3
MDS Models and Measures of Fit

MDS models are defined by specifying how given similarity or dissimilarity data, the proximities p_{ij}, are mapped into distances of an m-dimensional MDS configuration $\mathbf{X}$. The mapping is given by a *representation function* $f(p_{ij})$ that specifies how the proximities should be related to the distances $d_{ij}(\mathbf{X})$. In practice, one usually does not attempt to strictly satisfy f. Rather, what is sought is a configuration (in a given dimensionality) whose distances satisfy f as closely as possible. The condition "as closely as" is quantified by a badness-of-fit measure or *loss function*. The loss function is a mathematical expression that aggregates the representation errors, $e_{ij} = f(p_{ij}) - d_{ij}(\mathbf{X})$, over all pairs (i, j). A normed sum-of-squares of these errors defines *Stress*, the most common loss function in MDS. How Stress should be evaluated is a major issue in MDS. It is discussed at length in this chapter, and various criteria are presented.

3.1 Basics of MDS Models

In this section, MDS models are defined and discussed on a level sufficient for most practical applications. In later chapters, we revisit some of the relevant issues in greater detail.

Assume that measures of similarity or dissimilarity, for which we use the general term *proximity*, p_{ij}, are given for the pairs (i, j) of n objects. Some examples for such proximities were discussed in Chapter 1: similarities of crimes, assessed by the correlations of their frequencies over different U.S.

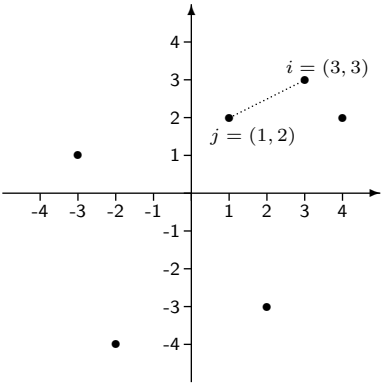

FIGURE 3.1. A Cartesian plane with some points; the length of the line segment connecting points i and j is the (Euclidean) distance of points i and j.

states; correlations among attitudes towards political protest behaviors; direct ratings of the overall similarity of pairs of different countries; and similarity judgments on one-spoked wheels. All of these cases are examples of measures of similarity, because the higher a correlation (or a rating of similarity), the more similar the objects i and j. However, instead of asking for judgments of similarity, it is just as easy—or even easier—to ask for judgments of dissimilarity, for example, by presenting a rating scale ranging from 0 = no difference to 10 = very dissimilar.

Coordinates in the MDS Space

MDS attempts to represent proximities by distances among the points of an m-dimensional configuration $\mathbf{X}$, the MDS space. The distances can be measured by a ruler, up to a certain level of precision, and if the MDS space is at most three-dimensional. But distances can also be *computed* with arbitrary precision, and this can be done in a space of arbitrarily high dimensionality. Computation is made possible by *coordinating* the MDS space. The most common such coordination is first to define a set of m directed axes that are perpendicular to each other and intersect in one point, the *origin* O. These axes—in the applied context often called *dimensions*—are then divided up into intervals of equal length so that they represent, in effect, a set of perpendicular "rulers".

Each point i, then, is uniquely described by an m-tuple $(x_{i1}, x_{i2}, \ldots, x_{im})$, where x_{ia} is i's projection onto dimension a. This m-tuple is point i's *coordinate vector*. The origin O is given the coordinates $(0, 0, \ldots, 0)$. Figure 3.1 shows some points and their coordinate vectors in a *Cartesian plane*, that is, in a plane coordinated by a set of perpendicular dimensions.

Computing Distances

Given a Cartesian space, one can compute the distance between any two of its points, i and j. The most frequently used and the most natural distance function is the Euclidean distance. It corresponds to the length of the straight line[1] segment that connects the points i and j. Figure 3.1 shows an example.

The Euclidean distance of points i and j in a two-dimensional configuration $\mathbf{X}$ is computed by the following formula:

$$d_{ij}(\mathbf{X}) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2}. \tag{3.1}$$

Thus, $d_{ij}(\mathbf{X})$ is equal to the square root of the sum of the intradimensional differences $x_{ia} - x_{ja}$, which is simply the Pythagorean theorem for the length of the hypotenuse of a right triangle. For Figure 3.1, thus, formula (3.1) yields $d_{ij} = \sqrt{(3-1)^2 + (3-2)^2} = \sqrt{5}$. Formula (3.1) can also be written as

$$d_{ij}(\mathbf{X}) = \left[\sum_{a=1}^{2}(x_{ia} - x_{ja})^2\right]^{1/2}, \tag{3.2}$$

which can easily be generalized to the m-dimensional case as

$$d_{ij}(\mathbf{X}) = \left[\sum_{a=1}^{m}(x_{ia} - x_{ja})^2\right]^{1/2}. \tag{3.3}$$

MDS Models and Their Representation Functions

MDS maps proximities p_{ij} into corresponding distances $d_{ij}(\mathbf{X})$ of an MDS space $\mathbf{X}$. That is, we have a *representation function*

$$f : p_{ij} \to d_{ij}(\mathbf{X}), \tag{3.4}$$

where the particular choice of f specifies the *MDS model*. Thus, an MDS model is a proposition that given proximities, after some transformation f, are equal to distances among points of a configuration $\mathbf{X}$:

$$f(p_{ij}) = d_{ij}(\mathbf{X}). \tag{3.5}$$

[1]The term "straight" corresponds to what we mean by straight in everyday language. In Euclidean geometry, a straight line can be drawn by tracing with a pen along a ruler. More generally, a straight line is the *shortest* path (geodesic) between two points. The notion of straightness, therefore, presupposes a distance measure. With different distance measures, straight lines often do not look straight at all. An example is the straight line between points 4 and 6 in Figure 1.7, which consists of the two dashed line segments that *look* like a "corner" line.

The distances $d_{ij}(\mathbf{X})$ in (3.4) and (3.5) are always unknowns. That is, MDS must find a configuration $\mathbf{X}$ of predetermined dimensionality m on which the distances are computed. The function f, on the other hand, can either be completely specified or it can be restricted to come from a particular class of functions. Shepard (1957), for example, collected similarities p_{ij} for which he predicted, on theoretical grounds, that they should be related to distances in an unknown two-dimensional space $\mathbf{X}$ by the exponential function. That is, it was hypothesized that $p_{ij} = \exp[-d_{ij}(\mathbf{X})]$. Similarly, Thurstone (1927) predicted that choice probabilities p_{ij} should be equal to unknown distances between points i and j on a line ("scale values") after transforming the p_{ij}s by the inverse normal distribution function. This choice of f, again, was theoretically justified.

In most applications of MDS, there is some looseness in specifying f. That is, for example, f is only restricted to be "some" exponential function or "some" linear function. The exact parameters of these functions are not specified. An important case is *interval MDS*. It is defined by

$$p_{ij} \to a + b \cdot p_{ij} = d_{ij}(\mathbf{X}), \tag{3.6}$$

for all pairs (i, j). The parameters a and b are free and can be chosen such that the equation holds. Another case is *ordinal MDS*, where f is restricted to be a monotone function that preserves the order of the proximities. That means—assuming, for simplicity, that the proximities are dissimilarity scores—that

$$\text{if } p_{ij} < p_{kl}, \text{ then } d_{ij}(\mathbf{X}) \leq d_{kl}(\mathbf{X}). \tag{3.7}$$

If $p_{ij} = p_{kl}$, (3.7) requires no particular relation of the corresponding distances. This is known as the *primary approach* to tied proximities, where ties can be "broken" in the corresponding distances. The *secondary approach* to ties requires that if $p_{ij} = p_{kl}$, then also $d_{ij} = d_{kl}$. The primary approach is the default in most ordinal MDS programs. A slight modification of (3.7) is to replace the relation $\leq$ by $<$. The first relation specifies a *weak* monotone function f, the second one a *strong* monotone function. Most often, ordinal MDS is used with a weak monotone function.

How should one choose a particular representation function? If no particular f can be derived by theoretical reasoning, one often restricts f to a particular class of functions on the basis of the scale level of the proximities. For example, if the proximities are direct similarity ratings on, say, pairs of nations, one might feel that only their rank-order yields reliable information about the respondent's true cognitions. Differences ("intervals") between any two ratings, in contrast, would not represent any corresponding psychological quantities. Under these assumptions, there is no reason to insist that these intervals be faithfully represented by distances in the MDS space. Moreover, a weak scale level makes it easier to approximately represent the essential information in an MDS space of low dimensionality.

Conversely, starting from an MDS model, one can choose a representation function g in the regression hypothesis $g : d_{ij}(\mathbf{X}) \rightarrow p_{ij}$. This hypothesis needs to be tested against the data. One can pick any g: if it leads to a model that is empirically satisfied—and provided that the model does not hold for formal reasons only—one has shown a nontrivial empirical regularity. No further justification is needed for picking a particular function g.

3.2 Errors, Loss Functions, and Stress

MDS models require that each proximity value be mapped *exactly* into its corresponding distance. This leaves out any notion of error. But empirical proximities always contain noise due to measurement imprecision, unreliability, sampling effects, and so on. Even the distances used in Table 2.1 are not completely error-free, because reading off values from a ruler only yields measures of limited precision. Hence, one should not insist, in practice, that $f(p_{ij}) = d_{ij}(\mathbf{X})$, but rather that $f(p_{ij}) \approx d_{ij}(\mathbf{X})$, where $\approx$ can be read as "as equal as possible". Given that the proximities contain some error, such approximate representations make even *better* representations—more robust, reliable, replicable, and substantively meaningful ones—than those that are formally perfect, because they may smooth out noise.

If one has a theory about the proximities, one would be interested to see how well this theory is able to explain the data, and so a best-possible MDS representation (of some sort) is sought. If the error of representation is "too large," one may reject or modify the theory, but obviously one first needs to know how well the theory accounts for the data. Any representation that is precise enough to check the validity of this theory is sufficiently exact. A perfect representation is not required.

Further arguments can be made for abandoning the equality requirement in $f(p_{ij}) = d_{ij}(\mathbf{X})$. Computerized procedures for finding an MDS representation usually start with some initial configuration and improve this configuration by moving around its points in small steps ("iteratively") to approximate the ideal model relation $f(p_{ij}) = d_{ij}(\mathbf{X})$ more and more closely. As long as the representation is not perfect, one only has $f(p_{ij}) \approx d_{ij}(\mathbf{X})$, where $\approx$ means "equal except for some small discrepancy".

The Stress Function

To make such notions as "almost", "nearly", and so on, more precise, we employ the often used statistical concept of error. A (squared) *error of representation* is defined by

$$e_{ij}^2 = [f(p_{ij}) - d_{ij}(\mathbf{X})]^2. \tag{3.8}$$

Summing e_{ij}^2 over all pairs (i, j) yields a badness-of-fit measure for the entire MDS representation, *raw Stress*,

$$\sigma_r = \sigma_r(\mathbf{X}) = \sum_{(i,j)} [f(p_{ij}) - d_{ij}(\mathbf{X})]^2. \tag{3.9}$$

The raw Stress value itself is not very informative. A large value does not necessarily indicate bad fit. For example, suppose that the dissimilarities are road distances between cities in kilometers. Suppose that an MDS analysis on these data yields $\sigma_r(\mathbf{X}_1) = .043$. Redoing the analysis with dissimilarities expressed in meters yields the same solution, but on a scale that is 1000 times as large, and so one gets $\sigma_r(\mathbf{X}_2) = 43,000$. This does not mean that $\mathbf{X}_2$ fits the data worse than $\mathbf{X}_1$; it merely reflects the different calibration of the dissimilarities. To avoid this scale dependency, σ_r can, for example, be normed as follows,

$$\sigma_1^2 = \sigma_1^2(\mathbf{X}) = \frac{\sigma_r(\mathbf{X})}{\sum d_{ij}^2(\mathbf{X})} = \frac{\sum [f(p_{ij}) - d_{ij}(\mathbf{X})]^2}{\sum d_{ij}^2(\mathbf{X})}. \tag{3.10}$$

Taking the square root of σ_1^2 yields a value known as *Stress-1* (Kruskal, 1964a). The reason for using σ_1 rather than σ_1^2 is that σ_1^2 is almost always very small in practice, so σ_1 values are easier to discriminate. Thus, more explicitly,

$$\text{Stress-1} \ = \sigma_1 = \sqrt{\frac{\sum [f(p_{ij}) - d_{ij}(\mathbf{X})]^2}{\sum d_{ij}^2(\mathbf{X})}}. \tag{3.11}$$

The summations extend over all p_{ij} for which there are observations. Missing data are skipped. In the typical case of symmetric proximities, where $p_{ij} = p_{ji}$ (for all i, j), it suffices to sum over one half of the data-distance pairs only. Obviously, $\sigma_1 = 0$ only if $d_{ij}(\mathbf{X}) = f(p_{ij})$.

Minimizing Stress-1 always requires finding an optimal $\mathbf{X}$ in a given dimensionality m. Moreover, if f is only specified up to certain free parameters, then optimal values for these parameters must also be found. This problem typically is solved by regressing the proximities onto the distances computed on $\mathbf{X}$. In interval MDS, one uses linear regression, in ordinal MDS monotone regression (see Section 9.2). The regression yields transformed proximities, $f(p_{ij})$s, that are "approximated distances" or "d-hats" ($\widehat{d}_{ij}$s) also referred to as *disparities* in the MDS-literature.

3.3 Stress Diagrams

Loss functions such as Stress are indices that assess the mismatch of (admissibly transformed) proximities and corresponding distances. Stress is, in a way, similar to a correlation coefficient, except that it measures the badness-of-fit rather than the goodness-of-fit. Experienced researchers know that

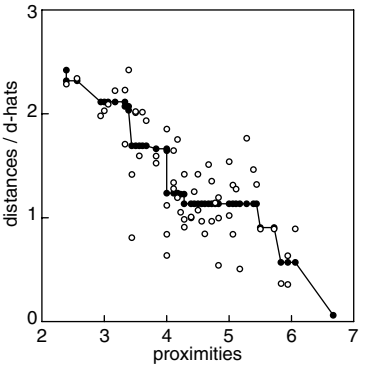

FIGURE 3.2. Shepard diagram for MDS solution shown in Fig. 1.5.

correlations can be high or low for various reasons. For example, a correlation can be artificially high because of outliers. It can also be misleadingly low because the regression trend is not linear. What one usually does to study such questions is to take a look at the scatter diagram.

Exactly the same approach is also customary in MDS. The most informative scatter diagram plots proximities on the X-axis against the corresponding MDS distances on the Y-axis. Typically, a regression line that shows how proximities and approximated distances ($\widehat{d}_{ij}$s) are related is also shown. This plot is known as a *Shepard diagram*.

Figure 3.2 gives an example. The Shepard diagram exhibits, as open circles, the similarities of Table 1.3 plotted against the corresponding distances of Figure 1.5. The filled circles represent the $(p_{ij}, \widehat{d}_{ij})$ pairs. They all lie on a monotonically descending line, as requested by the ordinal MDS model used to scale these data. The vertical distance of each (p_{ij}, d_{ij}) point (open circle) from the $(p_{ij}, \widehat{d}_{ij})$ point (filled circle) represents the error of representation for this particular proximity, e_{ij}. The Y-axis of the Shepard diagram has two labels: distances (d_{ij}s) and approximated distances ($\widehat{d}_{ij}$s).

What can be learned from this Shepard diagram? First, it gives an overall impression of the scatter around the representation function. In Figure 3.2, one notes that there is quite a bit of scatter around the monotone regression curve. The vertical distances of the points from the step function (e_{ij}s) are generally quite large, and thus $\sigma_1 = .186$. Then, one notes that there are no real outliers, although some points contribute relatively much to Stress. The most prominent case is the point with coordinates $(3.44, 0.82)$. Its error or "residual," which enters the Stress function quadratically, is -0.877, and the second greatest residual is only 0.636. One finds in Table 1.3 that there are two dissimilarity estimates of 3.44, one for India vs. France and one for Brazil vs. Egypt. The MDS program keeps track of each and every proximity and informs us that the large residual is related to the pair India–France. Hence, this observation is explained worst by

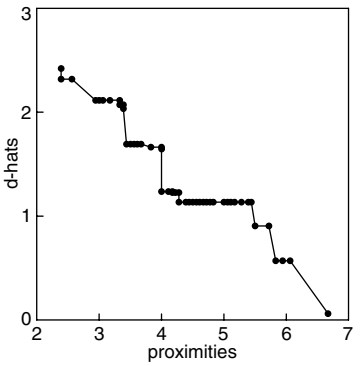

FIGURE 3.3. A transformation plot (scatter diagram of proximities vs. d-hats) for the MDS solution shown in Fig. 1.5.

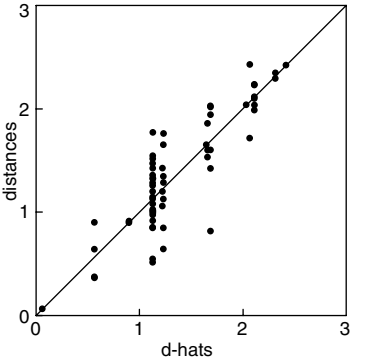

FIGURE 3.4. A residual plot (scatter diagram of d-hats vs. distances) for the MDS solution shown in Fig. 1.5.

the MDS space in Figure 1.5, possibly because it brings in an additional dimension.

A Shepard diagram is particularly informative in the case of ordinal MDS. This model requires a monotone representation function f, but its particular shape is left open. It is often interesting to see which shape it acquires in scaling real data. [Indeed, this question motivated the invention of ordinal MDS (see Chapter 17).] In Figures 3.2 and 3.3, we note that the regression curve is roughly linear, although it shows a number of marked steps.

Some MDS programs also provide scatter plots of the $\widehat{d}_{ij}$s vs. the corresponding d_{ij}s. Figure 3.4 gives an example for the data in Table 1.3. The points in such a plot scatter around the bisector from the lower left-hand corner to the upper right-hand corner. If Stress is zero, they all lie on this bisector; otherwise, they do not. The vertical distance of the points from the bisector corresponds to the error of approximation, but the horizontal distances have the same magnitude, $|e_{ij}|$. The outlier discussed above, the proximity for France vs. India, has coordinates 0.815 on the vertical axis and 1.69 on the horizontal axis. It lies farthest from the bisector. Generally, what one studies in such plots is the distribution of the points around this bisector for possible outliers, anomalies, gaps, and so on.

3.4 Stress per Point

In the previous section, we have looked at how well each proximity p_{ij} or its transformation $\widehat{d}_{ij}$ is fitted by the corresponding distance d_{ij}. The error for one particular proximity is the vertical distance between $\widehat{d}_{ij}$ and the d_{ij}

TABLE 3.1. Squared error for the solution in Figure 1.5 of the similarity ratings for 12 nations (Wish, 1971). The last row (and column) contains the average per row (or column) and is called Stress per point.

Nation		1	2	3	4	5	6	7	8	9	10	11	12	SPP
Brazil	1	–	.02	.24	.09	.00	.08	.08	.02	.00	.00	.07	.00	.05
Congo	2	.02	–	.01	.07	.00	.03	.00	.04	.01	.00	.00	.05	.02
Cuba	3	.24	.01	–	.09	.01	.05	.05	.02	.01	.00	.00	.00	.04
Egypt	4	.09	.07	.09	–	.01	.02	.07	.01	.01	.00	.08	.00	.04
France	5	.00	.00	.01	.01	–	.23	.21	.17	.01	.02	.01	.01	.06
India	6	.08	.03	.05	.02	.23	–	.00	.04	.03	.01	.01	.00	.04
Israel	7	.08	.00	.05	.07	.21	.00	–	.04	.00	.00	.00	.02	.04
Japan	8	.02	.04	.02	.01	.17	.04	.04	–	.10	.01	.00	.02	.04
China	9	.00	.01	.01	.01	.01	.03	.00	.10	–	.00	.00	.06	.02
USSR	10	.00	.00	.00	.00	.02	.01	.00	.01	.00	–	.04	.00	.01
U.S.A	11	.07	.00	.00	.08	.01	.01	.00	.00	.00	.04	–	.00	.02
Yugoslavia	12	.00	.05	.00	.00	.01	.00	.02	.02	.06	.00	.00	–	.01
Stress per point		.05	.02	.04	.04	.06	.04	.04	.04	.02	.01	.02	.01	.03

in the Shepard diagram. Instead of looking at a single error only, it may be more interesting to consider all errors of one object to all others. We examine the definition of raw Stress in (3.9) more closely. Clearly, raw Stress is a sum of the squared errors over all pairs of objects. Table 3.1 contains the squared error for the solution in Figure 1.5 of the similarity ratings for twelve nations (Wish, 1971). Note that for convenience, this table shows the squared errors below and above the diagonal, although because of the symmetry the errors below (or above) the diagonal would suffice. Now, a simple measure to indicate how badly each individual point is fitted can be obtained by averaging the squared errors between the current object and all other objects. We call this measure *Stress per point* and it is shown in the last column (and the last row) of Table 3.1. For example, the Stress per point for France can be obtained by averaging all the squared errors in the row of France in Table 3.1. Equivalently, the same value is obtained by averaging column 5 for France in this table. An additional feature of Stress per point is that their average equals the total Stress. Because Stress per point is defined on the squared errors, we must square σ_1 to compare it with the average Stress per point. In the previous section, we found that $\sigma_1 = .186$, so that $\sigma_1^2 = .0346 \approx .03$ is the same value indeed as the element in the lower right-hand corner of Table 3.1.

Several conclusions can be drawn from this table. First, most points are fitted rather well by this solution, because their Stress per point is reasonably low. Second, the best fitting points are Yugoslavia and the USSR, followed by U.S.A., China, and Congo. Third, the worst fitting points are France and Brazil. When interpreting the solution, this information should be kept in mind. Apparently, the MDS solution in Figure 1.5 is not very well able to represent the points for France and Brazil. Their Stress per

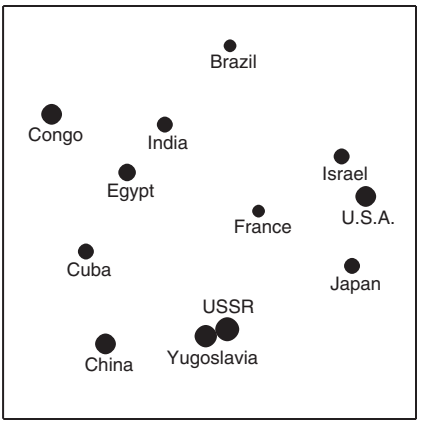

FIGURE 3.5. Bubble plot of fit per point derived from Stress per point of the similarity nations data of Wish (1971). Big bubbles indicate points with good fit, small bubbles indicate points with poor fit.

point is relatively high because there is quite some difference between the distances and the transformed data with all other countries. It can be verified in Table 3.1 that the high Stress per point for France is caused, in particular, by high errors of France with India, Israel, and Japan. A high Stress per point indicates that we cannot be certain about the exact location of this point. It may be an indication that an additional dimension is needed for these points to reduce the error.

To inspect the Stress per point graphically, it is simpler to switch to the *fit per point* that is defined as one minus the Stress per point. Generally, the fit per point is a value between zero and one. Usually, the fit per point is close to one. In our example, the fit per point varies between .99 for Yugoslavia and USSR and .94 for France. In Figure 3.5, the fit per point is expressed by the radius of the bubble representing the point. The centers of the bubbles are the locations of the points, just as in Figure 1.5. To avoid too little discrimination in the size of the bubbles, we linearly transformed the radii such that the worst fitting point (France) has a radius twice as small as the best fitting point (Yugoslavia). It can be seen in Figure 3.5 that the best fitting points (with the largest bubble) are mostly located around the edges (with the exception of Brazil) and that the worst fitting points are located towards the center (such as, for example, France). To interpret the solution, Figure 3.5 shows immediately which points should be emphasized in the interpretation because of their good fit per point, and which points should not be emphasized because of their bad fit.

3.5 Evaluating Stress

How should one evaluate the Stress of a given MDS solution? One approach is to study the Shepard diagram. It shows the number of points that have to be fitted, the optimal regression line, the size of the deviations, possible outliers, and systematic deviations from the requested regression line. Thus, Shepard diagrams are highly informative. Nevertheless, it is customary to condense all of this information into a single number, Stress.

In ordinal MDS, *any* matrix of proximities p_{ij} $(i < j)$ can be represented, with zero Stress, in $m = n - 2$ dimensions (see Chapter 19). However, such perfect solutions are not desired, as we saw above. Therefore, one seeks an MDS representation with considerably fewer dimensions. The problem is how to choose the "proper" dimensionality. Scaling with too few dimensions may distort the true (reliable) MDS structure due to *over-compression* or may lead to technical problems (see Chapter 13). Being too generous on dimensions may, on the other hand, blur the MDS structure due to *overfitting* noise components. If information is available about the reliability of the data, one should choose a dimensionality whose Stress corresponds to the random component of the data. Inasmuch as this information is rarely given, one has to resort to other criteria.

Simple Norms for Stress

Beginners in multivariate data analysis typically ask for simple (often overly so) norms. In MDS, a number is requested so that whenever Stress is less than that benchmark value, the MDS solution should be considered acceptable. Guttman (in Porrat, 1974) proposes such a norm for a coefficient closely related to Stress: he required that the coefficient of alienation K should be less than 0.15 for an acceptably precise MDS solution. He later added that what he had in mind when he made this proposal were "the usual circumstances" (Guttman, personal communication). [Note that here and in the following, we are considering ordinal MDS only.]

It is easy to see that such circumstances are important. Any global fit measure will be low, for example, when the number of points n is small relative to the dimensionality of the space, m. Guttman thus assumed for the $K < 0.15$ rule that n "clearly" exceeds m (as another rule of thumb, at least fourfold: Rabinowitz, 1975; Kruskal & Wish, 1978). Conversely, if n is much larger than m (more than 10 times as large, say), higher badness-of-fit values might also be acceptable.

Another rough criterion is to pick that solution "for which further increase in $[m]$ does not significantly reduce Stress" (Kruskal, 1964a, p. 16). To find that m, one should first compute MDS solutions for different dimensionalities (e.g., for $m = 1, 2, \ldots, 5$) and then plot the resulting Stress values (on the Y-axis) against the m-values (on the X-axis). If the points in this diagram are connected by a line, starting at $m = 1$ and ending at

$m = \max$, one obtains a *scree plot*. (An example of a scree plot is given in Figure 4.5.)

The curve in a scree plot is generally monotonically decreasing, but at an increasingly slower rate with more and more dimensions (convex curve).[2] What one looks for is an *elbow* in this curve, a point where the decrements in Stress begin to be less pronounced. That point corresponds to the dimensionality that should be chosen. The rationale of this choice is that the elbow marks the point where MDS uses additional dimensions to essentially only scale the noise in the data, after having succeeded in representing the systematic structure in the given dimensionality m.

For the Stress-1 coefficient σ_1 using ordinal MDS, Kruskal (1964a), on the basis of his "experience with experimental and synthetic data" (p. 16), suggests the following benchmarks: .20 = poor, .10 = fair, .05 = good, .025 = excellent, and .00 = perfect.[3] Unfortunately, such criteria almost inevitably lead to misuse by suggesting that only solutions whose Stress is less than .20 are acceptable, or that all solutions with a Stress of less than .05 are good in more than just a formal sense. Neither conclusion is correct. An MDS solution may have high Stress simply as a consequence of high error in the data, and finding a precise representation for the data does not imply anything about its scientific value.

Obviously, one needs more systematic insights into how Stress depends on the number of points, the dimensionality of the MDS solution, the kind and amount of error in the proximities, the type of the underlying true configuration, and so on. Computer simulation studies can help to answer such questions. In the following, we consider some such studies.

Stress for Random Data

The most extreme case that can be studied is concerned with the "nullest of all null hypotheses" (Cliff, 1973), that is, with the question of whether the Stress for some given data is significantly lower than for random data. Stenson and Knoll (1969) and Klahr (1969) compute the distribution of Stress values for ordinal MDS under H_0 as follows: (a) pick some values for n, the number of the points, and m, the dimensionality of the MDS space; (b) randomly insert the numbers $1, 2, 3, \ldots, \binom{n}{2}$ into the cells of a lower-half proximity matrix; (c) use ordinal MDS on these proximities and

[2] An exception to that rule can result, for example, when the MDS computer program does not succeed in finding the optimal solution for some dimensionality. The scree test can, therefore, occasionally be useful to identify such suboptimal solutions.

[3] A Stress value of .20, say, is often written as 20%. Why this language became popular is not entirely clear. However, if one replaces Stress by squared Stress, then one can show that, for example, 20% (squared) Stress means that 80% of the variance of the d-hats is explained by the distances (see Section 11.1).

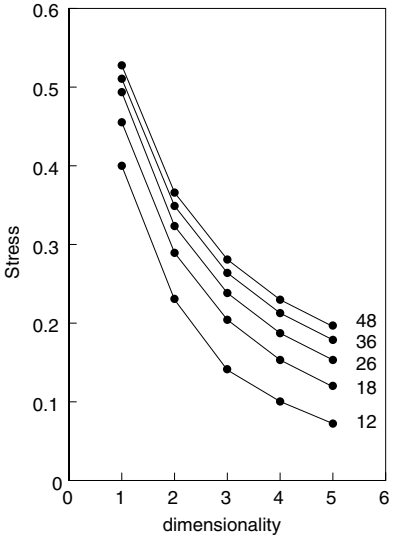

FIGURE 3.6. Average Stress for random proximities among n objects, represented via ordinal MDS in different dimensionalities (Spence & Ogilvie, 1973).

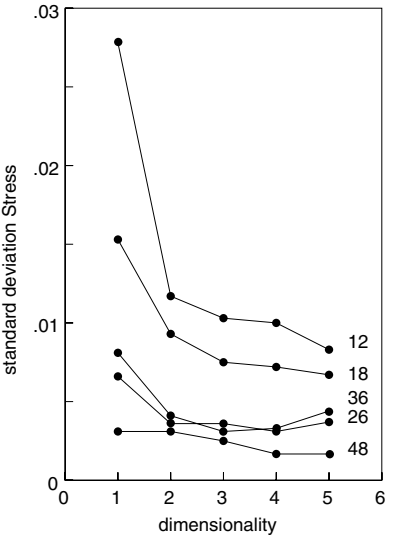

FIGURE 3.7. Standard deviations for curves in Fig. 3.6.

compute Stress; and (d) repeat the above for many permutations of the data, so that a distribution of Stress values results.

These simulations show that if n grows, then expected Stress also grows and its variance becomes smaller; if m grows, then expected Stress becomes smaller. If the data contain ties, the primary approach leads to lower Stress (because ties are optimally broken) than the secondary (where ties in the data must be preserved in the distances); the more ties there are, the larger the difference.

Spence and Ogilvie (1973) conduct a similar investigation for $n = 12, 13, \ldots, 48$ points and $m = 1, 2, \ldots, 5$ dimensions, a useful range for many practical purposes. Figure 3.6 shows the average Stress curves for various n values, using ordinal MDS. The curves indicate again that Stress depends on n and m. One also notes that each additional dimension reduces Stress increasingly less. The confidence intervals of the expected Stress values are quite narrow, as the standard deviations of the Stress distributions in Figure 3.7 show. The standard deviations are so small that lowering the curves by about 0.03 should result in reliable cutoff values for testing this H_0.

Spence (1979) has shown that one can closely approximate the curves in Figure 3.6 and curves interpolated therein for $n = 12, 13, \ldots, 48$ by the

formula

$$\sigma_1 = .001(a_0 + a_1 m + a_2 n + a_3 \ln(m) + a_4 \sqrt{\ln(n)}), \tag{3.12}$$

where $a_0 = -524.25, a_1 = 33.8, a_2 = -2.54, a_3 = -307.26$, and $a_4 = 588.35$. A comparison of the results with those from Stenson and Knoll (1969) shows very good agreement, so that formula (3.12) can be used to estimate expected "random" Stress for the range $n = 10, \ldots, 60$ and $m = 1, \ldots, 5$.

We show that the Stress values in all real-data MDS applications discussed in this book lie definitely under the values expected for H_0. This also shows that this kind of null hypothesis represents a very small hurdle indeed. On the other hand, if one does not even succeed in rejecting this H_0, then it seems unreasonable to study the MDS representation further.

The Hefner Model

Simulations that study the distribution of Stress for random data (of some sort) are useful from a data-analytic point of view. They do not attempt to simulate an MDS model in the sense of a psychological theory about similarity judgments. If MDS is used in this way, then one also needs a more explicit model for what is meant by the "random" component of the data.

Consider the similarity-of-nations example in Section 1.3. We may want to assume that a respondent arrives at his or her overall similarity judgment by first computing the distance of two nations in his or her system of dimensions or *perceptual space*, and then mapping this distance into the response format provided by the researcher. Moreover, we could postulate that the perceptual space is not static, but that its points "oscillate" about their characteristic position over time. The oscillations could be due to unsystematic variations in attention, fluctuating discrimination thresholds, activation and decay processes on the memory traces, and so on. Under these conditions, the respondent would compute a distance at each point in time, but these distances would not fit together in a plane, because each distance depends on the particular positions of the points at time t, and these positions are not constant over time. An observed proximity, after transformation by f, is thus conceived as $f(p_{ij}) = d_{ij}^{(e)} = \sum_{a=1}^{m} [x_{ia}^{(e)} - x_{ja}^{(e)})^2]^{1/2}$, where $x_{ia}^{(e)} = x_{ia} + e_{ia}$ and e_{ia} is a value from the random distribution of point i.

For the "error" terms e_{ia}, one can postulate a particular distribution over time. A commonly used assumption is that the points oscillate symmetrically in all directions of the MDS space around their characteristic (true) locations. It is usually assumed that these distributions are normal, of equal size, and uncorrelated among each other, so that e_{ia} is modeled as

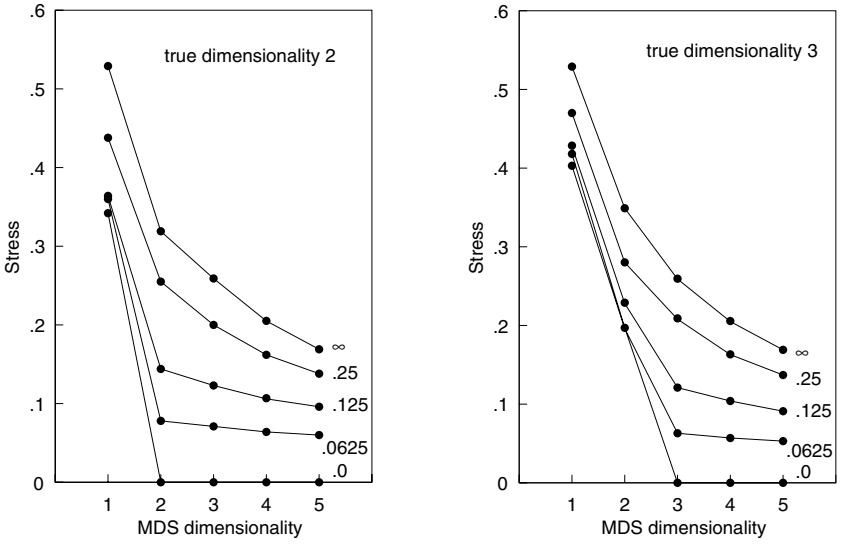

FIGURE 3.8. Expected Stress for distances in evenly scattered 2D (left panel) and 3D (right panel) configurations of 36 points with random error ranging from .0 to ∞, when represented in 1D through 5D (Spence & Graef, 1974).

a value sampled randomly from $N(0, \sigma^2)$. These definitions constitute the Hefner (1958) generalization of the Thurstone (1927) Case-V model.

Wagenaar and Padmos (1971) and Spence and Graef (1974) report simulation studies based on the Hefner model. They randomly pick n points from within a unit (hyper-)square or (hyper-)disk, and add error components sampled from $N(0, \sigma^2)$ to each of its coordinates. This leads to error-affected distances that are subsequently taken as proximities.

In contrast to the study by Spence and Ogilvie (1973) described above, this simulation allows one to specify the true (underlying) MDS space as the point configuration used in computing the proximities. Wagenaar and Padmos (1971) simulate this case for $n = 12, 18, 26, 36$; in $m = 1, \ldots, 4$ dimensions; and with error variances of $\sigma = 0.0, 0.0625, 0.125, 0.25$, and ∞ (i.e., pure random data).

Figure 3.8 shows the Stress curves obtained for proximities computed from 36 points in 2D and 3D MDS spaces, respectively, and represented in MDS spaces of one to five dimensions. One notes that all Stress curves are convex downwards. The upper curves in both diagrams almost have the same shape: they result from the condition of pure error. For the other conditions, we note elbows in the Stress curves for MDS dimensionalities of 2 and 3, respectively, that is, for the true dimensionalities of the underlying MDS spaces. These elbows are most pronounced in the error-free case, but are washed out with more and more error.

How large is the error component in these studies? One can check, by computer simulation, that the absolute difference of an error-affected distance (computed in an evenly scattered configuration of points within a unit disk) and the corresponding true distance, $|d_{ij}^{(e)} - d_{ij}|$, can be expected to be somewhat larger than the σs utilized by Spence and Graef (1974). That is, for example, for $m = 2$ and $\sigma = 0.25$, one finds that the expected absolute difference is 0.27, whereas for $\sigma = 0.0625$ and $m = 3$ it is 0.07. An error of judgment of about 25% does not seem excessive for many data in the social sciences. This may explain why elbows in scree plots are virtually never observed in practice, because, for $\sigma = 0.25$ or smaller σs, they are not obvious in Figure 3.8 either.

For real data, Spence and Graef (1974) propose comparing the Stress values for MDS solutions in different dimensionalities with their simulation curves in order to determine both the portion of error as well as the true dimensionality of the observations.

If one has an independent estimate of the error component in the data, the true dimensionality may be found by identifying that simulation curve among all those for the given error level that most closely matches the Stress curve for the given data. If the true dimensionality is known, one can proceed analogously for the error level. The conclusion depends, however, on the validity of the simulated error model.

Taking a closer look at the Hefner model, one notes that the normal error distribution is only a convenient approximation, because it puts no restrictions on the range of the point oscillations. Apart from that, however, the Hefner model has some interesting properties. It implies that error-affected distances tend to over-estimate true distances, because, by expanding the definition of $d_{ij}^{(e)}$, $E[(d_{ij}^{(e)})^2] = d_{ij}^2 + 2m\sigma^2$. Indeed, the error-affected distances are distributed as the noncentral χ^2 distribution (Suppes & Zinnes, 1963; Ramsay, 1969). Thus, a true distance of zero will only be over-estimated; small true distances can be expected to be more often over- than under-estimated; and the larger the true distance, the more balanced over- and under-estimation. This is a plausible model that prevents distance estimates from becoming negative.

Empirically, however, one often finds that dissimilarity judgments for very similar objects are more reliable than those for very dissimilar objects. Ramsay (1977), therefore, suggests making the error on the distances proportional to their size. In one particular model, the true distances are multiplied by a random factor whose logarithm has a normal distribution with mean 0 and standard deviation σ. This leads to a log-normal distribution for the error-affected distances where: (a) $d_{ij}^{(e)} \geq 0$; (b) the larger the true distance, the larger the noise; and (c) error-affected distances are

more likely to be over-estimated than under-estimated.[4] These properties seem to hold for many empirical contexts. However, what remains less clear is how this error model could be conceived in terms of what is going on in the psychological space.

Recovering Distances Under Noise

Simulation studies with error-perturbed distances do not assess how precisely the true distances are recovered by ordinal MDS. This question is investigated in the early days of MDS by Young (1970), Sherman (1972), Isaac and Poor (1974), and Cohen and Jones (1973), among others. Young (1970) proceeds as follows. (a) A true configuration with dimensionality t is defined by randomly sampling point coordinates. This yields true distances and, after adding error to the point coordinates, error-perturbed distances, as above. (b) The error-perturbed distances are monotonically transformed. (c) The resulting values are taken as data for an ordinal MDS procedure. (d) An MDS representation is then assessed with respect to the degree to which it recovers the true distances.

Young's simulations for different numbers of points, error levels, and monotone transformations—always setting $m = t$, so that the MDS analysis is in the true dimensionality—show that the precision of recovered distances grows with the number of points, and decreases with a higher error level in the data and with larger dimensionality of the solution space. This is intuitively plausible, because the isotonic regions in ordinal MDS shrink dramatically as a function of the number of points. Indeed, in the distances-among-cities example of Chapter 2, we found that the distances of the original map were almost perfectly reconstructed in a metric sense by the ordinal 2D MDS solution.

The effect of error on recovery precision is also easy to understand. More error simply reduces the correspondence of true distances and proximities. However, the harmful effect of error on recovery decreases with more points, because, with many points, the error-affected distances randomly over-estimate and under-estimate the true distances in so many ways that the effect of error on the configuration is balanced out and the solution essentially reconstructs the true distances. Stress, on the other hand, *in-*

[4]This error model, and related ones, is incorporated into the program MULTISCALE (see Appendix A). MULTISCALE does not minimize a loss function such as Stress. Rather, it tries to find that configuration $\mathbf{X}$ which, given a particular error model, maximizes the likelihood to yield $d_{ij}^{(e)}$s that correspond to the observed dissimilarities (maximum likelihood estimation). Given that the assumed error model holds, this allows one to determine confidence regions for the points and to make a number of inferential decisions, such as one on the proper dimensionality. Maximum likelihood MDS methods also exist for the Hefner error model (Zinnes & MacKay, 1983) and for ordinal MDS (Takane & Carroll, 1981).

creases when the number of points goes up, other conditions being equal! Cox and Cox (1990) even showed, by simulation, that Stress is an almost perfectly linear function of noise, given some special circumstances such as $m = t = 2$, but independently of the spatial pattern of points (ranging from extremely regular patterns through complete spatial randomness to cluster-like aggregations of points) and also independently of n. Cox and Cox (1992) report similar results for $m > 2$, but without such a strong linear relation between Stress and noise.

These findings have important practical implications. Global fit indices such as Stress are closely related to the proportion of error in the data. They are largely useless as measures of how well an MDS solution represents the "true" structure of the data. Therefore, it is quite possible that one obtains an MDS representation that has high Stress but that, nevertheless, is highly reliable over replications of the data. This means that a given Stress value should always be evaluated against some rough estimate of how much error is contained in the data.

An interesting further investigation on recovering true MDS spaces by means of ordinal MDS is presented by Sherman (1972), who studied, in particular, the effects of *over-* and *under-compression*. These notions refer to the question of whether the MDS dimensionality (m) is smaller or greater than the dimensionality of the space from which the proximities were derived (t). Sherman finds that picking the wrong dimensionality ($m \neq t$) has a pronounced effect: although Stress goes down monotonically when m goes up, the metric determinacy is best when $m = t$ and decreases with the extent of both over- and under-compression. There are slight differences though: under-compression, especially when there are many points in a relatively low-dimensional space, is somewhat less serious. This again shows that lower Stress (as a consequence of higher dimensionality) does not imply better metric recovery.

Summary on Stress

Stress is a badness-of-fit measure that depends, as we saw, on many factors. Some of them are:

- n, the number of points: the higher n, the higher Stress in general;
- m, the dimensionality of the MDS space: the higher m, the lower Stress;
- the error in the data: more error means higher Stress;
- the number of ties in the data (for ordinal MDS with weak monotonicity): more ties allow for lower Stress in general;
- the number of missing data: more missing data lead to lower Stress, in general;

TABLE 3.2. Average recovery coefficients, r^2s, for proximities related to true distances by $p_{ij} = d_{ij}^k$, under choice of different MDS models (Green, 1974).

Power k	Ratio MDS	Interval MDS
1.2	.99	.99
2.2	.94	.99
3.2	.85	.97
4.2	.78	.96
5.2	.72	.94

- the MDS model: interval MDS generally leads to higher Stress than ordinal MDS, particularly if f is markedly nonlinear and/or has major steps.

All of these criteria are mechanical ones. They are not sufficient for evaluating Stress, nor are they always important. Kruskal (1964a) writes: "A second criterion lies in the interpretability of the coordinates. If the m-dimensional solution provides a satisfying interpretation, but the $(m+1)$-dimensional solution *reveals no further structure* [our emphasis], it may be well to use only the m-dimensional solution" (p. 16). It is in this sense that Guttman (personal communication) called Stress a mere "technical measure." A measure of scientific significance, in contrast, would take into account the degree to which an MDS solution can be brought into a meaningful and replicable correspondence with prior knowledge or with theory about the scaled objects.

3.6 Recovering True Distances by Metric MDS

So far, we have investigated the performance of ordinal MDS only. In metric MDS, many of the above questions can be answered rather directly. For example, for interval MDS and error-free proximities, increasing the number of points has no effect on the goodness of recovery. If we scale under $t = m$, we can expect that Stress is zero for any n. Moreover, the correlation of the true and the recovered distances should be one. In ordinal MDS, in contrast, we cannot easily infer from the obtained Stress value how high the metric recovery is. This depends, among other things, on n, because the number of points is related to the size of the isotonic regions. If the data are not error-free, then interval MDS succeeds in representing somewhat more error variance in general when n is small, so that the metric recovery is likely to be less than perfect. If n grows, then both Stress and metric recovery go up, just as in ordinal MDS. Thus, it can be seen that the behavior of metric MDS is quite predictable without simulation studies.

The situation is not as easily diagnosed if we ask how well interval MDS does if the true relation between proximities and distances is not linear.

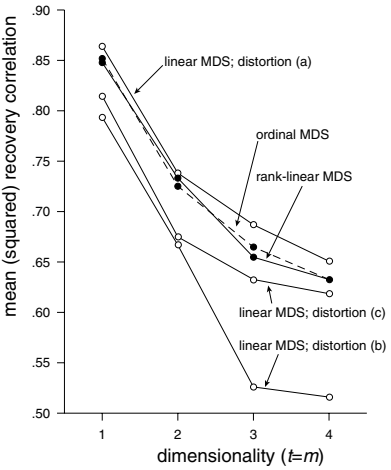

FIGURE 3.9. Recovery performance of MDS under choice of different models, number of dimensions, and distortions on proximities (after Weeks & Bentler, 1979).

Some answers are given by Green (1974). He selects $n = 10, 20$, and 30 points in $t = 2$ and $t = 3$ dimensions. Distances were computed and transformed into proximities by the function $p_{ij} = d_{ij}^k$, with $k = 1.2$, 2.2, 3.2, 4.2, and 5.2. Interval and ratio MDS were used to recover the underlying configurations from these proximities. The recovery coefficients in Table 3.2 show that ratio MDS is quite robust against such monotonic distortions of the function relating distances and proximities, as long as they are not extremely nonlinear. Interval MDS is almost unaffected by these (appreciable) nonlinear transformations.

Green (1974) demonstrates further that if we first substitute ranking numbers for the p_{ij} values, and then use ratio or interval MDS on these numbers, recovery is even better. The idea of *rank-interval MDS* was studied in more detail by Weeks and Bentler (1979). They used the following parameters for their simulations: $n = 10$, 20, 30; $t = 1$, 2, 3, 4; and $e = 0.25$, 0.75, 2.0, defined as the proportion of the error variance to the variance of the true distances. The proximities were derived from the error-perturbed distances by (a) $p_{ij} = d_{ij}^{(e)}$, (b) $p_{ij} = [d_{ij}^{(e)}]^4$, (c) $p_{ij} = [d_{ij}^{(e)}]^{1/4}$, or (d) $p_{ij} = \text{rank}[d_{ij}^{(e)}]$. Condition (d) is Green's ranking number substitution, and condition (a) simply means that the error-perturbed distances were taken directly as data, without any further distortions. These data were represented by both ordinal and interval MDS. The dimensionality of the solution space, m, varied from 1 to 6.

Figure 3.9 shows the main result of the study. The various curves are defined by the average values of the (squared) metric determinacy coefficient under the different conditions. As expected, all curves drop as $t = m$

goes up, because the higher the dimensionality, the more error variance can be represented by MDS, which negatively affects the metric determinacy of the solution. Ordinal MDS leads to the same recovery curve under all conditions. For (a), interval MDS does slightly better than ordinal MDS, but its recovery performance is definitely worse than that of ordinal MDS under the nonlinear distortions (b) and (c), as expected. However, with the ranking number substitutions, interval MDS leads to virtually the same recovery curve as ordinal MDS, as one can see from comparing the two lines with the solid black points in Figure 3.9. (Note that ranking number substitutions make all of the data sets used by Weeks and Bentler (1979) equivalent.) This replicates the finding of Green (1974). The "linearizing" effect of ranking number substitutions was also known to Lingoes (1965), who used this method in constructing initial configurations for ordinal MDS procedures.

Two conclusions can be derived from these results. (1) If proximities and distances are related in a linear way, then the metric information contained in the data is only marginally more powerful than the ordinal information contained in the data for recovering the true distances. (2) If proximities and data are related in a monotonic way, then ordinal and rank-interval MDS can be expected to lead to essentially the same solutions. This is important insofar as metric MDS methods are more robust in a numerical sense; that is, they generally are more likely to yield globally optimal solutions and are less likely to produce degenerate solutions (see Chapter 13).

3.7 Further Variants of MDS Models

The generic model relation (3.4) leaves room for many variants of MDS models not discussed so far. The most obvious way to generate such models is to specify the representation function f in different ways. There are many possibilities, and some of them are considered in Chapter 9. Further possibilities arise out of considering particular patterns of missing data. A whole model class, called unfolding, is discussed at length in Chapters 14 to 16. Then, one could partition the proximities into subsets, and specify independent fs or even different fs for each such subset rather than just one single f for all proximities as in (3.4).

At this point, we need not go into such models. We introduce, however, one generalization of (3.4) that allows us to introduce some notions useful for further classifying MDS models. Assume that we have more than one proximity for each pair (i, j). Such a case can arise, for example, if the data collection is replicated K times or if there are K persons, each giving rise to one set of proximities. In such a case, the proximities can be given three indices, p_{ijk} $(i, j = 1, \ldots, n; k = 1, \ldots, K)$. This means that they can be

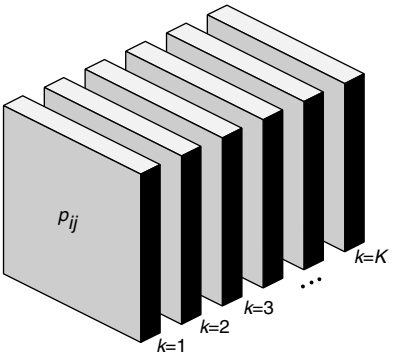

FIGURE 3.10. Symbolic representation of a three-way proximity array; $r = 1$ indicates replication 1.

collected in a *three-way* data array, as illustrated in Figure 3.10. This array can be conceived as a "deck" of K proximity matrices, where each "card" comes from one replication or person.

One could analyze such data by scaling each replication separately and then comparing or aggregating the different MDS solutions (see Chapter 20), or by first averaging the proximities over the replications and then scaling the aggregated scores. Another possibility is the following model,

$$f : p_{ijk} \rightarrow d_{ij}(\mathbf{X}), \tag{3.13}$$

for all pairs (i, j) and all ks, given that p_{ijk} is nonmissing. Note that this model relation differs from (3.4) only with respect to the proximities: it maps K proximities p_{ijk} (rather than just a single p_{ij}) into just one distance d_{ij}.

The three-way proximity block in Figure 3.10 suggests further possibilities for MDS models. Carroll and Arabie (1980) developed a taxonomy for MDS models according to which the three-way data in Figure 3.10 would also be characterized as *two-mode* data: "A mode is defined as a particular class of entities. ... Entities could be, for example, subjects, stimuli, test items, occasions, experimental conditions, geographical areas, or components of a 'multiattribute stimulus'. ... A K-way array is defined as the Cartesian product of a number of modes, some of which may be repeated. For example, an array associated with three-way multidimensional scaling might be of the form $A \times B \times B$, where A denotes subjects, and B stimuli" (p. 610). Hence, the "ways" of a proximity array refer, in a sense, to the number of subscripts of its proximities, whereas the "modes" distinguish whether these ways are qualitatively different ones.

There exist particular MDS models for three-way two-mode proximities, especially those where the "third" way denotes different individuals (see Chapters 21 and 22). There are also special models for two-way two-mode proximities, where one mode represents individuals and the other denotes

choice objects (see Chapter 17). Typical MDS data are, however, two-way one-mode proximities such as item intercorrelations or direct similarity ratings.

3.8 Exercises

Exercise 3.1 Consider the configuration in Figure 3.1. Compute the Euclidean distances among its $n = 6$ points.

(a) From these distances, generate dissimilarities by adding random error to each value. That is, $\delta_{ij} = d_{ij} + e_{ij}$, where e_{ij} is a value taken from a normal distribution $N(0, \sigma)$. (Alternatively, add random error to the point coordinates and then compute the distances. This may be easier to do within your statistics package.) Use different σs to simulate data with small, medium, and large error components. Run ordinal MDS with these dissimilarities. Compare the MDS solutions to Figure 3.1 and check the ability of ordinal MDS to recover the d_{ij}s from the δ_{ij}s.

(b) Repeat (a) using interval MDS.

(c) Repeat with $n = 20$ and $n = 40$ points that you choose at random in the plane shown in Figure 3.1, that is, with points (x, y), where $x, y \in [-4, +4]$.

Exercise 3.2 Suppose that the solution of Exercise 2.4 is given by the coordinates

	Dim 1	Dim 2
Red	0	3
Orange	0	0
Green	4	0
Blue	6	6

(a) Make a scatter plot of these points. Compute the distances between the points.

(b) Summarize the results in a table that has as its rows the six pairs of colors. Then, add a column that contains the proximity data for these pairs (see Exercise 2.4). Add a second column with the corresponding distances, computed from the table above. Finally, order the rows so that the row with the smallest proximity value is on top, and the row with the largest proximity at the bottom. Does the rank-order of the proximities match the rank-order of the distances? What do you conclude about the quality of the MDS solution?

Exercise 3.3 Consider data from 13 stock market indices of 784 daily measures from January 1, 1995, to December 31, 1997 (Groenen & Franses, 2000). From these data, the so-called return values are derived by taking the difference of the log of two subsequent index values. A correlation matrix of these stock market indices is given below.

Stock market		1	2	3	4	5	6	7	8	9	10	11	12	13
1	Brus	1.00												
2	CBS	.62	1.00											
3	DAX	.64	.69	1.00										
4	DJ	.29	.36	.21	1.00									
5	FTSE	.52	.69	.54	.38	1.00								
6	HS	.43	.40	.50	.11	.35	1.00							
7	Madrid	.51	.61	.57	.31	.59	.33	1.00						
8	Milan	.49	.50	.60	.15	.41	.37	.47	1.00					
9	Nikkei	.25	.28	.29	.04	.24	.33	.24	.23	1.00				
10	Sing	.34	.26	.36	.05	.25	.67	.26	.29	.29	1.00			
11	SP	.28	.35	.20	.96	.37	.09	.29	.14	.05	.04	1.00		
12	Taiwan	.04	.05	.07	-.03	.03	.15	.05	.07	.10	.19	-.03	1.00	
13	VEC	.52	.71	.62	.33	.63	.37	.61	.45	.25	.27	.32	.04	1.00

Now, the question is how different (or similar) the fluctuations are among the indices of the 13 stock markets.

(a) Use a computer program to do an interval MDS in 1 to 6 dimensions. Make a scree plot of the Stress values. Motivate your choice for the dimensionality of the solution.

(b) Can the Stress values be compared to the ones obtained for random data (see Figure 3.6) and the Hefner model? Explain why.

(c) Inspect Stress diagrams of your solution. What can you say about the fit? Do all points fit equally well?

(d) Interpret the solution. Can you distinguish groups of stock markets that have similar fluctuations?

(e) In what stock markets should you invest your money, if you want to spread the risks of your investment? Motivate.

(f) Redo the analysis with an ordinal transformation. Is the resulting configuration different? Compare the Shepard plots or the transformation plots. Is the difference in Stress small or large? Explain why this is so.

Exercise 3.4 Use the solution you like best from the previous exercise and compute the Stress per point and the fit per point. Produce a bubble plot that shows the fit per point either by hand or by a graphics program. Which are the worst fitting points? Which are the best fitting points? Interpret the solution again. Is it different from your first interpretation?

Exercise 3.5 Run an ordinal MDS analysis with your MDS program on the data from Table 2.3. The Stress of the resulting MDS solution is most likely not equal to zero even though we know that the distances were measured on a flat map.

(a) Explain why Stress is not zero.

(b) Try to get your MDS program to come up with a smaller Stress value.

(c) Compare the solution generated under the program's default settings with any one that has an even lower Stress. What do you conclude?