ỨNG DỤNG PHÂN TÍCH NHÓM TRONG LÂM HỌC

(Xem chi tiết Chương 8 trong cuốn sách: **Những cách thức và phương pháp xử lí số liệu trong lâm học**).

1. Khái niệm về phân tích nhóm

Phân tích nhóm (Grouping Analysis) là phân chia một tập dữ liệu lớn thành nhiều nhóm nhỏ dựa trên các biến phân loại định lượng (Quantitative classification Variables). Biến phân loại còn được gọi là biến lập nhóm, biến dự đoán, biến độc lập.

Những biến có ý nghĩa phân loại được sử dụng để xây dựng các hàm lập nhóm (Discriminant, Grouping, Classification Functions).

Hàm lập nhóm tuyến tính tổng quát có dạng như Hàm 1; trong đó $Z^{(k)}$ là hàm của nhóm thứ k (k = 1, 2…); $X_1^{(k)}, X_2^{(k)},..., X_i^{(k)}$ là biến phân loại thứ i của hàm lập nhóm thứ k; a, b, c…f là các hệ số của hàm lập nhóm. Các hệ số của hàm lập nhóm được xác định theo phương pháp lập nhóm. Hàm lập nhóm có thể bao gồm nhiều biến phân loại. Hàm lập nhóm tối ưu chỉ bao gồm một số biến phân loại hữu ích nhất hay khác biệt rõ rệt nhất giữa các nhóm.

$$Z^{(k)} = a\times X_1^{(k)} + b\times X_2^{(k)} + c\times X_3^{(k)} + \ldots + f\times X_i^{(k)} \qquad (1)$$

2. Mục đích và mục tiêu của phân tích nhóm

Mục đích của phân tích nhóm là cung cấp các hàm để dự đoán các đối tượng hay trường hợp mới thuộc về các nhóm khác nhau. Các đối tượng mới là các đối tượng nằm ngoài tập dữ liệu để xây dựng các hàm lập nhóm.

Phân tích nhóm có ba mục tiêu cơ bản: (1) Xác định những biến độc lập (biến dự đoán, biến phân loại) khác nhau nhiều nhất giữa các nhóm; (2) Xác định sự khác biệt có ý nghĩa thống kê giữa hai hay nhiều nhóm; (3) Xây dựng các hàm lập nhóm để phân loại những trường hợp mới vào các nhóm khác nhau.

3. Những ứng dụng của phân tích nhóm trong lâm học

(1) Phân loại (rừng, cấp lập địa, cấp nguy cơ cháy rừng, cấp sản lượng rừng…).

(2) Kiểm định sự khác nhau giữa các nghiệm thức hay đối tượng dựa trên nhiều biến số khác nhau.

(3) Xác định những tập hợp biến hữu ích nhất để phân loại hoặc so sánh giữa các đối tượng.

(4) So sánh vai trò của những tập hợp biến dự đoán khác nhau được sử dụng để xây dựng hàm lập nhóm.

(5) Tìm kiếm những giá trị bất thường hay ngoại lai.

4. Những điều kiện trong phân tích nhóm

Phân tích nhóm đòi hỏi phải thỏa mãn 6 điều kiện cơ bản.

(1) Số nhóm phân loại từ 2 trở lên.

(2) Mỗi nhóm phải có hai biến phân loại (Biến dự đoán) trở lên. Các biến phân loại này phải nhận các giá trị liên tục.

(3) Các biến phân loại là các biến định lượng và khác nhau rõ rệt giữa các nhóm. Nếu các biến phân loại là các biến định danh, thì chúng phải được mã hóa thành các biến giả (Dummy variables) hoặc biến nhị phân (0 và 1).

(4) Các biến phân loại của mỗi nhóm phải có phân bố chuẩn.

(5) Phương sai của các biến phân loại phải bằng nhau giữa các nhóm.

(6) Ma trận hiệp phương sai trong nội bộ nhóm phải bằng nhau giữa các nhóm.

5. Các bước cơ bản trong phân tích nhóm

Bước 1. Phân tích đối tượng nghiên cứu. Mục đích của bước này là chọn các biến dự đoán thích hợp (biến phân loại, biến độc lập); kiểm định sự phù hợp của các hàm lập nhóm; xác định phương pháp xây dựng các hàm lập nhóm và đánh giá ý nghĩa hay vai trò của các biến dự đoán trong các hàm lập nhóm.

Bước 2. Xác định mục tiêu phân loại nhóm và độ chính xác của các hàm phân loại nhóm. Nếu mục tiêu nghiên cứu là sắp xếp các trường hợp vào những nhóm khác nhau, thì hàm lập nhóm có thể được xây dựng từ một tập hợp biến nào đó. Nếu mục tiêu nghiên cứu là so sánh sự khác nhau giữa các nhóm đã biết trước, thì hàm lập nhóm được xây dựng từ những nhóm đã biết này. Những nhóm đã biết trước là những nhóm đã được định nghĩa rõ ràng. Chẳng hạn: (a) Các nghiệm thức

thí nghiệm trong gieo ươm; (b) Rừng trồng từ một loài cây gỗ nào đó theo ba phương thức khác nhau; (c) Rừng trồng trên các đất khác nhau...

Nếu mục tiêu nghiên cứu là xây dựng các hàm lập nhóm để xác định các trường hợp mới thuộc về những nhóm đã biết trước, thì các bước xây dựng hàm lập nhóm được bắt đầu từ những biến của các nhóm đã biết này. Chẳng hạn: (a) Phân cấp sinh trưởng cho những cây hình thành rừng trồng thuần loài đồng tuổi; (b) Phân chia các ô mẫu theo những kiểu QXTV đã biết. Trong trường hợp này, trước hết xây dựng các hàm lập nhóm tuyến tính Fisher. Sau đó sử dụng các hàm lập nhóm này để phân loại những trường hợp mới thuộc về những nhóm khác nhau. Một trường hợp mới thuộc về thành viên của một nhóm nào đó nếu hàm lập nhóm tại nhóm này nhận được giá trị lớn nhất.

Bước 3. Phân tích các kết quả phân loại nhóm. Kết quả phân loại (Classification Results) cung cấp các thông tin về kết quả phân loại các trường hợp vào từng nhóm (Bảng 1).

Bảng 1. Kết quả phân loại (Classification Results)

Kết quả phân loại		Nhóm	Predicted Group Membership		Total
			1	2	
Original (Ban đầu)	Count	1	110		110
		2		102	102
	%	1	100,0		100
		2		100,0	100
Cross-validated[b] (Kiểm định chéo)	Count	1	110		110
		2		102	102
	%	1	100,0		100
		2		100,0	100

a. 100.0% of original grouped cases correctly classified.

b. 100.0% of cross-validated grouped cases correctly classified.

Mức độ thành công của việc phân loại các trường hợp vào từng nhóm được thống kê bằng tần số và tỉ lệ phần trăm.

Bảng kết quả bao gồm kết quả phân loại ban đầu (Original Classification Results) và kết quả kiểm định độ chính xác của các trường hợp được phân loại ban đầu (Cross-validated Grouped Cases).

Bảng này cũng báo cáo: (a) Số trường hợp và tỉ lệ phần trăm số trường hợp bị phân loại nhầm vào mỗi nhóm; (b) Tỉ lệ phần trăm trung bình đối với số trường hợp được phân loại đúng.