[Mid-Western Educational Researcher](https://scholarworks.bgsu.edu/mwer)

Volume 9
Issue 4 *Regression Analyses*

Article 7

1996

The Use of Multiple Regression Models to Determine if Conjoint Analysis Should Be Conducted on Aggregate Data

John W. Fraas
Ashland University

Isadore Newman
The University of Akron

The Use of Multiple Regression Models to Determine if Conjoint Analysis Should Be Conducted on Aggregate Data

John W. Fraas, Ashland University
Isadore Newman, The University of Akron

Abstract

Conjoint analysis is a statistical procedure often used by marketing researchers to measure the relative importance of various characteristics of a product or service as perceived by consumers. During the past ten years, conjoint analysis has been used to estimate consumers' preferences for many different types of products and services including educational services. In a conjoint analysis study, a researcher must determine whether the product factor estimates, which are used to measure consumer preferences, should be calculated and interpreted for each respondent or the respondents collectively. The purpose of this article is to demonstrate how a researcher can use multiple regression models to determine whether it is appropriate to analyze and interpret the aggregate data by examining the factor-respondents interaction effects. A hypothetical example is used to clarify how this technique can be used.

It is a common task for marketing researchers to attempt to measure the relative importance of various attributes of a product or a service as viewed by its consumers. In the mid-1970s, marketing researchers began to use conjoint analysis as a tool to measure the relative importance of a product's attributes as perceived by its current and prospective buyers. As noted by Hair, Anderson, Tatham, and Black (1995), "conjoint analysis is best suited for understanding consumers' reactions to and evaluations of predetermined attribute combinations that represent potential products or services" (p. 557).

Wittink and Cattin (1989) documented the widespread use of conjoint analysis by various companies and institutions. It should be noted, however, that conjoint studies have not been restricted to just the business arena. Such studies have been conducted for educational institutions as well (Fraas & Paugh, 1990). With the increased pressure on educational institutions to market their services, this type of information may provide valuable information on how to market or possibly change an institution's educational service to match the preferences of its potential and current students.

One issue that a researcher who utilizes conjoint analysis will face is whether it is appropriate to analyze the respondents' data collectively. That is, the researcher has the choice of fitting the model to the aggregate of the consumers' responses or fitting a model to each of the respondent's data separately. As noted by Hair et al. (1995), "unless the researcher is definitely dealing with a population exhibiting homogeneous behavior with respect to the factors, aggregate analysis should not be used" (p. 579).

Thus, a researcher who has collected data in a conjoint study must decide whether the population exhibits homogenous behavior with respect to the factors. The purpose of this article is to demonstrate how a researcher can use multiple regression models to determine if, in fact, the population exhibits such responses.

Conjoint Analysis

As previously stated, conjoint analysis is an analytical procedure used to measure the relative importance of various characteristics of a product or service as perceived by consumers. The reader is encouraged to refer to Hair et al. (1995) for an excellent discussion of the steps that a researcher should follow when conducting a conjoint study. Only a brief discussion of the essential components of a conjoint analysis study is presented here.

When conducting conjoint analysis, the researcher must first identify the key decision criteria, that is, the factors that are involved in the choice process. Once these factors are identified, the researcher must determine the number of characteristics, which is referred to as the number of levels, that each factor will contain. Based on the number of factors and their number of levels, the researcher creates the various hypothetical products. These hypothetical products or services will consist of different combinations of factor characteristics.

The number of factors and factor levels dictate the number of hypothetical products that the respondents must evaluate. When the number of different products or services is not prohibitively large, a full-factorial design could be used. Often, however, a conjoint study will include a significantly large number of factors and/or factor levels. In such a case, the total number of possible hypothetical products would far exceed a respondent's endurance to evaluate each one. To illustrate, if a study involved four factors with four levels each, a full-factorial design would require the respondent to evaluate 256 or (4 x 4 x 4 x 4) different services. In studies such as this, the researcher often assumes that a model that contains only the main effects will be an adequate model and then uses a fractional-factorial design. The use of such a design would reduce the required number of hypothetical products from 256 to 16.

Once the hypothetical products are designed, the researcher has the respondents evaluate them by rating each one on a given point scale, such as a l-to-10 scale. The researcher uses a conjoint analysis computer program to estimate the preference or utility associated with each value of each factor. Two such computer programs are produced by Bretton-Clark (1988a, 1988b, & 1988c) and SPSS (1990). These preference estimates are often referred to as part-worth values. The part-worth values, which can be estimated through a regression procedure, are used to judge the importance and type of influence that the factor has on the consumers' preferences.

As previously mentioned, a key question that the researcher must address before the part-worth values are estimated is: Should the part-worth values be estimated and interpreted for each respondent or for the respondents collectively? The following section discusses, with the aid of a simplified hypothetical example, a technique that a researcher could use to judge whether the population is sufficiently homogeneous with respect to the influence of each factor on the respondents' ratings to analyze their data collectively.

The Importance of the Factor-Respondents Interaction Effects

If it is appropriate to conduct a conjoint analysis on the ratings of the respondents as a group, the researcher must assume that the factors do not interact with the respondents. That is, the relative contributions of the factors remain constant across the respondents. If this is not the case, an aggregate analysis of the ratings of the hypothetical products could be misleading. In such a case, segmentation of the respondents would be informative.

To determine if indeed factor-respondents interaction effects are present, a researcher can use multiple regression models that utilize person variables. Person variables, which are discussed by Pedhazur (1977), Williams (1977) and McNeil, Newman, & Kelly (in press), contain zero and one values. For a given person variable, a zero indicates that the corresponding criterion value was not obtained from the person represented by this person variable. And a value of one indicates that the corresponding criterion value was given by this person.

Fraas and Newman (1989) have demonstrated that a multiple regression model that contains person variables can be used to generate part-worth values that are identical to the values estimated by traditional conjoint analysis computer programs. Thus, the use of person variables will not change the part-worth estimates. To determine whether the factor-respondents interaction effects are present, the person vectors must be used along with the factor variables to generate variables that, when used in conjunction with regression models, will estimate the amount of variation in the criterion variable that is associated with those interaction effects. How such variables are generated and incorporated into the appropriate regression models can best be explained through the use of a simple hypothetical conjoint study.

Hypothetical Conjoint Study

Assume that a researcher has decided to use conjoint analysis to estimate the part-worth values for two factors that relate to the students' preferences for certain offerings in a continuing education program. The first factor, which consists of two levels, deals with the number of sessions that the class would meet per week. Since this factor contains two levels, the variable that represents this factor would contain zero and one values. A value of zero represents a class that would meet one day a week for four hours. The other level, which is represented by a value of one, indicates that the class would meet two days a week for two hours each day.

The second factor, which also consists of two levels, deals with the location of the class. A value of zero for this location variable indicates that the class would be held in Ashland, Ohio. A value of one indicates that the class would meet in Medina, Ohio. The four hypothetical continuing education classes, which will be evaluated by the prospective students, are listed in Table 1.

Table 1.
Hypothetical Continuing Education Classes

Hypothetical Courses	Number of Class Meetings Per Week	Class Location
Course 1	Two	Medina
Course 2	Two	Ashland
Course 3	One	Medina
Course 4	One	Ashland

In this example, we are assuming that three prospective students are asked to rate the four hypothetical classes using a 10-point scale. It should be noted that for such a study, the number of prospective students would normally far exceed three. For the sake of demonstrating our proposed technique in a clear manner, however, the number of respondents is limited to three.

The hypothetical rating given to each of the four hypothetical classes by each prospective student is listed in Table 2 under the variable entitled Y. The values for the factors that contain the information regarding the Number of class meetings per week and the Location that corresponds to these ratings are listed in Table 2 under the symbols N_1 and L_1, respectively. In addition, the values for the three-person variables are also listed in Table 2 under the symbols P_1, P_2, and P_3.

Table 2.
Variables Used in the Regression Analyses

Variables and Values							
Y	N_1	N_2	L_1	L_2	P_1	P_2	P_3
10	1	0	1	0	1	0	0
8	1	0	0	1	1	0	0
6	0	1	1	0	1	0	0
4	0	1	0	1	1	0	0
5	1	0	1	0	0	1	0
3	1	0	0	1	0	1	0
6	0	1	1	0	0	1	0
7	0	1	0	1	0	1	0
8	1	0	1	0	0	0	1
6	1	0	0	1	0	0	1
5	0	1	1	0	0	0	1
4	0	1	0	1	0	0	1

Interaction Variables

As previously mentioned, interaction variables must be generated and included in the multiple regression model to determine if factor-respondents interaction effects exist. These interaction terms are generated by first creating a companion variable for each factor variable in the study. The values contained in the variable that serves as the companion variable to the Number of meetings per week variable are zero when the value for N_1 is one. And the value is one when the value for N_1 is zero. The values contained in the variables that serve as the companion variable to the location variable are generated in the same manner. The companion variables for the Number of meetings per week and the Location variables, which are represented by the symbols of N_2 and L_2 respectively, are listed in Table 2.

The next step in generating the variables required to test the factor-respondents interaction effects requires that each of the person variables (P_1, P_2 and P_3) be multiplied by each of the two factor variables (N_1 and L_1) and by each of the two companion variables (N_2 and L_2). The 12 interaction variables, which are formed by multiplying the factor and companion variables by the person variables, are listed in Table 3. The values contained in each interaction variable are listed in Table 4.

Table 3.
Variables Used to Generate the 12 Interaction Variables

Interaction Variables

$X_{10} = N_1 * P_1$
$X_{11} = N_1 * P_2$
$X_{12} = N_1 * P_3$
$X_{13} = N_2 * P_1$
$X_{14} = N_2 * P_2$
$X_{15} = N_2 * P_3$
$X_{16} = L_1 * P_1$
$X_{17} = L_1 * P_2$
$X_{18} = L_1 * P_3$
$X_{19} = L_2 * P_1$
$X_{20} = L_2 * P_2$
$X_{21} = L_2 * P_3$

Table 4.
Interaction Variables Used in the Regression Analyses

Variables and Values											
X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}	X_{20}	X_{21}
1	0	0	0	0	0	1	0	0	0	0	0
1	0	0	0	0	0	0	0	0	1	0	0
0	0	0	1	0	0	1	0	0	0	0	0
0	0	0	1	0	0	0	0	0	1	0	0
0	1	0	0	0	0	0	1	0	0	0	0
0	1	0	0	0	0	0	0	0	0	1	0
0	0	0	0	1	0	0	1	0	0	0	0
0	0	0	0	1	0	0	0	0	0	1	0
0	0	1	0	0	0	0	0	1	0	0	0
0	0	1	0	0	0	0	0	0	0	0	1
0	0	0	0	0	1	0	0	1	0	0	0
0	0	0	0	0	1	0	0	0	0	0	1

Regression Models

Two regression models are required to statistically test the factor-respondents interaction effects. One model, which is referred to as the Full Model contains the eight linearly independent interaction variables. It should be noted that these eight variables are designed to measure the amount of variation in the ratings associated with the main effects of number of meetings and location, and the respondents as well as the factor-respondents interaction effects. The Full Model would be as follows:

$$Y = a + b_{10}X_{10} + b_{13}X_{13} + b_{14}X_{14} + b_{15}X_{15} + b_{17}X_{17} + b_{18}X_{18} + b_{19}X_{19} + b_{20}X_{20} + e$$

The other model, which is referred to as the Restricted Model, contains the two variables that represent the Number of classes per week factor, the Location factor, and the two linearly independent person variables. The Restricted Model is as follows:

$$Y = a + b_1N_1 + b_2L_1 + b_3P_1 + b_4P_2 + e$$

When these regression models were used to analyze the respondents' ratings of the hypothetical classes, the R^2 values for the Full Model and the Restricted Model are .943 and .390 respectively. The difference between these two R^2 values, which is equal to .553, is attributed to the factor-respondents interaction effects.

This difference between the two R^2 values is statistically tested with an *F* test by using the following formula:

$$F= \frac{(R_F^2 - R_R^2) / dfn}{(1 - R_F^2) / dfd}$$

where:

1. R_F^2 represents the R^2 value for the Full Model.

2. R_R^2 represents the R^2 value for the Restricted Model.

3. dfn represents the value that is equal to the difference between the number of linearly independent variables in the Full Model minus the number of linearly independent variables in the Restricted Model.

4. dfd represents the value that is equal to the number of cases minus the quantity one plus the total number of linearly independent variables contained in the Full Model. It should be noted that the number of cases is equal to the number of respondents multiplied by the number of hypothetical products.

Since the number of cases in this example is 12 or (3 respondents x 4 products) and the Full and Restricted Models contain eight and four linearly independent variables respectively, the dfn value is 4 or (dfn = 8 - 4) and the dfd value is 3 or (dfd = 12 - 9). Thus, the *F* value for the difference between the R^2 values (.553) would be calculated as follows:

$$F= \frac{(.943 - .390) / 4}{(1 - .943) / 3} = 7.28$$

The probability for this *F* value is .067.

Since the denominator degrees of freedom value will be quite small when this technique is used, the power of the *F* test used to test for the factor-respondents interaction effects will tend to be low. To increase the power of this test, we suggest that a researcher use a liberal alpha level, such as .25. This practice is similar to the one used by researchers who set a high alpha value, usually at the .50 level, when attempting to determine if an outlier exists in a regression analysis with the *F* tests of Cook's distance measures (Neter, Wasserman & Kutner, 1985).

Since the probability value of this *F* test ($F = 7.28$, $p = .067$) is less than the alpha level of .25, evidence exists that would allow the researcher to conclude that it may be misleading to aggregate the data when interpreting the influences of the various factors on the respondents' ratings. That is, the researcher may find it more informative to segment the respondents based on the differing influences of the factors on their ratings.

Discussion

In this article we have attempted to describe a technique that a researcher could use to determine whether respondents' ratings, which are obtained in a conjoint study, should or should not be analyzed and interpreted for each respondent. This technique is based on the position that the data should not be aggregated when factor-respondents interaction effects exist.

To determine if respondent-factor interaction effects are present, the researcher would design two regression models. The first model would include a series of variables that measure the amount of variation in the ratings that is associated with the factor-respondents interaction effects as well as the variation associated with the factors and the respondents. The second model contains only the variables that will measure the amount of variation in the ratings associated with the factors and the respondents.

The difference between the R^2 values of the Full and Restricted Models is equal to the proportion of variation in the ratings of the hypothetical products that is associated with the factor-respondents interaction effects. This difference is statistically tested with an *F* test. Since the denominator degrees of freedom value will tend to be small in this procedure, the researcher would compare the probability of this *F* value to a liberal alpha value of, possibly .25. If the probability value of the *F* test is less than the alpha value, the researcher would question the appropriateness of analyzing the aggregate data-set and interpreting the results.

Only one simple type of conjoint analysis was used in this article to demonstrate how this technique could be used to assist the researcher in determining whether it is appropriate to aggregate the data. The reader should be aware that the technique, as presented in this article, may need to be modified if it is to be appropriately applied to more complicated conjoint studies.

It is important to note that the appropriateness of using this technique to determine whether the analysis should be conducted on the aggregate data is not restricted to just conjoint studies. We believe that this technique would be equally valuable for researchers who utilize repeated measures designs and the study involves more than one respondent. One of the major problems researchers encounter when analyzing aggregate scores for respondents with repeated measures is the possibility of not being able to identify critical incidences. For example, consider a repeated measures study in which the scores for every respondent in the repeated measures design exhibits a roller coaster effect over time, but the peaks and valleys occur in different time periods. When one aggregates such scores, the analysis may appear to be linear, although not one of the respondent's possesses scores that are, in fact, linear over time. The specific techniques that could assist a researcher in determining whether it is appropriate to analyze the aggregate data in such studies, although beyond the scope of this article, are similar to the technique that we have discussed.

References

Bretton-Clark (1988). *Conjoint Analyzer* [Computer Software]. New York: Bretton-Clark.

Bretton-Clark (1988). *Conjoint Designer* [Computer Software]. New York: Bretton-Clark.

Bretton-Clark (1988). *Simgraf.* New York: Bretton-Clark.

Fraas, J.W. & Newman, I. (1989, October). *Conjoint analysis: A study of the effects of using person variable.* Paper presented at the meeting of the Mid-Western Educational Research Associatio, Chicago, IL.

Fraas, J.W. & Paugh, R. (1990, November). *Student perceptions of the relative importance of selected attributes of an institution of higher education: A conjoint approach.* Paper presented at the meeting of the Southern Marketing Association Conference, Orlando, FL.

Hair, J.F. Jr., Anderson, R.E., Tatham, R.D., & Black, W.L. (1995). *Multivariate data analysis with readings, (4th ed.).* Englewood Cliffs, NJ: Prentice Hall.

McNeil, K., Newman, I. & Kelly, F.J. (In press). *Testing research hypotheses with the general linear model.* Carbondale, IL: Southern Illinois University Press.

Neter, J., Wasserman, W. & Kutner, M.H. (1985). *Applied linear statistical models.* Homewood, IL: Richard D. Irwin.

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

SPSS, Inc. (1990). *SPSS Categories* [Computer Software]. Chicago, IL.

Williams, J.D. (1977). A note on coding the subjects effect in treatments X subject designs. *Multiple linear regression viewpoints, 8*(1), 32-35.

Wittink, D.R., & Cattin, P. (1989). Commercial use of conjoint analysis: An update. *Journal of Marketing, 53*, 91-96.

Reviewers for 1996

We wish to acknowledge with thanks the services of the following reviewers who helped with our peer-review process during the 1996 year:

Anthony L. Ambrosio
Wichita State University

Deborah L. Bainer
Ohio State University - Mansfield

Mary K. Bendixen-Noe
Ohio State University - Newark

Russell N. Carney
Southwest Missouri State University

Josue Cruz
University of South Florida - Tampa

Orpha Duell
Wichita State University

Mary Ann Flowers
Cleveland State University

W. Gregory Gerrick
Ashland University

Stephen G. Jurs
University of Toledo

Christopher L. Kleinke
University of Alaska - Anchorage

Joseph C. Kush
Duquesne University

Susie Lamborn
University of Wisconsin - Milwaukee

Ronald N. Marso
Bowling Green State University

James V. Mead
University of Dayton

A. William Place
University of Dayton

John T. Pohlmann
Southern Illinois University

Diane Pollard
University of Wisconsin - Milwaukee

Diane L. Schallert
University of Texas - Austin

Greg Schraw
University Of Nebraska - Lincokn

Mary R. Sudzina
University of Dayton

Joan Timm
University of Wisconsin - Oshkosh

Betty Lou Whitford
University of Louisville

Kenneth Wodtke
University of Wisconsin - Milwaukee