**Statistics 850** **yandell** **Spring 1995**

## Homework # 6—Due Mon 3 Apr

Scientists who study social insects, such as wasps and ants, have long noted that queens and workers can have very different sizes and shapes. For some species, the queens are simply larger, suggesting that they continue to grow with the same basic "plan" as workers. They are just fed for longer. An entomologist on campus examined a wasp species which does not follow this pattern. Your job is to provide statistical evidence to support the claim that queens are not simply larger workers – they likely differed in development at a very early stage. The data consist of measurements on 50 workers and 50 queens for `p=13` multiple responses. The measurements have names which code where they were taken – head (`H`), thorax (`T`) or gonadium (`G`) – and what kind of measurement – width (`W`), height (`H`) and length (`L`). The gonadium has two length and 3 height measurements.

1. The first task is to reduce the "dimensionality" of the data.
(a) Overall Test: Conduct univariate analyses on all responses to assess worker-queen differences. Conduct a multivariate ANOVA for same. Report test results.
(b) Use a stepwise discriminant analysis procedure (e.g. stepdisc or ANCOVA) to reduce to a subset of significant responses as discriminators. Briefly summarize these results.
(c) Head width (`HW`) and gonadium width (e.g. `G1Wa`) appear to be the most significant discriminators, as found in part b. Plot them against each other to note that one is larger while the other is smaller in queens than workers. Use analyses so far, along with correlations and partial correlations (adjusting for group) among the multiple responses to argue that these two responses appear to capture the main features.
(d) Conduct a canonical discriminant analysis to examine the "best" linear combination of responses to discriminate between workers and queens. How much better does it do than any one response? What is the canoncial variate correlation (see Total Canonical Structure in SAS output) with the two responses considered in (c)? Why is one positive and one negative?
(e) Plot the canonical variate against the two responses in (c). Briefly interpret in light of (d) above.

2. The scientist thought it might be important to first remove "size" in order to isolate "shape" differences between workers and queens. He suggested using `TL`, as it is the poorest discriminator – that is, the distributions of `TL` substantially overlap for queens and workers.
(a) Conduct an analysis of covariance of the two responses `HW` and `G1Wa` on size (`TL`) and group. Are there group differences after adjusting for size?
(b) Remove the effect of size and conduct a stepwise DA. Compare results with 1(b).
(c) Conduct canonical DA for the residuals with size removed. How do the canoncial variate correlations compare with 1(d)? (d) Plot canonical variate and the residuals for the two responses from 1(c) against each other. Briefly compare and contrast with earlier plots.

The data can be found in `hwk6.dat`, with ideas on analysis shown in `hwk6.sas` and `hwk6.prt`. [It is much easier with current tools to use SAS for this problem.]