	22.4	31	19.4		156	24.9	27	22.9		49	17.1	22	19.3	
No	709	77.6	129	80.6	.4490	471	75.1	91	77.1	.7292	238	82.9	92	80.7	.7028
DOES FAMILY SHARE RECREATION ACTIVITIES (P210)															
Yes	818	89.5	145	90.6		566	90.3	106	89.8		252	87.8	105	92.1	
No	96	10.5	15	9.4	.7705	61	9.7	12	10.2	.9832	35	12.2	9	7.9	.2865
IS THERE A FAMILY DOCTOR (P225)															
Yes	653	71.4	104	65.0		471	75.1	84	71.2		182	63.4	69	60.5	
No	261	28.6	56	35.0	.1200	156	24.9	34	28.8	.4330	105	36.6	45	39.5	.6710

Table 6-4 (continued)

Financial and Social Characteristics: Pre-Center Comparisons for the Actual Population by All Participants/All Controls, Completer-Participants/Completer-Controls, Resignee-Participants/Resignee Controls

VARIABLE DESCRIPTION	RESIGNEE PARTICIPANTS		RESIGNEE CONTROLS		F RATIO (signif)	COMPLETER PARTICIPANTS		COMPLETER CONTROLS		F RATIO (signif)	ALL PARTICIPANTS		ALL CONTROLS		F RATIO (signif)
	MEAN	SD	MEAN	SD		MEAN	SD	MEAN	SD		MEAN	SD	MEAN	SD	
AVERAGE NUMBER OF COMMUNITY ACTIVITIES ENGAGED IN BY HOH & SPOUSE COMBINED (SOCTOT)	.54	.87	.48	.84	.35	.53	.86	.62	.93	.29	.45	.79	.37	.68	.32
NUMBER OF COMMUNITY ACTIVITIES ENGAGED IN BY HOHs (SOCHOH)	.49	.88	.46	.86	.69	.51	.89	.57	.95	.52	.44	.82	.32	.70	.16
NUMBER OF COMMUNITY ACTIVITIES ENGAGED IN BY SPOUSE (SOCSPS)	.39	.73	.28	.70	.14	.37	.80	.45	.76	.40	.24	.61	.27	.65	.72

gives a good match to the completer and resignee participants. These results are similar for the other four variables as well.

Consider next "Number of weeks unemployed last year," one of the pre-program characteristics not used in defining to control subgroup. The participant-completers were unemployed an average of 18 weeks while the average resignee spent 24 weeks without a job. The average control family again falls somewhere in between, at 22 weeks. The completer-control and resignee-control subgroups yield figures of 19 weeks and 24 weeks respectively, again an excellent match to the completer and resignee partipants. The resulting goodness of match is similar in almost all other instances where the overall control group was fairly similar to the group of all participant families.

Finally, let us consider a variable such as "Sex, Head of Family," on which the controls differed significantly from the participants. While the controls consisted of 29% female headed families, females constituted the head of only 20% of the participant families. This 9% difference is significant at well beyond the .05 level. The quality of match given by the control subgroups is also a function of the quality of match between all participants and all controls as now will be seen.

The breakdown of the 20% female-headed participant families becomes 21% and 17% respectively for the completer-participants and resignee-participants. Although the control families have a larger proportion of female headed families than either of these participant groups, the control subgroups still perform in the desirable direction. Just as the completer-participants have a higher percentage of female headed families than the resignee-participants, the completer-control subgroup also has a larger percentage of female-headed families than does the resignee-control subgroup (30% compared to 25%). Thus, the significant difference between the completer-controls and completer-participants on this characteristic is in part attributable to the fact that control families consist of a higher proportion of female-headed families.

7. Summary and Implications for Future Applications

In this paper we applied Goodman's general nonparametric approach for the analysis of qualitative/categorical variables to classify the control popoulation into estimated completers and estimated resignees from the Mountain Plains Career Education Program. The method performed well as judged by similar comparisons between the completers and the estimated completer-controls and between the resignees and the estimated resignee-controls as compared to all the participants and all controls comparison.

Recently, a number of articles have compared the discriminant analysis estimater with that of logit analysis (Halperin, et. al. 1970, McFadden, 1976 and Efron, 1975).

Although it was shown by Halperin, et. al. that the discriminant analysis estimator can lead to severe biases when one of the discriminators is a dichotomy, all of these papers assume no interaction effects. In the case that interaction effects do exist, both the classical discriminant analysis estimator and the linear logit model are subject to major distortions depending upon the amount of interaction that exists. Generally, it might be expected that interaction effects are present and as a general rule one should test for them before postulating their nonexistence. In this paper we uncovered the existence of a rather strong race/number of children interaction effect.

One limitation of the research report here is that we did not compare alternative methods for partitioning the control group. Another was that we did not examine the effect of dichotomizing the variables. These omissions were due to time contraints. Finally, although our model was interpreted as determining the probability in a causal sense, we did not adequately deal with the concept of measurement error.

A more intuitively pleasing formulation might posit the probability of completing the program as a causal function of (F_1) Socio- Economic Status, (F_2) Receptivity to traditional teaching methods and (F_3) Propensity to remain in one place. All 5 of the observed discriminators can be assumed to be indicators of F_1. Race and Education can be assumed to be fallible measures of F_2 while F_3 might be measured by Housing Status and Number of Children. In Magidson (1977a), the formulation of models using unobserved constructs is recommended in dealing with the general nonequivalent control group problem. For an application of this approach see Magidson (1977b). With respect to latent structure models, see Lazarsfeld and Henry (1968) and Goodman (1974).

We suggest the need for research on nonlinear classification techniques.

FOOTNOTES

[1] Analyses were conducted separately by sex and marital status. Thus, married female participants were compared to married female controls and unmarried female participants were compared to female controls who were also unmarried. See Bale and Molitor (1976).

REFERENCES

Anderson, T. 1974. An Introduction to Multivariate Statistical Analysis. Wiley: New York.

Bale, R. and Park, D. 1976. An Initial Exploration of Analysis Procedures and Problems Related to the Evaluation of Mountain-Plains Program Impacts in the Family Domain, Report No. AAI-76-79, A Preliminary draft working paper, Abt Associates Inc., Cambridge, Mass.

Bale, R. and Molitor, J. 1976. Impacts of the Mountain-Plains Program in Five Domains of Quality of Life, Report No. AAI-76-216, A draft working paper, Abt Assoicates Inc., Cambridge, Mass.

Campbell, D. and Stanley, J. 1966. Experimental and Quasiexperimental Designs for Research, Chicago: Rand McNally.

Efron, B. 1975. "The Efficiency of Logistic Regression Compared to Normal Discriminant Analysis," Journal of the American Statistical Association, 70(352), pp. 892-898.

Goodman, L. 1971. "The Analysis of Multidimensional Contingency Tables: Stepwise Procedures and Direct Estimation Methods for Building Models for Multiple Classifications," Technometrics, 13, pp. 33-61.

________ 1972. "A General Model for the Analysis of Surveys," American Journal of Sociology, 77, pp. 1035-1086.

________ 1974. "Exploratory Latent Strucutre Analysis Using Both Identificable and Unidentifiable Moedls," Biometrika, 61, pp. 215-231.

Halperin, M. Blackwelder, W. and J. Verter, 1971. "Estimation of the Multivariate Logistic Risk Function: A Comparison of the Discriminant Function and Maximum Likelihood Approaches," Journal of Chronic Diseases, 24, pp. 125-158

Lazarsfeld, and Henry, 1968. Latent Structure Analysis, Houghton Miflin: Boston.

Magidson, J. 1976. The Multivariate Analysis of Qualitative Variance: Analyzing the Probability of an Event as a Function of Observable Variables, unpublished doctoral thesis, Northwestern University (June).

________ 1977a. Multivariate Methods for Dealing with Unobserved Variables in Educational Research, Unsolicited Research Proposal, number AAI-2885, submitted to the National Institute of Education. Abt Associates Inc., Cambridge, Mass.

________ 1977b. "Towards a Causal Model Approach for Assessing Program Impacts in the Nonequivalent Control Group Situation: A General Alternative to ANCOVA," to be presented at the American Sociological Association Meetings, Chicago (September).

McFadden, C. 1976. "A Comment on Discriminant Analysis 'Versus' Logit Analysis," Annals of Economic and Social Measurement, 5/4, pp. 511-523.

Truett, J. Cornfield, J. and Kannel, W. 1967. "A Multivariate Analysis of the Risks of Coronary Heart Disease in Framingham," Journal of Chronic Diseases, 20, pp. 511-524.