Sensory Science
Theory and Applications in Foods
Procrustes Analysis and Its Applications to Free-Choice and Other Sensory Profiling

Dolores C. Oreskovich, Barbara P. Klein, and John W. Sutherland

University of Illinois at Urbana-Champaign
Urbana, Illinois

INTRODUCTION

Descriptive analysis, also referred to as quantitative sensory (Powers, 1988), consensus (Tunaley et al., 1988), or fixed-choice (MacFie, 1987) profiling, is the method by which information about products has been obtained by sensory scientists for many years. Descriptive analysis provides a detailed description of the perceived quantitative and qualitative characteristics of a product. A number of descriptive analysis techniques have been developed, including flavor profiling (Cairncross and Sjöstrom, 1950; Caul, 1957), texture profiling (Brandt et al., 1963; Szczesniak, 1963; Szczesniak et al., 1963; Muñoz, 1986), Quantitative Descriptive Analysis™ (Stone et al., 1974; Stone and Sidel, 1985), and Spectrum™ (Meilgaard et al., 1988), and described extensively by these and other investigators (Piggott and Canaway, 1981; Powers, 1988). Each descriptive analysis method varies in the approach used for obtaining product informa-
tion. A particular method is chosen based on the questions to be answered and the product being evaluated. All descriptive analysis methods are similar in that potential panelists are recruited, screened, selected, and trained. However, the number of panelists used, length of training, and type of products evaluated can vary. After training, panelists evaluate products (usually in a controlled environment), using established experimental procedures specific for each method. The data collected are then analyzed using statistical techniques.

Descriptive analysis is confined to the assessment of a product's attributes, usually considered independently of one another. Therefore, the only way a product is judged by a panelist is through individual assessment of each attribute. By collectively examining assessments of each attribute, a "profile" of the product can be provided, as exemplified by the Quantitative Descriptive Analysis™ "spider web" (Stone et al., 1974). Although a profile of a product can provide information about each attribute, little consideration is given to the impression created when attributes interact when descriptive analysis is used. The flavor profile "amplitude" attempts to measure these interactions. In the past, descriptive analysis data have been analyzed using Analysis of Variance (ANOVA), a univariate statistical approach in which information is obtained about samples or products by pooling panelists' responses. Univariate statistics can provide information about the ability of panelists to collectively identify and quantify attributes, but disclose no information about relationships existing among attributes. This limitation makes it difficult to obtain information about the overall perception of a product.

More recently, multivariate analyses such as multidimensional scaling, factor analysis, and principal component analysis have been used to interpret data from descriptive analysis (Powers, 1984). The use of multivariate analysis is restricted by sample size. Meilgaard et al. (1988) suggested that the sample size be at least the square of the number of responses. Despite sample size restrictions, multivariate analysis takes into account the correlation between responses and therefore can provide additional information about the differences between products and panelists that univariate analysis of the same data may not.

Several descriptive analysis methods have been examined closely, improved upon, and ultimately matured into well-accepted and established techniques in sensory science. Although some of the methods are proprietary, attempts are being made to describe standardized procedures (see Chapter 12). Descriptive analysis and its applications have expanded and been improved upon as a result of extensive research and is now widely accepted as a standard methodology in the sensory science area (IFT, 1981; Stone and Sidel, 1985; Meilgaard et al., 1988).

Just as the benefits of descriptive analysis have been documented, so have the limitations. Descriptive analysis may require a considerable amount of time to recruit, screen, and train panelists to evaluate specific products. Once panelists are trained, the panel must be maintained over a lengthy period of time. Not only is panel training and maintenance time consuming, it can also be expensive. In addition, most descriptive analysis methods suggest that testing be done in environmentally controlled facilities, which can further escalate costs.

Since descriptive analysis relies on individuals to obtain information, it is limited by the ability of those individuals to perform the requested tasks, as well as to find words to adequately express their perception of the product. Once words are found, it may be difficult to obtain complete agreement among panelists on their interpretation and meaning. In addition, some individuals may perceive the same stimuli differently or may vary in their sensitivity to a particular attribute. The latter may be reflected in scoring differences in intensity values given by panelists. Panelists may also differ in the range of the scale they use. When product differences are large, the variance between products may be great enough to override the variance arising from these scoring and scaling differences (Powers, 1984). But when products are similar, these variances can no longer be ignored. Training helps panelists become consistent in performing repeated evaluations of the same product and thus promotes panel consistency. However, differences in the way a panelist evaluates a product from session to session or over replications can never be totally eliminated. Interpretation of the data from descriptive analysis requires the use of statistical techniques that can identify differences between samples in the presence of differences between individuals. In most cases the panelist-to-panelist differences are ignored and one or more of the panelists that differ are eliminated.

In this chapter, an alternative to the more conventional descriptive analysis procedures, free-choice profiling, is described and its usefulness is addressed. For analyzing free-choice profiling data, a multivariate statistical method, Procrustes analysis, has been rediscovered and applied specifically to sensory profiling data. This statistical technique will be discussed as a means of analyzing data obtained from free-choice profiling, as well as the more traditional descriptive analysis data. A historical and theoretical background of Procrustes analysis, as well as techniques and applications of Procrustes analysis to sensory data, are discussed below. A step-by-step example of the application of the basic principles of Procrustes analysis to descriptive sensory data is given.

FREE-CHOICE PROFILING IN SENSORY TESTING

Free-choice profiling, also referred to as free word association (Moskowitz and Howard, 1982; Lyon, 1987), is a relatively new technique used to obtain information about a product. Free-choice profiling differs from conventional descriptive profiling in that products are evaluated by members of a panel who
describe perceived qualities of that product using their own individual list of terms, rather than a common scorecard. Terms are not shared or used collectively by panel members and may be mutually exclusive. The number of terms used by panelists may also vary, based on panelist experience and familiarity with the product. Terms must be defined and understood by their originator to ensure consistency in their use. Free-choice profiling is similar to traditional descriptive methods in that panelists must be able to detect differences between similar products, verbally describe specific attributes of products, and quantify those attributes. Moskowitz and Howard (1982) were able to use free word association successfully for market research purposes. Free-choice profiling, however, was first developed by Williams and coworkers (1981) and applied by Williams and Langron (1984) for the evaluation of commercial port wines. This technique has been further described by Arnold and Williams (1986).

Free-choice profiling has a number of procedures in common with descriptive analysis and, therefore, can be considered another type of descriptive analysis. Panelists are recruited, selected, and trained in scale usage, impartiality of judgments, and consistent term usage so that reproducible results may be obtained. However, the degree of training is less, because difficulties associated with achieving agreement among panelists do not have to be confronted.

Since free-choice profiling is fairly new, it has had only limited use and therefore procedural guidelines have not been established or standardized. Standardization of guidelines comes only from extensive testing of established methods obtained through sound research and, finally, acceptance by potential users as an adequate technique for obtaining product information.

Researchers who have used free-choice profiling thus far have varied in the characteristics evaluated, products tested, selection and screening procedures, training practices, scaling and sampling presentation, facilities, and testing conditions as well as analysis of the data. Different examples of how free-choice profiling has been applied to obtain sensory information about a product are discussed below.

Characteristics Evaluated

Similar to other descriptive analysis methods, free-choice profiling can be used to describe a product in terms of one or more characteristics. Free-choice profiling has been used to describe products in terms of a single characteristic by Williams and Arnold (1985) in the evaluation of the aroma of coffees, or Marshall and Kirby (1988) in the evaluation of texture of cheeses, similar to texture or flavor profiling. Free-choice profiling can be used to describe a product in terms of a number of characteristics, such as appearance, flavor, aroma, texture, or any combination of these, similar to the Spectrum™ or

PROCRUSTES ANALYSIS

Quantitative Descriptive Analysis™ methods. For example, several investigators have used free-choice profiling for the evaluation of appearance in addition to aroma, flavor, or texture (Williams and Langron, 1984; McEwan et al., 1989; Guy et al., 1989). The characteristics being judged may be restricted by the sensory analyst, but the number of terms used by each panelist is limited only by their perceptual and descriptive skills.

Products Tested

Although some traditional descriptive methods may be restricted to the evaluation of specific characteristics of a product, i.e., texture and flavor profiling, none is restricted to a specific class of products; neither is free-choice profiling. Free-choice profiling studies have dealt with a variety of products. Free-choice profiling has been used most extensively in the evaluation of beverages, such as wines (Williams and Langron, 1984), coffees (Williams and Arnold, 1985), whiskies (Guy et al., 1989), and meat products, such as chicken patties (Lyon, 1987), frankfurters (Oreskovich et al., 1990), and pork roasts (Oreskovich et al., in preparation). In addition to beverages and meats, other products evaluated include chocolates (McEwan et al., 1989), cheeses (Marshall and Kirby, 1988), and yogurts (Dijksterhuis and Punter, 1990).

Screening and Selection of Panelists

Traditionally, panelists for descriptive analysis are selected on a number of criteria depending on the descriptive method to be used. The criteria could include ability to discriminate between products, communicate perceived perceptions, product usage, task comprehension, availability, interest, health, attitude, level of confidence, etc. Lyon (1987) based selection of free-choice profile panelists on availability and willingness to participate, as well as ability to discriminate and verbally communicate perceptions. Marshall and Kirby (1988) used criteria described above, such as availability and health, and others to screen and select free-choice profiling panelists. Individuals were given a short questionnaire concerning likes, dislikes, and the extent of their texture vocabulary. Panelists were selected based on their responses. Oreskovich et al. (1990) used methods similar to those described for traditional descriptive analysis. They selected subjects based on responses to questions pertaining to food, interest, availability, and health. These subjects were then asked to complete a series of acuity tests including odor matching, basic taste identity, texture ranking, and a series of triangle tests. Panelists were further screened based on their performance. In other studies, however, a screening process was not used and selection was based on past experience of the panelists. Williams and Langron (1984)
grouped panelists into expert and nonexpert wine tasters, using their familiarity with wine tasting as a criterion. Williams and Arnold (1985) selected panelists who had previous experience in descriptive profiling to evaluate coffee aroma. In contrast, McEwan et al. (1989) and Guy et al. (1990) selected subjects with no previous experience in sensory profiling.

Studies with free-choice profiling have routinely used from 8 to 20 panelists in evaluation of products. These numbers are comparable to conventional descriptive analysis techniques. Guy et al. (1989), on the other hand, recruited 100 subjects to study the usefulness of free-choice profiling by consumers. It is possible to develop panels of a larger size than used in other descriptive methods because free-choice profiling does not require extensive training. It has not been established that larger numbers are necessary or desirable because of the difficulty in interpreting the data.

Training of Panelists

The training of panelists is the major factor that distinguishes descriptive analysis from other sensory testing methods and is common to all conventional descriptive techniques. A training period of about 6 months is not uncommon for most descriptive methods, and the time is usually product dependent. A flavor profile panel, for example, may require approximately 60 hours of training and 100 hours of practice per panelist (ASTM, in preparation). About 3 months is required for training a Spectrum™ panel. It is suggested for training a Quantitative Descriptive Analysis™ panel that six to eight one-hour sessions are needed for the development of terms or attributes, as well as additional sessions to orient panelists to new products. For most descriptive analysis methods, the training procedures are well standardized. Standards or references are routinely used to acquaint panelists with product attributes and intensities and the ranges that might be encountered during testing. In comparison, free-choice profiling studies vary considerably in training protocols. In some studies, panelists received no training, while other investigators used extensive training and held numerous practice sessions. Reference samples were sometimes presented to train or familiarize free-choice profile panelists with products or ingredients.

Even investigators who frequently use free-choice profiling do not follow fixed training procedures. In the evaluation of wines by expert and nonexpert wine tasters, panelists were merely introduced to the concept of free-choice profiling and scaling with no mention of any training (Williams and Langron, 1984). Panelists were asked to identify different attributes of wine singularly or in combination with other attributes and to assign intensity values to them. Unlabeled samples were then introduced, and panelists were asked to assign their own descriptive terms. Terms used by each individual were defined and categorized into appearance, aroma, and flavor groups. A list of terms used by three expert and nonexpert tasters are given in Tables 13.1 and 13.2. These two groups used different terms to describe the same wines, based on their experience. Clearly, more terms were developed by the more experienced taster. Similarly, panelists had no formal training in the evaluation of the aromas of six varieties of coffee (Williams and Arnold, 1985) or in the appearance, flavor, and texture of five samples of milk chocolate (McEwan et al., 1989). However, in the evaluation of chocolates, eight samples, five of which were the real samples to be tested, were provided to help in the development of terms by individual panelists.

In a study reported by Marshall and Kirby (1988), panelists were trained in the free-choice profile technique to evaluate the texture of five unflavored processed cheese analogs over 10 30-minute sessions during which panelists, as a group, developed general guidelines to assess the texture of samples both manually and orally. Samples of cheese analogs that varied in moisture and fat content (similar to the model cheeses to be evaluated), as well as samples of different varieties of natural cheeses, were presented to encourage the individual development of texture terms. Panelists were also trained in the use of scales.

### Table 13.1

<table>
<thead>
<tr>
<th>Expert</th>
<th>Appearance</th>
<th>Aroma</th>
<th>Flavor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taster 1</td>
<td>Depth</td>
<td>Cleaness</td>
<td>Cleaness</td>
</tr>
<tr>
<td></td>
<td>Fresh</td>
<td>Fresh</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Brightness</td>
<td>Fruiteness</td>
<td>Body</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Richness</td>
<td>Grip</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Smoothness</td>
<td>Round</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Concentration</td>
<td>Aftertaste</td>
</tr>
<tr>
<td>Taster 2</td>
<td>Tawny</td>
<td>Clean</td>
<td>Body</td>
</tr>
<tr>
<td></td>
<td>Ruby</td>
<td>Fruity</td>
<td>Firmness</td>
</tr>
<tr>
<td></td>
<td>Purple</td>
<td>Green</td>
<td>Coarse</td>
</tr>
<tr>
<td></td>
<td>Cloudy</td>
<td></td>
<td>Tannin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hard</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Crisp</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sour</td>
</tr>
<tr>
<td>Taster 3</td>
<td>Red</td>
<td>Fruit</td>
<td>Tannin</td>
</tr>
<tr>
<td></td>
<td>Blue</td>
<td>Esters</td>
<td>Acid</td>
</tr>
<tr>
<td></td>
<td>Brown</td>
<td>Wood</td>
<td>Sweetness</td>
</tr>
<tr>
<td></td>
<td>Intensity</td>
<td>Oloroso</td>
<td>Chocolate</td>
</tr>
<tr>
<td></td>
<td>Spirit</td>
<td>Body</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burnt</td>
<td>Green</td>
<td></td>
</tr>
</tbody>
</table>

TABLE 13.2 Descriptive Terms Used by Three Inexperienced Wine Tasters

<table>
<thead>
<tr>
<th>Nonexpert</th>
<th>Appearance</th>
<th>Aroma</th>
<th>Flavor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taster 1</td>
<td>Clarity</td>
<td>Volatile acidity</td>
<td>Acid</td>
</tr>
<tr>
<td></td>
<td>Intensity</td>
<td>Fruiteness</td>
<td>Sweetness</td>
</tr>
<tr>
<td></td>
<td>Redness</td>
<td>Woodiness</td>
<td>Smoothness</td>
</tr>
<tr>
<td></td>
<td>Yellowness</td>
<td>Alcoholic Strength</td>
<td>Body</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Overall intensity</td>
<td>Aftertaste</td>
</tr>
<tr>
<td>Taster 2</td>
<td>Intensity</td>
<td>Burnt</td>
<td>Astringent</td>
</tr>
<tr>
<td></td>
<td>Red</td>
<td></td>
<td>Acidy</td>
</tr>
<tr>
<td></td>
<td>Brown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taster 3</td>
<td>Color</td>
<td>Smell</td>
<td>Taste</td>
</tr>
<tr>
<td></td>
<td>Red</td>
<td></td>
<td>Acidity</td>
</tr>
<tr>
<td></td>
<td>Brown</td>
<td></td>
<td>Tannin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Softness</td>
</tr>
</tbody>
</table>


After 6 days of training, panelists were asked to define their terms and to eliminate any that had similar meanings. Individual ballots were compiled following the developed guidelines, and terms were placed under two subheadings, manual and oral. The remaining four training sessions were used by panelists to refine their ballots.

More extensive training was provided to panelists evaluating meat products. Lyon (1987) used free word association to develop a list of terms to describe the taste and aroma of both fresh chicken patties and stored chicken patties that had been reheated. A 10-member panel was selected and trained on basic taste and odor recognition, as well as threshold perception. Training to help generate terms took place over an 8-week period in which panelists were presented with chicken patties that varied according to methods used for cooking and reheating.

Oreskovitch et al. (1990) used a free-choice profile panel to evaluate the flavor and texture of two samples of frankfurters differing in fat content. Subjects selected for the free-choice profile panel practiced using a 15-cm continuous line scale during each training session. Panelists were offered several brands of commercial frankfurters differing in fat, meat type, and salt content, as well as frankfurters processed at the university meat science laboratory to help generate terms. Terms were then defined and incorporated into individual score sheets. Panelists were asked to generate values representing their perceived intensity for each attribute on a control frankfurter. Six one-hour sessions over 3 weeks were used to train panelists to evaluate frankfurters and to become familiar with their own intensity values given to the control sample.

In an ongoing study by Oreskovitch et al. (in preparation), similar criteria were used to generate terms in the training of panelists for the evaluation of texture and flavor of whole muscle pork products. Panelists determined intensity values for a control sample of pork to use for the comparison with other products. After terms were generated and defined, panelists were trained in their use for 20 hours over a 5-week period.

In the majority of free-choice profiling studies, panelists develop terms individually with no suggestion of terms being added or eliminated by an outside mediator. Gay et al. (1989), however, added two terms, smoothness and maturity, to all consumer free-choice profiling ballots for whiskies. This was done to assess whether consumers could evaluate important attributes even if they did not generate the descriptors themselves. What the consumers characterized as maturity of whiskies, the trained panel characterized as smooth, sweet, vanilla, and malty. In addition, the consumers' perception of smooth, mellow, and sweet in the whiskies were characterized as estery and woody by the trained descriptive panel. This emphasizes the necessity for panelists to use their own terms, even if they are not as specific as the investigator would like.

Scale/Scorecard and Replications

Scales and scorecards used in descriptive analysis are usually very specific and are dependent upon the method used. For example, Quantitative Descriptive Analysis™ uses a 6-inch unstructured scale, while Spectrum™ scales are similar, being 15 cm in length. A 15-cm unstructured line scale is adopted in many descriptive techniques. Final scorecards are derived by consensus with guidance of a group leader. Flavor profiling has a unique scale that does not adapt easily to statistical analysis. In its earliest format, the scorecard was individualized with panelists reporting flavor notes as they appeared. Texture profiling depends on a standardized series of references to align panelists' judgments, and evaluations are made relative to the reference scales.

In free-choice profiling, no standardized scaling method or scorecard has emerged. The number of terms on an individual scorecard is idiosyncratic and reflects the subjects ability to describe the product attributes. Category and unstructured line scales (6.5–15 cm long) are commonly used. In the evaluation of wines (Williams and Langron, 1984) and coffees (Williams and Arnold, 1985), six-point category scales were incorporated in the score sheet. Marshall and Kirby (1988) used a 12-cm line scale. A 6.5-cm continuous line scale was used for the evaluation of chocolate samples (McEwan et al., 1989).

In Lyon's (1987) study of chicken patties, a nine-point category scale was
used in one phase and a 10-cm line scale in another. A 15-cm continuous line scale was employed for the evaluation of frankfurters and pork roasts (Oreskovitch et al., 1990; Oreskovitch et al., in preparation). The diversity of scales and scorecards in free-choice profiling is evident.

Three or four replications by each judge for each treatment is normally recommended for any sensory testing to ensure reliable results (Lamond, 1977), but the actual number applied is the choice of the sensory analyst. Complex products with minor differences may require more replications. Similarly, the number of samples presented at one time is dependent on the product’s characteristics and the judgment of the analyst.

The free-choice profiling studies reflect the same diversity in numbers of samples and replications seen in other testing methods. Since free-choice panelists are not extensively trained, the number of samples presented at one time is usually limited to two or three. Samples are often presented monadically (Williams and Langron, 1984; Williams and Arnold, 1985) or in pairs (Marshall and Kirby, 1988). Generally the number of replications was three or four (Williams and Arnold, 1985; Lyon, 1987; Marshall and Kirby, 1988; McEwan et al., 1989).

In studies by Oreskovitch et al. (1990; in preparation), for each of four sessions, four coded samples of frankfurters (two treatment and two control) were randomly presented to panelists. Seven replications of eight different pork sample variations were presented in 14 sessions over 7 weeks. In addition to the scorecards, each panelist was given his or her intensity scores for the control product at the beginning of each session. This helped to maintain consistency in the panelist’s evaluation.

Facilities and Testing Conditions

It is accepted practice in descriptive analysis to present samples to panelists using standard procedures under conditions in which lighting, temperature, humidity, odors, and sounds can be controlled to minimize distractions and potential biases. Similar practices have been implemented in many of the free-choice profiling studies. Testing of samples by panelists was done under controlled conditions utilizing individual sensory booths to ensure independent judgments. However, in the evaluation of whiskies by consumers trained in the free-choice profiling technique (Guy et al., 1989), samples were evaluated at their homes and ballots returned by mail.

Statistical Analysis

The procedures used in the aforementioned studies vary considerably, illustrating the lack of uniformity in the use of free-choice profiling at the present time. Unlike more conventional descriptive analysis techniques, the data collected cannot be analyzed and compared using univariate or traditional multivariate analysis because of the lack of uniformity in scorecards among panelists. Thus, the information must be examined differently.

Analysis of free-choice profiling data differs from that associated with descriptive analysis in that it is nicely interpreted using Procrustes analysis. Individual panelist’s responses to a set of samples can be geometrically arranged into a single configuration where scores for each sample can be defined as a point in space, with each attribute defining a different dimension for each attribute in the space. Configurations of individual panelist’s responses to each sample can then be matched and compared using Procrustes analysis. Procrustes analysis can also be used to analyze descriptive analysis data, but was not considered as an alternative statistical method until Harries and MacFie (1976) used it to examine the textural characteristics of beef roasts. It should be noted that Procrustes analysis can provide sample, panelist, as well as attribute information.

PROCRUSTES ANALYSIS

The advantages and disadvantages of using free-choice profiling or more conventional descriptive analysis techniques have been suggested above. Since its inception, free-choice profiling has been thought to have some advantages over conventional descriptive profiling. Free-choice profiling is less time consuming, since panelists require less training. Panelists may feel more at ease with their judgments because it is not necessary for them to agree upon a common list of descriptors. In addition, the frustration associated with trying to force agreement among panelists is thought to be alleviated. Similar to descriptive analysis, free-choice profiling is not exempt from dealing with potentially large variations that can be associated with any method that uses individuals as analytical tools. However, by using Procrustes analysis to interpret free-choice profiling data, many of these panelist-to-panelist variations can be reconciled. As with any statistical tool, there are certain limitations and disadvantages to Procrustes analysis.

Background

The name “Procrustes” was first used in 1962 by Hurley and Cattell to describe the matching of configurations. “Procrustes,” a Greek term meaning “to beat out,” described a Greek innkeeper in Attica who seized travelers and tied them to the iron framework of his beds where he stretched the legs of short men and cut the legs of tall men so that they would all fit his beds. The name
"Procrustes" conceptually describes, to some extent, what is done with data in a Procrustes analysis.

In conventional descriptive analysis as well as free-choice profiling, panelists evaluate samples by assigning to them intensity values for various attributes. The values can be placed in a matrix in which samples and attributes are represented by rows and columns, respectively. Each element in the matrix is the intensity value perceived for that sample and attribute. A single matrix is created for each panelist as seen in Fig. 13.1. The matrices for several panelists can be represented geometrically, with each sample defining a point in space, and each attribute corresponding to a different dimension. A panelist's responses to a set of samples then defines a configuration of points in that space. Procrustes analysis is a multivariate statistical tool in which the configurations of individual panelists can be matched and compared. Procrustes analysis uses geometrical transformations, whereby the configurations are adjusted to match each other as closely as possible. Transformations allow for reexpression of the configurations. The closeness of the matrices to one another can be characterized by summing the squared distances between the points (matrices) and a common centroid (target or consensus) matrix. This sum of squares is referred to as the Procrustes statistic.

Procrustes transformation procedures are aimed at optimizing a particular criterion. Criteria are considered to be indirect measures of the similarities between the transformed and target matrix and are usually defined or selected by the experimenter. Three criteria commonly used with Procrustes analysis are least-squares, inner product, and a consensus criterion. The Procrustes statistic described earlier is used as a statistical check to determine if the designated criterion is being met.

<table>
<thead>
<tr>
<th>Attribute 1</th>
<th>Attribute 2</th>
<th>Attribute 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>5.9</td>
<td>7.1</td>
</tr>
<tr>
<td>Sample 2</td>
<td>4.2</td>
<td>3.1</td>
</tr>
<tr>
<td>Sample 3</td>
<td>6.0</td>
<td>7.0</td>
</tr>
<tr>
<td>Sample 4</td>
<td>3.9</td>
<td>2.9</td>
</tr>
</tbody>
</table>

**FIG. 13.1** Data matrix. Matrix configuration of simulated data set for panelist 1. Rows represent samples and columns represent attributes. The numbers within each column and row represent the intensity values assigned by a single panelist.

Initially, Procrustes analysis was used to match one configuration to a target configuration, which has been defined as the pair-wise approach. Moiser (1939) and Green (1952) first used the analysis for the comparison of different types of statistical analyses. Green (1952) described the matching of configurations as a transformation in which one matrix is rotated to a target matrix under specified constraints. Specifically, Green (1952) restricted the matrices to having the same number of columns and to be of "full column rank." The "column rank of a matrix" is the number of linearly independent columns in the matrix. If the column rank is equal to the number of columns in the matrix, the matrix is said to have "full column rank."

Green (1952) used a least-squares criterion to develop three analytical methods for obtaining transformed matrices. The least-squares criterion minimizes the distances between corresponding points in the configurations (Procrustes-PC, 1989). Hurley and Cattell (1962), who were the first to coin the term "Procrustes," used a rotation matrix to transform one matrix of responses to match a designated target matrix for hypothesis testing.

Cliff (1966) transformed matrices for matching and for fitting to a target matrix, as described above. However, he restricted these to orthogonal transformations. Schönemann (1966) derived his own solution for matching matrices and referred to it as the "orthogonal Procrustes problem." This solution was a more generalized approach in that it was not restricted by Green's (1952) restraints in which matrices had to be of "full column rank." The matching of one configuration to a target configuration as described by Cliff (1966) and Schönemann (1966) was later referred to as the "two-sided" orthogonal Procrustes problem by Schönemann (1968).

Schönemann and Carroll (1970) found the matching of matrices helpful when comparing different methods of multidimensional scaling. Schönemann and Carroll (1970) presented a least-squares technique designed to be programmable for use with computers for fitting matrices by rotation, translation, and central dilation. A scaling factor for central dilation that allowed for the expansion or contraction of data points, taking differences in spread between configurations into account, was also included.

Instead of rotating one matrix to fit another, it became possible to rotate many matrices to fit a common centroid matrix: generalized Procrustes analysis. This centroid matrix was first introduced by Kristof and Wingersky (1971) and is a key element in an iterative procedure in which the least-squares criterion is met without any consideration to translation and scaling. Gower in 1975 described the centroid matrix as representing the average or consensus configuration. Unlike Kristof and Wingersky (1971), Gower included translation and scaling constraints and provided an elaborate starting procedure in which matrices were initially standardized prior to rotation. In addition, he provided a computational technique by which results from Procrustes analysis can be summarized in an analysis of variance format.
Ten Berge (1977) made modifications, specifically the rotation and scaling steps, of Gower's method. Ten Berge and Kroll (1984) derived transformations for multiple matrices having different numbers of columns and proposed an inner product criterion by which angles between corresponding point vectors were minimized.

Peay (1988) used a different criterion: consensus. The consensus criterion maximizes the total variance of the common centroid configuration. Peay (1988) was able to combine and expand features of techniques derived and discussed by other investigators that incorporated scaling and matching multiple asymmetric matrices (unequal column numbers). One computer program that is available for Procrustes analysis uses the consensus criterion (Procrustes-PC, 1989).

It is often suggested that principal component analysis be used in place of Procrustes analysis and vice versa. Principal component analysis is also a multivariate technique. Data for analysis is arranged in a matrix as described in the Procrustes procedure. However, the purpose of applying principal component analysis to a set of data is to reduce the number of dimensions (attributes) in which samples may be found, with minimal loss of information. Principal component analysis determines the orientation or directionality of a single set of data, which can indicate relationships between attributes and differences between samples. This differs from Procrustes analysis which matches a minimum of two sets of data. As explained by Piggott and Sharman (1986), principal component analysis searches for a sequence of linear combination of attributes that can account for the maximum variation associated with the data. The number of axes or principal components needed to characterize the majority of variability for a single data set is dependent upon how correlated the original set of variables are.

Principal component analysis is applied to only one data set at a time. This data set may be a very large set in which raw data are entered. Alternatively, it is most often applied to a data set consisting of sample means averaged across panelists and over replications for each attribute (Powers, 1984). Use of a single large data set is not recommended (MacFie, 1987) because the resulting configuration is very crowded and may be difficult to interpret.

Whether panelist effects are identified by using principal component analysis or a univariate analysis method, variations associated with panelists should be accounted for prior to any analysis to determine differences between samples. Because generalized Procrustes analysis and principal component analysis provide the experimenter with very different information, it would not be fair to suggest that one be used over the other. In fact, it has been suggested that principal component analysis be used in conjunction with Procrustes analysis. By applying principal component analysis to the final configuration obtained by Procrustes analysis, a visual representation of the configuration can be constructed that allows for easier interpretation.

Early investigators (Gower, 1966) mentioned using principal coordinate analysis following Procrustes analysis. Principal coordinate analysis is similar to principal component analysis in that the number of dimensions present in the final configuration obtained by applying Procrustes analysis to the data can be reduced. These two analyses differ in that principal coordinate analysis starts with a matrix of distances and principal component analysis with a covariance or correlation matrix (Piggott and Sharman, 1986). Gower (1966) suggested that principal coordinate analysis be used as a matter of computational convenience especially when there are more attributes than samples. This was suggested at a time when computer analysis was not as advanced as it is today. Because of this advancement either principal coordinate or principal component analysis may be used on data following Procrustes analysis, each providing similar information.

**Approaches**

**Pair-wise** In the pair-wise Procrustes analysis approach, pairs of configurations are compared. These configurations may represent panelists, samples, or attributes. For simplicity, discussion will be restricted to panelist configurations. As illustrated in Fig. 13.2, samples are represented by points and the attrib-

![Diagram](image)

FIG. 13.2 Pair-wise Procrustes analysis. Pairs of panelist configurations are transformed; for the first pair, panelist 1 is matched to a target configuration. As a result of the analysis, a transformed panelist 1 configuration was computed. Letters (A, B, C) represent three samples. Subscripts (1, 2, n) represent panelists, and subscript T represents the target configuration. Bold axes within each configuration represent the dimensions in which samples could be found.
utes represent the dimensions. In the pair-wise approach, one panelist configuration is usually designated as a "target" to which other panelist configurations are compared. The target is a matrix of values for each attribute that is set by the investigator. The target matrix can be based on theoretical expectations, previous experience or obtained by some mathematical computation of existing values for identical attributes (Cliff, 1987). The configuration (matrix) representing a panelist is mathematically transformed in Procrustes analysis to form a new transformed panelist configuration, which more closely matches the target configuration. Through Procrustes analysis, the distance between a pair of configurations, referred to as the Procrustes statistic, is obtained. In the pair-wise approach, a Procrustes statistic can be calculated between the newly transformed panelist configuration and the target configuration.

A Procrustes statistic is computed by summing the squared distances between the corresponding points of two configurations, whether it be between a pair of configurations representing panelists or between a configuration and a target. A Procrustes statistic computed for a pair of panelists can be illustrated in the form of a lower triangle matrix (Table 13.3), as in data taken from Banfield and Harries (1975). A low Procrustes statistic, as obtained from expert panelists 1 and 2 (0.513), for example, depicts how similar these two panelists were in their perceptions of the differences between samples. A high value such as that

<table>
<thead>
<tr>
<th>Expert</th>
<th>Inexperienced</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2 3 4 5</td>
</tr>
<tr>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>0.513 0.000</td>
</tr>
<tr>
<td>3</td>
<td>0.566 0.522 0.000</td>
</tr>
<tr>
<td>4</td>
<td>0.591 0.569 0.514 0.000</td>
</tr>
<tr>
<td>5</td>
<td>0.502 0.390 0.624 0.546 0.000</td>
</tr>
</tbody>
</table>

Inexperienced

<table>
<thead>
<tr>
<th>6 7 8 9 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.747 0.764 0.810 0.776 0.752 0.000</td>
</tr>
<tr>
<td>1.149 1.085 1.088 1.104 1.161 1.153 0.000</td>
</tr>
<tr>
<td>1.412 1.449 1.483 1.517 1.513 1.589 1.608 0.000</td>
</tr>
<tr>
<td>1.184 1.169 1.253 1.296 1.173 1.318 1.536 1.504 0.000</td>
</tr>
<tr>
<td>1.078 0.987 1.041 1.065 1.065 1.113 1.289 1.486 1.569 0.000</td>
</tr>
</tbody>
</table>

Source: Banfield and Harries, 1975.

PROCRUSTES ANALYSIS

between inexperienced panelists 7 and 8 (1.608) shows a large disagreement between this pair of panelists in determining differences between samples.

After Procrustes analysis has been applied to the data, principal coordinate or principal component analysis can be used to provide a visual image. This is important since Procrustes analysis provides information in multidimensional space, which makes it difficult, if not impossible, to visualize. Although either analysis is appropriate to use following Procrustes analysis, in that each can provide a visualization in two dimensions, principal coordinate analysis has been used most often. Interpretation of the Procrustes data is then easier because the number of dimensions in which the points can be found can be reduced without sacrificing any valuable information. Principal coordinate analysis can be used to produce a reduced set of axes from each configuration, while attempting to keep the distances between panelists and between attributes fixed or constant. Newly constructed axes for panelists can then be plotted to give a graphical display of their relative positions allowing for easier interpretation of the original configuration.

Generalized The generalized approach is based on the pair-wise concept, but differs in that all configurations are compared as a group (Fig. 13.3) to a target configuration and then a new consensus configuration is derived. The target configuration can vary depending upon the experimenter and is derived by computing the mean of individual configurations or it can be the configuration

![FIG. 13.3 Generalized Procrustes analysis. Several panelist configurations are transformed to more closely match a target configuration. As a result of the analysis, transformed panelist's configurations were computed. Letters (A, B, C) represent three samples. Subscripts (1, 2, n) represent panelists, and subscript T represents the target configuration. Bold axes within each configuration represent the dimensions in which samples could be found.](image-url)
for a particular individual. The consensus configuration, with samples represented as points, is most often derived from individual configurations by the following process. The dimensions of the consensus configuration are equal to the number of attributes used for the evaluation of samples. When generalized Procrustes analysis is applied to conventional descriptive analysis data, the number of attributes is equal. However, in free-choice profiling the number of attributes may be different for each panelist. To obtain the consensus configuration for free-choice profiling data, columns of zeros are added to individual matrices so that all panelist configurations will have equal dimensionality. Individual configurations are then mathematically matched to a target configuration, similar to the pair-wise approach. Matching is achieved through a series of iterative steps, by transforming individual configurations to the target configuration in such a way that intersample distances for individual configurations are maintained (Fig. 13.4). The consensus configuration is then computed as the mean of all transformed individual configurations. The consensus configuration is now the new "target" to which transformed individual configurations are matched. This iterative process continues until the distance between the newly transformed configurations and the consensus configuration is minimized as measured by the Procrustes statistic. The Procrustes statistic can be determined at any transformation point and is calculated for a pair of configurations as in the pair-wise approach, such as between a panelist and a consensus configuration. The distance will be minimized to a point where continued transformations result in a minimal change in the Procrustes statistic meeting some predetermined tolerance. A final consensus configuration can then be computed (Fig. 13.5).

Similar to the pair-wise approach, principal coordinate or principal component analysis can be applied to transformed individual configurations, as well as the final consensus configuration to obtain those attributes or factors that can clearly differentiate between samples. These attributes or dimensions can account for decreasing proportions of total variance. Just as in the pair-wise approach, a graphical two-dimensional display of the consensus configuration can be obtained by plotting sample scores on the first two axes. These new axes can be interpreted in terms of each panelist's own initial attributes list in one of two ways: by calculating correlations of the original attributes for each panelist with the newly transformed attributes of the consensus or by obtaining a rotation matrix for each panelist, which transforms his or her original centered configuration to the consensus (Arnold and Williams, 1986).

When generalized Procrustes analysis is applied to free-choice or conventional descriptive analysis data, panelists' configurations are matched in such a way that the relative positions between samples for each individual are unchanged. As a result, a consensus configuration is computed that can be used in place of untransformed sample means for comparing other analytical measurements or sensory data (Langron et al., 1984).

**FIG. 13.4** Iterative procedure for generalized Procrustes analysis. In the generalized Procrustes approach, transformed configurations for each panelist goes through a series of iterative steps in which configurations are continually matched and compared to newly computed target configurations. The iterative process continues until the distance between the newly transformed configurations and the target (consensus) configuration is minimized as measured by the Procrustes statistic to a point where continued transformations result in a minimal change meeting some predetermined tolerance. Letters (A, B, C) represent three samples. Subscripts (1, 2, n) represent panelists, and subscript T represents the target configuration. Bold axes within each configuration represent the dimensions in which samples could be found.

**FIG. 13.5** Final transformed panelist and consensus configuration. Transformed panelist configurations can be superimposed to form one configuration. In addition, a final consensus configuration can be computed as the mean of all finally transformed panelists' configurations. Letters (A, B, C) represent three samples. Subscripts (1, 2, n) represent panelists, and hollow letters represent the final configuration for each sample known as the consensus configuration. Bold axes within each configuration represent the dimensions in which samples could be found.
ILLUSTRATION OF PROCRUSTES ANALYSIS

Transformation steps of Procrustes analysis can include initialization, rotation/reflection, and scaling. In the various Procrustes applications that are to be presented later in this chapter (see Applications), the exact steps and the order in which they have been applied has varied greatly. Although the authors of this chapter feel that all the steps are important, the experimenter must decide the relevance of each transformation step for the particular data to be analyzed. In addition, the way in which each transformation is computed can also vary. Each transformation attempts to reconcile differences between the panelist and target configurations. To demonstrate this, a simulated data set will be used to illustrate each transformation step.

Three panelists used conventional descriptive analysis to evaluate three texture attributes (hardness, cohesiveness, and springiness) of four different whole wheat bread formulations. All attributes were evaluated using a 15-cm continuous line scale. A response for each panelist was constructed, where the rows represent different breads (A, B, C, D) and columns represent the three texture attributes (Table 13.4). Table 13.4 shows the intensity values of each attribute given by panelist 1 in the evaluation of the four breads. Each row of the matrix is the coordinates for a point representing each bread sample, with each attribute representing a dimension in the space in which a particular sample can be found. The different shapes in Fig. 13.6 represent the different bread formulations: sphere-bread 1, cube-bread 2, cone-bread 3, and cylinder-bread 4. The different shading of these shapes as well as labels designated as P1–P3 represent panelists, where light gray is panelist 1 (P1), middle gray is panelist 2 (P2), and black is panelist 3 (P3). All panelist configurations have been plotted in three-dimensional space in which the x axis represents hardness, the y axis cohesiveness, and the z axis springiness. The size of the symbol indicates its position, with symbols that appear larger being closer to the viewer. For panelist 1, the scores appear to be more spread out while scores for panelist 3 appear more condensed. It is difficult to determine differences between the treatments; however, breads 1 and 3 appear to be similar. Procrustes statistics and scaling factors for each panelist at each transformation step are found in Table 13.5. Procrustes statistics were calculated based on original data between each panelist and the initial consensus configuration. The initial consensus configuration was calculated as the mean of the three panelists' configurations. As previously mentioned, a large Procrustes statistic can indicate disagreement between a panelist and the target or consensus. In Table 13.5, a large Procrustes statistic for panelist 1 (35.002) compared to values for panelists 2 and 3 indicates that the configuration for panelist 1 is much further away from the initial consensus configuration. Panelist 3, however, appears to be the closest to the initial consensus configuration as suggested by a low Procrustes statistic (14.346).

TABLE 13.4 Simulated Data Set for Panelist 1 in the Evaluation of Whole Wheat Bread Formulations

<table>
<thead>
<tr>
<th>Samples</th>
<th>Hardness</th>
<th>Cohesiveness</th>
<th>Springiness</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5.9</td>
<td>7.1</td>
<td>6.9</td>
</tr>
<tr>
<td>B</td>
<td>4.2</td>
<td>3.1</td>
<td>7.0</td>
</tr>
<tr>
<td>C</td>
<td>6.0</td>
<td>7.0</td>
<td>2.9</td>
</tr>
<tr>
<td>D</td>
<td>3.9</td>
<td>2.9</td>
<td>3.1</td>
</tr>
</tbody>
</table>
TABLE 13.5 Procrustes Statistics Calculated Before and After Each Transformation and Scaling Factors for Iteration 1

<table>
<thead>
<tr>
<th>Panelist</th>
<th>Original</th>
<th>Initialization without normalization</th>
<th>Rotation and reflection</th>
<th>Scaling</th>
<th>Scaling factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>35.002</td>
<td>8.968</td>
<td>3.202</td>
<td>1.573</td>
<td>0.865</td>
</tr>
<tr>
<td>2</td>
<td>17.766</td>
<td>6.948</td>
<td>4.179</td>
<td>3.069</td>
<td>0.869</td>
</tr>
<tr>
<td>3</td>
<td>14.346</td>
<td>10.344</td>
<td>4.980</td>
<td>1.939</td>
<td>1.306</td>
</tr>
<tr>
<td>Total</td>
<td>67.113</td>
<td>26.260</td>
<td>12.362</td>
<td>6.581</td>
<td></td>
</tr>
</tbody>
</table>

Initialization

Translation is the first transformation step in Procrustes analysis. It has been suggested that translated matrices be initially standardized prior to rotation and reflection. This standardization process was also referred to as “initial scaling” and later referred to as “normalization” by Peay (1988). In the development and explanation of Procrustes analysis, most researchers refer to both these steps, translation and standardization, as simply “translation.” To minimize any possible confusion, we will refer to this two-step process as “initialization” with the first step referred to as “translation,” and the second “normalization.” Note that the normalization step should not be confused with “normalizing the data” by converting sets of scores so that they are normally distributed, and that the term “scaling” is reserved for a later step in the analysis.

Translation

The translation step involves the moving of individual configurations about the origin (zero). Translated matrices are computed by subtracting from each individual matrix its matrix of column means. Each element of the translated matrix is then the deviation from its mean. The resulting matrix can also be referred to as a “deviation” or “shifted” matrix. The translation step can account for the variation due to different levels of the scale being used by different panelists. It can remove the effect of individual panelists who consistently under- or overscore a particular attribute. These scoring differences may reflect an individual’s response to the perception of different stimuli or a difference in an individual’s sensitivity to a particular attribute. In the initial data (Fig. 13.6) all the configurations were positive and grouped together. After translation, the configurations are centered about the origin. It appears that scores for panelist 2 (Fig. 13.7) are much more spread out than they appeared in the initial configuration (Fig. 13.6). Panelist 1, however, appears to have shifted the most after translation, and panelist 3 the least. The consensus was calculated as the mean of all panelist’s translated configurations. The Procrustes statistic was calculated between each panelist’s translated configuration and the new consensus. The Procrustes statistic (Table 13.5) for panelist 1 confirms what can be seen in Fig. 13.7. The difference between the Procrustes statistic for panelist 1 initially (35.002) and after translation (8.968) is 26.034, which is much greater than for panelist 2 with a difference of 10.818 and panelist 3 of 4.002. This large difference confirms that panelist 1 was more affected by the translation step than panelist 2 or 3. It can therefore be inferred that panelist 1 overscored or consistently assigned higher intensity values to all the bread samples in comparison to panelists 2 and 3.

Normalization

Normalization is a computational procedure that adjusts a matrix so that its elements are of comparable magnitude to other normalized matrices. The matrices are then expressed on a common basis so that no one measurement, because of its magnitude, can unduly influence the comparison of...
samples. After normalization, elements of translated matrices for each panelist are converted so that all values fall between -1 and 1. The application of the normalization step is optional but usually follows the translation step. Gower (1975) and Peay (1988) suggested that normalization be applied to translated matrices that were unusually disproportionate in magnitude prior to transformation by rotation and reflection. Such incommensurable translated matrices may present themselves when comparing panel matrices derived from different sensory methodologies or when comparing matrices obtained from sensory assessment to matrices derived from analytical measurements. Peay (1988) was able to derive a more generalized version of Gower's normalization step, in that it could be used for sets of translated matrices that did not necessarily have identical column order. When the normalization process is applied, an adjustment must be made to the final transformed matrices if they are to be compared relative to the units of the initial matrices. Because scores for panelists in our simulated data set were of similar magnitude, normalization was not applied to the panelist configurations.

Rotation/Reflection

The next step in the analysis is to optimally fit or match initialized configurations to the target configuration by using the rotation step. The target configuration to be matched in this illustration was calculated as the mean of the translated panelist configurations. This step is accomplished by rotating panelist configurations to most closely match the target configuration. Rotation matrices are calculated and then multiplied by the initialized panelist configurations to obtain transformed configurations. Rotation procedures may differ based, for example, on whether they allow a particular configuration to be reflected. When generalized Procrustes analysis is applied to free-choice profiling data, the rotation/reflection transformation procedure typically has a very dramatic effect. This is so, because this is the first step in comparing individual configurations that may have different dimensions or attributes associated with them.

Rotation and reflection can account for variation associated with individual panelists' different interpretations of the same term. This type of variation can come about when a particular descriptor that may be difficult to define or interpret is used. For instance, in this example, the terms cohesiveness and hardness in the evaluation of bread may be confused by a panelist. In the evaluation of tea, the terms bitter and astringency have been commonly confused. This variation is reflected in the panelist's scores resulting in a changed orientation of the configuration compared to other panelists. After rotation/reflection the configurations appear to be better matched (Fig. 13.8). The different breads seem to be much more differentiated after rotation than after translation. Bread 1 is characterized as being much harder yet as springy as bread 2. Bread 2 appears to be much softer than bread 1 and 3. Bread 3 appears to be more cohesive than the other breads. Bread 4 was given low intensity values for all three attributes. The Procrustes statistics for the panelists were calculated between each panelist configuration and the new consensus. The new consensus was calculated as the mean of newly rotated panelist configurations. The Procrustes statistic (Table 13.5) for panelist 2 confirms what is seen in Fig. 13.8. The difference between the Procrustes statistic for panelist 2 after translation (6.948) and after rotation (4.179) is 2.769 in comparison to 5.748 and 5.364 for panelists 1 and 3, respectively. The small difference in the Procrustes statistic for panelist 2 in comparison to the other panelists confirms that panelist 2 was the least affected by the rotation/reflection step. These results suggest that panelist 2 was the most consistent within himself in his use of the scorecard.

**FIG. 13.8** Initialized and rotated configuration. Illustrated, rotated, and reflected configuration of simulated data set presented in three-dimensional space. Shapes represent breads 1–4. Sphere-bread, 1; cube-bread, 2; cone-bread, 3; and cylinder-bread, 4. The different shading of these shapes as well as labels P1–P3 represent panelists, where light gray is panelist 1 (P1), middle gray is panelist 2 (P2), and black is panelist 3 (P3). Axes represent attributes 1–3: X-hardness, Y-cohesiveness, and Z-springiness. The size of the symbol indicates its position with symbols that appear larger being closer to the viewer.
Scaling

Scaling is the last step on the Procrustes analysis procedure, and again, there are many alternative procedures that can be used to "scale" the data. One alternative is not to scale the data at all (Schönemann and Carroll, 1970; Gower, 1975; Ten Berge, 1977; Langron, 1983). If it is decided to use the scaling step, then scaling can occur in two directions. The points in a particular configuration can either be expanded or contracted. Dilation is another term used to describe this expansion. When the expansion or stretching occurs about some center point, it is referred to as "central dilation." Scaling is optimized by computing scaling factors that meet some predetermine criteria and are then applied to the translated, rotated, and reflected configurations. Scaling factors can indicate the dispersion of a set of data points for a particular panelist. One panelist may use a very small range of the scale, while another may use a much larger range to express differences between samples on a particular attribute. A large scaling factor will indicate that a panelist used a small range of the scale while a small number will indicate the opposite. When scaling is applied, this variation associated with different scale usage by panelists can be addressed. If the distance between points is small, for instance, that distance can be enlarged; however, if the value is large, it can be decreased depending upon how the criteria can best be met. If the scaling factor is greater than 1, the distance of the points from the origin are increased (stretched) and if they are less than 1, they are decreased (shrunk) to match the target configuration. The configurations appear as shown in Fig. 13.9. Panelist 2 appears to have moved very little relative to the amount of movement by panelists 1 and 3. The Procrustes statistic for each panelist to a new consensus was calculated. The new consensus was calculated as the mean of the newly scaled panelist's configurations. The Procrustes statistics (Table 13.5) for the panelists confirm what is seen in Fig. 13.9. The difference between the Procrustes statistic for panelist 3 after rotation (4.980) and after scaling (1.939) is 3.041 in comparison to a value of 1.629 and 1.110 for panelists 1 and 2, respectively. This confirms that panelist 2 was least affected by scaling and panelist 3 was the most affected. In addition to the Procrustes statistic, scaling factors have been calculated for each of the panelists. Panelist 3 had a scaling factor greater than 1 (1.306) indicating that a limited range of the scale was used, whereas panelists 1 and 2 with scaling factors less than 1 (0.865 and 0.869, respectively) used a much larger range of the scale.

Iterative Procedure

The iterative procedure begins at the point where individual panelists' configurations have been normalized and a consensus configuration has been computed. The consensus configuration is the mean of all newly normalized panelists' configurations. When all normalized configurations have been rotated and scaled (referred to as newly transformed configurations), a single iteration is complete. After the first iteration, a new consensus configuration is derived as the mean of all newly transformed panelists' configurations. It is this new consensus configuration or target that newly transformed configurations computed from iteration one are rotated and scaled to match. For our simulated data, five iterations were computed, and panelist configurations are shown in Fig. 13.10. When examining Fig. 13.10, the data appear to be closer to the origin (zero) in general. In addition, there seems to be greater variation between panelists in evaluating bread 4. Breads 1 and 2 had similar scores for springiness, although they differed in hardness. Hardness scores for bread 3 were comparable to that of bread 1; however, cohesive scores were higher. Table 13.6 shows the Procrustes statistic after each iterative step. Results of each iterative step show a
decrease in the sum of the Procrustes statistic over all panelists from 4.653 to 3.252. The point at which a change in the Procrustes statistic between iterations is less than a predetermined tolerance is when the iterative process is terminated. In this example, the difference in the total Procrustes statistic after iteration 4 (3.543) to iteration 5 (3.252) was 0.291. This difference is less than our predetermined tolerance of 0.300. Gower (1975) stopped the iterative process when differences in the sum of the Procrustes statistics were less than 0.0001 when examining data from three panelists’ evaluations of nine beef carcasses. Scaling factors computed after the fifth iteration (Table 13.6) show similar trends as scaling factors calculated after iteration 1 (Table 13.5).

The final consensus configuration is shown in Fig. 13.11 in which all transformed panelists’ configurations have now been combined to give a relative comparison of the samples. The final consensus configuration is the mean of all finally transformed configurations and can be used in place of the untransformed panel mean. The consensus configuration shows similar sample arrangement as Fig. 13.9.

### TABLE 13.6 Procrustes Statistics Calculated Before and After Iterations 1–5 and Final Scaling Factors

<table>
<thead>
<tr>
<th>Panelist</th>
<th>Iteration 1</th>
<th>Iteration 2</th>
<th>Iteration 3</th>
<th>Iteration 4</th>
<th>Iteration 5</th>
<th>Scaling Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.573</td>
<td>0.960</td>
<td>0.714</td>
<td>0.593</td>
<td>0.495</td>
<td>0.656</td>
</tr>
<tr>
<td>2</td>
<td>3.069</td>
<td>2.579</td>
<td>2.319</td>
<td>2.130</td>
<td>1.969</td>
<td>0.648</td>
</tr>
<tr>
<td>3</td>
<td>1.939</td>
<td>1.113</td>
<td>0.897</td>
<td>0.820</td>
<td>0.788</td>
<td>1.560</td>
</tr>
<tr>
<td>Total</td>
<td>6.581</td>
<td>4.653</td>
<td>3.930</td>
<td>3.543</td>
<td>3.252</td>
<td></td>
</tr>
</tbody>
</table>

**FIG. 13.10** Panelist configurations after five iterations. Final configuration after five iterations of simulated data set presented in three-dimensional space. Shapes represent breads 1–4. Sphere-bread, 1; cube-bread, 2; cone-bread, 3; and cylinder-bread, 4. The different shading of these shapes as well as labels P1–P3 represent panelists, where light gray is panelist 1 (P1), middle gray is panelist 2 (P2), and black is panelist 3 (P3). Axes represent attributes 1–3: X-hardness, Y-cohesiveness, and Z-springiness. The size of the symbol indicates its position with symbols that appear larger being closer to the viewer.

**FIG. 13.11** Final consensus configuration. Final consensus configuration of simulated data set presented in three-dimensional space. Shapes represent breads 1–4. Sphere-bread, 1; cube-bread, 2; cone-bread, 3; and cylinder-bread, 4. Axes represent attributes 1–3: X-hardness, Y-cohesiveness, and Z-springiness. The size of the symbol indicates its position with symbols that appear larger being closer to the viewer.
Table 13.7 gives the final rotation matrix for each panelist, which when multiplied to the initial panelist matrices, best matches the final consensus configuration. By examining the final rotation matrices, information about how each panelist used the terms to determine sample differences can be obtained. Those values for each attribute closest to 1 indicate good agreement with the transformed attributes. For example, the first transformed attribute was best described by panelists 2 and 3 as hardness because values of 0.984 for panelist 2 and 0.981 for panelist 3 were obtained. Panelist 1, however, used a combination of both hardness and cohesiveness to differentiate between the samples as suggested by a lower value of 0.862 for hardness and a higher absolute value for cohesiveness (−0.504). The second transformed attribute was cohesiveness in which panelists 1 and 2 had similar meaning with values of 0.857 and 0.985, respectively. Panelist 3 however may have perceived this attribute much differently or reversed the scale for this attribute as noted by the negative value (−0.973). The last transformed attribute was springiness. Springiness of the samples appears to be evaluated similarly by all the panelists as noted by the closeness of the values (0.988, 0.991, and 0.992, respectively) for the three panelists. If the panelists had used different, rather than common terms for the evaluations as in free-choice profiling, Procrustes analysis would allow one to determine which descriptors were used similarly.

Procrustes Statistic

The Procrustes statistic is not considered a transformation step, but when it is calculated, it can give an indication of the relative orientation of configurations from a target or consensus configuration as shown in our illustration. The Procrustes statistic can be calculated at any point in the transformation procedure: initially, after initialization, rotation/reflection, scaling or after each iterative procedure. As in the simulated data, calculating the Procrustes statistic

for each panelist after each step of the transformation can indicate the degree to which a particular transformation has affected each individual’s configuration. The Procrustes statistic calculated before and after translation can give an indication as to how panelists used the scale. For instance in our simulated data, a large difference between the Procrustes statistic value for panelist 1 compared to the other panelists indicated that a different level of the scale was used by panelist 1 than by panelist 2 or 3. In addition, after the last iteration, a large Procrustes statistic can indicate that a panelist is perceiving the samples differently than the other panelists.

The Procrustes statistic can be calculated not only for panelists as described above but also for samples. The Procrustes statistic for samples is calculated as the sum of squared distances for all panelists from the consensus positioning to an individual panelist’s final positioning for each sample. A small Procrustes statistic value for samples indicates a greater agreement among panelists with respect to the positioning of a given sample.

Applications of Procrustes Analysis to Sensory Data

Procrustes analysis, either pair-wise or generalized, can be an effective way of obtaining information about samples, their attributes, and the panelists assessing them. Although Procrustes analysis has most often been used for free-choice profiling data, it has other applications as well. In most of the studies described here, Procrustes analysis was used for multiple reasons: (1) to assess panelists’ performance; (2) to describe sample differences; (3) to relate instrumental data with sensory; (4) to compare methodologies; and (5) for acceptance and preference testing. These various applications will be described here.

Panelist Performance

Procrustes analysis has been used to provide performance information about individuals, which can be helpful in determining proper panel selection and training. Panelist performance information allows sensory scientists to measure a panelist’s consistency over replications and sessions. Researchers have found Procrustes analysis useful in determining panelist performance, as well as identifying discrepancies among panelists in their interpretation of terms and in their use of scales. Several examples of how Procrustes analysis has been used to obtain such information are given below.

Banfield and Harries (1975) compared the visual assessments of beef carcasses made by experts, trainees, and inexperienced panelists. Pair-wise Procrustes

---

**TABLE 13.7** Final Rotation Matrices for Individual Panelists

<table>
<thead>
<tr>
<th>Panelist</th>
<th>First</th>
<th>Transformed attributes</th>
<th>Second</th>
<th>Third</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H</td>
<td>C</td>
<td>S</td>
<td>H</td>
</tr>
<tr>
<td>1</td>
<td>0.862</td>
<td>−0.504</td>
<td>0.044</td>
<td>0.493</td>
</tr>
<tr>
<td>2</td>
<td>0.984</td>
<td>0.143</td>
<td>−0.108</td>
<td>−0.151</td>
</tr>
<tr>
<td>3</td>
<td>0.981</td>
<td>−0.194</td>
<td>0.011</td>
<td>−0.191</td>
</tr>
</tbody>
</table>

*Transformed attributes include hardness (H), cohesiveness (C), and springiness (S).*
analysis followed by principal coordinate analysis was used to provide information about choosing and training panelists. The Procrustes statistics for comparison of each pair of panelists (see Table 13.3) estimates the agreement within or among the groups. In this case, the large values for the five inexperienced panelists (6–10) show that they were less consistent among themselves and with the other group. The size of the Procrustes statistic can indicate whether a potential panelist approaches the ability of a trained judge. Visualizing panelists’ scores, using principal component analysis, while they are being trained can also be helpful in determining training progress.

When sensory profiling is used to describe product characteristics, the results are usually expressed as panel means. However, in free-choice profiling the panelists use different vocabularies, so Procrustes analysis provides a consensus configuration. The steps used in achieving the consensus (e.g., translation, rotation/reflection, and scaling) enables the experimenter to identify panelists’ differences in vocabulary usage, and to determine if panelists are in agreement. Through replications, panelists who are not consistent in the use of their scales can be identified (Langron, 1983).

Williams and Langron (1984) confirmed the effectiveness of generalized Procrustes analysis in interpretation of free-choice profiling data when port wines were evaluated by experts and nonexperts. The Procrustes statistics, in addition to principal coordinate analysis, showed that, overall, color was used more consistently by both expert and nonexpert tasters than aroma or flavor in differentiating between samples. When dimensions for individual panelist configurations were examined, however, it appeared that several panelists used all three attributes (color, aroma, and flavor) to differentiate between samples. Thus, a multidimensional configuration, such as that achieved with a Procrustes consensus configuration, indicates how panelists make their evaluations and distinguish between samples.

Danzart (1988) reported that Procrustes analysis was useful in detecting those descriptive analysis panelists with similar perceptions of different products. Danzart (1988) was able to group panelists who had similar perceptions in their evaluations of leg and breast of chickens from 12 farms for juiciness, tenderness, and flavor.

Replication of sample evaluations by free-choice profiling panelist provides a means of measuring consistency of responses (Marshall and Kirby, 1988). Comparison of the consensus configuration for panelists or samples can be visualized by principal coordinate analysis, and the Procrustes statistic for each individual panelist can identify anyone who is different. By examining differences in replications with Procrustes analysis, variations in panelist performance can be identified and accounted for (Harries and MacFie, 1976; Williams et al., 1981; Williams et al., 1984; Williams and Arnold, 1985; MacFie, 1987; Tunaley et al., 1988; McEwan et al., 1989).

Sample Information

One of the most important applications of Procrustes analysis is its use in distinguishing those attributes by which samples can be most clearly differentiated. This is in addition to monitoring panelist performance in evaluating the same samples.

Harries and MacFie (1976) used a subset of data from the Banfield and Harries (1975) study of 180 beef carcasses cited above to determine the interrelationship between attributes. Ten attributes were used by expert and inexperienced evaluators to visually assess the carcasses. Generalized Procrustes analysis of the data followed by principal coordinate analysis revealed that several of the attributes (points) were in close proximity. Close association of attributes suggests that some attributes can be combined or eliminated without affecting the evaluator’s ability to discriminate between the carcasses. For example, of the 10 attributes used by expert evaluators, attributes representing conformation of the rump and buttock were grouped similarly, as were the conformation of the loin and forerib. Harries and MacFie (1976) suggested that two of these four attributes be eliminated. For the inexperienced panelist a single attribute, overall conformation, was used in place of three other attributes (proportion of muscle and conformation of the buttock and rump) to discriminate between carcasses. By using Procrustes analysis, attributes for which panelists are most discriminating can be identified.

Differentiation between samples using generalized Procrustes analysis and principal coordinate analysis can be obtained from sensory profiling data of any type (texture, descriptive, flavor, free-choice). It is particularly useful when multiple attributes are evaluated and panelists use them differently to discriminate between samples. Examples can be found in a number of studies of products, including chocolate (McEwan and Thomson, 1988), sweeteners (Tunaley et al., 1988), and wines (Williams and Langron, 1984).

Sensory Data Versus Instrumental Measurements

Generalized Procrustes analysis can be applied to either instrumental data and/or sensory data to obtain a consensus configuration to which the other is compared. The comparisons between instrumental and sensory measurements provide a means of testing the reliability of instruments to provide sensory information about samples.

Williams and Langron (1983) were able to relate analytical color measurements to the appearance evaluation of eight commercial wines by free-choice profile panelists by using multiple regression. Vectors obtained from the correlations of the tristimulus color measurements (L, a, b) and the derived variables
(saturation and hue angle) with the transformed attributes were superimposed on the consensus sample plot (Arnold and Williams, 1986). By examining the length of these vectors and the direction in which they were pointing, information about the relationship of the samples with the tristimulus color measurements could be examined. Even though panelists used different words to describe the appearance of the ports, they were still able to evaluate similar color properties of the samples.

Daget et al. (1983) compared the texture properties of gel systems by both quantitative descriptive analysis and by mechanical measurements. Data from sensory and mechanical measurements were examined using canonical analysis and principal component analysis and then compared using Procrustes analysis. By applying Procrustes analysis to the data, it was determined to what extent each method was reproducible. In addition, information about the sensitivity of each method was provided, as well as how related terms used by panelists described the physical properties of the gels.

Marshall and Kirby (1988) examined the composition (fat and moisture) of unflavored cheese analogs and related these variables to the texture data obtained by free-choice profiling. Calculated consensus scores for each cheese for the first two principal component axes of the consensus configuration were derived from the generalized Procrustes analysis for each replicate. These scores were then regressed separately against the percent of fat and moisture content of nonfat solids. The generalized Procrustes analyses of each individual's calculated scores gave information about each panelist's sensitivity to changes in composition. The calculated consensus scores for each replicate were related to the known moisture content of nonfat solids and fat content of cheeses, and suggested that panelists were most sensitive to changes in the moisture content of nonfat solids.

Sensory Methodology Comparisons

Consensus configurations for samples evaluated by different sensory methods can be compared through the use of the Procrustes analysis. This includes comparisons of descriptive analytical techniques, as well as different mathematical treatments of the same data. Procrustes analysis can be used to validate results obtained by different methods.

Langron et al. (1984) showed that the consensus configuration from generalized Procrustes analysis was more meaningful than untransformed panel means from sensory profiling of an apple cultivar during storage. This further suggests that descriptive analysis data could be subjected to general Procrustes analysis.

In a study of three methods of evaluating coffees, Williams and Arnold (1984) noted that generalized Procrustes analysis was helpful when comparing conventional descriptive profiling with free-choice profiling. Grouping of samples was similar with both methods, and free-choice profile panelists seemed to be in greater agreement than descriptive panelists. However, this was not clearly demonstrated.

McEwan et al. (1989) suggested that "conventional" free-choice profiling of chocolates was preferred to a structured free-choice technique based on Kelly's repertory grid method (Kelly, 1955). The conclusion was based on the similarity of results obtained by the two methods and the greater simplicity of the conventional method.

Procedures for conducting sensory tests have also been compared using Procrustes analysis. For example, in a study by Williams et al. (1984), Procrustes analysis was used to determine how the assessment of aroma and flavor of 10 wines by 10 panelists differed when the ability to see the product was varied by presenting samples in red or clear glass. Generalized Procrustes analysis was applied to scores averaged over replications for each panelist and for each attribute. Sample and panelists consensus configurations were obtained for each of five evaluations (appearance in red glass; aroma and flavor in red and clear glass). Pair-wise Procrustes analysis was then applied to the five sample configurations in which appearance was compared with aroma and flavor evaluated in clear and red glass. The Procrustes statistic was calculated and principal coordinate analysis applied to provide a visual representation of the relationship between the five evaluations. Results showed that the panel aroma and flavor scores differed depending upon whether the appearance of the wines could be evaluated. Scores from individuals varied, indicating how each panelist may have been influenced differently.

Acceptance and Preference Testing

The ultimate goal of all sensory evaluations is to relate panel evaluations of a product and its attributes to the way consumers perceive and finally accept that product. Once this relationship is found, the important product attributes may be optimized.

Laslett and Bremer (1979) determined the relationship between scores for six attributes (aroma, off-aroma, flavor, off-flavor, toughness, and moistness) of fish minces and acceptability as evaluated by the same panel. Generalized Procrustes analysis was used to determine if the panelists were evaluating the fish minces in similar ways. The consensus configuration for the panelists indicated that the panelists were similar so the mean attribute scores could be used to determine the relationship between acceptance and sensory scores.

Williams et al. (1988) addressed some of the problems encountered when trying to relate sensory information with acceptance data. They found that by using Procrustes analysis, a sample consensus configuration could be derived in
which dimensions that relate to acceptability could be superimposed. In addition, subgroups of panelists who perceive samples similarly could also be determined. The importance of bridging the gap between objective information and hedonic information in determining what aspects of a product are significant to consumers when they select products and how such characteristics may be optimized was reemphasized.

Benedict et al. (1988) discussed a new procedure using generalized Procrustes analysis integrated with a technique called "natural grouping" in which individual subjects divided a set of products based on their perceived similarities until no further separations were possible. After each partitioning each individual described the attributes which separated each group of products. By collecting data in this manner a completely individualized perceptual configuration was obtained for each individual for the same set of products. These configurations are referred to as "perceptual maps" and can provide some insight into consumers' perceptions of food products. Generalized Procrustes analysis was used to construct a consensus of all the individual configurations which allowed for easier interpretation of the data that took into account individual variation.

CONCLUSIONS

Procrustes analysis is a multivariate statistical tool that allows individual panelists configurations, such as those obtained from free-choice profiling, to be matched and compared by initialization, rotation/reflection, and scaling. The final transformed configurations obtained from this analysis can be combined to form a consensus configuration that can be used in place of sample means that are traditionally averaged across panelists and replications.

Free-choice profiling differs from conventional descriptive analysis in the use of terms, in the development of individual scorecards, and in the analysis of the data. Free-choice profiling is newer than other more conventional descriptive analysis techniques, so procedures have not been standardized. However, researchers have reported shorter training times than that required by other profiling methods. In addition, because panelists are evaluating samples on an individual basis, variations associated with forcing conformity in term and scale use by panel members are alleviated.

By using free-choice profiling and analyzing the resulting data using Procrustes analysis, sample information, as well as panelist information can be provided. Panelists responses can then be evaluated and outliers identified. One note of caution in using free-choice profiling and applying Procrustes analysis to data is that a false sense of security may be obtained from its use. The analyst must still examine the data and judge whether it is meaningful within the context of the test. Sensory professionals must not be misled by the power of this statistical tool. Training is indeed very important in getting accurate and reliable information about a product. Free-choice profile panelists still must be trained in the use of scales, and to be consistent.

Procrustes analysis is a powerful technique that has not only been used to analyze free-choice profiling data, but descriptive analysis data as well. Other statistical methods used in conjunction with Procrustes analysis, such as principal component and principal coordinate analysis, can help interpret results obtained from applying Procrustes analysis to sensory data. From its application, information about the performance of panelists and differences and similarities between samples can be identified. In addition, it has been used to compare different sensory methodologies. Recently, Procrustes analysis has been helpful in providing insight to consumer acceptance and preferences when compared to trained panel responses. Procrustes analysis can be a valuable tool to the sensory scientist who understands not only the way in which it can be applied to sensory data, but also its limitations.

ACKNOWLEDGMENTS

Special appreciation goes to Donald J. Colby for construction of the three-dimensional data plots with the cooperation of the National Center for Supercomputing Applications and the Beckman Institute on the University of Illinois Urbana-Champaign campus. Dolores Oreskovich gratefully acknowledges the support of the Ford Foundation, sponsor of the Ford Foundation Doctoral Fellowship, the National Pork Council Women, sponsor of The American Home Economics National Pork Fellowship, and the National Hispanic Scholarship Fund. This work was part of Project No. 60-326 of the Agricultural Experiment Station, College of Agriculture, University of Illinois at Urbana-Champaign.

REFERENCES


Mosier, C. I. 1939. Determining a simple structure when loadings for certain tests are known. Psychometrika 4: 149.

Sens. Studies 1: 55.

Oreskovich, D. C., Klein, B. P., and Sutherland, J. W. 1990. Variances
associated with descriptive analysis and free-choice profiling of frankfurters.
Annual Meeting of the Institute of Food Technologists, Anaheim, CA. Ab-
stracts.

Oreskovich, D. C., Klein, B. P., and Sutherland, J. W. In preparation. Develop-
ment of a standardized language for evaluating the flavor and texture of pork
by descriptive analysis and free-choice profiling.

optimal agreement. Psychometrika 53: 199.

Piggot, J. R. and Canaway, P. R. 1981. Finding the word for it—Methods and

Piggot, J. R. and Sharman, K. 1986. Methods to aid interpretation of multi-
dimensional data. In Statistical Procedures in Food Research (J. R. Piggott,

Food Technol. 38(6): 74.

Science, London.

Procrustes-PC V2.0. Computer software. Oliemans Punter and Partners, Inc. SG
Utrecht, The Netherlands.

Schönenmann, P. H. 1966. A generalized solution of the orthogonal Procrustes

Psychometrika 33: 19.

Schönenmann, P. H. and Carroll, R. M. 1970. Fitting one matrix to another under
choice of a central dilation and a rigid motion. Psychometrika 35: 245.

Orlando, FL.

Sensory evaluation by quantitative descriptive analysis. Food Technol.

28: 385.

standard rating scales for mechanical parameters of texture and correlation
between the objective and the sensory methods of texture evaluation. J. Food
Sci. 28: 397.

TenBerge, J. M. F. 1977. Orthogonal Procrustes rotation for two or more

agreement for two or more matrices of different column orders. Psychometrika
49: 49.

the relationship between preference and sensory characteristics of nine

Williams, A. A. and Arnold, G. M. 1984. A new approach to the sensory
analysis of foods and beverages. In Progress in Flavour Research 1984,
Proceedings of the 4th Weurman Flavour Research Symposium (J. Adda,

coffees characterised by conventional profiling, free choice profiling and

Evaluating masters' performance in the profiling of foods and beverages. In

Williams, A. A. and Carter, C. S. 1977. A language and procedure for the
sensory assessment of Cox's Orange Pippin apples. J. Sci. Food Agric. 28:
1090.

analysis. In Flavour of Distilled Beverages: Origin & Development (J. R.

the evaluation of commercial ports. J. Sci. Food Agric. 35: 558.


physical and sensory data in food acceptance studies. Food Quality and
Preference 1: 25.