

Multivariate Data Analysis

Exercises 2

Released 31 October 2011, work submitted by ~Tuesday 15 November 2011 will be marked and returned. Work submitted after solutions are made available will not be marked

1) The data given in file *dogmandibles.** (in various formats) are extracted, via Manly (1994), from Higham etc (1980), *J.Arch.Sci*, 149–165. The file contains 9 measurements of various dimensions of the mandibles of 5 canine species as well as records of the sex and the species, eleven variables in total. These are

X_1 : length of mandible

X_2 : breadth of mandible

X_3 : breadth of articular condyle

X_4 : height of mandible below first molar

X_5 : length of 1st molar

X_6 : breadth of 1st molar

X_7 : length between 1st to 3rd molar inclusive (1st to 2nd for Cuons)

X_8 : length between 1st to 4th premolar inclusive

X_9 : breadth of lower canine

X_{10} : gender (1 \equiv male, 2 \equiv female, 3 \equiv unknown)

X_{11} : species (1 \equiv modern dog from Thailand, 2 \equiv Golden Jackal,
3 \equiv Cuon, 4 \equiv Indian Wolf, 5 \equiv Prehistoric Thai dog)

All measurements are in mm; molars, premolars and canines are types of teeth; an articular condyle is the round knobby bit in a joint; a Cuon, or Red Dog, is a wild dog indigenous to south east Asia and notable for lacking one pair of molars.

- i) Ignoring the group structure, what interpretations can be given to the first two principal components?
- ii) Construct a display of the measurements on the first two crimcoords, using different symbols for the five different groups.



- iii) If the linear discriminant analysis were performed on the data after transformation to the full set of nine principal components what differences (if any) would there be in the plot on crimcoords and the eigenvalues and eigenvectors of the matrix $W^{-1}B$?
- iv) Which group is separated from the other four by the first crimcoord?
- v) Which group is separated from the other four by the second crimcoord?
- vi) Which group is separated from the other four by the third crimcoord?
- vii) What features of the mandibles provide discrimination between the various species?

Notes: if you do this in Minitab then you need to use the slicktrick outlined on p135 of the course notes and use the balanced manova facility within anova to obtain the eigenvectors and then cut&paste these into the worksheet, followed by a couple of steps of matrix manipulation. In S-PLUS the required plots can be obtained from the lda menu. Interpretation of the crimcoords is a little easier after looking at the interpretations of the PCs.

- 2) * The question of prime interest in the study of canines was related to an investigation of the origin of the prehistoric dogs. Try calculating the discriminant analysis based on the four groups of modern canines and then plot the prehistoric cases on the same coordinate system a (c.f. informal data classification method (iii) on p140 of course notes) and seeing to which of the modern groups the majority of the prehistoric are closest.

(The interpretation of the results of this exercise are **within** the scope of PAS465; the required computer skills to produce it are useful but a little beyond the scope of PAS465, i.e. if you do not attempt it ensure that you look carefully at the printed solution in due course.)

