CHAPTER 3

Multidimensional Scaling

3.1 Introduction

Multidimensional scaling is one of several multivariate techniques that aim to reveal the structure of a data set by plotting points in one or two dimensions. The basic idea can be motivated by a geographical example. Suppose we are given the distances between pairs of cities and are asked to reconstruct the two-dimensional map from which those distances were derived. We could attempt to do this by a process of trial and error by moving points about on a sheet of paper until we got the distances right. A procedure that does this automatically is called multidimensional scaling (MDS). The "multi" part of the name refers to the fact that we are not restricted to constructing maps in one or two dimensions.

This simple example differs in two important ways from the typical MDS problem. In the first place, there is no ambiguity about what we mean by the "distance" between two cities (measured in miles or kilometres in a straight line), whereas in the typical MDS problem there is often a degree of arbitrariness in the definition of distance which, in some cases, may be based on subjective assessments rather than precise measurement. Secondly, we know that the cities can be located on a two-dimensional map (provided that the curvature of the earth and other topographical features can be ignored), whereas in the typical MDS problem we would have little idea how many dimensions would be necessary in order to reproduce, even approximately, the given distances between objects of interest. Indeed one of the prime objects of the analysis will be to discover whether such a representation is possible in a small number of dimensions. Unless this can be done, preferably in one or two dimensions, we shall not be able to take advantage of the eye's ability to spot patterns in the plots. Even if it turns out that more than two dimensions are necessary, the main way we can view the points is by projecting them onto two-dimensional space.

The input data for MDS is in the form of a distance matrix representing the distances between pairs of objects. We have already discussed the construction of such matrices in Chapter 2 and there is nothing to add here. However, whereas the choice between distance and proximity was largely a matter of indifference in cluster analysis, distance is the prime concept in MDS. Thus although we may start with a proximity or similarity matrix, it may need to be converted to a distance matrix in the course of the analysis; the output will be expressed in terms of distance.

As we have said, MDS is used to determine whether the distance matrix may be represented by a map or configuration in a small number of dimensions such that distances on the map reproduce, approximately, the original distance matrix $\{\delta_{ij}\}$. For example, we would aim to have the two objects that are closest together according to the distance matrix closest together on the map, and so on. As we have posed the problem, the distances on the map would be in the same metric (scale of measurement) as the original δ_{ij}s. This is often known as *classical* multidimensional scaling. However, it is often the case, particularly in social science research, that the values of the δ_{ij}s may be interpreted only in an ordinal sense as if, for example, the distances come from subjective similarity ratings. In such cases, it may be more reasonable only to attempt to produce a map on which the distances have the right rank order. This is called *ordinal* or *non-metrical* multidimensional scaling. In this chapter, we shall be mainly concerned with ordinal MDS. In the second example in Section 3.2 below, students were asked to rate the degree of similarity between pairs of countries on a nine-point scale. Similarity, here, is a subjective thing for which there is no natural underlying "space" reflected in the similarities. Part of the interest in the analysis is to try to uncover which attributes of the countries appear to carry weight in the students' judgement of similarity.

Returning to classical scaling, suppose that we have four cities labelled A, B, C, and D and that the distances (in hundreds of miles) between the pairs of cities are as given by the following matrix:

$$\begin{matrix} A \\ B \\ C \\ D \end{matrix} \begin{pmatrix} - & & & \\ 2 & - & & \\ 1 & 3 & - & \\ 5 & 3 & 6 & - \end{pmatrix}$$

Using multidimensional scaling (or by inspection), it is possible to represent this distance matrix exactly in one dimension. A possible solution is given in Figure 3.1.

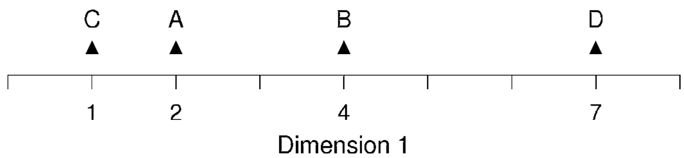

Figure 3.1 *A one-dimensional configuration of four cities using classical MDS*

We shall denote the distance between objects i and j in the above configuration by d_{ij} and in this case, these distances are precisely equal to the δ_{ij}s. In classical MDS, we seek a configuration such that the d_{ij}s, the inter-point distances in the configuration, will be approximately equal to the corresponding δ_{ij}s, as given in the distance matrix; whereas in ordinal MDS, the object is only to find a configuration such that the d_{ij}s are in the same rank order as the corresponding δ_{ij}s.

Given the Euclidean distances between n objects, it is always mathematically possible to find a configuration in $(n-1)$ dimensions that matches perfectly, but this would be of little use. Our aim will be to obtain a fairly good *approximate* representation in a small number of dimensions.

Measures of similarity between variables

We have already remarked in Section 2.7 that one can reverse the roles of variables and objects. Instead of clustering objects, which was our main concern, we could have clustered variables. This duality arises with all analyses that start from a data matrix. If we wished to carry out an MDS analysis on variables, we would need measures of similarity between columns of the data matrix instead of between the rows.

3.2 Examples

Reproducing a two-dimensional map from air distances between pairs of cities

MDS was carried out to determine whether a two-dimensional map could be produced from a matrix of pairwise distances between ten cities in Europe and Asia. The dissimilarity or distance matrix is shown in Table 3.1.

Table 3.1 *Distances between ten cities in air miles*

	London	Berlin	Oslo	Moscow	Paris	Rome	Beijing	Istanbul	Gibraltar	Reykjavik
London	–									
Berlin	570	–								
Oslo	710	520	–							
Moscow	1550	1000	1020	–						
Paris	210	540	830	1540	–					
Rome	890	730	1240	1470	680	–				
Beijing	5050	4570	4360	3600	5100	5050	–			
Istanbul	1550	1080	1520	1090	1040	850	4380	–		
Gibraltar	1090	1450	1790	2410	960	1030	6010	1870	–	
Reykjavik	1170	1480	1080	2060	1380	2040	4900	2560	2050	–

The solution from a classical MDS in two dimensions is shown in Figure 3.2.

The MDS has mapped points in two-dimensional space such that the "straight line" (Euclidean) distances between the points d_{ij} match the observed distances δ_{ij}. The d_{ij}s are very close to the (rescaled) δ_{ij}s. They are not precisely equal because the δ_{ij}s are not "straight line" distances but distances across the surface of a sphere.

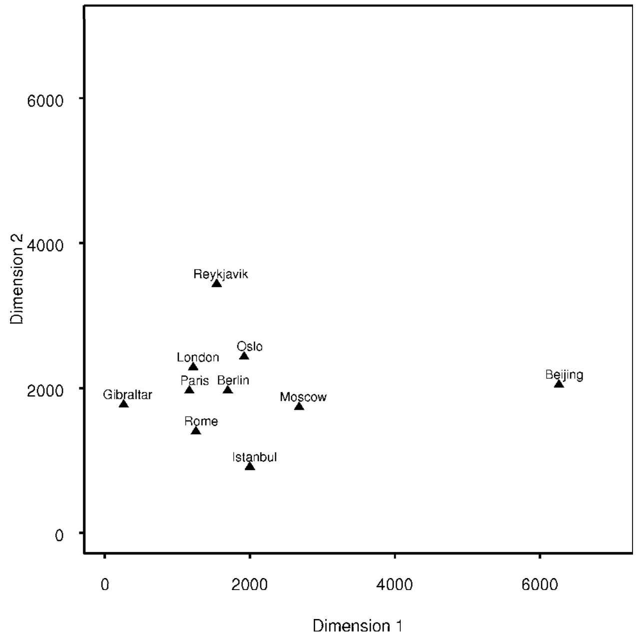

Figure 3.2 *Two-dimensional plot of 10 cities from a classical MDS*

Figure 3.2 is recognisable as a map of Europe and Asia. However, in general a configuration may need to be rotated and/or reflected in order to clarify the interpretation. Three important points about interpreting MDS solutions are:

i) The configuration can be reflected without changing the inter-point distances.

ii) The inter-point distances are not affected if we change the origin by adding or subtracting a constant from the row or the column coordinates.

iii) The set of points can be rotated without affecting the inter-point distances. This comes to the same thing as rotating the axes.

We must therefore be prepared to look for the most meaningful set of axes when interpreting an MDS solution. This idea will become clearer when we come to the next example. To summarise, the interpretation we put upon any MDS solution must be invariant under reflection, translation, and rotation.

An attempt to determine the dimensions underlying similarity judgements for pairs of 12 countries

In 1968, a group of 18 students was asked to rate the degree of similarity between each pair of 12 countries on a scale from 1 ("very different") to 9 ("very similar"). The study is described in Kruskal and Wish (1978), but our analysis is slightly different. The mean similarity ratings were calculated across students to obtain the similarity matrix in Table 3.2.

Table 3.2 *Subjective similarities between pairs of 12 countries*

	Brazil	Congo	Cuba	Egypt	France	India	Israel	Japan	China	Russia	USA	Yugo-slavia
Brazil	–											
Congo	4.83	–										
Cuba	5.28	4.56	–									
Egypt	3.44	5.00	5.17	–								
France	4.72	4.00	4.11	4.78	–							
India	4.50	4.83	4.00	5.83	3.44	–						
Israel	3.83	3.33	3.61	4.67	4.00	4.11	–					
Japan	3.50	3.39	2.94	3.83	4.22	4.50	4.83	–				
China	2.39	4.00	5.50	4.39	3.67	4.11	3.00	4.17	–			
Russia	3.06	3.39	5.44	4.39	5.06	4.50	4.17	4.61	5.72	–		
USA	5.39	2.39	3.17	3.33	5.94	4.28	5.94	6.06	2.56	5.00	–	
Yugo-slavia	3.17	3.50	5.11	4.28	4.72	4.00	4.44	4.28	5.06	6.67	3.56	–

Ordinal MDS was applied to this similarity matrix, because the similarities are based on subjective judgements. The resulting solution in two dimensions is shown in Figure 3.3 below.

We have to consider whether we can identify what is varying as we move along the two axes. Thus, for example, what do those countries on the right of the diagram have more of than those on the left, or those at the top than those at the bottom? Nothing very obvious seems to emerge from such comparisons but we must remember that the orientation is arbitrary and maybe the message will be clearer if we consider other rotations. The dotted axes shown on Figure 3.3 correspond to a rotation that does seem to have an interpretation in terms of meaningful variables. Kruskal and Wish (1978), note that variation in the direction of the axis that runs from bottom left to top right corresponds to a tendency to be pro-Western or pro-Communist. Those at the top right are the more pro-Communist and those at the bottom left are the more pro-Western. Variation in the direction at right angles separates the developed (top left) from the developing (bottom right) countries. It thus appears that when making their judgements in 1968, the students were taking account, consciously or unconsciously, of two types of difference, and the analysis has helped us to identify what those two dimensions were.

It is worth adding two cautionary remarks about this example. The similarities were obtained by averaging the assessments of the 18 students. Implicitly, therefore, we are assuming that all are using the same two dimensions and that they are giving them the same relative weight. This may not be the case and it would be useful to have a method of discovering whether this was

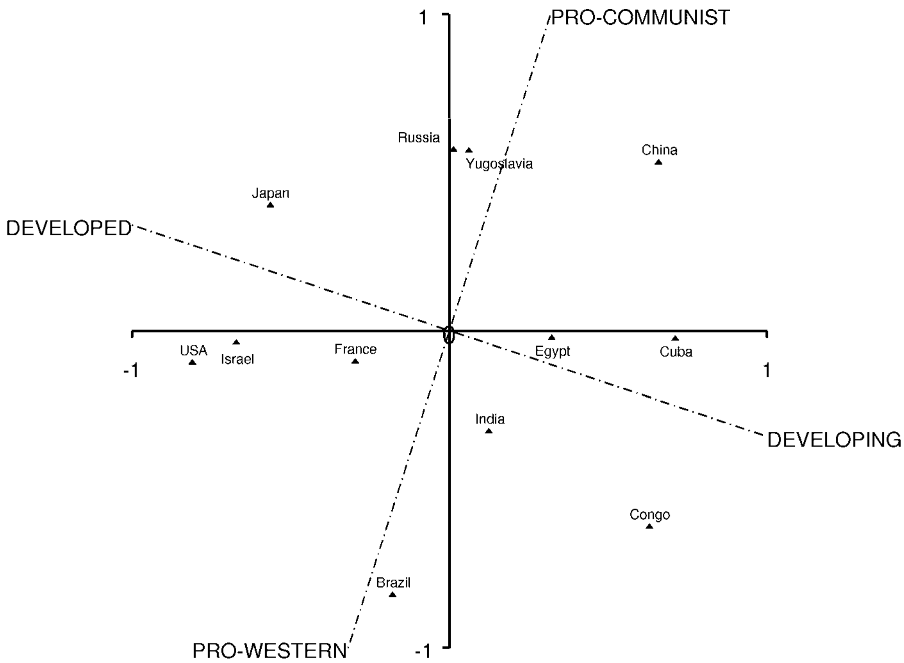

Figure 3.3 *Two-dimensional plot of countries from ordinal MDS*

true. Such methods, known as Individual Scaling or Three-Way Scaling, are available but are outside the scope of this book (see, for example, Borg and Groenen (2005) or Kruskal and Wish (1978)).

The second remark is that the identification of interpretable axes for a plot is not always the best way of discerning interesting patterns. It may be that we can identify clusters of points which have practical significance, as in the acoustic confusion example in Section 3.7, or, as in the colour data example in Section 3.6, the clue may be in the "horseshoe" shape of the two-dimensional plot.

3.3 Classical, ordinal, and metrical multidimensional scaling

We now pose the problem of multidimensional scaling in more formal terms so that we can outline the algorithms used to arrive at a solution.

Classical scaling

In *classical* MDS, the aim is to find a configuration in a low number of dimensions such that the distances between the points in the configuration, d_{ij}, are close in value to the observed distances δ_{ij}. The method treats the distances as Euclidean distances. We saw in Chapter 2 how to go from a data matrix to a Euclidean distance matrix; here we have to go in the reverse direction and recover the data matrix from the distances. We cannot recover everything because information about location and orientation is lost in the process of calculating distances, but we can determine the configuration. This problem can be tackled algebraically, and it turns out that the solution gives us a series of approximations starting with one dimension, then two, and so on. It also happens, however, that the mathematics involved is equivalent to that for another problem for which the solution is already known. This establishes an interesting link with principal components analysis that we shall discuss in Chapter 5. We shall return to this link in that chapter but we can prepare the ground by expressing the classical MDS problem in a slightly different way. If we start with an $n \times p$ data matrix, we first construct a distance table and then might seek to find a two- or three-dimensional map on which the inter-point distances are as close as possible to the original distances. Another way of putting this is to say that we are looking for a new data matrix, with two or three columns, which is close to the original matrix in the sense that it gives rise to (nearly) the same distance matrix.

Having found a solution, we may wish to have a measure of how good the fit is. This would be particularly useful for helping us to judge how many dimensions are necessary to get a good enough fit. An obvious way to do this is to look at the sum of squares $\sum_{i<j}(d_{ij} - \delta_{ij})^2$. (This is mathematically appropriate since the fits obtained are best in a least squares sense.) However, the simple sum of squares depends on the scale in which the distances are measured. It is, therefore, preferable to normalize the sum of squares and, in order to reduce it to the same units as the distances, to take the square root. Our goodness-of-fit measure is then

$$\sqrt{\frac{\sum_{i<j}(d_{ij} - \delta_{ij})^2}{\sum_{i<j} d_{ij}^2}}. \tag{3.1}$$

This measure is called the *stress* or, sometimes, the normalised stress. There are other ways of calculating a normalised stress measure. For example, an alternative measure of stress may be obtained by replacing d_{ij} with δ_{ij} in the denominator of equation (3.1). Values of stress that are close to zero would indicate that the MDS solution is a good fit to the original δ_{ij}s.

Ordinal (non-metrical) scaling

Very often it is not the actual value of δ_{ij} that is important or meaningful, but its value in relation to the distances between other pairs of objects. This is particularly true when the δ_{ij}s are the result of an experiment where subjects are asked to give their subjective assessments of the distance between objects. In such cases, the δ_{ij}s can be interpreted only in an ordinal sense. In *ordinal* MDS, the aim is to find a configuration such that the d_{ij}s are in the same rank order as the original δ_{ij}s. So, for example, if the distance apart of objects 1 and 3 rank fifth among the δ_{ij}s then they should also rank fifth in the MDS configuration. The emphasis in this chapter, as noted in Section 3.1, is on ordinal MDS.

In ordinal MDS, we construct fitted distances, often called *disparities*, $\hat{d}_{ij}$, from the d_{ij}s such that the $\hat{d}_{ij}$s are in the same rank order as the δ_{ij}s (for dissimilarities) or reverse rank order (for similarities). We can think of the $\hat{d}_{ij}$s as "smoothed" versions of the d_{ij}s. This smoothing process is carried out using a method called least-squares monotonic regression ("monotonic" means that the regression curve is either non-decreasing or non-increasing). Using this method, the d_{ij}s are regressed on the δ_{ij}s. In a plot of d_{ij} versus δ_{ij}, we would like to see a monotonic curve (one where the lines joining adjacent points are flat/increasing if δ_{ij} are dissimilarities or flat/decreasing if δ_{ij} are similarities). If the d_{ij}s and the δ_{ij}s have the same rank order, then the plot will show such a monotonic curve and the d_{ij}s will not require any smoothing. Usually, however, there will be some departures from monotonicity and some smoothing will be necessary. The aim of monotonic regression is to fit a monotonic curve to the points (d_{ij}, δ_{ij}), while making the sum of squared vertical deviations as small as possible (as in least-squares linear regression). The point on the monotonic curve, $\hat{d}_{ij}$, is the fitted or predicted value of d_{ij} from the monotonic regression. In judging how good the fit is, we are now interested in how close the distances, d_{ij}, are to the disparities, $\hat{d}_{ij}$, rather than the observed distances, δ_{ij}. This is because we are only aiming to reproduce the rank order of the observed distances and not the distances themselves. Hence, our measure of fit is obtained by cleverly replacing δ_{ij} by $\hat{d}_{ij}$ in the formula for the stress ($\hat{d}_{ij}$ and δ_{ij} having the same rank order). Thus in ordinal MDS, the stress is calculated as

$$\sqrt{\frac{\sum_{i<j}(d_{ij} - \hat{d}_{ij})^2}{\sum_{i<j} d_{ij}^2}}. \tag{3.2}$$

This is also known as Kruskal's stress, type I (which we shall refer to simply as stress). The optimum configuration is determined by minimising this measure of stress or some variant of it.

The points (δ_{ij}, d_{ij}) are shown by a cross in Figure 3.4. Note that while the first and second points (counting from left to right) follow a monotonic pattern, the third does not. To achieve monotonicity, the values of d_{ij} for the second and third points are replaced by their mean. Similarly, the values of

d_{ij} for the fourth and fifth points are replaced by their mean. This leads to the monotonic regression curve consisting of the series of solid lines shown in the plot. The vertical dotted lines represent the distances $d_{ij} - \hat{d}_{ij}$.

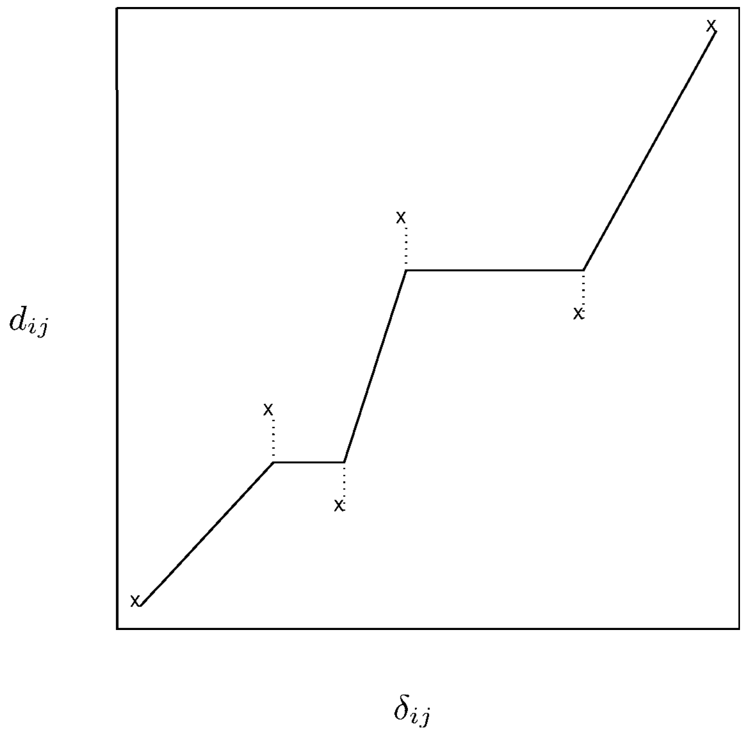

Figure 3.4 *Example of monotonic regression*

Metrical scaling

Classical scaling could be described as metrical scaling since, in contrast to non-metrical scaling, the fitted and original distances are expressed in the same metric. However, the term metrical scaling usually seems to be reserved for something which may most naturally be thought of as related to non-metrical (ordinal) scaling in another way. In classical scaling, we supposed that the distances were Euclidean distances. In ordinal scaling, we made use only of the rank order of the δ_{ij}s. This was tantamount to assuming that we had to make a monotonic transformation of the δ_{ij}s to turn them into Euclidean distances. In metrical scaling, we assume that they can be transformed into Euclidean distances by some other parametric transformation. In some fields, there may be good reasons for supposing that such transformations exist, but we are not aware of any convincing arguments for introducing them in social science applications. However, we mention two special cases because they are closely linked to classical scaling. *Interval scaling* refers to the case where it is supposed that a linear transformation will turn the δ_{ij}s into Euclidean distances. Instead of fitting a monotonic regression to the distances to obtain the disparities, we would now fit a linear regression. The disparities would then become the points on the regression line instead of points on the monotonic regression curve. The formula for stress remains the same except that the $\hat{d}_{ij}$s would now be obtained from the least squares regression line. In the special case of *ratio scaling*, when the regression goes through the origin, we are back to the situation we faced in classical scaling because multiplying the δ_{ij}s by a constant does not change the metric — if they were Euclidean before they will be Euclidean afterwards and vice versa. The difference here lies in the function which is being minimised. The Kruskal's stress formula applied in this case aims to achieve the closest degree of proportionality between the given distances and those fitted. Classical scaling aims to achieve the closest fit in a least squares sense. The two methods will often give very similar results and we shall use ratio scaling in one of the examples below.

3.4 Comments on computational procedures

Given the number of dimensions, k, the aim of MDS is to find a configuration in k dimensions such that the stress criterion used is minimised.

Most ordinal MDS computer packages start with an initial configuration in k dimensions, and then iteratively improve the configuration by moving the points short distances in such a manner as to reduce the stress slightly on each iteration. When further changes to the configuration do not reduce the stress (or not by more than some pre-specified tolerance level), the procedure ends and that configuration is the MDS solution. Typically, the method of steepest descent is used. Kruskal and Wish (1978) give the analogy of a blindfolded parachutist trying to find the lowest point in a terrain by following the gradient downhill.

Unfortunately, it is possible that a *local* minimum rather than the *global* minimum will be found. Repeating the process with different starting configurations to see whether the same minimum is found is one way of checking for this, but there is no absolute guarantee that there may not be some even smaller minimum lurking in a region of the space which has not been explored.

The MDS solution achieved depends on

i) the choice of initial configuration

ii) the stress criterion used

For example, the program PROXSCAL (available in SPSS), with which many of the calculations in this chapter were done, arrives at a solution which minimises a stress function with d_{ij} replaced with δ_{ij} in the denominator of the formula for Kruskal's stress type I. There are other variants of stress which measure the differences between the distances and the disparities in slightly different ways.

Full discussion of such computational issues is outside the scope of this book, but the reader should be aware that different packages may give slightly different solutions. If the solutions are very different, this suggests that either there is no strong structure in the data, or that at least one of the solutions is a local rather than a global optimum, or that complete convergence has not been achieved for one or both solutions.

3.5 Assessing fit and choosing the number of dimensions

There are a number of ways of assessing the fit of a MDS solution. One method involves comparing the stress obtained for the solution with the guidelines shown in Table 3.3. These were developed by Kruskal (1964) and are based on empirical experience rather than theoretical criteria. These should always be used flexibly with an eye on the interpretability of the solution to which they lead.

Table 3.3 *Guidelines for assessing fit using stress*

Stress (Kruskal's type I)	Assessment of fit
0.20	poor
0.05	good
0.00	perfect

Another method that may be used to choose the number of dimensions is to examine a scree plot in which the stress is plotted against the number of dimensions. As the number of dimensions increases the stress decreases, but there is a trade-off between improving fit and reducing the interpretability of the solution. In the scree plot, we look for an "elbow" which is the point at which increasing the number of dimensions has little further effect on the stress. Again there is a strong subjective element in using this method, but experience shows that it often works well. See, for example, Figure 3.5 below.

There are also a number of useful diagnostic plots. In the case of ordinal scaling, the plots involve all pairs of δ_{ij}, d_{ij} and $\hat{d}_{ij}$, that may be examined to evaluate the fit of a MDS solution.

i) Plot of d_{ij} (the inter-point distance in the configuration) versus $\hat{d}_{ij}$ (the disparity or fitted value of d_{ij} obtained from the monotonic regression on δ_{ij}). If the MDS solution is a good fit, this plot should show a linear relationship with a 45 degree slope and only a small amount of scatter about the line. If little smoothing of the d_{ij}s was necessary to produce the $\hat{d}_{ij}$s, then they should be in almost the same rank order and close in value since they are measured on the same scale. See, for example, Figure 3.7.
ii) Plot of d_{ij} (the inter-point distance in the configuration) versus δ_{ij} (the observed distance or dissimilarity or similarity). If the solution is a good fit, d_{ij} and δ_{ij} should have approximately the same (or the reverse) rank

order and this plot should show a monotonic curve, either increasing (for dissimilarities) or decreasing (for similarities). See, for example, Figure 3.8.

iii) Plot of $\hat{d}_{ij}$ (the disparity or fitted value of the inter-point distance, d_{ij}) versus δ_{ij} (the observed distance or dissimilarity or similarity). The $\hat{d}_{ij}$s are the "smoothed" versions of the d_{ij}s constructed to have the same rank order as δ_{ij} (for dissimilarities) or reverse rank order (for similarities). If a large amount of smoothing were required to achieve a monotonic curve (that is, if the solution were a poor fit), this plot would show a number of large horizontal steps where the smoothing took place. When the fit is good there will only be small steps. See, for example, Figure 3.9.

For metrical scaling, the $\hat{d}_{ij}$s are made to be proportional to the δ_{ij}s. Therefore, the plots involving δ_{ij} are redundant, leaving only the plot of d_{ij} versus $\hat{d}_{ij}$ to be examined.

3.6 A worked example: dimensions of colour vision

We now illustrate these ideas and methods on an example which was originally analysed by other means before the development of multidimensional scaling methods.

An experiment was conducted where subjects were asked to look at a screen which had two circular opaque glass windows. These windows were lit from two projectors behind the screen. Different colour filters could be inserted in the projectors. Fourteen colour filters were used, transmitting light of wavelengths $434m\mu$ to $674m\mu$. Each stimulus was combined with each other stimulus in a random order. The subjects were then asked to rate the degree of "qualitative similarity" between each pair of colour filters on a five-point scale. Further details, and the original analysis, will be found in Ekman (1954). The similarity matrix constructed by Ekman is given in Table 3.4. An ordinal MDS of these similarities was carried out.

This is a case where we might guess in advance that a one-dimensional solution would be possible because the difference in wavelength between two colours is a continuous metric measuring how far apart the colours are. However, the scree plot given in Figure 3.5 shows that there is a big reduction in stress in passing from one to two dimensions, so there must be other factors which come into play when making subjective assessments of colour. The "elbow" at two dimensions indicates that there is little reduction in stress after two dimensions. Therefore, we select a two-dimensional solution. This solution has stress of 0.03 (3%) which according to Kruskal's guidelines is a good fit.

In the two-dimensional configuration (Figure 3.6), the points appear on a curve to give a "horseshoe" effect — a common phenomenon. At one extreme, are the violets (colours 1 and 2) and at the other are the reds (colours 11-14). As we go round the horseshoe, we encounter the colours in strict order of wavelength. However, it appears that subjects were making more subtle

Table 3.4 *Similarities between colours based on subjective judgements*

Colour	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	–	.86	.42	.42	.18	.06	.07	.04	.02	.07	.09	.12	.13	.16
2	.86	–	.50	.44	.22	.09	.07	.07	.02	.04	.07	.11	.13	.14
3	.42	.50	–	.81	.47	.17	.10	.08	.02	.01	.02	.01	.05	.03
4	.42	.44	.81	–	.54	.25	.10	.09	.02	.01	.01	.01	.02	.04
5	.18	.22	.47	.54	–	.61	.31	.26	.07	.02	.02	.01	.02	.01
6	.06	.09	.17	.25	.61	–	.62	.45	.14	.08	.02	.02	.02	.01
7	.07	.07	.10	.10	.31	.62	–	.73	.22	.14	.05	.02	.02	.01
8	.04	.07	.08	.09	.26	.45	.73	–	.33	.19	.04	.03	.02	.02
9	.02	.02	.02	.02	.07	.14	.22	.33	–	.58	.37	.27	.20	.23
10	.07	.04	.01	.01	.02	.08	.14	.19	.58	–	.74	.50	.41	.28
11	.09	.07	.02	.01	.02	.02	.05	.04	.37	.74	–	.76	.62	.55
12	.12	.11	.01	.01	.01	.02	.02	.03	.27	.50	.76	–	.85	.68
13	.13	.13	.05	.02	.02	.02	.02	.02	.20	.41	.62	.85	–	.76
14	.16	.14	.03	.04	.01	.01	.01	.02	.23	.28	.55	.68	.76	–

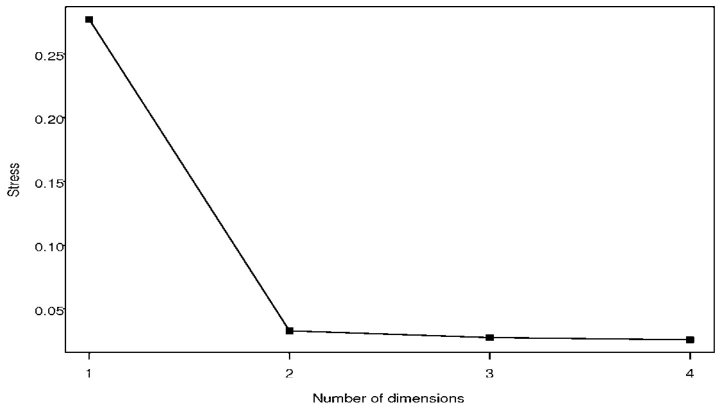

Figure 3.5 *Scree plot of stress by number of dimensions, colour data*

judgements in that reds are seen as closer to violets than to greens (colours 6-8), even though reds and greens are closer in terms of their wavelengths. Reference back to Table 3.4 confirms that this is not an accidental artefact of the MDS solution. There is clearly some other aspect of the perception of colour influencing the subject's comparisons than is conveyed by wavelength alone.

The three diagnostic plots are typical of what one finds with a reasonably good fit. On Figure 3.7, the points lie close to the 45 degree line; the curve in Figure 3.8 shows marked monotonicity, and Figure 3.9 has horizontal steps of short length reflecting the near monotonicity shown by the previous figure.

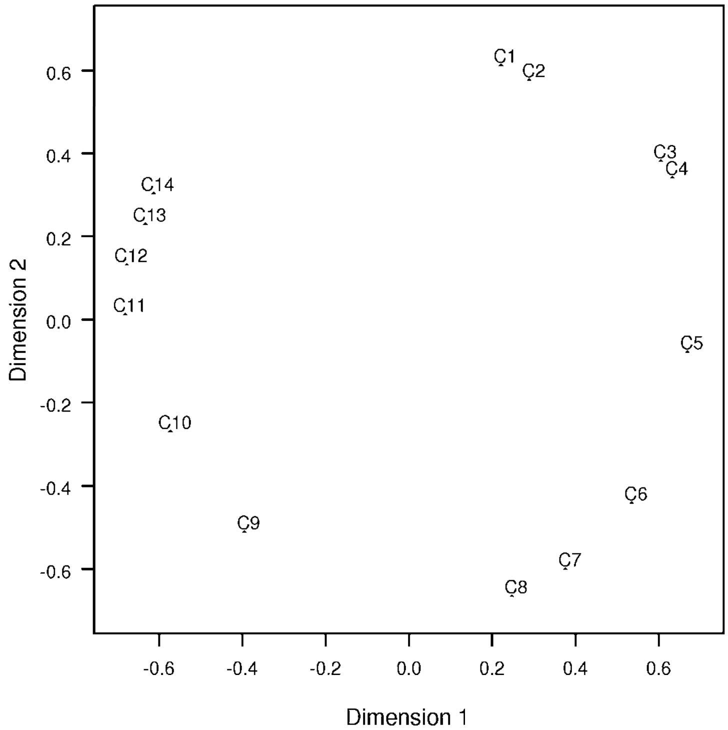

Figure 3.6 *Two-dimensional configuration plot from an ordinal MDS of colour data*

Before leaving this example, it is interesting to return to the one-dimensional solution plotted in Figure 3.10. The colours do not appear in order of their wavelength though there is a clear separation between the "blue" end of the spectrum (colours with low numbers) and the "red" (colours with high numbers). Within those two groups, however, there seems to be some inversion of the order one would have expected. The fit, of course, was not good in this case. The stress was 0.28, which on Kruskal's criterion, indicates a poor fit.

The clear conclusion of our analysis is that colour perception involves more than is conveyed by the wavelength of the light. To return to the title of Ekman's paper, there appear to be two dimensions of colour vision.

3.7 Additional examples and further work

In this section, we give four further examples to illustrate the methods. We shall not carry out an exhaustive analysis on any of them, but focus on particularly interesting features which the individual examples show. You are invited to use these examples to explore the other options available in the various software packages. Two of the examples have already occurred in the chapter on cluster analysis, and our main interest in these cases will be to compare the two methods when applied to the same data.

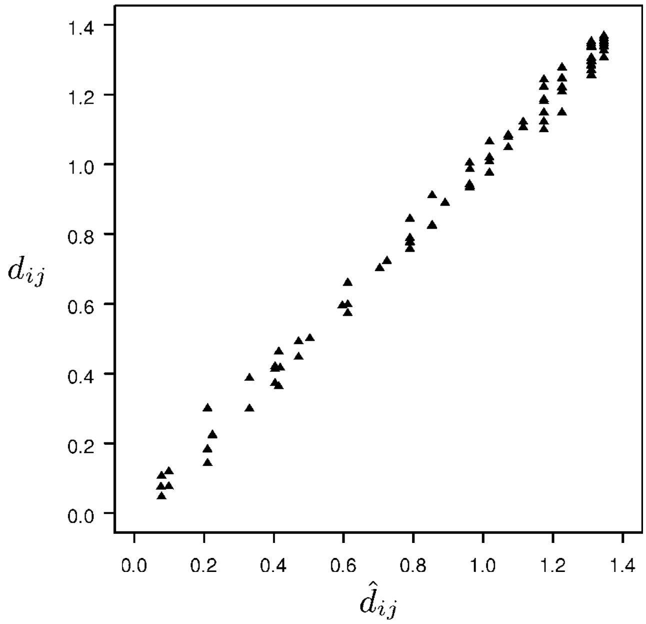

Figure 3.7 *Plot of* d_{ij} *(inter-point distance in the configuration) versus* $\hat{d}_{ij}$ *(fitted value of* d_{ij}*) from a two-dimensional ordinal MDS of colour data*

Economic and demographic indicators for 25 countries

Table 3.5 shows the values of five economic and demographic indicators for a sample of 25 countries. The data refer to 1990 and they come from the United Nations Statistical Yearbook of 1997. The indicators are annual percentage population growth rate (Increase), life expectancy in years (Life), infant mortality rate per 1000 (IMR), total fertility rate (TFR), and Gross Domestic Product per capita in US dollars (GDP).

Ratio MDS was applied to these data. Since the data are in the form of a data matrix, the first stage of a MDS is to convert the data to a distance matrix showing the pairwise distances between countries. Since the variables differ greatly in terms of their variances, the variables are first standardized to have a variance of 1. Euclidean distances are then computed. Since we apply ratio MDS, the fitted distances will be proportional to the actual distances. You should try ordinal scaling and compare the results.

One aim of a MDS of these data might be to determine whether countries can be placed on a scale of development based on these five indicators. Therefore, the one-dimensional solution is of particular interest. Developed countries are generally characterised by low growth rate, high life expectancy, low infant mortality, low fertility and high GDP. If countries can be located on a single dimension of development, developed countries should be placed at one extreme with less developed countries (characterised by high growth

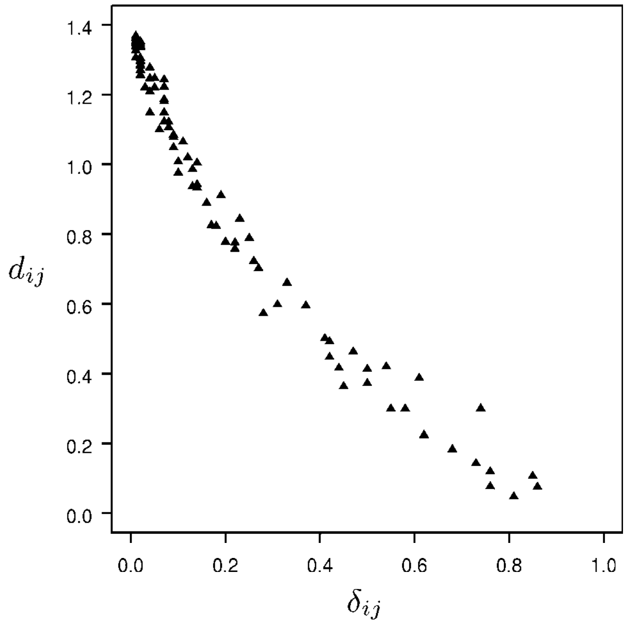

Figure 3.8 *Plot of* d_{ij} *(inter-point distance in the configuration) versus* δ_{ij} *(observed similarity) from a two-dimensional ordinal MDS of colour data*

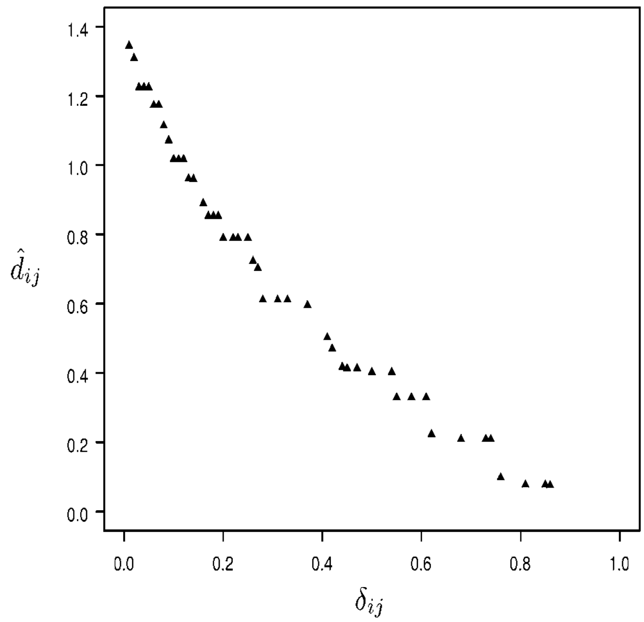

Figure 3.9 *Plot of* $\hat{d}_{ij}$ *(fitted value of* d_{ij}*) versus* δ_{ij} *(observed similarity) from a two-dimensional ordinal MDS of colour data*

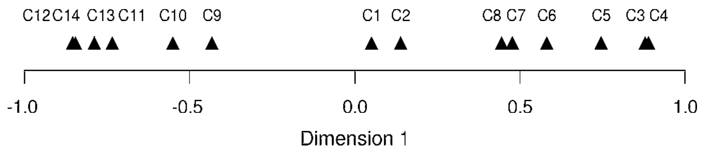

Figure 3.10 *MDS solution for colour data in one dimension*

Table 3.5 *Economic and demographic indicators for 25 countries, 1990, UN Statistical Yearbook of 1997*

Country	Increase	Life	IMR	TFR	GDP
Albania	1.2	69.2	30	2.9	659.91
Argentina	1.2	68.6	24	2.8	4343.04
Australia	1.1	74.7	7	1.9	17529.98
Austria	1.0	73.0	7	1.5	20561.88
Benin	3.2	45.9	86	7.1	398.21
Bolivia	2.4	57.7	75	4.8	812.19
Brazil	1.5	64.0	58	2.9	3219.22
Cambodia	2.8	50.1	116	5.3	97.39
China	1.1	66.7	44	2.0	341.31
Colombia	1.7	66.4	37	2.7	1246.87
Croatia	−1.5	67.1	9	1.7	5400.66
El Salvador	2.2	63.9	46	4.0	988.58
France	0.4	73.0	7	1.7	21076.77
Greece	0.6	75.0	10	1.4	6501.23
Guatemala	2.9	62.4	48	5.4	831.81
Iran	2.3	67.0	36	5.0	9129.34
Italy	−0.2	74.2	8	1.3	19204.92
Malawi	3.3	45.0	143	7.2	229.01
Netherlands	0.7	74.4	7	1.6	18961.90
Pakistan	3.1	60.6	91	6.2	385.59
Papua New Guinea	1.9	55.2	68	5.1	839.03
Peru	1.7	64.1	64	3.4	1674.15
Romania	−0.5	66.6	23	1.5	1647.97
USA	1.1	72.5	9	2.1	21965.08
Zimbabwe	4.4	52.4	67	5.0	686.75

rate, low life expectancy, high infant mortality, high fertility and low GDP) placed at the other extreme.

The stress (Kruskal type I) value for the one-dimensional solution was 0.17, suggesting a poor fit. The locations of the countries on a single dimension are given in Table 3.6. We find that the countries lie approximately where we would expect. At one extreme, we have the less developed mainly African and Asian countries, while at the other we have European countries, the USA and Australia. Since the one-dimensional fit is poor, however, you should go on to examine a two-dimensional solution to see whether a second dimension improves the fit and adds any new insight into the structure of the data.

Table 3.6 *Coordinate for each country from a one-dimensional ratio MDS of Economic and demographic indicators (arranged in increasing order)*

Country	Coordinate
Malawi	−2.027
Benin	−1.616
Cambodia	−1.414
Zimbabwe	−1.302
Pakistan	−1.133
Bolivia	−0.798
Papua New Guinea	−0.783
Guatemala	−0.706
El Salvador	−0.344
Peru	−0.277
Iran	−0.167
Brazil	−0.112
Colombia	0.036
China	0.188
Albania	0.220
Argentina	0.327
Romania	0.786
Greece	0.921
Australia	1.049
USA	1.105
Netherlands	1.158
Austria	1.164
Croatia	1.167
France	1.230
Italy	1.328

The stress value for the two-dimensional solution is 0.05, indicating a much better fit than the one-dimensional solution. Figure 3.11 shows the plot of d_{ij} versus $\hat{d}_{ij}$. The strong linear relationship between the distances in the configuration and the smoothed distances is a further indication that the data are well represented in two dimensions. As noted in Section 3.5, with ratio MDS the other two diagnostic plots, involving δ_{ij}, are redundant since δ_{ij} and $\hat{d}_{ij}$ have been made to be proportional.

The two-dimensional configuration is shown in Figure 3.12. The location of countries on dimension 1 is almost the same as in the one-dimensional solution. Dimension 1 could be interpreted as a measure of overall development. On the second dimension, Romania and Croatia stand out from the other countries. If you look at the profiles of these countries in Table 3.5, you can see that they both have characteristics associated with developed countries (low growth rate, moderately high life expectancy, fairly low infant mortality and very low fertility), which places them on the left-hand side of the first dimension together with other developed countries. However, they have very low GDPs compared to other developed countries. Those countries located at the other end of the second dimension generally have high GDPs. Thus, the second dimension is largely a function of GDP.

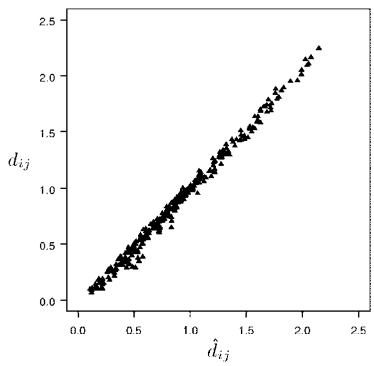

Figure 3.11 *Plot of d_{ij} versus $\hat{d}_{ij}$ from a two-dimensional ratio MDS of economic and demographic indicators*

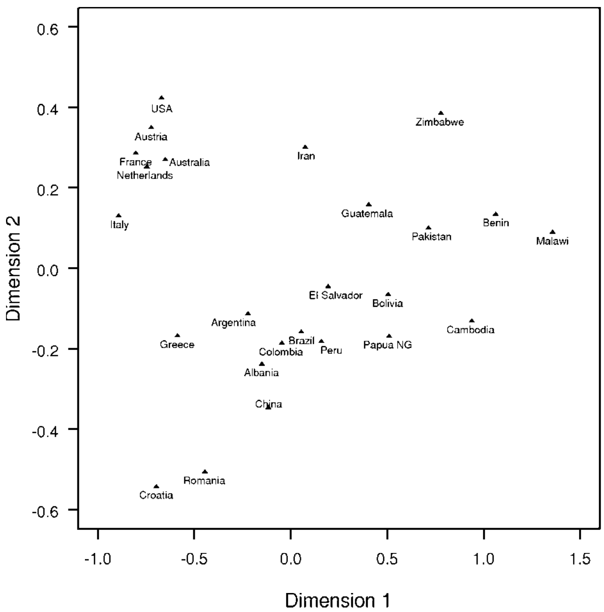

Figure 3.12 *Configuration of countries from a two-dimensional ratio MDS of economic and demographic indicators*

Persian archers

In Chapter 2, similarities between pairs of 24 archers (Table 2.13) were analysed using cluster analysis. The data are described in Section 2.8. The similarities may also be analysed using ordinal MDS.

From the scree plot in Figure 3.13, there is a suggestion of an elbow at two dimensions, indicating that a two-dimensional solution may be adequate, but three or four dimensions might improve the representation of the dissimilarities between the archers. The configuration for the two-dimensional solution is plotted in Figure 3.14.

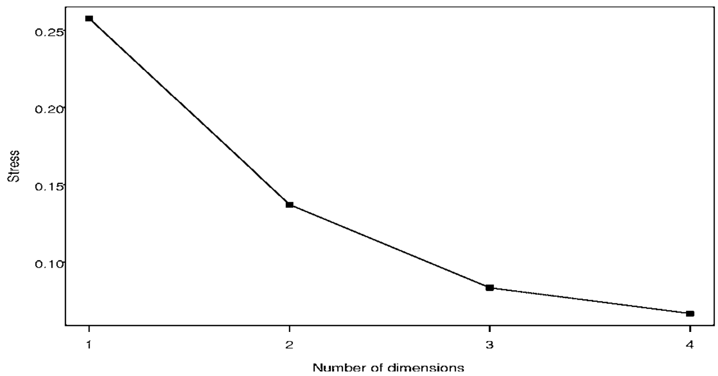

Figure 3.13 *A scree plot for an ordinal MDS of data on 24 Persian Archers*

Archaeologists want to know how the bas-reliefs were carved. Were they the work of a single sculptor, several independent sculptors, or of one or more teams of sculptors?

Figure 3.14 shows five archers (20 to 24) clustered together to the left of centre near the bottom; eight archers (1-8) spread out upwards and slightly diagonally on the left; the remaining archers (9 to 19) are spread out on the right. Roaf (1983), p. 14-16, as we noted in Chapter 2, concluded that there could have been three teams of sculptors. One working on the top section of the staircase (1 to 8), another on the centre section (9 to 19) and a third on the bottom section (20 to 24), these last five being so similar that they could be the work of a single sculptor. Within this broad clustering into groups, adjacent archers on the staircase tend to be close to each other in the configuration.

You may suggest explanations of why archers 1 to 8 are strung out in a line in Figure 3.14, why archer 2 appears relatively close to archers 20 to 24, and why archer 12 is distant from the others. Then turn to Figure 3.15 where lines have been added joining points (archers) with similarity of 15 or more. Such additions to the plot of the MDS solution can clarify whether the relative positions of individual points in the configuration reflect their true similarities. Points close together on the map but with low similarities will be major contributors to the stress.

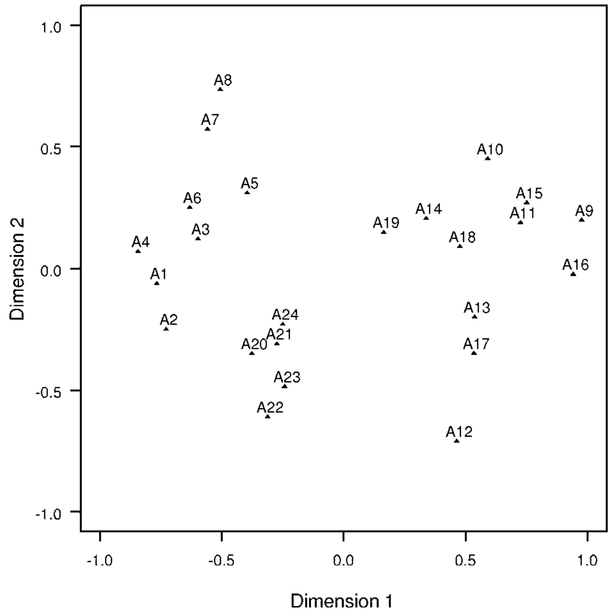

Figure 3.14 *24 Persian archers plotted in the two-dimensional space found through ordinal MDS*

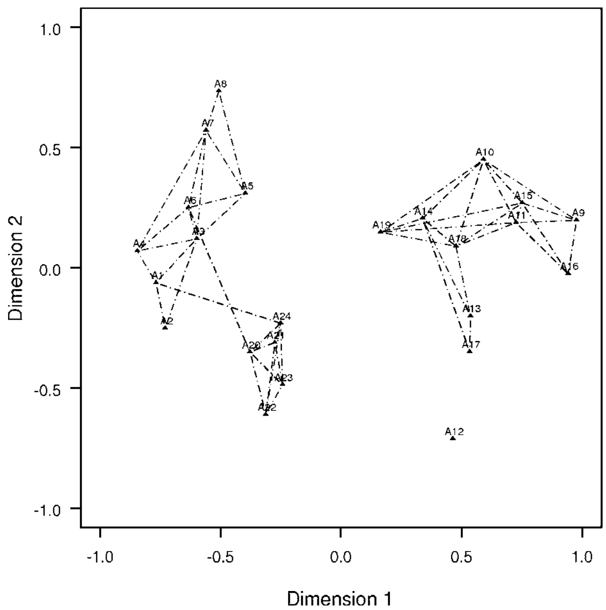

Figure 3.15 *24 Persian archers plotted in the two-dimensional space found through ordinal MDS, with lines drawn between pairs of archers with similarity* $\geq$ *15*

Dialect words in 25 English villages

The data set on which this example is based was given in Section 2.5 where it formed the basis for demonstrating the various techniques of cluster analysis. We showed there that there was a fairly clear cluster structure in which the villages in each cluster were close together geographically as one would have expected. Given that the villages can be represented on a map in two dimensions, it is natural to ask whether one would obtain a similar map if linguistic similarity were used as a measure of distance. We would then be able to see whether the pattern of villages on the linguistic map was similar to their geographical situations. If this turned out to be the case, we would infer that the result of easier interchange between villages close together led to them having more words in common. But major topographical features, like rivers, roads and railways might make for greater similarity along the main lines of communication. There are no mountain barriers in that part of England, but a river like the Trent might well prove a barrier to easy communication.

Bearing these points in mind, you should carry out an ordinal MDS on the similarities in two dimensions. The stress is 0.14 which is not a particularly good fit in two dimensions, but given the particular interest of the two-dimensional plot in this case, it is given in Figure 3.16.

This should be compared with the map in Figure 2.6. The orientation is not the conventional one with north at the top of the diagram. The most northerly village on the map is V4 which occurs on the extreme right of the figure. The orientation will therefore be approximately correct if we rotate the figure anti-clockwise through 90 degrees. In that case, the Huntingdon village (V22) will be on the right, as for the conventional view. Rotating the figure has thus produced something fairly close to the map given in Chapter 2. This is shown in Figure 3.17.

A careful comparison of the "map" provided by your analysis with the true map will show a fairly good, but by no means exact, correspondence. This suggests that geographical factors play a major role in explaining the distribution of dialect words. It must be remembered, of course, that the measure of linguistic similarity we have used is based on a fairly small sample (60) of words.

In view of the relatively poor fit of the two-dimensional map, it is worth looking at the diagnostic plot of d_{ij} versus $\hat{d}_{ij}$. This is given in Figure 3.18. Although the fit is not as good as in some of the other examples, there is a broad correspondence between the d_{ij} and the $\hat{d}_{ij}$.

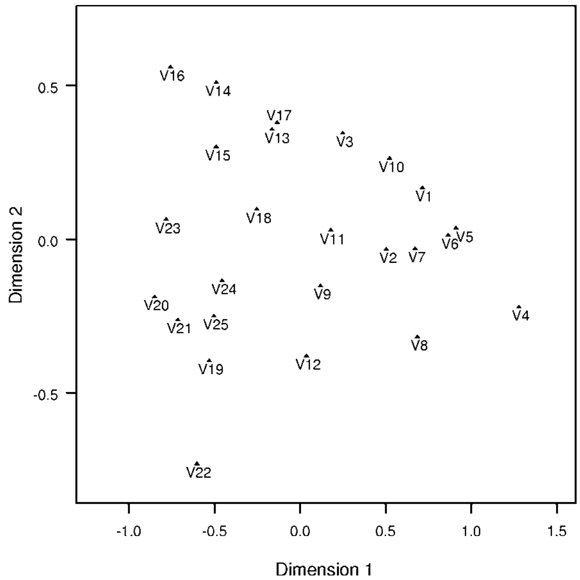

Figure 3.16 *Two-dimensional representation of 25 English villages*

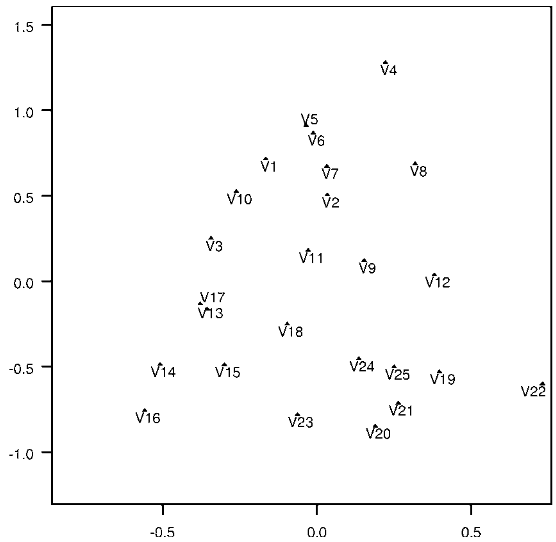

Figure 3.17 *The points on Figure 3.16 in a more conventional orientation*

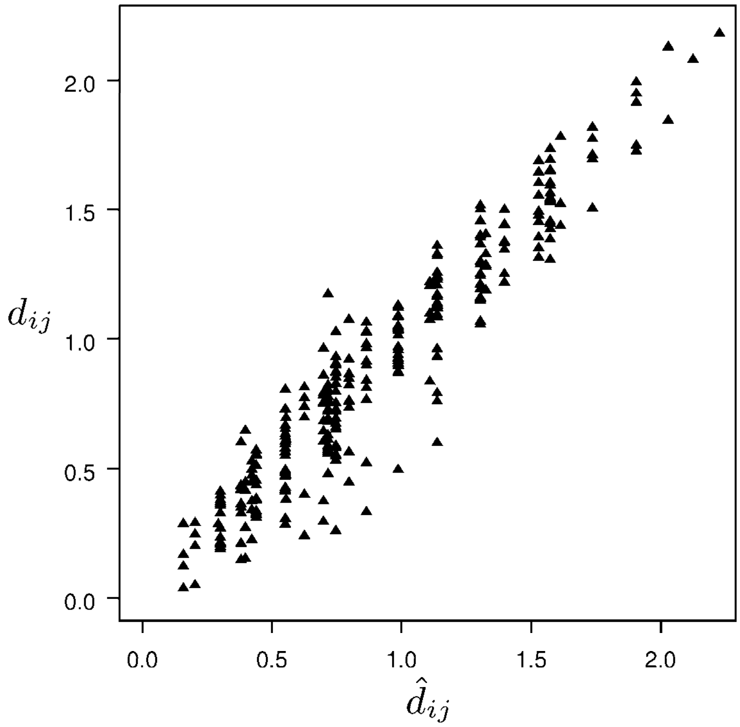

Figure 3.18 *Plot of* d_{ij} *versus* $\hat{d}_{ij}$ *for the dialect data*

Acoustic confusion of letters of the alphabet

In psychological experiments on memory, subjects may be asked to listen to and remember letters of the alphabet in some sequence. There is a risk that they may fail to give the right letters, not because of a failure of memory, but because they did not hear them clearly. Conrad (1964) reports the results of an experiment to investigate acoustic confusion in identifying letters of the alphabet. Three hundred post office employees wrote down the letters they thought they heard when letters were spoken against a background noise at a rate of one every five seconds.

Morgan (1973) calculated the similarities given in Table 3.7 by averaging the number of times the first letter was confused with the second, and the number of times the second was confused with the first, each letter being presented a total of 1440 times. The object of using MDS is to discover what led to letters being confused with each other.

Figure 3.19 shows the minimum values of Kruskal's stress type I for one- through six-dimensional solutions obtained from an ordinal MDS. There is unfortunately no clear elbow, and it is not until you come to the four-dimensional solution that the stress falls below 0.1. A two-dimensional solution will not be adequate, but that does not mean that it will be of no use.

Table 3.7 *Similarities between letters (average of number of times each was confused with the other), acoustic data*

	letter	w	g	c	q	p	t	b	d	e	u	v	h	f
1	w	*	6	6	8	8	10	35	27	18	30	21	18	13
2	g	6	*	41	142	185	128	182	151	242	222	172	5	3
3	c	6	41	*	73	385	274	203	90	129	78	81	22	8
4	q	8	142	73	*	446	265	137	106	118	153	61	32	0
5	p	8	185	385	446	*	786	237	235	283	95	125	27	13
6	t	10	128	274	265	786	*	227	201	287	40	72	32	18
7	b	35	182	203	137	237	227	*	322	379	139	290	18	8
8	d	27	151	90	106	235	201	322	*	418	101	252	17	5
9	e	18	242	129	118	283	287	379	418	*	190	174	53	15
10	u	30	222	78	153	95	40	139	101	190	*	426	28	6
11	v	21	172	81	61	125	72	290	252	174	426	*	18	4
12	h	18	5	22	32	27	32	18	17	53	28	18	*	81
13	f	13	3	8	0	13	18	8	5	15	6	4	81	*
14	s	7	7	20	10	7	15	4	9	23	8	4	194	824
15	x	3	3	11	3	7	7	3	5	25	15	1	191	483
16	l	38	6	2	7	6	2	9	6	11	3	3	16	41
17	j	13	20	16	19	26	14	35	24	10	31	25	23	13
18	k	21	5	11	20	45	19	13	12	16	25	23	43	37
19	m	25	25	18	10	33	15	21	16	72	28	12	18	35
20	n	39	34	26	12	29	20	23	27	112	35	31	55	40
21	a	83	39	11	11	16	20	27	28	26	38	26	104	19
22	o	77	27	5	9	13	40	14	10	27	25	20	50	28
23	i	22	13	8	13	13	15	15	14	114	55	9	4	7
24	r	9	10	3	10	1	10	3	5	19	5	6	8	18
25	y	16	12	12	5	4	9	8	7	12	8	11	4	12
26	z	97	5	8	14	34	10	26	14	10	12	21	53	121

	letter	s	x	l	j	k	m	n	a	o	i	r	y	z
1	w	7	3	38	13	21	25	39	83	77	22	9	16	97
2	g	7	3	6	20	5	25	34	39	27	13	10	12	5
3	c	20	11	2	16	11	18	26	11	5	8	3	12	8
4	q	10	3	7	19	20	10	12	11	9	13	10	5	14
5	p	7	7	6	26	45	33	29	16	13	13	1	4	34
6	t	15	7	2	14	19	15	20	20	40	15	10	9	10
7	b	4	3	9	35	13	21	23	27	14	15	3	8	26
8	d	9	5	6	24	12	16	27	28	10	14	5	7	14
9	e	23	25	11	10	16	72	112	26	27	114	19	12	10
10	u	8	15	3	31	25	28	35	38	25	55	5	8	12
11	v	4	1	3	25	23	12	31	26	20	9	6	11	21
12	h	194	191	16	23	43	18	55	104	50	4	8	4	53
13	f	824	483	41	13	37	35	40	19	28	7	18	12	121
14	s	*	575	60	40	41	44	49	42	44	24	20	15	120
15	x	575	*	13	8	15	11	15	9	7	5	11	6	78
16	l	60	13	*	74	68	115	76	203	101	86	193	123	47
17	j	40	8	74	*	222	46	106	161	87	14	18	118	150
18	k	41	15	68	222	*	82	144	246	101	13	27	31	80
19	m	44	11	115	46	82	*	846	151	339	83	65	69	48
20	n	49	15	76	106	144	846	*	360	89	77	52	58	58
21	a	42	9	203	161	246	151	360	*	594	20	36	28	26
22	o	44	7	101	87	101	339	89	594	*	54	56	22	53
23	i	24	5	86	14	13	83	77	20	54	*	292	164	7
24	r	20	11	193	18	27	65	52	36	56	292	*	194	30
25	y	15	6	123	118	31	69	58	28	22	164	194	*	41
26	z	120	78	47	150	80	48	58	26	53	7	30	41	*

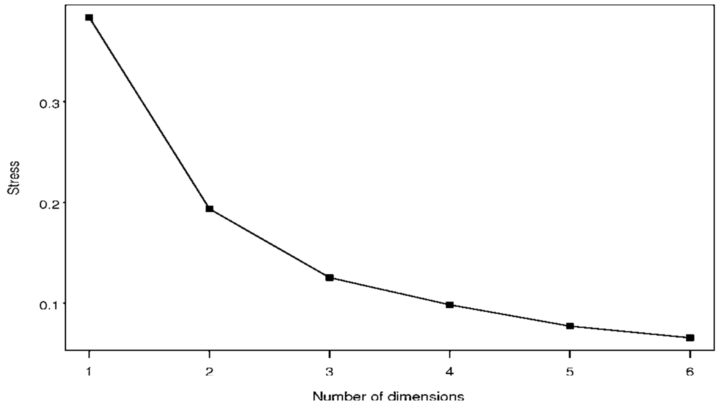

Figure 3.19 *A scree plot of stress against the number of dimensions used for the acoustic data*

The configuration of letters for the two-dimensional solution is shown in Figure 3.20.

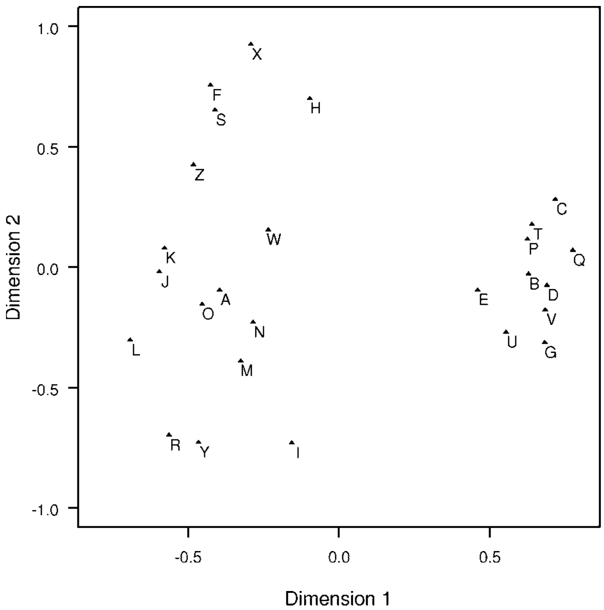

Figure 3.20 *Two-dimensional configuration of acoustic data from an ordinal MDS*

The tightest cluster consists of c, t, p, b, e, d, v, g, and q, u. Referring back to Table 3.7, you can see that these letters including q and u all do have relatively high similarities with each other. You can easily see that most members of this group share the "ee" sound, and it is presumably that fact which leads to them frequently being confused. It is not so obvious why q and u also come in this group, although q and u do have something in common.

This example shows that even when there is a rather poor fit, some meaning can still be extracted from the analysis. You may care to investigate solutions in three or more dimensions to see whether further meaningful groupings occur.

3.8 Further reading

Borg, I. and Groenen, P. J. F. (2005). *Modern Multidimensional Scaling* (2nd ed.). New York: Springer-Verlag.

Cox, T. F. and Cox, M. A. A. (2001). *Multidimensional Scaling* (2nd ed.). London: Chapman and Hall/CRC.

Everitt, B. S. and Rabe-Hesketh, S. (1997). *The Analysis of Proximity Data.* London: Arnold.

Kruskal, J. B. and Wish, M. (1978). *Multidimensional Scaling.* Series Quantitative Applications in the Social Sciences, Number 11. Sage Publications.