Comparison of Conjoint Analysis, Multiple Regression Models with Person Vectors and Profile Analysis to Assess Important Factors Used to Select Colleges

Isadore Newman
The University of Akron

John Fraas
Ashland University

The purpose of this study was to investigate the relative effectiveness of the traditional conjoint analysis approach to the multiple regression approach that includes person vectors profiles analysis. It was expected that the more sophisticated models would increase the effectiveness in terms of its shrinkage estimates and the accuracy of its predictability of two holdout groups. The data source consisted of a sample of 100 students who rated eight colleges on five attributes--quality of education, financial aid, quality of dorm life, student/faculty relations, and social aid.

Comparison of Conjoint Analysis, Multiple Regression
Models with Person Vectors and Profile Analysis
to Assess Important Factors Used to Select Colleges

Introduction

In recent years, many colleges and universities have faced
increased competition for students. Thus, it has been
increasingly important for an institution of higher education to
be able to identify what factors are important to the students
who chose to enroll in the institution.

Marketing research (Cattin & Wittink, 1982) has identified
conjoint analysis as a very useful statistical technique in which
one is interested in having the clients, students, or consumers
prioritize a variety of items. Two other approaches also seem to
be appropriate to use when attempting to assess the selection
process of college-bound students: (1) multiple regression
models with person vectors (Frass & Newman, 1989); and (2)
profile analysis.

Objectives

This paper attempted to compare the ability of conjoint
analysis, multiple regression models with person vectors, and
profile analysis to produce information that could be used by
college and university personnel to determine which factors were
important to students when selecting a university or what type of
students selected a given type of university.

Data Collection

The research instrument used to collect the data analyzed in this study focused on five institutional attributes reported to be of significance to students who martriculated to Ashland University. This list of attributes was developed through literature reviews (Tierney 1980; Traynor, 1981; Kuh, Coomers, & Lindquist, 1984; Conant, Brow, & Mokwa, 1985), discussion with program advisors and students, and from the past experiences of admissions recruiters.

The five attributes included in this study were financial aid, social life, quality of dorm life, student-faculty relationships, and quality of education. Each of the five attributes had two levels. The two levels that were formed for each attribute were assigned a value of 0 or 1 in order to allow the researchers to quantitatively form hypothetical universities with various combinations of attribute levels. The attributes, levels, and values assigned to each level were as follows:

1. Quality of education

    a) reputation is not well known - 0
    b) reputation is well known - 1

2. Student/Faculty relationships

    a) faculty are accessible if sought - 0
    b) faculty are extremely accessible - 1

3. Quality of dorm life

    a) below my expectations - 0
    b) above my expectatons - 1

4.   Financial aid

       a)   little financial need is met - 0
       b)   most financial need is met - 1

5.   Social Life

       a)   few social activities are available - 0
       b)   many social activities are available - 1

Five attributes with two levels each would allow 32 different university profiles to be formed. With the assumption that interaction effects are negligible, the main effects could be estimated with only eight orthogonal arrays. The eight orthogonal arrays used in this study which were formed with the aid of the computer software entitled Conjoint Designer (Bretton-Clark, 1987), were listed in Table 1.

In addition to the eight orthogonal arrays, two arrays were designed to provide a means of assessing the degree of predictive validity. (See Table 1.) These two arrays were referred to as the "holdout universities" because they were not included in the estimation procedures.

The questionnaire was administered during the second week of the fall term of 1987 to freshman students enrolled in a freshman seminar course. The responses of 100 of the students were used in this study. See Fraas and Paugh (1989) for

4

Conjoint Analysis

The analysis conducted by the use of a software package (Bretton-Clark, 1987) produces a set of five regression coefficients plus a constant term for each student. That is, a separate regression analysis was performed on the data of each of the 100 students.

Each of the regression coefficients generated by the conjoint analysis for a given student indicated what would happen to the respondent's ratings of the universities when the attribute changed from the "zero" level to the "one" level. To illustrate the point, consider the regression coefficient value of 2.0 recorded for the financial attribute for respondent 1. If financial aid was to increase from the "little need being met" category to the "most need being met" category, the respondent's ratings of the universities would increase by 2.0 points on the 1 to 10 scale used on the questionnaire.

A relative importance figure was calculated for each attribute by dividing the sum of the five average regression coefficients into each of the average regression values. The five relative importance figures generated by this procedure were expressed as percentages.

Table 1

Orthogonal Arrays Used for Conjoint Analysis and Multiple Linear Regression Models

Universities	Quality of Education	Student/ Faculty Relation- ships	Quality of Dorm Life	Financial Aid	Social Life
A	0	0	0	0	0
B	1	0	0	1	1
C	1	1	1	1	0
D	0	1	1	0	1
E	0	1	0	1	0
F	1	1	0	0	1
G	1	0	1	0	0
H	0	0	1	1	1
Holdout Universities					
I	1	1	1	0	1
J	1	1	0	1	0

Note. Each characteristic is composed of two levels. The zero value indicates the presence of the lower of the two levels.

Results of the Conjoint Analysis

The relative importance figures indicated that financial aid was the most important attribute with a value of 26.24%. Financial aid was followed in importance by the quality of dorm life (21.29%), the quality of education (20.84%), the student/faculty relationships (16.63%), and the social life (15%). (See Table 2.)

Predictive Validity

The observed and predicted ratings for the holdout universities were used to provide two estimates of the ability of the results of the conjoint analysis to predict student ratings. The first estimate was a correlation coefficient for the predicted and observed ratings. The second estimate was an average absolute difference value for the difference between the predicted and observed ratings. The correlation coefficient value and the average absolute difference for the observed and predicted ratings were .37 and 1.87, respectively.

Multiple Linear Regression Model
With a Surrogate Person Variable

Model Structure

The second approach used to analyze the survey information required the construction of a multiple linear regression model that included a surrogate person variable. Before such a model is presented, however, a discussion of a model that includes the actual person variables may prove helpful. The variables included in the model that used person variables (Model 1) were as follows:

Y = ratings of the eight hypothetical universities (values ranged from 1 to 10)

X1 = quality of education

0 = "low" level; 1 = "high" level)

X2 = student/faculty relationship

(0 = "low" level; 1 = "high" level)

X3 = quality of dorm life

(0 = "low" level; 1 = "high" level)

X4 = financial aid

(0 = "low" level; 1 = "high" level)

X5 = social life

(0 = "low" level; 1 = "high" level)

P1 = respondent 1

> (1 if from respondent 1; 0 otherwise)

P2 = respondent 2

> (1 if from respondent 2; 0 otherwise)

P99= respondent 99

> (1 if from respondent 99; 0 otherwise)

The structure of the regression model with person variables was:

$$Y = aU + b1X1 = b2X2 = b3X3 = bb4X4 = b5X5 = b6P1 = b7P2 = \ldots$$
$$b104P99 = e \qquad \text{(model 1)}$$

The use of the person variable required by Model 1 is not practical due to their large number. Thus a multiple linear regression model designed to include a surrogate person variable was used. This surrogate person variable measured the impact of the 99 person variables required by Model 1.[1]

Table 2

Conjoint Analysis Results

Characteristic	Average Regression Coefficient	% of Relative Importance
Financial Aid	1.775	26.24
Quality of Dorm Life	1.440	21.29
Quality of Education	1.410	20.84
Student/Faculty Relationships	1.125	16.63
Social Life	1.015	15.00

Correlation coefficient between the predicted and observed ratings of the holdout universities = .37

Average absolute difference between the predicted and observed ratings of the holdout universities = 1.87

The value of the surrogate person variables was composed of an average rating for each person. The surrogate variables was represented in Model 2 by "X6." The values for this variable ranged from 2.625 to 8.5 for the 100 students.

The multiple regression model with the surrogate person variable (Model 2) used to analyze the survey information was as follows:

$$Y = aU - b1X2 - b2X2 - b3X3 - b4X4 - b5X5 - b6X6 - e \quad \text{(Model 2)}$$

The regression coefficients for the university attributes that were generated by Model 2 were equal to the average regression coefficients for the conjoint analysis (See Table 3).

Before the regression coefficients could be statistically tested, the standard errors had to be corrected for the appropriate degrees of freedom. The number of degrees of freedom was 695, which was equal to the sample size of 100 (number of students) multiplied by 8 (number of colleges) minus 6 (number of attributes plus one). Each of the regression coefficients for the university attributes was statistically significant at the .01 level. The multiple correlation coefficient was .764; and the R^2 value was .58.

Predictive Validity

The regression coefficients generated by Model 2 were used to predict the ratings of the holdout universities. The .

correlation coefficient for the predicted and observed ratings was .76. The average absolute difference between the predicted The same procedure applied to the second half of the data set resulted in a correlation coefficient value of .74 between the observed and predicted ratings. Again, this value shows little shrinkage (1.7%) from the multiple correlation coefficient of 753 for Model 2.

---

1 Refer to Pedhazur (1977), Williams (1977; 1980), Fraas McDougall (1983), and Williams and Williams (1985a; 1985b) for discussions of a surrogate variable used to measure the amount of variation in the dependent variable associated with a set of person variables.

Comparison of the Results

The estimated impact of the university attributes on the student ratings by the conjoint analysis, and the multiple linear regression model with a surrogate person variable were identical. For both procedures, the order of importance was as follows: (1) financial aid, (2) quality of dorm life, (3) quality of education, (4) student/faculty relationships, and (5) quality of social life.

The multiple linear regression model with the surrogate person variable, however, produced a correlation coefficient value of .76 for the predicted and observed ratings of the holdout universities, as compared to the value of only .37 for the conjoint analysis.

The multiple linear regression model with the surrogate person variable also produced a lower average absolute difference between the predicted and observed ratings for the holdout universities than did the conjoint analysis. The average absolute difference values were 1.50 and 1.87.

The low R^2 values of the regression models that used the clusters as the independent variables indicated that the clusters were unable to explain the variation in the university ratings to any high degree. For this data set, the cluster information was of little assistance in identifying the importance of university characteristics as viewed by various groups of students.

Quannal Analysis

The following description of quannal analysis is heavily based on Vantubergen (1966) and Newman and Carolyn Benz (1988). The third data analysis procedure applied to the data set was quannal analysis. The purpose of using this procedure was to determine whether certain types of people could be identified that favored different types of schools.

The factor analysis computer program used in this study was QUANNAL Vantubergen, 1966). This program places squared multiple correlation values in the principle diagonal as commonality estimates and conducts a Q-analysis. This approach is appropriate for the purpose of differentiating between people int erms of the shape of their profiles.

Five steps are used in a Q factor analysis.

Step 1 - An intercorrelation matrix is formed by correlating every person's ratings of the items with every other person's rating of items.

Thus, the eight ratings for respondent 1 were correlated with the ratings of the other 99 respondents. The same procedure was followed for each respondent.

Step 2 - The matrix of intercorrelations if submitted to factor analysis so that "persons" are variables and

items are observations. A principal axis solution is obtained. This result is submitted to a varimax rotation which produces orthogonal factors. On this basis, a factor represents a grouping of persons around a common pattern of sorting the items. Hence, a factor represents a type of "person" (Vantubergen, 1966).

Sub. No.	Two Factor Solution			Sub. No.	Three Factor Solution			
	I	II	h		I	II	III	h^2
1.	.22	.83	.95	1.	.30	.87	.03	.85
2.	.92	.17	.88	2	.87	.16	.39	.93
3.	.98	-.13	.97	3.	.84	-.16	.50	.98
4.	.75	.49	.81	4.	.33	.37	.86	.99
5.	.82	.19	.71	5.	.37	.05	.90	.95
6.	-.06	.90	.82	6.	-.04	.91	.03	.83
7.	.86	.09	.76	7.	.97	.14	.14	.99
8.	.17	.92	.88	8.	-.02	.87	.39	.91
% Total Var.	48	34	82	% Total Var.	34	32	27	93

The factor analytic model constructs hypothetical types of "persons" based on the way the actual people interviewed rated the items. One can group people by assigning them to the type that they are most like, i.e., the factor on which they have the highest loading.

Step 3 - Each pattern of items associated with each
by weighting each item response of each item response
of each of the persons most highly associated with a
given factor by the degree to which they are loaded on
that factor, the greater is the weight. These weighted
responses are summed across each item separately. This
procedure produces an item array of weighted responses for
each factor in the rotated factor analysis solution
selected. The arrays of weighted repsonses are then
converted to z-scores (Vantubergen, 1966).

Hypothetical types constructed by the factor analytic model
is based on a weighted pattern of the items (hypothetical types).
The more a person's rating is like the hypothetical type, the
more weight it receives in the average. The specific weight
given is calculated as follows:

$$\text{weight} = \frac{r}{1 - r_2} \qquad \text{where: } r = \text{loading}$$

The weighted average is called an item factor array.

The persons used to estimate an array are highly associated
with that type, but they are not associated to a high degree with
any of the other types. For the persons selected, the square of
the loading on that factor should approach the communality h_2.
The arrays of weighted item ratings are converted to z scores.
The array of z scores for each type is called the factor array.

16

Step 4 - The arrays of item z - scores for each factor (factor arrays) are ordered from most rejected for each factor. This provides a hierarchy of item acceptance for each factor or type of "persons" (Vantubergen, 1966).

The following are examples of hypothetical types of "persons" that the factor analytic model would construct:

| Items | Types | | |
	I	II	III
University 1	1.02	- .24	.72
University 2	1.53	1.03	1.54
University 3	.42	.31	-1.03
University 4	- .06	.32	- .51
University 5	-1.08	-1.35	-1.54
University 6	.80	1.20	.5
University 7	-1.20	.02	- .6
University 8	.70	1.50	2.0

When ordered in terms of the z-scores, the factor array becomes a hierarchy of items that are rated for each of the factors or types. The following is an example of the first typology (Type I):

Z-Score	Item
1.53	University 2
1.02	University 1
.80	University 6
.70	University 8
.42	University 3
- .06	University 4
-1.08	University 5
-1.20	University 7

Similar results were obtained for each type.

Step 5 - The arrays of item z-scores (factor arrays) for each type are compared by subtraction for each pair of factors. This produces arrays of difference scores for each pair of factors. This provides the basis for differentiating one factor or type of person from another Vantugergen, 1966).

This is accomplished by comparing the types dealing with the following questions:

1. What items differentiate one type from another type?

2. What items differentiate one type from all other types?

3. What items or areas of agreement seem to cut across all of the types?

Question 1 is dealt with by comparing the array for all types taken two at a time. The Z-scores for each pair of universities are subtracted and ranked according to absolute differences. To illustrate, consider the following:

	Type II	Type I-Type II	
1.02	- .24	1.26	University 1
-1.20	.02	1.22	University 7
.70	1.80	.80	University 8
1.53	1.03	.50	University 2
.80	1.20	.40	University 6
-1.08	-1.35	.27	University 5
- .06	.32	.38	University 4
.43	.31	.12	University 3

Similar analyses are conducted for all other comparisons.

Question 2. Question 2 was addressed by examining those items that are higher (or lower) in the array for one type than they are in the arrays for all other types. This process is similar to the process followed in Question 1. That is, the Z scores of Type I are compared to the average Z scores for Types II and III.

Question 3. To the extent that the z-scores for all types are nearly equal, one assumes agreement. A consensus item would be one in which the difference between the largest z-score given that item by one of the types and the smallest z score is less than 1.00. In our example, the consensus items would be the following:

Rating of Universities	Maximum Difference	Average Z-Scores Across Types
University 5	.46	1.32
University 2	.50	1.37
University 6	.70	.83
University 4	.83	.08

The average Z-scores of the consensus items and the Z-scores of the differentiation items, which resulted from addressing Questions 1, 2, 3, are used to describe the types. That is, the universities corresponding to the aforementioned Z-scores are used to identify types.

Results of Quannal Analysis

Three Q-factor analyses were computed. One analysis was based upon the ratings of the eight universities, the second on demographic variables, and the third on the university and demographic variables together. On all three of the Q-factor analyses, only one typology emerged.

In the first analysis, all of the 100 subjects were identified in Type I. In the second analysis, 99 of the 100 were identified in Type I. In the analysis combining the universities and demographic variables, 98 of the subjects were identified in Type I. As one can see form these results, only one type consistently emerged; therefore, we were unable to use differences in types as predictor variables. A multiple regression analysis by Fraas on the impact of the demographic variable of the data further validates the homogeneity of this sample.

Since we were in a desperate search for more than one type, it was suggested that we try a cluster approach, which tends to produce more than one type. Ward's (1963) clustering program takes a set of N objects, which are measured on a number of different variables, and attempts to optimally group them from N to N-1, etc. The groupings are based upon maximizing the average intergroup distance, while minimizing the average intragroup distance.

The approach begins by defining each object as a group. These N groups are then reduced by one, until all persons have been classified into one of two groups. More detail of this approach can be found in SAS, as well as Veldman (1967).

Using the clustering program, three cluster analyses were completed. When using a cluster analysis, one has to decide on the number clusters one wants in the solution. The decision used for this study was that no cluster would contain less than five people.

The first cluster analysis, using the universities' ratings and the three demographics, produced four clusters with 27 people in cluster one, 56 in cluster two, 11 in cluster three, and 6 in cluster four. These four clusters accounted for 61% of the variance for all groupings. The second cluster analysis, based upon universities' ratings, produced three clusters with an R^2 equal to .55, with 58 individuals in cluster one, 36 in cluster two, and 7 in cluster three. The third cluster analysis, based

upon demographics alone, produced only two clusters with almost everyone loading on cluster one. Therefore, it was not considered.

The four clusters produced by the first cluster analysis were used as predictor variables to predict the ratings of each of the eight universities, the eight regression equations

Table 3

Multiple linear Regression Results for Model 2

Variable	Regression Coefficients	T Value
X_1	1.410	12.21*
X_2	1.125	9.74*
X_3	1.440	12.47*
X_4	1.775	15.37*
X_5	1.015	8.79*
X_6	1.000	
Constant	-3.38	

n = 800
R^2 = .58
df_d = 695

* Statistically significant at the .01 level.

produced the following values: .12, .27, .17, .18, .18, .26, .34, and .28. When the clusters from the second cluster analysis containing three clusters, were used as predictor variables, they yielded the following R^2 values: .03, .18, .14, .15, .16, .20, .30, and .18. Since the use of cross-validation procedures would produce even lower values, those procedures were not implemented.

Discussion

The conjoint analysis and the multiple regression model with a surrogate person vector produced identical estimates for the five university attributes. The multiple regression procedure that incorporated a surrogate person vector was better able to predict the holdout universities. Thus, these results seem to imply that if a university administration wants to obtain information on which university attributes are most important to their students, either conjoint analysis or a multiple regression model with a surrogate variable is an appropriate procedure.

With this data set the Q-factor analysis failed to provide useful information. The classifying of student by type did not allow for a high degree of explanation of the ratings of the various hypothetical universities. The use of Q-factor analysis, however, may provide insight into the university selection process by students if various groups are identifiable.

Three points should be noted with regard to future research. First, a multiple linear regression model with a surrogate person vector is a valuable procedure to use to determine which university attributes are important to students when selecting a university. The inclusion of the surrogate person variable did improve the researchers' ability to predict the ratings of the holdout universities. Further studies in this area with more detailed attributes would be informative.

Second, unless various groups of students rate the universities differently, Q-factor analysis obviously will not provide useful informaiton. If such groups exist, however, the information may provide university administrators with some insight into what type of students prefer their particular university.

Third, the conjoint and regression analyses are really asking different questions that the Q-factor analysis. The conjoint and regression analyses are attempting to determine which of the university characteristics are most important. The Q-factor analysis attempts to determine if there are various typologies based on the students' university ratings. This third point leads to an often discussed conclusion. Determining the preferable research method is dependent upon the question of interest. In other words, the research question has to dictate the methodology.

References

Bretton-Clark. (1987). Conjoint analyzer. New York.

Conant, J., Brown, J., & Mokwa, M. (1985), Summer). Students are important consumers: Assessing satisfaction in a higher educaton context. Journal of Marketing Education, 13-20.

Fraas, J. W. & Paugh, R. (1989). Student perceptions of the relative importance of selected attributes of an institution of higher education: a conjoint approach. Ashland, Ohio: Ashland University, 1989. (ERIC Document Reproduction Service No. ED 312 960)

Fraas, J. W., & McDougall, W. R. (1983). The use of one full full MLR novel to conduct multiple comparisons in a repeated measures design: An industrial application. Multiple Linear Regression Viewpoints, 12(1), 42-55.

Fraas, J. W. & Newman, I. Conjoint Analysis: A study of the effects of using person var ables. Ashlan, Ohio: Ashland University, 1989. (ERIC Document Reproduction Service No. ED 312 961)

Green, P., Goldbert, S., & Montemajor, M. (1981). A hybrid utility estimation model for conjoint analysis. estimation model for conjoint analysis. Journal of Marketing, 45, 33-41.

Green, P., Carroll, J. D., & Goldberg, M. (1981). A general approach to product design optimization via conjoint analysis. Journal of Marketing, 45, 17-37.

Green, P., & Strinivasan, V. (1978). Conjoinjt analysis in consumer research: Issues and outlook. Journal of Consumer Research, 5, 103-124.

Kuh, G., Comes, M., & Lundquist, I. (1984, Winter). What prospective students really need to know about institutional quality. College and University, 167-175

Leigh, T., Mackay, D., & Summer, J. (1981). Realiability and validity of conjoint analysis and self-explicated weights: A comparison. Journal of Marketing Research, 21, 456-462.

McNeil, K., Kelly, F., & McNeil, J. (1975. esting research hypotheses using multiple linear regressio. Carbondale, IL: Southern Illinois University ess.

Newman, I. & Benz, C. R. Multivariate evaluation design: A suggested technique for mental health systems. April 25, 1987. Ohio Academy of Science, Canton, Ohio.

Pedhazur, E. J. (1977). Coding subjects in repeated measures designs. Psychological Bulletin, 84, 298-305.

Tierney, M. L. (1984, November. Ten questions to ask when choosing a college. Money, 133-134.

Vantubergen, G. N. (1966). Quannal: A computer program for Q analysis. Mass Communication Research Bureau, School of Journalism, University of Iowa. Mimeo.

Veldman, D. (1967). Fortran Programming for the behavioral sciences. New York: Holt, Rinehart, & Winston.

Ward, J. d. (1963). Hierarchical grouping to potimize an objective function. American Statistical Association Journal, 58, 236-244.

Williams, J. D. (1989). Multiple comparisons in higher
    dimension designs. Monograph Series #5. <u>Multiple Linear
    Regression Viewpoints,</u> <u>10</u>(3).
Williams, J. D., & Williams, J. A. (1985a). Testing hypotheses
    in a repeated measures design on employee attitudes with
    large samples. <u>Multiple Linear Regression Viewpoints,</u>
    <u>13</u>(2), 1-20.
Wliams, J. D., & Williams, J. A. (1985b). Testing hypotheses
    in a repeated measures design: An example. <u>Multiple
    Regressions Viewpoints,</u> <u>13</u>(2), 35-46.

rev.042291