

Dissimilarities and transformations

Quiz

The goal of quizzes is to help you learn.

Compare your answers to the list of correct answers at the end of the quiz.

Dissimilarities – Generalities

1. In the Q mode of analysis, one first computes an association coefficient (which is a S or D coefficient) among the objects. In the R language, the objects are the rows of the data matrix. – True, False.
2. In the R mode of analysis, one first computes an association coefficient (e.g. a correlation coefficient) among the variables. In the R language, the variables (or descriptors) are the columns of the data matrix. – True, False.
3. A similarity S and a dissimilarity D matrix have zeros on the main diagonal. – True, False.
4. Functions of the R language that use Q-mode coefficients, e.g. for clustering or ordination, require dissimilarity (D) matrices, not similarity (S) matrices. – True, False.

Dissimilarities – Metric and Euclidean properties

5. A dissimilarity coefficient is metric if it meets the four metric properties, including the triangular inequality for all possible triplets of points in the D matrix. – True, False.
6. A dissimilarity coefficient is semimetric if it violates the triangular inequality for all possible triplets of points in the D matrix. – True, False.
7. A D coefficient is Euclidean if it always produces D matrices that can be fully represented in Euclidean space without distortion. – True, False.
8. A non-Euclidean dissimilarity matrix is identified by the criterion that principal coordinate analysis (PCoA) of that matrix produces some negative eigenvalues. – True, False.

Dissimilarities – Converting S to D

9. The following conversion can be used: $D = (1 - S)$ and $D = \sqrt{1 - S}$. There is no reason to choose one or the other transformation. – True, False.
10. In most cases, \sqrt{D} or $\sqrt{1 - S}$ turns a non-Euclidean matrix to Euclidean. – True, False.

Dissimilarities – The double zero problem

11. Ecologists prefer to remove double zeros from the calculation of (dis)similarity coefficients because double zeros have no clear, unambiguous ecological interpretation. – True, False.

Dissimilarities – Coefficients for binary data

12. In double-zero *symmetrical* coefficients, like the simple matching coefficient, double zeros affect the S or D value. – True, False.

13. In double-zero *asymmetrical* coefficients, like the Jaccard and Sørensen coefficients, double zeros are not included in the formulas and do not affect the *S* or *D* values. – True, False.

14. Double-zero *asymmetrical* coefficients are well suited to the analysis of community composition or other types of frequency data, e.g. gene frequencies. – True, False.

Dissimilarities – Coefficients for quantitative physical descriptors

15. The Euclidean distance computed on physical data that have different physical units does not make sense because the distances have un-interpretable physical dimensions. – True, False.

16. For descriptors with different physical units, the Euclidean distance computed on standardized descriptors makes sense; the distances then have no physical units. – True, False.

17. The Gower coefficient was designed to handle descriptors with different physical units and of mixed precision levels. – True, False.

18. Some R functions propose forms of the Gower coefficient that can handle missing values and give different weights to the descriptors. – True, False.

Dissimilarities – Coefficients for quantitative community composition data

19. The asymmetrical Ružička and percentage difference (aka Bray-Curtis) dissimilarities are Euclidean indices. – True, False.

20. In the Ružička and percentage difference dissimilarities, the data are not scaled by rows. The differences in productivity of the sites are taken into account in the calculation of the dissimilarities. – True, False.

21. The distance between species profiles and the Hellinger, chord, and chi-square distances are Euclidean indices. – True, False.

22. In these indices, the scaling by rows removes the differences in productivity of the sites from the data. They are not taken into account in the calculation of the dissimilarities. – True, False.

23. These four indices are insensitive to double zeros. Thus they are asymmetrical indices. – True, False.

24. The Jaccard and Sørensen coefficients are, respectively, the binary forms of the Ružička and percentage difference; the Ochiai coefficient is the binary form of the Hellinger and chord distances. – True, False.

Transformations for community composition data

25. The symmetrical indices: distance between species profiles, Hellinger, chord and chi-square distances can be computed in two steps: transform the data in a way that makes the dissimilarity insensitive to double zeros, then compute the Euclidean distance. – True, False.

26. These transformations and corresponding distances are equivalent and interchangeable. They produce the same PCoA ordinations of the sites. – True, False.

27. Without calculation of the Euclidean distance, the transformed data matrices can be used as input into linear methods of data analysis that preserve the Euclidean distance, such as PCA, RDA and *k*-means partitioning. – True, False.

Correct answers to the quiz about Dissimilarities and transformations –

1. True

2. True

3. False

4. True

5. True

6. False

7. True

8. True

9. False

10. True

11. True

12. True

13. True

14. True

15. True

16. True

17. True

18. True

19. False

20. True

21. True

22. True

23. True

24. True

25. True

26. False

27. True