

STA254

CORRESPONDENCE ANALYSIS AND RELATED METHODS

Michael Greenacre

Department of Statistics, Stanford University, fall 2008

Week 9: Homework exercises (to be handed in December 3)

Using the `ca` package in R, load the data set `wg93`:

```
data(wg93)
```

1. Look at a part of the dataset to familiarize yourself with its contents: there are four question responses (same as the four questions on environment & science) and three additional demographic variables, sex, age and education.
2. Do a multiple correspondence analysis (indicator matrix version) on the question responses (first four columns only), using function `mjca`.
3. Plot the results and comment on the configuration and the eigenvalues.
4. Repeat the analysis on the first three questions only, omitting question D, and comment on how the results change.

Note: there seems to be a bug in the `plot()` function when one uses the `supcol=` option in `mjca()`. You can get do an MCA with the three demographic variables as supplementary by analyzing the Burt matrix, however, which is one of the results of MCA.

So try this (optional extra):

5. Store the Burt matrix for all 7 variables, e.g. `Burt<-mjca(wg93)$Burt`
6. Check out the contents of `Burt` and then do the simple CA (using `ca()` function) on the submatrix of the Burt matrix that contains the questions crosstabulated with the questions and the demographics, declaring the demographics as supplementary.

I have put Chapter 20 of my book on the website to assist you with the scaling aspects of MCA.