# CHAPTER **S-4**

# Analyzing Nominal Data with Correspondence Analysis

### LEARNING OBJECTIVES

**Upon completing this chapter, you should be able to do the following:**

- Understand the basics of perceptual mapping with nonmetric data.
- Select between a decompositional or compositional approach.
- Explain correspondence analysis as a method of perceptual mapping.

## CHAPTER PREVIEW

In Chapter S-3, we discussed the traditional decompositional approaches (multidimensional scaling, MDS) to perceptual mapping, but what about compositional techniques? In the past, compositional approaches relied on traditional multivariate techniques such as discriminant and factor analysis. Recent developments, however, combine aspects of both methods and MDS to form potent new tools for perceptual mapping.

Correspondence analysis (CA) is a related perceptual mapping technique with similar objectives. CA infers the underlying dimensions that are evaluated as well as the positioning of objects, but it follows a quite different approach. First, instead of using overall evaluations of similarity or preference concerning the objects, each object is evaluated (in nonmetric terms) on a series of attributes. Then, based on this information, CA develops the dimensions of comparison between objects and places each object in this dimensional space to allow for comparisons among both objects and attributes simultaneously.

## KEY TERMS

Before starting the chapter, review the key terms to develop an understanding of the concepts and terminology used. Throughout the chapter the key terms appear in **boldface.** Other points of emphasis in the chapter and key term cross-references are *italicized.*

**Chi-square value** Method of analyzing data in a *contingency table* by comparing the actual cell frequencies to an expected cell frequency. The expected cell frequency is based on the marginal probabilities of its row and column (probability of a row and column among all rows and columns).

**Compositional method** An approach to perceptual mapping that derives overall *similarity* or *preference* evaluations from evaluations of separate attributes by each respondent. With compositional methods, separate attribute evaluations are combined (composed) into an overall evaluation. The most common examples of compositional methods are factor analysis and discriminant analysis.

**Contingency table** Cross-tabulation of two nonmetric or categorical variables in which the entries are the frequencies of responses that fall into each cell of the matrix. For example, if three brands were rated on four attributes, the brand-by-attribute contingency table would be a three-row by four-column table. The entries would be the number of times a brand (e.g., Coke) was rated as having an attribute (e.g., sweet taste).

**Correspondence analysis (CA)** *Compositional method* to perceptual mapping that is based on categories of a *contingency table.* Most applications involve a set of *objects* and attributes, with the results portraying both objects and attributes in a common *perceptual map.* A minimum of three attributes and three objects is required to derive a multidimensional map.

**Cross-tabulation table** See *contingency table.*

**Decompositional method** Perceptual mapping method associated with MDS techniques in which the respondent provides only an overall evaluation of *similarity* or preference between *objects.* This set of overall evaluations is then decomposed into a set of dimensions that best represent the objects' differences.

**Dimensions** Features of an *object.* A particular object can be thought of as possessing both perceived/subjective dimensions (e.g., expensive, fragile) and *objective* dimensions (e.g., color, price, features).

**Inertia** A relative measure of *chi-square* used in correspondence analysis. The total inertia of a *cross-tabulation table* is calculated as the total chi-square divided by the total frequency count (sum of either rows or columns). Inertia can then be calculated for any row or column category to represent its contribution to the total.

**Mass** A relative measure of frequency used in *correspondence analysis* to describe the size of any single cell or category in a *cross-tabulation.* It is defined as the value (cell or category total) divided by the total frequency count, making it the percentage of the total frequency represented by the value. As such, the total mass across rows, columns, or all cell entries is 1.0.

**Multiple correspondence analysis** Form of *correspondence analysis* that involves three or more categorical variables related in a common perceptual space.

**Object** Any stimulus that can be compared and evaluated by the respondent, including tangible entities (product or physical object), actions (service), sensory perceptions (smell, taste, sights), or even thoughts (ideas, slogans).

**Objective dimension** Physical or tangible characteristics of an *object* that have an objective basis of comparison. For example, a product has size, shape, color, weight, and so on.

**Perceptual map** Visual representation of a respondent's perceptions of *objects* on two or more *dimensions*. Usually this map has opposite levels of dimensions on the ends of the $X$ and $Y$ axes, such as "sweet" to "sour" on the ends of the $X$ axis and "high-priced" to "low-priced" on the ends of the $Y$ axis. Each object then has a spatial position on the perceptual map that reflects the relative *similarity* or *preference* to other objects with regard to the dimensions of the perceptual map.

**Similarity** Data used to determine which *objects* are the most similar to each other and which are the most dissimilar. Implicit in similarities measurement is the ability to compare all pairs of objects.

## WHAT IS PERCEPTUAL MAPPING?

Perceptual mapping is a set of techniques that attempt to identify the perceived relative image of a set of objects (firms, products, ideas, or other items associated with commonly held perceptions). The objective of any perceptual mapping approach is to use consumer judgments of similarity or preference to represent the objects (e.g., stores or brands) in a multidimensional space.

The three basic elements of any perceptual mapping process are defining objects, defining a measure of similarity, and establishing the dimensions of comparison. **Objects** can be any entity that can be evaluated by respondents. Although we normally think of tangible objects (e.g., products, people, etc.), an object can also be something very intangible (e.g., political philosophies, cultural beliefs, etc.).

The second aspect of perceptual mapping is the concept of **similarity**—a relative judgment of one object versus another. A key characteristic of similarity is that it can be defined differently by each respondent. All that is required is that the respondent can make a comparison between the objects and form some perception of similarity.

With the similarity judgments in hand, the perceptual mapping technique then forms **dimensions** of comparison. Dimensions are unobserved characteristics that allow for the objects to be arrayed in a multidimensional space (**perceptual map**) that replicates the respondent's similarity judgments. Similar to factors formed in factor analysis (see Chapter 3) because of their combination of a number of specific characteristics, here dimensions differ in that the underlying characteristics that "define" the dimensions are unknown.

Readers are encouraged to also read Chapter S-3 to gain a broader perspective on the overall perceptual mapping process and specific MDS techniques. Although there are quite specific differences between correspondence analysis and multidimensional scaling, they share a number of common objectives and approaches to perceptual mapping.

## PERCEPTUAL MAPPING WITH CORRESPONDENCE ANALYSIS

**Correspondence analysis (CA)** is an increasingly popular interdependence technique for dimensional reduction and perceptual mapping [1, 2, 3, 5, 8]. It is also known as optimal scaling or scoring, reciprocal averaging, or homogeneity analysis. When compared to MDS techniques, correspondence analysis has three distinguishing characteristics.

First, it is a **compositional method** rather than a decompositional approach, because the perceptual map is based on the association between objects and a set of descriptive characteristics or attributes specified by the researcher. A **decompositional method**, like the MDS techniques discussed in Chapter S-3, requires that the respondent provide a direct measure of similarity. Second, most applications of CA involve the correspondence of categories of variables, particularly those measured on nominal measurement scales. This correspondence is then the basis for developing perceptual maps. Finally, the unique benefits of CA lie in its abilities for simultaneously representing rows and columns, for example, brands and attributes, in joint space.

### Differences from Other Multivariate Techniques

Among the compositional techniques, factor analysis is the most similar because it defines composite dimensions (factors) of the variables (e.g., attributes) and then plots objects (e.g., products) on their scores on each dimension. In discriminant analysis, products can be distinguished by their profiles across a set of variables and plotted in a dimensional space as well. Correspondence analysis extends beyond either of these two compositional techniques:

- CA can be used with nominal data (e.g., frequency counts of preference for objects across a set of attributes) rather than metric ratings of each object on each object. This capability enables CA to be used in many situations in which the more traditional multivariate techniques are inappropriate.
- CA creates perceptual maps in a single step, where variables and objects are simultaneously plotted in the perceptual map based directly on the association of variables and objects. The relationships between objects and variables are the explicit objective of CA.

We first examine a simple example of CA to gain some perspective on its basic principles. Then we discuss each of the six stages of the decision-making process introduced in Chapter 1. The emphasis is on those unique elements of CA as compared to the decompositional methods of MDS in Chapter S-3.

## A SIMPLE EXAMPLE OF CA

Let us examine a simple situation as an introduction to CA. In its most basic form, CA examines the relationships between categories of nominal data in a **contingency table,** the cross-tabulation of two categorical (nonmetric) variables. Perhaps the most common form of contingency table would be cross-tabulating objects and attributes (e.g., most distinctive attributes for each product or product sales by demographic category). CA can be applied to any contingency table and portray a perceptual map relating the categories of each nonmetric variable in a single perceptual map.

In our simple example, we compare the product sales of three products across a single demographic variable (age). The cross-tabulated data (see Table S4-1) portray the sales figures for products A, B, and C broken down by three age categories (young adults, who are 18 to 35 years old; middle age, who are 36 to 55 years old; and mature individuals, who are 56 or older).

## Utilizing Cross-Tabulated Data

What can we learn from cross-tabulated data? First, we can look at the column and row totals to identify the ordering of categories (highest to lowest). But more important, we can view the relative sizes of each cell of the contingency table reflecting the amount of each variable for each object. Comparing the cells may identify patterns reflecting associations among certain objects and attributes.

Viewing Table S4-1, we see that product sales vary across products (product C has the highest total sales of 100 units; product B the lowest sales, 40 units) and age groups (middle-age group buys 90 units; the young adults, 40 units). To identify any pattern to the sales we need to be able to state that a certain group (e.g., young adults) buys more or less of a certain product. But to do this we need to have some way to define what the "expected" sales would be so that we can say that the actual sales amount is more or less. For example, the middle-age group purchases 40 units of both product A and C. Do we then assess that this group has equal preference for the two products, or should the fact that product C was generally a more popular product with 100 units of total sales than was product A, with only 80 total units, have some impact on how we judge the unit sales of the two products among this age group?

To identify these patterns of "more or less" for any cell in the table, we need two more elements to help in quantifying the amount of "more or less," as well as portraying it graphically. The first is a means of standardizing the cell counts to make them comparable and then to develop a means of portraying the values of each cell.

**STANDARDIZING FREQUENCY COUNTS** First is a standardized measure of the cell counts that simultaneously considers the differences in row and column totals. We can directly compare the cells when all the row and column totals are equal, which is rarely the case. Instead, the rows and column totals are usually unequal. In this case, we need a measure that compares each cell value to an expected value that reflects the specific row and column totals of that cell.

**TABLE S4-1 Cross-Tabulated Data Detailing Product Sales by Age Category**

| | *Product Sales* | | | |
|---|---|---|---|---|
| **Age Category** | **A** | **B** | **C** | **Total** |
| Young adults (18–35 years old) | 20 | 20 | 20 | 60 |
| Middle age (36–55 years old) | 40 | 10 | 40 | 90 |
| Mature individuals (56+ years old) | 20 | 10 | 40 | 70 |
| Total | 80 | 40 | 100 | 220 |

In our product sales example, let us examine the sales to the young adults group. As we can see, this group purchased equal amounts of each product (20 units each). But is this what we would expect? How can we tell if they actually prefer one product over another by purchasing it more when compared to all the other age groups? A simple way would be to calculate what we expect product sales to be in direct proportion to overall product sales across all the groups. Product A had 36 percent of total sales (36% = 80 ÷ 220), whereas product B had 18 percent and product C had 45 percent. We can then calculate an "expected" sales amount for each product by applying these percentages to the 60 unit total sales by the young adults group. This would give us an expected sales of just over 21 units for product A (36% × 60 units), while we would expect only about 11 units for product B and about 27 units for product C. We can now see that this group purchases slightly less than expected for both product A (21.8 units expected versus an actual of 20 units) and product C (27 units expected sales versus actual sales of 20 units), but purchases substantially more than expected for product B (actual sales of 20 units when expected sales are only 11 units).

This illustrates a simple way to calculate the amount of the differences (i.e., the amount that purchases are "more or less" than expected), but we still need a way to "standardize" the differences across age groups and products. As we will discuss later, the chi-square value (a variant of the process we just described) will be used as the standardized measure of comparison.

**PORTRAYING EACH CELL** Once we have a standardized measure of the differences between expected and actual values for each cell (representing a distinct combination of rows and columns), we then need a method for portraying each cell in a single perceptual map. The task is to portray all of the associations between the rows and columns (age groups and products in our example). We do this by assigning each category of the rows and columns a separate symbol. Then, when the standardized values are higher than expected for a cell (a specific row/column combination) we would expect that the symbols for that row and column would be located closer together, whereas the symbols for cells with standardized values much lower than expected would be more widely separated. The challenge is to develop a perceptual map that best portrays all of the associations represented by the cells of the contingency table.

In our product example, we would start with six symbols (three symbols for the age groups and three symbols for the products). We would then use the standardized measures of difference discussed earlier to help position the symbols to represent the associations between age groups and products. Going back to the young adult age group, we would expect the symbol for product B to be located closer to the young adult symbol than either product A or C, with C the farthest removed from the young adult symbol because it had the largest difference of actual sales being lower than expected sales. This process is repeated for all the age group/product combinations, and then the perceptual map is drawn so that the symbol positions best reflect the values of the standardized measure across all of the age group/product combinations.

In the following sections we discuss how CA calculates a standardized measure of association based on the cell counts of the contingency and then the process whereby these associations are converted into a perceptual map.

## Calculating a Measure of Association or Similarity

Correspondence analysis uses one of the most basic statistical concepts, chi-square, to standardize the cell frequency values of the contingency table and form the basis for association or similarity. **Chi-square** is a standardized measure of actual cell frequencies compared to expected cell frequencies. In cross-tabulated data, each cell contains the values for a specific row–column combination (e.g., sales of a specific product in a specific age group). Thus, the **chi-square value** is a measure of association between the row and column categories. Higher levels of association, just like higher levels of similarity, should be represented as closer together in the perceptual map than those with lower levels of association. Readers wishing to review the process of calculating chi-square values can refer to the Basic Stats appendix on the text's Web sites (accessed through cengagebrain.co.uk or www.mvstats.com).

The chi-square value can be easily transformed into a similarity measure. The process of calculating the chi-square (squaring the difference) removes the direction of the similarity. To restore the directionality, we use the sign of the original difference, but reverse it to make it more intuitive. This must be done because the difference in the chi-square calculation is defined as expected minus actual values. This makes negative differences represent those situations that we denote as greater association—where actual counts exceed expected counts. In order to make the similarity measure like those used in MDS (i.e., positive/larger values are greater association and negative/smaller values are less association) we reverse the sign of the original difference and apply it to the chi-square value. The result is a measure that acts just like the similarity measures used in earlier MDS examples. Negative values indicate less association (similarity) and positive values indicate greater association.

Table S4-2 illustrates the calculation of the chi-square value and its transformation into a similarity measure. For each cell, the actual and expected values are given along with the difference. Then the chi-square value is shown along with the transformed (signed) value. For example, the top row shows the actual sales for the young adults across the three products (20 units each) as well as the expected sales (21.82, 10.91, and 27.27 units, respectively). The differences for products A and C are positive, meaning that expected sales were greater than actual sales (a negative association), whereas the difference was positive for Product B. The chi-square values act as the standardized measure of difference. The final value in each cell is the signed chi-square value, where positive values represent greater similarity between the age group/product combination and negative values are lower similarity.

**TABLE S4-2 Calculating Chi-Square as Similarity Values for Cross-Tabulated Data**

| | *Product Sales* | | | |
|---|---|---|---|---|
| **Age Category** | **A** | **B** | **C** | **Total** |
| Young Adults | | | | |
| Actual sales | 20 | 20 | 20 | 60 |
| Expected sales[a] | 21.82 | 10.91 | 27.27 | 60 |
| Difference[b] | 1.82 | □ 9.09 | 7.27 | — |
| Chi-square value[c] | .15 | 7.58 | 1.94 | |
| Signed chi-square value[d] | □ .15 | 7.58 | □ 1.94 | |
| Middle Age | | | | |
| Actual sales | 40 | 10 | 40 | 90 |
| Expected sales | 32.73 | 16.36 | 40.91 | 90 |
| Difference | □ 7.27 | 6.36 | .91 | — |
| Chi-square value | 1.62 | 2.47 | .02 | |
| Signed chi-square value | 1.62 | □ 2.47 | □ .02 | |
| Mature Individuals | | | | |
| Actual sales | 20 | 10 | 40 | 70 |

| | | | | |
|---|---|---|---|---|
| Expected sales | 25.45 | 12.73 | 31.82 | 70 |
| Difference | 5.45 | 2.73 | −8.18 | — |
| Chi-square value | 1.17 | .58 | 2.10 | |
| Signed chi-square value | −1.17 | −58 | 2.10 | |
| Total | | | | |
| Sales | 80 | 40 | 100 | 220 |
| Expected sales | 80 | 40 | 100 | 220 |
| Difference | — | — | — | — |

[a]Expected sales = (Row total × Column total) ÷ Overall total

Example: $\text{Cell}_{\text{Young Adults, Product A}}$ = (60 × 80) ÷ 220 = 21.82

[b]Difference = Expected sales − Actual sales

Example: $\text{Cell}_{\text{Young Adults, Product A}}$ = 21.82 − 20.00 = 1.82

[c] $$\text{Chi-square value} = \frac{\text{Difference}^2}{\text{Expected sales}}$$

Example: $\text{Cell}_{\text{Young Adults, Product A}}$ = $1.82^2 \div 21.82$ = .15

[d]Signed chi-square value is the chi-square value with a reversed sign of the difference between expected and actual sales.

## Creating the Perceptual Map

The similarity (signed chi-square) values provide a standardized measure of association, much like the similarity judgments used in MDS methods. With this association/similarity measure, CA creates a perceptual map by using the standardized measure to estimate orthogonal dimensions upon which the categories can be placed to best account for the strength of association represented by the chi-square distances.

As done in MDS techniques, we first consider a lower-dimensional solution (e.g., one or two dimensions) and then expand the number of dimensions and continue until we reach the maximum number of dimensions. In CA, the maximum number of dimensions is one less than the smaller of the number of rows or columns.

Looking back at the signed chi-square values in Table 11-2, which act as a similarity measure, the perceptual map should place certain combinations with positive values closer together on the perceptual map (e.g., young adults close to product B, middle age close to product A, and mature individuals close to product C). Moreover, certain combinations should be farther apart given their negative values (young adults farther from product C, middle age farther from product B, and mature individuals farther from product A).

In our example, we can only have two dimensions (the smaller of the number of rows or columns minus one, or 3 □ 1 □ 2), so the two-dimensional perceptual map is as shown in Figure S4-1.

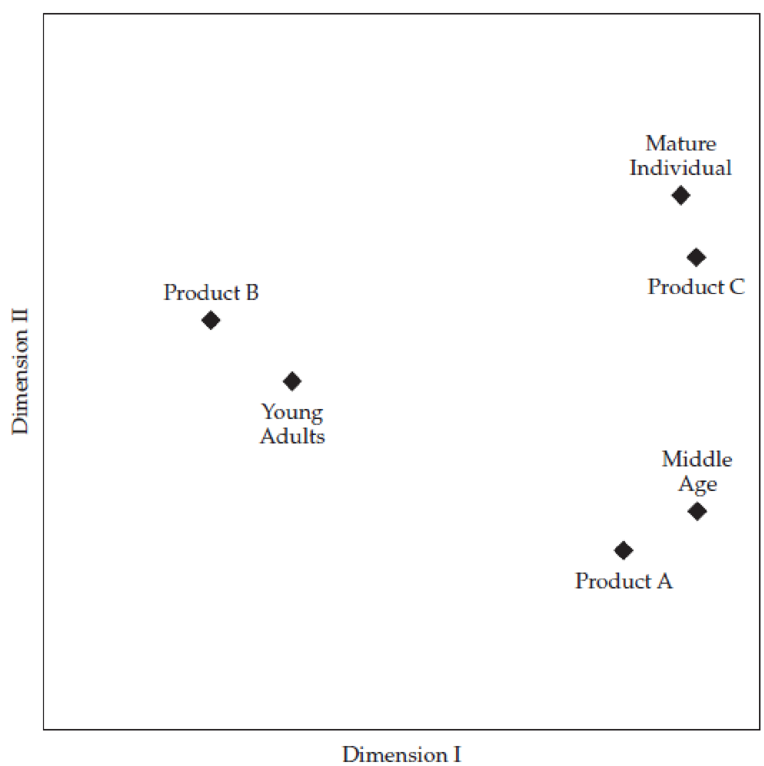


**FIGURE S4-1 Perceptual Map from Correspondence Analysis**

Corresponding to the similarity values just described, the positions of the age groups and products represent the positive and negative associations between age group and products very well. The researcher can examine the perceptual map to understand the product preferences among age groups based on their sales patterns.

In contrast to MDS, however, we can relate age groups with product positions. We can see that we now have an additional tool that can relate different characteristics (in this case products and the age characteristics of their buyers) in a single perceptual map. Just as easily, we could have related products to attributes, benefits, or any other characteristic. We should note, however, that in CA the positions of objects are now dependent on the characteristics they are associated with.

## A DECISION FRAMEWORK FOR CORRESPONDENCE ANALYSIS

Correspondence analysis and the issues associated with a successful analysis can be viewed through the model-building process introduced in Chapter 1. In the following sections we examine the unique issues associated with correspondence analysis methods across the six stages of the decision process.

## STAGE 1: OBJECTIVES OF CA

Researchers are constantly faced with the need to quantify the qualitative data found in nominal variables. CA differs from other MDS techniques in its ability to accommodate both nonmetric data and nonlinear relationships. It performs dimensional reduction similar to multidimensional scaling and a type of perceptual mapping, in which categories are represented in the multidimensional space. Proximity indicates the level of association among row or column categories. CA can address either of two basic objectives:

**1.** *Association among only row or column categories.* CA can be used to examine the association among the categories of just a row or column. A typical use is the examination of the categories of a scale, such as the Likert scale (five categories from "strongly agree" to "strongly disagree") or other qualitative scales (e.g., excellent, good, poor, bad). The categories can be compared to see whether two can be combined (i.e., they are in close proximity on the map) or whether they provide discrimination (i.e., they are located separately in the perceptual space).

**2.** *Association between both row and column categories.* In this application, interest lies is portraying the association between categories of the rows and columns, such as our example of product sales by age group. This use is most similar to the previous example of MDS and has propelled CA into more widespread use across many research areas.

The researcher must determine the specific objectives of the analysis because certain decisions are based on which type of objective is chosen. CA provides a multivariate representation of interdependence for nonmetric data not possible with other methods. With a compositional method, the researcher must be aware that the results are based on the descriptive characteristics (e.g., object attributes or respondent characteristics) of the objects included in the analysis. This contrasts with decompositional MDS procedures in which only the overall measure of similarity between objects is needed.

## STAGE 2: RESEARCH DESIGN OF CA

Correspondence analysis requires only a rectangular data matrix (cross-tabulation) of nonnegative entries. The most common type of input matrix is a contingency table with specific categories defining the rows and columns. In creating the table, several issues emerge, relating to the nature of the variables and categories comprising the rows and columns:

**1.** The rows and columns do not have predefined meanings (i.e., attributes do not always have to be rows, and so on) but instead represent the responses to one or more categorical variables. The categories in both rows and columns, however, must have specific meaning for interpretation purposes.

2. The categories for a row or column need not be a single variable but can represent any set of relationships. A prime example is the "pick any" method [6, 7], in which respondents are given a set of objects and characteristics. A common application is where a set of objects (e.g., products) are rated on a set of characteristics (e.g., attributes). Here the rows can be the individual products and each attribute is a separate column. The respondents then indicate which objects, if any, are described by each characteristic. The respondent may choose any number of objects for each characteristic and the **cross-tabulation table** is the total number of times each object was described by each characteristic.
3. The cross-tabulation may occur for more than two variables in a multiway matrix form. In these cases, **multiple correspondence analysis** is employed. In a procedure quite similar to two-way analysis, the additional variables are fitted so that all the categories are placed in the same multidimensional space.

The generalized nature of the types of relationships that can be portrayed in the contingency table makes CA widely applicable. Its increased usage in recent years is a direct result of the continued development of approaches for using this format for analyzing new types of relationships.

## STAGE 3: ASSUMPTIONS OF CA

Correspondence analysis shares with the more traditional MDS techniques a relative freedom from assumptions. The use of strictly nonmetric data in its simplest form (cross-tabulated data) represents linear and nonlinear relationships equally well. The lack of assumptions, however, must not cause the researcher to neglect the efforts to ensure the comparability of objects and, because it is a compositional technique, the completeness of the attributes used.

## STAGE 4: DERIVING CA RESULTS AND ASSESSING OVERALL FIT

With a cross-tabulation table, the frequencies for any row–column combination of categories are related to other combinations based on the marginal frequencies. As depicted in our simple example, correspondence analysis uses this basic relationship in three steps to create a perceptual map:

1. Calculate a conditional expectation (the expected cell count) that represents the similarity or association between row and column categories.
2. Once obtained, compute the differences between the expected and actual cell counts and convert them into a standardized measure (chi-square). Using these results as a distance metric makes them comparable to the input matrices used in the MDS approaches discussed earlier.
3. Using estimation techniques similar to MDS to convert the measure of similarity between categories (i.e., the signed chi-square), a series of dimensional solutions (one-dimensional, two-dimensional, etc.) is created where possible. The dimensions simultaneously relate the rows and columns in a single joint plot. The result is a representation of categories of rows and/or columns (e.g., brands and attributes) in the same plot.

### Determining Impact of Individual Cells

It should be noted that two specific terms, developed in correspondence analysis, describe the properties of the frequency values and their relative contribution to the analysis:

- The first term is **mass**, which is first defined for any single entry in the cross-tabulation table as the percentage of the total represented by that entry. It is calculated as the value of any single entry divided by *N* (the total for the table, which equals the sum of either the rows or columns). Thus, the sum of all table entries (cells) equals 1.0. We can also calculate the mass of any row or column category by summing across all entries. This result represents the contribution of any row or column category to the total mass.

- The second measure is **inertia**, which is defined as the total chi-square divided by $N$ (the total of the frequency counts). In this way, we have a relative measure of chi-square that can be related to any frequency count.

With these similarities to MDS comes a similar set of problems, focused on two primary issues in assessing overall fit: assessing the relative importance of the dimensions and then identifying the appropriate number of dimensions. Each of these issues is discussed in the following section.

**ASSESSING THE NUMBER OF DIMENSIONS** Eigenvalues, also known as singular values, are derived for each dimension and indicate the relative contribution of each dimension in explaining the variance in the categories. Similar to factor analysis, we can determine the amount of explained variance both for individual dimensions and the solution as a whole. Some programs, such as those of SPSS, calculate the inertia for a dimension, which also measures explained variation and is directly related to the eigenvalue.

The maximum number of dimensions that can be estimated is one less than the smaller of the number of rows or columns. For example, with six columns and eight rows, the maximum number of dimensions would be five, which is six (the number of columns) minus one.

The researcher selects the number of dimensions based on the overall level of explained variance desired and the incremental explanation gained by adding another dimension. In assessing dimensionality, the researcher is faced with trade-offs, much as with other MDS solutions or even factor analysis (Chapter 3):

- Each dimension added to the solution increases the explained variance of the solution, but at a decreasing amount (i.e., the first dimension explains the most variance, the second dimension the second greatest, etc.).
- Adding dimensions increases the complexity of the interpretation process; perceptual maps of

greater than three dimensions become increasing complex to analyze.

The researcher must balance the desire for increased explained variance versus the more complex solution that may affect interpretation. A rule of thumb is that dimensions with inertia (eigenvalues) greater than .2 should be included in the analysis.

### Model Estimation

A number of computer programs are available to perform correspondence analysis. Among the more popular programs are ANACOR and HOMALS, available with SPSS; PROC CORRESP in SAS; CORRESP from NewMDSX [25]; CORRAN and CORRESP from PC-MDS [10]; and MAPWISE [9]. A large number of specialized applications have emerged in specific disciplines such as ecology, geology, and many of the social sciences.

## STAGE 5: INTERPRETATION OF THE RESULTS

Once the dimensionality has been established, the researcher is faced with two tasks: interpreting the dimensions to understand the basis for the association among categories and assessing the degree of association between categories, either within a row/column or between rows and columns. In doing so, the researcher gains an understanding of the underlying dimensions upon which the perceptual map is based along with the derived association of any specific set of categories.

### Defining the Character of the Dimensions

If the researcher is interested in defining the character of one or more dimensions in terms of the row or column categories, descriptive measures in each software program indicate the association of each category with a specific dimension. For example, in SPSS the inertia measure (used to assess the degree of explained variance) is decomposed across the dimensions. Similar in character to factor loadings, these values represent the extent of association for each category individually with each dimension. The researcher can then name each **objective dimension** in terms of the categories most associated with that dimension.

In addition to representing the association of each category with each dimension, the inertia values can be totaled across dimensions in a collective measure. In doing so, we gain an empirical measure of the degree to which each category is represented across all dimensions. In concept, this measure is similar to the communality measure of factor analysis (see Chapter 3).

### Assessing Association Among Categories

The second task in interpretation is to identify a category's association with other categories, which can be done visually or through empirical measures. No matter which approach is used, the researcher must first select the types of comparison to be made and then the appropriate normalization for the selected comparison. The two types of comparison are:

1. *Between categories of the same row or column.* Here the focus is on only rows or columns, such as when examining the categories of a scale to see whether they can be combined. These types of comparisons can be made directly from any correspondence analysis.
2. *Between rows and columns.* An attempt to relate the association between a row category and a column category. This type of comparison, which is the most common, relates categories across dimensions (e.g., in our earlier example, product sales most associated with age categories). At this time, however, some debate centers on the appropriateness of comparing between row and column categories. In a strict sense, distances between points representing categories can only be made within a row or a column. It is deemed inappropriate to directly compare a row and column category. It is appropriate to make generalizations regarding the dimensions and each category's position on those dimensions. Thus, the relative positioning of row and column categories can be defined within those dimensions, but they should not be directly compared.

   Some computer programs provide for a normalization procedure to allow for this direct comparison. If only row or column normalization is available, alternative procedures are proposed to make all categories comparable [2, 9], yet disagreement still remains as to their success [4]. In the cases for which direct comparisons are not possible, the general correspondence still holds and specific patterns can be distinguished.

Research objectives may focus on either evaluation of the dimensions or comparison of the categories, and the researcher is encouraged to make both interpretations because they reinforce each other. For example, comparing row versus column categories can always be complemented with an understanding of the nature of the dimensions to provide a more comprehensive perspective on the positioning of categories rather than just the specific comparisons. Likewise, evaluating the specific category comparisons can provide specificity to the interpretation of the dimensions.

## STAGE 6: VALIDATION OF THE RESULTS

The compositional nature of correspondence analysis provides more specificity for the researcher with which to validate the results. In doing so, the researcher should strive to assess two key questions concerning generalizability of two elements:

- *Sample.* As with all MDS techniques, an emphasis must be made to ensure generalizability through split- or multisample analyses.
- *Objects.* The generalizability of the objects (represented individually and as a set by the categories) must also be established. The sensitivity of the results to the addition or deletion of a category can be evaluated. The goal is to assess whether the analysis is dependent on only a few objects and/or attributes.

---

**RULES OF THUMB S4-1**

**Correspondence Analysis**

- Correspondence analysis (CA) is best suited for exploratory research and is not appropriate for hypothesis testing

- CA is a form of compositional technique that requires specification of both objects and attributes to be compared
- Correspondence analysis is sensitive to outliers, which should be eliminated prior to using the technique
- The number of dimensions to be retained in the solution is based on:
  - Dimensions with inertia (eigenvalues) greater than .2
  - Enough dimensions to meet the research objectives (usually two or three)
- Dimensions can be "named" based on the decomposition of inertia measures across a dimension:
  - These values show the extent of association for each category individually with each dimension
  - They can be used for description much like loadings in factor analysis

---

In either instance, the researcher must understand the true meaning of the results in terms of the categories being analyzed. The inferential nature of correspondence analysis, like other MDS methods, requires strict confidence in the representativeness and generalizability of the sample of respondents and the objects (categories) being analyzed.

## OVERVIEW OF CORRESPONDENCE ANALYSIS

Correspondence analysis presents the researcher with a number of advantages, ranging from the generalized nature of the input data to development of unique perceptual maps:

- The simple cross-tabulation of multiple categorical variables, such as product attributes versus brands, can be represented in a perceptual space. This approach enables the researcher either to analyze existing responses or to gather responses at the least restrictive measurement type, the nominal or categorical level. For example, the respondent need rate only yes or no for a set of objects on a number of attributes. These responses can then be aggregated in a cross-tabulation table and analyzed. Other techniques, such as factor analysis, require interval ratings of each attribute for each object.

- CA portrays not only the relationships between the rows and columns, but also the relationships between the categories of either the rows or the columns. For example, if the columns were attributes, multiple attributes in close proximity would all have similar profiles across products, forming a group of attributes quite similar to a factor from principal components analysis.
- CA can provide a joint display of row and column categories in the same dimensionality. Certain program modifications allow for interpoint comparisons in which relative proximity is directly related to higher association among separate points [1, 9]. When these comparisons are possible, they enable row and column categories to be examined simultaneously. An analysis of this type would enable the researcher to identify groups of products characterized by attributes in close proximity.

With the advantages of CA, however, come a number of disadvantages or limitations.

- The technique is descriptive and not at all appropriate for hypothesis testing. If the quantitative relationship of categories is desired, methods such as log-linear models are suggested. CA is best suited for exploratory data analysis.
- CA, as is the case with many dimensionality-reducing methods, has no method for conclusively determining the appropriate number of dimensions. As with similar methods, the researcher must balance interpretability versus parsimony of the data representation.
- The technique is quite sensitive to outliers, in terms of either rows or columns (e.g., attributes or brands). Also, for purposes of generalizability, the problem of omitted objects or attributes is critical.

Overall, correspondence analysis provides a valuable analytical tool for a type of data (nonmetric) normally not the focal point of multivariate techniques. Correspondence analysis also provides the researcher with a complementary compositional technique to MDS for addressing issues where direct comparison of objects and attributes is preferable.

## ILLUSTRATION OF CORRESPONDENCE ANALYSIS

An alternative to attribute-free perceptual mapping is correspondence analysis (CA), a compositional method based on nonmetric measures (frequency counts) between objects and/or attributes. In this attribute-based method, the perceptual map is a joint space, showing both attributes and firms in a single representation. Moreover, the positions of firms are relative not only to the other firms included in the analysis, but also to the attributes selected.

To demonstrate the use of CA, we examine data HBAT gathered in a series of interviews with company representatives from a cross-section of potential customers. This is the same sample used to gather data for the MDS techniques discussed in Chapter S-3 although different types of data are used in each technique. Many of the issues, particularly in the initial stages, have a great deal of overlap between CA and MDS. Data sets for both techniques are available on the text's Web site (accessed through cengagebrain.co.uk and www.mvstats.com).

### Stage 1: Objectives of Perceptual Mapping

The primary purpose of perceptual mapping is to understand how a firm's image compares to other firms in the market. In this case, the result is a perceptual map of HBAT and the major competitors in the market. With CA, more emphasis is placed on assessing the dimensions of evaluation because a firm's position in the perceptual map can be compared on both the dimensions and specific attributes.

A critical decision for any perceptual mapping analysis is the selection of the objects to be compared. Although direct similarity judgments are not made in CA, the inclusion or exclusion of objects can have a marked impact. For example, excluding a firm with distinguishing characteristics unique to other firms may help reveal firm-to-firm comparisons or even dimensions not otherwise detected. Likewise, the exclusion of distinctive or otherwise relevant firms may affect the results in a similar manner.

In our example, the objects of study are HBAT and its nine major competitors. To understand the perceptions of these competing firms, mid-level executives of firms representing potential customers are surveyed on their perceptions of HBAT and the competing firms. The resulting perceptual maps hopefully portray HBAT's positioning in the marketplace.

### Stage 2: Research Design of the CA Study

With the objectives defined for the perceptual mapping analyses, HBAT researchers must next address a set of decisions focusing on research design issues that define the methods used and the specific firms to be studied. By doing so, they also define the types of data that need to be collected to perform the desired analyses. Each of these issues is discussed in the following section.

**SELECTING FIRMS FOR ANALYSIS** In selecting firms for analysis, the researcher must address two issues. First, are all of the firms comparable and relevant for the objectives of this study? Second, is the number of firms included enough to portray the dimensionality desired? The design of the research to address each issue is discussed here.

This study includes nine competitors, plus HBAT, representing all of the major firms in this industry and collectively having more than 85 percent of total sales. Moreover, they are considered representative of all of the potential segments existing in the market. All of the remaining firms not included in the analysis are considered secondary competitors to one or more of the firms already included.

**COLLECTING ATTRIBUTE DATA** A unique characteristic of correspondence analysis is the use of nonmetric data to portray relationships between categories (objects or attributes). A common approach to data presentation is the use of a cross-tabulation matrix relating the attributes (represented as rows) to the ratings of objects/firms (the columns). The values represent the number of times each firm is rated as being characterized by that attribute. Thus, higher frequencies indicate a stronger association between that object and the attribute in question. Nonmetric ratings were gathered by asking each respondent to pick the firms best characterized by each attribute. As with the "pick any" method [6, 7], the respondent could pick any number of firms for each attribute.

The HBAT image study is composed of in-depth interviews with 18 mid-level management personnel from different firms. From the research objectives, the primary goal is to understand the similarities of firms based on firms' attributes. Eight of the 10 attributes identified as composing the four factors in Chapter 3 were selected for this study. The eight attributes included were $X_6$, Product Quality; $X_8$, Technical Support; $X_{10}$, Advertising; $X_{12}$, Salesforce Image; $X_{13}$, Competitive Pricing; $X_{14}$, Warranty & Claims; $X_{16}$, Order & Billing; and $X_{18}$, Delivery Speed. Two of the attributes from the original set of 10 were eliminated in this analysis. First, $X_7$, relating to E-Commerce, was not used because about one-half of the firms did not have an e-commerce presence. Also, $X_9$, Complaint Resolution, which is largely experience-based, was also omitted because evaluation by noncustomers was difficult for the respondents.

In the HBAT study, binary firm ratings were gathered for each firm on each of the eight attributes (i.e., a yes–no ratings of each firm on each attribute). The individual entries in the cross-tabulation table are the number of times a firm is rated as possessing a specific attribute. Respondents could choose any number of attributes as characterizing each firm. Table 11-3 contains the responses for a single individual in the HBAT study and then provides the complete cross-tabulation table for all respondents. The HBAT_CORRESP_INDIV data set contains the responses for the individuals and the HBAT_CORRESP data set has the final cross-tabulation data set used in the analysis.

## Stage 3: Assumptions in Perceptual Mapping

The assumptions of CA deal primarily with the comparability and representativeness of the objects being evaluated and the respondents. The techniques themselves place few limitations on the data, but their success is based on several characteristics of the data.

With regard to the sample, the sampling plan emphasized obtaining a representative sample of HBAT customers. Moreover, care was taken to obtain respondents of comparable position and market knowledge. Because HBAT and the other firms serve a fairly distinct market, all the firms evaluated in the perceptual mapping should be known, ensuring that positioning discrepancies can be attributed to perceptual differences among respondents.

## Stage 4: Estimating a Correspondence Analysis

The data preparation and estimation procedure for correspondence analysis is similar in some regards to the multidimensional scaling process, with some notable exceptions. In the following sections we discuss the issues involved in calculating similarity and determining the dimensionality of the solution. The Correspondence Analysis procedure in SPSS was used to perform this analysis. Note that correspondence analysis is available in a number of more specialized programs, as discussed earlier.

**TABLE S4-3 Individual Respondent Ratings and Cross-Tabulated Frequency Data of Attribute Descriptors for HBAT and Nine Competing Firms**

RESPONDENT # 1

| | *Firm* | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Variables** | **HBAT** | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** | **I** |
| $X_6$ Product Quality | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 |
| $X_8$ Technical Support | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 |
| $X_{10}$ Advertising | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 |
| $X_{12}$ Salesforce Image | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 |
| $X_{13}$ Competitive Pricing | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 |
| $X_{14}$ Warranty & Claims | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 |
| $X_{16}$ Order & Billing | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 |
| $X_{18}$ Delivery Speed | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |

OVERALL TOTAL

| | *Firm* | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Variables** | **HBAT** | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** | **I** |
| $X_6$ Product Quality | 6 | 6 | 14 | 10 | 11 | 8 | 7 | 4 | 14 | 4 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $X_8$ Technical Support | 15 | 18 | 9 | 2 | 3 | 15 | 16 | 7 | 8 | 8 |
| $X_{10}$ Advertising | 15 | 16 | 15 | 11 | 11 | 14 | 16 | 12 | 14 | 14 |
| $X_{12}$ Salesforce Image | 4 | 3 | 1 | 13 | 9 | 6 | 3 | 18 | 2 | 10 |
| $X_{13}$ Competitive Pricing | 15 | 14 | 6 | 4 | 4 | 15 | 14 | 13 | 7 | 13 |
| $X_{14}$ Warranty & Claims | 7 | 18 | 13 | 4 | 9 | 16 | 14 | 5 | 4 | 16 |
| $X_{16}$ Order & Billing | 14 | 14 | 10 | 11 | 11 | 14 | 12 | 13 | 10 | 14 |
| $X_{18}$ Delivery Speed | 16 | 13 | 8 | 13 | 9 | 17 | 15 | 16 | 6 | 12 |

**CALCULATING THE SIMILARITY MEASURE** Correspondence analysis is based on a transformation of the chi-square value into a metric measure of distance, which acts as the similarity measure. The chi-square value is calculated as the actual frequency of occurrence minus the expected frequency. Thus, a negative value indicates, in this case, that a firm was rated less often than would be expected. The expected value for a cell (any firm–attribute combination in the cross-tabulation table) is based on how often the firm was rated on other attributes and how often other firms were rated on that attribute. (In statistical terms, the expected value is based on the row [attribute] and column [firm] marginal probabilities.)

Table S4-4 contains the transformed (metric) chi-square distances for each cell of cross-tabulation from Table S4-3. High positive values indicate a strong degree of correspondence between the attribute and firm, and negative values have the opposite interpretation. For example, the high values for HBAT and firms A and F with the technical support attribute ($X_8$) indicate that they should be located close together on the perceptual map if possible. Likewise, the high negative values for firms C and D on the same variable would indicate that their position should be far from the attribute's location.

**DETERMINING THE DIMENSIONALITY OF THE SOLUTION** Correspondence analysis tries to satisfy all of these relationships simultaneously by producing dimensions representing the chi-square distances. To determine the dimensionality of the solution, the researcher examines the cumulative percentage of variation explained, much as in factor analysis, and determines the appropriate dimensionality. The researcher balances the desire for increased explanation in adding additional dimensions versus interpretability by creating more complexity with each dimension added.

**TABLE S4-4 Measures of Similarity in Correspondence Analysis Chi-Square Distances**

| | | *Firm* | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Variables** | | **HBAT** | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** | **I** |
| $X_6$ | Product Quality | □ 1.02 | □ 1.28 | 2.37 | 1.27 | 1.71 | □ .73 | □ .83 | □ 1.59 | 2.99 | □ 1.66 |
| $X_8$ | Technical Support | 1.24 | 1.69 | □ .01 | □ 2.14 | □ 1.76 | .72 | 1.32 | □ 1.07 | .10 | □ .85 |
| $X_{10}$ | Advertising | .02 | □ .13 | .76 | □ .01 | .04 | □ .73 | .07 | □ .60 | 1.07 | □ .20 |
| $X_{12}$ | Salesforce Image | □ 1.27 | □ 1.83 | □ 2.08 | 3.19 | 1.53 | □ .86 | □ 1.73 | 4.07 | □ 1.42 | .97 |
| $X_{13}$ | Competitive Pricing | 1.08 | .40 | □ 1.10 | □ 1.52 | □ 1.48 | .57 | .59 | .65 | □ .36 | .53 |
| $X_{14}$ | Warranty & Claims | □ 1.32 | □ 1.49 | 1.15 | □ 1.54 | .23 | .81 | .55 | □ 1.80 | □ 1.44 | 1.39 |
| $X_{16}$ | Order & Billing | .19 | □ .19 | □ .30 | .37 | .42 | □ .30 | □ .54 | .08 | .20 | .23 |
| $X_{18}$ | Delivery Speed | .68 | □ .51 | □ .95 | .95 | □ .27 | .40 | .20 | .86 | □ 1.15 | □ .37 |

Table S4-5 contains the eigenvalues and cumulative and explained percentages of variation for each dimension up to the maximum of seven. A two-dimensional solution in this situation explains 86 percent of the variation, whereas increasing to a three-dimensional solution adds only an additional 10 percent. In comparing the additional variance explained in relation to the increased complexity in interpreting the results, a two-dimensional solution was deemed adequate for further analysis.

## Stage 5: Interpreting CA Results

With the number of dimensions defined, the researcher must proceed with an interpretation of the derived perceptual map. In doing so, at least three issues must be addressed: positioning of row and/or column categories, characterization of the dimensions, and assessing the goodness-of-fit of individual categories. Each will be discussed in the following sections.

**RELATIVE POSITIONING OF CATEGORIES** The first task is to assess the relative positions of the categories for the rows and columns. In doing so, the researcher can assess the association between categories in terms of their proximity in the perceptual map. Note that the comparison should only be between categories within the same row or column.

**TABLE S4-5 Determining the Appropriate Dimensionality in Correspondence Analysis**

| Dimension | Eigenvalue (Singular Value) | Inertia (Normalized Chi-Square) | Percentage Explained | Cumulative Percentage |
|---|---|---|---|---|
| 1 | 27666 | 07654 | 53.1 | 53.1 |
| 2 | 21866 | 04781 | 33.2 | 86.3 |
| 3 | 12366 | 01529 | 10.6 | 96.9 |
| 4 | 05155 | 00266 | 1.8 | 98.8 |
| 5 | 02838 | 00081 | 6 | 99.3 |
| 6 | 02400 | 00058 | 4 | 99.7 |
| 7 | 01951 | 00038 | 3 | 100.0 |

The perceptual map shows the relative proximities of both firms and attributes (see Figure S4-2). If we focus on the firms first, we see that the pattern of firm groups is similar to that found in the MDS results. Firms A, E, F, and I, plus HBAT form one group; firms C and D and H and B form two other similar groups. However, the relative proximities of the members in each group differ somewhat from the MDS solution. Also, firm G is more isolated and distinct, and firms F and E are now seen as more similar to HBAT.

In terms of attributes, several patterns emerge. First, $X_6$ and $X_{13}$, the two variables that are negatively related, appear at opposite extremes of the perceptual map. Moreover, variables shown to have high association (e.g., forming factors) fall in close proximity ($X_{16}$ and $X_{18}$, $X_8$, and $X_{14}$). Perhaps a more appropriate perspective is an attribute's contribution to each dimension, as discussed in the next section.

**INTERPRETING THE DIMENSIONS** It may be helpful to interpret the dimensions if row or column normalizations are used. For these purposes, the inertia (explained variation) of each dimension can be attributed among categories for rows and columns.

Table S4-6 provides the contributions of both sets of categories to each dimension. For the attributes, we can see that $X_{12}$ (Salesforce Image) is the primary contributor to dimension I, and $X_8$ (Technical Support) is a secondary contributor. Note that these two attributes are extreme in terms of their location on dimension I (i.e., highest or lowest values on dimension I). Between these two attributes, 86 percent of dimension I is accounted for. A similar pattern follows for dimension II, for which $X_6$ (Product Quality) is the primary contributor, followed by $X_{13}$ (Competitive Pricing), which when combined account for 83 percent of the inertia of dimension II. If we shift our focus to the 10 firms, we see a somewhat more balanced situation, with three firms (A, C, and G) contributing above the average of 10 percent. For the second dimension, four firms (B, D, G, and H) have contributions above average.

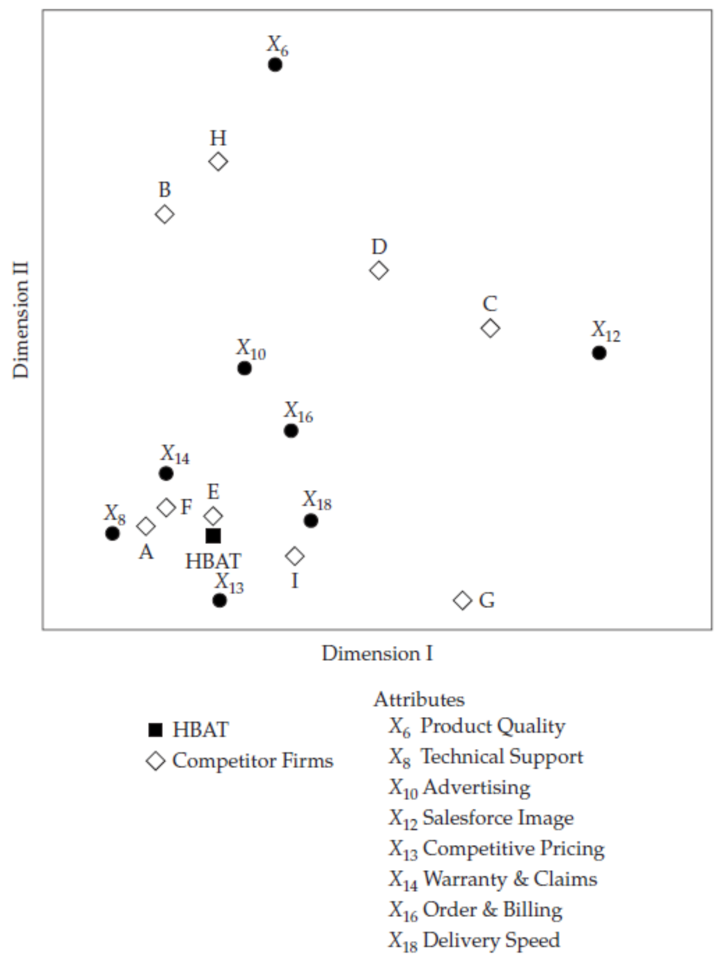


**FIGURE S4-2 Perceptual Mapping with Compositional Methods: Correspondence Analysis**

Although the comparisons in this example are between both sets of categories and are not restricted to a single set of categories (either row or column), these measures of contribution demonstrate the ability to interpret the dimension when desired.

**ASSESSING FIT FOR CATEGORIES** One final measure provides an assessment of fit for each category. Comparable to squared factor loadings in factor analysis (see Chapter 3 for a more detailed discussion), these values represent the amount of variation in the category accounted for by the dimension. The total value represents the total amount of variation across all dimensions, with the maximum possible being 100 percent.

Table S4-6 also contains fit values for each category on each dimension. As we can see, the total fit values range from a high of 99.1 for $X_6$ (Product Quality) and $X_{12}$ (Salesforce Image) to a low of .372 for $X_{14}$ (Warranty & Claims). Among the attributes, only $X_{14}$ has a value below 50 percent and only two firms (HBAT and firm I) fall below this value. Even though these values are somewhat low, they still represent a substantial enough explanation to retain them in the analysis and deem the analysis of sufficient practical significance.

**TABLE S4-6  Interpreting the Dimensions and Their Correspondence to Firms and Attributes**

| | *Coordinates* | | *Contribution to Inertia*[a] | | *Explanation by Dimension (Fit)*[b] | | |
|---|---|---|---|---|---|---|---|
| **Object** | **I** | **II** | **I** | **II** | **I** | **II** | **Total** |
| **Attribute** | | | | | | | |
| $X_6$ Product Quality | .044 | 1.235 | .001 | .689 | .002 | .989 | .991 |
| $X_8$ Technical Support | □.676 | □.285 | .196 | .044 | .789 | .111 | .901 |
| $X_{10}$ Advertising | □.081 | .245 | .004 | .045 | .093 | .678 | .772 |
| $X_{12}$ Salesforce Image | 1.506 | 0.298 | .665 | .033 | .961 | .030 | .991 |
| $X_{13}$ Competitive Pricing | □.202 | □.502 | .018 | .142 | .138 | .677 | .816 |
| $X_{14}$ Warranty & Claims | □.440 | □.099 | .087 | .006 | .358 | .014 | .372 |
| $X_{16}$ Order & Billing | .115 | .046 | .007 | .001 | .469 | .058 | .527 |
| $X_{18}$ Delivery Speed | .204 | □.245 | .022 | .040 | .289 | .330 | .619 |
| **Firm** | | | | | | | |
| HBAT | □.247 | □.293 | .024 | .042 | .206 | .228 | .433 |
| A | □.537 | □.271 | .125 | .040 | .772 | .156 | .928 |
| B | □.444 | .740 | .063 | .224 | .294 | .648 | .942 |
| C | 1.017 | .371 | .299 | .050 | .882 | .093 | .975 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| D | .510 | .556 | .074 | .111 | .445 | .418 | .863 |
| E | □.237 | □.235 | .025 | .031 | .456 | .356 | .812 |
| F | □.441 | 38□.209 | .080 | .023 | .810 | .144 | .954 |
| G | .884 | □.511 | .292 | .123 | .762 | .201 | .963 |
| H | □.206 | .909 | .012 | .289 | .049 | .748 | .797 |
| I | .123 | □.367 | .006 | .066 | .055 | .390 | .446 |

[a]Proportion of dimension's inertia attributable to each category.

[b]Proportion of category variation accounted for by dimension.

**OVERVIEW OF CA** These and other comparisons highlight the differences between MDS and CA methods and their results. CA results provide a means for directly comparing the similarity or dissimilarity of firms and the associated attributes, whereas MDS allows only for the comparison of firms. However, the CA solution is conditioned on the set of attributes included. It assumes that all attributes are appropriate for all firms and that the same dimensionality applies to each firm. Thus, the resulting perceptual map should always be viewed only in the context of both the firms and attributes included in the analysis.

Correspondence analysis is a quite flexible technique that is applicable to a wide range of issues and situations. The advantages of the joint plot of attributes and objects must always be weighed against the inherent interdependencies that exist and the potentially biasing effects of a single inappropriate attribute or firm, or perhaps more important, the omitted attribute of a firm. Yet CA still provides a powerful tool for gaining managerial insight into the relative position of firms and the attributes associated with those positions.

## Stage 6: Validation of the Results

Perhaps the strongest internal validation of this analysis is to assess the convergence between the results from the separate decompositional and compositional techniques. Each technique employs different types of consumer responses, but the resulting perceptual maps are representations of the same perceptual space and should correspond. If the correspondence is high, the researcher can be assured that the results reflect the problem as depicted. The researcher should note that this type of convergence does not address the generalizability of the results to other objects or samples of the population.

The decompositional perceptual map from Chapter S-3 and the CA results (Figure S4-2) are shown in Figure S4-3. The comparison will examine the relative positioning of objects and interpret the axes. First, let us examine the positioning of firms. When the perceptual maps in Figures S4-3 are rotated to obtain the same perspective, they both show two groups of firms: firms B, H, D, and C versus firms E, F, G, and I. Although the relative proximity varies between maps, we see a consistent relationship for HBAT being associated with firms A and I in each perceptual map. Given that the objective of CA is to define firm positions as a result of differences, it will generate more distinctiveness in its perceptual maps.

The interpretation of axes and distinguishing characteristics also shows similar patterns in the two perceptual maps. For the perceptual map from MDS, we should rotate the axes to obtain a clearer interpretation. Dimension I becomes associated with customer service and product value ($X_6$, $X_{13}$, $X_{16}$, and $X_{18}$), whereas dimension II reflects marketing and technical support ($X_8$, $X_{10}$, and $X_{12}$). The remaining attributes are not associated strongly with either axis. In comparing to CA we must reorient the axes because the dimensions flip between the two analyses. The firm groupings remain essentially the same, but they are in different positions on the perceptual map. In CA, the dimensions reflect somewhat the same elements, with the highest loadings being $X_{18}$ (Delivery Speed) on dimension I and $X_{12}$ (Salesforce Image) on dimension II. This compares quite favorably with the decompositional results except that the other attributes are somewhat more diffused on the dimensions.

Although some differences do exist, the similarity in the two results does provide some internal validity to the perceptual maps. Some attributes may have perceptual differences, but the overall patterns of firm positions and evaluative dimensions are supported by both approaches. As the comparison shows, MDS and CA are complementary methods in the understanding of consumer perceptions. MDS determines position based on overall judgments, whereas CA positions firms according to the selected set of attributes. These differences do not make either approach better or optimal but instead must be understood by the researcher to ensure selection of the method most suited to the research objectives.

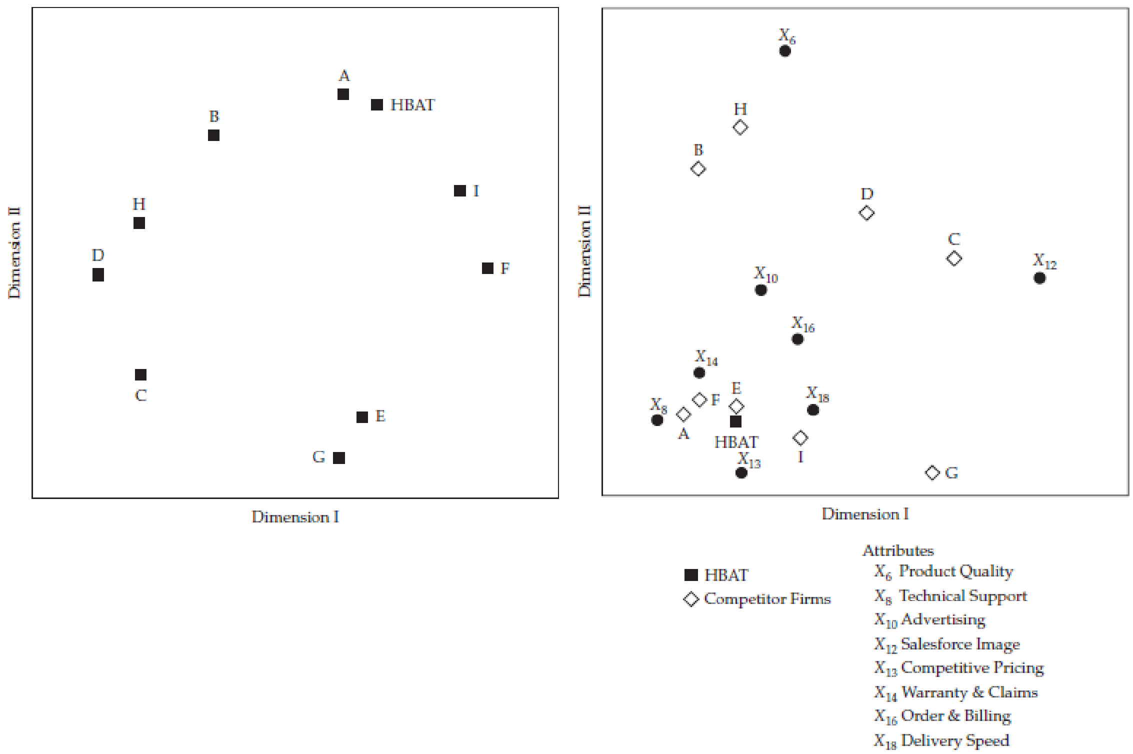


**FIGURE S4-3 A Comparison of Perceptual Maps from Correspondence Analysis and MDS**

## Summary

Correspondence analysis is a form of perceptual mapping that identifies relationships between objects based on the respondents' evaluations. As a compositional approach it differs from the MDS techniques described in Chapter S3, but it still results in a perceptual map that represents the relative position of objects and attributes. This chapter helps you to do the following:

**Understand the basics of perceptual mapping.** Perceptual mapping is a type of procedure that provides the perceived relative image of a set of objects (firms, products, ideas, or other items associated with commonly held perceptions). The purpose is to translate consumer judgments of similarity into distances represented in multidimensional space. To perform a perceptual mapping analysis, the researcher must address three issues: (1) develop a measure of similarity between the entire set of objects to be analyzed, (2) employ the appropriate techniques based on the type of data collected (e.g., nonmetric contingency tables for CA), and (3) identify and interpret the axes of the perceptual map in terms of perceptual and/or objective attributes.

**Select between a decompositional or compositional approach.** Perceptual mapping techniques can be classified into one of two types based on the nature of the responses obtained from the individual concerning the object: (1) the decompositional method measures only the overall impression or evaluation of an object and then attempts to derive spatial positions in multidimensional space that reflect these perceptions (It uses either similarity or preferences data and is the approach typically associated with MDS.) and (2) the compositional method, typified by correspondence analysis, utilizes attribute perceptions of each object to develop a measure of similarity. Perceptual mapping can be performed with both compositional and decompositional techniques, but each approach has specific advantages and disadvantages that must be considered in view of the research objectives. If perceptual mapping is undertaken either as an exploratory technique to identify unrecognized dimensions or as a means of obtaining comparative evaluations of objects when the specific bases of comparison are unknown or undefined, the decompositional or attribute-free approaches are the most appropriate. In contrast, if the research objectives include the portrayal among objects on a defined set of attributes, then the compositional techniques are the preferred alternative.

**Explain correspondence analysis as a method of perceptual mapping.** Correspondence analysis (CA) is an interdependence technique that has become increasingly popular for dimensional reduction and perceptual mapping. Correspondence analysis has three distinguishing characteristics: (1) it is a compositional technique, rather than a decompositional approach, because the perceptual map is based on the association between objects and a set of descriptive characteristics or attributes specified by the researcher; (2) its most direct application is portraying the correspondence of categories of variables, particularly those measured in nominal measurement scales, which is then used as the basis for developing perceptual maps; and (3) the unique benefits of CA lie in its abilities for representing rows and columns, for example, brands and attributes, in joint space. Overall, correspondence analysis provides a valuable analytical tool for a type of data (nonmetric) that often is not the focal point of multivariate techniques. Correspondence analysis also provides the researcher with a complementary compositional technique to MDS for addressing issues where direct comparison of objects and attributes is preferable.

Perceptual mapping provides visual representations emphasizing the relationships between the stimuli under study. But, as with all multivariate techniques, the researcher must be cautious when using this technique because misuse is common. The researcher should become familiar with the technique before using it and should view the output as only the first step in the determination of perceptual information.

## Questions

1. Compare and contrast CA and MDS techniques.
2. Describe how correspondence, or association, is derived from a contingency table.
3. Describe the methods for interpretation of categories (row or column) in CA. Can categories always be directly compared based on proximity in the perceptual map?

## Suggested Readings

A list of suggested readings illustrating issues and applications of multivariate techniques in general is available on the Web accessed through cengagebrain.co.uk or www.mvstats.com.

## References


1. Carroll, J. Douglas, Paul E. Green, and Catherine M. Schaffer. 1986. Interpoint Distance Comparisons in Correspondence Analysis. *Journal of Marketing Research* 23 (August): 271–80.
2. Carroll, J. Douglas, Paul E. Green, and Catherine M. Schaffer. 1987. Comparing Interpoint Distances in Correspondence Analysis: A Clarification. *Journal of Marketing Research* 24 (November): 445–50.
3. Greenacre, Michael J. 1984. *Theory and Applications of Correspondence Analyses.* London: Academic Press.
4. Greenacre, Michael J. 1989. The Carroll-Green-Schaffer Scaling in Correspondence Analysis: A Theoretical and Empirical Appraisal. *Journal of Marketing Research* 26 (August): 358–65.
5. Hoffman, Donna L., and George R. Franke. 1986. Correspondence Analysis: Graphical Representation of Categorical Data in Marketing Research. *Journal of Marketing Research* 23 (August): 213–27.
6. Holbrook, Morris B., William L. Moore, and Russell S. Winer. 1982. Constructing Joint Spaces from Pick-Any Data: A New Tool for Consumer Analysis. *Journal of Consumer Research* 9 (June): 99–105.
7. Levine, Joel H. 1979. Joint-Space Analysis of “Pick-Any” Data: Analysis of Choices from an Unconstrained Set of Alternatives. *Psychometrika* 44 (March): 85–92.
8. Lebart, Ludovic, Alain Morineau, and Kenneth M. Warwick. 1984. *Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices.* New York: Wiley.
9. Market ACTION Research Software, Inc. 1989. *MAPWISE: Perceptual Mapping Software.* Peoria, IL: Business Technology Center, Bradley University.
10. Smith, Scott M. 1989. *PC–MDS: A Multidimensional Statistics Package.* Provo, UT: Brigham Young University.