Data Theory and Scaling Methods

In: Data Theory and Dimensional Analysis

By: William G. Jacoby

Pub. Date: 2011
Access Date: January 29, 2018
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks
Print ISBN: 9780803941786
Online ISBN: 9781412983860
DOI: http://dx.doi.org/10.4135/9781412983860
Print pages: 39-74

Data Theory and Scaling Methods

The main purpose for a theory of data is to rationalize the use of scaling procedures. In other words, there are many different scaling methods available to the researcher. But which one is appropriate in any given situation? Data theory is intended to bring a comprehensive perspective to bear on this question. The choice of a scaling procedure always depends upon the researcher's interpretation of the observations—that is, the nature of the data.

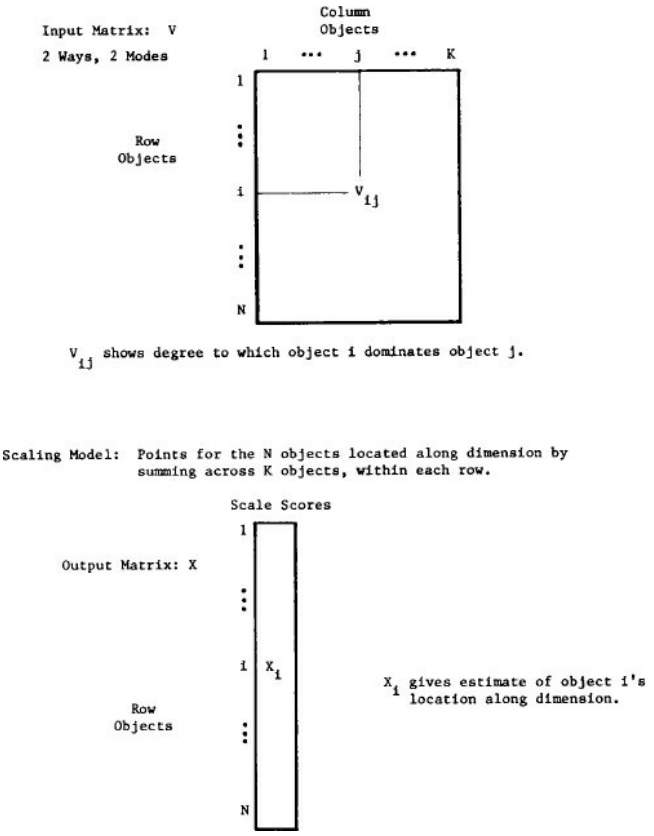
In this chapter, we will examine the various types of data, and discuss some of the scaling procedures that are appropriate for each one. Note that the emphasis will be on scaling *models* rather than scaling *techniques*. That is, we will focus on abstract representations of variability within observations, rather than the specific details involved in constructing these representations. For the latter, readers will be referred to other sources, such as the appropriate volumes in this monograph series.

Single Stimulus Data

Single stimulus data are, by far, the most common of the four types. Therefore, many different scaling procedures have been devised to handle them. We will examine three models in this section: the summated rating model, the cumulative model, and the factor analysis model.

Summated Rating Scales. Summated rating scales simplify the representation of empirical observations by summing across the levels in at least one of the ways of a multiway, multimode data matrix. Figure 5.1 shows the most common situation: An *N* by *K* matrix is scaled by summing across the columns, within rows of the matrix. This results in the *N* by 1 matrix of empirical scale score for the objects represented by the rows. There are several versions of summated rating scales; the best known are Likert scales and Thurstone's method of equal-appearing intervals (McIver and Carmines, 1981). Although these are usually treated as different scaling *techniques*, they are actually different manifestations of the same basic model.[9]

Figure 5.1. Summated Rating Scale



From a geometric perspective, the summated rating approach assumes that one set of points varies systematically with respect to the dimension, whereas the second set of points fluctuates randomly. For each object in the first point set, we sum across all of the objects in the second point set, so that the fluctuations cancel each other out, thereby providing an accurate estimate of the first point's location along the underlying dimension. Note that the points representing the objects in the nonscaled set are not fixed along the dimension; they can vary markedly from one row of the data matrix to the next.

From a measurement perspective, the individual items that are summed to produce the scale are ordinal-level functions of the latent dimension. In other words, the numeric values assigned to the rows of the data matrix are all monotonically related to the underlying characteristic. Summing across the items implies summing across the functions, as well. The summed monotone functions should be *linear*, because the idiosyncracies of the item-specific monotone functions should cancel each other out. And, as explained back in Chapter 2, if there is a specific function (linear, in this case) between the underlying characteristic and an empirical set of numeric assignments, then interval-level measurement has been achieved.

The summated rating model depends entirely on the assumption of random fluctuations across the items that are summed to create the scale. The criteria established for selecting items used

in Thurstone and Likert scales (e.g., Edwards, 1957) can be viewed as strategies for trying to insure the randomness of interitem differences. If this assumption is met, then the summated rating approach is a very powerful scaling technique. However, this same assumption leads to the two limitations of the sum-mated rating model.

First, the method assumes that *all* errors are attributable to random fluctuations in one set of points. However, they could actually occur for other reasons, such as the simultaneous influence of several underlying dimensions. But the summated rating model dismisses the possibility of multidimensionality on a priori grounds. In strictly practical terms, this means that the scale could appear to fit the data quite well (i.e., the reliability coefficient would take on a value close to 1.0), even though the “true” sources of variation stem from several underlying dimensions. Accordingly, the summated rating approach is very useful as a scaling technique, but poor as a scaling criterion.

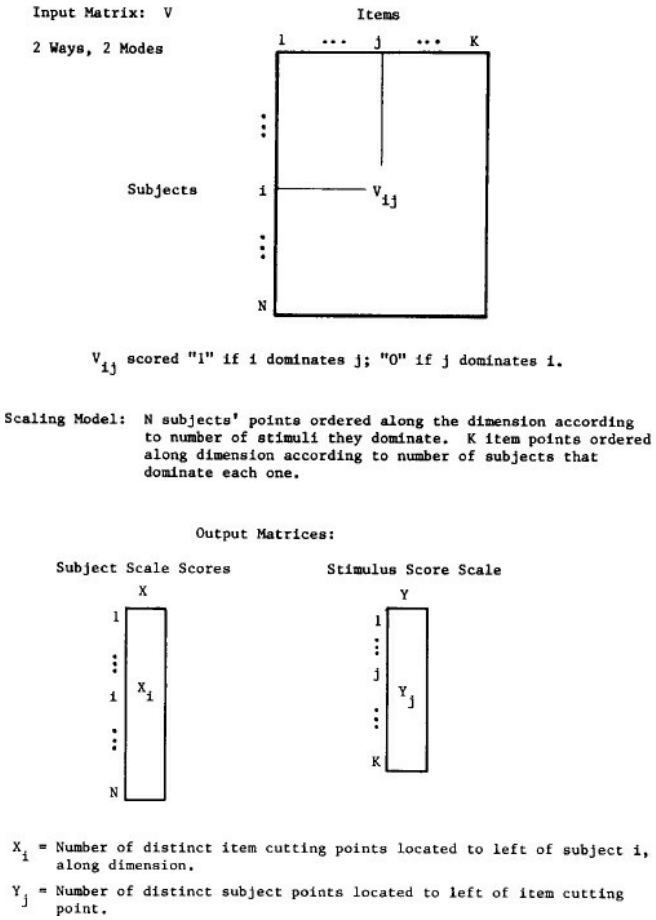
The second limitation of the summated rating approach is that it only scales a single set of points, out of the two sets that always constitute single stimulus data. This must be the case, because the second point set is only assumed to vary randomly; therefore, precise estimates for these items' locations would be meaningless. Exactly *which* point set is scaled depends entirely upon the researcher's assumptions about the data and the analytic objectives. It is possible to scale subjects, while assuming that items are randomly different replications (as in the Likert scaling approach), but it is equally possible to scale items, by assuming that subjects only differ randomly with respect to the items (as in Thurstone's equal-appearing intervals approach). In any event, the ability to “collapse” one of the ways of the data matrix presupposes that any differences across levels within that way are uninteresting in substantive terms.

Cumulative Scales. In contrast to the preceding approach, the cumulative scaling model does locate both sets of points along the underlying dimension. This is accomplished by using a more stringent scaling model. The most common version of the cumulative model is a Guttman scale (McIver and Carmines, 1981), which is usually applied to a rectangular data matrix showing *N* subjects' responses to *K* dichotomous items (see Figure 5.2). The items and the subjects are both assumed to occupy fixed positions along the underlying dimension. This assumption, combined with the dominance relationship implied by single stimulus data, leads to an ordering of subject and item points based upon the cumulation of responses. A cumulative scale simplifies the information contained in the data by reducing the single *N* by *K* input matrix to two separate vectors: An *N* by 1 vector of subject scale locations, and a *K* by 1 vector of the item scale locations.

In geometric terms, each cell entry in the dichotomous data matrix provides information about a single subject-item point pair. If the cell entry is 1 then the subject point is located to the right of (i.e., it dominates) the item point along the underlying dimension. If the entry is 0 then the item point is located to the right of the subject point (that is, the item dominates the subject). Each point is located along the dimension by summing the number of 1s contained within its row or column of the data matrix. Thus the larger the number of positive responses given by a subject, the farther to the right that point is located, because it dominates a larger number of items. And the larger the number of positive responses given to an item, the farther to the left that point is located, because the item is dominated by a larger number of subjects.

A Guttman scale of dichotomous data is the simplest operationalization of the cumulative scaling model. But the model can be generalized to handle a variety of other situations. For example, the cumulative approach can easily be adapted for polychotomous, ordinal items (Coombs, 1964; Edwards, 1957; Torgerson, 1958). The point representing an item is actually a "cutting point" along the underlying dimension, in the sense that it serves as the boundary between negative and positive responses on the item. In other words, if subjects give negative responses to the item, then their points are located to the left of the cutting point, whereas positive responses correspond to subject points on the right of the cutting point. Dichotomies only have a single cutting point per item, so locating the cutting points seems to be equivalent to locating the items themselves. Polychotomous items simply have more than one cutting point; there is a separate cutting point between each pair of adjacent response categories. If an item has *Q* categories, then there will be *Q* - 1 cutting points for that item.

Figure 5.2. Cumulative (Guttman) Scale

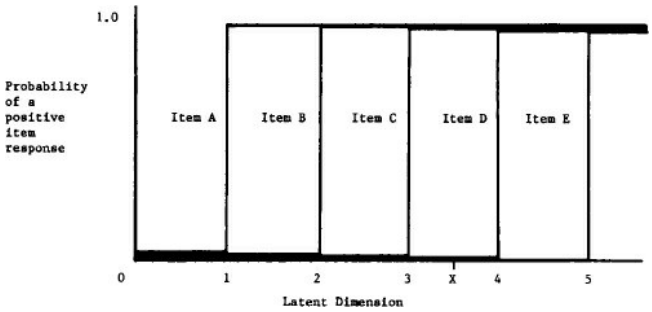


In geometric terms, each item is still a replication of the underlying dimension. However, a polychotomy divides the dimension into *Q* segments, rather than just two segments, as was the case with dichotomous items. A subject's score on an item will automatically pin down that subject's point location relative to all of the item's cutting points; thus it is easy to break each item up into *Q* - 1 dichotomies. The scalogram is then constructed exactly as it was in the case of dichotomous items. The only difference is that the columns of the data matrix now explicitly represent cutting points, rather than the items themselves. If an item has more than two possible responses, it will require more than a single column in the matrix.[10]

The most serious limitation of the traditional Guttman scale approach is that it is based upon a strictly deterministic model. It makes no provisions to explain deviations from perfectly scalable response patterns; it has no theory of scaling errors. Scaling errors occur when a subject gives a negative response to an item with a point that is located to the left of the subject's own point, or when a positive response is given to an item with a point located to the right of the subject's point. In either case, such responses are contradictions, given the geometry of the scaling solution. In the standard Guttman scaling analysis, errors are unexplainable; we merely acknowledge that such responses do exist, and try to develop methods for assigning scale

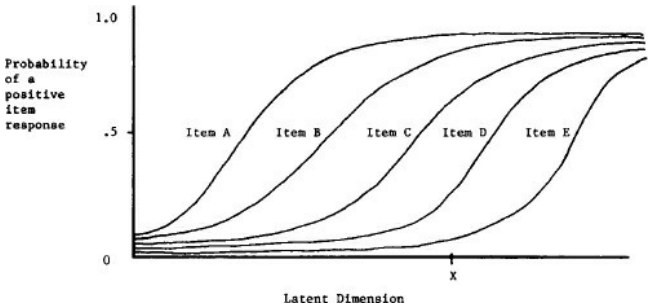
scores to observations that exhibit these kinds of errors. But they remain as gaps in our explanation of the variability in the data.

Figure 5.3. Trace Lines for Items in Deterministic Scaling Model



In order to deal with exactly this limitation, probabilistic forms of the cumulative model have been developed. Figures 5.3 and 5.4 show the difference between the deterministic and probabilistic versions of a cumulative scale. The figures give two versions of the trace lines for five items that form a scale. An item's trace lines show the probability of a positive response on that item, at each position along the underlying dimension. Figure 5.3 shows the deterministic model. If a subject's point is located at a higher position (i.e., farther to the right) than an item point, then the subject *must* give a positive response to that item. There is absolutely no provision for a negative response on an item that falls to the left of the subject point. For example, the point labeled X in Figure 5.3 represents a subject; this person must give a positive response to items A, B, and C because the trace lines for those items show that the probability of a positive response, at the scale position corresponding to X, is identically 1.0 for all three of the items.

Figure 5.4. Trace Lines for Items in Probabilistic Scaling Model



But this is too stringent a standard for the real world, precisely because scaling models never fit perfectly. Therefore, the trace lines in Figure 5.4 show a probabilistic version of the same scale. Here, the probability of a positive response for any item increases as one moves to higher positions along the dimension. Thus a subject with a point located at position X in the figure is *likely* to give positive responses to items A, B, and C, and negative responses to items D and E. However, there are still nonzero probabilities of the opposite responses on each of the five

items. The varying scale locations of the items are still clearly discernible in the different positions of the separate item trace lines. The cumulative nature of the model is represented by the relations among the probabilities. For any arbitrary subject point along the dimension, the following will always be true:

Prob(A = 1) ≥ Prob(B = 1) ≥ Prob(C = 1) ≥ Prob(D = 1) ≥ Prob(E = 1)

Thus the probabilities are:

(1)
monotonically related to the position along the underlying dimension; as the item possesses more of the attribute corresponding to the latent dimension, the trace line is shifted to the right (i.e., to a higher numerical position)
(2)
monotonically related to each other, at any position along the latent dimension; regardless of the subject's location, a positive response on item A is more likely than a positive response on B, and so on

There are several techniques available for estimating the probabilistic version of the cumulative scale model, including Rasch models (Andrich, 1988), Mokken scales (Mokken, 1970; Mokken and Lewis, 1982; Niemoller and van Schuur, 1983), and latent class models (McCutcheon, 1987). They all represent improvements over the deterministic approach because they explicitly allow for empirical deviations from the “perfect” form of the cumulative scale. Furthermore, they all evaluate the fit of the scaling model against the explicit null hypothesis that the separate item responses are statistically independent of each other. That would be the situation if no latent dimension existed to impose structure on the data. This kind of explicit hypothesis testing is preferable to the ad hoc decisions and rules of thumb (e.g., “the coefficient of reproducibility must be greater than .90”) that usually guide judgments about the fit of the cumulative model to a data set.

In strictly practical terms, probabilistic cumulative scaling techniques often lead to the conclusion that a latent dimension exists when the opposite conclusion would have been reached using the traditional Guttman scaling standards. This illustrates the trade-off between the model and the data: As the researcher is willing to make more explicit assumptions about the nature of the errors, it becomes possible to accept greater discrepancies between the perfect scale patterns implied by the scaling model and the empirical data themselves. Thus probabilistic versions of the cumulative scaling model are not only more realistic than the traditional Guttman approach, they also allow scales to be constructed more readily from error-

laden data.

Let us next consider the problem of multidimensionality and the cumulative scaling model. Stated simply, the question is, can scaling procedures based upon this model recover the dimensional structure underlying a set of data when the latter are simultaneously affected by *several* independent sources of variability? The general answer to this question is no (Coombs, 1964). The goodness of fit measures used to test the cumulative scaling model can indicate when a single dimension is inadequate to represent the data. However, they cannot be used to discern the reasons that this might occur (i.e., is it lack of structure, measurement error, or a more complicated dimensional structure?). Furthermore, if a unidimensional model is applied to a situation involving multidimensional phenomena, then the scaling technique will usually not successfully estimate *any* of the underlying attributes with any degree of accuracy (Coombs, 1964).[11]

In general, the cumulative scaling approaches work best when applied to data that conform to a unidimensional model in which both sets of objects contained in the data vary systematically according to a single attribute. If the researcher suspects either (1) that only one of the object sets contains systematic variation, or (2) that there are multiple sources of variability in a set of observations, then some other scaling approach would probably prove more useful.

Factor Analysis. Factor analysis is another procedure that can be used to scale single stimulus data. It models responses on a set of empirical items as manifestations of variability along one or more latent dimensions. Thus factor analysis has objectives that are similar to, but more general than, those in summated rating scales and cumulative scales. However, the method employed to estimate the factor analysis model is markedly different from the other two. In this section, we will examine the basic geometry of the factor analysis model. The methods used to estimate this model from empirical data are covered in detail in many sources, including Kim and Mueller (1978a, 1978b) and Long (1983).

As shown in Figure 5.5, factor analysis is applied to an N by K data matrix (usually, observations by variables). The first step of the analysis is usually to obtain correlations between the levels of one mode—usually the variables. These correlations are arranged in a square matrix of order K. The entries in the main diagonal are called *communalities;* they represent the proportion of the total variance in each observed variable that is "explained by" the latent dimensions.[12] The scaling technique decomposes (or "factors") the information in the correlation matrix, to produce a new K by M "factor pattern matrix."[13] This shows the K observed variables' dependencies on M underlying dimensions (which are usually called

"factors").[14] The scaling analysis sometimes stops at this point. But, it can proceed by generating an *M* by *K* matrix of "factor scoring coefficients." The latter estimate the *M* latent dimensions as linear functions of the observed variables. The original *N* by *K* data matrix can be multiplied by the factor scoring matrix to produce an N by *M* matrix of "factor scores." These estimate the other mode's (i.e., the observations, if the variables were factored) locations along the underlying factors. In this manner, a factor analysis can be used to scale both sets of points contained in a set of single stimulus data.

The geometric model for factor analysis represents the objects to be scaled (usually, the variables) as vectors within a space formed by the elements in the other mode of the data matrix (an "observation space"). Figure 5.6 shows a simple example of a standard scatter diagram. Here, the axes represent variables (thus, a variable space), and the plotted points represent two observations. Of course, these points are located in the space according to the observations' values on the two variables. Now consider Figure 5.7. Here, the axes and the points are switched. Each observation is an axis, and the plotted points represent the variables. Furthermore, the plotted point is now shown as a vector—a directed line segment, beginning at the origin of the space and terminating at the plotted point. The locations of the variable vectors are determined by their values on each of the observation-axes.

Figure 5.5. Factor Analysis

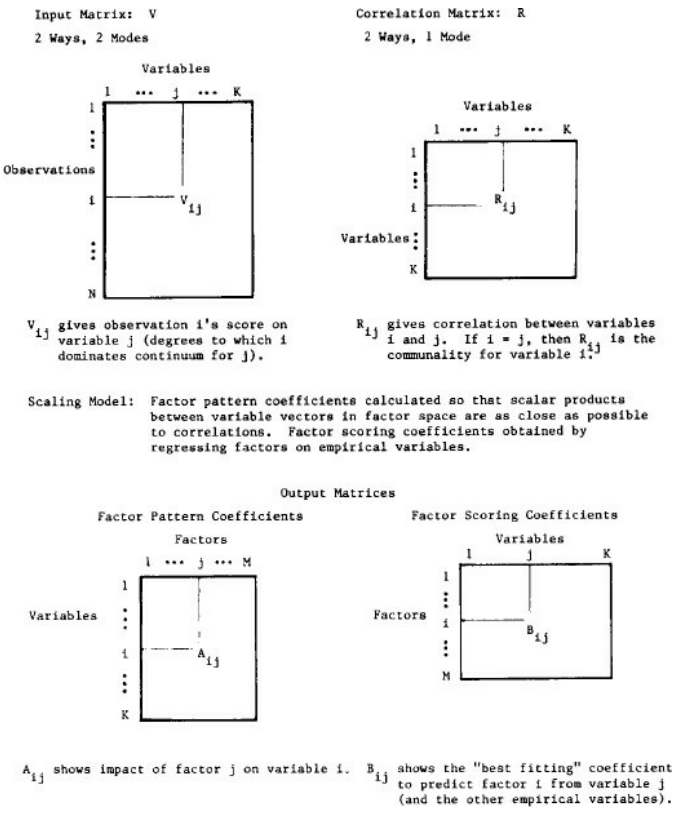
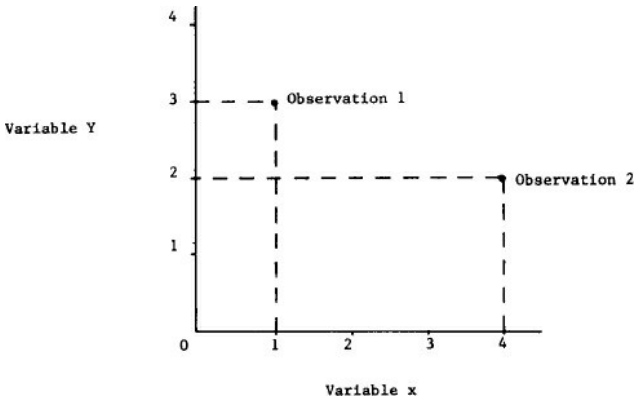
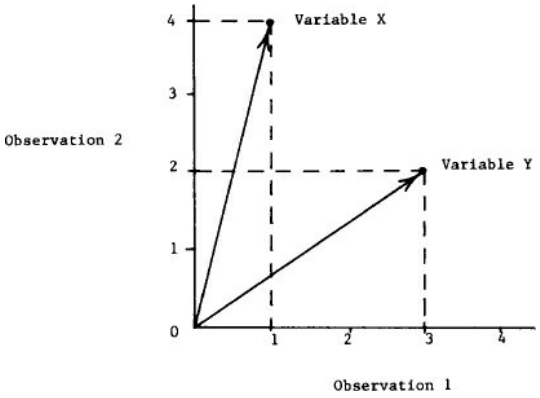


Figure 5.6. Vector Representation of Observation Points in Variable Space



Of course, Figures 5.6 and 5.7 present exactly the same information; they merely show it in different ways. This is an extremely simple example, because there are only two observations. Generally, there should be *N* perpendicular axes, and each variable vector would have coordinates on each of these axes. Thus the orientation of each vector completely summarizes the information contained in that variable.

Figure 5.7. Vector Representation of Variable Vectors in Observation Space



When the variable values are expressed as deviations from their respective means, and when each value is divided by the square root of *N* (the total number of axes in the space), some very useful results occur. First, the variable means are all located at the origin of the space. Second, the squared length of each vector is equal to the variance of the corresponding variable; of course, this also means that the length is the standard deviation. Third, the cosine of the angle between any two vectors is equal to the correlation between those two variables; the smaller the angle, the higher the correlation, and vice versa. A right angle (90 degrees) means that the two variables are uncorrelated (or orthogonal).[15]

Geometrically, the factor analysis begins with *K* variable vectors located within the *N*-space formed by the observations. The objective of the scaling technique is to simplify the representation of the variables: Instead of *K* vectors in an *N*-space (where *N* is probably a very

large number), we try to locate the K variable vectors within a subspace defined by M new vectors, or "factors" (where M is much smaller than N and, it is hoped, smaller than K). Of course, it will probably be impossible to find a factor subspace that completely contains the variable vectors. This means that the factors cannot account for all of the variance in the observed variables. The factor analysis model handles this by assuming that each observed variable has its own unique component. The component contributes to the variable's variance, but it is entirely separate from (i.e., uncorrelated with) the factors and the other observed variables. This unique component plays a very important role, because it accounts for scaling errors in a factor analysis solution; it explains why the observed variables are not perfectly predictable on the basis of the underlying factors alone.

Figures 5.8 and 5.9 show the geometry of a hypothetical factor analysis solution, involving four observed variables (X_1 through X_4) and two uncorrelated factors (F_1 and F_2). Figure 5.8 shows the subspace formed by the factors.[16] The four vectors shown in this subspace are not the variables themselves. Instead, they are "shadows" of the variables, formed by taking the perpendicular projections from the tips of the respective variable vectors into the factor space. The squared lengths of these shadow vectors are the communalities. The shadow vectors' coordinates on the factor axes show the linear dependencies of the variables on the factors; the larger the value of each coordinate, the stronger the impact of that factor on that variable. These coordinates are called *factor pattern coefficients*. Figure 5.8 shows a plot of the factor pattern for these variables and factors; it contains a great deal of information. Thus we can see that the factors account for twice as much variance in X_1 as X_2, because the X_1 vector is twice as long as that for X_2. At the same time, the second (vertical) factor has a much stronger influence on X_1 than the first factor; the b_{12} coordinate is farther away from the origin (i.e., its numeric value is greater) than the b_{11} coordinate. Figure 5.9 shows the entire vector for variable X_2, and its relation to the factor space. The full vector is contained in a three-dimensional space; two of its dimensions are formed by the F_1, F_2 plane, and the third corresponds to U_2, the unique component for variable X_2. The component's impact on X_2 is measured by the coordinate on the U_2 axis (d_2). In order to graph the X_2 vector completely, we must take all three axes into account; this is the geometric equivalent of saying that the total variance in an observed variable depends upon its unique component, as well as the factors. Figures 5.8 and 5.9 clearly show that the observed variables' dependencies on the factors are modeled by the orientations of the variable vectors relative to the factor axes; the closer the former comes to the latter, the stronger the impact of the factors on the variables.

Figure 5.8. Geometric Model for Factor Analysis (Two-Factor Space, Containing Vectors for Empirical Variables—Factors Are Orthogonal)

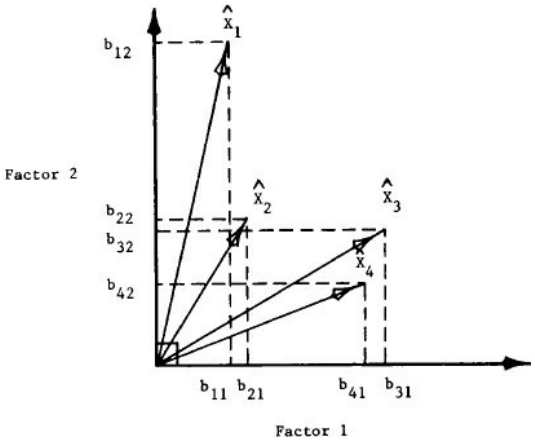
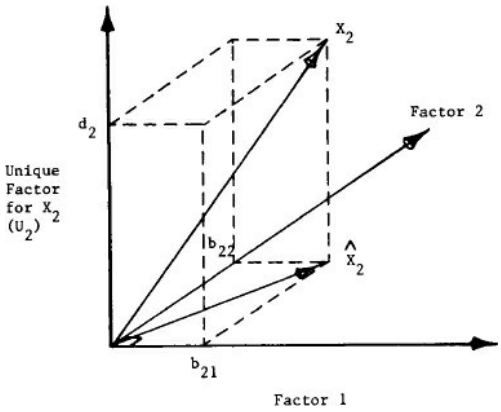


Figure 5.9. Geometric Model for Factor Analysis—Space for Variable X_2



But how do we determine the subspace containing the factor axes relative to the variable vectors? The factor subspace is located by using the following criterion: The scalar products between each pair of shadow vectors should come as close as possible to the observed correlations between the respective pairs of variables.[17] Ideally, the scalar products should equal the correlations. In this sense, the factor space explains the correlations between the empirical variables. And it divides the total observed variance in each variable into two parts: (a) The communality, which lies in the factor space and is therefore shared with the other variables; and (b) the uniqueness, which is outside of and uncorrelated with the factor space—therefore, it is the part of the variable's variance that is not shared with any other variables (observed or unobserved).

The scalar product criterion for fitting the factor model leads to some of the troublesome indeterminacies involved in this model. The scalar product of two vectors depends entirely upon the orientation of those vectors relative to each other, and *not* upon the vectors' locations relative to the axes of the space. This means that for any arbitrary number of factors there is an

infinite number of factor solutions, corresponding to all of the possible orientations of the factor axes relative to the shadow vectors (i.e., the rotation problem). Similarly, the factor axes may be at right angles to each other (i.e., an orthogonal factor solution), or they may form some other angles (an oblique solution). Once again, the scalar products between the variables' shadow vectors will remain unchanged.

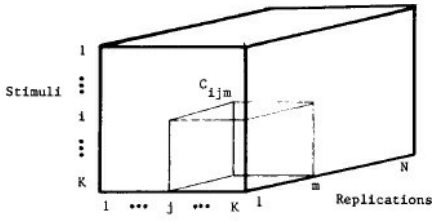
Factor analysis can be used to determine whether *K* variables and *N* observations can be modeled as points in *some M*-dimensional factor space. However, the exact nature of this space must be discerned on the basis of criteria that are essentially separate from the factor analysis itself (e.g., simple structure guidelines for rotation, and substantive theory for correlated versus uncorrelated factors). This inability to use an "objective procedure" to pin down a precise scaling solution is often very frustrating (Steiger and Schonemann, 1978). However, the problem is inescapable, and it should not cause potential users to overlook a research tool that can be very useful: Factor analysis still provides useful insights into the common sources of variability underlying objects.

Stimulus Comparison Data

Stimulus comparison data are relatively uncommon in nonexperimental settings.[18] And when they do arise, stimulus comparisons are often uninteresting for scaling purposes. After all, if the objects are compared to each other using a common standard, then the very existence of the standard implies that the objects can be located along a dimension. As a result, further measurement or scaling operations may be unnecessary. It is only when the stimulus comparisons fail to provide enough information in themselves that scaling methods must be employed. Although there are a few methods for analyzing stimulus comparison data, they are among the least-used scaling models considered in this chapter.

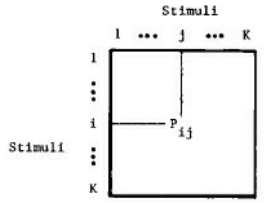
Scaling Methods for Stimulus Comparison Data. The method of paired comparisons, initially developed by L. L. Thurstone, uses a scaling model that is appropriate for stimulus comparisons (e.g., Edwards, 1957). The model assumes that the stimulus objects are represented as probability distributions, rather than fixed points, along the latent dimension. When one object is compared to another, single points are sampled from each of their distributions. The point estimate for each object corresponds to the mean of that object's distribution. Thurstone's contribution was a method for estimating the relative locations of the stimulus means, using the observed probabilities of one stimulus dominating another.

Figure 5.10. Paired Comparisons Scale (Input Matrix)



C_{ijm} is "1" if stimulus i dominates stimulus j on replication m, "0" if j dominates i on that replication.

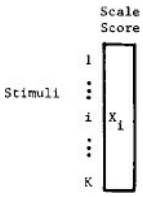
Figure 5.11. Paired Comparisons Scale (Probability Matrix)



P_{ij} is the probability that stimulus i dominates stimulus j.

Scaling Model: Distance between stimulus points i and j are a function of probability, P_{ij}.

Figure 5.12. Paired Comparisons Scale (Output Matrix)

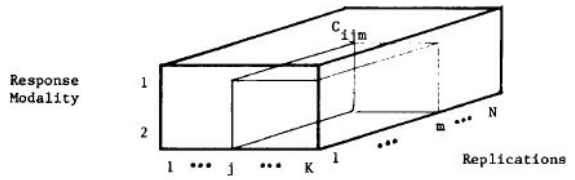


X_i gives estimated location of point for stimulus i, along latent dimension.

The method of paired comparisons begins with a three-way, two-mode matrix of dichotomous data, as shown in Figure 5.10. The first two ways correspond to the *K* stimuli, so there are *K* levels in each way. The cell entries show whether the row stimulus dominates the column stimulus, or vice versa. The third way, and second mode, consists of *N* replications for each pair of stimuli. In many applications, the stimuli pairs correspond to *N* subjects. The scaling technique begins by collapsing the data matrix across the levels of the third way, resulting in the two-way, single-mode matrix shown in Figure 5.11. Now, the cell entries are interpreted as the probabilities that the row stimuli dominate the column stimuli. Thurstone's method is applied directly to this matrix of probabilities to produce a square matrix of estimated distances between the *K* stimulus points (which are, themselves, the means of the distributions). Finally, one stimulus is arbitrarily placed at the origin of the dimension, and the scale values of the remaining stimuli are calculated according to their distances from this stimulus; this produces the *K* by 1 vector of scale scores shown in Figure 5.12. The resultant scale is related to the probabilities, usually by the inverse of the normal density function (i.e., *Z*-scores are obtained by “working backwards” from the probabilities). The fact that there is a specific relationship between observations and scores, along with the arbitrary location of the origin, means that the paired comparison method produces interval level measurement.

Psychophysical magnitude scaling is another approach that can be used for stimulus comparison data. This model is based upon Steven's (1957) law of ratio estimation; specific scaling techniques based upon this model are covered by Lodge (1981). Ideally, the data for a psychophysical magnitude scaling analysis would consist of *two separate* three-way, three-mode matrices; the first matrix would contain observations on a set of calibration stimuli, whereas the second matrix would contain the observations pertaining to the substantive stimulus objects. The interpretation of the two matrices is essentially identical (see Figure 5.13). The first way and mode of the matrix contains two levels, corresponding to the two separate response modalities. The second way and mode contains *K* levels, one for each stimulus. The third way and mode corresponds to replications of the magnitude judgments; there are *N* levels in this way, often corresponding to separate subjects. The cell entries in the data matrix contain judgments about the relative magnitudes of two stimuli: The magnitude of the column stimulus is expressed as a ratio of the magnitude of a reference stimulus. For each of the *K* stimuli, the magnitude judgment is repeated once for each response modality and for each replication.

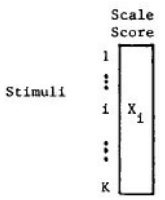
Figure 5.13. Psychophysical Magnitude Scale



C_{ijm} is the ratio of stimulus j's magnitude to the magnitude of the reference stimulus, measured according to response modality l, on replication m.

Scaling Model: Stimulus points are located along dimension by raising each C_{ijk} to power given by characteristic exponent for response modality i. Input matrix is collapsed by taking geometric means across modes for replications and response modalities.

Output Matrix: X



X_i gives location of point for stimulus i, based upon ratio comparison to a specified reference stimulus.

*There would be two separate input matrices-- one for calibration stimuli and one for the substantive stimuli. Both input matrices would be the same shape.

In data theory terms, the magnitude judgments express the degree to which one stimulus dominates another (i.e., the reference stimulus). Geometrically, the stimuli are modeled as points along a latent dimension corresponding to the specified attribute (i.e., the basis for the magnitude judgments). The stimuli are located by arbitrarily fixing the point for the reference

stimulus; the values for the remaining stimulus points are determined by their ratios to the reference stimulus. The scale values are related to the magnitude judgments by a specific power function (which is, itself, determined by the characteristic exponents for the response modalities). Furthermore, the use of ratio judgments implies that the origin of the dimension is fixed. Therefore, psychophysical magnitude scales are believed to provide ratio level measurement of underlying stimulus characteristics.

Assessment of Scaling Methods for Stimulus Comparisons. In substantive and practical terms, the empirical observations required to perform a psychophysical magnitude scaling analysis are very different from those required for the method of paired comparisons. However, from a data theory perspective, the differences between them are quite simple: With paired comparisons, we need dominance relations between every pair of stimuli. In magnitude scaling, each stimulus is compared only against a single reference stimulus. Apart from this, the two scaling models proceed in a similar manner, using dominance relations between pairs of stimuli to provide information about the relative locations of points along a dimension.

These two stimulus comparison methods are very powerful scaling techniques. However, they do have at least two problems that may limit their usefulness in certain situations. The first concerns the dimensionality of the attribute being measured. Both methods require the analyst to specify the *nature* of the dimension in advance; in other words, one stimulus is compared to another on the basis of a clearly defined characteristic. This can be a weakness if the researcher is uncertain about the existence or nature of the substantive property itself. In the most extreme case, the researcher might specify a criterion for the comparisons with which the subjects are completely unfamiliar. Any scaling results from such an analysis would be meaningless. This problem would probably be manifested as a poor fit between the scale and the data, but the source of the problem would not be clear: It could be the wrong specification, a lack of structure, or a multidimensional structure within the data. Thus these methods are problematic when applied as scaling criteria.

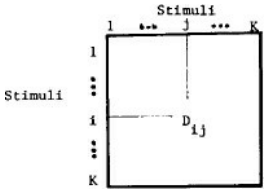
The second problem is more practical in nature. Stimulus comparison methods tend to be heavily data intensive. In other words, they require a great deal of empirical information before it is possible to estimate the stimulus locations along the latent dimension. Data collection is always costly, and researchers must consider whether these costs can justify the benefits that the empirical scales may provide. In many situations, it may simply be more efficient to obtain different data and use other scaling methods. This seems to be the main reason for the infrequency of paired comparison analyses, and it surely inhibits the wider application of psychophysical magnitude scaling techniques as well.

Similarities Data

Similarities data are usually used for multidimensional scaling (MDS). In an MDS analysis, the entries in the data matrix are modeled as distances between points in a space (e.g., Arabie, Carroll, and DeSarbo, 1987; Kruskal and Wish, 1978). The greater the proximity (or "similarity") between two objects, the *smaller* the distance between their points. In order to avoid this inverse relation between the data values and the geometric model, it is common to reverse the entries in the data matrix, changing similarities into *dissimilarities*.[19] After such a transformation is carried out, the numerical magnitudes of the data values are directly related to interpoint distances in the scaled space (larger values correspond to larger distances). Reversing the data values causes no problems, because the transformation preserves all of the information contained in the original data matrix.

Different Variations of MDS. There are many different varieties of MDS (e.g., Young, 1987). From a data theory perspective, their differences involve the interpretation of the entries in the data matrix, as well as the shape of the data matrix itself. Perhaps the most fundamental difference involves the distinction between metric and nonmetric MDS analyses. Stated simply, a metric analysis assumes that the input data are measured at the interval or ratio level, whereas a nonmetric analysis permits ordinal and even nominal data. In a metric MDS, the data values are transformed into interpoint distances according to a specific function (e.g., distances may be linearly related to dissimilarities); in a nonmetric MDS, the transformation from data to model can follow any monotonic form, so long as increasing dissimilarity does not correspond to *decreasing* distances. Note that the metric-nonmetric distinction applies *only* to the input data. Both forms of MDS produce a metric model in their output:

Figure 5.14. Multidimensional Scaling of Two-Way Matrix (Input Matrix -2 Ways, 1 Mode)



The distances between the object points are ratio level, whereas the object coordinates in the space are interval level.

The simplest data matrix for an MDS analysis has two ways and one mode, as shown in Figure 5.14. The single mode has *K* levels, corresponding to the stimulus objects. Each cell entry measures the dissimilarity between the objects represented in that row and that column. The matrix may or may not be symmetric about its main diagonal. If the entries are symmetric (that is, the dissimilarity between objects A and B equals the dissimilarity between B and A), then the

data can be transformed directly into distances (which are, by definition, symmetric). Figure 5.15 shows the results from an MDS analysis of a two-way, one-mode symmetric matrix of dissimilarities data. The output matrix is *K* by *M*, given the coordinates of the *K* stimulus points along the *M* axes in the space. Of course, the coordinates can be used to recover the interpoint distances, using the familiar Euclidean distance formula. This is the simplest variation; it is sometimes called *classical multidimensional scaling* or CMDS.

What if the entries in the data matrix are asymmetric, so that the dissimilarity between A and B is *not* equal to that between B and A? The course of action depends entirely upon the researcher's interpretation of the asymmetry. In many cases, the asymmetry is *substantively* uninteresting. For example, it may simply represent measurement error, or fluctuations in judgments. In that case, the asymmetry would not be a part of the geometric model used to represent the data. Any empirical asymmetries would be eliminated *before* the scaling analysis actually began, perhaps by taking the mean of the entries above and below the main diagonal of the matrix. The result would be a symmetric dissimilarities matrix, and the rest of the analysis would proceed as a CMDS. On the other hand, the asymmetries may, in fact, be substantively interesting. For example, there may be some kind of order effects, in which the nature of the first stimulus affects the response to the second stimulus. In that case, the researcher probably would want to include them in the geometric model produced by the scaling analysis; this would be an *asymmetric multidimensional scaling analysis* (or AMDS). One way of doing so is to analyze the data using Young's ASYMSCAL model.[20] This produces results that are diagrammed in Figure 5.16. There is still a *K* by *M* matrix of stimulus point coordinates. But now, there is also a *K* by *M* matrix of stimulus weights. When calculating the distance from, say, A to B, A's weights are used to distort the differences between A and B, along each of the *M* dimensions. Because the weights for A can be different from the weights for B, the overall distances can be asymmetric. Although this example shows the flexibility of the MDS approach, it should be noted that there are very few substantive applications of the ASYMSCAL model.

Figure 5.15. Multidimensional Scaling of Two-Way Matrix (Classical MDS)

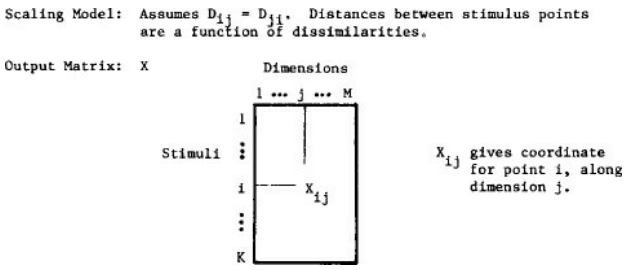


Figure 5.17 shows a three-way, two-mode matrix of dissimilarities. The first two ways and the first mode are all identical to the previous situation. The third way (which is also the second

mode) consists of repeated observations on the dissimilarities. Thus each entry in this matrix (say, d_{ijn}) represents the dissimilarity between stimulus *i* and stimulus *j*, as measured by data source *n.*

How are the objects in the third way (and second mode) incorporated into the scaling analysis? Once again, this depends entirely upon the researcher's interpretation of the information supplied by these objects. The differences across the data sources may not be substantively interesting; the third way would then be treated as simple replications of the original dissimilarities among the *K* stimuli. For example, there may be similarity judgments given by *N* judges, drawn from a homogeneous population. In that case, a *replicated MDS* or RMDS analysis could be used to collapse across the third way of the matrix, producing the standard *K* by *M* matrix of stimulus point coordinates shown in Figure 5.18. The practical advantage of RMDS is that we presumably have more reliable estimates of the dissimilarities between the stimuli.

Figure 5.16. Multidimensional Scaling of Two-Way Matrix (Assymetric MDS)

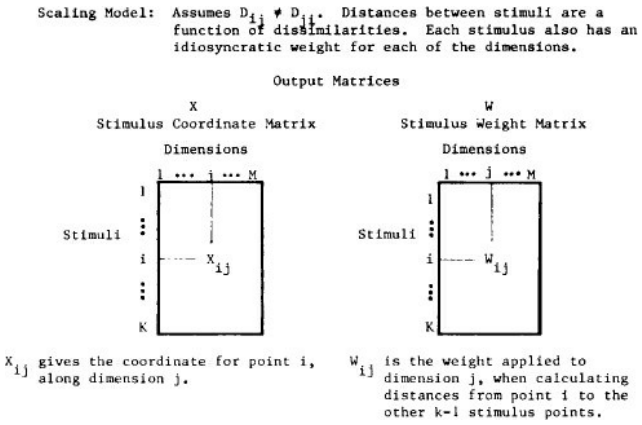
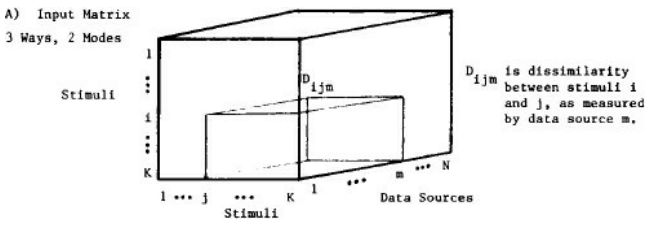


Figure 5.17. Multidimensional Scaling of Two-Way Matrix (Input Matrix—3 Ways, 2 Modes)



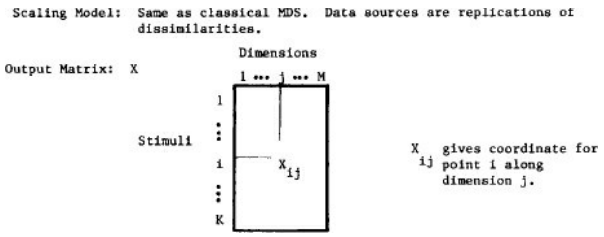
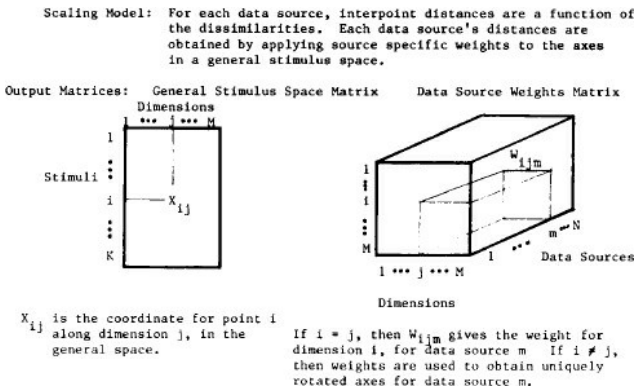


Figure 5.18. Multidimensional Scaling of Two-Way Matrix (Replicated MDS)

If the *N* data sources do vary in substantively interesting ways, then the researcher would want to incorporate these differences into the geometric model. The most common strategy for doing so is *weighted multidimensional scaling* or WMDS (Arabie, Carroll, and DeSarbo, 1987). There are several different forms of WMDS, but they all produce the kind of information shown in Figure 5.19. As usual, there is a *K* by *M* matrix of stimulus coordinates. But now, there is also a separate *M* by *M* weight matrix for each of the *N* data sources; this is shown as a single *M* by *M* by *N* matrix in the figure. The nature of the geometric model is determined by the contents of the weight matrices. If they are all identity matrices, then the results reduce to the RMDS model. If all of the weight matrices only have nonzero entries in their main diagonals, then the result is the familiar INDSCAL model (Carroll and Chang, 1970); for each data source, the interpoint differences along each axis are stretched or shrunk according to the values of the weights. If the weight matrices are symmetric, then the result is the IDIOSCAL (Carroll and Chang, 1970) or GEMSCAL model (Young, 1987); each data source corresponds to a different rotation of the stimulus space, along with the differential stretching and shrinking of each dimension.[21] These variations do not exhaust the possibilities, by any means. However, the INDSCAL model is really the only version of WMDS that has received widespread use in substantive analyses.

Figure 5.19. Multidimensional Scaling of Two-Way Matrix (Weighted MDS)



Multidimensional Scaling and Factor Analysis. There are some important parallels between MDS and factor analysis. Both scaling strategies use a two-way, one-mode data matrix (dissimilarities for MDS and correlations for factor analysis) in order to generate a

multidimensional configuration of object points. Indeed, the correlation matrix is often interpreted as a matrix of proximities (even though it is constructed from single stimulus data). What, then, distinguishes between factor analysis and MDS? Why use one technique rather than the other? The fundamental difference between them lies in the ways they model the entries in their respective data matrices. Factor analysis represents correlations as *scalar products*; the latter are functions of vector lengths (i.e., the communalities) and the angles between the vectors. MDS models dissimilarities as *distances*. Thus the two scaling strategies differ in the kind of "geometric picture" they construct for the input data. All of the other differences between them are subordinate to this fundamental distinction.[22]

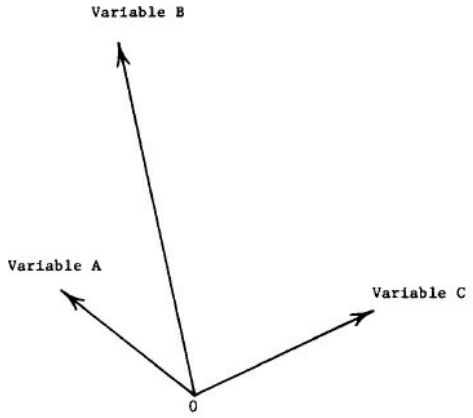
A researcher's decision to use factor analysis or MDS should be based entirely upon the most reasonable interpretation of the data. If the entries in the data matrix are best represented as scalar products, then factor analysis should be used; if the data can be accurately modeled as distances, then MDS is appropriate. Correlation coefficients (and directly related measures like covariances) are about the only kind of data that are routinely interpreted as scalar products. In contrast, there are many kinds of empirical observations that could be interpreted as similarities data. These include: direct similarity judgments; physical distances; joint probabilities (i.e., how frequently two events occur together); and conditional probabilities (i.e., given that one event has occurred, how frequently does the other occur?). Useful and detailed discussion of such measures are included in a variety of sources, including Schiffman, Reynolds, and Young (1981), Coxon (1982), and Davison (1983), as well as the volume by Kruskal and Wish (1978) in this monograph series.

It is important to emphasize that correlation coefficients *do not* work very well when they are interpreted as similarities data. Figure 5.20 shows vectors representing three variables: *A*, *B*, and *C*. Assume that we have correlations between these variables. An MDS analysis would use the input data to estimate the distances between the three terminal points of the vectors. From the figure, it is easy to see that these distances are ordered as follows: $d_{AC} < d_{AB} < d_{BC}$. Of course, these distances are not known a priori; the only information the researcher possesses is the set of correlations. But these values correspond to the *cosines* of the angles between the vectors. Obviously, there is a problem: The highest correlation (which MDS would model as the smallest distance) is between A and B; the second highest is between B and C; the smallest correlation (modeled as the largest distance) is that between A and C. Thus the correlations would seriously distort the distances between the points. This simple example shows that correlations *cannot* be modeled accurately as distances; therefore, they should not generally be used as input data for an MDS analysis.[23] Instead, factor analysis is the preferred method for analyzing a matrix of correlation coefficients.

Profile Distances. As an alternative to correlations, profile distances are sometimes used to convert single stimulus data into similarities. Assume that we want to perform an MDS analysis on a set of objects, using data in a *K* by *Q*, objects by variables matrix. The Euclidean distance formula can be used to calculate dissimilarities between objects by taking the square root of the sum of square differences, across the variables, for every pair of observations in the data matrix.[24] The resultant values (one value for each pair of observations) are, in fact, distances between the objects in a *Q*-dimensional variable space. MDS can be used to recover these distances.

An MDS analysis based upon profile distances can be used to reduce the sheer amount of data facing the researcher. But in many cases, the main objective of the analysis is to discern the underlying structure among a set of objects, without imposing the researcher's own ideas about how the objects differ among themselves. Profile distance data are much *less* useful for this purpose, because the researcher has decided on the variables that will go into the profile. In doing so, it is easy to influence the nature of the dimensions that will be obtained from the analysis. The dimensions of the space in the scaling result will tend to correspond to the variables with the largest variances—that is, the variables that exhibit the largest differences in their values across the objects. And, of course, these variables are selected by the researcher in the first place. Thus profile distances can be used to *summarize* a great deal of information succinctly about the objects under investigation. However, they are much less likely to produce new insights about the *sources* of variability among the objects.

Figure 5.20. Variable Vectors, Correlations, and Interpoint Distances



True Distances:

Dist (A,C)* < Dist (A,B) < Dist (B,C)

$\angle AOB$ = 39 degrees, $\cos \angle AOB = r_{AB} = .78$

$\angle AOC$ = 116 degrees, $\cos \angle AOC = r_{AC} = -.44$

$\angle BOC$ = 77 degrees, $\cos \angle BOC = r_{BC} = .22$

Incorrect Distances, estimated from correlations:

Dist (A,B) < Dist (B,C) < Dist (A,C)

*Dist (A,C) means "the distance between the A point and the C point."

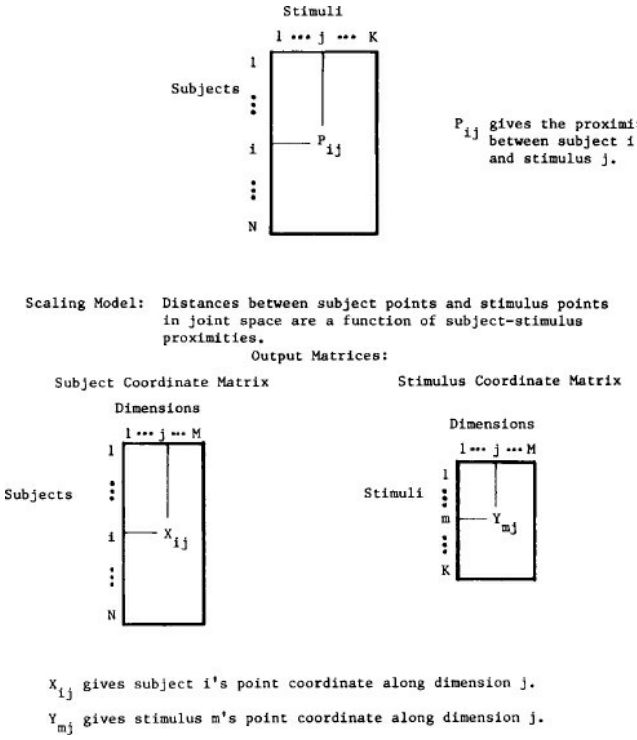
Preferential Choice Data

Preferential choice data are usually analyzed with the unfolding model (e.g., Coombs, 1964; McIver and Carmines, 1981). As shown in Figure 5.21, the simplest data matrix would contain two ways and two modes; the modes are often called “subjects” and “stimuli,” although they could easily correspond to other types of objects. The cell entries measure the proximities between the respective row and column objects. An unfolding analysis produces two separate matrices: The first matrix give the coordinates for a set of points representing the subjects. The second matrix gives the coordinates for points representing the stimuli. Both of these point sets are contained in the same “joint space.” The subject and stimulus points are located relative to each other, according to the following rule: The greater the preference of subject *i* for stimulus *j*, the smaller the distance between the point representing *i* and the point representing *j*.[25] There are many different procedures for fitting the unfolding model to empirical preferential choice data, and the variety among them is far too great to be covered in the monograph. Thompson (1988) provides a brief but comprehensive review of the different approaches to unfolding.

Problems in Multidimensional Unfolding. The unfolding model focuses entirely on representing proximity information in the data as distances within a space; this model transcends the

dimensionality of the space itself. If subjects use a single criterion to judge the stimuli, then a single dimension is appropriate to array the subject and stimulus points. If subjects use several criteria simultaneously to form their judgments, then a multidimensional space will be needed to locate the subject and stimulus points.

Figure 5.21. Unfolding Analysis



Even though the multidimensional unfolding *model* is well developed, virtually all of the algorithms for fitting this model to data are problematic. The difficulty arises because there is a relatively small amount of empirical data (the *N* by *K* data matrix) used to estimate a fairly large amount of information (the *N* by *M* matrix of subject point coordinates, along with the *K* by *M* matrix of stimulus point coordinates). As a result, there are often many different point configurations that can provide an equally acceptable fit to the data. Therefore, an estimated multidimensional joint space can only be regarded with caution. In extreme situations, degenerate scaling solutions can result, in which all of the objects in one of the data modes are located at a single point in the space. The only solution to this problem is to increase the amount of data relative to the number of points. One approach is to obtain replicated preferential choice data; algorithms based upon this approach seem to work quite well (e.g., DeSarbo and Carroll, 1985). But replicated data are often unavailable, especially in nonexperimental situations.

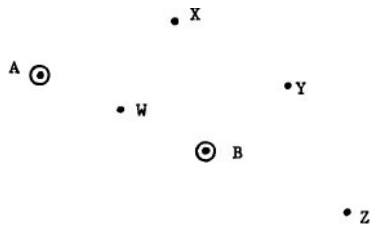
An alternative approach is to perform an *external* unfolding analysis. This assumes that one of the point configurations is already known (obtained from prior information of some kind), and

the other set of points is unfolded “into” the space that already contains the first set of points (Carroll, 1972). This is a much more tractable problem, because there are far fewer point coordinates that need to be estimated, relative to the amount of input data.[26]

Of course, the external unfolding approach still begs the question of the first point configuration. The obvious solution is to have coordinates for the stimuli obtained from a previous MDS analysis that used suitable similarities data for the stimuli. This could be obtained from information that is completely separate from the data used in the external unfolding analysis. Or, there are at least two methods of using the preference data themselves in order to obtain the similarities data.

First, Rabinowitz (1976) developed the “line-of-sight” measure (LOS) for exactly this purpose. The basic idea behind this method is shown in Figure 5.22. The figure depicts a simple space, containing two stimulus points (A and B) and four subjects' ideal points (W, X, Y, and Z). Assume that the values of the preference data are exactly equal to the distances between the ideal points and the stimulus points; this means that the data have been preprocessed so that *decreasing* distances correspond to *increasing* preferences. For each subject, we can sum the preference values across the two stimuli (that is, simply add Ws preference value for A to W's preference value for B, and do the same thing for the other three subjects). Similarly, we can take the difference between the two preference values given by each subject. Visual inspection of the figure will verify that the distance between A and B is measured by the *smallest sum* and also by the *largest difference*. Thus the summed preference values for subject W and the absolute difference of the preference values for subject Z both measure the distance between A and B. This result occurs because the points for W and Z lie precisely on the line connecting A and B. Of course, none of these point locations are known beforehand. Rabinowitz (1976) showed that, given certain assumptions about the distribution of the points, the *rank-order* of the distances between pairs of stimuli can be obtained by taking the smallest sums and the largest differences of the preference ratings for each pair of stimuli (even though none of the subject points may lie on the line connecting a pair of stimulus points). The actual method is a bit more complex, but this brief description should give the basic idea. The LOS measure is calculated for the stimuli, used as input to an MDS algorithm, and the results provide the stimulus configuration for the external unfolding analysis.

Figure 5.22. The Line-of-Sight Approach for Obtaining Similarities Data From Preferences



Dist (A,B) = 35

P:

		Stimuli A	Stimuli B
Subjects	W	16	19
	X	29	26
	Y	50	20
	Z	65	30

	Sums ($P_A + P_B$)	Differences ($\lvert P_A - P_B \rvert$)
W	35*	3
X	55	3
Y	70	30
Z	95	35**

*Smallest sum

**Largest difference

Rodgers and Young (1981) suggest another strategy, called “successive unfolding.” This approach proceeds in two steps. First, the *N* subjects' preference data values for each stimulus are interpreted as a profile for that stimulus; they are used to calculate profile distance measures for every pair of stimuli. This aggregates the preference data across the rows of the original data matrix, producing a matrix of similarities data, which is input to an MDS routine. The MDS routine provides a stimulus configuration, and the preference data are used a second time to locate the subject points relative to the stimulus points.

Dichotomous Preferential Choice Data. Recall from the discussion in Chapter 3 that it is possible to have dichotomous preferential choice data. Such data can be analyzed by a variant of the unfolding model, which is sometimes called “proximity scaling” (e.g., Coombs, 1964; Coombs and Smith, 1973; Weisberg, 1972, 1974). Figures 5.23, 5.24, and 5.25 show a simple example of a proximity scale. The top part of the figure depicts a two-way, two-mode matrix of dichotomous data. Assume that the rows are subjects, the columns are stimuli, and that the cell entries indicate whether the subject *chooses* the stimulus (1) or fails to choose the stimulus (0). Consistent with the unfolding model, we assume that subjects choose the most proximal

stimuli to themselves. If this is the case, and if choices are based upon a single, common criterion (i.e., they are unidimensional), then it should be possible to permute the columns of the data matrix to obtain a solid band of 1s in each row; the width of the band varies, because subjects can choose as many stimuli as they want. This is carried out in Figure 5.24. The order of the stimuli along the dimension is given by the order of the columns in the matrix. Geometrically, this corresponds to the ordering of the stimulus points along the latent dimension (Figure 5.25). It is as if each ideal point is surrounded by "boundaries of acceptance." If a stimulus point falls within these boundaries, then it is chosen by that subject; otherwise it is not chosen. Note that the ideal points are not located very precisely with this model, because we can only pin them down to their respective sets of chosen stimuli.[27]

Figure 5.23. Hypothetical Proximity Scale Analysis (Input Matrix)

		Stimuli			
		A	B	C	D
Subjects	1	1	1	0	1
	2	0	0	1	1
	3	1	1	1	1
	4	1	1	0	0
	5	1	0	1	1
	6	1	0	0	1

Figure 5.24. Hypothetical Proximity Scale Analysis (Scaling Model—Rearranged Matrix

		Stimuli			
		C	D	A	B
Subjects	1	0	1	1	1
	2	1	1	0	0
	3	1	1	1	1
	4	0	0	1	1
	5	1	1	1	0
	6	0	1	1	0

Figure 5.25. Hypothetical Proximity Scale Analysis (Output Matrix)

C	1*
D	2
A	3
B	4

*Entries give rank order of stimulus points along the latent dimension.

There are several possible ways to fit the proximity model to empirical data. The most obvious is simply to permute the rows and columns of the data matrix by hand until the best-fitting pattern is found. This is often not too difficult, if the number of stimuli and distinct response profiles are both relatively small. Alternatively, computer software for unfolding analysis (e.g., ALSCAL) usually permits the use of dichotomous data as input. Finally, the MUDFOLD (*m*ultiple *unid*imensional un*fold*ing) approach (van Schuur, 1984) was designed specifically for this purpose.[28] Dichotomous preferential choice data are fairly common. Therefore, these scaling approaches can be very useful. These give researchers a powerful model for analyzing the

structure contained within behavior that is qualitative in nature.

Interpretations of Data

There is never any single correct type of data that must be extracted from a given set of empirical observations. The interpretation of the data is always based on a combination of substantive considerations (which interpretation of the observations makes the most sense?) and analytic objectives (which scaling procedure will produce the kind of information desired?). To illustrate how the researcher takes an active part in interpreting empirical observations, consider a simple example: A set of *N* students, each of whom gives a response on *K* different test questions. The entries in the resultant data matrix are dichotomous (0 for incorrect, and 1 for correct responses).

How should these observations be interpreted? The answer depends on a number of factors. From one perspective, the responses could be single stimulus data (a correct answer means that the student's ability dominates the difficulty level of the test item, and vice versa). From another perspective, the responses could be dichotomous preferential choice data (a correct response means that the student's skills are proximal to those required to answer the question successfully).

For the moment, we will accept the single stimulus interpretation (as most teachers do when they test their students). This narrows down the range of possible scaling strategies, but we still need to decide how to interpret the two elements of each observation. The students are fairly straightforward—we want to represent them as points along a dimension, arrayed according to their test performance. Performance is measured by the number of correct answers; the more correct answers given by a student, the greater the extent to which that student's point dominates the item points, and hence it is located at a numerically higher position along the continuum. But what about the test items? Are these items equally good measures of the skills that are being tested, so that they only differ randomly among themselves? If so (or if we are simply uninterested in variability across items), we can use a scaling procedure that only locates the student points on the dimension (i.e., a summated rating scale); thus, students' test scores are simply based on the summed number of 1s in each row of the data matrix. On the other hand, the differences across the items may be substantively interesting (e.g., in the development of a battery of standard items that can be used for future testing purposes). In that case, we would want to use a method that simultaneously locates points for the items as well as the stimuli (e.g., a cumulative scale).

Further decisions must be made if the data do not meet the scaling criterion for the method that

is chosen—in other words, if the empirical responses are not consistent with a single dimension. One possibility would be to move to a less stringent scaling model; that is, one that includes a wider variety of response patterns in its definition of "structure within the data." Thus, if the data do not fit a cumulative scale, we could try a summated rating scale (thereby implicitly concluding that the test items do not, in fact, vary systematically among themselves, even though the students may still do so). Or, if the summated rating model does not fit (i.e., the reliability is too low), then we could try the unfolding model (which incorporates a wider variety of response patterns as perfectly scalable observations). Of course, in doing the latter, we are implicitly giving up on the objective of *ranking* the students according to their respective skill levels; the proximity model admits that students may possess different combinations of skills (gauged by which items they answer correctly), but it makes it more difficult to say that some students possess *more* of the underlying trait than other students.

Of course, we could continue on with this example. But the overall point of the discussion is probably already clear. The nature of data is never predetermined. Instead, the data depend upon the researcher's *interpretation* of the observations. Differing interpretations lead to the application of different scaling procedures which, in turn, directly affect the kind of information that is extracted from the analysis. The researcher must decide which interpretation is most appropriate, and work accordingly. This is an important *creative* component of empirical research in the social sciences.

http://dx.doi.org/10.4135/9781412983860.n5