Biological data analysis and chemometrics (27411)

Cluster analysis day

1. Make the cluster analysis on paper (UPGMA, single linkage, complete linkage)

2. Make the same analysis using the program **NTSYS** on PC

3. Binary datamatrix (sm.txt). Perform cluster analyses on the objects using the binary (qualitative) distances: Simple matching (SM), Jaccard (J) and Yule (Y).
For each of those try the UPGMA, Single linkage and complete linkage. Compare the cophenetic correlation coefficient of all 9 analyses

(The SM datafile contains 5 species, 51 isolates (objects) in all, *Penicillium brevicompactum*, 1-6, *P. crustosum*, 7-16, *P. echinulatum*, 17-15, *P. discolor*, 26-38 and *P. expansum*, 39-51) . These fungal cultures have been examined in the microscope (variable 1-11), and further extracted with organic solvents and analyzed using HPLC with diode array detection (the secondary metabolites detected are variable 12-126)

For examn:
4a. Quantitative datamatrix (vin.txt). Remember to transpose the matrix, because **NTSYS** regard columns as objects and rows as variables unlike the other program we use: **UNSCRAMBLER**. Perform cluster analysis using interval (quantitative) data with the distance coefficient Euclidean distance. Compare the results when using raw data and standardized data. Compare the results using the standardized data with those using CORR coefficient (Product-moment correlation). Calculate the cophenetic correlation for all three.

4b. Quantitative datamatrix (vin.txt). Work on the untransposed matrix to see whether any of the variables are correlated (using CORR), i.e. cluster the variables using the objects as features. Which variables are most clearly correlated? Will COSINE give a different result?

.