STA254

CORRESPONDENCE ANALYSIS AND RELATED METHODS

Michael Greenacre

Department of Statistics, Stanford University, fall 2008

Week 3: Homework exercises (not to be handed in)

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. Read in the data EU from the Excel spreadsheet EU.xls.
2. Perform the part of the principal component analysis (PCA) given in your course notes (see R script for week 3), leading to a plot of the 12 countries.
3. Revise the section on regression biplots and calculate the five biplot vectors for the column variables of the EU data, and add them to the above plot.

Take a look at:

http://www.snl.salk.edu/~shlens/pub/notes/pca.pdf

Notice that most authors weight each case by $1/(n-1)$, not $1/n$, because of the definition of the sample variance as

$$\frac{1}{n-1}\sum_i (x_i - \bar{x})^2$$

(this is the unbiased estimate of the variance which you need in statistical hypothesis testing). I prefer the weighting of $1/n$ for both the variance calculation and the SVD.