Course 27411 Exercises in correspondence analysis (CA), 12/4 2010 (using **NTSYS**)

First: Install NTSYS, which consists of an editor, the program itself, and a help function from J:\install\chemistry

Secondly: In My Computer - Control Panel - Regional and Language Settings, set the decimal setting to the English (UK) system, as NTSYS may occasionally have problems with numbers such as 1,236 (IT SHOULD BE 1.236), as the comma indicates a blank in some cases (and so division by 0 may occur, giving severe problems)

In NTSYS the columns are regarded as the objects and the row variables (opposite UNSCRAMBLER)

In NTSYS data files **should** start with a "commando line": x - y – z - w, where x is the type of matrix (1 for a rectangular matrix), y is number of rows, z is number of columns and the last number w is 0 if there are no missing values. Data can be imported from a text file or excel (often excel is the best). If you use excel import data : "Open file in grid" (else open file as text) in NTEDIT and safe as a NTS file. Then you can go to the real program NTSYS and today we will use Ordination: Correspondence analysis and fill in input file, and names for the output files: column and row matrix and remember to tick of "show details". The first report you see is the eigenvalues and % variation explained by each CA axis + relative frequency and distance for the center of the plot (deviations from the expected) for all rows and columns.

If you use text files, you will add an L after the column number: and then in the next row the names of the objects (the L will not be used if you import from an excel file):

Example:

1 4 10L 0

A B C D E F G H I J

8 7 9 13 6 12 9 7 11 6

5 6 3 3 7 10 5 7 5 7

11 13 12 9 8 18 17 21 22 13

10 18 21 9 7 17 18 26 24 18

Exercise 1:

Use the questionnaire results from Lebart et al. (1984) to see the different applications of correspondence analysis. The rows are 23 occupations and columns represent 15 advantages of these occupations according to the questionnaire (data file Lebart.txt in ascii text format)

1. How much of the variation is explained by the first axes? How many axes are needed to explain most of the variance in the data (80%, 90%)?
2. Which row and column factors are most deviating from the expected?
3. Make a biplot in graphics, with both objects and variables. Which occupations are related to which advantages? (to see names on biplots, use Options – Plot Options and add labels to both object and variable points)

Exercise 2 (form examn):

Use correspondence analysis to analyze the EU data (EU.txt) from EU Government conference 1996

1. How much variance is explained by the first three CA axes? How many CA axes are necessary to explain the most important information in the data?
2. How many axes can maximally be extracted from these data? (Please note that the "first" CA axis is always of eigenvalue one (1), as you loose one degree of freedom as you make column and row summations, so this is completely disregarded)
3. Which countries are most "extreme"?
4. Are there any groupings of the countries?
5. Are the any correspondence between certain countries and certain
6. If time permits please analyze the same data in PCA. What are the differences between CA and PCA?

The Government conference on EU i 1996

Row factors (Variables):

1. INT: The fundamental opinion on more integration in EU (General)
2. EXP: The fundamental opinion on expansion of EU (Institutions)
3. VOT: New placement of votes in the ministerial council (Institutions)
4. RUL: New rules on the chairmanship of EU (Institutions)
5. POW: More power to the European Commission (Institutions)
6. STR: Enforcement of the European Parliament (Institutions)
7. BES: Enhanced use of the consiliarity principle (for EU Parliament) (Institutions)
8. SUB: Enhancement of the subsidiarity principle? (Institutions)

9. BUD: Treatment of budget problems at the government conference? (finances)
10. HOU: More power to the European parliament in the “household” budget=? (finances)
11. MAJ: Enhanced use of majority decisions? (First column)
12. OMU: European Monitary Union (ØMU) at the government conference? (First column)
13. COM: New treaty-based competences to EU (new areas)? (First column)
14. ENV: Fortification of the environment and the social dimension? (First column)
15. FOR: Gradual movement to majority decisions in foreign policy? (Second column)
16. MRX: Election of a mister X to represent EU in foreign policy? (Second column)
17. FIN: Financing of the common foreign policy over the EU budget? (Second column)
18. WEU: Merging of EU and the Western Union? (Defence)
19. OVE: Transfer of between-state to the “over-state” collaboration? (Third column)

FEW: Fewer EU commisarians? (Institutions) has not been included, but can be included if you wish so!