

PHÂN TÍCH THỐNG KÊ NHIỀU CHIỀU MAT3452

Bài tập tuần 2: Phân phối chuẩn một chiều

Câu 1. Tập dữ liệu **hills** trong thư viện MASS gồm thời gian kỷ lục của 35 cuộc đua trên các ngọn đồi ở Scotland vào năm 1984. Với mỗi cuộc đua trên từng ngọn đồi, bộ dữ liệu cung cấp độ dài đường đua (dặm), độ cao đạt được (feet) và thời gian chạy của người dành chiến thắng (phút).

- (i) Tìm giá trị trung bình và độ lệch tiêu chuẩn của độ dài các đường đua, độ cao đạt được và thời gian chạy của người chiến thắng.
- (ii) Các biến trong bộ dữ liệu có tuân theo phân phối chuẩn không?
- (iii) Dùng hàm pairs để vẽ ma trận biểu đồ tán xạ cho bộ dữ liệu. Nêu nhận xét, sát và thời gian dự đoán theo mô hình ở ý (iv). Bình phương hiệu số đó có phân phối chuẩn không?

Câu 2. Tập dữ liệu **iris** trong thư viện datasets gồm một số đặc điểm của các bông hoa diên vĩ. Với mỗi một bông hoa, bộ dữ liệu cung cấp độ dài, độ rộng của cánh hoa và đài hoa (cm) và loài.

- (i) Tìm giá trị trung bình và độ lệch tiêu chuẩn của độ dài, độ rộng cánh hoa và đài hoa .
- (ii) Các biến trong bộ dữ liệu có tuân theo phân phối chuẩn không?
- (iii) Vẽ đồ thị hàm mật độ và hàm phân phối của các biến có phân phối chuẩn trên cùng một khung hình. Đặt tên đồ thị, tên và độ dài các trục phù hợp.
- (iv) Với mức ý nghĩa 5%, có thể nói độ rộng trung bình của đài hoa nhỏ hơn 3.5 không?
- (v) Tìm a , biết xác suất để độ rộng đài hoa nhỏ hơn a là 70%.
- (vi) Độ rộng trung bình của đài hoa của loài setosa và versicolor có thực sự khác nhau không? Mức ý nghĩa 1%.

Câu 3. Sinh 30 mẫu ngẫu nhiên, mỗi mẫu gồm 20 quan sát có phân phối chuẩn với kỳ vọng 158 và phương sai 8.

- (i) Tìm khoảng tin cậy 98% cho giá trị trung bình của các mẫu ngẫu nhiên đó.
- (ii) Mẫu thứ bao nhiêu có khoảng tin cậy rộng nhất/hẹp nhất?
- (iii) Tìm mẫu thứ i biết giá trị trung bình của mẫu này có giá trị lớn nhất. In ra màn hình "Mẫu thứ i có giá trị trung bình lớn nhất".
- (iv) Vẽ các khoảng tin cậy ở ý (i) thỏa mãn các khoảng đó chứa giá trị trung bình của mẫu gồm toàn bộ các quan sát trong 30 mẫu ngẫu nhiên được sinh ban đầu. Ký hiệu hai đầu mút mỗi khoảng bằng chấm tròn đỏ, độ dài khoảng có màu xanh lá cây; đặt tên đồ thị là "Khoảng tin cậy cho GTTB của các mẫu ngẫu nhiên", trục Ox là "Khoảng tin cậy cho GTTB", trục Oy là "Các mẫu ngẫu nhiên".