STA254

CORRESPONDENCE ANALYSIS AND RELATED METHODS

Michael Greenacre

Department of Statistics, Stanford University, fall 2008

Week 2: Homework exercises (to be handed in Tuesday Oct. 7)

Slides, supporting material and R scripts can be found at:

> http://www.econ.upf.edu/~michael/stanford

(Refer to Homework for week 1 for details about reading data into R).

1. Compute the chi-square distances between the 30 sites for the five species "abcde" data set (taken from the Excel file on the website). You can calculate the distances any way you like (for example, you can use the provided function chidist).

2. Perform hierarchical scaling on the distance matrix, using either complete linkage or Ward clustering (your choice).

3. Perform classical multidimensional scaling on the distance matrix.

4. Hand in your R code and results on paper.

An R challenge (optional – high difficulty for R beginners):

(Refer to Week 1's R scripts). You will see that in Computing Step 4, as well as in the provided function braycurtis, the Bray-Curtis dissimilarity is calculated in a double 'for' loop, calculating each element of the dissimilarity matrix one at a time, whereas the chi-square distance is calculated using matrix–vector multiplications, with no sign of looping over the row and column indices. Using 'for' loops is, in fact, very inefficient in R, so we do our best to avoid them. The Bray-Curtis index can be calculated without using loops. Can you work out a way to do this?