

**ĐỀ THI THỰC HÀNH:  
THỐNG KÊ NHIỀU CHIỀU**

Ngày 18 tháng 6 năm 2013

### **SINH VIÊN CHÚ Ý:**

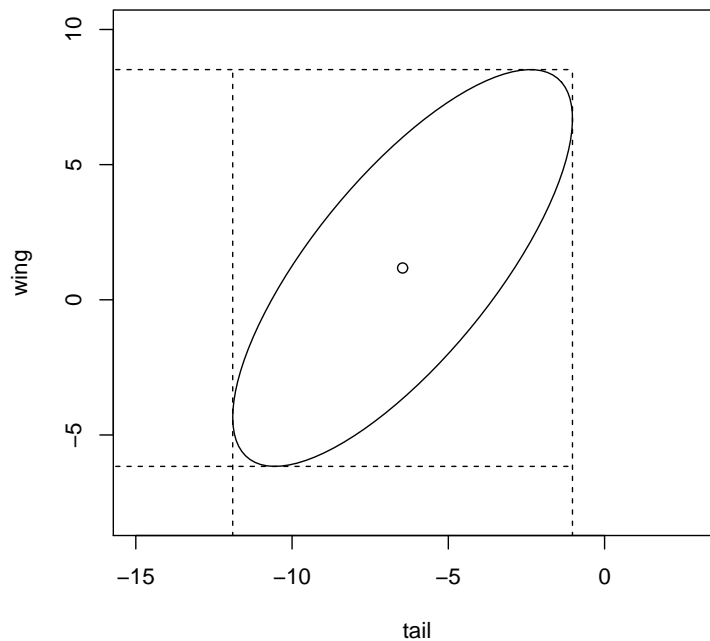
- Làm bài theo nhóm bài tập đã chọn.
- Bài nộp gồm 2 file: file code **R** và file word trình bày các nhận xét về các ý do đề bài yêu cầu. File word chỉ nêu các nhận xét và các kết quả dùng cho nhận xét, không cần đánh tất cả công thức hay kết quả chạy trong **R** vào.
- Đặt tên file theo tên các thành viên trong nhóm, ví dụ nhóm của Nguyễn Văn Xanh - Nguyễn Thị Hoài Thương - Tô Quang Vinh đặt tên các file nộp như sau: **Thuong-Vinh-Xanh.R** và **Thuong-Vinh-Xanh.doc**.
- Các nhóm có thể gửi mail để hỏi về đề thi (nếu chưa rõ ý), cách trình bày và nhận xét.
- Nộp bài qua email: [hvha@hcmus.edu.vn](mailto:hvha@hcmus.edu.vn). Tiêu đề ghi: "Nop bai thi thuc hanh thong ke nhieu chieu (...)" (... là tên của các bạn trong nhóm làm bài).

**Hạn nộp: 23g59 ngày 06/07/2013**

**Bài 1.** Khảo sát dữ liệu về chiều dài đuôi  $X_1$  (mm) và cánh  $X_2$  (mm) ở loài chim ưng (hook-billed kite). File `bird-males.txt` và `bird-females.txt` chứa số liệu về chiều dài đuôi và cánh của 45 chim ưng trống và 45 chim ưng mái.

Mô tả:

- *tail.m*: chiều dài đuôi của chim ưng trống.
  - *wing.m*: chiều dài cánh của chim ưng trống.
  - *tail.f*: chiều dài đuôi của chim ưng mái.
  - *wing.f*: chiều dài cánh của chim ưng mái.
- (a) Vẽ đồ thị phân tán của  $(X_1, X_2)$  ứng với chim ưng trống và mái trên cùng 1 đồ thị. Dựa vào đồ thị, hãy kiểm tra liệu có điểm outlier trong dữ liệu của chim ưng trống hay không? Nếu có hãy loại bỏ điểm outlier (Chú ý là 1 điểm trong không gian  $p$ -chiều có dạng  $(x_1, \dots, x_p)$ ). Vẽ lại đồ thị phân tán sau khi bỏ điểm outlier.
- (b) Với mức ý nghĩa 5%, kiểm định giả thuyết  $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ , trong đó  $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$  lần lượt là véc-tơ trung bình mẫu về chiều dài đuôi và cánh của chim trống và chim mái,  $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2})$ . Yêu cầu: xuất các véc-tơ trung bình mẫu, các ma trận hiệp phương sai mẫu, thống kê  $T^2$ , thống kê  $F$  và  $p$ -giá trị ứng với thống kê  $F$ .
- (c) Xác định miền tin cậy 95% cho  $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$  và các khoảng tin cậy đồng thời cho các thành phần của  $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ . Yêu cầu: vẽ miền tin cậy 95% (vẽ ellip) và các khoảng tin cậy đồng thời cho  $\mu_{i1}, \mu_{i2}$  trên cùng đồ thị (như hình bên dưới), xuất ra véc-tơ chứa các khoảng tin cậy đồng thời.



- (d) Chim trống hay chim mái lớn hơn?

**Bài 2.** Bài tập **8.28** sách Johnson & Wichern (trang 479-480). Dữ liệu lấy trong file `mali-family-farm.txt`. Đối với câu (b), yêu cầu xuất đầy đủ các thông tin sau:

- Véc-tơ trung bình mẫu và độ lệch chuẩn (tạo thành bảng).
- Ma trận hiệp phương sai mẫu và ma trận hệ số tương quan mẫu.
- Các trị riêng và các véc-tơ riêng tương ứng.
- Bảng kết quả tổng hợp của phân tích thành phần chính.
- Bảng mô tả hệ số tương quan giữa thành phần chính thứ 1, thứ 2 và các biến  $X_1, \dots, X_9$ .
- Đồ thị screeplot.