

## Multivariate Data Analysis: Tasks for Week11

(Released 08/12/11, see §9.0–§9.5 & revision of §4.4–§4.7)

[Note that these questions are more substantial than on previous task sheets. Question 1 is a past examination question. Question 3 is only of benefit to those wanting more practice on PCA interpretation and practical data analysis]

- 1) An archaeologist wishes to distinguish pottery from two different sources on the basis of its chemical composition. Measurements by Neutron Activation Analysis of the concentrations in parts per million of trace elements Cr and V in 19 samples of pottery from Tell el-Amarna gave mean results of 2.3 and 6.7, respectively, with sample variances 0.62 and 1.41 and covariance 0.09. Similar measurements on 23 samples from Memphis gave mean results of 2.9 and 5.9 with sample variances 0.7 and 1.36 and sample covariance 0.08.
  - i) Assuming that these measurements are adequately modelled by bivariate Normal distributions with a common variance, calculate the linear discriminant rule for distinguishing Amarna from Memphis pottery on the basis of the concentrations of Cr and V.
  - ii) Prove that the estimated probabilities of misclassifying Memphis pottery as Amarna and *vice versa* are the same using this rule.
  - iii) By how much is this misclassification probability an improvement over those using each of the elements separately?
  - iv) What advice would you give to the archaeologist in the light of these results?



- 2) Referring to the data set *dogmandibles.\** (including the Prehistoric Thai dogs (group 5 on  $X_{11}$ ))
- i) Using STAT>MULTIVARIATE>DISCRIMINANT in Minitab or lda() in S-PLUS look at the discrimination between the 5 species (using the nine measurements) and estimate the classification rate. [In S-PLUS it is easy to find the cross-validation (or jackknife) estimate of classification rate using the CV=T option].
  - ii) Perform the discriminant analysis just on the first four [modern] species and then use this to classify the prehistoric Thai dogs.
  - iii) Compare the results of these analyses with the results of the more informal exploratory analyses with Crimcoords in Exercises 2.
- 3) The datafile CLAYPOTS has 272 observations on the trace element content of clay samples from pots found at various archaeological sites around the Aegean. Column 1 gives the group number (i.e. archaeological site for most of the pots) and columns 2–9 give the amounts of 9 trace elements (which have been labelled A to I) found in samples of clay from the pots. It is suggested that before investigating the specific questions below it is advisable to do some exploratory analysis with PCA etc. Groups 1, 3 and 4 are from known sources; groups 2 and 5 are from unknown sources but are believed to come from one or other of 1,3 or 4.
- i) Construct a display on crimcoords of groups 1,3 and 4 and add in the points from groups 2 and 5.
  - ii) Which are the best classifications of these pots?

