Multidimensional scaling and seriation

Seriation: an archeological example

Multidimensional scaling can be used to pick out one-dimensional structure in a dataset. The most common example is seriation, where we want to ascertain the chronological ordering of the data. Note that although multidimensional scaling can be used to order the data in time, the direction of time must be determined independently. Even when the data is one-dimensional, it need not lie on the axis of the first dimension but can sometimes lie on a curve. If our two-dimensional plot of the data shows the items lying along a curve in the plane, then their order along the curve may correspond to their order along a dimension, time for example, that has not been recorded or is unknown. This idea has been used in archaeology, one of the early application areas for multidimensional scaling. In an early archaeological application, a distance matrix for a set of tombs was obtained by recording the presence/absence of several pottery types for each tomb, then distances were obtained using one of the methods described in Chapter 10. On the multidimensional scaling map obtained from this distance matrix, the tombs formed a shallow curve, and their order along this curve was taken to be their temporal order.

Seriation: continuing the alternative medical example

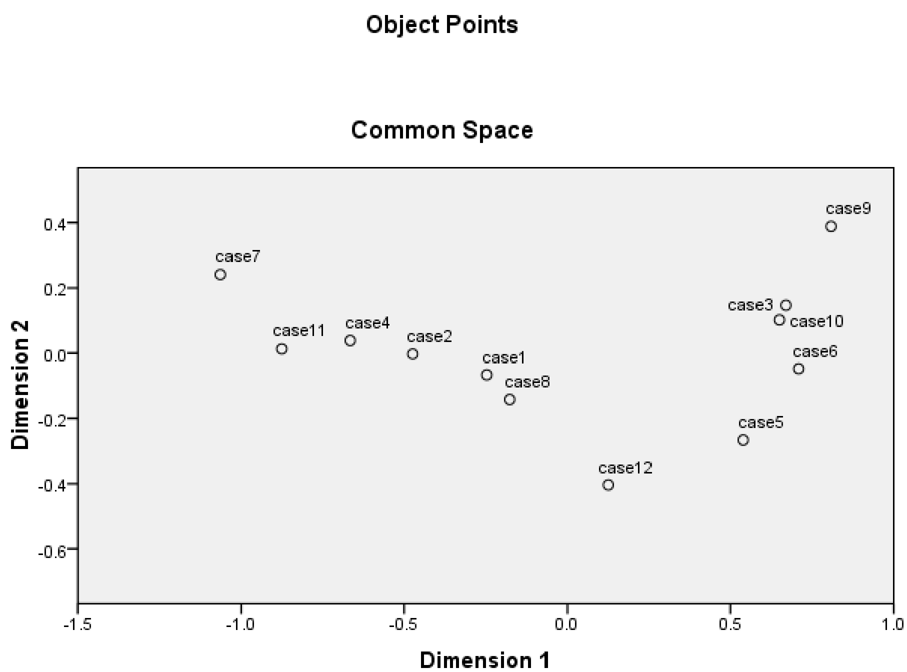
Could this idea have an application in medicine? Well perhaps: imagine this. Suppose we ask the some of our MS patients yes/no questions about how they see their lives in the future (e.g., do you expect to take holidays abroad? will you make the major decisions about your living environment? are better treatments on the way? From the answers, we calculate the simple matching coefficient for each pair of patients (this is discussed in Chapter 10). We could use 1 – the matching coefficient as a distance measure. Table 11.6 shows the (fabricated) results of an experiment like this.

Table 11.6
'Distances' between pairs of patients calculated from 'optimism' questions (mds.distancec.sav)

case	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	0.00	.	.	.	.	.	.	.	.	.	.	.
2	0.13	0.00	.	.	.	.	.	.	.	.	.	.
3	0.41	0.49	0.00	.	.	.	.	.	.	.	.	.
4	0.22	0.10	0.57	0.00	.	.	.	.	.	.	.	.
5	0.36	0.46	0.17	0.57	0.00	.	.	.	.	.	.	.
6	0.44	0.53	0.08	0.54	0.13	0.00	.	.	.	.	.	.
7	0.36	0.29	0.78	0.20	0.74	0.81	0.00	.	.	.	.	.
8	0.04	0.18	0.39	0.26	0.32	0.41	0.40	0.00	.	.	.	.
9	0.52	0.57	0.15	0.65	0.32	0.20	0.86	0.52	0.00	.	.	.
10	0.43	0.49	0.02	0.58	0.19	0.10	0.80	0.40	0.12	0.00	.	.
11	0.29	0.18	0.67	0.09	0.64	0.70	0.12	0.33	0.74	0.67	0.00	.
12	0.22	0.33	0.32	0.43	0.19	0.32	0.57	0.17	0.47	0.35	0.50	0.00

Seriation: understanding output from the medical example

The resulting map is shown in SPSS Output 11.7. The stress values and DAF suggest that a two-dimensional map is a good fit. Let's suppose one of the staff working with the patients suggests that the curve could represent a continuum according to the patients' optimism. If we list the cases in order, case 7, case 11, case 4,…..case 9 we can then consider whether it would be reasonable to see them as arranged in order of their optimism. We may also consider whether the two dimensions of the map reflect different aspects optimism (perhaps optimism about the course of their illness and optimism about their ability to cope).



SPSS Output 11.7. Map of cases from distance matrix in Table 10.7