PGS. TS. Nguyễn văn Thêm, TS. Đào Thị Thùy Dương

TS. Trần Thị Ngoan, ThS. Nguyễn Xuân Hùng

_____oOo_____

NHỮNG CÁCH THỨC VÀ PHƯƠNG PHÁP XỬ LÝ SỐ LIỆU TRONG LÂM HỌC

NHÀ XUẤT BẢN NÔNG NGHIỆP

Hà Nội, quý II/2021

LỜI NÓI ĐẦU

Cuốn sách "**NHỮNG CÁCH THỨC VÀ PHƯƠNG PHÁP XỬ LÝ SỐ LIỆU TRONG LÂM HỌC**" được biên soạn theo yêu cầu của đào tạo Đại học và sau Đại học đối với chuyên ngành lâm học của Trường Đại học lâm nghiệp – Cơ sở II tại tỉnh Đồng Nai. Mục đích cơ bản của cuốn sách này là cung cấp cho người học những kỹ thuật xử lý các thông tin trong lâm học.

Nội dung của cuốn sách này bao gồm 10 chương. Chương 1 giới thiệu về những kiểu nghiên cứu trong lâm học. Chương 2 giới thiệu về phương pháp phân tích kết cấu và đa dạng loài cây gỗ. Chương 3 giới thiệu về phương pháp phân tích cấu trúc rừng. Chương 4 giới thiệu về phương pháp phân tích mối liên hệ giữa những loài cây gỗ trong quần xã thực vật rừng. Chương 5 giới thiệu phân tích hồi quy trong lâm học. Chương 6 giới thiệu về phương pháp xây dựng mô hình sinh trưởng và sản lượng rừng. Chương 7 giới thiệu về phương pháp đánh giá lập địa rừng bằng chỉ số. Chương 8 giới thiệu về phương pháp phân tích nhóm và ứng dụng trong lâm học. Chương 9 giới thiệu về chỉ số mật độ và chỉ số cạnh tranh của quần thụ. Chương 10 giới thiệu về những phương pháp phân tích phương sai 1 và 2 nhân tố và áp dụng trong lâm học. Nội dung của mỗi chương bao gồm 2 phần; trong đó phần 1 trình bày khái quát về lý thuyết, còn phần 2 là ứng dụng một số phần mềm thống kê để xử lý các thông tin trong lâm học.

Biên soạn cuốn sách này là PGS. TS. Nguyễn Văn Thêm – Hội khoa học và kỹ thuật lâm nghiệp Tp. Hồ Chí Minh, TS. Đào Thị Thùy Dương, TS. Trần Thị Ngoan và ThS. Nguyễn Xân Hùng - Giảng viên Trường Đại học lâm nghiệp – Cơ sở II tại tỉnh Đồng Nai. PGS. TS. Nguyễn Văn Thêm biên soạn các chương 1, 3, 5, 6 và 8. TS. Đào Thị Thùy Dương biên soạn chương 2 và 9. TS. Trần Thị Ngoan biên soạn chương 7. ThS. Nguyễn Xuân Hùng biên soạn chương 4 và 10. Cuốn sách này được viết cho những sinh viên Đại học và sau Đại học đã có kiến thức cơ bản về lâm học và thống kê toán học.

Cuối cùng, chúng tôi nhận thấy rằng việc biên soạn một tài liệu với những nội dung có liên quan đến nhiều lĩnh vực khoa học khác nhau sẽ không thể tránh khỏi những thiếu sót. Vì thế, chúng tôi rất mong nhận được những trao đổi và ý kiến đóng góp chân tình của bạn đọc. Những ý kiến đóng góp xin gởi về địa chỉ: TS. Đào Thị Thùy Dương, giảng viên Trường Đại học lâm nghiệp – Cơ sở II tại tỉnh Đồng Nai. Email: daothuyduongvfu2@gmail.com.

Đồng Nai, *tháng 04/2021*

Các tác giả

MỤC LỤC

LỜI NÓI ĐẦU i
DANH SÁCH BẢNG ii
DANH SÁCH HÌNH viii
DANH SÁCH PHỤ LỤC ix
DANH SÁCH CÁC CHỮ VIẾT TẮT xiii
Chương 1. PHƯƠNG PHÁP NGHIÊN CỨU TRONG LÂM HỌC 1
1.1. Những kiểu nghiên cứu trong lâm học 1
1.2. Các bước cơ bản trong nghiên cứu lâm học 2
1.3. Những công cụ nghiên cứu trong lâm học 2
1.4. Phương pháp thu thập số liệu trong lâm học 3
1.5. Lập kế hoạch thu thập số liệu 13
1.6. Phương pháp phân tích số liệu trong lâm học 16
1.7. Độ tin cậy của kết quả nghiên cứu 20
Chương 2. PHÂN TÍCH KẾT CẤU VÀ ĐA DẠNG LOÀI CÂY GỖ 23
2.1. Khái niệm về kết cấu loài cây gỗ 23
2.2. Phương pháp xác định vai trò sinh thái của loài cây gỗ trong QXTV ... 24
2.3. Phương pháp so sánh sự tương đồng về loài cây gỗ giữa các QXTV ... 40
2.4. Phân tích đa dạng loài cây gỗ trong các QXTV 41
Chương 3. PHÂN TÍCH CẤU TRÚC RỪNG 58
3.1. Khái niệm về cấu trúc rừng 58
3.2. Phân tích cấu trúc rừng bằng phương pháp trắc đồ rừng 59
3.3. Phân tích cấu trúc rừng bằng các mô hình toán 60
Chương 4. PHÂN TÍCH MỐI LIÊN HỆ GIỮA CÁC LOÀI CÂY GỖ 83
4.1. Giới thiệu chung 83
4.2. Phân tích sự kết nhóm giữa các loài cây gỗ 85
4.3. Tương quan giữa các biến có thứ bậc và áp dụng trong lâm học 101
Chương 5. PHÂN TÍCH HỒI QUY TRONG LÂM HỌC 112
5.1. Giới thiệu chung 112
5.2. Hồi quy tuyến tính và ứng dụng trong lâm học 115
5.3. Hàm hồi quy bậc 2 và ứng dụng trong lâm học 129
5.4. Hàm hồi quy Sigmoid và Logit Gauss và ứng dụng trong lâm học 136
5.5. Hàm hồi quy Logistic và ứng dụng trong lâm học 144
5.6. Hàm hồi quy Poisson và ứng dụng trong lâm học 153
Chương 6. MÔ HÌNH SINH TRƯỞNG VÀ SẢN LƯỢNG RỪNG 162
6.1. Khái niệm về sinh trưởng, tăng trưởng và sản lượng rừng 162
6.2. Phương pháp mô hình hóa sinh trưởng của cây gỗ và rừng 165
6.3. Mô hình sinh trưởng quần thụ 166
6.4. Phân tích sai lệch của các hàm sinh trưởng 171
6.5. Phương pháp chọn các hàm sinh trưởng và sản lượng rừng 174

6.6. Phương pháp phân tích hồi quy phi tuyến tính và hồi quy Ridge.....186
Chương 7. ĐÁNH GIÁ LẬP ĐỊA RỪNG BẰNG CHỈ SỐ 197
7.1. Tình hình chung 197
7.2. Phương pháp đánh giá chất lượng lập địa rừng 198
Chương 8. PHÂN TÍCH NHÓM VÀ ỨNG DỤNG TRONG LÂM HỌC.....210
8.1. Khái quát về phân tích nhóm 210
8.2. Phân tích mô hình hai nhóm 216
8.3. Phân tích mô hình nhiều nhóm 225
Chương 9. CHỈ SỐ MẬT ĐỘ QUẦN THỤ VÀ CHỈ SỐ CẠNH TRANH 231
9.1. Chỉ số mật độ quần thụ 231
9.2. Xác định sự cạnh tranh giữa những cây gỗ trong quần thụ 239
9.3. Chỉ số hình thái thân và tán lá của quần thụ 248
Chương 10. PHÂN TÍCH PHƯƠNG SAI 251
10.1. Khái quát chung 251
10.2. Các mô hình phân tích phương sai.....251
10.3. Ứng dụng phân tích phương sai trong lâm học.....260
TÀI LIỆU THAM KHẢO.....285
PHỤ LỤC.....289

DANH SÁCH CÁC BẢNG

Bảng 1.1. Phân cấp độ phong phú của các loài cây thân thảo sống dưới tán rừng theo phương pháp của Druze. 6
Bảng 1.2. Phân phối kết quả điều trị bệnh phấn trắng trên 200 cây Keo lai. 7
Bảng 1.3. Phân phối kết quả điều trị bệnh phấn trắng theo 3 loại thuốc khác nhau trên 205 cây Keo lai. 8
Bảng 1.4. Phân bố số cây theo cấp đường kính đối với ưu hợp Dầu con rái (*Dipterocarpus alatus* Roxb). 8
Bảng 2.1. Phân cấp độ phong phú của các loài cây thân thảo sống dưới tán rừng theo phương pháp của Druze. 26
Bảng 2.2. Kết cấu loài cây gỗ của ưu hợp Dầu con rái được xác định theo hai tham số N% và G%.. 35
Bảng 2.3. Kết cấu loài cây gỗ của ưu hợp Dầu con rái được xác định theo theo ba tham số N%, G% và V%.....35
Bảng 2.4. Đặc trưng thống kê những chỉ số đa dạng loài cây gỗ đối với ưu hợp Dầu con rái.....52
Bảng 2.5. Hồ sơ đa dạng loài cây gỗ đối với hai ưu hợp Dầu con rái. Mỗi ưu hợp Dầu con rái bao gồm 3 ô tiêu chuẩn 0,25 ha.....53
Bảng 2.6. Hồ sơ đa dạng loài cây gỗ đối với 6 QXTV khác nhau. Mỗi QXTV bao gồm 5 ô tiêu chuẩn 0,25 ha. 56
Bảng 3.1. Phân bố của *Potentilla acaulis* trên 100 phân đoạn, mỗi phân đoạn 5 ô dạng bản và so sánh với phân bố nhị thức. 67
Bảng 3.2. Kiểm định phân bố cây tái sinh Dầu con rái trên mặt đất theo mô hình phân bố Poisson**.** 68
Bảng 3.3. Xác định phân bố cây tái sinh Dầu song nàng trên mặt đất bằng tiêu chuẩn T.....69
Bảng 3.4. Kiểu phân bố của cây gỗ trên mặt đất. Trường hợp J = 3. 70
Bảng 3.5. Kiểu phân bố của cây gỗ trên mặt đất. Trường hợp J = 4. 70
Bảng 3.6. Kiểm định phân bố trên mặt đất đối với cây tái sinh Dầu song nàng theo phân bố Thomas và Neyman. 73
Bảng 3.7. Sự phụ thuộc của phương sai theo kích thước ô thống kê. 73
Bảng 3.8. Phân cấp mức độ hỗn giao của các loài cây gỗ trong quần thụ. 76
Bảng 3.9. Mức độ hỗn giao của QXTV được xác đinh từ 5 điểm quan sát. 77
Bảng 3.10. Phân cấp mức độ đồng đều về kích thước của các cây gỗ trong QXTV.78
Bảng 3.11. Mức độ đồng đều về kích thước của các cây gỗ trong QXTV được xác định từ 3 điểm quan sát. 78
Bảng 3.12. Mức độ khác biệt về đường kính của các cây gỗ trong QXTV được xác định từ 3 điểm quan sát. 79
Bảng 3.13. Phân chia các kiểu phân bố cây trên mặt đất. 81
Bảng 4.1. Bảng phân tích mối quan hệ giữa hai loài cây gỗ. 87

Bảng 4.2. Mối quan hệ giữa *Luzula multriflora* và *Leucandenum vulgare*. 90
Bảng 4.3. Kết quả kiểm định mối quan hệ giữa Trắc và Hà nu. 95
Bảng 4.4. Kiểm định ảnh hưởng của Chò xót đến quan hệ giữa Trắc và Hà nu. 98
Bảng 4.5. Kiểm định ảnh hưởng của Chò xót và Lành ngạnh đến mối quan hệ giữa Trắc và Hà nu. 99
Bảng 4.6. Phân bố cấp mật độ cây tái sinh tự nhiên của Dầu con rái theo các cấp độ tàn che của tán rừng. 103
Bảng 4.7. Kết quả kiểm định mối quan hệ giữa mật độ cây tái sinh của Dầu con rái với cấp độ tàn che của tán rừng. 105
Bảng 4.8. Phân bố số cây Dầu con rái theo tình trạng bệnh và lập địa. 107
Bảng 4.9. Kiểm định quan hệ giữa mức độ bị bệnh của Dầu con rái với tuổi. 109
Bảng 4.10. Kết quả kiểm định mối quan hệ giữa mức độ bị bệnh của Dầu con rái với chỉ số SI. 1109
Bảng 4.11. Kiểm định mối quan hệ giữa mức độ bị bệnh của Dầu con rái với tuổi trong mỗi cấp chỉ số SI. 111
Bảng 5.1. Phân tích biến động (ANOVA) đối với hàm hồi quy tuyến tính đơn. ...117
Bảng 5.2. Các hệ số hồi quy và kiểm định sai lệch của hồi quy tuyến tính đơn. ...118
Bảng 5.3. Phân tích biến động đối với hàm hồi quy đa biến tuyến tính. 119
Bảng 5.4. Kết quả ước lượng các tham số và phân tích biến động đối với hàm hồi quy đa biến tuyến tính 119
Bảng 5.5. Mối quan hệ giữa tăng trưởng bề rộng vòng năm của Du sam với nhiệt độ không khí của các tháng trong năm. 128
Bảng 5.6. Phân tích vai trò của nhiệt độ không khí đối với tăng trưởng bề rộng vòng năm của Du sam. 129
Bảng 5.7. Hàm phản hồi của cây con Dầu song nàng 2 năm tuổi đối với hàm lượng mùn (X_1%) và pH_{H2O} (X_2) ở tầng đất mặt (0 – 30 cm) của đất Bazan. 134
Bảng 5.8. Những tham số sinh thái đối với sinh trưởng của cây con Dầu song nàng 2 năm tuổi. 134
Bảng 5.9. Hàm phản hồi của cây con Dầu song nàng 2 năm tuổi đối với hàm lượng mùn (X_1%) và pH_{H2O} (X_2) ở tầng đất mặt (0 – 30 cm) của đất Bazan được ước lượng bằng hồi quy Ridge. 135
Bảng 5.10. Hàm Sigmoid biểu diễn xác suất bắt gặp cây tái sinh Dầu con rái dựa theo độ ẩm ở tầng đất mặt. 140
Bảng 5.11. Hàm Logit Gauss biểu diễn xác suất bắt gặp cây tái sinh Dầu con rái dựa theo độ ẩm ở tầng đất mặt. 140
Bảng 5.12. Hàm Sigmoid biểu diễn xác suất bắt gặp cây tái sinh Dầu con rái dựa theo pH ở tầng đất mặt. 141
Bảng 5.13. Hàm Logit Gauss biểu diễn xác suất bắt gặp cây tái sinh Dầu con rái dựa theo pH ở tầng đất mặt. 142
Bảng 5.14. Hàm phản hồi của cây tái sinh Dầu con rái đối với độ ẩm (X_1) và pH_{H2O} (X_2) ở tầng đất mặt (0 – 30 cm). 143
Bảng 5.15. Những tham số sinh thái đối với độ bắt gặp cây tái sinh Dầu con rái. 144

Bảng 5.16. Tỷ lệ số cây họ Sao Dầu trong các cấp đường kính ở hai trạng thái rừng $IIIA_2$ và $IIIA_3$ 147
Bảng 5.17. Kết quả của hàm Logistic biểu diễn mối quan hệ giữa tỷ lệ số cây họ Sao Dầu trong các cấp đường kính ở hai trạng thái rừng $IIIA_2$ và $IIIA_3$. ...148
Bảng 5.18. Độ bắt gặp loài cây quan tâm trên hai loại đất khác nhau. 150
Bảng 5.19. Kết quả phân tích mối quan hệ giữa độ bắt gặp loài cây quan tâm theo ba biến X_1, X_2 và X_3. 152
Bảng 5.20. Tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột theo cấp D và loại đất. ...155
Bảng 5.21. Số liệu để xây dựng các hàm ước lượng tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột theo các cấp D và loại đất.....156
Bảng 5.22. Hàm ước lượng tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột trên đất bazan và đất xám phù sa cổ. 157
Bảng 5.23. Hàm ước lượng tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột theo cấp đường kính. Hàm hồi quy: $\lambda_{iJ} = \exp(b_0 + b_1*D)$.....158
Bảng 5.24. Hàm ước lượng tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột theo loại đất và cấp đường kính. Hàm hồi quy: $\lambda_{iJ} = \exp(b_0 + b_1*J + b_2*D)$. 159
Bảng 6.1. Quá trình sinh trưởng chiều cao của rừng Đước từ 2 – 12 tuổi. 163
Bảng 6.2. Tăng trưởng định kỳ 2 năm về chiều cao của rừng Đước 2 – 12 tuổi. ..163
Bảng 6.3. Quá trình sinh trưởng trữ lượng gỗ của rừng *Pinus sylvestris* 100 tuổi. 164
Bảng 6.4. Một số hàm thực nghiệm thường được ứng dụng để mô tả quá trình sinh trưởng của cây gỗ và rừng. 171
Bảng 6.5. Số liệu trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi. 174
Bảng 6.6. Mô hình ước lượng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi bằng bốn hàm khác nhau. 175
Bảng 6.7. Những đặc trưng sinh trưởng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi được xác định bằng bốn hàm khác nhau. 175
Bảng 6.8. Phân tích hồi quy tương quan M = f(A) đối với rừng Thông (*Pinus sylvestris*) 140 tuổi bằng hàm Korf. Trường hợp cố định tham số m. 178
Bảng 6.9. Phân tích hồi quy tương quan M = f(A) đối với rừng Thông (*Pinus sylvestris*) 140 tuổi bằng hàm Korf. Trường hợp cố định tham số c. 180
Bảng 6.10. So sánh sai lệch của ba phương pháp ước lượng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi bằng hàm Korf. 181
Bảng 6.11. Những đặc trưng sinh trưởng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi được xác định bằng hàm Korf. 181
Bảng 6.12. Thể tích thân cây bình quân của rừng trồng Keo lai trên ba cấp lập địa khác nhau tại tỉnh Đồng Nai. 183
Bảng 6.13. Ma trận tương quan giữa thể tích thân cây bình quân của rừng trồng Keo lai với bốn biến A, A^2, (D^2H) và (SI/H).....184
Bảng 6.14. Kết quả phân tích hàm ước lượng thể tích thân cây bình quân của rừng Keo lai theo bốn biến A, A^2, (D^2H) và (SI/H) bằng hồi quy bình phương nhỏ nhất.....184
Bảng 6.15. Kết quả phân tích hàm ước lượng thể tích thân cây bình quân của rừng Keo lai theo bốn biến A, A^2, (D^2H) và (SI/H) bằng hồi quy Ridge.....185

Bảng 6.16. Sai lệch của hàm ước lượng thể tích thân cây bình quân của rừng Keo lai theo phương pháp bình phương nhỏ nhất và hồi quy Ridge.....185
Bảng 6.17. Hàm trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi được xây dựng bằng hàm Korf ở dạng tuyến tính hóa với M = 2000 m^3/ha.....187
Bảng 6.18. Kết quả ước lượng tham số b_0 và b_1 của hàm Korf ở dạng tuyến tính hóa với M = 2000 m^3/ha. 188
Bảng 6.19. Kết quả xây dựng hàm ước lượng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi bằng hàm Korf.....189
Bảng 6.20. Kết quả xây dựng hàm ước lượng thể tích thân cây bình quân của rừng Keo lai 10 tuổi bằng hồi quy Ridge với tham số điều chỉnh bằng 0.....192
Bảng 6.21. Kết quả xây dựng hàm ước lượng thể tích thân cây bình quân của rừng Keo lai 10 tuổi bằng hồi quy Ridge với tham số điều chỉnh bằng 0,1.....194
Bảng 6.22. Kết quả xây dựng hàm ước lượng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi bằng hàm Korf.....196
Bảng 7.1. Đặc trưng thống kê chiều cao của những cây ưu thế và đồng ưu thế của rừng Tràm trồng từ cấp tuổi 2 – 14 ở khu vực Tây Nam Bộ.....202
Bảng 7.2. Ước lượng H_D của rừng Tràm trồng ở khu vực Tây Nam Bộ bằng hàm SI = f(A) khi giả định A_0 = 6 - 14 năm.....204
Bảng 7.3. Sai lệch ước lượng H_D của rừng Tràm ở khu vực Tây Nam Bộ bằng hàm SI = f(A) khi giả định A_0 = 6 - 14 năm 204
Bảng 7.4. Các hàm chỉ số SI đối với rừng Tràm tại khu vực Tây Nam Bộ. 205
Bảng 7.5. Đặc trưng chiều cao của những cây ưu thế và đồng ưu thế không tham gia xây dựng hàm chỉ số SI đối với rừng Tràm ở khu vực Tây Nam Bộ. 206
Bảng 7.6. So sánh điểm chặn (a) và độ dốc (b) của ba hàm chỉ số SI với điểm chặn (a') và độ dốc (b') của ba hàm H_D = f(A) đối với những cây mẫu không tham gia xây dựng hàm chỉ số SI.....207
Bảng 7.7. Biểu cấp lập địa đối với rừng Tràm trồng tại khu vực Tây Nam Bộ. 208
Bảng 8.1. Các thống kê mô tả trong hai nhóm 219
Bảng 8.2. Kiểm định ngang bằng trung bình của các biến giữa hai nhóm.....219
Bảng 8.3. Kiểm định ý nghĩa của các biến số dự tuyển. 220
Bảng 8.4. Tóm tắt hàm lập nhóm hợp quy. 221
Bảng 8.5. Các hệ số của hàm lập nhóm hợp quy chuẩn hóa và chưa chuẩn hóa ...221
Bảng 8.6. Các hệ số tương quan giữa các biến dự đoán với các biến của hàm lập nhóm hợp quy chuẩn hóa.....222
Bảng 8.7. Hàm lập nhóm hợp quy chưa chuẩn hóa ở trung tâm nhóm.....222
Bảng 8.8. Hàm ở trung tâm nhóm (Functions at Group Centroids) 222
Bảng 8.9. Các hệ số của hàm phân loại (Classification Function Coefficients) ...222
Bảng 8.10. Xác định nhóm đối với 4 cây mới theo hai hàm lập nhóm. 223
Bảng 8.11. Kết quả phân loại (Classification Results).....224
Bảng 8.12. Đặc trưng thống kê D, H và D_T của rừng Đước trồng 16 tuổi.....225
Bảng 8.13. Kiểm định ngang bằng D, H và D_T giữa 5 cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi. 226
Bảng 8.14. Kiểm định các biến ở trong mô hình.....226

Bảng 8.15. Kiểm định các biến không ở trong mô hình. 227
Bảng 8.16. Kiểm định ý nghĩa của các hàm phân loại hợp quy. 227
Bảng 8.17. Các hệ số của các hàm phân loại hợp quy chuẩn hóa. 227
Bảng 8.18. Các hệ số của các hàm phân loại hợp quy chưa chuẩn hóa. 227
Bảng 8.19. Điểm số trung bình của các hàm hợp quy ở trung tâm 5 cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi. 228
Bảng 8.20. Các hàm phân loại 5 cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi. 228
Bảng 8.21. Kết quả phân loại 5 cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi. 221
Bảng 8.22. Phân cấp sinh trưởng đối với 5 cây mới theo 5 hàm lập nhóm. 229
Bảng 8.23. Đặc trưng của rừng Đước 16 tuổi theo 5 cấp sinh trưởng thực tế.....229
Bảng 8.24. Đặc trưng của rừng Đước 16 tuổi theo 5 cấp sinh trưởng được phân loại theo những hàm lập nhóm. 229
Bảng 9.1. Đánh giá mức độ cạnh tranh giữa các cây gỗ trong quần thụ. 232
Bảng 9.2. Mối quan hệ giữa các số đo chỉ số mật độ quần thụ. 233
Bảng 9.3. Một số đặc trưng của rừng Đước trồng từ 11 – 25 tuổi tại khu vực Biển Tây thuộc tỉnh Cà Mau. 234
Bảng 9.4. Xây dựng hàm ước lượng đặc trưng của rừng Đước 11 – 25 tuổi. 235
Bảng 9.5. Chỉ số mật độ của rừng Đước được xác định theo phương pháp của Réineke (1933). 235
Bảng 9.6. Chỉ số mật độ của rừng Đước trồng từ 11 – 25 tuổi được xác định theo phương pháp của Curtis (1982). 237
Bảng 9.7. Mật độ và cường độ tỉa thưa đối với rừng Đước từ 11 - 25 tuổi được xác định theo phương pháp của Curtis (1982). 237
Bảng 9.8. Chỉ số không gian tương đối của rừng Đước trồng từ 11 - 25 tuổi.....238
Bảng 9.9. Ước lượng chỉ số không gian tương đối của rừng Đước 11 - 25 tuổi....238
Bảng 9.10. Mật độ của rừng Đước ở giới hạn dưới và giới hạn trên, số cây tỉa thưa (N_{TT}, cây/ha) và cường độ tỉa thưa theo N_L và N_U. 239
Bảng 9.11. Phân bố mật độ và diện tích tán theo các cấp H của rừng Đước ở tuổi 18 và 20 tại khu vực Biển Tây thuộc tỉnh Cà Mau. 242
Bảng 9.12. Những hàm ước lượng chỉ số cạnh tranh tán theo cấp chiều cao đối với rừng Đước trồng ở tuổi 18 và 20. 242
Bảng 9.13. Chỉ số cạnh tranh tán theo cấp chiều cao đối với rừng Đước 18 tuổi. Đơn vị tính: 1,0 ha. 243
Bảng 9.14. Chỉ số cạnh tranh tán theo cấp chiều cao đối với rừng Đước 20 tuổi. Đơn vị tính: 1,0 ha. 244
Bảng 9.15. Mật độ và chỉ số cạnh tranh tán theo cấp H của UhSaoDau<40%. 245
Bảng 9.16. Mật độ và chỉ số cạnh tranh tán theo cấp H của UhSaoDau>40% 245
Bảng 9.17. Những hàm ước lượng chỉ số cạnh tranh tán theo cấp chiều cao đối với hai ưu hợp cây họ Sao Dầu. 246
Bảng 9.18. Chỉ số cạnh tranh tán theo cấp chiều cao đối với $UhSaoDau_{<40\%}$.....246
Bảng 9.19. Chỉ số cạnh tranh tán theo cấp chiều cao đối với $UhSaoDau_{>40\%}$.....246

Bảng 9.20. Chỉ số cạnh tranh tán đối với những loài cây gỗ trong $UhSaoDau_{<40\%}$.248
Bảng 9.21. Chỉ số cạnh tranh đối với những loài cây gỗ trong $UhSaoDau_{>40\%}$.. ...249
Bảng 9.22. Những tiêu chuẩn đánh giá tính ổn định của rừng theo tỷ lệ H/D. 249
Bảng 9.23. Những tiêu chuẩn đánh giá tính ổn định của rừng theo tỷ lệ L_T/H. 250
Bảng 9.24. Tỷ lệ chiều cao và đường kính thân cây Đước ở các cấp tuổi. 250
Bảng 10.1. Kiểm định ảnh hưởng của địa hình đến sinh trưởng của Sao đen. 262
Bảng 10.2. Kiểm định ảnh hưởng của độ tàn che đến sinh trưởng chiều cao của cây con Dầu song nàng 8 tháng tuổi. Mô hình $H = \beta_0 + \beta_1X_1 + \beta_2$*Khối. 264
Bảng 10.3. Kiểm định ảnh hưởng của Super lân đến sinh trưởng chiều cao của cây con Gõ đỏ 3 tháng tuổi. Mô hình $H = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3$.....267
Bảng 10.4. Kiểm định ảnh hưởng của phân đạm và Super lân đến sinh trưởng chiều cao của cây con Dầu song nàng 6 tháng tuổi. 270
Bảng 10.5. Kiểm định ảnh hưởng của phân đạm và Super lân đến sinh trưởng chiều cao của cây con Sao đen 6 tháng tuổi. 272
Bảng 10.6. Kiểm định ảnh hưởng của kích thước bầu và phân tổng hợp NPK đến sinh trưởng chiều cao của Dầu con rái 8 tháng tuổi. 275
Bảng 10.7. Kiểm định ảnh hưởng của phương thức xử lý đất đến sinh trưởng chiều cao của ba loài (Dầu con rái, Sao đen, Vên vên) 8 tháng tuổi.....278
Bảng 10.8. Kiểm định ảnh hưởng của phân tổng hợp NPK đến sinh trưởng của cây con Gõ đỏ 6 tháng tuổi. 281
Bảng 10.9. Kiểm định ảnh hưởng của phân tổng hợp NPK đến sinh trưởng của cây con Gõ đỏ 6 tháng tuổi. 284

DANH SÁCH CÁC HÌNH

Hình 1.1. Sơ đồ bố trí ô mẫu theo phương pháp hệ thống. (a) Bố trí ô mẫu theo những số định trước. (b) Bố trí ô mẫu theo một cự ly nhất định. (c) Bố trí ô mẫu theo mắt lưới. 11
Hình 1.2. Ba kiểu phân bố số cây theo cấp H của rừng trồng thuần loài đồng tuổi. 19
Hình 2.1. Những thông tin về các loài cây gỗ và độ phong phú của chúng.....36
Hình 2.2. Xác định kết cấu loài cây gỗ theo ba biến N, G và V. 36
Hình 2.3. Các đặc trưng thống kê N, G và V đối với các loài cây gỗ. 36
Hình 2.4. Tóm tắt các đặc trưng thống kê N, G và V đối với các loài cây gỗ 37
Hình 2.5. Báo cáo các đặc trưng thống kê N, G và V đối với các loài cây gỗ. 37
Hình 2.6. Khay báo cáo các biến N, G và V đối với các loài cây gỗ. 37
Hình 2.7. Thống kê các biến N, G và V đối với các loài cây gỗ. 38
Hình 2.8. Đồ thị biểu diễn hồ sơ đa dạng loài cây gỗ của Rényi đối với hai ưu hợp Dầu con rái. 54
Hình 2.9. Đồ thị biểu diễn hồ sơ đa dạng loài cây gỗ của Rényi đối với sáu QXTV khác nhau. 57
Hình 3.1. Phân bố N/D của rừng trồng thuần loài đồng tuổi. 61
Hình 3.2. Những kiểu phân bố của các loài cây gỗ trên mặt đất.....64
Hình 3.3. Sơ đồ bố trí hệ thống ô mẫu để phân tích phân bố cây trên mặt đất. 65
Hình 3.4. Sơ đồ bố trí các ô mẫu để phân tích phân bố cây trên mặt đất.....68
Hình 3.5. Mối quan hệ giữa phương sai với kích thước khối ô mẫu.....74
Hình 4.1. Các loài cây gỗ và độ bắt gặp của chúng. 94
Hình 4.2. Xác định sự kết nhóm giữa các loài cây gỗ. 94
Hình 4.3. Kiểm định mối quan hệ giữa các loài cây gỗ. 95
Hình 4.4. Xây dựng bảng chéo để phân tích quan hệ giữa ba loài cây gỗ. 97
Hình 4.5. Xây dựng bảng chéo để phân tích ảnh hưởng của Chò xót và Lành ngạnh đến mối quan hệ giữa Trắc và Hà nu. 98
Hình 4.6. Số liệu tái sinh của Dầu con rái và độ tàn che tán rừng. 104
Hình 4.7. Xây dựng Bảng 5*5 giữa 5 cấp mật độ cây tái sinh của Dầu con rái và 5 cấp độ tàn che của tán rừng.....104
Hình 4.8. Xác định hệ số tương quan giữa mật độ cây tái sinh của Dầu con rái với cấp độ tàn che của tán rừng.....105
Hình 4.9. Số liệu Dầu con rái theo tình trạng bệnh và lập địa. 107
Hình 4.10. Xây dựng Bảng chéo 3 cấp bệnh và 3 cấp A của Dầu con rái. 108
Hình 4.11. Xác định mối quan hệ giữa 3 cấp bệnh với tuổi.....108
Hình 4.12. Xây dựng bảng chéo biểu thị mối quan hệ giữa mức bị bệnh với tuổi Dầu con rái theo ba cấp chỉ số SI.....110
Hình 4.13. Xác định quan hệ giữa 3 cấp bệnh với tuổi trong mỗi cấp chỉ số SI....110
Hình 5.1. Các dạng đường cong phản hồi được lập theo biến môi trường. 114

Hình 5.2. Những đường cong phản hồi được phân chia từ đường cong hai đỉnh bằng cách giới hạn khoảng thu mẫu. 115
Hình 5.3. Mối quan hệ giữa Y với X.....116
Hình 5.4. Chỉ số Kd của Du sam và chỉ số $T_1 – T_{12}$ từ 1980 - 2013. 126
Hình 5.5. Số liệu để xác định tương quan Kd = f(T_{1-12}).....127
Hình 5.6. Ma trận tương quan Kd = f(T_{1-12}).....127
Hình 5.7. Kỹ thuật phân tích hồi quy tương quan riêng phần. 128
Hình 5.8. Đường cong phản hồi của cây gỗ và rừng với môi trường. 130
Hình 5.9. Số liệu để xây dựng hàm phản hồi H = f(X_1, X_2).....131
Hình 5.10. Xác định hàm phản hồi H = f(X_1). 132
Hình 5.11. Kết quả phân tích hàm phản hồi H = f(X_1).....132
Hình 5.12. Xác định hàm phản hồi H = f(X_1, X_2, X_1*X_2).....133
Hình 5.13. Kết quả phân tích hàm phản hồi H = f(X_1, X_2, X_1*X_2).....133
Hình 5.14. Đồ thị biểu diễn sự biến đổi chiều cao của cây con Dầu song nàng 2 năm tuổi theo hàm lượng mùn ở tầng đất mặt. 134
Hình 5.15. Đồ thị biểu diễn sự biến đổi chiều cao của cây con Dầu song nàng 2 năm tuổi theo pH_{H2O} ở tầng đất mặt.....135
Hình 5.16. Số liệu để xây dựng hàm phản hồi P = f(X_1, X_2). 138
Hình 5.17. Xác định hàm phản hồi P = f(X_1).....138
Hình 5.18. Kết quả phân tích hàm hồi quy P = f(X_1) theo hàm Sigmoid.....138
Hình 5.19. Chọn hàm hồi quy P = f(X_1) theo hàm Logit Gauss. 139
Hình 5.20. Kết quả phân tích hàm hồi quy P = f(X_1) theo hàm Logit Gauss.....139
Hình 5.21. Xác định hàm phản hồi P = f(X_1, X_2, X_1*X_2). 142
Hình 5.22. Kết quả phân tích hàm phản hồi H = f(X_1, X_2, X_1*X_2).....143
Hình 5.23. Số liệu để xây dựng hàm hồi quy Logistic. 147
Hình 5.24. Xác định hàm hồi quy Logistic. 148
Hình 5.25. Chọn hàm hồi quy Logistic để mô tả quan hệ P = f(J, D).....148
Hình 5.26. Số liệu để xây dựng hàm hồi quy Y = f(X_i). 150
Hình 5.27. Xác định hàm hồi quy Y = f(X_1). 151
Hình 5.28. Xây dựng hàm hồi quy Y = f(X_1, X_2).....151
Hình 5.29. Xác định hàm hồi quy Y = f(X_1, X_2, X_3).....152
Hình 5.30. Số liệu để xây dựng hàm hồi quy Poisson λ_{iJ} = f(D, J).....156
Hình 5.31. Xác định hàm hồi quy λ_{iJ} = exp(b_0 + b_1*J). 157
Hình 5.32. Xác định hàm hồi quy λ_{iJ} = exp(b_0 + b_1*J + b_2*D).....159
Hình 5.33. Xác định những biến giải thích phù hợp nhất. 160
Hình 5.34. Tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột theo các cấp đường kính khác nhau trên đất bazan và đất xám phù sa cổ. 160
Hình 5.35. Đồ thị biểu diễn quan hệ giữa tham số tỷ lệ λ_i với biến giải thích X...161
Hình 6.1. Tăng trưởng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 100 tuổi. 164
Hình 6.2. Các pha sinh trưởng của cây gỗ và quần thụ thuần loài đồng tuổi.....170
Hình 6.3. Đồ thị so sánh sai lệch của hàm Korf và hàm lũy thừa được sử dụng để phân tích sinh trưởng trữ lượng gỗ của rừng Thông ba lá 40 tuổi. 172

Hình 6.4. Đồ thị so sánh lượng tăng trưởng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi được ước lượng bằng bốn hàm khác nhau. 176
Hình 6.5. Đồ thị biểu diễn sự biến đổi của sai số ước lượng trữ lượng gỗ của rừng *Pinus sylvestris* 140 tuổi theo hàm Korf. Trường hợp cố định tham số m. 179
Hình 6.6. Đồ thị biểu diễn sự biến đổi của sai số ước lượng trữ lượng gỗ của rừng *Pinus sylvestris* 140 tuổi theo hàm Korf. Trường hợp cố định tham số c. .181
Hình 6.7. Đồ thị so sánh lượng tăng trưởng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi được ước lượng bằng hàm Korf; trong đó ba tham số được ước lượng theo ba phương pháp khác nhau. 182
Hình 6.8. Số liệu đầu vào để ước lượng ba tham số m, b và c ban đầu của hàm Korf. 187
Hình 6.9. Xây dựng hàm Korf ở dạng tuyến tính: $Y = b_0 - b_1*X$. 188
Hình 6.10. Giả định ba tham số m, b và c ban đầu của hàm Korf. 189
Hình 6.11. Kết quả của hàm ước lượng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi bằng hàm Korf. 190
Hình 6.12. Số liệu đầu vào để xây dựng hàm thể tích thân cây Keo lai bằng kỹ thuật hồi quy Ridge. 191
Hình 6.13. Xây dựng hàm thể tích bằng hồi quy Ridge. 191
Hình 6.14. Kết quả phân tích hàm thể tích bằng hồi quy Ridge. 191
Hình 6.15. Đồ thị biểu diễn yết tố khuyếch đại phượng sai. 193
Hình 6.16. Điều chỉnh tham số thích hợp trong hồi quy Ridge. 194
Hình 6.17. Xác định các hệ số hồi quy chuẩn hóa trong hồi quy Ridge 194
Hình 7.1. Đồ thị biểu diễn sự biến đổi chiều cao cây tầng trội theo tuổi rừng Tràm trồng ở khu vực Tây Nam Bộ. 203
Hình 7.2. Đồ thị so sánh chiều cao của những cây kiểm tra và ba cấp chỉ số SI đối với rừng Tràm trồng từ 2 – 14 tuổi ở khu vực Tây Nam Bộ. 207
Hình 7.3. Đường cong chiều cao cây tầng trội (H_D, m) và đường cong chỉ số SI (m) đối với rừng Tràm tại khu vực Tây Nam Bộ. 208
Hình 8.1. Đặc trưng của hai quần thụ 1 và 2. 216
Hình 8.2. Đường dẫn vào Menu phân tích hàm lập nhóm. 217
Hình 8.3. Phân tích hàm lập nhóm tuyến tính Fisher 217
Hình 8.4. Những thống kê trong phân tích hàm lập nhóm. 218
Hình 8.5. Xác định những thông tin của hàm lập nhóm. 218
Hình 8.6. Phương pháp chọn biến từng bước. 219
Hình 8.7. Chọn phương pháp phân tích hàm lập nhóm. 219
Hình 8.8. Xác định những thống kê của các trường hợp. 223
Hình 9.1. Đồ thị biểu diễn mật độ hiện còn và mật độ tại tuổi cơ sở của rừng Đước được xác định theo phương pháp của Réineke (1933). 236
Hình 9.2. Đồ thị biểu diễn chỉ số cạnh tranh tán theo cấp chiều cao đối với rừng Đước 18 tuổi. 243
Hình 9.3. Đồ thị biểu diễn chỉ số cạnh tranh tán theo cấp chiều cao đối với rừng Đước 20 tuổi. 244

Hình 9.4. Đồ thị biểu diễn chỉ số cạnh tranh tán theo cấp chiều cao đối với hai ưu hợp cây họ Sao Dầu. 247
Hình 9.5. Đồ thị biểu diễn sự tích lũy chỉ số cạnh tranh tán theo cấp chiều cao đối với hai ưu hợp cây họ Sao Dầu. 247
Hình 10.1. Sơ đồ bố trí thí nghiệm theo khối đầy đủ ngẫu nhiên. 255
Hình 10.2. Sơ đồ bố trí thí nghiệm theo kiểu bình phương la tinh. 256
Hình 10.3. Sơ đồ bố trí thí nghiệm hai nhân tố ngẫu nhiên hoàn toàn.....257
Hình 10.4. Sơ đồ bố trí thí nghiệm hai nhân tố theo khối đầy đủ ngẫu nhiên. 258
Hình 10.5. Sơ đồ bố trí thí nghiệm hai nhân tố theo kiểu bình phương la tinh.....259
Hình 10.6. Sơ đồ bố trí thí nghiệm hai nhân tố có lô phụ. 260
Hình 10.7. Đường kính của rừng Sao đen trên ba cấp địa hình. 260
Hình 10.8. Số liệu để xây dựng mô hình: $D = \beta_0 + \beta_1X_1 + \varepsilon$. 261
Hình 10.9. Xây dựng mô hình: $D = \beta_0 + \beta_1X_1 + \varepsilon$. 261
Hình 10.10. So sánh D của rừng Sao đen giữa ba địa hình. 261
Hình 10.11. Kiểm định thuần nhất D của rừng Sao đen giữa ba địa hình.....261
Hình 10.12. Chiều cao của Dầu song nàng 8 tháng tuổi. 263
Hình 10.13. Số liệu để xây dựng mô hình: $H = \beta_0 + \beta_1X_1 + \beta_2$*Khối. 264
Hình 10.14. Xây dựng mô hình: $H = \beta_0 + \beta_1X_1 + \beta_2$*Khối. 264
Hình 10.15. So sánh H của Dầu song nàng theo 5 cấp độ tàn che. 264
Hình 10.16. Kiểm định thuần nhất H của Dầu song nàng theo 5 độ tàn che. 264
Hình 10.17. Chiều cao của Gõ đỏ 3 tháng tuổi. 265
Hình 10.18. Số liệu để xây dựng mô hình: $H = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3$. 266
Hình 10.19. Xây dựng mô hình: $H = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3$.....266
Hình 10.20. So sánh H của Gõ đỏ theo 6 mức Super lân. 266
Hình 10.21. Kiểm định thuần nhất H của Gõ đỏ theo 6 mức Super lân.....261
Hình 10.22. Chiều cao của cây con Dầu song nàng 6 tháng tuổi. 268
Hình 10.23. Số liệu để xây dựng mô hình: $H = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2$. 269
Hình 10.24. Xây dựng mô hình: $H = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2$. 269
Hình 10.25. So sánh H Dầu song nàng 6 tháng theo ba mức đạm và Super lân. ...269
Hình 10.26. Kiểm định thuần nhất H của Dầu song nàng 6 tháng tuổi theo ba mức đạm và Super lân. 269
Hình 10.27. Chiều cao của cây con Sao đen 6 tháng tuổi. 271
Hình 10.28. Số liệu xây dựng mô hình: $H = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2 + \beta_4$*$X_3$.171
Hình 10.29. Xây dựng mô hình: $H = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2 + \beta_4$*$X_3$. 271
Hình 10.30. So sánh H của Sao đen 6 tháng tuổi theo ba mức đạm và Super lân. 272
Hình 10.31. Kiểm định thuần nhất H của Sao đen 6 tháng tuổi theo ba mức đạm và Super lân. 272
Hình 10.32. Chiều cao của cây con Dầu con rái 8 tháng tuổi. 273
Hình 10.33. Số liệu để xây dựng mô hình tuyến tính tổng quát theo kiểu thí nghiệm bình phương la tinh. 274
Hình 10.34. Xây dựng mô hình tuyến tính tổng quát theo kiểu thí nghiệm bình phương la tinh. 274

Hình 10.35. So sánh H của Dầu con rái 8 tháng tuổi theo 2 cấp bầu và 3 mức phân tổng hợp NPK. 274
Hình 10.36. Kiểm định thuần nhất H của Dầu con rái 8 tháng tuổi theo 2 cấp bầu và 3 mức phân NPK. 274
Hình 10.37. Chiều cao của Dầu con rái, Sao đen, Vên vên 8 tháng tuổi. 276
Hình 10.38. Số liệu để xây dựng mô hình tuyến tính tổng quát theo kiểu thí nghiệm lô chính – lô phụ. 277
Hình 10.39. Xây dựng mô hình tuyến tính tổng quát theo kiểu thí nghiệm lô chính – lô phụ. 277
Hình 10.40. So sánh H của Dầu con rái, Sao đen, Vên vên 8 tháng tuổi theo 3 phương thức xử lý đất. 277
Hình 10.41. Kiểm định thuần nhất H của Dầu con rái, Sao đen, Vên vên theo 3 phương thức xử lý đất. 277
Hình 10.42. Sinh trưởng của cây con Gõ đỏ 6 tháng tuổi dưới ảnh hưởng của phân tổng hợp NPK. 279
Hình 10.43. Số liệu để xây dựng mô hình: $Y_i = \beta_0 + \beta_1X_1 + \beta_2X_2$. 280
Hình 10.44. Xây dựng mô hình: $Y_i = \beta_0 + \beta_1X_1 + \beta_2X_2$. 280
Hình 10.45. So sánh D, H và số lá của Gõ đỏ 6 tháng tuổi theo các mức phân tổng hợp NPK. 280
Hình 10.46. Kiểm định thuần nhất D, H và số lá của Gõ đỏ 6 tháng tuổi theo các mức phân tổng hợp NPK 280
Hình 10.47. Sinh trưởng của cây con Gõ đỏ 6 tháng tuổi dưới ảnh hưởng của phân chuồng hoai và phân tổng hợp NPK. 282
Hình 10.48. Số liệu để xây dựng mô hình: $Y_i = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_2*X_3$. 283
Hình 10.49. Xây dựng mô hình: $Y_i = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_2*X_3$. 283
Hình 10.50. So sánh D, H và số lá của Gõ đỏ 6 tháng tuổi theo các mức phân chuồng và phân NPK. 283
Hình 10.51. Kiểm định thuần nhất D, H và số lá của Gõ đỏ theo các mức phân chuồng và phân NPK. 283

PHỤ LỤC

Phụ lục 1. Kết cấu loài cây gỗ của ưu hợp Dầu con rái được xác định theo theo chỉ số IVI với hai tham số N% và G%. 289
Phụ lục 2. Kết cấu loài cây gỗ của ưu hợp Dầu con rái được xác định theo theo chỉ số IVI với ba tham số N%, G% và V%. 290
Phụ lục 3. Chuỗi chỉ số bề rộng vòng năm chuẩn hóa của Du sam. 291
Phụ lục 4. Chỉ số nhiệt độ không khí của các tháng từ năm 1981 - 2013. 291
Phụ lục 5. Ảnh hưởng của hàm lượng mùn (X_1%) và pH_{H2O} (X_2) của đất Bazan nâu đỏ đến sinh trưởng chiều cao (H, m) của Dầu song nàng 2 năm tuổi.....292
Phụ lục 5. Ảnh hưởng của độ ẩm (X_1%) và pH_{H2O} (X_2) đến xác suất bắt gặp (P) cây tái sinh tự nhiên của Dầu con rái.. 293
Phụ lục 7. Ảnh hưởng của địa hình đến sinh trưởng của rừng Sao đen. Thí nghiệm 1 nhân tố được bố trí hoàn toàn ngẫu nhiên.....294
Phụ lục 8. Ảnh hưởng của độ tàn che đến sinh trưởng chiều cao (H, cm) của cây con Dầu song nàng 8 tháng tuổi. Thí nghiệm 1 nhân tố được bố trí theo khối ngẫu nhiên hoàn toàn. 295
Phụ lục 9. Ảnh hưởng của Super lân đến đến sinh trưởng chiều cao của cây con Gõ đỏ 3 tháng tuổi ở vườn ươm.....295
Phụ lục 10. Ảnh hưởng của phân đạm và Super lân đến sinh trưởng của cây con Dầu song nàng trong giai đoạn 6 tháng tuổi ở vườn ươm. Thí nghiệm hai yếu tố được bố trí ngẫu nhiên. 296
Phụ lục 11. Ảnh hưởng của phân đạm và Super lân đến sinh trưởng của cây con Sao đen trong giai đoạn 6 tháng tuổi ở vườn ươm. Thí nghiệm hai yếu tố được bố trí theo khối ngẫu nhiên hoàn toàn. 297
Phụ lục 12. Ảnh hưởng của kích thước bầu và phân tổng hợp NPK đến sinh trưởng chiều cao của Dầu con rái 8 tháng tuổi. Thí nghiệm được bố trí theo kiểu bình phương la tinh. 298
Phụ lục 13. Ảnh hưởng của phương thức xử lý đất đến sinh trưởng chiều cao của Dầu con rái, Sao đen và Vên vên 8 tháng tuổi. Thí nghiệm được bố trí theo kiểu lô chính - lô phụ. 298
Phụ lục 14. Ảnh hưởng của hàm lượng phân NPK đến sinh trưởng của Gõ đỏ 6 tháng tuổi. Thí nghiệm được bố trí theo khối ngẫu nhiên hoàn toàn.....299
Phụ lục 15. Ảnh hưởng của hàm lượng phân NPK và phân chuồng hoai đến sinh trưởng của Gõ đỏ 6 tháng tuổi. Thí nghiệm được bố trí theo khối ngẫu nhiên hoàn toàn. 299

NHỮNG CHỮ VIẾT TẮT

Chữ viết tắt	Tên gọi đầy đủ
A (năm)	Tuổi cây và quần thụ (rừng).
B (kg, tấn)	Sinh khối của cây gỗ và rừng.
Bias	Sai lệch.
Bias%	Sai lệch theo phần trăm.
BFNN	Phương pháp bình phương nhỏ nhất.
C (%)	Độ che phủ hay độ tàn che của tán lá theo chiều nằm ngang trên mặt đất.
CCI	Chỉ số cạnh tranh tán.
CV%	Hệ số biến động.
D (cm)	Đường kính cây ngang ngực.
D_{Bq} (cm)	Đường kính ngang ngực bình quân.
D_G (cm)	Đường kính toàn phương hay đường kính tương ứng với cây có tiết diện ngang bình quân.
$D_{Max} – D_{Min}$	Biên độ biến động đường kính thân cây.
D_T (m)	Đường kính tán.
F(%)	Tần số bắt gặp loài cây gỗ.
g (m2/ha)	Tiết diện ngang thân cây gỗ trên 1 ha.
G (m2/ha)	Tiết diện ngang của rừng Đước trên 1 ha.
H (m)	Chiều cao thân cây vút ngọn.
$H_{Max} – H_{Min}$	Biên độ biến động chiều cao thân cây.
H_D (m)	Chiều cao cây ưu thế và đồng ưu thế (tầng trội).
H_{DC} (m)	Chiều cao thân dưới cành lớn nhất còn sống.
ID	Chỉ số ưu thế về kích thước thân cây.
L_T (m)	Chiều dài tán cây.
$L_{TMax} – L_{TMin}$	Biên độ chiều dài tán cây.
MAE	Sai số tuyệt đối trung bình.
MAPE	Sai số tuyệt đối trung bình theo phần trăm.
ME	Sai số trung bình.
MPE	Sai số trung bình theo phần trăm.
M (m3/ha)	Trữ lượng gỗ trên 1 ha.
N (cây/ha)	Mật độ rừng trên 1 ha.
N_{HC} (cây/ha)	Mật độ rừng hiện còn hay thực tế trên 1 ha.
N_{TT} (cây/ha)	Mật độ tỉa thưa trên 1 ha.
N_{STT} (cây/ha)	Mật độ rừng sau tỉa thưa trên 1 ha.
OTC	Ô mẫu hay ô tiêu chuẩn.
ODB	Ô dạng bản.

PM (%)	Suất tăng trưởng trữ lượng gỗ.
QXTV	Quần xã thực vật rừng hay quần thụ.
R^2	Hệ số xác định.
RD	Chỉ số mật độ quần thụ.
RD35%	Chỉ số mật độ quần thụ ở giới hạn dưới của tỉa thưa.
RD50%	Chỉ số mật độ quần thụ ở giới hạn trên của tỉa thưa.
RS	Chỉ số không gian tương đối.
RS_L	Chỉ số không gian tương đối ở giới hạn dưới.
RS_U	Chỉ số không gian tương đối ở giới hạn trên.
S (loài)	Loài cây gỗ.
SCI	Chỉ số phức tạp về cấu trúc rừng.
SEE	Sai lệch chuẩn của ước lượng.
SI (m)	Chỉ số lập địa.
SSE	Tổng sai lệch bình phương.
SSE_{Min}	Tổng sai lệch bình phương nhỏ nhất.
TTR	Trạng thái rừng.
V (m^3/ha)	Thể tích thân cây.
Z (m^2)	Diện tích ô tiêu chuẩn.
ZY	Lượng tăng trưởng thường xuyên về đường kính, chiều cao, tiết diện ngang và trữ lượng gỗ của quần thụ.
ZYn	Lượng tăng trưởng bình quân theo định kỳ n năm về đường kính, chiều cao, tiết diện ngang và trữ lượng gỗ của quần thụ.
ΔY	Lượng tăng trưởng bình quân năm về về đường kính, chiều cao, tiết diện ngang và trữ lượng gỗ của quần thụ.

Chương 1
PHƯƠNG PHÁP NGHIÊN CỨU TRONG LÂM HỌC

1.1. Những kiểu nghiên cứu trong lâm học

Lâm học là khoa học nghiên cứu rừng và sự tương tác giữa rừng với môi trường sống của chúng, các quy luật tái sinh rừng, những đặc tính của quần thể và quần xã cây gỗ, động thái biến đổi của rừng theo không gian và thời gian, phân loại rừng.

Để hiểu về đời sống của các loài cây gỗ và rừng, lâm học sử dụng hai kiểu nghiên cứu. Kiểu thứ nhất là nghiên cứu sinh thái cá thể. Kiểu này nghiên cứu sự tương tác giữa một loài cây gỗ với môi trường sống của nó. Kiểu thứ hai là nghiên cứu sinh thái quần thể. Kiểu này nghiên cứu sự tương tác giữa nhiều loài cây gỗ với môi trường sống của chúng. Trên phạm vi không gian rộng lớn, lâm học cũng sử dụng những phương pháp nghiên cứu trong sinh thái cảnh quan để nghiên cứu các kiểu rừng và các mối quan hệ giữa chúng với môi trường. Kết quả của những nghiên cứu cảnh quan giúp cho nhà lâm học phân loại rừng và xây dựng những phương thức lâm sinh.

Hầu hết những nghiên cứu sinh thái quần thể chỉ tập trung vào nghiên cứu quần xã thực vật rừng (QXTV). Bởi vì cây gỗ cung cấp cho con người nhiều giá trị nhất, nên lâm học chỉ tập trung nghiên cứu những quần xã cây gỗ hay quần thụ. Tất cả những nghiên cứu này dựa trên phân tích cảnh quan rừng. Chẳng hạn: Phân loại thảm thực vật và phân loại đất có thể được sử dụng để phân loại cảnh quan rừng. Để nghiên cứu cảnh quan rừng, lâm học không chỉ sử dụng những kiến thức lâm học, mà còn cả những kiến thức về khí hậu, địa chất, địa lý, sinh học và đất...

Ranh giới giữa sinh thái quần thể và sinh thái cá thể là không rõ ràng. Điều này được giải thích là do một số thảm thực vật chỉ được hình thành từ một số loài cây gỗ hoặc chỉ một loài cây gỗ. Vì thế, khi nghiên cứu những thảm thực vật này, chúng ta có thể định nghĩa thảm thực vật hoặc là quần thể hoặc là quần xã. Nếu mục tiêu nghiên cứu là phân loại thảm thực vật, thì việc nghiên cứu này là nghiên cứu sinh thái quần xã. Nếu mục tiêu nghiên cứu là xác định đặc tính sinh thái của một số loài cây gỗ trong những hệ sinh thái khác nhau, thì đó là nghiên cứu sinh thái hệ sinh thái hay nghiên cứu cảnh quan rừng. Nếu mục tiêu nghiên cứu là xác định những tham số sinh thái (biên độ sinh thái, tối ưu sinh thái, phạm vi chống chịu) đối với một loài cây gỗ nào đó, thì nghiên cứu này là nghiên cứu sinh thái cá thể.

1.2. Các bước cơ bản trong nghiên cứu lâm học

Bước đầu tiên của nghiên cứu là xây dựng hệ thống những câu hỏi cần phải trả lời, các mục tiêu nghiên cứu và các giả thuyết nghiên cứu. Kế đến là xác định những cách thức và phương pháp nghiên cứu. Cuối cùng là xác định phương pháp phân tích số liệu, tóm tắt và giải thích kết quả nghiên cứu.

Rừng là hệ sinh thái rất phức tạp. Tính phức tạp biểu hiện ở chỗ rừng bao gồm rất nhiều thành phần (khí hậu, địa hình, đất, cây gỗ, cây bụi, thảm cỏ, hệ động vật và hệ vi sinh vật...) với những đặc tính rất khác nhau. Rừng phân bố trên không gian rộng lớn và khó đi lại. Rừng không chỉ bao gồm rừng tự nhiên với nhiều loài cây gỗ khác nhau, mà còn cả rừng nhân tạo. Để hiểu rõ những đặc tính của cây gỗ và rừng và mối quan hệ giữa chúng với môi trường, lâm học cần phải thu thập rất nhiều số liệu khác nhau. Phần lớn số liệu chỉ được thu thập bằng phương pháp quan sát và mô tả. Để nhận được những thông tin hữu ích và đáng tin cậy, lâm học cũng sử dụng những phương pháp nghiên cứu thực nghiệm. Nghiên cứu thực nghiệm trong lâm học là công việc rất khó khăn, đặc biệt là ở mức quần xã, hệ sinh thái và cảnh quan. Hầu hết những nghiên cứu ở các mức này là mô tả. Mức độ chính xác của số liệu tùy thuộc vào kế hoạch nghiên cứu, nơi thu mẫu (đối tượng thu mẫu) và thời gian thu mẫu. Bởi vậy, xác định chiến lược nghiên cứu là một vấn đề rất quan trọng. Nó không chỉ giúp cho nhà lâm học tiết kiệm thời gian, nhân lực và kinh phí, mà còn thu được số liệu có thể giải thích được.

1.3. Những công cụ nghiên cứu trong lâm học

Mục đích chính của các nghiên cứu trong lâm học là cung cấp những thông tin cho việc xây dựng lý thuyết về rừng và những phương thức lâm sinh. Ngoài ra, các kết quả nghiên cứu trong lâm học còn cung cấp những thông tin để hỗ trợ cho việc bảo tồn tài nguyên rừng và đánh giá việc sử dụng thảm thực vật rừng. Để đạt được những mục đích này, lâm học cần phải thu thập rất nhiều số liệu khác nhau. Bởi vì rừng bao gồm nhiều thành phần khác nhau (khí hậu, địa hình, đất, cây gỗ, cây bụi, thảm cỏ, hệ động vật và hệ vi sinh vật...), nên lâm học phải sử dụng nhiều phương pháp hay công cụ khác nhau để thu thập và phân tích số liệu. Một số phương pháp do lâm học sáng tạo ra. Chẳng hạn: Phương pháp phân chia tầng thứ và độ tàn che của tán rừng; Phương pháp phân chia mức độ ưu thế của các loài cây gỗ trong quần thụ; Phương pháp phân cấp sinh trưởng của những cây gỗ hình thành rừng; Phương pháp mô tả cấu trúc rừng bằng các trắc đồ...Một số phương pháp khác thuộc các chuyên ngành khác (khí hậu – thủy văn, địa lý, đất, thực vật, điều tra rừng...). Vì thế, khi nghiên cứu về rừng, nhà lâm học cần phải hợp tác với những nhà khoa học thuộc nhiều chuyên ngành khác nhau.

Cách thức thu thập dữ liệu được xác định bởi mục tiêu nghiên cứu. Một thiết kế thu mẫu tốt sẽ làm giảm các chi phí nghiên cứu và tránh phá hoại đối tượng nghiên cứu. Chiến lược rút mẫu là rất quan trọng, bởi vì kết qủa nghiên cứu phụ thuộc không chỉ vào mục tiêu và phương pháp phân tích số liệu, mà còn vào số liệu được sử dụng. Số liệu đúng cho phép giải thích kết quả với độ tin cậy cao. Mẫu nhỏ

làm giảm khả năng phân tích và có thể dẫn đến khó khăn cho giải thích kết quả. Độ lệch trong mẫu dẫn đến kết luận sai. Mẫu quá lớn lại gây ra dư thừa số liệu, tốn thời gian thu thập và xử lý, chi phí cao về tài chính. Vì thế, giai đoạn thu mẫu có ý nghĩa quyết định trong nghiên cứu lâm học.

Rừng bao gồm rất nhiều đặc tính khác nhau và chúng có quan hệ với nhau theo những cách thức rất phức tạp. Để phân tích những đặc tính của cây gỗ và rừng, lâm học cần phải thu thập rất nhiều biến số khác nhau. Một số biến số được thu ở dạng số, một số biến khác ở dạng biến định danh có thứ bậc hơn kém và không có thứ bậc hơn kém. Vì thế, khi xử lý và phân tích số liệu, lâm học cần đến kiến thức toán học. Tùy theo cách thức thu thập số liệu, lâm học sử dụng những phương pháp phân tích số liệu khác nhau. Nếu dữ liệu được thu thập ở dạng số, lâm học có thể xử lý bằng phân tích tương quan và hồi quy. Nếu số liệu được thu thập ở dạng biến định danh có thứ bậc hơn kém và không có thứ bậc hơn kém, lâm học có thể xử lý bằng phương pháp phân tích tương quan có thứ bậc hơn kém và không có thứ bậc hơn kém. Một cách thức tóm tắt số liệu về các loài cây gỗ và các kiểu QXTV là sử dụng phân tích nhóm. Phân tích nhóm dựa trên ý tưởng cho rằng các kiểu QXTV tồn tại và mỗi QXTV có một tổ hợp loài đặc trưng. Phân tích nhóm sử dụng cho mục đích mô tả các kiểu QXTV trong mối quan hệ với các kiểu lập địa. Kết cấu của QXTV thay đổi rất lớn theo các kiểu lập địa, nhưng thay đổi rất nhỏ trong cùng một kiểu lập địa. Phân tích nhóm đặc biệt hữu dụng cho việc định nghĩa các kiểu rừng, kiểu QXTV, kiểu lập địa và các phương thức xử lý rừng. Phân tích định vị và phân tích nhóm thường được sử dụng trong giai đoạn điều tra sinh thái thực nghiệm. Kết quả của giai đoạn này là công cụ hữu ích để xác định phương pháp nghiên cứu mối quan hệ giữa cây gỗ và rừng với môi trường. Những mối liên hệ này có thể được làm rõ bằng phương pháp phân tích hồi quy và tương quan. Phân tích định vị và phân tích nhóm là pha nghiên cứu đầu tiên, còn phân tích hồi quy là pha nghiên cứu sau. Phân tích định vị và phân tích nhóm được sử dụng để phân tích biến động của loài cây gỗ và QXTV theo môi tường. Chẳng hạn: Phân tích sự có mặt hay vắng mặt của một số loài cây gỗ quan tâm theo biến động của địa hình và đất. Nhưng khi cần phân tích rõ quan hệ giữa các loài cây gỗ và QXTV với môi trường, lâm học sử dụng phương pháp phân tích hồi quy. Trong lâm học, phương pháp phân tích hồi quy và tương quan đa biến thường được sử dụng để phát hiện những biến phản hồi (Biến phụ thuộc = Y) của cây gỗ và rừng với sự thay đổi của các nhân tố sinh thái (Biến dự doán = X_i). Để đạt được mục đích này, lâm học cần phải thu thập rất nhiều số liệu về cây gỗ và rừng và các nhân tố sinh thái. Đây là một khó khăn không nhỏ trong các nghiên cứu lâm học.

1.4. Phương pháp thu thập số liệu trong lâm học

1.4.1. Một số thuật ngữ trong lâm học

Trong lâm học, người ta sử dụng các thuật ngữ theo những quy ước nhất định. Thuật ngữ "Mẫu" được sử dụng theo nghĩa chung là việc chọn lựa những đơn vị mẫu thống kê. Thuật ngữ "Vị trí thu mẫu" biểu thị những nơi thu thập mẫu thống kê hay đối tương nghiên cứu. Trong một quần thụ, nhà lâm học có thể chọn một số

vị trí đặc trưng về không gian và thời gian để thu thập mẫu. Từ quần thụ này, nhà lâm học mô tả kết cấu loài cây gỗ, đa dạng loài cây gỗ, cấu trúc quần thụ, tính không đồng nhất của quần thụ và những thay đổi của nó theo thời gian. Ví dụ: (1) Một khoảnh rừng là từ dùng để chỉ những nghiên cứu về thảm thực vật; (2) Ô mẫu được các nhà lâm học sử dụng trong nghiên cứu kết cấu loài cây gỗ, đa dạng loài cây gỗ, cấu trúc quần thụ, sản lượng rừng. Thuật ngữ "Bắt gặp" và "Không bắt gặp" được sử dụng để chỉ sự có mặt hay vắng mặt loài cây gỗ quan tâm nào đó trong QXTV rừng. Thuật ngữ "Phong phú" được sử dụng để biểu thị số lượng cá thể hay mật độ (N, cây), tiết diện ngang (G, m^2), thể tích (V, m^3), sinh khối (B, kg hoặc tấn) của loài cây quan tâm trong QXTV rừng. Thuật ngữ "Độ che tán" biểu thị tỷ lệ (% hay phần mười) diện tích mặt đất nằm ngang được loài cây gỗ che chiếu ánh sáng. Thuật ngữ "Tổng độ che tán" biểu thị tỷ lệ (% hay phần mười) diện tích mặt đất nằm ngang được các loài cây gỗ che chiếu ánh sáng. Thuật ngữ "Độ che phủ" biểu thị tỷ lệ (% hay phần mười) diện tích mặt đất nằm ngang được quần thụ, thảm cây bụi, cây cỏ che phủ hay xâm chiếm. Thuật ngữ "Nhóm sinh thái" biểu thị những loài cây gỗ sống trong cùng không gian và môi trường và chúng có mối liên hệ với nhau. Thuật ngữ "Sự kết nhóm sinh thái" biểu thị mối quan hệ giữa các loài cây gỗ với nhau về mặt sinh thái. Thuật ngữ "Dạng sống" biểu thị những loài cây có sự tương đồng về hình thái, về chức năng sinh lý và tập tính sinh học. Theo thuật ngữ này, lâm học phân biệt 4 dạng sống cơ bản trong các QXTV: cây thân gỗ; cây thân bụi; cây thân thảo; cây thân leo. Thuật ngữ "Tầng phiến" biểu thị những loài cây thuộc những đơn vị phân loại khác nhau cùng chung sống trong một không gian nhất định. Theo thuật ngữ này, lâm học phân biệt 4 tầng phiến cơ bản: tầng phiến cây gỗ, tầng phiến cây thân bụi, tầng phiến cây thân thảo và tầng phiến cây thân leo. Thuật ngữ "Kiểu rừng" biểu thị một nhóm thảm thực vật rừng được phân chia dựa theo những dấu hiệu trội của môi trường (khí hậu, đất, địa hình) và thảm thực vật như dạng sống ưu thế trong tầng ưu thế sinh thái, mức độ che tán của tầng ưu thế sinh thái (kín, thưa), hình thái lá (lá rộng, lá kim) và vật hậu của tán lá (thường xanh, rụng lá). Thuật ngữ "Kiểu QXTV rừng" biểu thị một nhóm QXTV có sự tương đồng về kết cấu loài cây ưu thế và đồng ưu thế, về năng suất hoặc điều kiện môi trường sống (vị trí địa lý, khí hậu, địa hình, đất...). Thuật ngữ "Lâm phần" biểu thị rừng thuần nhất về loài cây gỗ, cây bụi, cây thân thảo, hệ động vật và vi sinh vật, tiểu khí hậu, địa hình và đất. Tính thuần nhất của rừng chỉ xuất hiện trong không gian nhỏ. Vì thế, thuật ngữ này chỉ sử dụng cho rừng trồng thuần loài đồng tuổi. Thuật ngữ "Lập địa rừng" biểu thị một tập hợp những điều kiện môi trường sống của rừng (khí hậu, địa hình, đất, thảm thực vật, hoạt động lâm sinh...) ở một nơi nhất định. Một cách hiểu khác là một khu vực được đánh giá theo khả năng tạo ra sản lượng rừng. Thuật ngữ "Loài cây gỗ ưu thế" biểu thị loài cây gỗ kiểm soát sự sống của những loài cây gỗ khác hoặc loài cây gỗ ưu thế về sinh thái. Thuật ngữ "Loài cây gỗ đồng ưu thế" biểu thị loài cây gỗ có kích thước hay vai trò sinh thái nhỏ hơn so với loài cây gỗ ưu thế và cùng với loài cây gỗ ưu thế kiểm soát sự sống của những loài cây gỗ khác. Thuật ngữ "Thảm thực vật rừng" biểu thị tập hợp toàn bộ hệ thực vật của hệ sinh thái rừng trong một cảnh quan nhất định.

1.4.2. Các thang đo

Để dễ dàng thu thập và xử lý số liệu, phân tích và trình bày kết quả nghiên cứu, trước hết nhà lâm học cần phải định nghĩa rõ ràng những thuật ngữ được sử dụng trong thu thập số liệu về cây gỗ và rừng.

(1) Các quan sát. Thuật ngữ này biểu thị các đặc tính của cây gỗ và rừng được nghiên cứu. Chẳng hạn: Đường kính thân cây ngang ngực (D, cm); chiều cao toàn thân (H, m); tiết diện ngang thân (G, m^2); trữ lượng gỗ (M, m^3); tình trạng sinh lực của cây gỗ...

(2) Các biến. Thuật ngữ này biểu thị các thông tin hay đặc điểm của từng đặc tính của cây gỗ và rừng. Mỗi biến được đặt một tên biến cụ thể. Để mô tả tên đầy đủ của một biến, người ta sử dụng nhãn biến. Mỗi giá trị của một biến cũng có thể được đặt một nhãn giá trị. Ví dụ 1: Đường kính thân cây gỗ là một biến số. Mỗi biểu hiện của biến D (cm) được chỉ định bằng một giá trị biến như sau: 1 = D < 10; 2 = D = 10 – 20; 3 = D = 20 – 30; 4 = D = 30 – 40, 5 = D = 40 – 50; 6 = D = 50 – 60; 7 = D > 60 cm. Ví dụ 2: Độ che phủ của thảm cỏ là một biến số. Giả sử biến số này có tên là ĐCP với nhãn đầy đủ là "Độ che phủ của thảm cỏ". Mỗi biểu hiện của biến ĐCP có thể được chỉ định bằng một giá trị biến như sau: 1 = Soc, 2 = Cop_3, 3 = Cop_2, 4 = Cop_1, 5 = Sol, 6 = Sp và 7 = Un. Ví dụ 3: Độ thường gặp một loài cây gỗ nào đó là một biến số. Giả sử biến số này có tên là F (Frequency) với nhãn đầy đủ là "Độ thường gặp". Mỗi biểu hiện của biến F có thể được chỉ định bằng một giá trị biến như sau: 1 = bắt gặp, 0 = không bắt gặp. Việc quy định đặt tên biến, nhãn biến và nhãn giá trị như trên nhằm giúp cho việc xử lý số liệu bằng toán học.

(3) Các thang đo. Khi lập kế hoạch nghiên cứu, nhà lâm học phải quyết định các thang đo đối với các đặc tính của cây gỗ và rừng. Các thang đo không chỉ đóng vai trò to lớn đối với thu thập và xử lý số liệu, mà còn phản ánh độ tin cậy của những kết quả nghiên cứu. Trong lâm học, người ta thường sử dụng 4 thang đo: định danh, thứ bậc, khoảng và tỷ lệ. Thang đo định danh biểu thị cho những biến số định tính. Thang đo thứ bậc và thang đo khoảng biểu thị cho những biến số định lượng.

(3.1) Thang đo định danh. Thang đo định danh biểu thị cho những biến mà các giá trị không có liên hệ với nhau. Những trị số của biến định danh chỉ thể hiện các nhóm phân loại không có thứ bậc hơn kém. Những giá trị đo được biểu thị bằng cấp. Mỗi cấp có thể được đánh số bằng số nguyên hay gán chuỗi dạng ngắn gọn cho những biểu hiện của một biến số. Ví dụ: Kiểu đất được chia thành nhiều cấp: đất cát, đất thịt, đất sét. Các giá trị được chỉ định vào các cấp này là 10, 11 và 17 hoặc X_1, X_2 và X_3. Bởi vì chúng không sắp xếp theo thứ bậc, nên chúng không có quan hệ với nhau. Các giá trị này chỉ biểu thị cho tên cấp. Vì thế, kiểu biến này không thể xử lý bằng các phép tính đại số.

(3.2) Thang đo thứ bậc. Đây là một thang đo định danh nhưng các giá trị của biến số có thứ bậc hơn kém. Các biến định danh có thứ bậc hơn kém còn được gọi là các biến phân loại (Classification Variables), các biến lập nhóm (Discriminant,

Grouping Variables) và các biến mã hóa (Coding Variables). Các biến định danh có thứ bậc hơn kém được sử dụng để phân loại các đối tượng khác nhau. Ví dụ 1: Một loại bệnh cây nào đó có thể được phân chia theo 5 cấp: 1 = bệnh rất nặng, 2 = bệnh nặng, 3 = bệnh trung bình, 4 = bệnh nhẹ và 5 = không có bệnh. Ví dụ 2: Khi nghiên cứu quan hệ giữa năng suất rừng trồng với 5 chỉ số lập địa (SI), thì các biểu hiện của biến SI có thể được ghi bằng 5 trị số 12, 14, 16, 18 và 20 m. Trong lâm học, chúng ta thường gặp rất nhiều biến định danh có thứ bậc hơn kém như cấp độ che tán, cấp năng suất, cấp sinh trưởng cây rừng, cấp che phủ của thảm cỏ, cấp bệnh cây, cấp cháy rừng…Lưu ý rằng, trong thang đo thứ bậc hơn kém, những con số có trị số cao hơn không có nghĩa là chúng luôn ở bậc cao hơn và ngược lại. Bảng 1.1 dẫn ví dụ phân chia 7 cấp độ phong phú của các loài cây thân thảo sống dưới tán rừng theo phương pháp của Druze. Bởi vì khoảng cách giữa 7 cấp độ che phủ của thảm cỏ là không giống nhau, nên chúng ta không thể tính toán bằng các phép tính đại số. Việc gán hay mã hóa những con số cho mỗi cấp độ che phủ của thảm cỏ chỉ mang ý nghĩa quy ước nhằm tạo thuận lợi cho việc ghi chép và xử lý số liệu.

Bảng 1.1. Phân cấp độ phong phú của các loài cây thân thảo sống dưới tán rừng theo phương pháp của Druze.

Thứ tự	Cấp che phủ của Druze	Phân chia theo hệ thống 7 cấp:		
		Mã hóa	Tiêu chuẩn đánh giá theo:	
			Số cá thể	Độ che phủ (%)
(1)	(2)	(3)	(4)	(5)
1	Soc(*)	6	Rất nhiều	76 - 100
2	Cop_3	5	Nhiều	51 - 75
3	Cop_2	4	Tương đối nhiều	21 - 50
4	Cop_1	3	Tương đối	5 - 20
5	Sp	2	Ít	1 - 4
6	Sol	1	Rất ít	< 1
7	Un	0	Chỉ một cá thể	< 0,2

(*) Các ký hiệu viết tắt từ chữ La tinh: *Socialis = Soc, Copiosus = Cop, Spars = Sp, Unicus = Un.*

(3.3) Thang đo khoảng. Đó là thang đo thứ bậc có khoảng cách đều nhau. Thang đo khoảng được chia theo một đơn vị ổn định. Do đó, sự khác biệt giữa giá trị này với giá trị khác có thể xác định được. Ví dụ: Chỉ tiêu D được phân chia theo khoảng cách 2 cm kèm theo tần số cây tương ứng. Từ số số liệu này, chúng ta có thể xác định được những đặc trưng thống kê mô tả (Trung bình, Min, Max, sai tiêu chuẩn = SEE và hệ số biến động = CV%).

(3.4) Thang đo tỷ lệ. Đó là thang đo khoảng với một điển Zero lấy làm gốc để so sánh. Ví dụ: Đo sự khác biệt giữa hai kiểu rừng bằng việc sử dụng G hay M để so sánh. Độ phong phú của loài cây gỗ thường được đo bằng số đo tỷ lệ với đơn vị đo là phần trăm hoặc phần mười.

Ngoài cách phân biệt các biến và thang đo trên đây, trong lâm học người ta còn phân chia các biến thành biến định tính và biến định lượng. Đặc tính định lượng

là những đặc tính có thể đếm hay cân đo được bằng một đơn vị vật lý nào đó. Ngược lại, đặc tính định tính là những đặc tính không thể đo đếm bằng một đơn vị vật lý nào đó.

(a) Đặc tính định tính. Trong lâm học, những dấu hiệu định tính có rất nhiều và chúng đem lại ý nghĩa rất lớn nhằm thuyết minh cho vấn đề nghiên cứu. Chẳng hạn: Khi mô tả QXTV, nhà lâm học có thể phải thu thập những đặc tính (Biến số) mà không thể đo đạc được như: (1) Màu sắc của lá, hoa và quả; (2) Độ bắt gặp của loài cây quan tâm; (3) Tình trạng sinh lực của cây gỗ; (4) Cây bị bệnh hay không bị bệnh; (5) Cây đã chết hay còn sống...Ngay cả một số đặc tính định lượng, vì lý do khó đo đếm, nhà lâm học cũng có thể xem chúng như những đặc tính định tính. Ví dụ: Số lượng hoa quả trên một cây gỗ; Mức độ rậm rạp của hệ thống lá; Độ che phủ của thảm cỏ trên mặt đất...

Để dễ dàng xử lý và so sánh các biến số định tính, nhà lâm học có thể ghi nhận bên cạnh chúng tần số bắt gặp. Tổng tần số của các sự kiện phải bằng tổng số đơn vị quan sát. Ví dụ: Nếu a, b và c là tần số của 3 sự kiện nào đó và N là tổng số đơn vị quan sát, thì (a + b + c) = N, còn a/N, b/N và c/N là tần suất của 3 sự kiện. Tần suất của mỗi nhóm đối tượng có thể biểu thị bằng đơn vị phần mười hoặc phần trăm. Trong trường hợp tần suất được biểu thị bằng đơn vị phần mười, thì tổng các tần suất của các nhóm phải bằng 1, nghĩa là (a/N + b/N +c/N) = 1. Ngược lại, nếu tần suất được biểu thị bằng đơn vị phần trăm, thì tổng các tần suất của các nhóm phải bằng 100% (Bảng 1.2). Tùy theo mục đích nghiên cứu, những đặc tính định tính được ghi vào bảng phân phối theo hai cách. Cách thứ nhất là bảng phân phối một chiều (Bảng 1.2). Cách thứ hai là bảng phân phối hai chiều (Bảng 1.3). Từ đó, nhà lâm học có thể tiến hành những xử lý và so sánh kết quả.

Bảng 1.2. Phân phối kết quả điều trị bệnh phấn trắng trên 200 cây Keo lai.

Kết quả điều trị	Tần số	Tần suất	Tỷ lệ (%)
(1)	(2)	(3)	(4)
Bệnh khỏi	105	0,525	52,5
Bệnh giảm	55	0,275	27,5
Tử vong	40	0,200	20,0
Tổng số	200	1,000	100,0

(i) Đặc tính định lượng. Các đặc tính định lượng bao gồm các đặc tính định lượng rời rạc và các đặc tính định lượng liên tục. Những đặc tính định lượng rời rạc là những đại lượng của một đặc tính được biểu thị bằng những số nguyên. Chẳng hạn: (a) Số cây trên một ô mẫu có kích thước nhất định; (b) Số cây gỗ mang bệnh; (c) Số cây tái sinh trên một ô mẫu 10 m^2. Trong bảng thống kê, chúng ta ghi trong một cột các trị số của đặc tính, còn cột bên cạnh ghi các tần số, tần suất, tỷ lệ (%) tương ứng với chúng. Ngoài ra, để phân tích chi tiết số liệu, chúng ta có thể tính các tần số tích lũy, tần suất tích lũy và tỷ lệ tích lũy bằng cách cộng dồn tần số, tần suất và tỷ lệ của nhóm trên với các nhóm dưới (Bảng 1.4).

Bảng 1.3. Phân phối kết quả điều trị bệnh phấn trắng theo 3 loại thuốc khác nhau trên 205 cây Keo lai.

Kết quả điều trị	Tần số cây theo 3 trị liệu:			Tổng số
	A	B	C	
(1)	(2)	(3)	(4)	(5)
Bệnh khỏi	35	38	24	97
Bệnh giảm	21	20	17	58
Tử vong	12	16	22	50
Tổng	68	74	63	205

Bảng 1.4. Phân bố số cây theo cấp đường kính đối với ưu hợp Dầu con rái (*Dipterocarpus alatus* Roxb). Nguồn: Đào Thị Thùy Dương, 2019.

TT	Cấp D (cm)	N (cây/ha)	N%	$N_{(Tích\ lũy)}$	$N\%_{(Tích\ lũy)}$
(1)	(2)	(3)	(4)	(5)	(6)
1	≤ 10	300	45,1	300	45,1
2	16	132	19,9	432	65,0
3	22	65	9,7	497	74,7
4	28	38	5,7	535	80,4
5	34	27	4,0	562	84,4
6	40	22	3,4	584	87,8
7	46	21	3,1	605	90,9
8	52	21	3,2	626	94,1
9	58	20	3,0	645	97,1
10	≥64	20	2,9	665	100,0
	Tổng số	665	100		

Đặc tính định lượng liên tục là những đặc tính định lượng lấy bất kỳ giá trị nào nằm trong một khoảng nào đó. Chẳng hạn: D, H, g, V của cây gỗ, G và M của QXTV rừng. Thế nhưng, do những giới hạn của các công cụ đo đạc, một số đặc tính định lượng liên tục vẫn chỉ được ghi nhận bằng các giá trị không liên tục. Chẳng hạn: Khi sử dụng thước đo chiều cao với độ chính xác 0,5 m, thì H của những cây gỗ từ 20 đến 22 m chỉ nhận được 5 giá trị 20,0; 20,5; 21,0, 21,5 và 22,0 m. Khi trình bày các số liệu định lượng liên tục dưới dạng bảng, trước hết phân chia các đặc tính thành các nhóm (tổ hay cấp) theo một cự ly (khoảng cách) nhất định. Sau đó xác định tần số, tần suất và tỷ lệ của các nhóm. Theo Brooks và Carruther, số nhóm (m) và cự ly giữa các nhóm (k) được tính tương ứng theo công thức 1.1 và 1.2; trong đó N là dung lượng quan sát, X_{Max} và X_{Min} là những trị số lớn nhất và nhỏ nhất của dãy số liệu quan sát.

$$m \geq 5*LogN \qquad (1.1)$$

$$k = \frac{(X_{Max} - X_{Min})}{m} \qquad (1.2)$$

Hoppman đề nghị tính cự ly giữa các nhóm (tổ) theo công thức 1.3.

$$d = \frac{(X_{Max} - X_{Min})}{2\sqrt[3]{N}} \qquad (1.3)$$

Lưu ý rằng, mặc dù số tổ và cự ly giữa các tổ được tính theo công thức 1.1 - 1.3 là khách quan, nhưng thực tế có thể nhận được kết quả kém ý nghĩa. Thật vậy, cự ly của mỗi cấp D thân cây gỗ được tính theo công thức 1.2 và 1.3 là 0,2 – 0,5 m, nhưng thực tế nhà lâm học đã sử dụng thước đo chiều cao với độ chính xác 1,0 m. Do đó, theo thước đo này nhà lâm học không thể tính được chính xác tần số tương ứng với từng cấp D. Vì vậy, tùy theo đối tượng và mục tiêu nghiên cứu, trong thực hành người ta nhận thấy số tổ dao động từ 6 - 12, còn cự ly giữa các tổ được xác định theo độ chính xác của các dụng cụ đo đạc. Nói chung, đối với rừng trồng thuần loài đồng tuổi, tùy theo tuổi (A, năm) và kích thước cây gỗ, mỗi cấp D được phân chia từ 2 – 4 cm, còn mỗi cấp H là 1 - 2 m. Đối với rừng tự nhiên hỗn loài, do D và H có biên độ rất lớn, nên mỗi cấp D được phân chia từ 4 – 8 cm, còn mỗi cấp H là 2 – 4 m tùy theo loại rừng. Khi các biến định lượng liên tục được phân chia theo tổ, thì chúng có thể được tập hợp thành các bảng số liệu. Nội dung của bảng số liệu bao gồm giá trị trung tâm hoặc ranh giới của mỗi tổ; tần số, tần suất và tỷ lệ của mỗi tổ (Bảng 1.4).

1.4.3. Chiến lược thu mẫu trong lâm học

Chiến lược thu mẫu là một khung thiết kế nghiên cứu. Để những dấu hiệu mô tả thảm thực vật có thể đưa vào xử lý thống kê, nhà lâm học cần phải thu thập chúng theo một phương pháp nhất định. Phương pháp thu thập dữ liệu phải được làm rõ trước các mục đích đặt ra. Mỗi thảm thực vật có nhiều QXTV và một QXTV cũng có nhiều đặc tính khác nhau. Việc mô tả tất cả những đặc tính của QXTV trên một lãnh thổ rộng lớn là hết sức khó khăn, tốn nhiều thời gian, nhân lực và kinh phí. Thay vì thế, người ta chỉ mô tả những đặc tính cơ bản nhất trên một bộ phận đại diện cho đối tượng nghiên cứu; sau đó suy diễn những đặc tính này cho toàn bộ các QXTV. Phương pháp nghiên cứu mà chỉ thực hiện trên một bộ phận không lớn của đối tượng và từ kết quả nhận được rút ra nhận định chung cho toàn bộ đối tượng nghiên cứu được gọi là phương pháp nghiên cứu mẫu. Tất cả đối tượng nghiên cứu được gọi là tổng thể, còn mẫu là một bộ phận của tổng thể được rút ra để nghiên cứu. Nhiệm vụ cơ bản của xử lý thống kê là trên cơ sở nghiên cứu các mẫu đưa ra những nhận định đúng cho tổng thể.

Khi nghiên cứu một thảm thực vật nào đó, tổng thể là toàn bộ các QXTV của thảm thực vật này, còn mẫu là tập hợp những ô mẫu nằm trong giới hạn của tổng thể đó. Các ô mẫu được quy ước theo những tiêu chuẩn nào đó được gọi là các ô tiêu chuẩn. Ô mẫu là một đơn vị thu mẫu được quy ước thống nhất trong nghiên cứu. Nếu đối tượng nghiên cứu là toàn bộ thảm thực vật của một lãnh thổ nào đó, thì tổng thể là toàn bộ các QXTV của lãnh thổ ấy, còn mẫu là tập hợp các QXTV trên các ô mẫu. Trong những nghiên cứu QXTV rừng, nhà lâm học thường sử dụng một chuỗi ô mẫu có hình dạng và kích thước nhất định. Để nhận được kết quả khách

quan và phản ánh đúng đối tượng nghiên cứu với độ chính xác mong muốn, thì mẫu đo đếm phải thỏa mãn một số điều kiện dưới đây.

a. Mẫu đo đếm là một đại diện tốt cho tổng thể. Điều này có nghĩa là trong tập hợp các phần tử của mẫu phải chứa tất cả những biến số của tổng thể cùng với những mối liên hệ định lượng giữa chúng.

b. Mỗi bộ phận của tổng thể phải có xác suất được chọn như nhau. Điều này có nghĩa là mỗi khoảnh rừng phải có xác suất được chọn để đo đếm như nhau.

Một thảm thực vật rừng bao gồm nhiều thành phần cấu thành. Vì thế, mỗi thành phần phải được rút theo tỷ lệ mà nó có trong đối tượng nghiên cứu. Trong một thảm thực vật rừng, diện tích của mỗi QXTV thường có biến động lớn nhất, do đó các nhà lâm học thường xác định tỷ lệ rút mẫu theo diện tích của QXTV. Điều này đặc biệt hữu ích khi đánh giá trữ lượng gỗ của mỗi QXTV và tính năng suất bình quân cho thảm thực vật trên một đơn vị diện tích. Để nhận được các mẫu đại diện tốt cho đối tượng nghiên cứu, lâm học có thể sử dụng 5 phương pháp bố trí ô mẫu.

1.4.3.1. Phương pháp chọn mẫu ngẫu nhiên

Thực chất của bố trí mẫu ngẫu nhiên là vị trí của mỗi ô mẫu là độc lập với những ô mẫu còn lại. Chọn mẫu ngẫu nhiên là phân chia ngẫu nhiên các nghiệm thức vào những đơn vị thí nghiệm hoặc chọn những mẫu ngẫu nhiên từ những đơn vị mẫu. Đây là phương pháp thu mẫu khách quan hơn so với những phương pháp thu mẫu không ngẫu nhiên. Thu mẫu ngẫu nhiên là một công cụ nghiên cứu hữu ích để đưa ra những kết luận với độ tin cậy mong muốn. Bởi vì thu mẫu ngẫu nhiên loại trừ được sai số hệ thống và cung cấp cơ sở cho việc áp dụng lý thuyết thống kê (kiểm định và ước lượng).

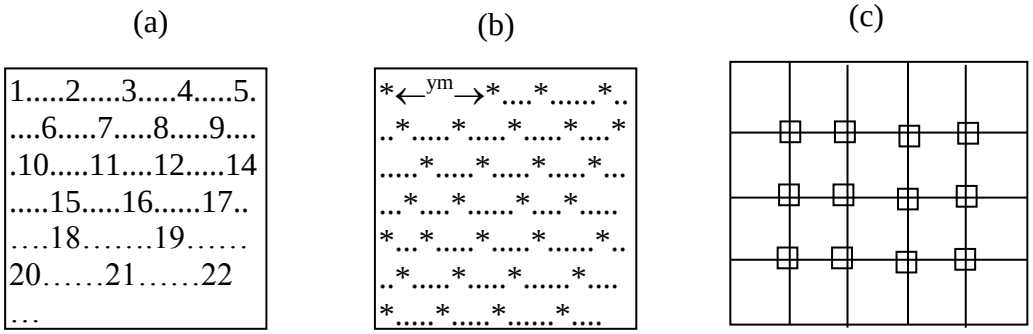
Trong thực hành, chọn mẫu ngẫu nhiên có thể được thực hiện bằng nhiều cách khác nhau. Giả thiết nhà lâm học sử dụng ô mẫu có kích thước s = 20*20 m (400 m^2), diện tích cần phải đo đếm là S ha, tổng diện tích khu vực nghiên cứu là W ha. Từ đó tính được dung lượng tổng thể là N = W/s; dung lượng mẫu (số ô mẫu) cần đo đếm là n = S/s và tỷ lệ rút mẫu là S/W hay n/N. Trên bản đồ rừng, chúng ta phân chia khu vực nghiên cứu thành hệ thống ô mẫu liên tiếp nhau theo dạng bàn cờ, mỗi ô mẫu có diện tích là s m^2. Sau đó đánh số thứ tự các ô mẫu và dùng bảng số ngẫu nhiên hoặc rút thăm để xác định những ô mẫu cần đo đếm. Khi chọn những ô mẫu theo phương pháp này, nhà lâm học có thể gặp trường hợp một số ô mẫu được chọn lặp lại 2, 3...lần. Về nguyên tắc chọn mẫu, chúng ta có mẫu đo lặp. Tuy vậy, để tăng tính đại diện của mẫu, nhà lâm học có thể bỏ qua những ô mẫu được chọn lặp lần 2, 3...Phương pháp này có ưu điểm lớn là mẫu đảm bảo hoàn toàn ngẫu nhiên. Mặc dù vậy, nó ít được sử dụng trong các nghiên cứu lâm học, bởi vì rừng là một tổng thể tự nhiên hết sức phức tạp, diện tích lớn, khó đi lại. Mặt khác, một số ô nghiên cứu được chọn có thể rơi vào những vị trí khó đến được. Chẳng hạn: Khe núi, đỉnh núi. Một số ô mẫu được chọn không đúng mục tiêu nghiên cứu. Chẳng

hạn: Nghiên cứu đặc tính của một quần thể loài cây gỗ trong một kiểu QXTV nhất định, nhưng một số ô mẫu không chứa quần thể loài này.

1.4.3.2. Phương pháp chọn mẫu hệ thống

Thực chất của phương pháp này là vị trí của các ô mẫu được xác định theo một nguyên tắc định trước. Giả thiết nhà lâm học cần nghiên cứu một QXTV rừng nào đó. Những thông tin cần phải thu thập bao gồm thành phần loài cây gỗ (S), N, G, M. Theo mục đích đặt ra, nhà lâm học cần thu thập một số ô mẫu có diện tích bằng nhau và chúng được bố trí theo một nguyên tắc định trước. Để thu thập được những ô mẫu đại diện cho tổng thể, nhà lâm học có thể bố trí chúng theo một trong những cách dưới đây.

(a) Bố trí ô mẫu hệ thống theo những số định trước. Trước hết, phân chia khu vực nghiên cứu thành những ô mẫu với kích thước bằng nhau và phân bố liên tục theo hệ thống ô bàn cờ. Mỗi ô được đánh số thứ tự từ trái sang phải và từ trên xuống dưới. Sau đó, căn cứ vào dung lượng mẫu cần quan sát, nhà lâm học quy ước cách rút mẫu hệ thống. Chẳng hạn: Thứ tự ô mẫu được rút ra nghiên cứu bắt đầu từ ô mẫu số 1, tiếp theo là ô mẫu số 3, 6...k (Hình 1.1a).



Hình 1.1. Sơ đồ bố trí ô mẫu theo phương pháp hệ thống. (a) Bố trí ô mẫu theo những số định trước. (b) Bố trí ô mẫu theo một cự ly nhất định. (c) Bố trí ô mẫu theo mắt lưới.

(b) Bố trí ô mẫu hệ thống trên các tuyến theo một cự ly nhất định. Để đảm bảo các ô mẫu phân bố đồng đều trên toàn bộ đối tượng nghiên cứu, nhà lâm học có thể phân chia khu vực nghiên vứu thành những tuyến song song và cách đều nhau. Khi biết tổng chiều dài các tuyến là S (m), dung lượng quan sát là k ô mẫu, thì cự ly giữa các ô mẫu là Y = S/k (m). Tiếp theo, nhà lâm học phân chia toàn bộ chiều dài của tất cả tuyến thành Y phân đoạn bằng nhau. Sau đó, quy ước ô mẫu đầu tiên được rút ra đo đếm bắt đầu từ một điểm nào đó trên tuyến thứ nhất. Theo đó, các ô mẫu thứ 2, 3...k được bố trí cách nhau một khoảng cách bằng Y m (Hình 1.1b).

(c) Bố trí ô mẫu hệ thống theo mắt lưới. Trước hết, phân chia đối tượng nghiên cứu thành những ô mẫu với kích thước bằng nhau theo dạng bàn cờ. Sau đó, căn cứ vào số lượng ô mẫu cần đo đếm, nhà lâm học thực hiện rút mẫu bằng cách

chọn các ô mẫu tại điểm giao nhau của các đường phân định ranh giới giữa các ô mẫu (Hình 1.1c).

So với phương pháp chọn mẫu ngẫu nhiên, phương pháp chọn mẫu hệ thống có ưu điểm là đơn giản, dễ thực hiện và các ô mẫu được chọn phân bố đồng đều trên đối tượng nghiên cứu...Nhược điểm của phương pháp chọn mẫu hệ thống là dễ bị ảnh hưởng do nhân tố chủ quan, đồng thời khó đánh giá độ chính xác của kết quả nghiên cứu. Tuy vậy, phương pháp chọn mẫu hệ thống vẫn thường được áp dụng trong những nghiên cứu thảm thực vật rừng.

1.4.3.3. Phương pháp chọn mẫu ngẫu nhiên - hệ thống

Để khắc phục những nhược điểm của phương pháp chọn mẫu ngẫu nhiên và phương pháp chọn mẫu hệ thống, nhà lâm học có thể áp dụng phương pháp chọn mẫu phối hợp giữa phương pháp chọn mẫu ngẫu nhiên và phương pháp chọn mẫu hệ thống. Trong trường hợp này, ô mẫu đầu tiên được chọn theo phương pháp ngẫu nhiên, còn các ô mẫu khác được chọn theo phương pháp hệ thống. Khi nghiên cứu trên những thảm thực vật rừng rộng lớn và phức tạp, phương pháp này cũng ít được áp dụng.

1.4.3.4. Phương pháp chọn mẫu theo khối

Các thảm thực vật rừng là những hệ sinh thái vô cùng phức tạp và chúng phân bố không đồng nhất theo không gian và thời gian. Ngoài ra, một thảm thực vật rừng có thể có nhiều kiểu rừng và mỗi kiểu rừng có nhiều kiểu QXTV với những đặc tính khác nhau. Vì thế, để thu được kết quả chính xác, nhà lâm học phải phân chia đối tượng nghiên cứu thành những khối riêng biệt. Mỗi khối có thể bao gồm một đơn vị phân loại rừng nhất định, một nhóm QXTV nhất định hoặc một giai đoạn diễn thế nhất định. Một đối tượng nghiên cứu được phân chia thành từng khối thuần nhất như thế sẽ giúp ích nhiều hơn trong nghiên cứu thảm thực vật rừng. Sau khi chia đối tượng nghiên cứu thành những khối thuần nhất, việc rút mẫu được thực hiện theo từng khối tương ứng với đặc điểm riêng của mỗi khối. Tùy theo diện tích của mỗi khối, dung lượng mẫu được rút theo một tỷ lệ nhất định. Cách tiến hành như vậy được gọi là chọn mẫu theo khối. Phương pháp này có những ưu điểm cơ bản sau đây:

a. Nó cho phép không chỉ phân tích rõ đặc điểm của từng đơn vị thảm thực vật rừng, mà còn so sánh và phân loại thảm thực vật rừng.
b. Nó cho phép xác định và đánh giá tần số bắt gặp loài cây gỗ, kiểu QXTV rừng và kiểu rừng.
c. Nó cho phép thực hiện những phân tích thống kê và suy diễn cho tổng thể.
d. Từ những thông tin về loài cây gỗ, kiểu QXTV và kiểu rừng, nhà lâm học có thể xây dựng các phương thức lâm sinh...

Phương pháp chọn mẫu theo khối cũng có một số hạn chế. Chẳng hạn: Các khối có thể cho những kết quả sai khác nhau rất lớn. Để giảm bớt sự sai khác này, nhà lâm học cần phải phân chia đối tượng nghiên cứu chi tiết hơn hoặc phải tăng khối lượng mẫu đo đếm. Điều đó lại dẫn đến làm tăng các chi phí về thời gian, nhân

lực và kinh phí...Để giảm tính phức tạp của vấn đề, nhà lâm học có thể chỉ thực hiện nghiên cứu trên những đối tượng phổ biến nhất và cũng chỉ thu thập những đặc tính cơ bản nhất của thảm thực vật rừng.

1.4.3.5. Phương pháp chọn mẫu điển hình

Về bản chất, mẫu điển hình là mẫu phản ánh mức trung bình của đối tượng nghiên cứu. Nếu chỉ xác định mức trung bình của đối tượng nghiên cứu, chúng ta có thể sử dụng phương pháp thu mẫu ngẫu nhiên hoặc phương pháp thu mẫu hệ thống. Tuy vậy, bởi vì hai phương pháp này có nhiều hạn chế, nên nhà lâm học thường sử dụng phương pháp thu mẫu điển hình. Mẫu điển hình là mẫu đại diện cho đối tượng nghiên cứu. Thế nhưng, thảm thực vật rừng là hệ sinh thái rất phức tạp và phân bố trong không gian rất rộng. Vì thế, nhà lâm học rất khó xác định những mẫu điển hình cho đối tượng nghiên cứu. Đối với một mẫu nhất định, nhà lâm học này cho là mẫu điển hình, còn nhà lâm học khác lại không công nhận điều đó. Chính điều này đã dẫn đến những tranh luận về kết quả nghiên cứu. Nói chung, phương pháp thu mẫu điển hình phụ thuộc vào chủ quan của nhà lâm học. Để hạn chế những khó khăn trong quá trình nghiên cứu và nhận được kết quả tốt, nhà lâm học cần thu thập nhiều mẫu khác nhau và sau đó tính bình quân cho đối tượng nghiên cứu. Về lý thuyết, số lượng mẫu càng nhiều, thì kết quả nhận được càng chính xác. Thế nhưng điều vừa nói lại dẫn đến những khó khăn trong quá trình nghiên cứu và phức tạp trong xử lý và báo cáo kết quả. Vì thế, khi nghiên cứu một QXTV rừng, nhà lâm học chỉ thu thập một số ô mẫu điển hình. Số lượng ô mẫu thường dao động từ 3 – 5.

Nói chung, về nguyên tắc, các nghiên cứu trong lâm học phải làm sao thu được những mẫu đại diện cao nhất cho đối tượng nghiên cứu với chi phí ít nhất về thời gian, nhân lực và kinh phí. Những nghiên cứu trong lâm học luôn hướng đến những mục tiêu nhất định. Chẳng hạn: (a) Xác định những đặc tính sinh thái của quần thể Dầu song nàng (*Dipterocarpys dyeri*) trong kiểu rừng kín thường xanh ẩm nhiệt đới ở tỉnh Đồng Nai; (b) Xác định những đặc tính sinh thái của ưu hợp Dầu song nàng trong kiểu rừng kín thường xanh ẩm nhiệt đới ở tỉnh Đồng Nai. Để đạt được hai mục tiêu nghiên cứu kể trên, chúng ta không thể áp dụng máy móc một phương pháp thu mẫu. Trong cả hai trường hợp này, chúng ta có thể áp dụng phương pháp thu mẫu hệ thống kết hợp với phương pháp thu mẫu ngẫu nhiên hoặc phương pháp thu mẫu theo khối và phương pháp thu mẫu điển hình. Phương pháp thu mẫu hệ thống được áp dụng để xác định phạm vi không gian bắt gặp quần thể loài và QXTV quan tâm. Từ phạm vi không gian này, chúng ta phân chia thành các khối; trong đó mỗi khối bao gồm một số đặc tính nào đó. Sau đó áp dụng phương pháp thu mẫu ngẫu nhiên hoặc phương pháp thu mẫu điển hình để thu thập những đặc tính quan tâm đối với từng mục tiêu nghiên cứu.

1.5. Lập kế hoạch thu thập số liệu

1.5.1. Thiết kế nghiên cứu

Sau khi đã xác định rõ mục tiêu nghiên cứu, nhà lâm học cần phải xây dựng sơ đồ tóm tắt quá trình nghiên cứu từ lúc bắt đầu cho tới khi kết thúc. Sơ đồ nghiên

cứu còn gọi là thiết kế thí nghiệm hay thiết kế nghiên cứu. Sơ đồ nghiên cứu cho biết tất cả các chuỗi hoạt động trong quá trình nghiên cứu. Sơ đồ nghiên cứu bao gồm 10 vấn đề chủ yếu sau đây.

Một là các câu hỏi và các giả thuyết nghiên cứu. Căn cứ vào mục tiêu nghiên cứu, điều kiện nghiên cứu (thời gian, nhân lực, kinh phí) và khả năng hiểu biết của mình về vấn đề nghiên cứu, nhà lâm học phải xác định rõ các câu hỏi kèm theo các giả thuyết nghiên cứu. Đây là bước quan trọng bậc nhất của bất kỳ nghiên cứu nào. Chẳng hạn: (1) Tổng thể nào được nghiên cứu? (2) Những câu hỏi nào cần được trả lời? (3) Những giả thuyết nào cần được kiểm định? (4) Những tham số nào cần được ước lượng? Hai là xác định chính xác số đối tượng nghiên cứu (đơn vị thí nghiệm; đơn vị thu mẫu). Ba là xác định các biến phản hồi (Biến phụ thuộc), các biến giải thích (Biến độc lập) và các nghiệm thức (Nhân tố và các mức của nhân tố). Bốn là thiết kế thủ tục phân phối (ngẫu nhiên, hệ thống…) các nghiệm thức vào các đơn vị thí nghiệm hoặc đơn vị thu mẫu. Năm là xác định phương pháp thu thập các đặc trưng của cây gỗ và rừng, phương pháp thu mẫu (ngẫu nhiên, hệ thống…) và tỷ lệ mẫu. Sáu là xác định độ chính xác mong muốn của các phép đo và tính đúng đắn của phương pháp thu thập mẫu. Bảy là xác định phương pháp toán thống kê sẽ áp dụng để xử lý số liệu. Tám là lập sơ đồ phân tích và cách thức mô tả kết quả nghiên cứu (bảng, biểu đồ, đồ thị). Chín là đánh giá các mục tiêu và tiêu chuẩn để làm rõ những hoạt động trên đây. Mười là xây dựng kế hoạch về thời gian, nhân lực và kinh phí nghiên cứu. Lưu ý rằng việc sắp xếp thứ tự 10 vấn đề trên đây chỉ là tương đối. Thật vậy, mặc dù các câu hỏi và các giả thuyết nghiên cứu đóng vai đóng vai trò quan trọng nhất, nhưng chúng chỉ được làm rõ một khi nhà nghiên cứu có đầy đủ điều kiện nghiên cứu (thời gian, nhân lực, kinh phí) và kiến thức về vấn đề nghiên cứu. Mặc khác, phương pháp bố trí thí nghiệm, phương pháp thu thập các biến phản hồi và các biến giải thích phụ thuộc vào phương pháp xử lý số liệu. Nếu áp dụng toán thống kê để xử lý và phân tích số liệu, thì nhà lâm học cần phải hiểu rõ những điều kiện của từng phương pháp xử lý số liệu. Vì thế, để xây dựng được sơ đồ nghiên cứu hoàn chỉnh và khả thi, trước hết nhà lâm học có thể bắt đầu từ phân tích điều kiện nghiên cứu (thời gian, nhân lực, kinh phí) và kiến thức về vấn đề nghiên cứu. Kế đến chọn những nội dung và phương pháp nghiên cứu thích hợp. Sau đó đặt ra những câu hỏi và những giả thuyết nghiên cứu.

1.5.2. Xác định kiểu nghiên cứu

Sau khi thiết kế sơ đồ nghiên cứu, một nhiệm vụ quan trọng khác là xác định kiểu nghiên cứu. Theo mục đích nghiên cứu, người ta phân chia các nghiên cứu dựa theo đặc tính của chúng. Cách phân loại này là rất quan trọng, bởi vì nó chỉ ra cho nhà lâm học biết được kiểu kết luận có thể rút ra từ việc áp dụng những kỹ thuật thống kê đối với số liệu quan tâm. Box và Snell (1981; Dẫn theo Jongman và ctv, 1987), các nghiên cứu được phân chia thành 5 kiểu.

Kiểu thứ 1 là những nghiên cứu thực nghiệm. Nhà nghiên cứu xây dựng và kiểm tra các giả thuyết nghiên cứu và xác định ảnh hưởng của những nhân tố thí nghiệm (Biến độc lập) đến biến phản hồi (Biến phụ thuộc) bằng thực nghiệm.

Kiểu thứ 2 là nghiên cứu thu thập số liệu thuần túy. Bằng cách nghiên cứu này, nhà nghiên cứu có thể khám phá và xác định được mối liên hệ giữa biến phụ thuộc và các biến độc lập. Tuy vậy, nhà nghiên cứu không thể kiểm tra được việc thu thập số liệu và không thể trả lời rõ các câu hỏi kèm theo độ tin cậy của kết quả nghiên cứu.

Kiểu thứ 3 là những điều tra mẫu. Bằng việc rút mẫu từ tổng thể, nhà nghiên cứu có thể thu được một kết quả tốt và xác định được sai số của kết quả nghiên cứu. Tuy vậy, cũng giống như nghiên cứu thu thập số liệu thuần túy, nhà nghiên cứu rất khó đưa ra lời giải thích rõ ràng cho mối quan hệ giữa các đặc tính của biến phụ thuộc và các biến độc lập. Mặt khác, nếu sử dụng mẫu nhỏ, thì các kết luận rút ra từ điều tra mẫu có thể phản ánh sai lệch đối tượng nghiên cứu.

Kiểu thứ 4 là những nghiên cứu có khả năng giải thích. Nhà nghiên cứu chọn lựa và đo đạc những mẫu chứa những biến số cần xem xét để tìm ra lời giải thích. Từ những mẫu đo đạc, nhà nghiên cứu nhận thấy một số sự kiện nhất định xuất hiện. Nếu tất cả những biến giải thích quan trọng đều được đo đạc, thì những nghiên cứu này có thể đưa ra những kết luận về đặc tính của những biến giải thích kèm theo độ tin cậy.

Kiểu thứ 5 là những nghiên cứu giải thích. Từ nghiên cứu mối quan hệ giữa những biến phản hồi và những biến giải thích, nhà nghiên cứu có thể xây dựng các mô hình để giải thích những nguyên nhân dẫn đến biến động của những biến phản hồi. Thông qua các hàm phản hồi, nhà nghiên cứu có thể trả lời rõ mỗi biến giải thích đóng góp bao nhiêu phần trăm trong biến động của những biến phản hồi. Mặt khác, các hàm phản hồi còn cho phép ước lượng và dự đoán những biến phản hồi từ những biến giải thích.

Trong lâm học, người ta phân chia các nghiên cứu thành hai kiểu. Kiểu thứ nhất là những nghiên cứu giải thích. Kiểu thứ hai là những nghiên cứu ứng dụng. Kiểu thứ nhất được sử dụng để xác định những biến giải thích ảnh hưởng lớn đến biến phản hồi. Kiểu thứ hai là những nghiên cứu với mục đích thực hành hay kỹ thuật. Mục đích cơ bản của các nghiên cứu trong lâm học là xác định quy luật sống của rừng để xây dựng lý thuyết về rừng và những phương thức lâm sinh. Lâm học là một môn sinh thái học ứng dụng. Phương pháp nghiên cứu cơ bản của lâm học là phương pháp quan sát và mô tả những hiện tượng xuất hiện ở rừng; sau đó thực nghiệm để kiểm chứng lại những hiện tượng này. Những hiện tượng đáp ứng tốt mục tiêu của lâm học sẽ được sử dụng để xây dựng những nguyên lý lâm sinh. Do nhiều hạn chế khác nhau, nên trong nhiều trường hợp nhà lâm học thường khó giải thích tất cả những hiện tượng quan sát được. Vì thế, bên cạnh những nghiên cứu của mình, lâm học cũng sử dụng những kết quả nghiên cứu của các khoa học khác (khí hậu – thủy văn rừng, thực vật học, sinh lý thực vật, địa lý, địa chất, đất, động vật, côn trùng và bệnh cây...).

1.6. Phương pháp phân tích số liệu trong lâm học

1.6.1. Giới thiệu chung

Phân tích số liệu trong lâm học bao gồm những bước cơ bản dưới đây.

(1) Kiểm tra số liệu thu thập được. Đây là thủ tục xác định chất lượng của những số liệu thu thập được. Kiểm tra số liệu nhằm làm rõ một số câu hỏi sau đây: (a) Những số ngoại lai nào xuất hiện? (b) Những số liệu nào bị thất lạc? (c) Sai số của số liệu như thế nào? (d) Số liệu có cần phải đo đạc bổ sung hay không?

(2) Kiểm tra số liệu để xác định phương pháp xử lý. Ở bước này, nhà lâm học cần kiểm tra những giả định mà phương pháp xử lý số liệu đòi hỏi. Ngoài ra, xác định phương pháp biến đổi các số liệu. Chẳng hạn: Những số liệu nào cần phải biến đổi như lấy log, căn bậc hai…

(3) Chuẩn bị dung lượng mẫu cho quá trình xử lý.

(4) Xác định phương pháp xử lý, phân tích và tổng hợp số liệu sau khi xử lý.

(5) Xác định công cụ xử lý số liệu. Chẳng hạn: Phần mềm máy tính được sử dụng để xử lý, phân tích và tổng hợp kết quả.

Dưới đây chỉ giới thiệu một số khía cạnh chung có liên quan đến sự biến đổi số liệu trước khi áp dụng những phương pháp xử lý số liệu. Đó là biến đổi dữ liệu và lập bảng tần số cho các biến để phân tích…Các chương sau trình bày những phương pháp xử lý số liệu đối với từng trường hợp cụ thể.

1.6.2. Biến đổi, chuẩn hoá và lấy trọng số

1.6.2.1. Biến đổi số liệu

Việc áp dụng những kỹ thuật tính toán để xử lý thông tin trong lâm học là việc làm rất quan trọng. Biến đổi số liệu là thay thế số đo này bằng số đo khác. Trong lâm học, biến đổi số liệu thường được sử dụng trong trường hợp các biến số không có quan hệ tuyến tính với nhau. Tùy theo mục đích biến đổi, chúng ta có thể sử dụng hai kiểu biến đổi số liệu. Kiểu thứ nhất bao gồm những biến đổi số liệu theo kiểu định tâm và kiểu chuẩn hoá. Kiểu biến đổi này phù hợp với việc tạo ra những biến so sánh. Chẳng hạn: (a) Biến đổi mức độ phong phú của các loài cây gỗ trong ô mẫu; (b) Biến đổi các giá trị của các nhân tố môi trường. Kiểu thứ hai bao gồm những biến đổi số liệu nhằm phát hiện quy luật phân bố hay mô hình nhất định. Kiểu thứ hai được áp dụng thích hợp cho việc xây dựng các mô hình hồi quy nhằm giải thích mối quan hệ giữa các đặc tính của cây gỗ và rừng với những nhân tố sinh thái và cho nhiều mục đích phân tích khác.

(a) Biến đổi số liệu theo kiểu định tâm. Kiểu biến đổi này được sử dụng để so sánh sự khác biệt giữa số đo của một mẫu (Y_{ki}) với số trung bình của n mẫu (Y_{BQ}) (Công thức 1.4). Giá trị Y_{ki}* và Y_{ki} tương ứng là giá trị Y thứ k của biến i sau khi biến đổi và trước khi biến đổi. Trong lâm học, dạng biến đổi này có thể áp dụng cho nhiều trường hợp khác nhau. Chẳng hạn: (1) So sánh sự khác biệt về kích thước cây con giữa các nghiệm thức khác nhau; (2) So sánh sự khác biệt về năng suất rừng

giữa các lập địa khác nhau; (3) So sánh sự khác biệt về kích thước cây gỗ và quần thụ giữa các phương thức lâm sinh. Nói chung, số liệu biến đổi theo kiểu định tâm chỉ thích hợp cho mục đích so sánh sự khác biệt giữa số đo một mẫu hoặc một số mẫu với số trung bình của tất cả các mẫu.

$$Y_{ki}^{*} = Y_{ki} - Y_{BQ} \qquad (1.4)$$

(b) Biến đổi số liệu đối với những giá trị phân bố theo quy luật nhất định. Trong phân tích hồi quy tuyến tính, giả định đặt ra là sai số của biến phụ thuộc là ngẫu nhiên và phân bố chuẩn. Nếu sai số ngẫu nhiên không tuân theo phân bố chuẩn, thì biến phụ thuộc có thể được biến đổi ở dạng Logarit (Công thức 1.5 và 1.6), dạng căn bậc 2 (Công thức 1.7) và dạng mũ (Công thức 1.8). Ở dạng Logarit, công thức 1.6 được sử dụng khi một số biến Y_{ki} nhận giá trị bằng Zero.

$$Y_{ki}^{*} = Ln(Y_{ki}) \qquad (1.5)$$

$$Y_{ki}^{*} = Ln(Y_{ki} + 1) \qquad (1.6)$$

$$Y^{*}_{ki} = Y_{ki}^{1/2} \qquad (1.7)$$

$$Y^{*}_{ki} = a^{Yki} \qquad (1.8)$$

Dạng biến đổi Logarit thường được sử dụng cho những mục đích sau đây: (a) Xác định những đặc trưng thống kê của phân bố chuẩn đối với những biến phân bố ở dạng Lognormal; (b) Xác định trọng số nhỏ ở loài ưu thế, còn trọng số lớn ở những loài có độ phong phú thấp; (c) Xác định phản hồi của những loài cây gỗ với một số biến môi trường. Dạng biến đổi số liệu bằng cách lấy căn bậc 2 được sử dụng cho hai mục đích: (1) Xác định những biến số có dạng phân bố Poisson; (2) Xác định trọng số nhỏ ở loài ưu thế, còn trọng số lớn ở những loài có độ phong phú thấp. Dạng biến đổi số liệu ở dạng mũ được sử dụng để làm tăng độ phong phú của loài ưu thế trong QXTV.

(c) Biến đổi số liệu theo thang định danh có thứ bậc hơn kém và không có thứ bậc hơn kém. Cách biến đổi theo thứ bậc hơn kém được sử dụng cho nhiều trường hợp khác nhau. Chẳng hạn: (1) Phân tích mối quan hệ giữa năng suất rừng (Y) với cấp độ dày tầng đất (X); (2) Phân tích mối quan hệ giữa năng suất rừng (Y) với các cấp chỉ số lập địa (SI); (3) Phân tích mối quan hệ giữa chiều cao (H, m) của cây con với cấp bệnh (X). Trong trường hợp này, biến phụ thuộc (Y) và biến độc lập (X) đều được phân chia thành các cấp theo thứ bậc hơn kém. Các cấp Y và X được mã hóa thành những số nào đó. Độ lớn của mỗi cấp Y và X thay đổi tùy theo trường hợp. Biến đổi số liệu theo thang định danh không có thứ bậc hơn kém được sử dụng cho nhiều trường hợp khác nhau. Chẳng hạn: Xác định mối quan hệ giữa độ bắt gặp (1) và không bắt gặp (0) loài cây gỗ quan tâm với những biến môi trường khác nhau. Hai cách biến đổi này có ưu điểm: (1) Xác định nhanh mối quan hệ giữa các đặc tính của cây gỗ và rừng với những yếu tố môi trường; (2) Làm giảm khối lượng công việc trong những nghiên cứu thảm thực vật. Hai cách biến đổi này có một số nhược điểm: (1) Kết quả phân tích phụ thuộc vào việc phân cấp các biến; (2) Phân cấp các biến làm mất đi nhiều thông tin hữu ích; (3) Số lượng và khoảng cách

giữa các cấp có thể ảnh hưởng đến những kết luận. Nói chung, nếu các đại lượng lấy giá trị liên tục, thì biến đổi số liệu theo thang định danh có thứ bậc là cần thiết.

1.6.2.2. Chuẩn hoá số liệu

Chuẩn hóa số liệu là loại bỏ các đơn vị đo của các biến phản hồi (Biến phụ thuộc) và các biến giải thích (Biến độc lập). Trong phân tích hồi quy đa biến, cách biến đổi này được sử dụng để xác định vai trò của các biến giải thích trong hàm phản hồi. Trong lâm học, chuẩn hóa các biến nhằm những mục đích sau đây: (1) Xác định khuynh hướng biến đổi của biến phản hồi theo các biến giải thích; (2) Xác định rõ mức độ hay cường độ ảnh hưởng của từng biến giải thích đối với biến phản hồi.

Khi chuẩn hoá số liệu, các giá trị Y_{ki} được xác định theo công thức 1.9; trong đó Y_{ki}^* là giá trị Y thứ k của biến i sau khi biến đổi, S_k là sai tiêu chuẩn của chuỗi số liệu, Y_{k+} là tổng các giá trị Y_{ki}, Y_{k+}/n là giá trị trung bình của n giá trị Y_{ki}. Những biến chuẩn hoá có trung bình bằng Zero và phương sai bằng 1.

$$Y_{ki}^* = \frac{Y_{ki} - \frac{Y_{k+}}{n}}{S_k} \qquad (1.9)$$

(a) Chuẩn hoá đối với mỗi điểm thu mẫu. Cách biến đổi này được áp dụng trong trường hợp độ phong phú của các loài cây gỗ được thu thập trên những ô mẫu có kích thước khác nhau. Bởi vì loài cây gỗ hiếm gặp có khuynh hướng xuất hiện ở những ô mẫu lớn, nên số đo sự tương đồng có thể bị sai lệch. Trong thực hành, trước hết tập hợp độ phong phú của từng loài cây gỗ và tính tổng độ phong phú của tất cả các loài cây gỗ trong một điểm thu mẫu. Kế đến tính độ phong phú tương đối theo đơn vị phần mười hoặc phần trăm đối với mỗi loài bằng cách chia độ phong phú của mỗi loài cho tổng độ phong phú của tất cả các loài.

(b) Chuẩn hoá đối với toàn bộ các điểm thu mẫu. Trước hết tổng hợp độ phong phú của mỗi loài cây gỗ trong tất cả các điểm thu mẫu. Kế đến tính độ phong phú tương đối của mỗi loài cây gỗ bằng cách chia độ phong phú của nó trong một điểm thu mẫu cho tổng độ phong phú của nó ở tất cả các điểm thu mẫu. Phương pháp này làm tăng độ phong phú của loài hiếm gặp, nhưng làm giảm độ phong phú của loài thường gặp. Cách biến đổi này chỉ được áp dụng trong những trường hợp số loài ở các điểm thu mẫu khác nhau không quá nhiều và các loài có độ phong phú khác nhau.

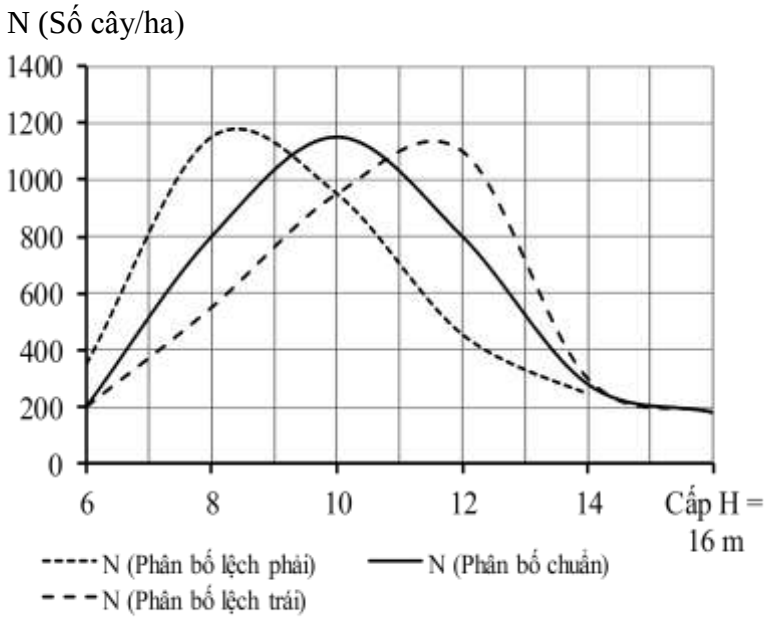
(c) Chuẩn hoá theo loài có độ phong phú lớn nhất. Cách chuẩn hoá này được thực hiện bằng cách chia độ phong phú của mỗi loài cây gỗ cho độ phong phú lớn nhất của loài cây gỗ nào đó. Khi áp dụng cách chuẩn hoá này, nhà lâm học cần phải lưu ý đến sự phong phú thành phần loài và kích thước ô mẫu. Khi sử dụng kích thước ô mẫu lớn, thì độ phong phú của những loài thường gặp có thể nhận giá trị lớn. Nếu một chỉ số được sử dụng với trọng số độ phong phú lớn, thì những điểm thu mẫu có nhiều điểm số bằng nhau sẽ trở nên khác nhau rất lớn so với những điểm có phạm vi điểm số lớn. Trong lâm học, cách chuẩn hoá này được áp dụng với mục

đích tạo ra dãy số liệu mới phụ thuộc không lớn vào kiểu số liệu (thể tích, tiết diện ngang, sinh khối, số lượng cá thể, độ che phủ...).

1.6.3. Phân bố tần số

Khi áp dụng toán thống kê để phân tích số liệu trong lâm học, điều quan trọng là phải biết quy luật phân bố của những đặc ở mức cây gỗ và quần thụ. Những đặc tính của cây gỗ và quần thụ thường phân bố theo hai kiểu liên tục và đứt quảng. Khi đại lượng nghiên cứu nhận những giá trị liên tục trong một khoảng nào đó, thì đại lượng ấy phân bố liên tụ. Chẳng hạn: Đường kính và chiều cao thân cây là những đại lượng nhận những giá trị liên tục. Khi đại lượng nghiên cứu nhận những giá trị xác định, thì đại lượng ấy phân bố đứt quãng. Chẳng hạn: Số quả trên cây; số cây trên một ô mẫu. Phân bố chuẩn đặc trưng cho những biến phân bố liên tục. Phân bố nhị thức và phân bố Poisson đặc trưng cho phân bố đứt quãng.

Phân bố của một số đặc trưng của quần thụ có thể ở dạng phân bố chuẩn hoặc dạng phân bố lệch chuẩn (Hình 1.2).



Hình 1.2. Ba kiểu phân bố số cây theo cấp chiều cao của rừng trồng thuần loài đồng tuổi.

Phân bố lệch chuẩn đặc trưng là phân bố Lognormal. Một biến số có phân bố Lognormal nếu Logarit của nó có phân bố chuẩn. Khi một biến số có phân bố lệch chuẩn, chúng ta có thể biến đổi nó về dạng phân bố chuẩn bằng cách lấy Logarit thập phân hoặc Logarit tự nhiên. Việc biến đổi số liệu như trên sẽ không làm sai lệch kết quả, bởi vì giá trị trung bình của biến số đã biến đổi có thể nhận được bằng cách lấy đối Logarit. Trị trung bình là một ước lượng của trung vị của phân bố Lognormal. Nếu một biến số có sai tiêu chuẩn lớn hơn giá trị trung bình hoặc giá trị lớn nhất lớn hơn 20 lần giá trị nhỏ nhất, thì biến số đó phân bố theo dạng

Lognormal. Trong lâm học, phân bố chuẩn và phân bố Lognormal thường được ứng dụng để mô tả phân bố số cây theo cấp đường kính (N/D) và cấp H (N/H). Phân bố nhị thức và phân bố Poisson thường được ứng dụng để mô tả phân bố số cây trên mặt đất.

1.7. Độ tin cậy của kết quả nghiên cứu

Về lý thuyết, một nghiên cứu phải tuân thủ những yêu cầu cơ bản sau đây: (1) Mục tiêu nghiên cứu rõ ràng; (2) Đối tượng nghiên cứu rõ ràng; (3) Những giả thuyết để kiểm định và những tham số ước lượng rõ ràng; (4) Rút mẫu ngẫu nhiên hoặc thiết kế nghiệm thức ngẫu nhiên; (5) Mức sai số của quyết định sai thích hợp; (6) Thiết kế nghiên cứu dựa trên những thủ tục chặt chẽ; (7) Sử dụng những phương pháp nghiên cứu thích hợp; (8) Những kết luận phải rõ ràng và chỉ liên quan đến những câu hỏi và mục tiêu nghiên cứu.

Một thiết kế nghiên cứu đáp ứng tốt những yêu cầu trên đây sẽ đưa ra được những kết luận rõ ràng về tổng thể và cung cấp cơ sở để giải thích những nguyên nhân về những khác biệt hay những ảnh hưởng và mối quan hệ giữa các biến số. Những chứng cứ về mối quan hệ nhân quả chỉ có tính thuyết phục cao khi các nghiệm thức được áp dụng cho đối tượng nghiên cứu. Nếu chỉ dựa trên những số liệu điều tra bằng phương pháp quan sát và mô tả, thì nhà lâm học không thể giải thích rõ ràng mối quan hệ nhân quả giữa các biến. Để chứng minh cho những nhận định trên đây, chúng xem xét hai phương pháp nghiên cứu ảnh hưởng của môi trường đến sinh trưởng của rừng Dầu con rái (*Dipterocarpus alatus*).

Phương pháp 1. Khi nghiên cứu ảnh hưởng của hàm lượng phân tổng hợp NPK (X) đến sinh khối (B) của rừng Dầu con rái, nhà lâm học thiết kế các lô thí nghiệm ngẫu nhiên trên nền đất đồng nhất; trong đó các lô thí nghiệm được bón phân tổng hợp NPK theo mức tăng dần từ 0, 5, 10,..., 50 g/gốc. Từ số liệu của thí nghiệm này, nhà lâm học nhận thấy: (1) Ở giới hạn thấp của phân tổng hợp NPK, sinh khối của rừng Dầu con rái gia tăng dần nhưng gia tăng rất chậm; (2) Kế đến khi tăng hàm lượng phân tổng hợp NPK, thì sinh khối của rừng Dầu con rái tăng lên rất nhanh và đạt lớn nhất ở mức X nào đó; (3) Khi tiếp tục gia tăng hàm lượng phân tổng hợp NPK lên mức rất cao, thì sinh khối của rừng Dầu con rái lại suy giảm dần. Trong trường hợp này, hàm phản hồi B = f(X) tồn tại ở dạng hàm bậc 2 với hệ số của X^2 mang giá trị âm. Bằng cách khai triển hàm này, nhà lâm học xác định được các tham số sinh thái như biên độ NPK thích hợp, hàm lượng NPK tối ưu và phạm vi chống chịu của Dầu con rái đối với hàm lượng phân tổng hợp NPK. Phương pháp bố trí thí nghiệm và xử lý số liệu như trên đây giúp cho nhà lâm học nhận được kết quả rõ ràng về ảnh hưởng của hàm lượng phân tổng hợp NPK đến sinh trưởng của rừng Dầu con rái.

Phương pháp 2. Khi nghiên cứu ảnh hưởng của hàm lượng đạm (X) ở lớp đất mặt dưới tán rừng tự nhiên đến sinh khối (B) của rừng Dầu con rái, nhà lâm học xác định B của rừng Dầu con rái kèm theo hàm lượng đạm trong đất. Sau đó, bằng phân tích hàm phản hồi B = f(X), nhà lâm học cũng nhận được mô hình B = f(X) tồn tại ở

dạng hàm bậc 2 với hệ số của X^2 mang giá trị âm. Bằng cách khai triển hàm này, xác định được các tham số sinh thái (biên độ sinh thái, tối ưu sinh thái và phạm vi chống chịu). Tuy vậy, kết luận đưa ra ở trường hợp này là kém tin trưởng. Nguyên nhân là vì sinh khối của rừng Dầu con rái không chỉ phụ thuộc vào hàm lượng đạm trong đất, mà còn nhiều nhân tố khác trong môi trường đất. Để xác định chính xác những nhân tố dẫn đến sự thay đổi sinh khối của rừng Dầu con rái ở trường hợp này, nhà lâm học cần phải xây dựng mô hình phản hồi dựa theo nhiều biến giải thích khác nhau.

Nói chung, độ tin cậy của những kết luận rút ra từ những nghiên cứu lâm học phụ thuộc vào phương pháp nghiên cứu. Các biến phản hồi phụ thuộc vào nhiều biến giải thích khác nhau. Bởi vì có rất nhiều biến giải thích, nên nhà lâm học không thể xác định chính xác mối quan hệ giữa các biến phản hồi với các biến giải thích. Vì thế, tính không phù hợp của những kết luận là do nhà lâm học không tuân thủ đúng những thủ tục nghiên cứu. Dưới đây là một số nguyên nhân dẫn đến những kết luận kém tin cây.

(1) Sử dụng không đúng phân tích thống kê. Những sai sót của việc lạm dụng phân tích thống kê là do: (a) Quá ít thí nghiệm và phân tích thí nghiệm không đúng; (b) Sử dụng không đúng nguyên lý và những tiêu chuẩn thống kê; (c) Xây dựng hàm phản hồi không đúng quy luật; (d) Thiếu hiểu biết về những mối liên hệ giữa cây gỗ và rừng với điều kiện môi trường...

(2) Sử dụng số liệu điều tra thuần túy để làm chứng cứ khoa học. Số liệu thu thập từ những cuộc điều tra đơn giản không thể cho ra kết luận tin tưởng.

(3) Những số liệu không thể giải thích là do thu mẫu trệch. Nếu mẫu không được rút ngẫu nhiên, thì nhà lâm học có thể nhận được những kết luận kém tin cậy về tổng thể. Sử dụng phương pháp rút mẫu điển hình thường dẫn đến kết luận kém tin cậy.

(4) Những nguồn không thể giải thích khác. Những kết luận sai có thể là do nhà lâm học đặt sai giả thuyết; sử dụng sai phương pháp phân tích thống kê; sử dụng không phù hợp mức ý nghĩa thống kê...

Tính phức tạp của những nghiên cứu trong lâm học biểu hiện ở chỗ: (1) Những biến nghiên cứu thường có biến động mạnh; (2) Sự tương tác phức tạp giữa các biến giải thích và giữa các biến giải thích với biến phản hồi; (3) Thiếu chắc chắn về nguyên nhân của những mối liên hệ. Những khó khăn này càng tăng lên khi nghiên cứu trong điều kiện ở ngoài trời. Biến động mạnh của những biến thu thập ở ngoài trời là do chúng chịu ảnh hưởng của nhiều nhân tố khác nhau. Ngoài ra, những nhân tố môi trường thường thay đổi theo không gian và thời gian. Chẳng hạn: Độ phong phú của các loài cây gỗ trong QXTV có mối quan hệ không chỉ với pH, nước, nhiệt độ, mà còn với hàm lượng khoáng chất trong đất và sinh vật khác...Nhiều biến giải thích có quan hệ với nhau hay có hiện tương cộng tuyến tính. Chẳng hạn: (a) Nước có ảnh hưởng đến hàm lượng chất khoáng và nhiệt độ đất; (b) Thành phần khoáng ảnh hưởng đến pH; (c) Các biến khí hậu có quan hệ với nhau;

(d) Các đặc tính của cây gỗ và rừng cũng có quan hệ với nhau...Những biến phản hồi cũng có hiện tượng cộng tuyến tính tương tự như các biến giải thích. Vì thế, nhà lâm học không thể dễ dàng xác định và giải thích rõ ràng những hiện tượng quan sát trong điều kiện nghiên cứu ở ngoài trời

Một trong những phương pháp phân tích số liệu về rừng là sử dụng phương pháp phân tích hồi quy và tương quan đa biến. Nhưng muốn áp dụng tốt kỹ thuật phân tích hồi quy đa biến, nhà lâm học cần phải tuân thủ đúng phương pháp bố trí thí nghiệm và phương pháp thu thập các dữ liệu theo chỉ dẫn của thống kê toán học.

Chương 2
PHÂN TÍCH KẾT CẤU VÀ ĐA DẠNG LOÀI CÂY GỖ

2.1. Khái niệm về kết cấu loài cây gỗ

Rừng bao gồm nhiều loài thực vật khác nhau (cây gỗ, cây thân bụi, cây thân thảo, cây thân leo...); trong đó cây gỗ là đối tượng nghiên cứu của lâm học. Sở dĩ lâm học chỉ tập trung nghiên cứu các loài cây gỗ là vì chúng là thành phần chính của rừng. Chúng cung cấp cho con người nhiều sản phẩm đa dạng (gỗ, củi, nhựa, hoa, quả...) để sử dụng vào đời sống xã hội. Ngoài ra, tài nguyên gỗ còn có ý nghĩa kinh doanh.

Trong lâm học, thuật ngữ kết cấu loài cây gỗ (Tree Species Composition) biểu thị thành phần loài cây gỗ và tỷ lệ độ phong phú của mỗi loài cây gỗ trong một QXTV. Thành phần loài cây gỗ là số loài cây gỗ (S, loài) bắt gặp trong QXTV. Tỷ lệ độ phong phú của mỗi loài cây gỗ trong QXTV được biểu thị bằng đơn vị phần mười hoặc phần trăm. Tổng tỷ lệ độ phong phú của các loài cây gỗ trong QXTV là 1 hoặc 100%.

Kết cấu loài cây gỗ khác với kết cấu hệ thực vật. Kết cấu hệ thực vật biểu thị thành phần loài cây và tỷ lệ độ phong phú của mỗi loài cây trong QXTV. Bởi vì lâm học chỉ tập trung nghiên cứu kết cấu loài cây gỗ, nên kết cấu QXTV được hiểu là kết cấu quần xã cây gỗ hay kết cấu quần thụ (Stand). Khi phân tích kết cấu loài cây gỗ của rừng, nhà lâm học tập trung làm rõ những câu hỏi dưới đây.

(1) Mức độ giàu có về loài cây gỗ của QXTV như thế nào? Câu hỏi này được trả lời thông qua số loài (S, loài) cây gỗ bắt gặp trong QXTV và chỉ số giàu có về loài.

(2) Vai trò sinh thái của mỗi loài cây gỗ trong QXTV như thế nào? Vai trò sinh thái của loài cây gỗ phản ánh mức độ cải biến môi trường của nó. Về lý thuyết, mức độ cải biến môi trường của mỗi loài cây gỗ và nhóm loài cây gỗ được đánh giá thông qua khả năng hấp thu và thải năng lượng và vật chất vào môi trường. Thông thường, nguồn năng lượng và vật chất mà mỗi loài cây gỗ thu vào và thải ra môi trường là những biến rất khó đo đạc tại rừng. Dựa trên quan điểm cho rằng loài cây gỗ nào có độ phong phú lớn hơn thì loài đó cải biến môi trường lớn hơn, câu hỏi 2 được trả lời thông qua phân tích độ phong phú của mỗi loài cây gỗ.

(3) Loài cây gỗ nào đóng vai trò ưu thế sinh thái trong QXTV? Đó là loài cây gỗ có độ phong phú lớn nhất trong QXTV.

(4) Loài cây gỗ nào đóng vai trò đồng ưu thế sinh thái? Đó là loài cây gỗ có độ phong phú lớn thứ 2 trong QXTV.

(5) Những loài cây gỗ nào là những loài phụ thuộc? Đó là những loài cây gỗ có độ phong phú thấp hơn so với loài cây gỗ đồng ưu thế sinh thái và đời sống của chúg phụ thục vào những loài cây gỗ ưu thế và đồng ưu thế.

(6) Những loài cây gỗ nào có ý nghĩa kinh doanh? Đó là những loài cây gỗ đáp ứng được yêu cầu của thị trường.

(7) Những loài cây gỗ nào cần được bảo tồn? Đó là những loài cây gỗ quý, hiếm hoặc có giá trị cao về khoa học và kinh tế nhưng chúng đang có nguy cơ bị tuyệt chủng.

(8) Vai trò của những nhóm loài cây gỗ trong QXTV như thế nào? Nhóm loài cây gỗ có thể được phân chia theo mục đích nghiên cứu và mục đích sử dụng rừng. Nếu mục đích nghiên cứu là cung cấp thông tin về các đơn vị phân loại thực vật, thì nhóm loài cây gỗ có thể được phân chia theo 7 đơn vị (Giới, Ngành, Lớp, Bộ, Họ, Chi, Loài). Theo mục đích sử dụng rừng hay mục đích kinh doanh, các loài cây gỗ được phân chia theo hai nhóm: Nhóm loài cây gỗ có ý nghĩa kinh doanh và nhóm loài cây gỗ phi kinh tế. Về bản chất, không có loài cây gỗ nào là loài phi kinh tế. Nhà kinh doanh rừng sử dụng thuật ngữ "Loài cây gỗ phi kinh tế" chỉ ngụ ý rằng cho đến nay thị trường không có nhu cầu về những sản phẩm của những loài cây gỗ này.

Phân tích kết cấu loài cây gỗ mang lại nhiều ý nghĩa khác nhau. Về khoa học, những nghiên cứu về kết cấu loài cây gỗ trong những QXTV khác nhau cung cấp những thông tin cho phân loại rừng, xác định vùng sống của các loài cây gỗ, đánh giá vai trò sinh thái của các loài cây gỗ và xây dựng những nguyên lý lâm sinh. Về thực tiễn, những thông tin về kết cấu loài cây gỗ giúp ích cho việc phân chia các đơn vị kinh doanh rừng, chọn lựa những loài cây kinh doanh, bảo tồn những loài cây gỗ quý, hiếm hoặc có giá trị cao về kinh tế.

2.2. Phương pháp xác định vai trò sinh thái của các loài cây gỗ trong QXTV

2.2.1. Độ phong phú của loài cây gỗ, cây bụi và cây thân thảo

Trong lâm học, thuật ngữ "Độ phong phú" của loài cây gỗ được sử dụng theo hai nghĩa khác nhau. Một là, theo nghĩa rộng, độ phong phú biểu thị khối lượng của loài cây gỗ trong QXTV trên ô mẫu. Chẳng hạn: G (m^2), M (m^3), B (kg hoặc tấn) của loài cây gỗ trên diện tích thu mẫu hoặc trên một đơn vị diện tích (1 ha). Hai là, theo nghĩa hẹp, độ phong phú phản ánh số lượng cá thể hay mật độ (N, cây) của loài cây gỗ trên ô mẫu, tần số bắt gặp loài (F) trên ô mẫu, tổng độ che chiếu của tán lá theo chiều nằm ngang (C, m^2) trên diện tích thu mẫu hoặc trên một đơn vị diện tích (1 ha). Độ phong phú là một đặc tính quan trọng của loài cây gỗ, bởi vì nó thuyết minh cho vai trò sinh thái của mỗi loài cây gỗ trong sự hình thành QXTV. Dưới đây giới thiệu một số chỉ tiêu biểu thị độ phong phú của các loài cây gỗ, cây bụi và thảm cỏ trong các QXTV rừng.

(1) Mật độ của loài cây gỗ. Chỉ tiêu này biểu thị số lượng cá thể của loài cây gỗ sống hay định cư trên một diện tích nhất định, thường là 1ha. Đối lập với thuật

ngữ mật độ là diện tích trung bình mà một cá thể sử dụng. Mật độ của loài cây gỗ trên 1 ha được xác định theo công thức 2.1; trong đó s (m^2) = diện tích ô mẫu, S = diện tích 1 ha (10.000 m^2), n_i = số cây trong ô mẫu.

$$N_i \text{ (Cây/ha)} = n_i * \left(\frac{S}{s}\right) \qquad (2.1)$$

Trong lâm học, mật độ là chỉ tiêu đơn giản nhất được sử dụng để đánh giá vai trò sinh thái của loài cây gỗ trong QXTV. Tuy nhiên, mật độ không phải luôn luôn là chỉ tiêu đánh giá đúng vai trò sinh thái của một loài cây gỗ trong QXTV. Điều này được giải thích bởi vai trò sinh thái của một loài cây gỗ phụ thuộc không chỉ vào mật độ, mà còn vào nhiều chỉ tiêu khác như hình dạng và độ lớn của nó, trạng thái cá thể và điều kiện lập địa...Thật vậy, giả thiết 1 ha rừng có 15 cây Dầu con rái (*Dipterocarpus alatus* Roxb) với đường kính ngang ngực (D, cm) là 50 cm và chiều cao toàn thân (H, m) là 30 m; 15 cây Dền đỏ (*Xylopia vielana* Pierre) với D = 40 cm và H = 25 m. Theo chỉ tiêu mật độ, hai loài cây này đóng vai trò sinh thái tương tự như nhau trong QXTV. Thế nhưng, nếu căn cứ vào kích thước (D, H), thì V của Dầu con rái (2,7 m^3/ha) lớn hơn 1,9 lần so với Dền đỏ (1,4 m^3/ha). Từ ví dụ này cho thấy, chỉ tiêu mật độ tự nó phản ánh không chính xác độ phong phú và vai trò sinh thái của mỗi loài cây gỗ trong QXTV. Chỉ tiêu mật độ chỉ có ý nghĩa khi kích thước của các loài cây gỗ đem so sánh là tương đồng với nhau.

(2) Độ che phủ của loài cây gỗ. Chỉ tiêu này biểu thị phần trăm hoặc phần mười diện tích mặt đất nằm ngang trong QXTV trên ô mẫu mà mỗi loài cây gỗ che phủ. Người ta phân biệt hai loại độ che phủ là độ che phủ thực và độ che chiếu. Độ che phủ thực của một loài cây gỗ là phần trăm hoặc phần mười diện tích mặt đất nằm ngang mà gốc của nó che phủ. Ngược lại, độ che chiếu của loài cây gỗ là phần trăm hoặc phần mười diện tích mặt đất nằm ngang mà tán lá của nó che chiếu. Độ che phủ của cây bụi thường chỉ được xác định theo diện tích tán che trên mặt đất trong ô mẫu. Độ che phủ của cây gỗ và cây bụi thường được xác định khá chính xác. Trái lại, độ che phủ của thảm cỏ là chỉ tiêu rất khó xác định chính xác. Trong lâm học, độ phong phú của thảm cỏ thường được đánh giá theo nhiều chỉ tiêu khác nhau: (a) Trọng lượng (tươi, khô không khí hoặc khô tuyệt đối); (b) Phần trăm diện tích mặt đất được loài xâm chiếm; (c) Số cây trên đơn vị diện tích...Căn cứ vào số lượng cá thể và phần trăm diện tích mặt đất được loài chiếm đoạt trên 1 đơn vị diện tích (1 m^2), Druze đã phân chia độ phong phú của thảm cỏ thành 7 cấp (Bảng 2.1). Để dễ dàng áp dụng toán học trong phân tích số liệu, số cá thể xuất hiện trên ô mẫu được mã hóa theo thứ bậc tăng dần từ 0 – 6.

(3) Trọng lượng hoặc khối lượng của loài cây gỗ. Chỉ tiêu này biểu thị trọng lượng (kg/m^2, tấn/ha) hoặc khối lượng (m^3/ha) của loài cây gỗ trên một đơn vị diện tích hay thể tích nhất định. So với khối lượng của loài cây gỗ, trọng lượng của loài có ý nghĩa sinh thái cao hơn. Thật vậy, trọng lượng của một loài cây gỗ cho biết khối lượng vật chất và năng lượng được nó hấp thụ và thải ra môi trường. Thông qua trọng lượng của các loài cây gỗ, nhà lâm học không chỉ xây dựng được mô hình

biểu diễn chu trình và động thái chuyển hóa vật chất và năng lượng trong QXTV, mà còn làm rõ mối quan hệ giữa các loài cây gỗ với môi trường.

Bảng 2.1. Phân cấp độ phong phú của các loài cây thân thảo sống dưới tán rừng theo phương pháp của Druze.

Thứ tự	Cấp che phủ	Phân chia theo hệ thống 7 cấp:		
		Mã hóa	Tiêu chuẩn đánh giá theo:	
			Số cá thể	Độ che phủ (%)
(1)	(2)	(3)	(4)	(5)
1	Soc(*)	6	Rất nhiều	76 - 100
2	Cop_3	5	Nhiều	51 - 75
3	Cop_2	4	Tương đối nhiều	21 - 50
4	Cop_1	3	Tương đối	5 - 20
5	Sp	2	Ít	1 - 4
6	Sol	1	Rất ít	< 1
7	Un	0	Chỉ một cá thể	< 0,2

(*) Viết tắt từ chữ La tinh: *Socialis = Soc, Copiosus = Cop, Spars = Sp, Unicus = Un.*

(4) Độ thường gặp (F = Frequency). Chỉ tiêu này biểu thị phần mười hoặc phần trăm số ô mẫu bắt gặp loài cây quan tâm mà không phụ thuộc vào độ phong phú của nó. Độ thường gặp của một loài cây gỗ được xác định theo công thức 2.2; trong đó n là số ô mẫu bắt gặp loài cây quan tâm, N là tổng số ô mẫu nghiên cứu. Về bản chất, độ thường gặp biểu thị xác suất bắt gặp loài cây quan tâm trên những ô mẫu có kích thước hay diện tích nhất định. Thứ bậc độ ưu thế của các loài cây gỗ được đánh theo thứ bậc cao thấp của độ thường gặp. Chỉ tiêu F nhận giá trị cao nhất bằng 1 theo đơn vị phần mười hoặc 100% theo đơn vị phần trăm.

$$F = (\frac{n}{N})*100 \qquad (2.2)$$

Trong thực hành, để dễ dàng xử lý số liệu bằng toán học, hai dấu hiệu "Bắt gặp loài" và "Không bắt gặp loài" trong ô mẫu được mã hóa bằng hai biến nhị phân tương ứng là 1 và 0. Chỉ tiêu F có một số ưu điểm. Một là nó được đo đạc và tính toán rất dễ dàng đối với những QXTV không phong phú về loài cây gỗ. Hai là sử dụng chỉ tiêu F cho phép xác định nhanh loài cây quan tâm có mặt trong khu vực nghiên cứu hay không. Ba là chỉ tiêu F được sử dụng để xác định gần đúng sự phân bố của các loài cây gỗ trong QXTV. Nếu một loài cây gỗ nào đó có xác suất bắt gặp lớn, thì sự phân bố của nó trong QXTV là đồng đều. Bốn là chỉ tiêu F được sử dụng để phân loại QXTV rừng. Từ so sánh hai dấu hiệu "Bắt gặp" và "Không bắt gặp" những loài cây gỗ ưu thế và đồng ưu thế, nhà lâm học có thể phân chia thảm thực vật rừng thành những đơn vị nhỏ hơn như kiểu rừng và kiểu QXTV. Năm là chỉ tiêu F được sử dụng để xác định nhanh biên độ sinh thái của mỗi loài cây gỗ. Chẳng hạn: Thông qua dấu hiệu "Bắt gặp" và "Không bắt gặp" loài cây gỗ trên những độ cao khác nhau trong một vùng sinh thái, nhà lâm học có thể xác định giới hạn độ cao mà loài cây gỗ này xuất hiện. Sáu là chỉ tiêu F có thể được sử dụng để xác định

gần đúng mật độ của các loài cây gỗ trong QXTV. Theo Gleason (1922; Dẫn theo Vasilevich, 1969), khi biết số ô mẫu nghiên cứu (q) và tổng số cá thể của loài trên các ô mẫu (N, cây), thì độ thường gặp (F) của loài được xác định theo hàm 2.3.

$$F = 1 - (1 - \frac{1}{q})^N \qquad (2.3)$$

Theo Blackman (1935, 1942; Dẫn theo Vasilevich), khi biết độ thường gặp F của loài cây gỗ, thì số cá thể bình quân (m) trên một ô mẫu được xác định theo hàm 2.4.

$$m = - Ln(1 - F) \qquad (2.4)$$

Lưu ý rằng cả hai cách tính trên đây chỉ cho kết quả tin cậy khi phân bố của loài cây gỗ trên mặt đất tuân theo luật phân bố ngẫu nhiên. Nhưng thông thường phân bố của các loài cây trên mặt đất không phải là ngẫu nhiên, bởi vì môi trường của QXTV không phải là môi trường thuần nhất. Nói chung, việc xác định số cá thể bình quân (m) trên một ô mẫu thông qua độ thường gặp (F) chỉ nhận được kết quả chính xác khi loài có mật độ thấp.

Chỉ tiêu F có một số nhược điểm cơ bản. Một là độ thường gặp của một loài cây gỗ phụ thuộc rất lớn vào số lượng ô mẫu. Ngoài ra, chúng ta cũng rất khó xác định biến động của độ thường gặp của một loài cây gỗ theo diện tích QXTV. Để xác định biến động của F, chúng ta có thể xác định F cục bộ bằng cách chia đối tượng nghiên cứu thành nhiều khối. Sau đó xác định F của loài cây gỗ trên từng khối và tìm biến động của F giữa các khối. Hai là độ thường gặp của một loài cây gỗ phụ thuộc rất nhiều vào kích thước ô mẫu. Khi ô mẫu có kích thước càng lớn, thì xác suất bắt gặp loài cây quan tâm càng lớn. Nguyên nhân là vì kích thước ô mẫu càng lớn thì nó càng tiến dần đến toàn bộ không gian mà loài cây gỗ quan tâm có mặt. Vì thế, khi xác định độ bắt gặp của một loài cây gỗ, nhà lâm học cần phải tìm kích thước ô mẫu thích hợp, đồng thời phải sử dụng thống nhất những kích thước và hình dạng ô mẫu cho những QXTV tương đồng. Ba là sử dụng độ thường gặp của loài cây gỗ trên ô mẫu có thể đánh giá sai biên độ sinh thái hay điều kiện môi trường hình thành nó. Sự vắng mặt của loài cây quan tâm trong khu vực nghiên cứu cũng chưa đủ căn cứ để nói rằng môi trường hay lập địa này là không thích hợp với nó. Thật vậy, trong biên độ sinh thái của loài cây gỗ quan tâm, vì nhiều nguyên nhân khác nhau, nên nó đã vắng mặt trên ô mẫu.

Hiện nay đang tồn tại những ý kiến trái ngược nhau về ý nghĩa của chỉ tiêu độ thường gặp. Một số nhà sinh thái học cho rằng độ thường gặp vừa là chỉ tiêu biểu thị độ phong phú của loài, vừa biểu thị sự phân bố của loài trong QXTV. Nếu chỉ dựa vào chỉ tiêu này, chúng ta không thể đánh giá chính xác mức độ phong phú và hình thái phân bố của loài cây gỗ.

Khi sử dụng những chỉ tiêu biểu thị vai trò sinh thái của các loài cây gỗ trong QXTV rừng, nhà lâm học cần chú ý đến một số điểm dưới đây.

(1) Mặc dù những chỉ tiêu biểu thị độ phong phú của các loài cây gỗ có quan hệ chặt chẽ với nhau, nhưng mỗi chỉ tiêu có ý nghĩa khác nhau. Mật độ của loài là chỉ tiêu được đo đạc dễ dàng nhưng kém ý nghĩa nhất. Chỉ tiêu này chỉ có ý nghĩa cao khi các loài đem so sánh có kích thước và khả năng trao đổi năng lượng và chất tương đồng với nhau. Độ che tán phản ánh tốt mức độ lợi dụng không gian sống của các loài cây gỗ. Độ che thực là chỉ tiêu ổn định và được đo đạc dễ dàng hơn so với độ che tán. Mặc dù vậy, so với độ che thực, độ che tán có ý nghĩa sinh thái lớn hơn. Nguyên nhân là vì độ che tán phản ánh phần trăm hoặc phần mười diện tích mặt đất được loài này hay loài khác sử dụng để hấp thu ánh sáng và ngăn cản ánh sáng đạt đến sàn rừng. Độ che tán thay đổi theo các loài cây gỗ, theo mùa, theo tuổi và môi trường sống. Mặt khác, nhà lâm học rất khó xác định chính xác độ che tán của các loài cây gỗ. Trái lại, độ che thực (G của các loài cây gỗ) là chỉ tiêu dễ đo đạc và tính toán. Vì thế, độ che thực thường được sử dụng nhiều hơn so với độ che tán. Chỉ tiêu trọng lượng hay sinh khối cho biết loài cây gỗ sử dụng không gian nơi ở với cường độ mạnh đến mức nào. Trong thực tế, sinh khối của loài cây gỗ là chỉ tiêu khó xác định tại rừng. Vì thế, nhà lâm học không sử dụng sinh khối để đánh giá kết cấu loài cây gỗ trên các ô mẫu. Độ thường gặp là một chỉ tiêu dễ đo đạc và thường được sử dụng nhiều trong phân loại QXTV. Thế nhưng, độ thường gặp của loài cây gỗ trong QXTV thay đổi không chỉ theo số lượng và kích thước ô mẫu, mà còn theo kiểu phân bố của nó trên mặt đất. Khi sử dụng kích thước ô mẫu khác nhau, thì độ thường gặp của một loài cây gỗ nhận những giá trị khác nhau. Vì thế, khi thuyết minh độ thường gặp của một loài nào đó, nhà lâm học cần phải chỉ rõ số lượng, kích thước và phương pháp bố trí ô mẫu.

(2) Tất cả những chỉ tiêu thuyết minh độ phong phú của loài cây gỗ chỉ có ý nghĩa khi đặt chúng trong mối quan hệ với số lượng, kích thước và phương pháp bố trí ô mẫu. Mật độ, tiết diện ngang và thể tích thân của các loài cây gỗ được tính trên ô mẫu 10 m^2 sẽ khác hoàn toàn so với ô mẫu có diện tích 100, 1000 và 10.000 m^2. Nguyên nhân là vì các chỉ tiêu này phân bố không đồng đều trên mặt đất.

(3) Trong thực tế và trong những nghiên cứu khoa học, độ phong phú của cây bụi thường được đánh giá theo số lượng cá thể, chiều cao trung bình (cm) của thân cây, độ che phủ của tán lá nằm ngang trên mặt đất. Độ phong phú của các loài cây thân thảo có thể được đánh giá theo một trong những chỉ tiêu sau đây: (a) Số lượng cá thể trên đơn vị diện tích (số cây/1 m^2); (b) Độ thường gặp của loài (F); (c) Độ che phủ của các cơ quan (thân cây, tán lá) trên mặt đất; (d) Trọng lượng của loài trên đơn vị diện tích (m^2).

2.2.2. Những chỉ số biểu thị mức độ ưu thế của các loài cây gỗ

Khi độ phong phú của các loài cây gỗ được xác định theo những chỉ tiêu khác nhau, thì vai trò (sinh thái, kinh tế) của chúng cũng được đánh giá khác nhau. Vì thế, để tránh tranh luận không cần thiết và đánh giá đúng độ phong phú của các loài cây gỗ, nhiều nhà lâm học đề nghị sử dụng một chỉ số tổng hợp chung. Dưới đây giới thiệu môt số phương pháp thường được sử dụng để phân tích và đánh giá độ phong phú của các loài cây gỗ.

(1) Chỉ số ưu thế Kayama. Theo Kayama (1961; Dẫn theo Vasilevich, 1969), vai trò sinh thái của một loài cây gỗ trong QXTV được xác định theo công thức 2.5; trong đó F là độ thường gặp của loài cây gỗ, C là độ che tán của loài cây gỗ trên mặt đất, N là mật độ của loài cây gỗ, B là sinh khối của loài cây gỗ. Tất cả bốn chỉ tiêu F, C, D và B trong công thức 2.5 đều được tính theo đơn vị phần trăm.

$$\frac{(F + C + N)}{3}*B \qquad (2.5)$$

(2) Chỉ số ưu thế của Curtis và McIntosh. Theo Curtis và McIntosh (1951; ; Dẫn theo Vasilevich, 1969), vai trò sinh thái của loài cây gỗ trong QXTV được biểu thị bằng chỉ số giá trị quan trọng (IVI = Importance Value Index). Chỉ số IVI của loài cây gỗ là tổng hoặc giá trị bình quân của độ thường gặp tương đối (F%), mật độ tương đối (N%) và "Độ ưu thế tương đối" (G%). Theo định nghĩa này, chỉ số IVI của loài i trong QXTV được tính theo công thức 2.6; trong đó F_i, N_i và G_i tương ứng là độ thường gặp tương đối, mật độ tương đối và độ ưu thế tương đối của loài i.

$$IVI_i = (F_i\% + N_i\% + G_i\%) \qquad (2.6a)$$

$$IVI = \frac{(F\% + N\% + G\%)}{3} \qquad (2.6b)$$

Độ thường gặp tương đối (F%) của loài cây gỗ thứ J là tỷ lệ phần trăm giữa số ô mẫu bắt gặp loài J (f_J) và tổng số ô nghiên cứu (K) (Công thức 2.7).

$$F\% = (f_J/K)*100 \qquad (2.7)$$

Mật độ tương đối của loài cây gỗ thứ J (N_J%) là tỷ lệ phần trăm giữa số cá thể của loài J (n_J, cây) và tổng số cá thể của tất cả các loài trong QXTV trên ô mẫu (N, cây) (Công thức 2.8).

$$N_J\% = (n_J/N)*100 \qquad (2.8)$$

Độ ưu thế tương đối của loài J (G_J%) là tỷ lệ phần trăm giữa tổng tiết diện ngang thân của loài J (G_J, m^2) và tổng tiết diện ngang thân (G, m^2) của tất cả các loài cây gỗ trong QXTV trên ô mẫu (Công thức 2.9). Tiết diện ngang thân ở vị trí ngang ngực của cây i thuộc loài J (g_{iJ}, m^2) được tính theo công thức 2.10; trong đó D_{iJ} (cm) = đường kính cây i thuộc loài J. Giá trị G_J được tính theo công thức 2.11; trong đó g_{iJ} = tiết diện ngang thân của cây i thuộc loài J

$$G_J\% = (G_J/G)*100 \qquad (2.9)$$

$$g_{iJ} = 0{,}00007854*D_{iJ}^2 \qquad (2.10)$$

$$G_J = \sum_{(i=i,nj)} g_{iJ} \qquad (2.11)$$

Đối với tất cả các loài cây gỗ trong QXTV, tổng (F% + N% + G%) = 300%. Nếu F, N và G được tính theo đơn vị phần mười, thì tổng các hệ số của chúng là 1.

Ngày nay các nhà sinh thái học và lâm học ở nhiều nước sử dụng chỉ số IVI của Curtis và McIntosh để đánh giá không chỉ vai trò sinh thái của các loài cây gỗ

trong QXTV, mà còn phân loại thảm thực vật rừng. Tuy vậy, chỉ số IVI có một số nhược điểm dưới đây.

Một là giá trị IVI thay đổi tùy theo số lượng, kích thước và phương pháp bố trí ô mẫu. Kích thước ô mẫu càng lớn thì độ bắt gặp của một loài cây gỗ nào đó càng cao. Mặt khác, nếu sử dụng ô mẫu với số lượng và phương pháp bố trí ô mẫu khác nhau, thì kết quả nhận được chỉ số IVI cũng khác nhau.

Hai là giá trị F% chỉ có ý nghĩa khi các loài cây gỗ phân bố ngẫu nhiên trên mặt đất. Thật vậy, giả thiết hai loài gỗ A và B có kích thước như nhau và mỗi loài có 15 cá thể. Tổng số 5 ô mẫu nghiên cứu. Loài A phân bố đều trong 5 ô mẫu với mỗi ô mẫu 3 cây. Theo số liệu này, loài A có F = 100% (100% = (5 ô mẫu bắt gặp loài A/tổng số 5 ô mẫu nghiên cứu)*100. Trái lại, loài B phân bố không đồng đều và chỉ bắt gặp trong 3 ô mẫu với mỗi ô mẫu 5 cây. Theo số liệu này, loài B có F = 60% (60% = (3 ô mẫu bắt gặp loài B/tổng số 5 ô mẫu nghiên cứu)*100. Từ kết quả tính toán này, kết luận đưa ra là độ phong phú của loài A (100%) lớn hơn độ phong phú của loài B (60%). Đây là kết luận sai lầm bởi vì hai loài A và B có kích thước như nhau và mỗi loài có 15 cá thể.

Ba là, khi hai cây gỗ có D như nhau nhưng H khác nhau, thì V thân của chúng là khác nhau. Vì thế, nếu chỉ sử dụng G, thì chỉ số IVI phản ánh không chính xác khối lượng của các loài cây gỗ trong QXTV.

Bốn là chỉ số F% khó xác định đối với rừng tự nhiên hỗn loài nhiệt đới. Nói chung, những nhược điểm của chỉ số IVI chính là do hạn chế của phương pháp xác định độ thường gặp loài.

(3) Chỉ số ưu thế của Thái Văn Trừng và Vũ Tự Lập. Thái Văn Trừng và Vũ Tự Lập (1978; 1998) đã đề xuất chỉ số IVI của các loài cây gỗ trong QXTV rừng được xác định theo giá trị bình quân của 3 tham số N%, G% và thể tích thân tương đối (V%) (Công thức 2.12). Giá trị V% là tỷ lệ giữa V của từng loài cây gỗ và tổng thể tích của tất cả các loài trong QXTV. Đối với loài J, V_J được xác định theo công thức 2.13; trong đó g_{iJ} = tiết diện ngang thân của cây i thuộc loài J, H_{iJ} = chiều cao của cây i thuộc loài J, F = hình số (F = 0,5 đối với cây gỗ ở rừng trồng thuần loài đồng tuổi; F = 0,45 đối với cây gỗ ở rừng tự nhiên hỗn loài).

$$IVI = \frac{(N\% + G\% + V\%)}{3} \qquad (2.12)$$

$$V_J = g_{iJ} * H_{iJ} * F \qquad (2.13)$$

Phương pháp của Thái Văn Trừng và Vũ Tự Lập (1978; 1998) có nhiều ưu điểm. Một là cả ba chỉ tiêu (N%, G%, V%) đều được đo đạc và tính toán dễ dàng. Sử dụng chỉ tiêu V% thay cho chỉ tiêu F% không chỉ làm giảm khó khăn trong tính toán, mà còn phản ánh rõ khối lượng của mỗi loài cây gỗ. Thật vậy, bằng việc sử dụng những phần mềm thống kê thông dụng (MINITAB; STATGRAPHICS; SPSS), nhà lâm học có thể dễ dàng xác định ba tham số (N%, G%, V%). Mặt khác, hai loài cây gỗ có N và D bằng nhau, loài nào có H lớn hơn thì V và sinh khối của

loài đó lớn hơn. Như vậy, thay thế F% bằng V% đã phản ánh rõ hơn vai trò của loài trong QXTV. Hai là ba chỉ tiêu (N%, G%, V%) không phụ thuộc vào kiểu phân bố của các loài cây gỗ trên mặt đất. Ba là vai trò (sinh thái, kinh tế) của các loài có thể được đánh giá và so sánh không chỉ trên từng ô mẫu, mà còn trên nhiều ô mẫu. Ưu điểm thứ ba là rất đáng quan tâm, bởi vì trong thực tế những biện pháp lâm sinh được xây dựng không chỉ xuất phát từ đặc điểm chung của toàn bộ QXTV, mà còn của từng QXTV hay khoảnh rừng. Tuy vậy, phương pháp này cũng có một số nhược điểm. Một là vai trò của loài cây gỗ thay đổi tùy theo số lượng và kích thước ô mẫu. Hai là vai trò của loài có thể nhận giá trị khác nhau nếu sử dụng phương pháp rút mẫu khác nhau.

2.2.3. Phương pháp thu thập số liệu trong phân tích kết cấu loài cây gỗ

2.2.3.1. Xác định số lượng ô mẫu

Theo nguyên lý thống kê, khi biết hệ số biến động (CV%) và sai số theo yêu cầu của nghiên cứu (P%), thì dung lượng mẫu thích hợp được xác định gần đúng theo công thức 2.14.

$$n = \frac{CV^2}{P^2} \qquad (2.14)$$

Khi phân tích kết cấu loài cây gỗ của QXTV, một khó khăn lớn của giai đoạn thiết kế nghiên cứu là chọn loài nào hay nhóm loài nào để tính số lượng ô mẫu cần thu thập. Một cách giải quyết hợp lý là xác định số lượng ô mẫu theo biến động của chỉ số IVI của loài ưu thế. Để làm rõ vấn đề đặt ra, trước khi thiết kế nghiên cứu, nhà lâm học cần phải điều tra sơ bộ kết cấu loài cây gỗ của QXTV dựa trên số lượng ô mẫu nào đó, thông thường là 30 ô mẫu. Từ 30 ô mẫu này, tính kết cấu loài cây gỗ và xác định loài cây gỗ ưu thế trong các QXTV. Tiếp theo tính hệ số biến động đối với chỉ số IVI của loài ưu thế. Sau đó tính dung lượng mẫu cần thu thập. Thông thường, yêu cầu đối với sai số của số liệu điều tra là 5%. Trên đây là những bước cơ bản nhằm đảm bảo nhận được kết quả nghiên cứu với độ tin cậy nhất định.

2.2.3.2. Xác định phương pháp bố trí ô mẫu

Phương pháp bố trí ô mẫu trong nghiên cứu kết cấu loài cây gỗ là mối quan tâm của các nhà lâm học. Theo nguyên lý thống kê, các ô mẫu có thể được bố trí ngẫu nhiên, hệ thống, hệ thống – ngẫu nhiên, khối và điển hình. Để làm rõ vai trò của các loài cây gỗ trong QXTV, lâm học thường sử dụng ba phương pháp rút mẫu: hệ thống, khối và điển hình. Phương pháp rút mẫu hệ thống được sử dụng để xây dựng bản đồ về các đơn vị phân loại rừng (Kiểu rừng, kiểu QXTV). Trong mỗi vùng sinh thái (Ecozone), phương pháp rút mẫu theo khối được sử dụng để phân chia thảm thực vật thành những kiểu QXTV khác nhau. Sau khi phân chia các kiểu QXTV riêng biệt, nhà lâm học có thể sử dụng một trong ba phương pháp chọn mẫu (ngẫu nhiên, hệ thống và điển hình) để thu thập những đặc trưng của mỗi kiểu QXTV.

2.2.3.3. Xác định kích thước và hình dạng ô mẫu

Kích thước và hình dạng ô mẫu là mối quan tâm của các nhà lâm học. Về lý thuyết, kích thước ô mẫu lớn bao trùm lên không gian sống của nhiều loài cây gỗ. Trái lại, kích thước ô mẫu nhỏ chỉ bao trùm lên không gian sống của một số loài cây gỗ. Theo nguyên lý thống kê, để xác định số lượng và kích thước ô mẫu thích hợp, trước hết phân chia rừng thành những kiểu QXTV thuần nhất. Kế đến chọn một khoảnh rừng mẫu và thống kê đầy đủ các đặc tính của rừng (S, N, G, M...). Số liệu này được xem là số liệu thực của kiểu QXTV. Tiếp đó bố trí các ô mẫu với kích thước và số lượng khác nhau và phân tích kết cấu loài cây gỗ trong mỗi kiểu QXTV. Sau đó so sánh thành phần loài cây gỗ được xác định theo những ô mẫu với kích thước khác nhau với số liệu thực. Thành phần loài cây gỗ và số lượng ô mẫu được xác định theo các đường cong tích lũy loài. Cuối cùng, căn cứ vào độ chính xác theo yêu cầu của nghiên cứu, xác định kích thước và số lượng ô mẫu thích hợp.

Bởi vì rừng là hệ sinh thái vô cùng phức tạp và khó đi lại, nên nhà lâm học rất khó xác định chính xác số lượng và kích thước ô mẫu để phân tích kết cấu loài cây gỗ. Vì thế, các nhà lâm học khác nhau đã sử dụng số lượng và kích thước ô mẫu khác nhau. Khi phân tích cấu trúc và đa dạng loài cây gỗ của rừng tự nhiên ở đảo Andaman (Ấn độ), Tripathi và ctv (2004) đã sử dụng những ô mẫu 15*15 m. Peet và Christensen (1998) đã sử dụng những ô mẫu 0,10 ha để xác định kết cấu và đa dạng loài cây gỗ đối với rừng thứ sinh ở miền bắc Carolinna (Mỹ). Ở trung Quốc, Wang và ctv (2006), Lü và ctv (2010) đã sử dụng ô mẫu 1,0 ha để phân tích kết cấu và đa dạng loài cây gỗ của rừng mưa á nhiệt đới. Ở Malaysia, Suratman (2012) đã sử dụng những ô mẫu 0,20 ha để xác định kết cấu loài cây gỗ, cấu trúc quần thụ và đa dạng loài cây gỗ của rừng Sao Dầu ở Vườn quốc gia Pahang. Ở Việt Nam, nhiều tác giả (Nguyễn Văn Trương, 1984; Nguyễn Văn Thêm, 1992; Thái Văn Trừng, 1998; Vũ Tiến Hinh, 2012; Nguyễn Văn Thêm và Nguyễn Tuấn Bình, 2017; Nguyễn Văn Thêm và Vũ Mạnh, 2017; Lê Văn Long, 2019; Đào Thị Thùy Dương, 2019) đã sử dụng kích thước ô mẫu thay đổi từ 0,1 – 1,0 ha để phân tích kết cấu loài cây gỗ, cấu trúc quần thụ và đa dạng loài cây gỗ của rừng tự nhiên hỗn loài. Theo Thái Văn Trừng (1998), kích thước và hình dạng ô mẫu thay đổi tùy theo kiểu rừng. Kích ô mẫu thích hợp là 0,25 ha đối với kiểu thảm thực vật cây thân bụi; 0,5 ha đối với kiểu rừng thưa và rừng nửa thường xanh; 1,0 ha đối với kiểu rừng kín thường xanh. Các ô mẫu thường có dạng hình chữ nhật, hình vuông và hình tròn; trong đó dạng hình tròn có nhiều ưu điểm hơn.

Tái sinh rừng được thống kê trong những ô dạng bản với kích thước từ 1 – 100 m^2. Những ô dạng bản từ 1 – 4 m^2 được sử dụng để thống kê hạt giống và những cây mạ có H < 50 cm. Những ô dạng bản có kích thước từ 4 – 100 m^2 được sử dụng để thống kê những cây tái sinh thuộc cấp H > 100 cm. Ô dạng bản thường có dạng hình vuông. Các ô dạng bản được bố trí theo phương pháp hệ thống trên các tuyến song song trong QXTV trên ô mẫu. Khi phân tích phân bố cây tái sinh trên mặt đất, các ô dạng bản được bố trí theo hệ thống ô bàn cờ (Nguyễn Văn Thêm, 1992, 2010).

2.2.3.4. Thu thập thông tin trong các ô mẫu

Để phân tích kết cấu loài cây gỗ trong QXTV, những thông tin cần thu thập trong các ô mẫu bao gồm thành phần loài cây gỗ (S, loài), D (cm), H (m), đường kính tán (D_T, m) và chiều dài tán (L_T, m) của từng cây gỗ. Nếu thu thập sinh khối của mỗi loài, trước hết cắt phần trên và phần dưới mặt đất của các cây gỗ; sau đó cân đo trọng lượng (tươi, khô không khí hoặc khô tuyệt đối) của chúng trên ô mẫu. Thế nhưng, trọng lượng của các loài cây gỗ là một chỉ tiêu rất khó đo đạc chính xác, nhất là những cây gỗ to lớn. Trong thực tế, thay vì đo đạc trực tiếp sinh khối của các loài cây gỗ tại rừng, người ta xác định sinh khối của các thành phần ở mức cây gỗ và quần thụ thông qua mối quan hệ giữa sinh khối (khô, tươi) với một số nhân tố dễ đo đạc khác. Chẳng hạn: Trọng lượng thân, cành và lá cây gỗ và cây bụi có thể được ước lượng thông qua mối quan hệ giữa chúng với D và H thân. Khối lượng và độ che chiếu của các cơ quan dưới mặt đất là một chi tiêu rất quan trọng, nhưng việc xác định chúng cũng hết sức khó khăn. Vì thế, những thành phần này cũng được xác định thông qua mối quan hệ giữa chúng với các cơ quan trên mặt đất. Tuy nhiên, vì tính phức tạp trong việc xác định trọng lượng hoặc độ che chiếu của các cơ quan thực vật nằm sâu dưới đất, nên trong thực tế người ta thường chỉ tính độ che chiếu và trọng lượng của các cơ quan trên mặt đất.

Để phân tích mối quan hệ giữa kết cấu loài cây gỗ với các nhân tố sinh thái, những thông tin cần được thu thập bao gồm khí hậu (ánh sáng, nhiệt độ, độ ẩm...), địa hình (độ cao, độc dốc, hướng dốc...), đất (độ dày tầng đất, độ chặt, độ xốp, độ ẩm, thành phần cơ giới, tính chất lý hóa tính...) và những hoạt động lâm sinh.

2.2.3.5. Phân tích và báo cáo kết cấu loài cây gỗ

Khi phân tích kết cấu loài cây gỗ, trước hết nhà lâm học cần xác định rõ phương pháp xác định chỉ số IVI của các loài cây gỗ. Việc chọn phương pháp nào là tùy thuộc vào quan điểm của từng nhà lâm học. Một câu hỏi đặt ra là kết cấu loài cây gỗ cần được báo cáo như thế nào? Về nguyên tắc, để hiểu rõ về kết cấu loài cây gỗ của các QXTV, số liệu báo cáo phải bao gồm đầy đủ thông tin về loài cây gỗ và các chỉ tiêu thuyết minh mức độ phong phú của chúng. Để dễ dàng tổng hợp và phân tích số liệu, kết cấu loài cây gỗ của các QXTV cần được báo cáo ngắn gọn dưới dạng "Bảng kết cấu loài cây gỗ" hoặc "Công thức kết cấu loài cây gỗ". Theo quy ước của lâm học, bảng và công thức kết cấu loài cây gỗ đều phải cung cấp đầy đủ các thông tin theo thứ tự từ những loài cây gỗ quý, hiếm hoặc có giá trị cao về kinh tế đến loài cây gỗ ưu thế, loài cây gỗ đồng ưu thế và một số loài cây gỗ phụ thuộc có chỉ số IVI cao. Yêu cầu này được đặt ra nhằm giúp cho các nhà lâm học và các nhà kinh doanh rừng nhận được đầy đủ các thông tin về kết cấu loài cây gỗ của các QXTV.

Khi báo cáo kết cấu loài cây gỗ ở dạng công thức, tên đầy đủ của các loài cây gỗ được viết tắt bằng tên tiếng Việt hoặc bằng tên khoa học. Tên viết tắt là 1, 2 hoặc 3 kí tự đầu của tên loài. Ví dụ: Sao đen = Sđ; Dầu rái = Dr; Dầu song nàng = Dsn. Khi hai hay nhiều loài cây gỗ có cùng tên viết tắt như nhau, thì tên của những loài

sau được viết thêm một kí tự. Ví dụ: Sao đen = Sđ; Săng đen = Sađ; Bằng lăng ổi = Bl; Bằng lăng cườm = Blc. Khi báo cáo kết cấu loài cây gỗ ở dạng công thức, lâm học quy ước chỉ báo cáo những loài cây gỗ quý, hiếm hoặc có giá trị cao về kinh tế, loài cây gỗ ưu thế, loài cây gỗ đồng ưu thế và một số loài cây gỗ phụ thuộc có chỉ số IVI ≥ 4%. Lưu ý rằng tổng các chỉ số IVI của tất cả các loài phải bằng 1 hoặc 100%.

Kết cấu loài cây gỗ được báo cáo rõ ràng nhất ở dạng bảng. Đối với những QXTV chỉ bao gồm dưới 25 loài cây gỗ, bảng kết cấu loài cây gỗ cần được trình bày đầy đủ số lượng loài cây gỗ và các chỉ tiêu thuyết minh độ phong phú của chúng (S, N, F, G, V). Đối với những QXTV có trên 25 loài cây gỗ, bảng kết cấu loài cây gỗ chỉ báo cáo tóm tắt những thông tin cơ bản của những loài cây gỗ quý, hiếm hoặc có giá trị cao về kinh tế, loài cây gỗ ưu thế, loài cây gỗ đồng ưu thế và một số loài cây gỗ phụ thuộc có chỉ số IVI ≥ 4%. Một cách khác là bảng kết cấu loài cây gỗ báo cáo tóm tắt các thông tin của 10 – 20 loài cây gỗ có chỉ số IVI từ cao nhất trở xuống. Quy định này có liên quan đến kích thước của trang giấy với khổ A_4. Nói chung, kết cấu loài cây gỗ được báo cáo ở dạng bảng có nhiều ưu điểm hơn so với dạng công thức. Lý do là vì bảng kết cấu loài cây gỗ cung cấp nhiều thông tin (S, N, F, G, V) và dễ theo dõi hơn.

Dưới đây dẫn ví dụ về phân tích kết cấu loài cây gỗ của ưu hợp Dầu con rái theo hai phương pháp tính chỉ số IVI của các loài cây gỗ. Phương pháp 1 tính chỉ số IVI của các loài cây gỗ theo hai tham số N% và G% (Phụ lục 1). Phương pháp 2 tính chỉ số IVI của các loài cây gỗ theo ba tham số N%, G% và V% (Phụ lục 2) của Thái Văn Trừng (1998). Số liệu tính toán bao gồm 1 ô mẫu với kích thước 0,25 ha. Ưu hợp Dầu con rái bao gồm 25 loài cây gỗ; trong đó Dầu con rái là loài ưu thế sinh thái, Bằng lăng ổi (*Lagerstroemia calyculata* Kurz) là loài đồng ưu thế sinh thái. Bảng 2.2 và Bảng 2.3 ghi lại tóm tắt kết cấu loài cây gỗ được xác định theo hai phương pháp kể trên.

Kết cấu loài cây gỗ của các QXTV cần được thuyết minh theo thứ tự từ tổng số họ, số chi và số loài cây gỗ bắt gặp, loài cây gỗ ưu thế sinh thái, loài cây gỗ đồng ưu thế sinh thái, những loài cây gỗ phụ thuộc và những loài cây gỗ quý, hiến. Mỗi nhóm loài này cần thuyết minh các thông tin về F và F%, N và N%, G và G%, M và M% và chỉ số IVI trung bình. Sự khác biệt về chỉ số IVI của các loài cây gỗ trong những QXTV khác nhau được so sánh bằng đường cong chỉ số IVI.

2.2.3.6. Phân tích kết cấu loài cây gỗ bằng phần mềm SPSS

Kết cấu loài cây gỗ của QXTV có thể được xác định dễ dàng bằng phần mềm thống kê SPSS for Window. Cách thức phân tích kết cấu loài cây gỗ của QXTV bằng phần mềm thống kê SPSS được thực hiện theo 5 bước dưới đây.

Bước 1. Mở SPSS và ghi tên loài, chi và họ bằng tiếng Việt và tiếng La Tinh vào 5 cột liên tiếp nhau. Những cột khác ghi các thông tin biểu thị kích thước thân cây (D, H, V...) (Hình 2.1). Quy ước này nhằm tạo điều kiện thuận lợi cho việc thống kê nhanh số loài, số chi và số họ cây gỗ.

Bảng 2.2. Kết cấu loài cây gỗ của ưu hợp Dầu con rái được xác định theo hai tham số N% và G%. Đơn vị tính: 1 ha.

TT	Loài cây gỗ	N (cây)	G (m^2)	V (m^2)	N%	G%	IVI%
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(9)
1	Dầu con rái	148	9,8	97,7	20,8	24,0	22,4
2	Bằng lăng ổi	108	10,4	103,8	15,2	25,4	20,3
3	Cầy	80	7,7	74,5	11,2	18,8	15,0
4	Trường lá nhỏ	68	3,7	38,9	9,6	8,9	9,3
5	Trâm vỏ đỏ	84	2,5	19,7	11,8	6,0	8,9
6	Cám	36	1,3	10,8	5,1	3,2	4,2
	Cộng 6 loài	524	35,4	345,4	73,7	86,3	80,0
19	Loài khác	188	5,5	45,1	26,3	13,7	20,0
25	Tổng số	712	40,9	390,4	100	100	100

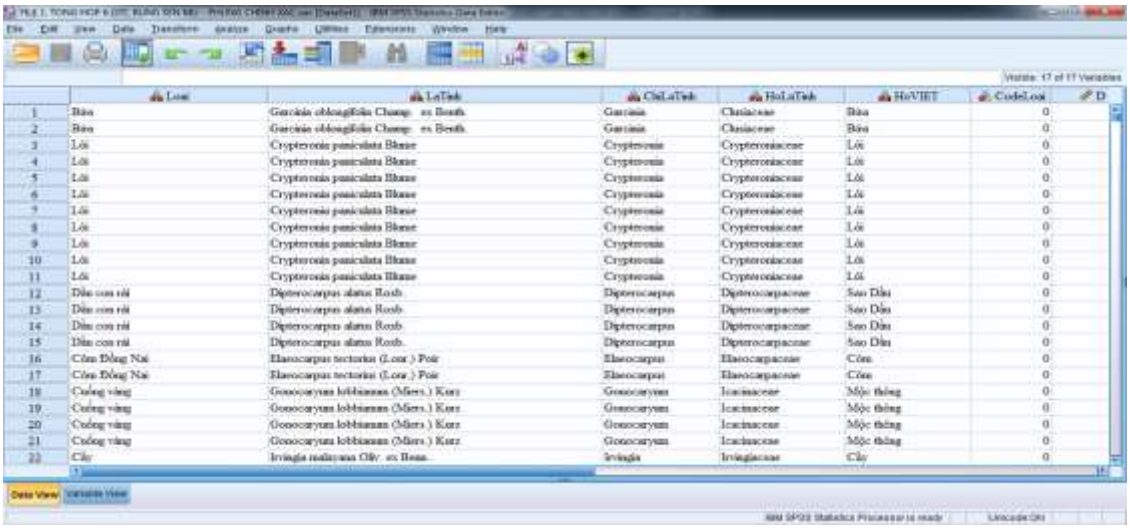
Bảng 2.3. Kết cấu loài cây gỗ của ưu hợp Dầu con rái được xác định theo theo ba tham số N%, G% và V%. Đơn vị tính: 1 ha.

TT	Loài cây gỗ	N(cây)	G(m^2)	V(m^2)	N%	G%	V%	IVI%
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	Dầu con rái	148	9,8	97,7	20,8	24,0	25,0	23,3
2	Bằng lăng ổi	108	10,4	103,8	15,2	25,4	26,6	22,4
3	Cầy	80	7,7	74,5	11,2	18,8	19,1	16,4
4	Trường lá nhỏ	68	3,7	38,9	9,6	8,9	10,0	9,5
5	Trâm vỏ đỏ	84	2,5	19,7	11,8	6,0	5,0	7,6
6	Cám	36	1,3	10,8	5,1	3,2	2,8	3,7
	Cộng 6 loài	524	35,4	345,4	73,7	86,3	88,5	82,9
19	Loài khác	188	5,5	45,1	26,3	13,7	11,5	17,1
25	Tổng số	712	40,9	390,4	100	100	100	100

Bước 2. Từ bảng tính ở Hình 2.1, mở Analyze > Reports > OLAP Cubes. Đến đây một cửa sổ xuất hiện. Tại cửa sổ này, chọn G, V và Loài; sau đó chuyển chúng vào ô "Summary Variables" và Grouping Variables (Hình 2.2). Thủ tục này cho phép xác định các đặc trưng thống kê đối với N, G và M của từng loài cây và tất cả các loài cây trong ô mẫu.

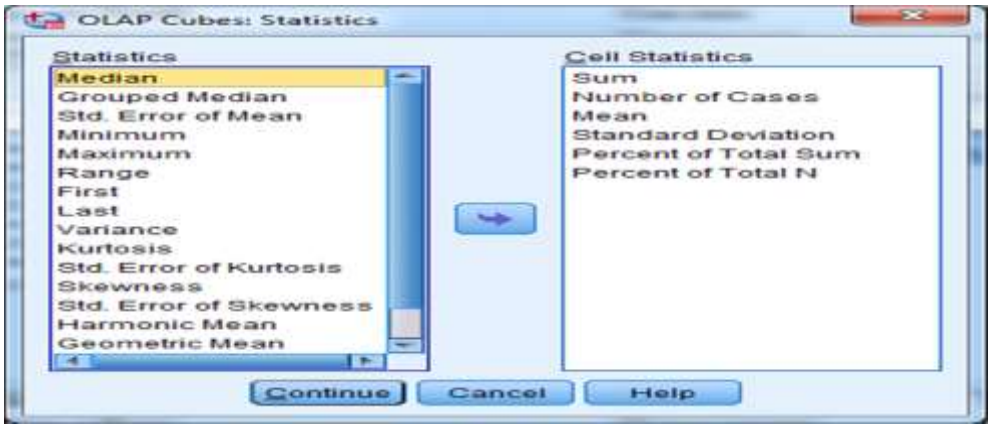
Bước 3. Xác định các đặc trưng thống kê đối với N, G và V của từng loài cây gỗ và tất cả các loài cây gỗ trong ô mẫu. Trước hết chọn Statistics ở Hình 2.2. Đến đây một cửa sổ xuất hiện như Hình 2.3. Kế đến, tại cửa sổ này chọn và chỉ giữ lại

hai thống kê Number of Cases (Số cây) và Sum (Tổng N, G và V) trong ô "Cell Statistics" (Hình 2.4).

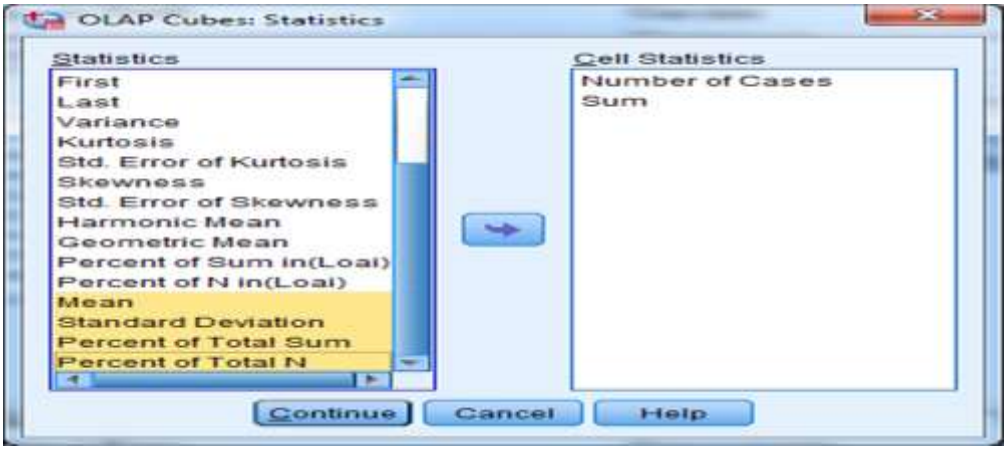


Hình 2.1. Những thông tin về các loài cây gỗ và độ phong phú của chúng.

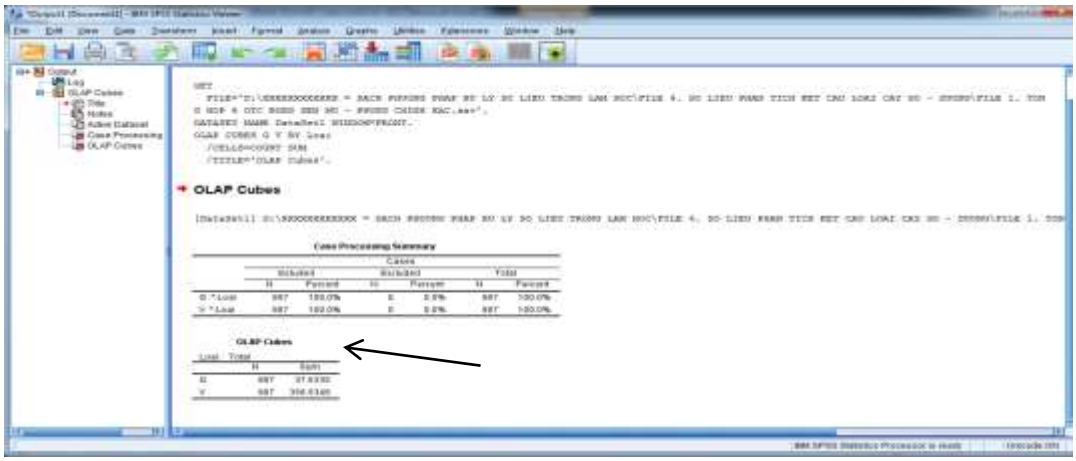
Hình 2.2. Xác định kết cấu loài cây gỗ theo ba biến N, G và V.



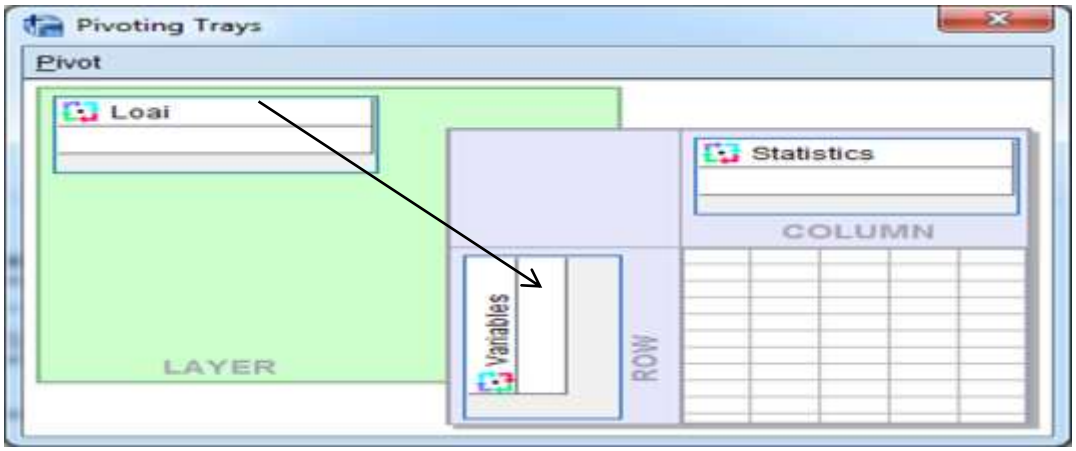
Hình 2.3. Các đặc trưng thống kê N, G và V đối với các loài cây gỗ.



Hình 2.4. Tóm tắt các đặc trưng thống kê N, G và V đối với các loài cây gỗ.

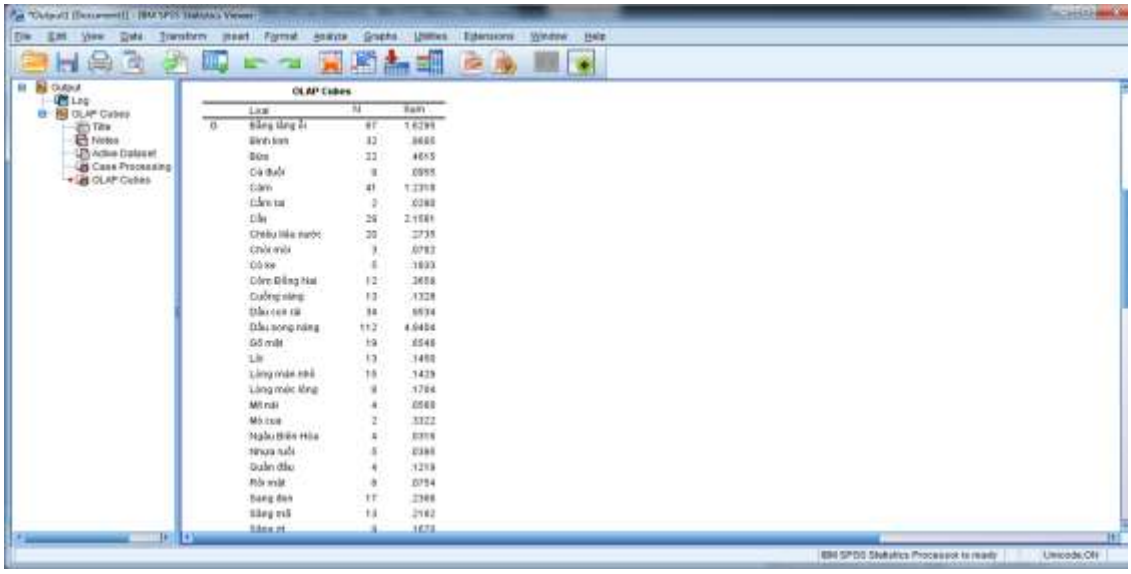


Hình 2.5. Báo cáo các đặc trưng thống kê N, G và V đối với các loài cây gỗ.



Hình 2.6. Khay báo cáo các biến N, G và V đối với các loài cây gỗ.

Bước 4. Thống kê N, G và V của từng loài cây gỗ và tất cả các loài cây gỗ trong ô mẫu. Để đạt được mục đích này, trước hết chọn Continue ở Hình 2.4. Sau đó chọn OK để nhận kết quả tính toán (Hình 2.5). Tại cửa sổ này, chọn và nhấp đúp con trỏ chuột vào OLAP Cubes. Đến đây một cửa sổ xuất hiện như ở Hình 2.6. Tại cửa sổ ở Hình 2.6, chọn biến "Loài" và giữ chặt con trỏ chuột và chuyển nó vào ô Variables. Kết quả nhận được thông tin về N, G và M của từng loài cây gỗ và tổng số loài cây gỗ (Hình 2.7).



Hình 2.7. Thống kê các biến N, G và V đối với các loài cây gỗ.

Bước 5. Xác định kết cấu loài cây gỗ. Trước hết Copy ba cột (Loài, N, Sum) và chuyển chúng vào 1 File trong Microsorft Word. Kế đến lại Copy ba cột này ở File trong Microsorft Word và dán chúng vào 1 File trong bảng tính của Microsorft Excel. Tiếp theo tính toán N, G, M và chỉ số IVI của từng loài trên ô mẫu và chuyển đổi ra đơn vị 1 ha. Sau đó sắp xếp các loài theo thứ tự giảm dần của chỉ số IVI. Thủ tục này tạo điều kiện thuận lợi cho việc tổng hợp và báo cáo các thông tin về loài cây gỗ ưu thế, loài cây gỗ đồng ưu thế và các loài cây gỗ phụ thuộc. Bằng cách thay thế biến "Loài" bằng biến "Chi" và biến "Họ" và thực hiện 5 bước như trên, chúng ta nhận được các đặc trưng thống kê đối với N, G và M của từng chi và họ.

2.2.3.7. Thảo luận chung về phương pháp phân tích kết cấu loài cây gỗ

(a) Số lượng, kích thước và phương pháp bố trí ô mẫu. Mục đích chủ yếu của phân tích kết cấu loài cây gỗ là cung cấp thông tin về loài cây gỗ của các kiểu QXTV để so sánh các kiểu QXTV khác nhau. Ngoài ra, những thông tin về kết cấu loài cây gỗ cũng là căn cứ hữu ích cho quản lý rừng và kinh doanh rừng. Kết cấu loài cây gỗ không chỉ phụ thuộc vào số lượng và kích thước ô mẫu, mà còn vào phương pháp bố trí ô mẫu. Khi phân tích kết cấu loài cây gỗ, vấn đề quan trọng là thống kê chính xác thành phần loài cây gỗ và độ phong phú của chúng trong đối tượng nghiên cứu. Độ chính xác của thống kê thành phần loài cây gỗ gia tăng theo

số lượng ô mẫu và kích thước ô mẫu. Phương pháp bố trí ô mẫu ngẫu nhiên cho phép phân tích và đánh giá kết cấu loài cây gỗ của QXTV khách quan và chỉ rõ mức độ tin cậy của kết quả. Tuy vậy, việc làm này lại dẫn đến những khó khăn cho việc thu thập số liệu. Vì thế, khi nghiên cứu về kết cấu loài cây gỗ trong những QXTV rừng, nhà lâm học cần phải báo cáo rõ kích thước ô mẫu, số lượng ô mẫu và phương pháp bố trí ô mẫu. Cấu trúc của những QXTV rừng thay đổi tùy theo mức độ đa dạng về loài cây gỗ. Vì thế, kích thước mẫu nhỏ được sử dụng để đánh giá kết cấu và đa dạng loài cây gỗ của những QXTV có nhiều loài. Ngược lại, kích thước mẫu lớn được sử dụng để đánh giá kết cấu và đa dạng loài cây gỗ của những QXTV có nhiều loài hiếm gặp.

Để đảm bảo thu thập chính xác thành phần loài cây gỗ của QXTV, nhà lâm học phải sử dụng công cụ thống kê toán học. Thế nhưng, thảm thực vật rừng là một hệ thống tự nhiên vô cùng phức tạp; trong đó bao gồm vô số QXTV mà ranh giới của chúng không rõ ràng. Nếu máy móc làm theo toán học, thì nhà lâm học phải thu thập rất nhiều ô mẫu. Điều này là không khả thi đối với nhiều nghiên cứu trong lâm học. Mặt khác, nếu phân chia các đơn vị kinh doanh rừng theo từng QXTV, thì nhà quản lý rừng cũng không thể quản lý tốt các QXTV. Để đảm bảo tính khả thi trong một dự án nghiên cứu và tính đơn giản trong quản lý rừng, thảm thực vật rừng cần được phân chia thành các kiểu QXTV dựa theo những tiêu chuẩn trội. Sau đó xác định số lượng ô mẫu cho từng kiểu QXTV dựa theo nguyên lý thống kê. Thế nhưng, do ranh giới của các kiểu QXTV là không rõ ràng, nên nhà lâm học không thể dễ dàng xác định đúng vị trí của các ô mẫu bằng phương pháp chọn mẫu ngẫu nhiên. Để khắc phục khó khăn này, các nhà lâm học thường sử dụng phương pháp chọn mẫu điển hình. Điều này được giải thích là do phương pháp này không chỉ tránh được việc đi trệch mục tiêu nghiên cứu, mà còn làm giảm những tính toán và các chi phí về thời gian, nhân lực và kinh phí. Số lượng ô mẫu thích hợp thường chỉ được xác định dựa theo đường cong tích lũy loài. Đường cong này được xây dựng dựa theo đồ thị biểu diễn mối quan hệ giữa số loài (S) với số lượng ô mẫu (N_i). Số lượng ô mẫu thích hợp được xác định tương ứng với số loài tiệm cận dần đến một giá trị ổn định. Kích thước ô mẫu thay đổi tùy theo mục đích nghiên cứu và diện tích khu vực nghiên cứu. Thông thường, lâm học sử dụng kích thước ô mẫu 0,1 – 1,0 ha để phân tích cây trưởng thành; 1 – 100 m^2 để phân tích tái sinh rừng. Hình dạng ô mẫu có nhiều dạng, nhưng thông thường là hình chữ nhật hay hình vuông. Khi so sánh kết cấu loài cây gỗ thay đổi theo môi trường và các phương thức lâm sinh, thì nhà lâm học cần sử dụng các ô mẫu có kích thước và hình dạng tương tự như nhau.

Khi sử dụng các kích thước ô mẫu khác nhau, thì kết quả so sánh phải được chuyển về cùng kích thước, thông thường là 1 ha. Những biến N, G, V có thể chuyển đổi dễ dàng bằng cách nhân với hệ số. Trái lại, số loài cây gỗ không thể chuyển đổi sang đơn vị 1 ha, bởi vì số loài cây gỗ không chỉ phụ thuộc vào số lượng và kích thước ô mẫu, mà còn cả phương pháp bố trí ô mẫu. Cùng một diện tích thu mẫu, nếu phân chia nhỏ thành nhiều ô mẫu, thì số loài bắt gặp nhiều hơn.

(b) Phương pháp xác định kết cấu loài cây gỗ. Sử dụng phương pháp xác định kết cấu loài cây gỗ khác nhau dẫn đến những kết luận khác nhau. Chẳng hạn: Cùng một chuỗi số liệu như nhau, nhà lâm học này xác định chỉ số IVI của các loài cây gỗ theo N%, F% và G%, còn nhà lâm học khác lại tính theo N%, G% và V%. Chính điều đó đã dẫn đến những tranh luận về độ tin cậy của kết quả nghiên cứu. Để kết quả nghiên cứu dễ được chấp nhận, các nhà sinh thái học và lâm học thường sử dụng những chỉ tiêu và tiêu chuẩn được mọi người thừa nhận. Quy ước này nhằm mục đích báo cáo kết quả thống nhất và tránh những tranh luận không cần thiết. Nói chung, khi báo cáo về kết cấu và đa dạng loài cây gỗ, nhà lâm học phải quy ước rõ thành phần cây gỗ được phân chia theo hệ thống phân loại nào. Tương tự, khi phân tích so sánh kết cấu và đa dạng loài cây gỗ theo kiểu rừng, kiểu lập địa, dạng địa hình và loại đất…, nếu sử dụng hệ thống phân loại rừng khác nhau, nhà lâm học phải quy ước rõ hệ thống phân chia những đối tượng này.

2.3. Phương pháp so sánh sự tương đồng về loài cây gỗ giữa các QXTV

Sự tương đồng hay sự giống nhau về thành phần loài cây gỗ giữa các QXTV được so sánh bằng hệ số tương đồng. Dưới đây giới thiệu 4 hệ số tương đồng thường được các nhà sinh thái học và lâm học sử dụng.

Theo Jaccard (1901; Dẫn theo Vasilevich, 1969), hệ số tương đồng (C_J%) giữa hai $QXTV_i$ và $QXTV_J$ được xác định theo công thức 2.15; trong đó S_{ij} = số loài cây gỗ giống nhau của hai $QXTV_i$ và $QXTV_J$, S_i = số loài cây gỗ của $QXTV_i$, S_J = số loài cây gỗ của $QXTV_J$. Hệ số không tương đồng hay hệ số khác nhau giữa hai QXTV là $1 – C_J$.

$$C_J = \frac{S_{ij}}{(S_i + S_J - S_{iJ})}*100 \qquad (2.15)$$

Sorensen (1948; Dẫn theo Vasilevich, 1969) xác định hệ số tương đồng (C_S%) giữa hai QXTV theo công thức 2.16; trong đó S_{iJ} = số loài cây gỗ giống nhau của hai $QXTV_i$ và $QXTV_J$, S_i = số loài cây gỗ của QXTVi, S_J = số loài cây gỗ của $QXTV_J$. Hệ số không tương đồng hay hệ số khác nhau giữa hai QXTV là $1 – C_S$.

$$C_S = \frac{2*S_{iJ}}{(S_i + S_J)}*100 \qquad (2.16)$$

Lưu ý rằng, cùng nguồn số liệu như nhau nhưng chúng ta nhận được hai hệ số C_J và C_S là khác nhau. Ví dụ: $QXTV_i$ và $QXTV_j$ tương ứng có 35 loài và 40 loài. Số loài giống nhau giữa hai QXTV này là 25. Giải: Hệ số tương đồng của Jaccard C_J = (25/(35 + 40 - 25)) = 50,0%. Hệ số tương đồng của Sorensen C_S = (2*25/(35 + 40)) = 66,7%. Vì thế, khi phân tích sự tương đồng về loài cây gỗ giữa các QXTV, nhà lâm học cần phải báo cáo rõ phương pháp tính hệ số tương đồng.

Bray và Curtis (1957; Dẫn theo Vasilevich, 1969) xác định hệ số không tương đồng (C_{BC}%) giữa hai QXTV theo công thức 2.17; trong đó T_i và T_J tương ứng là tổng số cá thể của các loài ở $QXTV_i$ và $QXTV_J$, T_{iJ} là tổng số cá thể nhỏ nhất của mỗi loài trong hai QXTV này. Hệ số C_{BCiJ} nhận giá trị giữa 0 và 1. Khi C_{BCiJ} =

0, thì hai QXTV có thành phần loài cây gỗ giống nhau. Khi $C_{BCiJ} = 1$, thì hai QXTV không có thành phần loài cây gỗ nào giống nhau.

$$C_{BCiJ} = \left(1 - \frac{2*T_{iJ}}{T_i + T_J}\right)*100 \qquad (2.17)$$

Ví dụ: Ba loài cây A, B và C ở $QXTV_i$ có số cá thể tương ứng là 15, 12 và 7. Ba loài cây A, B và C ở $QXTV_J$ có số cá thể tương ứng là 10, 6 và 8. Giải: T_i = (15 + 12 + 7) = 34. T_J = (10 + 6 + 8) = 24. T_{iJ} = (10 + 6 + 7) = 23. C_{BCiJ} = (1 – (2*23)/(34 + 24) = 0,207 hay 20,7%. Kết luận: Hai QXTV này chỉ khác nhau 20,7%. Nói cách khác, hai QXTV này giống nhau 79,3%.

Koch (1957; Dẫn theo Vasilevich, 1969) đề xuất hệ số tương đồng (IBD = Index of Biologic Dispersity) giữa hai QXTV được xác định theo công thức 2.18; trong đó S là tổng số loài giống nhau trong n ô mẫu của hai QXTV, T = (S_1+ S_2) là tổng số loài của QXTV 1 và QXTV 2. Nếu hai QXTV không có các loài chung, thì T = S và IBD = 0. Trong trường hợp tất cả các loài của hai QXTV là giống nhau, thì IBD = 100%.

$$IBD = \frac{(T - S)*100}{(n - 1)*S} \qquad (2.18)$$

Hệ số tương đồng của Jaccard và Sorensen được ứng dụng rộng rãi trong phân tích kết cấu loài cây gỗ và phân loại rừng. Lý do là vì hai hệ số này dễ tính toán, cho kết quả khá rõ ràng, đáng tin cậy hơn về mức độ tương đồng giữa các QXTV. Hai phương pháp này có thiếu sót là chỉ dựa vào sự có mặt của các loài mà không kể đến độ phong phú của chúng. Ngoài ra, cùng một số liệu như nhau, nhưng hệ số tương đồng của Jaccard và Sorensen lại nhận kết quả khác nhau. Vì thế, khi báo cáo kết quả, nhà lâm học cần nói rõ phương pháp xác định hệ số tương đồng.

Phân tích sự tương đồng giữa thành phần cây trưởng thành và thành phần cây tái sinh dưới tán rừng là cơ sở để đánh giá tính ổn định của QXTV. Khi hệ số C_S giữa thành phần cây trưởng thành và thành phần cây tái sinh lớn hơn 60%, thì thành phần loài cây gỗ của các QXTV rừng là ổn định theo thời gian. Trái lại, khi hệ số C_S < 60%, thì QXTV là chưa ổn định hay đang trong quá trình phát triển để đạt đến thế ổn định với môi trường bên ngoài.

2.4. Phân tích đa dạng loài cây gỗ trong các QXTV

2.4.1. Khái niệm về đa dạng sinh vật

Hiện nay có nhiều định nghĩa về đa dạng sinh học. Các nhà sinh học định nghĩa đa dạng sinh học là tổng số nguồn gen, tổng số loài và tổng số hệ sinh thái của một khu vực nhất định. Gaston (2000) định nghĩa đa dạng sinh vật là sự khác nhau ở tất các mức tổ chức sinh học - đó là gien, cá thể và hệ sinh thái. Đa dạng gien biểu thị sự đa dạng trong quần thể sinh vật hoặc giữa các quần thể sinh vật. Chẳng hạn đa dạng về những thành phần axit nucleic. Đa dạng loài biểu thị số lượng loài sinh vật trong một môi trường nhất định hay nơi ở nhất định. Đa dạng hệ sinh thái biểu thị số lượng hệ sinh thái trong một khu vực nhất định. Nhiều nhà sinh

thái học (Kimmins, J. P., 1998; Gaston, 2000; Magurran, 1988, 2004) cho rằng, bởi vì sự sống trên trái đất tồn tại chủ yếu ở dạng loài, nên thuật ngữ "Đa dạng sinh vật" thường được sử dụng với nghĩa là "Đa dạng loài".

Đa dạng sinh vật trong một cảnh quan hay một khu vực địa lý nhất định là đa dạng Gamma (γ - Diversity). Đa dạng Gamma biểu thị sự giàu có về loài của những quần xã sinh vật khác nhau trong một cảnh quan hay một khu vực địa lý nhất định. Đa dạng Gamma được xác định bằng cách tính đa dạng của tất cả các mẫu thuộc tất cả các quần xã gộp lại. Đa dạng Gamma bằng đa dạng Alpha (α - Diversity) cộng với đa dạng Beta (β -Diversity). Đa dạng Alpha là đa dạng sinh vật trong một vi môi trường sống nhất định hoặc một quần xã sinh vật nhất định. Đa dạng Alpha là giá trị bình quân của những chỉ số đa dạng loài được tính từ nhiều ô mẫu trong một quần xã sinh vật nhất định. Đa dạng Beta là đa dạng loài của nhiều quần xã sinh vật trong những môi trường khác nhau hay cảnh quan khác nhau. Đa dạng Beta được xác định bằng cách gộp chung nhiều quần xã trong những môi trường khác nhau. Nói chung, giá trị của các chỉ số đa dạng Alpha, Beta và Gamma thay đổi tùy theo phạm vi đối tượng thu mẫu (Whittaker, 1972, 1977).

Đa dạng sinh vật thay đổi tùy theo môi trường. Những yếu tố ảnh hưởng đến đa dạng sinh vật có thể được phân chia thành một số nhóm. Nhóm 1 là những yếu tố biểu hiện trên phạm vi không gian rộng lớn như yếu tố địa lý (vĩ độ, kinh độ, độ cao) và độ sâu của biển và đại dương. Nhóm 2 bao gồm những yếu tố có liên quan đến vĩ độ như biến động khí hậu. Nhóm 3 là những rối loạn trong môi trường, sự cách ly về địa lý và những khác biệt về thành phần môi trường vô cơ. Nhóm 4 là những yếu tố thuộc về những sinh vật như số lượng loài và khả năng cạnh tranh giữa các loài, sự thay đổi về không gian, cấu trúc và giai đoạn diễn thế của quần xã sinh vật. Vì thế, nghiên cứu và giải thích mối quan hệ giữa đa dạng sinh vật với những yếu tố môi trường là mối quan tâm to lớn của các nhà sinh thái học và lâm học.

2.4.2. Tầm quan trọng của những nghiên cứu về đa dạng sinh vật

Sau Hội nghị về trái đất vào năm 1992 tại Rio de Janeiro (Braxin), đa dạng sinh vật đã trở thành mối quan tâm không chỉ đối với các nhà sinh thái học và lâm học, mà còn cả các nhà môi trường, các nhà lập kế hoạch và các nhà quản lý. Nhiều nước là thành viên của Công ước về đa dạng sinh học. Theo công ước này, mỗi quốc gia đều có nghĩa vụ và trách nhiệm ghi nhận những loài cây và động vật xuất hiện ở nước mình, đồng thời đánh giá chính xác đa dạng sinh vật và xác định những chiến lược quản lý bền vững để bảo tồn đa dạng sinh vật.

Hiện nay các nước phải tập trung nghiên cứu đa dạng sinh vật bởi nhiều lý do khác nhau. Một là mỗi quốc gia cần phải bảo tồn đa dạng sinh vật. Nguyên nhân là vì nhiều hệ sinh thái rừng tự nhiên đã bị khai thác kiệt hoặc có nguy cơ biến mất. Hai là con người là một trong những thành phần cấu thành mạng lưới những hệ sinh thái. Sự sống của con người phụ thuộc vào không khí sạch, nguồn nước sạch, nguồn thức ăn an toàn và thuốc chữa bệnh...Vì sự sống còn, con người đã tác động đến hầu

hết các hệ sinh thái. Kết quả đã làm suy yếu hoặc làm biến mất nhiều hệ sinh thái mà sự sống của con người phụ thuộc vào đó. Ba là đa dạng sinh vật không chỉ là mối quan tâm to lớn của các nhà sinh học, sinh thái học, lâm học và môi trường, mà còn cả các nhà quản lý và hoạch định chính sách. Bốn là hiện nay khoa học và thực tiễn vẫn còn những thảo luận khác nhau xung quanh những số đo đa dạng sinh vật. Trước đây các nhà sinh thái học đã đề xuất rất nhiều chỉ số và những mô hình để đo đạc đa dạng sinh vật tùy theo môi trường, nơi sinh sống và vị trí khác nhau của sinh vật. Tuy vậy, phần lớn những số đo đa dạng không phải là những chỉ số đo đa dạng thực, mà chỉ là những entropy. Vì thế, nghiên cứu những chỉ số đa dạng và phương pháp đánh giá chính xác đa dạng vẫn cần phải được đặt ra. Năm là những thông tin về kết cấu loài cây gỗ, cấu trúc quần thụ, đa dạng loài cây gỗ và các quần xã thực vật là cơ sở khoa học cho quy hoạch rừng và bảo tồn đa dạng sinh vật rừng.

Nói chung, chỉ số đa dạng là một công cụ để đo đạc và so sánh sức khỏe của các hệ sinh thái ở những nơi khác nhau; đánh giá sự thay đổi đa dạng sinh vật ở một nơi nào đó theo thời gian. Việc đánh giá có thể giúp xác định những nguyên nhân dẫn đến sự thay đổi đa dạng sinh vật.

2.4.3. Những số đo đa dạng sinh vật

Theo Magurran (2004), hiện nay có rất nhiều số đo đa dạng sinh vật và chúng ta rất khó quyết định phương pháp đo đa dạng nào là phương pháp tốt nhất. Một cách thức để chọn số đo đa dạng sinh vật là kiểm định hiệu lực của phương pháp với số liệu thực tế. Về mặt khoa học, một chỉ số đa dạng tốt phải có khả năng phân loại được những hệ sinh thái và những vị trí khác nhau. Ngoài ra, nó không phụ thuộc vào kích thước ô mẫu và thành phần đa dạng được đo đạc, đồng thời được sử dụng rộng rãi và dễ hiểu. Trong sinh thái học, đa dạng sinh vật của một khu vực nào đó được xác định thông qua ba số đo hay ba thành phần đa dạng. Một là sự giàu có về loài (Species Richness). Hai là đa dạng về loài (Species Diversity). Ba là tính đồng đều về phân bố độ phong phú hay độ ưu thế của các loài sinh vật (Evenness of Species Abundance).

Sự giàu có về loài sinh vật biểu thị số lượng loài sinh vật bắt gặp trong những mẫu đo đếm. Sự giàu có về loài sinh vật chỉ đơn giản cho biết số loài sinh vật của đối tượng nghiên cứu. Sự giàu có về loài sinh vật không tính đến độ phong phú của các loài sinh vật. Hệ sinh thái nào càng nhiều loài sinh vật thì hệ sinh thái đó càng giàu có về loài sinh vật.

Mức độ đồng đều về độ phong phú của các loài sinh vật biểu thị phân bố độ phong phú tương đối của các loài sinh vật trong quần xã ở một khu vực nào đó. Sự đồng đều về độ phong phú của các loài sinh vật được tính toán nhằm mục đích so sánh sự tương đồng về kích thước quần thể (độ ưu thế) của mỗi loài sinh vật bắt gặp ở một khu vực nào đó.

2.4.3.1. Những số đo sự giàu có về loài sinh vật

Sự giàu có về loài sinh vật biểu thị số loài (S, loài) có trong quần xã sinh vật. Theo Whittaker (1972; 1977), sự giàu có về loài có thể được đo bằng chỉ số của

Margalef (Công thức 2.19) và chỉ số Menhinick (Công thức 2.20). Ở công thức 2.19 và 2.20, S = số loài trong quần xã; N = tổng số cá thể của S loài trong quần xã.

$$I_{Margalef} = \frac{S - 1}{Ln(N)} \quad (2.19)$$

$$I_{Menhinick} = \frac{S}{\sqrt{N}} \quad (2.20)$$

Cả hai chỉ số Margalef và Menhinick đều thay đổi theo kích thước và số lượng ô mẫu. Chỉ số Margalef nhận giá trị nhỏ nhất bằng 0 khi S = 1. Trong các QXTV rừng, hai chỉ số này thường nhận giá trị lớn hơn 1. Theo Magurran (2004), hai chỉ số Margalef và Menhinick được sử dụng để so sánh sự giàu có về loài tùy theo vị trí, quần xã, hệ sinh thái và kích thước ô mẫu.

2.4.3.2. Những số đo đa dạng loài sinh vật

(a) Những số đo đa dạng loài sinh vật bằng phương pháp tham số. Phần này giới thiệu bốn phương pháp đo đa dạng loài thông dụng: (1) Phương pháp chuỗi Logarit; (2) Phương pháp thống kê Q; (3) Phương pháp thống kê Jackknife; (4) Phương pháp của Chao.

(a_1) Phương pháp chuỗi Logarit. Ronald Fisher (1943; Dẫn theo Magurran, 2004) là người đầu tiên mô tả sự phong phú của các loài trong các quần xã sinh vật bằng chuỗi đường cong Logarit (Công thức 2.21). Ở công thức 2.21, αx là loài chỉ có 1 cá thể; $\alpha x^2/2$ là số loài có 2 cá thể; $\alpha x^3/3$ là loài có 3 cá thể; $\alpha x^k/k$ là loài có k cá thể. Giá trị x và α đều là hằng số; trong đó x nhận giá trị giữa 0 và 1.

$$\alpha x, \alpha x^2/2, \alpha x^3/3, \alpha x^4/4 \ldots \alpha x/n \quad (2.21)$$

Hai tham số α và S biểu thị cấu trúc quần xã. Giữa hai tham số này có mối quan hệ như hàm 2.22; trong đó giá trị α được xác định theo hàm 2.23. Khoảng tin cậy của α được xác định theo công thức 2.24. Giá trị x phụ thuộc vào tỷ lệ N/S. Trong thực hành, người ta thường lấy giá trị x > 0,99 nhưng không bao giờ x > 1. Khi tỷ lệ (N/S) > 20, thì x = 0,99. Vì x thường nhận giá trị xấp xỉ 1, nên α biểu thị số loài chỉ có 1 cá thể. Ví dụ: Khi N = 1000 và x = 0,99, thì S = 47.

$$S = \alpha * Ln(1 + N/\alpha) \quad (2.22)$$

$$\alpha = \frac{N*(1 - x)}{x} \quad (2.23)$$

$$Var(\alpha) = \frac{Ln(2)*\alpha}{[Ln(x/(1 - x) - 1)]^2} \quad (2.24)$$

Theo mô hình chuỗi log, Hayek và Buzas (1997; Dẫn theo Magurran, 2004) đã xây dựng mô hình ước lượng số loài bắt gặp (S) ở một khu vực theo hàm 2.25; trong đó N = tổng số cá thể của S loài. Ví dụ: Khi N = 1000 và x = 0,99, thì S = 47.

$$(S/N) = [(1 - x)/x]*[-Ln(1 - x)] \quad (2.25)$$

Theo Magurran (2004), phân bố chuỗi Log của Fisher chỉ thích hợp với những chuỗi số liệu mà độ phong phú của các loài biểu thị bằng số cá thể. Chuỗi Log không thích hợp với những chuỗi số liệu mà độ phong phú của loài biểu thị bằng sinh khối. Mặt khác, chỉ số đa dạng α của Fisher không thích hợp với trường hợp quần xã có S rất lớn và mật độ cao. Trái lại, chỉ số này chỉ thích hợp với trường hợp quần xã có S nhỏ và mật độ thấp. Nói chung, chỉ số đa dạng Alpha theo chuỗi Log không phụ thuộc vào kích thước ô mẫu. Ngoài ra, nó được báo cáo bằng những số đo đa dạng rõ ràng, thậm chí độ phong phú của loài không phân bố theo chuỗi Log và ít chịu ảnh hưởng bởi độ phong phú của những loài có mặt ở nhiều ô mẫu. Vì thế, chỉ số đa dạng Alpha của Fisher có thể được sử dụng trong phân loại các QXTV rừng.

(a_2) Phương pháp dựa trên thống kê Q. Kempton và Taylor (1976, 1978; Dẫn theo Magurran, 2004) đã đề xuất sử dụng thống kê Q làm số đo đa dạng. Chỉ số này là số đo độ dốc tứ phân vị (Q = Inter-Quartile Slope) của đường cong tích lũy độ phong phú của các loài. Hàm ước lượng độ dốc tứ phân vị (Q_{Slope}) có dạng như hàm 2.26; trong đó các tứ phân vị được tính theo công thức 2.27. Ở công thức 2.26 và 2.27, S = tổng số loài trong mẫu; $\sum(n_r)$ = tổng số loài giữa các tứ phân vị; Q_1 và Q_2 tương ứng là tứ phân vị 25% và 75%; n_{R1} = số loài rơi vào cấp Q_1; n_{R2} = số loài rơi vào cấp Q_2.

$$Q_{Slope} = [\frac{n_{R1}}{2} + \sum(n_r) + \frac{n_{R2}}{2}]/Ln(Q_2/Q_1) \qquad (2.26)$$

$$Q_1 = S/4;\ Q_2 = 0,5*S;\ Q_3 = 0,75*S \qquad (2.27)$$

(a_3)Phương pháp dựa trên thống kê Jackknife. Dựa trên thống kê Jackknife, Burnham và Overton (1978, 1979; Dẫn theo Magurran, 2004) đã xây dựng hai phương pháp ước lượng số loài (S) trong quần xã bằng phương pháp lấy mẫu lặp, tương ứng gọi là phương pháp Jackknife 1 và Jackknife 2. Theo hai phương pháp này, số loài (S) trong quần xã được xác định tương ứng theo công thức 2.28 và 2.29; trong đó $S_{(Mẫu)}$ là số loài thực tế trong ô mẫu, Q_1 và Q_2 tương ứng là số loài chỉ có ở 1 ô mẫu và 2 ô mẫu, m là tổng số ô mẫu.

$$S_{Jack\ 1} = S_{(Mẫu)} + Q_1(m - 1)/m \qquad (2.28)$$

$$S_{Jack\ 2} = S_{(Mẫu)} + [(Q_1(2m-3)/m) - (Q_2(m-2)^2/m(m-1))] \qquad (2.29)$$

(a_4) Phương pháp tham số của Chao. Chao (1984; Dẫn theo Magurran, 2004) đã xây dựng hai phương pháp ước lượng số loài trong quần xã. Sau này, khi áp dụng hai phương pháp này để nghiên cứu đa dạng sinh vật ở rừng mưa nhiệt đới Costa Rica, Chazdon và ctv (1998) đã gọi hai phương pháp này là Chao 1 và Chao 2. Theo phương pháp Chao 1 và Chao 2, số loài (S) trong QXTV được ước lượng tương ứng theo công thức 2.30 và 2.31. Ở công thức 2.30 và 2.31, $S_{(Mẫu)}$ là số loài thực tế thu được trong ô mẫu; F_1 và F_2 tương ứng là số loài chỉ có 1 và 2 cá thể; Q_1 và Q_2 tương ứng là số loài chỉ xuất hiện trong 1 ô mẫu và 2 ô mẫu.

$$S_{Chao\ 1} = S_{(mẫu)} + \frac{F_1^2}{2F_2} \quad (2.30)$$

$$S_{Chao\ 2} = S_{(mẫu)} + \frac{Q_1^2}{2Q_2} \quad (2.31)$$

Nói chung, ưu điểm của phương pháp tham số là số loài được ước lượng dựa theo quy luật phân bố số loài của quần xã và kết quả có thể kiểm tra bằng những tiêu chuẩn thống kê. Nhược điểm của phương pháp tham số là tính toán phức tạp.

(b) Những số đo đa dạng loài bằng phương pháp phi tham số. Phần này giới thiệu 9 phương pháp đo đa dạng sinh vật thông dụng nhất.

(b_1) Chỉ số đa dạng của Simpson. Theo Simpson (1949; Dẫn theo Magurran, 2004), nếu quần xã rất lớn (số cá thể là vô hạn, N_∞), thì việc loại bỏ một cá thể của loài nào đó cũng không ảnh hưởng đáng kể đến xác suất bắt gặp loài đó lần thứ 2. Vì thế, xác suất bắt gặp loài i lần thứ nhất là P_i ($P_i = n_i/N$), còn xác suất bắt gặp loài i lần thứ 2 là P_i^2. Nếu quần xã có dân số nhỏ, thì việc loại bỏ một cá thể của loài nào đó có ảnh hưởng đáng kể đến xác suất bắt gặp loài đó lần thứ 2. Vì thế, xác suất bắt gặp loài i lần thứ nhất là $P_i = n_i/N$, còn xác suất bắt gặp loài i lần thứ 2 là $P_i = (n-1)/(N – 1)$. Trong cả hai trường hợp kể trên, chỉ số Simpson là tổng P_i của tất cả các loài. Nói chung, khi quần xã có nhiều loài và dân số lớn, chỉ số Simpson được xác định theo công thức 2.32. Trái lại, khi quần xã có ít loài và dân số nhỏ, chỉ số Simpson được xác định theo công thức 2.33. Ở công thức 2.32 và 2.33, N = tổng số cá thể của S loài; n_i = số cá thể của loài thứ i. Trong những tài liệu sinh thái học, chỉ số D tính theo công thức 2.32 và 2.33 còn được gọi tương ứng là chỉ số λ và λ'. Về bản chất, chỉ số đa dạng Simpson bao gồm sự giàu có về loài (Σ^S_i) và độ phong phú của các loài (ni/N), nên chỉ số này còn được gọi là chỉ số ưu thế của các loài.

$$D_{Simpson} = \Sigma^S_{i=1}(ni/N)^2 = \Sigma^S_{i=1}p_i^2 \quad (2.32)$$

$$D_{Simpson} = \Sigma^S_{i=1}\left(\frac{n_i(n_i - 1)}{N(N - 1)}\right) \quad (2.33)$$

Chỉ số Simpson (D) càng nhỏ thì mức độ đa dạng càng cao. Trái lại, chỉ số D càng lớn thì mức độ đa dạng càng thấp. Khi D = 1 thì không đa dạng. Vì thế, để logic và dễ hiểu hơn, Gini-Simpson đã biến đổi chỉ số D thành (1 – D) và (1/D). Chỉ số (1 – D) cũng nhận giá trị từ 0 đến 1. Khác với chỉ số D, nếu (1 – D) càng lớn thì tính đa dạng càng cao. Chỉ số nghịch đảo Simpson (1/D) là một cách biến đổi khác để khắc phục thiếu sót của chỉ số D. Chỉ số 1/D nhận giá trị nhỏ nhất bằng 1. Giá trị 1/D = 1 cho biết quần xã chỉ có 1 loài. Nếu quần xã có ít loài và một số loài có độ phong phú lớn, thì chỉ số D nhận giá trị nhỏ, còn chỉ số 1/D nhận giá trị lớn. Nói chung, khi 1/D nhận giá trị càng lớn thì đa dạng loài càng cao. Giá trị 1/D lớn nhất bằng số lượng loài có mặt trong mẫu. Ví dụ: Mẫu có 5 loài thì ($1/D_{Max}$) = 5. Trong thực tế, cả ba chỉ số D, 1 – D và 1/D đều có thể được tính toán và báo cáo. Để tính 3 chỉ số này, cần phải biết số loài (S), số cá thể của từng loài (n_i) và tổng số cá thể của các loài (N). Ngoài ra, để ước lượng chính xác đa dạng loài của các quần xã, chúng

ta cần phải thu thập nhiều ô mẫu. Nói chung, ba chỉ số này nhận những giá trị khác nhau nhưng cùng phản ánh một mức đa dạng. Vì thế, điều quan trọng là chọn chỉ số nào để so sánh tính đa dạng của các quần xã sinh vật. Khi loài có nhiều cá thể, thì trọng số sẽ lớn. Những loài hiếm gặp thì D ít thay đổi. Nói chung, chỉ số đa dạng hay chỉ số ưu thế Simpson được sử dụng để xác định đa dạng của những quần xã sinh vật ở một môi trường nhất định (Đa dạng Alpha). Trong sinh thái học và lâm học, ba chỉ số D, 1 – D và 1/D đều có thể được sử dụng để đánh giá tác động môi trường và những hoạt động lâm sinh.

(b_2) Chỉ số đa dạng Shannon – Weiner. Shannon – Weiner (1948 – 1949; Dẫn theo Magurran, 2004) tính chỉ số đa dạng theo công thức 2.34; trong đó H = chỉ số đa dạng, S = số loài, $P_i = (n_i/N)$, n_i = số cá thể của loài i, N = tổng số cá thể của các loài.

$$H = -\Sigma^{S}_{i=1} P_i * Log(P_i) \qquad (2.34)$$

Giá trị của chỉ số H phụ thuộc vào việc sử dụng Logarit cơ số 2, cơ số Neper hay cơ số 10. Trong lý thuyết thông tin, người ta thường sử dụng $Log_{(2)}$. Trong sinh thái học, người ta thường sử dụng Logarit cơ số Neper và H được kí hiệu là H'. Theo Magurran (2004), giá trị H' thường nằm trong khoảng [1,5 – 3,5]; ít khi H' > 4,5. Theo Clarcke và Warwick (2001), chỉ số đa dạng Shannon – Weiner là một trong những chỉ số đa dạng được sử dụng rộng rãi nhất. Chỉ số này được sử dụng để so sánh đa dạng giữa những hệ sinh thái khác nhau (Đa dạng Beta) và giữa những cảnh quan khác nhau (Đa dạng Gamma).

(b_3) Chỉ số đa dạng của McIntosh. McIntosh (1967; Dẫn theo Magurran, 2004) đề xuất chỉ số đa dạng (MI) được tính theo công thức 2.35; trong đó n_i = số cá thể của loài i, N = tổng số cá thể của các loài.

$$MI = 1 - \sqrt{n_i^2/N} \qquad (2.25)$$

(b_4) Chỉ số đa dạng của Berger-Parker. Berger - Parker (1970; Dẫn theo Magurran, 2004) đã xây dựng chỉ số đa dạng (I_{BP}) theo số đo nghịch đảo của chỉ số ưu thế lớn nhất P_{Max} (Công thức 2.36). Chỉ số P_{Max} là tỷ lệ giữa mật độ lớn nhất của loài nào đó và tổng mật độ của tất cả các loài (Công thức 2.37).

$$I_{BP} = 1/P_{Max} \qquad (2.36)$$

$$P_{Max} = (N_{Max}/N) \qquad (2.37)$$

Khi số cá thể của các loài trong quần xã bằng nhau, thì tỷ lệ độ phong phú của các loài đạt cao nhất (P_{iMax}). Khi chỉ số I_{BP} tăng lên thì tính đa dạng của quần xã sinh vật cũng tăng lên.

(b_5) Chỉ số đa dạng của Hill. Chỉ số đa dạng tổng quát của Hill (1973) có dạng như hàm 2.38; trong đó S = số loài, P_i = độ phong phú tương đối của loài i.

$$D_a = \sum_{i=1}^{S} (P_i^a)^{(1/1-a)} \qquad (2.38)$$

Mark Hill (1973) đã đề xuất các số đo $N_a = \exp(H_a)$. Sau này người ta gọi các số đo N_a là những số của Hill. Khi a = 0, thì D_a = S (số loài). Khi a = 0,999, thì

$D_a = H'$. Khi a = 2, thì $D_a = D_{Simpson}$. Từ công thức 2.31, Hill đã đề xuất ba chỉ số đa dạng N_1, N_2 và N_∞ tương ứng được tính theo công thức 2.39, 2.40 và 2.41. Ba chỉ số N_1, N_2 và N_∞ tương ứng với chỉ số đa dạng thực, chỉ số nghịch đảo của chỉ số đa dạng Simpson và chỉ số đa dạng của Berger-Parker.

$$N_1 = \exp(H') \qquad (2.39)$$

$$N_2 = \frac{1}{\sum_{i=1}^{S} P_i^2} \qquad (2.40)$$

$$N_\infty = \frac{1}{P_{Max}} \qquad (2.41)$$

Những số đo đa dạng của Hill có những ưu điểm sau đây: (1) Chúng có những chú giải thống nhất; (2) Chúng thay đổi rất nhỏ đối với những loài hiếm gặp và thay đổi mạnh đối với những loài thường gặp; (3) Chúng biểu thị không chỉ số loài, mà còn cả số cá thể của các loài; (4) Về mặt toán học, chúng được sử dụng để xác định các số đo đa dạng như $N_0 = S$ (Số loài ở đối tượng nghiên cứu), $N_1 = \exp(H')$ (H' là chỉ số đa dạng Shannon-Weiner), $N_2 = 1/D$ (Chỉ số ưu thế Simpson). Nói chung, các chỉ số Simpson và Shannon – Weiner được tính toán dễ dàng. Vì thế, chỉ số đa dạng sinh vật chỉ cần tính theo các chỉ số đa dạng Simpson và Shannon – Weiner.

(b_6) Chỉ số đa dạng của Margalef. Chỉ số đa dạng của Margalef (1958) có dạng như công thức 2.42; trong đó $P_i = D_{Simpson}$.

$$D_{Margalef} = \frac{(1 - P_i)}{P_i} \qquad (2.42)$$

(b_7) Chỉ số đa dạng của Patil-Taillie. Patil-Taillie (1976; Dẫn theo Giavelli và ctv, 1986) đề xuất chỉ số đa dạng được xác định theo công thức 2.43; trong đó P_T = chỉ số đa dạng, N = tổng số cá thể trong mẫu; n_i = số cá thể của loài i.

$$P_T = 1 + \frac{\sum_{i=1}^{S}(n_i/N)^2 * Log_{10}(n_i/N)}{1 - (n_i/N)} \qquad (2.43)$$

(b_8) Chỉ số đa dạng của Brillouin. Theo Clarcke và Warwick (2001), Brillouin đề xuất chỉ số đa dạng được tính theo công thức 2.44.

$$I_{Brillouin} = \ln(N!) - \sum \ln(n_i!)N \qquad (2.44)$$

Clarcke và Warwick (2001) và Magurran (2004) cho rằng chỉ số Brillouin không thích hợp để ước lượng sự đa dạng của một quần xã dựa trên một mẫu đo đếm ngẫu nhiên. Chúng chỉ thích hợp cho những trường hợp đã thống kê đầy đủ độ phong phú của toàn bộ các loài hình thành quần xã hoặc những mẫu đo đếm điển hình. Ngoài ra, chỉ số Brillouin cũng không dễ tính toán, ngay cả với máy tính. Tính phức tạp này có liên quan đến phép tính Ln(N!) và Ln(n_i!).

(b_9) Chỉ số hiếm. Chỉ số hiếm (IR) được sử dụng để xác định những loài hiếm gặp trong những quần xã sinh vật. Chỉ số IR được xác định theo công thức

2.45; trong đó N = tổng số ô đo đếm, n = tổng số ô bắt gặp loài quan tâm. Giá trị IR thay đổi từ 0 – 100%. Khi IR = 78 – 95%, thì loài ít bắt gặp trong QXTV. Khi IR = 95 - 97%, thì loài rất ít bắt gặp trong QXTV. Khi IR > 97%, thì loài cực kỳ ít bắt gặp trong QXTV.

$$IR = (1 - (n/N))*100 \qquad (2.45)$$

Nói chung, ưu điểm của phương pháp pháp phi tham số là đơn giản trong tính toán và không cần phải biết quy luật phân bố số loài trong quần xã sinh vật. Nhược điểm của phương pháp phi tham số là kết quả không chỉ ra được mức ý nghĩa thống kê.

2.4.3.3. Những số đo mức độ đồng đều về độ phong phú của các loài

Một trong những thành phần quan trọng của đa dạng sinh vật là phân bố độ phong phú của các loài trong quần xã. Đa dạng sinh vật phụ thuộc không chỉ vào số loài, mà còn vào phân bố độ phong phú của các loài. Vì thế, để so sánh đa dạng sinh vật giữa các quần xã, các nhà sinh thái học không chỉ sử dụng số loài và chỉ số đa dạng loài, mà còn cả chỉ số đồng đều (Evenness Index). Chỉ số này đo mức độ đồng đều đối với độ phong phú của các loài trong quần xã. Dưới đây giới thiệu một số chỉ số thông dụng được sử dụng để đo phân bố độ phong phú của các loài.

*(1) Chỉ số đồng đều của Pielou***.** Pielou (1969; Dẫn theo Magurran, 2004) đề xuất chỉ số đo sự đồng đều về độ phong phú của các loài theo công thức 2.46.

$$J' = H'/H'_{Max} \qquad (2.46)$$

$$H'_{Max} = -\sum^{S}_{i=1}(1/S)*\ln(1/S) = \ln(S)$$

(2) Chỉ số đồng đều của Simpson. Simpson (1949; Dẫn theo Magurran, 2004) đã xác định sự đồng đều về độ phong phú của các loài trong quần xã bằng cách chia chỉ số ưu thế cho số loài (Công thức 2.47).

$$E_{Simpson} = (D_{Simpson}/S) \qquad (2.47a)$$

$$E_{Simpson} = (1 - D_{Simpson})/(1 - D_{SimpsonMax}) \qquad (2.47b)$$

Chỉ số $E_{Simpson}$ phụ thuộc chặt chẽ vào số lượng loài trong quần xã. Vì thế, chỉ số đồng đều về độ phong phú của các loài trong quần xã có thể được xác định bằng cách chia chỉ số đa dạng của Simpson ở dạng nghịch đảo cho số loài (Công thức 2.48).

$$E_{1/D} = (1/D)/S \qquad (2.48)$$

*(3) Chỉ số đồng đều của Shannon – Weiner***.** Theo Shannon-Weiner (1948, 1949; Dẫn theo Magurran, 2004), khi chỉ số đa dạng H' tiến đến 1, thì phân bố độ phong phú của các loài trong quần xã rất đồng đều. Ngược lại, khi chỉ số đa dạng H' tiến đến 0, thì phân bố độ phong phú của các loài trong quần xã rất không đồng đều. Khi tất cả các loài có độ phong phú bằng nhau (đồng đều), thì $\sum^{S}_{i=1}P_i = 1/S$ và H' = Ln(S). Vì thế, Shannon-Weiner đã xác định chỉ số đồng đều theo công thức 2.49. Từ công thức 2.49 cho thấy, nếu các loài có độ phong phú bằng nhau, thì $E_{Shannon}$ =

1. Khi độ phong phú của các loài khác nhau (loài hiếm, loài nhiều), thì $E_{Shannon}$ sẽ giảm. Vì thế, giá trị H' của những ô mẫu được so sánh với $H'_{Max} = 1$. Nếu độ phong phú tập trung rất cao vào một số loài, thì chỉ số H' sẽ tiến đến zero. Nếu đối tượng nghiên cứu chỉ có 1 loài, thì H' = 0.

$$E_{Shannon} = H'/\ln(S) \qquad (2.49)$$

(4) Chỉ số đồng đều của Hill. Hill (1973) đã đề xuất những số đo sự đồng đều về phân bố độ phong phú của các loài theo công thức 2.43. Từ công thức 2.43, Hill đã xây dựng 4 số đo sự đồng đều (Công thức 2.50 -2.54). Nói chung, những số đo của Hill có ưu điểm là chúng không chỉ đo đa dạng, mà còn đo cả sự đồng đều.

$$E_{a,b} = (N_a/N_b) \text{ với } a > b \qquad (2.50)$$

$$N_{10} = (N_1/S) \qquad (2.51)$$

$$N_{10}' = \left(\frac{N_1 - 1}{S - 1}\right) \qquad (2.52)$$

$$N_{21} = (N_2/N_1) \qquad (2.53)$$

$$N_{21'} = \left(\frac{N_2 - 1}{N_1 - 1}\right) \qquad (2.54)$$

(5) Chỉ số đồng đều của McIntosh. McIntosh (1951; Dẫn theo Magurran, 2004) đã đề xuất số đo sự đồng đều về phân bố độ phong phú của các loài theo công thức 2.55; trong đó S = số loài, N = tổng số cá thể của S loài và n_i = số cá thể của loài i.

$$E = \frac{\sqrt{n_i^2}}{\sqrt{((N - S + 1)^2 + S - 1)}} \qquad (2.55)$$

(6) Chỉ số đồng đều của Whittaker. Whittaker (1972) xác định số đo sự đồng đều về phân bố độ phong phú của các loài theo công thức 2.56; trong đó S = số loài, N_i = độ phong phú của loài ưu thế cao nhất, N_S = độ phong phú của loài ưu thế thấp nhất.

$$E_W = \frac{S}{\mathrm{Ln}(N_i) - \mathrm{Ln}(N_S)} \qquad (2.56)$$

Để đánh giá tính không thuần nhất về phân bố của các loài cây gỗ theo không gian hoặc tính không thuần nhất về môi trường trong những quần xã sinh vật, Whittaker (1972) đã sử dụng chỉ số đa dạng β (Công thức 2.57); trong đó S là tổng số loài bắt gặp ở đối tượng nghiên cứu, s là số loài bình quân bắt gặp trong một ô mẫu của đối tượng nghiên cứu. Từ công thức 2.57 cho thấy, khi S = s thì các loài phân bố đồng đều trong quần xã. Khi S >> s thì các loài phân bố không đồng đều trong quần xã. Nói cách khác, khi S = s thì môi trường trong quần xã là thuần nhất (đồng đều). Khi S >> s thì môi trường trong quần xã là không thuần nhất.

$$\beta \text{ - Whittaker} = (S/s) \qquad (2.57)$$

(7) Chỉ số đồng đều của Heip. Heip và ctv (1974) đã đề xuất công thức đo sự đồng đều về phân bố độ phong phú của các loài theo công thức 2.58; trong đó H' là chỉ số đa dạng Shannon-Weiner, S = số loài.

$$E = \frac{\exp(H') - 1}{S - 1} \qquad (2.58)$$

(8) Chỉ số đồng đều của Sheldon. Sheldon (Dẫn theo Magurran, 2004) đã đo sự đồng đều về phân bố độ phong phú của các loài theo công thức 2.59; trong đó trong đó H' là chỉ số đa dạng Shannon-Weiner, S = số loài.

$$E = \frac{\exp(H')}{S} \qquad (2.59)$$

(9) Chỉ số đồng đều của Brillouin. Brillouin (Dẫn theo Magurran, 2004) đã xác định sự đồng đều về phân bố độ phong phú của các loài theo công thức 2.60.

$$E_{Br} = (I_{Br}/I_{BrMax}) \qquad (2.60)$$

Khi các loài có độ phong phú bằng nhau, thì chỉ số I_{Br} nhận giá trị cao nhất (Ibr_{Max}). Về nguyên tắc, khi độ phong phú của các loài bằng nhau, thì số cá thể trung bình của mỗi loài (n) được thay thế bằng N/S. Theo đó, giá trị ΣLn(ni!) sẽ bằng S*Ln(n!). Tuy nhiên, khi tính n bằng tỷ lệ N/S, thì n có thể nhận giá trị lẻ. Điều này gây khó khăn cho việc tính Ln(ni!). Để khắc phục tình trạng này, giá trị N/S cần phải được làm tròn. Ngoài ra, bởi vì S*n có thể nhỏ hơn N, nên giá trị Ibr_{Max} được xác định theo công thức 2.61; trong đó với r = N – S*n.

$$Ibr_{Max} = (Ln(N!) – r*Ln((n + 1)!) - (S - r)*Ln(n!))/N \qquad (2.61)$$

2.4.3.4. Phương pháp so sánh và đánh giá đa dạng loài cây gỗ

Xác định phương pháp so sánh và đánh giá đa dạng loài cây gỗ của các QXTV rừng là một vấn đề quan trọng. Về lý thuyết, phương pháp so sánh đa dạng loài cây gỗ của các QXTV là rất đa dạng. Nguyên tắc chung là chọn những phương pháp dễ tính toán và trả lời rõ ràng nhiều câu hỏi đặt ra. Từ nhiều tài liệu tham khảo cho thấy, việc so sánh và đánh giá đa dạng loài cây gỗ của các QXTV rừng được thực hiện theo các bước cơ bản sau dây.

Bước 1. Xác định những đặc trưng thống kê của các thành phần đa dạng theo từng đối tượng nghiên cứu (QXTV rừng). Những đặc trưng thống kê thường được báo cáo là trung bình, nhỏ nhất, lớn nhất, sai tiêu chuẩn, hệ số biến động. Bảng 2.4 dẫn ví dụ báo cáo tóm tắt những chỉ số đa dạng loài cây gỗ đối với ưu hợp Dầu con rái.

Bước 2. Đánh giá kết quả phân tích đa dạng loài cây gỗ. Theo Magurran (1988, 2004), nếu $d_{Margalef}$ < 2, 2 – 8 và > 8, thì mức độ giàu có về loài cây gỗ tương ứng ở mức thấp, trung bình và cao. Giá trị J' dao động từ 0 – 1; trong đó J' càng gần 1 thì độ phong phú của các loài cây gỗ càng đồng đều. Đa dạng loài cây gỗ ở mức thấp, trung bình, cao và rất cao tương ứng với H' < 2, H' = 2 – 3, H' = 3 - 4 và H' > 4. Khi loài nào có chỉ số 1 – λ' càng cao, thì mức độ ưu thế của loài đó càng cao. Đa dạng

β biểu thị biến động thành phần loài theo điều kiện môi trường. Khi chỉ số β của QXTV rừng nhận giá trị càng gần 1, thì thành phần loài cây gỗ phân bố trong QXTV càng đồng đều hay môi trường của QXTV biến động càng nhỏ. Trái lại, khi chỉ số β của QXTV rừng nhận giá trị càng lớn, thì thành phần loài cây gỗ phân bố càng không đồng đều hay môi trường biến động càng lớn.

Bảng 2.4. Đặc trưng thống kê những chỉ số đa dạng loài cây gỗ đối với ưu hợp Dầu con rái. Đơn vị tính: 0,25 ha. Nguồn: Đào Thị Thùy Dương, 2019.

Thống kê	S	N	d	J'	H'	1-λ'	β-Whittaker
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Số ô mẫu	9	9	9	9	9	9	9
Trung bình	23	162	4,28	0,82	2,56	0,89	2,39
±S	2,0	12,3	0,39	0,06	0,19	0,03	
CV%	8,7	7,6	9,1	7,7	7,3	3,6	
Nhỏ nhất	20	147	3,68	0,70	2,20	0,82	
Lớn nhất	25	178	4,81	0,90	2,80	0,92	
Lớn nhất-Nhỏ nhất	5	31	1,13	0,20	0,60	0,10	

Bước 3. So sánh các thành phần đa dạng. Sự khác nhau về ba thành phần đa dạng giữa các đối tượng nghiên cứu (QXTV rừng) có thể so sánh bằng phương pháp phân tích biến động (ANOVA) và hồ sơ đa dạng của Rényi. Phương pháp ANOVA có ưu điểm là dễ tính toán bằng các phần mềm thống kê và trả lời kết quả kèm theo độ tin cậy rõ ràng. Phương pháp ANOVA có nhược điểm là phải thực hiện nhiều so sánh và khó chọn được chỉ tiêu để phân loại các QXTV rừng. Mặt khác, do các thành phần đa dạng loài cây gỗ của các QXTV rừng có biên độ biến động (Max – Min) khá rộng và trùng lấp lên nhau, nên phương pháp này phân chia không rõ ràng sự khác nhau về các thành phần đa dạng giữa các QXTV rừng.

Để xác định sự khác biệt về các thành phần đa dạng giữa các QXTV rừng, người ta đề nghị sử dụng hồ sơ đa dạng của Rényi (Công thức 2.62). Ở công thức 2.62, H_α là hồ sơ đa dạng của Rényi; $P_i = (ni/N)$ với ni là số cá thể của loài i, N là tổng số cá thể của các loài; α nhận giá trị = 0; 0,25; 0,5; 1, 2, 3,.., ∞; S = số loài cây gỗ.

$$H\alpha = Ln(\sum_{i=1, S}(P_i^{\alpha})/(1 - \alpha) \qquad (2.62)$$

Hồ sơ đa dạng của Rényi có một số tính chất đáng lưu ý. Khi α = 0, thì $\exp(H_0) = S$. Khi α = 1, thì Hα = H'. Khi α = 2, thì $H\alpha = Ln(1/D_{Simpson})$. Khi α = ∞, thì H_∞ = -Ln(1/Berger - Parker) và $1/\exp(H_\infty) = P_{Max}$ (Tỷ lệ độ phong phú của loài có độ phong phú cao nhất). Lưu ý rằng khi α = 1 và α = ∞ thì $Ln(H_\alpha) = 0$. Vì thế, khi tính H' và P_{Max}, giá trị α = 1 thay bằng α = 0,999, còn α = ∞ thay bằng α = 100.

Hồ sơ đa dạng của Rényi là những đường cong cung cấp các thông tin về ba thành phần đa dạng loài (sự giàu có về loài, sự đồng đều về phân bố độ phong phú của các loài và đa dạng loài). Ưu điểm của hồ sơ đa dạng của Rényi là các đường cong đa dạng biểu thị rõ ràng thứ bậc cao thấp của các QXTV theo mức đa dạng.

Mặt khác, một số thành phần đa dạng (S, H', $1/D_{Simpson}$, P_{Max}) có thể xác định từ tính chất của hồ sơ đa dạng của Rényi. Nhược điểm của hồ sơ đa dạng của Rényi là nó không cho biết rõ các thành phần đa dạng và độ tin cậy của các kết luận. Nhược điểm này được khắc phục bằng cách tính các đặc trưng thống kê đối với các thành phần đa dạng.

Khi vẽ đường cong đa dạng loài, trục hoành là giá trị α, còn trục tung là hồ sơ đa dạng của Rényi (H_α). Trục hoành bên trái cho biết mức độ giàu có về loài, còn trục hoành bên phải cho biết mức độ đồng đều về phân bố độ phong phú của các loài cây gỗ. Sự khác biệt về đa dạng loài cây gỗ giữa hai hay nhiều QXTV được so sánh dựa theo vị trí cao thấp của đường cong hồ sơ đa dạng loài. Theo hồ sơ đa dạng, QXTV nào có vị trí hồ sơ đa dạng cao hơn thì QXTV đó đa dạng hơn.

Bảng 2.5 và Hình 2.8 dẫn ví dụ so sánh hồ sơ đa dạng loài cây gỗ đối với hai ưu hợp Dầu con rái. Mỗi ưu hợp Dầu con rái bao gồm 3 ô tiêu chuẩn 0,25 ha; trong đó số loài cây gỗ của ưu hợp 1 và 2 tương ứng là 41 và 44 loài. Chỉ số đa dạng Shannon H' ở ưu hợp 1 và 2 tương ứng là 3,2 và 3,0.

Bảng 2.5. Hồ sơ đa dạng loài cây gỗ đối với hai ưu hợp Dầu con rái. Mỗi ưu hợp Dầu con rái bao gồm 3 ô tiêu chuẩn 0,25 ha.

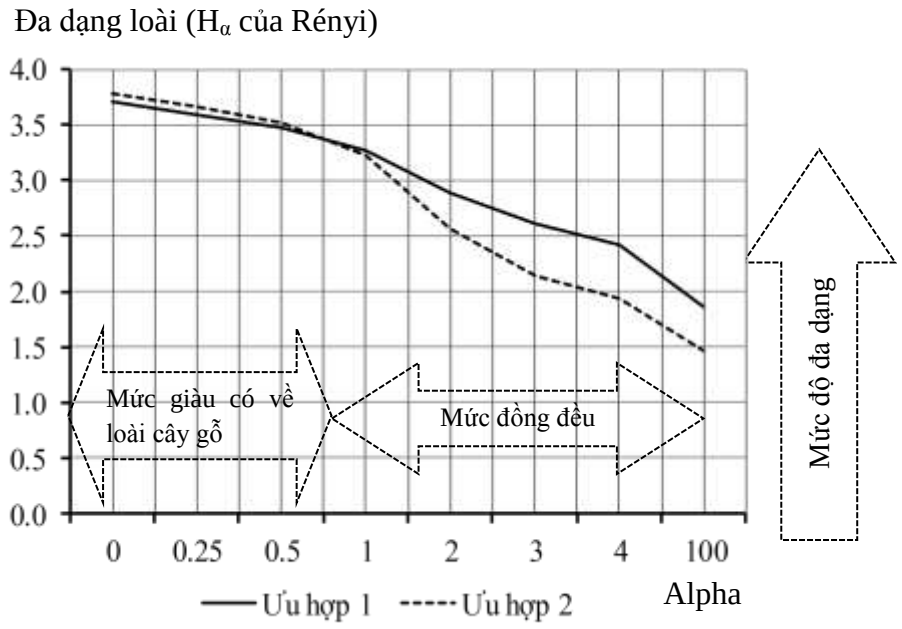
TT	Giá trị α	Giá trị Hα của hai ưu hợp Dầu con rái:	
		Ưu hợp 1	Ưu hợp 2
(1)	(2)	(3)	(4)
1	0	3,714	3,784
2	0,25	3,588	3,660
3	0,5	3,472	3,524
4	0,999	3,265	3,218
5	2	2,890	2,558
6	3	2,606	2,147
7	4	2,416	1,937
8	100	1,865	1,472

Từ số liệu ở Bảng 2.5 cho thấy, đối với ưu hợp 1, khi α = 0, thì $\exp(H_0)$ = exp(3,714) = 41 loài. Khi α = 0,999, thì H_1 = 3,26. Đối với ưu hợp 2, khi α = 0, thì $\exp(H_0)$ = exp(3,784) = 44 loài. Khi α = 0,999, thì H_1 = 3,21. Như vậy, hai thành phần S và H' được tính từ hồ sơ đa dạng nhận kết quả tương tự như số liệu từ phân tích thống kê. Hồ sơ đa dạng ở Hình 2.8 chỉ ra rằng ưu hợp 2 có số loài cây gỗ lớn hơn so với ưu hợp 1, nhưng phân bố độ phong phú của các loài cây gỗ (Chỉ số đồng đều) và đa dạng loài cây gỗ (Chỉ số đa dạng) lại nhỏ hơn.

2.4.3.5. Thảo luận chung về phương pháp xác định chỉ số đa dạng loài cây gỗ

(a) Các chỉ số đa dạng có thể được xác định theo những phương pháp tham số và phương pháp phi tham số. Phương pháp tham số phụ thuộc vào quy luật phân bố của các thành phần đa dạng và tính toán phức tạp. Việc tìm ra quy luật này là một việc làm khó khăn. Phương pháp phi tham số không đòi hỏi phải biết quy luật

phân bố của các thành phần đa dạng và tính toán đơn giản. Mặt khác, từ một nguồn số liệu như nhau, nếu tính toán từ những chỉ số đa dạng khác nhau, thì kết quả nhận được cũng khác nhau. Tương tự, nếu sử dụng số đo độ phong phú khác nhau (N, G, M, chỉ số IVI) để tính chỉ số đa dạng loài cây gỗ, thì kết quả nhận được cũng khác nhau. Vì thế, chọn phương pháp đo đa dạng và số đo độ phong phú của loài là một vấn đề cần được quan tâm.



Hình 2.8. Đồ thị biểu diễn hồ sơ đa dạng loài cây gỗ của Rényi đối với hai ưu hợp Dầu con rái.

(b) Chọn những chỉ số thích hợp để đo đa dạng sinh vật. Các nhà sinh thái học đã xây dựng rất nhiều phương pháp xác định đa dạng sinh vật. Mỗi phương pháp đều có những ưu điểm và nhược điểm riêng. Vấn đề đặt ra là chọn những chỉ số đa dạng nào để giải quyết tốt những mục tiêu nghiên cứu. Theo Magurran (2004), những chỉ số đa dạng sinh vật thích hợp phải đảm bảo những yêu cầu cơ bản sau đây:

(1) Chúng là những chỉ số đơn giản và dễ tính toán. Ngày nay người ta đã xây dựng những chỉ số đa dạng dễ tính toán trong những gói thống kê. Bởi vì khi xây dựng những gói thống kê, các nhà sinh thái học đã chọn và cập nhật những chỉ số đa dạng thích hợp nhất.

(2) Chúng phù hợp với đối tượng nghiên cứu. Điều đó cho phép giải quyết không chỉ những vấn đề về đa dạng sinh học, mà còn cả những vấn đề khác nhau trong sinh thái học và lâm học.

(3) Chúng cho phép ước lượng chính xác số loài, bởi vì sự giàu có về loài là một thành phần của đa dạng sinh vật.

(4) Chúng cho phép phân loại những quần xã sinh vật khác nhau thành những nhóm quần xã tương đồng. Điều đó không chỉ có ý nghĩa về khoa học, mà còn cả ứng dụng trong thực tiễn.

(5) Chúng không hoặc ít phụ thuộc vào kích thước mẫu. Điều đó cho phép xử lý số liệu từ những nguồn số liệu khác nhau. Chỉ số α của Fisher và chỉ số Simpson đáp ứng tốt yêu cầu này.

(6) Chúng là những chỉ số đã được sử dụng rộng rãi trong nghiên cứu đa dạng. Chỉ số đa dạng Shannon-Weiner (H') là một chỉ số đa dạng sinh vật thông dụng. Chỉ số H' cho phép so sánh đa dạng sinh vật của một quần xã thay đổi theo thời gian và giữa những quần xã khác nhau. Tuy vậy, chỉ số H' lại phụ thuộc vào kích thước mẫu. Khó khăn này được khắc phục bằng cách chuyển chỉ số H' thành exp(H').

(7) Chúng cho phép giải thích độ ưu thế của các loài trong quần xã. Chỉ số Simpson và Berger-Paker là hai chỉ số đơn giản, dễ tính toán và đáp ứng tốt yêu cầu này.

(8) Chúng cho phép xác định cả ba thành phần đa dạng (sự giàu có về loài, sự đồng đều về phân bố độ phong phú của các loài và chỉ số đa dạng).

Nói chung, hai chỉ số đa dạng Simpson và Shannon-Weiner và hai chỉ số đồng đều Pielou và Shannon-Weiner là những số đo đa dạng được các nhà sinh thái học sử dụng rộng rãi. Các chỉ số đa dạng Simpson (D, 1 - D và 1/D) được sử dụng không chỉ nhằm mục đích phân tích đa dạng và sự đồng đều về phân bố độ phong phú của những loài cây gỗ trong quần xã, mà còn so sánh mức độ ưu thế của các loài trong quần xã. Chỉ số Shannon-Weiner (H) được sử dụng nhằm mục đích: (a) Phân tích đa dạng của các quần xã; (b) So sánh sự đồng đều về phân bố độ phong phú của các loài trong quần xã; (c) So sánh đa dạng hay sự phức tạp giữa các quần xã; (d) Phân chia những quần xã tương đồng. Ngoài ra, những chỉ số này đều là những chỉ số đơn giản và rất dễ tính toán; đồng thời chúng đã được xây dựng thành phần mềm tính toán trong các gói thống kê.

(c) Những thông tin về đặc điểm lâm học và đa dạng loài cây gỗ của những QXTV rừng có thể thay đổi tùy theo vị trí thu mẫu và đối tượng thu mẫu. Một đặc tính quan trọng của rừng mưa nhiệt đới là tính không thuần nhất theo không gian và thời gian. Vì thế, cùng một kiểu rừng giống nhau, nếu đối tương và ranh giới hay vị trí thu mẫu khác nhau, thì kết quả nhận được cũng khác nhau. Do đó, khi báo cáo về đặc điểm lâm học và đa dạng cây gỗ của các QXTV rừng, nhà lâm học cần phải định nghĩa chính xác kiểu rừng, điều kiện môi trường, vị trí thu mẫu và phương pháp thu mẫu. Việc này là rất quan trọng, bởi vì kết quả chỉ ra kết cấu và đa dạng loài cây gỗ của các QXTV rừng định cư trong khu vực nào, kiểu rừng nào và môi trường nào. Tóm lại, khi báo cáo về đặc điểm lâm học và đa dạng loài cây gỗ của rừng, nhà lâm học cần phải báo cáo rõ phương pháp thu thập và xử lý số liệu. Một phương pháp thu mẫu tốt sẽ nhận được kết qủa tốt. Ngoài ra, sử dụng những phương pháp thông dụng sẽ tránh được những tranh luận không cần thiết.

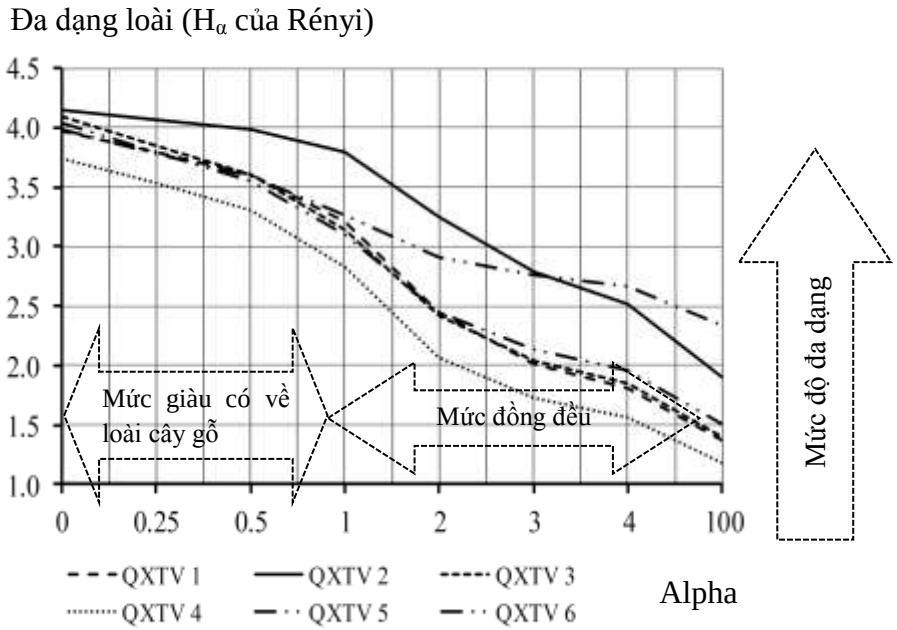
(d) Thứ bậc cao thấp của đa dạng thay đổi tùy theo số đo đa dạng. Từng chỉ số đa dạng không thể cung cấp thông tin chính xác về thứ bậc cao thấp của đa dạng. Nếu sử dụng chỉ số đa dạng khác nhau, thì kết quả phân loại đa dạng khác nhau. Chẳng hạn một vị trí có H' cao nhưng J' nhỏ. Mặc dù sử dụng nhiều chỉ số đa dạng cho phép đánh giá đúng thứ bậc cao thấp của đa dạng, nhưng việc tính toán lại gặp khó khăn. Khó khăn này được giải quyết bằng hồ sơ đa dạng của Rényi.

(e) Chọn phương pháp thích hợp để so sánh đa dạng sinh vật giữa các QXTV rừng. Sự khác biệt về đa dạng loài cây gỗ giữa các QXTV có thể được so sánh bằng phương pháp ANOVA và hồ sơ đa dạng của Rényi. Để thấy rõ tính phức tạp của phương pháp ANOVA, chúng ta xem xét trường hợp so sánh đa dạng loài cây gỗ của 6 QXTV khác nhau. Sáu thành phần đa dạng cần so sánh là S, N, $d_{Margalef}$, J', H' và $D_{Simpson}$. Khi phân tích thống kê, giả thiết chúng ta chỉ báo cáo: (1) Những đặc trưng thống kê của 6 thành phần đa dạng; (2) Bảng ANOVA đối với 6 thành phần đa dạng; (3) Bảng so sánh sự khác biệt của 6 thành phần đa dạng (S, N, $d_{Margalef}$, J', H' và $D_{Simpson}$) giữa 6 QXTV.

Bảng 2.6. Hồ sơ đa dạng loài cây gỗ đối với 6 QXTV khác nhau. Mỗi QXTV bao gồm 5 ô tiêu chuẩn 0,25 ha.

TT	Giá trị α	Quần xã thực vật:					
		1	2	3	4	5	6
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	0	3,970	4,143	4,094	3,738	4,043	3,989
2	0,25	3,792	4,069	3,849	3,535	3,796	3,782
3	0,5	3,602	3,987	3,604	3,309	3,547	3,585
4	0,99	3,205	3,793	3,147	2,828	3,096	3,267
5	2	2,444	3,254	2,427	2,067	2,453	2,909
6	3	2,017	2,784	2,043	1,724	2,128	2,755
7	4	1,812	2,514	1,847	1,560	1,954	2,671
8	100	1,375	1,909	1,406	1,189	1,505	2,334

Ở ví dụ này, mặc dù phân tích ANOVA được tính toán đơn giản bằng các phần mềm thống kê, nhưng kết quả xử lý cung cấp nhiều thông tin và các thông tin này được báo cáo rất phức tạp. Tính phức tạp biểu hiện ở chỗ, chúng ta phải báo cáo 1 hoặc 6 bảng thông kê mô tả tùy theo cách lập bảng; 6 hoặc 1 bảng phân tích ANOVA tùy theo phương pháp phân tích; 180 cặp so sánh giữa 6 QXTV. Tính phức tạp sẽ tăng lên khi đối tượng so sánh trên 6 QXTV. Trái lại, sử dụng phương pháp phân tích đa dạng theo hồ sơ đa dạng của Rényi làm giảm tính phức tạp trong tính toán và báo cáo kết quả. Theo phương pháp này, chúng ta chỉ báo cáo 1 hoặc 6 bảng thông kê mô tả tùy theo cách lập bảng; 1 bảng hồ sơ đa dạng (Bảng 2.6) và 1 sơ đồ đa dạng (Hình 2.9).



Hình 2.9. Đồ thị biểu diễn hồ sơ đa dạng loài cây gỗ của Rényi đối với sáu QXTV khác nhau.

Chương 3
PHÂN TÍCH CẤU TRÚC RỪNG

3.1. Khái niệm về cấu trúc rừng

Cấu trúc rừng biểu thị các thành phần của rừng và sự tổ chức sắp xếp của các thành phần theo không gian (đứng, nằm ngang) và thời gian (tuổi rừng). Cấu trúc rừng biểu hiện ở hai khía cạnh. Một là phân bố của các thành phần của rừng theo không gian (chiều ngang, chiều đứng) và thời gian (năm). Đặc tính này có thể bao gồm ba số đo sau đây: (a) Số loài cây gỗ (S); (b) Sự đa dạng về kích thước của cây gỗ (D, H, DT...); (c) Biến động của các thành phần (S, N, G, M) theo không gian. Hai là tính phức tạp về cấu trúc rừng. Đặc tính này phản ánh sự đa dạng của các thành phần (N, S, D, H, G, M) của rừng (Nguyễn Văn Thêm, 2002).

Xác định cấu trúc rừng là một vấn đề quan trọng để đánh giá tính đa dạng về môi trường, đa dạng loài và đa dạng kích thước cây gỗ và quần thụ. Tính không đồng nhất của cấu trúc rừng theo chiều đứng và chiều ngang có quan hệ với đa dạng loài và tính không ổn định của môi trường. Những hoạt động lâm sinh có thể làm thay đổi cấu trúc rừng. Vì thế, nghiên cứu cấu trúc rừng không chỉ góp phần cung cấp các thông tin về rừng, mà còn là công cụ hữu ích cho việc xây dựng những giải pháp quản lý, bảo vệ và khai thác những lợi ích của rừng.

Cấu trúc rừng được phân chia thành cấu trúc sinh thái và cấu trúc hình thái. Cấu trúc sinh thái phản ánh mối quan hệ giữa các thành phần của rừng (N, S, D, H, G, M) với các điều kiện môi trường sống (khí hậu, địa hình, đất, sinh vật, hoạt động lâm sinh). Cấu trúc hình thái phản ánh biến động của các thành phần của rừng (N, S, D, H, G, M) theo không gian và thời gian mà không xét đến điều kiện môi trường sống. Cấu trúc rừng có thể được mô tả bằng phương pháp trắc đồ rừng và những mô hình toán. Phương pháp mô tả cấu trúc rừng bằng các trắc đồ có một số ưu điểm. Một là trắc đồ rừng giúp cho nhà lâm học nhận thấy rõ vị trí của các loài cây (gỗ, bụi, cỏ) trong tán rừng từ thấp đến cao. Thông tin này giúp ích cho việc bố trí vị trí của các loài cây theo chiều đứng trong trồng rừng hỗn giao và nuôi dưỡng rừng. Hai là các trắc đồ rừng được mô tả theo thời gian giúp cho nhà lâm học xác định động thái biến đổi của rừng theo thời gian và ảnh hưởng của các phương thức lâm sinh. Ba là các trắc đồ rừng được mô tả theo các QXTV rừng khác nhau là công cụ tốt cho phân loại các QXTV và phân loại rừng. Bốn là các trắc đồ rừng cũng là công cụ tốt để phân tích ảnh hưởng của tán rừng đến đời sống của những cây tầng dưới (cây bụi, cây thân thảo) và thực vật ngoại tầng (cây thân leo, cây phụ sinh). Phương pháp trắc đồ rừng có một số nhược điểm cơ bản. Một là trắc đồ rừng mô tả không chính xác kích thước (thân, cành, lá) của các loài cây gỗ. Hạn chế này là do hạn chế về kích thước trang giấy vẽ và công cụ vẽ trắc đồ rừng (Bút). Hai là phương pháp này

không cho phép định lượng các đặc trưng thống kê đối với các thành phần rừng. Ba là các trắc đồ rừng chỉ được vẽ "Tùy hứng". Nhược điểm này xảy ra là do độ chính xác của các trắc đồ phụ thuộc vào năng lực, cảm hứng và trạng thái sức khỏe của nhà lâm học. Mặc dù có một số thiếu sót, nhưng phương pháp trắc đồ rừng vẫn được các nhà lâm học sử dụng trong những trường hợp sau đây: (a) Mô tả các đặc tính của rừng; (b) So sánh và đánh giá biến động của rừng tùy theo môi trường và những hoạt động lâm sinh; (c) Phân loại thảm thực vật rừng thành những đơn vị khác nhau (kiểu rừng, kiểu QXTV...).

Để định lượng các thành phần của rừng, các nhà lâm học sử dụng các mô hình toán để xác định cấu trúc rừng theo chiều nằm ngang và theo chiều thẳng đứng. Theo chiều nằm ngang, lâm học phân tích phân bố số cây (N) theo cấp D (N/D), phân bố N theo cấp G (N/G), phân bố N theo cấp V (N/V), phân bố cây trên mặt đất. Theo chiều thẳng đứng, lâm học phân tích phân bố của các loài cây gỗ (S, loài) theo cấp H (S/H) và phân bố N/H. Phương pháp mô tả cấu trúc rừng bằng các mô hình toán có một số ưu điểm. Một là nó cho phép nhà lâm học ước lượng và dự đoán động thái biến đổi của rừng theo thời gian. Chẳng hạn: (a) Sự biến đổi N, G và M theo tuổi rừng (A, năm); (b) Sự biến đổi N cây tái sinh theo tuổi, cấp kích thước và theo diện tích. Hai là nó giúp cho nhà lâm học xác định rõ những chỉ tiêu kỹ thuật trong chặt nuôi dưỡng rừng (cường độ tỉa thưa và sản lượng gỗ thu hoạch trung gian...) và khai thác rừng (cường độ khai thác, chu kỳ khai thác, sản lượng gỗ thu hoạch chính...). Ba là nó giúp cho nhà lâm học phân chia chỉ số lập địa rừng (SI) và đánh giá chất lượng lập địa rừng. Bốn là nó giúp cho nhà lâm học xây dựng biểu quá trình sinh trưởng và biểu sản lượng rừng.

3.2. Phân tích cấu trúc rừng bằng phương pháp trắc đồ rừng

Richards và Davis (1933 – 1934; Dẫn theo Richards, 1965) đề xuất cấu trúc hình thái của rừng có thể mô tả bằng các trắc rừng. Để mô tả cấu trúc rừng hỗn loài tự nhiên bằng phương pháp trắc đồ, Thái Văn Trừng (1978; 1998) đề nghị sử dụng dải rừng có bề dài 50 – 60 m và bề rộng 10 m. Dải rừng này được bố trí điển hình trong ô mẫu, thường ở khoảng giữa ô mẫu. Thông tin thu thập trong dải vẽ trắc đồ rừng là loài cây gỗ và kích thước của chúng (D, cm; H, m; chiều cao dưới cành = H_{DC}, m; đường kính tán = D_T, m ở vị trí rộng nhất; chiều dài tán lá = L_T). Vị trí cây gỗ theo chiều nằm ngang trong dải vẽ trắc đồ được xác định bằng hệ tọa độ Đề-Các; trong đó gốc tọa độ được chọn ở góc dưới bên trái của dải vẽ. Trắc đồ rừng được vẽ trên giấy kẻ li. Chiều đứng của trang giấy biểu diễn chiều cao thân cây gỗ từ 8 m trở lên (Quy ước quốc tế) theo tỷ lệ 1/200 (1 cm = 200 cm). Chiều ngang của trang giấy biểu diễn đường kính thân ngang ngực theo tỷ lệ 1/100 (1 cm = 100 cm). Hình chiếu nằm ngang của tán cây gỗ được bố trí ở dưới dải vẽ trắc đồ đứng theo tỷ lệ 1/200. Cây bụi và cây gỗ non có H < 8 m được vẽ riêng theo tỷ lệ 1/50 (1 cm = 50 cm).

Dựa theo hình thái và sự phân bố của các loài cây gỗ theo các tầng (lớp) tán khác nhau, Thái Văn Trừng (1998) đã phân chia cấu trúc đứng của rừng tự nhiên hỗn loài nhiệt đới thành 5 tầng: (1) Tầng vượt tán (A_1); (2) Tầng ưu thế sinh thái (A_2); (3) Tầng dưới tán (A_3); (4) Tầng cây bụi (B); (5) Tầng cỏ quyết (C). Tầng

vượt tán hay tầng trên tán rừng là lớp nằm ngang của những cây gỗ ưu thế và đồng ưu thế trong quần thụ. Những cây gỗ ở tầng vượt tán tiếp nhận ánh sáng hoàn toàn ở mọi phía. Tầng ưu thế sinh thái là tầng tập trung những cây gỗ có kích thước trung bình. Tầng dưới tán rừng bao gồm những cây gỗ phân bố ở dưới tầng ưu thế sinh thái. Tầng cây bụi và tầng cỏ quyết nằm dưới tán rừng được gọi chung là cây tầng thấp. Những loài cây phân bố ở 5 tầng tán rừng tiếp nhận ánh sáng với cường độ và chất lượng khác nhau. Nói chung, hình thái và sự phân bố của các loài cây gỗ theo các tầng thay đổi tùy theo điều kiện môi trường và kiểu rừng.

3.3. Phân tích cấu trúc rừng bằng các mô hình toán

3.3.1. Phân tích phân bố số cây theo cấp tuổi và cấp kích thước

Đối với rừng trồng thuần loài đồng tuổi, đồ thị mô tả sự suy giảm mật độ quần thụ (N, cây/ha) theo tuổi (A, năm) thường có dạng đường cong Sigmoid. Vì thế, phân bố N/A thường được mô tả tốt bằng hàm số mũ giảm (Hàm 3.1). Ở hàm 3.1, tham số m biểu thị mật độ trồng rừng ban đầu hoặc số cây ở cấp A bắt đầu nghiên cứu; tham số b biểu thị tỷ lệ hay tốc độ suy giảm số cây sau mỗi cấp A; tham số k là số cây ở cấp A cao nhất được nghiên cứu. Mô tả phân bố N/A theo Hàm 3.1b có ưu điểm là đường cong không cắt trục hoành khi rừng đã đạt đến tuổi cao.

$$N = m*\exp(-b*A) \qquad (3.1a)$$

$$N = m*\exp(-b*A) + k \qquad (3.1b)$$

Phân bố N/D và phân bố N/H của rừng trồng thuần loài đồng tuổi có thể mô tả bằng các hàm phân bố chuẩn, Lognormal, Gamma, Weibull, Richards và khoảng cách. Hàm mật độ xác suất của phân bố chuẩn có dạng như hàm 3.2; trong đó μ và σ tương ứng là trung bình và độ lệch chuẩn của biến X (X = D, H).

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{1}{2}(\frac{X - \mu}{\sigma})^2] \qquad (3.2)$$

Hàm mật độ xác suất của phân bố Lognormal có dạng như hàm 3.3; trong đó μ và σ tương ứng là trung bình và độ lệch cuẩn của biến X (X = D, H).

$$f(x) = \frac{1}{X\sigma\sqrt{2\pi}} \exp[-\frac{(LnX - \mu)^2}{2\sigma^2}] \qquad (3.3)$$

Hàm mật độ xác suất của phân bố Gamma có dạng như hàm 3.4; trong đó tham số α đặc trưng cho hình dạng hay độ lệch của đường cong hàm mật độ. Tham số β phản ánh mức độ co dãn của đường cong hàm mật độ.

$$f(x) = \frac{(X/\beta)^{\alpha-1}\exp(-X/\beta)}{\beta\Gamma(\alpha)} \qquad (3.4)$$

Hàm mật độ xác suất của phân bố Weibull 3 tham số có dạng như hàm 3.5; trong đó a = tham số vị trí, b = tham số tỷ lệ, c = tham số hình dạng, X = D và H, $a < X < \infty$, tham số b, c > 0.

$$f(a,b,c) = (c/b)*((X - a)/b)^{c-1}*\exp(-(X - a)/b)^c) = 0 \qquad (3.5a)$$

Hàm phân bố xác suất tích lũy của phân bố Weibull 3 tham số có dạng hàm 3.5b. Đường cong của hàm (3.5b) có dạng đường cong Sigmoid.

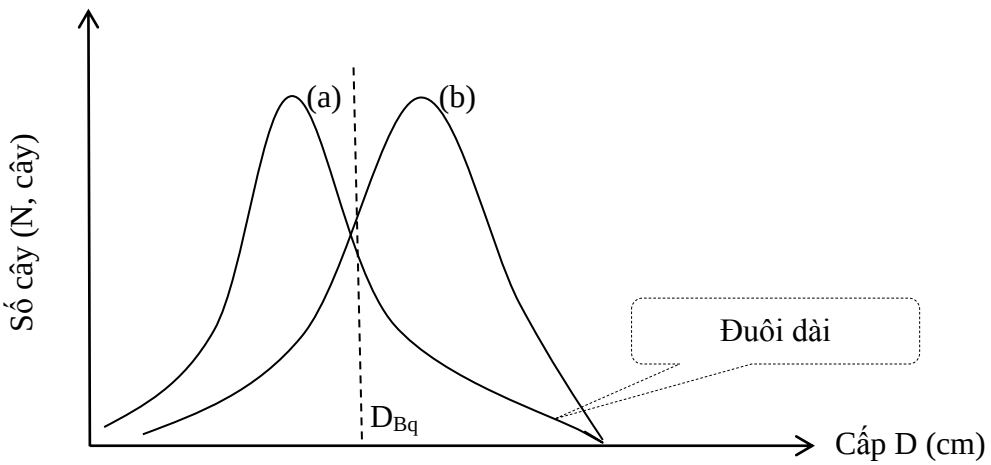
Y = 1-exp(-((X – a)/b))^c (3.5b)

Điểm uốn trên trục hoành: X = b*(1 -1/c)^1/(c + a).

Điểm uốn trên trục tung: y = 1 – exp(1/(c – 1)).

X = D (cm); a < X < ∞.

Hàm phân bố Weibull 3 tham số chỉ có ý nghĩa khi (X – a)/b > 0. Điều đó xảy ra khi b > 0 và X ≥ a. Do đó giá trị của tham số a ≤ cấp D_{Min}. Nếu X < a, thì hàm Weibull là vô nghĩa. Trong thực tế, tham số b nhận giá trị gần bằng trung vị (Me); tham số c chỉ độ lệch (Sk) của phân bố N/D. Khi c = 3,6 thì phân bố N/D tiệm cận phân bố chuẩn (S_k = 3,0). Khi tham số c tiến đến 1, thì phân bố N/D chuyển thành phân bố lệch với Mo dịch chuyển về bên trái so với giá trị trung bình (D_{Bq}) và đuôi dài lệch về bên phải (S_k < 0) (Hình 3.1a). Trái lại, khi tham số c > 3,6, thì phân bố N/D chuyển thành phân bố lệch với Mo dịch chuyển về bên phải so với giá trị trung bình và đuôi dài lệch về bên trái (S_k > 0) (Hình 3.1b). Khi c nhận giá trị rất lớn hoặc rất nhỏ thì phương sai sẽ giảm.



Hình 3.1. Phân bố N/D của rừng trồng thuần loài đồng tuổi.

Hàm mật độ xác suất và hàm xác suất tích lũy của phân bố Weibull 2 tham số tương ứng có dạng hàm (3.6a) và hàm (3.6b); trong đó b = tham số tỷ lệ, c = tham số hình dạng, b < 0 và c > 0, X = giá trị giữa của các cấp D. Đường cong của hàm (3.6b) có dạng đường cong Sigmoid.

$f_{(X)}$ = (c/b)(X/b)^(c-1)exp(-(X/b)^c) (3.6a)

$F_{(X)}$ = 1 – exp(-bX^c) (3.6b)

Điểm uốn trên trục hoành: X = b*(1 -1/c)^1/(c + a).

Điểm uốn trên trục tung: Y = 1 - exp(1/c – 1).

Hàm mật độ xác suất của hàm phân bố Richards có dạng hàm (3.7a). Hàm phân bố xác suất tích lũy (F_X) có dạng như hàm 3.7b; trong đó q = tham số vị trí, p = tham số tỷ lệ và r = tham số hình dạng (p, q > 0; r < 0).

$$f(x) = (-r/p)\exp(-(X - q)/p)(1 + \exp(-(X - q)/p))^{(r - 1)} \quad (3.7a)$$

$$F_{(X)} = 1 + \exp(-(x - q)/p))^{-r} \quad (3.7b)$$

Điểm uốn trên trục hoành: X = pln(-r) + q.

Điểm uốn trên trục tung: Y = ((r - 1)/r)^r.

Hàm phân bố khoảng cách có dạng như hàm 3.8; trong đó a và b là tham số của phân bố khoảng cách, X = 0, 1, 2,…, k.

P(x) = a với x = 0

$$P(x) = (1 - a)(1 - b)*b^{(X - 1)} \text{ với } X \geq 1 \quad (3.8)$$

Ngày nay các tham số của hàm phân bố chuẩn, Lognormal, Gamma và Weibull có thể được xác định dễ dàng bằng những phần mềm thống kê thông dụng (MINITAB; STATGRAPHICS; SPSS).

Các tham số của hàm 3.1 và ba hàm (Weibull, Richards, khoảng cách) được xác định dễ dàng bằng phương pháp hồi quy phi tuyến tính của Marquatdt. Sai lệch của những hàm này so với số liệu thực nghiệm được đánh giá bằng hệ số xác định (r^2) (Công thức 3.9); sai lệch chuẩn của ước lượng (SEE) (Công thức 3.10); sai số tuyệt đối trung bình (MAE) (Công thức 3.11); sai số tuyệt đối trung bình theo phần trăm (MAPE) (Công thức 3.12) và tổng sai lệch bình phương (SSE) (Công thức 3.13). Ở công thức 3.9 – 3.13, N_{TN} là số cây thực tế trong mỗi cấp D và cấp H, N_{Bq} là số cây bình quân trong mỗi cấp D và cấp H, N_{UL} là số cây ước lượng trong mỗi cấp D và cấp H, n là số cấp D và cấp H, p là số tham số trong mô hình. Mô hình phân bố N/D và phân bố N/H thích hợp được chọn theo tiêu chuẩn SSE_{Min}.

$$R^2 = \left[1- \frac{\sum(N_{TN} - N_{UL})^2}{\sum(N_{TN} - N_{Bq})^2}\right]*100 \quad (3.9)$$

$$SEE = \sqrt{\sqrt{\sum(N_{TN} - N_{UL})^2/(n - p)}} \quad (3.10)$$

$$MAE = \left|((N_{TN} - N_{UL})/n))\right| \quad (3.11)$$

$$MAPE = \frac{MAE*100}{N_{TN}} \quad (3.12)$$

$$SSE = \sum(N_{TN} - N_{UL})^2 \quad (3.13)$$

Đối với rừng tự nhiên hỗn loài, sự suy giảm mật độ quần thụ theo cấp D được mô tả tốt bằng hàm số mũ giảm (Hàm 3.1) và hàm phân bố Beta (Hàm 3.14).

$$f(x) = [(D - a)^{p-1}(b - D)^{q-1}]/B(p, q)(b - a)^{p+q-1} \quad (3.14)$$

với $a \leq D \leq b$; p, q > 0.

Đối với hàm phân bố Beta (3.14), tham số p và q là tham số hình dạng 1 và 2; tham số a và b tương ứng là D_{Min} và D_{Max}; (b – a) là phạm vi biến động D; B(p, q) là hàm phân bố Beta. Hai tham số p và q được xác định tương ứng theo công thức

3.15 và 3.16; trong đó D^* và S^{*2} được xác định tương ứng theo công thức 3.17 và 3.18, SEE là sai số chuẩn của D.

$$p = D^*((D^*(1 - D^*)/S^{*2}) - 1) \quad (3.15)$$

$$q = 1 - D^*((1 - D^*)/S^{*2}) - 1) \quad (3.16)$$

$$D^* = (D_{bq} - a)/(b - a) \quad (3.17)$$

$$S^{2*} = SEE^2/(b - a)^2 \quad (3.18)$$

Sai lệch của hàm 3.14 so với phân bố thực nghiệm cũng được đánh giá theo R^2, SEE, MAE, MAPE, SSE. Mô hình phân bố N/D thích hợp được xác định theo tiêu chuẩn SSE_{Min}.

Phân bố N/H của rừng tự nhiên hỗn giao khác tuổi chưa bị rối loạn do khai thác, cháy và gió bão cũng tồn tại ở dạng đường cong 1 đỉnh gần đối xứng. Dạng phân bố này cũng được mô tả tốt bằng hàm phân bố chuẩn, Lognormal, Gamma, Weibull, Richards và phân bố khoảng cách.

3.3.2. Phân bố của các loài cây gỗ trên mặt đất

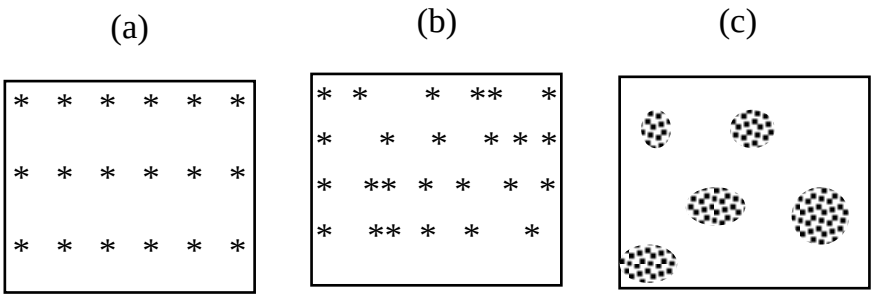
3.3.2.1. Ý nghĩa phân tích phân bố của các loài cây gỗ trên mặt đất

Các loài cây gỗ phân bố trên mặt đất có thể là đồng đều, không đồng đều và ngẫu nhiên. Ở đây có hai câu hỏi cần được đặt ra. Một là bằng cách nào có thể xác định đặc điểm phân bố của các loài cây gỗ trên mặt đất? Câu hỏi này được trả lời thông qua quan sát hoặc ứng dụng các mô hình toán để phân tích phân bố cây gỗ trên mặt đất. Hai là phân bố của các loài cây gỗ trên mặt đất chịu ảnh hưởng bởi những nguyên nhân nào? Sự phân bố của các loài cây gỗ trên mặt đất là do nhiều nguyên nhân khác nhau. Một là sự hình thành từng cụm cây ở một nơi nào đó có liên quan đến đặc tính sinh học của loài. Chẳng hạn: (a) Một số loài sinh sản vô tính bằng chồi gốc; (b) Một số loài có hạt giống nặng và phát tán theo trọng lực trong phạm vi không gian nhất định quanh gốc cây mẹ. Hai là tính không thuần nhất của môi trường trong QXTV. Chẳng hạn: (a) Địa hình không bằng phẳng; (b) Độ che phủ của cây bụi không đồng đều; (c) Động vật sử dụng hạt cây gỗ là nguồn thức ăn. Ba là sự kết nhóm sinh thái giữa các loài cây gỗ.

Nói chung, xác định rõ những nguyên nhân làm phát sinh các kiểu phân bố của các loài cây gỗ trong QXTV cho phép nhà lâm học hiểu rõ đặc tính sinh thái của các loài cây gỗ, đặc điểm tái sinh rừng, diễn thế rừng, điều kiện môi trường của QXTV và mối quan hệ qua lại giữa các loài cây gỗ...Từ đó nhà lâm học có thể xây dựng những biện pháp xử lý rừng.

3.3.2.2. Các kiểu phân bố của các loài cây gỗ trên mặt đất

Các loài cây gỗ có thể phân bố trên mặt đất theo ba kiểu: điều hòa; ngẫu nhiên; cụm. Phân bố điều hòa biểu hiện ở chỗ mật độ của loài cây gỗ biến động rất nhỏ trên những ô mẫu có kích thước bằng nhau. Nói cách khác, số lượng cá thể của loài có xác suất bằng nhau trên một khoảng cách hay diện tích bằng nhau (Hình 3.2a). Kiểu phân bố này không xuất hiện ở rừng tự nhiên mà chỉ có trong những hệ sinh thái rừng do con người tạo ra.



Hình 3.2. Những kiểu phân bố của các loài cây gỗ trên mặt đất.
(a) Phân bố điều hòa; (b) Phân bố ngẫu nhiên; (c) Phân bố cụm.

Phân bố ngẫu nhiên xuất hiện khi điều kiện môi trường trong QXTV là đồng nhất (Hình 3.2b). Trên quan điểm thống kê, tính đồng nhất của môi trường biểu hiện rõ trong những trường hợp sau đây: (1) Mỗi cá thể của loài cây gỗ quan tâm có xác suất bắt gặp như nhau tại bất kỳ vị trí nào trong QXTV; (2) Xác suất bắt gặp một loài cây gỗ quan tâm ở một nơi nào đó không phụ thuộc vào những cá thể cùng loài hoặc khác loài. Theo quan điểm này, phân bố ngẫu nhiên cũng được hiểu là phân bố đều. Kiểu phân bố này cũng rất ít gặp trong các hệ sinh thái rừng tự nhiên.

Phân bố cụm còn gọi là phân bố theo đám. Kiểu phân bố này biểu hiện ở chỗ những cá thể của cùng một loài cây gỗ hình thành những cụm ở một số khu vực nào đó trong QXTV, nhưng lại vắng mặt ở những nơi khác (Hình 3.2c). Đây là kiểu phân bố thường gặp trong các hệ sinh thái rừng tự nhiên. Kiểu phân bố cụm có liên quan đến tính không thuần nhất của môi trường và đặc tính sinh học - sinh thái của các loài cây gỗ. Đây là kiểu phân bố thu hút sự chú ý lớn nhất của các nhà sinh thái học và lâm học. Bởi vì thông qua việc làm rõ nguyên nhân hình thành kiểu phân bố cụm, nhà lâm học có thể hiểu rõ đặc tính sinh thái của các loài cây gỗ. Từ đó nhà lâm học có thể xây dựng những biện pháp xử lý rừng.

3.3.2.3. Phương pháp phân tích phân bố của các loài cây gỗ trên mặt đất

Tùy theo mục tiêu nghiên cứu, nhà lâm học có thể sử dụng những phương pháp khác nhau để xác định phân bố của các loài cây gỗ trên mặt đất. Phương pháp đơn giản nhất là quan sát bằng mắt. Tuy vậy, kết quả của phương pháp này phụ thuộc vào chủ quan của nhà lâm học. Một phương pháp khác là ứng dụng những mô hình phân bố xác suất. Phương pháp này nhận được kết quả khách quan với độ tin cậy cao. Tuy vậy, phương pháp này đòi hỏi nhiều thời gian, nhân lực và kinh phí. Dưới đây giới thiệu một số phương pháp có thể được ứng dụng trong phân tích phân bố cây gỗ trên mặt đất.

(1) Xác định phân bố cây trên mặt đất theo mô hình phân bố nhị thức. Phân bố nhị thức có dạng như hàm 3.19; trong đó $P_{(k)}$ là xác suất bắt gặp k ô mẫu (k = 0, 1, 2...n) xuất hiện loài cây quan tâm (Công thức 3.20), còn q là xác suất không bắt gặp loài cây quan tâm (q = 1 - p).

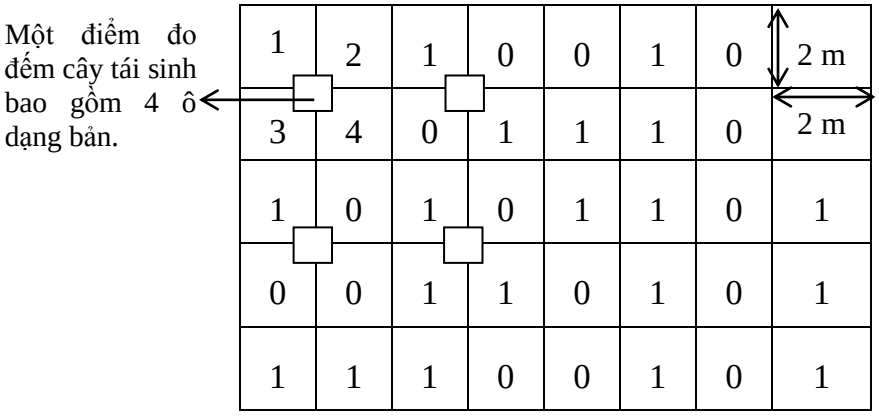
$$(p + q)^n \qquad (3.19)$$

$$P_{(k)} = C(n, k)p^k q^{n-k} \text{ với } C(n, k) = \frac{n!}{k!(n - k)!} \qquad (3.20)$$

Để dễ dàng bố trí ô mẫu và tính toán, người ta thường chọn n = 4 hoặc 6 ô mẫu. Khi n = 4 ô mẫu, thì xác suất bắt gặp 0, 1, 2, 3 và 4 ô mẫu có loài cây quan tâm là $P_0 = q^4$; $P_1 = 4pq^3$; $P_2 = 6p^2q^2$; $P_3 = 4p^3q$ và $P_4 = p^4$. Những bước tính toán tiếp theo bao gồm xác định tần số xuất hiện 0, 1, 2, 3, 4 ô mẫu có loài cây quan tâm và kiểm định giả thuyết về dạng phân bố bằng tiêu chuẩn χ^2 .

Trong thưc hành, khi ứng dụng phân bố nhị thức để phân tích phân bố của các loài cây gỗ trên mặt đất, các ô mẫu có thể được bố trí theo một số phương pháp dưới đây.

Phương pháp 1. Giả thiết nghiên cứu phân bố cây tái sinh của một loài cây gỗ nào đó trên mặt đất. Trước hết, phân chia đối tượng nghiên cứu (QXTV) thành những tuyến song song cách đều. Mỗi tuyến có bề rộng từ 2 - 4 m tùy theo loài cây gỗ và tuổi hay kích thước của chúng. Kế đến phân chia các tuyến thành N phân đoạn; trong đó chiều dài của mỗi phân đoạn cũng bằng 2 - 4 m. Cách bố trí các phân đoạn như vậy tạo thành những ô mẫu có dạng hình vuông (Hình 3.3).



Hình 3.3. Sơ đồ bố trí hệ thống ô mẫu để phân tích phân bố cây tái sinh trên mặt đất. Ký hiệu: 1 = bắt gặp cây tái sinh, 0 = không bắt gặp cây tái sinh.

Hình dạng ô mẫu có thể là hình vuông, hình chữ nhật hoặc hình tròn, còn kích thước ô mẫu là tùy ý. Mặc dù vậy, để đơn giản trong tính toán, số lượng ô mẫu tại 1 điểm quan sát chỉ nên dao động từ 4 – 6. Để nhận được kết quả chính xác và khách quan, số phân đoạn cần bố trí phải đủ lớn, thông thường từ 100 – 500 phân đoạn. Tiếp theo phân chia mỗi phân đoạn hay mỗi ô mẫu thành n ô dạng bản. Nếu sử dụng ô mẫu 2*2 m (4 m^2) và phân chia thành 4 ô dạng bản, thì kích thước mỗi ô dạng bản là 1 m^2. Thông tin thu thập trong mỗi ô dạng bản là độ bắt gặp cây tái sinh (bắt gặp = 1; không bắt gặp = 0).

Phương pháp 2. Trong đối tượng nghiên cứu, bố trí các ô mẫu theo phương pháp chọn mẫu ngẫu nhiên hoặc theo phương pháp chọn mẫu hệ thống. Giả thiết bố

trí 200 điểm để nghiên cứu phân bố cây tái sinh của một loài cây gỗ nào đó. Tại mỗi điểm, kích thước ô mẫu là 2*3 m (6 m^2). Mỗi ô mẫu được phân chia thành 6 ô dạng bản; mỗi ô dạng bản 1 m^2. Các ô dạng bản cũng có thể được bố trí liền nhau theo dạng bàn cờ hoặc theo 1 dãy liên tục. Thông tin thu thập trong mỗi ô dạng bản là độ bắt gặp (1, 0) cây tái sinh.

Từ hai cách bố trí ô mẫu trên đây, chúng ta có thể xác định được tỷ lệ số ô dạng bản bắt gặp loài cây quan tâm (p) và tỷ lệ số ô dạng bản không bắt gặp loài cây quan tâm (q = 1 - p). Gọi $p_{(K)}$ là xác suất mà các phân đoạn của tuyến chứa k ô dạng bản bắt gặp loài cây quan tâm. Theo nhị thức Niutơn, chúng ta có P(k = 0, 1…n) = $C^k_n p^k q^{n-k}$. Tần số lý thuyết của những cụm ô xuất hiện k ô dạng bản bắt gặp loài cây quan tâm là N*P(k). Sự phù hợp của phân bố nhị thức so với số liệu thực tế được kiểm định bằng tiêu chuẩn χ^2 (Công thức 3.21); trong đó F_{TN} = tần số cây thực nghiệm; F_{UL} = tần số cây ước lượng.

$$\chi^2 = \sum_{i=1}^{n} \frac{(F_{TN} - F_{UL})^2}{F_{UL}} \qquad (3.21)$$

Kiểm định χ^2 chỉ cho kết quả chính xác khi tần số của các cấp (tổ, nhóm) lớn hơn hoặc bằng 5. Giả thuyết H^0: Phân bố của loài cây quan tâm trên mặt đất là ngẫu nhiên. Quy tắc quyết định: Nếu $\chi^2 < \chi^2_{05}$ (hoặc P > α = 0,05) với số bậc tự do là f = m - 2 (m = số cấp sau khi gộp để tần số của các cấp lớn hơn hoặc bằng 5), thì phân bố của loài cây quan tâm trên mặt đất phù hợp với phân bố nhị thức. Ngược lại, $\chi^2 > \chi^2_{05}$ (hoặc P < α = 0,05), thì giả thuyết bị bác bỏ.

Bảng 3.1 dẫn ví dụ phân tích phân bố của loài *Potentilla acaulis* trên một tuyến bao gồm 500 ô dạng bản liền nhau. Kích thước mỗi ô dạng bản là 1m^2. Tuyến được phân chia thành 100 phân đoạn với mỗi phân đoạn 5 ô dạng bản. Theo số liệu của Bảng 3.1, p = 0,362; q = 0,638; n = 5; N = 100. Để tìm xác suất của những điểm bắt gặp 0, 1, 2, 3, 4 và 5 ô dạng bản bắt gặp loài cây này, chúng ta phải khai triển nhị thức: $(0,362 + 0,638)^5$. Giả thuyết H^0: Phân bố của loài *Potentilla acaulis* trên mặt đất là ngẫu nhiên. Kết quả tính toán ở Bảng 3.1 cho thấy $\chi^2 = 41,5 > \chi^2_{05} = 16,3$. Từ đó có thể nhận định rằng phân bố trên mặt đất của loài *Potentilla acaulis* không tuân theo luật phân bố ngẫu nhiên mà tuân theo luât phân bố cụm. Điều đó được giải thích là do tần số ước lượng những phân đoạn có 1, 2, 3 ô dạng bản bắt gặp loài cây này lớn hơn rất nhiều so với tần số thực nghiệm. Lưu ý rằng những phân tích thống kê trên đây đơn thuần chỉ là toán học. Về mặt lâm học, chúng ta cố gắng tìm lời giải thích vì sao loài cây này phân bố thành từng cụm trên mặt đất? Mặt khác, khi báo cáo kết quả, chúng ta chỉ viết "Kết quả phân tích thống kê cho thấy loài *Potentilla acaulis* phân bố thành từng đám trên mặt đất ($\chi^2 = 41,5 > \chi^2_{05} = 16,3$ hoặc P < 0,05)".

Bảng 3.1. Phân bố của *Potentilla acaulis* trên 100 phân đoạn, mỗi phân đoạn 5 ô dạng bản và so sánh với phân bố nhị thức. Nguồn: Vasilevich, 1969.

Số ô bắt gặp *P. acaulis*	0	1	2	3	4	5
Số phân đoạn của tuyến	23	26	22	12	10	7
Tần số ước lượng	10,6	29,9	34,0	19,3	5,5	0,6
χ^2	14,5	0,51	4,24	2,75	19,5	
$\chi^2_{0,05}$			16,3			

(2) Xác định phân bố cây trên mặt đất theo mô hình phân bố Poisson. Phân bố Possion có dạng như hàm 3.22; trong đó e = 2,7182, m = số cây trung bình trên ô mẫu, k = số cây bắt gặp trong một ô mẫu.

$$P(k) = \frac{e^{-m} m^k}{k!} \qquad (3.22)$$

Khi k = 5, thì P(k = 0) = exp(-m); P(k = 1) = m*exp(-m); P(k = 2) = $\frac{e^{-m} m^2}{2}$; P(k = 3) = $\frac{e^{-m} m^3}{6}$; P(k = 4)= $\frac{e^{-m} m^4}{24}$; P(k = 5) = $\frac{e^{-m} m^5}{120}$.

Giả thiết nghiên cứu phân bố trên mặt đất đối với một loài cây gỗ nào đó bằng cách áp dụng mô hình phân bố Possion. Trong thực hành, chúng ta bố trí N ô mẫu. Để nhận được kết quả chính xác và khách quan, số lượng ô mẫu phải đủ lớn, thường từ 100 – 500. Các ô mẫu có thể bố trí liên tục trên các tuyến song song cách đều hoặc theo dạng bàn cờ. Kích thước của các ô mẫu phải thống nhất và thay đổi tùy theo loài cây, tuổi hay kích thước thân cây. Kích thước của các ô mẫu cũng cần giới hạn để dễ tính toán. Đối với cây tái sinh có H < 50 cm, do mật độ cao, nên kích thước của các ô mẫu thường chỉ dao động từ 1 – 4 m^2. Đối với cây tái sinh có H > 100 cm, do mật độ thấp, nên kích thước của các ô mẫu thường dao động từ 4 – 10 m^2. Đối với cây gỗ trưởng thành, do mật độ thấp, nên kích thước của các ô mẫu thường dao động từ 10 – 50 m^2.

Thông tin thu thập trong mỗi ô mẫu là số lượng cá thể của loài cây quan tâm. Mỗi ô mẫu có thể bắt gặp 0, 1, 2...k cá thể của loài cây quan tâm. Từ tài liệu này, chúng có thể phân tích đặc điểm phân bố của loài cây quan tâm trên mặt đất bằng cách áp dụng mô hình phân bố Poisson. Để kiểm định sự phù hợp của phân bố Poisson với phân bố số cây trên mặt đất, chúng ta thực hiện các bước sau đây: (1) Xác định tần số bắt gặp 0, 1, 2...k cây trên ô mẫu; (2) Kiểm định sự phù hợp của phân bố Poisson với số liệu thực nghiệm theo tiêu chuẩn χ^2 với số bậc tự do bằng số cấp sau khi gộp (điều kiện là tần số của mỗi cấp phải lớn hơn 5) trừ 2.

Bảng 3.2 ghi lại phân bố trên mặt đất đối với cây tái sinh Dầu con rái (*Dipterocrapus alatus*) ở cấp H < 50 cm. Số liệu nghiên cứu bao gồm 400 ô dạng bản. Kích thước ô dạng bản là 1 m^2. Kết quả tính toán từ công thức 3.22 cho thấy m = 0,98; N = 400 ô mẫu và χ^2_{TT} = 4,9. Vì χ^2 = 4,9 < $\chi^2_{0.05}$ = 5,9, nên phân bố trên mặt đất đối với cây tái sinh Dầu con rái tuân theo phân bố Poisson. Nói khác đi, cây tái sinh Dầu con rái phân bố ngẫu nhiên trên mặt đất. Hiện tượng này xảy ra được giải

thích là do: (a) Mật độ quần thể cao và ra nhiều quả; (b) Quả phát tán vừa theo trọng lực vừa theo gió; (c) Cây bụi và thảm cỏ phát triển rất kém trên mặt đất rừng.

Bảng 3.2. Kiểm định phân bố cây tái sinh Dầu con rái trên mặt đất theo mô hình phân bố Poisson**.**

Số cây/ô mẫu	0	1	2	3	4	Tổng
Tần số ô mẫu thực nghiệm	158	140	62	32	8	400
Tần số ô mẫu ước lượng	150,1	147,1	72,1	24,0	6,0	399
χ^2	0,393	0,362	1,642	1,985	0,496	4,9

(3) Xác định phân bố cây trên mặt đất theo kiểm định T. Kiểm định T được tính theo công thức 3.23; trong đó n_1 và n_2 tương ứng là số ô mẫu bắt gặp (1) và không bắt gặp (0) loài cây quan tâm, K là số cụm ô mẫu lặp lại liên tiếp ở dạng bắt gặp và không bắt gặp loài cây quan tâm. Khi biết T, chúng ta so sánh T với $T_{0.5}$ ở mức ý nghĩa P = 0,05 hoặc 0,01. Khi giá trị T < - 2, T = 0 ÷ ± 2 và T > 2, thì phân bố của loài cây quan tâm trên mặt đất tương ứng có dạng phân bố cụm, ngẫu nhiên và đồng đều (Terenchiev, 1964; Dẫn theo Vasilevich, 1969).

$$T = \frac{K - \dfrac{2n_1 n_2}{n_1 + n_2} + 1}{\sqrt{\dfrac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}}} \qquad (3.23)$$

Tuyến I	1	1	0	1	0	0	0	1
II	0	1	1	0	1	0	1	1
III	1	1	0	0	1	0	1	0
IV	0	1	1	0	0	0	1	1
V	1	1	0	1	1	0	1	0

Hình 3.4. Sơ đồ bố trí các ô mẫu để phân tích phân bố cây trên mặt đất. Kí hiệu: (1) và (0) tương ứng là ô mẫu bắt gặp và không bắt gặp loài cây quan tâm.

Trong thực hành, khi áp dụng phương pháp này, các ô mẫu được bố trí liên tục trên các tuyến song song cách đều dưới tán rừng. Bề rộng của tuyến tùy thuộc vào loài cây gỗ. Đối với cây tái sinh, thông thường bề rộng của tuyến là 2 đến 4 m, còn chiều dài của tuyến được bố trí tùy ý. Tiếp theo, phân chia các tuyến thành những phân đoạn đều nhau với chiều dài mỗi phân đoạn bằng bề rộng của tuyến.

Mỗi phân đoạn là một ô mẫu (Hình 3.4). Thông tin về loài cây quan tâm trong mỗi ô mẫu bao gồm hai dấu hiệu bắt gặp (1) và không bắt gặp (0).

Bảng 3.3 dẫn ví dụ xác định phân bố trên mặt đất đối với cây tái sinh Dầu song nàng (*Dipterocarpus dyeri*). Tổng số 4 tuyến điều tra. Kích thước mỗi ô mẫu là 1 m^2 (1*1 m). Số ô mẫu bố trí trên các tuyến thay đổi từ 80 - 160. Những tính toán từ công thức 3.23 (Bảng 3.3) cho thấy, T < -2 nên cây tái sinh Dầu song nàng phân bố trên mặt đất theo kiểu phân bố cụm.

Bảng 3.3. Xác định phân bố cây tái sinh Dầu song nàng trên mặt đất bằng tiêu chuẩn T. Nguồn: Nguyễn Văn Thêm, 1992.

Chỉ tiêu thống kê	Thứ tự tuyến thí nghiệm:			
	1	2	3	4
1. Tổng số ô mẫu (N)	160	80	111	115
2. Số ô mẫu bắt gặp loài (n_1)	103	36	57	42
3. Số ô mẫu không bắt gặp loài (n_2)	57	44	54	73
4. Tổng số lô (K)	60	22	40	34
5. Kiểm định T	- 2,5	+ 4,1	- 3,14	- 4,1
6. Giá trị $T_{0,05}$	2,58	2,66	2,63	2,63

*(4) Xác định phân bố cây trên mặt đất theo chỉ số đồng góc***.** Ba kiểu phân bố (đồng đều, ngẫu nhiên, cụm) của cây gỗ trên mặt đất trong QXTV có thể được xác định theo phương pháp chỉ số đồng góc (Uniform Angle Index = U). Chỉ số U đo mức độ đồng đều về góc hợp bởi cây cơ sở i và J cây xung quanh. Chỉ số U (Công thức 3.24) là tỷ lệ giữa các góc (W_J) nhỏ hơn góc tiêu chuẩn (W_0); trong đó k là số cây xung quanh cây cơ sở, W_J là góc hợp bởi cây cơ sở i và k cây xung quanh. Góc tiêu chuẩn hay góc cơ sở $W_0 = 360^0/(k + 1)$; trong đó k là số cây xung quanh phân bố gần nhất với cây cơ sở. Trong đánh giá kết quả nghiên cứu, nếu $W_J < W_0$, thì $W_J = 1$. Trái lại, nếu $W_J > W_0$, thì $W_J = 0$. Để dễ dàng phân tích kết quả, người ta thường chọn J = 3 hoặc 4.

$$U_i = \frac{\sum_{(J=1,\,k)} W_J}{k} \qquad (3.24)$$

Để đánh giá phân bố cây trên mặt đất, chúng ta bố trí N điểm nghiên cứu và tính giá trị U_i trung bình. Khi J = 3 và J = 4, thì kiểu phân bố của cây gỗ trên mặt đất được đánh giá tương ứng theo Bảng 3.4 và Bảng 3.5.

(5) Xác định phân bố của cây gỗ trên mặt đất theo chỉ số quần tụ. Kiểu phân bố cây gỗ trên mặt đất (đồng đều, ngẫu nhiên, cụm) có thể được xác định bằng chỉ số quần tụ. Theo Clark và Evans (1954; Dẫn theo Pommerening, A., 2002), chỉ số quần tụ được xác định theo công thức 3.25; trong đó R = chỉ số quần tụ, N = mật độ của loài cây gỗ quan tâm trên ô mẫu, Z = diện tích ô mẫu, R_{Bq} = khoảng cách trung bình từ cây cơ sở đến cây gần nhất, E(r) = khoảng cách trung bình kỳ vọng từ cây cơ sở đến cây gần nhất đối với phân bố ngẫu nhiên (Công thức 3.26).

Bảng 3.4. Kiểu phân bố của cây gỗ trên mặt đất. Trường hợp J = 3.

U_i	Số cây có W_i khác với W_0	Kiểu phân bố
(1)	(2)	(3)
0	0 cây gần nhất với $W_i < W_0$.	Rất đồng đều
0,33	1 cây gần nhất có $W_i < W_0$.	Đồng đều
0,67	2 cây gần nhất có $W_i < W_0$.	Ngẫu nhiên
1,0	3 cây gần nhất có $W_i < W_0$.	Cụm

Bảng 3.5. Kiểu phân bố của cây gỗ trên mặt đất. Trường hợp J = 4.

U_i	Số cây có W_i khác với W_0	Kiểu phân bố
(1)	(2)	(3)
0	0 cây gần nhất với $W_i < W_0$.	Rất đồng đều
0,25	1 cây gần nhất có $W_i < W_0$.	Đồng đều
0,50	2 cây gần nhất có $W_i < W_0$.	Ngẫu nhiên
>0,75	3 cây gần nhất có $W_i < W_0$.	Cụm

$$R = R_{Bq}/E(r) \quad (3.25)$$

$$E(r) = \frac{1}{\sqrt{N/Z}} \quad (3.26)$$

Quy tắc quyết định:

Nếu R > 1, thì cây gỗ phân bố đều.

Nếu R = 1, thì cây gỗ phân bố ngẫu nhiên.

Nếu R < 1, thì cây gỗ phân bố cụm.

(6) Xác định phân bố của cây gỗ trên mặt đất theo chỉ số Whitford. Theo Whitford (1948; Dẫn theo Dẫn theo Vasilevich, 1969), kiểu phân bố của một loài cây gỗ trên mặt đất trong QXTV được xác định theo tỷ lệ giữa độ phong phú (A) và độ bắt gặp (F) của nó (Công thức 3.27). Độ phong phú của loài cây gỗ thứ i (A_i) trên 1 ô mẫu là tỷ lệ gữa số cá thể của loài i (N_i) và tổng số ô mẫu bắt gặp loài i (T_i). Độ thường gặp của loài cây gỗ thứ i (F_i) được xác định theo công thức 3.28; trong đó T_i = số ô mẫu bắt gặp loài i, T = tổng số ô mẫu quan sát.

$$A_i/F_i = \frac{(N_i/T_i)}{F_i} \quad (3.27)$$

$$F_i = \frac{T_i}{T} * 100 \quad (3.28)$$

Quy tắc quyết định:

Nếu tỷ lệ A/F < 0,025, thì loài cây gỗ phân bố đồng đều.

Nếu tỷ lệ A/F = 0,025 - 0,05, thì loài cây gỗ phân bố ngẫu nhiên.

Nếu tỷ lệ A/F > 0,05, thì loài cây gỗ phân bố theo đám.

(7) Phân tích phân bố cây gỗ trên mặt đất theo mô hình phân bố cụm. Xác định những mô hình biểu diễn những đặc tính phân bố của cây gỗ trên mặt đất theo dạng phân bố cụm là mối quan tâm của các nhà lâm học. Neyman và Thomas (1949;

Dẫn theo Vasilevich, 1969) đã xây dựng những mô hình để biểu diễn phân bố của cây gỗ trên mặt đất theo dạng phân bố cụm.

Hàm mật độ xác suất của phân bố Neyman có dạng như hàm 3.29; trong đó k là số cây bắt gặp trong ô mẫu, m_1 = giá trị bình quân của các cụm cây trong một ô mẫu, m_2 = số cây bình quân trong 1 cụm. Giá trị m_2 và m_1 được tính theo công thức 3.30 và 3.31; trong đó X_{Bq} và S^2 tương ứng là số cây bình quân và phương sai đối với số cây trong ô mẫu.

$$P(X = k + 1) = \frac{m_1 m_2 e^{-m_2}}{k + 1} \sum_{t=0}^{t=k} \frac{m_2^t}{t!} P(X = k - t) \quad (3.29)$$

$$m_2 = \frac{(S^2 - X_{Bq})}{X_{Bq}} \quad (3.30)$$

$$m_1 = \frac{X_{Bq}}{m_2} \quad (3.31)$$

Hàm mật độ xác suất của phân bố Thomas có dạng như hàm 3.32; trong đó k là số cây bắt gặp trong ô mẫu, m = số cụm cây trong ô mẫu.

$$P(k) = \sum_{r=1}^{k} \frac{(m^r e^{-m} (r\lambda)^{k-r} e^{-\lambda r})}{r!(k - r)!}; \quad (3.32)$$

Giá trị bình quân (μ_1) của phân bố Thomas được tính theo công thức 3.33; trong đó m là số cụm cây trong ô mẫu, còn $1 + \lambda$ là số cây bình quân/cụm. Phương sai (μ_2) của phân bố Thomas được xác định theo công thức 3.34. Hai tham số m và λ được xác định tương ứng theo công thức 3.35 và 3.36; trong đó n_0 và n_1 tương ứng là số ô mẫu không có cây và số ô mẫu chỉ bắt gặp 1 cây, N là tổng số ô mẫu được điều tra.

$$\mu_1 = m*(1 + \lambda) \quad (3.33)$$

$$\mu_2 = m*(1 + 3\lambda + \lambda^2) \quad (3.34)$$

$$m = - Ln(\frac{n_0}{N}) \quad (3.35)$$

$$\lambda = - Ln(\frac{n_1}{m*n_0}) \quad (3.36)$$

Khi ứng dụng hai phân bố Neyman và Thomass để mô tả phân bố cây trên mặt đất, trước hết xác định xác suất $P_{(K)}$. Đối với phân bố Neyman (Hàm 3.29), xác suất $P_{(K)}$ được khai triển như sau:

$$P_{(X=0)} = \exp(- m_1(1 - e^{-m_2})); \; P_{(X=1)} = m_1 m_2 \exp(- m_2)* P_0;$$

$$P_{(X=2)} = \frac{m_1 m_2 \exp(- m^2)}{2} *(P_1 + m_2 P_0);$$

$$P_{(X=3)} = \frac{m_1 m_2 \exp(- m^2)}{3} *(P_2 + m_2 P_1 + \frac{m_2^2}{2} P_0)\ldots$$

Đối với phân bố Thomass (Hàm 3.32), xác suất $P_{(K)}$ được khai triển như sau:

$Po = \exp(-m)$; $P_1 = m*\exp(-m)*\exp(-\lambda)$;

$P_2 = m*\exp(-m)*\lambda*\exp(-\lambda) + \frac{m^2 e^{-m}}{2!}(\exp(-\lambda))^2 \ldots$

Kế đến ước lượng tần số ô mẫu xuất hiện k cây bằng cách nhân tổng số ô mẫu (N) với P(K = 0,1, 2...). Sau đó kiểm định sự phù hợp của phân bố Neyman và phân bố Thomass với số liệu thực nghiệm bằng tiêu chuẩn χ^2 với số bậc tự do f = n – 3; trong đó n là số tổ sau khi gộp để tần số lớn hơn 5. Nếu $\chi_{TT}^2 > \chi^2_{05}$, thì phân bố của loài cây quan tâm không phù hợp với phân bố Neyman và Thomass. Ngược lại, nếu $\chi_{TT}^2 < \chi^2_{05}$, thì phân bố của loài cây quan tâm phù hợp với phân bố Neyman và Thomass.

Bước tiếp theo là xác định nguyên nhân mà loài cây này phân bố cụm và kích thước trung bình của những cụm cây. Những nguyên nhân của phân bố cụm có thể là do (a) Mật độ quần thể cao; (b) Cây mẹ sinh sản mạnh; (c) Kiểu cách phát tán quả và hạt; (d) Tình trạng phát triển của cây bụi và thảm cỏ dưới tán rừng; (e) Hoạt động của động vật...

Lưu ý rằng, nếu chỉ sử dụng các ô mẫu được bố trí ngẫu nhiên trong QXTV, thì chúng ta rất khó xác định độ lớn của một cụm cây. Kích thước của các cụm cây có thể được xác định bằng cách đo vẽ chi tiết ranh giới của chúng. Đây là một công việc khó khăn, tốn nhiều thời gian và nhân lực. Trong thực hành, khi ứng dụng các mô hình toán (Nhị thức, Possion, Neyman và Thomas) để phân tích phân bố cây trên mặt đất, chúng ta cần phải thu thập số lượng ô mẫu đủ lớn, thường từ 400 – 1600. Để kết hợp phân tích kích thước của các cụm cây, các ô mẫu cần được bố trí theo dạng bàn cờ.

Theo Greig - Smith (1952; Dẫn theo Vasilevich, 1969), độ lớn của các cụm cây có thể được xác định theo các bước dưới đây.

Bước 1. Tại khu vực nghiên cứu, thiết lập một hệ thống các ô mẫu khép kín theo dạng bàn cờ sao cho có thể liên kết các ô cơ sở ở lân cận thành từng khối K = 2, 4, 8, 16, 32, 64, 128, 256...ô mẫu. Mỗi ô mẫu là một ô cơ sở hay ô đơn vị.

Bước 2. Xác định mật độ (N_i), số cây bình quân (N_{Bq}) và phương sai (S^2) tương ứng với mỗi khối ô mẫu khác nhau.

Bước 3. Xây dựng đồ thị biểu diễn sự biến đổi của S^2 theo kích thước khối (K). Khi xây dựng đồ thị, trục tung đặt S^2, còn trục hoành đặt kích thước khối (K).

Bước 4. Nối các điểm giao nhau giữa S^2 và K tương ứng. Kết quả nhận được một đường cong có nhiều đỉnh. Các đỉnh này tương ứng với diện tích bình quân của từng đám cá thể. Bảng 3.6 và 3.7 và Hình 3.5 dẫn ví dụ phân tích kích thước các cụm cây tái sinh tự nhiên của Dầu song nàng. Kiểu phân bố cụm của Dầu song nàng được mô tả gần đúng bằng mô hình phân bố Thomas và Neyman.

Bảng 3.6. Kiểm định phân bố trên mặt đất đối với cây tái sinh Dầu song nàng theo phân bố Thomas và Neyman. Nguồn: Nguyễn Văn Thêm, 1992.

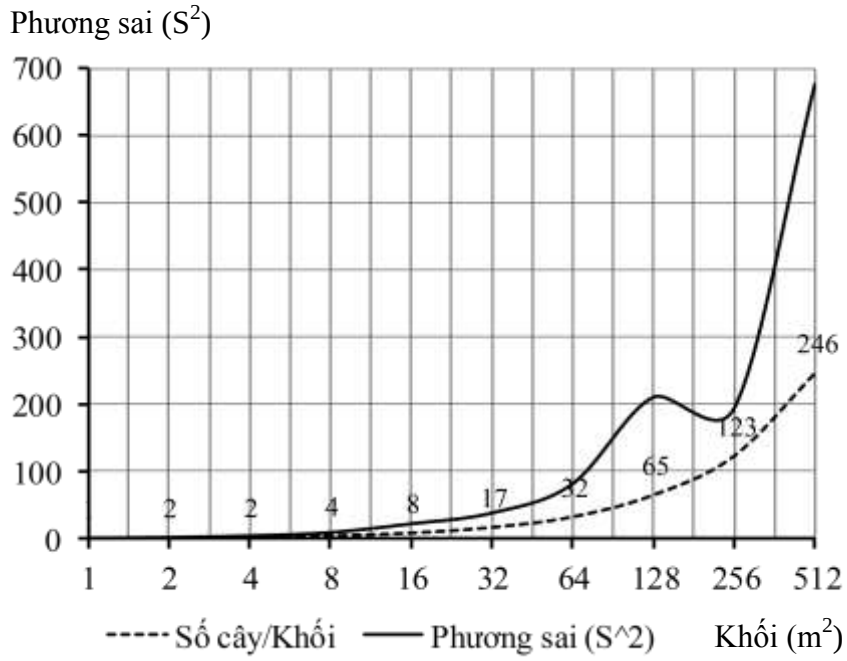
Số cây/ô mẫu	Số ô mẫu bắt gặp	Số ô mẫu ước lượng theo phân bố:	
		Thomas	Neyman
0	159	159,0	158,2
1	113	113,0	113,9
2	71	72,3	71,2
3	36	37,0	37,0
4	19	16,8	17,1
5	6	6,9	7,2
6	3	2,6	2,8
≥7	2	0,1	0,003
Tổng số	409	409	409
χ^2		0,555	0,363
χ^2_{05}		7,81	7,81
Các đặc trưng	X_{Bq} = 1,2 S^2 = 2,1	m = 0,945; λ = 0,285 X_{Bq} = 1,22; S^2 = 1,88	m_1 = 2,3045 m_2 = 0,5315

Theo phương pháp của Greig – Smith (1952), xác định được kích thước bình quân của các cụm cây tái sinh Dầu song nàng là 128 m^2; trong đó trung bình một cụm là 65 cây. Giả thiết hình thái của một cụm cây là hình tròn. Từ đó xác định được đường kính trung bình của một cụm cây tái sinh Dầu song nàng là 11 – 12 m. Kích thước này gần bằng đường kính tán trung bình của một cây mẹ (7 - 12 m). Quả Dầu song nàng rất to và thường phát tán theo trọng lực. Ngoài ra, Dầu song nàng thường tái sinh theo lỗ trống. Vì thế, phân bố cây tái sinh theo dạng cụm có liên quan đến kiểu cách tái sinh tự nhiên của Dầu song nàng.

Bảng 3.7. Sự phụ thuộc của phương sai theo kích thước ô thống kê.

TT	Khối (m^2)	Số khối (n)	X_{Bq} (cây/ô mẫu)	S^2	$W = S^2/X_{Bq}$
10	512	2	246	676,0	2,75
9	256	4	123	194,0	1,58
8	128(*)	10	64,8	210,4	3,23
7	64	25	32,28	81,6	2,53
6	32	50	16,64	38,0	2,28
5	16	100	8,32	21,9	2,63
4	8	200	4,15	9,0	2,17
3	4	400	2,10	4,1	1,95
2	2	800	1,64	1,72	1,65
1	1	1600	0,52	0,73	1,40

(*) Độ lớn của một cụm cây tái sinh.



Hình 3.5. Mối quan hệ giữa phương sai với kích thước khối ô mẫu.

Sau khi xác định kiểu phân bố trên mặt đất (điều hòa, ngẫu nhiên và cụm) của những loài cây quan tâm, một nhiệm vụ quan trọng khác là tìm kiếm những nguyên nhân dẫn đến kiểu phân bố này. Trong sinh thái học và lâm học, người ta đặc biệt quan tâm đến những nguyên nhân dẫn đến loài cây này hay loài cây khác phân bố theo cụm, còn các dạng phân bố khác ít được chú ý. Để làm rõ nguyên nhân ảnh hưởng đến kiểu phân bố cụm của các loài cây gỗ, nhà lâm học cần phải thu thập những thông tin về môi trường sống của loài cây quan tâm, đặc biệt là vi môi trường trong mỗi ô mẫu. Về cơ bản, phân bố cụm của các loài cây là do những nguyên nhân sau đây: (1) Tính không thuần nhất của địa hình và đất; (2) Sự phát triển mạnh của cây bụi và thảm cỏ; (3) Quan hệ qua lại giữa các loài cây; (4) Đặc tính sinh học của loài (sinh sản vô tính hay hữu tính); (5) Ảnh hưởng của độ tàn che tán rừng; (6) Kiểu cách tái sinh của loài cây quan tâm; (7) Sự cạnh tranh trong loài và giữa các loài cây...Nói chung, xác định rõ những nguyên nhân của các kiểu phân bố cây trên mặt đất giúp cho nhà lâm học giải thích bản chất của hiện tượng và xây dựng các phương thức lâm sinh.

Khi áp dụng các mô hình toán để phân tích phân bố cây trên mặt đất, về lý thuyết, độ lớn của các ô mẫu bằng bao nhiêu cũng được. Nhưng trong thực tiễn, ý nghiã của các kết luận và chi phí nghiên cứu phụ thuộc rất nhiều vào cách chọn lựa kích thước ô mẫu. Sử dụng ô mẫu qúa nhỏ hoặc qúa lớn đều làm giảm ý nghĩa của các kết luận. Thật vậy, khi sử dụng ô mẫu quá lớn, thì nó có thể bao trùm lên nhiều không gian không có cây. Ngược lại, khi sử dụng ô mẫu quá nhỏ, thì nhiều ô mẫu lại rơi vào chỗ trống. Ở trường hợp đầu phân bố cây có thể là phân bố cụm, nhưng chúng ta lại kết luận là phân bố đồng đều. Ngược lại, ở trường hợp sau, mật độ cây

thưa thớt, nhưng chúng ta lại kết luận là phân bố cụm. Ngoài ra, kích thước ô mẫu lớn chứa rất nhiều cây. Điều đó lại gây ra những khó khăn cho việc thống kê số cây trong ô mẫu, tính toán phức tạp và tốn nhiều thời gian hơn. Vì vậy, khi nghiên cứu phân bố cây trên mặt đất, kích thước của các ô mẫu phải thống nhất và thay đổi tùy theo loài cây, tuổi hay kích thước thân cây. Kích thước của các ô mẫu cũng cần giới hạn để dễ tính toán. Đối với cây tái sinh có H < 50 cm, do mật độ cao, nên kích thước của các ô mẫu thường chỉ dao động từ 1 – 4 m^2. Đối với cây tái sinh có H > 100 cm, do mật độ thấp, nên kích thước của các ô mẫu thường dao động từ 4 – 10 m^2. Đối với cây gỗ trưởng thành, do mật độ thấp, nên kích thước của các ô mẫu thường dao động từ 10 – 50 m^2. Diện tích này bằng diện tích trung bình của hình chiếu tán nằm ngang của một cây trưởng thành. Thông tin thu thập trong mỗi ô mẫu là số lượng cá thể của loài cây quan tâm.

3.3.3. Phân tích tính phức tạp của rừng

3.3.3.1. Khái niệm về tính phức tạp của rừng

Tính phức tạp của rừng phản ánh sự đa dạng về loài cây, tuổi, kích thước cây gỗ và quần thụ, trạng thái sức sống của cây gỗ và quần thụ, những mối quan hệ giữa các loài cây gỗ với nhau và giữa chúng với môi trường. Tính phức tạp về cấu trúc rừng là do rừng bao gồm nhiều đặc tính khác nhau. Các đặc tính của rừng có thể được phân chia thành 4 nhóm. Nhóm 1 là những đặc tính về các loài cây hình thành rừng. Đặc tính này biểu thị sự giàu có về các loài cây. Nhóm 2 là sự đa dạng về tuổi và kích thước của các cây gỗ. Đặc tính này xuất hiện là do sự đa dạng về loài cây và nhiều loài cây tái sinh liên tục dưới tán rừng. Nhóm 3 là sự đa dạng về trạng thái sức sống của cây gỗ và quần thụ. Nhóm 4 là sự đa dạng về những mối quan hệ giữa các loài cây và giữa chúng với môi trường. Phân tích tính phức tạp của rừng không chỉ giúp cho nhà lâm học hiểu về rừng, mà còn xây dựng những phương thức lâm sinh và phân loại rừng.

3.3.3.2. Những chỉ số biểu thị tính phức tạp của rừng

Rừng bao gồm rất nhiều đặc tính khác nhau. Một số đặc tính dễ đo đạc. Chẳng hạn: (a) Ở mức cây gỗ, các đặc tính dễ đo đạc là D, H, V, D_T, L_T. (b) Ở mức quần thụ, các đặc tính dễ đo đạc là N, G và M. Trái lại, QXTV rừng có nhiều đặc tính rất khó đo đạc chính xác. Chẳng hạn: số loài cây (S); sinh khối (B)...Vì thế, để đơn giản trong đo đạc, tính toán, so sánh, đánh giá và phân loại QXTV và rừng, các nhà sinh thái học và lâm học đề nghị sử dụng các chỉ tiêu tổng hợp. Tính phức tạp về cấu trúc rừng có thể được đánh giá bằng chỉ số phức tạp về cấu trúc (SCI = Structural Complexity Index). Các chỉ số SCI là một dạng toán học. Chúng được sử dụng để biểu thị sự phức tạp về cấu trúc quần thụ. Chúng tóm tắt ảnh hưởng của hai hoặc nhiều đặc tính của quần thụ.

Các chỉ số SCI có thể được xác định bằng hai phương pháp. Một là các chỉ số SCI được xác định từ tổng số cây trong ô mẫu. Hai là các chỉ số SCI được xác định từ k cây gần nhất với cây cơ sở. Các chỉ số SCI có thể ở dạng tổng điểm số của các đặc tính và dạng tích số giữa các đặc tính của QXTV.

Nói chung, các chỉ số SCI được xây dựng theo 5 bước. Bước 1 là xác định những đặc tính cơ bản của rừng được sử dụng để đánh giá tính phức tạp về cấu trúc. Bước 2 là phát triển những hệ thống số đo các đặc tính cấu trúc của rừng. Bước 3 là thu thập những đặc tính của rừng. Bước 4 là xác định phương pháp tính các chỉ số SCI. Nếu sử dụng chỉ số SCI ở dạng tổng các đặc tính, thì chúng ta cần phải phân chia thang điểm và cho điểm đối với mỗi đặc tính. Những điểm số được cho từ cao xuống thấp. Đặc tính quan trọng có điểm số cao nhất. Điểm số ở dạng tuyệt đối hoặc tương đối. Phương pháp tổng có ưu điểm là chỉ số SCI có thể bao gồm nhiều đặc tính định lượng và định tính. Nhược điểm cơ bản nằm ở phương pháp cho điểm. Chỉ số SCI ở dạng tích các đặc tính có ưu điểm là tính toán dễ dàng. Nhược điểm cơ bản là khó đánh giá đối với các chỉ tiêu định tính. Bước 5 là tổng hợp điểm số của các đặc tính để nhận được điểm số của các chỉ số SCI. Tính phức tạp của các QXTV và rừng được sắp xếp theo thứ tự giảm dần điểm số của các chỉ số SCI.

Dưới đây giới thiệu một số phương pháp xác định chỉ số SCI đối với QXTV và rừng dựa theo một và nhiều biến số khác nhau.

(1) Những chỉ số phức tạp dựa trên một biến số. Gadow và Hui (2007) xác định chỉ số hỗn giao của các loài cây gỗ trong QXTV theo công thức 3.37; trong đó M_i là chỉ số hỗn giao tại điểm i trong QXTV, k là số loài cây gỗ gần nhất với cây cơ sở, V_J là loài cây gỗ phân bố bên cạnh cây cơ sở. Giá trị $M_i = 0 – 1$. Tại mỗi điểm quan sát, nếu loài cây gỗ bên cạnh khác với loài cây gỗ cơ sở, thì $V_j = 1$. Trái lại, nếu loài cây gỗ bên cạnh giống với loài cây gỗ cơ sở, thì $V_j = 0$. Nếu xác định sự hỗn giao giữa một loài cây gỗ quan tâm với những loài cây gỗ khác, thì chọn loài cây gỗ quan tâm ở một số điểm khác nhau trong QXTV.

$$M_i = \frac{\sum_{(i=1,\,k)} V_i}{k} \qquad (3.37)$$

Để dễ dàng tính toán và nhận được kết quả với độ tin cậy cao, người ta thường chọn k = 3 hoặc 4 cây gỗ phân bố xung quanh cây cơ sở và giá trị M_i được tính bình quân từ N điểm khác nhau trong QXTV, ít nhất N = 3 – 5 điểm. Khi phân tích và đánh giá kết quả nghiên cứu, mức độ hỗn giao giữa các loài cây gỗ được đánh giá theo Bảng 3.8.

Bảng 3.8. Phân cấp mức độ hỗn giao của các loài cây gỗ trong quần thụ.
(a) Trường hợp chọn 3 cây gần nhất với cây cơ sở.

Mi	Số loài cây giống với loài cây cơ sở	Mức độ hỗn giao
(1)	(2)	(3)
0	3 cây gần nhất cùng loài với cây cơ sở.	Không
0,33	2 cây gần nhất cùng loài với cây cơ sở.	Thấp
0,67	1 cây gần nhất cùng loài với cây cơ sở.	Trung bình
1,0	0 cây gần nhất cùng loài với cây cơ sở.	Cao

(b) Trường hợp chọn 4 cây gần nhất với cây cơ sở.

Mi	Số loài cây giống với loài cây cơ sở	Mức độ hỗn giao
(1)	(2)	(3)
0	4 cây gần nhất cùng loài với cây cơ sở.	Không
0,25	3 cây gần nhất cùng loài với cây cơ sở.	Thấp
0,50	2 cây gần nhất cùng loài với cây cơ sở.	Trung bình
0,75	1 cây gần nhất cùng loài với cây cơ sở.	Cao
1,0	4 cây gần nhất không cùng loài với cây cơ sở.	Rất cao

Bảng 3.9 dẫn ví dụ về mức độ hỗn giao của QXTV được xác đinh từ 3 điểm quan sát. Từ đó cho thấy, giá trị M_i trung bình là 0,375 cho biết các loài cây gỗ của QXTV này hỗn giao ở mức trung bình. Trong thực hành, đối với rừng tự nhiên nhiệt đới, chúng ta có thể phân tích sự hỗn giao giữa hai loài cây gỗ ưu thế sinh thái và đồng ưu thế sinh thái với một vài loài cây gỗ có ý nghĩa kinh doanh.

Bảng 3.9. Mức độ hỗn giao của QXTV được xác đinh từ 3 điểm quan sát.

Quan sát	Loài cây gần nhất với cây cơ sở				Tổng	MI_i	Hỗn giao
1	0	0	0	0	0	0,00	Không
2	0	0	0	1	1	0,25	Thấp
3	0	0	1	1	2	0,50	Trung bình
Trung bình						0,38	Trung bình

Sự phân hóa về kích thước cây gỗ là mối quan tâm của nhà lâm học. Kích thước cây gỗ (D, H, G, V) là những biến định lượng. Vì thế, tính phức tạp của rừng có thể được xác định bằng việc phân tích đặc trưng thống kê (Trung bình, sai tiêu chuẩn, Max, Min, hệ số biến động...) đối với những biến định lượng. Sự đa dạng về D của QXTV có thể được phân chia thành 2 kiểu. Kiểu thứ 1 là sự khác biệt về biên độ D của các cây gỗ trong quần thụ. Kiểu thứ 2 là sự đồng đều về phân bố số cây trong các cấp D. Một cách khác là phân tích sự khác biệt về kích thước cây gỗ ở những vị trí khác nhau trong QXTV. Sau đó tổng hợp lại để nhận được chỉ số phức tạp về cấu trúc của QXTV.

Theo Aguirre và ctv (2003), sự khác biệt về kích thước cây gỗ ở những vị trí khác nhau trong QXTV được xác định theo chỉ số ưu thế về kích thước thân cây (Dominance Index = DI). Chỉ số DI phản ánh mức độ khác biệt về kích thước (D, H, G, V) của những cây bên cạnh so với cây cơ sở. Chỉ số DI là tỷ lệ kích thước của cây cơ sở với k cây gần nhất (Công thức 3.38). Ở công thức 3.38, nếu kích thước của cây gỗ bên cạnh nhỏ hơn so với kích thước của cây gỗ cơ sở, thì $V_j = 0$. Trái lại, nếu kích thước của cây gỗ bên cạnh lớn hơn so với kích thước của cây gỗ cơ sở, thì $V_j = 1$.

$$DI_i = \frac{\sum_{(i=1,\,k)} V_i}{k} \qquad (3.38)$$

Để dễ tính toán và nhận được kết quả với độ tin cậy cao, chọn 3 hoặc 4 cây gỗ phân bố xung quanh cây cơ sở và giá trị DI_i được tính bình quân từ N điểm khác nhau trong QXTV, ít nhất N = 3 – 5 điểm. Nếu phân tích sự khác biệt về kích thước của loài cây gỗ quan tâm với những loài cây gỗ khác, thì chọn loài cây quan tâm ở một số vị trí khác nhau. Trong đánh giá kết quả nghiên cứu, mức độ đồng đều hay mức độ phân hóa về kích thước giữa các cây gỗ trong QXTV và rừng được đánh giá theo Bảng 3.10.

Bảng 3.10. Phân cấp mức độ đồng đều về kích thước của các cây gỗ trong QXTV. Trường hợp chọn 4 cây gần nhất với cây cơ sở.

DI_i	Số cây có kích thước khác với cây cơ sở	Mức độ đồng đều
(1)	(2)	(3)
0	4 cây gần nhất nhỏ hơn cây cơ sở.	Ưu thế trội
0,25	3 cây gần nhất nhỏ hơn cây cơ sở.	Ưu thế
0,50	2 cây gần nhất nhỏ hơn cây cơ sở.	Trung bình
0,75	1 cây gần nhất nhỏ hơn cây cơ sở.	Bị chèn ép
1,0	4 cây gần nhất lớn hơn cây cơ sở.	Bị chèn ép mạnh

Bảng 3.11 dẫn ví dụ xác định mức độ đồng đều về D thân cây từ 3 điểm quan sát. Từ đó cho thấy, giá trị DI trung bình là 0,375 cho biết D thân cây gỗ trong QXTV này phân hóa ở mức trung bình.

Bảng 3.11. Mức độ đồng đều về kích thước của các cây gỗ trong QXTV được xác định từ 3 điểm quan sát.

Quan sát	Kích thước cây gần nhất				Tổng	DI_i	Mức độ đồng đều
1	0	0	0	0	0	0,00	Ưu thế trội
2	0	0	0	1	1	0,25	Ưu thế
3	0	0	1	1	2	0,50	Trung bình
Trung bình						0,38	Trung bình

Kint (2003) đã xác định chỉ số khác biệt về D (T_D) theo tỷ lệ D của cây cơ sở và những cây xung quanh (Công thức 3.39). Ở công thức 3.39, d_{ij} là tỷ lệ D của cây nhỏ (D_{Min}) và D của cây lớn (D_{Max}) trong cặp cây cơ sở (i) và cây gần nhất (j); k là số cây gần nhất với cây cơ sở. Để dễ tính toán và nhận được kết quả với độ tin cậy cao, chọn k = 3 hoặc 4 cây gỗ phân bố xung quanh cây cơ sở. Giá trị T_D nằm trong khoảng 0 – 1. Trong đánh giá kết quả, mức độ khác biệt về D của những cây gỗ tại mỗi vị trí trong QXTV được đánh giá theo 4 cấp: thấp ($T_D \leq 0,3$); trung bình (T_D = 3 ÷ 0,5); cao (T_D = 0,5 ÷ 0,7) và rất cao (T_D = 0,7 ÷ 1,0). Để nhận được kết quả khách quan và phản ánh đúng sự khác biệt về D của các cây gỗ trong QXTV, chỉ số T_D được tính bình quân từ N điểm khác nhau trong QXTV, ít nhất là 3 điểm (Bảng 3.12).

$$T_D = \frac{\sum_{(i=1,\,k)}(1 - d_{ij})}{k} \qquad (3.39)$$

Bảng 3.12. Mức độ khác biệt về đường kính của các cây gỗ trong QXTV được xác định từ 3 điểm quan sát.

D_i (cm)	D_J (cm) của 4 cây so sánh				Giá trị d_{iJ}	Khác biệt D
	1	2	3	4		
10	24	22	8	6	0,40	Trung bình
22	32	17	18	15	0,32	Trung bình
45	38	40	32	25	0,44	Trung bình
Trung bình					0,39	Trung bình

Khi xác định sự đa dạng về D của cây gỗ trong quần thụ, Clifford và Stephenson (1975; dẫn theo Kint, 2003) đã sử dụng chỉ số $d_{Margalef}$, chỉ số H' của Shannon, chỉ số Simpson và chỉ số Berger - Parker. Các chỉ số này được sử dụng để đo mức độ phong phú và sự phân bố đồng đều đối với G ở các cấp D. Chỉ số $d_{Margalef}$ được xác định theo công thức 3.40; trong đó F = số cấp D, G = tiết diện ngang quần thụ (m^2/ha). Phạm vi biến đổi của chỉ số $d_{Margalef}$ từ 0 - ∞; trong đó $d_{Margalef}$ = 0 khi F = 1 và $d_{Margalef}$ = ∞ khi F = ∞.

$$d_{Margalef} = \frac{(F - 1)}{Ln(G)} \qquad (3.40)$$

Chỉ số Shannon (1948) được xác định theo công thức 3.41; trong đó P_i là tỷ lệ giữa tiết diện ngang của từng cấp D (g_i, m^2) và tổng tiết diện ngang của quần thụ (G, m^2) trên ô mẫu. Phạm vi biến đổi của chỉ số H' từ 0 – Ln(F); trong đó F = số cấp D.

$$H' = - \Sigma^{F}_{i = 1} P_i * Ln(P_i) \qquad (3.41)$$

Chỉ số Simpson (1949) được xác định theo công thức 3.42; trong đó P_i là tỷ lệ giữa tiết diện ngang của từng cấp D (g_i, m^2) và tổng tiết diện ngang của quần thụ (G, m^2) trên ô mẫu, F = số cấp D. Phạm vi biến đổi của chỉ số $D_{Simpson}$ từ 0 – 1.

$$D_{Simpson} = 1 - \Sigma^{F}_{i = 1} P_i^2 \qquad (3.42)$$

Chỉ số Berger - Parker (1970) được sử dụng để xác định sự ưu thế về độ phong phú của các cấp D. Chỉ số này được xác định theo công thức 3.43; trong đó g_{iMax} là tiết diện ngang của cấp D có tiết diện ngang lớn nhất, G là tổng tiết diện ngang của quần thụ (G, m^2). Phạm vi biến đổi của chỉ số D_{BP} từ 0 – 1; trong đó D_{BP} = 0 khi số cấp D = 1 và D_{BP} = 1 khi tiết diện ngang của các cấp D bằng nhau.

$$D_{BP} = 1 - (\frac{g_{iMax}}{G}) \qquad (3.43)$$

Sự đồng đều về D của các cây gỗ trong QXTV được xác định theo chỉ số đồng đều của Shannon (Công thức 3.44); trong đó H' là chỉ số Shannon, F là số cấp D. Nếu các loài có độ phong phú bằng nhau, thì E = 1. Khi độ phong phú của các loài khác nhau, thì E sẽ giảm. Vì thế, giá trị H' của những ô mẫu được so sánh với

H'_{Max} = 1. Nếu độ phong phú tập trung rất cao vào một số loài, thì chỉ số H' sẽ tiến đến Zero. Nếu đối tượng nghiên cứu chỉ có 1 loài, thì H' = 0.

$$E_{Shannon} = H'/Ln(F) \quad (3.44)$$

(b) Những chỉ số phức tạp dựa trên nhiều biến số. Khi phân tích mức độ hỗn giao của các loài cây gỗ trong QXTV trên các ô mẫu, Nguyễn Văn Trương (1984) đã xác định chỉ số hỗn giao (HG) theo công thức 3.45; trong đó S và N tương ứng là số loài cây gỗ và mật độ quần thụ trên ô mẫu.

$$HG = (S/N) \quad (3.45)$$

Holdridge (1967; Dẫn theo Cintrón và Schaeffer-Novelli, 1984) đã xây dựng chỉ số SCI ở dạng tích số giữa số loài cây gỗ (S), mật độ (N, cây), chiều cao (H, m) và tiết diện ngang (G, m^2) của quần thụ trên ô mẫu (Công thức 3.46); trong đó 10^k (k = 1 – 6) là hệ số chuyển SCI về giá trị nhỏ.

$$SCI = (S*N*H*G)/10^k \quad (3.46)$$

Jaehne và Dohrenbusch (1997; Dẫn theo Neumann and Starlinger, 2001) đã xây dựng chỉ số phức tạp về cấu trúc (SD) từ 4 đặc tính: biến động D (S_d); biến động số loài (Ss); biến động không gian sống của cây (Y) và biến động diện tích tán che trên mặt đất (S_C) (Công thức 3.47).

$$SD = (S_d*Ss*Y*S_C) \quad (3.47)$$

Neumann và Starlinger (2001) đã sử dụng cả hai chỉ số SCI và SD để đánh giá tính phức tạp của một số kiểu rừng ở Áo. Füldner (1995; Dẫn theo Pommerening, 2003) đã xác định đa dạng cấu trúc rừng theo ba chỉ số: TM, M và W. Chỉ số TM, M và W tương ứng biểu thị sự khác biệt về D, loài cây gỗ và kiểu phân bố của cây gỗ trong quần thụ. Chỉ số TM được tính theo công thức 3.48; trong đó k = số cây so sánh với cây cơ sở, d_{ij} = tỷ lệ giữa D của cây nhỏ (i) và cây lớn (J) trong cặp cây cơ sở và cây gần nhất với cây cơ sở.

$$TM_i = \frac{\sum_{(i=1,\,k)}(1 - d_{ij})}{k} \quad (3.48)$$

Để dễ tính toán và so sánh sự khác biệt về D của các cây gỗ trong QXTV, người ta thường chọn k = 3 hoặc 4. Mặc khác, để nhận được kết quả khách quan và phản ánh gần đúng sự khác biệt về D của các cây gỗ trong QXTV, chỉ số TM_i được tính bình quân từ N điểm khác nhau, ít nhất N = 3 – 5 điểm. Trong đánh giá kết quả, sự khác biệt về D được đánh giá theo 4 cấp: thấp ($T_D \leq 0,3$); trung bình ($T_D = 3 \div 0,5$); cao ($T_D = 0,5 \div 0,7$) và rất cao ($T_D = 0,7 \div 1,0$).

Chỉ số M được xác định theo công thứ 3.49; trong đó V_{ij} biểu thị sự khác biệt về loài cây gỗ, k = số cây so sánh với cây cơ sở. Giá trị V_{ij} = 1 nếu cây cơ sở (i) và cây so sánh (j) thuộc cùng 1 loài. Ngược lại, V_{ij} = 0 nếu hai cây (i) và (j) không thuộc cùng 1 loài. Chỉ số M nhận giá trị từ 0 – 1. Trong đánh giá kết quả, nếu M_i < 0,3, M_i = 0,3 – 0,5 và M_i > 0,5, thì sự khác biệt về loài cây gỗ tương ứng ở mức thấp, trung bình và cao.

$$M_i = \frac{\sum_{(i=1,\,k)} V_{ij}}{k} \qquad (3.49)$$

Chỉ số W được xác định theo công thức 3.50; trong đó W_{ij} là góc hợp bởi cặp cây cơ sở (i) và cây so sánh (j), k là số cây xung quanh (j) phân bố gần nhất với cây cơ sở (i)

$$W_i = \frac{\sum_{(i=1,\,k)} W_{ij}}{k} \qquad (3.50)$$

Nếu k cây gỗ phân bố đồng đều trên mặt đất, thì góc tiêu chuẩn hay góc cơ sở (W_0) = $360^0/(k + 1)$. Trong đánh giá kết quả nghiên cứu, nếu $W_j < W_0$, thì $W_j = 1$. Trái lại, nếu $W_j > W_0$, thì $W_j = 0$. Để dễ dàng phân tích kết quả, người ta thường chọn k = 3 hoặc 4. Khi đánh giá phân bố cây trong QXTV, bố trí N điểm nghiên cứu và tính giá trị W_i trung bình. Khi k = 3 và 4, thì kiểu phân bố cây trên mặt đất được đánh giá tương ứng theo Bảng 3.13.

Bảng 3.13. Phân chia các kiểu phân bố cây trên mặt đất.

(a) Trường hợp chọn 3 cây gần nhất với cây cơ sở.

W_i	Số cây có W_i khác với W_0	Kiểu phân bố
(1)	(2)	(3)
0	0 cây gần nhất với $W_i < W_0$.	Rất đồng đều
0,33	1 cây gần nhất có $W_i < W_0$.	Đồng đều
0,67	2 cây gần nhất có $W_i < W_0$.	Ngẫu nhiên
1,0	3 cây gần nhất có $W_i < W_0$.	Cụm

(b) Trường hợp chọn 4 cây gần nhất với cây cơ sở.

W_i	Số cây có W_i khác với W_0	Kiểu phân bố
(1)	(2)	(3)
0	0 cây gần nhất với $W_i < W_0$.	Rất đồng đều
0,25	1 cây gần nhất có $W_i < W_0$.	Đồng đều
0,50	2 cây gần nhất có $W_i < W_0$.	Ngẫu nhiên
>0,75	3 cây gần nhất có $W_i < W_0$.	Cụm

Pastorella và Paletto (2013) đã xác định mức độ phức tạp về cấu trúc quần thụ (SCI) bằng tổng số điểm của ba chỉ số TM_i, M_i và W_i (Công thức 3.51); trong đó k_i (i = 1, 2, 3) là trọng số của ba chỉ số TM_i, M_i và W_i. Ba trọng số này nhận giá trị tương ứng là 0,2; 0,5 và 0,3. Căn cứ vào phạm vi biến động của chỉ số SCI, các QXTV được phân chia thành 4 mức độ phức tạp về cấu trúc: rất phức tạp, phức tạp, trung bình và ít phức tạp.

$$SCI = (TM*k_1) + (M*k_2) + (W_i*k_3) \qquad (3.51)$$

Nói chung, trong các tài liệu sinh thái học và lâm học, người ta còn sử dụng rất nhiều đặc tính để xây dựng các chỉ số phức tạp về cấu trúc rừng. Ở mức cây gỗ, các đặc tính thường được sử dụng là D, H, G, V, B, diện tích tán che. Ở mức quần thụ, các đặc tính thường được sử dụng là N, S, G, M, số cấp D và cấp H của quần

thụ, số lượng cây chết đứng, tình trạng phát triển của cây bụi và thảm cỏ dưới tán rừng, bề dày của lớp vật rụng trên sàn rừng. Rừng tự nhiên hỗn loài nhiệt đới là những hệ sinh thái rất phức tạp. Vì thế, để đơn giản trong phân tích và so sánh giữa các kiểu QXTV và phân loại rừng, cấu trúc của các QXTV và rừng thường được mô tả bằng các chỉ số cấu trúc.

Chương 4
PHÂN TÍCH MỐI QUAN HỆ GIỮA CÁC LOÀI CÂY GỖ

4.1. Giới thiệu chung

Rừng bao gồm nhiều loài cây gỗ khác nhau; trong đó mỗi loài cây gỗ đòi hỏi những điều kiện sống khác nhau. Mặt khác, khi chung sống bên nhau, các loài cây gỗ có sự tương tác với nhau như che bóng lẫn nhau và cạnh tranh với nhau về nguồn sống (ánh sáng, nước, chất khoáng...). Sự tương tác giữa các loài cây gỗ có thể ở dạng cạnh tranh với nhau và ở dạng hỗ trợ lẫn nhau. Căn cứ vào mối quan hệ giữa các loài cây gỗ trong QXTV, các nhà sinh thái học và lâm học phân chia chúng thành ba nhóm: (1) Những loài cây gỗ không có quan hệ với nhau hay độc lập với nhau; (2) Những loài cây gỗ có quan hệ yếu với nhau; (3) Những loài cây gỗ có quan hệ chặt chẽ với nhau. Sự đa dạng về các mối quan hệ giữa các loài cây đã tạo ra sự đa dạng về kết cấu loài cây gỗ và cấu trúc rừng. Trong sinh thái học và lâm học, người ta gọi những loài cây gỗ chung sống trong một nơi ở và có yêu cầu môi trường sống tương tự như nhau là nhóm sinh thái (Ecological Group). Mối quan hệ giữa các loài cây gỗ trong cùng nhóm sinh thái được gọi là sự kết nhóm sinh thái (Ecological Group Association).

Nghiên cứu các nhóm sinh thái mang lại những ý nghĩa khác nhau. Về lý luận, kết quả của nghiên cứu này giúp cho nhà lâm học hiểu về đời sống của các loài cây gỗ và rừng. Các phương thức nuôi dưỡng và khai thác rừng đều hướng đến mục tiêu tạo ra kết cấu loài cây gỗ tối ưu và cấu trúc quần thụ hợp lý. Để đạt được hai mục tiêu này, lâm học cần hiểu rõ những nhóm sinh thái và mối quan hệ giữa các loài cây trong nhóm sinh thái. Vì thế, phân chia các loài cây gỗ thành các nhóm sinh thái khác nhau là cơ sở khoa học cho việc xây dựng những phương thức lâm sinh, quản lý rừng và bảo tồn những loài cây gỗ quý, hiếm hoặc có giá trị cao về kinh tế.

Về toán học, người ta phân biệt hai kiểu phụ thuộc khác nhau. Một là sự phụ thuộc hàm số. Đó là giá trị của một đặc tính nào đó được xác định chính xác thông qua những đặc tính khác. Chẳng hạn: Thể tích thân cây (V) phụ thuộc vào D và H. Mối quan hệ này được xác định bằng các hàm số. Hai là sự phụ thuộc xác suất. Đó là một đặc tính thay đổi có thể nhận được nhiều giá trị ngẫu nhiên của những đặc tính khác. Chẳng hạn: Tại cấp D = 20 cm, giá trị H có thể là 18, 20 và 22 m. Mối quan hệ này được nghiên cứu bằng phương pháp phân tích hồi quy và tương quan.

Rừng là một hệ thống rất phức tạp; trong đó mỗi quần thể loài bị kiểm soát bởi nhiều quần thể loài khác và nhiều nhân tố môi trường khác nhau. Sự phụ thuộc giữa các loài cây gỗ trong QXTV có thể được nghiên cứu trên ba mức độ không gian khác nhau. Trên mức độ địa lý, lâm học nghiên cứu sự cùng chung sống của các loài cây gỗ trong một địa phương và đánh giá sự trùng hợp của chúng trong các

vùng địa lý. Trên mức độ QXTV, lâm học nghiên cứu: (a) Khả năng của các loài cây gỗ tham gia vào cùng một QXTV và những nhân tố kiểm soát sự sống của chúng; (b) Sự trùng hợp biên độ sinh thái của các loài cây gỗ. Trên mức giữa các loài cây gỗ trong cùng QXTV, lâm học nghiên cứu quan hệ qua lại giữa các loài cây gỗ.

Mặc dù sự phụ thuộc giữa các loài cây gỗ được nghiên cứu trên ba mức độ không gian khác nhau, nhưng ranh giới giữa mức QXTV và các loài trong cùng QXTV là rất khó phân định. Nguyên nhân là vì khi tăng kích thước ô mẫu, chúng ta đã chuyển từ việc nghiên cứu quan hệ giữa các loài trong một vi môi trường thành quan hệ giữa chúng trong QXTV. Ở mức QXTV, nhà lâm học quan tâm đến các loài cây gỗ phân bố ở tầng phiến nào, dạng sống nào và vi môi trường nào. Ở mức các QXTV, nhà lâm học không chỉ xác định những loài cây gỗ quan tâm thuộc về một QXTV hay nhiều QXTV, mà còn cả sự tương đồng về điều kiện sống của các QXTV.

Bằng phân tích hồi quy và tương quan, nhà lâm học có thể xác định được sự kết nhóm giữa các loài cây gỗ với nhau và mối quan hệ giữa chúng với môi trường. Khi phân tích sự kết nhóm giữa các loài cây gỗ, thì biến phản hồi và biến giải thích là do quy ước của nhà nghiên cứu. Nếu mục tiêu nghiên cứu là xác định sự biến đổi độ phong phú của loài A dựa theo sự biến đổi độ phong phú của loài B, thì độ phong phú của loài A là biến phản hồi, còn độ phong phú của loài B là biến giải thích. Trái lại, nếu mục tiêu nghiên cứu là xác định sự biến đổi độ phong phú của loài B dựa theo độ phong phú của loài A, thì độ phong phú của loài B là biến phản hồi, còn độ phong phú của loài A là biến giải thích. Thuật ngữ "Phản hồi" biểu thị phản ứng hay sự thay đổi những đặc tính của cây gỗ và rừng đối với sự thay đổi của các nhân tố môi trường. Thuật ngữ "Biến giải thích" biểu thị những nhân tố môi trường được sử dụng để giải thích sự thay đổi của những đặc tính ở mức cây gỗ và rừng.

Mối quan hệ giữa các loài cây gỗ có thể được nghiên cứu bằng nhiều phương pháp khác nhau. Một là phương pháp xây dựng bảng, biểu đồ và đồ thị tương quan. Hai là lập bảng chéo và xác định mối quan hệ bằng kiểm định χ^2. Ba là sử dụng phương pháp phân tích hồi quy và tương quan. Nói chung, tùy theo mục đích nghiên cứu và phương pháp thu thập số liệu, trình tự nghiên cứu mối quan hệ giữa các loài cây gỗ được thực hiện theo 4 bước: (1) Xác định sự tồn tại của mối quan hệ giữa các loài cây gỗ; (2) Xác định khuynh hướng và cường độ của mối quan hệ giữa các loài cây gỗ; (3) Phân tích hồi quy để xác định mức độ ảnh hưởng của loài cây gỗ này đến những loài cây gỗ khác; (4) Sử dụng các hàm hồi quy để ước lượng và dự đoán biến phản hồi. Nói chung, tùy theo mục đích nghiên cứu, phân tích sự kết nhóm giữa các loài cây gỗ trong QXTV có thể được thực hiện đầy đủ theo 4 bước hay chỉ một số bước. Khi mục tiêu nghiên cứu là xác định những loài cây gỗ có quan hệ với nhau, thì nghiên cứu chỉ dừng lại ở bước 1. Nếu mục tiêu nghiên cứu là xác định sự tương tác giữa các loài cây gỗ trong các nhóm sinh thái, thì quá trình nghiên cứu phải trải qua 3 bước từ 1 – 3. Nếu mục tiêu nghiên cứu là xác định các

nhóm sinh thái và động thái biến đổi của chúng theo thời gian, thì quá trình nghiên cứu phải thực hiện qua 4 bước.

4.2. Phân tích sự kết nhóm giữa các loài cây gỗ

4.2.1. Các kiểu kết nhóm giữa các loài cây gỗ

Trong sinh thái học và lâm học, người ta phân chia sự kết nhóm giữa hai hay nhiều loài cây gỗ trong QXTV theo bốn dạng khác nhau.

Một là sự kết nhóm âm. Kiểu kết nhóm này phản ánh sự gia tăng độ phong phú của loài cây gỗ này dẫn đến sự suy giảm độ phong phú của loài cây gỗ khác. Bản chất của hiện tượng này là sự cạnh tranh giữa các loài cây gỗ trong QXTV.

Hai là sự kết nhóm dương. Kiểu kết nhóm này phản ánh sự gia tăng độ phong phú của loài cây gỗ này kéo theo sự gia tăng độ phong phú của loài cây gỗ khác. Bản chất của hiện tượng này là sự cộng sinh giữa các loài cây gỗ hoặc sự hỗ trợ lẫn nhau giữa các loài cây gỗ trong quá trình sống. Từ mối quan hệ dương giữa các loài cây gỗ, chúng ta xác định được những loài cây gỗ thuộc cùng một nhóm sinh thái.

Ba là không tồn tại sự kết nhóm. Điều đó có nghĩa là sự thay đổi độ phong phú của loài cây gỗ này độc lập với sự thay đổi độ phong phú của loài cây gỗ khác. Bản chất của hiện tượng này là các loài cây gỗ phân bố độc lập với nhau. Sự cùng có mặt của chúng ở một nơi ở nào đó có thể là do chúng có biên độ sinh thái giống nhau.

Bốn là sự phụ thuộc phức tạp. Mối quan hệ này phản ánh các loài cây gỗ không chỉ có quan hệ với nhau mà còn quan hệ với nhiều nhân tố khác nhau.

4.2.2. Phương pháp phân tích sự kết nhóm giữa các loài cây gỗ

4.2.2.1. Phương pháp hồi quy và tương quan

Phương pháp này được áp dụng trong trường hợp độ phong phú của các loài cây gỗ được đo bằng những đơn vị đo nào đó. Mối quan hệ giữa các loài cây gỗ có thể được phân tích bằng hồi quy và tương quan đơn biến và đa biến. Chi tiết của phân tích hồi quy và tương quan đơn biến và đa biến được giới thiệu ở Chương 5. Về cơ bản, phân tích hồi quy và tương quan bao gồm 5 bước cơ bản dưới đây.

Bước 1. Thu thập số liệu độ phong phú của những loài cây gỗ quan tâm trong QXTV. Độ phong phú của chúng có thể là D, H, G, V và B; trong đó ba biến quan trọng lá G, V và B.

Bước 2. Phân tích mối quan hệ giữa từng cặp loài cây gỗ. Mối quan hệ giữa hai loài cây gỗ được xác định bằng phân tích tương quan giữa độ phong phú của chúng. Kết quả phân tích tương quan cho biết cường độ và khuynh hướng của mối quan hệ giữa hai loài cây gỗ. Cường độ của mối quan hệ giữa hai loài cây gỗ được đánh giá theo độ lớn của hệ số tương quan. Khuynh hướng của mối quan hệ giữa hai loài cây gỗ được xác định theo dấu của hệ số tương quan.

Bước 3. Xây dựng mô hình hồi quy. Khi đặc tính của các loài cây gỗ quan tâm biểu hiện mối quan hệ rõ rệt, thì sử dụng phân tích hồi quy để xác định mô hình biểu diễn mối quan hệ giữa chúng. Về lý thuyết, các biến phản hồi và các biến giải thích tồn tại mối quan hệ rất phức tạp. Nếu chỉ phân tích mối quan hệ giữa hai loài cây gỗ, thì xây dựng mô hình hồi quy đơn biến. Nếu phân tích mối quan hệ giữa một loài cây gỗ với nhiều loài cây gỗ khác, thì xây dựng mô hình hồi quy đa biến. Trình tự phân tích hồi quy bắt đầu từ xác định hình thái của mối quan hệ. Điều này có thể được thực hiện bằng phương pháp biểu đồ. Kế đến xây dựng mô hình biểu diễn mối quan hệ giữa các đặc tính của các loài cây gỗ. Mục đích của phân tích hồi quy không chỉ là xác định mô hình biểu diễn tốt khuynh hướng của mối quan hệ giữa biến phản hồi và các biến giải thích, mà còn nhận được sai số ước lượng và sai số dự đoán thấp nhất. Thông thường dạng mô hình thích hợp được xác định bằng biểu đồ hoặc từ những nghiên cứu trước đó. Tuy vậy, với cùng số liệu như nhau, nhà nghiên cứu này sử dụng mô hình A, còn nhà nghiên cứu khác lại chọn mô hình B. Vì thế, kết quả ước lượng và dự đoán có thể khác nhau. Các hệ số và các thống kê sai lệch của các hàm hồi quy được ước lượng bằng phương pháp bình phương nhỏ nhất hoặc phương pháp phân tích hồi quy và tương quan phi tuyến tính. Mối quan hệ và sai lệch của các hàm hồi quy so với số liệu thực nghiệm được đánh giá theo hệ số xác định (r^2) (Công thức 4.1); sai số tuyệt đối trung bình (MAE) (Công thức 4.2); sai số tuyệt đối trung bình theo phần trăm (MAPE) (Công thức 4.3); tổng sai lệch bình phương (SSE) (Công thức 4.4); sai lệch chuẩn của ước lượng (SEE) (Công thức 4.5). Ở công thức 4.1 – 4.5, Y_{TN} là biến phản hồi thực tế; Y_{Bq} là biến phản hồi thực tế bình quân; Y_{UL} là biến phản hồi ước lượng; p là số tham số trong mô hình. Mô hình phù hợp nhất được chọn theo tiêu chuẩn SSE_{Min} hoặc SEE_{Min}.

$$R^2 = [1- \frac{\sum(Y_{TN} - Y_{UL})^2}{\sum(Y_{TN} - Y_{Bq})^2}]*100 \qquad (4.1)$$

$$MAE = \left| ((Y_{TN} - Y_{UL})/n)) \right| \qquad (4.2)$$

$$MAPE = (MAE/Y_{TN})*100 \qquad (4.3)$$

$$SSE = \sum(Y_{TN} - Y_{UL})\text{^}2 \qquad (4.4)$$

$$SEE = \sqrt{SSE/(n - p)} \qquad (4.5)$$

Bước 4. Phân tích mô hình hồi quy để xác định vai trò của các biến giải thích đối với biến phản hồi. Để xác định vai trò của các biến giải thích, trước hết xây dựng hàm hồi quy chuẩn hóa. Biến giải thích nào có hệ số hồi quy chuẩn hóa cao hơn cho biết biến đó đóng vai trò lớn hơn đối với sự thay đổi của biến phản hồi. Mức độ đóng góp của mỗi biến giải thích được xác định bằng đơn vị phần mười hoặc đơn vị phần trăm. Chẳng hạn: Hệ số của biến X_1 = +1,2, còn hệ số của biến X_2 = -1,05. Điều đó có nghĩa là khi biến X_1 tăng lên 1 đơn vị, thì biến phản hồi Y tăng lên 1,2 lần hay 20%. Trái lại, khi biến X_2 tăng lên 1 đơn vị, thì biến phản hồi Y giảm 1,05 lần hay giảm 5%.

Bước 5. Ước lượng và dự đoán sự thay đổi của biến phản hồi. Ước lượng là xác định các giá trị của biến phản hồi (Y) từ các biến giải thích (X_i); trong đó cả hai biến này đều nằm trong phạm vi đã thu thập được. Dự đoán là xác định các giá trị Y từ các biến X_i; trong đó cả hai biến này đều nằm ngoài phạm vi đã thu thập được. Độ chính xác của ước lượng và dự đoán không chỉ phụ thuộc vào dạng mô hình được chọn, mà còn vào phương pháp xây dựng mô hình hồi quy.

Phân tích mối quan hệ giữa nhiều loài cây gỗ bằng phương pháp hồi quy và tương quan đòi hỏi các biến được định lượng bằng các số đo rõ ràng. Để nhận được kết quả với độ tin cậy cao, nhà lâm học cần phải sử dụng dung lượng mẫu quan sát lớn. Đây là một khó khăn không nhỏ, bởi vì dung lượng mẫu quan sát lớn kéo theo những tốn kém về thời gian, nhân lực và kinh phí. Những khó khăn này sẽ tăng lên khi nhà lâm học định lượng sinh khối của các loài cây gỗ với kích thước lớn. Vì thế, phương pháp này ít được áp dụng để phân tích sự kết nhóm giữa các loài cây gỗ trong QXTV rừng.

4.2.2.2. Phương pháp phân tích tương quan không có thứ bậc

Phương pháp này được áp dụng khi độ phong phú của các loài cây gỗ chỉ được định lượng bằng độ thường gặp (1 = bắt gặp, 0 = không bắt gặp). Khuynh hướng và cường độ của mối quan hệ giữa các loài cây gỗ được xác định bằng phân tích tương quan không có thứ bậc. Nếu hai hay nhiều loài cây gỗ có quan hệ chặt chẽ với nhau, thì xác suất bắt gặp chúng ở cùng một nơi ở sẽ nhận giá trị lớn. Phương pháp này được thực hiện theo 2 bước. Bước 1 là kiểm định tính độc lập giữa các loài cây gỗ bằng thống kê χ^2. Bước 2 là phân tích khuynh hướng và cường độ kết nhóm giữa các loài cây gỗ bằng các hệ số kết nhóm.

(a) Kiểm định tính độc lập giữa các loài cây gỗ. Tính độc lập giữa các loài cây gỗ trong QXTV được kiểm định bằng cách lập Bảng chéo R*C (R = số hàng, C = số cột) và kiểm định tính độc lập bằng thống kê χ^2. Khi kiểm định mối quan hệ giữa hai loài cây gỗ, thì lập Bảng chéo 2*2 (Bảng 4.1); trong đó (1) = bắt gặp loài A và B; (0) không bắt gặp loài A và B; a = số ô mẫu cùng xuất hiện loài A và loài B; b = số ô mẫu chỉ xuất hiện loài B; c = số ô mẫu chỉ xuất hiện loài A; d = số ô mẫu không bắt gặp cả hai loài A và B; (a + b) = tổng số ô mẫu bắt gặp loài B; (a + c) = tổng số ô mẫu bắt gặp loài A; N = tổng số ô mẫu nghiên cứu.

Bảng 4.1. Bảng phân tích mối quan hệ giữa hai loài cây gỗ.

		Loài cây A		Tổng theo cột
	Bắt gặp	(1)	(0)	
Loài cây B	(1)	a (a’)	b (b’)	(a + b)
	(0)	c (c’)	d (d’)	(c + d)
Tổng theo hàng		(a + c)	(b + d)	N

Bốn tần số ô mẫu ước lượng (a’, b’, c’ và d’) được xác định theo công thức 4.6 – 4.9. Nói chung, các tần số ước lượng trong mỗi ô của Bảng 2*2 được xác định

bằng cách nhân tổng số ô mẫu ở cuối cột với tổng số ô mẫu ở cuối hàng và chia cho tổng số ô mẫu nghiên cứu.

$$a' = (a + b)*(a + c)/N \quad (4.6)$$

$$b' = (a + b) - a' = (b + d)*(a + b)/N \quad (4.7)$$

$$c' = (a + c) - a' = (a + c)*(c + d)/N \quad (4.8)$$

$$d' = (b + d) - b' = (b + d)*(c + d)/N \quad (4.9)$$

Sau khi xác định các tần số ô mẫu ước lượng từ Bảng 2*2, tính độc lập của hai loài cây gỗ được kiểm định bằng thống kê χ^2 (Công thức 4.10).

$$\chi^2 = \frac{(ad - bc)^2 N}{(a + b)(c + d)(a+c)(b + d)} \quad (4.10)$$

Kiểm định χ^2 được tính theo công thức 4.10 chỉ trong trường hợp tần số ô mẫu trong Bảng 2*2 đều lớn hơn 5 hay 75% số ô có tần số lớn hơn 5. Khi điều kiện này không thỏa mãn, thì kiểm định χ^2 được hiệu chỉnh theo kiểm định χ^2 của Yates (Công thức 4.11); trong đó $\sum R_i$ = Tổng tần số ở cuối hàng, $\sum C$ = Tổng tần số ở cuối cột, N = Tổng số quan sát.

$$\chi^2_{\text{Yates}} = \sum_{(i = 1, n)} * \frac{(|F_{\text{Quan sát}} - F_{\text{Ước lượng}}| - 0.5)^2}{F_{\text{Ước lượng}}} \quad (4.11)$$

$$F_{\text{Ước lượng}} = \frac{\sum R_i * \sum C_i}{N}$$

Giả thuyết H_0+: Hai loài cây gỗ không có quan hệ với nhau.

Đối thuyết H_{0-}: Hai loài cây gỗ có quan hệ với nhau.

Quy tắc quyết định. Nếu $\chi^2 > \chi^2_{0.05}$ hoặc P < 0,05, thì hai loài cây gỗ có quan hệ với nhau. Ngược lại, nếu $\chi^2 < \chi^2_{(0.05)}$ hoặc P > 0,05, thì hai loài cây gỗ không có quan hệ với nhau

(b) Xác định cường độ kết nhóm giữa các loài cây gỗ. Khi hai loài cây gỗ có kết nhóm với nhau, thì bước tiếp theo là xác định cường độ kết nhóm và khuynh hướng kết nhóm giữa chúng. Hai đặc tính này được đo bằng các hệ số kết nhóm. Cường độ kết nhóm giữa các loài cây gỗ được đánh giá theo độ lớn của hệ số kết nhóm. Nếu hệ số kết nhóm tồn tại và mang giá trị dương, thì hai loài cây gỗ có quan hệ hỗ trợ lẫn nhau. Trái lại, nếu hệ số kết nhóm tồn tại và mang giá trị âm, thì hai loài cây gỗ cạnh tranh với nhau về nguồn sống trong cùng nơi ở.

Dưới đây giới thiệu một số phương pháp xác định cường độ và khuynh hướng kết nhóm giữa hai loài cây gỗ.

(b_1) Hệ số Forbes. Theo Forbes (1925; Dẫn theo Vasilevich, 1969), hệ số đo cường độ của mối quan hệ giữa hai loài cây gỗ được tính theo công thức 4.12. Trong trường hợp đo cường độ kết nhóm giữa hai loài cây gỗ, hệ số F cho biết tỷ lệ số ô mẫu cùng bắt gặp cả hai loài so với trị số kỳ vọng. Nếu hai loài cây gỗ phụ thuộc hoàn toàn vào nhau (a = N), thì F = 1. Khi hai loài cây gỗ có mối quan hệ

nghịch (a = 0), thì F = 0. Nhược điểm của hệ số F là không cho biết chiều hướng của mối quan hệ.

$$F = \frac{aN}{(a + b)(a + c)} = \frac{a}{a'} \qquad (4.12)$$

(b_2) Hệ số Pearson. Hệ số Pearson (Dẫn theo Vasilevich, 1969) được tính theo công thức 4.13. Thiếu sót cơ bản của hệ số này là không cho biết khuynh hướng của mối quan hệ giữa các loài là dương hay âm. Để hệ số R nhận giá trị bằng ± 1, thì (a + b) = (a + c). Điều này chỉ xảy ra khi độ thường gặp của hai loài cây gỗ là bằng nhau, nghĩa là chúng phân bố đều trong QXTV.

$$R = \frac{ad - bc}{\sqrt{(a + b)(a + c)(b + d)(c + d)}} \qquad (4.13)$$

(b_3) Hệ số Yule. Yule (1920; Dẫn theo Vasilevich, 1969) đề xuất hệ số đo cường độ của mối quan hệ giữa hai loài cây gỗ theo công thức 4.14. Hệ số Q nhận giá trị từ -1 đến +1.

$$Q = \frac{ad - bc}{ad + bc} \qquad (4.14)$$

(b_4) Hệ số Cole. Theo Cole (1949; Dẫn theo Vasilevich, 1969), mối quan hệ giữa hai loài cây gỗ được xác định theo công thức 4.15. Công thức (4.15) chỉ được sử dụng trong trường hợp mối quan hệ giữa hai loài cây gỗ là dương, nghĩa là khi ad > bc. Ngược lại, khi hai loài cây gỗ có mối quan hệ âm, thì hệ số C được xác định theo công thức 4.16. Khi a > d thì hệ số C được xác định theo công thức 4.17.

$$C = \frac{(ad - bc)}{(a + b)(b + d)} \qquad (4.15)$$

$$C = \frac{(ab - bc)}{(a + b)(a + c)} \qquad (4.16)$$

$$C = \frac{(ad - bc)}{(b + d)(c + d)} \qquad (4.17)$$

Bảng 4.2 dẫn ví dụ phân tích mối quan hệ giữa hai loài cây *Luzula multriflora* và *Leucandenum vulgare*. Từ số liệu ở Bảng 4.2, chúng ta có:

Hệ số Fobes: F = (66*500)/(120*180) = 1,54.

Hệ số Pearson: R = (66 * 266 - 54*114)/$\sqrt{(120*180*320*380)}$ = + 0,22.

Hệ số Yule: Q = (66*266 - 54*114)/(66*266 + 54*114) = + 0,48.

Hệ số Cole: C = (66*266 - 54*114)/(120*320) = + 0,30.

(b_5) Hệ số Lambda. Hệ số Lambda (λ) được xác định theo công thức 4.18. Ở công thức 4.18, sai số trước là những sai lầm phạm phải khi dự đoán các trị số của biến phụ thuộc mà không kể đến biến độc lập. Sai số sau là những sai lầm phạm phải khi dự đoán các trị số của biến phụ thuộc mà có kể đến biến độc lập.

Bảng 4.2. Mối quan hệ giữa *Luzula multriflora* và *Leucandenum vulgare*. Nguồn: Vasilevich, 1969.

Loài cây		*L. multiflora*		Tổng theo cột
	Bắt gặp	(1)	(0)	
L. vulgare	(1)	66 (43)	54 (77)	120
	(0)	114 (137)	266 (243)	380
Tổng theo hàng		180	320	500

$$\lambda = \frac{\text{(Sai số trước - Sai số sau)}}{\text{Sai số trước}} \qquad (4.18)$$

(b_6) Hệ số Phi và Cramer'V. Hệ số Phi và Cramer (V) được xác định theo công thức 4.19; trong đó N là tổng số ô mẫu nghiên cứu, K là số hàng hoặc số cột nhỏ nhất của bảng R*C. Giá trị χ^2 được xác định theo công thức 4.10.

$$V = \sqrt{\frac{\chi^2}{N(K - 1)}} \qquad (4.19)$$

(b_7) Hệ số ngẫu nhiên. Hệ số ngẫu nhiên (C) được xác định theo công thức 4.20; trong đó N là tổng số ô mẫu nghiên cứu. Giá trị χ^2 được xác định theo công thức 4.10.

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}} \qquad (4.20)$$

Ba hệ số λ, V và C đều nhận các giá trị từ 0 ÷ 1. Giá trị 0 cho biết biến phụ thuộc không có quan hệ với biến độc lập. Trong trường hợp này, biến phụ thuộc không thể dự đoán bằng biến độc lập. Ngược lại, ba hệ số này tồn tại và càng gần bằng ±1 cho biết hai biến có quan hệ chặt chẽ với nhau. Vì thế, biến phụ thuộc có thể được dự đoán bằng biến độc lập. Khi áp dụng vào phân tích sự kết nhóm giữa các loài cây gỗ, chúng ta nói "Độ bắt gặp của loài cây gỗ này có thể được dự đoán bằng độ bắt gặp của loài cây gỗ khác". Mức độ giảm tỷ lệ sai số khi sử dụng biến này để dự đoán biến kia được đánh giá theo số đo tính đối xứng (Symmetric Measures).

Ba hệ số λ, V và C này cung cấp 2 nhóm số đo khác nhau. Nhóm một là số đo xu hướng (Directional Measures) của mối quan hệ giữa các biến. Nhóm hai là số đo tính đối xứng (Symmetric Measures) hay sự tương đồng của các biến. Những hệ số đo xu hướng cho biết mức độ giảm tỷ lệ sai số khi sử dụng biến này để dự đoán biến kia. Nhóm này bao gồm hệ số Lambda, hệ số Goodman và Kruskal's tau và hệ số không chắc chắn của mối quan hệ (Uncertainty Coefficient). Nếu cả ba hệ số này đều cho kết quả như nhau, thì kết luận nhận được càng chính xác. Những số đo tính đối xứng không chỉ cho biết mức độ quan hệ giữa các biến, mà còn cả mức giảm tỷ lệ trong sai số dự đoán. Nhóm này bao gồm hệ số Phi, hệ số ngẫu nhiên

(Contingency Coefficient) và hệ số Cramer's V. Những hệ số này có thể tìm thấy trong những phần mềm thống kê SPSS và STATGRAPHICS.

4.2.2.3. Phương pháp phân tích mối quan hệ giữa nhiều loài cây gỗ

Rừng tự nhiên hỗn loài nhiệt đới bao gồm nhiều loài cây gỗ khác nhau. Một loài cây gỗ nào đó không chỉ có quan hệ với một loài cây gỗ khác, mà còn với rất nhiều loài cây gỗ khác. Vì thế, phân tích mối quan hệ giữa hai loài cây gỗ với nhau chưa thể làm rõ tính phức tạp của mối quan hệ giữa các loài cây gỗ trong QXTV. Vấn đề này chỉ được làm rõ bằng phương pháp phân tích mối quan hệ giữa nhiều loài cây gỗ trong QXTV.

Phương pháp hồi quy và tương quan riêng phần trong hồi quy đa biến là công cụ hữu ích để phân tích mối quan hệ giữa các loài cây gỗ trong QXTV. Bản chất của phương pháp này là tính hệ số tương quan khi các giá trị của một hoặc một số biến độc lập được cố định. Điều đó cho phép loại trừ ảnh hưởng của một số nhân tố đến quan hệ của các đại lượng nghiên cứu.

Giả sử nghiên cứu trương quan giữa ba đại lượng. Trong trường hợp này, hệ số tương quan riêng phần được xác định theo công thức 4.21. Ở công thức 4.21, $r_{12(3)}$ là hệ số tương quan giữa đại lượng 1 và 2 khi cố định đại lượng 3; r_{12}, r_{13} và r_{23} tương ứng là hệ số tương quan cặp giữa đại lượng 1 và 2, 1 và 3, 2 và 3.

$$r_{12(3)} = \frac{(r_{12} - r_{13}\, r_{23})}{\sqrt{(1 - r^2_{13})(1 - r^2_{23})}} \qquad (4.21)$$

Phương pháp phân tích tương quan riêng phần không chỉ cho phép xác định tương quan cặp giữa các loài cây gỗ, mà còn cả mức ảnh hưởng lẫn nhau giữa các loài cây gỗ trong các các nhóm sinh thái. Về mặt sinh thái học và lâm học, nếu các hệ số tương quan riêng phần khác nhau không rõ rệt, thì sự kết nhóm này không phải là do quan hệ qua lại giữa các loài cây gỗ. Sự cùng có mặt của các loài cây gỗ trong QXTV này là do chúng có sự tương đồng về biên độ sinh thái. Đó là phương pháp luận để phân tích kết cấu loài cây gỗ và cấu trúc của QXTV rừng.

Theo Cole (1957; Dẫn theo Vasilevich, 1969), khi phân tích sự kết nhóm của nhiều loài cây gỗ bằng tương quan riêng phần, trình tự nghiên cứu bao gồm một số bước sau đây.

Bước 1. Phân tích mối quan hệ giữa từng cặp loài cây gỗ trên tất cả các ô mẫu thu thập được. Hệ số kết nhóm này được gọi là hệ số kết nhóm đầy đủ. Điều này có nghĩa là hệ số kết nhóm giữa hai loài cây gỗ trên tất cả các ô mẫu bắt gặp hay không bắt gặp các loài cây gỗ khác.

Bước 2. Phân tích ảnh hưởng của loài cây gỗ thứ ba đến mối quan hệ giữa loài cây gỗ 1 và loài cây gỗ 2. Để làm rõ quan hệ này, các ô mẫu được phân chia thành hai phần; trong đó phần thứ nhất là những ô mẫu bắt gặp loài cây cây gỗ thứ 3, còn phần thứ 2 là những ô mẫu không bắt gặp loài cây gỗ thứ 3. Sau đó thực hiện 2 phân tích riêng biệt: (a) Xác định hệ số kết nhóm giữa hai loài 1 và 2 trên những ô mẫu bắt gặp loài thứ 3 (Kí hiệu C_1); (b) Xác định hệ số kết nhóm giữa hai loài 1 và

2 trên những ô mẫu không bắt gặp loài thứ 3 (Kí hiệu C_2). Các hệ số kết nhóm tính theo cách phân tích này được gọi là hệ số kết nhóm riêng phần. So sánh hai hệ số kết nhóm riêng phần C_1 và C_2 không chỉ cho phép đánh giá quan hệ giữa hai loài 1 và 2, mà còn xác định ảnh hưởng của loài thứ 3 đến sự phân bố của hai loài 1 và 2. Sự khác nhau giữa hai hệ số kết nhóm riêng phần C_1 và C_2 được kiểm định bằng thống kê T (Công thức 4.22); trong đó δ_{C1} (Công thức 4.23) là sai số của hệ số kết nhóm trên những ô mẫu bắt gặp loài thứ 3, δ_{C2} (Công thức 4.24) là sai số của hệ số kết nhóm trên những ô mẫu không bắt gặp loài thứ 3. Ở công thức 4.23 và 4.24, N_1 là tổng số ô mẫu bắt gặp loài thứ 3; N_2 là tổng số ô mẫu không bắt gặp loài thứ 3; a_1, b_1, c_1 và d_1 là tần số ô mẫu trong bảng chéo dùng để tính quan hệ giữa loài 1 và loài 2 trên những ô bắt gặp loài 3; a_2, b_2, c_2 và d_2 là tần số ô mẫu trong bảng chéo dùng để tính quan hệ giữa loài 1 và loài 2 trên những ô mẫu vắng mặt loài 3.

$$T = \frac{C_2 - C_1}{\sqrt{\delta_{C1} + \delta_{C2}}} \qquad (4.22)$$

$$\delta_{C1} = \sqrt{\frac{(a_1+b_1)*(b_1+d_1)}{(N_1*(a_1+c_1)*(c_1+d_1))}} \qquad (4.23)$$

$$\delta_{C2} = \sqrt{\frac{(a_2+b_2)*(b_2+d_2)}{(N_2*(a_2+c_2)*(c_2+d_2))}} \qquad (4.24)$$

Nếu $T < T_{(0,05)}$ thì sự khác biệt giữa hai hệ số kết nhóm riêng phần là không có ý nghĩa. Điều này có nghĩa là sự xuất hiện của loài thứ 3 ảnh hưởng không đáng kể đến sự phân bố của loài 1 và loài 2. Ngược lại, nếu $T > T_{(0.05)}$ thì sự xuất hiện của loài thứ 3 ảnh hưởng đáng kể đến sự phân bố của loài 1 và loài 2.

Trong trường hợp phân tích sự kết nhóm của 4 loài cây gỗ trở lên, cách thức phân tích cũng được thực hiện tương tự như phân tích quan hệ giữa 3 loài cây gỗ. Tuy vậy, chúng ta cần phải tính toán qua rất nhiều bước khác nhau. Giả thiết nghiên cứu sự kết nhóm giữa 4 loài cây gỗ (Sao đen, Dầu con rái, Cầy, Cám). Để xác định các nhóm sinh thái, trình tự tính toán bao gồm 4 bước.

Bước 1. Phân tích mối quan hệ giữa từng cặp loài: (1) Sao đen và Dầu con rái; (2) Sao đen và Cầy; (3) Sao đen và Cám; (4) Dầu con rái và Cầy; (5) Dầu con rái và Cám; (6) Cầy và Cám.

Bước 2. Phân tích quan hệ giữa một loài cây gỗ với 2 loài cây gỗ khác: (1) = Sao đen – (Dầu con rái + Cầy); (2) Sao đen – (Dầu con rái + Cám); (3) Sao đen – (Cầy + Cám); (3) Dầu con rái – (Sao đen + Cầy); (4) Dầu con rái – (Sao đen + Cám); (5) Dầu con rái – (Cầy và Cám); (7) Cầy – (Sao đen + Dầu con rái); (8) Cầy – (Sao đen + Cám); (9) Cầy – (Dầu con rái + Cám); (10) Cám – (Sao đen + Dầu con rái); (11) Cám – (Sao đen + Cầy); (12) Cám – (Dầu con rái + Cầy).

Bước 3. Phân tích quan hệ giữa một loài cây gỗ với 3 loài cây gỗ khác: (1) Sao đen - (Dầu con rái + Cầy + Cám); (2) Dầu con rái – (Sao đen + Cầy + Cám); (3) Cầy – (Sao đen + Dầu con rái + Cám); (4) Cám – (Sao đen + Dầu con rái + Cám).

Phân tích mối quan hệ giữa từng cặp loài cây gỗ được gọi là phân tích đơn biến. Phân tích mối quan hệ giữa một loài cây gỗ với 2 hay nhiều loài cây gỗ khác được gọi là phân tích đa biến. Khi xác định quan hệ giữa 1 loài cây gỗ với tổ hợp 2 loài cây gỗ khác, thì những ô mẫu bắt gặp cả hai loài cây gỗ được mã hóa bằng 1. Trái lại, những ô mẫu chỉ bắt gặp 1 loài cây gỗ hoặc không bắt gặp cả 2 loài cây gỗ được mã hóa bằng 0. Cách mã hóa này cũng được áp dụng cho trường hợp nghiên cứu quan hệ đa biến giữa 1 loài cây gỗ với tổ hợp 3 hay nhiều loài cây gỗ khác. Trong phân tích số liệu, các mối quan hệ trên đây cũng được lập thành Bảng chéo 2*2. Sau đó xác định các mối quan hệ bằng kiểm định χ^2.

Bước 4. Phân tích mối quan hệ của các loài cây gỗ trong các nhóm sinh thái. Vấn đề này được làm rõ thông qua phân tích xu hướng và cường độ kết nhóm (dấu và độ lớn của hệ số kết nhóm) và tỷ lệ giảm sai số trong việc sử dụng loài này để dự đoán mức độ bắt gặp loài khác. Mối quan hệ của các loài cây gỗ trong các nhóm sinh thái cũng được xác định theo ba hệ số λ (Công thức 4.18), hệ số V (Công thức 4.19) và hệ số C (Công thức 4.20).

Bước 5. Xác định các nhóm sinh thái. Nhóm sinh thái bao gồm những loài có mối quan hệ dương với nhau. Trong trường hợp ở ví dụ này, nhóm sinh thái có thể bao gồm 2, 3 và 4 loài. Nói chung, số lượng nhóm sinh thái ít hay nhiều phụ thuộc vào mục tiêu nghiên cứu. Những nhóm sinh thái được chọn là những nhóm cây gỗ có hệ số kết nhóm cao hoặc nhóm chứa những loài cây gỗ ưu thế và đồng ưu thế.

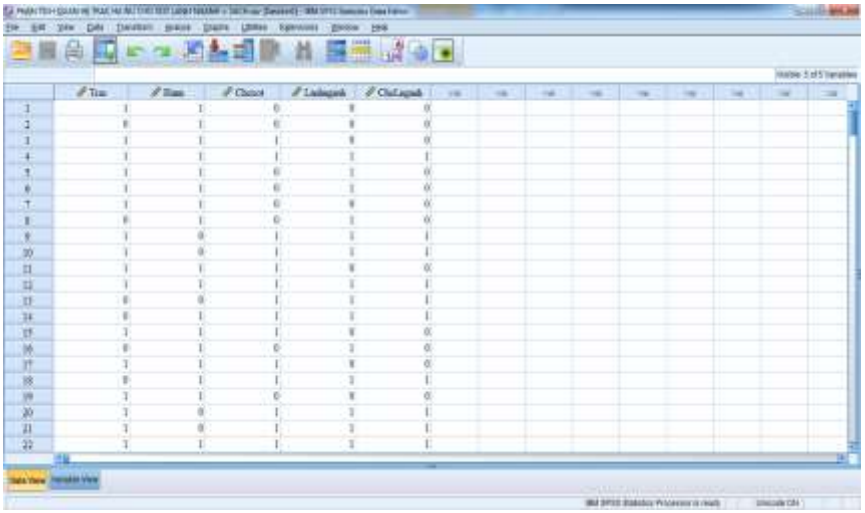
4.2.2.4. Phân tích sự kết nhóm giữa các loài cây gỗ bằng phần mềm SPSS

Giả sử phân tích sự kết nhóm sinh thái giữa 4 loài cây gỗ: Trắc (*Dalbergia cochinchinensis*), Hà nu (*Ixonathes reticulata* Jack), Chò xót (*Schima superba* Gard.& Champ) và Lành ngạnh (*Cratoxylum maingayi* Dyers in Hook.f). Độ bắt gặp của 4 loài cây gỗ này trong QXTV được xác định từ 250 ô mẫu với kích thước 200 m^2. Các ô mẫu được bố trí trên các tuyến song song; mỗi tuyến cách nhau 300 m. Trong phần xử lý số liệu, xác định: (1) Mối quan hệ giữa từng cặp loài cây gỗ; (2) Ảnh hưởng của Chò xót và Lành ngạnh đến mối quan hệ giữa Trắc và Hà nu.

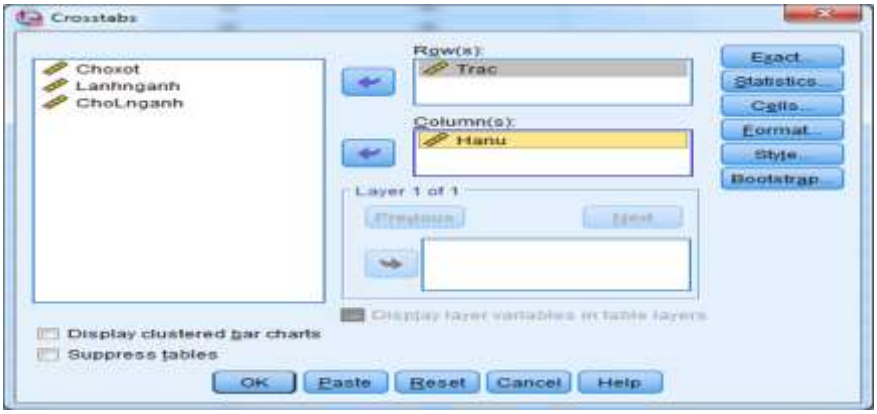
Nội dung 1. Phân tích mối quan hệ giữa từng cặp loài cây gỗ. Các cặp loài cây gỗ cần phân tích là Trắc – Hà nu; Trắc – Chò xót; Trắc – Lành ngạnh; Hà nu + Chò xót; Hà nu – Lành ngạnh. Giả thuyết H_0^+: Hai loài cây gỗ không kết nhóm với nhau. Đối thuyết H_0^-: Hai loài cây gỗ kết nhóm với nhau. Kiểm định này được thực hiện qua ba bước.

Bước 1. Trước hết mở SPSS và ghi tên 4 loài cây gỗ và độ bắt gặp của chúng vào 5 cột liên tiếp nhau (Hình 4.1).

Bước 2. Xây dựng Bảng chéo 2*2 giữa Trắc và Hà nu theo đường dẫn: Analyze > Descriptive Statistics > Crosstabs. Đến đây một cửa sổ xuất hiện. Tại cửa sổ này, chọn Trắc và Hà nu và chuyển chúng vào hai ô Row(s) và Column(s) (Hình 4.2). Đối với những cặp loài khác, cách giải quyết cũng được thực hiện tương tự.



Hình 4.1. Các loài cây gỗ và độ bắt gặp của chúng.



Hình 4.2. Xác định sự kết nhóm giữa các loài cây gỗ.

Bước 3. Xác định mối quan hệ, xu hướng và cường độ của mối quan hệ giữa Trắc và Hà nu. Để đạt được mục đích này, trước hết chọn Statistics ở Hình 4.2. Đến đây một cửa sổ Crosstabs xuất hiện. Kế đến tại cửa sổ này chọn kiểm định Chi – square (χ^2) và 4 thống kê định danh ở ô Nominal (Contingency Coefficient; Phi and Cramer's V; Lambda; Uncertainty Coefficient). Bốn thống kê này tính toán xu hướng và cường độ của mối quan hệ giữa các loài (Hình 4.3). Lưu ý rằng chọn cả 4 thống kê này để đảm bảo độ tin cậy của kết quả nhận được.

Bước 4. Xác định kết quả kiểm định tính độc lập của hai loài Trắc và Hà nu bằng kiểm định χ^2. Để đạt mục đích này, trước hết chọn Continue ở Hình 4.3; sau đó chọn OK ở Hình 4.2. Đến đây chương trình báo cáo 4 kết quả (Bảng 4.3). Bảng 4.3a báo cáo tóm tắt xử lý các trường hợp (Case Processing Summary). Nếu số liệu đưa vào bảng SPSS là đúng (Valid), thì kết quả nhận được N = 250 hay 100%. Khi N khác 100%, thì kiểm tra lại sai sót đối với số liệu đầu vào.

Hình 4.3. Kiểm định mối quan hệ giữa các loài cây gỗ.

Bảng 4.3. Kết quả kiểm định mối quan hệ giữa Trắc và Hà nu.

4.3a. Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Trac * Hanu	250	100.0%	0	0.0%	250	100.0%

4.3b. Trac * Hanu Crosstabulation

		Hà nu		Total
		0	1	
Trắc	0	26	34	60
	1	28	162	190
Total		98	54	250

4.3c. Chi - Square Tests

	Value	df	Significance	Sig.	Sig.(1-sided)
Pearson Chi-Square	22.020[a]	1	.000		
Continuity Correction	20.364	1	.000		
Likelihood Ratio	19.907	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear	21.932	1	.000		
N of Valid Cases	250				

4.3d. Directional Measures

			Value	Error[a]	Appro. T[d]	Significance
Nominal by Nominal	Lambda	Symmetric	.000	.000		
		Trắc Dependent	.000	.000		
		Hà nu Dependent	.000	.000		
	Goodman and Kruskal	Trắc Dependent	.088	.041		.000[c]
		Hà nu Dependent	.088	.041		.000[c]
	Uncertainty	Symmetric	.074	.034	2.171	.000[e]

Coefficient	Trắc Dependent	.072	.033	2.171	.000[e]
	Hà nu Dependent	.076	.035	2.171	.000[e]

4.3e. Symmetric Measures

		Value	Significance
Nominal by Nominal	Phi	0.297	0.000
	Cramer's V	0.297	0.000
	Contingency Coefficient	0.285	0.000
N of Valid Cases		250	

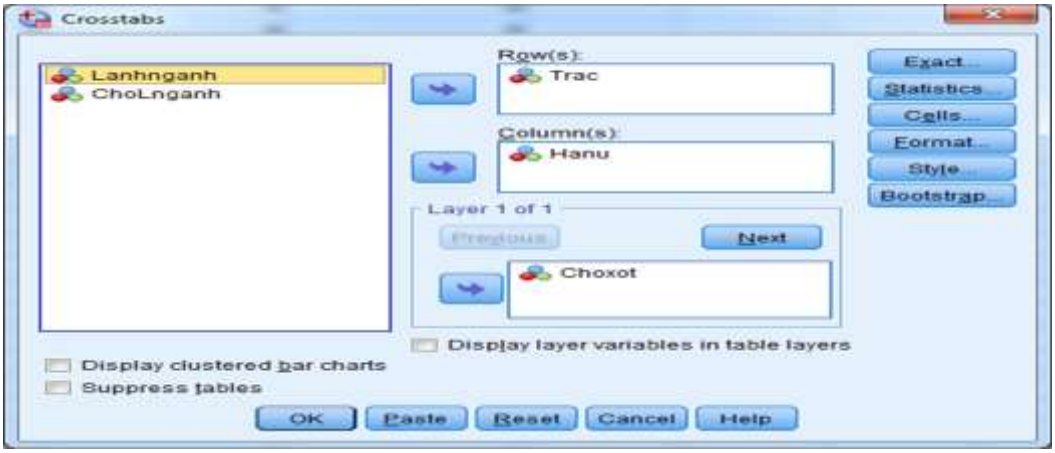
Bảng 4.3b báo cáo tần số ô mẫu bắt gặp (1) và không bắt gặp (0) Trắc và Hà nu. Bảng 4.3c báo cáo 5 kiểm định χ^2 (Pearson; Continuity Correction; Likelihood Ratio; Fisher's Exact; Linear-by-Linear). Các kiểm định này đều báo cáo giá trị (Value) và mức ý nghĩa thống kê (Significance). Nếu các tần số trong 4 ô của bảng 2*2 nhỏ hơn 5, thì mức ý nghĩa của kiểm định được xác định theo kiểm định Fisher. Bởi vì P (Significance) < 0, 01, nên Trắc và Hà nu có quan hệ với nhau. Bảng 4.3d báo cáo kết quả phân tích xu hướng của mối quan hệ giữa Trắc và Hà nu bằng ba hệ số Lambda, Goodman và Kruskal tau và hệ số không chắc chắn của mối quan hệ (Uncertainty Coefficient). Thực chất của Bảng 4.3d cho biết mức giảm sai số khi sử dụng độ bắt gặp của Hà nu để dự đoán độ bắt gặp của Trắc và ngược lại. Tiêu chuẩn Lambda = 0 cho biết nếu sử dụng độ bắt gặp của Trắc để dự đoán độ bắt gặp của Hà nu, thì mức giảm sai số dự đoán là 0. Trái lại, theo tiêu chuẩn Goodman và Kruskal và Uncertainty Coefficient, mức giảm sai số dự đoán độ bắt gặp của Trắc từ độ bắt gặp của Hà nu tương ứng là 8,8% và 7,2%.

Bảng 4.3e báo cáo kết quả phân tích cường độ của mối quan hệ giữa Trắc và Hà nu bằng ba hệ số (Phi, Cramer's V và Contingency Coefficient). Từ ba hệ số này cho thấy, Trắc và Hà nu có quan hệ với nhau (P < 0,01) và cường độ quan hệ khá chặt chẽ (Phi = 0,297, Cramer's V = 0,297 và Contingency Coefficient = 0,285). Ba hệ số này đều mang giá trị dương chứng tỏ hai loài Trắc và Hà nu hỗ trợ lẫn nhau trong quá trình sống.

Nội dung 2. Phân tích ảnh hưởng của Chò xót đến mối quan hệ giữa Trắc và Hà nu. Những phân tích ở nội dung 1 cho thấy Trắc và Chò xót có quan hệ với nhau. Tuy vậy, QXTV này không chỉ có hai loài Trắc và Hà nu, mà còn nhiều loài cây gỗ khác. Vì thế, những phân tích ở bước này nhằm trả lời câu hỏi: Trắc có quan hệ thực sự với Hà nu hay là nó có quan hệ với Chò xót? Để làm rõ câu hỏi này, các ô mẫu được phân chia thành 2 nhóm. Nhóm 1 bao gồm những ô mẫu bắt gặp Trắc và Hà nu nhưng không bắt gặp Chò xót. Nhóm 2 bao gồm những ô mẫu bắt gặp cả ba loài Trắc, Hà nu và cả Chò xót. Ở mỗi nhóm, phân tích riêng rẽ quan hệ giữa Trắc và Hà nu. Sau đó so sánh các hệ số kết nhóm riêng phần giữa hai nhóm để làm rõ vai trò của Chò xót trong mối quan hệ giữa Trắc và Hà nu. Khi hai hệ số kết nhóm trong hai nhóm này đều tồn tại, thì Trắc và Hà nu quan hệ thực sự với nhau. Nói cách khác, sự có mặt của Chò xót trong QXTV không ảnh hưởng đến mối quan

hệ giữa Trắc và Hà nu. Trái lại, nếu hệ số kết nhóm giữa Trắc và Hà nu chỉ tồn ở nhóm có mặt Chò xót, thì Chò xót ảnh hưởng đến mối quan hệ giữa Trắc và Hà Nu. Tất cả những phân tích trên đây được xử lý theo 3 bước dưới đây.

Bước 1. Xây dựng các Bảng chéo 2*2. Để đạt được mục đích này, tại cửa sổ ở Hình 4.2, chọn Chò xót và chuyển vào cửa sổ Layer 1 of 1 (Hình 4.4).



Hình 4.4. Xây dựng bảng chéo để phân tích quan hệ giữa ba loài cây gỗ.

Bước 2. Kiểm định ảnh hưởng của Chò xót đến mối quan hệ giữa Trắc và Hà nu. Giả thuyết H_0^+: Sự có mặt của Chò xót trong QXTV không ảnh hưởng đến mối quan hệ giữa Trắc và Hà nu. Giả thuyết này bị bác bỏ khi mức ý nghĩa (P) ở những ô mẫu bắt gặp Chò xót nhỏ hơn 0,05. Để kiểm định giả thuyết này, trước hết chọn Statistics ở Hình 4.4. Đến đây một cửa sổ Crosstabs xuất hiện. Kế đến tại cửa sổ này chọn kiểm định Chi – square (χ^2) và 5 thống kê (Pearson; Continuity Correction; Likelihood Ratio; Fisher's Exact; Linear-by-Linear) biểu thị xu hướng và cường độ của mối quan hệ giữa các loài giống như ở Hình 4.3. Sau đó chọn OK để nhận kết quả phân tích ảnh hưởng của Chò xót đến mối quan hệ giữa Trắc và Hà nu. Kết quả báo cáo 4 nội dung: (1) Tóm tắt xử lý các trường hợp (Case Processing Summary); (2) Bảng chéo Trắc * Hà nu * Chò xót; (3) Các số đo xu hướng của mối quan hệ (Directional Measures); (4) Các số đo tính đối xứng hay cường độ của mối quan hệ (Symmetic Measures). Để làm rõ câu hỏi "Trắc có quan hệ thực sự với Hà nu hay là nó quan hệ với Chò xót?", chúng ta phân tích mối quan hệ giữa Trắc và Hà nu trong hai nhóm ô mẫu bắt gặp (1) và không bắt gặp (0) Chò xót (Bảng 4.4).

Ở nhóm ô mẫu không bắt gặp (0) Chò xót, cả ba hệ số Phi, Cramer's V và Contingency Coefficient đều nhận giá trị bằng 0,218 với P = 0,028. Điều đó cho biết Trắc và Hà nu có quan hệ với nhau. Ở nhóm ô mẫu bắt gặp (1) Chò xót, cả ba hệ số Phi, Cramer's V và Contingency Coefficient đều nhận giá trị bằng 0,305 với P < 0,01. Điều đó cho biết Chò xót ảnh hưởng đến mối quan hệ giữa Trắc và Hà nu.

Bảng 4.4. Kiểm định ảnh hưởng của Chò xót đến mối quan hệ giữa Trắc và Hà nu.

Symmetric Measures

Chò xót			Value	Significance
0	Nominal by Nominal	Phi	.218	.028
		Cramer's V	.218	.028
		Contingency Coefficient	.213	.028
	N of Valid Cases		101	
1	Nominal by Nominal	Phi	.305	.000
		Cramer's V	.305	.000
		Contingency Coefficient	.292	.000
	N of Valid Cases		149	
Total	Nominal by Nominal	Phi	.297	.000
		Cramer's V	.297	.000
		Contingency Coefficient	.285	.000
	N of Valid Cases		250	

Bước 3. Kiểm định ảnh hưởng của Chò xót và Lành ngạnh đến mối quan hệ giữa Trắc và Hà nu. Giả thuyết H_0^+: Sự có mặt của Chò xót và Lành ngạnh trong QXTV không ảnh hưởng đến mối quan hệ giữa Trắc và Hà nu. Giả thuyết này bị bác bỏ khi mức ý nghĩa (P) ở những ô mẫu bắt gặp Chò xót và Lành ngạnh nhỏ hơn 0,05. Để kiểm định giả thuyết này, trước hết xây dựng Bảng chéo 2*2 (Hình 4.5).

Hình 4.5. Xây dựng bảng chéo để phân tích ảnh hưởng của Chò xót và Lành ngạnh đến mối quan hệ giữa Trắc và Hà nu.

Lưu ý rằng, những ô mẫu cùng bắt gặp Chò xót và Lành ngạnh được mã hóa = 1. Trái lại, những ô mẫu không bắt gặp cả Chò xót và Lành ngạnh hoặc chỉ bắt gặp 1 trong 2 loài được mã hóa = 0. Sau đó phân tích ảnh hưởng của Chò xót và Lành ngạnh đến mối quan hệ giữa Trắc và Hà nu theo 2 bước trên đây. Kết quả

phân tích (Bảng 4.5) cho thấy, ở nhóm ô mẫu không bắt gặp (0) Chò xót và Lành ngạnh, cả ba hệ số Phi, Cramer's V và Contingency Coefficient đều nhận giá trị bằng 0,258 với P < 0,01. Điều đó cho biết Trắc và Hà nu có quan hệ với nhau. Ở nhóm ô mẫu bắt gặp (1) Chò xót và Lành ngạnh, cả ba hệ số Phi, Cramer's V và Contingency Coefficient đều nhận giá trị bằng 0,297 với P < 0,01. Điều đó cho biết Chò xót và Lành ngạnh ảnh hưởng đến mối quan hệ giữa Trắc và Hà nu.

Bảng 4.5. Kiểm định ảnh hưởng của Chò xót và Lành ngạnh đến mối quan hệ giữa Trắc và Hà nu.

Symmetric Measures

Chò xót + Lành ngạnh			Value	Significance
0	Nominal by Nominal	Phi	.258	.001
		Cramer's V	.258	.001
		Contingency Coefficient	.249	.001
	N of Valid Cases		171	
1		Phi	.346	.002
		Cramer's V	.346	.002
		Contingency Coefficient	.327	.002
	N of Valid Cases		79	

4.2.3. Phương pháp thu mẫu trong phân tích sự kết nhóm giữa các loài cây gỗ

Độ chính xác của phân tích sự kết nhóm giữa các loài cây gỗ phụ thuộc vào phương pháp thu mẫu. Trị số và dấu của hệ số kết nhóm thay đổi tùy thuộc vào kích thước ô mẫu. Khi thay đổi kích thước ô mẫu, thì sự kết nhóm của các loài cây gỗ có thể thay đổi. Nói chung, xác định kích thước ô mẫu thích hợp trong nghiên cứu sự kết nhóm giữa các loài cây gỗ là một vấn đề rất phức tạp. Tính phức tạp biểu hiện ở chỗ, kích thước ô mẫu lớn có thể bao trùm lên nhiều loài, còn kích thước ô mẫu nhỏ chỉ bao trùm lên một số loài hoặc chỉ một loài. Khi kích thước ô mẫu càng lớn, thì độ thường gặp của các loài cây gỗ càng cao. Khi kích thước ô mẫu càng nhỏ, thì độ thường gặp của các loài cây gỗ càng nhỏ. Trong cả hai trường hợp này, kết quả phân tích sự kết nhóm giữa các loài cây gỗ đều đưa đến kết luận kém tin cậy. Kích thước ô mẫu lớn còn làm gia tăng khối lượng đo đạc và tính toán. Vì thế, xác định kích thước ô mẫu thích hợp để nhận được kết quả chính xác và khách quan là một vấn đề cần được quan tâm.

Về lý thuyết, kích thước ô mẫu thích hợp phải phản ánh đúng quan hệ giữa các loài cây gỗ. Hai loài cây gỗ chỉ tồn tại mối quan hệ với nhau khi hệ thống tán hoặc hệ thống rễ của chúng giao nhau cả theo chiều đứng và chiều ngang. Trong thực tế, sự giao tán giữa các loài cây gỗ là dấu hiệu có thể nhận biết bằng mắt hoặc bằng các trắc đồ rừng. Vì thế, diện tích hình chiếu tán nằm ngang của các loài cây gỗ sống gần nhau là cơ sở cho việc chọn kích thước ô mẫu. Kích thước tán cây gỗ thay đổi theo tuổi và môi trường sống. Nói chung, khi đến tuổi trưởng thành, kích

thước tán của các cây gỗ theo chiều nằm ngang thường dao động từ 10 – 15 m. Vì thế, nếu nghiên cứu sự kết nhóm của 4 – 5 loài cây gỗ, kích thước ô mẫu được chọn dao động từ 200 – 400 m^2. Hình dạng ô mẫu có thể là hình tròn, hình vuông hoặc hình chữ nhật. Các ô mẫu có thể được bố trí ngẫu nhiên hoặc đồng đều trên những tuyến song song.

Về mặt toán học, kiểm định χ^2 chỉ chính xác khi các tần số trong các ô của Bảng chéo C*R lớn hơn 5. Rừng tự nhiên hỗn loài nhiệt đới bao gồm rất nhiều loài cây gỗ và chúng phân bố không đồng đều trên mặt đất. Vì thế, để đảm bảo đúng yêu cầu của kiểm định χ^2, nhà lâm học cần phải thu thập rất nhiều ô mẫu. Tuy vậy, số lượng ô mẫu bằng bao nhiêu là một vấn đề rất khó xác định. Lý do là vì chúng ta không biết rõ độ bắt gặp của các loài cây quan tâm trong giai đoạn thiết kế nghiên cứu. Trong thực hành, số lượng ô mẫu cần thu thập gia tăng dần sao cho đáp ứng tốt yêu cầu của kiểm định χ^2. Nói chung, khi nghiên cứu quan hệ giữa 4 loài cây gỗ trong những QXTV thuộc rừng mưa nhiệt đới, thì số lượng ô mẫu cần thu thập không nhỏ hơn 500.

Đối tượng nghiên cứu sự kết nhóm giữa các loài cây gỗ là một vấn đề đáng quan tâm. Trong lâm học, kết quả nghiên cứu sự kết nhóm giữa các loài cây gỗ chỉ có ý nghĩa khi đối tượng nghiên cứu là thảm thực vật ở giai đoạn ổn định hoặc ở những giai đoạn cuối của chuỗi diễn thế sinh thái tiến về cao đỉnh (Climax). Thảm thực vật ổn định là kết quả của quá trình cạnh tranh, đào thải tự nhiên và tiến hoá lâu dài của các loài cây gỗ. Thông qua những quá trình này, các loài cây gỗ đã chọn được nơi ở thích hợp và những loài cùng chung sống bên nhau. Những thảm thực vật đang trong quá trình diễn thế ở giai đoạn đầu và giữa là những thảm thực vật chưa ổn định về thành phần loài cây gỗ và cấu trúc. Chúng không phải là đối tượng tốt của những nghiên cứu về sự kết nhóm giữa các loài cây gỗ. Những thảm thực vật đã phát triển ổn định nhưng hiện tại đang gặp những rối loạn (khai thác, lửa…) cũng không phải là đối tượng tốt của những nghiên cứu về sự kết nhóm giữa các loài cây gỗ.

Bên cạnh mối quan tâm đến đối tượng nghiên cứu và kích thước ô mẫu thích hợp, nhà lâm học cũng cần lưu ý đến việc chọn loài cây gỗ và số lượng loài cây gỗ nghiên cứu. Về mặt khoa học, những nghiên cứu sự kết nhóm của bất kể tập hợp loài cây gỗ nào trong các QXTV rừng đều mang lại những ý nghĩa lớn. Kiến thức này giúp cho nhà lâm học không chỉ hiểu rõ về đời sống của các loài cây gỗ và rừng, mà còn xây dựng những nguyên lý lâm sinh. Về thực tiễn, những nghiên cứu về sự kết nhóm giữa các loài cây gỗ đều hướng đến những yêu cầu của kinh doanh rừng. Mặt khác, xác định quan hệ giữa nhiều loài cây gỗ cũng đòi hỏi nhiều thời gian, nhân lực, kinh phí và tính toán rất phức tạp. Vì thế, khi phân phân tích sự kết nhóm giữa các loài cây gỗ ở rừng tự nhiên hỗn loài nhiệt đới, phần lớn các nghiên cứu chỉ phân tích sự kết nhóm của khoảng dưới 5 loài cây gỗ. Những loài cây gỗ được quan tâm là những loài cây gỗ ưu thế, đồng ưu thế và những loài cây gỗ có tỷ trọng cao trong QXTV.

4.3. Tương quan giữa các biến có thứ bậc và áp dụng trong lâm học

4.3.1. Phương pháp phân tích mối quan hệ giữa các biến có thứ bậc

Thang đo thứ bậc là thang đo định danh, nhưng các giá trị của một biến số có thứ bậc hơn kém. Các biến định danh có thứ bậc hơn kém còn được gọi là các biến phân loại (Classification Variables), các biến lập nhóm (Grouping Variables) và các biến mã hóa (Coding Variables). Vì thế, các biến định danh có thứ bậc hơn kém được sử dụng để phân loại các đối tượng khác nhau.

Trong lâm học, thang đo thứ bậc không chỉ được sử dụng để đo những đặc tính của cây gỗ và rừng, mà còn cả một số đặc tính của môi trường. Chẳng hạn: (1) Năng suất rừng trồng có thể được phân chia thành 5 cấp: 1 = rất thấp; 2 = thấp; 3 = trung bình; 4 = cao; 5 = rất cao. (2) Chất lượng của lập địa có thể được phân chia thành 5 cấp: 1 = rất xấu; 2 = xấu; 3 = trung bình; 4 = tốt; 5 = rất tốt. (3) Bệnh của cây gỗ có thể được phân chia theo 5 cấp: 1 = bệnh rất nặng; 2 = bệnh nặng; 3 = bệnh trung bình; 4 = bệnh nhẹ và 5 = không có bệnh.

Bằng cách phân chia các đặc tính của rừng và môi trường theo thứ bậc hơn kém hay theo nhóm, nhà lâm học có thể xác định được mối quan hệ giữa chúng. Mối quan hệ này có thể được phân tích bằng cách xây dựng thành bảng ngẫu nhiên gồm R*C và thực hiện các kiểm định về mối quan hệ giữa biến dòng với biến cột.

Phương pháp phân tích mối quan hệ giữa các biến có thứ bậc được áp dụng với 2 điều kiện: (1) Các biến định danh có thứ bậc hơn kém; (2) Các biến có thể lập thành nhóm (Cấp) trong bảng chéo R*C. Trình tự phân tích mối quan hệ giữa các biến có thứ bậc hơn kém bao gồm 2 bước.

Bước 1. Xác định tương quan giữa hai hay nhiều biến lập nhóm. Mối quan hệ giữa các biến lập nhóm có thể được xác định bằng hệ số tương quan hạng của Spearman, hệ số tương quan Pearson và kiểm định quan hệ tuyến tính bằng tiêu chuẩn χ^2. Hệ số tương quan Pearson đòi hỏi hai đại lượng quan sát phải có phân bố chuẩn. Hệ số tương quan Spearman không đòi hỏi hai đại lượng quan sát có phân phối chuẩn. Vì thế, hệ số tương quan Spearman thường được sử dụng để xác định mối quan hệ giữa hai biến có thứ bậc hơn kém.

Bước 2. Xác định xu hướng và cường độ của mối quan hệ giữa hai hay nhiều biến lập nhóm. Xu hướng của mối quan hệ phản ánh các đặc tính nghiên cứu biến đổi tương đồng hay không tương đồng với nhau. Nếu các cặp giá trị của hai đặc tính nghiên cứu cùng tăng hoặc cùng giảm, thì hai đặc tính này biến đổi tương đồng với nhau. Chẳng hạn: Hai giá trị thứ 2 và thứ 3 ở đặc tính A tương ứng là 5 và 7, còn ở đặc tính B tương ứng là 4 và 6. Trái lại, nếu các cặp giá trị của hai đặc tính nghiên cứu biến đổi khác nhau, thì hai đặc tính này biến đổi không tương đồng với nhau. Chẳng hạn: Hai giá trị thứ 10 và thứ 11 ở đặc tính A tương ứng là 6 và 8, còn ở đặc tính B tương ứng là 8 và 3. Khi hai giá trị liên tiếp của một đặc tính bằng nhau, thì cặp giá trị này được gọi là cặp biến bị ràng buộc. Chẳng hạn: Hai giá trị thứ 8 và thứ 9 ở đặc tính A đều là 5 và 5. Nói khác đi, khi các hạng của biến Y có xu hướng tăng cùng với các hạng của biến X (Y và X đồng biến), thì các cặp biến ấy được gọi là

phù hợp hay tương đồng. Ngược lại, khi các hạng của biến Y tăng lên còn hạng của biến X giảm xuống (Y và X nghịch biến), thì các cặp biến ấy được gọi là không phù hợp hay không tương đồng. Khi số cặp phù hợp bằng với số cặp không phù hợp, thì hai biến số ấy không có quan hệ với nhau hay không tương đồng với nhau.

Xu hướng và cường độ của mối quan hệ giữa hai hay nhiều biến lập nhóm có thứ bậc hơn kém được xác định bằng các hệ số tương quan dưới đây.

(a) Hệ số tương quan Somers'D và Conditional Gamma. Hai hệ số tương quan này đo xu hướng của mối quan hệ giữa hai biến Y và X. Chúng được áp dụng cho hai biến lập nhóm có thứ bậc hơn kém và vai trò của biến độc lập và biến phụ thuộc là rõ ràng. Hệ số tương quan Somers'D được tính theo công thức 4.25. Hệ số tương quan Gamma được tính theo công thức 4.26. Ở công thức 4.25 và 4.26, D = hệ số tương quan Somers'D; G = hệ số tương quan Gamma; P = số cặp biến tương đồng với nhau, nghĩa là cùng tăng hoặc cùng giảm; Q = số cặp biến không tương đồng với nhau, nghĩa là chúng tăng và giảm khác nhau; Ty = những cặp bị ràng buộc bởi yếu tố Y. Hai hệ số Somers và Gamma nhận giá trị = -1 đến 1.

$$D_Y = \frac{(P - Q)}{(P + Q + Ty)} \qquad (4.25)$$

$$G = \frac{P - Q}{P + Q} \qquad (4.26)$$

(b) Hệ hệ số tương quan Goodman và Kruskal's Gamma, Kendall tau b, Kendall tau c. Ba hệ số tương quan này đo cường độ của mối quan hệ giữa hai biến Y và X. Hệ số tương quan Goodman và Kruskal's Gamma (Gamma) được tính theo công thức 4.26. Hệ số tương quan Kendall tau b và Kendall tau c được tính tương ứng theo công thức 4.27 và 4.28. Ở công thức 4.27 và 4.28, r_B và r_C = hệ số tương quan Kendall tau b và Kendall tau c; P = số cặp biến tương đồng với nhau; Q = số cặp không tương đồng với nhau; Tx = những cặp bị ràng buộc bởi yếu tố X; Ty = những cặp bị ràng buộc bởi yếu tố Y; m = số nhỏ nhất trong hai số cột và số hàng của bảng R*C; N = tổng số cặp quan sát Y và X. Hệ số Kendall = 0 đến 1.

$$r_B = \frac{P - Q}{\sqrt{(P + Q + T_X)(P + Q + Ty)}} \qquad (4.27)$$

$$r_C = \frac{2m*(P - Q)}{N^2*(m - 1)} \qquad (4.28)$$

Ba hệ số tương quan G, r_B và r_C nhận giá trị bằng 0 khi P = Q. Sự tồn tại của 4 hệ số tương quan (Somers'd, Gamma, Kendall tau b, Kendall tau c) được kiểm định bằng thống kê T –Student (Công thức 4.29); trong đó r = hệ số tương quan, P = số cặp biến tương đồng với nhau, Q = số cặp không tương đồng với nhau, N = tổng số cặp quan sát Y và X.

$$T = r* \sqrt{\frac{P + Q}{N(1 - r^2)}} \qquad (4.29)$$

Giả thuyết H_0: Hệ số tương quan (Somers'd, Gamma, Kendall tau b, Kendall tau c) = 0 (không tồn tại).

Đối thuyết H_0^-: Hệ số tương quan (Somers'd, Gamma, Kendall tau b, Kendall tau c) khác 0 (tồn tại).

Quy tắc quyết định. Khi P < 0,05, thì giả thuyết H_0 bị bác bỏ. Ngược lại, khi P > 0,05, thì giả thuyết H_0 được chấp nhận.

4.3.2. Phân tích quan hệ giữa các biến có thứ bậc bằng phần mềm SPSS

Giả sử phân tích mối quan hệ giữa tái sinh tự nhiên của Dầu con rái với độ tàn che của tán rừng. Mật độ cây tái sinh (N, cây/ha) tự nhiên của Dầu con rái được phân chia thành 5 cấp: 1 = rất kém (<1000 cây/ha); 2 = kém (1000 - 2000 cây/ha); 3 = trung bình (2000 – 3000 cây/ha); 4 = tốt (3000 – 4000 cây/ha); 5 = rất tốt (> 4000 cây/ha). Độ tàn che (C) của tán rừng được phân chia theo 5 cấp: 1 = C < 0,4; 2 = 0,4 – 0,5; 3 = 0,5 – 0,6; 4 = 0,7 – 0,8 và 5 = C > 0,9. Tổng số ô dạng bản được điều tra là 250. Số liệu tái sinh của Dầu con rái và cấp độ tàn che được tổng hợp thành Bảng 5*5 (Bảng 4.6). Để phân tích mối quan hệ giữa cấp tái sinh tự nhiên của Dầu con rái với cấp độ tàn che tán rừng bằng phần mềm SPSS, trình tự các bước xử lý như dưới đây.

Bảng 4.6. Phân bố cấp mật độ cây tái sinh tự nhiên của Dầu con rái theo các cấp độ tàn che của tán rừng.

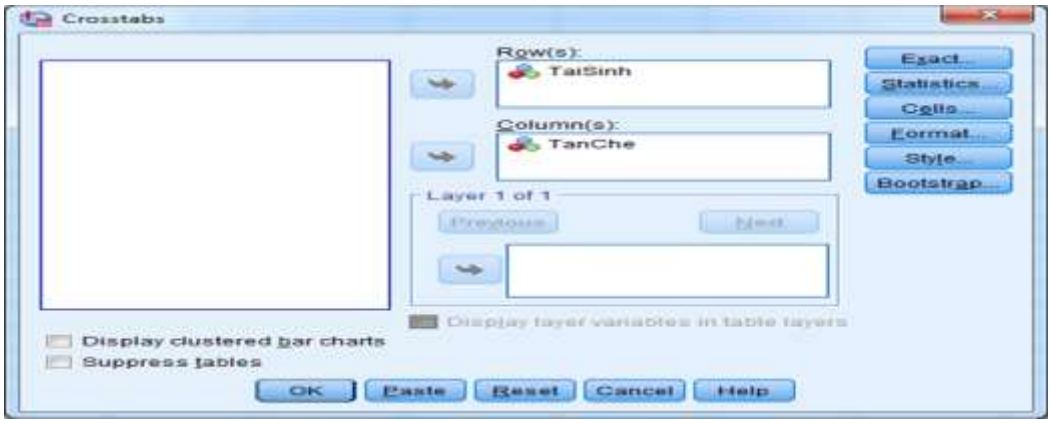
			Cấp độ tàn che tán rừng					Tổng số
			1	2	3	4	5	
Cấp tái sinh của Dầu con rái	1	N thực nghiệm	8					8
		N ước lượng	1,2	2,4	2,0	1,3	1,1	8,0
	2	N thực nghiệm	29	8	4	4	12	57
		N ước lượng	8,4	17,1	14,4	9,6	7,5	57,0
	3	N thực nghiệm		63		4	4	71
		N ước lượng	10,5	21,3	17,9	11,9	9,4	71,0
	4	N thực nghiệm		4	59		17	80
		N ước lượng	11,8	24,0	20,2	13,4	10,6	80,0
	5	N thực nghiệm				34		34
		N ước lượng	5,0	10,2	8,6	5,7	4,5	34,0
Tổng số		N thực nghiệm	37	75	63	42	33	250
		N ước lượng	37.0	75,0	63,0	42,0	33,0	250,0

Bước 1. Nhập số liệu vào bảng tính. Khi nhập số liệu vào bảng tính, cột 1 ghi 5 cấp mật độ cây tái sinh của Dầu con rái, còn cột 2 là 5 cấp độ tàn che tán rừng theo số mã hóa (Hình 4.6).

Bước 2. Xây dựng Bảng chéo 5*5 giữa 5 cấp mật độ cây tái sinh của Dầu con rái và 5 cấp độ tàn che tán rừng. Trước hết vào đường dẫn: Analyze > Descriptive Statistics > Crosstabs. Đến đây một cửa sổ xuất hiện. Tại cửa sổ này,

chọn 5 cấp mật độ cây tái sinh của Dầu con rái và 5 cấp độ tàn che tán rừng và chuyển chúng vào hai ô Row(s) và Column(s) (Hình 4.7).

Hình 4.6. Số liệu tái sinh của Dầu con rái và độ tàn che tán rừng.



Hình 4.7. Xây dựng Bảng 5*5 giữa 5 cấp mật độ cây tái sinh của Dầu con rái và 5 cấp độ tàn che của tán rừng.

Bước 3. Xác định mối quan hệ giữa mật độ cây tái sinh của Dầu con rái và cấp độ tàn che của tán rừng. Giả thuyết H_0^+: Mật độ cây tái sinh của Dầu con rái không có quan hệ với độ tàn che của tán rừng. Giả thuyết này bị bác bỏ khi mức ý nghĩa (P) của các hệ số tương quan (Spearman, Kendall tau b, Kendall tau c, Somers'd, Gamma) nhỏ hơn 0,05. Để kiểm định giả thuyết này, trước hết chọn Statistics ở Hình 4.7. Đến đây xuất hiện cửa sổ Crosstabs. Kế đến tại cửa sổ này chọn Correlaton và 4 hệ số (Kendall tau b, Kendall tau c, Somers'd, Gamma) đo xu hướng và cường độ của mối quan hệ giữa mật độ cây tái sinh của Dầu con rái và cấp độ tàn che của tán rừng (Hình 4.8). Sau đó chọn OK để nhận kết quả phân tích mối quan hệ giữa mật độ cây tái sinh của Dầu con rái và cấp độ tàn che tán rừng.

Hình 4.8. Xác định hệ số tương quan giữa mật độ cây tái sinh của Dầu con rái với cấp độ tàn che của tán rừng.

Kết quả báo cáo 4 nội dung (Bảng 4.7): (1) Tóm tắt xử lý các trường hợp (Case Processing Summary); (2) Bảng chéo 5*5; (3) Các số đo xu hướng của mối quan hệ (Directional Measures); (4) Các số đo tính đối xứng hay cường độ của mối quan hệ (Symmetic Measures).

Bảng 4.7. Kết quả kiểm định mối quan hệ giữa mật độ cây tái sinh của Dầu con rái với cấp độ tàn che của tán rừng.

4.7a. Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Tái sinh * Tàn che	250	100.0%	0	0.0%	250	100.0%

4.7b. Directional Measures

			Value	Error	T	P
Ordinal by Ordinal	Somers' d	Symmetric	0.567	0.053	10.841	0.000
		TaiSinh Dependent	0.555	0.053	10.841	0.000
		TanChe Dependent	0.580	0.053	10.841	0.000

4.7c. Symmetric Measures

		Value	Error	T	P
Ordinal by Ordinal	Kendall's tau-b	0.568	0.053	10.841	0.000
	Kendall's tau-c	0.541	0.050	10.841	0.000
	Gamma	0.631	0.055	10.841	0.000
	Spearman Correlation	0.594	0.060	11.630	0.000
Interval by Interval	Pearson's R	0.543	0.057	10.183	0.000
N of Valid Cases		250			

Bảng 4.7a trình bày tóm tắt xử lý các trường hợp. Bảng 4.7b báo cáo xu hướng của mối quan hệ giữa mật độ cây tái sinh tự nhiên của Dầu con rái với 5 cấp độ tàn che của tán rừng bằng hệ số tương quan Somers. Hệ số Somers = 0,567 với T = 10,84 và P < 0,001 cho biết mật độ cây tái sinh tự nhiên của Dầu con rái có quan hệ dương khá chặt chẽ với độ tàn che của tán rừng. Bảng 4.7c báo cáo 5 hệ số đo cường độ của mối quan hệ giữa mật độ cây tái sinh tự nhiên của Dầu con rái với độ tàn che của tán rừng. Ba hệ số (Kendall tau b, Kendall tau c, Gamma) nhận giá giá trị khá lớn, dao động từ 0,541 (Kendall's tau-c) đến 0,631 (Gamma). Hệ số Somers mang giá trị dương và ba hệ số (Kendall tau b, Kendall tau c, Gamma) nhận giá trị khá cao chứng tỏ mật độ cây tái sinh tự nhiên của Dầu con rái gia tăng dần theo sự gia tăng độ tàn che của tán rừng. Về bản chất, độ tàn che của tán rừng càng lớn thì Dầu con rái tái sinh tự nhiên càng mạnh. Từ thống kê Somers'D cho thấy, nếu sử dụng độ tàn che của tán rừng để ước lượng mật độ cây tái sinh tự nhiên của Dầu con rái, thì tỷ lệ giảm sai số là 55,5% (TaiSinh Dependent = 0,555). Từ thống kê Gamma cho thấy, nếu sử dụng độ tàn che của tán rừng để ước lượng mật độ cây tái sinh tự nhiên của Dầu con rái, thì tỷ lệ giảm sai số là 63,1% (Gamma = 0,631). Tỷ lệ giảm sai số dự đoán khác nhau là do sử dụng các hệ số tương quan khác nhau.

4.3.3. Phân tích tương quan thứ bậc bằng bảng chéo nhiều chiều

Bảng chéo nhiều chiều là tập hợp của những bảng chéo hai chiều; trong đó mỗi biến số là một chiều. Số liệu dùng trong phân tích bao gồm một biến hàng, một biến cột và một biến lớp (Layer Variable). Biến lớp còn được gọi là biến điều khiển hay biến lập nhóm (Grouping Variable). Từ số liệu của bảng chéo nhiều chiều, chúng ta có thể xây dựng các bảng phụ 2 chiều cho từng trị số của biến lớp hoặc kết hợp giữa hai hoặc lớn hơn hai biến lớp. Mục đích của kiểm định trong nội bộ các lớp là cung cấp những thông tin về mối quan hệ giữa hai biến trong nội bộ các lớp.

Kiểm định trong nội bộ các lớp bao gồm hai nội dung. Một là xác định mối quan hệ và cường độ của mối quan hệ giữa hai biến số. Hai là xác định cường độ của mối quan hệ giữa hai biến số trong nội bộ các lớp. Điều này có nghĩa là phân tích ảnh hưởng của yếu tố thứ 3 đến mối quan hệ của hai biến ban đầu.

Giả sử nghiên cứu đặt ra 2 mục tiêu: (1) Xác định quan hệ giữa mức độ bị bệnh của Dầu con rái với tuổi của nó. (2) So sánh tình trạng bệnh của Dầu con rái theo 3 chỉ số lập địa khác nhau (SI). Số liệu thu thập là 180 cây Dầu con rái ở 3 cấp tuổi (A = 2, 4 và 6 năm) và 3 chỉ số SI (14, 18 và 22 m). Trong trường hợp này, lập địa là yếu tố lớp (Layer) trong một bảng quan hệ giữa mức độ bị bệnh của Dầu con rái với tuổi của của nó. Tình trạng bệnh của Dầu con rái theo 3 cấp tuổi và 3 chỉ số SI được tổng hợp ở Bảng 4.8. Để phân tích mối quan hệ giữa mức độ bị bệnh của Dầu con rái với tuổi và so sánh tình trạng bệnh theo 3 lập địa khác nhau bằng phần mềm SPSS, trình tự các bước xử lý như dưới đây.

Bước 1. Nhập số liệu vào bảng tính. Khi nhập số liệu vào bảng tính, cột 1 ghi 3 cấp chỉ số SI, còn cột 2 và cột 3 tương ứng là tuổi (A, năm) và mức độ bị bệnh (Hình 4.9).

Bảng 4.8. Phân bố số cây Dầu con rái theo tình trạng bệnh và lập địa.

Chỉ số SI			Cấp A (năm)			Tổng theo hàng
			2	4	6	
14	Bệnh	1	12	5	0	17
		2	5	7	7	19
		3	3	8	13	24
	Tổng số		20	20	20	60
18	Bệnh	1	3	3	1	7
		2	12	5	4	21
		3	5	12	15	32
	Tổng số		20	20	20	60
22	Bệnh	1	5	4	5	14
		2	7	6	6	19
		3	8	10	9	27
	Tổng số		20	20	20	60
Tổng theo cột	Bệnh	1	20	12	6	38
		2	24	18	17	59
		3	16	30	37	83
	Tổng số		60	60	60	180

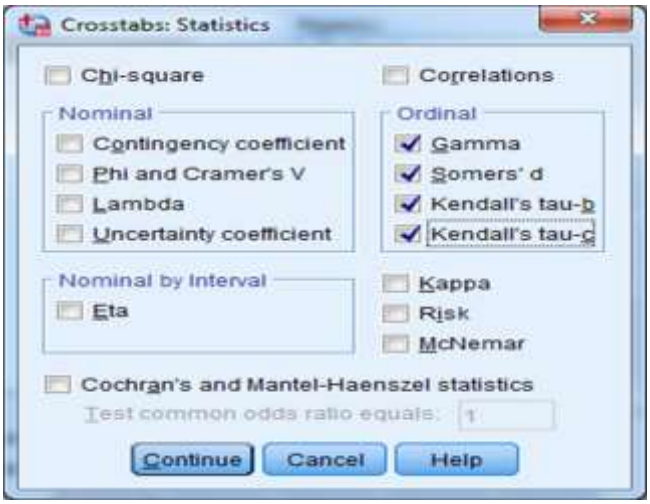
Hình 4.9. Số liệu Dầu con rái theo tình trạng bệnh và lập địa.

Bước 2. Xây dựng bảng chéo mô tả mối quan hệ giữa cấp bệnh theo cấp A. Trước hết vào đường dẫn: Analyze > Descriptive Statistics > Crosstabs. Đến đây một cửa sổ xuất hiện. Tại cửa sổ này, chọn biến "Bệnh" và biến "A" và chuyển chúng vào hai ô Row(s) và Column(s) (Hình 4.10).

Bước 3. Xác định mối quan hệ giữa mức độ bị bệnh của Dầu con rái với tuổi. Giả thuyết H_0^+: Mức độ bị bệnh của Dầu con rái không phụ thuộc vào tuổi. Giả thuyết này bị bác bỏ khi mức ý nghĩa (P) của các hệ số tương quan (Kendall tau b, Kendall tau c, Somers'd, Gamma) nhỏ hơn 0,05. Để kiểm định giả thuyết này, trước

hết chọn Statistics ở Hình 4.10. Đến đây một cửa sổ Crosstabs xuất hiện. Kế đến tại cửa sổ này chọn 4 hệ số (Kendall tau b, Kendall tau c, Somers'd, Gamma) biểu thị xu hướng và cường độ của mối quan hệ giữa mức độ bị bệnh của Dầu con rái với tuổi (Hình 4.11). Sau đó chọn OK để nhận kết quả phân tích mối quan hệ giữa mức độ bị bệnh của Dầu con rái với tuổi (Bảng 4.9).

Hình 4.10. Xây dựng Bảng chéo 3 cấp bệnh và 3 cấp A của Dầu con rái.



Hình 4.11. Xác định mối quan hệ giữa 3 cấp bệnh với tuổi.

Bảng 4.9a báo cáo xu hướng của mối quan hệ giữa mức độ bị bệnh của Dầu con rái với tuổi bằng hệ số tương quan Somers. Hệ số Somers = 0,273 với T = 4,472 và P < 0,001 cho biết mức độ bị bệnh của Dầu con rái có xu hướng gia tăng theo tuổi. Bảng 4.9b báo cáo 4 hệ số đo cường độ của mối quan hệ giữa mức độ bị bệnh của Dầu con rái với tuổi. Bốn hệ số này khá lớn và dao động từ 0,267 (Kendall's tau-c) đến 0,410 (Gamma). Hệ số Somers và bốn hệ số này đều mang giá trị dương chứng tỏ mức độ bị bệnh của Dầu con rái gia tăng theo tuổi. Từ thống kê Gamma

cho thấy, nếu sử dụng tuổi để ước lượng mức độ bị bệnh của Dầu con rái, thì tỷ lệ giảm sai số là 41,0% (Gamma = 0,410).

Bảng 4.9. Kiểm định mối quan hệ giữa mức độ bị bệnh của Dầu con rái với tuổi.

4.9a. Directional Measures

			Value	Error	T	P
Ordinal by Ordinal	Somers' d	Symmetric	0.273	0.061	4.472	0.000
		Benh Dependent	0.267	0.060	4.472	0.000
		ANam Dependent	0.280	0.062	4.472	0.000

4.9b. Symmetric Measures

		Value	Error	T	P
Ordinal by Ordinal	Kendall's tau-b	0.274	0.061	4.472	0.000
	Kendall's tau-c	0.267	0.060	4.472	0.000
	Gamma	0.410	0.087	4.472	0.000
	N of Valid Cases	180			

Bước 4. Kiểm định mối quan hệ giữa mức độ bị bệnh của Dầu con rái với chỉ số lập địa. Giả thuyết H_0^+: Mức độ bị bệnh của Dầu con rái không phụ thuộc vào lập địa. Giả thuyết này bị bác bỏ khi mức ý nghĩa (P) của các hệ số tương quan (Kendall tau b, Kendall tau c, Somers'd, Gamma) nhỏ hơn 0,05. Để kiểm định giả thuyết này, trước hết từ cửa sổ ở Hình 4.10, bỏ nhân tố A và thay bằng chỉ số SI. Đến đây một cửa sổ Crosstabs xuất hiện. Kế đến tại cửa sổ này chọn Statistic và 4 hệ số (Kendall tau b, Kendall tau c, Somers'd, Gamma) biểu thị xu hướng và cường độ của mối quan hệ giữa mức độ bị bệnh của Dầu con rái với chỉ số SI. Sau đó chọn OK để nhận kết quả phân tích mối quan hệ giữa mức độ bị bệnh của Dầu con rái với chỉ số SI (Bảng 4.10). Hệ số Somers = 0,046 với P = 0,513 và hệ số Gamma = 0,07 với P = 0,513 chứng tỏ mức độ bị bệnh của Dầu con rái có quan hệ không rõ rệt với chỉ số SI.

Bảng 4.10. Kết quả kiểm định mối quan hệ giữa mức độ bị bệnh của Dầu con rái với chỉ số SI.

4.10a. Directional Measures

			Value	Error	T	P
Ordinal by Ordinal	Somers' d	Symmetric	0.046	0.070	0.654	0.513
		Benh Dependent	0.044	0.068	0.654	0.513
		SI Dependent	0.047	0.071	0.654	0.513

4.10b. Symmetric Measures

		Value	Error	T	P
Ordinal by Ordinal	Kendall's tau-b	0.046	0.070	0.654	0.513
	Kendall's tau-c	0.044	0.068	0.654	0.513
	Gamma	0.070	0.106	0.654	0.513
	N of Valid Cases	180			

Bước 5. Kiểm định mối quan hệ giữa mức độ bị bệnh của Dầu con rái với tuổi trong từng cấp chỉ số lập địa. Giả thuyết H_0^+: Mức độ bị bệnh của Dầu con rái trong mỗi cấp chỉ số SI không phụ thuộc vào tuổi. Giả thuyết này bị bác bỏ khi mức ý nghĩa (P) của các hệ số tương quan (Kendall tau b, Kendall tau c, Somers'd, Gamma) nhỏ hơn 0,05. Để kiểm định giả thuyết này, trước hết mở Analyze > Descriptive Statistics > Crosstabs. Đến đây một cửa sổ xuất hiện. Tại cửa sổ này, chọn biến "Bệnh" và biến "A" và chuyển chúng vào hai ô Row(s) và Column(s), còn biến chỉ số SI chuyển vào ô Layer 1 of 1 (Hình 4.12). Đến đây một cửa sổ Crosstabs xuất hiện. Kế đến tại cửa sổ này chọn Statistic và bốn hệ số (Kendall tau b, Kendall tau c, Somers'd, Gamma) biểu thị xu hướng và cường độ của mối quan hệ giữa mức độ bị bệnh của Dầu con rái với tuổi trong mỗi cấp chỉ số SI (Hình 4.13). Sau đó chọn OK để nhận kết quả phân tích mối quan hệ giữa mức độ bị bệnh của Dầu con rái với tuổi trong mỗi cấp chỉ số SI (Bảng 4.11).

Hình 4.12. Xây dựng bảng chéo biểu thị mối quan hệ giữa mức bị bệnh với tuổi Dầu con rái theo ba cấp chỉ số SI.

Hình 4.13. Xác định mối quan hệ giữa 3 cấp bệnh với tuổi trong mỗi cấp chỉ số SI.

Bảng 4.11a báo cáo khuynh hướng của mối quan hệ giữa mức độ bị bệnh của Dầu con rái với tuổi trong mỗi cấp chỉ số SI. Hệ số Somers chỉ tồn tại và nhận giá trị dương ở cấp chỉ số SI = 14 m (0,484; P < 0,01) và cấp chỉ số SI = 18 m (0,344; P < 0,01) và không tồn tại ở cấp chỉ số SI = 22 m (0,273; P = 0,820). Ba hệ số tương quan (Kendall tau b, Kendall tau c, Gamma) ở cấp chỉ số SI = 14 m và SI = 18 m đều nhận giá trị khá cao, dao từ 0,322 đến 0,677. Điều đó chứng tỏ mức độ bị bệnh của Dầu con rái ở hai cấp lập địa này gia tăng dần theo tuổi.

Bảng 4.11. Kiểm định mối quan hệ giữa mức độ bị bệnh của Dầu con rái với tuổi trong mỗi cấp chỉ số SI.

11.a. Directional Measures

Chỉ số lập địa				Value	Error	T	P
14	Ordinal by Ordinal	Somers' d	Symmetric	0.484	0.086	5.5	0.000
			Benh Dependent	0.482	0.087	5.5	0.000
			ANam Dependent	0.487	0.086	5.5	0.000
18	Ordinal by Ordinal	Somers' d	Symmetric	0.344	0.101	3.4	0.001
			Benh Dependent	0.322	0.094	3.4	0.001
			ANam Dependent	0.370	0.110	3.4	0.001
22	Ordinal by Ordinal	Somers' d	Symmetric	0.026	0.116	0.2	0.820
			Benh Dependent	0.026	0.114	0.2	0.820
			ANam Dependent	0.027	0.118	0.2	0.820
Tổng	Ordinal by Ordinal	Somers' d	Symmetric	0.273	0.061	4.5	0.000
			Benh Dependent	0.267	0.060	4.5	0.000
			ANam Dependent	0.280	0.062	4.5	0.000

11b. Symmetric Measures

Chỉ số lập địa				Value	Error	T	P
14	Ordinal by Ordinal	Kendall's tau-b		.484	.086	5.540	.000
		Kendall's tau-c		.482	.087	5.540	.000
		Gamma	Zero-Order	.677	.102	5.540	.000
	N of Valid Cases			60			
18	Ordinal by Ordinal	Kendall's tau-b		.345	.101	3.431	.001
		Kendall's tau-c		.322	.094	3.431	.001
		Gamma	Zero-Order	.520	.140	3.431	.001
	N of Valid Cases			60			
22	Ordinal by Ordinal	Kendall's tau-b		.026	.116	.227	.820
		Kendall's tau-c		.026	.114	.227	.820
		Gamma	Zero-Order	.040	.177	.227	.820
	N of Valid Cases			60			
Total	Ordinal by Ordinal	Kendall's tau-b		.274	.061	4.472	.000
		Kendall's tau-c		.267	.060	4.472	.000
		Gamma	Zero-Order	.410	.087	4.472	.000
			First-Order Partial	.420			
	N of Valid Cases			180			

Chương 5
PHÂN TÍCH HỒI QUY TRONG LÂM HỌC

5.1. Giới thiệu chung

Kiến thức về mối quan hệ giữa rừng với môi trường sống của chúng là cơ sở cho việc xây dựng lý thuyết về rừng và các phương thức lâm sinh, phân vùng lập địa rừng và phân loại thảm thực vật rừng. Đời sống của các loài cây gỗ và rừng bị kiểm soát bởi rất nhiều nhân tố sinh thái. Các nhân tố sinh thái tác động rất phức tạp đến cây gỗ và rừng. Nhà lâm học chẳng thể kiểm soát được tất cả các nhân tố sinh thái. Vì thế, nhà lâm học chỉ tập chung nghiên cứu những nhân tố sinh thái chủ yếu mà con người có thể kiểm soát được thông qua những hoạt động lâm sinh.

Khi bị tác động bởi các nhân tố sinh thái, cây gỗ và rừng phản ứng lại thông qua sự biến đổi các đặc tính của chúng. Trong sinh thái học và lâm học, người ta gọi sự phản ứng của cây gỗ và rừng đối với sự biến đổi của những nhân tố sinh thái là sự phản hồi. Các phản hồi của cây gỗ và rừng biểu hiện ở sự thay đổi các đặc tính của chúng. Một số đặc tính của cây gỗ và rừng có thể định lượng bằng những số đo nào đó. Chẳng hạn D (cm); H (m), g (m^2), V (m^3), B (kg, tấn) ở mức cây gỗ; N, G, M và B ở mức quần thụ. Một số đặc tính của cây gỗ và rừng không thể định lượng bằng những số đo nào đó. Chẳng hạn: Độ bắt gặp loài; Sức sống của cây gỗ và rừng; Màu sắc (lá, vỏ...) của cây gỗ. Do nhiều hạn chế khác nhau, nên một số đặc tính định lượng ở mức cây gỗ và rừng cũng được chuyển thành đặc tính định tính. Chẳng hạn: Số lượng hoa, quả và hạt trên tán cây gỗ; Mật độ lá trên tán cây...

Bằng phân tích hồi quy, nhà lâm học có thể mô tả những phản hồi của cây gỗ và rừng bằng một hàm số của những nhân tố sinh thái. Hàm số này được gọi là hàm phản hồi; trong đó biến phụ thuộc (Y) là những đặc tính của cây gỗ và rừng, còn các biến độc lập (X_i) là những nhân tố sinh thái. Trong lâm học, các biến phụ thuộc được gọi là biến phản hồi, còn các biến độc lập là biến giải thích hay biến nguyên nhân. Các hàm phản hồi biểu diễn phản ứng của các loài cây gỗ và rừng đối với sự thay đổi của những nhân tố sinh thái theo cách thức nhân quả nào đó. Tuy vậy, những nguyên nhân dẫn đến những thay đổi ở cây gỗ và rừng không thể suy luận đơn giản từ phân tích hồi quy. Phân tích hồi quy chỉ trả lời rõ một số câu hỏi sau đây: (1) Loài cây gỗ và rừng có quan hệ với những nhân tố sinh thái trong chuỗi lập địa này hay không? Câu hỏi này được trả lời thông qua kiểm định sự tồn tại của các hàm phản hồi. (2) Những biến môi trường nào ảnh hưởng rõ rệt, còn những biến môi trường nào ảnh hưởng không rõ rệt đến cây gỗ và rừng? Câu hỏi này được trả lời thông qua phân tích hồi quy và tương quan riêng phần. (3) Những biến môi trường nào ảnh hưởng tốt, còn những biến môi trường nào ảnh hưởng xấu đến cây gỗ và rừng? Câu hỏi này được trả lời thông qua phân tích dấu của các hệ số hồi quy riêng phần.

Trong lâm học, mục đích của phân tích hồi quy là mô tả những phản hồi của các loài cây gỗ và rừng như là một hàm của một hoặc nhiều nhân tố sinh thái. Kết quả của phân tích hồi quy được sử dụng để giải quyết những vấn đề sau đây:

(a) Phân tích vai trò của các nhân tố sinh thái và xác định những nhân tố sinh thái chủ yếu ảnh hưởng đến các loài cây gỗ và rừng.

(b) Ước lượng các tham số sinh thái của các loài cây gỗ và rừng. Chẳng hạn: Biên độ sinh thái; Tối ưu sinh thái; Phạm vi chống chịu; Vùng tử vong.

(c) Xây dựng các hàm phản hồi để ước lượng và dự đoán những thay đổi của các loài cây gỗ và rừng theo sự thay đổi của một hoặc nhiều biến môi trường.

(d) Xây dựng các hàm để dự đoán những thay đổi của môi trường dựa trên những đặc tính của các loài cây gỗ và rừng.

Các hàm phản hồi ở mức cây gỗ và rừng là rất đa dạng. Sự đa dạng của các hàm phản hồi không chỉ là do sự đa dạng của các biến phản hồi và các biến giải thích, mà còn do phương pháp thu thập các đặc tính ở mức cây gỗ và rừng. Tuy theo mục đích nghiên cứu, hai biến này được thu thập bằng những phương pháp khác nhau. Nếu các biến phản hồi và các biến giải thích đều được định lượng, thì công cụ xử lý thích hợp nhất là phương pháp phân tích hồi quy và tương quan đơn biến hoặc đa biến. Phương pháp này cũng được áp dụng trong trường hợp các biến phản hồi được đo bằng hai biến nhị phân (Bắt gặp = 1; Không bắt gặp = 0), còn các biến giải thích được định lượng. Trái lại, nếu cả hai biến phản hồi và biến giải thích chỉ là biến định danh không có thứ bậc hơn kém (Biến nhị phân), thì công cụ xử lý thích hợp là kiểm định χ^2 để xác định mối quan hệ giữa các biến phản hồi và các biến giải thích. Hạn chế của phương pháp này là không thể giải thích chính xác mức độ và chiều hướng ảnh hưởng của các nhân tố sinh thái. Tóm lại, tùy theo mục đích nghiên cứu, nhà lâm học có thể áp dụng các phương pháp khác nhau để xử lý số liệu nhằm trả lời những câu hỏi đặt ra.

Hồi quy đối với một biến giải thích định lượng bao gồm việc làm phù hợp số liệu với một đường cong. Vấn đề là ở chỗ, nhà lâm học cần chọn dạng đường cong nào là thích hợp với tập số liệu nghiên cứu. Để giải quyết vấn đề vừa nói, nhà lâm học có thể sử dụng đồ thị phân tán để mô tả quan hệ giữa các biến phản hồi với các biến giải thích hoặc dựa theo lý thuyết. Trong phân tích hồi quy, chữ X biểu thị cho biến giải thích (Biến môi trường), còn Ey biểu thị cho biến phản hồi kỳ vọng của biến phản hồi Y. Hình 5.1 mô tả những dạng đường cong phản hồi.

(a) Dạng hằng số. Giá trị Ey bằng một hằng số (Hình 5.1a) cho biết Ey không phụ thuộc X.

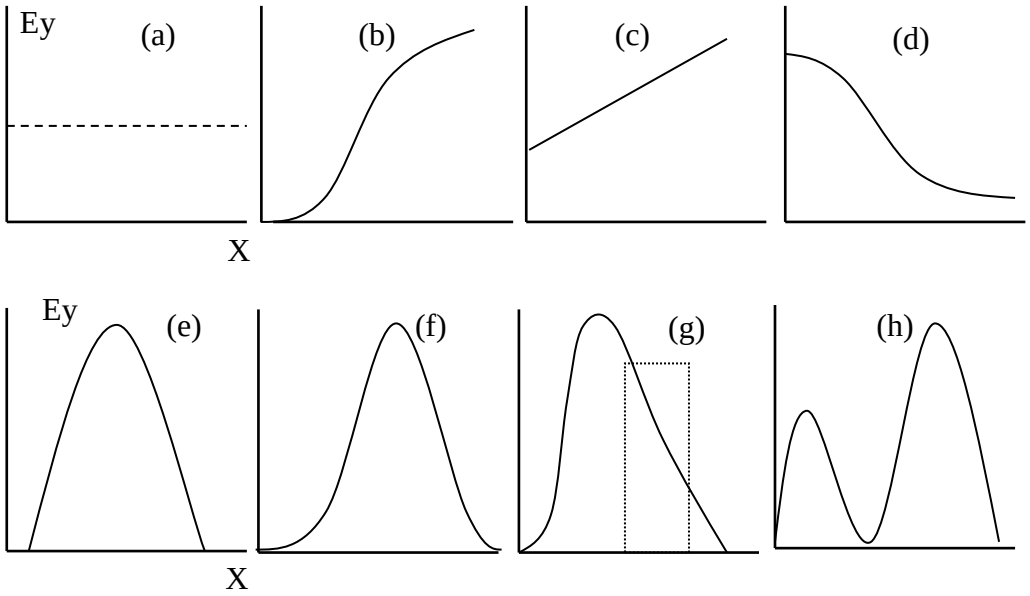
(b) Dạng tăng hoặc giảm đều. Giá trị Ey tăng hoặc giảm cùng với sự gia tăng giá trị X. Dạng này bao gồm dạng đường thẳng và đường cong Sigmoid (Hình 5.1b, c, d).

(c) Dạng một đỉnh. Thoạt đầu Ey tăng lên cùng với sự gia tăng giá trị X; tiếp theo đạt cực đại và sau đó giảm. Dạng này bao gồm đường cong Parabol với 1 cực

đại (Hình 5.1e) và đường cong có dạng hình chuông (Hình 5.1f). Giá trị X mà tại đó Ey đạt lớn nhất được gọi là mốt hoặc tối ưu. Giá trị tối ưu không phải là duy nhất khi đường cong có đỉnh bằng (Hình 5.1g). Đường cong một đỉnh có thể đối xứng (Tối ưu là điểm đối xứng) hoặc bất đối xứng (Hình 5.1g).

(d) Dạng hai đỉnh. Thoạt đầu Ey tăng lên cùng với sự gia tăng giá trị X; kế đến đạt cực đại; tiếp theo giảm đến một cực tiểu; sau cực tiểu lại tăng lên đến cực đại mới; cuối cùng giảm liên tục (Hình 5.1h).

(e) Những dạng khác. Giá trị Ey có hình thái khác với những kiểu đường cong trên đây.



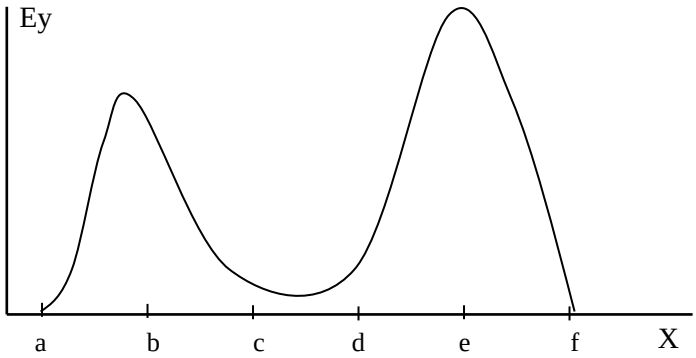
Hình 5.1. Các dạng đường cong phản hồi được lập theo biến môi trường.

(a) Đường cong hằng số; (b) Đường cong Sigmoid tăng; (c) Đường thẳng; (d) Đường cong Sigmoid giảm; (e) Đường cong Parabol; (f) Đường cong đối xứng hay đường cong Gauss; (g) Đường cong bất đối xứng hay lệch và một hàm khối; (h) Đường cong hai đỉnh.

Trên đây liệt kê những dạng đường cong theo mức độ phức tạp của chúng. Khi đường cong phản hồi ở dạng hằng số, thì Ey không phụ thuộc X hay biến giải thích không ảnh hưởng đến biến phản hồi Ey. Từ dạng đường cong hai đỉnh, bằng cách giới hạn khoảng thu mẫu khác nhau, nhà lâm học có thể nhận được những đường cong khác nhau. Thật vậy, đường cong tăng hoặc giảm đều là những trường hợp đặc biệt của đường cong một đỉnh. Khi tối ưu nằm ngoài khoảng thu mẫu, thì đường cong một đỉnh tăng dần hoặc giảm dần trong khoảng này (Hình 5.2). Những đường cong có 1 cực tiểu cũng có thể rơi vào trường hợp đặc biệt của đường cong hai đỉnh.

Rừng là hệ sinh thái rất phức tạp. Tính phức tạp biểu hiện ở chỗ, các đặc tính của rừng phân bố không đồng đều theo không gian và thời gian. Vì thế, các hàm phản hồi có hình thái rất phức tạp. Nhà lâm học không thể dễ dàng thu được các

thông tin để xây dựng các hàm phản hồi phức tạp. Để làm giảm tính phức tạp của các hàm phản hồi, nhà lâm học có thể phân chia rừng thành những đơn vị nhỏ hơn. Sau đó thu thập số liệu trong từng đơn vị nhỏ này và xây dựng hàm phản hồi dựa theo những hàm hồi quy đơn giản. Phần dưới đây giới thiệu một số phương pháp có thể áp dụng để phân tích các hàm phản hồi ở mức cây gỗ và rừng.



Hình 5.2. Những đường cong phản hồi được phân chia từ đường cong hai đỉnh bằng cách giới hạn khoảng thu mẫu.

(a) Đường cong hai đỉnh trong khoảng X = [a, f].
(b) Đường cong một đỉnh khi X = [a, c] và X = [d, f].
(c) Đường cong giảm đều khi X = [b, c] và [e, f].
(d) Đường cong tăng đều khi X = [c, e].
(e) Đường thẳng nằm ngang khi X = [c, d].

5.2. Hồi quy tuyến tính và ứng dụng trong lâm học

5.2.1. Hồi quy tuyến tính đơn biến

Hàm số mô tả quan hệ giữa Y với X kèm theo sai số (e) được gọi là hàm hồi quy tuyến tính đơn biến (Hàm 5.1).

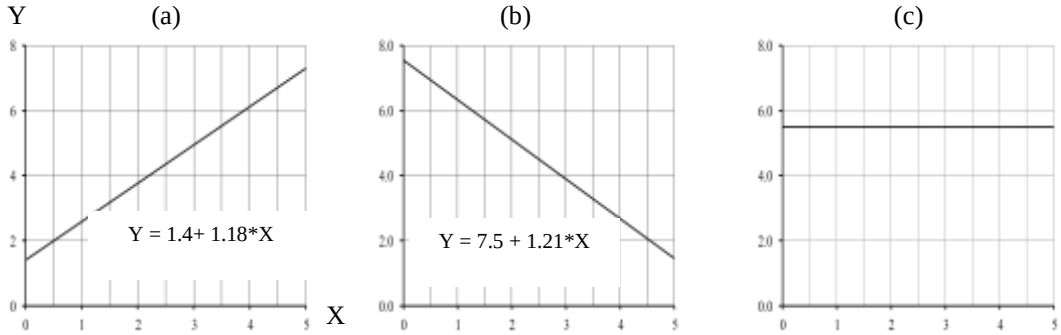
$$Y = b_0 + b_1*X + e \qquad (5.1)$$

Trong mô hình hồi quy tuyến tính đơn, Y là hàm số tuyến tính của X; b_0 và b_1 là các tham số của mô hình; e là biến ngẫu nhiên biểu thị sai số của mô hình. Sai số cho biết biến động của Y không thể giải thích được bằng mối quan hệ tuyến tính giữa Y với X.

Những giả định của hồi quy tuyến tính đơn: (1) Sai số kỳ vọng E(e) = 0; (2) Phương sai của e (S^2) bằng nhau đối với mọi X; (3) Các giá trị của e là độc lập; Sai số e có phân bố N(0, σ^2).

Đồ thị của hàm hồi quy tuyến tính đơn (Hình 5.3) là một đường thẳng; trong đó b_0 là điểm chặn (Intercept) của đường hồi quy trên trục tung, b_1 là độ dốc (Slope) của đường hồi quy, E(Y) là giá trị kỳ vọng của Y tương ứng với một giá trị X nhất định. Trong hồi quy tuyến tính đơn, hệ số b_1 là ước lượng sự thay đổi của Y theo một đơn vị thay đổi của X. Hình 5.3a cho biết Y có mối quan hệ dương với X, nghĩa

là Y và X là đồng biến. Hình 5.3b cho biết Y có mối quan hệ âm với X; nghĩa là Y và X là nghịch biến. Hình 5.3c cho biết Y không có mối quan hệ với X.



Hình 5.3. Mối quan hệ giữa Y với X.

Để xác định hàm hồi quy tuyến tính $Y = b_0 + b_1X$, chúng ta cần phải ước lượng hai tham số b_0 và b_1. Hai tham số này được ước lượng theo phương pháp bình phương nhỏ nhất (Công thức 5.2 và 5.3); trong đó n = tổng số quan sát, X_i = giá trị của biến độc lập tương ứng với quan sát i, Y_i = giá trị của biến phụ thuộc tương ứng với quan sát i, X_{Bq} = giá trị trung bình của biến độc lập, Y_{Bq} = giá trị trung bình của biến phụ thuộc.

$$b_1 = \frac{\sum_{(i=1,n)}(X_i - X_{Bq})(Y_i - Y_{Bq})}{\sum_{(i=1,n)}(X_i - X_{Bq})^2} \qquad (5.2)$$

$$b_0 = Y_{Bq} - b_1X_{Bq} \qquad (5.3)$$

Mức độ chặt chẽ của mối quan hệ giữa Y với X được đánh giá theo hệ số xác định (Công thức 5.4), SSR = tổng sai lệch bình phương do hồi quy (Công thức 5.5), SSE = tổng sai số bình phương (Công thức 5.6), SST = tổng sai lệch toàn bộ (Công thức 5.7), SST_{Corr} = tổng sai lệch thực hay hiệu chỉnh (Công thức 5.8).

$$r^2 = [1- \frac{\sum(Y_{TN} - Y_{UL})^2}{\sum(Y_{TN} - Y_{Bq})^2}]*100 \qquad (5.4)$$

$$SSR = \sum_{(i=1,n)}(Y_{UL} - Y_{Bq})^2 \qquad (5.5)$$

$$SSE = \sum_{(i=1,n)}(Y_{UL} - Y_{TN})^2 \qquad (5.6)$$

$$SST = SSR + SSE \qquad (5.7)$$

$$SST_{(Corr)} = \sum_{(i=1,n)}(Y_{TN} - Y_{Bq})^2 \qquad (5.8)$$

Hệ số tương quan (Kí hiệu r) là căn bậc hai của hệ số r^2 (Công thức 5.9). Hệ số r lấy giá trị từ - 1 đến +1. Khi r = 0 thì Y và X không có quan hệ với nhau. Khi r = -1 thì Y và X có quan hệ tuyến tính âm rất chặt chẽ. Khi r = +1 thì Y và X có quan hệ tuyến tính dương rất chặt chẽ.

$$r_{xy} = (\text{Dấu của } b_1)\sqrt{1 - \frac{SSE}{SST}} = \sqrt{r^2} \qquad (5.9)$$

Sự tồn tại của hàm $Y = b_0 + b_1*X$ được kiểm định bằng thống kê T và thống kê F. Thống kê T (Công thức 5.10 và 5.11) kiểm định sự tồn tại của tham số b_1. Giả thuyết H_0: $b_1 = 0$. Đối thuyết H_a: $b_1 \neq 0$. Ở mức ý nghĩa α (α = 0,05 hoặc 0,01), bác bỏ H_0 khi $P < \alpha$.

$$T = \frac{b_1}{S_{b1}} \qquad (5.10)$$

$$S_{b1} = \frac{S}{\sqrt{\sum(X_i - X_{Bq})^2}} \qquad (5.11)$$

Kiểm định F được tính theo công thức 5.12; trong đó MSR = hồi quy bình phương trung bình (Công thức 5.13), MSE = sai số trung bình bình phương (Công thức 5.14). Ở công thức 5.13 và 5.14, p = số tham số của hàm hồi quy.

$$F = \frac{MSR}{MSE} \qquad (5.12)$$

$$MSR = \frac{SSR}{p} \qquad (5.13)$$

$$MSE = \frac{SSE}{n - p} \qquad (5.14)$$

Quy tắc quyết định: (1) Sử dụng mức ý nghĩa α (α = 0,05 hoặc 0,01), bác bỏ H_0 khi $F > F\alpha$; trong đó Fα dựa trên phân bố F với 1 độ tự do ở tử số và n - 2 độ tự do ở mẫu số. (2) Ở mức ý nghĩa α (α = 0,05 hoặc 0,01), bác bỏ H_0 khi $P < \alpha$.

Ngày nay việc xây dựng và kiểm định sai lệch của hàm hồi quy tuyến tính với số liệu thực nghiệm được xác định dễ dàng bằng các phần mềm thống kê. Bảng 5.1 và Bảng 5.2 dẫn phân tích biến động (ANOVA) và kết quả báo cáo các thống tin của hàm hồi quy tuyến tính đơn bằng phần mềm Statgraphics.

Bảng 5.1. Phân tích biến động (ANOVA) đối với hàm hồi quy tuyến tính đơn.

Nguồn biến động	SS	Df	MS	F
Model = Regression (Hồi quy)	SSR	2	$MSR = \frac{SSR}{2}$	$\frac{MSR}{MSE}$
Residual Error (Sai số)	SSE	n - 2	$MSE = \frac{SSE}{n - 2}$	
Total (Tổng)	SST	n	MST = SST/n	
Corrected Total (Tổng hiệu chỉnh)	SST(Corr.)	n-1	MST_{Corr}	

Ghi chú: SS = Tổng bình phương; Df = Độ tự do; MS = Trung bình bình phương.

5.2.2. Hồi quy tuyến tính đa biến

5.2.2.1. Hàm hồi quy tuyến tính đa biến

Hàm hồi quy tuyến tính đa biến là hàm của 2 hay nhiều biến độc lập (X_p) để giải thích biến động của biến phụ thuộc (Y). Khi áp dụng vào lâm học, biến Y gọi là biến phản hồi, còn các biến X_p là biến giải thích hay biến dự đoán. Thuật ngữ "Biến phản hồi" biểu thị những thay đổi của cây gỗ và rừng để phản ứng lại những thay

đổi của các nhân tố sinh thái. Các biến X_p được sử dụng để giải thích những nguyên nhân gây ra những thay đổi ở cây gỗ và rừng.

Bảng 5.2. Các hệ số hồi quy và kiểm định sai lệch của hàm hồi quy tuyến tính đơn.

Parameter	Least Squares Estimate	Standard Error	T Statistic	P-Value
(1)	(2)	(3)	(4)	(5)
Intercept	0.107527	0.0924052	1.16364	0.2827
Slope	1.19181	0.0075080	158.723	0.0000

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
(1)	(2)	(3)	(4)	(5)	
Model	162.717	1	162.717	25192.83	0.0000
Residual	0.04521	7	0.00645		
Total (Corr.)	162.762	8			

Cột 4 = (Cột 2/Cột 3); Cột 5 = Hàng 1 của cột 4/Hàng 2 của cột 4.

Mô hình hồi quy đa biến tổng quát có dạng như hàm 5.15; trong đó β_0, β_1, β_2 …β_p = các tham số của mô hình; e = sai số của mô hình.

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \ldots + \beta_pX_p + e \qquad (5.15)$$

Hàm hồi quy đa biến tuyến tính ước lượng có dạng như hàm 5.16.

$$Y_{UL} = b_0 + b_1X_1 + b_2X_2 + \ldots + b_pX_p + e \qquad (5.16)$$

Những giả định cơ bản của hàm hồi quy đa biến tuyến tính:

(a) Sai số e là một biến ngẫu nhiên có phân bố $N(0, S^2)$.

(b) Các biến X_p là độc lập và phương sai bằng nhau.

5.2.2.2. Các bước phân tích hàm hồi quy tuyến tính đa biến

Bước 1. Xác định các thành phần sai lệch của mô hình hồi quy (SST, SSR và SSE) tương tự như công thức 5.5 – 5.8.

Bước 2. Tính hệ số xác định (R^2) và hệ số tương quan (R) tương ứng theo công thức 5.17 và 5.18.

$$R^2 = (1 - \frac{SSE}{SST})*100 \qquad (5.17)$$

$$R = \sqrt{1 - \frac{SSE}{SST}} = \sqrt{R^2} \qquad (5.18)$$

Bước 2. Kiểm định sự tồn tại của hàm hồi quy đa biến tuyến tính bằng thống kê T và thống kê F. Thống kê F kiểm định sự tồn tại của hàm hồi quy. Thống kê F được xác định theo công thức 5.19; trong đó MSR, MSE và sai số chuẩn của ước lượng (SEE) được tính tương ứng theo công thức 5.20, 5.21 và 5.22.

$$F = \frac{MSR}{MSE} \qquad (5.19)$$

$$MSR = \frac{SSR}{p} \quad (5.20)$$

$$S^2 = MSE = \frac{SSE}{n - p} \quad (5.21)$$

$$SEE = \sqrt{MSE} = \sqrt{S^2} \quad (5.22)$$

Giả thuyết H_0: Không tồn tại hàm hồi quy đa biến tuyến tính. Đối thuyết H_a: Tồn tại hàm hồi quy đa biến tuyến tính. Quy tắc quyết định: Bác bỏ H_0 nếu $F > F\alpha$ ($\alpha = 0,05$ hoặc 0,01) hoặc $P < \alpha$; trong đó $F\alpha$ dựa trên phân bố F với p độ tự do ở tử số và n - p - 1 độ tự do ở mẫu số, p là số tham số của hàm hồi quy.

Bước 3. Kiểm định sự tồn tại của các biến độc lập bằng thống kê T. Nếu kiểm định F chỉ ra Y và các X_i tồn tại mối quan hệ thực sự, thì tiếp theo kiểm định từng biến độc lập có ý nghĩa hay không. Mỗi kiểm định T là một kiểm định mức ý nghĩa riêng biệt. Đối với từng biến độc lập, kiểm định T_i được tính theo công thức 5.23; trong đó S_{bi} được tính theo công thức 5.24.

$$T_i = \frac{b_i}{S_{bi}} \quad (5.23)$$

$$S_{bi} = \frac{S}{\sqrt{\sum(X_{ij} - X_{Bqj})^2}} \quad (5.24)$$

Giả thuyết H_0: $\beta_i = 0$. Đối thuyết H_a: $\beta_i \neq 0$. Quy tắc quyết định: Bác bỏ H_0 nếu $T > T_{\alpha/2}$ hoặc $T < - T_{\alpha/2}$ ($\alpha = 0,05$ hoặc 0,01); trong đó $T_{\alpha/2}$ dựa trên phân bố T với n - p độ tự do. Sử dụng mức ý nghĩa P: Bác bỏ H_0 nếu $P < \alpha$.

Bảng 5.3 và Bảng 5.4 dẫn phân tích biến động (ANOVA) và kết quả báo cáo các thống tin của hàm hồi quy đa biến tuyến tính bằng phần mềm Statgraphics.

Bảng 5.3. Phân tích biến động đối với hàm hồi quy đa biến tuyến tính.

Nguồn biến động	SS	Df	MS	F
Model = Regression (Hồi quy)	SSR	p	$MSR = \frac{SSR}{p}$	$\frac{MSR}{MSE}$
Residual Error (Sai số)	SSE	n-p	$MSE = \frac{SSE}{n - p}$	
Total (Tổng)	SST	n	MST	
Corrected Total (Tổng hiệu chỉnh)	$SST_{Corr.}$	n-1	$MST_{Corr.}$	

Bảng 5.4. Kết quả ước lượng các tham số và phân tích biến động đối với hàm hồi quy đa biến tuyến tính: $Y = b_0 + b_1*X_1 + b_2X_2$.

Parameter	Estimate	Standard Error	T Statistic	P-Value
(1)	(2)	(3)	(4)	(5)
CONSTANT	0.028859	0.00500063	5.77107	0.0012
X_1	-3.33119	1.13338	-2.93917	0.0260
X_2	0.000002	8.1963E-8	24.0346	0.0000

Phân tích ANOVA

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
(1)	(2)	(3)	(4)	(5)	(6)
Model	0.0425534	2	0.0212767	501.00	0.0000
Residual	0.0002548	6	0.0000425		
Total	0.0428082	9	0.0047565		
Total (Corr.)	0.0428082	8			

Bước 4. Kiểm định cộng tuyến tính. Giả định của hồi quy tuyến tính đa biến là các biến X_p độc lập với nhau và phương sai bằng nhau. Trong mối quan hệ giữa cây gỗ và rừng với các nhân tố sinh thái, hầu hết các biến độc lập có quan hệ với nhau ở một mức độ nào đó. Chẳng hạn: Đối với hàm M = f(khí hậu), các biến khí hậu (nhiệt độ, mưa, nắng…) lại tồn tại mối quan hệ chặt chẽ với nhau. Khi các biến độc lập có quan hệ chặt chẽ với nhau, thì việc sử dụng các biến độc lập để ước lượng và dự đoán biến phụ thuộc sẽ bị sai lệch. Mặt khác, nếu dựa vào các biến độc lập có quan hệ với nhau, thì nhà lâm học không thể xác định đúng những nguyên nhân chính ảnh hưởng đến cây gỗ và rừng. Nói chung, cộng tuyến tính xuất hiện khi hệ số tương quan giữa hai biến độc lập lớn hơn 0,7. Trong thực hành, việc kiểm định hiện tượng cộng tuyến tính giữa các biến độc lập và chọn những biến độc lập hữu ích nhất để xây dựng hàm hồi quy ước lượng phải trải qua nhiều bước tính toán rất phức tạp. Hiện nay nhiều phần mềm thống kê (SPSS, MINITAB, STATGRAPHICS) cho phép thực hiện các kiểm định cộng tuyến tính và chọn những biến độc lập hữu ích nhất để xây dựng hàm hồi quy ước lượng. Dưới đây giới thiệu các bước phân tích hồi quy và tương quan tuyến tính đa biến bằng phần mềm thống kê SPSS và Statgraphics.

Bước 1. Thăm dò mối quan hệ giữa Y với các X_i và xác định sơ bộ những biến X_i đưa vào mô hình. Mục đích của bước này là xác định tương quan giữa Y với các X_i và tương quan giữa các X_i với nhau thông qua một ma trận tương quan. Ở bước này, các biến X_i được chọn để xây dựng mô hình hồi quy là những biến tồn tại hệ số tương quan cao với biến Y. Nguyên nhân là vì chúng ảnh hưởng lớn đến kết quả ước lượng và dự đoán sự thay đổi của biến Y. Lưu ý rằng, theo nguyên lý sinh thái học và lâm học, anh hưởng của các nhân tố sinh thái đến cây gỗ và rừng là đồng thời và tổng hợp. Theo nguyên lý này, mặc dù một số biến môi trường không tồn tại mối quan hệ chặt chẽ với biến phản hồi (Y), nhưng điều đó không có nghĩa là chúng không ảnh hưởng đến biến Y. Biến phản hồi quan hệ không chặt chẽ với biến giải thích có thể là do nhiều nguyên nhân khác nhau. Chẳng hạn: Chọn sai đối tượng nghiên cứu; dung lượng mẫu nhỏ; đo đạc thiếu chính xác...Trong sinh thái học và lâm học, bước xử lý này chỉ nhằm mục đích xác định những nhân tố sinh thái ảnh hưởng rõ rệt đến sinh trưởng của cây gỗ và rừng.

Bước 2. Phân tích hồi quy tuyến tính đa biến. Mục đích của bước này là nhằm xác định các hệ số hồi quy riêng phần (Partial Regression Coefficients); kiểm định quan hệ tuyến tính giữa biến Y với các biến X_i; xác định các hệ số hồi quy và kiểm định cộng tuyến tính (Collinearity Diagnostics). Tầm quan trọng của từng biến

X_i trong mô hình được xác định bằng phân tích các hệ số tương quan trật tự không (Zero-order Correlations); các hệ số hồi quy chuẩn hóa (Standardized Regression Coefficients); hệ số tương quan từng phần (Part Correlation) và hệ số tương quan riêng phần (Partial Correlation).

Hệ số tương quan trật tự không là hệ số tương quan giữa Y với từng X_i nhưng chưa loại bỏ ảnh hưởng tuyến tính của các biến độc lập khác. Các hệ số tương quan này phản ánh không chính xác tầm quan trọng của từng biến độc lập trong hàm hồi quy tuyến tính đa biến. Hệ số hồi quy chuẩn hóa Beta (Beta Standardized Coefficients) được xây dựng nhằm xác định mức ý nghĩa của từng biến X_i trong hàm hồi quy tuyến tính đa biến. Các hệ số hồi quy chuẩn hóa mô tả hồi quy giữa biến chuẩn hóa Y với các biến chuẩn hóa X_i. Bởi vì các hệ số hồi quy từng phần B_k (mỗi hệ số được điều chỉnh theo các biến độc lập khác) là các hệ số không được chuẩn hóa (các biến được đo đạc bằng các thang đo khác nhau), nên chúng không thể giải thích rõ tầm quan trọng của từng biến trong mô hình. Hệ số chuẩn hóa Beta được tính theo công thức 5.25; trong đó B_k là hệ số hồi quy từng phần của biến k, S_k là độ lệch chuẩn của biến độc lập thứ k, S_y là độ lệch chuẩn của biến Y. Hệ số $Beta_k$ và hệ số B_k đều phụ thuộc vào số lượng biến đưa vào mô hình và quan hệ giữa các biến độc lập. Vì thế, chúng phản ánh vẫn chưa chính xác tầm quan trọng của từng biến số độc lập.

$$Beta_k = B_k(S_k/Sy) \qquad (5.25)$$

Hệ số tương quan từng phần đo mức độ quan hệ giữa Y với biến số độc lập thứ i sau khi đã loại bỏ ảnh hưởng tuyến tính của các biến độc lập khác đối với biến số độc lập thứ i. Hệ số tương quan riêng phần đo mức độ quan hệ giữa biến Y với biến độc lập thứ i sau khi đã loại bỏ ảnh hưởng tuyến tính của các biến độc lập khác đối với cả biến Y lẫn biến độc lập thứ i.

Bình phương của hệ số tương quan từng phần được tính theo công thức 5.26. Bình phương của hệ số tương quan riêng phần được tính theo công thức 5.27. Ở công thức 5.26 và 5.27, R^2 là bình phương của hệ số tương quan đa biến khi mô hình bao gồm tất cả các biến X_i; $R^2_{(i)}$ là bình phương của hệ số tương quan đa biến khi tất cả các biến đều được đưa vào mô hình trừ biến i. Bởi vì mẫu số của $P_{Ri}^2 \leq 1$, nên $P_{Ri}^2 > R^2_{Change}$. Thực chất của việc tính hệ số R^2_{Change} là nhằm xác định biến độc lập thứ i đưa vào mô hình có làm thay đổi đáng kể hệ số tương quan đa biến hay không.

$$R^2_{Change} = R^2 - R^2_i \qquad (5.26)$$

$$P_{Ri}^2 = \frac{R^2 - R^2_i}{1 - R^2_i} \qquad (5.27)$$

Thống kê cộng tuyến tính (Collinearity Statistics) đo mức độ quan hệ tuyến tính giữa các biến độc lập (X_i). Để đo mức độ cộng tuyến tính, người ta sử dụng thống kê độ chấp nhận (Tolerance) và yếu tố khuyếch đại phương sai (Variance Inflation Factor = VIF). Độ chấp nhận của biến X_i được tính theo công thức 5.28; trong đó R^2_i = hệ số tương quan giữa biến độc lập thứ i với các biến độc lập khác.

$$X_i = 1 - R^2_i \qquad (5.28)$$

Khi biến độc lập thứ thứ i có quan hệ yếu (R^2_i nhỏ) với các biến độc lập khác, thì độ chấp nhận cao. Ngược lại, khi biến độc lập thứ thứ i có quan hệ chặt chẽ (R^2_i lớn) với các biến độc lập khác, thì độ chấp nhận thấp. Nói chung, độ chấp nhận của biến thứ i được tính nhằm xác định biến này có ý nghĩa trong mô hình hay không. Tiêu chuẩn độ chấp nhận mặc định là 0,0001. Vì thế, nếu một biến được đưa vào mô hình, thì độ chấp nhận của nó phải lớn hơn 0,0001.

Yếu tố khuyếch đại phương sai của biến X_i (VIF_i) là trị số nghịch đảo của độ chấp nhận (Công thức 5.29). Nói chung, VIF_i là phương sai của hệ số hồi quy thứ i. Khi VIF_i lớn, thì hệ số hồi quy thứ i biến động mạnh. Sự biến động mạnh của hệ số hồi quy thứ i cho biết biến này có ý nghĩa kém trong mô hình.

$$VIF_i = \frac{1}{1 - R^2_i} \qquad (5.29)$$

Kiểm định cộng tuyến tính (Collinearity Diagnostics) còn được xác định theo các số đặc trưng (Eigenvalue), các chỉ số điều kiện (Condition Index = CI) và phần đóng góp trong phương sai (Variance Proportions). Giá trị đặc trưng (Eigenvalue) cho biết số lượng chiều (Dimensions) hay số lượng biến độc lập trong mô hình. Khi giá trị đặc trưng ≅ 0 thì các biến Xi có quan hệ chặt chẽ với nhau. Các chỉ số CI là căn bậc 2 của các tỷ số giữa giá trị đặc trưng lớn nhất ($Eigenvalue_{Max}$) và từng giá trị đặc trưng kế tiếp với từng biến. Nếu chỉ số CI của một biến nào đó lớn hơn 15, thì biến số đó có hiện tượng cộng tuyến tính với những biến khác. Các phần đóng góp trong phương sai (Variance Proportions) là các phần phương sai của những ước lượng được giải thích bởi từng bộ phận có liên quan đến từng trị số Eigenvalue. Nếu một biến số nào đó có hiện tượng cộng tuyến tính với các biến độc lập khác, thì nó đóng góp lớn trong phương sai của các biến khác. Nói chung, khi một biến độc lập có quan hệ tuyến tính với những biến độc lập khác, thì VIP >> 1 và CI >> 15.

Bước 3. Xây dựng mô hình hồi quy đa tuyến tính. Bước này được thực hiện sau khi chọn được những biến dự đoán (X = độc lập, giải thích) có quan hệ chặt chẽ với biến phụ thuộc (Y). Trước khi xây dựng mô hình, chúng ta cần phân tích các biến được đưa vào và loại ra khỏi hàm hồi quy và những thống kê của các biến không được đưa vào mô hình.

(a) Những biến được đưa vào và loại ra khỏi mô hình. Khi có nhiều biến độc lập được chọn để xây dựng mô hình, chúng ta cần biết những biến nào tạo ra một mô hình tốt (sai lệch ước lượng và sai lệch dự đoán đều nhỏ). Những biến đưa vào và loại ra khỏi mô hình được xác định theo phương pháp chọn biến từng bước (Stepwise Method). Mục đích của chọn biến từng bước là nhằm xác định những biến dự đoán hữu ích để đưa vào mô hình. Trước khi xem xét các thủ tục xây dựng mô hình, chúng ta cần phân tích kết quả khi đưa vào hoặc loại bỏ các biến trong mô hình cũng như các thống kê đối với các biến không ở trong mô hình. Trong lâm học, việc làm này là rất quan trọng bởi vì chúng ta cần biết không chỉ vai trò của nhân tố sinh thái chủ yếu, mà còn cả nhân tố sinh thái thứ yếu.

(b) Phân tích một biến được đưa vào hay loại ra khỏi mô hình. Một biến được đưa vào mô hình khi nó làm thay đổi đáng kể hệ số xác định của mô hình tuyến tính đa biến (R^2). Để biết một biến đưa vào mô hình làm thay đổi đáng kể hệ số R^2, người ta tính hệ số tương quan từng phần R^2_{Change}. Giả thuyết Ho: R^2_{Change} = 0. Giả thuyết này được kiểm định bằng thống kê F riêng phần (F_{Change}). Thống kê F_{Change} được tính theo công thức 5.30; trong đó N = dung lượng mẫu, p = số lượng biến độc lập, q = số lượng biến được đưa vào mô hình ở từng bước, R^2 = bình phương của hệ số tương quan giữa biến phụ thuộc với những viến độc lập ở bước trước. Quy tắc quyết định: Bác bỏ H_0 nếu P (Sig. Fchange) < α (α = 0,05 hoặc 0,01). Giả thuyết R^2_{Change} = 0 cũng tương đồng với giả thuyết các tham số b_i = 0. Giả thuyết này được kiểm định bằng thống kê T. Bác bỏ H_0 nếu P > 0,05.

$$F_{Change} = \frac{R^2_{Change}(N - p - 1)}{q(1 - R^2)} \qquad (5.30)$$

(c) Phân tích những thống kê của những biến không ở trong mô hình. Mục đích của việc này là xác định những biến nào sẽ được đưa vào mô hình ở bước tiếp theo. Điều này được xác định bằng các thống kê Beta In, T và hệ số tương quan riêng phần. Thống kê Beta In tương ứng với mỗi biến là hệ số hồi quy chuẩn hóa của mỗi biến khi nó được đưa vào ở bước tiếp theo. Giả thuyết H_0: Beta In = 0. Giả thuyết này được kiểm định bằng thống kê T. Khi mức ý nghĩa của thống kê T nhỏ (P < 0,05 hay 0,01), thì giả thuyết H_0 bị bác bỏ. Khi giả thuyết H_0 bị bác bỏ, thì biến ấy sẽ được đưa vào mô hình ở bước sau.

(d) Xác định số lượng biến dự đoán tối ưu. Về lý thuyết, khi đưa vào mô hình càng nhiều biến, thì hệ số xác định (R^2) càng cao. Mặt khác, hệ số R^2 sẽ không bao giờ giảm khi các biến dự đoán tiếp tục được đưa vào mô hình. Nhưng một mô hình với nhiều biến dự đoán không có nghĩa là nó sẽ phù hợp hơn với tổng thể. Một mô hình nhiều biến có thể dẫn đến làm tăng sai số chuẩn của ước lượng và kết quả dự đoán trệch. Vì thế, xác định những biến dự đoán tối ưu là cần thiết. Để đạt được mục đích này, người ta sử dụng hệ số R^2 điều chỉnh (Adjusted $R^2 = R^2_a$) (Công thức 5.31); trong đó n = kích thước mẫu hay số cặp biến Y và X, p = số biến dự đoán đưa vào mô hình. Từ công thức 5.31 cho thấy, khi p tăng lên thì R^2_a giảm xuống. Hệ số R^2_a không nhất thiết tăng lên khi đưa thêm một biến vào mô hình. Vì thế, R^2_a là thước đo mức độ phù hợp của các biến độc lập trong mô hình.

$$R^2_a = 1 - (1 - R^2)*\frac{n - 1}{n - p - 1} \qquad (5.31)$$

Bước 4. Xác định phương pháp chọn biến. Trong phân tích hồi quy tuyến tính đa biến, các biến có thể được chọn theo ba phương pháp.

(a) Phương pháp loại bỏ dần từng biến (Backward selection). Phương pháp này khởi đầu với tất cả các biến được đưa vào mô hình. Sau đó, chương trình sẽ tuần tự loại bỏ dần từng biến thông qua tiêu chuẩn loại trừ (Removal Criteria). Đây là phương pháp mặc định của chương trình. Để loại trừ dần từng biến, chương trình

sử dụng hai tiêu chuẩn F_{Ra} (F to Remove) và xác suất tối đa của F_{Ra} (Probability of F to remove).

(a1) Tiêu chuẩn F_{Ra} (F to Remove). Một biến được để lại trong mô hình khi $F_{Ra} > F_{Ra}$ tiêu chuẩn (F_{Min}). Ngược lại, một biến bị loại ra khỏi mô hình khi $F_{Ra} < F_{Min}$. Giá trị F_{Min} = 2,71. Giá trị F_{Min} có thể được thay đổi theo yêu cầu của nhà nghiên cứu.

(a2) Tiêu chuẩn xác suất tối đa của F_{Ra} (Probability of F to remove). Một biến được để lại trong mô hình khi xác suất của F tính phải nhỏ hơn xác suất tiêu chuẩn và ngược lại. Giá trị xác suất tiêu chuẩn mặc định là 0,10.

(b) Phương pháp đưa dần từng biến vào (Forward Selection). Phương pháp này khởi đầu với mô hình không có biến nào. Sau đó chọn dần từng biến đưa vào mô hình. Ở bước thứ nhất, biến dự đoán đầu tiên là biến có hệ số tương quan cao nhất. Ở những bước tiếp theo, biến nào có hệ số tương quan cục bộ cao nhất sẽ được đưa dần vào. Lưu ý rằng một biến X_i được đưa vào mô hình khi nó làm cho hệ số R^2 thay đổi đáng kể hoặc hệ số $b_i \neq 0$. Giả thuyết $b \neq 0$ được kiểm định theo thống kê F. Thống kê F được xác định theo hai tiêu chuẩn $F_{Vào}$ (F to enter) và xác xuất của $F_{Vào}$ (Probability of F to enter).

(b1) Tiêu chuẩn $F_{Vào}$ (F to enter). Giá trị mặc định của F to enter là 3,84. Vì thế, một biến được đưa vào khi $F_{Tính} > 3,84$. Khi một biến không thoả mãn điều kiện này, thì biến ấy sẽ bị loại ra ở bước sau. Giá trị $F_{Vào}$ có thể thay đổi theo yêu cầu của nhà nghiên cứu.

(b2) Tiêu chuẩn xác xuất của $F_{Vào}$ (Probability of F to enter). Xác xuất mặc định của $F_{Vào}$ = 0,05. Vì thế, khi một biến được đưa vào mô hình thì P < 0,05 (mặc định). Giá trị P có thể thay đổi theo yêu cầu của nhà nghiên cứu. Chẳng hạn: P = 0,01 - 0,10.

Nếu những điều kiện trên đây được thoả mãn, thì những bước tiếp theo sẽ được thực hiện tiếp. Ngược lại, khi những điều kiện này không được thoả mãn, thì việc chọn biến sẽ dừng.

Lưu ý rằng chương trình không báo cáo kết quả của thống kê F mà đưa ra thống kê T. Hai thống kê này là tương đồng với nhau vì $T^2 = F$. Khi chọn tiêu chuẩn $F_{Vào}$, thì kết quả có thể khác với việc chọn xác xuất của $F_{Vào}$. Điều này xảy ra là vì, khi đưa các biến vào mô hình, số bậc tự do của tổng bình phương phần dư (SS) sẽ giảm xuống, nhưng bậc tự do của hồi quy lại tăng lên. Vì thế, khi cho một giá trị F cố định, thì mức ý nghĩa sẽ thay đổi theo số lượng biến đưa vào. Khi dung lượng quan sát lớn, thì sự khác biệt này sẽ giảm.

(c) Phương pháp chọn biến từng bước (Stepwise selection). Phương pháp này là sự kết hợp giữa hai phương pháp đưa dần vào và loại bỏ dần từng biến độc lập. Đây là phương pháp thông dụng nhất. Trước hết, chương trình bắt đầu chọn biến thứ 1. Khi biến thứ 1 thỏa điều kiện vào ($F_{Vào}$ = FIN và P = PIN), thì chương trình sẽ tiếp tục chọn biến thứ 2. Biến thứ 2 được chọn là biến có hệ số tương quan

riêng phần cao nhất. Khác với thủ tục đưa dần vào từng biến, thủ tục chọn biến từng bước xem xét có nên loại bỏ biến thứ nhất hay không dựa theo tiêu chuẩn F_{Ra} (F to remove) hoặc xác suất của F_{Ra}. Cách thức này giống như thủ tục loại trừ dần từng biến. Ở những bước kế tiếp, những biến nào không ở trong mô hình sẽ được xem xét để đưa dần vào. Nếu biến đưa vào mô hình ở bước sau tốt hơn, thì các biến ở trong mô hình trước đó được xem xét lại để loại ra. Khi không còn biến nào thỏa điều kiện loại ra, thì chương trình sẽ dừng.

Trong phương pháp chọn biến từng bước, một biến giải thích có thể được đưa vào và loại ra nhiều lần. Mặt khác, ý nghĩa của mỗi biến giải thích phụ thuộc vào thứ tự mà nó được đưa vào mô hình. Để tránh hiện tượng này, xác suất vào phải nhỏ hơn xác suất ra hay $F_{IN} > F_{OUT}$. Thủ tục sẽ dừng khi không còn biến giải thích nào thỏa mãn điều kiện vào và ra.

Trên đây là lý thuyết của phương pháp chọn biến từng bước trong phân tích hồi quy tuyến tính đa biến. Khi áp dụng phương pháp này để xây dựng các hàm phản hồi của cây gỗ và rừng, nhà lâm học cần lưu ý một số vấn đề dưới đây.

(1) Xác định thứ tự ảnh hưởng của các biến giải thích. Bởi vì ý nghĩa của mỗi biến giải thích phụ thuộc vào thứ tự mà nó được đưa vào mô hình, nên việc chọn biến nào đưa vào trước và sau phải thay đổi tùy theo mục tiêu nghiên cứu. Chẳng hạn: Mục tiêu nghiên cứu là xác định những tháng mà nhiệt độ không khí ảnh hưởng lớn đến tăng trưởng bề rộng vòng năm của Thông ba lá (*Pinus kesiya* Royle ex Gordon). Trong trường hợp này, nhà lâm học phải sử dụng phương pháp chọn biến từng bước để lọc ra các tháng theo thứ tự từ ảnh hưởng thấp đến ảnh hưởng lớn hoặc ngược lại. Trái lại, mục tiêu nghiên cứu là xây dựng hàm phản hồi để giải thích ảnh hưởng của nhiệt độ không khí trong năm đến tăng trưởng bề rộng vòng năm của Thông ba lá. Giả sử nhiệt độ không khí tháng 2 và tháng 4 ảnh hưởng tương tự như nhau đến tăng trưởng bề rộng vòng năm của Thông ba lá. Nếu sử dụng phương pháp chọn biến từng bước và đưa nhiệt độ không khí tháng 4 vào trước tháng 2, thì hàm phản hồi không bao gồm tháng 2. Điều đó xảy ra là vì chương trình ưu tiên chọn biến đưa vào trước. Cách chọn biến này là không hợp lý bởi vì cả hai nhân tố này đều ảnh hưởng như nhau đến tăng trưởng bề rộng vòng năm của Thông ba lá. Vì thế, hàm phản hồi phải bao gồm cả nhiệt độ không khí tháng 2 và tháng 4. Khi sử dụng hàm để ước lượng và dự đoán tăng trưởng bề rộng vòng năm của Thông ba lá, thì thứ tự từng biến là không quan trọng.

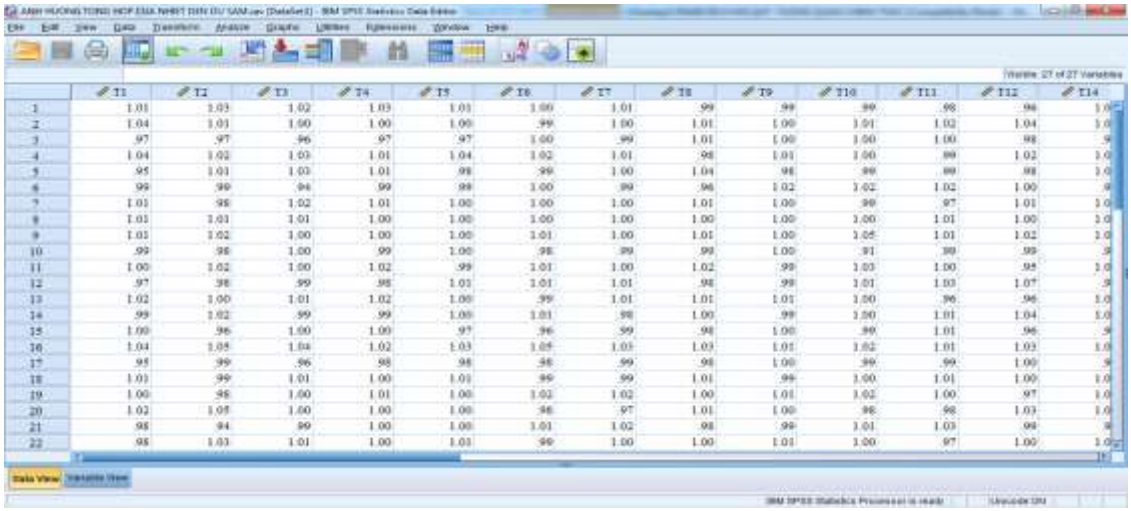
(2) Chọn cách báo cáo kết quả trong phân phân tích hồi quy tuyến tính đa biến. Các phần mềm thống kê cho phép người sử dụng chọn lựa hai cách báo cáo kết quả phân tích hàm hồi quy tuyến tính đa biến. Cách thứ nhất chỉ báo cáo tóm tắt kết quả phân tích hàm hồi quy thích hợp nhất. Ưu điểm của cách báo cáo này là giúp cho nhà lâm học thu được những hàm phản hồi thích hợp để sử dụng vào việc ước lượng và dự đoán biến phản hồi từ những biến dự đoán. Cách thứ hai báo cáo chi tiết các thông tin của từng bước phân tích hàm hồi quy. Điều này là rất hữu ích, bởi vì những thông tin này giúp ích cho nhà lâm học hiểu rõ về vai trò của từng biến

giải thích trong hàm phản hồi. Vì thế, tùy theo mục đích đặt ra, nhà lâm học có thể chọn cách báo cáo nào.

5.2.2.3. Ứng dụng phân tích hồi quy trong lâm học

Cây gỗ và rừng chịu ảnh hưởng tổng hợp của nhiều nhân tố sinh thái khác nhau; trong đó một số nhân tố sinh thái giữ vai trò chủ đạo, còn một số nhân tố sinh thái khác chỉ giữ vai trò thứ yếu. Để xác định rõ vai trò của những nhân sinh thái này, lâm học sử dụng phương pháp phân tích hồi quy tuyến tính. Hồi quy tuyến tính đơn được sử dụng để phân tích phản hồi của cây gỗ và rừng với một biến giải thích. Hồi quy tuyến tính đa biến được sử dụng để phân tích phản hồi của cây gỗ và rừng với hai hay nhiều biến giải thích. Dưới đây dẫn nghiên cứu của Nguyễn Văn Nhẫn (2018) về ảnh hưởng của nhiệt độ không khí (T^0C) đến tăng trưởng bề rộng vòng năm (ZD, mm) của Du sam (*Keteleeria evelyniana* Master) tại khu vực Đức Trọng thuộc tỉnh Lâm Đồng. Mục tiêu của nghiên cứu này là xác định những tháng mà nhiệt độ không khí ảnh hưởng rõ rệt đến tăng trưởng bề rộng vòng năm của Du sam. Để loại bỏ ảnh hưởng của tuổi cây (A, năm) và biến động nhiệt độ theo năm, các giá trị ZD và nhiệt độ không khí của các tháng trong năm được chuyển tương ứng thành chỉ số bề rộng vòng năm (Kd) và chỉ số nhiệt độ (T) theo phương pháp bình quân di động 3 năm. Chuỗi số liệu Kd và T được thu thập 33 năm từ 1980 – 2013 (Phụ lục 3 và 4). Để xác định mối quan hệ giữa chỉ số Kd với các chỉ số T của 12 tháng trong năm (T_{1-12}) bằng phần mềm SPSS, các bước thực hiện như dưới đây.

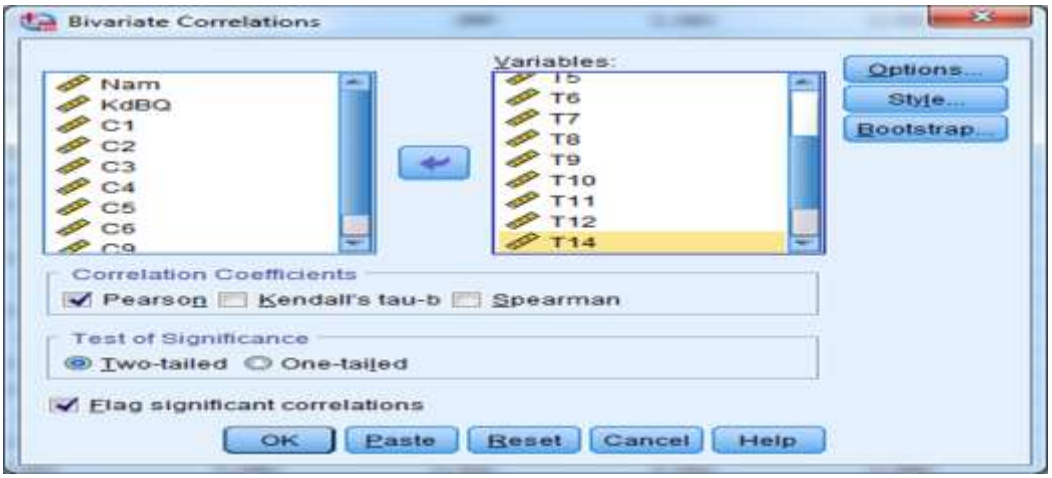
Bước 1. Ghi các biến Kd và T_1 – T_{12} từ năm 1980 – 2013 vào 13 cột khác nhau (Hình 5.4).



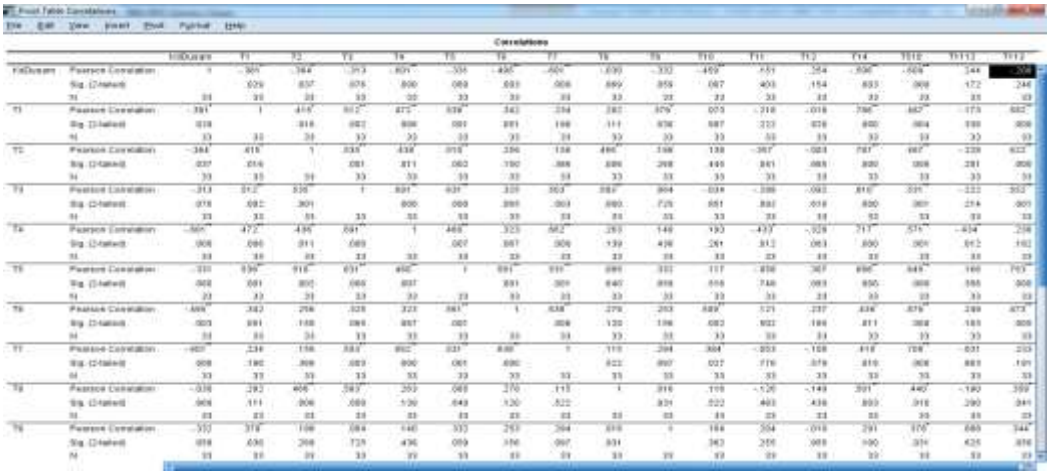
Hình 5.4. Chỉ số Kd của Du sam và chỉ số T_1 – T_{12} từ 1980 - 2013.

Bước 2. Phân tích tương quan giữa chỉ số Kd với các chỉ số T_1 – T_{12}. Bước này được đặt ra nhằm xác định khuynh hướng và cường độ của mối quan hệ Kd = f(T_i). Để đạt mục đích này, trước hết vào đường dẫn: Analyze > Correlate > Bivariate. Đến đây xuất hiện cửa sổ Bivariate Correlations. Tại cửa sổ này, chọn Kd

và $T_1 – T_{12}$ rồi chuyển chúng vào ô Variables (Hình 5.5). Sau đó chọn OK để nhận ma trận tương quan Kd = f(T_{1-12}) (Hình 5.6).



Hình 5.5. Số liệu để xác định tương quan Kd = f(T_{1-12}).



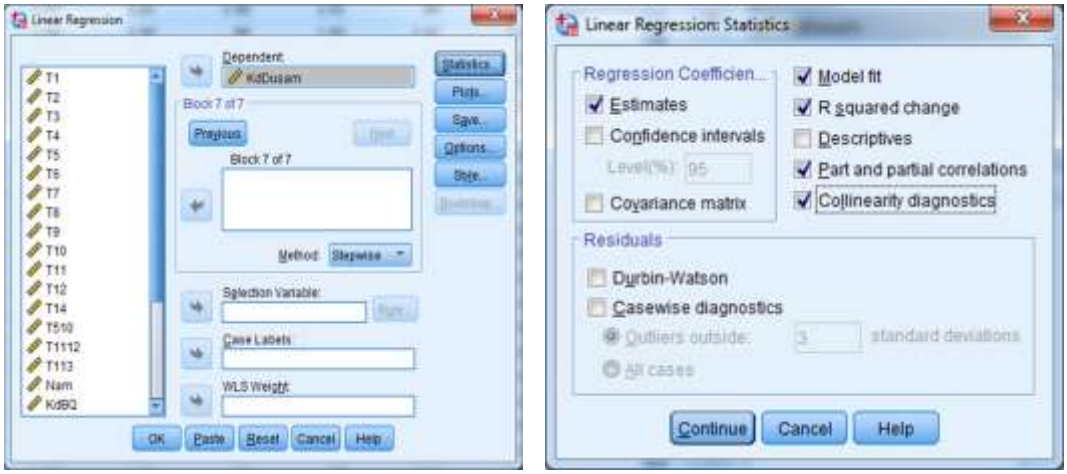
Hình 5.6. Ma trận tương quan Kd = f(T_{1-12}).

Bảng 5.5 tóm tắt kết quả phân tích mối quan hệ giữa chỉ số Kd của Du sam với T không khí của các tháng trong năm. Ở Bảng 5.5, r là hệ số tương quan, P là mức ý nghĩa thống kê, N là số năm nghiên cứu, Tháng (1 – 4), (5 – 10), (11 – 12) và (11 – 3) là tổ hợp nhiệt độ trung bình của các tháng. Từ số liệu ở Bảng 5.5 cho thấy, nhiệt độ không khí nâng cao vào các tháng 1 – 10 và tháng 11 năm trước đến tháng 3 năm sau đều dẫn đến sự suy giảm (hệ số r âm) tăng trưởng bề rộng vòng năm của Du sam. Trái lại, sự gia tăng nhiệt độ không khí vào tháng 11 và 12 lại dẫn đến sự nâng cao (hệ số r dương) tăng trưởng bề rộng vòng năm của Du sam. Tuy vậy, chỉ số Kd chỉ biểu hiện mối quan hệ rõ rệt với T_1 (r = -0,381; P = 0,029), T_2 (r = -0,364; P = 0,037), T_4 (r = -0,601; P = 0,000), T_6 (r = -0,495; P = 0,003), T_7 (r = -0,601; P < 0,001), T_{10} (r = -0,459; P = 0,007), T_{1-4} (r = -0,506; P = 0,003) và T_{5-10} (r = -0,609; P < 0,001).

Bảng 5.5. Mối quan hệ giữa tăng trưởng bề rộng vòng năm của Du sam với nhiệt độ không khí của các tháng trong năm.

Tháng	r	P	N	Tháng	r	P	N
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	-0,381	0,029	33	9	-0,332	0,059	33
2	-0,364	0,037	33	10	-0,459	0,007	33
3	-0,313	0,076	33	11	0,151	0,403	33
4	-0,601	0,000	33	12	0,254	0,154	33
5	-0,331	0,060	33	1 - 4	-0,506	0,003	33
6	-0,495	0,003	33	5 - 10	-0,609	0,000	33
7	-0,601	0,000	33	11 - 12	0,244	0,172	33
8	-0,030	0,869	33	11 - 3	-0,208	0,246	33

Bước 3. Phân tích vai trò của nhiệt độ không khí đối với tăng trưởng bề rộng vòng năm của Du sam. Kết quả phân tích ở bước 2 chỉ cho biết khuynh hướng (dấu của hệ số tương quan r) và cường độ (độ lớn của hệ số tương quan r) của mối quan hệ giữa chỉ số Kd với T. Để xác định rõ vai trò của 6 nhân tố (T_1, T_2, T_4, T_6, T_7 và T_{10}) đối với tăng trưởng bề rộng vòng năm của Du sam, chúng ta phải sử dụng kỹ thuật phân tích hồi quy tương quan riêng phần. Để đạt mục đích này, trước hết vào đường dẫn: Analyze > Regression Linear. Đến đây xuất hiện cửa sổ hồi quy tuyến tính (Linear Regression). Tại cửa sổ này, chọn Kd và chuyển vào ô Dependent. Kế đến chọn lần lượt T_1, T_2, T_4, T_6, T_7 và T_{10} và chuyển chúng vào ô Independent(s). Lưu ý rằng, sau khi đưa một biến T_i vào ô Independent(s), thì phải chọn Next và tiếp tục đưa vào biến khác. Sau khi đưa hết các biến giải thích vào đó ô Independent(s), chọn Stepwise và Statistic và đánh dấu các thống kê như Hình 5.7. Ở đây sử dụng tiêu chuẩn $F_{Vào}$ = 3,84 và F_{ra} = 2,71. Cuối cùng chọn OK để nhận kết quả phân tích hồi quy tương quan riêng phần (Bảng 5.6).



Hình 5.7. Kỹ thuật phân tích hồi quy tương quan riêng phần.

Bảng 5.6. Phân tích vai trò của nhiệt độ không khí đối với tăng trưởng bề rộng vòng năm của Du sam.

(5.6a) Thống kê sự thay đổi của các biến T trong 6 mô hình(*).

Model	R	SEE	Change Statistics				
			R Square Change	F Change	df1	df2	Sig. F Change
1	0,381[a]	0,07866	0,145	5,253	1	31	0,029
2	0,443[b]	0,07753	0,051	1,915	1	30	0,177
3	0,617[c]	0,06918	0,185	8,675	1	29	0,006
4	0,684[d]	0,06529	0,087	4,561	1	28	0,042
5	0,699[e]	0,06517	0,021	1,106	1	27	0,302
6	0,734[f]	0,06306	0,050	2,839	1	26	0,104

(*) Mô hình 1: $Kd = a_0 + a_1T_1$.
Mô hình 2: $Kd = a_0 + a_1T_1 + a_2T_2$.
Mô hình 3: $Kd = a_0 + a_1T_1 + a_2T_2 + a_3T_4$
Mô hình 4: $Kd = a_0 + a_1T_1 + a_2T_2 + a_3T_4 + a_4T_6$
Mô hình 5: $Kd = a_0 + a_1T_1 + a_2T_2 + a_3T_4 + a_4T_6 + a_5T_7$
Mô hình 6: $Kd = a_0 + a_1T_1 + a_2T_2 + a_3T_4 + a_4T_6 + a_5T_7 + a_6T_{10}$.

(5.6b) Ý nghĩa của từng biến T trong mô hình.

Model		Standardized Coefficients Beta	T	Sig.	Collinearity Statistics Tolerance	VIF
	(Constant)		6,516	0,000		
	T_1	-0,114	-0,679	0,503	0,633	1,580
	T_2	-0,112	-0,697	0,492	0,682	1,466
6	T_4	-0,269	-1,182	0,248	0,342	2,924
	T_6	-0,049	-0,232	0,819	0,401	2,492
	T_7	-0,246	-1,001	0,326	0,294	3,404
	T_{10}	-0,264	-1,685	0,104	0,720	1,390

Từ số liệu ở Bảng 5.6a cho thấy, ba biến (T_1, T_4 và T_6) đưa vào mô hình Kd = f(T_i) là có ý nghĩa lớn ($P < 0,05$). Trái lại, ba biến (T_2, T_7 và T_{10}) đưa vào mô hình Kd = f(T_i) là kém ý nghĩa ($P > 0,05$). Để đơn giản, mô hình Kd = f(T_i) có thể chỉ bao gồm ba biến (T_1, T_4 và T_6). Nếu mô hình bao gồm tất cả 6 biến (T_1, T_2, T_4, T_6, T_7 và T_{10}) (Bảng 5.6b), thì tăng trưởng bề rộng vòng năm của Du sam phụ thuộc nhiều nhất vào T_4 (hệ số hồi quy chuẩn hóa |-0,269|); kế đến là T_{10} (hệ số hồi quy chuẩn hóa |-0,264|) và T_7 (hệ số hồi quy chuẩn hóa |0,246|); thấp nhất là T_6 (hệ số hồi quy chuẩn hóa |-0,049|. Nói chung, sự nâng cao nhiệt độ không khí có ảnh hưởng xấu đến tăng trưởng bề rộng vòng năm của Du sam.

5.3. Hàm hồi quy bậc hai và ứng dụng trong lâm học

5.3.1. Khái quát về hàm hồi quy bậc hai

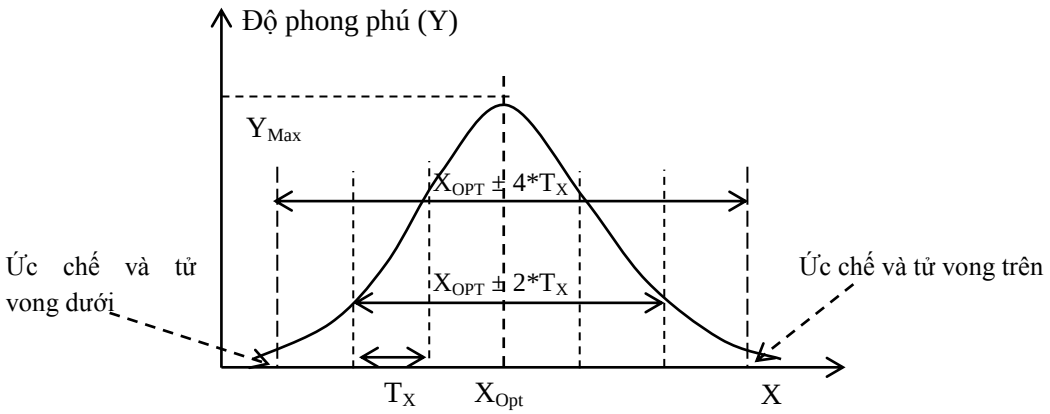
Hàm hồi quy bậc 2 có dạng như hàm 5.32; trong đó Y = biến phản hồi kỳ vọng; X = biến giải thích; b_0, b_1 và b_2 = các tham số; e = sai số. Các tham số của hàm bậc 2 được xác định bằng phương pháp bình phương nhỏ nhất.

$$Y = b_0 + b_1X + b_2X^2 + e \qquad (5.32)$$

Khi hệ số của X^2 mang giá trị âm, thì đồ thị của hàm bậc 2 có dạng hình chuông với 1 đỉnh nằm trên trục hoành. Khi hệ số của X^2 mang giá trị dương, thì đồ thị của hàm bậc 2 có dạng hình chuông với 1 đỉnh nằm dưới trục hoành.

5.3.2. Ứng dụng hàm hồi quy bậc hai trong lâm học

Theo định luật Shelfords (1913; Dẫn theo Nguyễn Văn Thêm, 2002), khi các nhân tố sinh thái khác là đồng nhất, thì phản hồi của cây gỗ và rừng đối với sự thay đổi của một nhân tố sinh thái tuân theo hàm bặc 2. Về mặt toán học, khi một đại lượng có phân bố chuẩn, thì xác suất bặt gặp các giá trị nằm trong khoảng $X_{Bq} \pm S$, $X_{Bq} \pm 2*S$ và $X_{Bq} \pm 3*S$ (với X_{Bq} = giá trị bình quân, S = sai tiêu chuẩn) tương ứng là 68%, 95% và 99%. Theo quy luật này, các nhà sinh thái học và lâm học xem các giá trị của một nhân tố sinh thái phân bố trong ba khoảng $X_{Bq} \pm S$, $X_{Bq} \pm 2*S$ và $X_{Bq} \pm 4*S$ tương ứng là vùng tối ưu sinh thái, biên độ sinh thái (Vùng hoạt động bình thường) và phạm vi chống chịu (Vùng sống) của cây gỗ và rừng. Giá trị trung tâm (X_{Bq}) của vùng tối ưu sinh thái là tối ưu sinh thái (Hình 5.8). Tối ưu sinh thái là giá trị của biến giải thích đảm bảo cho cây gỗ và rừng sống tốt nhất. Biên độ sinh thái là phạm vi của biến giải thích đảm bảo cho cây gỗ và rừng sống bình thường. Phạm vi chống chịu là ranh giới của biến giải thích đảm bảo cho cây gỗ và rừng có thể sống. Các giá trị $X < (X_{Bq} - 4*S)$ và $X > (X_{Bq} + 4*S)$ là vùng bị ức chế và tử vong. Trong hai vùng này, xác suất bắt gặp cây gỗ và rừng còn sống là dưới 0,1%. Độ phong phú lớn nhất (Y_{Max}) là giá trị của biến phản hồi cao nhất tại giá trị của biến giải thích tối ưu (X_{Opt}).



Hình 5.8. Đường cong phản hồi của cây gỗ và rừng với môi trường.

Theo Jongman và ctv (1987), khi khảo sát các hàm phản hồi bậc 2, chúng ta có thể xác định được các phản hồi (Y_i) của cây gỗ và rừng, tối ưu sinh thái (Công thức 5.33), vùng tối ưu sinh thái (Công thức 5.34), vùng tác động bình thường của nhân tố sinh thái (Công thức 5.35), phạm vi chống chịu (Công thức 5.36) và phản hồi lớn nhất (Công thức 5.37). Ở các công thức 5.33 - 5.36, giá trị $\pm T_X$ là biên độ sinh thái.

$$X_{OPT} = \frac{b_1}{2b_2} \qquad (5.33)$$

$X_{68\%} = X_{OPT} \pm T_X$ (5.34)

$X_{95\%} = X_{OPT} \pm 2*T_X$ (5.35)

$X_{99,9\%} = X_{OPT} \pm 4*T_X$ (5.36)

với $T_X = \frac{1}{\sqrt{2b_2}}$

$Y_{Max} = b_0 + b_1X_{OPT} - b_2X_{OPT}^2$ (5.37)

Giả thiết phân tích ảnh hưởng của hàm lượng mùn (X_1%) và pH_{H2O} (X_2) ở tầng đất mặt (0 – 30 cm) của đất Bazan nâu đỏ đến sinh trưởng chiều cao (H, m) của Dầu song nàng 2 năm tuổi. Số liệu về H và hai biến X_1 và X_2 được ghi lại ở Phụ lục 5. Trong trường hợp này, chúng ta có thể xây dựng 3 hàm phản hồi khác nhau.

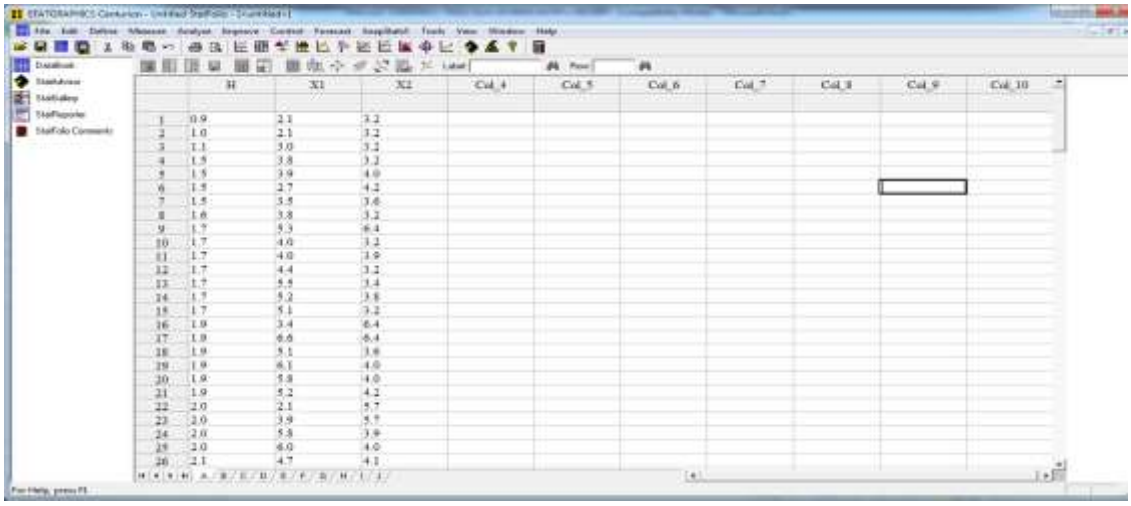
Hàm 1: $H = a_0 + a_1X_1 - a_2X_1^2$. Hàm này cho phép xác định phản hồi của Dầu song nàng với hàm lượng mùn trong đất.

Hàm 2: $H = b_0 + b_1X_2 - b_2X_2^2$. Hàm này cho phép xác định phản hồi của Dầu song nàng với trạng thái pH_{H2O} trong dung dịch đất.

Hàm 3: $H = c_0 + c_1X_1 - c_2X_1^2 + c_3X_2 - c_4X_2^2 + c_5X_1*X_2$. Hàm này cho phép xác định phản hồi của Dầu song nàng với cả hai nhân tố X_1 và X_2. Nói cách khác, hàm này cho phép xác định ảnh hưởng tổng hợp của hai nhân tố X_1 và X_2 đến sinh trưởng của Dầu song nàng.

Trong thực hành, chúng ta có thể xây dựng ba hàm phản hồi trên đây bằng phần mềm Statgraphics. Các bước xây dựng hàm phần hồi như dưới đây.

Bước 1. Ghi ba biến H, X_1 và X_2 vào ba cột khác nhau (Hình 5.9).



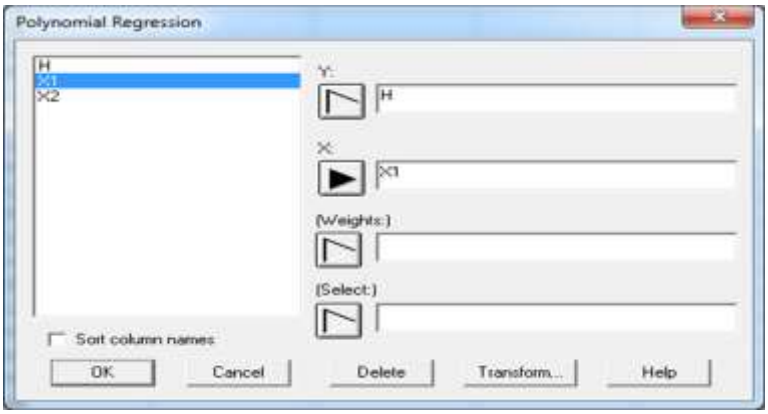
Hình 5.9. Số liệu để xây dựng hàm phản hồi $H = f(X_1, X_2)$.

Bước 2. Xây dựng hàm phản hồi $H = a_0 + a_1X_1 - a_2X_1^2$. Trước hết vào đường dẫn: Improve > Regression Analysis > One Factor > Polynorminal Regression. Đến đây một cửa sổ xuất hiện. Tại cửa sổ này, chọn H và X_1 rồi chuyển chúng vào ô Y và ô X (Hình 5.10).

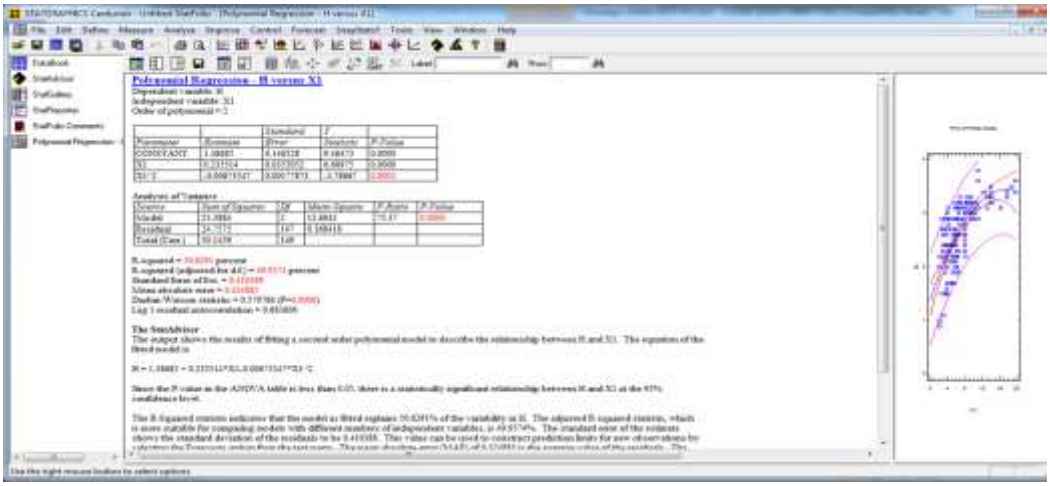
Bước 3. Xác định các thông tin của hàm phản hồi $H = a_0 + a_1X_1 - a_2X_1^2$. Để đạt được mục đích này, chọn OK ở Hình 5.10. Đến đây chương trình báo cáo kết quả phân tích hồi quy $H = f(X_1)$ (Hình 5.11).

Bước 4. Xây dựng hàm phản hồi $H = b_0 + b_1X_2 - b_2X_2^2$. Các bước xây dựng hàm phản hồi này cũng được thực hiện tương tự như hàm phản hồi $H = f(X_1)$.

Bước 5. Xây dựng hàm phản hồi $H = c_0 + c_1X_1 - c_2X_1^2 + c_3X_2 - c_4X_2^2 + c_5X_1*X_2$. Ở các bước 1 – 4, chúng ta chỉ xây dựng hàm phản hồi riêng rẽ với từng nhân tố X_1 và X_2. Theo lý thuyết lâm học, cây gỗ và rừng chịu ảnh hưởng đồng thời và tổng hợp của nhiều nhân tố sinh thái. Mặt khác, sự tương tác giữa các nhân tố sinh thái cũng ảnh hưởng đến cây gỗ và rừng. Vì thế, để phân tích ảnh hưởng của từng nhân tố X_1 và X_2 và ảnh hưởng tổng hợp của hai nhân tố này, chúng ta cần phải xây dựng hàm $H = f(X_1, X_2, X_1*X_2)$ dưới dạng hàm 5.38. Hàm này cũng có thể được mở rộng cho 3 hay nhiều nhân tố giải thích khác.



Hình 5.10. Xác định hàm phản hồi $H = f(X_1)$.

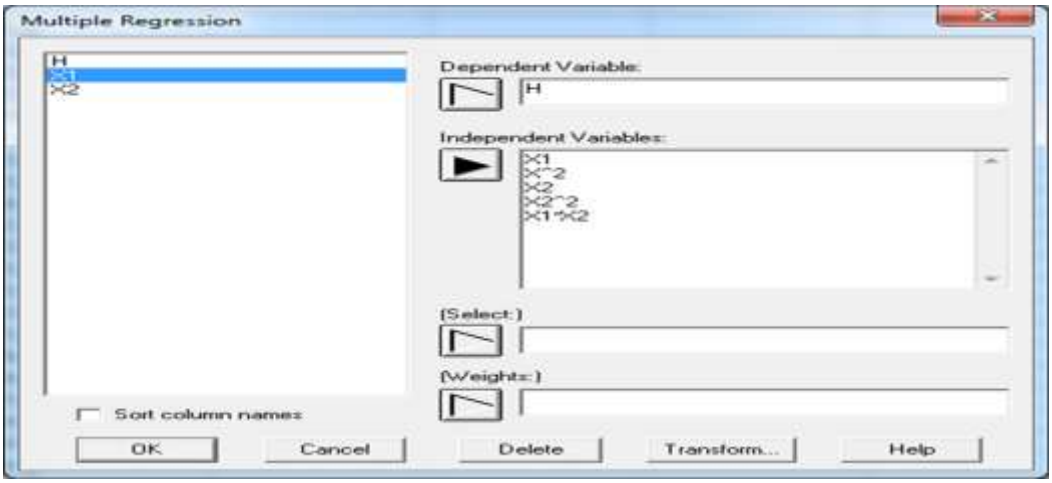


Hình 5.11. Kết quả phân tích hàm phản hồi $H = f(X_1)$.

$$H = c_0 + c_1X_1 - c_2X_1^2 + c_3X_2 - c_4X_2^2 + c_5X_1*X_2 \qquad (5.38)$$

Để xây dựng hàm 5.38, trước hết vào đường dẫn: Improve > Regression Analysis > Multiple Factor > Multiple Regression. Đến đây một cửa sổ xuất hiện. Tại cửa sổ này, chọn H, X_1, X_2 và X_1*X_2 và chuyển chúng vào ô Y và ô X (Hình 5.12). Sau đó chọn OK để nhận kết quả phân tích hồi quy H = f(X_1, X_2, X_1*X_2) (Hình 5.13).

Lưu ý rằng, khi cần phân tích từng nhân tố, thì chúng ta thực hiện riêng rẽ các bước xử lý từ 1 - 4. Nếu chỉ cần tính toán các tham số sinh thái (T_X, X_{Opt}, $X_{95\%}$, $X_{99,9\%}$), H_{Max} và phân tích ảnh hưởng tổng hợp của hai hay nhiều nhân tố sinh thái, thì chúng ta chỉ cần thực hiện bước 1 và bước 5.



Hình 5.12. Xác định hàm phản hồi H = f(X_1, X_2, X_1*X_2).

Bước 6. Phân tích các tham số sinh thái đối với biến X_1 và X_2. Trước hết Copy hàm phản hồi ở Hình 5.13 và chuyển vào File Word. Kế đến Copy kết quả hàm phản hồi ở File Word và chuyển vào File Excel. Tại đây, tính toán T_X, X_{Opt}, $X_{95\%}$, $X_{99,9\%}$ và H_{Max} theo các công thức 5.33 - 5.37.

Hình 5.13. Kết quả phân tích hàm phản hồi H = f(X_1, X_2, X_1*X_2).

Bước 7. Báo cáo và phân tích các hàm phản hồi. Phần này có thể báo cáo riêng rẽ theo từng nhân tố sinh thái hoặc báo cáo chung cho nhiều nhân tố sinh thái. Bảng 5.7 dẫn báo cáo tóm tắt các kết quả phân tích hàm phản hồi $H = f(X_1, X_2, X_1*X_2)$. Kết quả ở Bảng 5.7 cho thấy hàm phản hồi $H = f(X_1, X_2, X_1X_2)$ tồn tại thực sự ($P < 0,01$). Các tham số $c_0 – c_5$ cũng tồn tại thực sự ($P < 0,01$). Vì thế, hàm $H = f(X_1, X_2, X_1X_2)$ được sử dụng để xác định các tham số sinh thái đối với cây con Dầu song nàng 2 tuổi và ước lượng H theo hai biến X_1 và X_2 (Bảng 5.8; Hình 5.14 và 5.15).

Bảng 5.7. Hàm phản hồi của cây con Dầu song nàng 2 năm tuổi đối với hàm lượng mùn (X_1%) và pH_{H2O} (X_2) ở tầng đất mặt (0 – 30 cm) của đất Bazan.

Tham số	Ước lượng	SEE	T	P
(1)	(2)	(3)	(4)	(5)
Hằng số	-5,12170	0,87445	-5,85703	0,000
X_1	0,36271	0,06398	5,66920	0,000
X_1^2	-0,00561	0,00155	-3,61387	0,000
X_2	2,55110	0,37195	6,85869	0,000
X_2^2	-0,22781	0,03798	-5,99882	0,000
X_1*X_2	-0,03666	0,01184	-3,09743	0,002
R^2	66,2%			
SEE	0,34			
MAE	0,27			
P-Value	<0,001			

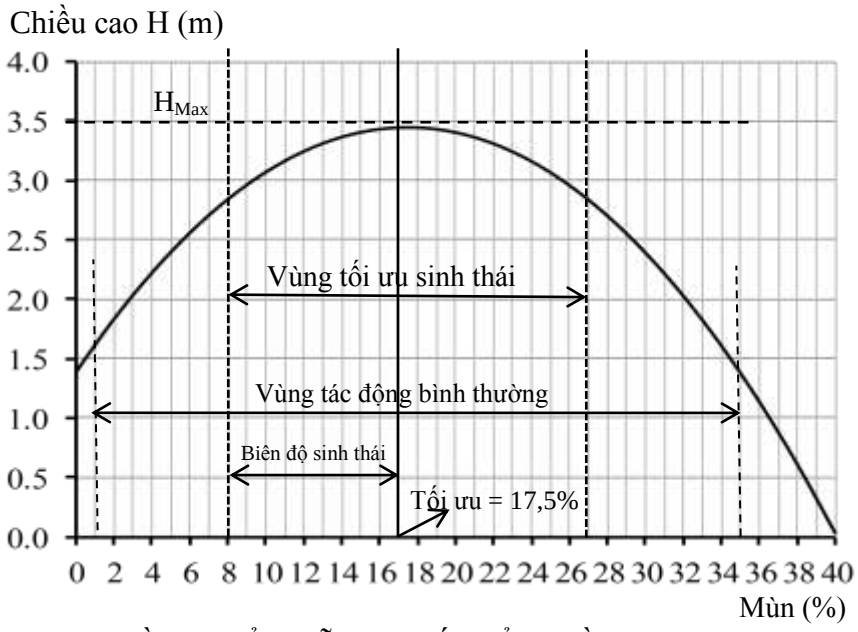
Bảng 5.8. Những tham số sinh thái đối với sinh trưởng của cây con Dầu song nàng 2 năm tuổi.

Nhân tố	X_{OPT}	T	$X_{Bq} \pm T$	$X_{Bq} \pm 2*T$	$X_{Bq} \pm 4*T$	H_{Max} (m)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mùn (%)	17,6	9,6	8 - 27	0 - 37	0 - 56	3,5
pH_{H2O}	5,2	1,5	3,7 - 6,6	2,2 - 8,1	0 - 10	3,5

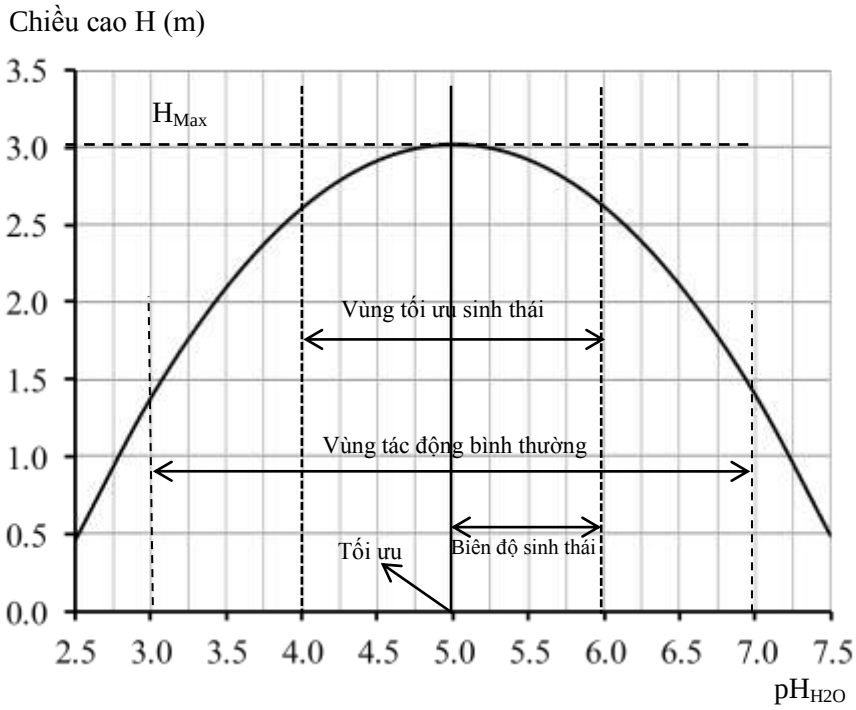
Phân tích số liệu ở Bảng 5.8 cho thấy cây con Dầu song nàng chỉ sống sót trên môi trường đất có hàm lượng mùn dao động từ 0 – 56% và $pH_{H2O} < 10$. Nó sống bình thường trên môi trường đất có hàm lượng mùn dao động từ 0 – 37% và $pH_{H2O} = 2,2 - 8,1$. Nó sống tốt trên môi trường đất có hàm lượng mùn dao động từ 8 – 27% và $pH_{H2O} = 3,7 – 6,6$. Mặt khác, sự phối hợp giữa mùn và pH_{H2O} ảnh hưởng rõ rệt ($P < 0,01$) đến sinh trưởng của cây con Dầu song nàng.

Lưu ý rằng muốn biết rõ sự phối hợp giữa hai nhân tố mùn và pH_{H2O} ảnh hưởng tốt hay xấu đến sinh trưởng của cây con Dầu song nàng, trước hết chúng ta phải chuẩn hóa hai nhân tố mùn và pH_{H2O}; sau đó xây dựng hàm $H = f(X_1, X_2, X_1*X_2)$ bằng phương pháp hồi quy từng bước. Kết quả nhận được hệ số của $c_5 = 0,0081$ (Bảng 5.9). Điều đó có nghĩa là sự phối hợp của 2 nhân tố (X_1*X_2) ảnh

hưởng tốt đến sinh trưởng H của cây con Dầu song nàng. Khi X_1*X_2 gia tăng 1 đơn vị, thì H của cây con Dầu song nàng tăng lên khoảng 1% (lấy tròn).



Hình 5.14. Đồ thị biểu diễn sự biến đổi chiều cao của cây con Dầu song nàng 2 năm tuổi theo hàm lượng mùn ở tầng đất mặt.



Hình 5.15. Đồ thị biểu diễn sự biến đổi chiều cao của cây con Dầu song nàng 2 năm tuổi theo pH_{H2O} ở tầng đất mặt.

Bảng 5.9. Hàm phản hồi của cây con Dầu song nàng 2 năm tuổi đối với hàm lượng mùn (X_1%) và pH_{H2O} (X_2) ở tầng đất mặt (0 – 30 cm) của đất Bazan.

Tham số	Ước lượng	R^2(%)	SEE	MAE	P-Value
(1)	(2)	(3)	(4)	(5)	(6)
Hằng số	1,14467	47,5	0,41	0,32	< 0,001
X_1	0,09181				
X_1^2	-0,00152				
X_2	0,21850				
X_2^2	-0,01455				
X_1*X_2	0,00810				

5.4. Hàm hồi quy Sigmoid và Logit Gauss và ứng dụng trong lâm học

5.4.1. Khái quát về hồi quy Sigmoid và Logit Gaus

Hàm hồi quy Sigmoid mô tả mối quan hệ giữa biến phản hồi (Y) ở dạng nhị phân (1 và 0) và các biến giải thích (X) ở dạng định lượng. Biến phản hồi kỳ vọng là xác suất xuất hiện một sự kiện tương ứng với các biến giải thích. Trong trường hợp một biến giải thích, hàm phản hồi có dạnh như hàm 5.39 và 5.40. Đồ thị biểu diễn xác suất của hàm 5.39 có dạng một đường cong hình chữ S. Xác suất lấy giá trị giữa 0 và 1. Các tham số của hàm 5.39 được ước lượng theo nguyên lý hợp lý tối đa.

$$Log_e[P/(1 - P)] = b_0 + b_1X \qquad (5.39)$$

$$P = \frac{\exp(b_0 + b_1X)}{1 + \exp(b_0 + b_1X)} \qquad (5.40)$$

Khi thay thế thành phần dự đoán tuyến tính ở hàm 5.39 bằng hàm bậc 2, chúng ta có hàm Logit Gauss (Hàm 5.41 và 5.42); trong đó tham số b_2 nhận giá trị âm.

$$Log_e[P/(1 - P)] = b_0 + b_1X - b_2X^2 \qquad (5.41)$$

$$P = \frac{\exp(b_0 + b_1X - b_2X^2)}{1 + \exp(b_0 + b_1X - b_2X^2)} \qquad (5.43)$$

Trong trường hợp sử dụng hai biến dự đoán, hàm Logit Gauss 2 biến số có dạng như hàm 5.44 và 5.45; trong đó hai tham số b_2 và b_4 nhận giá trị âm.

$$Log_e[P/(1 - P)] = b_0 + b_1X_1 - b_2X_1^2 + b_3X_2 - b_4X_2^2 \qquad (5.44)$$

$$P = \frac{\exp(b_0 + b_1X_1 - b_2X_1^2 + b_3X_2 - b_4X_2^2)}{[1 + \exp(b_0 + b_1X_1 - b_2X_1^2 + b_3X_2 - b_4X_2^2)]} \qquad (5.45)$$

Khi xét đến sự tương tác giữa hai biến dự đoán, hàm Logit Gauss 2 biến số có dạng như hàm 5.46 và 5.47.

$$Log_e[P/(1 - P)] = b_0 + b_1X_1 - b_2X_1^2 + b_3X_2 - b_4X_2^2 + b_5X_1*X_2 \quad (5.46)$$

$$P = \frac{\exp(b_0 + b_1X_1 - b_2X_1^2 + b_3X_2 - b_4X_2^2 + b_5X_1*X_2)}{[1 + \exp(b_0 + b_1X_1 - b_2X_1^2 + b_3X_2 - b_4X_2^2 + b_5X_1*X_2)]} \quad (5.47)$$

5.4.2. Ứng dụng hàm hồi quy Logit Gauss trong lâm học

Trong lâm học, hàm hồi quy Logit Gauss được ứng dụng để phân tích biến động độ phong phú của các loài cây gỗ trong mối quan hệ với biến động của các nhân tố sinh thái. Bằng cách xây dựng các hàm phản hồi ở dạng hàm 5.41, 5.44 và 5.46, nhà lâm học xác định được biên độ của nhân tố sinh thái ($\pm T_i$), vùng tối ưu sinh thái, tối ưu sinh thái (U_i), biên độ sinh thái, biên độ sống sót và xác xuất bắt gặp độ phong phú cao nhất (P_{iMax}). Sáu tham số này được xác định theo 6 hàm 5.48 - 5.53.

Biên độ sinh thái: $T_i = 1/\sqrt{2b_2}$ (5.48)

Vùng tối ưu sinh thái: $U_i \pm T_i$ (5.49)

Tối ưu sinh thái: $U_i = (b_1/2b_2)$ (5.50)

Vùng tác động bình thường: $U_i \pm 2T_i$ (5.51)

Biên độ sống sót: $U_i \pm 4T_i$ (5.52)

$$P_{iMax} = \frac{\exp(b_0 + b_1U_i + b_2U_i^2)}{1 + \exp(b_0 + b_1U_i + b_2U_i^2)} \quad (5.53)$$

Cây gỗ và rừng chịu ảnh hưởng của nhiều nhân tố sinh thái khác nhau. Vì thế, mô hình hồi quy Logit Gauss cần được xây dựng với nhiều biến giải thích khác nhau. Phương pháp phân tích hàm Logit Gauss với nhiều biến số giúp ích cho nhà lâm học phân tích ảnh hưởng tổng hợp của nhiều nhân tố sinh thái đến cây gỗ và rừng; trong đó biến phản hồi ở dạng bắt gặp và không bắt gặp, còn biến giải thích ở dạng định lượng. Bằng cách chọn biến từng bước, nhà lâm học có thể phân tích và lọc ra những tập hợp biến giải thích hữu ích nhất. Đây là phương pháp nghiên cứu được áp dụng rộng rãi trong lâm học.

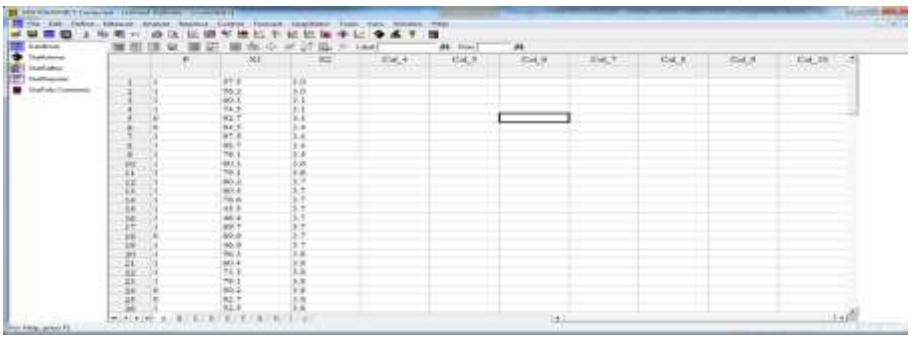
Đào Thị Thùy Dương (2018) đã áp dụng phương pháp này để phân tích ảnh hưởng của độ ẩm (X_1%) và pH_{H2O} ở tầng đất mặt (0 – 30 cm) của đất xám trên đá hoa cương đến tái sinh tự nhiên của Dầu con rái. Tổng số điểm quan sát là 120. Tại mỗi điểm quan sát, lập ô mẫu với kích thước 2*2 m. Độ phong phú của cây tái sinh Dầu con rái trong mỗi ô mẫu được thu thập bằng hai dấu hiệu: bắt gặp = 1 và không bắt gặp = 0. Độ ẩm (X_1%) và pH_{H2O} (X_2) được xác định bằng máy đo nhanh. Số liệu thu thập độ bắt gặp cây tái sinh tự nhiên của Dầu con rái và hai biến giải thích X_1 và X_2 được ghi lại ở Phụ lục 6. Từ nguồn số liệu này, dưới đây dẫn lại các bước xây dựng và chọn hàm phản hồi $P = f(X_1, X_2)$ thích hợp từ hàm Sigmoid và hàm Logit Gauss. Công cụ xử lý là phần mềm Statgraphics.

Bước 1. Copy ba biến P, X_1 và X_2 rồi chuyển chúng vào ba cột khác nhau (Hình 5.16).

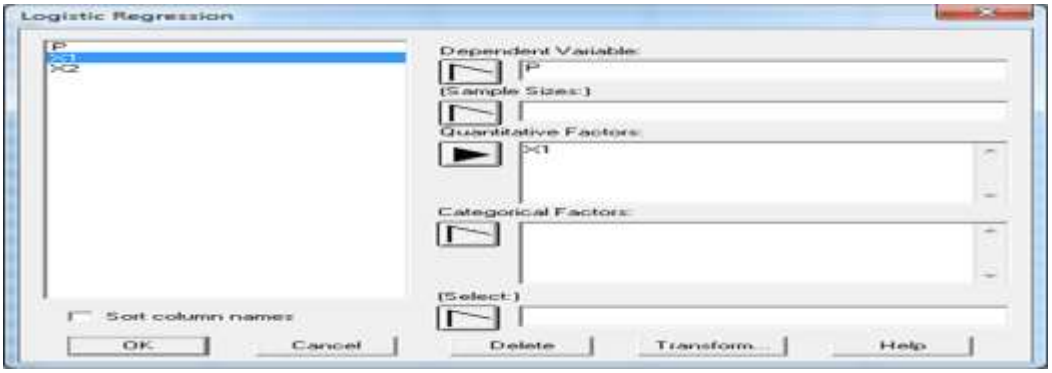
Bước 2. Xây dựng hàm phản hồi $P = f(X_1)$. Trước hết vào đường dẫn: Improve > Regression Analysis > Attribute Data > Logistic Regression. Đến đây

một cửa sổ xuất hiện. Tại cửa sổ này, chọn P và X_1 rồi chuyển chúng vào ô Dependent Variable và ô Quantitative Factor (Hình 5.17).

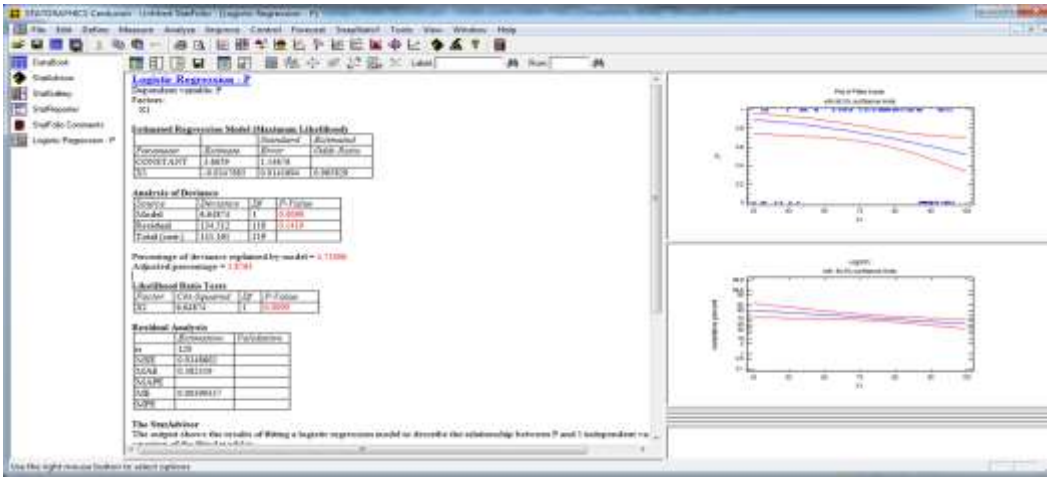
Bước 3. Kiểm định hàm phản hồi P = f(X_1) theo hàm Sigmoid. Để đạt được mục đích này, chọn OK ở Hình 5.17 để nhận kết quả phân tích hàm hồi quy P = f(X_1) theo hàm Sigmoid (Hình 5.18).



Hình 5.16. Số liệu để xây dựng hàm phản hồi P = f(X_1, X_2).



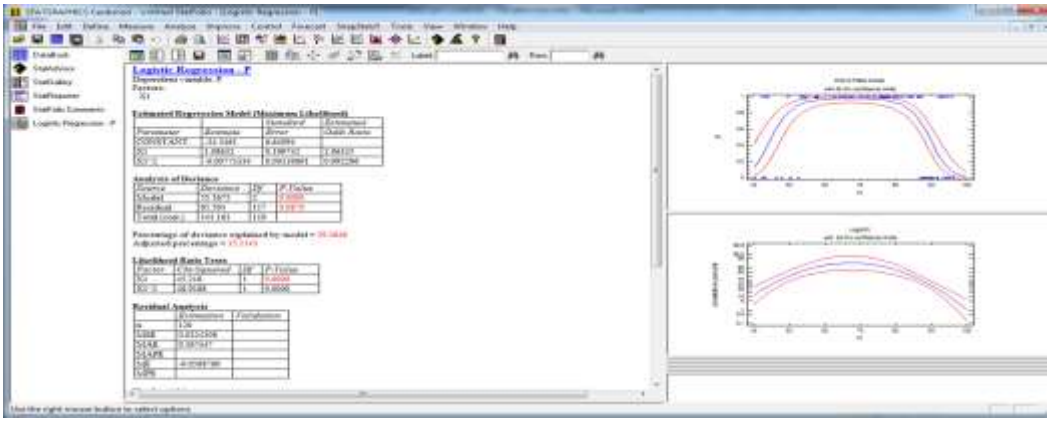
Hình 5.17. Xác định hàm phản hồi P = f(X_1).



Hình 5.18. Kết quả phân tích hàm hồi quy P = f(X_1) theo hàm Sigmoid.

Bước 4. Kiểm định hàm phản hồi P = f(X_1) theo hàm Logit Gauss. Để đạt được mục đích này, đặt con trỏ chuột tại cửa sổ ở Hình 5.18 và nhấn chuột phải rồi chọn Analysis Option. Đến đây xuất hiện cửa sổ Logistic Regression Options. Tại cửa sổ này, chọn Second Order (Hình 5.19) và sau đó chọn OK. Đến đây chương trình báo cáo kết quả phân tích hàm hồi quy P = f(X_1) theo hàm Logit Gauss (Hình 5.20).

Hình 5.19. Chọn hàm hồi quy P = f(X_1) theo hàm Logit Gauss.



Hình 5.20. Kết quả phân tích hàm hồi quy P = f(X_1) theo hàm Logit Gauss.

Bước 5. Phân tích và chọn dạng hàm phản hồi thích hợp. Các bước 1 – 4 chỉ xây dựng hàm phản hồi riêng rẽ với từng nhân tố X_1 và X_2. Ở bước này, tập hợp các thống kê phản ánh sự phù hợp của hàm Sigmoid và hàm Logit Gauss với số liệu thực nghiệm (Bảng 5.10 và 5.11). So sánh các thống kê cho thấy, hai hàm Sigmoid và Logit Gauss đều phản ánh rõ mối quan hệ ($P < 0,01$) giữa xác suất bắt gặp (P) cây tái sinh Dầu con rái với biến X_1. Hệ số tương quan ở hàm Sigmoid (4,7%) thấp hơn so với hàm Logit Gauss (39,3%). Sai số bình phương trung bình ở hàm Sigmoid (MSE = 0,034) cao hơn 1,54 lần so với hàm Logit Gauss (MSE = 0,022%).

Theo tiêu chuẩn MSE_{Min}, hàm Logit Gauss là hàm phù hợp để mô tả mối quan hệ P = f(X_1).

Bảng 5.10. Hàm Sigmoid biểu diễn xác suất bắt gặp cây tái sinh Dầu con rái dựa theo độ ẩm ở tầng đất mặt.

Estimated Regression Model (Maximum Likelihood)

Parameter	Estimate	Standard Error	Estimated Odds Ratio
CONSTANT	3.6659	1.14676	
X_1	-0.0347685	0.0141694	0.965829

Analysis of Deviance

Source	Deviance	Df	P-Value
Model	6.64874	1	0.0099
Residual	134.512	118	0.1419
Total (corr.)	141.161	119	

Percentage of deviance explained by model = 4.71006

Adjusted percentage = 1.8764

Likelihood Ratio Tests

Factor	Chi-Squared	Df	P-Value
X_1	6.64874	1	0.0099

Residual Analysis

	Estimation	Validation
n	120	
MSE	0.0348602	
MAE	0.382319	
MAPE		
ME	0.00399337	
MPE		

Bảng 5.11. Hàm Logit Gauss biểu diễn xác suất bắt gặp cây tái sinh Dầu con rái dựa theo độ ẩm ở tầng đất mặt.

Estimated Regression Model (Maximum Likelihood)

Parameter	Estimate	Standard Error	Estimated Odds Ratio
CONSTANT	-34.3461	6.64094	
X_1	1.08632	0.199742	2.96335
X_1^2	-0.00773334	0.00139691	0.992296

Analysis of Deviance

Source	Deviance	Df	P-Value
Model	55.5675	2	0.0000
Residual	85.593	117	0.9870
Total (corr.)	141.161	119	

Percentage of deviance explained by model = 39.3648

Adjusted percentage = 35.1143

Likelihood Ratio Tests

Factor	Chi-Squared	Df	P-Value
X	45.218	1	0.0000
X^2	48.9188	1	0.0000

Residual Analysis

	Estimation	Validation
n	120	
MSE	0.0224308	
MAE	0.367447	
MAPE		
ME	-0.0568769	

Để xây dựng hàm hồi quy P = f(X_2), chúng ta thực hiện các bước từ 1 – 4 tương tự xây dựng hàm P = f(X_1). Bảng 5.12 và 5.13 ghi lại kết quả làm phù hợp hàm Sigmoid và hàm Logit Gauss với số liệu thực nghiệm. So sánh các thống kê ở Bảng 5.12 và 5.13 cho thấy, hai hàm Sigmoid và Logit Gauss đều phản ánh rõ mối quan hệ (P < 0,01) giữa P bắt gặp cây tái sinh Dầu con rái với biến X_2. Tuy vậy, hệ số tương quan ở hàm Sigmoid (33,3%) thấp hơn so với hàm Logit Gauss (41,2%). Giá trị MSE ở hàm Sigmoid (0,0134) cao hơn 1,05 lần so với hàm Logit Gauss (0,0128%). Theo tiêu chuẩn MSE_{Min}, hàm Logit Gauss là hàm phù hợp để mô tả mối quan hệ P = f(X_2).

Bảng 5.12. Hàm Sigmoid biểu diễn xác suất bắt gặp cây tái sinh Dầu con rái dựa theo pH ở tầng đất mặt.

Estimated Regression Model (Maximum Likelihood)

Parameter	Estimate	Standard Error	Estimated Odds Ratio
CONSTANT	8.5039	1.39692	
X2	-1.59816	0.287354	0.202269

Analysis of Deviance

Source	Deviance	Df	P-Value
Model	47.0086	1	0.0000
Residual	94.1519	118	0.9481
Total (corr.)	141.161	119	

Percentage of deviance explained by model = 33.3015

Adjusted percentage = 30.4679

Likelihood Ratio Tests

Factor	Chi-Squared	Df	P-Value
X2	47.0086	1	0.0000

Residual Analysis

	Estimation	Validation
n	120	
MSE	0.0134467	
MAE	0.257969	
MAPE		
ME	0.0365655	

Bảng 5.13. Hàm Logit Gauss biểu diễn xác suất bắt gặp cây tái sinh Dầu con rái dựa theo pH ở tầng đất mặt.

Estimated Regression Model (Maximum Likelihood)

Parameter	Estimate	Error	Odds Ratio
CONSTANT	-16.9995	7.49276	
X2	9.26971	3.2791	10611.7
X2^2	-1.11436	0.347841	0.328125

Analysis of Deviance

Source	Deviance	Df	P-Value
Model	58.0895	2	0.0000
Residual	83.071	117	0.9925
Total (corr.)	141.161	119	

Percentage of deviance explained by model = 41.1514

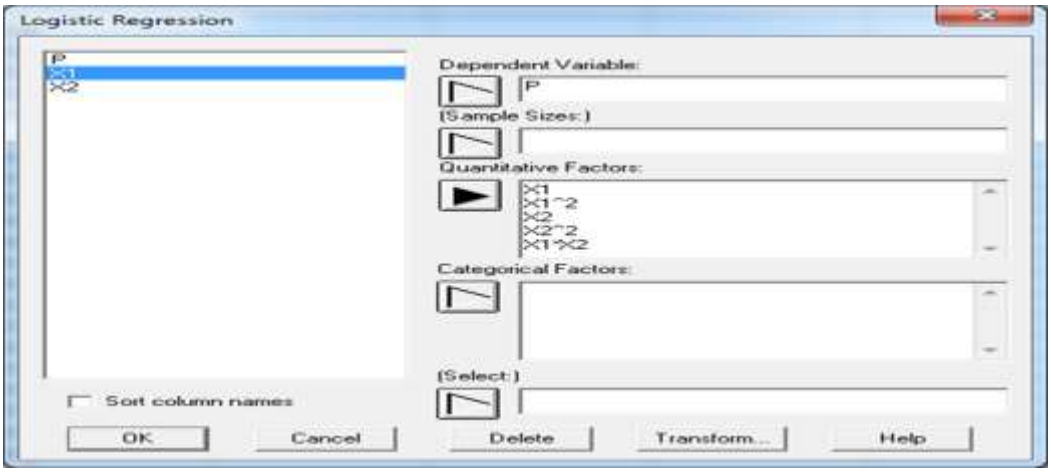
Adjusted percentage = 36.9009

Likelihood Ratio Tests

Factor	Chi-Squared	Df	P-Value
X2	8.06207	1	0.0045
X2^2	11.0809	1	0.0009

Residual Analysis

	Estimation	Validation
n	120	
MSE	0.0128413	
MAE	0.244499	
MAPE		
ME	-0.008827	
MPE		

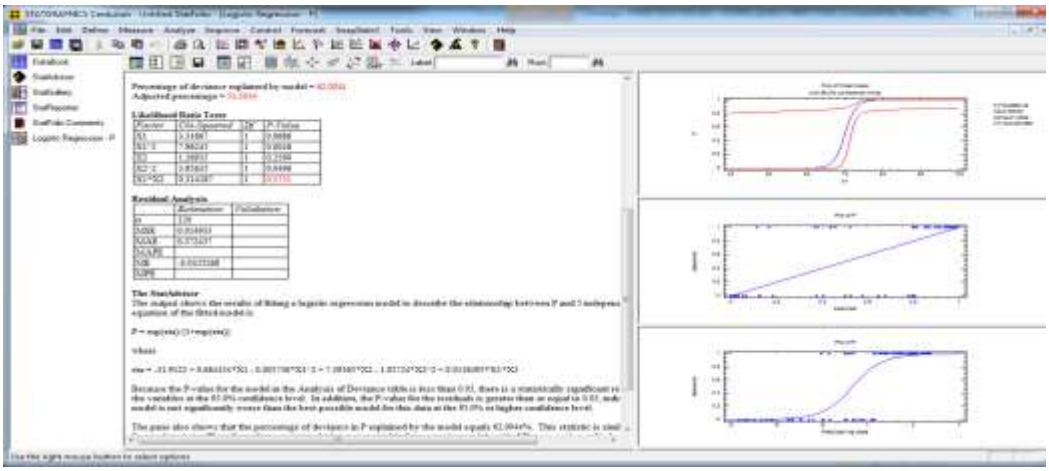


Hình 5.21. Xác định hàm phản hồi P = f(X_1, X_2, X_1*X_2).

Bước 6. Xây dựng hàm phản hồi P = f(X_1, X_2, X_1*X_2). Kết quả ở bước 5 cho thấy độ bắt gặp cây tái sinh Dầu con rái phụ thuộc vào hai biến X_1 và X_2. Vì thế, để phân tích ảnh hưởng tổng hợp của hai nhân tố này, chúng ta xây dựng hàm P = f(X_1,

X_2, X_1*X_2) dưới dạng hàm 5.46. Để đạt mục đích này, tại cửa sổ ở Hình 5.17, chọn P, X_1, X_2 và X_1*X_2 và chuyển chúng vào ô Dependent Variable và ô Independent Variable X (Hình 5.21).

Bước 7. Xác định các thông tin của hàm phản hồi P = f(X_1, X_2, X_1*X_2). Để đạt được mục đích này, chọn OK ở Hình 5.21. Đến đây chương trình báo cáo kết quả phân tích hồi quy P = f(X_1, X_2, X_1*X_2) (Hình 5.22).



Hình 5.22. Kết quả phân tích hàm phản hồi H = f(X_1, X_2, X_1*X_2).

Bước 8. Phân tích hàm phản hồi P = f(X_1, X_2, X_1*X_2). Kết quả ở bước 7 nhận được hàm phản hồi P = f(X_1, X_2, X_1*X_2) cùng với những thống kê tương quan và sai lệch của hàm này (Bảng 5.14). Từ số liệu ở Bảng 5.14 cho thấy, hàm này tồn tại (P < 0,01) với hệ số xác định khá cao (R^2 = 62,0%) và MSE = 0,0149. Hệ số của X_1*X_2 tồn tại và mang giá trị dương chứng tỏ sự tương tác của hai nhân tố X_1 và X_2 ảnh hưởng tốt đến tái sinh của Dầu con rái. Vì thế, hàm này được sử dụng để mô tả mối quan hệ P = f(X_1, X_2, X_1*X_2).

Bảng 5.14. Hàm phản hồi của cây tái sinh Dầu con rái đối với độ ẩm (X_1) và pH_{H2O} (X_2) ở tầng đất mặt (0 – 30 cm).

Estimated Regression Model (Maximum Likelihood)

Parameter	*Estimate*	*Standard Error*	*Estimated Odds Ratio*
CONSTANT	-31.9122	22.0352	
X1	0.684454	0.343979	1.98269
X1^2	-0.005736	0.00198953	0.99428
X2	7.39565	6.55016	1628.88
X2^2	-1.05724	0.567799	0.347415
X1*X2	0.0156395	0.0279627	1.01576

Analysis of Deviance

Source	*Deviance*	*Df*	*P-Value*
Model	87.5257	5	0.0000
Residual	53.6348	114	1.0000

Total (corr.)	141.161	119

Percentage of deviance explained by model = 62.0044
Adjusted percentage = 53.5034
Residual Analysis

	Estimation	*Validation*
n	120	
MSE	0.014933	
MAE	0.372437	
MAPE		
ME	-0.0325268	
MPE		

Từ hàm $P = f(X_1, X_1^2)$ ở Bảng 5.11 và $P = f(X_2, X_2^2)$ ở Bảng 5.13, bằng cách áp dụng 6 hàm 5.48 - 5.53, xác định được tối ưu sinh thái, biên độ sinh thái, vùng tối ưu sinh thái, vùng tác động bình thường của nhân tố sinh thái, biên độ chống chịu và xác xuất bắt gặp cây tái sinh Dầu con rái ở mức cao nhất (Bảng 5.15). Từ số liệu ở Bảng 5.15 cho thấy cây tái sinh Dầu con rái có thể sống trên môi trường đất có độ ẩm dao động từ 38 – 100% và $pH_{H2O} < 7,0$. Nó sống bình thường trên môi trường đất có độ ẩm dao động từ 54 – 86% và $pH_{H2O} = 3,0 – 6,0$. Nó sống tốt trên môi trường đất có độ ẩm dao động từ 62 – 78% và $pH_{H2O} = 3,5 – 4,8$.

Bảng 5.15. Những tham số sinh thái đối với độ bắt gặp cây tái sinh Dầu con rái.

Nhân tố	U_{OPT}	T	U ± T	U ± 2*T	U ± 4*T	P_{Max}
(1)	(2)	(3)	(4)	(5)	(6)	(7)
X_1 = Độ ẩm (%)	70	8	62 - 78	54 - 86	38 – 100	0,978
$X_2 = pH_{H2O}$	4,2	0,7	3,5 – 4,8	3 – 6	0 - 7	0,907

5.5. Hàm hồi quy Logistic và ứng dụng trong lâm học

5.5.1. Khái quát về hàm hồi quy Logistic

Hồi quy Logistic được sử dụng để làm phù hợp mô hình hồi quy mà trong đó biến phụ thuộc (Y) đặc trưng cho một sự kiện với chỉ 2 kết quả. Đối với các biến Y ở dạng tỷ lệ của một sự kiện nào đó, hàm Logistic có dạng hàm tuyến tính (Công thức 5.54); trong đó hệ số b_i mô tả mối quan hệ giữa Y với các biến dự đoán (X_i).

$$E_{(Y)} = \frac{1}{1 + \exp[-(b_0 + \sum_{i=1}^{k} b_i X_i)]} \qquad (5.54)$$

Đối với các biến ngẫu nhiên Y ở dạng nhị phân (1, 0), hàm Logistic có dạng hàm tuyến tính (Công thức 5.55); trong đó $P_{(Y = 1)}$ là xác suất bắt gặp sự kiện Y, b_i là hệ số mô tả mối quan hệ giữa $P_{(Y = 1)}$ với các biến dự đoán (X_i).

$$P_{(Y = 1)} = \frac{1}{1 + \exp[-(b_0 + \sum_{i=1}^{k} b_i X_i)]} \qquad (5.55)$$

Để đo đạc ảnh hưởng của các biến X_i đến xác suất bắt gặp Y ở hai hay nhiều đối tượng khác nhau, người ta sử dụng số đo tỷ lệ sai khác (Odds Ratio = OR). Sự

sai khác (Odds) là tỷ lệ giữa xác suất bắt gặp và không bắt gặp một sự kiện nào đó. Sự sai khác của việc dự đoán sự kiện D dựa theo các biến X_i được biểu diễn theo công thức 5.56. Chẳng hạn: Khi P(D) = 0,25, thì Odds(D) = (0,25/(1 - 0,25) = 1/3. Kết quả này có nghĩa là xác suất bắt gặp sự kiện D bằng 1/3 so với xác suất không bắt gặp sự kiện D.

$$Odds(D) = \frac{P_{(D)}}{1 - P_{(D)}} \qquad (5.56)$$

Tỷ lệ sai khác (OR) giữa hai đối tượng là tỷ lệ của hai sự sai khác. Tỷ lệ sai khác giữa 2 đối tương A và B (OR_{A-B}) được tính theo công thức 5.57.

$$OR_{(A - B)} = Odds(D_A)/Odds(D_B) = [\frac{P_{(DA)}}{1 - P_{(DA)}}]/[\frac{P_{(DB)}}{1 - P_{(DB)}}] \qquad (5.57)$$

Chẳng hạn: Giả thiết so sánh nguy cơ bị bệnh vàng lá giữa hai nhóm cây được bón phân chuồng hoai (PC) và không được bón phân chuồng hoai (NPC). Tỷ lệ cây bị bệnh vàng lá ở nhóm PC là 0,25, còn nhóm NPC là 0,10. Từ đó $OR_{(NPC-PC)}$ = (0,25/1- 0,25)/(0,10/1 - 0,10) = (1/3)/(1/9) = 3.

Khi biến đổi ở dạng Logit, hàm Logistic (5.54) được viết ở dạng hàm 5.58.

$$Logit\ (Odds) = Ln[\frac{P(_{Y=1})}{1 - P(_{Y=1})}] = b_0 + \sum^{k}_{i=1} b_i X_i \qquad (5.58)$$

Hàm Logistic có thể được sử dụng để so sánh tỷ lệ sai khác giữa các nhóm dựa trên một tập hợp các biến dự đoán $X_1, X_2, \ldots, X_k$. Nhóm A bao gồm các biến dự đoán $X_A = (X_{A1}, X_{A2}, \ldots, X_{Ak})$, còn nhóm B bao gồm các biến dự đoán $X_B = (X_{B1}, X_{B2}, \ldots, X_{Bk})$. Khi k = 3, thì hàm Logit có dạng như hàm 5.59.

$$Logit(P/1 - P) = b_0 + b_1X_1 + b_2X_2 + b_3X_3 \qquad (5.59)$$

Tỷ lệ sai khác giữa hai nhóm được tính theo công thức 5.60.

$$OR_{(XA - XB)} = Odds(X_A)/Odds(X_B)$$

$$Odds(X_A) = exp(b_0 + \sum^{k}_{i=1} b_i X_A)$$

$$Odds(X_B) = exp(b_0 + \sum^{k}_{i=1} b_i X_B)$$

$$OR_{(XA - XB)} = exp[\sum^{k}_{i=1}(X_A - X_B)b_i] \qquad (5.60)$$

Tỷ lệ sai khác giữa hai đối tượng theo mỗi biến giải thích được xác định theo công thức 5.61.

$$OR_{(XA - XB)} = exp(b_i) \qquad (5.61)$$

Nói chung, hồi quy Logistic được sử dụng cho 2 kiểu số liệu: (1) Biến phụ thuộc (Y) là tỷ lệ xuất hiện của một sự kiện nào đó; (2) Biến phụ thuộc (Y) được đo bằng hai biến nhị phân (1 và 0). Khi biến phụ thuộc là tỷ lệ của một sự kiện nào đó, thì các hệ số của hàm Logistic được ước lượng bằng phương pháp bình phương nhỏ nhất có trọng số (Weighted Least Squares = WLS). Phương pháp này kiểm định sự tồn tại của hàm hồi quy bằng tiêu chuẩn F. Sự tồn tại của các hệ số hồi quy được kiểm định bằng tiêu chuẩn T. Khi biến phụ thuộc là hai biến nhị phân, thì các hệ số

của hàm Logistic được ước lượng bằng phương pháp hợp lý tối đa (Maximum Likelihood = MLE). Phương pháp này xác định sự tồn tại của hàm hồi quy và các hệ số hồi quy bằng kiểm định Likelihood.

Nói chung, hai phương pháp MLE và WLS báo cáo kết quả phân tích hồi quy khác nhau. Đối với phương pháp MLE, sự tồn tại của mô hình được kiểm định bằng tiêu chuẩn F. Giả thuyết: Mô hình không tồn tại. Khi $P < 0,05$ thì mô hình tồn tại. Ý nghĩa của các biến trong mô hình được kiểm định bằng thống kê sai lệch (Residual hay Deviation). Giả thuyết: Các biến độc lập có ý nghĩa trong mô hình. Khi $P < 0,01$ hoặc 0,05, thì các biến X_i không có ý nghĩa. Trong trường hợp này, các biến độc lập cần được biến đổi sang dạng khác như căn bậc 2, logarit...Các hệ số của các biến độc lập (X) được ước lượng bằng phương pháp Maximum Likelihood. Mức độ chặt chẽ của mối quan hệ giữa biến phụ thuộc và các biến độc lập được xác định bằng phần trăm sai lệch (Percentage of deviance explained by model). Ở phương pháp WLS, sử dụng kiểm định F theo phương pháp tổng bình phương nhỏ nhất để kiểm định sự tồn tại của mô hình. Sự tồn tại của các biến độc lập (X_i) được kiểm định bằng tiêu chuẩn T. Mức độ chặt chẽ của mối quan hệ giữa biến phụ thuộc và các biến độc lập được xác định bằng hệ số xác định (R^2). Hai phương pháp MLS và WLS đều tính tỷ lệ sai khác tương tự như nhau.

5.5.2. Ứng dụng hàm hồi quy Logistic trong lâm học

5.5.2.1. Trường hợp biến phụ thuộc ở dạng tỷ lệ

Trong lâm học, một số đặc tính của cây gỗ và rừng có thể được đo đạc bằng những giá trị tỷ lệ. Chẳng hạn: (a) Tỷ lệ số cây bị bệnh ở những cấp tuổi hay cấp kích thước khác nhau; (b) Tỷ lệ số cây gỗ quý trong các cấp D; (c) Tỷ lệ số cây tái sinh bị cụt ngọn trong các cấp tuổi hay cấp H...Đối với những trường hợp này, nhà lâm học có thể sử dụng hàm Logistic để mô tả mối quan hệ giữa tỷ lệ xuất hiện một sự kiện và biến giải thích.

Giả sử chúng ta cần so sánh tỷ lệ số cây họ Sao Dầu theo cấp D ở hai trạng thái rừng $IIIA_2$ và $IIIA_3$. Số liệu thu thập được ghi lại ở Bảng 5.16; trong đó D = cấp D, N = tổng số cây theo cấp D, K = số cây họ Sao Dầu theo cấp D, J = 1 (TTR = $IIIA_3$) và J = 2 (TTR = $IIIA_2$). Để làm rõ vấn đề này, chúng ta có thể xây dựng 3 mô hình phân tích sau đây: (1) $P = a_0 + a_1J$; (2) $P_2 = b_0 + b_1D$; (3) $P_3 = c_0 + c_1J + c_3D$. Ở đây P = tỷ lệ số cây họ Sao Dầu, J = TTR, D = cấp D. Theo mô hình 3, chúng ta cần kiểm định giả thuyết H_0: Tỷ lệ số cây họ Sao Dầu (P) trong các cấp D ở trạng thái rừng $IIIA_3$ lớn hơn trạng thái rừng $IIIA_2$. Mối quan hệ giữa P với J và D được mô tả bằng hàm Logistic. Khi sử dụng phần mềm Statgraphics, hàm hồi quy Logistic được xây dựng theo 3 bước.

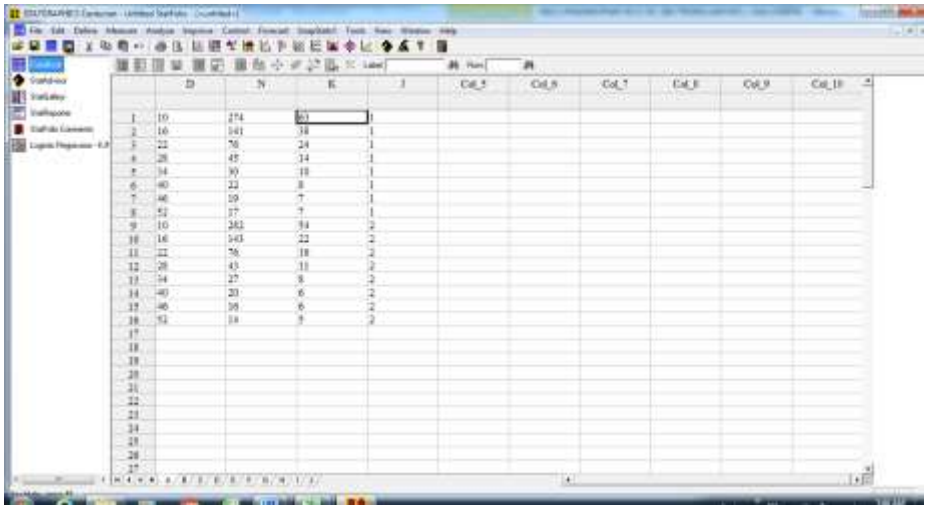
Bước 1. Copy 4 biến D, N, K và J rồi chuyển chúng vào 4 cột khác nhau (Hình 5.23).

Bước 2. Xây dựng hàm mô tả quan hệ P = f(J, D). Trước hết vào đường dẫn: Improve > Regression Analysis > Attribute Data > Logistic Regression. Đến đây

một cửa sổ xuất hiện. Tại cửa sổ này, chọn K/N, N, J và D rồi chuyển chúng vào ô Dependent Variable, Sample sizes và Quantitative Factor (Hình 5.24).

Bảng 5.16. Tỷ lệ số cây họ Sao Dầu trong các cấp đường kính ở hai trạng thái rừng $IIIA_2$ và $IIIA_3$.

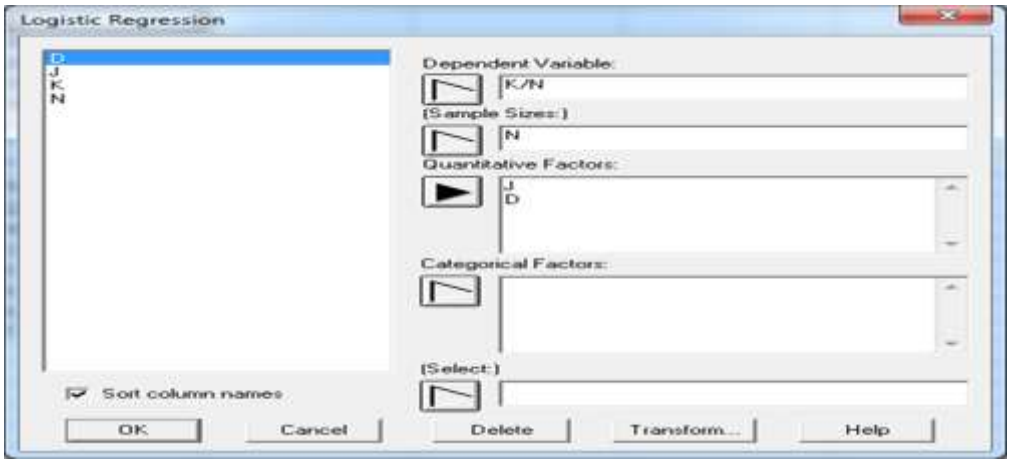
D (cm)	N (Cây)	K (Cây)	J	N (Cây)	K (Cây)	J
(1)	(2)	(3)	(4)	(4)	(5)	(6)
10	274	63	1	282	54	2
16	141	38	1	143	22	2
22	76	24	1	76	16	2
28	45	14	1	43	11	2
34	30	10	1	27	8	2
40	22	8	1	20	6	2
46	19	7	1	16	6	2
52	17	7	1	14	5	2



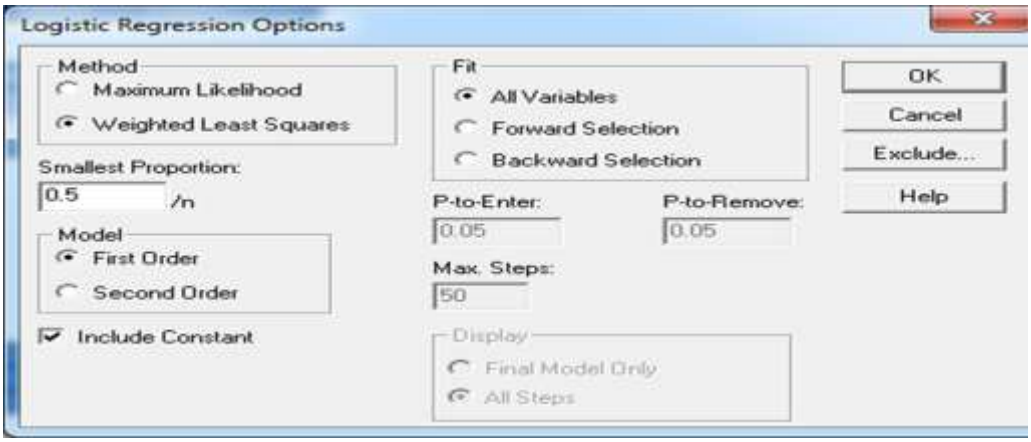
Hình 5.23. Số liệu để xây dựng hàm hồi quy Logistic.

Bước 3. Xác định phương pháp xây dựng hàm hồi quy Logistic: $P_3 = c_0 + c_1J + c_3D$. Để đạt được mục đích này, chọn OK ở Hình 5.24. Đến đây xuất hiện một cửa sổ như Hình 5.25. Tại cửa sổ này, chọn First Order Model, Weighted Least Squares Mothod, Smallest Proportion (Mặc định) và OK. Kết quả nhận được hàm hồi quy Logistic (Bảng 5.17). Từ số liệu ở Bảng 5.17 cho thấy, hàm này tồn tại ($F = 45,87$; $P < 0,001$) với hệ số xác định khá cao ($R^2 = 87,58\%$) và MSE = 0,2373. Hệ số của J và D đều tồn tại ($P < 0,01$); trong đó hệ số của D mang giá trị dương, còn hệ số của J mang giá trị âm. Điều đó chứng tỏ tỷ lệ số cây họ Sao Dầu gia tăng theo

cấp D; trung bình 1,022 lần hay 2,2% (1,022 = exp(b_i) = exp(0,0219)). So với TTR $IIIA_3$, tỷ lệ số cây họ Sao Dầu ở TTR $IIIA_2$ chỉ bằng 0,699 lần hay thấp hơn 30,1% (0,699 = exp(b_i) = exp(-0,35755)).



Hình 5.24. Xác định hàm hồi quy Logistic.



Hình 5.25. Chọn hàm hồi quy Logistic để mô tả quan hệ P = f(J, D).

Bảng 5.17. Kết quả của hàm Logistic biểu diễn mối quan hệ giữa tỷ lệ số cây họ Sao Dầu trong các cấp đường kính ở hai trạng thái rừng $IIIA_2$ và $IIIA_3$.

Estimated Regression Model (Weighted Least Squares)

Parameter	Estimate	Standard Error	t Statistic	P-Value	Estimated Odds Ratio
CONSTANT	-1.03489	0.114888	-9.00785	0.0000	
J	-0.357551	0.06566	-5.4455	0.0001	0.699387
D	0.0219058	0.00279737	7.83086	0.0000	1.02215

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	21.7728	2	10.8864	45.87	0.0000
Residual	3.08503	13	0.23731		
Total (Corr.)	24.8579	15			

R-Squared = 87.5893 percent
R-Squared (adjusted for d.f.) = 85.68 percent
Standard Error of Est. = 0.487145
Mean absolute error = 0.0886143

Type III Sums of Squares

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
J	7.03706	1	7.03706	29.65	0.0001
D	14.5524	1	14.5524	61.32	0.0000
Residual	3.08503	13	0.23731		

Residual Analysis

	Estimation	Validation
n	16	
MSE	0.23731	
MAE	0.088614	
MAPE	41.646	
ME	3.377E-16	
MPE	-5.54519	

5.5.2.2. Trường hợp biến phụ thuộc ở dạng nhị phân

Trong lâm học, một số đặc tính của cây gỗ và rừng có thể được đo đạc bằng độ bắt gặp (1, 0). Chẳng hạn: (a) Độ bắt gặp cây tái sinh của những loài cây gỗ quý trong những lớp H khác nhau; (b) Độ bắt gặp loài cây gỗ quan tâm trong mối quan hệ với một số nhân tố sinh thái...Khi nghiên cứu về mối quan hệ giữa rừng và đất, nhà lâm học có thể đặt ra một số câu hỏi. Một là độ bắt gặp loài cây gỗ quan tâm có thay đổi theo loại đất hay không? Hai là tính chất của mỗi loại đất thay đổi ảnh hưởng đến độ phong phú của loài cây gỗ quan tâm như thế nào? Đối với trường hợp này, hàm Logistic là một công cụ hữu ích để giúp nhà lâm học trả lời rõ hai câu hỏi đặt ra trên đây. Dưới đây dẫn ví dụ áp dụng hàm Logistic để phân tích mối quan hệ giữa rừng và đất.

Giả sử nghiên cứu đặt ra 2 mục tiêu: (1) Phân tích độ phong phú (Y) của loài cây gỗ quan tâm trên hai loại đất A và B (X_1); (2) Xác định ảnh hưởng của độ ẩm (X_2) và pH_{H2O} (X_3) của hai loại đất A và B đến độ phong phú (Y) của loài cây gỗ quan tâm. Để đạt được hai mục tiêu này, trước hết bố trí những tuyến cắt ngang qua hai loại đất A và B. Để loại trừ những ảnh hưởng của trạng thái QXTV, hai loại đất A và B phải phân bố trên cùng trạng thái QXTV. Kế đến bố trí những ô mẫu với kích thước 20*20 m trên các tuyến. Khoảng cách giữa hai ô mẫu kế cận trên mỗi tuyến là 100 m. Tiếp đến tại mỗi ô mẫu xác định độ bắt gặp (1) và không bắt gặp (0) loài cây gỗ quan tâm; đồng thời dùng máy đo hai biến X_2 và X_3 tại trung tâm ô mẫu. Để đảm bảo độ tin cậy của kết quả nghiên cứu, số lượng ô mẫu trên mỗi loại đất cần được thu thập đủ lớn (≥ 100 ô mẫu). Trong phần xử lý số liệu, chúng ta có thể xây dựng 7 mô hình khác nhau (Hàm 5.62 – 5.68); trong đó P là tỷ lệ ô mẫu bắt gặp loài cây gỗ quan tâm; 1 - P là tỷ lệ ô mẫu không bắt gặp loài cây gỗ quan tâm; Y = (P/1-P); X_1, X_2 và X_3 tương ứng là loại đất, độ ẩm và pH_{H2O} của đất.

Mô hình 1. $Ln(P/1 - P) = a_0 + a_1X_1$ (5.62)

Mô hình 2. $Ln(P/1 - P) = b_0 + b_1X_2$ (5.63)

Mô hình 3. $Ln(P/1 - P) = c_0 + c_1X_3$ (5.64)

Mô hình 4. $Ln(P/1 - P) = d_0 + d_1X_1 + d_2X_2 + d_3X_1X_2$ (5.65)

Mô hình 5. $Ln(P/1 - P) = e_0 + e_1X_1 + e_2X_3 + e_3X_1X_3$ (5.66)

Mô hình 6. $Ln(P/1 - P) = f_0 + f_1X_2 + f_2X_3 + f_3X_2X_3$ (5.67)

Mô hình 7. $Ln(P/1 - P) = h_0 + h_1X_1 + h_2X_2 + h_3X_3$
$+ h_4X_1X_2 + h_5X_1X_3 + h_6(X_2X_3) + h_7(X_1X_2X_3)$ (5.68)

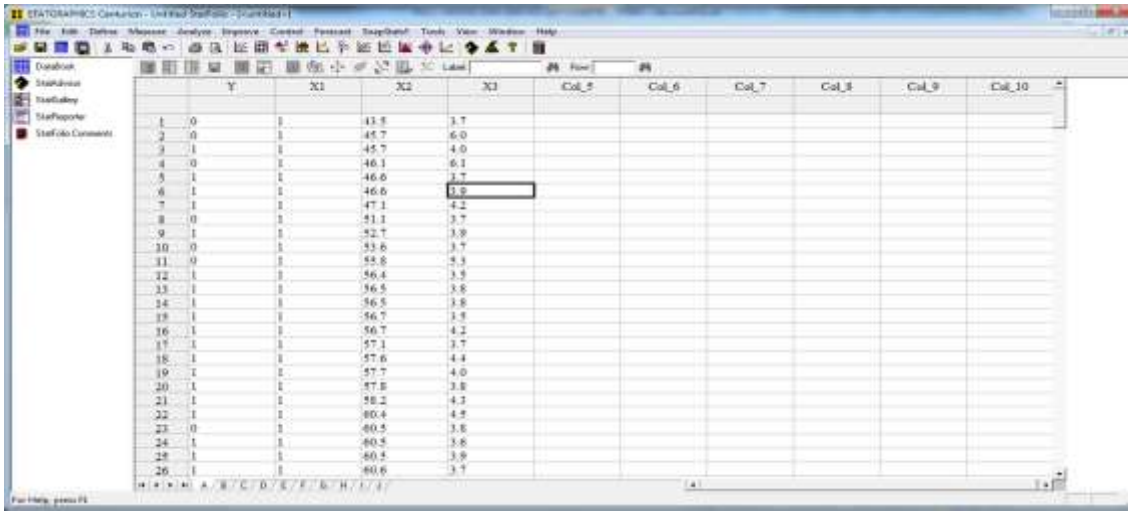
Ba mô hình 5.62 ÷ 5.64 tương ứng phân tích mối quan hệ $Y = f(X_1)$, $Y = f(X_2)$ và $Y = f(X_3)$. Kết quả của ba mô hình này trả lời những câu hỏi về vai trò của từng nhân tố sinh thái. Bốn mô hình 5.65 – 5.68 được sử dụng để phân tích ảnh hưởng tổng hợp của 2 và 3 nhân tố đến độ phong phú của loài cây quan tâm. Đối với phần mềm thống kê Statgraphics, các hàm hồi quy trên đây được xây dựng theo những bước dưới đây.

Bước 1. Lập bảng số liệu trong bảng tính Excell (Bảng 5.18). Cột 1 là thứ tự ô mẫu; Cột 2 là biến phụ thuộc Y (1 = bắt gặp và 0 = không bắt gặp); Cột 3 là loại đất X_1 (1 = đất A và 0 = đất B); Cột 4 và 5 tương ứng là hai biến X_2 và X_3.

Bảng 5.18. Độ bắt gặp loài cây quan tâm trên hai loại đất khác nhau.

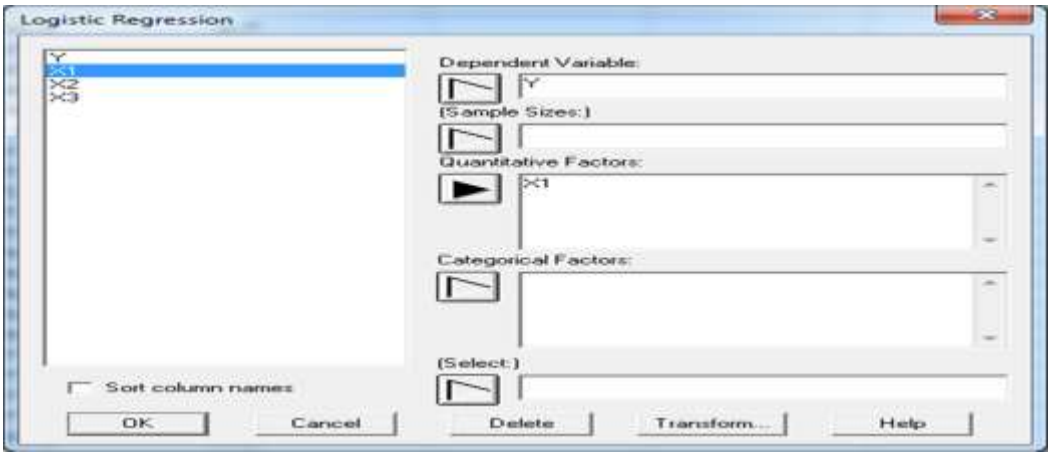
TT	Độ bắt gặp (Y)	Loại đất (X_1)	Độ ẩm (X_2)	pH_{H2O} (X_3)
(1)	(2)	(3)	(4)	(5)
1	1	1	70	6.0
2	0	1	78	3.5
3	0	1	72	4.0
...	...	...	...	...
101	1	0	65	5,5
102	0	0	58	3.2
...	...	...	...	...

Bước 2. Copy 4 biến Y, X_1, X_2 và X_3 rồi chuyển chúng vào 4 cột khác nhau của bảng tính Statgraphics (Hình 5.26).



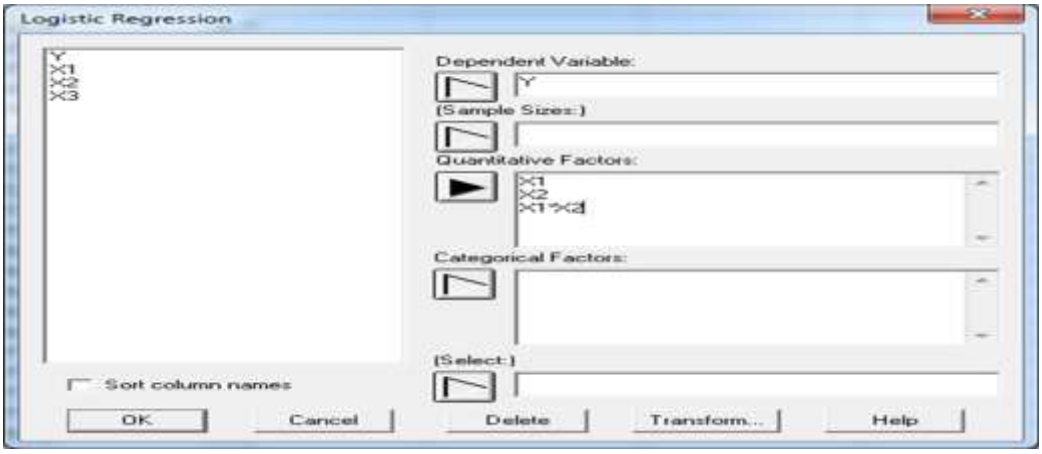
Hình 5.26. Số liệu để xây dựng hàm hồi quy $Y = f(X_i)$.

Bước 3. Xây dựng hàm phản hồi quy Y = f(X_1) (Mô hình 5.62). Trước hết vào đường dẫn: Improve > Regression Analysis > Attribute Data > Logistic Regression. Đến đây một cửa sổ xuất hiện. Tại cửa sổ này, chọn Y và X_1 rồi chuyển chúng vào ô Dependent Variable và Quantitative Factor (Hình 5.27). Sau đó chọn OK để nhận hàm hồi quy Logistic Y = f(X_1). Hai mô hình Y = f(X_2) (Mô hình 5.63) và Y = f(X_3) (Mô hình 5.64) cũng được thực hiện tương tự như bước 3.



Hình 5.27. Xác định hàm hồi quy Y = f(X_1).

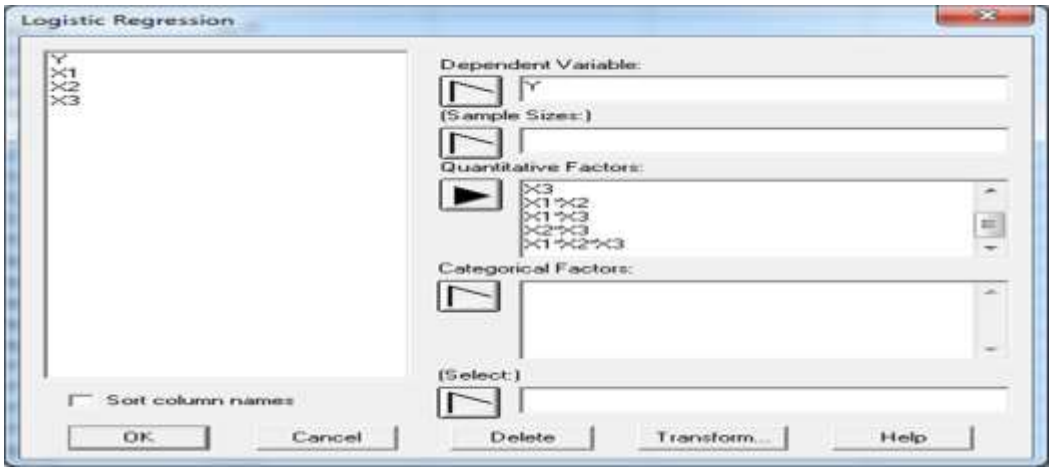
Bước 4. Xây dựng hàm hồi quy Y = f(X_1, X_2) (Mô hình 5.65). Để đạt được mục đích này, chọn Y, X_1 và X_2 rồi chuyển chúng vào ô Dependent Variable và Quantitative Factor (Hình 5.28). Sau đó chọn OK để nhận hàm hồi quy ước lượng. Hai mô hình Y = f(X_1, X_3) (Mô hình 5.66) và Y = f(X_2, X_3) (Mô hình 5.67) cũng được xâ dựng tương tự như hàm hồi quy Y = f(X_1, X_2).



Hình 5.28. Xây dựng hàm hồi quy Y = f(X_1, X_2).

Bước 5. Xây dựng hàm phản hồi quy Y = f(X_1, X_2, X_3) (Hàm 5.68). Trước hết vào đường dẫn: Improve > Regression Analysis > Attribute Data > Logistic Regression. Đến đây chọn Y, X_1, X_2 và X_3 rồi chuyển chúng vào ô Dependent Variable và Quantitative Factor (Hình 5.29). Sau đó chọn OK để nhận được hàm

hồi quy ước lượng. Bảng 5.19 là định dạng kết quả báo cáo hàm hồi quy Logistic Y = f(X_1, X_2, X_3).



Hình 5.29. Xác định hàm hồi quy Y = f(X_1, X_2, X_3).

Bảng 5.19. Kết quả phân tích mối quan hệ giữa độ bắt gặp loài cây quan tâm theo ba biến X_1, X_2 và X_3.

Estimated Regression Model (Maximum Likelihood)

Parameter	Estimate	Standard Error	Estimated Odds Ratio
CONSTANT	0.6576	8.9253	
X1	12.9005	11.9516	400528.0
X2	0.1610	0.1372	1.1746
X3	0.3858	2.0125	1.4708
X1*X2	-0.2537	0.1704	0.7759
X1*X3	-2.3929	2.6206	0.0914
X2*X3	-0.0411	0.0310	0.9597
X1*X2*X3	0.0518	0.0378	1.0532

Analysis of Deviance

Source	Deviance	Df	P-Value
Model	42.2582	7	0.0000
Residual	205.145	242	0.9589
Total (corr.)	247.403	249	

Percentage of deviance explained by model = 17.0807

Adjusted percentage = 10.6135

Likelihood Ratio Tests

Factor	Chi-Squared	Df	P-Value
X1	1.15148	1	0.2832
X2	1.3902	1	0.2384
X3	0.036584	1	0.8483
X1*X2	2.23549	1	0.1349
X1*X3	0.823946	1	0.3640
X2*X3	1.78956	1	0.1810

X1*X2*X3	1.8972	1	0.1684

Residual Analysis

	Estimation	Validation
n	250	
MSE	0.018172	
MAE	0.306967	
MAPE		
ME	0.008582	
MPE		

Nói chung, các hàm hồi quy Logistic là rất đa dạng. Bảy hàm từ 5.62 – 5.68 được xây dựng để trả lời chi tiết về vai trò của từng nhân tố sinh thái và ảnh hưởng tổng hợp của chúng đến độ phong phú của loài cây gỗ quan tâm. Nếu mục tiêu nghiên cứu là xác định ảnh hưởng tổng hợp của ba nhân tố X_1, X_2 và X_3 đến độ phong phú của loài cây gỗ quan tâm, thì chúng ta chỉ cần xây dựng 1 hàm hồi quy Logistic ở dạng hàm 5.68. Ở ví dụ trên đây, nếu xét đến cả trạng thái rừng (X_4), thì các hàm hồi quy cần phải được xây dựng nhiều hơn.

Những bước phân tích tiếp theo là sử dụng hàm Logistic để so sánh: (a) Sự sai khác giữa độ bắt gặp loài cây gỗ quan tâm ở hai loại đất; (b) Sự sai khác giữa độ bắt gặp loài cây gỗ quan tâm theo mỗi nhân tố sinh thái trên cùng một loại đất; (c) Sự sai khác giữa độ bắt gặp loài cây gỗ quan tâm theo mỗi nhân tố sinh thái trên hai loại đất khác nhau. Tỷ lệ sai khác giữa hai đối tượng được xác định theo công thức 5.60. Tỷ lệ sai khác giữa hai đối tượng theo mỗi biến giải thích được xác định theo công thức 5.61.

5.6. Hàm hồi quy Poisson và ứng dụng trong lâm học

5.6.1. Khái quát về hàm hồi quy Poisson

Hồi quy Poisson mô tả mối quan hệ giữa biến phụ thuộc Y với các biến dự đoán X_i. Biến Y là các số nguyên không âm. Các biến X_i có thể là biến nhị phân (1, 0) hoặc biến định lượng. Mô hình phân bố xác suất Poisson có dạng như hàm 5.69; trong đó Y_i = tần số xuất hiện sự kiện Y thứ i, T_i = dung lượng mẫu quan sát sự kiện Y_i, λ_i = tỷ lệ của biến phụ thuộc (Y_i) thứ i ($\lambda_i = Y_i/T_i$), $\lambda_i T_i$ = giá trị trung bình của Y_i.

$$p(Y_i) = \frac{\exp(-\lambda_i T_i)*(\lambda_i T_i)^{Yi}}{Y_i!} \qquad (5.69)$$

Ở hàm 5.69, hai thành phần Y_i và T_i đã biết. Tham số λ_i được ước lượng theo các biến dự đoán X_i. Tham số λ_i và các biến dự đoán X_i có quan hệ với nhau theo dạng Logarit tuyến tính (Hàm 5.70).

$$Ln(\lambda) = b_0 + b_1X_1 + b_2X_2 + ... + b_kX_k \qquad (5.70a)$$

$$\text{Hay } \lambda = \exp(b_0 + b_1X_1 + b_2X_2 + ... + b_kX_k) \qquad (5.70b)$$

Các tham số b_i ở hàm 5.70 phản ánh tỷ lệ (phần mười hay %) thay đổi của biến Y_i khi biến X_i thay đổi 1 đơn vị. Khi biết λ_i, thì biến phụ thuộc $Y_i = \lambda_i T_i$. Tỷ lệ

thay đổi của biến phụ thuộc Y_i theo mỗi đơn vị của biến dự đoán X_i được xác định theo công thức 5.71; trong đó RR_i = tỷ lệ thay đổi (Rate Ratio), λ_{i1} = tỷ lệ bắt gặp, λ_{i0} = tỷ lệ không bắt gặp.

$$RR_i = \frac{\lambda_{i1}}{\lambda_{i0}} = \exp(b_i) \qquad (5.71$$

Khi xây dựng hàm hồi quy Poisson, các tham số b_i được ước lượng theo phương pháp Maximum Likelihood. Hàm hồi quy Poisson có thể xây dựng ở dạng bậc 1 và bậc 2. Dạng bậc 1 chỉ bao gồm ảnh hưởng chính. Dạng bậc 2 bao gồm ảnh hưởng chính của các biến giải thích và sự tương tác giữa 2 biến giải thích. Khi mô hình bao gồm nhiều biến giải thích, thì hồi quy Poisson có thể bao gồm tất cả các biến đưa vào mô hình hoặc chỉ bao gồm một số biến giải thích có ý nghĩa lớn. Ở trường hợp sau, những biến giải thích có ý nghĩa lớn xác định theo phương pháp phân tích hồi quy tương quan từng bước.

Khi sử dụng phần mền Statgraphics để xây dựng hàm hồi quy Poisson, kết quả báo cáo những thống kê như dưới đây.

(1) Mô hình hồi quy ước lượng (Estimated Regression Model). Phần này bao gồm các tham số của mô hình (b_i), sai số chuẩn của từng hệ số trong mô hình (SEE = Standard Error), tỷ lệ thay đổi của biến phụ thuộc Y_i theo mỗi đơn vị của biến X_i (RR_i = Rate Ratio). Biến X_i nào có giá trị RR_i càng lớn thì biến đó ảnh hưởng càng lớn đến biến phụ thuộc Y. Ví dụ: Giá trị $RR_{(X1)}$ = 1,12 và $RR_{(X2)}$ = 1,25 cho biết khi biến X_1 và biến X_2 thay đổi 1 đơn vị, thì biến Y thay đổi tương ứng 1,12 lần hay 12% và 1,25 lần hay 25%.

(2) Phân tích sai lệch (Analysis of Deviance). Bảng này báo cáo: (a) Sai lệch của mô hình (Model Deviance); (b) Sai lệch còn lại sau khi làm phù hợp mô hình (Residual Deviance), nghĩa là sai lệch của mô hình do sử dụng các biến X_i để ước lượng và dự đoán biến Y. Giả thuyết 1: Mô hình không tồn tại. Nếu $P < 0,05$ thì mô hình tồn tại. Giả thuyết 2: Các biến X_i có ý nghĩa để ước lượng và dự đoán Y_i. Theo nghĩa của sinh thái học và lâm học, các biến X_i ảnh hưởng rõ rệt đến biến Y_i hay chúng là nguyên nhân gây ra những thay đổi ở biến Y_i. Nếu $P > 0,05$ thì các biến X_i giải thích rõ biến động của biến Y_i. Trái lại, nếu $P < 0,05$ thì các biến X_i giải thích không rõ rệt biến động của biến Y_i. Trong trường hợp sau, chúng ta cần chọn những biến giải thích khác hoặc biến đổi các biến X_i ban đầu sang dạng khác (Căn bậc 2, Logarit...).

(3) Phần trăm sai lệch được giải thích bởi mô hình (Percentage of Deviance explained by model). Thống kê này có ý nghĩa tương tự như hệ số xác định (R^2).

(4) Kiểm định sự tồn tại hay ý nghĩa của các biến giải thích (Likelihood Ratio Tests). Sự tồn tại của các biến giải thích được kiểm định bằng tiêu chuẩn χ^2 với 1 bậc tự do. Giả thuyết H_0: Các biến giải thích không tồn tại trong mô hình. Nếu $P < 0,05$, thì giả thuyết bị bác bỏ.

(5) Phân tích các sai lệch (Residual Analysis). Phần này báo cáo dung lượng mẫu quan sát (n), sai số bình phương trung bình (MSE), sai số tuyệt đối trung bình (MAE), sai số tương đối trung bình theo phần trăm (MAPE), độ lệch trung bình (Bias hay ME) và độ lệch trung bình theo phần trăm (MPE). Khi hàm hồi quy thích hợp, nhà lâm học sử dụng mô hình này để xác định các biến X_i ảnh hưởng lớn đến Y_i. Ngoài ra, mô hình có thể được sử dụng để ước lượng và dự đoán Y_i theo các biến X_i.

(6) Các đồ thị biểu diễn quan hệ giữa Y_i với các biến X_i.

5.6.2. Ứng dụng hàm hồi quy Poisson trong lâm học

Trong lâm học, hồi quy Poisson được sử dụng để mô tả sự xuất hiện của những sự kiện hiếm gặp trong các QXTV rừng. Chẳng hạn: (a) Số cây tái sinh của một loài cây gỗ hiếm xuất hiện trong những QXTV; (b) Số cây bị bệnh rỗng ruột ở các cấp D khác nhau. Dưới đây dẫn ví dụ ứng dụng hồi quy Poisson để phân tích và so sánh tỷ lệ số cây Bằng lăng ổi (*Lagerstroemia calyculata* Kurz) mọc trên đất Bazan và đất xám phù sa cổ bị bệnh rỗng ruột. Số liệu thu thập được ghi lại ở Bảng 5.20; trong đó D = cấp D (cm), N_{ij} = số cây quan sát ở cấp D_i trên đất J, Y_{ij} = số cây bị bệnh rỗng ruột ở cấp D_i trên đất J, λ_{ij} = tỷ lệ số cây bị bệnh rỗng ruột ở cấp D_i trên đất J, J = loại đất (0 = đất bazan, 1 = đất xám phù sa cổ), RR_i = tỷ lệ sai khác về bệnh rỗng ruột ở cấp D_i trên đất xám phù sa cổ so với đất Bazan.

Nghiên cứu này đặt ra 2 mục tiêu: (1) Xây dựng những hàm ước lượng tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột theo cấp D và loại đất; (2) So sánh tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột theo cấp D và loại đất khác nhau. Để đạt được 2 mục tiêu kể trên, trong phần xử lý số liệu, xây dựng 3 hàm hồi quy. Hàm thứ nhất mô tả ảnh hưởng của loại đất đến tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột (Hàm 5.72). Hàm thứ hai mô tả ảnh hưởng của cấp D đến tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột (Hàm 5.73). Hàm thứ ba mô tả ảnh hưởng của loại đất và cấp D đến tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột (Hàm 5.74).

Bảng 5.20. Tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột theo theo các cấp D và loại đất khác nhau.

D (cm)	Đất Bazan				Đất xám phù sa cổ				RR_i
	N_{ij}	Y_{ij}	J	λ_{ij}	N_{ij}	Y_{ij}	J	λ_{ij}	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
10	4000	8	0	0,002	3500	7	1	0,002	1,00
16	4500	10	0	0,002	4500	10	1	0,002	1,00
22	5000	15	0	0,003	5500	13	1	0,002	0,67
28	5500	22	0	0,004	6500	15	1	0,002	0,50
34	6000	28	0	0,005	7500	19	1	0,003	0,60
40	6500	32	0	0,005	8500	25	1	0,003	0,60
46	7000	45	0	0,006	9500	30	1	0,003	0,50
52	7500	46	0	0,006	10500	37	1	0,004	0,67

$$\lambda_{iJ} = \exp(b_0 + b_1 * J) \qquad (5.72)$$

$\lambda_{iJ} = \exp(b_0 + b_1*D)$ (5.73)

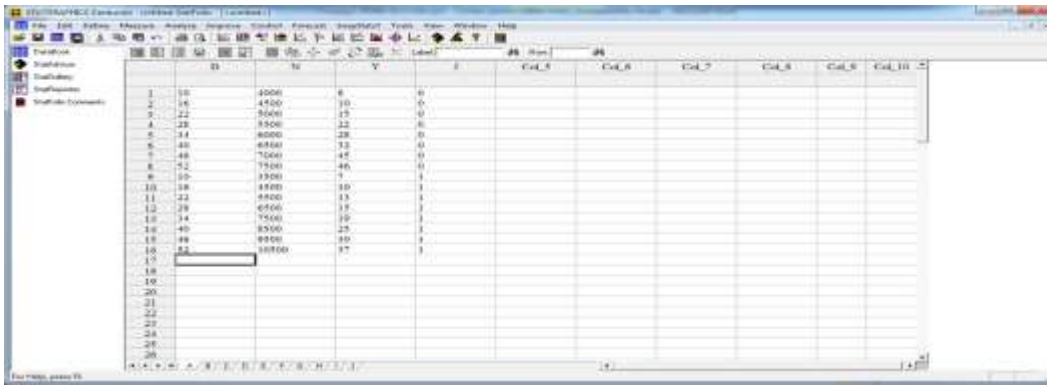
$\lambda_{iJ} = \exp(b_0 + b_1*J + b_2*D)$ (5.74)

Sau đó sử dụng những hàm này để so sánh tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột theo loại đất và cấp D. Khi sử dụng phần mềm Statgraphics, các hàm hồi quy này được xây dựng theo những bước dưới đây.

Bước 1. Tập hợp số liệu (Bảng 5.21). Cột 1 là cấp D; Cột 2 là dung lượng mẫu quan sát trong các cấp D; Cột 3 là số cây Bằng lăng ổi bị bệnh rỗng ruột trong các cấp D; Cột 4 là mã hóa loại đất J (0 = đất bazan; 1 = đất xám phù sa cổ). Cột 5 là tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột theo cấp D và loại đất.

Bảng 5.21. Số liệu để xây dựng các hàm ước lượng tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột theo các cấp D và loại đất.

Cấp D	N_{ij}	Y_{ij}	J	λ_{ij}
(1)	(2)	(3)	(4)	(5)
10	4000	8	0	0,002
16	4500	10	0	0,002
22	5000	15	0	0,003
28	5500	22	0	0,004
34	6000	28	0	0,005
40	6500	32	0	0,005
46	7000	45	0	0,006
52	7500	46	0	0,006
10	3500	7	1	0,002
16	4500	10	1	0,002
22	5500	13	1	0,002
28	6500	15	1	0,002
34	7500	19	1	0,003
40	8500	25	1	0,003
46	9500	30	1	0,003
52	10500	37	1	0,004

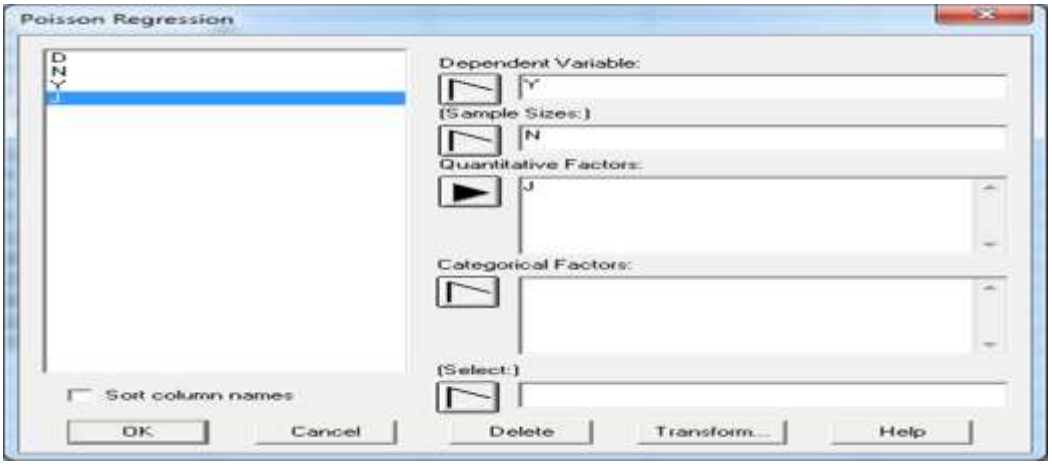


Hình 5.30. Số liệu để xây dựng hàm hồi quy Poisson $\lambda_{iJ} = f(D, J)$.

Bước 2. Copy 4 biến D, N, Y và J rồi chuyển chúng vào 4 cột khác nhau của bảng tính Statgraphics (Hình 5.30).

Bước 3. Xây dựng hàm hồi quy $\lambda_{iJ} = \exp(b_0 + b_1*J)$ (Hàm 5.72). Trước hết vào đường dẫn: Improve > Regression Analysis > Attribute Data > Piosson Regression. Đến đây một cửa sổ xuất hiện. Tại cửa sổ này, chọn Y, N và J rồi chuyển chúng vào ô Dependent Variable, Sample Sizes và Quantitative Factor (Hình 5.31). Sau đó chọn OK để nhận kết quả (Bảng 5.22). Tương tự, bằng cách thay thế J = D, nhận được hàm hồi quy $\lambda_{iJ} = \exp(b_0 + b_1*D)$ (Bảng 5.23).

Bước 4. Xây dựng hàm hồi quy $\lambda_{iJ} = \exp(b_0 + b_1*J + b_2D)$ (Hàm 5.74). Tại cửa sổ ở Hình 5.30, chọn D rồi chuyển nó vào ô Quantitative Factor (Hình 5.32). Sau đó chọn OK để nhận kết quả (Bảng 5.24).



Hình 5.31. Xác định hàm hồi quy $\lambda_{iJ} = \exp(b_0 + b_1*J)$.

Bảng 5.22. Hàm ước lượng tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột trên đất bazan và đất xám phù sa cổ. Hàm hồi quy: $\lambda_{iJ} = \exp(b_0 + b_1*J)$.

Estimated Regression Model (Maximum Likelihood)

Parameter	Estimate	Standard Error	Estimated Rate Ratio
CONSTANT	-5.40852	0.0696733	
J	-0.47473	0.106135	0.622053

$\lambda_{iJ} = \exp(-5.40852 - 0.47473*J)$

Analysis of Deviance

Source	Deviance	Df	P-Value
Model	20.26	1	0.0000
Residual	30.8843	14	0.0058
Total (corr.)	51.1443	15	

Percentage of deviance explained by model = 39.6134

Adjusted percentage = 31.7924

Likelihood Ratio Tests

Factor	Chi-Squared	Df	P-Value
J	20.26	1	0.0000

Residual Analysis

	Estimation	Validation
n	16	
MSE	0.00003739	
MAE	0.000938437	
MAPE	29.0047	
ME	-4.33777E-17	
MPE	-10.8909	

Phân tích số liệu ở Bảng 5.22 cho thấy loại đất ảnh hưởng rõ rệt (r^2 = 39,6%; P < 0,01) đến tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột. So với đất bazan, tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột trên đất xám phù sa cổ chỉ bằng 62,2% (RR_i = 0,622053). Mô hình ước lượng tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột trên cả hai loại đất nhận sai số âm, trung bình -10,89% (MPE = -10,89%).

Bảng 5.23. Hàm ước lượng tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột theo cấp đường kính. Hàm hồi quy: $\lambda_{iJ} = \exp(b_0 + b_1*D)$.

Estimated Regression Model (Maximum Likelihood)

Parameter	Estimate	Standard Error	Estimated Rate Ratio
CONSTANT	-6.37087	0.169172	
D	0.0200481	0.00422851	1.02025

$\lambda_{iJ} = \exp(-6.37087 + 0.0200481*D)$

Analysis of Deviance

Source	Deviance	Df	P-Value
Model	23.6007	1	0.0000
Residual	27.5436	14	0.0163
Total (corr.)	51.1443	15	

Percentage of deviance explained by model = 46.1453

Adjusted percentage = 38.3243

Likelihood Ratio Tests

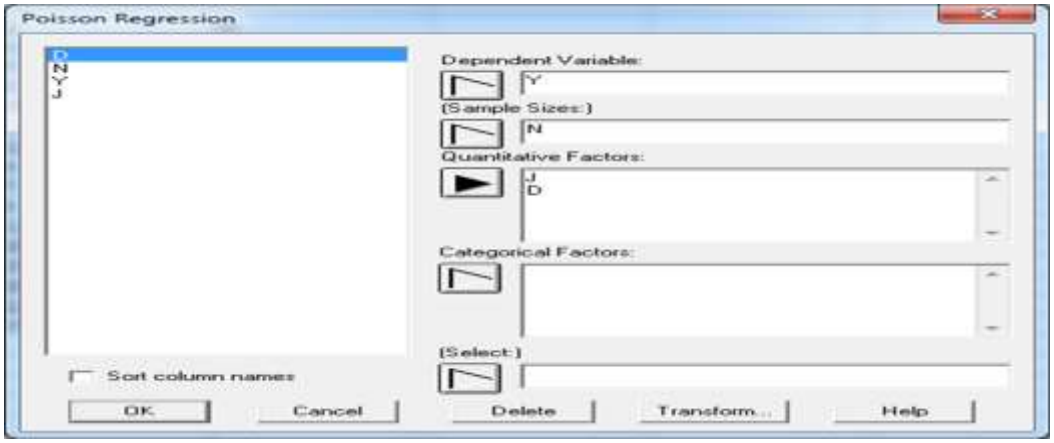
Factor	*Chi-Squared*	*Df*	*P-Value*
D	23.6007	1	0.0000

Residual Analysis

	Estimation	Validation
n	16	
MSE	0.0000339601	
MAE	0.00102659	
MAPE	26.5683	
ME	-0.00000302523	
MPE	-7.57822	

Số liệu ở Bảng 5.23 cho thấy tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột thay đổi rõ rệt theo kích thước thân cây (r^2 = 46,1%; P < 0,01). Trên cả hai loại đất, khi gia tăng một cấp D, thì tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột gia tăng trung bình 2,0% (RR_i = 1,02025). Sử dụng mô hình này để ước lượng tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột nhận sai số âm, trung bình -7,6% (MPE = -7,5782%).

Bước 5. Xác định những biến giải thích có ý nghĩa nhất trong hàm phản hồi quy $\lambda_{iJ} = \exp(b_0 + b_1*J + b_2D)$ (Hàm 5.74). Tại cửa sổ ở Hình 5.32, sau khi chọn OK, đặt con trỏ chuột tại bảng kết quả và nhấn nút chuột phải để chọn Analysis Option. Tại đây chọn Forward Selection ở cửa sổ Fit và All Steps ở ô Display (Hình 5.33). Lựa chọn này có nghĩa là xây dựng hàm hồi quy với hai biến giải thích J và D. Kết quả phân tích báo cáo chi tiết từng bước xử lý. Nếu chỉ cần xác định hàm hồi quy thích hợp cuối cùng, thì chọn Final Model Only. Sau đó, căn cứ vào những thống kê ở từng bước, phân tích ý nghĩa của từng biến trong mô hình và chọn những biến phù hợp nhất.



Hình 5.32. Xác định hàm hồi quy $\lambda_{iJ} = \exp(b_0 + b_1*J + b_2*D)$.

Bảng 5.24. Hàm ước lượng tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột theo loại đất và cấp đường kính. Hàm hồi quy: $\lambda_{iJ} = \exp(b_0 + b_1*J + b_2*D)$.

Estimated Regression Model (Maximum Likelihood)

Parameter	Estimate	Standard Error	Estimated Rate Ratio
CONSTANT	-6.16474	0.172815	
J	-0.508907	0.106311	0.601152
D	0.0212376	0.004228	1.02146

$\lambda_{iJ} = \exp(-6.16474 - 0.508907*J + 0.0212376*D)$

Analysis of Deviance

Source	Deviance	Df	P-Value	
Model	46.8081	2	0.0000	
Residual	4.33619	13	0.9871	*Hai biến J và D có ý nghĩa*
Total (corr.)	51.1443	15		

Percentage of deviance explained by model = 91.5217

Adjusted percentage = 79.7901

Likelihood Ratio Tests

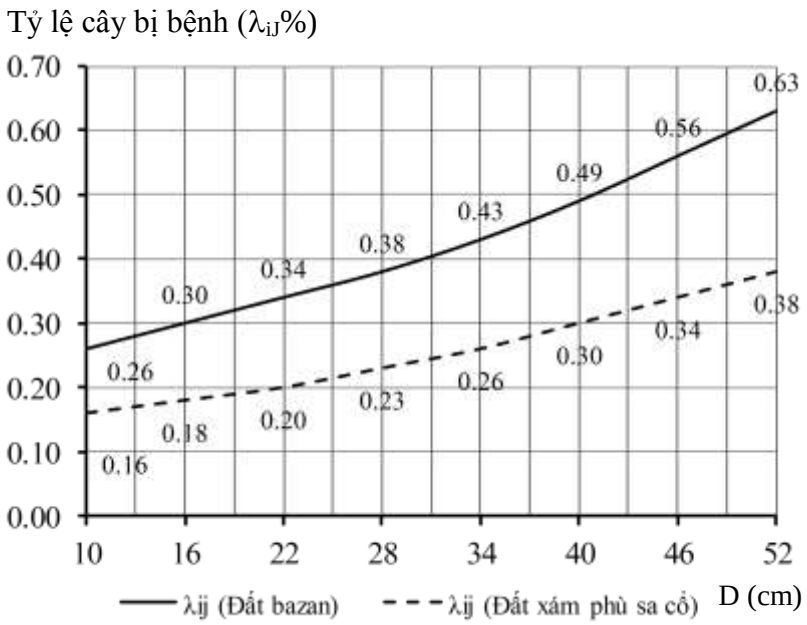
Factor	Chi-Squared	Df	P-Value
J	23.2074	1	0.0000
D	26.5481	1	0.0000

Residual Analysis

	Estimation	Validation

n	16
MSE	0.00000425654
MAE	0.000297143
MAPE	8.18885
ME	0.0000126394
MPE	-1.22672

Hình 5.33. Xác định những biến giải thích phù hợp nhất.

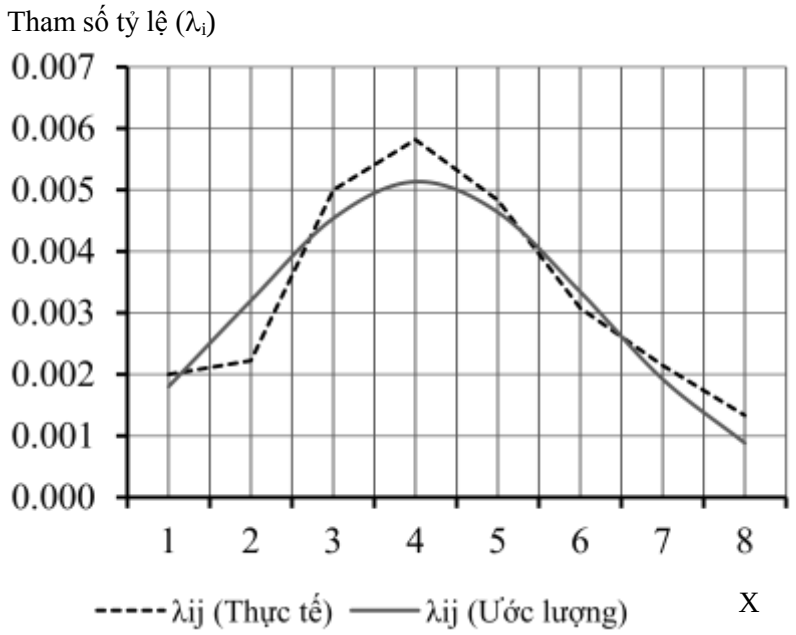


Hình 5.34. Tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột theo các cấp đường kính khác nhau trên đất bazan và đất xám phù sa cổ.

Số liệu ở Bảng 5.24 cho thấy tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột phụ thuộc chặt chẽ (R^2 = 91,5%; P < 0,001) vào loại đất và cấp D. Tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột trên đất xám chỉ bằng 60,1% (RR_i = 0,601152) so với đất bazan. Trên cả hai loại đất, tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột gia tăng theo cấp D, trung bình 2,1% (RR_i = 1,02146) (Hình 5.34). Sử dụng mô hình này để ước

lượng tỷ lệ số cây Bằng lăng ổi bị bệnh rỗng ruột nhận sai số âm, trung bình -1,2% (MPE = -1,22672).

Lưu ý rằng giữa biến phụ thuộc Y_i và các biến giải thích định lượng X_i có thể tồn tại mối quan hệ ở dạng bậc 2. Để kiểm định mối quan hệ này, tại cửa sổ ở Hình 5.33, đánh dấu Second Order và chọn OK để nhận kết quả. Sau đó phân tích và so sánh các thống kê tương quan và sai lệch giữa hàm bậc 1 và bậc 2 để chọn dạng hàm phù hợp. Hàm phù hợp được chọn theo tiêu chuẩn Max(R^2) và Min(MSE). Đồ thị mô tả quan hệ giữa tham số tỷ lệ λ_i với biến X_i có dạng như Hình 5.35.



Hình 5.35. Đồ thị biểu diễn mối quan hệ giữa tham số tỷ lệ λ_i với biến giải thích X.

Chương 6
MÔ HÌNH SINH TRƯỞNG VÀ SẢN LƯỢNG RỪNG

6.1. Khái niệm về sinh trưởng, tăng trưởng và sản lượng rừng

6.1.1. Sinh trưởng và tăng trưởng của cây gỗ và rừng

Sinh trưởng của cây gỗ và rừng biểu thị sự gia tăng kích thước của chúng theo thời gian. Tăng trưởng của cây gỗ và rừng biểu thị lượng gia tăng kích thước của chúng sau A năm. Kích thước của cây gỗ thường được quan tâm là đường kính thân ngang ngực (D, cm), chiều cao toàn thân (H, m), tiết diện ngang thân (g, m^3), thể tích thân (V, m^3) và sinh khối (B, kg hay tấn). Kích thước quần thụ thường được quan tâm là mật độ (N, cây/ha), tiết diện ngang (G, m^2/ha), trữ lượnng gỗ (M, m^3/ha) và sinh khối (B, tấn/ha). Trong lâm học, người ta phân biệt 3 loại tăng trưởng: hàng năm, trung bình năm và trung bình theo định kỳ n năm. Lượng tăng trưởng thường xuyên hàng năm biểu thị sự thay đổi về kích thước của cây gỗ (ZY, với Y = D, H, g, V) và quần thụ (ZY, với Y = G và M) sau 1 năm. Lượng tăng trưởng trung bình năm (ΔY) biểu thị sự thay đổi kích thước trung bình của cây gỗ và quần thụ sau thời kỳ A năm. Lượng tăng trưởng thường xuyên theo định kỳ n năm (ZYn) biểu thị sự thay đổi kích thước trung bình của cây gỗ và quần thụ theo định kỳ n năm.

Gọi Y_1 và Y_2 tương ứng là kích thước của cây gỗ và rừng ở thời kỳ bắt đầu (A_1, năm) và thời kỳ kết thúc (A_2, năm) quá trình nghiên cứu tăng trưởng. Thông thường, thời gian dãn cách giữa hai lần (n) xác định tăng trưởng của cây gỗ và rừng là 2, 5 và 10 năm tương ứng với cây mọc nhanh, trung bình và chậm. Ba đại lượng ZY, ΔY và ZY_n được xác định theo công thức 6.1 – 6.3.

$$ZY = (Y_A - Y_{A-1}) \quad (6.1)$$

$$\Delta Y = Y_A/A \quad (6.2)$$

$$ZYn = (Y_2 - Y_1)/(A_2 - A_1) \quad (6.3)$$

Hai đại lượng ZY và ΔY (Bảng 6.1) gia tăng dần và đạt cao nhất ở một tuổi nào đó; sau đó giảm dần theo A. Lượng tăng trưởng thường xuyên hàng năm (ZY) luôn nhận giá trị dương. Lượng tăng trưởng trung bình sau định kỳ n năm (ZYn) (Bảng 6.2) cũng gia tăng dần và đạt cao nhất ở tuổi nào đo; sau đó giảm dần theo A. Theo thời gian, ZYn có thể nhận giá trị bằng 0 hoặc âm. Đại lượng ZYn nhận giá trị âm khi thể tích thân cây bị mất mát do nhiều nguyên nhân khác nhau (Côn trùng, Nấm hại, Cây già cỗi và chết). Giá trị ZYn_{Max} là điểm uốn của đường cong tăng trưởng. Tuổi tương ứng với ZYn_{Max} cho biết thời kỳ mà từ đó trở đi tăng trưởng của rừng bị suy giảm. Trong lâm học, người ta xem đây là thời kỳ chuyển tiếp từ giai đoạn rừng non (thời kỳ sinh trưởng nhanh) sang giai đoạn rừng trung niên (thời kỳ sinh trưởng chậm).

Bảng 6.1. Quá trình sinh trưởng chiều cao hàng năm của rừng Đước từ 2 – 12 tuổi.

A (năm)	H (m)	ZH (m/năm)	ΔH (m/năm)	Ph(%)
(1)	(2)	(3)	(4)	(5)
2	1,5	0,77	0,77	50,0
3	3,6	2,05	1,20	57,1
4	5,5	1,88	1,37	34,3
5	7,0	1,56	1,40	22,2
6	8,3	1,27	1,38	15,3
7	9,3	1,04	1,33	11,2
8	10,2	0,86	1,27	8,5
9	10,9	0,72	1,21	6,6
10	11,5	0,61	1,15	5,3
11	12,1	0,53	1,10	4,4
12	12,5	0,46	1,04	3,6

Đại lượng ZY là một chỉ tiêu khó đo đạc, nhất là những loài cây sinh trưởng chậm. Trong lâm lâm học, ZYn được sử dụng để thay thế ZY nhằm tính lượng tăng trưởng theo phần trăm. Lượng tăng trưởng theo phần trăm là cơ sở để tính tốc độ gia tăng sản lượng rừng hay vốn rừng. Tuổi rừng mà tại đó ZYn = ΔY được gọi là tuổi thành thục số lượng (Bảng 6.3; Hình 6.1). Đó là tuổi khai thác rừng cho sản lượng thu hoạch trung bình năm cao nhất.

Bảng 6.2. Tăng trưởng định kỳ 2 năm về chiều cao của rừng Đước từ 2 – 12 tuổi.

A (năm)	H (m)	ZH (m/năm)	ΔH (m/năm)	Ph%
(1)	(2)	(3)	(4)	(5)
2	1,5	0,77	0,77	50,0
4	5,5	1,96	1,37	35,9
6	8,3	1,41	1,38	17,0
8	10,2	0,95	1,27	9,3
10	11,5	0,67	1,15	5,8
12	12,5	0,49	1,04	3,9

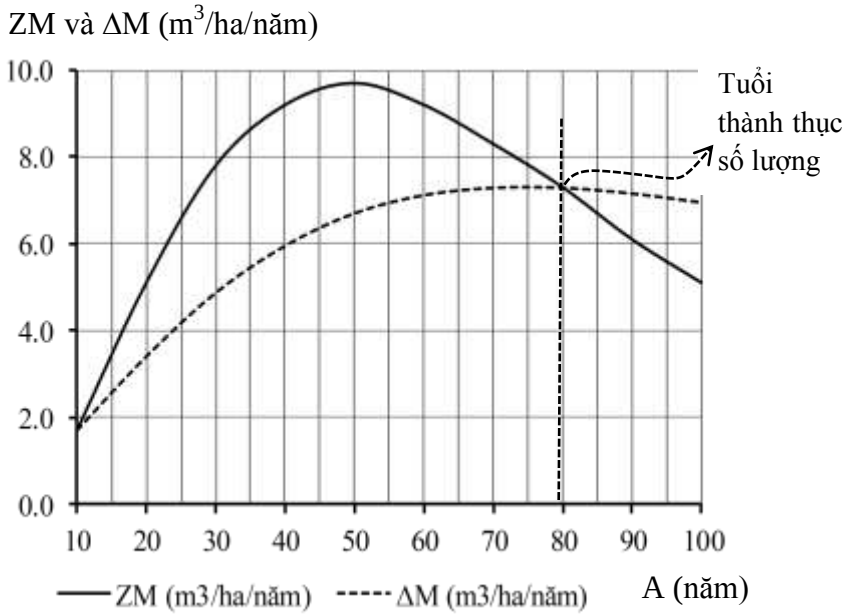
6.1.2. Sản lượng của cây gỗ và rừng

Sản lượng biểu thị kích thước của cây cá thể (D, H, g, V, B) và quần thụ (N, G, M, B) đạt đến tuổi nhất định (A, năm). Sản lượng là tổng khối lượng (gỗ, sinh khối…) mà cây gỗ và quần thụ tạo ra sau một khoảng thời gian nào đó (Công thức 6.4).

$$\text{Sản lượng} = A*\Delta Y = \sum_{(i=1,A)} ZY \qquad (6.4)$$

Bảng 6.3. Quá trình sinh trưởng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 100 tuổi.

A (năm)	M (m^3/ha)	ZM (m^3/ha/năm)	ΔM (m^3/ha/năm)	PM%
(1)	(2)	(3)	(4)	(5)
10	17	1,7	1,7	10,0
20	68	5,1	3,4	7,5
30	146	7,8	4,9	5,3
40	238	9,2	6,0	3,9
50	335	9,7	6,7	2,9
60	427	9,2	7,1	2,2
70	510	8,3	7,3	1,6
80	583	7,3	7,3	1,3
90	644	6,1	7,2	0,9
100	695	5,1	7,0	0,7



Hình 6.1. Tăng trưởng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 100 tuổi.

Sản lượng được phân chia thành sản lượng cây đứng hay cây còn sống và sản lượng thu hoạch. Sản lượng cây đứng là tổng sản lượng trên cây còn sống ở cuối thời kỳ nhất định. Sản lượng thu hoạch là tổng sản lượng thu hoạch được ở cuối thời kỳ nhất định. Trong lâm học, sản lượng gỗ và sinh khối là mối quan tâm lớn nhất. Chẳng hạn: Tổng sản lượng cây đứng của rừng Keo lai ở tuổi 10 trên cấp lập địa I là

250 m^3/ha, còn tổng sản lượng thu hoạch ở tuổi 10 là 200 m^3/ha. Sản lượng cây đứng luôn cao hơn sản lượng thu hoạch. Nguyên nhân là vì một bộ phận sản lượng của cây gỗ và rừng không được thu hoạch.

Thông thường, sản lượng được xác định theo luân kỳ kinh doanh đối với rừng hỗn giao khác tuổi hay chu kỳ kinh doanh đối với rừng thuần loài đồng tuổi. Chẳng hạn: Tổng sản lượng gỗ của rừng Thông ba lá (*Pinus kesiya* Royle ex Gordon) thu được sau thời gian 10 năm là 250 m^3/ha/10 năm. Sản lượng gỗ tạo ra sau 1 năm trên diện tích 1 ha được gọi là năng suất của rừng. Ở ví dụ trên, năng suất gỗ trung bình của rừng Thông ba lá là 25 m^3/ha/năm.

6.2. Phương pháp mô hình hóa sinh trưởng của cây gỗ và rừng

6.2.1. Mô hình sinh trưởng và sản lượng của cây gỗ và rừng

Mô hình sinh trưởng của cây cá thể mô tả quá trình sinh trưởng của nó theo thời gian. Chẳng hạn: Quá trình sinh trưởng của Y (Y = D, H, V…) theo thời gian (A, năm). Quá trình trình biến đổi này có thể biểu thị bằng hàm Y = f(A). Mô hình sinh trưởng của quần thụ mô tả quá trình biến đổi của nó theo thời gian. Chẳng hạn: Quá trình biến đổi của N, G và M theo A (năm). Quá trình trình biến đổi này có thể biểu thị bằng hàm Y = f(A). Để xây dựng các mô hình sinh trưởng ở mức cây cá thể và quần thụ, nguồn số liệu cần có là số liệu đo đếm từ những ô định vị hoặc số liệu từ điều tra rừng hàng năm hoặc số liệu từ những cây giải tích ở những tuổi khác nhau.

Các mô hình sinh trưởng mang lại nhiều ý nghĩa khác nhau. Về lâm học, các mô hình sinh trưởng được sử dụng để đánh giá chất lượng lập địa, xác định ảnh hưởng của các phương thức lâm sinh đến chất lượng lập địa và giải thích kết quả của những hoạt động lâm sinh (nuôi dưỡng và khai thác rừng). Trong điều tra rừng, các mô hình sinh trưởng được sử dụng để ước lượng và dự đoán sản lượng rừng, xây dựng biểu thể tích và biểu sản lượng rừng. Trong quản lý rừng, các mô hình sinh trưởng được sử dụng để xây dựng chiến lược về quản lý và khai thác rừng. Trong kinh tế lâm nghiệp, các mô hình sinh trưởng gỗ thu hoạch được sử dụng để phân tích hiệu quả kinh doanh rừng và xác định chu kỳ khai thác rừng tối ưu về kinh tế.

6.2.2. Các kiểu mô hình sinh trưởng

Các mô hình sinh trưởng có thể được phân chia thành mô hình sinh trưởng ở mức cây cá thể và mô hình sinh trưởng ở mức quần thu. Mô hình sinh trưởng ở mức cây cá thể biểu thị quá trình sinh trưởng của cây cá thể trong những điều kiện môi trường (lập địa) khác nhau. Chẳng hạn: Mô hình D = f(A), H = f(A), H = f(D) trên những lập địa khác nhau. Mô hình sinh trưởng ở mức quần thụ biểu thị quá trình sinh trưởng của quần thụ trong những điều kiện môi trường khác nhau. Chẳng hạn: Mô hình N = f(A), G = f(A) và M = f(A) trên những lập địa khác nhau. Các mô hình ở mức cây cá thể và quần thụ cung cấp những thông tin về quá trình biến đổi những đặc tính của chúng theo thời gian. Những thông tin này có thể được xây dựng thành biểu quá trình sinh trưởng của quần thụ.

6.2.3. Những thành phần của mô hình sinh trưởng

Những thành phần của mô hình sinh trưởng phụ thuộc vào đối tượng nghiên cứu. Ở mức cây cá thể, những thành phần của mô hình sinh trưởng bao gồm những thành phần mô tả sự biến đổi kích thước (D, H, g, V, B) theo tuổi (A, năm). Ở mức quần thụ, những thành phần của mô hình sinh trưởng bao gồm những thành phần mô tả sự biến đổi kích thước (N, G, M, B) theo A. Những thành phần của mô hình sinh trưởng cần phải được xác định ở những thời điểm khác nhau trong đời sống của cây cá thể và quần thụ.

6.2.4. Chọn mô hình sinh trưởng và mô hình sản lượng

Các mô hình sinh trưởng và mô hình sản lượng được chọn tùy theo mục tiêu nghiên cứu và mục đích sử dụng. Đối với mục đích nghiên cứu, các mô hình sinh trưởng và mô hình sản lượng được xây dựng để phân tích động thái biến đổi của những đặc tính ở mức cây gỗ và rừng theo thời gian và phân tích ảnh hưởng của môi trường đến cây gỗ và rừng. Ngoài ra, những mô hình sinh trưởng và mô hình sản lượng trong những môi trường khác nhau là cơ sở khoa học để xây dựng lý thuyết về rừng và những nguyên lý lâm sinh. Đối với mục đích sử dụng, các mô hình sinh trưởng và mô hình sản lượng được sử dụng để đánh giá chất lượng lập địa, kiểm định giả thuyết về tăng trưởng, ước lượng sản lượng gỗ, nghiên cứu biến động sản lượng gỗ, đánh giá các hoạt động lâm sinh, xây dựng phương thức quản lý rừng…

6.3. Mô hình sinh trưởng quần thụ

6.3.1. Khái quát chung

Đối tượng nghiên cứu của lâm học chủ yếu là cây gỗ hay quần thụ. Trước đây sinh trưởng và sản lượng của quần thụ được xác định thông qua các biểu quá trình sinh trưởng của quần thụ. Chúng được xây dựng dựa trên những nghiên cứu về đặc tính của quần thụ theo thời gian. Biểu sản lượng quần thụ tóm tắt quá trình biến đổi sản lượng quần thụ theo những cấp tuổi khác nhau. Biểu sản lượng quần thụ có thể bao gồm cả N, D, H, G, M, phân bố N/D, phân bố N/H, tăng trưởng (D, H, G và M) của quần thụ. Các biểu sản lượng rừng được xây dựng từ những số liệu đo lặp lại trên cây mẫu và ô mẫu định vị. Các biểu sản lượng rừng có thể được xây dựng cho những cấp lập địa khác nhau. Bởi vì quần thụ thay đổi theo thời gian, theo phương thức lâm sinh và môi trường (lập địa), nên các biểu sản lượng quần thụ phản ánh không chính xác cho mọi trường hợp.

Ngày nay biểu quá trình sinh trưởng trưởng và sản lượng quần thụ được xây dựng dựa trên các mô hình sinh trưởng. Các mô này được xây dựng với các biến dự đoán ở mức quần thụ (N, D, H, V, G, M, phân bố N/D và phân bố N/H).

6.3.2. Hàm sinh trưởng và sản lượng lý thuyết

Bertalanffy (1968; Dẫn theo Vanclay, 1994) đã giả định tăng trưởng của cây gỗ và quần thụ là sự khác biệt giữa quang hợp và hô hấp. Nếu biểu thị Y là quang hợp của cây gỗ và quần thụ, thì tăng trưởng có thể được mô tả bằng hàm (6.5); trong đó giá trị α và γ tương ứng là hằng số quang hợp và hằng số hô hấp. Giữa α và γ có quan hệ chặt chẽ với nhau. Trong thực hành, tham số γ = 1.

$$dY/dA = (b_1Y^{\alpha} - b_2Y^{\gamma}) \qquad (6.5)$$

Moser và Hall (1969; Dẫn theo Vanclay, 1994) đã giả định tăng trưởng M là một hàm của G. Lượng tăng trưởng G được dự đoán từ G quần thụ. Từ đó tăng trưởng M quần thụ có dạng hàm (6.6). Tăng trưởng M là tích phân hàm tăng trưởng G và bổ sung thêm trữ lượng gỗ.

$$\Delta M = \beta_1 M(\beta_2 G^{\beta 3-1} - \beta_4) \qquad (6.6)$$

García (1974; Dẫn theo Vanclay, 1994) đã xây dựng hàm sinh trưởng và sản lượng quần thụ dựa trên những giả định sau đây:

(a) Tăng trưởng trữ lượng gỗ (ΔM) tỷ lệ với tăng trưởng sinh khối khô (ΔB).
(b) Mật độ quần thụ (N, cây/ha) trên các lập địa rất cao và bằng hằng số.
(c) Hô hấp tỷ lệ với diện tích bề mặt thân cây. Diện tích bề mặt thân cây tỷ lệ với tích số (N*D*H); trong đó D và H là giá trị bình quân của quần thụ.

Từ những giả định trên đây, ΔM có thể xác định theo hàm (6.7).

$$\Delta M = b_0 - b_1(N*D*H) \qquad (6.7)$$

Giả định hình số (F) là hằng số. Trữ lượng quần thụ M = G*H*F; trong đó G = $(\pi/4)*N*D^2$. Giá trị D bình quân = $(2/\sqrt{(\pi*F)})*\sqrt{(M/(N*H))}$. Từ đó ΔM có thể biểu thị bằng hàm (6.8).

$$\Delta M = b_0 - b_1*\sqrt{(M*N*H)} \qquad (6.8)$$

6.3.3. Hàm sinh trưởng và sản lượng thực nghiệm

Trước khi xây dựng mô hình sinh trưởng, chúng ta cần phải xem xét những vấn đề sau đây:

(a) Mục đích xây dựng mô hình sinh trưởng?
(b) Số liệu đầu vào có phù hợp với việc xây dựng mô hình sinh trưởng?
(c) Số liệu đầu ra đòi hỏi như thế nào?
(d) Các nguồn số liệu để kiểm định độ tin cậy của các mô hình sinh trưởng?

Các mô hình sinh trưởng phải tuân thủ hai nguyên tắc cơ bản: (1) Mô hình sinh trưởng không bao gồm những biến số khó xác định; (2) Mô hình sinh trưởng đơn giản và dễ áp dụng trong thực tế.

Khi thiết kế mô hình sinh trưởng, nhà lâm học cần phải hiểu rõ những phương pháp xây dựng các mô hình sinh trưởng và những yêu cầu đối với việc xây dựng mô hình sinh trưởng. Trước khi xây dựng mô hình sinh trưởng, nhà lâm học

cần phải xác định hình thái của mô hình sinh trưởng thông qua các đồ thị, mối liên hệ giữa biến phụ thuộc với biến độc lập và sự phù hợp của mô hình sinh trưởng so với số liệu thực nghiệm. Các mô hình sinh trưởng bao gồm các biến phụ thuộc (biến phản hồi), các biến độc lập (biến dự đoán, biến giải thích) và các biến phân loại hay lập nhóm (Grouping Variables).

Các biến phụ thuộc (biến phản hồi) là các biến ước lượng và dự đoán. Chúng phản ánh sinh trưởng của cây cá thể và quần thụ. Các biến phản hồi có thể là D, H, V ở mức cây cá thể và N, G và M ở mức quần thụ. Các biến phản hồi cũng có thể là lượng tăng trưởng ở mức cây cá thể (ZD và ΔD; ZH và ΔH, Zg và Δg; ZV và ΔV; ZB và và ΔB) và quần thụ (ZG và ΔG; ZM và ΔM; ZB và và ΔB).

Các biến độc lập là các biến được chọn để ước lượng, dự đoán và giải thích sự biến đổi của các biến phản hồi. Các biến độc lập được chọn là các biến có ý nghĩa và dễ đo đạc. Các biến độc lập thông thường là tuổi (A), nhân tố khí hậu, độ cao địa hình, chỉ số lập địa (SI)...Tuy vậy, về mặt sinh học, các biến dự đoán thường không độc lập với nhau. Các biến độc lập hữu ích được xác định thông qua những kiểm định thống kê.

Biến phân loại là các biến nhị phân nhận giá trị 0 và 1. Những biến này dùng để "Ngắt" hay "Lọc" biến phản hồi theo những tình huống nhất định. Chẳng hạn: Xây dựng mô hình sinh trưởng của quần thụ trên hai cấp lập địa khác nhau.

Các biến phản hồi và các biến giải thích được chọn theo yêu cầu đặt ra. Những kiến thức về rừng cho phép nhà lâm học không chỉ xây dựng những giả thuyết, mà còn xác định các biến phản hồi, các biến giải thích và hình thái của những mô hình sinh trưởng. Mặc dù các mô hình toán phù hợp với số liệu, nghĩa là nhận sai lệch nhỏ, nhưng chúng bị bác bỏ vì phản ánh không chính xác quy luật sinh trưởng của cây gỗ và quần thụ. Chẳng hạn: Nếu hàm mô tả quá trình sinh trưởng trữ lượng gỗ (M, m^3/ha) của toàn bộ đời sống quần thụ có dạng hàm lũy thừa ($M = bA^c$), thì kết quả chỉ ra M tăng liên tục theo A. Kết luận này vi phạm nguyên lý sinh thái thực vật. Theo nguyên lý này, đồ thị của hàm M = f(A) phải có dạng đường cong Sigmoid. Vì thế, mô hình này bị bác bỏ. Nói chung, kiến thức về thực vật và rừng giúp cho nhà lâm học chọn các mô hình phù hợp với quy luật sống của cây gỗ và rừng.

Các hàm sản lượng ở mức quần thụ được xây dựng dựa trên giả định quần thụ đã được quản lý theo một phương thức lâm sinh nào đó. Trái lại, các hàm tăng trưởng có thể được xây dựng cho tất cả những quần thụ đã qua khai thác hoặc những phương thức lâm sinh nào đó.

Buell (1945; Dẫn theo Vanclay, 1994) đã ước lượng tăng trưởng của rừng hỗn giao khác tuổi bằng cách nhóm các quần thụ thành các cấp; trong đó mỗi cấp quần thụ có sự tương đồng về kết cấu loài cây gỗ, lịch sử lâm sinh, suất tăng trưởng và hình dạng thân cây. Tăng trưởng V thân cây cá thể được giả định là một hàm bậc 2 của D (Hàm 6.9). Tăng trưởng tổng trữ lượng gỗ của quần thụ (M) là tổng tăng trưởng thể tích (V) của N cây cá thể (Hàm 6.10).

$\Delta V = a + bD + cD^2$ (6.9)

$\sum\Delta M = aN + b\sum_{(i-1,n)}D + c\sum_{(i-1,n)}D^2$ (6.10)

Từ giả định tăng trưởng tiết diện ngang (ΔG) của rừng trồng đồng tuổi giảm dần theo tuổi, nhưng tăng dần theo chỉ số lập địa (SI = H ở tuổi cơ sở) và G quần thụ, Nelson (1963; Dẫn theo Vanclay, 1994) đã xây dựng hàm $\Delta G = f(A, S, G)$ theo dạng hàm (6.11).

$\Delta G = a + b(G/A) + cG^2 + d(G^2/A) + e(SI*G^2)$ (6.11)

Theo Clutter (1963; Dẫn theo Vanclay, 1994), hàm tăng trưởng M quần thụ có dạng hàm (6.12); trong đó M = trữ lượng quần thụ, SI = chỉ số lập địa, G = tiết diện ngang quần thụ, A = thời gian dự đoán M.

$Ln(M) = b_0 + b_1*SI + b_2Ln(G) + (b_3/A)$ (6.12)

Tăng trưởng của quần thụ khác tuổi có thể được xác định bằng cách xây dựng một hệ thống các hàm đối với các thành phần quần thụ. Sau đó liên kết các hàm này để ước lượng trữ lượng quần thụ. Theo đó, Furnival và Wilson (1971; Dẫn theo Vanclay, 1994) đã phát triển mô hình sinh trưởng và sản lượng rừng bằng hệ thống các hàm (6.13) – (6.18). Ở các hàm (6.13) – (6.18), H (m) = chiều cao trung bình của quần thụ; A (năm) = tuổi bình quân ở D ngang ngực của những cây hình thành quần thụ; D_g (cm) = đường kính của cây có tiết diện ngang bình quân; F = hình số; k = hằng số (π/40.000), M = trữ lượng quần thụ.

$Ln(H) = b_1 + b_2Ln(A)$ (6.13)

$Ln(N) = b_3 + b_4Ln(D_g)$ (6.14)

$Ln(G) = b_5 + b_6Ln(H)$ (6.15)

$Ln(F) = b_7 + b_8Ln(D_g)$ (6.16)

$Ln(G) = Ln(k) + Ln(N) + 2*Ln(D_g)$ (6.17)

$Ln(M) = Ln(F) + Ln(H) + Ln(G)$ (6.18)

Mendoza và Gumpal (1987; Dẫn theo Vanclay, 1994) đã xây dựng mô hình ước lượng trữ lượng gỗ của rừng Sao Dầu sau khai thác chọn ở Philippines theo dạng hàm 6.19; trong đó G_0 = tiết diện ngang của quần thụ tăng thêm sau khai thác, A = thời gian sau khai thác, SI = chỉ số lập địa được ước lượng theo H bình quân của những cây có D = 50 – 80 cm.

$Ln(M) = b_0 + b_1*G_0 + b_2*Ln(A) + b_3*(SI/A)$ (6.19)

Vanclay (1988; Dẫn theo Vanclay, 1994) đã xây dựng hàm tăng trưởng tiết diện ngang của rừng Thông (*Callitris* spp.) không đồng tuổi ở Queensland theo dạng hàm (6.20); trong đó SI là chỉ số lập địa, G là tiết diện ngang quần thụ. Chỉ số SI được ước lượng dựa theo mối quan hệ giữa H và D.

$Ln(\Delta G) = a + b*Ln(G) + c(G*SI) + d*G$ (6.20)

Trong nhiều tài liệu lâm học, người ta còn biểu diễn sự biến đổi thể tích thân trên những lập địa khác nhau bằng các hàm 6.21 – 6.23.

$V = b_0 + b_1*A + b_2*A^2 + b_3*(D^2*H) - b_4*(SI/H)$ (6.21)

$Ln(V+1) = b_0 + b_1*Ln(D) + b_2*Ln(H) + b_3*Ln(SI/H)$ (6.22)

$V = b_0 + b_1*(D^2*H) - b_2*(SI/H)$ (6.23)

Hai biến ZY và ΔY (Y = D, H, G, V, M) được mô tả bằng hàm 6.24 và 6.25.

$ZY = a_0 + a_1*A + a_2*A^2 + a_3*Y + a_4*Y^2$ (6.24)

$\Delta Y = b_0 + b_1*A + b_2*A^2 + b_3*Y + b_4*Y^2$ (6.25)

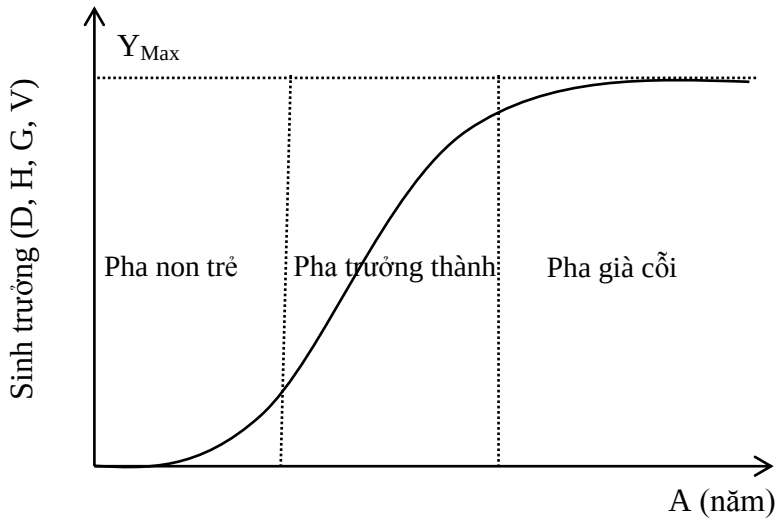
Nói chung, các hàm thực nghiệm là biểu thức mô tả mối quan hệ giữa biến phản hồi và biến giải thích (nguyên nhân). Các hàm thực nghiệm có nhiều dạng khác nhau như hàm tuyến tính và hàm phi tuyến tính. Các hàm tuyến tính không nhất thiết là các đường thẳng, nhưng các biến giải thích trong các hàm tuyến tính tồn tại ở dạng đường thẳng. Đối với hàm tuyến tính đa biến (Hàm 6.26), Y = biến phản hồi (D, H, V), X_i = biến giải thích hay biến dự đoán (A, SI), e = sai số.

$Y = b_0 + b_1X_1 + b_2X_2 + \ldots + b_kX_k + e$ (6.26)

Trong một số trường hợp, các biến phản hồi có thể ở các dạng chuyển đổi khác nhau (Y, Y^2, LnY…). Tương tự, các biến giải thích cũng có thể ở các dạng chuyển đổi khác nhau (X, X^2, LnX) hoặc ở dạng tương tác (X_iX_j).

6.3.4. Những hàm thực nghiệm tương đương với hàm lý thuyết

Theo quy luật, đường cong mô tả quá trình sinh trưởng trong toàn bộ đời sống của cây gỗ và quần thụ thuần loài đồng tuổi có dạng đường cong Sigmoid hay đường cong hình chữ S. Đường cong này có 2 điểm uốn, hai tiệm cận tại Y = 0 khi A = 0 và $Y = Y_{Max}$ khi A = ∞. Hai điểm uốn phân chia đời sống của cây gỗ và quần thụ thành ba pha: non trẻ, trưởng thành và già cỗi (Hình 6.2).



Hình 6.2. Các pha sinh trưởng của cây gỗ và quần thụ thuần loài đồng tuổi.

Pha non trẻ bắt đầu từ lúc rừng phát sinh cho đến khi rừng khép tán kín. Ở pha này, rừng sinh trưởng chậm. Trên đồ thị, đường cong biểu diễn quá trình sinh trưởng (Y = D, H, G, M) xuất phát từ gốc tọa độ, nghĩa là Y = 0 khi A = 0. Pha

trung niên kể từ khi rừng khép tán kín cho đến khi sinh sản mạnh. Ở pha này, chiều cao sinh trưởng chậm, còn đường kính và thể tích sinh trưởng nhanh. Trên đồ thị, đường cong sinh trưởng có dạng gần như đường thẳng với độ dốc lớn. Pha già cỗi (thành thục) kể từ sau khi rừng sinh sản mạnh cho đến khi ngừng sinh sản. Ở pha này, chiều cao ngừng tăng trưởng, còn D và V tăng trưởng rất chậm.

Theo quy luật sinh trưởng của cây gỗ và quần thụ, các hàm sinh trưởng thực nghiệm được chọn phải thỏa mãn 3 điều kiện: f(A) = 0 khi A = 0; 1 tiệm cận tại f(A) = Y_{Max} và 2 điểm uốn. Điểm uốn thứ nhất tương ứng với giai đoạn sinh trưởng chậm chuyển sang giai đoạn sinh trưởng nhanh (Giai đoạn rừng non chuyển sang giai đoạn rừng trung niên). Điểm uốn thứ hai tương ứng với giai đoạn sinh trưởng nhanh chuyển sang giai đoạn sinh trưởng chậm (Giai đoạn rừng trung niên chuyển sang giai đoạn rừng thành thục).

Nói chung, xác định những hàm thỏa mãn đầy đủ ba điều kiện trên đây là việc làm rất khó khăn. Trước đây nhiều nhà lâm học và điều tra rừng đã sử dụng những hàm thực nghiệm tương đương với hàm lý thuyết để biểu diễn gần đúng quá trình sinh trưởng của cây gỗ và quần thụ. Bảng 6.4 ghi lại một số hàm thực nghiệm thường được ứng dụng để mô tả quá trình sinh trưởng của cây gỗ và rừng.

Bảng 6.4. Một số hàm thực nghiệm thường được ứng dụng để mô tả quá trình sinh trưởng của cây gỗ và rừng.

TT	Tên hàm số	Năm	Hình thái của hàm
1	Gompertz	1825	Y = m*exp(-b*exp(-c*A))
2	Logistic	1838	Y = m/(1 + b*exp(-c*A))
3	Monomolecular		Y = m*(1 - b*exp(-c*A))
4	Schumarcher	1939	Y = m*exp(-b/A^c)
5	Korf	1929	Y = m*exp(-b*A^-c)
6	Korf	1939	Y = m*exp((b/(1-c))*A^(1-c))
7	Weibull		Y = m*(1 – exp(-b*A^c))
8	Bertalanffy	1957	Y = m*(1 – exp(-b*A)^3
9	Chapman – Richards	1959	Y = m*(1 – exp(-b*A))^c
10	Korsun - Strand	1964	Y = A^2/(b_0 + b_1*A + b_2*A^2)
11	Sloboda	1971	Y = m*exp(-b_1*exp(-b_2*A^b_3))
12	Lũy thừa		Y = m*A^b

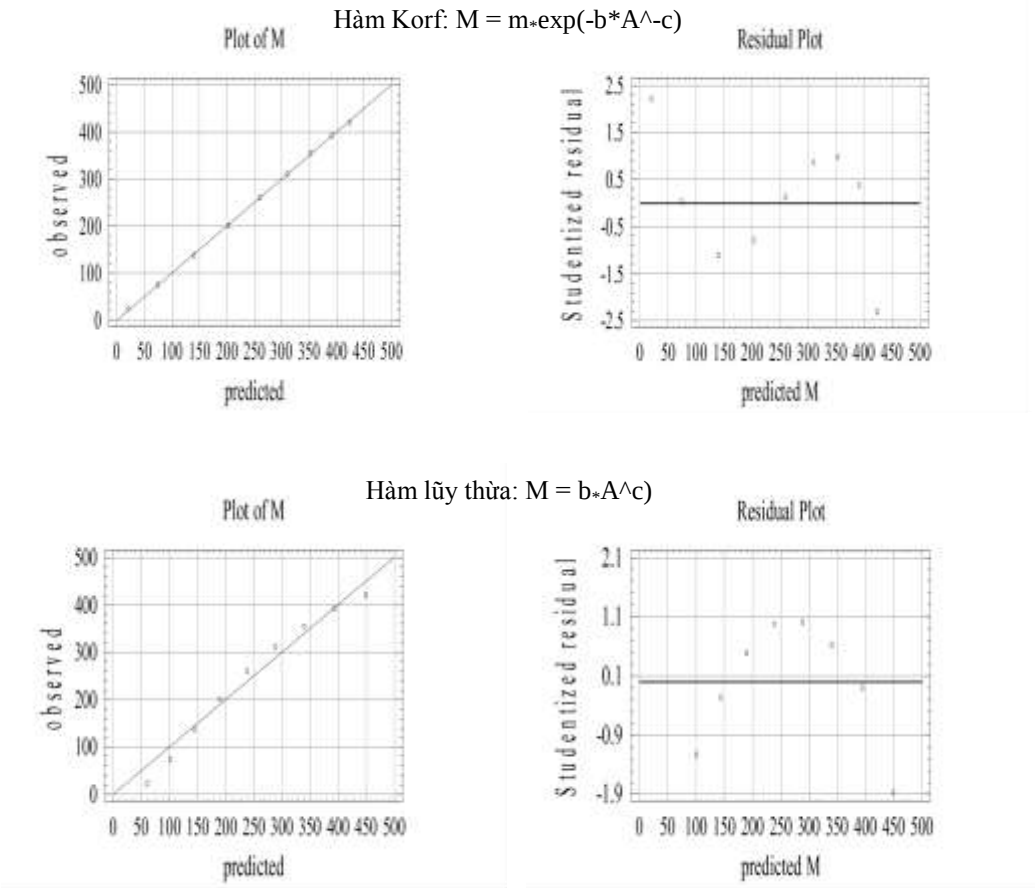
6.4. Phân tích sai lệch của các hàm sinh trưởng

Phân tích sai lệch là kiểm tra mức độ phù hợp của các hàm sinh trưởng so với số liệu thực tế. Từ đó, chọn hàm sinh trưởng tốt nhất. Về lý thuyết, sai số của một hàm thích hợp phân bố theo luật phân bố chuẩn và ngẫu nhiên. Trái lại, sai số của một hàm không thích hợp phân bố lệch chuẩn và không ngẫu nhiên.

Phân bố sai số có thể được kiểm tra bằng đồ thị sai lệch (Residual Plots) và đồ thị mô tả quan hệ giữa số liệu đầu vào (Observed Data) và số liệu đầu ra (Predicted Data). Nếu một hàm phù hợp, thì các điểm giao nhau giữa số liệu đầu vào và số liệu đầu ra nằm trên đường chéo của hình vuông. Nếu các điểm giao nhau

giữa số liệu đầu vào và số liệu đầu ra không nằm trên đường chéo của hình vuông, thì mô hình là kém phù hợp.

Hình 6.3 mô tả sai lệch của hàm Korf và hàm lũy thừa được sử dụng để ước lượng trữ lượng gỗ của rừng trồng Thông ba lá 40 tuổi. Từ Hình 6.3 cho thấy, so với hàm lũy thừa, hàm Korf phù hợp hơn bởi vì các điểm giao nhau giữa số liệu đầu vào và số liệu đầu ra đều nằm trên đường chéo của hình vuông.



Hình 6.3. Đồ thị so sánh sai lệch của hàm Korf và hàm lũy thừa được sử dụng để phân tích sinh trưởng trữ lượng gỗ của rừng Thông ba lá 40 tuổi.

Sai lệch của các hàm ước lượng sinh trưởng của cây gỗ và quần thụ có thể được đánh giá bằng nhiều tiêu chuẩn khác nhau; trong đó người ta thường sử dụng 10 tiêu chuẩn cơ bản: (1) Hệ số xác định = R^2 (Công thức 6.27); (2) Kiểm định F (Công thức 6.28); (3) Tổng sai lệch bình phương = SSE (Công thức 6.29); (4) Sai số chuẩn của ước lượng = SEE (Công thức 6.30); (5) Sai số tuyệt đối trung bình = MAE (Công thức 6.31); (6) Sai số tuyệt đối trung bình theo phần trăm = MAPE (Công thức 6.32); (7) Sai lệch trung bình = ME hay Bias (Công thức 6.33); (8) Chỉ số phù hợp (FI = Fit Index) (Công thức 6.34); (9) Tiêu chuẩn thông tin Akaike (Công thức 6.35); (10) Tiêu chuẩn U của Theil (Công thức 6.36). Ở công thức (6.27) – (6.36), Y_{TN} và Y_{UL} = giá trị của biến phụ thuộc thực nghiệm và ước lượng

từ hàm sinh trưởng; Y_{Bq} = Trung bình của biến phụ thuộc thực nghiệm; n = dung lượng quan sát; p = số lượng tham số của hàm sinh trưởng.

$$R^2 = [1- \frac{\sum(Y_{TN} - Y_{UL})^2}{\sum(Y_{TN} - Y_{Bq})^2}]*100 \quad (6.27)$$

$$F = \frac{MSR}{MSE} \quad (6.28)$$

MSR = SSE/p và MSE = SSE/(n - p).

$$SSE = \sum(Y_{TN} - Y_{UL})\text{^}2 \quad (6.29)$$

$$RMSE = SEE = \sqrt{SSE/(n - p)} \quad (6.30)$$

$$MAE = |((Y_{TN} - Y_{UL})/n))| \quad (6.31)$$

$$MAPE = (MAE/Y_{TN})*100 \quad (6.32)$$

$$Bias = ME = (Y_{TN} - Y_{UL}) \quad (6.33a)$$

$$Bias(\%) = MPE = \frac{(Y_{TN} - Y_{UL})}{Y_{TN}}*100 \quad (6.33a)$$

$$FI = 1 - \frac{(Y_{TN} - Y_{UL})^2}{(Y_{TN} - Y_{Bq})^2} \quad (6.34)$$

$$AIC = n*Ln(SSE/n) + 2*p \quad (6.35)$$

$$U = \frac{SEE}{\sqrt{Y^2_{UL}/n} + \sqrt{Y^2_{TN}/n}} \quad (6.36)$$

Kiểm định F là tiêu chuẩn tốt để xác định sự phù hợp của hàm sinh trưởng với nhiều biến dự đoán. Hệ số xác định R^2 = 0 – 1. Khi R^2 = 1, thì mô hình ước lượng phù hợp hoàn toàn với chuỗi số liệu. Tuy vậy, bởi vì R^2 không chỉ rõ sai số của mô hình, nên mô hình có R^2 = 1 không có nghĩa là mô hình tốt nhất. Độ lớn của R^2 gia tăng dần theo số lượng biến dự đoán trong các mô hình, mặc dù hệ số của một số biến không có ý nghĩa. Hệ số R^2 còn phụ thuộc vào sự chuyển đổi biến Y và X_i. Vì thế, hệ số R^2 là tiêu chuẩn kém chính xác để so sánh các mô hình khác nhau. Sai số trung bình (ME hay Bias) có thể nhận giá trị âm hoặc dương. Nếu n rất lớn thì ME nhận giá trị bằng Zero. Tiêu chuẩn ME thường được sử dụng để xác định sai số hệ thống của các hàm sinh trưởng. Nếu ME dương, thì giá trị dự đoán nhận sai số dương, nghĩa là giá trị dự đoán lớn hơn giá trị thực tế. Ngược lại, nếu ME âm, thì giá trị dự đoán nhận sai số âm, nghĩa là giá trị dự đoán nhỏ hơn giá trị thực tế. Bởi vì ME có thể nhận giá trị bằng Zero, nên nó cũng không thích hợp để so sánh các mô hình sinh trưởng. Tiêu chuẩn MAE và MAPE loại bỏ dấu của các sai lệch (ME, MPE). Vì thế, MAE và MAPE là tiêu chuẩn tốt để đánh giá và so sánh các mô hình. Mô hình tốt hơn là mô hình có MAE và MAPE nhỏ. Hạn chế của MAE là không cho biết chiều hướng của sai số. Chỉ số FI nhận giá trị từ 0 – 1. Hàm sinh trưởng thích hợp là hàm có điểm số FI lớn nhất. Tiêu chuẩn AIC là ước lượng sai lệch của hàm sinh trưởng. Hàm sinh trưởng thích hợp là hàm nhận AIC thấp nhất. Tuy vậy, tiêu chuẩn AIC chỉ thích hợp để so sánh các hàm sinh trưởng với ít biến số dự đoán. Tiêu chuẩn U của Theil là sai số tương đối và nhận giá trị từ 0 – 1. Hàm sinh trưởng thích hợp là hàm có điểm số U thấp nhất. Ba tiêu chuẩn SSE, MSE và SEE có quan hệ với nhau và chúng đều biểu thị sai số của mô hình. Vì thế, chúng là những tiêu

chuẩn thích hợp để đánh giá và so sánh các hàm sinh trưởng. Hàm sinh trưởng thích hợp là hàm có SSE_{Min}, MSE_{Min} và SEE_{Min}.

Quy tắc quyết định: Đối với hàm sinh trưởng tốt nhất hay hàm sinh trưởng tối ưu, $R^2 = 1$; SEE = 0; MAE = 0; MAPE = 0; SSE = 0; FI = 1; U = 0. Vì thế, khi chọn nhiều hàm sinh trưởng để so sánh, thì hàm sinh trưởng phù hợp nhất là hàm có R^2_{Max}, SEE_{Min}, MAE_{Min}, $MAPE_{Min}$, SSE_{Min}, FI_{Max}, U_{Min}. Tuy vậy, những thống kê này lại thay đổi tùy theo hàm sinh trưởng. Một hàm có R^2_{Max} nhưng SEE lại cao hơn hàm khác. Mặt khác, hai hàm có R^2 như nhau nhưng hàm này nhận SEE cao hơn hàm khác. Vì thế, nếu dựa theo tất cả những thống kê này, thì chúng ta rất khó chọn ra hàm phù hợp nhất. Mục đích của xây dựng hàm sinh trưởng là xác định hàm ước lượng và dự đoán biến phụ thuộc từ những biến dự đoán với sai lệch nhỏ nhất. Theo mục đích này, hàm sinh trưởng phù hợp nhất là hàm nhận SSE_{Min} hoặc SEE_{Min}.

6.5. Phương pháp chọn các hàm sinh trưởng và sản lượng rừng

6.5.1. Hàm sinh trưởng thích hợp phụ thuộc vào các hàm dự tuyển

Quá trình sinh trưởng của cây gỗ và rừng có thể được mô tả bằng một số hàm khác nhau. Nếu sử dụng những hàm dự tuyển khác nhau, thì kết quả ước lượng và dự đoán quá trình sinh trưởng của cây gỗ và rừng là khác nhau. Để chứng minh nhận định này, dưới đây dẫn ví dụ so sánh kết quả ước lượng M của rừng Thông (*Pinus sylvestris*) 140 tuổi bằng 4 hàm dự tuyển khác nhau (Korf, Gompertz, Korsun-Strand và Chapman-Richrads). Hình thái của 4 hàm này được dẫn ra ở Bảng 6.4. Trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi được dẫn ra ở Bảng 6.5. Các hệ số hồi quy và những thống kê sai lệch của 4 hàm này được xác định bằng phương pháp hồi quy phi tuyến tính của Marquardt (Bảng 6.6).

Bảng 6.5. Số liệu trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi.

A (năm)	M (m^3/ha)	A (năm)	M (m^3/ha)
(1)	(2)	(3)	(4)
10	20	80	590
20	75	90	655
30	155	100	695
40	235	110	730
50	325	120	765
60	415	130	800
70	520	140	835

Số liệu ở Bảng $6.6_{(b)}$ cho thấy bốn hàm này đều có hệ số xác định rất cao; trong đó thấp nhất ở hàm Korf (r^2 = 99,71%), cao nhất ở hàm Korsun – Strand (r^2 = 99,92%). Hàm Korf nhận những giá trị SSE, SEE, MAE và MAPE cao nhất (tương ứng 2923,2; 16,3 cm; 13,0 cm và 10,7%), nhỏ nhất ở hàm Korsun – Strand (tương ứng 861,0; 8,8 cm; 7,1 cm và 3,1%). Bằng cách khảo sát 4 hàm ở Bảng $6.6_{(a)}$, xác

định được những đặc trưng sinh trưởng của rừng Thông (*Pinus sylvestris*) 140 tuổi (Bảng 6.7; Hình 6.4).

Bảng 6.6$_{(a)}$. Mô hình ước lượng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi bằng bốn hàm khác nhau.

Hàm	Phương trình:	
(1)	(2)	
Korf	$M = 1505,17*\exp(-49,7197*A^{-0,900264})$	(6.37)
Gompertz	$M = 874,43*\exp(-4,38861*\exp(-0,0299387*A))$	(6.38)
Korsun-Strand	$M = A^2/(5,8326 - 0,020041*A + 0,0010548*A^2)$	(6.39)
Chapman-Richards	$M = 926,855*(1-\exp(-0,0224639*A))^{2,57605}$	(6.40)

Bảng 6.6$_{(b)}$. Tương quan và sai lệch của bốn hàm ước lượng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi.

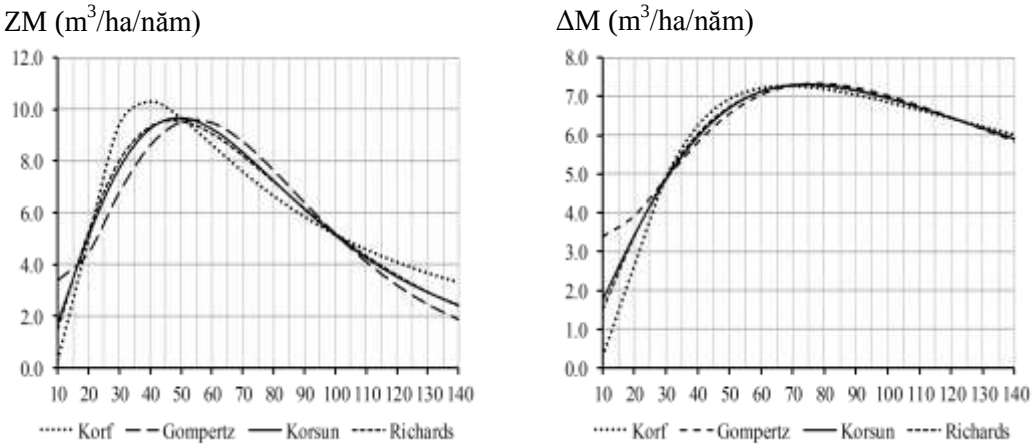
Thống kê	Hàm sinh trưởng:			
	Korf	Gompertz	Korsun-Strad	Chapman-Richards
(1)	(2)	(3)	(4)	(5)
r^2	99,71	99,89	99,92	99,90
SSE	2923,2	1080,3	861,0	1013,9
MSE	265,7	98,2	78,3	92,2
SEE	16,3	9,9	8,8	9,6
MAE	13,0	7,4	7,1	7,8
MAPE	10,7	6,6	3,1	4,2
Bias	2,02	-0,64	0,64	0,63
Bias%	7,84	-4,85	1,68	2,48
	(6.37)	(6.38)	(6.39)	(6.40)

Bảng 6.7. Những đặc trưng sinh trưởng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi được xác định bằng bốn hàm khác nhau.

Hàm	Tăng trưởng hàng năm:			Tăng trưởng bình quân năm:		
	ZM_{Max}	A (năm)	M (m^3/ha)	ΔM_{Max}	A (năm)	M (m^3/ha)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Korf	10,3	40	249,9	7,3	70	508,6
Gompertz	9,5	50	327,4	7,3	80	586,1
Korsun	9,7	50	334,8	7,3	80	510,5
Chapman	9,3	50	336,4	7,3	70	509,1

Phân tích số liệu ở Bảng 6.7 và Hình 6.4 cho thấy, nếu sử dụng 4 hàm (Korf, Gompertz, Korsun-Strand và Chapman - Richrads) để phân tích tăng trưởng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi, thì đại lượng ZM_{Max} đạt cao nhất ở hàm Korf (10,3 m^3/ha/năm), thấp nhất ở hàm Chapman – Richards (9,3 m^3/ha/năm). Thời điểm đạt ZM_{Max} hay điểm uốn của hàm Korf (40 năm) nhỏ hơn

một cấp so với ba hàm Gompertz, Korsun-Strand và Chapman-Richrads (50 năm). Mặc dù ΔM_{Max} ở cả 4 hàm này đều nhận giá trị tương tự như nhau (7,3 m^3/ha/năm), nhưng thời điểm đạt ΔM_{Max} ở hàm Korf và hàm Chapman-Richards (70 năm) nhỏ hơn một cấp so với hai hàm Gompertz và hàm Korsun-Strand (80 năm). Trữ lượng gỗ mà tại đó xuất hiện ZM_{Max} và ΔM_{Max} ở 4 hàm này cũng khác nhau khá lớn.



Hình 6.4. Đồ thị so sánh lượng tăng trưởng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi được ước lượng bằng bốn hàm khác nhau.

Từ những phân tích sai khác của 4 hàm dự tuyển trên đây cho thấy, với cùng nguồn số liệu như nhau, nếu hai nhà nghiên cứu chọn hai hàm khác nhau, thì kết quả ước lượng và dự đoán trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi là khác nhau. Bởi vì các nhà nghiên cứu có thể chọn những hàm dự tuyển khác nhau, nên kết quả ước lượng và dự đoán sinh trưởng của rừng cũng khác nhau. Đây là lý do dẫn đến những tranh luận về kết quả nghiên cứu. Ở đây kiến thức của nhà nghiên cứu và nguồn tài liệu tham khảo đóng vai trò quan trọng trong việc chấp nhận kết quả nghiên cứu.

6.5.2. Phương pháp phân tích hồi quy

6.5.2.1. Khái quát về phương pháp phân tích hồi quy

Phân tích hồi quy bao gồm ba phương pháp cơ bản. Một là phương pháp bình phương nhỏ nhất (Least Squares Method). Hai là phương pháp hồi quy phi tuyến tính (Nonlinear Regression Method). Ba là phương pháp hồi quy Ridge (Ridge Regression Method).

Phương pháp bình phương nhỏ nhất là một dạng phân tích hồi quy để xác định hàm hồi quy phù hợp nhất với một tập dữ liệu. Các hệ số của hàm hồi quy được xác định bằng phương pháp đại số tuyến tính. Phương pháp này có đặc tính là ước lượng phương sai không trệch và nhỏ nhất. Tuy vậy, đối với những hàm phi tuyến tính không thể tuyến tính hóa bằng phương pháp đại số, thì sai số sẽ lệch chuẩn và phương sai là hằng số.

Hồi quy phi tuyến tính là một dạng phân tích hồi quy; trong đó các tham số của hàm hồi quy được xác định theo nguyên lý SSE_{Min}. Để tìm các hệ số của hàm hồi quy, trước hết giả định những tham số của hàm hồi quy bằng những giá trị nào đó. Kế đến, thay những giá trị giả định vào hàm hồi quy và tính hệ số xác định (R^2) cùng với các thống kê sai lệch (Bias, SSE, SEE, MAE, MAPE...). Quá trình tính toán này được lặp đi lặp lại nhiều lần cho đến khi thỏa mãn tiêu chuẩn dừng (Stoping Category) hay tiêu chuẩn kiểm định tính phù hợp của mô hình. Mục đích của phân tích hồi quy là xác định hàm hồi quy với sai lệch nhỏ nhất. Theo đó, tiêu chuẩn dừng là SSE_{Min}. Số bước tính toán phụ thuộc vào những tham số giả định ban đầu. Nếu các tham số giả định ban đầu gần bằng tham số thực, thì số bước tính toán ít hơn so với các tham số giả định ban đầu khác xa với tham số thực. Hàm hồi quy có tham số thực là hàm hồi quy nhận SSE_{Min}. Bởi vì các bước tính toán rất lớn và mất nhiều thời gian, nên ngày nay người ta đã xây dựng những phần mềm thống kê để giải các hàm bằng phương pháp phân tích hồi quy phi tuyến tính. Các tham số của hàm hồi quy có thể được xác định theo phương pháp của Marquardt, Gauss – Newton và Steepest Descent; trong đó phương pháp của Marquardt là thông dụng nhất.

Hồi quy Ridge được thiết kế để giải những bài toán hồi quy đa biến mà các biến độc lập có quan hệ chặt chẽ với nhau (Cộng tuyến tính). Trong trường hợp này, những tham số của mô hình hồi quy đa biến được ước lượng theo phương pháp bình phương nhỏ nhất thông thường bị sai lệch lớn. Hồi quy Ridge không chỉ làm giảm những biến động của các hệ số của mô hình tuyến tính đa biến, mà còn phản ánh rõ chiều hướng của mối quan hệ giữa biến phụ thuộc (Y) với các biến độc lập (X_i). Điều đó cho phép xây dựng mô hình chính xác hơn. Các hệ số của X_i được xác định bằng hồi quy Ridge sẽ bị trệch, nhưng chúng lại có biến động nhỏ (VIF gần bằng 1). Tham số điều chỉnh (Ridge) kiểm soát độ lệch của các hệ số hồi quy. Tham số Ridge nhận giá trị từ 0 ÷ 0,1. Khi giá trị Ridge càng gần 0, thì các hệ số của mô hình càng không trệch. Khi giá trị Ridge càng gần 0,1, thì các hệ số của mô hình càng trệch. Vì thế, để làm giảm độ trệch của các hệ số hồi quy, chúng ta cần phải điều chỉnh tham số Ridge thích hợp. Để tìm tham số Ridge thích hợp, người ta dựa vào 2 dấu hiệu. Một là biến động của các hệ số hồi quy chuẩn hóa X_i (Standardized Regression Coefficients). Hai là yếu tố khuyếch đại phương sai (Variance Inflation Factors = VIF). Yếu tố khuyếch đại phương sai so sánh biến động của các hệ số hồi quy chuẩn hóa và hệ số hồi quy không chuẩn hóa. Khi VIF >> 1, thì biến động của các hệ số hồi quy chuẩn hóa lớn hơn so với hệ số hồi quy không chuẩn hóa. Khi VIF << 1, thì biến động của các hệ số hồi quy chuẩn hóa nhỏ hơn so với hệ số hồi quy không chuẩn hóa. Khi VIF = 1, thì biến động của các hệ số hồi quy chuẩn hóa bằng biến động của hệ số hồi quy không chuẩn hóa. Vì thế, nếu căn cứ vào VIF, thì tham số Ridge thích hợp tương ứng với VIF = 1.

6.5.2.2. Kỹ thuật phân tích hồi quy

Kỹ thuật phân tích hồi quy không chỉ ảnh hưởng đến những đặc tính của hàm hồi quy ước lượng, mà còn đến kết quả phân tích quá trình sinh trưởng của cây gỗ

và rừng. Để chứng minh cho nhận định này, dưới đây dẫn ví dụ xây dựng hàm sinh trưởng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi được ước lượng bằng hàm Korf (Hàm 6.41); trong đó ba tham số m, b và c được xác định bằng phương pháp bình phương nhỏ nhất và phương pháp hồi quy phi tuyến tính.

M = m*exp(-b*A^-c) (6.41)

Theo phương pháp bình phương nhỏ nhất, các hệ số của hàm Korf được xác định theo 2 phương pháp: (a) Cố định tham số m, thay đổi tham số b và c; (a) Cố định tham số c, thay đổi tham số m và b.

Khi cố định tham số m, thì hàm Korf được tuyến tính hóa như hàm 6.42.

Ln(Ln(m/M)) = Ln(b) - c*Ln(A) (6.42)

Đặt Ln(Ln(m/M)) = Y; Ln(b) = b_0; c = b_1 và Ln(A) = X.

Từ đó Y = b_0 - b_1*X.

Khi cố định tham số c, thì hàm Korf được tuyến tính hóa như hàm 6.43.

Ln(M) = Ln(m) - b*A^-c (6.43)

Đặt Ln(M) = Y; Ln(m) = c_0; b = c_1; 1/A^c = X.

Từ đó Y = c_0 - c_1*X

Bảng 6.8$_{(a)}$. Phân tích hồi quy tương quan M = f(A) đối với rừng Thông (*Pinus sylvestris*) 140 tuổi bằng hàm Korf. Trường hợp cố định tham số m.

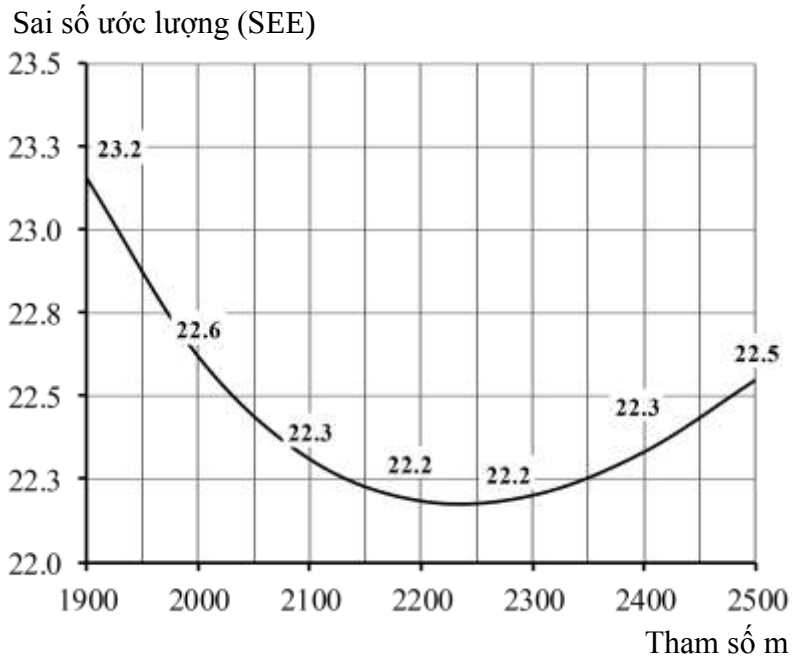
TT	Ba tham số:			Hàm
	m	b	c	
(1)	(2)	(3)	(4)	
1	2000	23,326	-0,6655	(6.44)
2	2100	22,4345	-0,6475	(6.45)
3	2200	21,6783	-0,6313	(6.46)
4	2300	21,0294	-0,6167	(6.47)
5	2400	20,4665	-0,6034	(6.48)
Phi tuyến	1505,17	49,7197	0,900264	(6.37)

Bảng 6.8$_{(b)}$. Thống kê sai lệch của hàm Korf. Trường hợp cố định tham số m.

m	r^2	SSE	SEE	MAE	MAPE	Bias	Hàm
(1)	(2)	(3)	(4)	(5)	(6)	(7)	
2000	99,28	5627,6	22,6	16,7	7,6	-0,34	(6.44)
2100	99,34	5475,0	22,3	16,7	7,4	-0,09	(6.45)
2200	99,38	5413,3	22,2	16,8	7,1	0,11	(6.46)
2300	99,42	5421,6	22,2	17,1	6,9	0,27	(6.47)
2400	99,44	5485,5	22,3	17,3	6,8	0,41	(6.48)
Phi tuyến	99,71	2923,2	16,3	13	10,7	2,02	(6.37)

Theo phương pháp bình phương nhỏ nhất, nếu giả định tham số m được chọn cố định trong khoảng từ 2000 – 2400, thì các hàm ước lượng M = f(A) của rừng

Thông (*Pinus sylvestris*) 140 tuổi có dạng như hàm 6.44 – 6.48 (Bảng 6.8 và Hình 6.5).



Hình 6.5. Đồ thị biểu diễn sự biến đổi của sai số ước lượng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi theo hàm Korf. Trường hợp cố định tham số m.

Phân tích số liệu ở Bảng $6.8_{(b)}$ và Hình 6.5 cho thấy, nếu giả định chọn tham số m tăng dần từ 2000 đến 2400, thì hệ số r^2 cũng tăng dần từ 99,28% đến 99,44%. Trái lại, giá trị SSE và SEE giảm dần và nhận giá trị nhỏ nhất tại m = 2200. Theo nguyên lý SSE_{Min}, tham số m thích hợp là 2200 và hàm ước lượng M = f(A) có dạng hàm 6.46 (Bảng 6.8_a). Hàm này nhận giá trị SSE_{Min} = 5413,3 và SEE = 22,2 m^3/ha.

Nếu tham số c được chọn cố định trong khoảng từ 0,20 – 0,80, thì hàm ước lượng M = f(A) của rừng Thông (*Pinus sylvestris*) 140 tuổi có dạng như hàm 6.49 – 6.55 (Bảng 6.9 và Hình 6.6).

Phân tích số liệu ở Bảng $6.9_{(b)}$ cho thấy, nếu giả định chọn tham số c tăng dần từ 0,2 đến 0,8, thì hệ số r^2 tăng dần từ 98,8% đối với tham số c = 0,2 và đạt lớn nhất 99,7% đối với tham số c = 0,4. Trái lại, giá trị SSE và SEE giảm dần từ c = 0,2 (tương ứng SSE = 69305; SEE = 79,4 m^3/ha) và nhận giá trị nhỏ nhất tại c = 0,5 (tương ứng SSE = 7995; SEE = 27,0 m^3/ha). Theo nguyên lý SSE_{Min}, tham số c thích hợp là 0,5 và hàm ước lượng M = f(A) có dạng như hàm 6.52 (Bảng 6.9_a).

Bảng 6.9$_{(a)}$. Phân tích hồi quy tương quan M = f(A) đối với rừng Thông (*Pinus sylvestris*) 140 tuổi bằng hàm Korf. Trường hợp cố định tham số c.

TT	Ba tham số của hàm Korf:			Hàm
	m	b	c	
(1)	(2)	(3)	(4)	
1	223708,7	-14,5149	0,2	(6.49)
2	22413,1	-13,8783	0,3	(6.50)
3	7094,7	-14,8109	0,4	(6.51)
4	3563,33	-16,7287	0,5	(6.52)
5	2256,9	-19,5331	0,6	(6.53)
6	1632,45	-23,2849	0,7	(6.54)
7	1284,0	-28,1332	0,8	(6.55)
Phi tuyến	1505,17	49,7197	0,900264	(6.37)

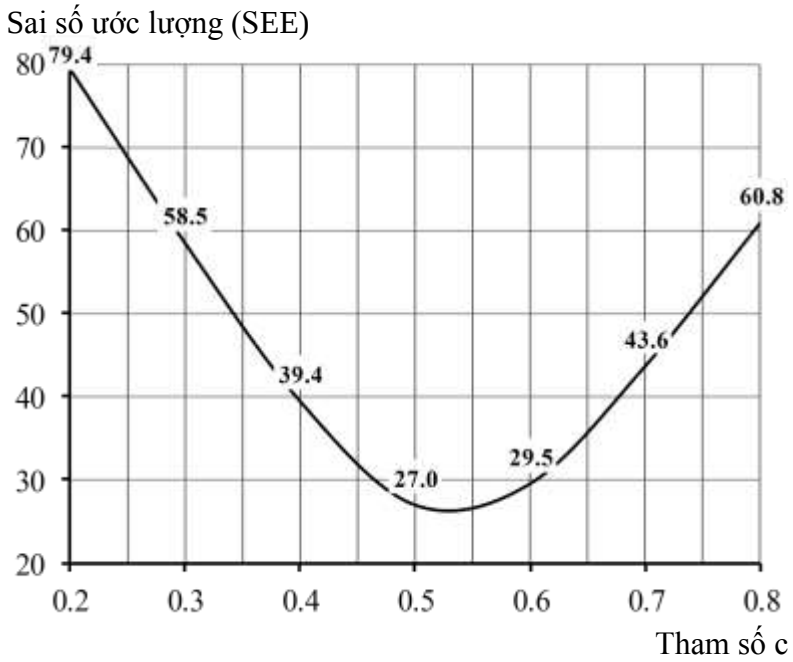
Bảng 6.9$_{(b)}$. Thống kê sai lệch của hàm Korf. Trường hợp cố định tham số c.

c	r^2	SSE	SEE	MAE	MAPE	Bias	Hàm
(1)	(2)	(3)	(4)	(5)	(6)	(7)	
0,2	98,8	69305,0	79,4	51,1	10,0	-11,3	(6.49)
0,3	99,4	37589,2	58,5	37,8	6,9	-6,7	(6.50)
0,4	99,7	17112,4	39,4	26,2	4,9	-1,6	(6.51)
0,5	99,6	7995,0	27,0	19,6	5,4	3,5	(6.52)
0,6	99,2	9580,7	29,5	23,0	7,6	8,7	(6.53)
0,7	98,5	20939,7	43,6	36,1	11,0	14,0	(6.54)
0,8	97,5	40721,5	60,8	49,0	14,4	19,1	(6.55)
Phi tuyến	99,71	2923,2	16,3	13	10,7	2,02	(6.37)

Đối với phương pháp hồi quy phi tuyến tính của Marquardt, ba tham số m, b và c của hàm Korf được điều chỉnh để nhận sai lệch nhỏ nhất (SSE_{Min} hay SEE_{Min}). Hàm ước lượng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi theo hàm Korf có dạng như hàm 6.37 (Bảng 6.6; 6.8 và 6.9). Hàm này nhận giá trị SSE_{Min} = 2923,2 và SEE = 16,3 m^3/ha.

So sánh hàm 6.37 với hàm 6.46 và hàm 6.52 (Bảng 6.8 và Bảng 6.9) cho thấy kỹ thuật phân tích hồi quy ảnh hưởng đến hệ số xác định và các sai lệch của hàm hồi quy ước lượng. Các tham số của hàm Korf được ước lượng bằng phương pháp bình phương nhỏ nhất nhận sai lệch lớn hơn so với phương pháp hồi quy phi tuyến tính (Bảng 6.10). Nếu sử dụng kỹ thuật bình phương nhỏ nhất và cố định tham số m, thì hàm hồi quy ước lượng nhận giá trị SSE và SEE tương ứng lớn hơn 1,85 lần và 1,36 lần so với phương pháp hồi quy phi tuyến tính. Nếu sử dụng kỹ thuật bình phương nhỏ nhất và cố định tham số c, thì hàm hồi quy ước lượng nhận giá trị SSE và SEE tương ứng lớn hơn 2,74 lần và 1,66 lần so với phương pháp hồi quy phi tuyến tính. Mặt khác, nếu các tham số của hàm Korf được ước lượng bằng

phương pháp bình phương nhỏ nhất và cố định tham số c, thì hàm hồi quy ước lượng nhận giá trị SSE và SEE lớn hơn tương ứng 1,48 lần và 1,22 lần so với phương pháp cố định tham số m.



Hình 6.6. Đồ thị biểu diễn sự biến đổi của sai số ước lượng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi theo hàm Korf. Trường hợp cố định tham số c.

Bảng 6.10. So sánh sai lệch của ba phương pháp ước lượng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi bằng hàm Korf.

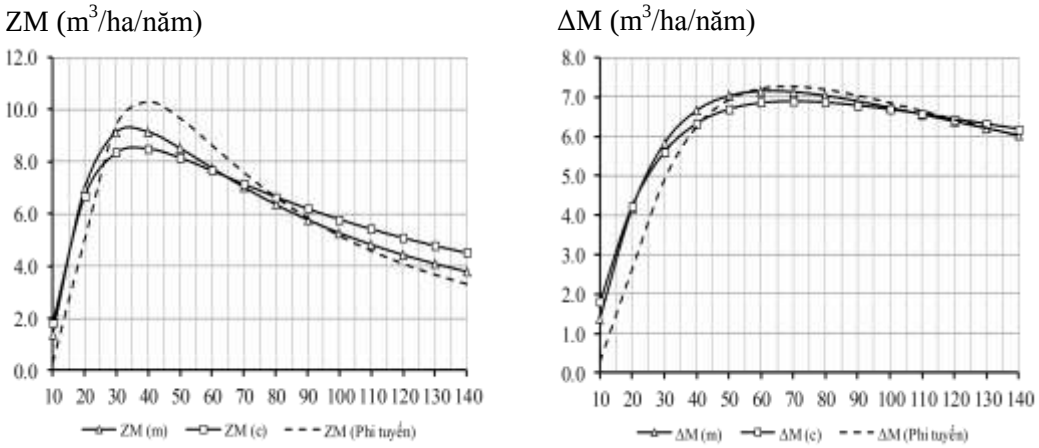
Phương pháp	SSE	SSE	$(SSE_{1,2})/(SSE_3)$	$(SEE_{1,2})/(SEE_3)$
(1)	(2)	(3)	(4)	(5)
(1) Cố định tham số m	5413	22,2	1,85	1,36
(2) Cố định tham số c	7995	27,0	2,74	1,66
(3) Hàm phi tuyến tính	2923	16,3	1,00	1,00

Kỹ thuật phân tích hồi quy cũng ảnh hưởng đến kết quả phân tích quá trình sinh trưởng của rừng Thông (*Pinus sylvestris*) 140 tuổi (Bảng 6.11 và Hình 6.7).

Bảng 6.11. Những đặc trưng sinh trưởng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi được xác định bằng hàm Korf.

Phương pháp	Tăng trưởng hàng năm:			Tăng trưởng bình quân năm:		
	ZM_{Max}	A (năm)	M (m^3/ha)	ΔM_{Max}	A (năm)	M (m^3/ha)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cố định m	9,1	40	266,3	7,1	60	429,0
Cố định c	9,5	40	253,0	6,9	70	482,5
Phi tuyến	10,3	40	249,9	7,3	70	508,6

Ba phương pháp ước lượng các tham số của hàm Korf đều cho thời điểm mà rừng Thông (*Pinus sylvestris*) đạt ZM_{Max} tương tự như nhau (40 tuổi). Thế nhưng, giá trị ZM_{Max} đạt cao nhất ở phương pháp hàm phi tuyến (10,3 m^3/ha/năm hay 100%), cao hơn tương ứng 13% và 8% so với phương pháp cố định tham số m và c. Hiện tượng này cũng xảy ra đối với đại lượng ΔM_{Max}. Thời điểm đạt ΔM_{Max} ở phương pháp cố định tham số m (70 năm) nhỏ hơn một cấp A so với phương pháp cố định tham số c (80 năm) và phương pháp hàm phi tuyến (80 năm). Trữ lượng gỗ mà tại đó xuất hiện ZM_{Max} và ΔM_{Max} ở ba phương pháp này cũng khác nhau khá lớn.



Hình 6.7. Đồ thị so sánh lượng tăng trưởng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi được ước lượng bằng hàm Korf; trong đó ba tham số được ước lượng theo ba phương pháp khác nhau.

Kỹ thuật phân tích hồi quy không chỉ ảnh hưởng đến chiều hướng của mối quan hệ giữa biến phụ thuộc (Y) với các biến độc lập (X_i), mà còn đến độ chính xác của hàm hồi quy ước lượng. Để chứng minh cho nhận định này, dưới đây dẫn ví dụ xây dựng hàm sinh trưởng thể tích thân (V, m^3) đối với cây bình quân của rừng trồng Keo lai theo số liệu ở Bảng 6.12. Trong ví dụ này, hàm thể tích thân cây bình quân của rừng trồng Keo lai được mô tả bằng hàm 6.56; trong đó SI là chỉ số lập địa.

$$V = b_0 + b_1*A + b_2*A^2 + b_3*(D^2*H) - b_4*(SI/H) \qquad (6.56)$$

Từ số liệu ở Bảng 6.12, xác định được hệ số tương quan giữa V với 4 biến A, A^2, (D^2H) và (SI/H) (Bảng 6.13). Số liệu ở Bảng 6.13 cho thấy giữa V và 4 biến dự đoán này tồn tại quan hệ chặt chẽ với nhau; trong đó V chỉ có mối quan hệ âm chặt chẽ (r = -0,630) với biến (SI/H).

Kết quả xây dựng hàm ước lượng thể tích thân cây bình quân của rừng Keo lai trên ba cấp lập địa bằng phương pháp bình phương nhỏ nhất (BFNN) và hồi quy Ridge được ghi lại tương ứng ở Bảng 6.14 và 6.15. Theo phương pháp BFNN (Bảng 6.14), V tồn tại quan hệ âm với A^2 (b_2 = -0,00003383), nhưng quan hệ dương

với (SI/H) (b_4 = 0,00049663). Kết quả này trái ngược với phân tích tương quan tuyến tính đơn giữa V với A^2 và (SI/H) (Bảng 6.13). Theo phương pháp hồi quy Ridge (Bảng 6.15), V tồn tại quan hệ dương với biến A^2 (b_2 = 0,00018885), nhưng quan hệ âm với biến (SI/H) (b_4 = -0,00179547). Kết quả này phù hợp với phân tích tương quan tuyến tính đơn giữa V với A^2 và (SI/H) (Bảng 6.13).

Bảng 6.12. Thể tích thân cây bình quân của rừng trồng Keo lai trên ba cấp lập địa khác nhau tại tỉnh Đồng Nai. Nguồn: Trần Thị Ngoan, 2019.

A(năm)	SI	V(m^3)	H (m)	D(cm)	A(năm)	SI	V(m^3)	H (m)	D (cm)
2	24	0,0104	7,2	6	2	16	0,0023	4,3	3,6
3	24	0,0362	10,9	9,1	3	16	0,0119	7,5	6,3
4	24	0,0757	14,1	11,6	4	16	0,0251	9,6	8,1
5	24	0,1284	16,5	13,9	5	16	0,0371	10,9	9,2
6	24	0,1591	17,6	14,9	6	16	0,0573	12,7	10,6
7	24	0,2085	19,5	16,3	7	16	0,058	12,6	10,7
8	24	0,2383	20,5	17	8	16	0,0734	13,7	11,6
9	24	0,2984	22	18,3	9	16	0,1098	15,7	13,1
10	24	0,326	22,7	18,9	10	16	0,1183	16,1	13,5
2	20	0,0047	5,6	4,5					
3	20	0,0225	9,3	7,8					
4	20	0,0462	11,9	9,8					
5	20	0,0727	13,7	11,5					
6	20	0,102	15,5	12,8					
7	20	0,1168	16,2	13,3					
8	20	0,1483	17,4	14,6					
9	20	0,1765	18,6	15,3					
10	20	0,1941	19,1	15,9					

Phân tích sai lệch của hàm thể tích (Bảng 6.16) cho thấy phương pháp BFNN nhận các sai lệch (Bias, Bias%, MAE, MAPE, SSE và SEE) nhỏ hơn so với phương pháp hồi quy Ridge. So với phương pháp BFNN, hai giá trị SSE và SEE ở hồi quy Ridge lớn hơn tương ứng 236,8 lần và 15,4 lần. Vì thế, hàm thể tích thân cây bình quân của rừng trồng Keo lai được xây dựng theo phương pháp BFNN cho kết quả ước lượng V với sai lệch nhỏ hơn. Bởi vì các hệ số của hàm hồi quy có biến động rất lớn (VIF >>), nên hàm V được ước lượng theo hồi quy BFNN cho kết quả dự đoán V với sai lệch lớn hơn so với hồi quy Ridge.

Bảng 6.13. Ma trận tương quan giữa thể tích thân cây bình quân của rừng trồng Keo lai với bốn biến A, A^2, (D^2H) và (SI/H).

		V	A	A^2	D^2H
	R	0,790**			
A	P	0,000			
	N	27	27		
	R	0,781**	0,983**		
A^2	P	0,000	0,000		
	N	27	27	27	
	R	0,988**	0,835**	0,807**	
D^2H	P	0,000	0,000	0,000	
	N	27	27	27	27
	R	-0,630**	-0,830**	-0,729**	-0,724**
SI/H	P	0,000	0,000	0,000	0,000
	N	27	27	27	27

Bảng 6.14. Kết quả phân tích hàm ước lượng thể tích thân cây bình quân của rừng Keo lai theo bốn biến A, A^2, (D^2H) và (SI/H) bằng hồi quy bình phương nhỏ nhất.

Parameter	*Estimate*	*Variance Inflation Factor*
CONSTANT	-0.00265890	
A	0.00053415	149.286
A^2	-0.00003383	97.1655
(D^2*H)	0.00004034	2.67289
(S/H)	0.00049663	10.7876

R-Squared = 99.9937 percent
R-Squared (adjusted for d.f.) = 99.9925 percent
Standard Error of Est. = 0.000768127
Mean absolute error = 0.000511543
Residual Analysis

	Estimation	*Validation*
n	27	
MSE	5.90019E-7	
MAE	0.0005115	
MAPE	0.653558	
ME	4.385E-17	
MPE	-0.0276901	

6.5.3. Thảo luận về phương pháp xây dựng và chọn hàm sinh trưởng

Sinh trưởng của cây gỗ và rừng có thể được mô tả bằng các hàm sinh trưởng. Một câu hỏi đặt ra là sử dụng hàm sinh trưởng khác nhau ảnh hưởng đến kết quả phân tích quá trình sinh trưởng của cây gỗ và rừng như thế nào? Về lý thuyết, các hàm sinh trưởng được chọn phải thỏa mãn 3 điều kiện: (a) f(A) = 0 khi A = 0; 1 tiệm cận tại f(A) = Y_{Max} và 2 điểm uốn. Điểm uốn thứ nhất tương ứng với giai đoạn sinh trưởng chậm chuyển sang giai đoạn sinh trưởng nhanh hay giai đoạn rừng non chuyển sang giai đoạn rừng trung niên. Điểm uốn thứ hai tương ứng với giai đoạn

sinh trưởng nhanh chuyển sang giai đoạn sinh trưởng chậm hay giai đoạn rừng trung niên chuyển sang giai đoạn rừng thành thục. Bởi vì nhiều hàm sinh trưởng cùng thỏa mãn ba điều kiện kể trên, nhưng đặc tính của các hàm là không giống nhau. Vì thế, nếu chọn những hàm dự tuyển khác nhau để mô tả quá trình sinh trưởng của cây gỗ và rừng, thì kết quả nhận được là khác nhau.

Bảng 6.15. Kết quả phân tích hàm ước lượng thể tích thân cây bình quân của rừng Keo lai theo bốn biến A, A^2, (D^2H) và (SI/H) bằng hồi quy Ridge.

Model Results for Ridge Parameter = 0.1

Parameter	*Estimate*	*Variance Inflation Factor*
CONSTANT	0.00167340	
A	0.00234954	0.55535
A^2	0.00018885	0.93988
(D^2*H)	0.00003234	1.49291
(S/H)	-0.00179547	1.39849

R-Squared = 91.9773 percent
R-Squared (adjusted for d.f.) = 90.5187 percent
Standard Error of Est. = 0.0118198
Mean absolute error = 0.00777656
Residual Analysis

	Estimation	*Validation*
n	27	
MSE	0.000139707	
MAE	0.00777656	
MAPE	9.77946	
ME	3.46945E-17	
MPE	-4.76243	

Bảng 6.16. Sai lệch của hàm ước lượng thể tích thân cây bình quân của rừng Keo lai theo phương pháp bình phương nhỏ nhất và hồi quy Ridge.

Hồi quy	Bias	Bias%	MAE	MAPE	SSE	SEE
BFNN(*) (1)	0,00	-0,03	0,00051	0,65	0,000013	0,00077
Ridge (2)	0,00	-4,76	0,00778	9,78	0,003074	0,0118
Tỷ lệ (2/1)	-0,4	171,3	15,2	15,0	236,8	15,4

(*) BFNN = Hồi quy bình phương nhỏ nhất.

Thông thường, hàm được sử dụng để phân tích quá trình sinh trưởng của cây gỗ và rừng được kiểm định từ một số hàm dự tuyển khác nhau và hàm hồi quy thích hợp được chọn theo tiêu chuẩn SSE_{Min} hoặc SEE_{Min}. Những hàm dự tuyển được chọn có liên quan đến kiến thức và nguồn tài liệu tham khảo của nhà nghiên cứu. Để hạn chế những tranh luận về kết quả nghiên cứu, tổng quan tài liệu nghiên cứu đóng vai trò quan trọng đối với việc chọn hàm sinh trưởng. Những hàm dự tuyển thường được chọn theo phương pháp biểu đồ hoặc từ những hàm đã được nhiều nhà nghiên cứu sử dụng. Một cách khác để thống nhất quan điểm là dựa vào mục đích của phân tích hồi quy. Về cơ bản, mục đích của phân tích hồi quy trong lâm học là xác định hàm ước lượng và dự đoán quá trình sinh trưởng của cây gỗ và rừng với

sai lệch nhỏ nhất. Theo quan điểm này, từ những hàm dự tuyển thỏa mãn 3 điều kiện của hàm sinh trưởng, nhà lâm học chọn ra một hàm thích hợp; trong đó tiêu chuẩn chấp nhận hàm thích hợp là SSE_{Min} hoặc SEE_{Min}.

Trong phân tích hồi quy, các tham số của hàm hồi quy có thể được xác định bằng phương pháp BFNN, hồi quy phi tuyến tính và hồi quy Ridge. Một câu hỏi đặt ra là kỹ thuật phân tích hồi quy ảnh hưởng đến kết quả phân tích quá trình sinh trưởng của cây gỗ và rừng như thế nào? Về cơ bản, nếu sử dụng kỹ thuật phân tích hồi quy khác nhau đối với cùng một hàm sinh trưởng, thì hàm hồi quy có các hệ số khác nhau. Điều đó dẫn đến kết quả phân tích quá trình sinh trưởng của cây gỗ và rừng cũng khác nhau. Để nhận được hàm ước lượng và dự đoán tốt, trước hết nhà lâm học cần sử dụng một số kỹ thuật phân tích hồi quy; sau đó chọn hàm hồi quy đáp ứng mục tiêu SSE_{Min}. Hàm hồi quy được xây dựng theo hai phương pháp BFNN và hồi quy phi tuyến tính cho sai số ước lượng nhỏ hơn so với hồi quy Ridge. Vì thế, hai phương pháp này được sử dụng để xây dựng hàm ước lượng những phản hồi của cây gỗ và rừng với các biến giải thích. Phương pháp hồi quy Ridge có ưu điểm là nó phản ánh rõ chiều hướng quan hệ giữa biến phản hồi và biến giải thích. Vì thế, khi xây dựng hàm để phân tích ảnh hưởng của các biến môi trường đến sinh trưởng của cây gỗ và rừng, thì hồi quy Ridge là phương pháp hữu dụng.

6.6. Phương pháp phân tích hồi quy phi tuyến tính và hồi quy Ridge

6.6.1. Phương pháp phân tích hồi quy phi tuyến tính

Hiện nay nhiều phần mềm thống kê cung cấp những phương pháp khác nhau trong phân tích hồi quy phi tuyến tính. Phần này chỉ giới thiệu phương pháp phân tích hồi quy phi tuyến tính bằng phần mềm thống kê STATGRAPHICS. Giả sử chúng ta cần xây dựng hàm sinh trưởng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi bằng hàm Korf (Hàm 6.57) theo số liệu đầu vào ở Bảng 6.5.

M = m*exp(-b*A^-c) (6.57)

Để xác định ba tham số m, b và c ban đầu bằng phương pháp phân tích hồi quy phi tuyến tính, các bước xử lý như dưới đây.

Bước 1. Tuyến tính hóa hàm Korf (Hàm 6.58 và 6.59).

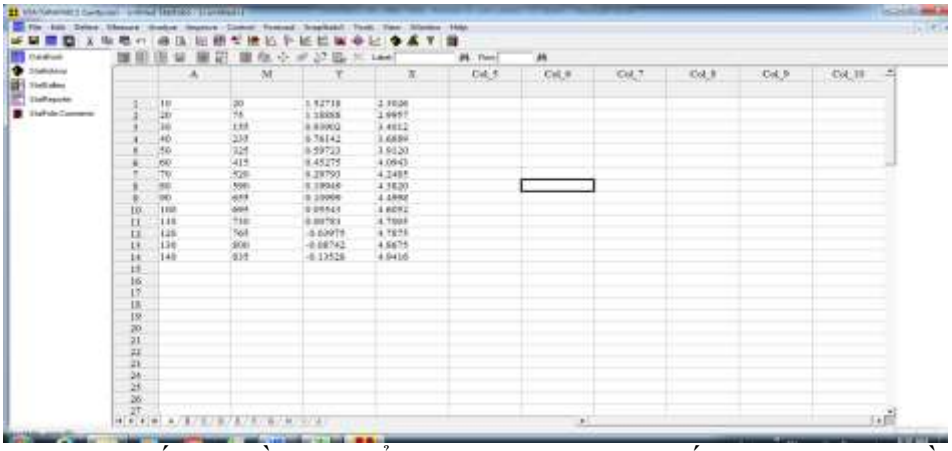
Ln(Ln(m/M)) = Ln(b) - c*Ln(A) (6.58)

Đặt Ln(Ln(m/M)) = Y; Ln(b) = b_0; c = b_1 và Ln(A) = X.

Từ đó Y = b_0 - b_1*X (6.59)

Bước 2. Giả định ba tham số m, b và c ban đầu. Trước hết giả định tham số m = 2000 (m^3/ha) và tính Ln(Ln(2000/M)) và Ln(A) (Bảng 6.17). Hai tham số b và c ban đầu được xác định bằng Hàm 6.59. Để đạt được mục đích này, trước hết Copy số liệu ở 4 cột của Bảng 6.17 và thả vào bảng tính của Statgraphics (Hình 6.8). Kế đến, để xác định hai tham số b_0 và b_1 của Hàm 6.59, vào đường dẫn: Improve > Regression Analysis > One Factor > Simple Regression. Đến đây xuất hiện một cửa sổ như Hình 6.9. Tại cửa sổ này, chuyển Y và X vào hai ô Y và X. Sau đó chọn OK để nhận hai tham số b_0 (Điểm chặn = Intercept) và b_1 (Độ dốc = Solpe) (Hàm 6.60;

Bảng 6.18). Tham số b của Hàm 6.57 được xác định bằng cách lấy đối Logarit cơ số Neper của tham số b_0, nghĩa là $\exp(b_0) = \exp(3,14957) = 23,32603$. Tham số m = 2000, còn tham số $c = b_1 = (-0,66552)$.



Hình 6.8. Số liệu đầu vào để ước lượng ba tham số m, b và c ban đầu của hàm Korf.

Bảng 6.17. Hàm trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi được xây dựng bằng hàm Korf ở dạng tuyến tính hóa với M = 2000 m^3/ha.

A (năm)	M (m^3/ha)	Ln(Ln(M/m))	Ln(A)
(1)	(2)	(3)	(4)
10	20	1,52718	2,3026
20	75	1,18888	2,9957
30	155	0,93902	3,4012
40	235	0,76142	3,6889
50	325	0,59723	3,9120
60	415	0,45275	4,0943
70	520	0,29793	4,2485
80	590	0,19949	4,3820
90	655	0,10999	4,4998
100	695	0,05543	4,6052
110	730	0,00783	4,7005
120	765	-0,03975	4,7875
130	800	-0,08742	4,8675
140	835	-0,13528	4,9416

Hình 6.9. Xây dựng hàm Korf ở dạng tuyến tính: $Y = b_0 - b_1*X$.

Bảng 6.18. Kết quả ước lượng tham số b_0 và b_1 của hàm Korf ở dạng tuyến tính hóa với M = 2000 m^3/ha.

Parameter	*Least Squares Estimate*	*Standard Error*	*T Statistic*	*P-Value*
Intercept	3.14957	0.0677435	46.4926	0.0000
Slope	-0.665525	0.0162485	-40.9592	0.0000

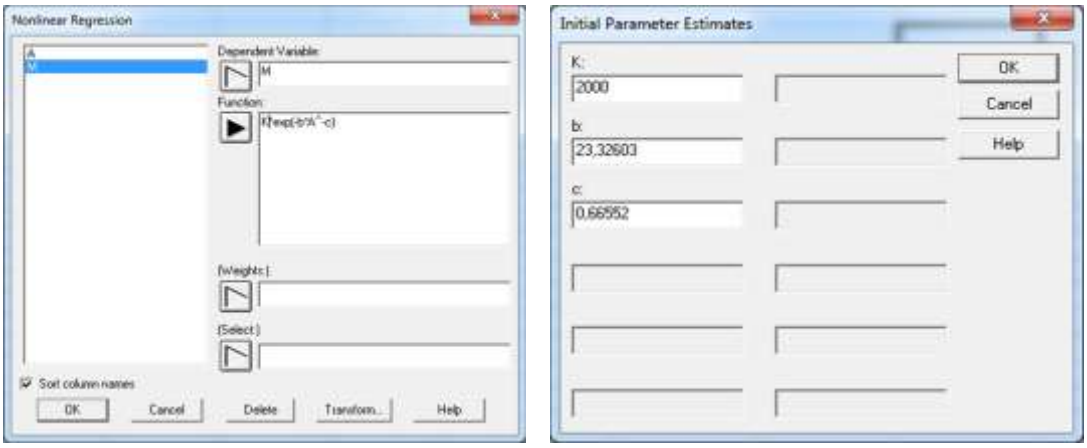
Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Model	3.45038	1	3.45038	1677.66	0.0000
Residual	0.02468	12	0.00205667		
Total (Corr.)	3.47506	13			

Correlation Coefficient = -0.996443
R-squared = 99.2898 percent
R-squared (adjusted for d.f.) = 99.2306 percent
Standard Error of Est. = 0.0453505
Mean absolute error = 0.0340777
Durbin-Watson statistic = 0.813269 (P=0.0019)
Lag 1 residual autocorrelation = 0.429124

$$Y = 3.14957 - 0.665525*X \qquad (6.60)$$

Bước 3. Xác định chính xác ba tham số m, b và c của Korf (Hàm 6.57) bằng hồi quy phi tuyến tính. Để đạt mục đích này, trước hết vào đường dẫn: Improve > Regression Analysis > Multiple Factor > Nonlinear Regression. Đến đây xuất hiện một cửa sổ như Hình 6.10. Tại cửa sổ này, chuyển M vào Dependent Variable. Sau đó ghi hàm Korf (6.57) vào ô Function. Lưu ý rằng các kí tự biểu diễn các tham số của hàm Korf ở ô Function phải khác với các kí tự ở đầu các cột của bảng tính ở Hình 6.8. Ở ví dụ này, bởi vì kí tự ở đầu cột 2 là M, nên hàm Korf được viết là K*exp(-b*A^-c). Kế đến chọn OK. Đến đây xuất hiện một cửa sổ; trong đó yêu cầu cho biết ba tham số m, b và c ban đầu (Initial Parameter Estimates) của hàm Korf. Ở đây thay tham số m ban đầu là K = 2000, còn tham số b và c tương ứng là 23,32603 và 0,66552 (Hình 6.10). Sau đó chọn OK để nhận kết quả của hàm hồi quy (Bảng

6.19; Hình 6.11). Các tham số của hàm ước lượng có thể được xác định theo phương pháp của Marquardt, Gauss – Newton và Steepest Descent. Trong ví dụ này, các tham số của hàm Korf được xác định theo phương pháp phân tích hồi quy phi tuyến tính của Marquardt. Kết quả báo cáo những thống kê sau đây: (a) Kết quả ước lượng; trong đó bao gồm các tham số của hàm, sai số và khoảng tin cậy của các tham số; (b) Bảng phân tích ANOVA; (c) Hệ số xác định (R-Squared); (d) Phân tích các sai lệch (Residual Analysis); (e) Hàm hồi quy.



Hình 6.10. Giả định ba tham số m, b và c ban đầu của hàm Korf.

Bảng 6.19. Kết quả xây dựng hàm ước lượng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi bằng hàm Korf.

Estimation Results

Parameter	*Estimate*	*Asymptotic Standard Error*	*Asymptotic Confidence Lower*	*95.0% Interval Upper*
k	1505.22	175.233	1119.53	1890.9
b	49.7126	17.4704	11.2604	88.1648
c	0.90022	0.108695	0.660987	1.13946

Analysis of Variance

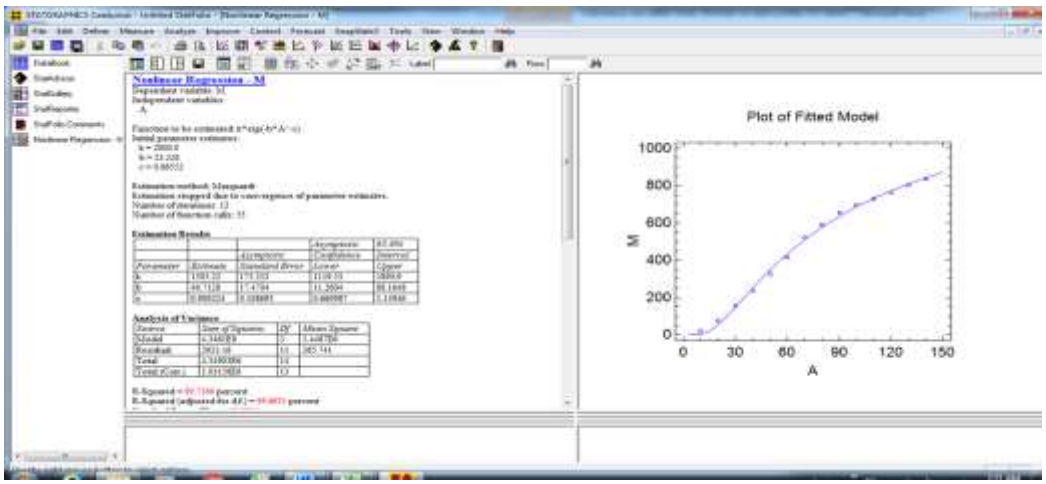
Source	*Sum of Squares*	*Df*	*Mean Square*
Model	4.3461E6	3	1.4487E6
Residual	2923.16	11	265.741
Total	4.34903E6	14	
Total (Corr.)	1.03158E6	13	

R-Squared = 99.7166 percent
R-Squared (adjusted for d.f.) = 99.6651 percent
Standard Error of Est. = 16.3016

Residual Analysis

	Estimation	*Validation*
n	14	
MSE	265.741	

MAE	13.0004
MAPE	10.7093
ME	2.01791
MPE	7.84303



Hình 6.11. Kết quả của hàm ước lượng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi bằng hàm Korf.

6.6.2. Phương pháp phân tích hồi quy Ridge

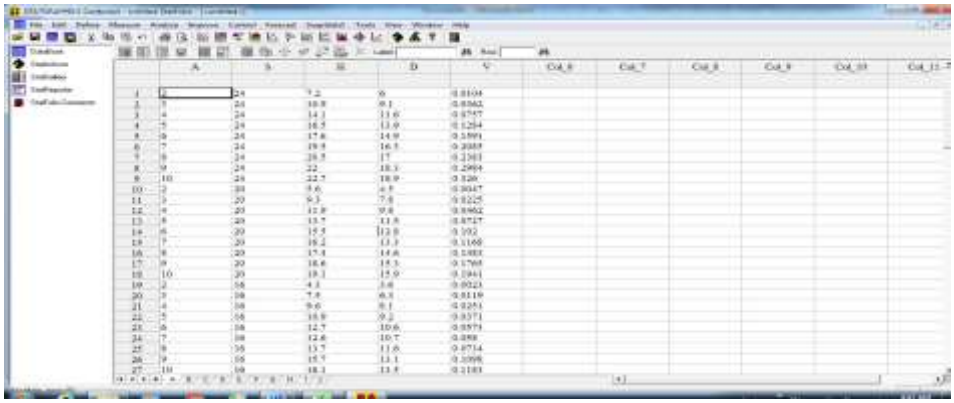
Hiện nay nhiều phần mềm thống kê cung cấp những phương pháp khác nhau trong phân tích hồi quy Ridge. Phần này chỉ giới thiệu phương pháp phân tích hồi quy Ridge bằng phần mềm thống kê STATGRAPHICS. Giả sử chúng ta cần xây dựng hàm thể tích thân cây bình quân của rừng trồng Keo lai 10 tuổi bằng hàm Hàm 6.61 theo số liệu đầu vào ở Bảng 6.12.

$$V = b_0 + b_1*A + b_2*A^2 + b_3*(D^2*H) - b_4*(SI/H) \qquad (6.61)$$

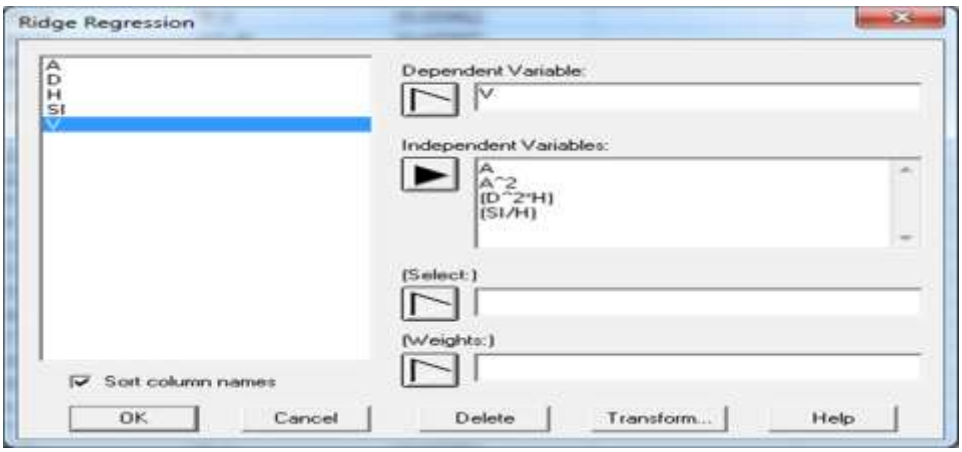
Để xác định 4 tham số (b_0, b_1, b_2, b_3 và b_4) của Hàm 6.61 bằng phương pháp phân tích hồi quy Ridge, các bước xử lý như dưới đây.

Bước 1. Copy A, SI, H, D và V ở Bảng 6.12 và dán vào bảng tính Statgraphics (Hình 6.12).

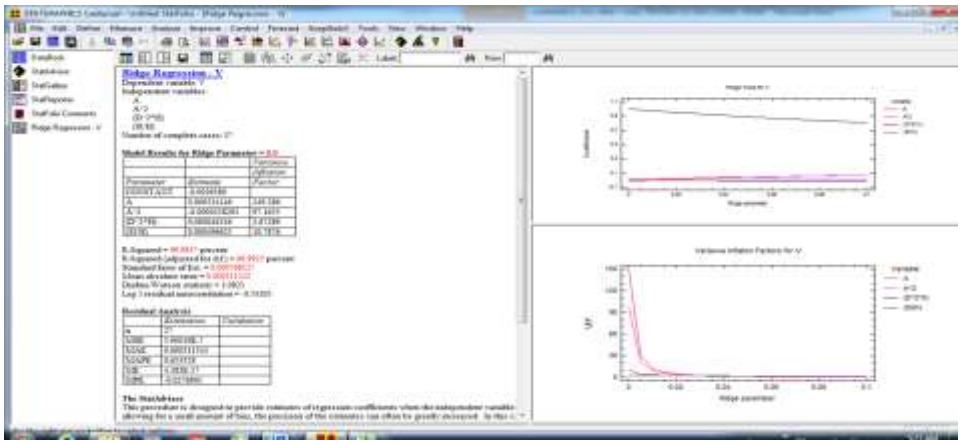
Bước 2. Vào đường dẫn: Improve > Regression Analysis > Multiple Factor > Ridge Regression. Đến đây xuất hiện một cửa sổ như Hình 6.13. Tại cửa sổ này, chuyển V vào Dependent Variable và chuyển A, A^2, (D^2*H) và (SI/H) vào Independent Variablres. Sau đó chọn OK để nhận kết quả của hàm V được ước lượng bằng phương pháp hồi quy Ridge (Hình 6.14 và Bảng 6.20).



Hình 6.12. Số liệu đầu vào để xây dựng hàm thể tích thân cây Keo lai bằng kỹ thuật hồi quy Ridge.



Hình 6.13. Xây dựng hàm thể tích bằng hồi quy Ridge.



Hình 6.14. Kết quả phân tích hàm thể tích bằng hồi quy Ridge.

Kết quả của mô hình ở bước 2 (Bảng 6.20) tương ứng với tham số Ridge hay Lambda = 0 (Model Results for Ridge Parameter = 0.0). Khi Lambda = 0, thì mô hình được giải theo phương pháp bình phương nhỏ nhất thông thường. Kết quả nhận được mô hình với các hệ số hồi quy không trệch. Khi giá trị Lambda càng gần Zero, thì các hệ số của mô hình càng không trệch. Tuy vậy, khi các biến X_i có hiện tượng cộng tuyến tính, thì các hệ số hồi quy có biến động lớn và kết quả dự đoán Y theo các X_i sẽ bị sai lệch lớn. Vì thế, để xây dựng mô hình dự đoán ít sai lệch hơn, chúng ta cần phải sử dụng Hồi quy Ridge.

Bảng 6.20. Kết quả xây dựng hàm ước lượng thể tích thân cây bình quân của rừng trồng Keo lai 10 tuổi bằng hồi quy Ridge với tham số điều chỉnh bằng 0.

Model Results for Ridge Parameter = 0.0

Parameter	*Estimate*	*Variance Inflation Factor*
CONSTANT	-0.0026589	
A	0.000534146	149.286
A^2	-0.0000338292	97.1655
(D^2*H)	0.000040336	2.67289
(SI/H)	0.000496625	10.7876

R-Squared = 99.9937 percent
R-Squared (adjusted for d.f.) = 99.9925 percent
Standard Error of Est. = 0.000768127
Mean absolute error = 0.000511543

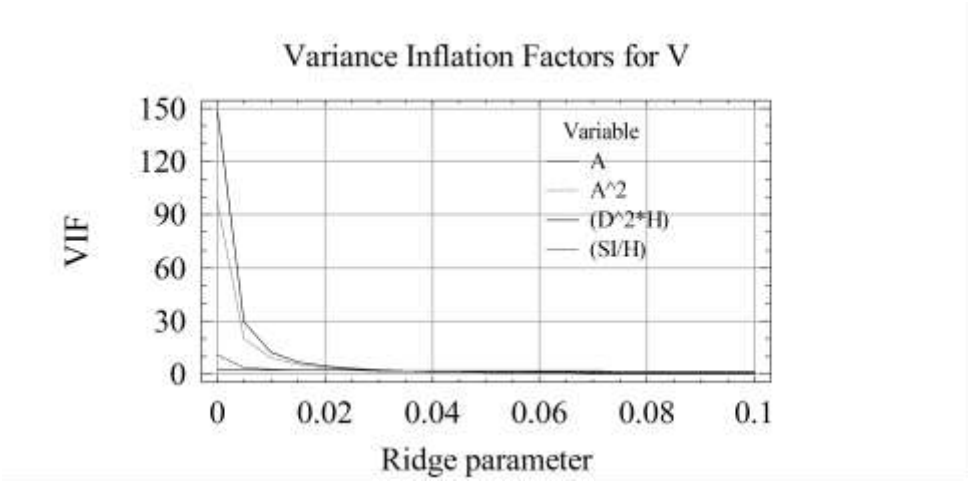
Residual Analysis

	Estimation	*Validation*
n	27	
MSE	5.90019E-7	
MAE	0.00051154	
MAPE	0.653558	
ME	4.385E-17	
MPE	-0.0276901	

Các hệ số của X_i được xác định bằng hồi quy Ridge sẽ bị trệch, nhưng chúng lại có biến động nhỏ. Tham số Lambda kiểm soát độ lệch của các hệ số hồi quy. Nó nhận giá trị từ 0 đến 0,1 (Mặc định Lambda = 0). Khi giá trị Lambda càng gần Zero, thì các hệ số của mô hình càng không trệch. Khi giá trị Lambda càng gần 0,1, thì các hệ số của mô hình càng trệch. Vì thế, để làm giảm độ trệch của các hệ số hồi quy, chúng ta cần phải xác định tham số Lamda thích hợp.

Bước 3. Xác định tham số Lamda thích hợp. Để tìm tham số Lambda thích hợp, người ta dựa vào 2 dấu hiệu. Một là biến động của các hệ số hồi quy chuẩn hóa X_i (Standardized Regression Coefficients). Hai là yếu tố khuyếch đại phương sai (Variance Inflation Factors = VIF). Khi tham số Lambda thay đổi từ 0 đến 0,1, thì các hệ số của X_i cũng thay đổi. Thông thường, các hệ số của X_i thay đổi rất lớn ở gần giá trị Lambda hay Ridge Parameter = 0; sau đó chúng thay đổi rất nhỏ (Hình 6.15).

Khi tham số Lambda bằng Zero, thì kết quả nhận được mô hình hồi quy không chuẩn hóa. Yếu tố khuyếch đại phương sai so sánh biến động của các hệ số hồi quy chuẩn hóa và hệ số hồi quy không chuẩn hóa (Mô hình với các biến độc lập không có quan hệ với nhau). Khi VIF >> 1, thì biến động của các hệ số hồi quy chuẩn hóa lớn hơn so với các hệ số hồi quy không chuẩn hóa. Khi VIF << 1, thì biến động của các hệ số hồi quy chuẩn hóa nhỏ hơn so với các hệ số hồi quy không chuẩn hóa. Khi VIF = 1, thì biến động của các hệ số hồi quy chuẩn hóa và các hệ số hồi quy không chuẩn hóa là bằng nhau. Vì thế, căn cứ để xác định tham số Lambda thích hợp là VIF = 1. Tóm lại, nếu căn cứ vào biến động của các hệ số hồi quy X_i (Ridge Trace), thì tham số Lamda thích hợp là giá trị Lambda nhỏ nhất tương ứng với các hệ số hồi quy ổn định (Hình 6.15).



Hình 6.15. Đồ thị biểu diễn yết tố khuyếch đại phượng sai.

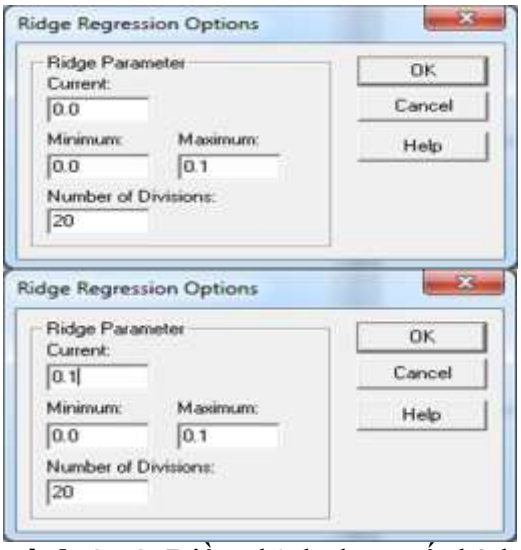
Nếu căn cứ vào VIF, thì tham số Lamda thích hợp tương ứng với VIF gần bằng 1. Theo quy tắc này, từ Hình 6.15 cho thấy, tham số điều chỉnh (Ridge Parameter) thích hợp là 0,1. Để nhận được hàm thể tích bằng hồi quy Ridge, tại cửa sổ ở Hình 6.14, nhấn con trỏ chuột phải và chọn Analysis Option. Tại đây xuất hiện một cửa sổ như Hình 6.16. Nếu chọn OK, chúng ta nhận được hàm V giống như phân tích hồi quy BFNN (Bảng 6.20). Để nhận được hàm thể tích bằng hồi quy Ridge, chúng ta thay tham số điều chỉnh (Ridge Parameter) là 0,1 (Hình 6.16). Để biết vai trò của từng biến giải thích, chọn các hệ số hồi quy chuẩn hóa (Standardized Regression Coefficients) trong Tables ở góc trên bên trái ở Hình 6.14 (Hình 6.17). Sau đó chọn OK để nhận được hàm V bằng hồi quy Ridge với tham số điều chỉnh thích hợp (Bảng 6.21).

6.6.3. Thảo luận về phân tích hồi quy phi tuyến tính và hồi quy Ridge

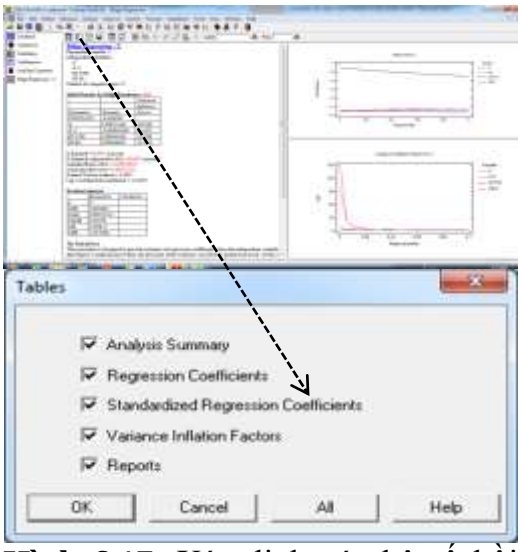
6.6.3.1. Phân tích hồi quy phi tuyến tính

Khi phân tích hồi quy bằng phương pháp hồi quy phi tuyến tính, thì số lượng các bước xác định hàm hồi quy phù hợp phụ thuộc vào những tham số được giả

định ban đầu. Ở ví dụ trên đây, khi cho tham số m = 2000, b = 23,32603 và c = 0,66552, thì hàm hồi quy phù hợp được xác định sau 55 bước. Trái lại, nếu cho tham số m = 900, b = 20 và c = 0.5, thì hàm hồi quy phù hợp được xác định sau 71 bước.



Hình 6.16. Điều chỉnh tham số thích hợp trong hồi quy Ridge.



Hình 6.17. Xác định các hệ số hồi quy chuẩn hóa trong hồi quy Ridge.

Bảng 6.21. Kết quả xây dựng hàm ước lượng thể tích thân cây bình quân của rừng trồng Keo lai 10 tuổi bằng hồi quy Ridge với tham số điều chỉnh bằng 0,1.

Model Results for Ridge Parameter = 0.1.

Parameter	*Estimate*	*Variance Inflation Factor*
CONSTANT	0.0016734	
A	0.00234954	0.555353
A^2	0.000188854	0.939883
(D^2*H)	0.00003234	1.49291
(SI/H)	-0.00179547	1.39849

R-Squared = 91.9773 percent
R-Squared (adjusted for d.f.) = 90.5187 percent
Standard Error of Est. = 0.0118198
Mean absolute error = 0.00777656

Residual Analysis

	Estimation	*Validation*
n	27	
MSE	0.000139707	
MAE	0.00777656	
MAPE	9.77946	
ME	3.46945E-17	
MPE	-4.76243	

Nếu giả định những tham số ban đầu không phù hợp, thì chương trình báo lỗi ở kết quả đầu ra. Để tránh tình trạng này, các tham số ban đầu của hàm hồi quy phi

tuyến tính được xác định gần đúng bằng hồi quy tuyến tính. Phương pháp này giúp hạn chế những khó khăn trong lúc chọn các tham số ban đầu. Khi thực hiện nhiều phân tích hồi quy có dạng tương tự như nhau, thì các tham số ban đầu có thể được xác định dễ dàng.

Hình thái của hàm hồi quy phù hợp phụ thuộc vào những tham số được giả định ban đầu. Khi giả định những tham số ban đầu khác nhau, thì các tham số của hàm hồi quy phù hợp cũng khác nhau. Tuy vậy, những thông tin về sai lệch của các hàm hồi quy (R^2, Bias, SSE, SEE, MAE, MAPE) thay đổi không lớn. Bởi vì phương pháp phân tích hồi quy phi tuyến tính dựa trên nguyên lý SSE_{Min}, nên các hàm hồi quy này đều cho kết quả ước lượng và dự đoán biến phụ thuộc tương tự nhau.

So với phương pháp phân tích hồi quy tuyến tính, phương pháp phân tích hồi quy phi tuyến tính cung cấp đầy đủ hơn về những sai lệch thực của các hàm hồi quy (R^2, Bias, SSE, SEE, MAE, MAPE). Điều đó không chỉ cho phép giảm bớt các bước tính toán, mà còn dễ dàng so sánh mức độ phù hợp của các hàm mà biến phụ thuộc và các biến độc lập ở dạng khác nhau (Y, LnY...; X, LnX...).

Về lý thuyết, khi phân tích hồi quy phi tuyến tính, thì một số tham số của hàm hồi quy có thể được cố định. Theo phương pháp này, hàm hồi quy thích hợp nhận được những tham số và những sai lệch khác với phương pháp thay đổi tất cả các tham số ban đầu. Để chứng minh nhận định này, Bảng 6.22 ghi lại kết quả xây dựng hàm ước lượng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi bằng hàm Korf theo hai kỹ thuật xử lý khác nhau. Trường hợp thứ nhất là giả định ba tham số ban đầu cùng thay đổi; trong đó m = 2000, b = 20 và c = 0,5. Trường hợp thứ hai là cố định m = 2000, thay đổi b và c. Hàm Korf ở trường hợp thứ 2 có dạng M = 2000*exp(-b*A^-c); trong đó giả định tham số ban đầu là b = 20 và c = 0,5. Từ số liệu ở Bảng 6.22 cho thấy, các tham số m, b và c của hai trường hợp này khác nhau rõ rệt. Giá trị SEE, MSE và MAE của trường hợp 2 tương ứng lớn hơn 10,4%, 22,2% và 13,9% so với trường hợp 1. Trái lại, giá trị MAPE, Bias và Bias(%) của trường hợp 2 tương ứng nhỏ hơn 28,8%, 85,3% và 50,3% so với trường hợp 1. Theo nguyên lý SSE_{Min}, phương pháp cùng thay đổi ba tham số ban đầu cho kết quả tốt hơn.

6.6.3.2. Phân tích hồi quy Ridge

Phương pháp hồi quy Ridge không chỉ cho phép tạo ra hàm hồi quy với các tham số có biến động nhỏ, mà nó còn giúp chúng ta xác định rõ khuynh hướng của mối quan hệ giữa biến phản hồi với các biến giải thích. Vì thế, khi cần dự đoán những biến phản hồi theo các biến giải thích nằm ngoài giới hạn của tập số liệu dùng để xây dựng hàm hồi quy, thì phương pháp hồi quy Ridge là một lựa chọn tốt. Phương pháp này cũng rất hữu ích trong trường hợp phân tích vai trò của các biến giải thích trong các hàm phản hồi. Nếu xây dựng hàm phản hồi bằng kỹ thuật BFNN và căn cứ vào dấu của các hệ số hồi quy, thì chúng ta sẽ giải thích sai vai trò của các biến giải thích. Chẳng hạn: Ở ví dụ trên, về lý thuyết, V thân cây bình quân

của rừng trồng Keo lai suy giảm theo sự suy giảm của chỉ số SI. Thế nhưng, bằng phương pháp BFNN, kết quả lại nhận được hệ số của biến (SI/H) mang giá trị dương. Điều này dẫn đến giải thích sai kết quả. Trong phân tích hồi quy Ridge, dấu của các biến giải thích phụ thuộc vào tham số điều chỉnh. Nếu tham số điều chỉnh càng gần Zero, thì kết quả phân tích hồi quy càng giống với phương pháp BFNN. Vì thế, khi xây dựng hàm phản hồi bằng phương pháp hồi quy Ridge, chúng ta cần phải dò tìm tham số điều chỉnh thích hợp. Ở ví dụ trên, V thân cây bình quân của rừng trồng Keo lai suy giảm theo sự suy giảm của chỉ số SI. Kết quả của hồi quy Ridge cho phép giải thích đúng mối quan hệ giữa V với chỉ số SI.

Bảng 6.22. Kết quả xây dựng hàm ước lượng trữ lượng gỗ của rừng Thông (*Pinus sylvestris*) 140 tuổi bằng hàm Korf.

Tham số và sai lệch	Thay đổi m, b và c (1)	Thay đổi b và c (2)	Tỷ lệ (%) (2)/(1)
(1)	(2)	(3)	
m	1505,05	2000	
b	-49,7205	-27,9103	
c	-0,9003	-0,7065	
SEE	16,3	18,0	10,4
MSE	265,7	324,8	22,2
MAE	13,0	14,8	13,9
MAPE	10,7	7,6	-28,8
Bias	2,0	0,3	-85,3
Bias(%)	7,8	3,9	-50,3

Chương 7
ĐÁNH GIÁ LẬP ĐỊA RỪNG BẰNG CHỈ SỐ

7.1. Tình hình chung

Một trong những nhiệm vụ của lâm học là tìm kiếm phương pháp xác định sự khác biệt giữa các điều kiện môi trường của các QXTV và kiểu rừng khác nhau. Một số môi trường giúp cho rừng sinh trưởng tốt. Trái lại, một số môi trường lại kìm hãm sinh trưởng của rừng. Sự khác biệt này có thể là do nhiều yếu tố khác nhau như khí hậu, địa hình, đất, phương thức lâm sinh. Những khác biệt này được phản ánh thông qua sự khác biệt về kết cấu loài cây gỗ, tăng trưởng và sản lượng rừng. Dự báo tăng trưởng và sản lượng rừng đòi hỏi đánh giá chính xác sự khác biệt giữa các điều kiện môi trường và giữa các phương thức lâm sinh. Ước lượng sai những điều kiện môi trường có thể dẫn đến sự sai lệch về ước lượng tăng trưởng và sản lượng rừng và những đặc tính khác của rừng.

Lập địa rừng (Forest Site) là tổng hợp những điều kiện sống (khí hậu, địa hình, đất, sinh vật, phương thức lâm sinh) của rừng ở một khu vực nhất định. Tùy theo tác động của điều kiện môi trường, cây gỗ và rừng sẽ có những phản ứng khác nhau. Bởi vì lập địa bao gồm rất nhiều nhân tố và khó đo đạc, nên lâm học sử dụng phản hồi của cây gỗ và rừng để đánh giá chất lượng lập địa. Theo quan điểm của lâm học, tình trạng sống của thảm thực vật rừng là dấu hiệu tổng hợp để đo phản ứng của cây gỗ và rừng với sự thay đổi của điều kiện môi trường. Tình trạng sống của thảm thực vật rừng phản ánh rõ nhất thông qua sản lượng gỗ hoặc tổng sinh khối. Tuy vậy, cả hai đặc tính này là những đặc tính khó đo đạc chính xác tại rừng. Vì thế, việc tìm kiếm những chỉ tiêu thay thế cho sản lượng gỗ là cần thiết. Chiều cao của quần thụ thuần loài đồng tuổi có quan hệ chặt chẽ với sản lượng gỗ. Chiều cao là chỉ tiêu dễ đo đạc. Do đó, từ hàm ước lượng sản lượng gỗ dựa theo H, chúng ta có thể ước lượng được sản lượng gỗ.

Chất lượng lập địa có thể được xác định bằng hai phương pháp. Một là phương pháp sử dụng điều kiện địa lý là trung tâm. Hai là phương pháp sử dụng thảm thực vật là trung tâm. Phương pháp thứ nhất giả định sản lượng gỗ phụ thuộc vào các điều kiện khí hậu, địa hình và đất. Phương pháp này sử dụng độ ẩm đất, trạng thái chất khoáng của đất và bức xạ hoạt tính quang hợp là số đo trực tiếp, còn số đo gián tiếp là dạng khí hậu, dạng đất và chỉ số thảm thực vật. Trái lại, phương pháp dựa vào thảm thực vật giả định rằng sản lượng gỗ là phản hồi của rừng đối với điều kiện lập địa. Phương pháp này sử dụng năng suất của lập địa là số đo trực tiếp và H quần thụ là số đo gián tiếp. Nói chung, những số đo trực tiếp ở cả hai phương pháp kể trên là những số đo rất khó xác định. Trái lại, những số đo gián tiếp có thể được xác định khá dễ dàng ở ngoài trời.

Phương pháp đánh giá chất lượng lập địa có thể được phân chia thành phương pháp định lượng và phương pháp định tính. Phương pháp định tính đánh giá năng suất lập địa dựa trên những số đo tình trạng môi trường diễn biến theo thời gian. Trái lại, phương pháp định lượng đánh giá năng suất lập địa dựa trên những số đo sản lượng rừng ở thời điểm hiện tại. Phương pháp định lượng đánh giá năng suất lập địa dựa trên những biến liên tục (H quần thụ), nhưng chúng là những số đo thực và có thể được ghép thành cấp lập địa. Tuy vậy, phương pháp này lại khó áp dụng đối với rừng hỗn loài khác tuổi.

7.2. Phương pháp đánh giá chất lượng lập địa rừng

7.2.1. Chất lượng của lập địa rừng

Chất lượng của lập địa rừng phản ánh khả năng sinh trưởng của rừng hoặc khả năng tạo ra sản lượng rừng. Vì thế, chất lượng lập địa còn được gọi là sản lượng của lập địa. Thông thường, lâm học đánh giá chất lượng lập địa thông qua sản lượng gỗ mà rừng tạo ra ở những tuổi nào đó. Chất lượng lập địa rừng có thể được xác định theo tiết diện ngang (G, m^2/ha), trữ lượng gỗ (M, m^3/ha), chiều cao trung bình (H_{Bq}, m), chiều cao trung bình của những cây ưu thế và đồng ưu thế (H_D m)...Trong lâm học, người ta còn dựa vào lượng tăng trưởng trung bình năm đối với trữ lượng gỗ (ΔM, m^3/ha/năm) để đánh giá năng suất của lập địa. Khi ΔM nhỏ hơn một giá trị giới hạn dưới nào đó, thì lập địa không tạo ra năng suất cao.

Các đại lượng G, M và H_{Bq} của rừng có thể thay đổi theo tuổi và những hoạt động lâm sinh (tỉa thưa, khai thác, bón phân), cháy, xói mòn đất…Vì thế, nếu tính theo những đại lượng khác nhau, thì chất lượng lập địa cũng khác nhau. Sản lượng của một loài cây gỗ thay đổi theo lập địa. Trên cùng một lập địa, sản lượng của loài cây gỗ này khác với sản lượng của loài cây gỗ khác. Vì thế, chất lượng của lập địa rừng được đánh giá theo loài cây gỗ.

7.2.2. Chỉ số lập địa rừng

7.2.2.1. Khái niệm về chỉ số lập địa rừng

Chất lượng lập địa rừng được biểu diễn bằng một chỉ số gọi là chỉ số lập địa rừng (Forest Site Index = SI). Chỉ số SI được các nhà lâm học sử dụng để xác định năng suất tương đối hay năng suất tiềm năng của một lập địa nhất định. Đó là khả năng của một vùng đất để loài cây gỗ sinh trưởng. Số đo này được sử dụng để mô tả điều kiện sinh trưởng của rừng trồng thuần loài đồng tuổi. Chỉ số SI phản ánh năng suất của loài cây, phương thức lâm sinh, điều kiện môi trường quá khứ và năng suất rừng tương lai.

Chỉ số SI phụ thuộc vào loài cây gỗ. Điều đó có nghĩa là chỉ số SI được xác định cho từng loài cây gỗ. Trên cùng một lập địa, chỉ số SI của loài cây gỗ này khác với chỉ số SI của loài cây gỗ khác. Chỉ số SI có một số giới hạn. (a) Chỉ số SI phụ thuộc vào những yếu tố ảnh hưởng đến tăng trưởng (mật độ, phương thức lâm sinh, chế độ quản lý, sâu bệnh, đặc tính di truyền, nguồn gốc rừng (hạt, chồi)); (b) Chỉ số SI không thể sử dụng cho rừng khác tuổi, rừng bị phá, những loài cây không thể

hiện rõ vòng năm; (c) Hàm chỉ số SI và đường cong chỉ số SI phụ thuộc vào phương pháp chọn hàm sinh trưởng; (d) Khi xây dựng đường cong chỉ số SI, việc ngoại suy lập địa và tuổi có thể mắc sai số lớn; (e) Chỉ số SI phụ thuộc vào phạm vi không gian xây dựng chỉ số SI và phương pháp xây dựng hàm sinh trưởng; (f) Sai số đo đạc sản lượng rừng; (g) Số lượng tuổi được đo đạc sản lượng rừng...

7.2.2.2. Những chỉ tiêu đo chỉ số lập địa rừng

Chỉ số SI có thể được xác định dựa theo mối quan hệ H = f(A), H = f(D), ΔD = f(D), ΔG = f(G), ΔM = f(M); trong đó phương pháp dựa vào mối quan hệ giữa H với tuổi rừng là thông dụng nhất. Trong lâm học, chỉ số SI được xác định theo chiều cao trung bình của những cây ưu thế và đồng ưu thế (H_D, m) của một loài cây gỗ nhất định ở tuổi cơ sở (A_0, năm) trên lập địa quan tâm. Đại lượng H_D còn được gọi là chiều cao cây tầng trội. Những cây ưu thế là những cây khỏe mạnh nhất và có chất lượng di truyền tốt nhất. Tán lá của chúng phát triển vượt lên trên những cây khác, đỉnh tán lá và xung quanh tán lá đều được hấp thu ánh sáng hoàn toàn. Những cây đồng ưu thế là những cây khỏe mạnh và có chất lượng di truyền tốt. Chúng phát triển kém hơn so với cây ưu thế và chỉ đỉnh tán lá được hấp thu ánh sáng hoàn toàn. Nói chung, hai nhóm cây này là những cây to lớn nhất trong quần thụ. Sự sống của chúng phụ thuộc vào chất lượng lập địa. Chúng có sản lượng gỗ lớn nhất và chất lượng gỗ tốt nhất. Vì thế, trong tỉa thưa rừng, những cây ưu thế và đồng ưu thế được giữ lại tiếp tục nuôi dưỡng cho đến kỳ thu hoạch chính.

Tuổi cơ sở được chọn tùy thuộc vào chu kỳ sống của rừng hoặc chu kỳ kinh doanh rừng. Thông thường tuổi cơ sở được chọn ở thời điểm mà tại đó sinh trưởng H của cây gỗ và quần thụ không phụ thuộc vào mật độ. Tuổi cơ sở có thể được chọn tương ứng với chu kỳ kinh doanh hoặc sau khi ΔM_{Max} (Trousdell và ctv., 1974; Teshome và ctv., 2000). Tuổi cơ sở cũng có thể được chọn ở thời điểm mà chiều cao của những tuổi khác được dự đoán từ hàm chỉ số SI tại tuổi cơ sở với sai lệch nhỏ nhất (Onyekwelu, J.C., 2003).

Sở dĩ sử dụng những cây ưu thế và đồng ưu thế để xây dựng chỉ số SI là vì: (a) Sức sinh trưởng của chúng phụ thuộc chủ yếu vào đặc tính di truyền và chất lượng lập địa; (b) Chiều cao của chúng (H_D, m) có quan hệ chặt chẽ với M quần thụ và ít phụ thuộc vào mật độ (N, cây/ha); (c) Khi tỉa thưa rừng, chúng được giữ lại tiếp tục nuôi dưỡng cho đến kỳ thu hoạch chính (Larsen, 1999).

Trước đây chỉ số SI được xác định từ H trung bình của 100 cây ưu thế và đồng ưu thế hoặc 100 cây có D lớn nhất trong quần thụ. Chỉ số SI cũng được xác định từ 20% số cây có H từ lớn nhất trở xuống. Phương pháp đầu có nhược điểm là không thích hợp với những loài cây gỗ có mật độ thưa (N < 200 cây/ha). Phương pháp sau có nhược điểm là số lượng cây phải đo đếm rất cao ở rừng có mật độ cao hoặc quá ít ở rừng có mật độ thấp. Ở Mỹ, chỉ số SI được xác định từ H của 5 - 10 cây có D lớn nhất trên diện tích ô mẫu 100 m^2. Đề xuất này dựa trên cơ sở: (a) Những cây ưu thế và đồng ưu thế phân bố không đồng đều trên mặt đất; (b) Những cây có H lớn nhất cũng có D lớn nhất; (c) Chiều cao trung bình của 5 – 10 cây này

có quan hệ chặt chẽ với trữ lượng quần thụ; (d) Số cây ít làm giảm đáng kể chi phí điều tra rừng.

7.2.2.3. Những phương pháp xác định chỉ số lập địa rừng

Chỉ số lập địa rừng được xác định theo 3 phương pháp khác nhau (Monserud, 1984; Larsen, 1999). Một là phương pháp đường cong chỉ số SI trung bình của quần thụ. Phương pháp này được sử dụng từ những năm 1940 trở về trước. Theo phương pháp này, trước hết xác định H_D (m) tại tuổi cơ sở (A_0). Kế đến vẽ đồ thị biểu diễn quan hệ $H_D = f(A_0)$. Sau đó nắn đường cong $H_D = f(A_0)$ để nhận được đường cong chỉ số SI. Phương pháp này có ưu điểm là đơn giản, dễ tính toán. Nhược điểm của phương pháp này là độ chính xác phụ thuộc vào chủ quan của con người và kết quả không được kiểm chứng bằng những tiêu chuẩn thống kê.

Hai là phương pháp đo đạc H_D của rừng ở những tuổi và môi trường khác nhau. Sau đó xây dựng hàm $H_D = f(A)$ để phân chia các chỉ số SI. Phương pháp này có ưu điểm là đơn giản; kết quả được kiểm chứng bằng những tiêu chuẩn thống kê; chi phí thấp về thời gian, nhân lực và kinh phí. Nhược điểm của phương pháp này là độ chính xác phụ thuộc vào dụng cụ đo đạc H_D và khó áp dụng cho những quần thụ không có nhiều cấp A.

Ba là phương pháp giải tích thân cây ưu thế và đồng ưu thế ở những môi trường khác nhau và làm phù hợp đường cong chỉ số SI theo cây giải tích. Phương pháp này ra đời vào những năm 1980 và tồn tại cho đến ngày nay. Phương pháp này có ưu điểm là H_D được đo đạc chính xác và kết quả được kiểm chứng bằng những tiêu chuẩn thống kê. Nhược điểm của phương pháp này là phức tạp và chi phí cao về thời gian, nhân lực và kinh phí.

Nói chung, hàm chỉ số SI không chỉ phụ thuộc vào kiểu tăng trưởng của loài cây gỗ và phương pháp thu thập số liệu, mà còn vào chủ quan của nhà nghiên cứu. Nếu sử dụng phương pháp thu thập số liệu khác nhau và những hàm khác nhau, thì đường cong chỉ số SI cũng khác nhau. Điều đó không chỉ dẫn đến giải thích kết quả khác nhau, mà còn đánh giá chất lượng lập địa khác nhau.

Để phản ánh đúng chất lượng lập địa, các hàm chỉ số SI phải được xây dựng từ những cây mẫu đại diện tốt nhất cho quần thụ ở lập địa quan tâm. Đây là công việc rất quan trọng. Những cây mẫu được chọn là những cây sinh trưởng tốt nhất. Tuổi cây và quần thụ được phân chia theo cấp. Khoảng cách giữa mỗi cấp A không nên quá lớn, thường là 2 năm đối với cây mọc nhanh, 5 năm đối với cây mọc trung bình và 10 năm đối với cây mọc chậm.

Nói chung, trình tự xác định chỉ số SI như sau: (a) Đối với mỗi loài cây, chọn ít nhất 10 cây ưu thế và đồng ưu thế cho mỗi cấp A trên mỗi lập đia. Những cây này phải có thân thẳng, tròn đều, không bị cụt ngọn hay hai thân, vòng năm rõ ràng…; (b) Giải tích và đo đạc chính xác H_D (m); (c) Xác định chính xác A quần thụ tại D ngang ngực bằng chặt hạ cây hoặc khoan tăng trưởng; (d) Sử dụng H_D, A và hàm sinh trưởng thích hợp để xác định chỉ số SI đối với môi trường nhất định; (e) Xác

định chỉ số SI trung bình từ tất cả cây mẫu. Đối với mỗi loài cây gỗ, chỉ số SI có thể được phân chia từ 3 – 7 cấp; trong đó mỗi cấp chỉ số SI được gọi là một cấp lập địa.

7.2.2.4. Ý nghĩa của chỉ số lập địa rừng

Chỉ số lập địa rừng được sử dụng trong lâm học, điều tra rừng, quản lý và kinh doanh rừng. Trong lâm học, chỉ số SI được sử dụng để mô tả và phân loại điều kiện môi trường, đánh giá hiệu quả của các kỹ thuật lâm sinh, chọn lập địa để trồng và nuôi rừng. Trong điều tra rừng, chỉ số SI được sử dụng để đánh giá năng suất rừng theo lập địa (tốt, trung bình, xấu), ước lượng tăng trưởng và xây dựng biểu sản lượng rừng. Trong quản lý rừng, chỉ số SI được sử dụng để phân chia nhỏ rừng theo các đơn vị để quản lý. Trong kinh doanh rừng, chỉ số SI được sử dụng để phân tích khả năng cung cấp gỗ của lập địa, ước lượng chu kỳ kinh doanh, quy mô rừng và đất rừng cần được xử lý, tuổi khai thác rừng và sản lượng gỗ thu hoạch từ những quần thụ khác nhau.

7.2.3. Phương pháp xây dựng biểu chỉ số lập địa rừng

Phần này giới thiệu phương pháp xây dựng biểu chỉ số lập địa rừng dựa trên đối tượng là rừng Tràm (*Melaleuca cajuputi* Powell). Biểu chỉ số SI và đường cong chỉ số SI được xây dựng thông qua 6 bước.

(1) Xác định phạm vi lập biểu chỉ số SI và thu số liệu để xây dựng hàm chỉ số SI.
(2) Xây dựng hàm $H_D = f(A)$ bình quân chung.
(3) Xác định tuổi cơ sở (A_0) để xây dựng hàm và đường cong chỉ số SI.
(4) Xác định số lượng các cấp chỉ số SI.
(5) Xây dựng các hàm chỉ số SI ở giữa các cấp chỉ số SI tại tuổi A_0.
(6) Kiểm định khả năng ứng dụng của các hàm chỉ số SI.
(7) Xây dựng biểu và đường cong chỉ số SI.

7.2.3.1. Phạm vi lập biểu chỉ số SI và số liệu để xây dựng hàm chỉ số SI

Phạm vi xây dựng biểu chỉ số lập địa đối với rừng Tràm là toàn bộ khu vực Tây Nam Bộ. Số liệu được thu thập tại ba tỉnh Long An, Kiên Giang và Cà Mau. Tổng số 98 cây mẫu ưu thế và đồng ưu thế từ 8 – 14 năm; trong đó 32 cây ở tỉnh Long An, 34 cây ở khu vực U Minh Thượng thuộc tỉnh Kiên Giang và 32 cây ở khu vực U Minh Hạ thuộc tỉnh Cà Mau. Số lượng cây mẫu này được phân phối vào ba cấp đất tốt, trung bình và xấu tương ứng là 33, 35 và 30 cây. Khả năng ứng dụng của biểu chỉ số SI được kiểm định từ 9 cây mẫu không tham gia xây dựng hàm chỉ số SI; trong đó mỗi cấp lập địa là 3 cây.

Những cây ưu thế và đồng ưu thế được chặt hạ ở vị trí cách cỗ rễ 20 cm. Thông tin thu thập trên những cây mẫu bao gồm D, H và V. Để đạt được mục đích này, tất cả cây mẫu sau khi chặt hạ được đo đạc D ngang ngực và H toàn thân bằng thước dây với độ chính xác 0,10 cm. Kế đến phân chia thân cây mẫu thành những phân đoạn có chiều dài từ 1,0 m, riêng đoạn ngọn còn khoảng 0,5 – 1,0 m. Đối với mỗi phân đoạn, đo đạc đường kính hai đầu lớn và nhỏ (D_{Max} và D_{Min}). Số liệu này là cơ sở để tính thể tích các phân đoạn trên thân cây bằng công thức kép tiết diện bình

quân. Sau đó cưa thớt giải tích ở các vị trí 0,0; 1,0; 1,3; 2,0; 3,0 (m)...và đáy đoạn ngọn. Các thớt giải tích được tập hợp theo từng cây giải tích; sau đó ghi chú thứ tự cây và vị trí thớt theo hướng về phía ngọn cây. Sau khi xử lý, bề rộng vòng năm của các thớt được xác định bằng kính lúp với độ chính xác 0,1 mm. Số liệu cây giải tích được tóm tắt ở Bảng 7.1. Phân tích số liệu ở Bảng 7.1 cho thấy giá trị H_{DBq} gia tăng dần từ cấp tuổi 2 (2,0 m) đến cấp tuổi 8 (8,0 m) và cấp tuổi 14 (11,5 m). Biên độ biến động của H_D dao động từ 2,0 m ở cấp tuổi 2 đến 6,0 m ở cấp tuổi 8 và 6,6 m ở cấp tuổi 10. Hệ số biến động (CV%) của H_D có khuynh hướng giảm dần từ cấp tuổi 2 (17,5%) đến cấp tuổi 10 (16,9%) và cấp tuổi 14 (13,9%). Nói chung, giá trị H_D của rừng Tràm trồng biến động khá lớn theo điều kiện lập địa.

Bảng 7.1. Đặc trưng thống kê chiều cao của những cây ưu thế và đồng ưu thế của rừng Tràm trồng từ 2 – 14 ở khu vực Tây Nam Bộ. Nguồn: Phạm Xuân Quý, 2010.

A(năm)	N(cây)	H_D(m)	±S	H_{DMin}	H_{DMax}	H_{DMax}-H_{DMin}	CV%
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2	98	2,0	0,4	1,5	3,5	2,0	17,5
4	98	4,2	0,7	2,8	6,4	3,6	17,6
6	98	6,3	1,1	4,0	8,5	4,5	17,8
8	98	8,0	1,4	5,0	11,0	6,0	17,8
10	80	9,5	1,6	6,0	12,4	6,6	16,9
12	6	10,4	1,8	7,6	11,5	3,9	17,3
14	6	11,5	1,6	8,1	13,0	4,9	13,9

7.2.3.2. Xây dựng những hàm chỉ số lập địa đối với rừng Tràm

Trong nghiên cứu này, hàm $H_0 = f(A)$ thích hợp được chọn ở dạng hàm Schumacher (1939) (Hàm 7.1). Biến đổi tuyến tính của hàm 7.1 có dạng hàm 7.2 – 7.5.

$H_D = m*exp(-b/A^c)$ (7.1)

$Ln(H_D) = Ln(m) - b/A^c$ (7.2)

Đặt $Ln(m) = b_0$ và $b = b_1$, chúng ta có:

$Ln(H_D) = b_0 + b_1/A^c$ (7.3)

$b_0 = Ln(H_D) – b_1/A^c$ (7.4)

$H_D = exp(b_0 – b_1/A^c)$ (7.5)

Theo định nghĩa, chỉ số SI là H_D tại tuổi cơ sở (A_0, năm). Theo đó, từ hàm (7.3) và (7.4), xác định được hàm SI = f(A) dưới dạng hàm (7.6) và b_0 dưới dạng hàm (7.7).

$Ln(SI) = b_0 + b_1/A_0^c$ (7.6)

$b_0 = Ln(SI) – b_1/A_0^c$ (7.7)

Thay b_0 ở hàm (7.7) vào hàm (7.3), chúng ta có hàm 7.8 – 7.10.

$Ln(H_0) = Ln(SI) – b_1/A_0^c + b_1/A^c$ (7.8)

$Ln(SI) = Ln(H_D) – b_1/A_0^c + b_1/A^c$ (7.9)

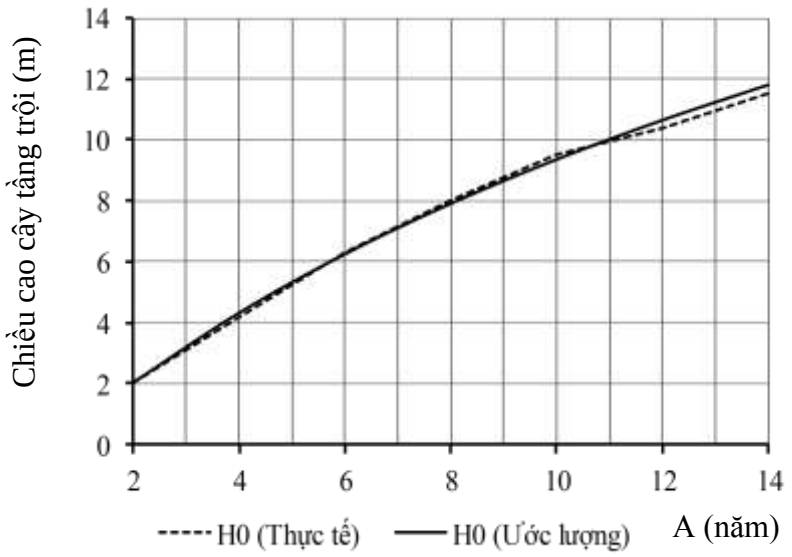
$SI = exp(Ln(H_D) – b_1(1/A^c - 1/A_0^c))$ (7.10)

Hàm (7.10) là hàm chỉ số SI tại tuổi A_0 hay H_D tại tuổi A_0.

Từ số liệu ở Bảng 7.1, bằng phương pháp phân tích hồi quy phi tuyến theo trọng số của Marquardt, xác định được hàm Schumacher (7.5) có dạng như hàm 7.11 (Hình 7.1). Hàm này nhận hệ số xác định rất cao (r^2 = 99,84%) và sai lệch nhỏ (MAPE = 3,3%). Vì thế, hàm 7.11 được sử dụng để ước lượng H_0 của rừng Tràm trồng từ cấp tuổi 2 – 14.

$H_D = exp(4,54876 – 4,77608/A^{0,314926})$ (7.11)

r^2 = 99,84%; SEE = ±1,12; MAE = 0,10; MAPE = 3,3%; SSE = 13,8.



Hình 7.1. Đồ thị biểu diễn sự biến đổi chiều cao cây tầng trội theo tuổi rừng Tràm trồng ở khu vực Tây Nam Bộ.

7.2.3.3. Xác định tuổi cơ sở để xây dựng hàm và đường cong chỉ số SI

Về lý thuyết, nếu chọn A_0 ở những thời điểm khác nhau, thì sai lệch dự đoán H_D tại tuổi A cũng khác nhau. Đối với rừng Tràm ở khu vực Tây Nam Bộ, biên độ tuổi dao động từ 2 – 14 năm. Vì thế, tuổi A_0 thích hợp được chọn trong khoảng 6 - 14 năm. Tuổi A_0 thích hợp là thời điểm mà hàm H_D tại A_0 được sử dụng để ước lượng H_D tại tuổi A (năm) với sai lệch nhỏ nhất (SSE_{Min}). Kết qủa kiểm định sai lệch dự đoán H_D từ tuổi 6 – 14 năm bằng hàm (7.11) được ghi lại ở Bảng 7.2 và Bảng 7.3. Từ số liệu ở Bảng 7.3 cho thấy, SSE nhận giá trị cao nhất tại tuổi 6 (2,350), thấp nhất tại tuổi 10 (0,189). Vì thế, tuổi 10 được chọn là tuổi cơ sở để phân chia các cấp chỉ số SI đối với rừng Tràm trồng ở khu vực Tây Nam Bộ.

7.2.3.4. Xác định số lượng chỉ số lập địa tại tuổi cơ sở

Từ số liệu ở Bảng 7.1 cho thấy, biên độ dao động của H_D tại tuổi 10 là 6,6 m. Nếu phân chia H_D (6,0 m) tại A_0 thành 3 cấp, thì mỗi cấp H_D là 2,0 m. Tại A_0 = 10

năm, H_D = 9,5 m (làm tròn 10 m). Vì thế, ba chỉ số SI tại A_0 nhận những giá trị tương ứng là 12,0 m đối với cấp lập địa I; 10,0 m đối với cấp lập địa II và 8,0 m đối với cấp lập địa III. Các chỉ số SI ở ranh giới giữa cấp lập địa I và II, II và III tương ứng là 11 và 9,0 m. Tương tự, chỉ số SI ở biên độ trên của cấp lập địa I là 13,0 m, còn biên độ dưới của cấp lập địa III là 7,0 m.

Bảng 7.2. Ước lượng H_D của rừng Tràm trồng ở khu vực Tây Nam Bộ bằng hàm SI = f(A) khi giả định A_0 = 6 - 14 năm.

A (năm)	H_{DTN}(m)	Giá trị ước lượng H_D (m) tại tuổi A_0 (năm):				
		6	8	10	12	14
(1)	(2)	(3)	(4)	(5)	(6)	(7)
6	6,3	8,0	7,0	6,9	6,6	6,6
8	8,0	9,8	8,6	8,4	8,1	8,1
10	9,5	11,0	9,7	9,5	9,1	9,1
12	10,4	11,9	10,5	10,3	9,9	9,8
14	11,5	12,6	11,1	10,9	10,5	10,4

Bảng 7.3. Sai lệch ước lượng H_D của rừng Tràm ở khu vực Tây Nam Bộ bằng hàm SI = f(A) khi giả định A_0 = 6 - 14 năm

A (năm)	Giá trị SSE tương ứng với tuổi A_0 (năm):				
	6	8	10	12	14
(1)	(2)	(3)	(4)	(5)	(6)
6	2,890	0,560	0,371	0,122	0,089
8	3,103	0,360	0,186	0,013	0,003
10	2,248	0,036	0,000	0,128	0,182
12	2,283	0,009	0,013	0,250	0,330
14	1,227	0,154	0,373	1,042	1,210
Trung bình	2,350	0,224	0,189	0,311	0,363

7.2.3.5. Xây dựng các hàm chỉ số lập địa đối với rừng Tràm

Các hàm chỉ số SI đối với rừng Tràm được chọn với độ dốc (b_1) bằng nhau. Từ hàm 7.11 cho thấy, độ dốc b_1 = -4,77608. Bằng cách thay thế b_1, A_0 và H_D ở giữa và ở giới hạn trên của các cấp lập địa I – III vào hàm 7.10, kết quả nhận được các hàm chỉ số SI (Bảng 7.4). Ở hàm 7.12 – 7.18 của Bảng 7.4, giá trị 0,48425 = 1/10^0,314926.

7.2.3.6. Kiểm định mức độ phù hợp của các hàm chỉ số lập địa

Mức độ phù hợp của các hàm chỉ số SI so với số liệu thực tế có thể được xác định theo hai phương pháp. Phương pháp thứ 1 là so sánh bằng đồ thị. Theo phương pháp này, nếu các đường cong mô tả sự biến đổi H_D của những cây không tham gia

xây dựng hàm $H_D = f(A)$ (Gọi là cây kiểm tra) thuộc mỗi cấp lập địa không vượt ra ngoài giới hạn dưới và giới hạn trên của cấp lập địa tương ứng, thì các hàm chỉ số SI là phù hợp. Ưu điểm của phương pháp đồ thị là đơn giản và dễ thực hiện. Nhược điểm của phương pháp đồ thị là kết quả không được kiểm chứng bằng những tiêu chuẩn thống kê.

Bảng 7.4. Các hàm chỉ số SI đối với rừng Tràm trồng tại khu vực Tây Nam Bộ.

Cấp SI	Hàm chỉ số SI tại tuổi A năm:	
(1)	(2)	
$I_{(Trên)}$	SI = exp((Ln(13) - 4,77608*(1/A^0,314926 - 0,48425)))	(7.12)
I	SI = exp((Ln(12) - 4,77608*(1/A^0,314926 - 0,48425)))	(7.13)
I - II	SI = exp((Ln(11) - 4,77608*(1/A^0,314926 - 0,48425)))	(7.14)
II	SI = exp((Ln(10) - 4,77608*(1/A^0,314926 - 0,48425)))	(7.15)
II - III	SI = exp((Ln(9) - 4,77608*(1/A^0,314926 - 0,48425)))	(7.16)
III	SI = exp((Ln(8) - 4,77608*(1/A^0,314926 - 0,48425)))	(7.17)
$III_{(Dưới)}$	SI = exp((Ln(7) - 4,77608*(1/A^0,314926 - 0,48425)))	(7.18)

Phương pháp thứ 2 là phương pháp hồi quy. Phương pháp này so sánh điểm chặn và độ dốc của các hàm $H_D = f(A)$ đối với những cây kiểm tra và các hàm chỉ số SI ở giữa các cấp chỉ số SI. Để đạt mục đích này, trước hết xây dựng các hàm $H_D = f(A)$ riêng rẽ đối với những cây kiểm tra theo từng cấp lập địa. Kế đến tuyến tính hóa các hàm ước lượng chỉ số SI ở giữa ba cấp lập địa và các hàm ước lượng $H_D = f(A)$ từ những cây kiểm tra tương ứng với ba cấp lập địa. Sau đó áp dụng phương pháp so sánh các đường hồi quy để kiểm định sự thuần nhất về điểm chặn và độ dốc của các hàm này. Giả định H_0^+: Không có khác biệt về điểm chặn và độ dốc của ba hàm chỉ số SI với ba hàm $H_D = f(A)$ đối với những cây kiểm tra. Giả định này được kiểm định bằng thống kê F ở mức ý nghĩa $\alpha = 0,05$. Nếu $P < 0,05$, thì giả thuyết bị bác bỏ. Trái lại, nếu $P > 0,05$, thì giả thuyết được cấp nhận.

Số liệu của những cây mẫu tham gia xây dựng các hàm chỉ số SI theo ba cấp lập địa được dẫn ra ở Bảng 7.1. Số liệu H_D của 9 cây mẫu không tham gia xây dựng hàm chỉ số SI được ghi lại ở Bảng 7.5; trong đó mỗi cấp lập địa là 3 cây.

Theo phương pháp đồ thị (Hình 7.2) cho thấy, đường cong chiều cao của những cây kiểm tra đều nằm trong giới hạn dưới và giới hạn trên của ba cấp chỉ số lập địa. Theo phương pháp hồi quy (Bảng 7.6) cho thấy, điểm chặn của ba hàm chỉ số SI và ba hàm $H_D = f(A)$ ở những cây kiểm tra có sự khác biệt rõ rệt ($P < 0,01$). Trái lại, độ dốc của chúng không có sự khác biệt rõ rệt ($P = 0,3649$ đối với cấp lập địa I; $P = 0,1840$ đối với cấp lập địa II và $P = 0,5386$ đối với cấp lập địa III). Điều đó chứng tỏ các hàm chỉ số SI (Hàm 7.12 - 7.18) là những hàm thích hợp để xây dựng biểu lập địa và đường cong chỉ số lập địa đối với rừng Tràm trồng từ 2 – 14 tuổi ở khu vực Tây Nam Bộ.

Bảng 7.5. Đặc trưng chiều cao của những cây ưu thế và đồng ưu thế không tham gia xây dựng hàm chỉ số SI đối với rừng Tràm trồng ở khu vực Tây Nam Bộ.

(a) Cấp lập địa I.

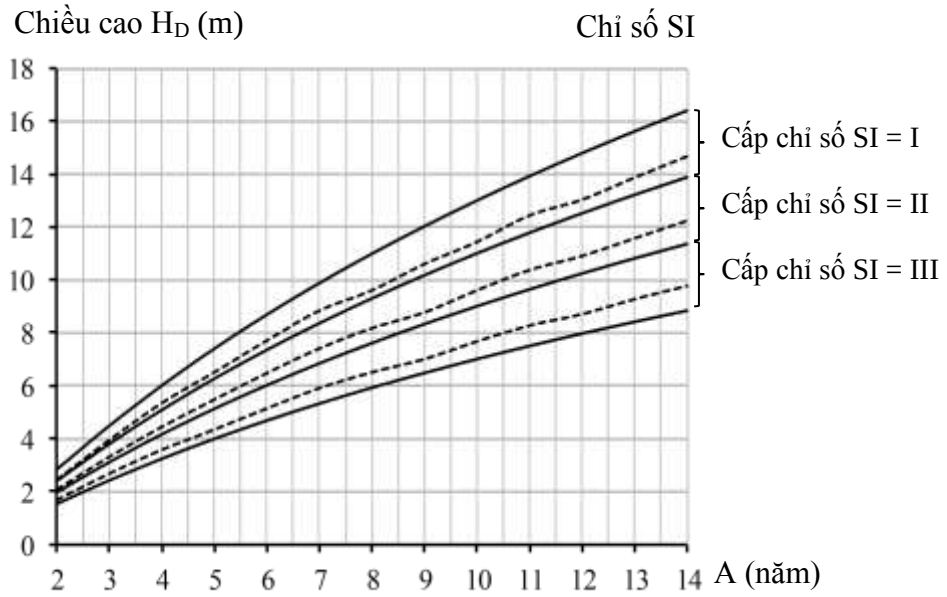
A (năm)	N (cây)	H_D (m)	±S	H_{DMin}	H_{DMax}	CV%
(1)	(2)	(3)	(4)	(5)	(6)	(7)
2	3	2,4	2,4	2,5	0,06	2,4
4	3	5,3	5,3	5,4	0,06	1,1
6	3	7,7	7,6	7,8	0,10	1,3
8	3	9,2	9,1	9,4	0,15	1,7
10	3	11,4	11,3	11,6	0,15	1,3
12	3	13,0	12,9	13,2	0,15	1,2
14	3	14,7	14,5	14,9	0,21	1,4

(b) Cấp lập địa II.

A (năm)	N (cây)	H_D (m)	±S	H_{DMin}	H_{DMax}	CV%
(1)	(2)	(3)	(4)	(5)	(6)	(7)
2	3	2,1	2,0	2,1	0,06	2,8
4	3	4,4	4,3	4,5	0,12	2,6
6	3	6,5	6,3	6,6	0,15	2,4
8	3	8,2	7,9	8,3	0,23	2,8
10	3	9,6	9,3	9,8	0,26	2,8
12	3	10,9	10,5	11,2	0,36	3,3
14	3	12,2	11,9	12,4	0,29	2,4

(c) Cấp lập địa III.

A (năm)	N (cây)	H_D (m)	±S	H_{DMin}	H_{DMax}	CV%
(1)	(2)	(3)	(4)	(5)	(6)	(7)
2	3	1,7	1,6	1,7	0,06	3,5
4	3	3,6	3,5	3,6	0,06	1,6
6	3	5,1	5,0	5,2	0,12	2,3
8	3	6,5	6,4	6,6	0,10	1,5
10	3	7,7	7,5	7,8	0,15	2,0
12	3	8,7	8,5	8,8	0,17	2,0
14	3	9,8	9,6	9,9	0,15	1,6



Hình 7.2. Đồ thị so sánh chiều cao của những cây kiểm tra và ba cấp chỉ số SI đối với rừng Tràm trồng từ 2 – 14 tuổi ở khu vực Tây Nam Bộ.

Bảng 7.6. So sánh điểm chặn (a) và độ dốc (b) của ba hàm chỉ số SI với điểm chặn (a') và độ dốc (b') của ba hàm $H_D = f(A)$ đối với những cây mẫu không tham gia xây dựng hàm chỉ số SI. Mô hình so sánh: $Ln(H_D) = a - b/A^{0,3149}$.

Tham số	Cấp lập địa:		
	I	II	III
(1)	(2)	(3)	(4)
a	4,7977	4,6154	4,3923
a'	4,7730	4,5913	4,3671
P	< 0,01	< 0,01	< 0,01
b	-4,7761	-4,7761	-4,7761
b'	-4,8194	-4,8024	-4,7925
P	0,3649	0,1840	0,5386
c	0,3149	0,3149	0,3149
c'	0,3149	0,3149	0,3149

7.2.3.7. Xây dựng biểu và các đường cong chỉ số SI đối với rừng Tràm

Biểu cấp lập địa (Cấp chỉ số SI) và các đường cong chỉ số SI được xây dựng bằng cách thay thế tuổi của rừng Tràm trồng từ 2 - 14 năm vào các hàm chỉ số SI (Hàm 7.12 - 7.18). Kết quả được ghi lại ở Bảng 7.7 và Hình 7.3.

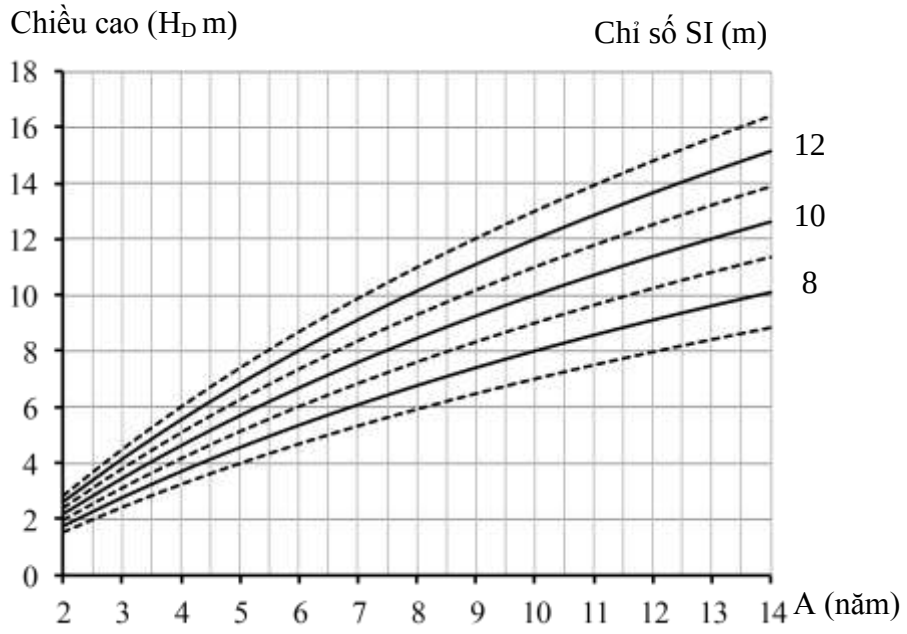
7.2.3.8. Đo chỉ số lập địa đối với rừng Tràm

Chỉ số lập địa của rừng Tràm trồng trên mỗi cấp lập địa là chiều cao trung bình của những cây ưu thế và đồng ưu thế (H_D, m) tại tuổi cơ sở (A_0, năm). Tuổi cơ sở của rừng Tràm trồng ở khu vực Tây Nam Bộ là 10 năm (A_0 = 10 năm). Chỉ số SI

tại A_0 = 10 năm trên cấp lập địa I là 12 m. Điều đó có nghĩa là khi rừng Tràm trồng đạt 10 tuổi, thì chiều cao trung bình của những cây ưu thế và đồng ưu thế trên cấp lập địa I là 12 m. Trong báo cáo kết quả, chỉ số lập địa được viết tắt là SI_{10} = 12 m. Khi mật độ trồng ban đầu như nhau và phương thức lâm sinh tương tự như nhau, thì năng suất gỗ của rừng Tràm trồng trên cấp lập địa I cao hơn so với cấp lập địa II và III.

Bảng 7.7. Biểu cấp lập địa đối với rừng Tràm trồng tại khu vực Tây Nam Bộ.

A (năm)	Chiều cao H_0 (m) của rừng Tràm trên ba cấp lập địa:						
	$I_{(Trên)}$	I	I - II	II	II - III	III	$III_{(Dưới)}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2	2,8	2,6	2,4	2,2	2,0	1,7	1,5
4	6,0	5,5	5,1	4,6	4,2	3,7	3,2
6	8,7	8,0	7,3	6,7	6,0	5,3	4,7
8	11,0	10,1	9,3	8,5	7,6	6,8	5,9
10	13,0	12,0	11,0	10,0	9,0	8,0	7,0
12	14,8	13,7	12,5	11,4	10,2	9,1	8,0
14	16,4	15,1	13,9	12,6	11,4	10,1	8,8



Hình 7.3. Đường cong chiều cao cây tầng trội (H_D, m) và đường cong chỉ số SI (m) đối với rừng Tràm tại khu vực Tây Nam Bộ.

Trong thực tế, khi xác định cấp lập địa cho một khoảnh rừng nào đó, trước hết đo H của 10 cây ưu thế và đồng ưu thế trong ô mẫu 100 – 200 m^2. Số lượng ô mẫu được xác định tùy theo diện tích của khoảnh rừng. Bởi vì giữa H và D có quan hệ chặt chẽ với nhau, nên H của 10 cây ưu thế và đồng ưu thế được chọn từ những cây có D lớn nhất trong ô mẫu. Kế đến xác định tuổi rừng thông qua lý lịch rừng hoặc bằng cách cắt thân ở vị trí 20 – 30 cm từ mặt đất trở lên và đếm tuổi. Sau đó sử dụng H_D và A của những cây này cùng với biểu chỉ số lập địa (Bảng 7.7) hoặc đồ thị cấp lập địa (Hình 7.3) để tra cấp chỉ số SI cho khoảnh rừng trên ô mẫu. Cấp chỉ SI là điểm giao nhau giữa H_D và A năm. Ví dụ: Tại A = 8 năm và H_D = 8,5 m, thì khoảnh rừng này thuộc cấp lập địa II (7,6 m < 8,5 m < 9,3 m).

Chương 8
PHÂN TÍCH NHÓM VÀ ỨNG DỤNG TRONG LÂM HỌC

8.1. Khái quát về phân tích nhóm

8.1.1. Định nghĩa phân tích nhóm

Phân tích nhóm (Grouping Analysis) là phân chia một tập dữ liệu thành nhiều nhóm dựa trên các biến phân loại định lượng. 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 8.1. Ở Hàm 8.1, $Z^{(k)}$ là hàm của nhóm thứ k; $X_1^{(k)}$, $X_2^{(k)}$,..., $X_i^{(k)}$ là biến phân loại thứ i của hàm lập nhóm thứ k (k = 1, 2…); a, b, c…f là các hệ số của hàm lập nhóm.

$$Z^{(k)} = a*X_1^{(k)} + b*X_2^{(k)} + c*X_3^{(k)} + \ldots + f*X_i^{(k)} \qquad (8.1)$$

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.

Để mô tả sự khác nhau giữa các nhóm, phương pháp phân tích nhóm xây dựng các hàm lập nhóm tuyến tính. Hàm này do Fisher đề xuất vào năm 1936, nên còn gọi là hàm lập nhóm tuyến tính Fisher. Khi chỉ có hai biến phân loại, thì hàm lập nhóm tuyến tính Fisher có dạng phương trình của một mặt phẳng. Khi số nhóm lớn hơn 2, các biến hợp quy (Canonical Variables) là trung tâm của phép phân tích nhóm. Biến hợp quy là sự kết hợp tuyến tính của các biến phân loại để làm khác nhau cực đại giữa các số trung bình của k nhóm theo một phương nào đó. Khi có k nhóm, thì số hàm hợp quy là k – 1. Biến hợp quy thứ nhất là sự kết hợp tuyến tính của các biến phân loại để mô tả sự khác nhau cực đại giữa các số trung bình của k nhóm theo phương thứ nhất. Biến hợp quy thứ 2 là sự kết hợp tuyến tính của các biến phân loại để mô tả sự khác nhau cực đại giữa các số trung bình của k nhóm theo phương thứ 2. Biến hợp quy thứ 3 là sự kết hợp tuyến tính của các biến phân loại để mô tả sự khác nhau cực đại giữa các số trung bình của k nhóm theo phương thứ ba vuông góc với hai phương 1 và 2. Biến hợp quy thứ k - 1 là sự kết hợp tuyến tính của các biến phân loại để mô tả sự khác nhau cực đại giữa các số trung bình của k nhóm theo một phương vuông góc với các phương trên. Mỗi biến hợp quy là một yếu tố để phân biệt các trung bình giữa các nhóm. Biến hợp quy còn được gọi là hàm phân loại hợp quy (Canonical Discriminant Function). Các hàm phân loại hợp quy có ba loại: (1) Các hàm phân loại hợp quy chuẩn hóa (Standardized Canonical Discriminant Functions); (2) Các hàm phân loại hợp quy chưa chuẩn hóa

(Unstandardized Canonical Discriminant Functions); (3) Các hàm phân loại hợp quy ở trung tâm nhóm (Functions at Group Centroids).

Hàm phân loại hợp quy chuẩn hóa được sử dụng để so sánh vai trò hay ý nghĩa của từng biến phân loại trong hàm hợp quy. Vai trò cao thấp của các biến phân loại được xác định theo giá trị tuyệt đối của các hệ số trong hàm lập nhóm. Hàm phân loại hợp quy chưa chuẩn hóa được sử dụng để tính điểm số phân loại cho từng trường hợp. Điểm số phân loại trung bình của các trường hợp trong mỗi nhóm tương ứng với mỗi hàm hợp quy được báo cáo qua hàm ở trung tâm nhóm. Các hàm ở trung tâm nhóm dùng để so sánh sự khác nhau giữa các nhóm. Các hàm hợp quy 1 và 2 được sử dụng để xây dựng đồ thị và bản đồ mô tả các nhóm. Kết quả phân tích nhóm cung cấp các hàm để phân loại các trường hợp mới vào những nhóm khác nhau.

Sự thành công của việc phân loại các nhóm phụ thuộc vào tổ hợp các biến được chọn để xây dựng hàm phân loại. Các hàm phân loại được xây dựng từ nhiều biến dự đoán cho kết quả thành công cao hơn so với các hàm phân loại từ ít biến dự đoán. Nguyên nhân là vì các đặc tính của nhóm được phản ánh thông qua nhiều biến khác nhau. Tuy vậy, hàm phân loại bao gồm nhiều biến dự đoán không chỉ làm tăng tính phức tạp trong đo đạc các biến, mà còn gây ra khó khăn cho việc ứng dụng vào thực tế. Vì thế, xác định những biến dự đoán hữu ích để xây dựng hàm phân loại là một việc làm cần thiết. Phương pháp chọn biến từng bước (Stepwise Method) cho phép chọn những biến dự đoán hữu ích nhất từ một tập hợp biến dự tuyển. Các biến dự tuyển phản ánh các đặc trưng của tổng thể. Chúng có thể được chọn để xây dựng hàm lập nhóm. Theo phương pháp chọn biến từng bước, trước hết xác định một biến tốt nhất từ một tập hợp biến dự tuyển, nghĩa là biến có trung bình khác nhau nhiều nhất giữa các nhóm. Tiếp theo chọn những biến tốt khác bằng cách lần lượt đưa từng biến dự tuyển vào phân tích. Các biến dự tuyển được đưa vào hay loại ra khỏi mô hình ở từng bước được kiểm định theo tiêu chuẩn Wilk's Lambda, tiêu chuẩn khoảng cách bình phương Mahalanobis (D^2), tiêu chuẩn tỷ lệ F nhỏ nhất (Smallest F ratio) và tiêu chuẩn Rao's V; trong đó hai tiêu chuẩn Wilk's Lambda và Mahalanobis được sử dụng nhiều nhất.

Theo tiêu chuẩn Wilk's Lambda, những biến được sử dụng để xây dựng hàm lập nhóm là những biến có giá trị F_{Ra} (F to Remove) > 2,71 (Mặc định). Trái lại, khi những biến có giá trị F_{Ra} < 2,71, thì chúng bị loại ra khỏi mô hình phân loại. Một biến có thể được đưa vào mô hình phân loại khi nó làm thay đổi thống kê Wilk's Lambda một cách có ý nghĩa. Sự thay đổi của thống kê Wilk's Lambda được kiểm định theo tiêu chuẩn F_{Change} (Công thức 8.2). Ở công thức 8.2, N = tổng số các trường hợp đưa vào phân loại, g = số nhóm phân loại, p = số biến số trong mô hình phân loại, λ_p = Wilk's Lambda ứng với p biến trước khi đưa thêm một biến vào hàm phân loại, λ_{p+1} = Wilk's Lambda ứng với p + 1 biến sau khi đưa thêm một biến vào hàm phân loại.

$$F_{Change} = \frac{N-g-p}{g-1} * \frac{1 - \lambda_{p+1}/\lambda_p}{\lambda_{p+1}/\lambda_p} \qquad (8.2)$$

Phương pháp khoảng cách bình phương của Mahalanobis dựa trên cơ sở của hàm lập nhóm tuyến tính (Công thức 8.3); trong đó m là số biến để xây dựng hàm lập nhóm; k là số nhóm phân loại hay số nghiệm thức so sánh (k = 1, 2, 3…); b_i là các tham số; $X_m^{(k)}$ là biến dự đoán thứ m thuộc nhóm phân loại k.

$$Z^{(k)} = b_1X_1^{(k)} + b_2X_2^{(k)} + ... + b_mX_m^{(k)} \qquad (8.3)$$

Hàm lập nhóm tuyến tính chỉ được xác định khi các nhóm khác nhau thực sự. Vì thế, giả thuyết H_0 đặt ra là các nhóm khác nhau không rõ rệt. Để kiểm định giả thuyết này, Mahalanobis sử dụng số đo khoảng cách D^2. Sự khác biệt giữa các nhóm được kiểm định theo tiêu chuẩn F (Công thức 8.4); trong đó n_1 và n_2 là dung lượng quan sát của nhóm 1 và nhóm 2. Khi $F > F_{05}$ hay $P < 0,05$ (Mặc định), thì hai nhóm khác nhau rõ rệt. Khi hai nhóm khác nhau rõ rệt, thì bước tiếp theo là xây dựng hàm lập nhóm tuyến tính Fisher và biểu đồ để mô tả quy luật phân bố của các trường hợp trong các nhóm. Sau đó hàm lập nhóm tuyến tính Fisher được sử dụng để phân loại các trường hợp mới vào các nhóm khác nhau.

$$F = \frac{n_1 * n_2}{n_1 + n_2} * \frac{n_1 + n_2 - m - 1}{n_1 + n_2 - 2} * D^2 \qquad (8.4)$$

8.1.2. Mục đích và mục tiêu của phân tích nhóm

Mục đích của phân tích nhóm là cung cấp những hàm để dự đoán những trường hợp mới thuộc về các nhóm khác nhau. Những trường hợp mới là những trường hợp không được sử dụng để xây dựng các hàm lập nhóm.

Phân tích nhóm có ba mục tiêu cơ bản: (1) Xác định những biến độc lập 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.

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

Phân tích nhóm có thể được sử dụng để giải quyết những vấn đề sau đây: (1) Phân loại số liệu thành hai hay nhiều nhóm; (2) Tìm kiếm một hàm số từ các biến số dự đoán (định lượng hoặc định tính) để phân chia các nhóm. Trong lâm học, phân tích nhóm được sử dụng để giải quyết một số vấn đề dưới đây.

(1) Phân loại (rừng, cấp lập địa, cấp nguy cơ cháy rừng, cấp năng suất rừng…). Để giải quyết những vấn đề đặt ra, người ta bắt đầu từ những đối tượng đã được phân loại sơ bộ theo một số đặc trưng nào đó. Tiếp theo xây dựng các hàm phân loại các đối tượng dựa trên những đặc trưng (biến số) của các đối tượng đã biết. Sau đó sử dụng các hàm phân loại này để phân chia những đối tượng mới thuộc về các nhóm khác nhau.

(2) Kiểm định sự khác nhau giữa các nghiệm thức dựa trên nhiều biến số khác nhau. Chẳng hạn: Phân tích ảnh hưởng của 4 phương thức lâm sinh đến sinh trưởng của rừng dựa trên các biến D, H, G, V và xác định những phương thức lâm sinh mang lại hiệu quả giống nhau nhiều nhất hoặc khác nhau nhiều nhất.

(3) Xác định những tập hợp biến hữu ích nhất cho việc phân loại hoặc so sánh giữa các đối tượng. Những đối tượng cần phân loại thường có rất nhiều đặc trưng khác nhau. Vì thế, tùy theo mục tiêu đặt ra, nhà nghiên cứu cần phải chọn ra những biến đặc trưng và có ý nghĩa nhất cho việc phân chia các nhóm.

(4) So sánh vai trò của những tập hợp biến dự đoán khác nhau được sử dụng để xây dựng hàm lập nhóm. Để giải quyết vấn đề đặt ra, người ta chọn những tập hợp biến phân loại khác nhau để xây dựng các hàm lập nhóm. Một tập hợp biến phân loại tốt hơn khi hàm phân loại được xây dựng từ những biến này cho kết quả phân loại tính bằng phần trăm cao hơn. Nói chung, không có phương pháp nào cho phép chọn được một tập hợp biến tốt nhất, mà chỉ có những phương pháp cho phép chọn được một tập hợp biến tốt hơn. Vì thế, khi phân loại các đối tượng, nhà nghiên cứu cần phải tìm kiếm những tập hợp biến cho kết quả phân loại cao hơn. Một hàm phân loại bao gồm nhiều biến dự đoán cho kết quả phân loại cao hơn so với hàm phân loại chỉ bao gồm ít biến. Tuy vậy, việc xây dựng những hàm nhiều biến không chỉ phức tạp, mà còn mất nhiều chi phí về nhân lực, thời gian, kinh phí và khả năng ứng dụng. Chính vì thế, tùy theo yêu cầu đặt ra, nhà lâm học cần xác định một tập hợp biến hữu ích cho việc phân loại các nhóm.

(5) Tìm kiếm những giá trị bất thường hay ngoại lai. Những giá trị bất thường làm tăng biến động trong một nhóm. Những giá trị này sẽ nhận các điểm số bất thường.

8.1.4. Những điều kiện trong phân tích nhóm

Phân tích nhóm đòi hỏi phải thỏa mãn 6 điều kiện cơ bản. Một là số nhóm phân loại từ 2 trở lên. Hai là mỗi nhóm phải có hai biến phân loại trở lên. Các biến phân loại này là những đại lượng nhận giá trị liên tục. Ba là các biến phân loại là các biến định lượng và khác nhau thực sự 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). Bốn là các biến phân loại của mỗi nhóm phải có phân bố chuẩn. Điều kiện này được kiểm định theo thống kê χ^2. Năm là phương sai của các biến phân loại phải bằng nhau giữa các nhóm. Điều kiện này được kiểm định theo thống kê Levence. Sáu là ma trận hiệp phương sai trong nội bộ nhóm phải bằng nhau giữa các nhóm. Điều kiện này được kiểm định bằng thống kê M của Box (Box's M Test).

8.1.5. Các bước cơ bản trong phân tích nhóm

Bước 1. Phân tích đối tượng nghiên cứu. Trong lâm học, đối đối tượng nghiên cứu có thể là rừng phân bố trên những môi trường khác nhau; rừng được xử lý theo những kỹ thuật lâm sinh khác nhau; các nghiệm thức trong gieo ươm và trồng rừng...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.

Kết quả phân loại các đối tượng phụ thuộc vào số lượng biến dự đoán được sử dụng để xây dựng hàm phân loại nhóm. Khi các biến dự đoán khác nhau rõ rệt giữa các nhóm, thì hàm lập nhóm cho kết quả phân loại với độ chính xác cao. Trái lại, nếu các biến dự đoán kém hữu ích cho việc phân loại, thì hàm lập nhóm cho kết quả phân loại với độ chính xác thấp. Vì thế, xác định những biến hữu ích cho việc phân loại là việc làm cần thiết. Sự khác nhau của các biến số dự đoán giữa các nhóm được kiểm định bằng tiêu chuẩn Wilk's Lambda (Công thức 8.5). Tiêu chuẩn Wilk's Lambda là tỷ lệ giữa tổng các độ lệch bình phương trong các nhóm (SSE_{Group}) với tổng các độ lệch bình phương của tất cả các nhóm gộp lại ($SSE_{Tổng}$).

$$\text{Wilk's Lambda} = \frac{SSR_{\text{Trong nhóm}}}{SSR_{\text{Tổng}}} \qquad (8.5)$$

Tiêu chuẩn Wilks's Lambda = 0 – 1. Giá trị 0 cho biết các biến dự đoán có ý nghĩa phân loại hoàn toàn. Khi Wilk's Lambda rất nhỏ hoặc P < 0,01, thì trung bình của các biến dự đoán khác nhau rõ rệt giữa các nhóm hay các nhóm khác nhau rõ rệt. Giá trị 1 cho biết các biến dự đoán không có ý nghĩa phân loại.

Hàm lập nhóm có thể được xây dựng bằng nhiều phương pháp khác nhau. Một là hàm lập nhóm bao gồm tất cả các biến dự đoán. Hai là hàm lập nhóm chỉ bao gồm những biến dự đoán hữu ích cho việc phân loại. Những biến hữu ích được xác định theo phương pháp hàm lập nhóm từng bước (Stepwise Discriminant Function Method). Hàm này có thể được xây dựng theo 4 phương pháp cơ bản: Wilk's Lambda, khoảng cách bình phương của Mahalanobis (Mahalanobis Squared Distance), tỷ lệ F nhỏ nhất (Smallest F ratio) và Rao's V.

Bước 2. Xác định mục tiêu phân loại nhóm và độ chính xác của các hàm phân loại nhóm. Khi phân tích nhóm, trước hết cần định nghĩa rõ ràng mục tiêu hay lý do của việc xây dựng các hàm lập nhóm. Phương pháp xây dựng các hàm lập nhóm thay đổi tùy theo mục tiêu nghiên cứu.

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. Tập hợp biến này được chọn ra từ những biến dự tuyển. Lưu ý rằng không có phương pháp nào cho chép chọn được một tập hợp biến tốt nhất, mà chỉ có những phương pháp cho chép chọn được một tập hợp biến tốt hơn. Vì thế, để xây dựng hàm phân loại các đối tượng, nhà lâm học cần xác định những tập hợp biến cho kết quả phân loại cao hơn. Để đạt mục tiêu này, các biến trong hàm phân loại được xác định từ những tập hợp biến dự tuyển khác nhau.

Nếu mục tiêu nghiên cứu là so sánh sự khác biệt 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...Trong trường hợp này, chúng ta chỉ chọn những tập hợp biến hữu ích cho việc so sánh các nhóm. Nếu các nhóm khác nhau không rõ rệt hoặc có nhiều nhóm gần giống nhau,

thì kết quả phân loại sẽ thấp. Ngược lại, khi các nhóm khác nhau rõ rệt, thì kết quả phân loại sẽ cao. Khi hai nhóm giao nhau một phần, thì tùy theo phần giao nhau mà chúng khác nhau nhiều hay ít. Nếu phần giao nhau càng lớn, thì hai nhóm càng giống 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 nhóm sinh trưởng cho những cây hình thành rừng 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 tại đó hàm lập nhóm tuyến tính Fisher nhận được giá trị lớn nhất. Sự thành công của việc phân loại ban đầu được báo cáo thông qua các thống kê cho từng trường hợp (Casewise Statistics), bản đồ phân loại (Territorial Map), đồ thị phân tán giữa các nhóm (All-groups Scatter Plot) và kết quả phân loại (Classification Results).

Bước 3. Phân tích các kết quả phân loại nhóm. Ở bước này xây dựng hai kiểu hàm là hàm lập nhóm hợp quy chuẩn hóa và hàm lập nhóm hợp quy chưa chuẩn hóa. Nếu có hai nhóm, thì chỉ có một hàm hợp quy. Nếu có k nhóm, thì có k – 1 hàm hợp quy. Hàm lập nhóm hợp quy chuẩn hóa được sử dụng để đánh giá tầm quan trọng hay ý nghĩa của từng biến trong hàm lập nhóm. Mức ý nghĩa của các biến dự đoán được đánh giá theo độ lớn tuyệt đối của các hệ số trong hàm lập nhóm. Hàm lập nhóm hợp quy chưa chuẩn hóa được sử dụng để xác định điểm số hay khoảng cách của mỗi trường hợp đến trung tâm nhóm. Điểm trung bình của mỗi nhóm được báo cáo thông qua các hàm ở trung tâm nhóm (Functions at Group Centroids). Khi có nhiều hàm hợp quy ($k > 2$), thì hai hàm hợp quy 1 và 2 được sử dụng để xây dựng các bản đồ và đồ thị phân chia các nhóm.

Sự khác nhau giữa các nhóm được đánh giá bằng thống kê Eigenvalues (Số đặc trưng) và hệ số tương quan. Thống kê Eigenvalues là tỷ số của tổng bình phương giữa các nhóm với tổng bình phương trong các nhóm. Tỷ lệ đóng góp của một hàm hợp quy theo một phương nào đó được đo bằng phần trăm phương sai. Phần trăm phương sai cộng dồn của các hàm hợp quy cho biết tỷ lệ đóng góp của chúng theo một phương nào đó. Lưu ý mỗi biến dự đoán là một phương.

Mối quan hệ giữa các điểm số lập nhóm được đo bằng hệ số tương quan hợp quy. Nếu chỉ có hai nhóm phân loại, thì hệ số tương quan hợp quy là hệ số tương quan Pearson. Khi có nhiều hơn hai nhóm phân loại, thì hệ số tương quan hợp quy là hệ số Eta hay tỷ tương quan. Đó là tỷ số của tổng bình phương giữa các nhóm với tổng bình phương toàn bộ. Nếu sử dụng phương pháp đưa tất cả các biến dự đoán vào hàm phân loại, thì sự khác nhau giữa các nhóm được xác định bằng tiêu chuẩn Wilk's Lambada và tiêu chuẩn χ^2. Khi $P < 0,05$ hoặc 0,01, thì trọng tâm giữa các nhóm là khác nhau. Nếu sử dụng phương pháp chọn biến từng bước (Stepwise

Method), thì những biến được chọn hay loại bỏ khỏi mô hình cũng được xác định bằng tiêu chuẩn Wilk's Lambda.

Khi sử dụng phương pháp chọn biến từng bước, thì sự khác nhau giữa các cặp nhóm được kiểm định bằng tiêu chuẩn F. Tiêu chuẩn F cho từng cặp nhóm tỷ lệ với khoảng cách Mahalanobis bình phương (D^2). Vì thế, độ lớn và mức ý nghĩa (P – value) của tiêu chuẩn F cho biết sự khác nhau giữa các nhóm. Thống kê F chỉ thích hợp với một cặp nhóm so sánh. Khi có nhiều cặp nhóm cần so sánh, vấn đề đa kiểm định sẽ phát sinh. Trong trường hợp này, sau khi mức ý nghĩa của các cặp so sánh được báo cáo, thì chúng phải được điều chỉnh bằng cách nhân mức ý nghĩa của mỗi cặp so sánh với số cặp được so sánh. Chẳng hạn: Khi so sánh 5 nhóm, thì số cặp so sánh là 10. Nếu mức ý nghĩa (P) của một cặp so sánh nào đó nhận giá trị là 0,01, thì xác suất được điều chỉnh là 10*0,01 = 0,10. Vì P = 0,10 > 0,05 nên cặp nhóm này khác nhau không có ý nghĩa thống kê.

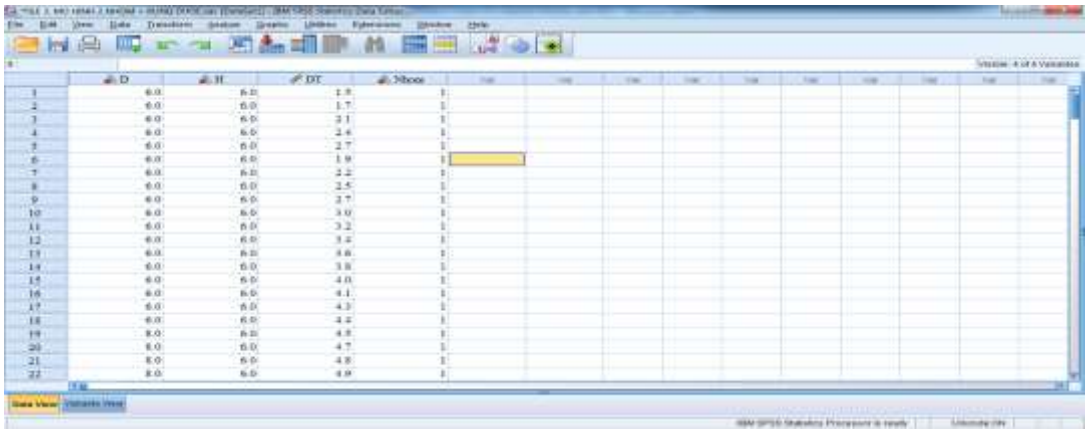
8.2. Phân tích mô hình hai nhóm

Giả sử mục tiêu nghiên cứu là xây dựng hàm lập nhóm để so sánh sự khác nhau giữa hai quần thụ Đước 16 tuổi. Hàm lập nhóm bao gồm 3 biến phân loại (D, H và D_T) của những cá thể đại diện cho hai quần thụ này. Trong trường hợp này, hàm lập nhóm có dạng như Hàm 8.6; trong đó k = 2 nhóm.

$$Z^{(k)} = a*D^{(k)} + b*H^{(k)} + c*D_T^{(k)} \qquad (8.6)$$

Để đạt mục tiêu này, hàm lập nhóm được xây dựng thông qua những bước dưới dây. Ở đây sử dụng công cụ tính toán là phần mềm thống kê SPSS.

Bước 1. Nhập các giá trị của 3 biến dự đoán (D, H, D_T) trong 3 cột liên tiếp của bảng tính SPSS. Cột thứ 4 nhập số nhóm phân loại. Trong trường hợp này, số nhóm phân loại là 2 quần thụ Đước 16 tuổi (Hình 8.1).

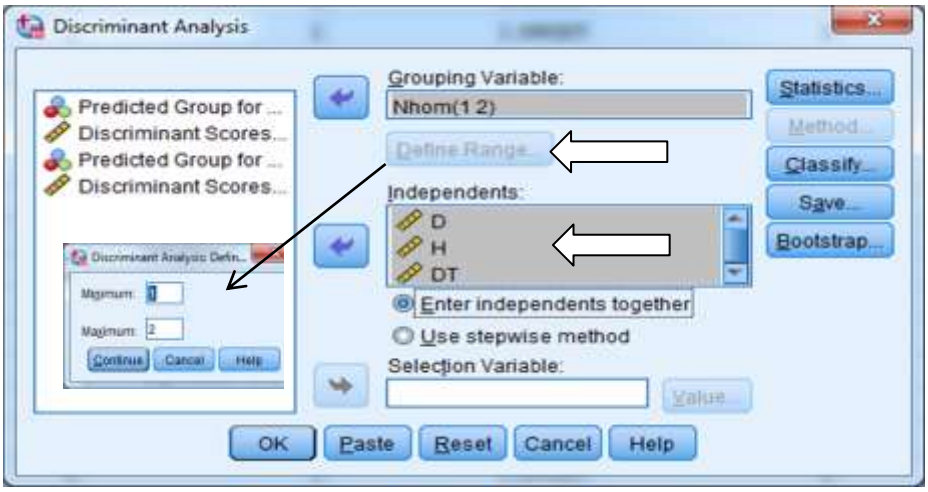


Hình 8.1. Đặc trưng của hai quần thụ 1 và 2.

Bước 2. Khai báo số nhóm và số lượng biến dự đoán của hàm lập nhóm tuyến tính Fisher theo Hàm 8.6. Trước hết mở tập số liệu của hai quần thụ Đước 16 tuổi. Kế đến, từ Menu chính, chọn Analyze ⇒ Classify ⇒ Discriminant (Hình 8.2).

Hình 8.2. Đường dẫn vào Menu phân tích hàm lập nhóm.

Tại cửa sổ Discriminant Analysis, chọn và chuyển số nhóm phân loại vào ô Grouping Variable. Kế đến chọn Define Range để định nghĩa phạm vi của các nhóm phân loại. Lưu ý số nhóm phân loại phải là số nguyên dương liên tục. Ở trường hợp này chỉ có hai quần thụ Đước 16 tuổi đưa vào phân tích. Để định nghĩa phạm vi của các nhóm, sau khi mở Define Range, nhập số 1 (Quần thụ 1) vào ô Minimum và số 2 (Quần thụ 2) vào ô Maximum. Tiếp theo chọn 3 biến dự đoán (D, H, D_T) và chuyển chúng vào ô Independents (Hình 8.3).

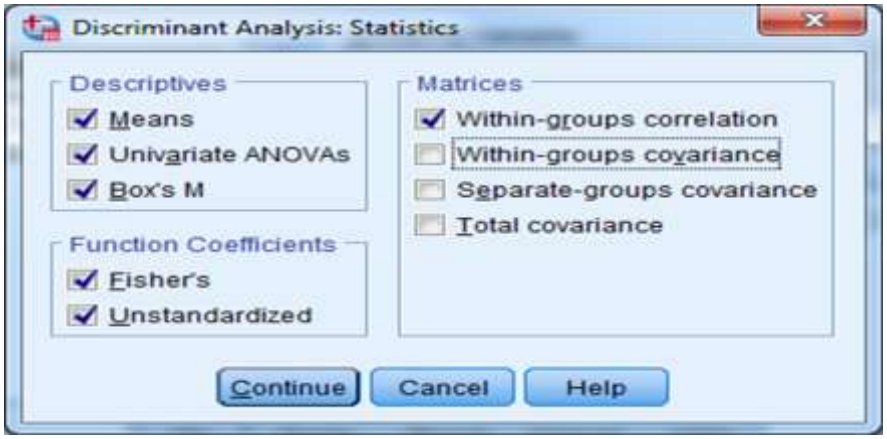


Hình 8.3. Phân tích hàm lập nhóm tuyến tính Fisher.

Bước 3. Xây dựng hàm lập nhóm tuyến tính Fisher theo Hàm 8.6. Từ cửa sổ Discriminant Analysis ở Hình 8.3, chọn Statistics và đánh dấu: (a) Các thống kê mô tả (Means; Univariable ANOVAs; Box's M = Kiểm định hiệp phương sai trong nội bộ nhóm); (b) Các hệ số của hàm phân loại (Hợp quy chuẩn hóa = Standardized Canonical Discriminant Function Coefficients; Hợp quy chưa chuẩn hóa =

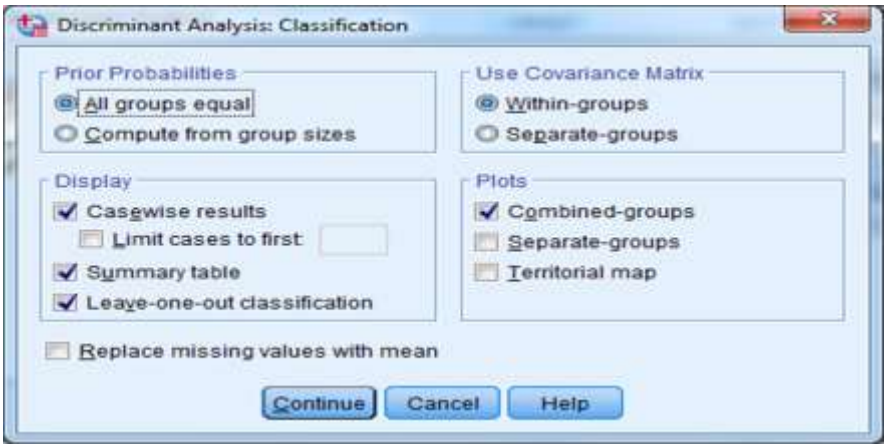
Unstandards Canonical Discriminant Function Coefficients và ma trận tương quan trong nhóm = Within-Group Correlation Matrices). Sau đó chọn Continue (Hình 8.4).

Bước 4. Xác định kết quả phân tích nhóm. Từ cửa sổ Discriminant Analysis ở Hình 8.3, chọn Classify và đánh dấu All Group Equal trong ô Prior Probabilities, Casewise Results và Summary Table trong ô Display, Within Groups trong ô Use Covariance Matrix (Hình 8.5). Sau cùng chọn OK để nhận kết quả phân tích nhóm.



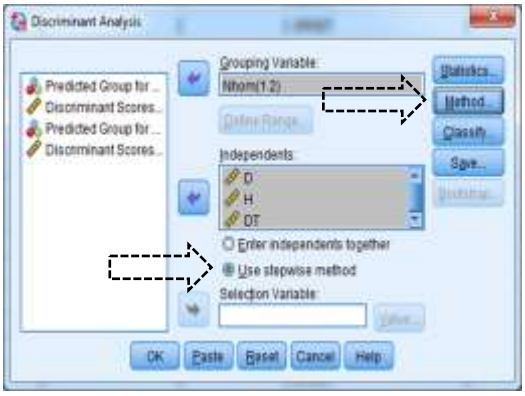
Hình 8.4. Những thống kê trong phân tích hàm lập nhóm.

Mục All Group Equal trong ô Prior Probabilities là xác suất ưu tiên cho mỗi nhóm. Nếu có k nhóm, thì xác suất ưu tiên của mỗi nhóm là (100/k)%. Ở ví dụ này, bởi vì chỉ có hai nhóm, nên xác suất ưu tiên của mỗi nhóm là (100/2) = 50%. Mục Casewise Results là kết quả phân tích nhóm; trong đó bao gồm số trường hợp phân tích, nhóm phân loại ban đầu (Actual Group), nhóm dự đoán (Predicted Group) và điểm số phân loại (Discriminant Scores) cho từng trường hợp.

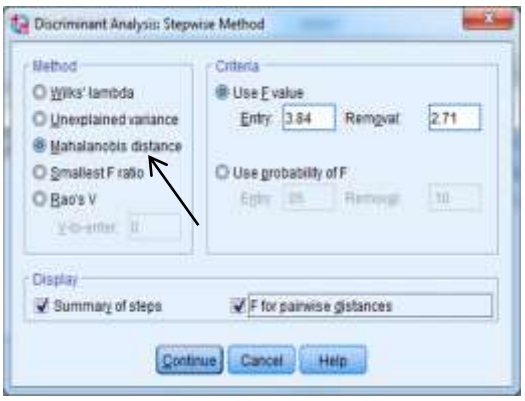


Hình 8.5. Xác định những thông tin của hàm lập nhóm.

Bốn bước trên đây được thực hiện khi hàm lập nhóm bao gồm tất cả 3 biến dự đoán (D, H và D_T). Nếu chỉ xây dựng hàm lập nhóm với những biến dự đoán hữu ích nhất, thì từ Hình 8.3 chọn phương pháp phân tích nhóm từng bước (Use Stepwise Method) theo phương pháp khoảng cách bình phương của Mahalanobis (Hình 8.6 và 8.7).



Hình 8.6. Phương pháp chọn biến từng bước.



Hình 8.7. Chọn phương pháp phân tích hàm lập nhóm.

Bước 5. Giải thích kết quả. Kết quả phân tích nhóm cung cấp nhiều nội dung khác nhau. Dưới đây chỉ trình bày tóm tắt những nội dung cơ bản.

Nội dung 1. Thống kê mô tả các nhóm. Phần này trước hết báo cáo các đặc trưng thống kê mô tả (trung bình, sai tiêu chuẩn, số mẫu của các biến đưa vào phân tích) đối với từng biến dự đoán trong mỗi nhóm (Bảng 8.1).

Bảng 8.1. Các thống kê mô tả trong hai nhóm (Group Statistics).

Biến dự đoán của 2 nhóm		Mean	Std. Deviation	Valid N (listwise)	
				Unweighted	Weighted
1	D	8,3	2,1	110	110
	H	8,0	2,0	110	110
	DT	5,0	1,7	110	110
2	D	14,2	2,2	102	102
	H	14,2	2,1	102	102
	DT	3,8	1,3	102	102
Total	D	11,2	3,6	212	212
	H	11,0	3,7	212	212
	DT	4,5	1,6	212	212

Kế đến kiểm định ngang bằng trung bình của các biến dự đoán giữa các nhóm (Bảng 8.2). Giả thuyết H_0: Trung bình của các biến dự đoán là không khác nhau giữa các nhóm. Giả thuyết này được kiểm định bằng tiêu chuẩn Wilks' Lambda và tiêu chuẩn F. Tiêu chuẩn Wilks' Lambda nhận giá trị từ 0 ÷ 1. Đối với mỗi biến phân loại, nếu giá trị Wilks' Lambda rất nhỏ hoặc P < 0,01, thì giả thuyết bị bác bỏ. Kết quả ở Bảng 8.2 cho thấy giá trị Sig hay P < 0,01 chứng tỏ ba biến (D,

H và D_T) khác nhau rất rõ rệt giữa hai quần thụ. Lưu ý rằng một biến số độc lập chỉ có ý nghĩa phân loại khi nó khác nhau thực sự giữa các nhóm.

Bảng 8.2. Kiểm định ngang bằng trung bình của các biến dự đoán giữa hai nhóm.

Biến dự đoán	Wilks' Lambda	F	df1	df2	Sig.
D	0,340	408,109	1	210	0,000
H	0,303	482,636	1	210	0,000
DT	0,860	34,221	1	210	0,000

Nội dung 2. Phân tích ý nghĩa của các biến dự tuyển hay biến ban đầu được đưa vào hàm phân loại theo phương pháp lập nhóm từng bước (Stepwise Method). Nội dung này phân tích và xác định những biến phân loại có ý nghĩa thực sự hay hữu ích để xây dựng hàm lập nhóm. Những biến phân loại có ý nghĩa được báo cáo ở mục phân tích những biến ở trong hàm lập nhóm (Variables Entered) và những biến bị loại ra khỏi hàm lập nhóm (Variables Removed). Ở ví dụ này, những biến được đưa vào hay loại ra khỏi hàm lập nhóm được kiểm định bằng phương pháp khoảng cách bình phương của Mahalanobis. Nếu sử dụng J biến dự tuyển, thì số bước kiểm định là J. Ở ví dụ này, hàm lập nhóm được xây dựng từ 3 biến dự tuyển (D, H, D_T), nên ý nghĩa của 3 biến này được kiểm định theo 3 bước. Giả thuyết H_0: Biến độc lập thứ nhất (Biến đưa vào phân tích đầu tiên) không có ý nghĩa phân loại. Nếu $P < 0,01$, thì giả thuyết bị bác bỏ. Những bước tiếp theo kiểm định ý nghĩa của các biến khác. Sau đó chương trình báo cáo kết quả kiểm định những biến dự tuyển ở trong hàm lập nhóm và những biến dự tuyển bị loại ra khỏi hàm lập nhóm. Ý nghĩa của các biến dự tuyển cũng được kiểm định bằng tiêu chuẩn Wilks' Lambda. Kết quả cho thấy ba biến (D, H và D_T) đều có ý nghĩa phân loại ($P < 0,01$) (Bảng 8.3).

Bảng 8.3. Kiểm định ý nghĩa của các biến số dự tuyển.

Step	Number of Variables	Lambda	df1	df2	df3	Exact F			
						Statistic	df1	df2	Sig.
1	1	0,303	1	1	210	482,6	1	210	0,000
2	2	0,086	2	1	210	1103,8	2	209	0,000
3	3	0,084	3	1	210	753,0	3	208	0,000

Nội dung 3. Phân tích hàm lập nhóm hợp quy (Canominal Discriminant Function Analysis). Phần này cung cấp các thông tin về kiểm định sự tồn tại của các hàm hợp quy; xác định các hệ số của hàm lập nhóm hợp quy chuẩn hóa và chưa chuẩn hóa; đo đạc mức ảnh hưởng hay vai trò của từng biến trong hàm lập nhóm. Những kiểm định này được xác định theo số đặc trưng (Eigenvalue) và tiêu chuẩn Wilks' Lambda (Bảng 8.4).

Thống kê Eigenvalue đo biến động về điểm số giữa các nhóm và quan hệ giữa các điểm số lập nhóm (Discriminant Scores). Nếu có hai nhóm, thì tương quan hợp quy (Canonical Correlation) là hệ số tương quan Pearson (R) giữa các điểm số lập nhóm của hai nhóm được mã hóa là 0 và 1. Nếu có ba nhóm trở lên, thì tương quan hợp quy là thống kê Eta (Tỷ tương quan η) trong phân tích phương sai 1 yếu

tố. Ở ví dụ này, tương quan hợp quy giữa các điểm số lập nhóm của hai nhóm là rất chặt chẽ (R = 0,957).

Bảng 8.4. Tóm tắt hàm lập nhóm hợp quy.

(a) Thống kê Eigenvalue.

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	10,862[a]	100,0	100,0	0,957

(b) Tiêu chuẩn Wilks' Lambda.

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	0,084	515,690	3	0,000

Tiêu chuẩn Wilks' Lambda là tỷ lệ của tổng biến thiên trong các điểm số lập nhóm không thể giải thích được bởi sự khác nhau giữa các nhóm. Ở ví dụ này, Wilks' Lambda = 0,084 cho biết khoảng 8,4% sự biến thiên của điểm số lập nhóm không thể giải thích được bằng sự khác nhau giữa các nhóm. Hai tiêu chuẩn Wilks' Lambda và Chi square đều kiểm định giả thuyết H_0: Điểm số trung bình hay trọng tâm của các nhóm là bằng nhau. Nếu P value < 0,01, thì điểm số trung bình của các nhóm là khác nhau. Ở ví dụ này, P < 0,01 cho biết điểm số trung bình của hai nhóm khác nhau rõ rệt.

Các hệ số của hàm lập nhóm hợp quy chuẩn hóa (Cột 1 - Bảng 8.5) được sử dụng để đánh giá vai trò hay mức độ ảnh hưởng của từng biến phân loại trong hàm lập nhóm. Vai trò của từng biến phân loại được đánh giá theo độ lớn tuyệt đối của các hệ số của hàm lập nhóm hợp quy chuẩn hóa. Ở ví dụ này, thứ bậc ảnh hưởng từ cao đến thấp là D_T, H và D. Vai trò của từng biến phân loại trong hàm lập nhóm hợp quy chuẩn hóa còn được đánh giá theo hệ số tương quan (Bảng 8.6). Theo hệ số tương quan, biến H có quan hệ với các biến của hàm lập nhóm hợp quy chuẩn hóa lớn hơn so với biến D và D_T.

Bảng 8.5. Các hệ số của hàm lập nhóm hợp quy chuẩn hóa và chưa chuẩn hóa.

Biến dự đoán	Standardized Function	Unstandardized Function
	(1)	(2)
D	0,526	0,247
H	1,267	0,617
D_T	-1,591	-1,053
(Constant)		-4,847

Hàm lập nhóm hợp quy chưa chuẩn hóa (Cột 2 - Bảng 8.5) được sử dụng để xác định điểm số của biến hợp quy (Canonical Variable Scores) cho từng trường hợp. Điểm số này không phụ thuộc vào dấu. Khi có hai nhóm, thì những điểm số âm thuộc về nhóm 1 hay hàm 1, còn điểm số dương thuộc về nhóm 2 hay hàm 2. Điểm số bình quân của mỗi nhóm được xác định bằng những hàm lập nhóm hợp quy chưa chuẩn hóa ở trung tâm nhóm (Functions at Group Centroids) (Bảng 8.7). Giá trị tuyệt đối của hàm lập nhóm hợp quy chưa chuẩn hóa ở trung tâm nhóm cho

biết mức độ khác nhau giữa các nhóm. Điểm số bình quân của mỗi nhóm được xác định theo hàm ở trung tâm nhóm (Bảng 8.8).

Bảng 8.6. Các hệ số tương quan giữa các biến dự đoán với các biến của hàm lập nhóm hợp quy chuẩn hóa.

Biến dự đoán	Function
	1
H	0,460
D	0,423
D_T	-0,122

Bảng 8.7. Hàm lập nhóm hợp quy chưa chuẩn hóa ở trung tâm nhóm.

Cấp sinh trưởng	Function	
	1	2
1	1,155	0,622
2	0,891	0,114
3	0,025	-0,173
4	-0,602	0,022
5	-1,771	0,505

Bảng 8.8. Hàm ở trung tâm nhóm (Functions at Group Centroids)

Nhóm	Functions at Group Centroids of 1
1	-3,159
2	3,406

Nội dung 4. Phân tích các hàm phân loại. Phần này phân tích xác suất ưu tiên của các nhóm và các hệ số của hàm lập nhóm. Ở ví dụ này chỉ có 2 quần thụ, nên xác suất ưu tiên của mỗi nhóm là 50%. Mỗi nhóm có một hàm phân loại tuyến tính Fisher. Ở ví dụ này có hai nhóm nên chỉ có hai hàm phân loại nhóm (Bảng 8.9). Hàm phân loại đối với quần thụ Đước (1) và (2) tương ứng có dạng như hàm 8.7 và 8.8.

Bảng 8.9. Các hệ số của hàm phân loại (Classification Function Coefficients)

Biến dự đoán	Các hệ số của nhóm	
	1	2
D	0,978	2,602
H	0,876	4,929
DT	0,106	-6,810
(Constant)	-8,540	-41,172

$$Z^{(1)} = 0,978*D + 0,876*H + 0,106*D_T = 8,540 \qquad (8.7)$$

$$Z^{(2)} = 2,602*D + 4,929*H - 6,810*D_T - 41,172 \qquad (8.8)$$

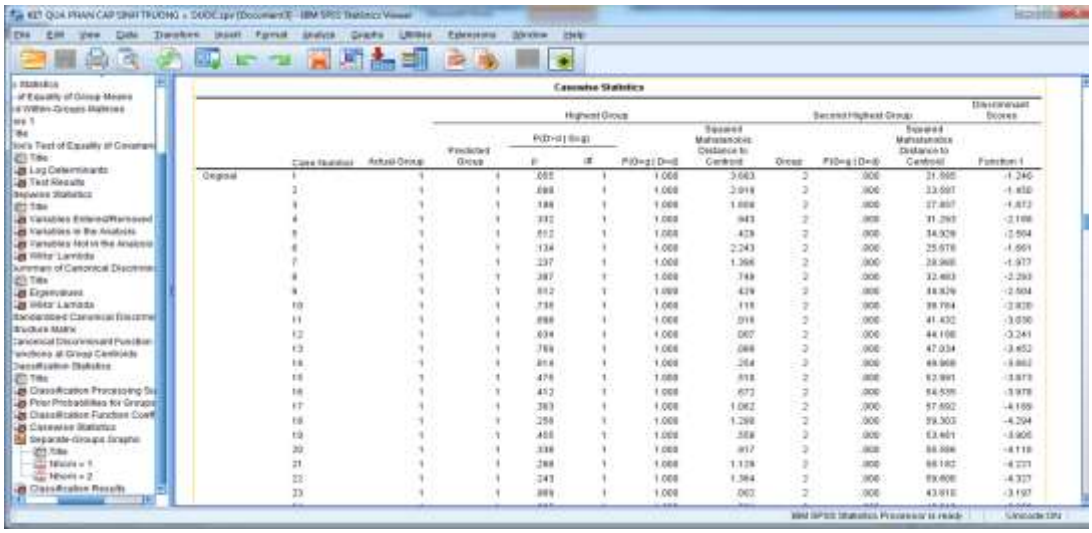
Hai hàm phân loại (Hàm 8.7 và 8.8) được sử dụng để tính khoảng cách khác nhau cực đại giữa hai quần thụ Đước. Hai hàm này cũng được sử dụng để phân loại những cây mới thuộc về quần thụ Đước khác nhau. Để biết một cây mới thuộc về quần thụ nào, trước hết đo đạc ba biến D, H và D_T của những cây cần phân loại

trong mỗi quần thụ được nghiên cứu. Kế đến nhân các hệ số của mỗi hàm với các giá trị của các biến dự đoán (D, H và DT). Kết quả nhận được khoảng cách khác nhau cực đại của mỗi quần thụ. Cây phân loại thuộc về quần thụ nhận giá trị khoảng cách lớn nhất. Bảng 8.10 dẫn ví dụ xác định nhóm đối với 4 cây Đước được tính toán từ hai hàm 8.7 và 8.8. Cây số 1 và 2 nhận giá trị khoảng cách lớn nhất ở hàm 1 (13), nên nó thuộc quần thụ 1. Cây số 3 và 4 nhận giá trị khoảng cách lớn nhất ở hàm 2 (tương ứng 27,4 và 48), nên nó thuộc quần thụ 2.

Bảng 8.10. Xác định nhóm đối với 4 cây mới theo hai hàm lập nhóm ở Bảng 8.9.

TT cây	Kích thước thân cây			Điểm số theo hàm:		Nhóm
	D (cm)	H (m)	D_T (m)	1	2	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	12	12	6,9	13,0	2,2	1
2	12	12	7	13,0	1,5	1
3	12	12	3,2	13,4	27,4	2
4	16	16	4,6	20,6	48,0	2

Nội dung 5. Phân tích những thống kê đối với các trường hợp (Casewise Statistics). Phần này so sánh các trường hợp được phân loại thực tế và được dự đoán. Bảng báo cáo này bao gồm 2 phần (Hình 8.8). Phần 1 cho biết các thông tin về kết quả phân loại lần đầu từ hàm lập nhóm (Highest Group). Phần 2 cung cấp các thống kê kiểm định lại các trường hợp phân loại lần đầu ở phần 1 có thể thuộc về nhóm hai hay không.



Hình 8.8. Xác định những thống kê của các trường hợp.

Phần 1 bao gồm 7 cột. Cột 1 là thứ tự các trường hợp được sử dụng để xây dựng hàm lập nhóm. Cột 2 và 3 tương ứng là nhóm thực tế và nhóm dự đoán của từng trường hợp. Cột 4 là xác suất có điều kiện (P(D>d | G=g)). Nó cho biết xác suất của điểm số quan sát thuộc nhóm đúng nhất. Cột 5 là độ tự do. Cột 6 là xác suất (P(G=g | D=d)) chấp nhận từng trường hợp thuộc về nhóm đã được phân loại. Cột 7 là khoảng cách Mahalanobis bình phương của mỗi trường hợp đến trung tâm nhóm. Khoảng cách Mahalanobis càng nhỏ cho biết trường hợp ấy càng gần trung tâm nhóm.

Phần 2 bao gồm 3 cột. Cột 8 – 10 là các thống kê kiểm định lại các trường hợp phân loại lần đầu ở phần 1 có thể thuộc về nhóm hai hay không (Second Highest Group). Cột 8 ghi lại nhóm của các trường hợp phân loại lần 2. Cột 9 cho biết xác suất (P(G=g | D=d)) mà một trường hợp cụ thể thuộc về nhóm hai. Nếu P < 0,01, thì trường hợp đó không thuộc về nhóm hai. Cột 10 là khoảng cách Mahalanobis bình phương của mỗi trường hợp đến trung tâm nhóm. Giá trị Mahalanobis bình càng nhỏ cho biết trường hợp ấy càng gần trung tâm nhóm. Cột 11 là các điểm số phân loại (Discriminant Scores). Điểm số phân loại các trường hợp được tính từ hàm lập nhóm hợp quy chưa chuẩn hóa (Unstandardized canonical discriminant functions; Hàm 2 của Bảng 8.5). Những trường hợp có điểm số phân loại âm cho biết chúng thuộc về nhóm 1. Ngược lại, những trường hợp có điểm số phân loại dương cho biết chúng thuộc về nhóm 2. Những trường hợp phân nhóm nhầm được ghi chú bằng dấu hoa thị (**) ở nhóm dự đoán (Predicted Group).

Nội dung 6. Kết quả phân loại (Classification Results). Phần này cung cấp những thông tin về kết quả phân loại các trường hợp vào từng nhóm (Bảng 8.11).

Bảng 8.11. Kết quả phân loại (Classification Results).

		Nhóm	Predicted Group Membership		Total
			1	2	
Original	Count	1	110		110
		2		102	102
	%	1	100,0		100
		2		100,0	100
Cross-validated[b]	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 tỷ lệ phần trăm. Bảng 8.11 bao gồm kết quả phân loại ban đầu (Original Classification Results) và kết quả kiểm định độ chính xác của phân loại ban đầu (Cross-validated Grouped Cases). Ở ví dụ này có 212 cây ban đầu được đưa vào phân loại; trong đó nhóm 1 là 110 cây, còn nhóm 2 là 102 cây. Kết quả dự đoán cho biết tất cả 110 cây được phân loại ban đầu vào nhóm 1 là chính xác

(100%). Tương tự, tất cả 102 cây được phân loại ban đầu vào nhóm 2 là chính xác (100%). Bảng này cũng báo cáo số cây và tỷ lệ phần trăm số cây đã phân loại nhầm vào mỗi nhóm. Kết quả còn báo cáo tỷ lệ phần trăm trung bình đối với những cây được phân loại đúng. Ở ví dụ này tất cả 212 cây (100%) đã được phân loại đúng vào hai nhóm.

8.3. Phân tích mô hình nhiều nhóm

8.3.1. Khái quát chung về mô hình nhiều nhóm

Những mô hình lớn hơn hai nhóm là những mô hình nhiều nhóm. Các bước xây dựng mô hình nhiều nhóm cũng tương tự như mô hình hai nhóm. Mô hình nhiều nhóm có thể được xây dựng theo một số phương pháp khác nhau. Một mô hình có thể bao gồm nhiều biến dự đoán hoặc chỉ bao gồm một vài biến dự đoán. Tổ hợp các biến dự đoán khác nhau cho kết quả phân loại khác nhau.

8.3.2. Ứng dụng mô hình nhiều nhóm trong phân cấp sinh trưởng cây rừng

Giả sử mục tiêu nghiên cứu là xác định hàm để phân chia 5 cấp sinh trưởng từ I – V đối với những cây hình thành rừng Đước (*Rhizophora apiculata*) trồng 16 tuổi. Các biến dự tuyển được sử dụng để phân chia 5 cấp sinh trưởng là D, H và D_T. Đặc trưng thống kê của rừng Đước trồng 16 tuổi được ghi lại ở Bảng 8.12.

Trong trường hợp này, hàm lập nhóm có dạng như Hàm 8.9; trong đó k = 5 cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi.

$$Z^{(k)} = a*D^{(k)} + b*H^{(k)} + c*D_T^{(k)} \qquad (8.9)$$

Bảng 8.12. Đặc trưng thống kê D, H và D_T của rừng Đước trồng 16 tuổi.

Chỉ tiêu	N (cây/ha)	Trung bình	Min	Max	±S	CV%
(1)	(2)	(3)	(4)	(5)	(6)	(7)
D (cm)	970	16,0	4,5	32,0	4,2	26,4
H (m)	970	14,5	3,0	30,5	4,2	28,8
D_T (m)	970	7,9	3,0	12,7	1,8	22,2
g (m^2/cây)	970	0,0215	0,0016	0,0804	0,0111	51,7
V (m^3/cây)	970	0,1782	0,0024	1,2265	0,1460	81,9

Trước khi xây dựng hàm phân loại 5 cấp sinh trưởng, chúng ta cần phải phân chia sơ bộ các cây của rừng Đước 16 tuổi vào 5 cấp sinh trưởng khác nhau. Đây là điều kiện để xây dựng hàm phân loại nhóm. Để thỏa mãn điều kiện này, chọn 1 biến phân loại nào đó và phân chia biến này thành 5 nhóm. Giả sử chọn biến D. Trước hết xác định các thống kê mô tả D. Kế đến xác định các hệ số D (KD) bằng cách chia D của từng cây (D_i) cho D bình quân (D_{Bq}) của rừng Đước 16 tuổi, nghĩa là $KD_i = D_i/D_{Bq}$. Tiếp theo phân chia chỉ số KD thành 5 cấp khác nhau. Ở ví dụ này, những cây thuộc cấp sinh trưởng I – V được phân chia sơ bộ theo 5 nhóm hệ số KD tương ứng như sau: $KD_I > 1,3$, $KD_{II} = 1,1 – 1,3$; $KD_{III} = 0,9 – 1,1$; $KD_{IV} = 0,7 - 0,9$; $KD_V < 0,7$. Sau đó mã hóa các hệ số này bằng các số nguyên liên tục từ 1 - 5.

Sau đó thực hiện các bước xây dựng và kiểm định 5 hàm phân cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi. Các bước xử lý ở đây cũng được thực hiện tương tự như mô hình 2 nhóm ở Mục 8.2. Kết quả xây dựng hàm phân loại 5 cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi được ghi lại tóm tắt ở Bảng 8.13 – 8.21. Bảng 8.13 cho biết kết quả kiểm định ngang bằng D, H và D_T giữa 5 cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi bằng tiêu chuẩn Wilks' Lambda. Khi P < 0,01, thì D, H và D_T khác nhau rõ rệt giữa 5 cấp sinh trưởng. Ở ví dụ này, giá trị P < 0,01, nên D, H và D_T khác nhau rõ rệt giữa 5 cấp sinh trưởng.

Bảng 8.13. Kiểm định ngang bằng D, H và D_T giữa 5 cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi.

Biến dự đoán	Wilks' Lambda	F	df1	df2	Sig.
(1)	(2)	(3)	(4)	(5)	(6)
D	0,083	2666,6	4	965	0,000
H	0,082	2700,8	4	965	0,000
D_T	0,922	20,281	4	965	0,000

Bảng 8.14 và 8.15 báo cáo kết quả kiểm định các biến ở trong hàm phân loại và những biến bị loại ra khỏi hàm phân loại bằng tiêu chuẩn độ chất nhận (Tolerance) và tiêu chuẩn F_{Ra} (F to Remoe). Khi một biến được sử dụng để xây dựng hàm phân loại, thì độ chấp nhận phải cao, cao nhất là 1,0. Ở ví dụ này, ba biến D, H và D_T đều nhận Tolerance rất cao, nên chúng được chấp nhận để xây dựng hàm phân loại cấp sinh trưởng. Theo tiêu chuẩn F_{Ra} (F to Remoe), một biến được sử dụng để xây dựng hàm phân loại khi F_{Ra} > 2,71 (Mặc định). Ở ví dụ này, vì F_{Ra} của ba biến D, H và D_T đều lớn hơn 2,71, nên chúng được chấp nhận để xây dựng hàm phân loại cấp sinh trưởng.

Bảng 8.14. Kiểm định các biến ở trong mô hình.

Step		Tolerance	F to Remove	Min. D Squared	Between Groups
(1)		(2)	(3)	(4)	(5)
1	H	1,000	2700,834		
2	H	0,975	2624,548	0,010	1 and 2
	D_T	0,975	13,728	3,815	2 and 3
3	H	0,127	15,423	3,882	2 and 3
	D_T	0,975	13,318	3,857	2 and 3
	D	0,128	12,090	4,133	2 and 3

Bảng 8.16 là kết quả kiểm định ý nghĩa của các hàm phân loại hợp quy. Các hàm này đo sự khác nhau giữa các nhóm. Giả định H_0: Các nhóm là khác nhau. Sự khác nhau giữa các nhóm được đo bằng tỷ lệ (%) phương sai cộng dồn. Tỷ lệ này càng gần 100% thì các nhóm càng khác nhau. Ở ví dụ này, phần trăm phương sai cộng dồn (Cumulative %) ở các hàm đều gần bằng 100%, nên các nhóm là khác nhau.

Bảng 8.15. Kiểm định các biến không ở trong mô hình.

Step		Tolerance	Min. Tolerance	F to Enter	Min. D Squared	Between Groups
(1)		(2)	(3)	(4)	(5)	(6)
0	D	1,000	1,000	2666,588	3,601	2 and 3
	H	1,000	1,000	2700,834	3,815	2 and 3
	D_T	1,000	1,000	20,281	,010	1 and 2
1	D	0,128	0,128	12,496	3,857	2 and 3
	D_T	0,975	0,975	13,728	4,133	2 and 3
2	D	0,128	0,127	12,090	4,173	2 and 3

Bảng 8.16. Kiểm định ý nghĩa của các hàm phân loại hợp quy.

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
(1)	(2)	(3)	(4)	(5)
1	11,618[a]	99,4	99,4	0,960
2	0,058[a]	0,5	99,9	0,235
3	0,014[a]	0,1	100,0	0,116

Bảng 8.17. Các hệ số của các hàm phân loại hợp quy chuẩn hóa.

Biến dự đoán	Function		
	1	2	3
(1)	(2)	(3)	(4)
D	0,457	-1,379	2,394
H	0,570	1,321	-2,408
D_T	-0,099	0,862	0,522

Bảng 8.17 là các hệ số của các hàm phân loại hợp quy chuẩn hóa. Chúng được sử dụng để xác định ý nghĩa của các biến dự đoán trong hàm phân loại. Ý nghĩa của các biến dự đoán được xác định theo độ lớn tuyệt đối của các hệ số của các hàm phân loại hợp quy chuẩn hóa. Ở ví dụ này, biến D và H có ý nghĩa lớn trong hàm phân loại thứ 3, còn biến D_T có ý nghĩa lớn trong hàm phân loại thứ 2.

Bảng 8.18 là các hệ số của các hàm phân loại hợp quy chưa chuẩn hóa. Chúng được sử dụng để tính điểm số của các trường hợp đưa vào phân loại. Điểm số trung bình của các hàm hợp quy ở trung tâm 5 nhóm được báo cáo ở Bảng 8.19.

Bảng 8.18. Các hệ số của các hàm phân loại hợp quy chưa chuẩn hóa.

Biến dự đoán	Function		
	1	2	3
(1)	(2)	(3)	(4)
D	0,376	-1,133	1,966
H	0,476	1,103	-2,011
D_T	-0,059	,510	0,309
(Constant)	-12,447	-1,906	-4,742

Bảng 8.20 là các hàm phân loại 5 cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi. Các hàm này được sử dụng để phân loại các cây mới vào 5 cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi.

Bảng 8.19. Điểm số trung bình của các hàm hợp quy ở trung tâm 5 cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi.

Cấp sinh trưởng	Function		
	1	2	3
(1)	(2)	(3)	(4)
1	6,207	-0,360	0,119
2	1,994	-0,004	-0,238
3	-0,012	0,248	0,054
4	-2,615	-0,072	0,044
5	-6,113	-0,342	-0,009

Bảng 8.20. Các hàm phân loại 5 cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi.

Biến dự đoán	Các hàm phân cấp sinh trưởng				
	1	2	3	4	5
(1)	(2)	(3)	(4)	(5)	(6)
D	14,816	12,128	11,663	11,027	9,916
H	0,808	-0,086	-1,351	-2,922	-4,781
D_T	1,193	1,512	1,849	1,836	1,888
(Constant)	-188,590	-117,833	-92,745	-63,078	-34,091

Bảng 8.21. Kết quả phân loại 5 cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi.

	Cấp sinh trưởng	Predicted Group Membership					Tổng
		1	2	3	4	5	
Count	1	127					127
	2		176				176
	3		30	369			399
	4				144		144
	5					124	124
%	1	100.0					100.0
	2		100.0				100.0
	3		7.5	92.5			100.0
	4				100.0		100.0
	5					100.0	100.0

Từ kết quả phân loại ở Bảng 8.21 cho thấy, tổng số 970 cây (100%) được sử dụng để xây dựng hàm phân loại 5 cấp cấp sinh trưởng; trong đó những cây thuộc cấp sinh trưởng 1, 2, 4 và 5 được phân loại chính xác 100%. Trái lại, tổng số cây cấp sinh trưởng 3 là 399 cây; trong đó số cây được phân loại chính xác là 369 (92,5%), còn lại 30 cây (7,5%) bị phân loại nhầm sang cấp sinh trưởng 2. Trung bình 96,9% (205 cây) số cây hình thành rừng Đước 16 tuổi được phân loại chính

xác vào 5 cấp sinh trưởng. Điều đó chứng tỏ các hàm ở Bảng 5.20 được sử dụng để phân loại 5 cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi. Bảng 8.22 dẫn ví dụ ứng dụng 5 hàm phân loại để xác định 5 cây mới thuộc về những cấp sinh trưởng khác nhau.

Bảng 8.22. Phân cấp sinh trưởng đối với 5 cây mới theo 5 hàm lập nhóm.

Kích thước thân cây			Điểm số theo hàm:					Cấp sinh trưởng
D (cm)	H (m)	D_T (m)	1	2	3	4	5	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
21,0	19,5	7,2	147	146	139	125	95	1
20,5	19,0	6,3	138	139	132	119	90	2
15,0	13,5	6,1	52	72	75	74	62	3
12,0	10,5	4,6	3	34	42	47	43	4
4,5	3,0	5,3	-113	-55	-35	-12	6	5

Bảng 8.23. Đặc trưng của rừng Đước 16 tuổi theo 5 cấp sinh trưởng thực tế. Đơn vị tính: 1 ha.

Cấp sinh trưởng	N (cây)	D (cm)	H (m)	G (m^2)	M (m^3)	N (%)	G (%)	M (%)	Bquân (%)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
I	127	23,4	21,7	5,5	61,1	13,1	26,5	35,4	25,0
II	176	18,3	16,9	4,6	39,2	18,1	22,2	22,7	21,0
III	399	16,0	14,5	8,0	58,7	41,1	38,5	33,9	37,9
IV	144	12,9	11,4	1,9	10,9	14,8	9,1	6,3	10,1
V	124	8,7	7,2	0,8	3,0	12,8	3,7	1,7	6,1
Tổng số	970	15,9	14,4	20,9	172,8	100,0	100,0	100,0	100,0

Bảng 8.24. Đặc trưng của rừng Đước 16 tuổi theo 5 cấp sinh trưởng được phân loại theo những hàm lập nhóm. Đơn vị tính: 1 ha.

Cấp sinh trưởng	N (cây)	D (cm)	H (m)	G (m^2)	M (m^3)	N (%)	G (%)	M (%)	Bquân (%)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
I	127	23,4	21,7	5,5	61,1	13,1	26,5	35,4	25,0
II	206	18,2	16,7	5,3	44,8	21,2	25,6	25,9	24,3
III	369	15,9	14,4	7,3	53,0	38,0	35,1	30,7	34,6
IV	144	12,9	11,4	1,9	10,9	14,8	9,1	6,3	10,1
V	124	8,7	7,2	0,8	3,0	12,8	3,7	1,7	6,1
Tổng số	970	15,8	14,3	20,9	172,8	100,0	100,0	100,0	100,0

Đặc trưng của rừng Đước 16 tuổi theo 5 cấp sinh trưởng thực tế được ghi lại tương ứng ở Bảng 8.23. Bảng 8.24 dẫn đặc trưng của rừng Đước 16 tuổi theo 5 cấp

sinh trưởng được phân loại theo những hàm ở Bảng 8.20. So sánh số liệu ở Bảng 8.23 và 8.24 cho thấy kết quả phân cấp sinh trưởng thực tế và dự đoán đối với những cây hình thành rừng Đước 16 tuổi là tương tự như nhau. Sự khác biệt chỉ xảy ra ở cấp sinh trưởng III; trong đó tỷ lệ sai khác bình quân về số cây là 7,5%. Điều đó chứng tỏ các hàm ở Bảng 5.20 được sử dụng để phân loại 5 cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi.

Chương 9
CHỈ SỐ MẬT ĐỘ QUẦN THỤ VÀ CHỈ SỐ CẠNH TRANH

9.1. Chỉ số mật độ quần thụ

9.1.1. Một số chỉ số mật độ quần thụ

Theo McArdle (1930; Dẫn theo Curtis, 1982), khi biết mật độ chuẩn (N_{Nor}, cây/ha) của quần thụ ở những A khác nhau trên lập địa nhất định (SI), thì chỉ số mật độ (N%) là tỷ lệ giữa mật độ quần thụ (N, cây/ha) và N_{Nor} (Công thức 9.1a). Khi biết tiết diện ngang chuẩn của quần thụ (G_{Nor}, m^2/ha) ở những A khác nhau trên lập địa nhất định (SI), thì chỉ số mật độ (N%) là tỷ lệ giữa G của quần thụ và G_{Nor} (Công thức 9.1b).

$$N\% = (N/N_{Nor})*100 \qquad (9.1a)$$

$$N\% = (G/G_{Nor})*100 \qquad (9.1b)$$

Theo Réineke (1933), khi rừng khép tán hoàn toàn, thì mối quan hệ giữa mật độ ở dạng Log(N) và đường kính trung bình quân phương (D_{QF}, cm) ở dạng Log(D_{QF}) có dạng hàm tuyến tính với độ dốc bằng -1,605 (Công thức 9.2). Đồ thị mô tả quan hệ giữa Log(N) với Log(D_{QF}) được gọi là đường mật độ lớn nhất.

$$Log(N) = -1{,}605*Log(D_{QF}) + k \qquad (9.2)$$

Chỉ số mật độ quần thụ (SDI) là thước đo mật độ tương đối để chuyển đổi mật độ quần thụ hiện tại thành mật độ quần thụ ở kích thước cơ sở. Quần thụ cơ sở là quần thụ có D_{QF} = 25 cm. Mật độ của quần thụ cơ sở được gọi là chỉ số SDI. Chỉ số SDI ở mỗi tuổi (A, năm) của rừng trồng thuần loài đồng tuổi là mật độ trung bình được xác định từ mối quan hệ giữa số cây và kích thước cây bình quân. Tại quần thụ cơ sở, Log(N) = Log(SDI). Theo định nghĩa này, chỉ số SDI được xác định theo hàm 9.3.

$$Log(SDI) = -1{,}605*Log(D_{QF}) + k \qquad (9.3)$$

Từ hàm 9.3, k = Log(SDI) + 1,605. Thay thế giá trị k vào hàm 9.2, ta có:

$$Log(N) = Log(SDI) + 1{,}605 - 1.605Log(D_{QF}) \qquad (9.4)$$

Chỉ số SDI của quần thụ nhất định được xác định theo công thức 9.5.

$$SDI = N*(25/D_{QF})\text{^}-1{,}605 = N*(D_{QF}/25)\text{^}1{,}605 \qquad (9.5)$$

Chỉ số SDI của một quần thụ nhất định có thể so sánh với chỉ số SDI_{Max}. Chỉ số SDI_{Max} là chỉ số mật độ của quần thụ mà các cây gỗ không giao tán với nhau hay tán phát triển tự do. Trong quần thụ này, tăng trưởng D, G và V của cây gỗ và quần thụ là lớn nhất. Căn cứ vào tỷ lệ (%) giữa chỉ số SDI của quần thụ nghiên cứu với chỉ số SDI_{Max}, mức độ cạnh tranh giữa những cây gỗ trong một quần thụ nghiên cứu được đánh giá theo 4 mức (Bảng 9.1).

Bảng 9.1. Đánh giá mức độ cạnh tranh giữa các cây gỗ trong quần thụ.

TT	(SDI/SDI_{Max})%	Cấp mật độ	Mức độ cạnh tranh
(1)	(2)	(3)	(4)
1	0 - 24	Thấp	Không cạnh tranh giữa các cây. Tăng trưởng mạnh nhất.
2	25 - 34	Trung bình	Cạnh tranh giữa các cây ở mức trung bình. Tăng trưởng trung bình.
2	35 - 55	Cao	Cây gỗ che phủ hoàn toàn mặt đất. Cạnh tranh mạnh giữa các cây gỗ. Tăng trưởng cao về thể tích thân.
3	> 56	Rất cao	Cây gỗ che phủ hoàn toàn mặt đất. Cạnh tranh rất mạnh giữa các cây gỗ. Tăng trưởng giảm mạnh. Nhiều cây bị đào thải.

Flewelling (1979; Dẫn theo Curtis, 1982) cho rằng thể tích thân cây bình quân (V, m^3/ha) được xác định gần đúng theo hàm 9.6a; trong đó N (cây/ha) là mật độ của quần thụ, a là tham số. Mối quan hệ giữa Ln(V) và Ln(N) là một đường thẳng với độ dốc bằng -3/2 (Hàm 9.6b).

$$V = a*N^{-3/2} \quad (9.6a)$$

$$Ln(V) = Ln(a) - 3/2*Ln(N) \quad (9.6b)$$

Khi biết N_{Max} trên lập địa nhất định, thì $Ln(V_{Max})$ và chỉ số RDI của quần thụ được xác định tương ứng theo hàm 9.6c và 9.6d; trong đó N_{TT} = mật độ thực tế.

$$Ln(V_{Max}) = Ln(a) - 3/2*Ln(N_{Max}) \quad (9.6c)$$

$$RDI = (N_{TT}/N_{Max}) \quad (9.6d)$$

Curtis (1982) xác định chỉ số mật độ (RD) ở mỗi tuổi (A, năm) của rừng trồng thuần loài đồng tuổi theo công thức 9.7; trong đó N, G và D_{QF} là mật độ, tiết diện ngang và đường kính trung bình quân phương.

$$RD = \frac{G}{\sqrt{D_{QF}}} \quad (9.7a)$$

$$RD = \frac{(0,00007854*N*D_{QF}\text{^}2)}{\sqrt{D_{QF}}} \quad (9.7b)$$

Khi quần thụ đạt đến tuổi A và D_{QF}, thì số cây tỉa thưa (N_{TT}, cây/ha) được xác định theo công thức 9.8. Số cây còn lại sau tỉa thưa (N_{STT}, cây/ha) được xác định theo công thức 9.9; trong đó N_{HC} là mật độ quần thụ hiện còn tại tuổi A, N_{TT} là số cây tỉa thưa tại tuổi A.

$$N_{TT} = \frac{RD}{(0,00007854*D_{QF}\text{^}3/2)} \quad (9.8)$$

$$N_{STT} = N_{HC} - N_{TT} \quad (9.9)$$

Theo Curtis (1982), để rừng che phủ kín lập địa và phát triển bình thường, mật độ sau tỉa thưa (N_{STT}) được để lại khoảng 50% so với chỉ số RD. Giới hạn dưới của tỉa thưa bằng 35% so với chỉ số RD. Giới hạn dưới được sử dụng để đảm bảo

mật độ quần thụ không tiếp tục bị suy giảm và lập địa không bị bỏ trống. Từ đó, số cây để lại nuôi dưỡng sau tỉa thưa (N_{STT}, cây/ha) được xác định theo công thức 9.10.

$$N_{(STT50\%)} = \frac{(0,5*RD)}{(0,00007854*D_{QF}\text{^}3/2)} \qquad (9.10a)$$

$$N_{(STT35\%)} = \frac{(0,35*RD)}{(0,00007854*D_{QF}\text{^}3/2)} \qquad (9.10b)$$

Các số đo chỉ số mật độ quần thụ của McArdle (N% và G%), Flewelling (RDI) và Curtis (RD) có mối quan hệ với nhau (Bảng 9.2).

Schröder và Gadow (1999) đề xuất chỉ số không gian tương đối (RS) của quần thụ. Đó là tỷ lệ giữa khoảng cách trung bình của các cây với chiều cao của những cây ưu thế và đồng ưu thế (H_D). Chỉ số RS của quần thụ ở tuổi A được xác định theo công thức 9.11. Ở công thức 9.11, N_A là mật độ quần thụ ở tuổi A; H_{DA} = chiều cao của những cây ưu thế và đồng ưu thế ở tuổi A.

$$RS = \frac{\sqrt{10000/N_A}}{H_{DA}} \qquad (9.11)$$

Bảng 9.2. Mối quan hệ giữa các số đo chỉ số mật độ quần thụ.

Mật độ	N% và G%	RDI	RD
(1)	(2)	(3)	(4)
Lớn nhất	150	1,0	100
Chuẩn	100	0,67	70
Mật độ tỉa thưa trên	75 - 80	0,5	50
Mật độ tỉa thưa dưới	50 - 60	0,3	35
Khép tán	30	0,15	20

N% và G% = McArdle (1930); RDI = Flewelling (1979); RD = Curtis (1982).

Chỉ số RS được sử dụng để xác định mật độ quần thụ suy giảm ổn định theo A. Giá trị RS biến đổi theo A dưới dạng đường cong hình chữ J ngược. Vì thế, hàm RS = f(A) được ước lượng theo hàm hàm 9.12; trong đó m là giới hạn trên của RS, b là hằng số, c là tham số hình dạng. Chỉ số RS là căn cứ để xác định biến động không gian sinh trưởng của quần thụ theo những cấp A khác nhau.

$$RS = m*\exp(b*A\text{^}-c) \qquad (9.12)$$

Tỉa thưa rừng được xác định theo biên độ giữa ranh giới trên của chỉ số RS (RS_U) và ranh giới dưới của chỉ số RS (RS_L). Chỉ số RS_U được thiết lập để đảm bảo cho quần thụ sinh trưởng bình thường và diện tích tán che phủ kín lập địa. Chỉ số RS_L được thiết lập để xác định mật độ quần thụ ở mức cao nhất.

Nếu giữ mật độ quần thụ sau tỉa thưa ở chỉ số RS_{UA}, thì số cây còn lại sau tỉa thưa tại tuổi A (N_{UA}) được xác định theo công thức 9.13; trong đó H_{DA} là chiều cao của những cây ưu thế và đồng ưu thế tại tuổi A. Trong trường hợp này, tuổi tỉa thưa (A_{TU}, năm) được xác định theo công thức 9.14; trong đó tham số m, b và c tương ứng là tham số tiệm cận, độ dốc và tham số hình dạng ở hàm 9.12.

$$N_{UA} = \frac{10.000}{(RS_{UA}*H_{DA})^2} \quad (9.13)$$

$$A_{TU} = (b/Ln(RS_U/m))\text{^}(1/c) \quad (9.14)$$

Nếu giữ mật độ quần thụ sau tỉa thưa ở chỉ số RS_L, thì số cây còn lại sau tỉa thưa tại tuổi A (N_{LA}, năm) được xác định theo công thức 9.15. Trong trường hợp này, tuổi tỉa thưa (A_{TL}, năm) được xác định theo công thức 9.16; trong đó tham số m, b và c tương ứng là tham số tiệm cận, độ dốc và tham số hình dạng ở hàm 9.12.

$$N_{LA} = \frac{10.000}{(RS_{LA}*H_{DA})^2} \quad (9.15)$$

$$A_{TL} = (b/Ln(RS_L/m))\text{^}(1/c) \quad (9.16)$$

Biên độ mật độ từ N_{UA} đến N_{LA} là số cây tỉa thưa tại tuổi A. Đây là điều kiện đảm bảo cho rừng sinh trưởng bình thường và lập địa được phủ kín.

9.1.2. Ứng dụng chỉ số mật độ trong tỉa thưa rừng

Để ứng dụng các chỉ số mật độ quần thụ trong nuôi dưỡng rừng, Bảng 9.3 dẫn ví dụ về đặc trưng của rừng Đước trồng từ 11 – 25 tuổi tại khu vực Biển Tây thuộc tỉnh Cà Mau. Ở Bảng 9.3, cột 1 là tuổi rừng (A, năm); cột 2 là mật độ hiện còn tại tuổi A (N_{HC}, cây/ha); cột 3 và 4 tương ứng là đường kính bình quân (D_{Bq}, cm) và đường kính bình quân quân phương tại tuổi A (D_{QF}, cm); cột 5 là chiều cao cây ưu thế và đồng ưu thế tại tuổi A (H_D, cm); cột 6 là tiết diện ngang tại tuổi A (G, m^2/ha); cột 7 là chỉ số mật độ theo phương pháp của Curtis (1982). Từ số liệu của Bảng 9.3, bằng phân tích hồi quy, xác định được các hàm ước lượng N = f(A), D_{QF} = f(A), H_D = f(A) và G = f(A) đối với rừng Đước từ 11 – 25 tuổi (Hàm 9.17 – 9.20) (Bảng 9.4). Các hàm này được sử dụng để ước lượng và dự đoán N, D_{QF}, H_D và G của rừng Đước.

Bảng 9.3. Một số đặc trưng của rừng Đước trồng từ 11 – 25 tuổi tại khu vực Biển Tây thuộc tỉnh Cà Mau.

A (năm)	H_D (m)	N (cây/ha)	D_{Bq} (cm)	D_{QF} (cm)	G (m^2/ha)	RD
(1)	(2)	(3)	(4)	(5)	(6)	(7)
11	9,4	4300	7,9	8,1	21,5	7,6
13	11,0	3800	9,5	9,7	29,6	9,4
15	12,7	3517	11,9	12,2	38,7	11,2
17	14,2	3158	13,5	13,7	48,5	13,1
19	15,7	3058	15,1	15,3	58,6	14,8
21	17,0	2958	17,3	17,5	68,8	16,5
23	18,2	2775	19,2	19,3	78,7	18,0
25	19,3	2575	20,5	20,7	88,1	19,3

Bảng 9.4a. Xây dựng hàm ước lượng những đặc trưng của rừng Đước trồng từ 11 – 25 tuổi.

Nhân tố	Hàm ước lượng:	
(1)	(2)	
Mật độ	N = 7840,48*exp(-0,125984*A) + 2313,66	(9.17)
H_D	H_D = 31,0458*exp(-2,9956*exp(-0,0949086*A))	(9.18)
D_{DF}	D_{QF} = 33,6686*exp(-3,33995*exp(-0,0773241*A))	(9.19)
G	G = 161,675*exp(-5,19275*exp(-0,0859168*A))	(9.20)

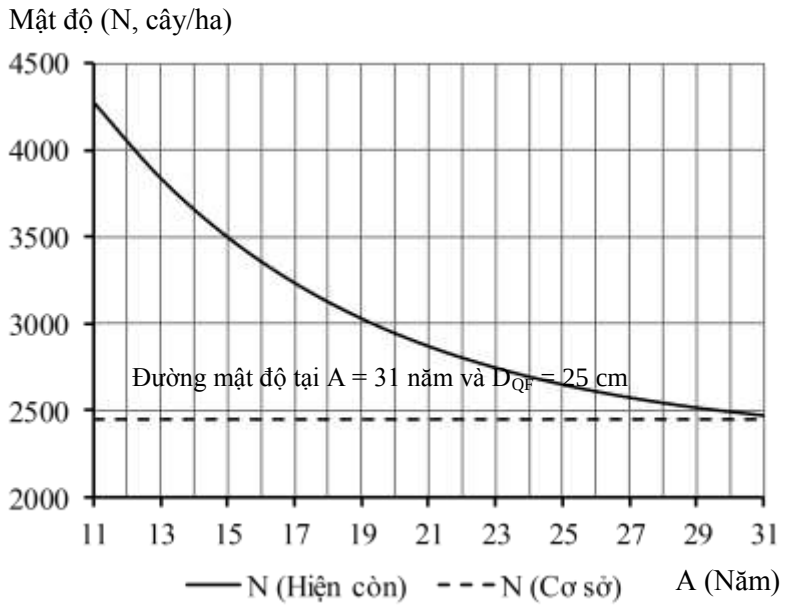
Bảng 9.4b. Tương quan và sai lệch của những hàm ước lượng những đặc trưng của rừng Đước trồng từ 11 – 25 tuổi.

Nhân tố	r^2(%)	±S	MAE	MAPE	SSE	
(1)	(2)	(3)	(4)	(5)	(6)	
Mật độ	98,9	68,2	47,3	1,5	23267	(9.17)
H_D	99,2	0,46	0,31	1,8	1,1	(9.18)
D_{DF}	99,7	0,25	0,15	1,2	0,32	(9.19)
G	99,2	2,5	1,88	3,7	30,7	(9.20)

(a) Xác định chỉ số mật độ của rừng Đước theo phương pháp của Réineke (1933). Những tính toán từ hàm 9.17 cho thấy D_{QF} = 25 cm xuất hiện tại cấp tuổi 31. Tại cấp tuổi 31 và D_{QF} = 25 cm, theo công thức 9.5, xác định được chỉ số SDI đối với rừng Đước từ 11 – 31 tuổi (Bảng 9.5; Hình 9.1). Giới hạn từ đường mật độ hiện tại và đường mật độ cơ sở là số cây tỉa thưa tiềm năng.

Bảng 9.5. Chỉ số mật độ của rừng Đước được xác định theo phương pháp của Réineke (1933).

A (năm)	N (cây)	Dg (cm)	SDI (cây)
(1)	(2)	(3)	(4)
11	4275	8,1	700
13	3838	9,9	868
15	3498	11,8	1048
17	3235	13,7	1232
19	3029	15,6	1421
21	2870	17,4	1604
23	2746	19,2	1798
25	2650	20,8	1973
27	2575	22,3	2143
29	2517	23,6	2295
31	2471	24,8	2440



Hình 9.1. Đồ thị biểu diễn mật độ hiện còn và mật độ tại tuổi cơ sở của rừng Đước được xác định theo phương pháp của Réineke (1933).

(b) Xác định chỉ số mật độ của rừng Đước trồng theo phương pháp của Curtis (1982). Trước hết, từ chỉ số RD ở Bảng 9.3, xây dựng hàm ước lượng RD = f(A) (Hàm 9.21). Sai lệch của hàm này rất nhỏ (MAPE < 0,6%) là vì chỉ số RD được ước lượng từ hàm D_{QF} = f(A) và hàm G = f(A).

RD = 104,665*exp(-9,69013*A^-0,543315) (9.21)

r^2 = 99,96%; S = ±0,09; MAE = 0,06; MAPE = 0,6%.

Kế đến thay thế A vào hàm 9.21 để xác định chỉ số RD của rừng Đước từ 11 – 25 tuổi (Bảng 9.6). Từ đó cho thấy, chỉ số RD của rừng Đước gia tăng dần từ tuổi 11 (RD = 7,5) đến tuổi 15 (RD = 11,3) và tuổi 25 (RD = 19,4).

Tiếp theo, từ công thức 9.8, xác định mật độ của rừng Đước để lại nuôi dưỡng ở giới hạn dưới tương ứng với 0,35*RD và giới hạn trên tương ứng với 0,5*RD (Bảng 9.7). Ở Bảng 9.7, cột 2 là mật độ hiện còn (N_{HC}, Cây/ha) tại tuổi 11 – 25; cột 4 và 5 tương ứng là N_{STT} và cường độ tỉa thưa (N%) ở giới hạn dưới tương ứng với 0,35*RD; cột 6 và 7 tương ứng là N_{STT} và N% tỉa thưa ở giới hạn trên tương ứng với 0,5*RD. Từ số liệu ở Bảng 9.7 cho thấy, nếu mong muốn rừng Đước từ tuổi 11 – 25 sinh trưởng bình thường và lập địa không bị bỏ trống, thì mật độ để lại sau tỉa thưa (N_{STT}) bằng 0,35*RD. Theo đó, mật độ hiện tại của rừng Đước từ tuổi 11 – 25 được điều chỉnh bằng phương pháp tỉa thưa với cường độ khoảng 65% số cây hiện còn ở mỗi tuổi. Nếu mong muốn rừng Đước từ tuổi 11 – 25 sinh trưởng tốt và che phủ kín lập địa, thì N_{STT} bằng 0,5*RD. Theo đó, mật độ hiện tại của rừng

Đước từ tuổi 11 – 25 được điều chỉnh bằng phương pháp tỉa thưa với cường độ khoảng 30% so với số cây hiện còn ở mỗi tuổi.

Bảng 9.6. Chỉ số mật độ của rừng Đước trồng từ 11 – 25 tuổi được xác định theo phương pháp của Curtis (1982).

A (năm)	$RD_{(Thực tế)}$	$RD_{(Ước lượng)}$	±S	Khoảng tin cậy 95%	
				$RD_{(Cận dưới)}$	$RD_{(Cận trên)}$
(1)	(2)	(3)	(4)	(5)	(6)
11	7,6	7,5	0,123	7,3	7,7
13	9,4	9,4	0,106	9,3	9,6
15	11,2	11,3	0,105	11,2	11,4
17	13,1	13,1	0,106	13,0	13,2
19	14,8	14,8	0,106	14,7	14,9
21	16,5	16,4	0,104	16,3	16,5
23	18,0	17,9	0,106	17,8	18,1
25	19,3	19,4	0,121	19,2	19,6

Bảng 9.7. Mật độ và cường độ tỉa thưa đối với rừng Đước từ 11 - 25 tuổi được xác định theo phương pháp của Curtis (1982).

A (năm)	N_{HC} (Cây/ha)	RD	N_{STT} theo $RD_{35\%}$		N_{STT} theo $RD_{50\%}$	
			cây/ha	N%	cây/ha	N%
(1)	(2)	(3)	(4)	(5)	(6)	(7)
11	4275	7,5	1458	65,9	2083	31,8
13	3838	9,4	1348	64,9	1926	29,7
15	3498	11,3	1241	64,5	1773	29,1
17	3235	13,1	1147	64,5	1639	29,1
19	3029	14,8	1068	64,7	1526	29,5
20	2945	15,6	1035	64,9	1478	29,7
23	2746	17,9	954	65,3	1362	30,5
25	2650	19,4	913	65,5	1305	31,1

(c) Xác định chỉ số mật độ của rừng Đước trồng theo chỉ số không gian tương đối của Schröder và Gadow (1999). Trước hết, từ hàm N = f(A) (Hàm 9.17), hàm H_D = f(A) (Hàm 9.18) và công thức 9.11, xác định được chỉ số RS đối với rừng Đước từ 11 - 25 tuổi (Bảng 9.8). Hai giá trị RS_L và RS_U được xác định theo giới hạn trên và giới hạn dưới của hàm N = f(A) và H_D = f(A). Kế đến xây dựng hàm ước lượng RD = f(A) (Hàm 9.22). Hệ số r^2 = 99,81% và MAPE = 0,66%. Hệ số r^2 rất cao và MAPE rất nhỏ là vì chỉ số RS được ước lượng từ hàm N = f(A) và hàm H_D = f(A).

Bảng 9.8. Chỉ số không gian tương đối của rừng Đước trồng từ 11 - 25 tuổi.

A (năm)	H_D (m)	N (cây/ha)	RS_L	RS	RS_U
(1)	(2)	(3)	(4)	(5)	(6)
11	10,8	4275	0,138	0,141	0,143
13	13,0	3838	0,123	0,124	0,126
15	15,1	3498	0,111	0,112	0,114
17	17,1	3235	0,102	0,103	0,105
19	19,0	3029	0,096	0,096	0,098
21	19,8	2945	0,090	0,093	0,092
23	22,1	2746	0,085	0,086	0,088
25	23,5	2650	0,081	0,083	0,085

RS = 0,0233755*exp(4,99083*A^-0,42687) (9.22)

r^2 = 99,81%; S = ±0,001; MAE = 0,0006; MAPE = 0,66%.

Hàm RS = f(A) ở cận dưới (RS_L) và cận trên (RS_U) tương ứng có dạng như Hàm 9.23 và 9.24.

RS_L = 0,0111045*exp(5,04248*A^-0,289169) (9.23)

r^2 = 99,97%; S = ±0,0004; MAE = 0,0002; MAPE = 0,24%.

RS_U = 0,0237387*exp(4,98636*A^-0,426554) (9.24)

r^2 = 99,94%; S = ±0,0005; MAE = 0,0003; MAPE = 0,36%.

Tiếp theo, thay thế A vào hàm 9.22 - 9.24 để xác định chỉ số RS, RS_U và RS_L của rừng Đước từ 11 – 25 tuổi (Bảng 9.9).

Bảng 9.9. Ước lượng chỉ số không gian tương đối của rừng Đước từ 11 - 25 tuổi.

A (năm)	N (cây/ha)	H_D (m)	Chỉ số không gian tương đối		
			RS_L	RS	RS_U
(1)	(2)	(3)	(4)	(5)	(6)
11	4275	10,8	0,138	0,140	0,143
13	3838	13,0	0,123	0,124	0,126
15	3498	15,1	0,111	0,112	0,114
17	3235	17,1	0,102	0,104	0,105
19	3029	19,0	0,096	0,097	0,098
21	2945	19,8	0,090	0,091	0,093
23	2746	22,1	0,085	0,087	0,088
25	2650	23,5	0,081	0,083	0,084

Sau đó, theo công thức 9.11 – 9.14, xác định mật độ của rừng Đước ở giới hạn dưới (N_L, cây/ha) và giới hạn trên (N_U, cây/ha), số cây tỉa thưa (N_{TT}, cây/ha) và cường độ tỉa thưa (N%) theo giới hạn N_L và giới hạn N_U (Bảng 9.10). Từ đó cho thấy, nếu mong muốn rừng Đước từ tuổi 11 – 25 sinh trưởng bình thường và lập địa không bị bỏ trống, thì mật độ của rừng Đước từ tuổi 11 – 25 cần được điều chỉnh bằng phương pháp tỉa thưa với cường độ khoảng 6% so với mật độ hiện tại.

Bảng 9.10. Mật độ của rừng Đước ở giới hạn dưới (N_L, cây/ha) và giới hạn trên (N_U, cây/ha), số cây tỉa thưa (N_{TT}, cây/ha) và cường độ tỉa thưa theo N_L và N_U.

A (năm)	N_L (Cây/ha)	N_U (Cây/ha)	N_{TT} (cây/ha)	Cường độ tỉa thưa theo:	
				N_L%	N_U%
(1)	(2)	(3)	(4)	(5)	(6)
11	4484	4205	279	6,2	6,6
13	3949	3737	213	5,4	5,7
15	3549	3368	181	5,1	5,4
17	3258	3090	168	5,1	5,4
19	3051	2886	165	5,4	5,7
21	3154	2972	182	5,8	6,1
23	2815	2639	176	6,3	6,7
25	2760	2573	187	6,8	7,3

9.2. Xác định sự cạnh tranh giữa những cây gỗ trong quần thụ

9.2.1. Khái niệm về sự cạnh tranh

Khi các cây gỗ lâm vào tình trạng thiếu nguồn sống (nước, ánh sáng, chất khoáng), thì chúng cạnh tranh với nhau. Những dấu hiệu phản ánh sự cạnh tranh giữa các cây gỗ là sự che tán, sự đan xen hệ rễ và sự phát thải các độc chất. Mức độ che tán phụ thuộc vào mật độ quần thụ (N, cây/ha), đường kính tán (D_T, m) và diện tích tán (S_T, m^2). Mật độ quần thụ được sử dụng để xác định mức độ lợi dụng không gian sống và tính toán mật độ tối ưu đối với quần thụ trên những lập địa khác nhau. Diện tích tán cây và đường kính thân cây có quan hệ chặt chẽ với nhau. Vì thế, diện tích tán và tiết diện ngang thân của mỗi loài cây gỗ phản ánh mức độ cạnh tranh giữa nó với những loài cây gỗ khác trên một lập địa nhất định. Hai chỉ tiêu này cũng được sử dụng để đánh giá mức độ xâm chiếm lập địa của các loài cây gỗ và rừng ở những tuổi và lập địa khác nhau. Phân tích ảnh hưởng của mức độ cạnh tranh đến tăng trưởng của các loài cây gỗ là cơ sở khoa học để xây dựng các những phương thức nuôi dưỡng và khai thác rừng.

9.2.2. Chỉ số cạnh tranh giữa các cây gỗ hình thành rừng

Các nhà sinh thái học và lâm học thường biểu thị sự cạnh tranh giữa các cây gỗ trong quần thụ bằng chỉ số cạnh tranh (CI = Competition Index). Mức độ cạnh tranh giữa các cây gỗ phản ánh mức độ xâm chiếm lập địa của chúng. Mức độ xâm

chiếm lập địa thường được đánh giá theo mức độ che phủ mặt đất của gốc cây và diện tích hình chiếu theo chiều nằm ngang của tán cây ở các tuổi khác nhau.

Glover và Hool (1979; Dẫn theo Daniel, 1986) đã xác định chỉ số CI của mỗi loài cây gỗ theo tỷ lệ D_{ij} và D_{Bq} (Công thức 9.25); trong đó D_{ij} là đường kính của cây i thuộc loài j, còn D_{Bq} là đường kính bình quân của quần thụ.

$$CI_i = \frac{\sum_{(i=1,\,k)} D_{iJ}}{D_{Bq}} \qquad (9.25)$$

Daniel và ctv (1986) đã xác định chỉ số CI của mỗi loài cây gỗ theo tỷ lệ Y_{ij} và Y_{Bq} (Công thức 9.26); trong đó Y_{ij} = H, G và diện tích tán (S_T, m^2) của cây i thuộc loài J.

$$CI = \frac{Y_{iJ}}{Y_{Bq}} \qquad (9.26)$$

Schröder và Gadow (1999) đề xuất chỉ số CI giữa những cây gỗ trong quần thụ được tính theo công thức 9.27.

$$CI = \frac{(1 - [1 - (BAL/G)])}{RS} \qquad (9.27)$$

Ở công thức 9.27; BAL là tiết diện ngang của cây gần nhất lớn hơn cây cơ sở; G = tiết diện ngang của những cây trong ô mẫu với bán kính bằng 11 m; RS là chỉ số không gian tương đối. Cây cơ sở là cây được chọn để xác định chỉ số CI. Chỉ số RS là tỷ lệ giữa khoảng cách trung bình của các cây với chiều cao trung bình của quần thụ (H_{Bq}, m). Chỉ số RS của quần thụ được xác định theo công thức 9.28; trong đó N là mật độ quần thụ trên ô mẫu.

$$RS = \frac{\sqrt{10000/N}}{H_{Bq}} \qquad (9.28)$$

Sự che phủ của hệ thống tán cây gỗ trên mặt đất phản ánh khả năng xâm chiếm lập địa rõ ràng nhất. Theo đó, chỉ số cạnh tranh của các cây gỗ trong quần thụ có thể được đánh giá theo chỉ số cạnh tranh tán (CCI = Crown Competition Index). Đối với rừng thuần loài đồng tuổi, chỉ số CCI ở mỗi cấp A là tỷ lệ giữa S_T và diện tích mặt đất (Z, m^2). Chỉ số CCI của cây thứ i tại cấp tuổi A (CCI_{iA}) được xác định theo công thức 9.29; trong đó S_{TiA} là diện tích tán của cây thứ i tại cấp tuổi A, còn Z = 10.000 m^2 hoặc diện tích của ô mẫu với kích thước nhất định. Chỉ tiêu S_{TiA} được xác định theo diện tích hình tròn với đường kính bằng D_{TiA} (Công thức 9.30). Tổng diện tích tán của rừng ở mỗi cấp A (S_{TA}, m^2/ha) được xác định theo công thức 9.31; trong đó N_A và D_{TA} tương ứng là mật độ và đường kính tán bình quân của quần thụ tại cấp tuổi A.

$$CCI_{iA} = \frac{S_{TiA}}{Z} \qquad (9.29)$$

$$S_{TiA} = 0{,}7854 * D_{TiA}^{\,2} \qquad (9.30)$$

$S_{TA} = N_A*(0,7854*D_{TA}^2)$ (9.31a)

$S_{TA} = \sum S_{TiA} = \sum 0,7854*D_{TiA}^2$ (9.31b)

Đối với rừng hỗn loài khác tuổi (tự nhiên, nhân tạo), chỉ số CCI của cây thứ i thuộc loài J (CCI_{ij}) được xác định theo công thức 9.32; trong đó S_{Tii} là diện tích tán của cây thứ i thuộc loài J, còn Z = 10.000 m^2 hoặc diện tích của ô mẫu với kích thước nhất định. Chỉ tiêu S_{TiJ} được xác định theo diện tích hình tròn với đường kính bằng D_{TiJ} (Công thức 9.33). Tổng diện tích tán của loài J (S_{Tj}) được xác định theo công thức 9.34; trong đó n_i là mật độ của loài J, còn D_{Tij} là đường kính tán của cây thứ i thuộc loài J.

$$CCI_{iJ} = \frac{S_{Tij}}{Z} \quad (9.32)$$

$S_{Tij} = 0,7854*D_{Tij}^2$ (9.33)

$S_{Tj} = \sum_{(i=1,\ ni)}(0,7854*D_{Tij}^2)$ (9.34)

Chỉ số CCI có thể được tổng hợp theo cấp D và cấp H khác nhau. Mô hình ước lượng chỉ số CCI theo cấp D và cấp H được mô tả bằng hàm 9.35; trong đó Y = cấp D và cấp H, N_i = số cây trong ở mỗi cấp D và cấp H. Ba tham số a, b và c được xác định bằng phương pháp bình phương nhỏ nhất. Sau đó xây dựng bảng và đồ thị để mô tả và so sánh chỉ số CCI theo cấp D và cấp D.

$CCI_i = a + b*N_i + c*(N_i*Y) + d*(N_i*Y)^2$ (9.35)

9.2.3. Phân tích chỉ số cạnh tranh tán đối với rừng trồng thuần loài

Phần này giới thiệu phương pháp phân tích chỉ số CCI đối với rừng trồng thuần loài theo số liệu từ rừng Đước trồng ở tuổi 18 và 20 năm tại khu vực Biển Tây thuộc tỉnh Cà Mau. Ở đây có 2 câu hỏi đặt ra: (1) Chỉ số CCI thay đổi theo cấp A của rừng Đước như thế nào? (2) Trong một cấp A của rừng Đước, chỉ số CCI thay đổi theo cấp H như thế nào? Để trả lời hai câu hỏi này, trước hết thống kê N, D_T, S_T và chỉ số CCI theo cấp A và cấp H của rừng Đước (Bảng 9.11). Kế đến, để dễ dàng ước lượng và sự đoán chỉ số CCI theo cấp A và cấp H, từ số liệu ở Bảng 9.11, xây dựng các hàm chỉ số CCI = f(N, H) đối với rừng Đước trồng ở tuổi 18 và 20 (Hàm 9.36 và 9.37; Bảng 9.12). Sau đó thay thế N và cấp H vào 2 hàm 9.36 và 9.37 để xác định chỉ số CCI ở các cấp H khác nhau đối với rừng Đước ở tuổi 18 và 20 (Bảng 9.13 và 9.14; Hình 9.2 và 9.3).

Đối với rừng Đước 18 tuổi (Bảng 9.13; Hình 9.2), chỉ số CCI gia tăng dần từ cấp H < 8 m (CCI = 0,074 hay 3,8%) và đạt cao nhất ở cấp H = 12 m (CCI = 0,594 hay 30,1%); sau đó giảm đến cấp H ≥ 18 m (CCI = 0,187 hay 9,5%). Hệ thống tán lá của rừng Đước 18 tuổi che phủ 1,97 lần diện tích mặt đất. Tổng diện tích tán của những cây thuộc các cấp H ≤ 12 m che phủ hoàn toàn diện tích mặt đất. So với tổng chỉ số CCI (1,97 hay 100%), tỷ lệ cạnh tranh tán của những cây từ cấp H = 8 m đến cấp H = 12 m là 49,2% (CCI = 0,972), còn lại những cây ở cấp H ≥ 14 m là 50,8% (CCI = 1,0).

Bảng 9.11. Phân bố mật độ và diện tích tán theo các cấp H của rừng Đước ở tuổi 18 và 20 tại khu vực Biển Tây thuộc tỉnh Cà Mau.

(9.11a). Số liệu N, D_T, S_T và chỉ số CCI của rừng Đước ở tuổi 18.

Cấp H (m)	N (cây/ha)	D_T (m)	S_T (m^2/ha)	CCI
(1)	(2)	(3)	(4)	(5)
8	163	2,4	739	0,074
10	513	2,8	3053	0,305
12	833	3,0	5932	0,593
14	635	3,3	5313	0,531
16	300	3,5	2873	0,287
18	173	3,7	1847	0,185
Tổng	2617		19757	1.976

(9.11b). Số liệu N, D_T, S_T và chỉ số CCI của rừng Đước ở tuổi 20.

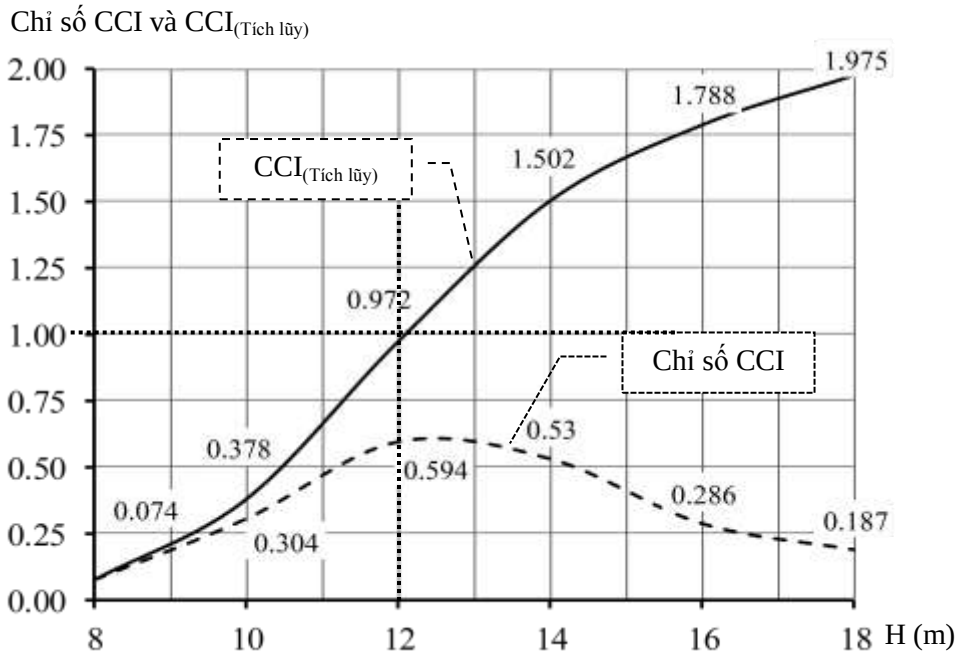
Cấp H (m)	N (cây/ha)	D_T (m)	S_T (m^2/ha)	CCI
(1)	(2)	(3)	(4)	(5)
10	233	2,6	1267	0,127
12	303	3,0	2196	0,220
14	614	3,2	5028	0,503
16	805	3,5	7592	0,759
18	449	3,6	4633	0,463
20	145	3,7	1558	0,156
Tổng	2550		22275	2.228

Bảng 9.12. Những hàm ước lượng chỉ số cạnh tranh tán theo cấp chiều cao đối với rừng Đước trồng ở tuổi 18 và 20. Đơn vị tính: 1,0 ha.

TT	Tham số	Tuổi rừng Đước:	
		18	20
(1)	(2)	(3)	(4)
1	a (Hằng số)	-0,0075161	-0,0023202
2	b*N	0,0001616	0,0004020
3	c*(NH)	0,0000337	0,0001176
4	d*(NH^2)	0,0000011	0,0000021
5	R^2 (%)	99,9	99,9
6	±S	0,0017	0,0078
7	MAE	0,0009	0,0037
8	MAPE (%)	0,3	1,5
	Hàm	(9.36)	(9.37)

Bảng 9.13. Chỉ số cạnh tranh tán theo cấp chiều cao đối với rừng Đước 18 tuổi. Đơn vị tính: 1,0 ha.

Cấp H (m)	CCI	CCI%	$CCI_{(Tích\ lũy)}$	$CCI\%_{(Tích\ lũy)}$
(1)	(2)	(3)	(4)	(5)
<8	0,074	3,8	0,074	3,8
10	0,304	15,4	0,378	19,1
12	0,594	30,1	0,972	49,2
14	0,530	26,8	1,502	76,0
16	0,286	14,5	1,788	90,5
>18	0,187	9,5	1,975	100,0
Tổng số	1,975	100,0		

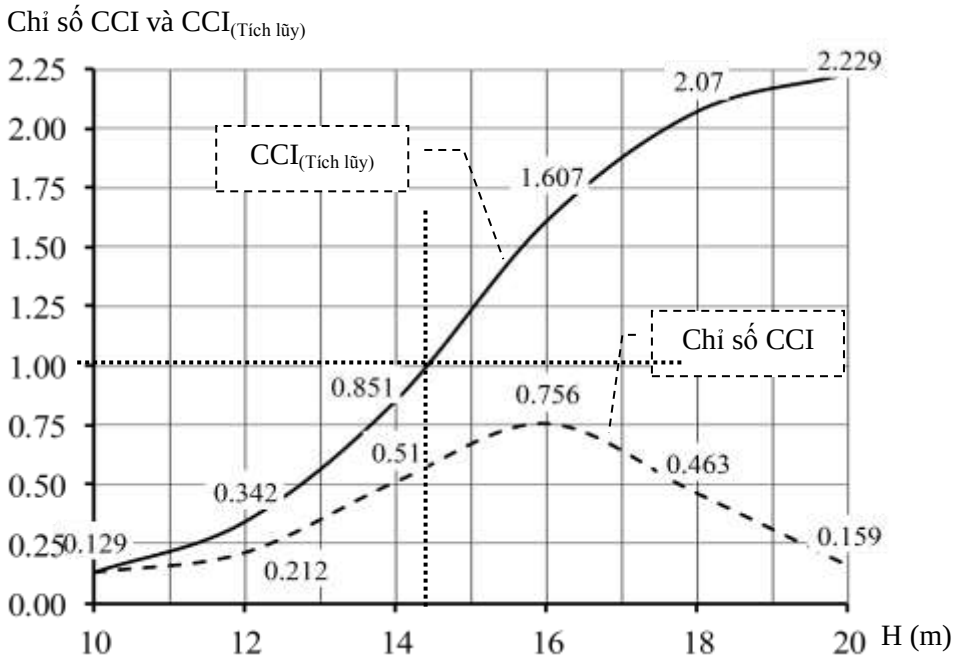


Hình 9.2. Đồ thị biểu diễn chỉ số cạnh tranh tán theo cấp chiều cao đối với rừng Đước 18 tuổi.

Đối với rừng Đước 20 tuổi (Bảng 9.14; Hình 9.3), chỉ số CCI gia tăng dần từ cấp H < 10 m (CCI = 0,129 hay 5,8%) và đạt cao nhất ở cấp H = 16 m (CCI = 0,756 hay 33,9%); sau đó giảm đến cấp H ≥ 20 m (CCI = 0,159 hay 7,1%). Hệ thống tán lá của rừng Đước 20 tuổi che phủ 2,2 lần diện tích đất rừng. Tổng diện tích tán của những cây thuộc các cấp H ≤ 14 m che phủ hoàn toàn diện tích mặt đất. So với tổng chỉ số CCI (2,2 hay 100%), tỷ lệ cạnh tranh tán từ cấp H = 10 m đến cấp H = 14 m là 38,2% (CCI = 0,851), còn lại những cây ở cấp H ≥ 16 m là 61,8% (CCI = 1,377).

Bảng 9.14. Chỉ số cạnh tranh tán theo cấp chiều cao đối với rừng Đước 20 tuổi. Đơn vị tính: 1,0 ha.

Cấp H (m)	CCI	CCI%	$CCI_{(Tích\ lũy)}$	$CCI\%_{(Tích\ lũy)}$
(1)	(2)	(3)	(4)	(5)
10	0,129	5,8	0,129	5,8
12	0,212	9,5	0,342	15,3
14	0,510	22,9	0,851	38,2
16	0,756	33,9	1,607	72,1
18	0,463	20,8	2,070	92,9
20	0,159	7,1	2,229	100,0
Tổng số	2,229	100,0		



Hình 9.3. Đồ thị biểu diễn chỉ số cạnh tranh tán theo cấp chiều cao đối với rừng Đước 20 tuổi.

Từ những phân tích trên đây cho thấy, rừng Đước trồng ở tuổi 18 và 20 có sự cạnh tranh tán rất mạnh. Hệ thống tán lá của rừng Đước trồng che phủ kín mặt đất từ 1,9 lần ở tuổi 18 đến 2,2 lần ở tuổi 20. Hệ thống tán lá của rừng Đước trồng ở tuổi 18 và 20 tập trung nhiều nhất tương ứng ở cấp H = 12 và 14 m. Hiện tượng này xảy ra là do rừng Đước được trồng với mật độ rất cao (10.000 cây/ha), rừng đang trong giai đoạn sinh trưởng nhanh và chưa qua tỉa thưa. Vì thế, để hạn chế sự cạnh tranh giữa những cây hình thành rừng Đước, biện pháp điều chỉnh tán bằng biện pháp tỉa thưa là cần thiết.

9.2.4. Phân tích chỉ số cạnh tranh tán đối với rừng tự nhiên hỗn loài

Phần này giới thiệu phương pháp phân tích cạnh tranh tán của rừng tự nhiên hỗn loài dựa trên số liệu từ hai ưu hợp cây họ Sao Dầu (UhSaoDau) ở khu vực Tân Phú thuộc tỉnh Đồng Nai. Chỉ số giá trị quan trọng (IVI) của cây họ Sao Dầu trong hai ưu hợp này tương ứng là nhỏ hơn 40% (UhSaoDau<40%) và lớn hơn 40% (UhSaoDau>40%). Ở đây có 3 câu hỏi đặt ra: (1) Chỉ số CCI thay đổi theo hai UhSaoDau như thế nào? (2) Trong mỗi UhSaoDau, chỉ số CCI thay đổi theo cấp H như thế nào? (3) Trong mỗi UhSaoDau, loài cây gỗ nào có mức cạnh tranh tán mạnh nhất? Để trả lời ba câu hỏi này, trước hết thống kê số loài cây gỗ (S, loài), N, D_T, S_T và chỉ số CCI theo loài cây gỗ và cấp H của hai UhSaoDau. Chỉ số CCI theo loài cây gỗ và cấp H của hai ưu này được dẫn ra ở Bảng 9.15 và 9.16. Kế đến, từ số liệu ở Bảng 9.15 và 9.16, xây dựng các hàm chỉ số CCI = f(N, H) (Hàm 9.38 và 9.39; Bảng 9.17). Tiếp theo thay thế cấp H vào hai hàm 9.38 và 9.39 để xác định chỉ số CCI theo các cấp H đối với hai UhSaoDau (Bảng 9.18 và 9.19; Hình 9.4 và 9.5).

Bảng 9.15. Mật độ và chỉ số cạnh tranh tán theo cấp H của UhSaoDau<40%.

H (m)	N (cây/ha)	S_T (m^2/ha)	CCI	$CCI_{(Tích\ lũy)}$
(1)	(2)	(3)	(4)	(5)
8	102	1855	0,185	0,185
12	268	4842	0,484	0,670
16	142	3901	0,390	1,060
20	75	3054	0,305	1,365
24	40	2237	0,224	1,589
28	20	1549	0,155	1,744
Tổng số	647	17438	1,744	

Bảng 9.16. Mật độ và chỉ số cạnh tranh tán theo cấp H của UhSaoDau>40%

H (m)	N (cây/ha)	S_T (m^2/ha)	CCI	$CCI_{(Tích\ lũy)}$
(1)	(2)	(3)	(4)	(5)
8	101	1.557	0,156	0,156
12	272	5.110	0,511	0,667
16	144	4.252	0,425	1,092
20	76	3.227	0,323	1,415
24	40	2.324	0,232	1,647
28	22	1.702	0,170	1,817
Tổng số	655	18.172	1,817	

Bảng 9.17. Những hàm ước lượng chỉ số cạnh tranh tán theo cấp chiều cao đối với hai ưu hợp cây họ Sao Dầu. Đơn vị tính: 1,0 ha.

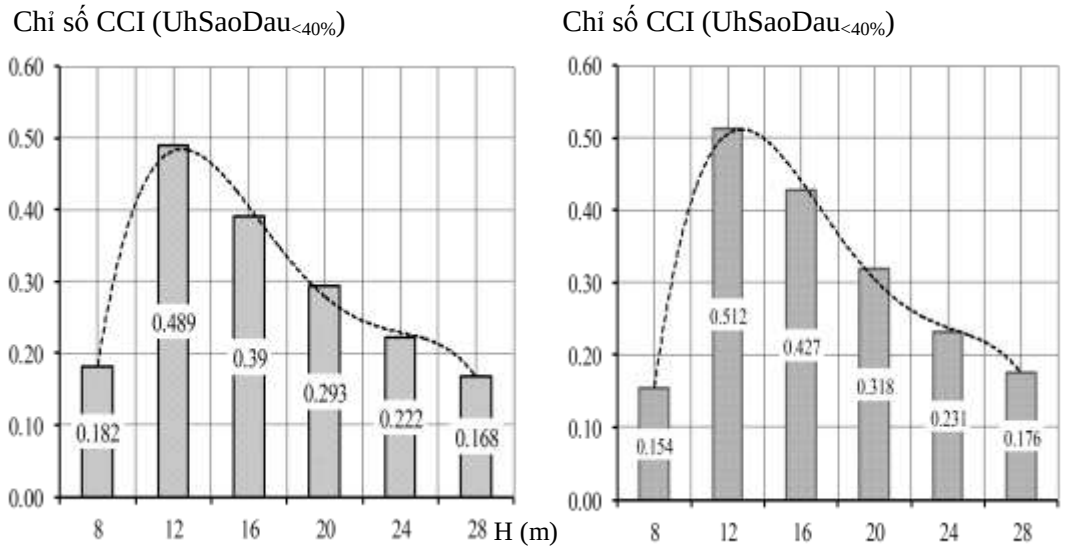
TT	Thống kê	$UhSaoDau_{<40\%}$	$UhSaoDau_{>40\%}$
(1)	(2)	(3)	(4)
1	a (Hằng số)	0,097784	0,091089
2	b*N	0,000696	0,001640
3	c*NH	0,000200	0,000318
4	d*NH^2	0,000002	0,000004
5	R^2 (%)	99,5	99,9
6	±Se	0,0136	0,0061
7	MAE	0,0095	0,0029
8	MAPE	2,9	1,3
Hàm		(9.38)	(9.39)

Bảng 9.18. Chỉ số cạnh tranh tán theo cấp chiều cao đối với $UhSaoDau_{<40\%}$. Đơn vị tính: 1,0 ha.

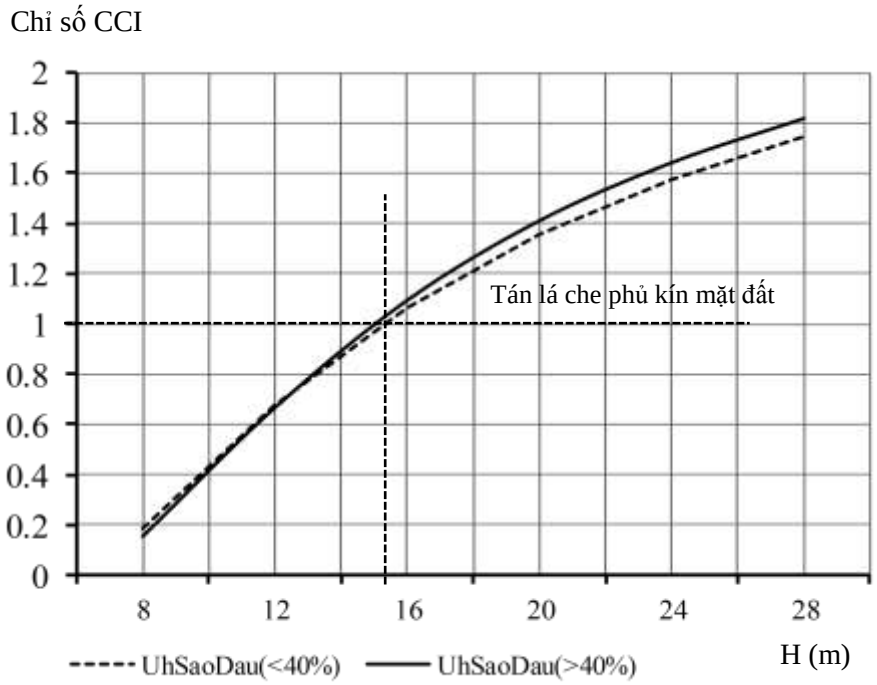
H (m)	CCI	CCI%	$CCI_{(Tích\ lũy)}$	$CCI\%_{(Tích\ lũy)}$
(1)	(2)	(3)	(4)	(5)
< 8	0,182	10,4	0.182	10,4
12	0,489	28,0	0.671	38,5
16	0,390	22,4	1.061	60,8
20	0,293	16,8	1.354	77,6
24	0,222	12,7	1.576	90,4
> 28	0,168	9,6	1.744	100,0
Tổng số	1,744	100,0		

Bảng 9.19. Chỉ số cạnh tranh tán theo cấp chiều cao đối với $UhSaoDau_{>40\%}$. Đơn vị tính: 1,0 ha.

H (m)	CCI	CCI%	$CCI_{(Tích\ lũy)}$	$CCI\%_{(Tích\ lũy)}$
(1)	(2)	(3)	(4)	(5)
< 8	0,154	8,5	0,154	8,5
12	0,512	28,2	0,666	36,6
16	0,427	23,5	1,093	60,1
20	0,318	17,5	1,410	77,6
24	0,231	12,7	1,641	90,3
> 28	0,176	9,7	1,817	100,0
Tổng số	1,817	100,0		



Hình 9.4. Đồ thị biểu diễn chỉ số cạnh tranh tán theo cấp chiều cao đối với hai ưu hợp cây họ Sao Dầu.



Hình 9.5. Đồ thị biểu diễn sự tích lũy chỉ số cạnh tranh tán theo cấp chiều cao đối với hai ưu hợp cây họ Sao Dầu.

Từ số liệu ở Bảng 9.18 và 9.19, Hình 9.4 và 9.5 cho thấy, chỉ số CCI đối với hai UhSaoDau thay đổi theo cấp H. Đối với nhóm UhSaoDau$_{<40\%}$, chỉ số CCI gia tăng dần từ cấp H < 8 m (0,182) và đạt cao nhất ở cấp H = 12 m (0,489); sau đó

giảm dần đến cấp H > 28 m (0,168). Đối với UhSaoDau$_{>40\%}$, chỉ số CCI cũng gia tăng dần từ cấp H < 8 m (0,154) và đạt cao nhất ở cấp H = 12 m (0,512); sau đó giảm dần đến cấp H > 28 m (0,176).

Nói chung, chỉ số CCI trong hai UhSaoDau này xảy ra mạnh nhất ở hai cấp H = 12 m và 16 m (50,4% đối với UhSaoDau$_{<40\%}$; 51,7% đối với UhSaoDau$_{>40\%}$). Hệ thống tán ở cả hai UhSaoDau này che phủ hoàn toàn diện tích mặt đất từ cấp H ≥ 16 m.

Bảng 9.20. Chỉ số cạnh tranh tán đối với những loài cây gỗ trong UhSaoDau$_{<40\%}$. Đơn vị tính: 1,0 ha.

TT	Loài cây gỗ	N (cây)	CCI	Tỷ lệ cạnh tranh tán:		
				N%	CCI%	Bình quân (%)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	Dầu song nàng	95	0,29	14,6	16,7	15,7
2	Trâm trắng	51	0,14	7,8	7,9	7,8
3	Cầy	45	0,13	7,0	7,6	7,3
4	Bình linh	35	0,10	5,4	5,6	5,5
5	Trường mật	36	0,08	5,6	4,3	4,9
6	Sao đen	27	0,08	4,1	4,5	4,3
	Cộng 6 loài	289	0,82	44,5	46,6	45,5
39	Loài khác	358	0,92	55,5	53,4	54,5
45	Tổng số	647	1,74	100	100	100

Chỉ số cạnh tranh đối với những loài cây gỗ trong hai UhSaoDau được dẫn ra ở Bảng 9.20 và 9.21. Đối với UhSaoDau$_{<40\%}$ (Bảng 9.20), so với chỉ số CCI của 45 loài cây gỗ (100%), 6 loài cây gỗ ưu thế và đồng ưu thế về cạnh tranh chiếm 46,6%; trong đó Dầu song nàng cạnh tranh tán mạnh nhất (16,7%), kế đến là Trâm trắng (7,9%), thấp nhất là Trường mật (4,3%). Chỉ số CCI của những loài cây gỗ khác (39 loài) là 53,4%; trung bình 1,4%/loài. Đối với UhSaoDau$_{>40\%}$ (Bảng 9.21), so với chỉ số CCI của 32 loài cây gỗ (100%), 5 loài cây gỗ ưu thế và đồng ưu thế về cạnh tranh chiếm 59,6%; trong đó Dầu song nàng cạnh tranh tán mạnh nhất (19,9%), kế đến là Vên vên (13,9%), thấp nhất là Cầy (7,8%). Chỉ số cạnh tranh của những loài cây gỗ khác (27 loài) là 40,4%; trung bình 1,5%/loài. Nói chung, những loài cây gỗ của họ Sao Dầu cạnh tranh tán lớn nhất; trong đó ưu thế là Dầu song nàng và Vên vên.

9.3. Chỉ số hình thái thân và tán lá của quần thụ

Hình thái thân cây và tán lá của các cây gỗ phản ánh tác động của những đặc tính quần thụ (tuổi, mật độ), những yếu tố môi trường (lập địa) và phương thức lâm sinh đến cây gỗ. Vì thế, phân tích hình thái cây gỗ là cơ sở cho việc đánh giá tính ổn định của cây gỗ và rừng đối với sự thay đổi của những yếu tố môi trường. Trong

lâm học, người ta quan tâm đến 4 chỉ số hay tỷ lệ sau đây; (1) Tỷ lệ chiều cao toàn thân (H, m) và đường kính thân ngang ngực (D, cm) (Công thức 9.40); (2) Tỷ lệ đường kính tán ở vị trí rộng nhất (D_T, m) và H (Công thức 9.41); (3) Tỷ lệ chiều dài tán (L_T, m) và H (Công thức 9.42); (4) Tỷ lệ chiều dài tán (L_T, m) và D_T (Công thức 9.43).

Bảng 9.21. Chỉ số cạnh tranh tán đối với những loài cây gỗ trong $UhSaoDau_{>40\%}$. Đơn vị tính: 1,0 ha.

TT	Loài cây gỗ	N (cây)	CCI	Tỷ lệ cạnh tranh tán:		
				N%	CCI%	Bình quân (%)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	Dầu song nàng	136	0,36	20,8	19,9	20,3
2	Vên vên	80	0,25	12,2	13,9	13,1
3	Làu táu trắng	63	0,18	9,6	9,7	9,6
4	Cầy	51	0,14	7,7	7,8	7,8
5	Bằng lăng ổi	39	0,15	5,9	8,3	7,1
	Cộng 5 loài	369	1,08	56,2	59,6	57,9
27	Loài khác	286	0,74	43,8	40,4	42,1
32	Tổng số	655	1,82	100	100	100

H/D (9.40)

D_T/H (9.41)

L_T/H (9.42)

L_T/D_T (9.43)

Tỷ lệ H/D được xác định theo hai phương pháp: (a) Không kể đến đơn vị đo; (b) Kể đến đơn vị đo. Khi không kể đến đơn vị đo, thì tỷ lệ H/D có thể nhận giá trị nhỏ hơn, bằng hoặc lớn hơn 1. Khi kể đến đơn vị đo, thì tỷ lệ H/D có thể nhận giá trị nhỏ hơn, bằng hoặc lớn hơn 100. Nói chung, về bản chất, hai cách tính tỷ lệ H/D này là như nhau.

Từ phân tích sự thay đổi của tỷ lệ H/D, Burschel và Huss (1997; Dẫn theo Ducey, 2009) đánh giá tính ổn định của rừng ở mỗi cấp A theo Bảng 9.22. Tỷ lệ H/D > 0,8 cho biết chiều cao phát triển quá mạnh, còn thân cây nhỏ. Hiện tượng này dẫn đến cây dễ bị đổ gãy do gió lớn hoặc do đất không ổn định. Tỷ lệ H/D < 0,8 cho biết đường kính thân cây lớn. Hiện tượng này đảm bảo cho cây đứng vững trước những tác động của gió lớn và đất không ổn định.

Bảng 9.22. Những tiêu chuẩn đánh giá tính ổn định của rừng theo tỷ lệ H/D.

H/D	> 1,0	0,8 – 1,0	0,45 – 0,8	< 0,45
Đánh giá	Rất không ổn định	Không ổn định	Ổn định	Cây mọc tự do

Dựa theo tỷ lệ L_T/H, Schutz (2001; Dẫn theo Ducey, 2009) đánh giá tính ổn định của rừng ở mỗi cấp A theo Bảng 9.23. Tỷ lệ $L_T/H < 0,5$ cho thấy chiều dài tán rất nhỏ so với chiều cao của cây. Chiều dài tán nhỏ dẫn đến hệ thống lá tiếp nhận ít ánh sáng và cây sinh trưởng kém.

Bảng 9.23. Những tiêu chuẩn đánh giá tính ổn định của rừng theo tỷ lệ L_T/H.

L_T/H	< 0,3	0,3 – 0,5	0,5 – 0,6	> 0,6
Đánh giá	Rất không ổn định	Không ổn định	Ổn định	Cây mọc tự do

Bảng 9.24 dẫn ví dụ về sự thay đổi hình thái thân cây Đước từ tuổi 11 – 25 tại khu vực Biển Tây tuộc tỉnh Cà Mau. Từ đó cho thấy, bốn tỷ lệ (H/D, L_T/H, D_T/H, L_T/D_T) khác nhau rất rõ rệt giữa các tuổi. Tỷ lệ H/D nhận giá trị thấp ở tuổi 11 - 13 và 21 -25 (1,0); nhận giá trị cao ở tuổi 15 - 19 (1,02). Tỷ lệ L_T/H gia tăng từ tuổi 11 (0,38) đến tuổi 25 (0,47). Tỷ lệ D_T/H có khuynh hướng giảm dần từ tuổi 11 (0,23) đến tuổi 25 (0,15). Tỷ lệ L_T/D_T gia tăng từ tuổi 11 (1,72) đến tuổi 25 (3,17).

Bảng 9.24. Tỷ lệ chiều cao và đường kính thân cây Đước ở các cấp tuổi.

A (năm)	N (cây)	H/D	L_T/H	D_T/H Max	L_T/D_T
(1)	(2)	(3)	(4)	(5)	(6)
11	516	1,00	0,38	0,23	1,72
13	456	1,01	0,38	0,20	1,96
15	422	1,02	0,39	0,18	2,28
17	379	1,02	0,41	0,17	2,48
19	367	1,02	0,42	0,16	2,67
21	355	1,01	0,44	0,15	2,90
23	333	1,01	0,46	0,15	3,06
25	309	1,00	0,47	0,15	3,17
Trung bình	392	1,01	0,42	0,17	2,53

Tỷ lệ H/D dao động từ 1,0 - 1,02 cho thấy chiều cao phát triển mạnh, còn đường kính thân nhỏ. Tỷ lệ L_T/H dao động từ 0,38 – 0,47 cho thấy chiều dài tán nhỏ. Hình thái thân cây như vậy chứng tỏ rừng Đước trồng từ tuổi 11 – 25 phát triển không cân đối. Hiện tượng này xảy ra là do mật độ cao dẫn đến những cây Đước phát triển mạnh chiều cao để thu nhận ánh sáng. Tuy vậy, rừng Đước trồng vẫn không bị đổ ngã do gió to và sóng lớn. Hiện tượng này được giải thích là do mật độ cao đã ngăn cản ảnh hưởng xấu của gió to và sóng lớn. Để nâng cao tính ổn định của rừng Đước, biện pháp điều chỉnh mật độ bằng cách tỉa thưa là cần thiết.

Chương 10
PHÂN TÍCH PHƯƠNG SAI

10.1. Khái quát chung

Phân tích phương sai (ANOVA) là một phương pháp xử lý số liệu rất hữu ích đối với nhiều nghiên cứu trong lâm học. Chẳng hạn: (a) Phân tích phản hồi (Biến phụ thuộc) của cây gỗ và rừng đối với sự thay đổi của một hoặc nhiều nhân tố sinh thái (Biến độc lập); (b) Phân tích phản hồi của cây tái sinh trong giai đoạn vườn ươm đối với sự thay đổi của thành phần hỗn hợp ruột bầu. Trong phân tích phương sai, những biến độc lập có thể là biến định danh và biến định lượng. Những biến định danh biểu thị cho những nhân tố sinh thái hoặc biến phân loại (Classification Variables). Những biến định lượng có thể được ghép thành nhóm hoặc khối (Block). Chẳng hạn: (a) Phân tích ảnh hưởng của độ dày tầng đất đến sản lượng rừng; (b) Phân tích ảnh hưởng ba phương thức nuôi dưỡng đến sản lượng rừng; (c) Phân tích ảnh hưởng của kích thước bầu đến sinh trưởng của cây con Gõ đỏ (*Afzelia xylocarpa* (Kurz) Craib) trong giai đoạn 6 tháng tuổi ở vườn ươm.

Tùy theo nhân tố tác động và kiểu bố trí thí nghiệm, nhà lâm học có thể xây dựng các mô hình phân tích phương sai khác nhau. Phân tích phương sai cũng có thể được thiết kế theo mô hình tuyến tính tổng quát (General Linear Model = GML). Phân tích mô hình tuyến tính tổng quát là công cụ rất mềm dẻo, đặc biệt đối với các mô hình bao gồm cả nhân tố phân loại và nhân tố định lượng. Phân tích phương sai (1 nhân tố, đa nhân tố) có thể được thực hiện cho 1 hoặc nhiều biến phụ thuộc. Điều kiện của phân tích phương sai là các biến phụ thuộc tuân theo phân bố chuẩn với phương sai bằng nhau. Trong phân tích phương sai, giả thuyết H_0: Các nghiệm thức là bằng nhau.

10.2. Các mô hình phân tích phương sai

10.2.1. Khái quát chung

Phân tích phương sai báo cáo nhiều nội dung khác nhau; trong đó nhà lâm học thường quan tâm đến những nội dung cơ bản sau đây: (1) Các đặc trưng thống kê mô tả đối với biến phản hồi; (2) Kiểm định phân bố chuẩn và ngang bằng phương sai; (3) Phân tích các biến động (ANOVA); (4) So sánh giữa các nghiệm thức và kiểm định phân hạng các nghiệm thức.

So sánh giữa các nghiệm thức bao gồm so sánh từng cặp nghiệm thức với nhau (Pairwise Comparisions) và kiểm định nhiều cặp (Pairwise Multiple Comparisions). Các kiểm định phân hạng các nghiệm thức (Post Hoc Range Tests) cho biết những giá trị trung bình của các nghiệm thức nào khác nhau có ý nghĩa thống kê. Các kiểm định này chỉ được áp dụng để so sánh ảnh hưởng giữa các nhân tố cố định đến một biến phụ thuộc. Khi có nhiều biến phụ thuộc, thì các kiểm định này được thực hiện cho từng biến riêng rẽ.

Kiểm định T được sử dụng để so sánh sự sai khác giữa hai nghiệm thức với nhau. Kiểm định T cho biết sự khác biệt giữa hai nghiệm thức ở mức xác suất sai lầm (α) cho trước, thông thường α = 0,05. Khi so sánh nhiều nghiệm thức, người ta sử dụng kiểm định nhiều cặp. Giả thiết có k cặp số trung bình cần so sánh ở mức ý nghĩa 5%, xác suất khác nhau rơi vào miền bác bỏ là 1 – 0,95^k. Chẳng hạn: Khi có 6 cặp nghiệm thức, thì số cặp so sánh ở mức ý nghĩa 5% là 15. Theo đó, xác suất khác nhau rơi vào miền bác bỏ là 0,265 = 1 – 0,95^6.

Khi có nhiều nghiệm thức nhưng chỉ cần so sánh từng cặp riêng rẽ với nhau, thì chúng ta có thể sử dụng các kiểm định LSD (Least Significance Difference), Bonferroni và Sidak. Nhưng nếu có nhiều nghiệm thức cần so sánh không chỉ từng cặp với nhau mà còn xếp hạng các nghiệm thức, thì chúng ta có thể sử dụng các kiểm định Tukey HSD, Scheffe, Duncan, Student - Newman - Keuls. Tất cả các kiểm định này đòi hỏi phương sai mẫu là bằng nhau. Kiểm định của Tukey sử dụng thống kê biên độ được Student hóa để xây dựng tất cả các so sánh cặp giữa các nghiệm thức và thiết lập tỷ lệ sai số thực nghiệm cho những cặp được chọn. Kiểm định Bonferroni sử dụng thống kê Student's T để điều chỉnh mức ý nghĩa quan sát dựa trên số lần so sánh. Khi có 5 lần so sánh với mức ý nghĩa chỉ định là 0,05, thì mức ý nghĩa cho 1 so sánh là 0,05/5 = 0,01. Vì thế, kiểm định Tukey được sử dụng để so sánh nhiều cặp trung bình. Khi chỉ có ít cặp trung bình cần so sánh thì kiểm định Bonferroni hữu dụng hơn. Các kiểm định Duncan, Student-Newman-Keuls (S-N-K) và Tukey's b là các kiểm định dựa trên hạng của các trung bình nhóm. Kiểm định LSD tương đương với kiểm định T. Hạn chế của kiểm định LSD là không có một kiểm định nào được điều chỉnh mức tin cậy cho nhiều so sánh.

Thực chất của phép so sánh giữa các nghiệm thức là xây dựng khoảng tin cậy đối với hiệu số của hai trung bình mẫu (C_i) và dựa vào khoảng tin cậy để kiểm định giả thuyết H_0. Nếu khoảng tin cậy C_i chứa giá trị Zero, thì giả thuyết H_0 được chấp nhận ở mức xác suất sai lầm P hoặc mức tin cậy 1 - P. Ngược lại, nếu khoảng tin cậy C_i không chứa giá trị Zero, thì giả thuyết H_0 bị bác bỏ ở mức xác suất sai lầm P hoặc mức tin cậy 1 - P. Ba mức xác suất sai lầm P thường được sử dụng là 0,05; 0,01 và 0,001. Dạng tổng quát của khoảng tin cậy của so sánh C_i được tính theo công thức 10.1. Ở công thức 10.1, C_i = hiệu số của hai trung bình tổng thể hay hai nhóm nghiệm thức; $C_i^{(*)}$ = ước lượng của C_i; S_{Ci}^* = độ lệch chuẩn của $C_i^{(*)}$ và được tính theo công thức 10.2 với n = dung lượng của một mẫu; MSE = sai số bình phương trung bình; $T_p^{(a)}(k)$ = đại lượng có phân phối Student hoặc phân phối F tùy theo kiểu so sánh; P = mức xác suất sai lầm; a = số nghiệm thức (với các trắc nghiệm không xếp hạng các nghiệm thức) hoặc hạng (L) của cặp nghiệm thức đem so sánh (với các trắc nghiệm có xếp hạng các nghiệm thức); k = bậc tự do của MSE.

$$C_i = C_i^{(*)} \pm S_{Ci}^* T_P^{(a)}(k) \qquad (10.1)$$

Đối với trắc nghiệm LSD, sự khác biệt lớn nhất của hai trung bình mẫu (i, j) được tính theo công thức 10.3; trong đó DF = độ tự do của MSE (N - k), N = tổng dung lượng của k mẫu.

$$S_{Ci}^* = \sqrt{\frac{2MSE}{n}} \text{ (khi dung lượng hai mẫu bằng nhau)} \qquad (10.2a)$$

$S_{Ci}^{*} = \sqrt{MSE*(1/n_1 + 1/n_2)}$ (khi dung lượng hai mẫu khác nhau) (10.2b)

$$LSD_{i,J} = T_{\alpha/2,\, DF}\sqrt{\frac{2*MSE}{n}} \qquad (10.3)$$

Quy tắc quyết định: Ở mức ý nghĩa α = 0,05 (Mặc định).

Nếu $T < T_{0.025,\, DF}$ thì trung bình của hai mẫu (nghiệm thức) là bằng nhau.

Nếu $T > T_{0.025,\, DF}$ thì trung bình của hai mẫuu là khác nhau.

Sự khác biệt giữa hai trung bình mẫu có thể được tính theo khoảng tin cậy (Công thức 10.4); trong đó n = dung lượng của 1 mẫu. Công thức 10.4(c,d) được áp dụng cho hai nghiệm thức có dung lượng mẫu khác nhau. Quy tắc quyết định: Ở mức ý nghĩa α = 0,025 (Mặc định), nếu khoảng tin cậy chứa Zero thì trung bình của hai nghiệm thức là bằng nhau.

$$(X_A - X_B) + T_{Crit}*\sqrt{\frac{2*MSE}{n}} \qquad (10.4a)$$

$$(X_A - X_B) - T_{Crit}*\sqrt{\frac{2*MSE}{n}} \qquad (10.4b)$$

$$(X_A - X_B) + T_{Crit}*\sqrt{MSE*(1/n_1 + 1/n_2)} \qquad (10.4c)$$

$$(X_A - X_B) - T_{Crit}*\sqrt{MSE*(1/n_1 + 1/n_2)} \qquad (10.4d)$$

Trắc nghiệm Duncan được tính theo công thức 10.5; trong đó $q_p^{(a)}(k)$ là trị số tra bảng lập sẵn dùng cho trắc nghiệm Duncan.

$$T_p^{(a)}(k) = \frac{q_p^{(a)}(k)}{\sqrt{2}} \qquad (10.5)$$

Trắc nghiệm Tukey được tính theo công thức 10.6; trong đó X_i và X_J là hai trung bình mẫu, MSE = sai số bình phương trung bình, n = dung lượng của một mẫu. Công thức 10.6a được áp dụng cho hai nghiệm thức có dung lượng mẫu bằng nhau. Công thức 10.6b được áp dụng cho hai nghiệm thức có dung lượng mẫu khác nhau.

$$q = \frac{X_i - X_J}{\sqrt{MSE/n}} \qquad (10.6a)$$

$$q = \frac{X_i - X_J}{\sqrt{2*MSE/(n_i + n_J)}} \qquad (10.6b)$$

Giá trị tiêu chuẩn của phân bố q (q_{Crit}) được tra từ bảng phân hạng của Tukey (Studentized Range Table) dựa trên mức ý nghĩa α, k nhóm và độ tự do là N – k; trong đó N = tổng dung lượng của k mẫu. Giả thuyết H_0: Trung bình của hai mẫu là không khác nhau. Quy tắc quyết định: Ở mức ý nghĩa α = 0,05 (Mặc định), nếu $q < q_{Crit}$ thì trung bình của hai nhóm là bằng nhau. Nếu $q > q_{Crit}$ thì trung bình của hai nhóm là khác nhau. Sự khác biệt giữa hai trung bình mẫu cũng có thể được xác định theo khoảng tin cậy (Công thức 10.7); trong đó n = dung lượng của 1 mẫu. Công thức 10.7(c,d) được áp dụng cho hai nghiệm thức có dung lượng mẫu khác nhau. Quy tắc quyết định: Ở mức ý nghĩa α = 0,05 (Mặc định), nếu khoảng tin cậy chứa Zero thì trung bình của hai nghiệm thức là bằng nhau.

$$(X_i - X_J) + (q_{Crit}*\sqrt{MSE/n}) \qquad (10.7a)$$

$$(X_i - X_J) - (q_{Crit}*\sqrt{MSE/n}) \qquad (10.7b)$$

$(X_i - X_J) + (q_{Crit} * \sqrt{2*MSE/(n_i+n_J)})$ (10.7c)

$(X_i - X_J) - (q_{Crit} * \sqrt{2*MSE/(n_i+n_J)})$ (10.7d)

10.2.2. Mô hình phân tích phương sai một nhân tố

10.2.2.1. Kiểu thí nghiệm ngẫu nhiên hoàn toàn

Đối với kiểu bố trí thí nghiệm ngẫu nhiên hoàn toàn (CRD = Complete Randomized Design), dung lượng mẫu (số liệu quan sát) có thể bằng nhau hoặc không bằng nhau và trật tự của các nghiệm thức không ảnh hưởng đến kết quả phân tích biến động. Ở kiểu thí nghiệm này có hai nguồn biến động: (a) Nguồn biến động giữa các nghiệm thức; (b) Nguồn biến động ngẫu nhiên của các số liệu được đo đạc ở mỗi nghiệm thức (Sai số thí nghiệm). Giả thuyết H_0: Các nghiệm thức hay nhân tố thí nghiệm ảnh hưởng đồng đều đến kết quả thí nghiệm. Đối thuyết H_0^-: Các nghiệm thức ảnh hưởng khác nhau đến kết quả thí nghiệm. Khi nguồn biến động giữa các nghiệm thức lớn hơn biến động ngẫu nhiên, thì giả thuyết H_0 bị bác bỏ.

Mô hình phân tích thống kê có dạng như hàm 10.8; trong đó Y_{ij} = biến phụ thuộc; μ = trung bình tổng thể; α_i = tham số đặc trưng cho ảnh hưởng của nhân tố thí nghiệm (i = 1, 2 …a); ε_{ij} = sai số ngẫu nhiên ($\varepsilon_{ij} \rightarrow N(0, \delta^2)$).

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad (10.8)$$

Mô hình tuyến tính tổng quát có dạng như hàm 10.9; trong đó X_1 = nhân tố thí nghiệm, ε = sai số ngẫu nhiên.

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon \quad (10.9)$$

10.2.2.2. Kiểu thí nghiệm theo khối đầy đủ ngẫu nhiên

Trong kiểu bố trí thí nghiệm theo khối đầy đủ ngẫu nhiên (RCBD = Randomized Complete Block Design), số liệu trong mỗi nghiệm thức được đo đạc từ những điều kiện khác nhau, còn số liệu của các nghiệm thức được đo đạc trong mỗi khối có điều kiện như nhau. Kiểu thí nghiệm theo khối đầy đủ ngẫu có thể được thiết kế theo sơ đồ ở Hình 10.1. Thí nghiệm RCBD có ba nguồn biến động: (1) Nguồn biến động của các nghiệm thức; (2) Sai số thí nghiệm; (3) Nguồn biến động theo khối nghiệm thức.

Giả thuyết H_0: Các nghiệm thức và các khối ảnh hưởng đồng đều đến kết quả thí nghiệm. Đối thuyết H_0^-: Các nghiệm thức và các khối ảnh hưởng không đồng đều đến kết quả thí nghiệm. Khi nguồn biến động của các nghiệm thức và khối lớn hơn biến động ngẫu nhiên, thì giả thuyết H_0 bị bác bỏ.

Mô hình phân tích thống kê có dạng như hàm 10.10; trong đó Y_{ij} = biến phụ thuộc, μ = trung bình tổng thể, α_i = tham số đặc trưng cho ảnh hưởng của nhân tố thí nghiệm (i = 1, 2 …a), β_j = tham số đặc trưng cho ảnh hưởng của khối (i = 1, 2 …r), ε_{ij} = sai số ngẫu nhiên ($\varepsilon_{ij} \rightarrow N(0, \delta^2)$).

$$Y_{ij} = \mu + \alpha_i + \beta_{ij} + \varepsilon_{ij} \quad (10.10)$$

Số lần lặp (Khối)	Nghiệm thức			
	1	2	…	k
1	Y_{11}	Y_{21}	…	Y_{k1}
…	…	…	…	…
r	Y_{1r}	…	…	Y_{kr}

Hình 10.1. Sơ đồ bố trí thí nghiệm theo khối đầy đủ ngẫu nhiên.

Khi sử dụng mô hình GLM để phân tích biến động 1 nhân tố theo kiểu thí nghiệm RCBD, mô hình tuyến tính tổng quát bao gồm hai kiểu. Kiểu thứ nhất bao gồm đầy đủ nhân tố thí nghiệm và nhân tố khối (Hàm 10.11); trong đó β_0 = điểm chặn (Intercept), X_1 = nhân tố thí nghiệm, X_2 = nhân tố khối, ε = sai số ngẫu nhiên. Kiểu thứ hai chỉ bao gồm nhân tố thí nghiệm, còn nhân tố khối được gộp vào thành phần sai số (Hàm 10.12); trong đó β_0 = điểm chặn (Intercept), X_1 = nhân tố thí nghiệm, ε = sai số ngẫu nhiên.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \qquad (10.11)$$

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon \qquad (10.12)$$

10.2.2.3. Kiểu thí nghiệm bình phương la tinh

Kiểu bình phương la tinh là kiểu thí nghiệm mà các nghiệm thức được bố trí ngẫu nhiên cả ở khối hàng lẫn khối cột; trong đó số nghiệm thức, số khối hàng và số khối cột là bằng nhau. Ví dụ: Nghiên cứu ảnh hưởng của 6 hàm lượng Super lân (0, 1, 2, 3, 4, 5% so với trọng lượng ruột bầu) đến sinh trưởng chiều cao (H, cm) của Dầu song nàng (*Dipterocarpus dyeri*) sau 8 tháng tuổi ở vườn ươm. Thí nghiệm được bố trí theo khối với 6 lần lặp lại. Tổng cộng 36 lô thí nghiệm. Các lô thí nghiệm này được phân phối theo sơ đồ bao gồm 6 hàng và 6 cột (Hình 10.2).

Kiểu bình phương la tinh bao gồm bốn nguồn biến động: (1) Nguồn biến động của các nghiệm thức; (2) Nguồn biến động ngẫu nhiên của các số liệu được thu thập ở mỗi nghiệm thức (Sai số thí nghiệm); (3) Nguồn biến động của các số liệu được thu thập ở mỗi nghiệm thức theo khối cột; (4) Nguồn biến động của các số liệu được thu thập ở mỗi nghiệm thức theo khối hàng. Giả thuyết H_0: Các nghiệm thức và các khối ảnh hưởng đồng đều đến kết quả thí nghiệm. Khi nguồn biến động của các nghiệm thức và khối lớn hơn biến động ngẫu nhiên, thì giả thuyết H_0 bị bác bỏ.

Mô hình phân tích thống kê có dạng như hàm 10.13; trong đó Y_{ij} = biến phụ thuộc, μ = trung bình tổng thể, α_i = tham số đặc trưng cho ảnh hưởng của nhân tố thí nghiệm (i = 1, 2 …a), β_j = tham số đặc trưng cho ảnh hưởng của khối cột (i = 1, 2 …r), β_k = tham số đặc trưng cho ảnh hưởng của khối hàng (j = 1…k), ε_{ijk} = sai số ngẫu nhiên ($\varepsilon_{ijk} \rightarrow N(0, \delta^2)$).

$$Y_{ij} = \mu + \alpha_i + \beta_j{*}\text{Khối cột} + \beta_k{*}\text{Khối hàng} + \varepsilon_{ij} \qquad (10.13)$$

		Khối hàng					
		1	2	3	4	5	6
Khối cột	1	A	B	C	D	E	F
	2	B	A	E	F	C	D
	3	C	F	A	E	D	B
	4	D	C	B	A	F	E
	5	E	D	F	C	B	A
	6	F	E	D	B	A	C

Kí hiệu: ABCDEF = 6 nghiệm thức; trong đó A = 0, B = 1, C = 2, D = 3, E = 4, F = 5% Super lân so với trọng lượng bầu.

Hình 10.2. Sơ đồ bố trí thí nghiệm theo kiểu bình phương la tinh.

Khi sử dụng mô hình GLM để phân tích biến động 1 nhân tố theo kiểu bình phương la tinh, mô hình tuyến tính tổng quát có dạng như Hàm 10.14a; trong đó β_0 = điểm chặn (Intercept), X_1 = nhân tố thí nghiệm, X_2 = nhân tố khối cột, X_3 = nhân tố khối hàng, β_1, β_2 và β_3 = hệ số phản ánh ảnh hưởng của X_1, X_2 và X_3, ε = sai số ngẫu nhiên. Nếu mô hình chỉ bao gồm nhân tố thí nghiệm, còn nhân tố khối (hàng và cột) được gộp vào thành phần sai số, thì mô hình tuyến tính có dạng như Hàm 10.14b. Ở đây β_0 là điểm chặn (Intercept), β_1 = hệ số phản ánh ảnh hưởng của X_1, ε = sai số ngẫu nhiên.

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \varepsilon \qquad (10.14a)$$

$$Y = \beta_0 + \beta_1X_1 + \varepsilon \qquad (10.14b)$$

10.2.3. Mô hình phân tích phương sai hai nhân tố

10.2.3.1. Kiểu thí nghiệm hoàn toàn ngẫu nhiên

Kiểu thí nghiệm hai nhân tố ngẫu nhiên hoàn toàn (RCD) có m lần lặp lại ở mỗi tổ hợp của nhân tố A và nhân tố B được thiết kế theo sơ đồ như Hình 10.3. Mô hình phân tích thống kê có dạng như hàm 10.15. Ở hàm 10.15, Y_{ijk} = biến phụ thuộc; μ = trung bình tổng thể được ước lượng từ trung bình mẫu Y_{bq}; α_i (i = 1…a) = tham số đặc trưng cho ảnh hưởng của nhân tố A; β_j (j = 1…b) = tham số đặc trưng cho ảnh hưởng của nhân tố B; $(\alpha\beta)_{ij}$ = tương tác giữa nhân tố A với nhân tố B; ε_{ijk} = sai số ngẫu nhiên; k = số lần lặp (k = 1…m). Giả thuyết đối với nhân tố A, H_a: α_i = 0 với mọi i; H_1: ít nhất có một $\alpha_i \neq 0$. Đối với nhân tố B, H_b: $\beta i = 0$ với mọi i; H_1: ít nhất có một $\beta_i \neq 0$. Đối với tổ hợp nhân tố A và B, H_{ab}: Không có sự tương tác giữa nhân tố A với nhân tố B.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \qquad (10.15)$$

Nhân tố A	Nhân tố B				
	1	2	3	…	b
A_1	Y_{111}	Y_{121}	Y_{131}		Y_{1b1}
	Y_{112}	Y_{122}	Y_{132}		Y_{1b2}
	Y_{11m}	Y_{12m}	Y_{13m}		Y_{1bm}
A_2	…				
A_3	Y_{311}	Y_{321}	Y_{331}		Y_{3b1}
	Y_{312}	Y_{322}	Y_{332}		Y_{3b2}
	Y_{31m}	Y_{32m}	Y_{33m}		Y_{3bm}

Hình 10.3. Sơ đồ bố trí thí nghiệm hai nhân tố ngẫu nhiên hoàn toàn.

Khi sử dụng mô hình GLM để phân tích biến động của 2 nhân tố theo kiểu thí nghiệm ngẫu nhiên hoàn toàn, mô hình tuyến tính tổng quát có dạng như hàm 10.16. Ở hàm 10.16, Y = biến phụ thuộc; β_0 = điểm chặn (Intercept); X_1 = nhân tố thí nghiệm 1; X_2 = nhân tố thí nghiệm 2; X_1*X_2 = tương tác giữa nhân tố nhân tố X_1 với nhân tố X_2; β_1, β_2 = hệ số phản ánh ảnh hưởng của X_1 và X_2; β_3 = hệ số phản ánh ảnh hưởng của cả hai nhân tố X_1 và X_2; ε = sai số ngẫu nhiên

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2 + \varepsilon \qquad (10.16)$$

10.2.3.2. Kiểu thí nghiệm theo khối đầy đủ ngẫu nhiên

Kiểu thí nghiệm hai nhân tố theo khối đầy đủ ngẫu nhiên hoàn toàn có m lần lặp lại ở mỗi tổ hợp của nhân tố A và nhân tố B được thiết kế theo sơ đồ như Hình 10.4. Mô hình phân tích thống kê có dạng như hàm 10.17; trong đó Y_{ijk} = biến phụ thuộc, μ = trung bình tổng thể được ước lượng từ trung bình mẫu (Y_{Bq}), αi (i = 1…a) = tham số đặc trưng cho ảnh hưởng của nhân tố thí nghiệm A, β_j (j = 1…b) = tham số đặc trưng cho ảnh hưởng của nhân tố thí nghiệm B, $(\alpha\beta)_{ij}$ = ảnh hưởng tương tác giữa nhân tố A với nhân tố B, k_k = ảnh hưởng của khối (k = 1…r), ε_{ijk} = sai số ngẫu nhiên. Giả thuyết đối với nhân tố A, H_a: $\alpha_i = 0$ với mọi i; H_1: ít nhất có 1 $\alpha_i \neq 0$. Đối với nhân tố B, H_b: $\beta_i = 0$ với mọi i; H_1: ít nhất có 1 $\beta i \neq 0$. Đối với tổ hợp nhân tố A và B, H_{ab}: Không có tương tác giữa nhân tố A với nhân tố B.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + k_k + (\alpha\beta)_{ij} + \varepsilon_{ijk} \qquad (10.17)$$

Khi sử dụng mô hình GLM để phân tích biến động của hai nhân tố theo khối đầy đủ ngẫu nhiên, mô hình tuyến tính tổng quát có dạng như hàm 10.18. Ở hàm 10.18, Y = biến phụ thuộc; β_0 = điểm chặn (Intercept); X_1 = nhân tố thí nghiệm 1; X_2 = nhân tố thí nghiệm 2; X_1*X_2 = tương tác giữa nhân tố X_2 và X_2; β_1, β_2 = hệ số phản ánh ảnh hưởng của X_1 và X_2, β_3 = hệ số phản ánh ảnh hưởng của cả hai nhân tố X_1 và X_2; X_3 = nhân tố khối; β_4 = hệ số phản ánh ảnh hưởng của khối; ε = sai số ngẫu nhiên. Nếu gộp nhân tố khối vào thành phần sai số, thì mô hình tuyến tính tổng quát có dạng như hàm 10.19.

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2 + \beta_4{*}X_3 + \varepsilon \qquad (10.18)$$

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2 + \varepsilon \qquad (10.19)$$

Khối	Nhân tố A	Nhân tố B				
		1	2	3	…	b
1	A_1	Y_{111}	Y_{121}	Y_{131}	…	Y_{1b1}
	A_2	Y_{211}	Y_{221}	Y_{231}	…	Y_{2b1}
	A_3	Y_{311}	Y_{321}	X_{331}	…	X_{3b1}
2	A_1	Y_{112}	Y_{122}	Y_{132}	…	Y_{1b2}
	A_2	Y_{212}	Y_{222}	Y_{232}	…	Y_{2b2}
	A_3	Y_{312}	Y_{322}	X_{332}	…	X_{3b2}
3	A_1	Y_{113}	Y_{123}	Y_{133}	…	Y_{1b3}
	A_2	Y_{213}	Y_{223}	Y_{233}	…	Y_{2b3}
	A_3	Y_{313}	Y_{323}	X_{333}	…	X_{3b3}

Hình 10.4. Sơ đồ bố trí thí nghiệm hai nhân tố theo khối đầy đủ ngẫu nhiên.

10.2.3.3. Kiểu thí nghiệm theo kiểu thí nghiệm bình phương la tinh

Giả thiết phân tích ảnh hưởng của hai nhân tố đến sinh trưởng chiều cao của Dầu song nàng 8 tháng tuổi: (1) Kích thước bầu với 2 mức cấp 15*25 cm và 20*30 cm; (2) Phân tổng hợp NPK với 3 mức 1%, 2% và 3% so với trọng lượng bầu. Tổng số 2*3 = 6 nghiệm thức. Sơ đồ thí nghiệm có thể được bố trí như Hình 10.5.

Thí nghiệm 2 nhân tố theo kiểu bình phương la tinh có 6 nguồn biến động: (1) Nguồn biến động của các nghiệm thức; (2) Sai số thí nghiệm; (3) Nguồn biến động gây ra bởi nhân tố X_1; (4) Nguồn biến động gây ra bởi nhân tố X_2; (5) Nguồn biến động của các số liệu được thu thập ở mỗi nghiệm thức theo khối cột; (6) Nguồn biến động của các số liệu được thu thập ở mỗi nghiệm thức theo khối hàng. Các giả thuyết H_0: (1) Các nhân tố độc lập (X_1 và X_2) ảnh hưởng đồng đều đến kết quả thí nghiệm; (2) Không có ảnh hưởng tương tác giữa các nhân tố độc lập đến kết quả thí nghiệm; (3) Các khối hàng và khối cột ảnh hưởng đồng đều đến kết quả thí nghiệm.

Mô hình phân tích thống kê có dạng như hàm 10.20; trong đó Y_{ijk} = biến phụ thuộc, μ = trung bình tổng thể được ước lượng từ trung bình mẫu (Y_{Bq}), α_i (i = 1…a) = tham số đặc trưng cho ảnh hưởng của nhân tố X_1, β_j (j = 1…b) = tham số đặc trưng cho ảnh hưởng của nhân tố X_2, $(\alpha\beta)_{ij}$ = tương tác giữa nhân tố X_1 với nhân tố X_2, γ_c = tham số đặc trưng cho ảnh hưởng của nhân tố cột, δ_h = tham số đặc trưng cho ảnh hưởng của nhân tố hàng, ε_{ijch} = sai số ngẫu nhiên.

$$Y_{ijch} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_c + \delta_h + \varepsilon_{ijch} \qquad (10.20)$$

		Khối hàng					
		1	2	3	4	5	6
Khối cột	1	A	B	C	D	E	F
	2	B	A	E	F	C	D
	3	C	F	A	E	D	B
	4	D	C	B	A	F	E
	5	E	D	F	C	B	A
	6	F	E	D	B	A	C

Hình 10.5. Sơ đồ bố trí thí nghiệm hai nhân tố theo kiểu bình phương la tinh.

Nếu sử dụng mô hình GLM để phân tích biến động của 2 nhân tố được bố trí theo kiểu bình phương la tinh, thì mô hình có dạng như Hàm 10.21. Ở Hàm 10.21, X_1 = nhân tố thí nghiệm 1; X_2 = nhân tố thí nghiệm 2; X_1*X_2 = tương tác của nhân tố thí nghiệm 1 và 2; X_3 = số lần lặp theo hàng; X_4 = số lần lặp theo cột; ε = sai số thí nghiệm. Nếu gộp những biến động theo khối hàng và theo khối cột vào thành phần sai số, thì mô hình GLM có dạng như Hàm 10.22.

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_1*X_2 + \varepsilon \qquad (10.21)$$

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2 + \varepsilon \qquad (10.22)$$

10.2.3.4. Kiểu thí nghiệm theo kiểu thí nghiệm lô chính – lô phụ

Thí nghiệm 2 nhân tố (A và B) theo kiểu lô chính (nhân tố A) và lô phụ (nhân tố B) có 6 nguồn biến động: (1)Nguồn biến động của nhân tố A, (2) Nguồn biến động của nhân tố B, (3) Sai số gây ra bởi nhân tố A, (4) Sai số gây ra bởi nhân tố B, (5) Nguồn biến động số liệu do lặp lại (Khối), (6) Nguồn biến động gây ra bởi sự tương tác giữa nhân tố A với nhân tố B. Các giả thuyết: (1) Từng nhân tố A và B ảnh hưởng đồng đều đến kết quả thí nghiệm; (2) Không có ảnh hưởng tương tác giữa nhân tố A với nhân tố B đến kết quả thí nghiệm; (3) Các khối hàng và cột ảnh hưởng đồng đều đến kết quả thí nghiệm. Giả sử cần phân tích ảnh hưởng của 3 phương thức lâm sinh (PTLS) đến sinh trưởng chiều cao (H, cm) của 3 loài (Dầu rái, Sao đen, Vên vên). Mỗi thí nghiệm được bố trí lặp lại 3 lần. Ở đây chọn lô chính là 3 PTLS, còn lô phụ là loài cây. Sơ đồ thí nghiệm có thể được bố trí như Hình 10.6.

Mô hình phân tích thống kê có dạng như hàm 10.23; trong đó Y_{ijk} = biến phụ thuộc, μ = trung bình tổng thể được ước lượng từ trung bình mẫu (Y_{Bq}), α_i (i = 1…a) = tham số đặc trưng cho ảnh hưởng của nhân tố chính A, β_j (j = 1…b) = tham số đặc trưng cho ảnh hưởng của nhân tố phụ B, $(\alpha\beta)_{ij}$ = tương tác giữa nhân tố A với nhân tố B, k_k = tham số đặc trưng cho ảnh hưởng của khối, ε_{ik} = sai số ngẫu nhiên của nhân tố A, ε_{ijk} = sai số ngẫu nhiên của nhân tố B.

$$Y_{ijk} = \mu + \alpha_i + k_k + \varepsilon_{ik} + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ij} \qquad (10.23)$$

Nghiệm thức	Số lần lặp lại						
	I			II	III		
	Dầu rái	Sao đen	Vên vên	...	Dầu rái	Sao đen	Vên vên
1	Y_{A11}	Y_{A21}	Y_{A31}	...	Y_{A13}	Y_{A23}	Y_{A33}
2	Y_{B11}	Y_{B21}	Y_{B31}	...	Y_{B13}	Y_{B23}	Y_{B33}
3	Y_{C11}	Y_{C21}	Y_{C31}	...	Y_{C13}	Y_{C23}	Y_{C33}

Hình 10.6. Sơ đồ bố trí thí nghiệm hai nhân tố có lô phụ.

Nếu sử dụng mô hình tuyến tính tổng quát để phân tích biến động của 2 nhân tố được bố trí theo kiểu lô chính – lô phụ, thì mô hình có dạng như Hàm 10.24. **Ở** Hàm 10.24, Y = biến phụ thuộc, X_1 = nhân tố thí nghiệm 1 (bố trí ở lô chính), X_2 = nhân tố thí nghiệm 2 (bố trí ở lô phụ), Khối = sai số của lô chính, ε = sai số thí nghiệm.

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2 + \beta_4\text{Khối} + \varepsilon \qquad (10.24)$$

10.3. Ứng dụng phân tích phương sai trong lâm học

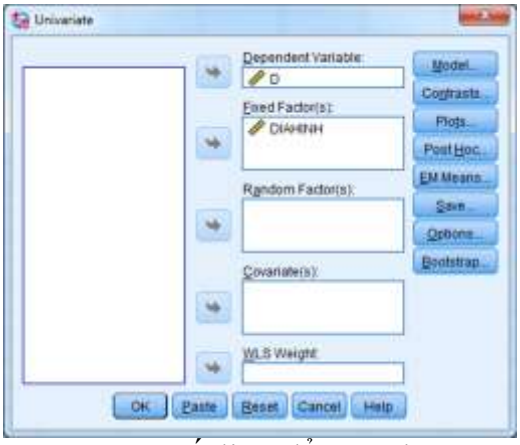
10.3.1. Phân tích phương sai một nhân tố

10.3.1.1. Kiểu thí nghiệm ngẫu nhiên hoàn toàn

Giả sử nghiên cứu ảnh hưởng của 3 cấp địa hình (chân, sườn, đỉnh) đến sinh trưởng D (cm) của rừng trồng Sao đen 35 tuổi. Kết quả thí nghiệm được ghi lại ở Phụ lục 7. Mô hình tuyến tính tổng quát: $D = \beta_0 + \beta_1X_1 + \varepsilon$; trong đó X_1 = nhân tố địa hình, ε = sai số ngẫu nhiên. Để kiểm định mô hình này, các bước xử lý số liệu bằng phần mềm SPSS for Window như dưới đây.

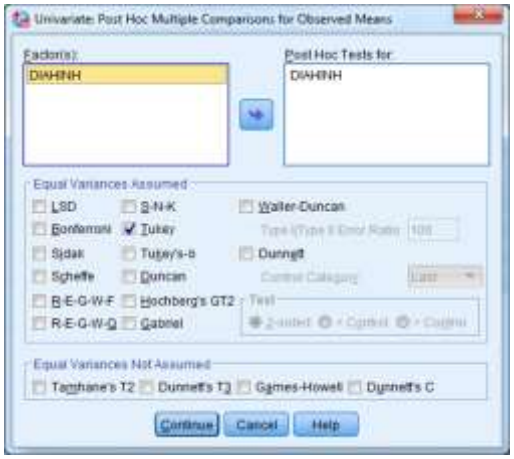
Hình 10.7. Đường kính của rừng Sao đen trên ba cấp địa hình.

Bước 1. Nhập số liệu vào bảng tính SPSS. Số liệu được nhập theo hai cột. Cột 1 là biến D theo thứ tự từ nghiệm thức 1 đến nghiệm thức 3. Cột 2 nhập mã biến độc lập (X_i) hay địa hình. Ở ví dụ này, mã số 1, 2 và 3 tương ứng là 1/3 sườn đỉnh, 1/3 sườn giữa và 1/3 sườn dưới (Hình 10.7).

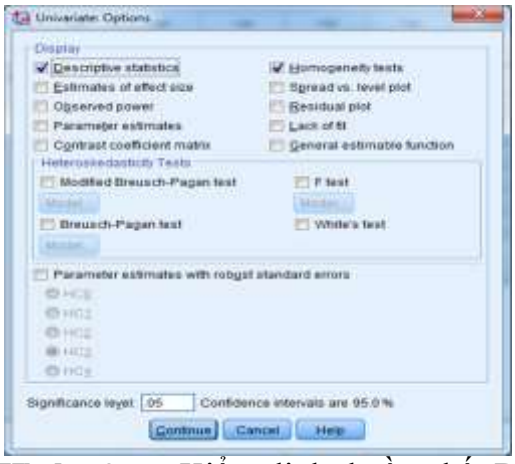


Hình 10.8. Số liệu để xây dựng mô hình: $D = \beta_0 + \beta_1 X_1 + \varepsilon$.

Hình 10.9. Xây dựng mô hình: $D = \beta_0 + \beta_1 X_1 + \varepsilon$.



Hình 10.10. So sánh D của rừng Sao đen giữa ba địa hình.



Hình 10.11. Kiểm định thuần nhất D của rừng Sao đen giữa ba địa hình.

Bước 2. Xây dựng mô hình tuyến tính tổng quát: $D = \beta_0 + \beta_1 X_1 + \varepsilon$. Để kiểm định mô hình này, từ menu chính, chọn Analyze > GLM: Univariate. Đến đây chọn biến D và chuyển nó vào Dependent Variable. Tương tự, chọn biến Địa hình và chuyển vào Fixed Factor (Hình 10.8). Để thực hiện các phân tích thống kê, từ cửa sổ Univariate ở Hình 10.8, chọn Model. Kế đến đánh dấu mục Custom và chuyển biến Địa hình từ ô Factor & Covariate sang ô Model (Hình 10.9). Tiếp đến chọn Continue ở Hình 10.9; sau đó chọn Post Hoc ở Hình 10.8. Khi xuất hiện cửa sổ Univariate Post Hoc Multiple Comparisons for Observed Means, chọn nhân tố Địa hình trong ô Factor và chuyển nó sang ô Post Hoc Tests for. Kế đến chọn kiểm định Tukey's Test và Continue (Hình 10.10). Mức tin cậy mặc định là 0,95. Sau đó kiểm định thuần nhất D của rừng Sao đen giữa ba địa hình. Để đạt mục đích này, tại cửa sổ Univariate: Option (Hình 10.8), chọn Homogeneity Test và Descriptive Statistics (Hình 10.11). Cuối cùng chọn Continue ở Hình 10.11 và chọn OK Hình 10.8 để nhận kết quả phân tích GLM (Bảng 10.1).

Bảng 10.1a. Kiểm định ảnh hưởng của địa hình đến sinh trưởng của rừng Sao đen.

Source	SS_{III}	df	MS	F	Sig.
(1)	(2)	(3)	(4)	(5)	(6)
Corr. Model (1)	416.144	2	208.072	6.235	0.004
Intercept (2)	70725.689	1	70725.68	2119.24	0.000
Nghiệm thức (3)	416.144	2	208.072	6.235	0.004
Error (4)	1401.667	42	33.373		
Total (5)	72543.500	45			
Corr. Total (6)	1817.811	44			

Bảng 10.1b. Kiểm định ảnh hưởng của địa hình đến sinh trưởng của rừng Sao đen.

Địa hình	N (Cây)	Nhóm	
		1	2
Đỉnh	15	35.567	
Chân	15	40.500	40.500
Sườn	15		42.867
Sig.		0.061	0.506

Ở Bảng 10.1, cột có nhãn Source cho biết các nguồn biến động của biến phụ thuộc. Corrected Model (1) biểu thị tổng các biến thiên của biến phụ thuộc do mô hình hay do nhân tố nghiên cứu gây ra (3). Intercept là điểm chặn của đường hồi quy giữa trung bình của các nghiệm thức. Giả thuyết H_0: Điểm chặn = 0. Nếu P < 0,05 (Mặc định), thì giả thuyết H_0 bị bác bỏ. Nghiệm thức (3) biểu thị biến động của biến phụ thuộc do nghiệm thức gây ra. Error (4) là sai số thí nghiệm hay biến động của những nhân tố không giải thích được. Total (5) là tổng bình phương của biến phụ thuộc (Y_i^2). Corrected Total (6) là tổng chênh lệch bình phương của bộ phận tạo ra sự biến thiên của biến phụ thuộc do mô hình (1) cộng với bộ phận tạo ra biến thiên của biến phụ thuộc do sai số (4), nghĩa là (6) = (1) + (4). Cột có nhãn SS_{III} cho biết tổng bình phương của các nguồn biến thiên theo kiểu III. Cột có nhãn df (Degree of Freedom) chỉ độ tự do của các thành phần trong mô hình. Cột có nhãn MS (Mean Square) biểu thị trung bình bình phương của các nguồn biến động và được tính bằng cách chia tổng bình phương của các nguồn biến thiên (Cột 2) cho độ tự do tương ứng (Cột 3). Cột có nhãn F (Cột 5) là tiêu chuẩn kiểm định giả thuyết H_0 cho rằng các mức của nhân tố nghiên cứu (Biến độc lập) ảnh hưởng đồng đều đến kết quả thí nghiệm (Biến phụ thuộc). Giá trị F = (Means Square/Error) là trung bình bình phương của các nguồn biến động chia cho sai số. Nếu giả thuyết H_0 là đúng thì F = 1. Nếu F >> 1 thì bác bỏ H_0 và chấp nhận H_0^-, nghĩa là các mức của nhân tố nghiên cứu ảnh hưởng không đồng đều đến biến phụ thuộc. Cột có nhãn (Sig) cho biết mức ý nghĩa của kiểm định hay xác suất chấp nhận giả thuyết H_0. Trong các sách thống kê, mức ý nghĩa của kiểm định thường được kí hiệu bằng chữ P hay P – value. Nếu P < 0,05 thì giả thuyết H_0 bị bác bỏ.

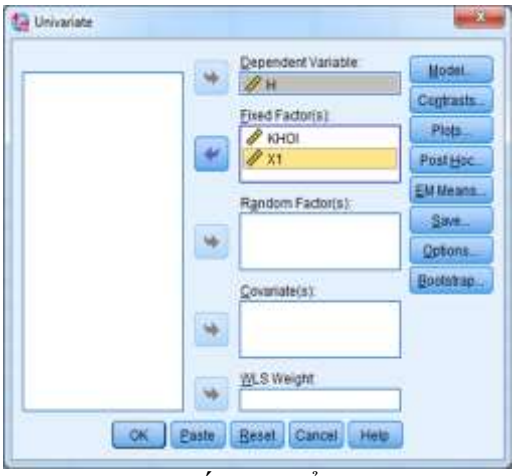
10.3.1.2. Kiểu thí nghiệm theo khối đầy đủ ngẫu nhiên

Giả sử cần xác định ảnh hưởng của 5 cấp độ tàn che đến sinh trưởng chiều cao (H, cm) của cây con Dầu song nàng 8 tháng tuổi trong giai đoạn vườn ươm. Thí nghiệm được bố trí theo kiểu khối đầy đủ ngẫu nhiên hoàn toàn với 5 nghiệm thức; mỗi nghiệm thức lặp lại 3 lần. Kết quả thí nghiệm được ghi lại ở Phụ lục 8. Mô hình tuyến tính tổng quát: $H = \beta_0 + \beta_1 X_1 + \beta_2$*Khối + ε; trong đó X_1 = nhân tố thí nghiệm, ε = sai số ngẫu nhiên. Để kiểm định mô hình này, các bước xử lý số liệu như dưới đây.

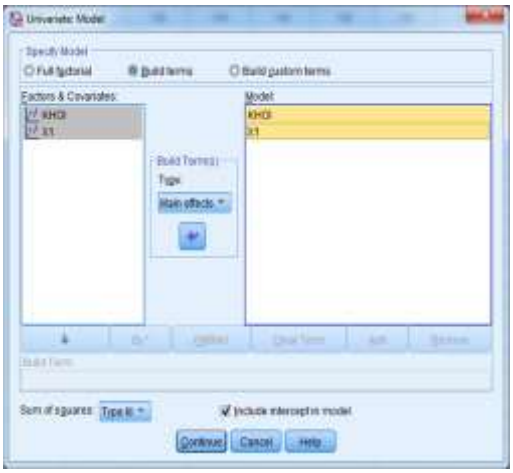
Bước 1. Nhập số liệu vào bảng tính SPSS. Số liệu được nhập vào bảng tính SPSS theo ba cột. Cột 1 nhập biến H theo thứ tự từ nghiệm thức 1 đến 5. Cột 2 nhập mã biến độc lập (X_i). Cột 3 nhập số lần lặp lại của mỗi nghiệm thức (Hình 10.12).

Hình 10.12. Chiều cao của Dầu song nàng 8 tháng tuổi.

Bước 2. Xây dựng mô hình tuyến tính tổng quát: $H = \beta_0 + \beta_1 X_1 + \beta_2$*Khối + ε. Để đạt mục đích này, từ menu chính, chọn Analyze > GLM: Univariate. Đến đây chọn biến H và chuyển nó vào Dependent Variable. Tương tự, chọn biến Khối và X_1 và chuyển vào Fixed Factor (Hình 10.13). Để thực hiện các phân tích thống kê, từ cửa sổ Univariate ở Hình 10.13, chọn Model. Kế đến đánh dấu mục Custom và chuyển biến Khối và X_1 từ ô Factor & Covariate sang ô Model (Hình 10.14). Tiếp đến chọn Continue ở Hình 10.14; sau đó chọn Post Hoc ở Hình 10.13. Khi xuất hiện cửa sổ Univariate Post Hoc Multiple Comparisons for Observed Means, chọn nhân tố X_1 trong ô Factor và chuyển nó sang ô Post Hoc Tests for. Tiếp theo chọn kiểm định Tukey's Test. Mức tin cậy mặc định là 0,95. Sau đó kiểm định thuần nhất H của Dầu song nàng giữa 5 cấp độ tàn che. Để đạt mục đích này, trước hết chọn Option ở Hình 10.13; sau đó chọn Homogeneity Test và Descriptive Statistics (Hình 10.16). Cuối cùng chọn Continue và chọn OK ở Hình 10.13 để nhận kết quả phân tích GLM (Bảng 10.2).

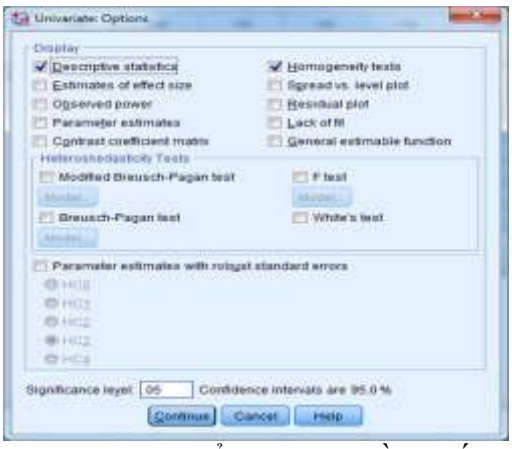


Hình 10.13. Số liệu để xây dựng mô hình: H = $\beta_0 + \beta_1 X_1 + \beta_2$*Khối.



Hình 10.14. Xây dựng mô hình: H = $\beta_0 + \beta_1 X_1 + \beta_2$*Khối.

Hình 10.15. So sánh H của Dầu song nàng theo 5 cấp độ tàn che.



Hình 10.16. Kiểm định thuần nhất H của Dầu song nàng theo 5 độ tàn che.

Bảng 10.2a. Kiểm định ảnh hưởng của độ tàn che đến sinh trưởng chiều cao của cây con Dầu song nàng 8 tháng tuổi. Mô hình H = $\beta_0 + \beta_1 X_1 + \beta_2$*Khối.

Source	SS_{III}	Df	MS	F	Sig.
(1)	(2)	(3)	(4)	(5)	(6)
Corrected Model	648.326	7	92.618	3.072	0.042
Intercept	145419.458	1	145419.458	4823.519	0.000
Khối	232.774	3	77.591	2.574	0.103
X_1	415.552	4	103.888	3.446	0.043
Error	361.776	12	30.148		
Total	146429.560	20			
Corrected Total	1010.102	19			

Bảng 10.2b. Kiểm định ảnh hưởng của độ tàn che đến sinh trưởng chiều cao của cây con Dầu song nàng 8 tháng tuổi ở vườn ươm. Mô hình: $H = \beta_0 + \beta_1X_1 + \varepsilon$.

Source	SS_{III}	Df	MS	F	Sig.
(1)	(2)	(3)	(4)	(5)	(6)
Corrected Model	415.552	4	103.888	2.621	0.077
Intercept	145419.458	1	145419.458	3668.811	0.000
X_1	415.552	4	103.888	2.621	0.077
Error	594.550	15	39.637		
Total	146429.560	20			
Corrected Total	1010.102	19			

Giải thích và nhận xét kết quả. Giả thuyết H_0: Độ tàn che ảnh hưởng đồng đều đến sinh trưởng H của cây con Dầu song nàng 8 tháng tuổi. Bởi vì P – value đối với Khối = 0,103 > 0,05 (Bảng 10.2a), nên Khối là đồng nhất. Trong trường hợp này, mô hình tuyến tính tổng quát được viết gọn ở dạng: $H = \beta_0 + \beta_1X_1 + \varepsilon$. Kết quả phân tích phương sai (Bảng 10.2b) cho thấy độ tàn che ảnh hưởng đồng đều (P = 0,077 > 0,05) đến sinh trưởng H của cây con Dầu song nàng 8 tháng tuổi.

10.3.1.3. Kiểu thí nghiệm bình phương la tinh

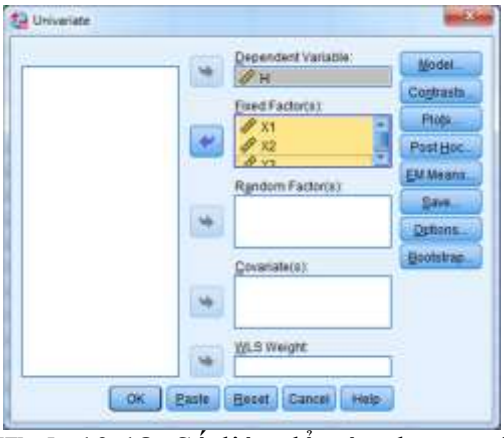
Giả thiết phân tích ảnh hưởng của 6 hàm lượng phân Super lân (X% = 0, 1, 2, 3, 4, 5 so với trọng lượng bầu) đến sinh trưởng chiều cao của Gõ đỏ 3 tháng tuổi. Kết quả thí nghiệm được ghi lại ở Phụ lục 9. Để giải quyết vấn đề đặt ra, các bước xử lý số liệu như dưới đây.

Bước 1. Nhập số liệu vào bảng tính SPSS. Số liệu được nhập theo 4 cột. Cột 1 là H theo thứ tự từ nghiệm thức 1 đến 6. Cột 2 nhập mã hàm lượng phân Super lân (X_1 = 0, 1, 2, 3, 4, 5%). Cột 3 và 4 nhập tương ứng khối hàng (X_2) và khối cột (X_3) (Hình 10.17).

Hình 10.17. Chiều cao của Gõ đỏ 3 tháng tuổi.

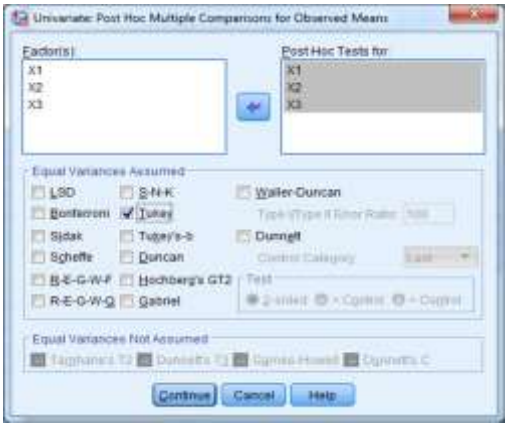
Bước 2. Xây dựng mô hình tuyến tính tổng quát: $H = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \varepsilon$; trong đó β_0 = điểm chặn (Intercept), X_1 = Hàm lượng phân Super lân, X_2 = Khối cột, X_3 = Khối hàng, β_1, β_2 và β_3 = hệ số phản ánh ảnh hưởng của X_1, X_2 và

X_3, ε = sai số ngẫu nhiên. Để đạt mục đích này, từ menu chính, chọn Analyze > GLM: Univariate. Đến đây chọn biến H và chuyển vào Dependent Variable. Tương tự, chọn Hàng và Cột và chuyển vào Fixed Factor (Hình 10.18). Để thực hiện các phân tích thống kê, từ cửa sổ ở Hình 10.18, chọn Model. Kế đến đánh dấu mục Custom và chuyển biến X_1, X_2 và X_3 từ ô Factor & Covariate sang ô Model (Hình 10.19). Tiếp đến chọn Continue ở Hình 10.19; sau đó chọn Post Hoc ở Hình 10.18. Khi xuất hiện cửa sổ Univariate Post Hoc Multiple Comparisons for Observed Means, chọn nhân tố X_1, X_2 và X_3 trong ô Factor và chuyển chúng sang ô Post Hoc Tests for. Tiếp theo chọn kiểm định Tukey's Test (Hình 10.20). Mức tin cậy mặc định là 0,95. Sau đó kiểm định thuần nhất H của Gõ đỏ giữa 6 hàm lượng phân Super lân. Để đạt mục đích này, tại cửa sổ Univariate: Option, chọn Homogeneity Test và Descriptive Statistics (Hình 10.21). Cuối cùng chọn Continue và chọn OK ở Hình 10.18 để nhận kết quả phân tích GLM (Bảng 10.3).

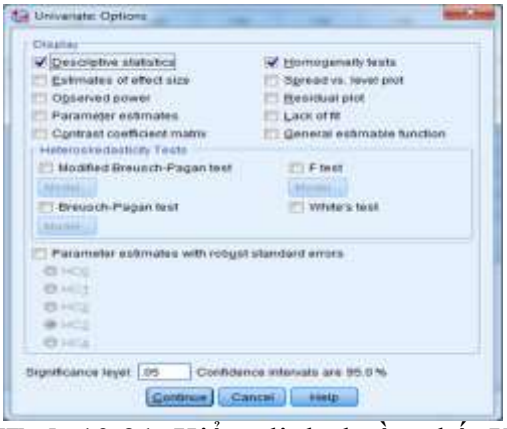


Hình 10.18. Số liệu để xây dựng mô hình: $H = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3$.

Hình 10.19. Xây dựng mô hình: $H = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3$.



Hình 10.20. So sánh H của Gõ đỏ theo 6 mức Super lân.



Hình 10.21. Kiểm định thuần nhất H của Gõ đỏ theo 6 mức Super lân.

Bảng 10.3a. Kiểm định ảnh hưởng của Super lân đến sinh trưởng chiều cao của cây con Gõ đỏ 3 tháng tuổi ở vườn ươm. Mô hình $H = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3$.

Source	SS_{III}	Df	MS	F	Sig.
(1)	(2)	(3)	(4)	(5)	(6)
Corrected Model	27047.0	15	1803.133	24.466	0.000
Intercept	59049.0	1	59049.000	801.208	0.000
Super lân (X_1)	24442.0	5	4888.400	66.328	0.000
Khối hàng (X_2)	2028.0	5	405.600	5.503	0.002
Khối cột (X_3)	577.0	5	115.400	1.566	0.215
Error	1474.0	20	73.700		
Total	87570.0	36			
Corrected Total	28521.0	35			

Bảng 10.3b. Kiểm định ảnh hưởng của Super lân đến sinh trưởng chiều cao của cây con Gõ đỏ 3 tháng tuổi ở vườn ươm. Mô hình: $H = \beta_0 + \beta_1X_1 + \varepsilon$.

Source	SS_{III}	Df	MS	F	Sig.
(1)	(2)	(3)	(4)	(5)	(6)
Corrected Model	24442.000	5	4888.400	35.953	0.000
Intercept	59049.000	1	59049.000	434.290	0.000
Super lân (X)	24442.000	5	4888.400	35.953	0.000
Error	4079.000	30	135.967		
Total	87570.000	36			
Corrected Total	28521.000	35			

Bảng 10.3c. Phân nhóm Super lân ảnh hưởng đồng đều đến sinh trưởng chiều cao của cây con Gõ đỏ 3 tháng tuổi ở vườn ươm.

Super lân (%)	N	Nhóm thuần nhất:			
		1	2	3	4
5	6	5.000			
4	6	15.333	15.333		
3	6		34.333	34.333	
2	6			43.500	
1	6				65.500
0	6				79.333
Sig.		0.645	0.081	0.749	0.337

Giải thích và nhận xét kết quả. Giả thuyết H_0: Hàm lượng phân Super lân từ 0 – 5% ảnh hưởng đồng đều đến sinh trưởng H của cây con Gõ đỏ 3 tháng tuổi. Bởi vì $P < 0,001$ (Bảng 10.13a), nên sự thay đổi hàm lượng phân Super lân từ 0 – 5% ảnh hưởng rõ rệt đến sinh trưởng H của cây con Gõ đỏ 3 tháng tuổi. Bảng 10.13b là mô hình không bao gồm yếu tố Cột và Hàng. Từ số liệu ở Bảng 10.13c cho thấy, bón phân Super lân từ 0 – 1% so với trọng lượng bầu đảm bảo cho cây con Gõ đỏ 3 tháng tuổi sinh trưởng tốt nhất.

10.3.2. Phân tích phương sai hai nhân tố

10.3.2.1. Kiểu thí nghiệm hoàn toàn ngẫu nhiên

Giả sử cần thăm dò sinh trưởng chiều cao (H, cm) của cây con Dầu song nàng trong giai đoạn 6 tháng tuổi ở vườn ươm dưới ảnh hưởng của 3 mức phân đạm (X_1 = 1, 2 và 3% so với trọng lượng ruột bầu); mỗi mức phân đạm được sổ sung thêm 3 mức Super lân (X_2 = 0,5, 1,0, 1,5% so với trọng lượng ruột bầu). Mỗi lô thí nghiệm sử dụng 15 cây 1 tuần tuổi đồng đều về hình thái và sức sống. Kết quả thí nghiệm được ghi lại ở Phụ lục 10. Các giả thuyết: (1) Các mức đạm ảnh hưởng đồng đều đến sinh trưởng H của Dầu song nàng; (2) Các mức Super lân ảnh hưởng đồng đều đến sinh trưởng H của Dầu song nàng; (3) Sự phối hợp của các mức đạm và Super lân ảnh hưởng đồng đều đến sinh trưởng H của Dầu song nàng. Để kiểm định các giả thuyết này, các bước xử lý như dưới đây.

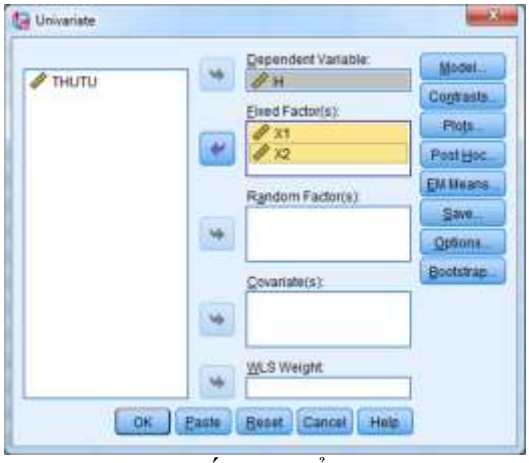
Bước 1. Nhập số liệu vào bảng tính SPSS. Số liệu được nhập theo 4 cột (Hình 10.22). Cột 1 nhập biến H. Cột 2 nhập thứ tự cây được đo H. Cột 3 và 4 nhập tương ứng biến X_1 (Đạm) và biến X_2 (Super lân).

Hình 10.22. Chiều cao của cây con Dầu song nàng 6 tháng tuổi.

Bước 2. Xây dựng mô hình tuyến tính tổng quát: $H = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2 + \varepsilon$; trong đó β_0 = điểm chặn (Intercept), X_1 = Đạm, X_2 = Super lân, X_1*X_2 = tương tác giữa hai yếu tố X_1 và X_2, β_1, β_2 và β_3 = hệ số phản ánh ảnh hưởng của X_1, X_2 và X_1X_2, ε = sai số ngẫu nhiên. Để đạt mục đích này, từ menu chính, chọn Analyze > GLM: Univariate. Đến đây chọn biến H và chuyển vào Dependent Variable. Tương tự, chọn X_1 và X_2 và chuyển vào Fixed Factor (Hình 10.23).

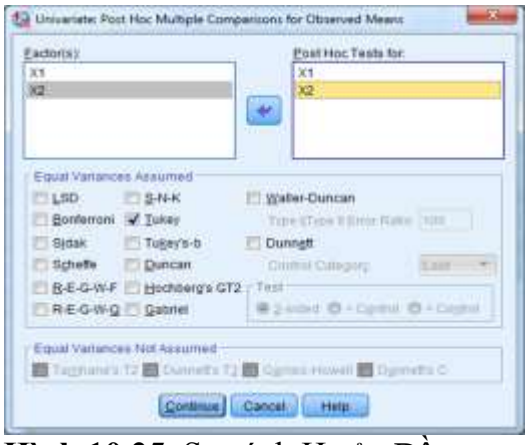
Để thực hiện các phân tích thống kê, từ cửa sổ Univariate ở Hình 10.23, chọn Model. Kế đến đánh dấu mục Custom và chuyển biến X_1 và X_2 từ ô Factor & Covariate sang ô Model (Hình 10.24). Tiếp đến chọn Continue ở Hình 10.24; sau đó chọn Post Hoc ở Hình 10.23. Khi xuất hiện cửa sổ Univariate Post Hoc Multiple Comparisons for Observed Means, chọn X_1 và X_2 trong ô Factor và chuyển nó sang ô Post Hoc Tests for. Tiếp theo chọn kiểm định Tukey's Test (Hình 10.25). Mức tin

cậy mặc định là 0,95. Sau đó kiểm định thuần nhất H của cây con Dầu song nàng giữa 3 hàm lượng đạm và Super lân. Để đạt mục đích này, tại cửa sổ Univariate: Option, chọn Homogeneity Test và Descriptive Statistics (Hình 10.26). Cuối cùng chọn Continue và chọn OK ở Hình 10.23 để nhận kết quả (Bảng 10.4).

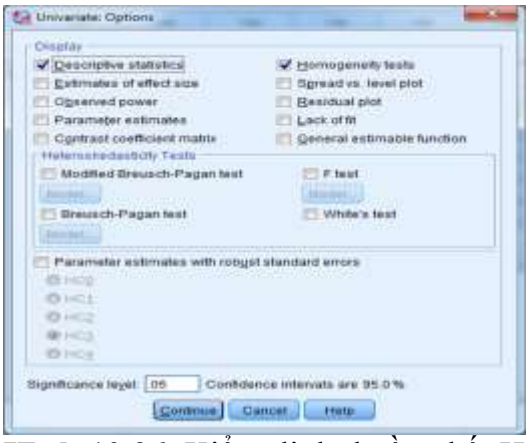


Hình 10.23. Số liệu để xây dựng mô hình: $H = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2$.

Hình 10.24. Xây dựng mô hình: $H = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2$.



Hình 10.25. So sánh H của Dầu song nàng 6 tháng tuổi theo ba mức đạm và Super lân.



Hình 10.26. Kiểm định thuần nhất H của Dầu song nàng 6 tháng tuổi theo ba mức đạm và Super lân.

Giải thích và nhận xét kết quả. Giả thuyết H_0: Sự phối hợp giữa hàm lượng phân đạm (X_1 = 1 – 3%) và Super lân (X_2 = 0,5, 1,0, 1,5%) ảnh hưởng đồng đều đến sinh trưởng H của cây con Dầu song nàng 6 tháng tuổi. Bởi vì $P < 0,05$ (Bảng 10.4a), nên sự thay đổi hàm lượng đạm từ 1 - 3% và Super lân từ 0,5 – 1,5% ảnh hưởng rõ rệt đến sinh trưởng H của cây con Dầu song nàng 6 tháng tuổi. Bởi vì $P = 1,0 > 0,05$ (Bảng 10.4a), nên sự phối hợp giữa 2 nhân tố X_1 và X_2 ảnh hưởng không rõ rệt đến sinh trưởng H của cây con Dầu song nàng 6 tháng tuổi. Từ số liệu ở Bảng

10.4c cho thấy, bón phân đạm từ 1 – 3% so với trọng lượng bầu đảm bảo cho cây con Gõ đỏ 3 tháng tuổi sinh trưởng tốt nhất.

Bảng 10.4a. Kiểm định ảnh hưởng của phân đạm và Super lân đến sinh trưởng chiều cao của cây con Dầu song nàng 6 tháng tuổi.

Source	SS_{III}	Df	MS	F	Sig.
(1)	(2)	(3)	(4)	(5)	(6)
Corrected Model	2829.161	8	353.645	37.509	0.000
Intercept	487765.601	1	487765.601	51734.980	0.000
Đạm (X_1)	2764.800	2	1382.400	146.625	0.000
Super lân (X_2)	64.361	2	32.181	3.413	0.036
X_1*X_2	0.000	4	0.000	0.000	1.000
Error	1187.948	126	9.428		
Total	491782.710	135			
Corrected Total	4017.109	134			

Bảng 10.4b. Phân nhóm đạm ảnh hưởng đồng đều đến sinh trưởng chiều cao của cây con Dầu song nàng 6 tháng tuổi ở vườn ươm.

Đạm (%)	N	Nhóm thuần nhất	
		1	2
1	45	56.91	
2	45	56.91	
3	45		66.51
Sig.		1.000	1.000

10.3.2.2. Kiểu thí nghiệm theo khối hoàn toàn ngẫu nhiên

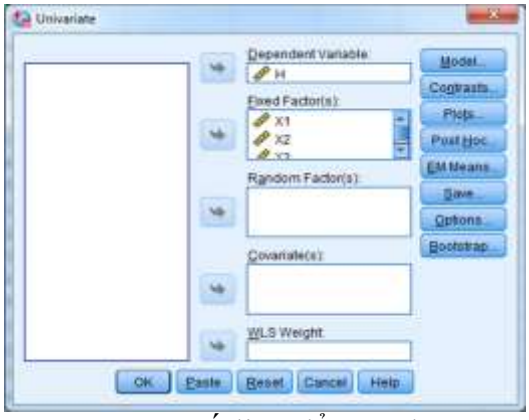
Giả sử cần phân tích ảnh hưởng của 3 mức đạm (X_1 = 1, 2 và 3% so với trọng lượng bầu) và 3 mức Super lân (X_2 = 0,5, 1,0, 1,5% so với trọng lượng bầu) đến sinh trưởng chiều cao (H, cm) của cây con Sao đen trong giai đoạn 6 tháng tuổi ở vườn ươm. Thí nghiệm bao gồm 9 nghiệm thức; mỗi nghiệm thức được lặp lại 5 lần (Khối). Tổng cộng có 45 lô thí nghiệm. Mỗi lô thí nghiệm sử dụng 49 cây đồng đều về hình thái và sức sống. Kết quả thí nghiệm được ghi lại ở Phụ lục 11. Các giả thuyết: (1) Các mức đạm ảnh hưởng đồng đều đến sinh trưởng H của Sao đen; (2) Các mức Super lân ảnh hưởng đồng đều đến sinh trưởng H của Sao đen; (3) Sự phối hợp của các mức đạm với các mức Super lân ảnh hưởng đồng đều đến sinh trưởng H của Sao đen. Để kiểm định các giả thuyết trên, các bước xử lý như dưới đây.

Bước 1. Nhập số liệu vào bảng tính SPSS. Số liệu được nhập theo 4 cột (Hình 10.27). Cột 1 nhập H (cm). Cột 2 và 3 nhập tương ứng nhân tố đạm (X_1) và Super lân (X_2). Cột 4 nhập số lần lặp của mỗi nghiệm thức (Khối = X_3).

Bước 2. Xây dựng mô hình tuyến tính tổng quát: $H = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2 + \beta_4*X_3 + \varepsilon$; trong đó β_0 = điểm chặn (Intercept), X_1 = Đạm, X_2 = Super lân, X_1*X_2 = tương tác giữa hai yếu tố X_1 và X_2, X_3 = Khối, β_1, β_2 và β_3 = hệ số phản ánh ảnh hưởng của X_1, X_2 và X_1X_2, ε = sai số ngẫu nhiên. Để đạt mục đích

này, từ menu chính, chọn Analyze > GLM: Univariate. Đến đây chọn biến H và chuyển vào Dependent Variable. Tương tự, chọn X_1, X_2 và X_3 và chuyển vào Fixed Factor (Hình 10.28).

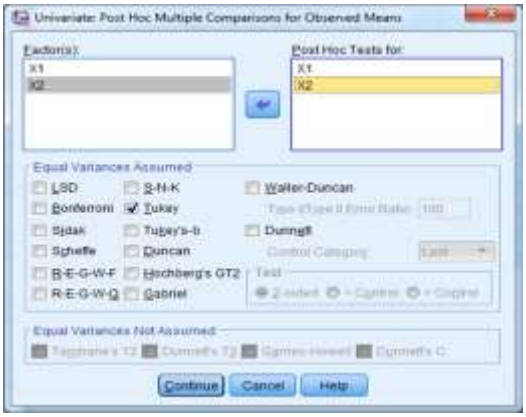
Hình 10.27. Chiều cao của cây con Sao đen 6 tháng tuổi.



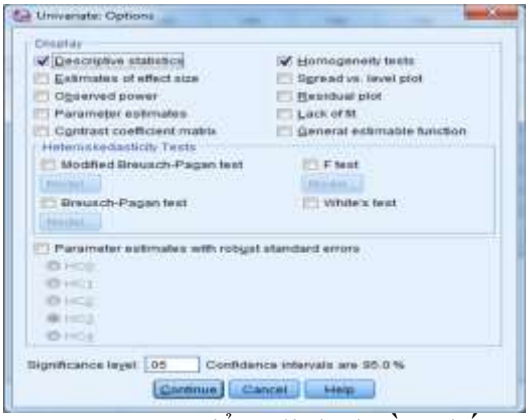
Hình 10.28. Số liệu để xây dựng mô hình: $H = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2 + \beta_4{*}X_3$.

Hình 10.29. Xây dựng mô hình: $H = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2 + \beta_4{*}X_3$.

Để thực hiện các phân tích thống kê, từ cửa sổ Univariate ở Hình 10.28, chọn Model. Kế đến đánh dấu mục Custom và chuyển biến X_1, X_2 và X_3 từ ô Factor & Covariate sang ô Model (Hình 10.29). Tiếp đến chọn Continue ở Hình 10.29; sau đó chọn Post Hoc ở Hình 10.28. Khi xuất hiện cửa sổ Univariate Post Hoc Multiple Comparisons for Observed Means, chọn X_1 và X_2 trong ô Factor và chuyển nó sang ô Post Hoc Tests for. Tiếp theo chọn kiểm định Tukey's Test (Hình 10.30). Mức tin cậy mặc định là 0,95. Sau đó kiểm định thuần nhất H của cây con Sao đen giữa 3 hàm lượng đạm và Super lân. Để đạt mục đích này, tại cửa sổ Univariate: Option, chọn Homogeneity Test và Descriptive Statistics (Hình 10.31). Cuối cùng chọn Continue ở Hình 10.31 và chọn OK Hình 10.28 để nhận kết quả phân tích GLM (Bảng 10.5).



Hình 10.30. So sánh H của Sao đen 6 tháng tuổi theo ba mức đạm và Super lân.



Hình 10.31. Kiểm định thuần nhất H của Sao đen 6 tháng tuổi theo ba mức đạm và Super lân.

Bảng 10.5a. Kiểm định ảnh hưởng của phân đạm và Super lân đến sinh trưởng chiều cao của cây con Sao đen 6 tháng tuổi.

Source	SS_{III}	Df	MS	F	Sig.
(1)	(2)	(3)	(4)	(5)	(6)
Corrected Model	10665.422	12	888.785	35.830	0.000
Intercept	333852.800	1	333852.800	13458.791	0.000
Đạm (X_1)	3040.533	2	1520.267	61.287	0.000
Super lân (X_2)	578.133	2	289.067	11.653	0.000
Khối (X_3)	257.422	4	64.356	2.594	0.055
X_1* X_2	6789.333	4	1697.333	68.426	0.000
Error	793.778	32	24.806		
Total	345312.000	45			
Corrected Total	11459.200	44			

Bảng 10.5b. Kiểm định ảnh hưởng của đạm và Super lân đến sinh trưởng chiều cao của cây con Sao đen 6 tháng tuổi ở vườn ươm. Trường hợp mô hình tuyến tính tổng quát ở dạng: $H = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2 + \varepsilon$.

Source	SS_{III}	Df	MS	F	Sig.
(1)	(2)	(3)	(4)	(5)	(6)
Corrected Model	10408.000	8	1301.000	44.555	0.000
Intercept	333852.800	1	333852.800	11433.315	0.000
Đạm (X_1)	3040.533	2	1520.267	52.064	0.000
Super lân (X_2)	578.133	2	289.067	9.900	0.000
X_1* X_2	6789.333	4	1697.333	58.128	0.000
Error	1051.200	36	29.200		
Total	345312.000	45			
Corrected Total	11459.200	44			

Bảng 10.5c. Phân nhóm đạm ảnh hưởng đồng đều đến sinh trưởng chiều cao của cây con Sao đen 6 tháng tuổi ở vườn ươm.

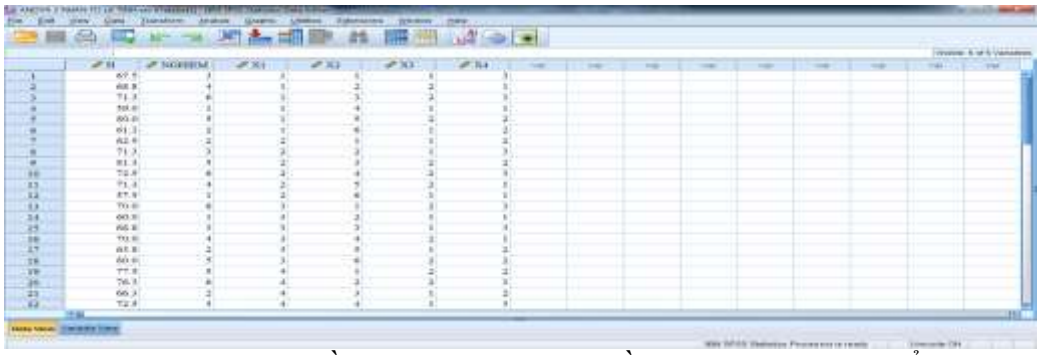
Đạm (%)	N	Phân nhóm thuần nhất		
		1	2	3
1	15	76.13		
3	15		86.00	
2	15			96.27
Sig.		1.000	1.000	1.000

Giải thích và nhận xét kết quả. Số liệu ở Bảng 10.5a cho thấy sự thay đổi hàm lượng phân đạm (X_1 = 1 – 3%) và Super lân (X_2 = 0,5, 1,0, 1,5%) ảnh hưởng rõ rệt (P < 0,01) đến sinh trưởng H của cây con Sao đen 6 tháng tuổi. Sự phối hợp giữa 2 nhân tố X_1 và X_2 cũng ảnh hưởng rõ rệt (P < 0,01) đến sinh trưởng H của cây con Sao đen 6 tháng tuổi. Bảng 10.5b là mô hình không bao gồm yếu tố Khối. Từ số liệu ở Bảng 10.5c cho thấy, bón phân đạm với tỷ lệ 2% so với trọng lượng bầu đảm bảo cho cây con Sao đen sinh trưởng tốt trong 6 tháng đầu ở vườn ươm.

10.3.2.3. Kiểu thí nghiệm bình phương la tinh

Giả thiết cần đánh giá ảnh hưởng của kích thước bầu và phân tổng hợp NPK đến sinh trưởng chiều cao của Dầu con rái 8 tháng tuổi ở vườn ươm. Kích thước bầu bao gồm 2 cấp: 15*25 cm và 20*30 cm. Phân tổng hợp NPK bao gồm 3 mức: 1, 2 và 3% so với trọng lượng bầu. Tổng số 2*3 = 6 nghiệm thức. Mỗi nghiệm thức được bố trí lặp lại 6 lần. Kết quả thí nghiệm được ghi lại ở Phụ lục 12. Các giả thuyết: (1) Sự thay đổi kích thước bầu ảnh hưởng đồng đều đến sinh trưởng H của Dầu con rái; (2) Các mức phân tổng hợp NPK ảnh hưởng đồng đều đến sinh trưởng H của Dầu con rái; (3) Sự phối hợp của các kích thước bầu với các mức phân tổng hợp NPK ảnh hưởng đồng đều đến sinh trưởng H của Dầu con rái. Để kiểm định các giả thuyết trên, các bước xử lý như dưới đây.

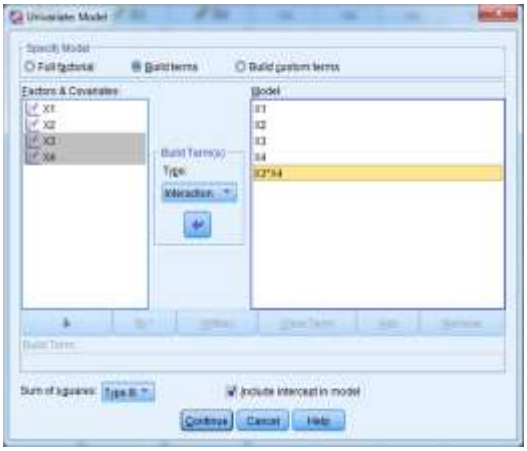
Bước 1. Nhập số liệu vào bảng tính SPSS. Số liệu được nhập theo 4 cột (Hình 10.32). Cột 1 là H (cm). Cột 2 là nghiệm thức. Cột 3 và 4 nhập tương ứng khối hàng (X_1) và khối cột (X_2). Cột 5 và 6 nhập tương ứng kích thước bầu (X_3) và phân tổng hợp NPK (X_4).



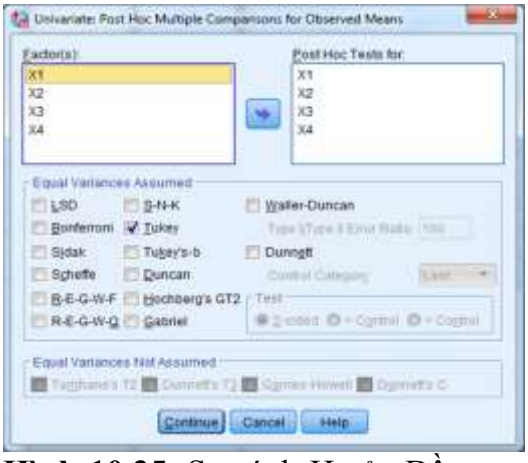
Hình 10.32. Chiều cao của cây con Dầu con rái 8 tháng tuổi.

Bước 2. Xây dựng mô hình tuyến tính tổng quát: $H = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4*X_4 + \beta_5*X_3*X_4 + \varepsilon$; trong đó β_0 = điểm chặn (Intercept), X_1 = Khối hàng, X_2 = Khối cột, X_3 = Kích thước bầu, X_4 = Phân tổng hợp NPK, X_3*X_4 = tương tác giữa hai yếu tố X_3 và X_4, β_1, β_2, β_3 và β_4 = hệ số phản ánh ảnh hưởng của X_1, X_2, X_3, X_4 và X_3X_4, ε = sai số ngẫu nhiên. Để đạt mục đích này, từ menu chính, chọn Analyze > GLM: Univariate. Đến đây chọn biến H và chuyển vào Dependent Variable. Tương tự, chọn X_1, X_2, X_3 và X_4 và chuyển vào Fixed Factor (Hình 10.33).

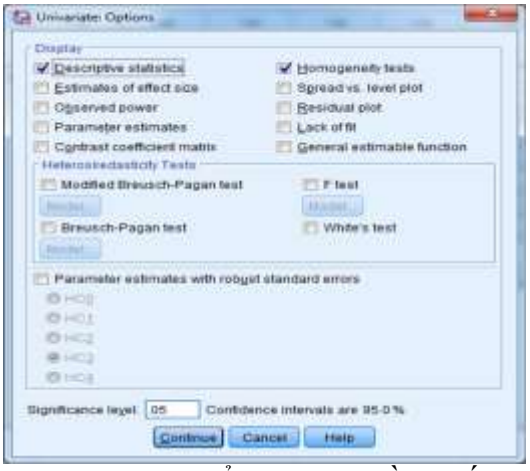
Hình 10.33. Số liệu để xây dựng mô hình tuyến tính tổng quát theo kiểu thí nghiệm bình phương la tinh.



Hình 10.34. Xây dựng mô hình tuyến tính tổng quát theo kiểu thí nghiệm bình phương la tinh.



Hình 10.35. So sánh H của Dầu con rái 8 tháng tuổi theo 2 cấp bầu và 3 mức phân tổng hợp NPK.



Hình 10.36. Kiểm định thuần nhất H của Dầu con rái 8 tháng tuổi theo 2 cấp bầu và 3 mức phân NPK.

Để thực hiện các phân tích thống kê, từ cửa sổ Univariate ở Hình 10.33, chọn Model. Kế đến đánh dấu mục Custom và chuyển biến X_1, X_2, X_3, X_4 và X_3X_4 từ ô Factor & Covariate sang ô Model (Hình 10.34). Tiếp đến chọn Continue ở Hình

10.34; sau đó chọn Post Hoc ở Hình 10.33. Khi xuất hiện cửa sổ Univariate Post Hoc Multiple Comparisons for Observed Means, chọn X_1, X_2, X_3, X_4 trong ô Factor và chuyển chúng sang ô Post Hoc Tests for. Tiếp theo chọn kiểm định Tukey's Test (Hình 10.35). Mức tin cậy mặc định là 0,95. Sau đó kiểm định thuần nhất H của cây con Dầu con rái theo 4 yếu tố (X_1, X_2, X_3, X_4). Để đạt mục đích này, tại cửa sổ Univariate: Option, chọn Homogeneity Test và Descriptive Statistics (Hình 10.36). Cuối cùng chọn Continue ở Hình 10.36 và chọn OK ở Hình 10.33 để nhận kết quả phân tích GLM (Bảng 10.6).

Bảng 10.6a. Kiểm định ảnh hưởng của kích thước bầu và phân tổng hợp NPK đến sinh trưởng chiều cao của Dầu con rái 8 tháng tuổi.

Source	SS_{III}	Df	MS	F	Sig.
(1)	(2)	(3)	(4)	(5)	(6)
Corrected Model	1851.784	15	123.452	40.236	0.000
Intercept	170280.023	1	170280.023	55498.948	0.000
Khối hàng (X_1)	8.709	5	1.742	0.568	0.724
Khối cột (X_2)	137.573	5	27.515	8.968	0.000
Kích thước bầu (X_3)	755.334	1	755.334	246.184	0.000
Phân NPK (X_4)	712.500	2	356.250	116.112	0.000
X_3*X_4	237.669	2	118.834	38.731	0.000
Error	61.363	20	3.068		
Total	172193.170	36			
Corrected Total	1913.147	35			

Bảng 10.6b. Kiểm định ảnh hưởng của kích thước bầu và phân tổng hợp NPK đến sinh trưởng chiều cao của Dầu con rái 8 tháng tuổi. Trường hợp mô hình tuyến tính tổng quát ở dạng: $H = \beta_0 + \beta_3X_3 + \beta_4X_4 + \beta_5*X_3*X_4 + \varepsilon$.

Source	SS_{III}	Df	MS	F	Sig.
(1)	(2)	(3)	(4)	(5)	(6)
Corrected Model	1705.502[a]	5	341.100	49.281	0.000
Intercept	170280.023	1	170280.023	24601.607	0.000
Kích thước bầu (X_3)	755.334	1	755.334	109.129	0.000
Phân NPK (X_4)	712.500	2	356.250	51.470	0.000
X_3 * X_4	237.669	2	118.834	17.169	0.000
Error	207.645	30	6.921		
Total	172193.170	36			
Corrected Total	1913.147	35			

Giải thích và nhận xét kết quả. Số liệu ở Bảng 10.6a cho thấy sự thay đổi kích thước bầu (X_3) và hàm lượng phân tổng hợp NPK (X_4) ảnh hưởng rõ rệt ($P < 0,01$) đến sinh trưởng H của cây con Dầu con rái 8 tháng tuổi. Sự phối hợp giữa 2 nhân tố X_3 và X_4 cũng ảnh hưởng rõ rệt ($P < 0,01$) đến sinh trưởng H của cây con Dầu con rái 8 tháng tuổi. Bảng 10.6b là mô hình không bao gồm yếu tố X_1 và X_2.

Từ số liệu ở Bảng 10.6c cho thấy, hàm lượng phân tổng hợp NPK = 2 - 3% so với trọng lượng ruột bầu đảm bảo cho cây con Dầu con rái sinh trưởng tốt trong 8 tháng đầu ở vườn ươm.

Bảng 10.6c. Phân nhóm phân tổng hợp NPK ảnh hưởng đồng đều đến sinh trưởng chiều cao của cây con Dầu con rái 8 tháng tuổi ở vườn ươm.

Phân NPK (%)	N	Phân nhóm thuần nhất	
		1	2
1	12	62.525	
3	12		71.275
2	12		72.525
Sig.		1.000	0.483

10.3.2.4. Kiểu thí nghiệm lô chính – lô phụ

Giả thiết cần đánh giá ảnh hưởng của ba phương thức xử lý đất đến sinh trưởng chiều cao của ba loài (Dầu con rái, Sao đen, Vên vên) 8 tháng tuổi. Mỗi thí nghiệm được lặp lại 4 lần. Ở đây chọn lô chính là 3 phương thức xử lý đất, còn lô phụ là 3 loài cây. Kết quả thí nghiệm được ghi lại ở Phụ lục 13. Giả thuyết H_0: Phương thức xử lý đất ảnh hưởng đồng đều đến sinh trưởng H của ba loài (Dầu con rái, Sao đen, Vên vên) trong giai đoạn 8 tháng tuổi. Để kiểm định các giả thuyết trên, các bước xử lý như dưới đây.

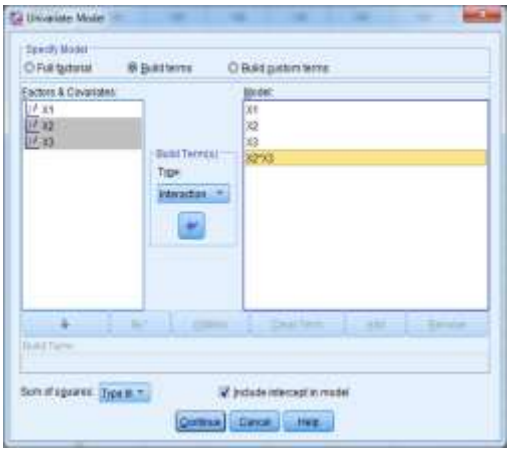
Bước 1. Nhập số liệu vào bảng tính SPSS. Số liệu được nhập theo 4 cột (Hình 10.37). Cột 1 là H (cm). Cột 2 là khối (X_1). Cột 3 là phương thức xử lý đất (X_2). Cột 4 là loài cây (X_3).

Hình 10.37. Chiều cao của Dầu con rái, Sao đen, Vên vên 8 tháng tuổi.

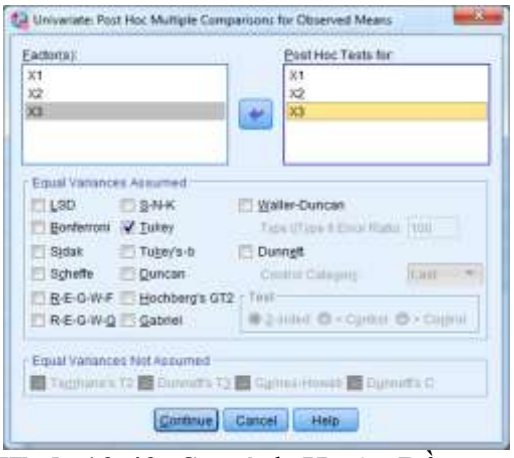
Bước 2. Xây dựng mô hình tuyến tính tổng quát: $H = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4*X_2*X_3 + \varepsilon$; trong đó β_0 = điểm chặn (Intercept), X_1 = Khối, X_2 = Phương thức xử lý đất, X_3 = Loài cây, X_2*X_3 = Tương tác giữa hai yếu tố X_2 và X_3, β_1, β_2, β_3 và β_4 = hệ số phản ánh ảnh hưởng của X_1, X_2, X_3 và X_2X_3, ε = sai số ngẫu nhiên. Để đạt mục đích này, từ menu chính, chọn Analyze > GLM: Univariate. Đến đây chọn biến H và chuyển vào Dependent Variable. Tương tự, chọn X_1, X_2 và X_3 và

chuyển vào Fixed Factor (Hình 10.38). Để thực hiện các phân tích thống kê, từ cửa sổ Univariate ở Hình 10.38, chọn Model. Kế đến đánh dấu mục Custom và chuyển biến X_1, X_2, X_3 và X_2X_3 từ ô Factor & Covariate sang ô Model (Hình 10.39). Tiếp đến chọn Continue ở Hình 10.39; sau đó chọn Post Hoc ở Hình 10.38. Khi xuất hiện cửa sổ Univariate Post Hoc Multiple Comparisons for Observed Means, chọn X_1, X_2, X_3 trong ô Factor và chuyển chúng sang ô Post Hoc Tests for. Tiếp theo chọn kiểm định Tukey's Test (Hình 10.40). Mức tin cậy mặc định là 0,95. Sau đó kiểm định thuần nhất H của loài cây theo nhân tố X_2. Để đạt mục đích này, tại cửa sổ Univariate: Option, chọn Homogeneity Test và Descriptive Statistics (Hình 10.41). Cuối cùng chọn Continue ở Hình 10.41 và chọn OK ở Hình 10.38 để nhận kết quả phân tích GLM (Bảng 10.7).

Hình 10.38. Số liệu để xây dựng mô hình tuyến tính tổng quát theo kiểu thí nghiệm lô chính – lô phụ.



Hình 10.39. Xây dựng mô hình tuyến tính tổng quát theo kiểu thí nghiệm lô chính – lô phụ.



Hình 10.40. So sánh H của Dầu con rái, Sao đen, Vên vên 8 tháng tuổi theo 3 phương thức xử lý đất.

Hình 10.41. Kiểm định thuần nhất H của Dầu con rái, Sao đen, Vên vên theo 3 phương thức xử lý đất.

Bảng 10.7a. Kiểm định ảnh hưởng của phương thức xử lý đất đến sinh trưởng chiều cao của ba loài (Dầu con rái, Sao đen, Vên vên) 8 tháng tuổi.

Source	SS_{III}	Df	MS	F	Sig.
(1)	(2)	(3)	(4)	(5)	(6)
Corrected Model	2992.361	11	272.033	22.161	0.000
Intercept	927048.028	1	927048.028	75520.413	0.000
Khối (X_1)	61.639	3	20.546	1.674	0.199
Xử lý đất (X_2)	476.722	2	238.361	19.418	0.000
Loài cây gỗ (X_3)	997.389	2	498.694	40.625	0.000
X_2*X_3	1456.611	4	364.153	29.665	0.000
Error	294.611	24	12.275		
Total	930335.000	36			
Corrected Total	3286.972	35			

Bảng 10.7b. Kiểm định ảnh hưởng của phương thức xử lý đất đến sinh trưởng chiều cao của ba loài (Dầu con rái, Sao đen, Vên vên) 8 tháng tuổi. Mô hình: $H = \beta_0 + \beta_2X_2 + \beta_3X_3 + \beta_4*X_2*X_3 + \varepsilon$.

Source	SS_{III}	Df	MS	F	Sig.
(1)	(2)	(3)	(4)	(5)	(6)
Corrected Model	2930.722[a]	8	366.340	27.765	0.000
Intercept	927048.028	1	927048.028	70260.482	0.000
Xử lý đất (X_2)	476.722	2	238.361	18.065	0.000
Loài cây gỗ (X_3)	997.389	2	498.694	37.796	0.000
X_2*X_3	1456.611	4	364.153	27.599	0.000
Error	356.250	27	13.194		
Total	930335.000	36			
Corrected Total	3286.972	35			

Bảng 10.7c. Phân nhóm phương thức xử lý đất ảnh hưởng đồng đều đến sinh trưởng chiều cao của ba loài (Dầu con rái, Sao đen, Vên vên) 8 tháng tuổi.

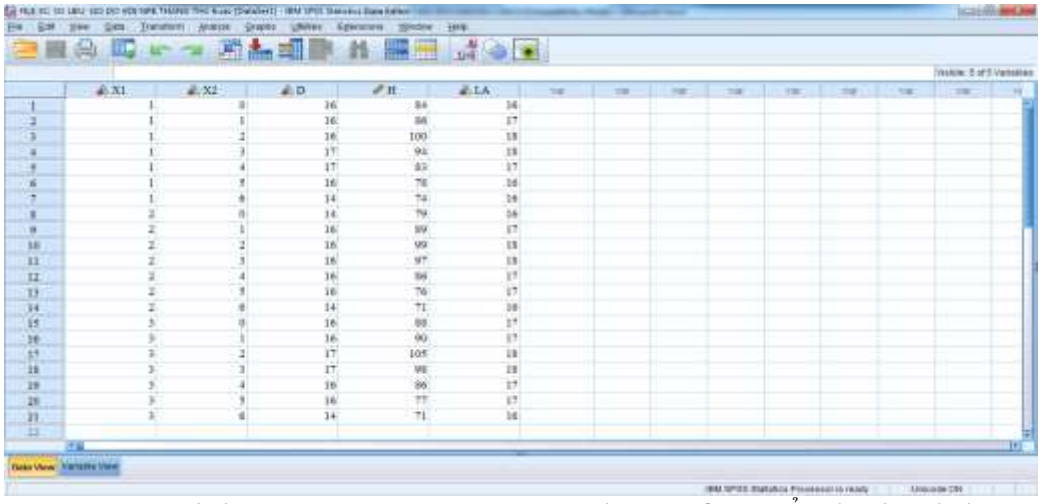
Xử lý đất	N	Nhóm thuần nhất	
		1	2
3	12	155.42	
2	12		162.17
1	12		163.83
Sig.		1.000	0.485

Giải thích và nhận xét kết quả. Số liệu ở Bảng 10.7a cho thấy phương thức xử lý đất (X_2) ảnh hưởng rõ rệt ($P < 0,01$) đến sinh trưởng chiều cao của ba loài (Dầu con rái, Sao đen, Vên vên) 8 tháng tuổi. Chiều cao của ba loài cây gỗ (X_3) khác nhau rõ rệt ($P < 0,01$). Bảng 10.7b là mô hình không bao gồm yếu tố X_1. Từ số liệu ở Bảng 10.7c cho thấy, phương thức xử lý đất 1 và 2 đảm bảo cho cây con của ba loài (Dầu con rái, Sao đen, Vên vên) sinh trưởng tốt trong 8 tháng đầu.

10.3.3. Phân tích phương sai một nhân tố với nhiều biến phụ thuộc

Giả sử cần phân tích ảnh hưởng của 7 mức phân tổng hợp NPK (X = 0, 1, 2,..., 6% so với trọng lượng bầu) đến sinh trưởng chiều cao (H, cm), D (cm) và số lá (La) của cây con Gõ đỏ trong giai đoạn 6 tháng tuổi ở vườn ươm. Thí nghiệm bao gồm 7 nghiệm thức và được bố trí theo khối ngẫu nhiên hoàn toàn. Mỗi nghiệm thức được lặp lại 3 lần (Khối). Tổng số 21 lô thí nghiệm. Mỗi lô thí nghiệm sử dụng 49 cây đồng đều về hình thái và sức sống. Kết quả thí nghiệm được ghi lại ở Phụ lục 14. Giả thuyết H_0: Các mức phân tổng hợp NPK ảnh hưởng đồng đều đến sinh trưởng D, H và số lá của Gõ đỏ trong giai đoạn 6 tháng tuổi ở vườn ươm. Để kiểm định các giả thuyết trên, các bước xử lý như dưới đây.

Bước 1. Nhập số liệu vào bảng tính SPSS. Số liệu được nhập theo 5 cột (Hình 10.42). Cột 1 nhập khối (X_1). Cột 2 nhập hàm lượng phân tổng hợp NPK (X_2). Cột 3, 4 và 5 nhập tương ứng D, H và số lá của cây con Gõ đỏ.



Hình 10.42. Sinh trưởng của cây con Gõ đỏ 6 tháng tuổi dưới ảnh hưởng của phân tổng hợp NPK.

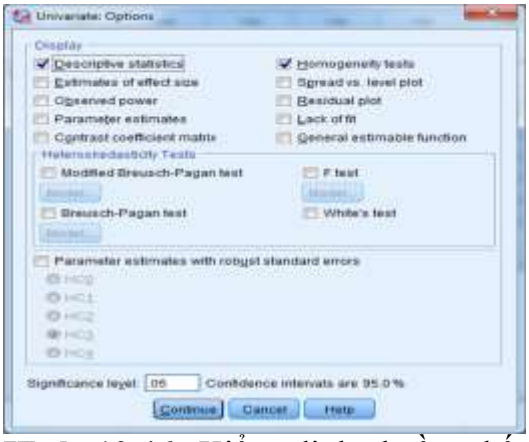
Bước 2. Xây dựng mô hình tuyến tính tổng quát: $Y_i = \beta_0 + \beta_1X_1 + \beta_2X_2 + \varepsilon$; trong đó Y = (D, H, số lá), β_0 = điểm chặn (Intercept), X_1 = Khối, X_2 = Phân tổng hợp NPK, β_1 và β_2 = hệ số phản ánh ảnh hưởng của X_1 và X_2, ε = sai số ngẫu nhiên. Để đạt mục đích này, từ menu chính, chọn Analyze > GLM: Multivariate. Đến đây chọn biến D, H và số lá và chuyển chúng vào Dependent Variable. Tương tự, chọn X_1 và X_2 và chuyển vào Fixed Factor (Hình 10.43). Để thực hiện các phân tích thống kê, từ cửa sổ Multivariate ở Hình 10.43, chọn Model. Kế đến đánh dấu mục Custom và chuyển X_1 và X_2 từ ô Factor & Covariate sang ô Model (Hình 10.44). Tiếp đến chọn Continue ở Hình 10.44; sau đó chọn Post Hoc ở Hình 10.43. Khi xuất hiện cửa sổ Post Hoc Multiple Comparisons for Observed Means, chọn X_1 và X_2 trong ô Factor và chuyển chúng sang ô Post Hoc Tests for. Tiếp theo chọn kiểm định Tukey's Test (Hình 10.45). Mức tin cậy mặc định là 0,95. Sau đó kiểm định thuần nhất D, H và số lá của cây con Gõ đỏ giữa 7 hàm lượng phân tổng hợp NPK. Để đạt mục đích này, tại cửa sổ Univariate: Option, chọn Homogeneity Test và

Descriptive Statistics (Hình 10.46). Cuối cùng chọn Continue ở Hình 10.46 và chọn OK ở Hình 10.43 để nhận kết quả phân tích GLM (Bảng 10.8).

Hình 10.43. Số liệu để xây dựng mô hình: $Y_i = \beta_0 + \beta_1X_1 + \beta_2X_2$.

Hình 10.44. Xây dựng mô hình: $Y_i = \beta_0 + \beta_1X_1 + \beta_2X_2$.

Hình 10.45. So sánh D, H và số lá của Gõ đỏ 6 tháng tuổi theo các mức phân tổng hợp NPK.



Hình 10.46. Kiểm định thuần nhất D, H và số lá của Gõ đỏ 6 tháng tuổi theo các mức phân tổng hợp NPK.

Giải thích và nhận xét kết quả. Số liệu ở Bảng 10.8a là kết quả phân tích phương sai đối với từng biến D, H và số lá. Từ đó cho thấy, sự thay đổi hàm lượng phân tổng hợp NPK ảnh hưởng rõ rệt ($P < 0,01$) đến sinh trưởng D, H và số lá của cây con Gõ đỏ 6 tháng tuổi ở vườn ươm. Bảng 10.8b là kết quả phân nhóm hàm lượng phân tổng hợp NPK ảnh hưởng đồng đều đến sinh trưởng đường kính của cây con Gõ đỏ 6 tháng tuổi ở vườn ươm. Bảng này cũng áp dụng để phân nhóm hàm lượng phân tổng hợp NPK ảnh hưởng đồng đều đến sinh trưởng H và số lá. Từ số liệu ở Bảng 10.8b cho thấy, hàm lượng phân tổng hợp NPK từ 1 – 4% so với trọng lượng ruột bầu đảm bảo cho cây con Gõ đỏ sinh trưởng tốt về D trong giai đoạn 6 tháng tuổi ở vườn ươm; trong đó tối ưu là 3%. Hai biến H và số lá cũng được phân tích tương tự như biến D.

Bảng 10.8a. Kiểm định ảnh hưởng của phân tổng hợp NPK đến sinh trưởng của cây con Gõ đỏ 6 tháng tuổi.

Source	Dependent	SS_{III}	df	MS	F	Sig
Corrected Model	D	15.32	8	1.92	15.4	0.000
	H	1882.60	8	235.32	51.0	0.000
	Lá	9.73	8	1.22	12.6	0.000
Intercept	D	5182.57	1	5182.57	41592.6	0.000
	H	155884.12	1	155884.12	33755.6	0.000
	Lá	6041.83	1	6041.83	62501.7	0.000
Khối (X_1)	D	0.30	2	0.15	1.2	0.336
	H	28.84	2	14.42	3.1	0.081
	Lá	0.13	2	0.06	0.7	0.537
Phân NPK (X_2)	D	15.03	6	2.50	20.1	0.000
	H	1853.76	6	308.96	66.9	0.000
	Lá	9.60	6	1.60	16.6	0.000
Error	D	1.50	12	0.13		
	H	55.42	12	4.62		
	Lá	1.16	12	0.10		
Total	D	5199.39	21			
	H	157822.13	21			
	Lá	6052.72	21			
Corrected Total	D	16.82	20			
	H	1938.01	20			
	Lá	10.89	20			

Bảng 10.8b. Phân nhóm hàm lượng phân tổng hợp NPK ảnh hưởng đồng đều đến sinh trưởng đường kính của cây con Gõ đỏ 6 tháng tuổi ở vườn ươm.

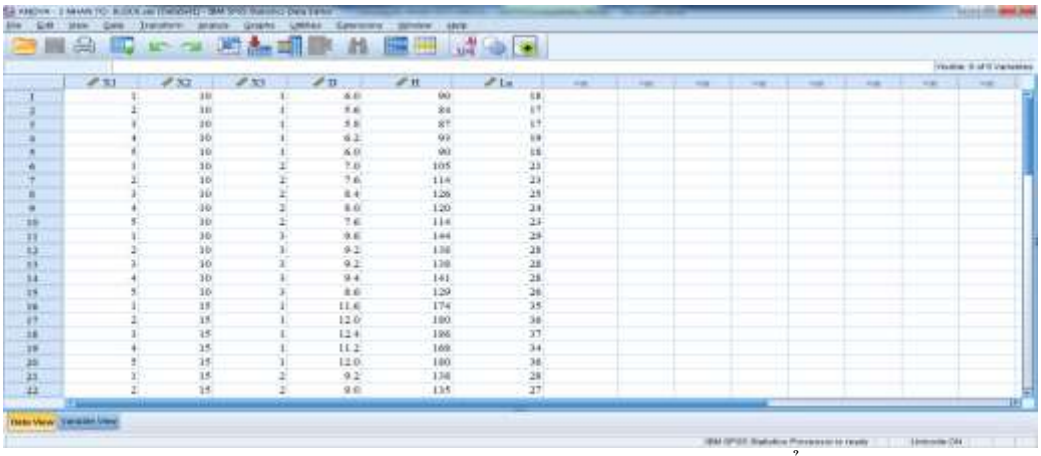
Phân NPK (%)	N	Phân nhóm thuần nhất		
		1	2	3
6	3	13.93		
0	3		15.07	
5	3		15.77	15.77
1	3			16.20
4	3			16.23
2	3			16.30
3	3			16.47
Sig.		1.000	0.266	0.266

10.3.4. Phân tích phương sai hai nhân tố với nhiều biến phụ thuộc

Giả sử cần phân tích ảnh hưởng của 3 mức phân chuồng hoai (X = 10, 15, 20% so với trọng lượng bầu) và 3 mức phân tổng hợp NPK (1, 2 và 3% so với trọng lượng bầu) đến sinh trưởng chiều cao (H, cm), D (cm) và số lá (La) của cây con Gõ đỏ trong giai đoạn 6 tháng tuổi ở vườn ươm. Thí nghiệm bao gồm 9 nghiệm thức và được bố trí theo khối ngẫu nhiên hoàn toàn. Mỗi nghiệm thức được lặp lại 5 lần

(Khối). Tổng số 45 lô thí nghiệm. Mỗi lô thí nghiệm sử dụng 49 cây đồng đều về hình thái và sức sống. Kết quả thí nghiệm được ghi lại ở Phụ lục 15. Giả thuyết H_0: (1) Các mức phân chuồng hoai ảnh hưởng đồng đều đến sinh trưởng D, H và số lá của Gõ đỏ trong giai đoạn 6 tháng tuổi ở vườn ươm. (2) Các mức phân tổng hợp NPK ảnh hưởng đồng đều đến sinh trưởng D, H và số lá của Gõ đỏ trong giai đoạn 6 tháng tuổi ở vườn ươm. (3) Sự phối hợp giữa các mức phân chuồng hoai và phân tổng hợp NPK ảnh hưởng đồng đều đến sinh trưởng D, H và số lá của Gõ đỏ trong giai đoạn 6 tháng tuổi ở vườn ươm. Để kiểm định các giả thuyết trên, các bước xử lý như dưới đây

Bước 1. Nhập số liệu vào bảng tính SPSS. Số liệu được nhập theo 6 cột (Hình 10.47). Cột 1 là khối (X_1). Cột 2 và 3 tương ứng là hàm lượng phân chuồng hoai (X_2) và phân tổng hợp NPK (X_3). Cột 4, 5 và 6 tương ứng là D, H và số lá của cây con Gõ đỏ.



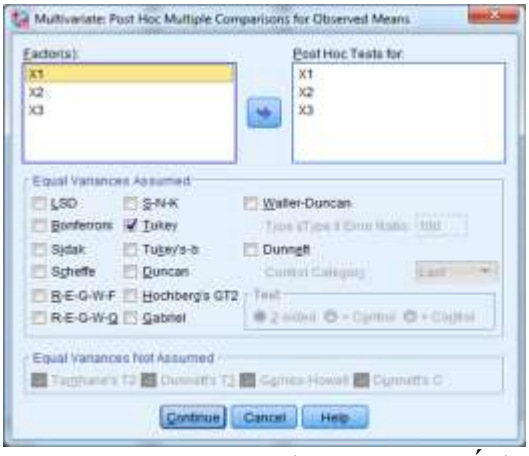
Hình 10.47. Sinh trưởng của cây con Gõ đỏ 6 tháng tuổi dưới ảnh hưởng của phân chuồng hoai và phân tổng hợp NPK.

Bước 2. Xây dựng mô hình tuyến tính tổng quát: $Y_i = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_2*X_3 + \varepsilon$; trong đó Y = (D, H, số lá), β_0 = điểm chặn (Intercept), X_1 = Khối, X_2 = Phân chuồng hoai, X_3 = Phân tổng hợp NPK, X_2*X_3 = Sự phối hợp giữa phân chuồng hoai và phân tổng hợp NPK, β_1, β_2, β_3 và β_4 = hệ số phản ánh ảnh hưởng của X_1, X_2, X_3 và X_2*X_3, ε = sai số ngẫu nhiên. Để đạt mục đích này, từ menu chính, chọn Analyze > GLM: Multivariate. Đến đây chọn biến D, H và số lá và chuyển chúng vào Dependent Variable. Tương tự, chọn X_1, X_2 và X_3 và chuyển chúng vào Fixed Factor (Hình 10.48). Để thực hiện các phân tích thống kê, từ cửa sổ Multivariate ở Hình 10.48, chọn Model. Kế đến đánh dấu mục Custom và chuyển X_1, X_2 và X_3 từ ô Factor & Covariate sang ô Model (Hình 10.49). Tiếp đến chọn Continue ở Hình 10.49; sau đó chọn Post Hoc ở Hình 10.48. Khi xuất hiện cửa sổ Post Hoc Multiple Comparisons for Observed Means, chọn X_1, X_2 và X_3 trong ô Factor và chuyển chúng sang ô Post Hoc Tests for. Tiếp theo chọn kiểm định Tukey's Test (Hình 10.50). Mức tin cậy mặc định là 0,95. Sau đó chọn kiểm định

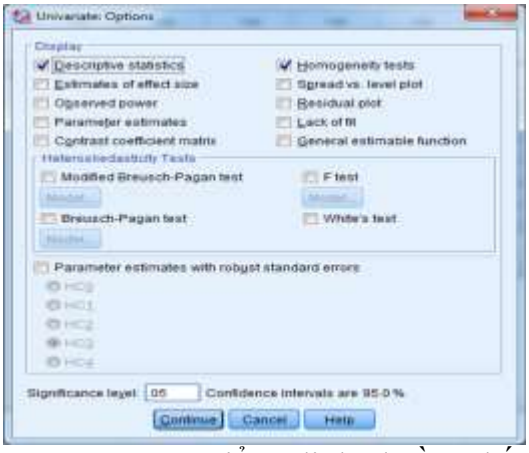
thuần nhất D, H và số lá của cây con Gõ đỏ giữa 3 hàm lượng phân chuồng và phân tổng hợp NPK. Để đạt mục đích này, tại cửa sổ Univariate: Option, chọn Homogeneity Test và Descriptive Statistics (Hình 10.51). Cuối cùng chọn Continue ở Hình 10.451 và chọn OK ở Hình 10.48 để nhận kết quả phân tích GLM (Bảng 10.9).

Hình 10.48. Số liệu để xây dựng mô hình: $Y_i = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_2*X_3$.

Hình 10.49. Xây dựng mô hình: $Y_i = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_2*X_3$.



Hình 10.50. So sánh D, H và số lá của Gõ đỏ 6 tháng tuổi theo các mức phân chuồng và phân NPK.



Hình 10.51. Kiểm định thuần nhất D, H và số lá của Gõ đỏ theo các mức phân chuồng và phân NPK.

Giải thích và nhận xét kết quả. Số liệu ở Bảng 10.9a cho thấy sự thay đổi hàm lượng phân chuồng hoai từ 10 – 20% và hàm lượng phân tổng hợp NPK từ 1 – 3% ảnh hưởng rõ rệt ($P < 0,01$) đến sinh trưởng D, H và số lá của cây con Gõ đỏ 6 tháng tuổi ở vườn ươm. Số liệu ở Bảng 10.9b dẫn ví dụ phân tích nhóm hàm lượng phân chuồng ảnh hưởng đồng đều đến sinh trưởng đường kính của cây con Gõ đỏ 6 tháng tuổi ở vườn ươm. Hai biến H và số lá cũng được phân tích tương tự như biến D.

Bảng 10.9a. Kiểm định ảnh hưởng của phân tổng hợp NPK đến sinh trưởng của cây con Gõ đỏ 6 tháng tuổi.

Source	Dependent	SS_{III}	df	MS	F	Sig
Corrected Model	D	106.7	12	8.9	35.8	0.000
	H	23997.2	12	1999.8	35.8	0.000
	La	968.3	12	80.7	35.8	0.000
Intercept	D	3338.5	1	3338.5	13458.8	0.000
	H	751168.8	1	751168.8	13458.8	0.000
	La	30160.6	1	30160.6	13371.7	0.000
X1	D	2.6	4	0.6	2.6	0.055
	H	579.2	4	144.8	2.6	0.055
	La	20.2	4	5.1	2.2	0.087
X2	D	30.4	2	15.2	61.3	0.000
	H	6841.2	2	3420.6	61.3	0.000
	La	270.2	2	135.1	59.9	0.000
X3	D	5.8	2	2.9	11.7	0.000
	H	1300.8	2	650.4	11.7	0.000
	La	51.4	2	25.7	11.4	0.000
X2 * X3	D	67.9	4	17.0	68.4	0.000
	H	15276.0	4	3819.0	68.4	0.000
	La	626.5	4	156.6	69.4	0.000
Error	D	7.9	32	0.2		
	H	1786.0	32	55.8		
	La	72.2	32	2.3		
Total	D	3453.1	45			
	H	776952.0	45			
	La	31201.0	45			
Corrected Total	D	114.6	44			
	H	25783.2	44			
	La	1040.4	44			

Bảng 10.9c. Phân nhóm hàm lượng phân chuồng ảnh hưởng đồng đều đến sinh trưởng đường kính của cây con Gõ đỏ 6 tháng tuổi ở vườn ươm.

Phân chuồng (%)	N	Phân nhóm thuần nhất		
		1	2	3
10	15	7.613		
20	15		8.600	
15	15			9.627
Sig.		1.000	1.000	1.000

TÀI LIỆU THAM KHẢO

1. Aguirre O., Hui, G., von Gadow K & Ziménez J., 2003. An analysis of spatial forest structure using neiggbourhood-based variables. Forest ecology and menagement 183(1): 137-145.
2. Cintrón, G.; Schaeffer-Novelli, Y., 1984. Methods for studying mangrove structure, In: Snedaker, S.C. (Ed.) (1984). The mangrove ecosystem: reSarch methods. Monographs on Oceanographic Methodology, 8. UNESCO: Paris. ISBN 978-9231021817. xv, 251 pp.
3. Curtis, R.O., 1982. A simple index of stand density for Douglas-fir. For. Sci. 28: 92 – 94.
4. Chao, A., 1984. Nonparametric estimation of the number of classes in a population. Scand. J.Stat. 11, 265 – 270.
5. Chazdon, R.L., Colwell, R.K., Denslow, J.S., and Guariguata, M.R., 1998. Statistical methods for estimating species richness of woody regenaration in primary and secondary rain forests of northeastern Costa Rica. In forest biodiversity research, monitoring and modelling: conceptual background amd old world case studies, pp 285 – 309. Paris: Parthenon Publishing.
6. Clarke, K.R. and Warwick, R.M., 2001. Changes in marine communities: an approach to statistical analysis and interpretation, 2nd edition, PRIMER-E: Plymouth.
7. Daniels, R.F., Burkhart, H.E., Clason, T.R., 1986. A comparison of competition measures for predicting growth of loblolly pine trees. Can. J. For. Res. 16, 1230– 1237.....
8. Ducey, M.J., 2009. Predicting crown size and shape from simple stand variables. J. Sustain. For. 28, 5-21pp.
9. Đào Thị Thùy Dương, 2019. Đặc điểm sinh thái tái sinh tự nhiên của Dầu rái (Dipterocarpus alatus Roxb) dưới tán rừng kín thường xanh ẩm nhiệt đới ở khu vực Tân Phú thuôc tỉnh Đồng Nai. Luận án tiến sĩ, Trường Đại học Nông LâmTp. Hồ Chí Minh, 155 trang.
10. Greig – Smith, P., 1964. Quantitative plant ecology. Ed. 2. London.
11. Giavelli, G., Rossi, O., Satore, F., 1986. Comparative evaluation of four species diversity indecies related to two specific ecologycal situations. Field studies 6 (1986), p 429-438.
12. Gaston, K.J., 2000. Global patterns in biodiversity. Nature 405 (6783): 220–227. DOI: 10.1038/35012228.
13. Gadow, K., Hui, G.Y., 2007. Characterizing forest spatial structure and diversity. Forestry studies, 46: 13-22.
14. Hill, M.O., 1973. Diversity and evenness: a unifying notation and its consequences. Ecology, 54 : 427-432.
15. Heip, C., 1974. A new index measuring evenness. Journal of the Marine Biological Association, 54, 37 – 555.

16. Jongman, R.H., ter Braak, C.J.F., van Tongeren, O.F.R., 1987. Data analysis in community and lanscape ecology. Centre for Agricultural Publishing and Documentation (Pudoc), Wageningen, 1987, 299 pp.....
17. Kimmins, J. P., 1998. Forest ecology, Prentice- Hall, upper Saddle River, New Jersey, 750 p.
18. Larsen, D.R., 1999. Site index. Natural Resource Biometrics. Construction of site index equations for Pinus sylvestris L. using permanent plot data in Sweden. The School of Natural Resources, University of Missouri-Columbia.
19. Kint, V., 2003. Structural development in ageing Scots pine (Pinus sylvestis) stands in Western Europe PhD thesis Ghent University, 194 pp. .....
20. Lü, X.T., Jin, J.X., Tang, J.W., 2010. Structure, tree species diversiy and composition of tropical seasonal rain forest in Xishuangbanna South-West China. Journal of tropical forest science 22 (3): 260 – 270 (2010) .....
21. Lê Văn Long, 2019. Đặc điểm lâm học của những loại hình quần xã thực vật thuộc kiểu rừng kín thường xanh ẩm nhiệt đới ở khu vực Tân Phú, tỉnh Đồng Nai. Luận án tiến sĩ, Trường Đại học Nông LâmTp. Hồ Chí Minh, 142 trang.
22. Margalef, F.R., 1958. Information theory in ecology. Gen. Syst, 3: 36-71.
23. Magurran, A.E, 1988. Ecologycal diversity and its measurement. Princeton NJ: Princeton University Press.
24. Magurran, A.E., 2004. Measuring biologycal diversity. Blackwell Sience Ltd., USA, 260 pages.
25. Monserud, R.A., 1984. Height growth and site index curves for inland Doughlas-fir based on stem analysis data and forest habitat type. Forest Science 30: 943–965.
26. Nguyễn Văn Trương, 1984. Quy luật cấu trúc rừng gỗ hỗn loại, Nxb. Khoa học và Kỹ thuật, Hà Nội, 200 trang.
27. Nguyễn Văn Thêm, 1992. Nghiên cứu tái sinh tự nhiên của Dầu song nàng (Dipterocarpus dyerii) trong kiểu rừng kín thường xanh và nửa rụng lá ẩm nhiệt đới ở Đồng Nai. Luận án Phó tiến sĩ khoa học nông nghiệp. Viện khoa học lâm nghiệp Việt Nam, 125 trang.
28. Nguyễn Văn Thêm, 2002. Sinh thái rừng. Nxb. Nông nghiệp, Hà Nội, 250 trang.
29. Nguyễn Văn Thêm, 2010. Phân tích số liệu quần xã thực vật rừng. Nxb. Nông nghiệp, Hà Nội, 397 trang.
30. Nguyễn Văn Thêm và Nguyễn Tuấn Bình, 2017. Chỉ số phức tạp về cấu trúc đối với rừng kín thường xanh ở Mã Đà, tỉnh Đồng Nai. Tạp chí KHKT Nông Lâm nghiệp, Trường Đại học Nông Lâm, Tp. Hồ Chí Minh, số 1/2017.
31. Nguyễn Văn Thêm và Vũ Mạnh, 2017. Mô hình hóa phân bố đường kính đối với rừng tự nhiên hỗn loài nhiệt đới ở khu vực Nam Cát Tiên thuộc tỉnh Đồng Nai bằng hàm phân bố mũ và phân bố Beta. Tạp chí khoa học Nông Lâm nghiệp, Trường ĐHNL. Tp. HCM, Số 2/2017.

32. Neumann, M., Starlinger, F., 2001. The significance of different indices for stand structure and diversity in forests. For. Ecol. Manage. 145, 91–106.
33. Nguyễn Văn Nhẫn, 2018. Ảnh hưởng của khí hậu đến tăng trưởng của Du sam, Bạch tùng và Đỉnh tùng ở khu vực Đà Lạt và Đức Trọng, tỉnh Lâm Đồng. Luận án Phó tiến sĩ khoa học nông nghiệp. Viện khoa học lâm nghiệp Việt Nam, 125 trang.
34. Onyekwelu, J.C., 2003. Choosing appropriate index age for estimating site index of Gmelina arborea timber plantations in Oluwa forest reserve. Food, Agriculture & Environment Vol.1(3&4) : 286-290.
35. Peet, R.K., Christensen, N.L., 1998. Changes in species diversity during secondary forest succession on the north Carolinna Piedmont. In Book "Diversity and Patern in Plant communities", edited by H.J. During, M.J.A Werger, H.J. Willems, 1998. SPB Academic Publishing bv, The Hague, The Netherlands. p 233-245.
36. Peet, R.K., Christensen, N.L., 1998. Changes in species diversity during secondary forest succession on the north Carolinna Piedmont. In Book "Diversity and Patern in Plant communities", edited by H.J. During, M.J.A Werger, H.J. Willems, 1998. SPB Academic Publishing bv, The Hague, The Netherlands. p 233-245.
37. Pommerening, A., 2002. Approaches to quantifying forest structures. Forestry Vol. 75, No 3, 2002.
38. Phạm Xuân Quý, 2010. Nghiên cứu một số đặc điểm lâm học của rừng Tràm *(Melaleuca cajuputi* Powell*)* trồng ở đồng bằng sông Cửu Long. Luận án tiến sỹ lâm nghiệp. Trường Đại học Nông Lâm Tp. Hồ Chí Minh, 135 trang.
39. Pastorella, F., Paletto, A., 2013. Stand structure indices as tools to support forest management: an application in Trentino forests (Italy). Journal of foest science, 59, 2013 (4): 159–168.
40. Richards, P.W., 1965. Rừng mưa nhiệt đới. Vương Tấn Nhị dịch, Nxb Khoa học và Kỹ thuật, Hà Nội.
41. Schröder, J., Gadow, K.V., 1999. Testing a new competition index for Maritime pine in northwestern Spain. Can. J. For. Res. 29, 280–283.
42. Suratman, M. N., 2012. Tree species diversity and forest stand structure of Pahang National Park, Malaysia. http://dx.doi.org/10.5772/50339.
43. Thái Văn Trừng, 1978. Thảm thực vật rừng Việt Nam trên quan điểm hệ sinh thái, Nxb. Khoa học và Kỹ thuật, Hà Nội.
44. Thái Văn Trừng, 1998. Những hệ sinh thái rừng nhiệt đới ở Việt Nam. Nxb. Khoa học và Kỹ thuật, Hà Nội, 566 trang.
45. Tripathi, K.P, Tripathi S., Selven, T., Kuma, K., Singh, Shanta Mehrotra., K.K., Pushpangadan, P., 2004. Conmmunity structure and species diversity of Saddle Peak forests in Andaman Island. Tropical ecology 45 (2): 241-250.
46. Trần Thị Ngoan, 2019. Ước lượng sinh khối và dự trữ carbon trên mặt đất đối với rừng trồng Keo lai (*Acacia auriculiformis*Acacia mangium*) ở tỉnh

Đồng Nai. Luận án tiến sĩ lâm nghiệp, Trường Đại học Nông Lâm Tp. Hồ Chí Minh, 130 trang.
47. Vasilevich, 1969. Những phương pháp thống kê trong địa thực vật, Nxb. Khoa học, Leningrad (Tiếng Nga).
48. Vanclay, J.K., 1994. Modelling forest growth and yield: Application to mixed tropical forests. Southern Cross University. ePublication@SCU.
49. Vũ Tiến Hinh, 2012. Điều tra rừng. Nxb. Nông nghiệp, Hà Nội, 196 trang.
50. Vũ Tiến Hinh, 2012. Phương pháp lập biểu thể tích cây đứng rừng tự nhiên ở Việt Nam. Nxb. Nông nghiệp, Hà Nội, 196 trang.
51. Whitford, P.B., 1948. Distribution of woodland plants in relation to succession and clonal growth. Ecology, 30, 1948, pp. 199-208.
52. Whittaker, R.H., 1972. Evolution and measurement of species diversity. Taxon 21, 213 – 251.
53. Whittaker, R.H, 1977. Evolution of species diversity in land communities. Evolutionury Biol 10, 1 – 67.
54. Wang, D.P., Ji, S.Y., Chen, F.P., Sing, F.W., Peng, S.L., 2006. Diversity and relationship with succession of natunally regenerated southern subtropical forests in Shenzen, China and its comparison with the zonal climax of Hong Kong. Forest ecology and Management 222: 384 – 390.

PHỤ LỤC

Phụ lục 1. Kết cấu loài cây gỗ của ưu hợp Dầu con rái được xác định theo theo chỉ số IVI với hai tham số N% và G%.

TT	Loài cây	N (cây)	G (m^2)	N%	G%	IV%
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	Dầu con rái	148	9,84	20,8	24,0	22,4
2	Bằng lăng ổi	108	10,41	15,2	25,4	20,3
3	Cầy	80	7,69	11,2	18,8	15,0
4	Trường lá nhỏ	68	3,66	9,6	8,9	9,3
5	Trâm vỏ đỏ	84	2,46	11,8	6,0	8,9
6	Cám	36	1,32	5,1	3,2	4,2
7	Sao đen	12	1,59	1,7	3,9	2,8
8	Bứa	28	0,95	3,9	2,3	3,1
9	Gõ đỏ	8	1,16	1,1	2,8	2,0
10	Cuống vàng	20	0,27	2,8	0,7	1,8
11	Quần đầu	20	0,18	2,8	0,4	1,6
12	Thị rừng	12	0,23	1,7	0,6	1,2
13	Dẻ	12	0,17	1,7	0,4	1,1
14	Chiếc tam lang	8	0,16	1,1	0,4	0,8
15	Dền đỏ	8	0,19	1,1	0,5	0,8
16	Vên vên	8	0,16	1,1	0,4	0,8
17	Nhọc	8	0,13	1,1	0,3	0,7
18	Côm	8	0,06	1,1	0,1	0,6
19	Re hương	8	0,06	1,1	0,1	0,6
20	Săng đen	8	0,05	1,1	0,1	0,6
21	Gạo	4	0,04	0,6	0,1	0,4
22	Mận rừng	4	0,04	0,6	0,1	0,4
23	Máu chó	4	0,03	0,6	0,1	0,4
24	Móng bò	4	0,03	0,6	0,1	0,4
25	Thành ngạnh	4	0,03	0,6	0,1	0,4
		712	40,9	100	100	100

Phụ lục 2. Kết cấu loài cây gỗ của ưu hợp Dầu con rái được xác định theo theo chỉ số IVI với ba tham số N%, G% và V%.

TT	Loài cây	N (cây)	G (m^2)	V (m^3)	N%	G%	V%	IV%

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	Dầu con rái	148	9,84	97,71	20,8	24,0	25,0	23,3
2	Bằng lăng ổi	108	10,41	103,79	15,2	25,4	26,6	22,4
3	Cầy	80	7,69	74,46	11,2	18,8	19,1	16,4
4	Trường lá nhỏ	68	3,66	38,90	9,6	8,9	10,0	9,5
5	Trâm vỏ đỏ	84	2,46	19,66	11,8	6,0	5,0	7,6
6	Cám	36	1,32	10,83	5,1	3,2	2,8	3,7
7	Sao đen	12	1,59	15,69	1,7	3,9	4,0	3,2
8	Bứa	28	0,95	7,47	3,9	2,3	1,9	2,7
9	Gõ đỏ	8	1,16	11,51	1,1	2,8	2,9	2,3
10	Cuống vàng	20	0,27	1,64	2,8	0,7	0,4	1,3
11	Quần đầu	20	0,18	0,78	2,8	0,4	0,2	1,2
12	Thị rừng	12	0,23	1,56	1,7	0,6	0,4	0,9
13	Dẻ	12	0,17	0,97	1,7	0,4	0,2	0,8
14	Chiếc tam lang	8	0,16	0,80	1,1	0,4	0,2	0,6
15	Dền đỏ	8	0,19	1,25	1,1	0,5	0,3	0,6
16	Vên vên	8	0,16	1,05	1,1	0,4	0,3	0,6
17	Nhọc	8	0,13	0,79	1,1	0,3	0,2	0,5
18	Côm	8	0,06	0,26	1,1	0,1	0,1	0,4
19	Re hương	8	0,06	0,29	1,1	0,1	0,1	0,4
20	Săng đen	8	0,05	0,19	1,1	0,1	0,0	0,4
21	Gạo	4	0,04	0,24	0,6	0,1	0,1	0,2
22	Mận rừng	4	0,04	0,21	0,6	0,1	0,1	0,2
23	Máu chó	4	0,03	0,14	0,6	0,1	0,0	0,2
24	Móng bò	4	0,03	0,14	0,6	0,1	0	0,2
25	Thành ngạnh	4	0,03	0,11	0,6	0,1	0	0,2
		712	40,9	390,4	100	100	100	100

Phụ lục 3. Chuỗi chỉ số bề rộng vòng năm chuẩn hóa của Du sam.

Năm	Chỉ số Kd chuẩn hóa									
	0	1	2	3	4	5	6	7	8	9

1820			0,91	1,20	0,95	1,01	0,98	0,95	0,99	1,12
1830	0,92	0,98	1,01	1,07	0,93	1,05	0,98	1,00	1,01	1,09
1840	0,78	1,04	1,09	0,90	0,93	1,19	0,93	0,96	1,10	0,87
1850	1,12	0,92	1,03	1,07	0,96	0,96	1,06	0,97	1,01	1,04
1860	0,86	1,16	0,91	1,00	1,03	1,03	1,01	0,95	1,12	0,88
1870	1,09	0,90	1,03	1,04	0,96	1,03	1,03	1,01	1,05	0,88
1880	1,01	1,05	0,87	1,19	0,92	1,02	0,85	1,12	1,00	1,03
1890	0,88	1,14	0,99	0,93	0,95	1,10	1,00	1,07	0,85	1,04
1900	1,01	0,96	1,07	0,95	1,11	0,92	0,91	1,07	1,01	1,04
1910	0,93	1,00	1,13	0,90	0,99	0,91	1,14	1,01	0,93	1,00
1920	1,04	1,02	1,00	0,94	1,08	0,93	1,13	0,95	0,89	1,09
1930	0,98	0,93	1,09	0,97	1,06	0,81	1,15	0,94	1,19	0,89
1940	1,01	0,85	1,12	0,96	0,97	0,97	1,21	0,87	0,93	1,04
1950	0,96	0,97	1,08	0,85	1,13	1,01	1,02	1,10	0,84	0,97
1960	1,01	1,02	1,01	1,08	0,81	1,13	0,97	1,13	0,86	1,09
1970	0,87	1,15	1,00	0,85	1,17	0,92	0,93	1,16	0,92	0,96
1980	1,07	0,93	1,20	0,83	1,08	0,98	0,97	1,04	0,95	1,13
1990	0,96	0,96	1,02	0,98	1,05	0,93	1,03	1,07	0,90	0,94
2000	1,05	1,03	0,98	0,96	1,15	0,89	1,00	1,06	0,91	1,01
2010	0,94	1,02	1,10	0,84						

Phụ lục 4. Chỉ số nhiệt độ không khí của các tháng từ năm 1981 - 2013.

Năm	Tháng trong năm											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
2013	1,01	1,03	1,02	1,03	1,01	1,00	1,01	0,99	0,99	0,99	0,98	0,96
2012	1,04	1,01	1,00	1,00	1,00	0,99	1,00	1,01	1,00	1,01	1,02	1,04
2011	0,97	0,97	0,96	0,97	0,97	1,00	0,99	1,01	1,00	1,00	1,00	0,98
2010	1,04	1,02	1,03	1,01	1,04	1,02	1,01	0,98	1,01	1,00	0,99	1,02
2009	0,95	1,01	1,03	1,01	0,98	0,99	1,00	1,04	0,98	0,99	0,99	0,98
2008	0,99	0,99	0,94	0,99	0,99	1,00	0,99	0,96	1,02	1,02	1,02	1,00
2007	1,01	0,98	1,02	1,01	1,00	1,00	1,00	1,01	1,00	0,99	0,97	1,01
2006	1,01	1,01	1,01	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,01	1,00
2005	1,01	1,02	1,00	1,00	1,00	1,01	1,00	1,01	1,00	1,05	1,01	1,02
2004	0,99	0,98	1,00	0,99	1,00	0,98	0,99	0,99	1,00	0,91	0,99	0,99
2003	1,00	1,02	1,00	1,02	0,99	1,01	1,00	1,02	0,99	1,03	1,00	0,95
2002	0,97	0,98	0,99	0,98	1,01	1,01	1,01	0,98	0,99	1,01	1,03	1,07
2001	1,02	1,00	1,01	1,02	1,00	0,99	1,01	1,01	1,01	1,00	0,96	0,96
2000	0,99	1,02	0,99	0,99	1,00	1,01	0,98	1,00	0,99	1,00	1,01	1,04

1999	1,00	0,96	1,00	1,00	0,97	0,96	0,99	0,98	1,00	0,99	1,01	0,96
1998	1,04	1,05	1,04	1,02	1,03	1,05	1,03	1,03	1,01	1,02	1,01	1,03
1997	0,95	0,99	0,96	0,98	0,98	0,98	0,99	0,98	1,00	0,99	0,99	1,00
1996	1,01	0,99	1,01	1,00	1,01	0,99	0,99	1,01	0,99	1,00	1,01	1,00
1995	1,00	0,98	1,00	1,01	1,00	1,02	1,02	1,00	1,01	1,02	1,00	0,97
1994	1,02	1,05	1,00	1,00	1,00	0,98	0,97	1,01	1,00	0,98	0,98	1,03
1993	0,98	0,94	0,99	1,00	1,00	1,01	1,02	0,98	0,99	1,01	1,03	0,99
1992	0,98	1,03	1,01	1,00	1,01	0,99	1,00	1,00	1,01	1,00	0,97	1,00
1991	1,03	0,99	1,00	0,99	1,00	1,00	0,99	1,01	1,00	0,99	1,00	1,00
1990	0,98	1,02	1,00	1,02	1,00	1,00	1,02	0,99	1,00	1,02	1,00	1,01
1989	1,01	0,94	0,97	0,98	0,98	0,99	0,98	0,99	0,99	0,99	1,01	1,00
1988	1,03	1,07	1,03	1,01	1,02	1,00	1,01	1,02	1,00	1,00	0,96	0,96
1987	0,98	0,96	1,00	1,00	1,01	1,01	1,01	1,00	1,02	1,02	1,07	1,03
1986	0,99	0,99	0,99	1,00	0,99	1,01	1,00	0,99	0,98	1,00	0,95	1,00
1985	1,01	1,03	1,00	0,98	1,00	1,00	0,99	1,01	1,02	1,00	1,04	0,99
1984	0,98	0,98	0,99	1,00	1,00	0,98	0,99	0,97	0,98	0,99	0,98	1,01
1983	1,05	1,02	1,03	1,03	1,01	1,02	1,02	1,02	1,02	1,00	0,97	0,99
1982	0,96	0,99	0,99	0,97	0,99	0,99	0,98	1,01	0,99	1,00	1,03	1,01
1981	0,99	1,01	1,00	1,01	1,01	0,99	1,01	0,99	1,01	1,00	1,00	0,99

Phụ lục 5. Ảnh hưởng của hàm lượng mùn (X_1%) và pH_{H2O} (X_2) của đất Bazan nâu đỏ đến sinh trưởng chiều cao (H, m) của Dầu song nàng 2 năm tuổi.

TT	H	X1	X2	TT	H	X1	X2	TT	H	X1	X2	TT	H	X1	X2
1	2,9	5,6	4,1	39	77	2,3	4,3	6,5	115	1,0	2,1	3,2	1	2,9	5,6
2	2,9	7,6	4,1	40	78	2,3	4,1	5,7	116	0,9	2,1	3,2	2	2,9	7,6
3	3,1	7,8	4,0	41	79	2,3	2,9	6,2	117	1,9	5,2	4,2	3	3,1	7,8
4	3,2	9,2	3,6	42	80	2,4	3,5	6,2	118	2,7	5,6	4,0	4	3,2	9,2
5	3,3	7,0	4,0	43	81	2,1	3,5	4,0	119	1,7	5,2	3,8	5	3,3	7,0
6	2,2	3,8	4,0	44	82	1,7	4,0	3,2	120	1,7	5,1	3,2	6	2,2	3,8
7	2,4	4,5	4,0	45	83	1,7	4,0	3,9	121	3,0	9,3	4,3	7	2,4	4,5
8	3,1	8,5	4,1	46	84	1,9	5,1	3,6	122	2,7	7,9	4,9	8	3,1	8,5
9	3,3	14,2	4,1	47	85	2,7	4,9	3,9	123	2,8	8,1	4,8	9	3,3	14,2
10	2,4	4,0	3,9	48	86	2,6	3,6	4,0	124	3,1	13,1	4,4	10	2,4	4,0
11	2,5	5,8	3,9	49	87	2,6	6,2	3,5	125	3,1	10,9	4,3	11	2,5	5,8
12	3,2	8,5	4,2	50	88	2,9	5,8	4,0	126	3,0	10,3	5,0	12	3,2	8,5
13	3,0	8,1	4,2	51	89	2,9	4,4	3,8	127	2,2	8,2	4,5	13	3,0	8,1
14	3,0	7,9	4,1	52	90	2,6	7,8	4,0	128	2,9	10,9	4,2	14	3,0	7,9
15	3,1	9,2	3,7	53	91	2,1	5,6	4,0	129	2,4	7,9	4,8	15	3,1	9,2
16	3,4	12,3	4,1	54	92	1,1	3,0	3,2	130	2,3	7,9	4,8	16	3,4	12,3
17	3,0	6,5	3,8	55	93	1,9	6,1	4,0	131	2,6	8,0	4,3	17	3,0	6,5
18	2,4	5,3	4,2	56	94	2,2	6,3	3,5	132	3,3	13,1	4,4	18	2,4	5,3

19	3,5	14,0	4,1	57	95	2,3	6,4	3,8	133	2,6	5,7	4,5	19	3,5	14,0
20	3,1	7,0	4,1	58	96	1,7	4,4	3,2	134	2,9	10,3	5,0	20	3,1	7,0
21	3,1	6,6	3,9	59	97	2,3	6,3	3,8	135	2,2	6,3	4,5	21	3,1	6,6
22	2,4	5,8	4,2	60	98	2,8	7,9	4,0	136	3,1	10,3	5,0	22	2,4	5,8
23	2,2	5,9	4,1	61	99	1,5	3,9	4,0	137	2,9	10,3	5,0	23	2,2	5,9
24	2,4	4,7	4,2	62	100	2,2	5,1	4,0	138	3,0	9,2	5,0	24	2,4	4,7
25	3,2	10,4	4,2	63	101	2,5	8,0	4,2	139	3,1	10,3	5,0	25	3,2	10,4
26	3,2	8,6	4,0	64	102	2,2	5,5	3,8	140	2,7	5,7	4,5	26	3,2	8,6
27	2,6	5,7	4,1	65	103	2,0	5,8	3,9	141	2,9	9,3	4,6	27	2,6	5,7
28	2,9	6,0	4,1	66	104	1,9	5,8	4,0	142	3,2	13,8	4,6	28	2,9	6,0
29	2,6	6,0	4,7	67	105	2,1	5,9	3,9	143	3,4	19,6	4,8	29	2,6	6,0
30	2,6	5,3	4,7	68	106	2,5	5,5	3,6	144	3,2	10,0	4,9	30	2,6	5,3
31	3,4	14,0	4,3	69	107	1,5	2,7	4,2	145	3,1	16,2	4,8	31	3,4	14,0
32	3,1	8,5	4,3	70	108	2,0	6,0	4,0	146	2,8	6,8	5,2	32	3,1	8,5
33	3,0	3,9	4,1	71	109	2,4	5,5	3,7	147	3,3	17,7	5,2	33	3,0	3,9
34	2,8	4,1	4,1	72	110	1,7	5,5	3,4	148	3,6	19,6	4,1	34	2,8	4,1
35	1,9	3,4	6,4	73	111	2,2	6,0	3,6	149	3,4	17,7	4,1	35	1,9	3,4
36	1,5	3,8	3,2	74	112	2,4	5,8	3,8	150	3,5	18,7	4,5	36	1,5	3,8
37	2,1	4,7	4,1	75	113	2,4	5,5	3,8					37	2,1	4,7
38	3,1	8,5	4,4	76	114	1,5	3,5	3,6					38	3,1	8,5

Phụ lục 6. Ảnh hưởng của độ ẩm (X_1%) và pH_{H2O} (X_2) đến xác suất bắt gặp (P) cây tái sinh tự nhiên của Dầu con rái.

TT	P	X1	X2	TT	P	X1	X2	TT	P	X1	X2	TT	P	X1	X2
1	1	57.5	3.0	31	1	56.5	3.9	61	1	87.2	4.0	91	1	84.8	4.8
2	1	56.2	3.0	32	1	56.9	3.9	62	1	65.5	4.1	92	1	86.1	4.8
3	1	60.3	3.1	33	1	60.5	3.9	63	1	70.2	4.1	93	1	87	4.8
4	1	74.5	3.1	34	1	80.1	3.9	64	1	78.4	4.1	94	1	58	5.0
5	0	92.7	3.4	35	1	80.6	3.9	65	1	78.5	4.2	95	0	45.5	5.4
6	0	94.5	3.4	36	1	80.7	3.9	66	1	80.7	4.2	96	0	91.2	5.4
7	1	97.5	3.4	37	0	94.7	3.9	67	1	89.4	4.2	97	0	96.2	5.4
8	1	98.7	3.4	38	1	95.2	3.9	68	1	90.2	4.2	98	1	57.4	5.5
9	1	79.1	3.5	39	0	90.2	3.9	69	1	90.2	4.2	99	0	45.9	6.0
10	1	60.3	3.6	40	0	90.7	3.9	70	1	82.3	4.4	100	0	46	6.0
11	1	79.1	3.6	41	0	90.8	3.9	71	1	82.7	4.4	101	0	92.2	6.0
12	1	60.2	3.7	42	1	60.5	3.9	72	1	46.4	4.4	102	0	93.2	6.0
13	1	60.4	3.7	43	1	60.8	3.9	73	1	70.5	4.4	103	0	97.2	6.0

14	1	78.6	3.7	44	1	65.6	3.9	74	1	80.9	4.4	104	0	97.2	6.0
15	1	45.5	3.7	45	1	65.6	3.9	75	1	95.5	4.5	105	1	57.6	6.1
16	1	46.4	3.7	46	1	67.2	3.9	76	1	68.1	4.6	106	1	60.9	6.1
17	1	89.7	3.7	47	1	68.1	3.9	77	1	68.4	4.6	107	0	46.1	6.3
18	0	89.8	3.7	48	1	78.6	3.9	78	1	70.4	4.6	108	0	55.6	6.3
19	1	46.9	3.7	49	1	78.8	3.9	79	1	87.8	4.7	109	0	93.2	6.3
20	1	56.3	3.8	50	1	90	3.9	80	1	56.5	4.7	110	0	97.2	6.3
21	1	60.4	3.8	51	1	60.3	4.0	81	1	69.1	4.7	111	0	43.3	6.5
22	1	73.3	3.8	52	1	70.7	4.0	82	1	82.5	4.7	112	0	49.3	6.6
23	1	79.1	3.8	53	1	78.5	4.0	83	1	89.5	4.7	113	0	50.9	6.6
24	0	90.2	3.8	54	1	81.1	4.0	84	1	89.6	4.7	114	0	53.4	6.6
25	0	92.7	3.8	55	1	65.5	4.0	85	0	91.4	4.8	115	0	93.4	6.6
26	1	52.5	3.8	56	1	69.1	4.0	86	1	60.9	4.8	116	0	94.7	6.6
27	1	56.3	3.8	57	1	89.1	4.0	87	1	75.4	4.8	117	0	96.2	6.6
28	1	95.9	3.8	58	1	96.5	4.1	88	1	79.7	4.8	118	0	97.5	6.6
29	1	76.7	3.8	59	1	70.3	4.0	89	1	83.6	4.8	119	0	98.2	6.6
30	1	79.9	3.8	60	1	77.1	4.0	90	1	83.7	4.8	120	0	99.2	6.6

Phụ lục 7. Ảnh hưởng của địa hình đến sinh trưởng của rừng Sao đen. Thí nghiệm 1 nhân tố được bố trí hoàn toàn ngẫu nhiên.

TT	D (cm)	Địa hình	TT	D (cm)	Địa hình	TT	D (cm)	Địa hình
1	27.5	1	16	48	2	31	49.5	3
2	25.5	1	17	49	2	32	45.5	3
3	26.5	1	18	48.5	2	33	46.5	3
4	28.5	1	19	50	2	34	42.5	3
5	27.5	1	20	46.5	2	35	43.5	3
6	32.5	1	21	43.5	2	36	42.5	3
7	35.5	1	22	42.5	2	37	39.5	3
8	39.5	1	23	45.5	2	38	40.5	3
9	37.5	1	24	38.5	2	39	37.5	3
10	35.5	1	25	39.5	2	40	38.5	3
11	45.5	1	26	47.5	2	41	36.5	3
12	43.5	1	27	37.5	2	42	37.5	3
13	43.5	1	28	36.5	2	43	35.5	3
14	44.5	1	29	35.5	2	44	34.5	3

15	40.5	1	30	34.5	2	45	37.5	3

Phụ lục 8. Ảnh hưởng của độ tàn che đến sinh trưởng chiều cao (H, cm) của cây con Dầu song nàng 8 tháng tuổi. Thí nghiệm 1 nhân tố được bố trí theo khối ngẫu nhiên hoàn toàn.

TT	H, cm	Nghiệm	Khối	TT	H, cm	Nghiệm	Khối
1	73.2	1	1	11	81.2	3	3
2	81.0	1	2	12	81.0	3	4
3	90.4	1	3	13	81.6	4	1
4	76.6	1	4	14	98.6	4	2
5	82.8	2	1	15	83.0	4	3
6	84.8	2	2	16	94.4	4	4
7	86.8	2	3	17	85.2	5	1
8	84.8	2	4	18	97.2	5	2
9	76.4	3	1	19	84.2	5	3
10	83.0	3	2	20	99.2	5	4

Phụ lục 9. Ảnh hưởng của Super lân đến đến sinh trưởng chiều cao của cây con Gõ đỏ 3 tháng tuổi ở vườn ươm.

TT	H (cm)	Lân	Hàng	Cột	TT	H (cm)	Lân	Hàng	Cột
1	67	1	1	1	19	24	4	1	4
2	46	2	2	1	20	2	6	2	4
3	47	3	3	1	21	13	5	3	4
4	28	4	4	1	22	90	1	4	4
5	15	5	5	1	23	42	3	5	4
6	7	6	6	1	24	86	2	6	4
7	62	2	1	2	25	9	5	1	5
8	62	1	2	2	26	54	3	2	5
9	5	6	3	2	27	35	4	3	5
10	43	3	4	2	28	3	6	4	5
11	42	4	5	2	29	81	2	5	5
12	23	5	6	2	30	105	1	6	5
13	30	3	1	3	31	5	6	1	6
14	15	5	2	3	32	19	4	2	6
15	65	1	3	3	33	57	2	3	6
16	61	2	4	3	34	17	5	4	6
17	8	6	5	3	35	87	1	5	6

18	58	4	6	3	36	45	3	6	6

Phụ lục 10. Ảnh hưởng của phân đạm và Super lân đến sinh trưởng của cây con Dầu song nàng trong giai đoạn 6 tháng tuổi ở vườn ươm. Thí nghiệm hai yếu tố được bố trí ngẫu nhiên.

TT	H, cm	TT	Đạm	Lân	TT	H, cm	TT	Đạm	Lân	TT	H, cm	TT	Đạm	Lân
1	55	1	1	1	23	61	8	1	2	45	55	15	1	3
2	51	2	1	1	24	57	9	1	2	46	55	1	2	1
3	53	3	1	1	25	59	10	1	2	47	51	2	2	1
4	57	4	1	1	26	61	11	1	2	48	53	3	2	1
5	55	5	1	1	27	56	12	1	2	49	57	4	2	1
6	65	6	1	1	28	54	13	1	2	50	55	5	2	1
7	52	7	1	1	29	60	14	1	2	51	65	6	2	1
8	54	8	1	1	30	59	15	1	2	52	52	7	2	1
9	55	9	1	1	31	55	1	1	3	53	54	8	2	1
10	51	10	1	1	32	61	2	1	3	54	55	9	2	1
11	58	11	1	1	33	60	3	1	3	55	51	10	2	1
12	57	12	1	1	34	56	4	1	3	56	58	11	2	1
13	56	13	1	1	35	57	5	1	3	57	57	12	2	1
14	59	14	1	1	36	55	6	1	3	58	56	13	2	1
15	61	15	1	1	37	59	7	1	3	59	59	14	2	1
16	57	1	1	2	38	61	8	1	3	60	61	15	2	1
17	54	2	1	2	39	55	9	1	3	61	57	1	2	2
18	57	3	1	2	40	57	10	1	3	62	54	2	2	2
19	53	4	1	2	41	54	11	1	3	63	57	3	2	2
20	61	5	1	2	42	56	12	1	3	64	53	4	2	2
21	58	6	1	2	43	61	13	1	3	65	61	5	2	2
22	56	7	1	2	44	59	14	1	3	66	58	6	2	2
67	56	7	2	2	92	61	2	3	1	113	71	8	3	2
68	61	8	2	2	93	63	3	3	1	114	67	9	3	2
69	57	9	2	2	94	67	4	3	1	115	69	10	3	2
70	59	10	2	2	95	65	5	3	1	116	71	11	3	2
71	61	11	2	2	96	75	6	3	1	117	65	12	3	2
72	56	12	2	2	97	62	7	3	1	118	63	13	3	2
73	54	13	2	2	98	64	8	3	1	119	70	14	3	2
74	60	14	2	2	99	65	9	3	1	120	69	15	3	2
75	59	15	2	2	89	59	14	2	3	121	65	1	3	3
76	55	1	2	3	90	55	15	2	3	122	71	2	3	3
77	61	2	2	3	91	65	1	3	1	123	70	3	3	3

78	60	3	2	3	92	61	2	3	1	124	65	4	3	3
79	56	4	2	3	93	63	3	3	1	125	67	5	3	3
80	57	5	2	3	94	67	4	3	1	126	65	6	3	3
81	55	6	2	3	95	65	5	3	1	127	69	7	3	3
82	59	7	2	3	96	75	6	3	1	128	71	8	3	3
83	61	8	2	3	97	62	7	3	1	129	65	9	3	3
84	55	9	2	3	98	64	8	3	1	130	67	10	3	3
85	57	10	2	3	99	65	9	3	1	131	64	11	3	3
86	54	11	2	3	107	64	2	3	2	132	65	12	3	3
87	56	12	2	3	108	67	3	3	2	133	71	13	3	3
88	61	13	2	3	109	63	4	3	2	134	69	14	3	3
89	59	14	2	3	110	71	5	3	2	135	65	15	3	3
90	55	15	2	3	111	67	6	3	2					
91	65	1	3	1	112	65	7	3	2					

Phụ lục 11. Ảnh hưởng của phân đạm và Super lân đến sinh trưởng của cây con Sao đen trong giai đoạn 6 tháng tuổi ở vườn ươm. Thí nghiệm hai yếu tố được bố trí theo khối ngẫu nhiên hoàn toàn.

TT	H, cm	Khối	Đạm	Lân	TT	H, cm	Khối	Đạm	Lân
1	60	1	1	1	24	82	4	2	2
2	56	2	1	1	25	84	5	2	2
3	58	3	1	1	26	100	1	2	3
4	62	4	1	1	27	80	2	2	3
5	60	5	1	1	28	78	3	2	3
6	70	1	1	2	29	76	4	2	3
7	76	2	1	2	30	74	5	2	3
8	84	3	1	2	31	104	1	3	1
9	80	4	1	2	32	96	2	3	1
10	76	5	1	2	33	98	3	3	1
11	96	1	1	3	34	90	4	3	1
12	92	2	1	3	35	92	5	3	1
13	92	3	1	3	36	90	1	3	2
14	94	4	1	3	37	84	2	3	2
15	86	5	1	3	38	86	3	3	2
16	116	1	2	1	39	80	4	3	2
17	120	2	2	1	40	82	5	3	2
18	124	3	2	1	41	78	1	3	3
19	112	4	2	1	42	80	2	3	3
20	120	5	2	1	43	76	3	3	3
21	92	1	2	2	44	74	4	3	3

22	90	2	2	2	45	80	5	3	3
23	96	3	2	2					

Phụ lục 12. Ảnh hưởng của kích thước bầu và phân tổng hợp NPK đến sinh trưởng chiều cao của Dầu con rái 8 tháng tuổi. Thí nghiệm được bố trí theo kiểu bình phương la tinh.

H(cm)	Nghiệm	Hàng	Cột	Bầu	NPK	H(cm)	Nghiệm	Hàng	Cột	Bầu	NPK
67.5	3	1	1	1	3	77.5	5	4	1	2	2
68.8	4	1	2	2	1	76.3	6	4	2	2	3
71.3	6	1	3	2	3	66.3	2	4	3	1	2
59	1	1	4	1	1	72.5	3	4	4	1	3
80	5	1	5	2	2	58.8	1	4	5	1	1
61.3	2	1	6	1	2	63.8	4	4	6	2	1
62.5	2	2	1	1	2	65	4	5	1	2	1
71.3	3	2	2	1	3	78.8	5	5	2	2	2
81.3	5	2	3	2	2	56	1	5	3	1	1
72.5	6	2	4	2	3	68.8	2	5	4	1	2
71.3	4	2	5	2	1	75	6	5	5	2	3
57.5	1	2	6	1	1	66.3	3	5	6	1	3
70	6	3	1	2	3	53.8	1	6	1	1	1
60	1	3	2	1	1	67.5	2	6	2	1	2
68.8	3	3	3	1	3	66.3	4	6	3	2	1
70	4	3	4	2	1	82.5	5	6	4	2	2
63.8	2	3	5	1	2	73.8	3	6	5	1	3
80	5	3	6	2	2	70	6	6	6	2	3

Phụ lục 13. Ảnh hưởng của phương thức xử lý đất đến sinh trưởng chiều cao của Dầu con rái, Sao đen và Vên vên 8 tháng tuổi. Thí nghiệm được bố trí theo kiểu lô chính - lô phụ.

TT	H (cm)	PTLS	Loài cây	Khối	TT	H (cm)	PTLS	Loài cây	Khối
1	150	1	1	1	19	163	2	2	3
2	154	1	1	2	20	159	2	2	4
3	159	1	1	3	21	177	2	3	1
4	151	1	1	4	22	171	2	3	2
5	156	1	2	1	23	170	2	3	3
6	160	1	2	2	24	174	2	3	4
7	162	1	2	3	25	153	3	1	1
8	159	1	2	4	26	164	3	1	2
9	178	1	3	1	27	162	3	1	3
10	179	1	3	2	28	158	3	1	4

11	183	1	3	3	29	152	3	2	1
12	175	1	3	4	30	161	3	2	2
13	159	2	1	1	31	158	3	2	3
14	149	2	1	2	32	154	3	2	4
15	151	2	1	3	33	149	3	3	1
16	148	2	1	4	34	153	3	3	2
17	165	2	2	1	35	151	3	3	3
18	160	2	2	2	36	150	3	3	4

Phụ lục 14. Ảnh hưởng của hàm lượng phân NPK đến sinh trưởng của Gõ đỏ 6 tháng tuổi. Thí nghiệm được bố trí theo khối ngẫu nhiên hoàn toàn.

Khối	NPK	D(cm)	H (m)	Số lá	Khối	NPK	D(cm)	H (m)	Số lá
1	0	15.6	83.7	16	2	4	16	85.5	17.1
1	1	16.2	86.3	17	2	5	16	76.4	16.9
1	2	16.1	99.7	18	2	6	13.9	70.9	16.2
1	3	16.6	93.7	18	3	0	15.5	88.2	16.8
1	4	16.5	82.8	17	3	1	16.2	89.6	17.3
1	5	15.6	77.5	16	3	2	16.5	104.7	18.1
1	6	14.1	73.7	16	3	3	16.5	97.5	17.7
2	0	14.1	79.3	15.9	3	4	16.2	86	16.7
2	1	16.2	89.2	17.4	3	5	15.7	77.3	16.6
2	2	16.3	99.1	17.6	3	6	13.8	71.4	16.1
2	3	16.3	96.8	17.8					

Phụ lục 15. Ảnh hưởng của hàm lượng phân NPK và phân chuồng hoai đến sinh trưởng của Gõ đỏ 6 tháng tuổi. Thí nghiệm được bố trí theo khối ngẫu nhiên hoàn toàn.

Khối	PC	NPK	D	H	Số lá	Khối	PC	NPK	D	H	Số lá
1	10	1	6	90	18	4	15	2	8.2	123	25
2	10	1	5.6	84	17	5	15	2	8.4	126	25
3	10	1	5.8	87	17	1	15	3	10	150	30
4	10	1	6.2	93	19	2	15	3	8	120	24
5	10	1	6	90	18	3	15	3	7.8	117	23
1	10	2	7	105	21	4	15	3	7.6	114	23
2	10	2	7.6	114	23	5	15	3	7.4	111	22
3	10	2	8.4	126	25	1	20	1	10.4	156	31
4	10	2	8	120	24	2	20	1	9.6	144	29
5	10	2	7.6	114	23	3	20	1	9.8	147	29
1	10	3	9.6	144	29	4	20	1	9	135	27
2	10	3	9.2	138	28	5	20	1	9.2	138	28
3	10	3	9.2	138	28	1	20	2	9	135	27
4	10	3	9.4	141	28	2	20	2	8.4	126	25
5	10	3	8.6	129	26	3	20	2	8.6	129	26
1	15	1	11.6	174	35	4	20	2	8	120	24
2	15	1	12	180	36	5	20	2	8.2	123	25

3	15	1	12.4	186	37	1	20	3	7.8	117	23
4	15	1	11.2	168	34	2	20	3	8	120	24
5	15	1	12	180	36	3	20	3	7.6	114	23
1	15	2	9.2	138	28	4	20	3	7.4	111	22
2	15	2	9	135	27	5	20	3	8	120	24
3	15	2	9.6	144	29						