

# Introducing the Market Research Analysis Application

Wayne E. Watson

## Abstract

Market research focuses on assessing the preferences and choices of consumers and potential consumers. A new component of SAS/STAT software in Release 6.11 of the SAS System is an application written in SAS/AF that provides statistical and graphical techniques for market research data analysis. The application allows you to employ statistical methods such as conjoint analysis, discrete choice analysis, correspondence analysis, and multidimensional scaling through intuitive point-and-click actions.\*

## Conjoint Analysis

Conjoint analysis is used to evaluate consumer preference. If products are considered to be composed of attributes, conjoint analysis can be used to determine what attributes are important to product preference and what combinations of attribute levels are most preferred.

Usually, conjoint analysis is a main-effects analysis of variance of ordinally-scaled dependent variables. Preferences are used as dependent variables, and attributes are used as independent variables. Often, a monotone transformation is used with the dependent variables to fit a model with no interactions.

As an example, suppose you have four attributes that you think are related to automobile tire purchase. You want to know how important each attribute is to consumers' stated preferences for a potential tire purchase. The four attributes under investigation are

- brand name
- expected tread mileage
- purchase price
- installation cost

The attributes of brand name, tread mileage, and purchase price have three possible values and installation cost has two values. The values for each attribute are:

---

\*For current documentation on the Market Research Application see SAS Institute Inc, *Getting Started with The Market Research Application*, Cary, NC: SAS Institute Inc., 1997, 56 pp. This paper was written and presented at SUGI 20 (1995) by Wayne E. Watson. This paper was also presented to SUGI-Korea (1995) by Warren F. Kuhfeld. Wayne Watson is a Research Statistician at SAS and wrote the Marketing Research Application which uses procedures and macros written by Warren F. Kuhfeld. Copies of this chapter (MR-2010B), the other chapters, sample code, and all of the macros are available on the Web [http://support.sas.com/resources/papers/tnote/tnote\\_marketresearch.html](http://support.sas.com/resources/papers/tnote/tnote_marketresearch.html). For help, please contact SAS Technical Support. See page 25 for more information.

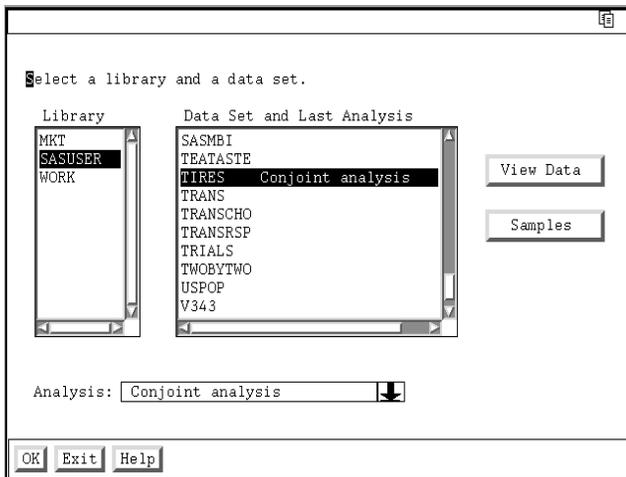


Figure 1. Selecting a Data Set and Analysis

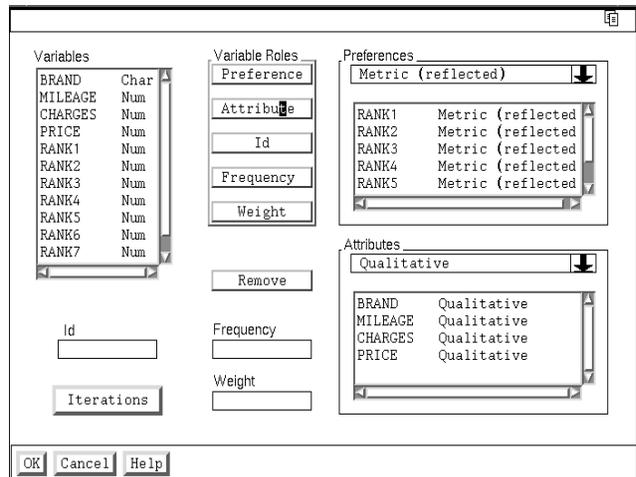


Figure 2. Conjoint Analysis Variable Selection

Brand: Michelin, Goodyear, Firestone  
 Tread Mileage: 40,000, 60,000, 80,000  
 Price: \$45.00, \$60.00, \$75.00  
 Installation Cost: \$0.00, \$7.50

Seven respondents are asked to rank in order of preference 18 out of the possible 54 combinations. Although rankings are used in this example, preference ratings are frequently used in conjoint analysis.

*Invoking the Application.* With the data in the SAS data set, SASUSER.TIRES, you can invoke the Market Research application and perform a conjoint analysis. The application is invoked by issuing the “market” command on any command line.

*Selecting a Data Set and Analysis.* The first window displayed requires you to select a data set and an analysis. Because your data set is SASUSER.TIRES, select SASUSER as the library in the left-hand list box and TIRES as the data set in the right-hand list box. Then, select an analysis by clicking on the down arrow to the right of the analysis name field below the list boxes and select “Conjoint Analysis” from the displayed popup menu. See Figure 1.

View the data by pressing the View Data button and then selecting “Data values.” The other selection under the View Data button, “Variable attributes,” displays information about each variable.

*Selecting Variables.* To proceed with the analysis once you have selected a data set and an analysis, press the OK button at the bottom of the window.

The analysis requires preference and attribute variables. The preference variables are the ranks from the seven respondents and the attribute variables are the four factors. See Figure 2.

You can choose to perform a metric or a non-metric conjoint analysis; the metric analysis uses the ranks as they are, while the non-metric analysis performs a monotone transformation on the ranks. To set the measurement type for the preferences, click on the down arrow in the Preferences box at the top right of the window. Select “Metric (reflected).” “Reflected” is used because the lowest rank value, 1, corresponds to the most preferred offering. If the highest preference value corresponded to the most

preferred offering, the “Metric” selection should be used instead.

To select preference variables, select RANK1, RANK2, ... RANK7 in the Variables list box on the left side of the window, and press the Preference button in the Variable Roles box.

Likewise, you must select a measurement type for the attribute variables you want to use. The default measurement type for attributes is Qualitative, which treats the variable as a set of dummy variables with the coefficients of the dummy variables summing to 0. In this way, the utility coefficients \* of each attribute sum to 0.

Use this measurement type for all four attribute variables, BRAND, MILEAGE, CHARGES, and PRICE. After selecting these four variables in the Variables list box, press the Attribute button in the Variable Roles box. Alternatively, you could use the “Continuous” measurement type for MILEAGE, CHARGES, or PRICE because these attributes are quantitative in nature.

To delete one or more of the Preference or Attribute variables, either double-click on each one in the appropriate right-hand list box or select them in any of the three list boxes and press the Remove button.

To obtain help about the window, press the Help button at the bottom of the window or click on any of the border titles on the window, for example, “Variables,” “Variable Roles,” “Preferences.”

Once the variables have been selected, press the OK button at the bottom of the window to perform the analysis. To change the analysis, return to the Variable Selection window by pressing the Variables button on the analysis main window.

*Results.* The first result is a plot of the relative importance of each attribute. Relative importance is a measure of importance of the contribution of each attribute to overall preference; it is calculated by dividing the range of utilities for each attribute by the sum of all ranges and multiplying by 100.

In the example, Tire Mileage is the most important attribute with an average relative importance of 49%. The box-and-whisker plot displays the first and third quartiles as the ends of the box, the maximum and minimum as the whiskers (if they fall outside the box), and the median as a vertical bar in the interior of each box. See Figure 3.

To display a selection of additional results, press the Results button on the window. The first selection, the Utilities Table window, displays the utility coefficients for each level of an attribute for all preferences (the dependent variables). The relative importance of each attribute is displayed separately for each preference variable. This table illustrates that BRAND is the most important attribute for RANK1, the first respondent, and Michelin is the most preferred brand, because it has the highest utility coefficient value. Thus, the first respondent preferred a 80,000 mile, \$45 Michelin with no installation charge.

After closing this window, you can view these results in graphical form by pressing the Results button again and selecting “Utilities plots.” The plot of the Brand utilities indicates that one respondent clearly prefers Michelin while the other respondents only mildly prefer one brand over another.

To change the plot from the BRAND to the MILEAGE attribute, select MILEAGE in the list box at the right. All but one person prefer longer over shorter mileage tires, and that one prefers the 60,000 mile tire. You can examine plots for the PRICE and CHARGES attributes in the same way.

---

\*Utility coefficients are estimates of the value or worth to a subject of each level of an attribute. The most preferred combination of attributes for a subject is the one with the attribute levels having the highest utility coefficient values for each attribute.

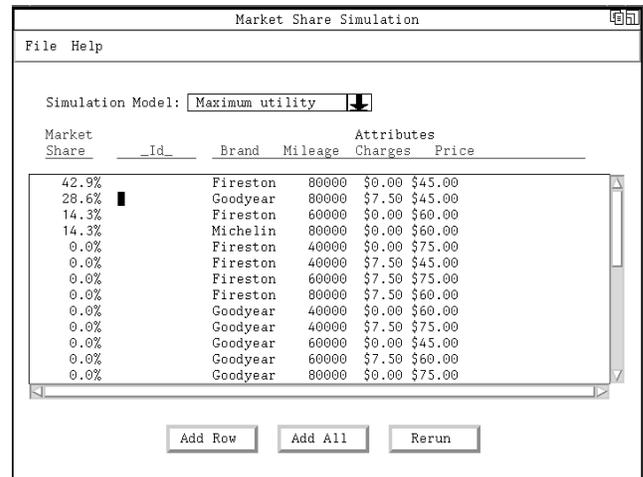
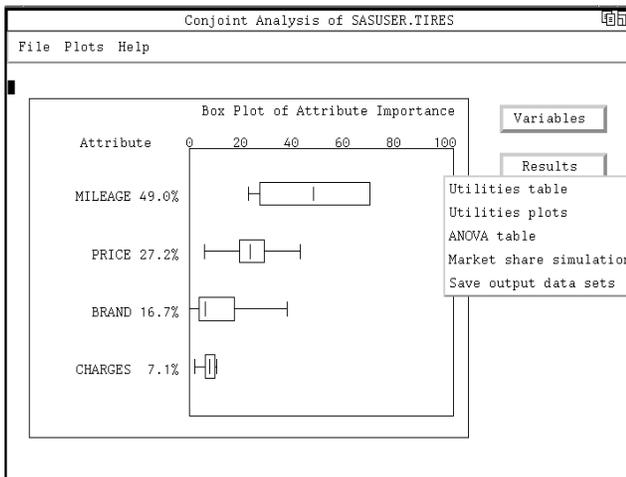


Figure 3. Plot of Relative Importance of Attributes      Figure 4. Estimating Market Share

*Estimating Market Share.* You also can calculate the expected market share for each tire purchase alternative in the sample. To do so, press the Results button and select “Market Share Simulation.” The entry in the table with the largest market share is the 80,000 mile, \$45 Firestone with no installation charge. It is expected to account for 42.9% of the market. The maximum utility simulation model, the default, was used to calculate the market share. You can choose from two other models: the logit model and the Bradley-Terry-Luce model. Click on the down arrow at the top of the window and select the desired model from the displayed list. See Figure 4.

Only 18 of the 54 possible tire purchase combinations were presented to the respondents. You may want to predict the expected market share of one or more of the combinations that were not present in the sample. To do so, press the Add Row button at the bottom of the window and fill in the observation in the top row of the table. Click on “-Select-” in each attribute column and select the desired level. If the observation that you create is a duplicate, a warning message is displayed. You can modify the contents of the Id column to contain a description of your own choice. After you have added some combinations, you can produce the expected market shares by pressing the Rerun button.

As an example an 80,000 mile, \$45 Michelin with no installation charges would be expected to have a 64.3% market share if it was the only combination added to the original sample. Adding combinations may change the estimated market share of the other combinations.

## Discrete Choice Analysis

Conjoint analysis is used to examine the preferences of consumers. The rationale for the use of preferences is that they indicate what people will choose to buy. Often in market research, the choices that consumers actually make are the behavior of interest. In these instances, it is appropriate to analyze choices directly using discrete choice analysis.

In discrete choice analysis, the respondent is presented with several choices and selects one of them. As in conjoint analysis, the factors that define the choice possibilities are called attributes. Here, they are called choice attributes to distinguish them from other factors, like demographic variables, that may be of interest but do not contribute to the definition of the choices. Each set of possible choices is called a choice set.

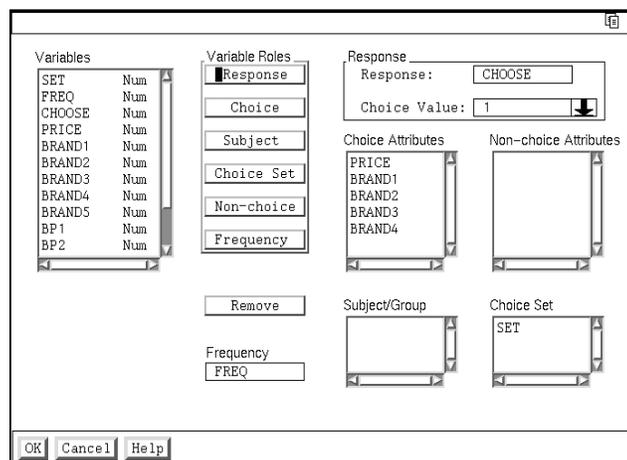


Figure 5. Discrete Choice Analysis Variable Selection

This example has choice possibilities defined by two attributes, price and brand. Five choice alternatives are presented at a time to a respondent, from which one alternative is chosen. Eight of these choice sets are presented, each one with a different set of five combinations of price and brand.

To change to a different data set or analysis, select “File → New dataset/analysis” on the main analysis window. Each time you change the data set or analysis or exit the application, you are asked if you want save the changes that you have made during the session. On the data set selection window, select the PRICE data set in the SASUSER library and then select “Discrete choice analysis.” To continue, press the OK button.

With the other analyses in the application, you would be taken directly to the appropriate variable selection window. With discrete choice analysis, a supplementary window is displayed to help you determine if your data are in the appropriate form.

With discrete choice analysis, the structure of the data is important and must be in one of several layouts. After specifying if your data are contained in one or two data sets and whether a frequency variable is used, you can view the appropriate layout by pressing the Examine button. The most important requirement of the data layout is that all choice alternatives must be included, whether chosen or not.

If your data are not in the proper form, they must be rearranged before proceeding with the analysis. If your data are in the proper form, continue with the analysis by pressing the OK button. If not, press the Cancel button.

On the Variable Selection window that appears next, you must select several required variables: a response variable, some choice attribute variables, and a subject variable. Optionally, you can also choose a frequency variable and some non-choice attribute variables. If you select a frequency variable, a subject variable is not necessary.

For this example, select CHOOSE as the response variable. You also must indicate which value of the variable represents a choice. Click on the down arrow to the right of “Choice Value:” and select 1 from the list. In this example the value 1 indicates the chosen alternative and the value 0 indicates the non-chosen alternatives. See Figure 5.

Next, select PRICE and BRAND1, BRAND2, ..., BRAND4 as Choice attributes. BRAND is a nominal variable with five levels. It can be represented as four dummy-coded variables. \*

Select FREQ as the frequency variable. The frequency variable contains the count of the number of times that a choice alternative was selected.

Because the data include more than one choice set, a Choice Set variable is needed; the choice set variable in this example is SET. After selecting the appropriate variables, press the OK button to perform the analysis.

On the analysis main window, a bar chart is displayed of the significances of each of the choice and non-choice attributes. The chart illustrates that PRICE, BRAND1, BRAND2, and BRAND4 are significant.

You can view other results by pressing the Results button and selecting “Statistics,” “Choice probabilities,” or “Residual plots” from the ensuing menu. Overall model fit statistics and parameter estimates for the attributes are available from the Statistics window. Probabilities for each choice alternative are available from the Choice Probabilities window. Plots of residual and predicted values are available from the Residual Plots window.

## Correspondence Analysis

Categorical data are frequently encountered in the field of market research. Correspondence analysis is a technique that graphically displays relationships among the rows and columns in a contingency table. In the resulting plot there is a point for each row and each column of the table. Rows with similar patterns of counts have points that are close together, and columns with similar patterns of counts have points that are close together.

The CARS data set in the SASUSER library is used as an example (also described in the *SAS/STAT User's Guide*). The CARS data are a sample of individuals who were asked to provide information about themselves and their cars. The pertinent questions for the example are country of origin of their car and their family status.

*Simple Correspondence Analysis.* Simple correspondence analysis analyzes a contingency table made up of one or more column variables and one or more row variables. To select a data set on which to perform a correspondence analysis, select “File → New dataset/analysis” on the main analysis window. First, select the CARS data set, then select “Correspondence analysis” as the analysis, and then press the OK button.

This example uses raw variables instead of an existing table. The desired type of analysis (simple correspondence analysis) and data layout (raw variables) are default selections on the Variable Selection window. Select ORIGIN, the country of origin of the car, as the column variable and MARITAL, family status, as the row variable to create the desired contingency table. See Figure 6.

---

\*Each dummy-coded variable has the value of 1 for a different level of the attribute. In this way, each dummy-coded variable represents the presence of that level and the absence of the other levels.

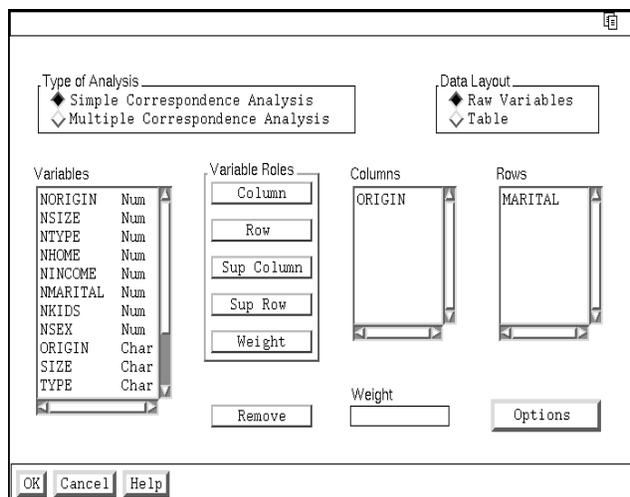


Figure 6. Simple Correspondence Analysis Variable Selection

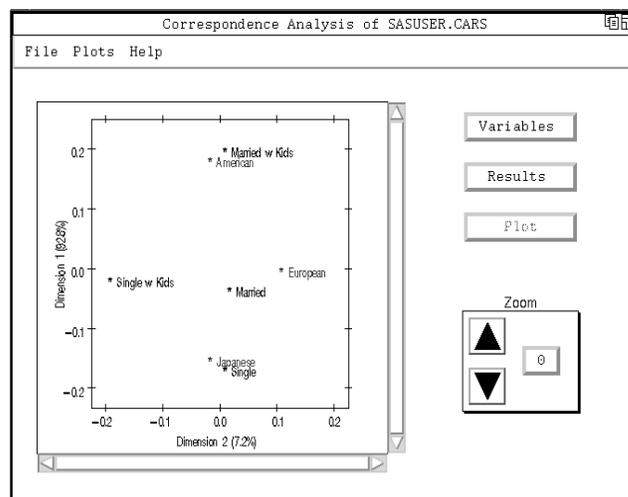


Figure 7. Correspondence Analysis Plot

*Plot.* The plot displays the column points and row points. The first example in the *SAS/STAT User's Guide* provides an interpretation of the plot. The interpretation has two aspects: what each dimension represents and what the relationship of the points in the dimensional space represents. An interpretation of the vertical dimension is that it represents the country of origin of the cars, with most of the influence coming from whether the car is American or Japanese. The horizontal dimension appears to represent “Single with kids” versus all of the other values. See Figure 7.

Although the row and column points are spread throughout the plot, “married” and “single” appear to be slightly more similar to each other than any of the other points. Keep in mind that distances between row and column points cannot be compared, only distances among row points and distances among column points. However, by treating the country-of-origin points as lines drawn from the 0,0 point and extending off the graph, you can see that the “Married with kids” point is closest to the American car line and the “Single” point is closest to the Japanese car line.

*Plot Controls.* To enlarge the plot, click on the up arrow in the zoom control box. To return the plot to its zero zoom state, click on the [0] button. If the plot is zoomed, you can move the plot left and right and up and down using the scroll bars.

*Results.* You can view other results by pressing the Results button and selecting “Inertia table,” “Statistics,” or “Frequencies.” The Inertia Table window lists the singular values and inertias for all possible dimensions in the analysis. The Statistics window displays tables of statistics that aid in the interpretations of the dimensions and the points: the row and column coordinates, the partial contributions to inertia, and the squared cosines. The Frequency Table window displays observed, expected, and deviation contingency tables and row and column profiles.

*Multiple Correspondence Analysis.* In a multiple correspondence analysis, only column variables are used. They are used to create a Burt table\* which is then used in the analysis.

The same data set can be used to illustrate multiple correspondence analysis. Return to the Variables Selection window by pressing the Variables button on the main analysis window. See Figure 8. Perform the following steps:

1. Remove the current column and row variables either by double-clicking on them or by selecting them and pressing the Remove button.
2. Select “Multiple Correspondence Analysis” in the Type of Analysis box in the upper left of the window.
3. Select the column variables ORIGIN, TYPE, SIZE, HOME, SEX, INCOME, and MARITAL by clicking on the ORIGIN variable and dragging through the list to the MARITAL variable, then press the Column button.
4. Press the OK button to perform the analysis.

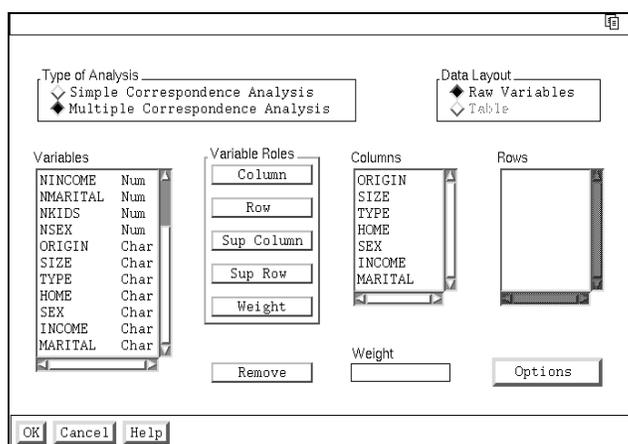


Figure 8. Multiple Correspondence Analysis Variable Selection

The distances between points of different variables can be interpreted in multiple correspondence analysis because they are all column points. However, the multiple correspondence analysis example has more dimensions (12) to interpret and examine than the single correspondence analysis example (2). The total number of dimensions can be examined in the inertia table, which is accessed from the Results button.

By default, a two-dimensional solution is computed. To request a higher dimensional solution, open the Variable Selection window, press the Options button, and select (or enter) the desired number of dimensions.

If you request a three-dimensional (or higher) solution, you can plot the dimensions two at a time by pressing the Plot button and selecting dimensions for the x axis and the y axis.

\*A Burt table is a partitioned symmetric matrix containing all pairs of crosstabulations among a set of categorical variables. For further explanation, see the *SAS/STAT User's Guide*

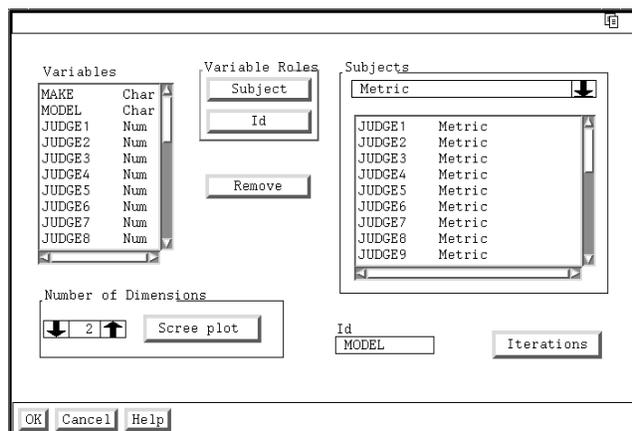


Figure 9. MDPREF Analysis Variable Selection

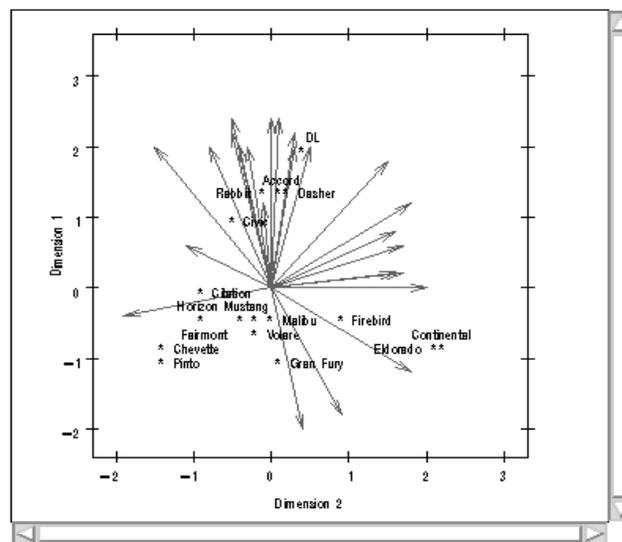


Figure 10. MDPREF Plot

## Multidimensional Preference Analysis

With conjoint analysis, respondents indicate their preferences for products that are composed of attributes determined by the experimenter. Sometimes, the data of interest may be preferences of existing products for which relevant attributes are not defined for the respondent. Multidimensional preference analysis (MDPREF) is used to analyze such data.

MDPREF is a principal component analysis of a data matrix whose columns correspond to people and whose rows correspond to objects, the transpose of the usual people by objects multivariate data matrix.

The CARPREF data set in the SASUSER library is used as an example (also described in the *SAS/S-TAT User's Guide*. It contains data about the preferences of 25 respondents for 17 cars. The preferences are on a scale of 0 to 9 with 0 meaning a very weak preference and 9 meaning a very strong preference. Select the data set and analysis as described in the preceding examples.

As in conjoint analysis, you can choose to perform a metric or non-metric analysis. Choose the measurement type by clicking the arrow in the upper right corner of the window and selecting the desired type. Other, less frequently used, types are available under the “Other” selection. The measurement type is used for all subsequently selected Subject variables. Infrequently, subject variables with different types may be used.

For the example, use the Metric measurement type. Select the preference ratings of each respondent, JUDGE1, JUDGE2, ..., JUDGE25, as Subject variables. Also, select MODEL as the Id variable. See Figure 9.

You also can set the number of dimensions for the analysis; the default is two. A scree plot of the eigenvalues is useful in determining an appropriate number of dimensions. To display the scree plot, press the Scree Plot button. The plot illustrates that the magnitude of the eigenvalues falls off for the first two dimensions; then the plot flattens out for the third and remaining dimensions. From this graph, two dimensions appear appropriate. After closing the Scree Plot window, press the OK button to perform the analysis. See Figure 10.

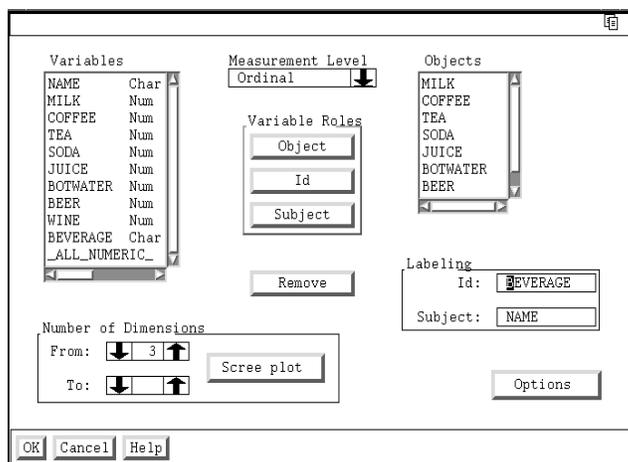


Figure 11. MDS Variable Selection

*Results.* The plot on the main analysis window contains points for the 17 car models and vectors for the 25 respondents. Interpretations of the two dimensions are 1) the vertical dimension separates foreign and domestic cars in the upper half and lower half, respectively, and 2) the horizontal dimension separates small cars and big cars in the left and right halves, respectively. Respondents prefer cars whose points are closest to their vector. Notice that there are a number of vectors in the upper right quadrant of the plot but there are no cars. This lack of available products to satisfy peoples' preferences indicates a possible niche to fill.

Other results are the “Initial Eigenvalue Plot,” “Final Eigenvalue Plot,” and “Configuration Table.” The Initial Eigenvalue plot is the same as the scree plot on the Variable Selection window. The Final Eigenvalue plot is also a scree plot; it differs from the initial plot only if a measurement type other than Metric is used. The Configuration Table contains the coordinates for the car points.

## Multidimensional Scaling

Multidimensional Scaling (MDS) takes subjects' judgments of either similarity or difference of pairs of items and produces a map of the perceived relationship among items.

For example, suppose you ask seven subjects to state their perceived similarity on a 1 to 7 scale for pairs of beverages, with 1 meaning very similar and 7 meaning very different. The beverages are milk, coffee, tea, soda, juice, bottled water, beer, and wine. Someone may state that their perceived similarity between coffee and tea is 3, somewhat similar, or 7, very different. There are 28 possible pairs of these eight beverages.

The data are ordered in an eight observation by eight variable matrix with one matrix (eight observations) for each subject. On the Data Set Selection window, select the BEVERAGE data set in the SASUSER library, then press the OK button. A message window informs you that MDS requires either similarity or distance data. Press the Continue button.

On the Variables Selection window, select the variables MILK, COFFEE, TEA, SODA, JUICE, BOTWATER, BEER, AND WINE as the objects. See Figure 11. It is crucial that the order of the objects is the same as their order in the rows of each matrix. In other words, from the above order, the upper left corner element in the matrix is MILK, MILK (which has a distance of zero) and the element to its

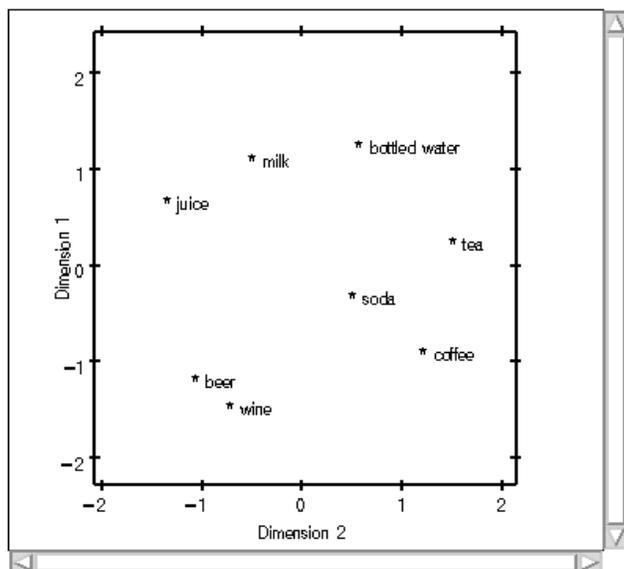


Figure 12. MDS Coordinates Plot

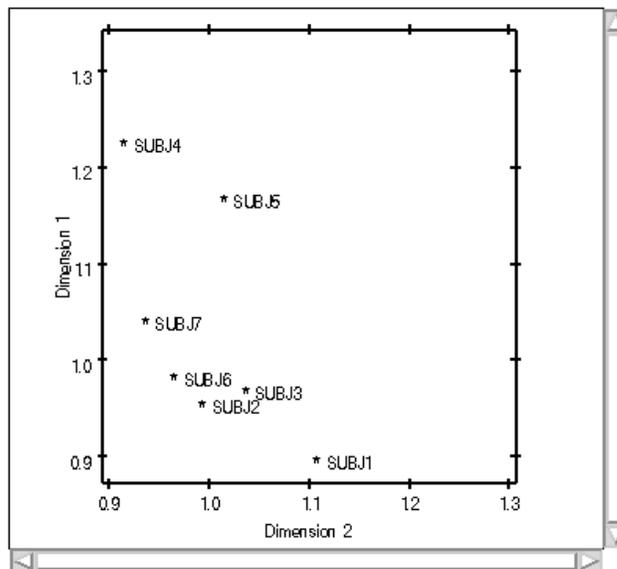


Figure 13. MDS Individual Coefficients Plot

right is MILK, COFFEE.

Also, select BEVERAGE, the beverage names, as the ID variable and NAME, the subject identifiers, as the SUBJECT variable. Because the objects are ordinally-scaled, the ordinal measurement level, the default, is appropriate for this example.

If you think that your subjects may use different perceptual schemes for judging similarity, you can choose to perform an individual differences analysis. Press the Options button and select “Individual Differences Analysis.” The data are distances, the default, because larger numbers represent more difference (less similarity). If the data were similarities, you would choose the appropriate selection on the Options window. To close the options window, press the OK button.

As in correspondence analysis and MDPREF analysis, you can set the number of dimensions for the solution. With MDS you have an extra capability; you can solve for several dimensional solutions in one analysis.

Choose a three-dimensional solution by entering a “3” in the input field to the right of the “From:” label or by clicking on the up arrow to its right until the number 3 appears in the input field. As with the other dimensional analyses, a scree plot may be useful in determining the appropriate number of dimensions. You can create the plot by pressing the Scree Plot button.

To continue with the analysis, press the OK button on the Variable Selection window.

*Results.* As with the correspondence analysis and MDPREF plots, interpreting the MDS plot has two parts: 1) finding a reasonable interpretation for each of the plot dimensions, and 2) finding a reasonable interpretation of the relationship of the points in the plot. See Figure 12.

The presence of bottled water, milk, and juice at the top of the plot and wine, beer, and coffee at the bottom of the plot might indicate a good for you/not so good for you interpretation for the vertical dimension, Dimension 1. The horizontal dimension, Dimension 2, does not have as clear an interpretation. Try to come up with your own interpretation that would have tea, coffee, and water on one side and juice, beer, and wine on the other.

Because you requested a three-dimensional solution, two other plots can be displayed: Dimensions 1 and 3 and Dimensions 2 and 3. To change which dimensions are plotted, press the Plot button and select the desired dimensions. Also, on this window you can choose to display the coefficients of the individual differences analysis instead of the coordinates. To do so, select “Coefficient” at the bottom of the window and press the OK button.

In an individual differences analysis, there is a common perceptual map for all subjects, but different subjects have different weights for each dimension. See Figure 13. SUBJ4 is found to be highest on the vertical axis and lowest on the horizontal axis. In other words, SUBJ4 weights whatever this dimension represents more than do the other subjects and it weights whatever dimension 2 represents less than the other subjects. If the good-for-you interpretation is appropriate for Dimension 1, then it plays a larger role in SUBJ4’s perceptual mapping of these beverages than it does for other subjects.

It is possible that SUBJ1, SUBJ2, SUBJ3, SUBJ6, and SUBJ7 may cluster together and SUBJ4 and SUBJ5 may be outliers. Additional subjects may sharpen this possible clustering or eliminate it. MDS is useful in market research for discovering possible perceptual perspectives used by consumers and for revealing possible market segments.

You can display other results by pressing the Results button. These results include Fit statistics, Configuration tables, Residual plots, and the Iteration history. The fit statistics are measures of how well the data fit the model. The Configuration tables contain the coordinates and, optionally, the individual difference coefficients that are used in the plots. The Residual plots allow you to assess the fit of the model graphically. The iteration history contains information about how many iterations were needed and how the criterion changed over the iterations.

## Summary

Investigators in the field of market research are interested in how consumers make decisions when they choose to buy products. What attributes are important? Do all people make decisions in the same way? If not, how do they differ? What are the perceptual schemes that people use in their purchasing decisions?

The analyses described in this paper can be used with many different types of data to investigate these questions. The Market Research application makes these analyses easy to use, and it is available in Release 6.11 and subsequent releases with the SAS/STAT product.

## Acknowledgments

I would like to thank Greg Goodwin, Warren Kuhfeld, Julie LaBarr, Donna Sawyer, and Maura Stokes for their thoughtful comments on this paper.