Wiley StatsRef:
Statistics Reference Online

Shepard Diagram

By *Jan De Leeuw*[1] *and Patrick Mair*[2]

Keywords: *fitting distances, multidimensional scaling, residuals, fit, scatterplot*

Abstract: The Shepard diagram is a version of the residual plot familiar from regression analysis that is useful in nonmetric scaling. We discuss its properties and illustrate it with an example.

In a general nonmetric scaling situation (*see* **Monotonic Regression: Basic**), using the Shepard–Kruskal approach, we have data $y_i, \ldots, y_n$ and a model $f_i(\theta)$ with a number of free parameters θ. Often, this is a nonmetric multidimensional scaling (|MDS) model (*see* **Multidimensional Scaling**), in which the model values are dissimilarities; however, linear models and inner product models can be and have been treated in the same way. We want to choose the parameters in such a way that the rank order of the model approximates the rank order of the data as well as possible.

In order to do this, we construct a loss function (*see* **Loss Function**) of the form

$$\sigma(\theta, \hat{y}) = \sum_{i=1}^{n} w_i(\hat{y}_i - f_i(\theta))^2$$

where the w_i are known weights. We then minimize σ over all $\hat{y}$ that are monotone with the data y and over the parameters θ.

After we have found the minimum, we can make a scatterplot (*see* **Scatterplots**) with the data y on the horizontal axis and the model values f on the vertical axis. This is what we would also do in linear regression (*see* **Linear Regression**) or nonlinear regression analysis (*see* **Nonlinear Regression-Introduction**; **Nonlinear Regression-Extensions**). In nonmetric scaling, however, we also have the $\hat{y}$, which are computed by monotone regression (*see* **Monotone Regression: Theory and Overview**; **Monotonic Regression: Basic**). We can add the $\hat{y}$ to vertical axis and use them to draw the best-fitting monotone step function (see **Step Function**) through the scatterplot. This shows the optimal scaling (*see* **Optimal Scaling**) of the data, in this case the monotone transformation of the data that best fits the fitted model values. The scatterplot with y and f, and $\hat{y}$ drawn in, is called the *Shepard diagram*.

Shepard diagrams are most widely used within the context of MDS where we minimize the *stress* loss function. We obtain a fitted dissimilarity matrix based on the configurations in a low-dimensional space. We can plot the observed dissimilarities against the fitted distances, which gives us the scatterplot. The regression function fitted into this scatterplot shows how the dissimilarities and the approximated distances are related to each other. In nonmetric MDS, this function is a monotone step function; in metric MDS a line.

[1]University of California, Los Angeles, CA, USA
[2]Harvard University, Cambridge, MA, USA

Update based on original article by Jan de Leeuw, Wiley StatsRef: Statistics Reference Online, © 2014, John Wiley & Sons, Ltd

DOI: 10.1002/9781118445112.stat06268.pub2

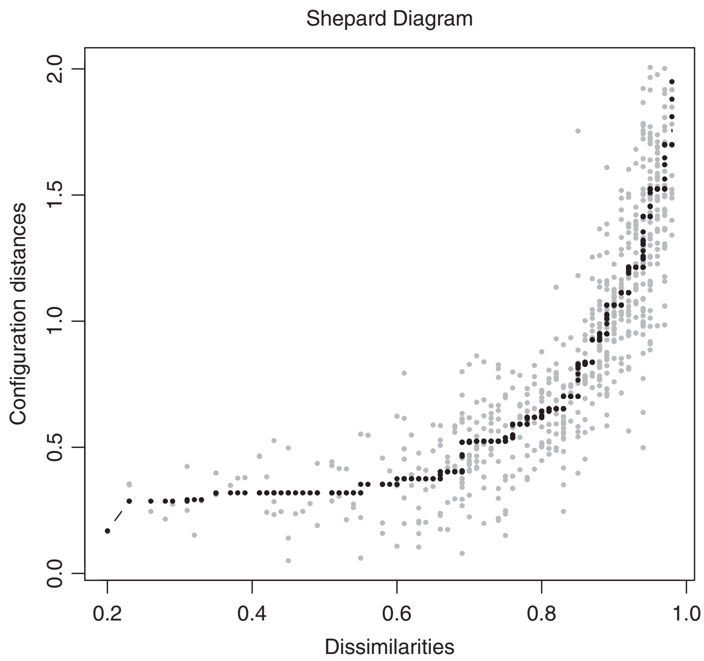

Figure 1. Shepard diagram Morse code data for nonmetric MDS.

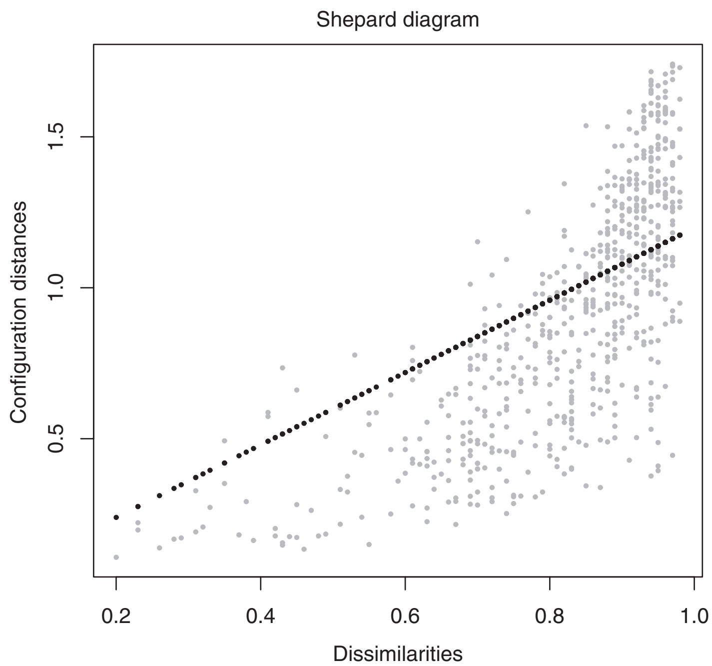

Figure 2. Shepard diagram Morse code data for metric MDS.

DOI: 10.1002/9781118445112.stat06268.pub2

The Shepard diagram gives a detailed insight into the fit structure[1]: The vertical line of each point from the regression function gives the residual of this particular point. This gives us an overall picture of the scatter around the regression function including the possibility to detect outliers. We can determine that individual dissimilarities are not well represented by the current MDS solution and may, for instance, bring in another dimension. This concept of looking at single residuals individually can be aggregated by averaging all the squared residuals between a single object and all the other ones. This is called *stress per point*.

We show an example from a nonmetric and metric MDS of the classical Rothkopf Morse code confusion data[2]. The stimuli are 36 Morse code signals. The scores are derived from confusion rates on 36 Morse code signals (26 for the alphabet; 10 for the numbers 0, … , 9). Each Morse code signal is a sequence of up to five "beeps." The beeps can be short (0.05 s) or long (0.15 s), and, when there are two or more beeps in a signal, they are separated by periods of silence (0.05 s). Rothkopf asked 598 subjects to judge whether two signals, presented acoustically one after another, were the same or not. The values are the average percentages with which the answer "Same!" was given in each combination of row stimulus i and column stimulus j, where either i or j was the first signal presented. The values are 1 minus the symmetrized confusion rates and are therefore dissimilarities.

In Figure 1, we show the Shepard diagram for the nonmetric MDS solution of the Morse code data.

The Shepard diagram in Figure 2 is based on a metric (ratio) MDS.

References

[1] Borg, I. and Groenen, P.J.F. (2005) *Modern Multidimensional Scaling: Theory and Applications*, 2nd edn, Springer-Verlag.

[2] Rothkopf, E.Z. (1957) A measure of stimulus similarity and errors in some paired associate learning. *J. Exp. Psychol.*, **53**, 94–101.

DOI: 10.1002/9781118445112.stat06268.pub2