PGS.TS. BẢO HUY

# TIN HỌC THỐNG KÊ TRONG LÂM NGHIỆP

NHÀ XUẤT BẢN KHOA HỌC VÀ KỸ THUẬT

# TIN HỌC THỐNG KÊ
## TRONG LÂM NGHIỆP

PGS.TS. BẢO HUY

# TIN HỌC THỐNG KÊ TRONG LÂM NGHIỆP

## SỬ DỤNG CÁC CHƯƠNG TRÌNH R, STATGRAPHICS, SPSS

*(Giáo trình đại học và sau đại học)*

**NHÀ XUẤT BẢN KHOA HỌC VÀ KỸ THUẬT**

# LỜI MỞ ĐẦU

**Tin học thống kê** có nghĩa là sử dụng tiến bộ tin học để xử lý các vấn đề thống kê chuyên ngành. Giáo trình này tập trung cho xử lý thống kê trong nghiên cứu, thử nghiệm, khảo sát, đánh giá trong lĩnh vực lâm nghiệp, quản lý tài nguyên thiên nhiên, tài nguyên rừng và môi trường rừng; trên cơ sở lựa chọn và xử lý theo các chương trình thống kê trên máy tính thích hợp, có hiệu quả.

Lâm nghiệp và quản lý tài nguyên môi trường rừng là một ngành, mà trong đó các nghiên cứu thực nghiệm, thống kê được sử dụng hầu như hoàn toàn. Thống kê trong ngành này được sử dụng bao gồm rút mẫu, phân tích đặc điểm và phân bố của mẫu, xác định dung lượng mẫu cần thiết; so sánh các thí nghiệm, thử nghiệm từ một công thức, nhân tố đến nhiều công thức, đa nhân tố; mô phỏng các quy luật phân bố của cá thể, quần thể loài trong lâm phần, hệ sinh thái; mô hình hóa các mối tương quan sinh học giữa các nhân tố điều tra cá thể, lâm phần, hệ sinh thái rừng; xác định kiểu dạng phân bố trên mặt đất rừng của cây rừng; ước lượng số cá thể, bầy đàn trong nghiên cứu đa dạng sinh học; nghiên cứu các mối quan hệ phức tạp giữa biến phụ thuộc và nhiều biến ảnh hưởng, mối quan hệ qua lại và ảnh hưởng tổng hợp của chúng đến biến phụ thuộc; nghiên cứu các mối quan hệ nhân quả trong hệ sinh thái và giữa các yếu tố của hệ sinh thái với các yếu tố của hệ xã hội nhân văn; thống kê còn được sử dụng để đánh giá tác động môi trường, thẩm định chéo các mô hình, thẩm  định các quy luật được mô phỏng.

Như vậy có thể thấy, nghiên cứu thực nghiệm lâm nghiệp, quản lý tài nguyên môi trường rừng áp dụng thống kê tin học vô cùng phong phú, đa dạng và mang lại các kết quả hết sức có ý nghĩa. Ít tìm thấy được một khảo sát, đánh giá, thử nghiệm, nghiên cứu nào trong lâm nghiệp mà không thể áp dụng thống kê hoặc không áp dụng thống kê để rút ra kết luận khách quan, có cơ sở khoa học và thực tế.

Giáo trình này không đi sâu vào lý thuyết toán xác suất thống kê, mà tập trung vào giải quyết các vấn đề thống kê được đặt ra trong thực tế nghiên cứu, thử nghiệm của ngành lâm nghiệp; trên cơ sở ứng dụng một cách tổng hợp các chương trình thống kê như: phần mềm mã nguồn mở R, các chương trình thống kê chuyên nghiệp như Statgraphics, SPSS và cả đơn giản, phổ biến 5 như là Excel.

Giáo trình gồm: 8 chương và phụ lục với 14 bộ dữ liệu thực hành toàn bộ các nội dung xử lý thống kê trên máy tính:

Chương 1: Tổng quan về ứng dụng tin học thống kê trong lâm nghiệp;

Chương 2: Khoa học rút mẫu thống kê và thiết kế các thử nghiệm lâm nghiệp;

Chương 3: Tin học thống kê mô tả và kiểm tra phân bố chuẩn, dung lượng của mẫu;

Chương 4: Tin học thống kê so sánh;

Chương 5: Tin học thống kê trong phân tích phương sai;

Chương 6: Tin học trong ứng dụng tiêu chuẩn phi tham số để so sánh các mẫu quan sát độc lập hoặc có liên hệ;

Chương 7: Tin học ứng dụng trong mô hình hóa rừng;

Chương 8: Tin học thống kê chuyên đề.

Trong đó, đối với đào tạo sau Đại học, có thể sử dụng toàn bộ nội dung để giới thiệu có hệ thống, nhưng tập trung vào các chương nâng cao như chương 1, 2, 7 và 8. Ở bậc Đại học giới thiệu toàn bộ nội dung có tính hệ thống nhưng giới hạn: i) Sử dụng phần mềm Excel hoặc Statgraphics hoặc SPSS; ii) Chương 7 chỉ giới hạn trong mô hình hóa theo hàm tuyến tính, hoặc tuyến tính hóa áp dụng phương pháp bình phương tối thiểu và chương 8 chỉ là lựa chọn nhằm nâng cao.

Với kinh nghiệm thực tế của mình, tác giả cố gắng đúc kết tất cả những kết quả nghiên cứu về ứng dụng thống kê tin học trong ngành lâm nghiệp và quản lý tài nguyên, môi trường rừng để minh họa trong giáo trình. Cuốn sách này không phải là lý thuyết thống kê; mà nó được viết, trình bày, thảo luận theo các chủ đề, xu hướng, nhu cầu áp dụng trong thực tiễn cả cho nghiên cứu, thực nghiệm và bố trí sản xuất thử trong thực tế; theo chiều hướng từ đơn giản đến phức tạp và đặc biệt là lựa chọn để ứng dụng công nghệ tin học trong xử lý dữ liệu và đưa ra kết luận khách quan, có độ tin cậy.

Hy vọng cuốn sách này sẽ giúp cho người đọc có thêm tài liệu để áp dụng và phát triển ứng dụng công nghệ thông tin, tin học trong phân tích, xử lý thống kê đa dạng trong ngành lâm nghiệp, quản lý tài nguyên môi trường rừng.

*Thành phố Corvallis, Bang Oregon, USA,*
Ngày 18 tháng 04 năm 2016
**Bảo Huy**

# MỤC LỤC

Chương 4

## TIN HỌC THỐNG KÊ SO SÁNH



Chương 5

## TIN HỌC TRONG PHÂN TÍCH PHƯƠNG SAI (ANALYSIS OF VARIANCE - ANOVA)



Chương 6

## TIN HỌC TRONG ỨNG DỤNG TIÊU CHUẨN PHI THAM SỐ ĐỂ SO SÁNH CÁC MẪU QUAN SÁT ĐỘC LẬP HOẶC CÓ LIÊN HỆ



Chương 7

## TIN HỌC ỨNG DỤNG TRONG MÔ HÌNH HÓA RỪNG (FOREST MODELLING)

Chương 8

## TIN HỌC THỐNG KÊ CHUYÊN ĐỀ

Chương 1

# TỔNG QUAN VỀ ỨNG DỤNG TIN HỌC THỐNG KÊ TRONG LÂM NGHIỆP

## 1.1 Phương pháp luận về tin học thống kê trong nghiên cứu khoa học lâm nghiệp

Khái quát về phương pháp luận trong tiếp cận ứng dụng tin học thống kê trong lâm nghiệp, quản lý tài nguyên và môi trường rừng được Vanclay (1994), Nguyễn Hải Tuất, (1982), Jayaraman (1999), Twery, (2004) khái quát và được tổng hợp lại sau đây là một tham khảo tốt cho kiến thức tổng quát về áp dụng khoa học thống kê tin học trong lâm nghiệp. Giống như các ngành khoa học khác, nghiên cứu lâm nghiệp cũng dựa trên phương pháp khoa học phổ biến là tiếp cận quy nạp – suy diễn. Trong lâm nghiệp các quy luật tự nhiên được quan sát và phát hiện ra các quy luật trên cơ sở quy nạp theo các quy luật phân bố, tương quan; đồng thời với một đối tượng rộng lớn là rừng, trong điều tra đánh giá tài nguyên chúng ta không thể đo đếm toàn bộ cá thể, do đó phương pháp rút mẫu và ước lượng thống kê – suy diễn cho tổng thể rộng lớn hơn với một độ tin cậy cho trước. Cả hai trường hợp này đều cần dùng đến các phương pháp thống kê từ rút mẫu, đến tính toán các đặc điểm của mẫu, ước lượng biến động, mô phỏng quy luật phân bố, tương quan và cuối cùng là thẩm định độ tin cậy của việc quy nạp – suy diễn.

Phương pháp thống kê luôn liên quan đến kiểm định một giả thuyết đặt ra (Nguyễn Hải Tuất, 1982), ví dụ qua quan sát thấy rằng các cây trồng ở ranh giới thường tốt hơn các cây bên trong, như vậy ở đây giả thuyết rằng các cây bên ngoài thu được nhiều ánh sáng hơn ở các hướng bên ngoài, và lúc này cần đo đếm số liệu sinh trưởng cây trong và ngoài để từ đó sử dụng thống kê đánh giá, thẩm định giả thuyết. Ở đây còn nhiều ví dụ khác như là giả thuyết bón phân sẽ tốt hơn cho cây trồng, mật độ trồng khác nhau sẽ cho sinh trưởng khác nhau,... từ đó cần thu thập mẫu và tính toán thống kê để khẳng định hay phủ nhận giả thuyết.

Hai đặc điểm chính của phương pháp khoa học thống kê là lặp lại và khách quan. Các thí nghiệm khi lặp đi lặp lại trong điều kiện tương tự không mang lại kết quả giống nhau, vì phải chịu sự biến động tính chất ngẫu nhiên. Đo lặp lại các cây rừng trồng cùng tuổi, cùng nguồn giống, gien nhưng chắc chắn sẽ cho giá trị sai khác bởi tác động ngẫu nhiên. Vì vậy, cần có đo thí nghiệm, đo đếm, điều tra thu thập số liệu lặp lại đủ lớn để bảo đảm phản ảnh đúng đối tượng quan sát, thí nghiệm; trong thống kê gọi là rút mẫu lặp lại với một độ tin cậy hoặc sai số cho trước. Các khoa học thống kê là hữu ích trong khách quan lựa chọn một mẫu, bố trí thí nghiệm nhằm bảo đảm rằng

dữ liệu thu được có có sai số từ biến động ngẫu nhiên, không phải do các yếu tố không kiểm soát được chi phối.

Hai khía cạnh thực tế chính của nghiên cứu khoa học thống kê là tập hợp các dữ liệu và giải thích các dữ liệu thu thập được. Các dữ liệu thu thập được cô đọng và thông tin hữu ích chiết xuất thông qua kỹ thuật của suy luận thống kê. Đặc biệt trong giai đoạn vừa qua với sự phát triển mạnh mẻ của tin học và máy tính đã làm thay đổi các nghiên cứu lâm nghiệp truyền thống. Các nghiên cứu mô phỏng, quy nạp, mô hình hóa là đặc biệt hữu ích và có giá trị trong quản lý tài nguyên thiên nhiên, vì nó thay cho các thí nghiệm quy mô lớn tốn kém và mất nhiều thời gian. Mô hình toán thống kê trên cơ sở ứng dụng tin học là một chủ đề ngày càng được áp dụng rộng rải, không chỉ để phản ảnh các mối quan hệ hình thái như đường kính – chiều cao cây mà còn chỉ ra sự tác động qua lại phức tạp của các thành phần bên trong hệ sinh thái rừng và ảnh hưởng của các nhân tố chủ đạo bên trong và ngoài hệ sinh thái rừng đến động thái, sinh trưởng, phát triển rừng (Twery, 2004; Vanclay, 1994).

Lý thuyết thống kê của các cuộc điều tra dựa trên lấy mẫu ngẫu nhiên, từ đó so sánh các quần thể khác nhau, nghiên cứu về sự phân bố mô hình của sinh vật hoặc cho việc tìm kiếm các mối quan hệ giữa các biến số. Các nghiên cứu khoa học như sinh thái rừng và sinh học động thực vật hoang dã nói chung là thuộc về thể loại này (Twery, 2004).

Theo các thử nghiệm lâm sinh trong trồng rừng và vườn ươm và thử nghiệm trong phòng thí nghiệm là những ví dụ điển hình các thí nghiệm về lâm nghiệp. Các thí nghiệm này nhằm phục vụ kiểm tra một hay nhiều giả thuyết trong điều kiện kiểm soát các yếu tố khác. Các nguyên tắc cơ bản của thí nghiệm là ngẫu nhiên, lặp lại và có kiểm soát các nhân tố ảnh hưởng là những điều kiện tiên quyết để có ước tính phù hợp và giảm sai số.

Thử nghiệm trên các trạng thái của một hệ thống với một mô hình theo thời gian được gọi là mô phỏng. Các mối tương quan giữa các yếu tố của một hệ thống sinh thái rừng là có thể diễn tả qua phương trình toán học và do đó tình trạng của hệ thống trong điều kiện thay thế là có thể dự đoán qua các mô hình toán học. Đây là một lĩnh vực khoa học còn rất mới mẻ, được áp dụng rất ít trong nước, tuy nhiên lại rất hữu ích trong quản lý tài nguyên cảnh quan sinh thái rừng, vì vậy cần quan tâm thúc đẩy, tổ chức nhiều hơn các nghiên cứu.

Cuối cùng (Jayaraman, 1999; Twery, 2004) cho thấy nghiên cứu liên quan đến rừng có từ cấp độ phân tử cho đến toàn bộ sinh quyển. Các nghiên cứu từ nhỏ đến lớn đều hoàn toàn có thể áp dụng thống kê để thẩm định giả thuyết, mô phỏng các quy luật của hệ thống sinh thái từ đơn giản đến phức tạp. Các dữ liệu đầu vào để xử lý chủ yếu thu được từ các phương pháp điều tra, thí nghiệm. Khung logic của cách tiếp cận khoa học và những suy luận thống kê ngày nay có thể không khác nhiều so với truyền thống, tuy nhiên sự phối hợp giữa khoa học thống kê lâm nghiệp với tiến bộ tin học đã mở ra những hướng nghiên cứu mới, có giá trị, có tầm bao quát rộng giúp cho việc xử lý, quản lý các hệ thống sinh thái rừng theo hướng bền vững.

## 1.2 Cơ sở để kết luận thống kê lâm nghiệp

Thống kê được áp dụng dựa vào một số cơ sở nền tảng chủ yếu như là: Xác suất thống kê, phân bố tần số, so sánh trung bình, phương sai, các mối quan hệ. Các kết luận thống kê dựa trên kiểm định các giả thuyết nhằm đánh giá sự sai khác giữa các trung bình, giữa các phương sai –

phân tích phương sai, đánh giá sai khác về tỷ lệ, đánh giá độ tin cậy của các mô hình tương quan – hồi quy, thẩm định sự phù hợp của các mô hình toán (Twery, 2004; Jayaraman, 1999). Tất cả các kết luận thống kê là dựa vào giả thuyết cho trước và kết quả kiểm định giả thuyết đó dựa vào phân tích dữ liệu quan sát và so sánh với giá trị của phân bố xác suất lý thuyết với sai số hoặc độ tin cậy cho trước.

Theo (Jayaraman, 1999; Nguyễn Hải Tuất, 1982) thì các bước chung liên quan đến kiểm định giả thuyết thử nghiệm như sau: (i) Nêu lên giả thuyết bằng không (Ho - Null hypothesis); (ii) Lựa chọn một tiêu chuẩn thống kê liên quan để kiểm tra giả thuyết $H_0$; (iii) Xác định mức ý nghĩa và kích thước mẫu; (iv) Phân bố mẫu; (v) Xác định phạm vi giá trị thống kê để bác bỏ $H_o$; (vi) Tính toán giá trị theo tiêu chuẩn thống kê đã lựa chọn và kiểm tra với giá trị lý thuyết để có kết luận thống kê.

*i) Giả thuyết $H_0$:* Đây là giả thuyết không có sự khác biệt. Ví dụ so sánh hai hay nhiều trung bình sinh trưởng đường kính của các rừng có nguồn gốc giống khác nhau; giả thuyết $H_0$ lúc này là chưa có sự khác biệt giữa các trung bình $\mu_1$, $\mu_2 \dots \mu_n$, hay nói khác các trung bình là gần như nhau và có nghĩa là nguồn gốc khác nhau chưa có sự khác biệt rõ rệt về đường kính cây trồng rừng. Công thức của $H_0$ như sau:

Trường hợp so sánh hai số trung bình thì $H_o$ sẽ là:

$$\text{Ho: } \bar{\mu}_1 = \bar{\mu}_2 \text{ hay Ho: } \bar{\mu}_1 - \bar{\mu}_2 = 0 \quad (1.1)$$

Trường hợp so sánh nhiều số trung bình, thì $H_o$ sẽ là:

$$\text{Ho} = \bar{\mu}_1 = \bar{\mu}_2 = \cdots = \bar{\mu}_n \quad (1.2)$$

Giả thuyết $H_0$ còn có thể hiểu là giả định là hai trung bình hoặc nhiều trung bình bằng nhau, hoặc hai hoặc nhiều phân bố tần số là cùng quy luật. Có nghĩa là khi chúng ta muốn so sánh sự khác biệt, thì giả thuyết $H_0$ được sử dụng để đối nghịch với giả thuyết $H_1$. Giả thuyết $H_1$ chỉ ra có sự khác biệt rõ rệt giữa các trung bình so sánh theo các công thức sau:

Trường hợp so sánh hai số trung bình thì $H_1$ sẽ là:

$$H_1: \bar{\mu}_1 \# \bar{\mu}_2 \text{ hay } H_1: \bar{\mu}_1 - \bar{\mu}_2 \# 0 \quad (1.3)$$

Trường hợp so sánh nhiều số trung bình, thì giả thuyết $H_1$ sẽ là có sự khác biệt ít nhất của hai số trung bình bất kỳ i và j trong dãy số trung bình so sánh:

$$H_1: \bar{\mu}_i \# \bar{\mu}_j \text{ hay } H_1: \bar{\mu}_i - \bar{\mu}_j \# 0 \quad (1.4)$$

Như vậy khi so sánh giả thuyết $H_1$ sẽ được chấp nhận (tức là có sự sai khác) nếu giả thuyết $H_0$ bị từ chối.

*ii) Lựa chọn một tiêu chuẩn thống kê để kiểm tra $H_o$:* Tiêu chuẩn thống kê để kiểm tra rất đa dạng, tùy thuộc vào $H_o$, số mẫu, công thức, nhân tố so sánh và phương pháp cũng như số lượng

mẫu thu thập được. Ví dụ kiểm tra sự sai khác của hai đến nhiều trung bình mẫu với dung lượng quan sát ngẫu nhiên, đủ lớn để phân bố mẫu tiệm cận chuẩn thì lúc này các tiêu chuẩn thống kê tham số như t, U, phân tích phương sai theo tiêu chuẩn F (ANOVA: Analysis of Variances) được áp dụng; ngược lại mẫu chưa đủ lớn hoặc chưa bảo đảm có phân bố chuẩn thì các tiêu chuẩn phi tham số có thể được sử dụng. Đánh giá hệ số tương quan, xác định và tiêu chuẩn F, so sánh hai hoặc nhiều dãy phân bố tần số bằng tiêu chuẩn $\chi^2$.

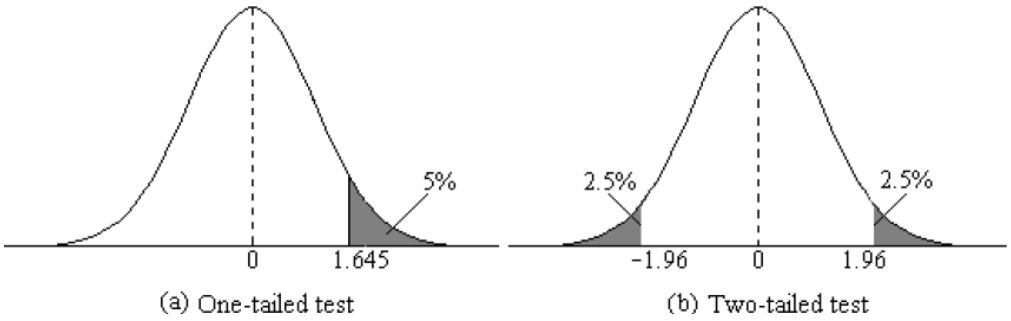
*iii) Xác định mức ý nghĩa thống kê và dung lượng mẫu:* Khi giả thuyết $H_0$ và tiêu chuẩn thống kê đã được lựa chọn, mức ý nghĩa thống kê (α) và dung lượng mẫu cần được xác định. Mức ý nghĩa thống kê α dùng để ra quyết dịnh chấp nhận hay bác bỏ giả thuyết $H_0$. Mức α thường được áp dụng là 0.05 hoặc 0.01 ứng với độ tin cậy 95% hoặc 90%. Nếu mức xác suất tính toán thông qua dữ liệu quan sát có $P_{value} < \alpha$ thì kết luận là bác bỏ $H_o$, hay nói khác chấp nhận giả thuyết $H_1$ là hai hoặc nhiều mẫu, dãy phân bố quan sát có sự khác biệt rõ rệt. Trên cơ sở mức ý nghĩa α, tính toán được dung lượng mẫu cần quan sát, bố trí thí nghiệm để phân bố mẫu tiệm cận chuẩn và có thế áp dụng các tiêu chuẩn thống kê tham số. Tuy nhiên trong thực tế một số điều tra hoặc thí nghiệm không thể thu thập hoặc bố trí thí nghiệm đủ lớn vì liên quan đến nguồn lực, thời gian có hạn; lúc này các tiêu chuẩn phi tham số cần được áp dụng và dung lượng mẫy thu thập, thí nghiệm cũng cần tuân theo yêu cầu của các tiêu chuẩn này.

*iv) Chọn dạng phân bố mẫu, mô hình toán quan hệ:* Một khi đã lựa chọn tiêu chuẩn thống kê để kiểm tra $H_o$, tiếp theo là xác định dạng phân bố của mẫu. Hay nó khác là để kiểm tra xem hai hoặc nhiều dãy phân bố có cùng chung một tổng thể (chấp nhận $H_o$). Ví dụ phân bố lý thuyết Weibull được lựa chọn để kiểm định giả thuyết là phân bố N/DBH có thể tuân theo quy luật này, tức là $H_0$ được chấp nhận; lúc này tiêu chuẩn $\chi^2$ có thể được áp dụng. Hay mối quan hệ giữa sinh khối và carbon cây rừng với các nhân tố điều tra cây như DBH, H, khối lượng thể tích gỗ (WD, $g/cm^3$) tuân theo hàm mũ power, lúc ngay tiêu chuẩn F dùng để kiểm tra hệ số xác định $R^2$ hoặc tiêu chuẩn phi tham số, hoặc tiêu chuẩn $\chi^2$ được áp dụng để kiểm tra phân bố giá trị lý thuyết so với quan sát có tuân theo mô hình toán hay không?

*v) Xác định phạm vi giá trị thống kê để bác bỏ $H_o$:* Giá trị xác suất lựa chọn để bác bỏ giả thuyết $H_0$ là α. Nếu giá trị xác suất của dữ liệu quan sát, thực nghiệm là $P_{value} < \alpha$ thì giả thuyết $H_0$ bị bác bỏ, hay nói khác chấp nhận giả thuyết ngược lại là $H_1$: Các trung bình, các dãy phân bố, các dãy dữ liệu ước lượng qua mô hình toán, phân bố,... có sự sai khác nhau rõ rệt hay không cùng một tổng thể. Như vậy giá trị α xác định trước quyết định đến việc chấp nhận hay bác bỏ Ho. Thông thường đối với các ngành kỹ thuật đòi hỏi độ chính xác cao thì α = 0.01 (1%), mức trung bình thường áp dụng trong lâm nghiệp thì α = 0.05 (5%) và trong trường hợp nghiên cứu các mối quan hệ sinh thái, xã hội phức tạp, chưa rõ ràng thì có thể chấp nhận α = 0.10 (10%).

*iv) Kết luận thống kê:* Kết luận thống kê được xác định thông qua phạm vi giá trị thống kê để bác bỏ $H_o$ nói trên. Trong thực tế thì còn cần phân biệt khi so sánh theo một (one tail) hay hai chiều (two tails). So sánh một chiều có nghĩa là các so sánh trung bình hoặc phân bố mẫu theo hướng lớn hơn hoặc nhỏ thua mẫu khác hay lý thuyết. So sánh hai chiều là không quan tâm đến các mẫu phải lớn hay nhỏ thua nhau mà chỉ quan tâm chúng có đồng nhất hay không (chấp nhận $H_o$) hay thực sự sai khác (chấp nhận $H_1$).

Hình 1.1 chỉ ra rằng cùng với ý nghĩa thống kê α = 5%, thì kiểm tra sai khác theo một chiều có vùng bác bỏ $H_o$ về một phía (phải hay trái), với giá trị thống kê ví dụ là t > 1.645; trong khi đó kiểm tra hai chiều có vùng bác bỏ $H_o$ nằm về cả hai phía của phân bố xác suất, ví dụ lúc này t < - 1.96 hoặc t > + 1.96.




**Hình 1.1.** Vùng bác bỏ giả thuyết $H_0$ với mức ý nghĩa thống kê α = 5%. (a) Kiểm tra theo một chiều và (b) Kiểm tra theo hai chiều *(Nguồn: Jayaraman, 1999)*

## 1.3 Giới thiệu các chương trình xử lý thống kê

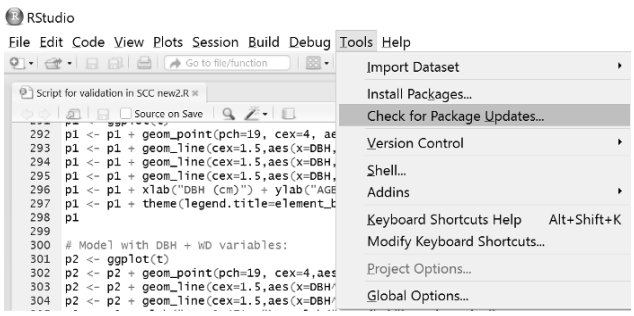
### 1.3.1 Chương trình mã nguồn mở R

Chương trình R là một chương trình mã nguồn mở, do vậy miễn phí và người sử dụng có thể trao đổi phát triển các ứng dụng. Ngay cả nước phát triển khoa học thống kê ứng dụng mạnh nhất như Hoa Kỳ, nơi đã sản xuất hàng loạt các phần mềm thống kê chuyên nghiệp, thì các nhà khoa học cũng ưa dùng R. Lý do đầu tiên là không có chí phí, thuận tiện trong đào tạo, giúp người học giảm chi phí mua phần mềm; nhưng lý do sâu xa nhất là nó cho phép người sử dụng tự do mở rộng các ứng dụng, trình bày các kết quả, đồ thị, biểu đồ theo mong muốn, không bị đóng khung như các phần mềm chuyên nghiệp.

R có khả năng áp dụng cho toàn bộ các vấn đề thống kê ứng dụng, mà còn có khả năng liên kết với hệ thống thông tin địa lý (GIS), xử lý ảnh viễn thám. Có thể nói sử dụng R trong thống kê là chuyên nghiệp hơn cả các phần mềm chuyên nghiệp hiện có.

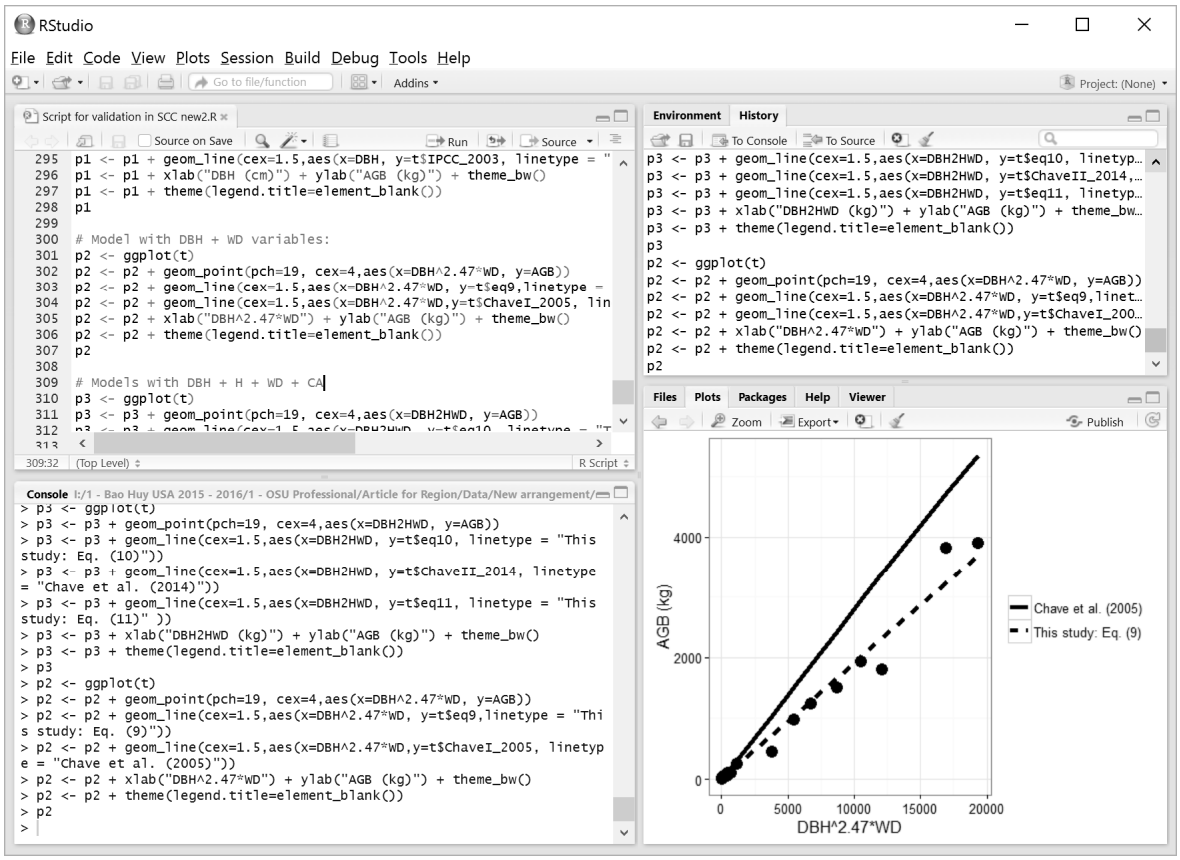
R là một phần mềm tự do, mã nguồn mở, được đóng góp bởi nhiều người tham gia (R-Core-Team, 2015). R là một ngôn ngữ lập trình và môi trường phần mềm dành cho tính toán và đồ họa thống kê. Đến nay R do Core Team chịu trách nhiệm phát triển. Tên của chương trình này một phần lấy từ chữ cái đầu của hai tác giả (Robert Gentleman và Ross Ihaka). Ngôn ngữ R đã trở thành một tiêu chuẩn để ứng dụng thống kê và được sử dụng rộng rãi để phát triển phần mềm thống kê và phân tích dữ liệu. R có các phiên bản dịch sẵn cho nhiều hệ điều hành khác nhau. R sử dụng giao diện dòng lệnh, tuy cũng có một vài giao diện đồ họa người dùng dành cho nó (Wikipedia, 2015). Trong tài liệu này giới thiệu ứng dụng R Studio, vì vậy trước hết cần cài đặt R, sau đó cài tiếp tục R Studio. Các chương trình này tải miễn phí từ web site của R: https://www.r-project.org/ và R Studio: https://www.rstudio.com/. Sau cài đặt, khởi động R sẽ có giao diện như Hình 1.2.

Khi bắt đầu hoặc định kỳ, cần kiểm tra để cập nhật được các gói chương trình cho Core Team và người dùng phát triển như sau:




- Cập nhật các gói chương trình (Packages): Tool/Check for Packages Updates

- Cài đặt một Package đã biết: Tool/Install Packages... Nhập tên của chương trình.




**Hình 1.2.** Giao diện chương trình R Studio

Gồm 4 cửa sổ: Trên trái: nơi viết các dòng lệnh, chương trình; dưới trái: nơi xuất ra các kết quả xử lý thống kê; trên phải: ghi lại lịch sử thực hiện các lệnh; dưới phải: có các tùy chọn như xuất ra đồ thị (Plot), thư mục, file, các gói chương trình đã có trong R (Packages), hướng dẫn (Help) dựa vào Search.

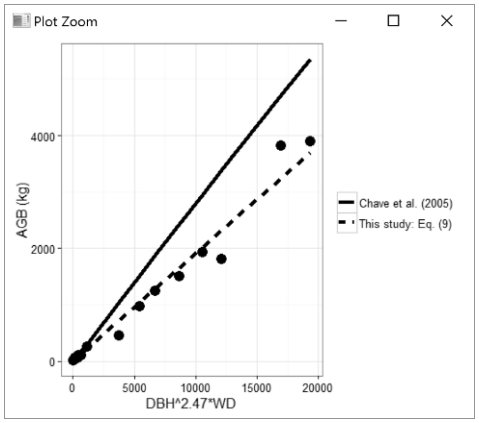
Dữ liệu đầu vào cho R để thuận tiện thường được nhập trong Excel dưới dạng file: *.txt. Trong R viết các lệnh để chỉ đường dẫn, thư mục và tên file như sau:

**Các dòng lệnh R để chỉ đường dẫn, thư mục và file dữ liệu dạng *.txt:**

```
# Xóa bộ nhớ:
rm(list=ls())
# Xóa các đồ thị:
dev.off()
```

```
# Chỉ đường dẫn đến dữ liệu (thay \ bằng / sử dụng Edit/Find):
setwd("I:/1 - Bao Huy USA 2015 - 2016/1 - OSU Professional/Article for Region/Data/New arrangement")
# Nhập file dữ liệu dạng *.txt:
t <- read.table("t_va.txt", header=T,sep="\t",stringsAsFactors = FALSE)
```

Để chép kết quả xử lý thống kê, chỉ quét chuột và kích chuột phải để chọn copy, sau đó dán (Paste) vào văn bản.




Để chép đồ thị trong R, trong của sổ kích vào nút Zoom để mở cửa sổ đồ thị riêng, sau đó kích chuột phải để copy và paste vào nơi thích hợp.

### 1.3.2 Chương trình Statgraphics

Đây là một phần mềm xứ lý thống kê chuyên nghiệp, bao gồm hầu hết các nội dung phân tích thống kê ứng dụng. Bao gồm các xử lý chính:

- Tạo lập cơ sở dữ liệu dưới dạng bảng tính

- Tính toán các đặc trưng mẫu, vẽ sơ đồ, đồ thị quan hệ, đồ thị phân bố mẫu

- So sánh hai hay nhiều mẫu bằng các tiêu chuẩn thống kê t, U, F và nhiều tiêu chuẩn phi tham số khác.

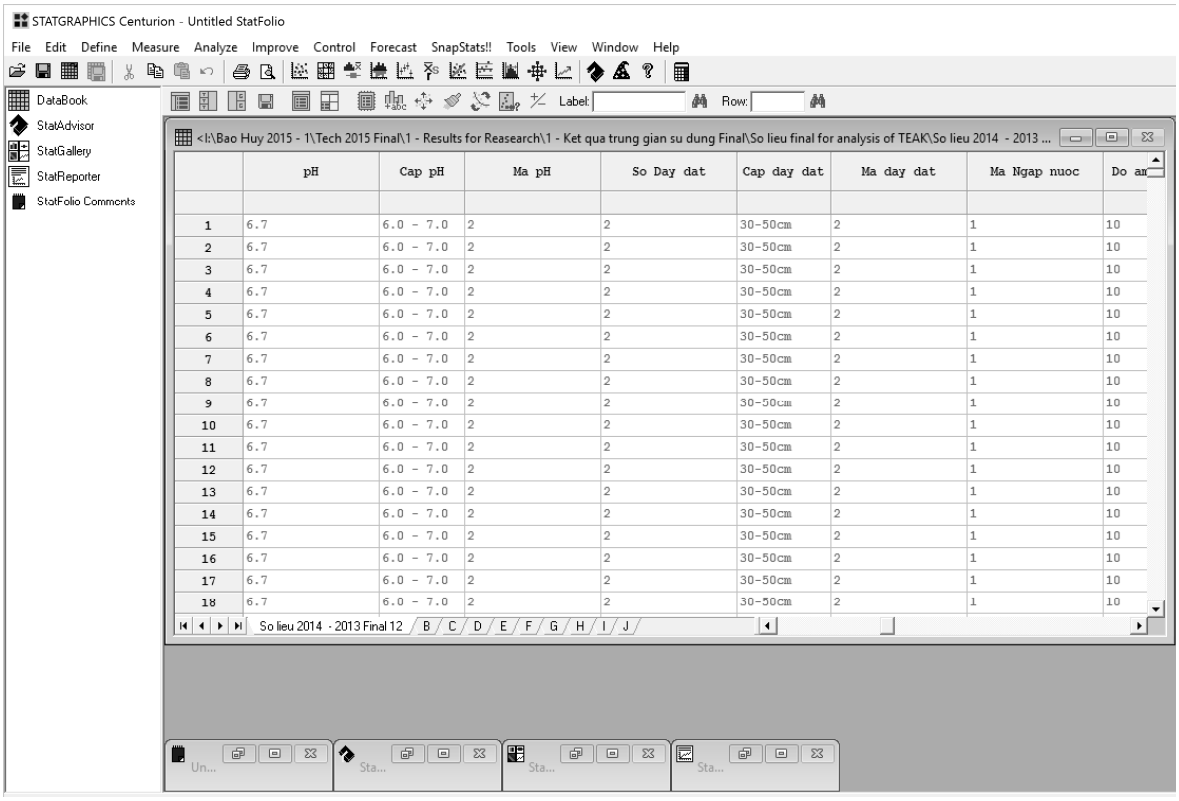
- Phân tích phương sai ANOVA.

- Kiểm tra tính chuẩn của dữ liệu và đổi biến số.

- Thiết lập các mô hình hồi quy tuyến tính hay phi tuyến tính từ một cho đến nhiều lớp, tổ hợp biến, có trọng số. Với cách xử lý đa dạng để chọn lựa được các biến ảnh hưởng đến một hậu quả (biến phụ thuộc).

Chi tiết về áp dụng Statgraphics tham khảo trong tài liệu hướng dẫn (StatPoint-Inc, 2005).

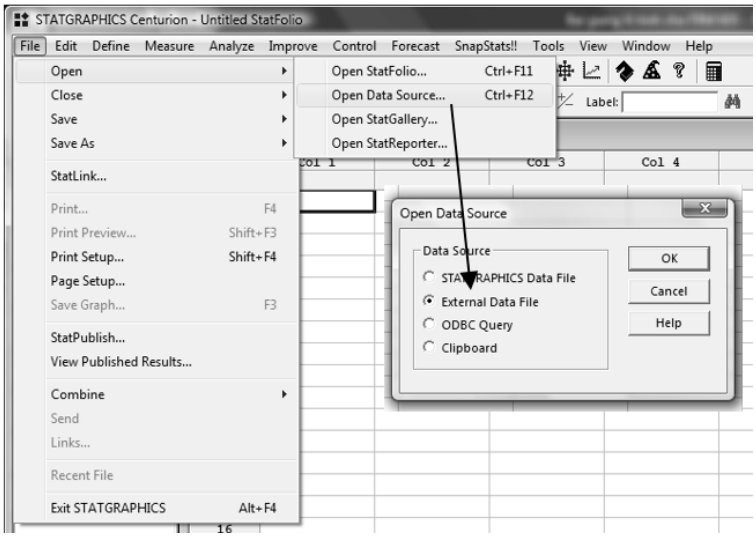
Giao diện chính của Statgraphics như ở Hình 1.3, trong đó có dòng Menu chính như các phần mềm thông thường, bên dưới là các cửa sổ bao gồm dữ liệu, các kết quả xử lý.




| | pH | Cap pH | Ma pH | So Day dat | Cap day dat | Ma day dat | Ma Ngap nuoc | Do am |
|---|---|---|---|---|---|---|---|---|
| 1 | 6.7 | 6.0 - 7.0 | 2 | 2 | 30-50cm | 2 | 1 | 10 |
| 2 | 6.7 | 6.0 - 7.0 | 2 | 2 | 30-50cm | 2 | 1 | 10 |
| 3 | 6.7 | 6.0 - 7.0 | 2 | 2 | 30-50cm | 2 | 1 | 10 |
| 4 | 6.7 | 6.0 - 7.0 | 2 | 2 | 30-50cm | 2 | 1 | 10 |
| 5 | 6.7 | 6.0 - 7.0 | 2 | 2 | 30-50cm | 2 | 1 | 10 |
| 6 | 6.7 | 6.0 - 7.0 | 2 | 2 | 30-50cm | 2 | 1 | 10 |
| 7 | 6.7 | 6.0 - 7.0 | 2 | 2 | 30-50cm | 2 | 1 | 10 |
| 8 | 6.7 | 6.0 - 7.0 | 2 | 2 | 30-50cm | 2 | 1 | 10 |
| 9 | 6.7 | 6.0 - 7.0 | 2 | 2 | 30-50cm | 2 | 1 | 10 |
| 10 | 6.7 | 6.0 - 7.0 | 2 | 2 | 30-50cm | 2 | 1 | 10 |
| 11 | 6.7 | 6.0 - 7.0 | 2 | 2 | 30-50cm | 2 | 1 | 10 |
| 12 | 6.7 | 6.0 - 7.0 | 2 | 2 | 30-50cm | 2 | 1 | 10 |
| 13 | 6.7 | 6.0 - 7.0 | 2 | 2 | 30-50cm | 2 | 1 | 10 |
| 14 | 6.7 | 6.0 - 7.0 | 2 | 2 | 30-50cm | 2 | 1 | 10 |
| 15 | 6.7 | 6.0 - 7.0 | 2 | 2 | 30-50cm | 2 | 1 | 10 |
| 16 | 6.7 | 6.0 - 7.0 | 2 | 2 | 30-50cm | 2 | 1 | 10 |
| 17 | 6.7 | 6.0 - 7.0 | 2 | 2 | 30-50cm | 2 | 1 | 10 |
| 18 | 6.7 | 6.0 - 7.0 | 2 | 2 | 30-50cm | 2 | 1 | 10 |

**Hình 1.3.** Giao diện chính của Statgraphics

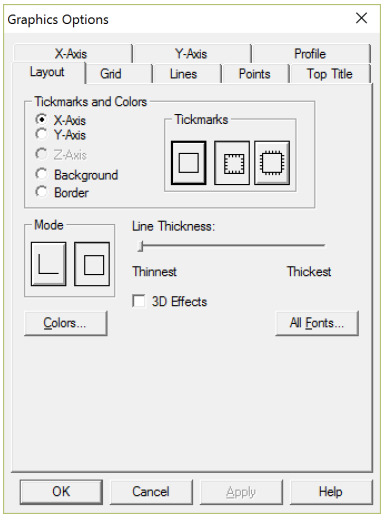
Trong Statgraphics Centurion, số liệu đầu vào có thể được nhập trực tiếp trong cửa sở bảng tính, song với các làm này đôi khi không thuận tiện trong các bước xử lý số liệu thô như đổi biến số, tính các biến trung gian, mã hóa biến số. Do đó, thông thường nên tạo lập cơ sở dữ liệu trong bảng tính Excel để có thể sử dụng những chức năng bảng tính mạnh của nó trong xử lý dữ liệu thô, tạo lập cơ sở dữ liệu; sau đó sẽ nhập vào Statgraphics Centurion để tính toán, thiết lập mô hình,... Cơ sở dữ liệu lập trong Excel cần lưu dưới dạng phiên bản của Excel 97 – 2003, vì nó chưa nhận được file Excel ở version sau đó.




Sau khi nhập dữ liệu trong Excel 97-2000, mở nó trong Statgrahics Centurion như sau: File/Open/Open Data Source; chọn External Data File – OK. Trong hộp thoại mở file, chọn kiểu file Excel và chọn file cần mở đã tạo trước đó.


Chọn các thông số để đọc file dữ liệu Excel: Chọn variable name nếu có tên các trường dữ liệu, chọn số thứ tự sheet dữ liệu trong excel, chọn dữ liệu bắt đầu từ hàng nào: Start row đến kết thúc End row; nhập giá trị cho ô dữ liệu thiếu: Missing value.

Các kết quả xử lý trong Statgraphics được xuất ra ở các cửa sổ, dùng chuột để copy và có thể dán kết quả đến các file khác. Có ba nút điều khiển thường sử dụng trong Statgraphics là: Input Dialog: Dùng gọi lại hộp thoại khai báo các biến để xử lý mà không phải nhập lại các biến đầu vào; Tables: Chọn các loại bảng kết quả phân tích muốn xuất ra; Graphs: Chọn các loại đồ thị.




Đối với các đồ thị, biểu đồ tạo ra, Statgraphics cho phép điều chỉnh tên, trục XY, kích thước đường vẽ, màu sắc,…

Kích chuột phải vào cửa sổ đồ thị để chọn menu Graphics Options. Từ đây có thể thay đổi định dạng đồ thị theo ý muốn.

Kết quả xử lý trong Statgraphics có thể được lưu cùng với cơ sở dữ liệu để thuận tiện cho việc theo dõi, tiếp tục xử lý. Việc lưu và mở file của Statgraphics theo như quản lý file máy tính thông thường.

### 1.3.3 Chương trình SPSS

Đây là một phần mềm chuyên dụng trong xử lý thống kê, bao gồm các chức năng gần giống như Statgraphics, tuy nhiên có ưu nhược điểm khi so sánh với nhau:

Ưu điểm SPSS so với Statgraphics:

- Mã hóa được biến số định tính

- Có các chức năng phân tích so sánh phi tham số

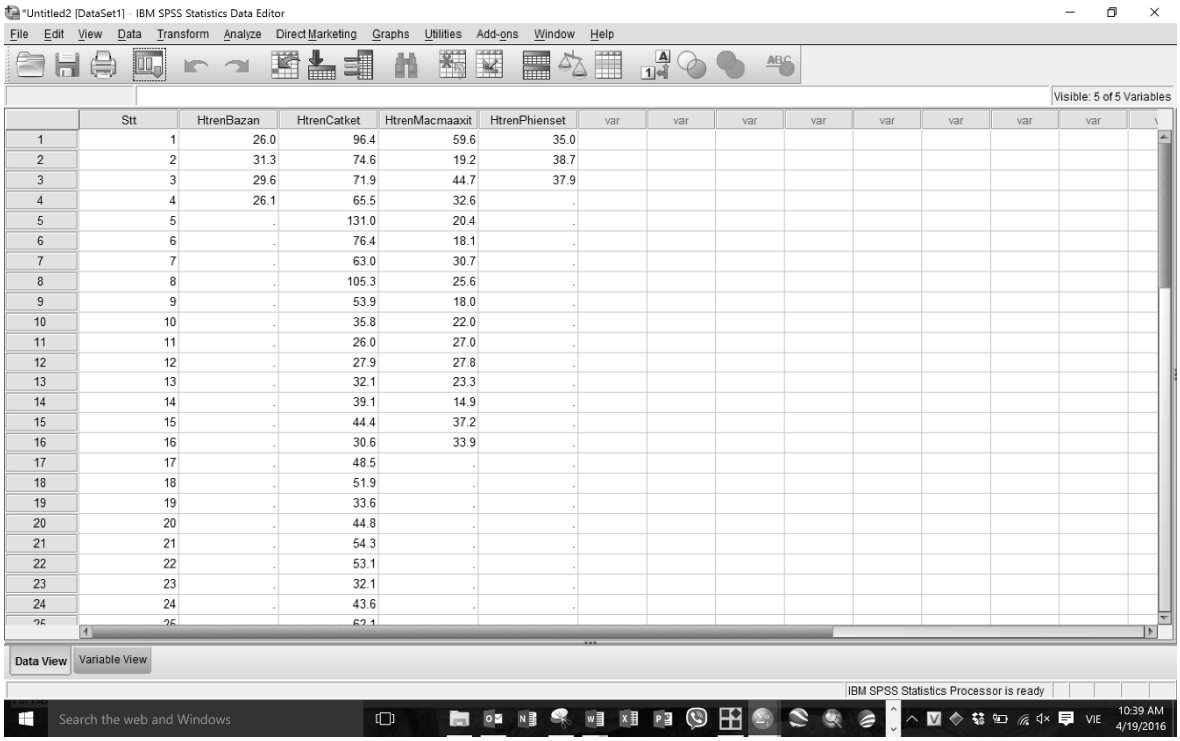
Nhược điểm SPSS so với Statgraphics:

- Không có tư vấn về kết quả phân tích thống kê

- Không đổi được biến số trực tiếp trong các hộp thoại khi phân tích thống kê

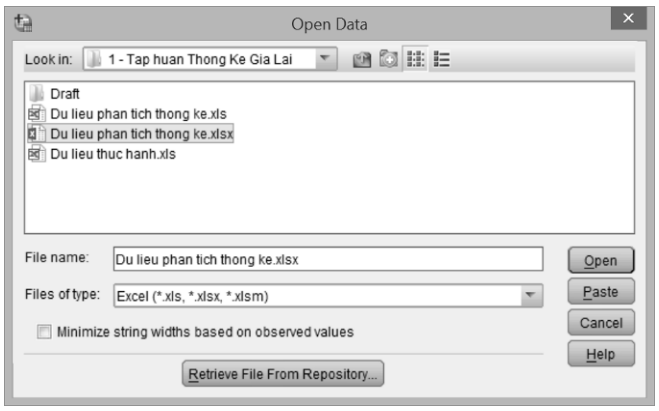
Tham khảo tài liệu (IBM, 2011) để tìm hiểu chi tiết các ứng dụng của IBM SPSS.

Giao diện chính của SPSS như ở Hình 1.4, cũng như Statgraphics, SPSS cũng có dòng Menu và các hộp thoại, dữ liệu và kết quả xử lý được xuất ra theo từng cửa sổ.

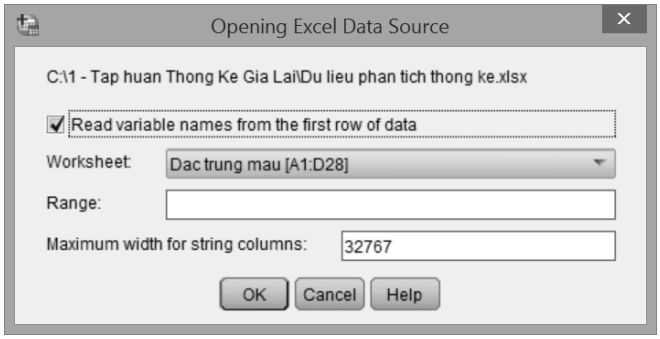



**Hình 1.4.** Giao diện chính của SPSS

Trong SPSS, số liệu đầu vào có thể được nhập trực tiếp từ cửa sổ bảng tính; song với các làm này đôi khi không thuận tiện trong các bước xử lý số liệu thô như đổi biến số, tính các biến trung gian. Do đó thông thường nên tạo lập cơ sở dữ liệu trong bảng tính Excel để có thể sử dụng những chức năng bảng tính mạnh của nó trong xử lý dữ liệu thô, tạo lập cơ sở dữ liệu; sau đó sẽ nhập vào SPSS để tính toán, thiết lập mô hình,...

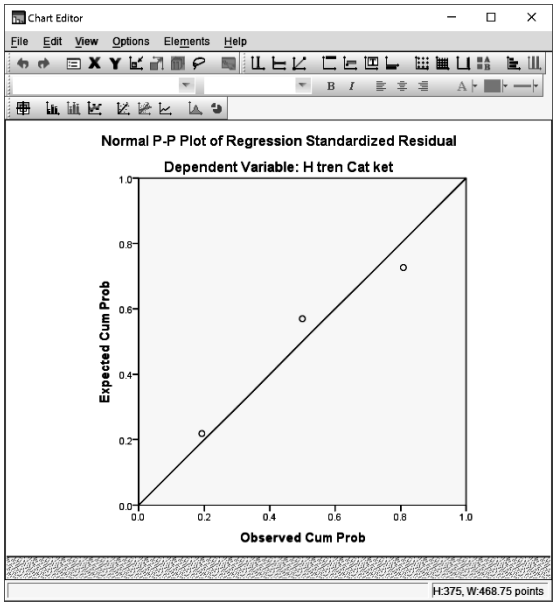



File/Open/Data. Trong hộp thoại mở file, chọn kiểu file Excel và chọn file cần mở đã tạo trước đó




Tiếp theo chọn row đầu tiên làm tên biến và Worksheet làm việc.

Các kết quả xử lý trong SPSS được xuất ra ở các cửa sổ, dùng chuột phải để chọn kết quả vừa xử lý "Select Last Output" và sau đó có thể dán đến file mong muốn.




Để thay đổi định dạng trong đồ thị SPSS, kích đôi chuột trái để mở cửa sổ Chart Editor có các chức năng điều chỉnh tiêu đề, trục XY, định dạng đường, màu sắc,…

Kết quả xử lý trong SPSS cũng có thể được lưu cùng với cơ sở dữ liệu để thuận tiện cho việc theo dõi, tiếp tục xử lý. Việc lưu và mở file của SPSS theo như quản lý file máy tính thông thường.

### 1.3.4 Chương trình thống kê trong Excel

Excel thiết kế sẵn chương trình để xử lý số liệu và phân tích thống kê cơ bản ứng dụng trong hộp thoại Data Analysis, bao gồm các xử lý chính như là:

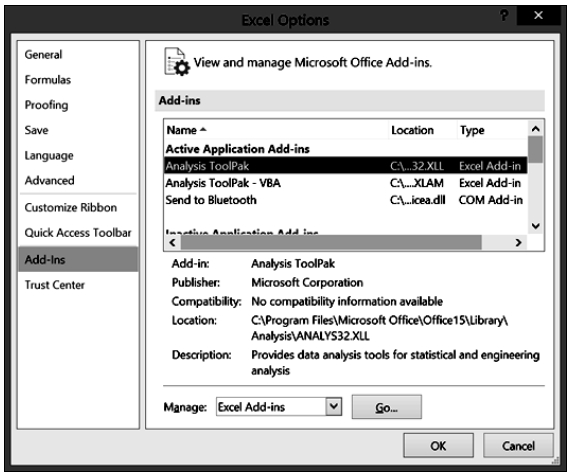
- Xử lý số liệu, tạo bảng tổng hợp dữ liệu: Sắp xếp, tính toán nhanh các bảng tổng hợp từ số liệu thô,...

- Tính toán giá trị lý thuyết của hầu hết các hàm xác suất thống kê như T, F, $\chi^2$

- Cung cấp hàng loạt các hàm tính tính toán thông dụng trong nhiều lĩnh vực toán học, kỹ thuật, tài chính,…

- Chức năng Data Analysis: Dùng để phân tích thống kê như phân tích các đặc trưng mẫu, vẽ biểu đồ phân bố, tiêu chuẩn t để so sánh sự sai khác, phân tích phương sai, ước lượng các tương quan hồi quy tuyến tính một đến nhiều biến số. Đối với hàm phi tuyến, thì áp dụng tuyến tính hóa để ước lượng. Excel không cung cấp phân tích hàm phi tuyến tính.

Nhìn chung đối với nhu cầu phân tích thống kê thông dụng, cơ bản không quá chuyên sâu thì Excel có thể đáp ứng tốt, bao gồm các phân tích thông kê và vẽ các đồ thị.




Để có chức năng phân tích thống kê "Data Analysis" trong Excel, tiến hành mở chế độ phân tích thống kê như sau: File/Option/Add-ins và chọn Analysis ToolPak – Go, sau đó kích chọn chức năng Analysis ToolPak trong hộp thoại - OK.

Một số hàm thống kê thông dụng trong Excel:

- Tính tổng: =Sum(dãy dữ liệu hoặc địa chỉ).

- Tổng bình phương: =Sumq(dãy dữ liệu hoặc địa chỉ).

- Trung bình: =Average(dãy dữ liệu hoặc địa chỉ).

- Lấy giá trị tuyệt đối: =Abs(dãy dữ liệu hoặc địa chỉ).

- Trị lớn nhất, nhỏ nhất: =Max(dãy dữ liệu hoặc địa chỉ), Min(dãy dữ liệu hoặc địa chỉ).

- Các hàm lượng giác: =Cos(dữ liệu), =Sin (dữ liệu), =tan(dữ liệu).

- Hàm mũ, log: =Exp(dữ liệu), =Ln(dữ liệu), =Log(dữ liệu).

- Căn bậc 2: =Sqrt(dữ liệu).

- Sai tiêu chuẩn mẫu chưa hiệu đính: =Stdevp(dãy dữ liệu hoặc địa chỉ); đã hiệu đính =Stdev(dãy dữ liệu hoặc địa chỉ).

- Phương sai mẫu chưa hiệu đính: =Varp(dãy dữ liệu hoặc địa chỉ); đã hiệu đính =Var(dãy dữ liệu hoặc địa chỉ).

- Giai thừa: =Fact(dữ liệu).

- Số Pi: =Pi().

Tính các giá trị lý thuyết trong Excel của một số hàm phân bố thống kê T, F, $\chi^2$:

- Hàm tính giá trị lý thuyết của phân bố T: =tinv(Probability, df)

- Hàm tính giá trị lý thuyết của phân bố $\chi^2$: =chiinv(Probability, df)

- Hàm tính giá trị lý thuyết của phân bố F: =finv(Probability, $df_1$, $df_2$)

Trong đó:

Probability: Mức xác suất sai, thông thường ở mức $\alpha$=0.05; 0.01 hay 0.001.

Degrees Freedom (df): Độ tự do, df = n – l, với l là số mẫu

Chương 2

# KHOA HỌC RÚT MẪU THỐNG KÊ VÀ THIẾT KẾ CÁC THỬ NGHIỆM LÂM NGHIỆP

## 2.1 Tính toán dung lượng mẫu trong điều tra, đánh giá

Trong điều tra tài nguyên thiên nhiên, xã hội, với tổng thể rất rộng lớn, không thể đo đếm, phỏng vấn toàn bộ. Do vậy rút mẫu cần được sử dụng. Rút mẫu để có thể xử lý thống kê là xác định số cá thể, số cây, số con, số ô mẫu, số người,… để điều tra, đo đếm, phỏng vấn lấy thông tin theo một độ tin cậy cho trước. Việc xác định số lượng mẫu, số ô mẫu, số cá thể, số người cần để điều tra, phỏng vấn, thu thập số liệu gọi là tính toán dung lượng mẫu. Dung lượng mẫu phụ thuộc vào đặc điểm và biến động của đối tượng nghiên cứu và độ tin cậy cho trước (Freese, 1976; Nguyễn Hải Tuất et al., 2006; Subedi et al., 2010; Huy et al., 2013).

Để xác định dung lượng mẫu, người ta cần rút mẫu thử trước, các mẫu này cần được bố trí ngẫu nhiên. Chỉ tiêu điều tra đánh giá để xác định dung lượng mẫu tùy thuộc vào mục đích nghiên cứu, ví dụ, để giám sát sinh khối rừng thì chỉ tiêu điều tra phải là sinh khối, hoặc để đánh giá về kinh tế hộ trong hoạt động lâm nghiệp thì chỉ tiêu điều tra trên hộ là thu nhập từ lâm nghiệp.

Có hai trường hợp chính để tính toán dung lượng mẫu (Nguyễn Hải Tuất et al., 2006; Huy et al., 2013):

- Đối tượng điều tra là một thể thống nhất, không có phân cấp, khối, loại…

- Đối tượng điều tra, đánh giá cần được phân chia theo cấp, khối, loại…

### 2.1.1 Xác định dung lượng mẫu khi không có phân cấp, khối, loại

Đây là trường hợp xác định dung lượng mẫu cho một đối tượng điều tra đánh giá không phân chia thành cấp, khối, loại nào cả. Như là xác định số lượng ô mẫu trong điều tra trữ lượng lượng rừng của một trạng thái rừng. Hoặc số cây trong ước tính sinh khối cây rừng của một kiểu rừng. Hoặc số cây để đánh giá sinh trưởng của một lô rừng theo một công thức thí nghiệm.

Các công thức xác định dung lượng mẫu không phân cấp, loại, khối,… theo một độ tin cậy cho trước như sau (Nguyễn Hải Tuất, 1982):

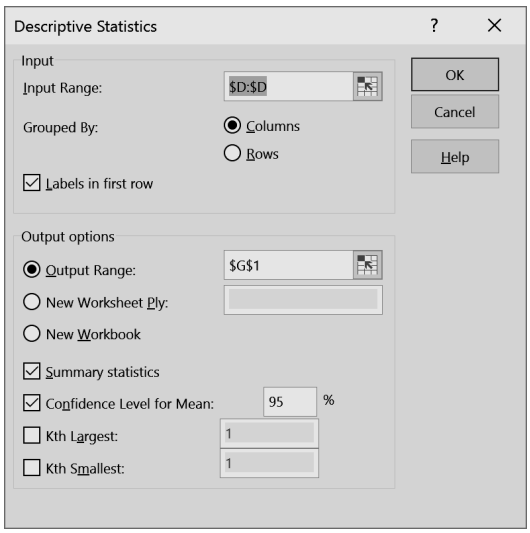
$$n_{ct} = \left(\frac{t * CV\%}{\Delta\%}\right)^2 \quad (2.1)$$

$$CV\% = \frac{S}{\overline{X}}100 \quad (2.2)$$

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \quad (2.3)$$

Trong đó: nct là dung lượng mẫu cần thiết để bảo đảm độ tin cậy của rút mẫu; CV% là hệ số biến động (Coefficient of Variation); S (Standard Deviation) là sai tiêu chuẩn của mẫu; $\overline{X}$ là trung bình của giá trị điều tra; $X_i$ là các giá trị điều tra; Δ% là sai số tương đối cho trước, thường từ 5 – 10%; t là gía trị của biến số t của hàm phân bố chuẩn được xác định theo mức ý nghĩa $P_{value}$ (α), với độ tin cậy là 95% thì $P_{value}$ = 0.05 và độ tự do df = n – 1, với n là số mẫu rút thử; t có thể dễ dàng xác định nhờ hàm tinv của excel = tinv($P_{value}$, df).

Sử dụng Dữ liệu 1 trong phụ lục để tính toán số ô mẫu cần thiết trong điều tra trữ lượng của một trạng thái rừng. Dữ liệu này thu được từ điều tra rút mẫu thử ngẫu nhiên 27 ô mẫu, mỗi ô tính các giá trị trung bình trong đó có trữ lượng là M ($m^3$/ha). Trong excel, sử dụng chức năng thống kê mô tả (Descriptive Statistics) để tính các chỉ tiêu thống kê để tính được $n_{ct}$.




Data/Data Anlysis và chọn Descriptive Statistics.

Trong hộp thoại thông kê mô tả (Descriptive Statistics), xác định:

- Input Range: Vùng ô dữ liệu.

- Grouped by: Dữ liệu xếp theo cột hay hàng.

- Label in first row: Đánh đấu nếu có tên của biến số ở hàng đầu.

- Output Range: Chỉ ô địa chỉ xuất ra kết quả.

- Summary statistics: Đánh dấu để có các chỉ tiêu thông kê.

- Confidence Level for Mean: Xác định độ tin cậy của ước lượng trung bình, thường là 95%.

**Bảng 2.1.** Kết quả tính toán các chỉ tiêu thống kê của rút mẫu thử

| M ($m^3$/ha) | |
|---|---|
| Mean (Trung bình) | 76.15 |
| Standard Error ($S_x$: Sai số của trung bình) | 4.60 |
| Standard Deviation (S: Sai tiêu chuẩn) | 23.93 |
| Sample Variance ($S^2$: Phương sai) | 572.66 |
| Count (n: số mẫu rút thử) | 27 |

Từ kết quả ở Bảng 2.1 tính dung lượng quan sát cần thiết với sai số cho trước Δ% = 10% như sau:

$$CV\% = \frac{23.93}{76.15}100 = 31.43\% \quad (2.4)$$

$$t = tinv(0.05, 26) = 2.06 \quad (2.5)$$

$$n_{ct} = \left(\frac{2.06 * 31.43\%}{10\%}\right)^2 = 42 \text{ ô} \quad (2.6)$$

Như vậy với minh họa này, cần điều tra tổng cộng 42 ô mẫu để ước lượng M có sai số 10%. Do đã rút mẫu thử 27 ô nên lúc này cần đo đếm bổ sung 42 – 27 = 15 ô mẫu.

### 2.1.2 Xác định dung lượng mẫu theo phân cấp, khối, loại

Khi tiến hành điều tra đánh giá một đối tượng được phân chia thành cấp, khối, như là điều tra tài nguyên rừng theo các trạng thái khác nhau, ước lượng sinh khối, trữ lượng rừng theo kiểu rừng, trạng thái rừng. Còn có các ví dụ khác như điều tra lâm nghiệp xã hội, cần xác định số hộ điều tra theo các nhóm kinh tế hộ. Lúc này dung lượng mẫu được phân chia theo cấp, khối, loại,... tùy theo yêu cầu phân cấp của điều tra, đánh giá.

Để xác định dung lượng mẫu theo cấp, loại,... một đợt điều tra rút mẫu ban đầu cần được tiến hành để ước tính sai tiêu chuẩn, biến động của chỉ tiêu điều tra của từng cấp, loại đã phân chia. Cần điều tra ban đầu khoảng 10 – 15 mẫu (tối ưu là 30) cho mỗi loại, cấp.

Tiến trình sau mô tả việc tiến hành xác định số ô mẫu cần thiết đối với điều tra sinh khối, carbon rừng trên mặt đất theo các trạng thái và kiểu rừng khác nhau (Freese, 1976; Subedi et al., 2010; Huy et al., 2013):

*Bước 1.* Xác định sai số % và độ tin cậy ước lượng của số trung bình. Thường sai số là 10% ứng với độ tin cậy ước lượng số trung bình là 95% là yêu cầu cần đạt được.

*Bước 2.* Lựa chọn địa điểm để lập 10 – 15 ô mẫu ban đầu (tốt hơn là 30 ô) cho mỗi trạng thái. Các ô mẫu này hoặc được bố trí ngẫu nhiên hoàn toàn hoặc được lựa chọn ngẫu nhiên từ mạng lưới ô mẫu bố trí trước theo lưới ô hình học. Các ô mẫu bố trí ngẫu nhiên trong mỗi trạng thái có thể sử dụng phần mềm của Hawths' Tool của ArcGIS (http://www.spatialecology.com/htools/tooldesc.php).

*Bước 3*. Ước tính sinh khối, carbon từng cây, từng ô, quy ra trên ha cho từng ô mẫu và trung bình carbon trên ha cho mỗi trạng thái rút mẫu thử.

*Bước 4.* Tính toán sai tiêu chuẩn về sinh khối, carbon (tấn/ha) cho tất cả các ô mẫu của từng trạng thái.

*Bước 5.* Tính toán số lượng ô mẫu cần thiết cho mỗi trạng thái rừng dựa vào số lượng ô mẫu tối đa của cả khu rừng và cho mỗi trạng thái theo các công thức sau:

$$N = \frac{A}{AP} \tag{2.7}$$

$$N_i = \frac{A_i}{AP} \tag{2.8}$$

$$n_{ct} = \frac{\left(\sum_{i=1}^{L} N_i \cdot S_i\right)^2}{\frac{N^2 \cdot E^2}{t^2} + \sum_{i=1}^{L} N_i \cdot S_i^2} \tag{2.9}$$

$$n_{ict} = n_{ct} \cdot \frac{N_i \cdot S_i}{\sum_{i=1}^{L} N_i \cdot S_i} \tag{2.10}$$

$$\overline{X} = \frac{1}{N} \sum_{i=1}^{L} N_i * \overline{X_i} \tag{2.11}$$

Trong đó: N = Số lượng ô mẫu tối đa trong vùng điều tra; $N_i$ = Số lượng ô mẫu tối đa của trạng thái i; $n_{ct}$ = Tổng số ô mẫu cần thiết trong vùng điều tra; $n_{ict}$ = Số ô mẫu cần thiết cho trạng thái I; A = Tổng diện tích của tất cả các trạng thái (ha); $A_i$ = Diện tích của mỗi trạng thái i (ha); AP = Diện tích ô mẫu (ha); i = Chỉ số của trạng thái từ 1 đến L; L = Tổng số trạng thái; $S_i$ = Sai tiêu chuẩn của trạng thái i; E = Sai số tuyệt đối với sai số tương đối cho trước. Với sai số tương đối là 10% thì E = 10% × $\overline{X}$, với $\overline{X}$ là bình quân chung sinh khối hoặc carbon/ha, $\overline{X_i}$ là trung bình sinh khối, carbon của trạng thái i; t = 2 ứng với mức ý nghĩa $P_{value}$ = 0.05. Hoặc chính xác hơn có thể xác định t qua hàm tinv trong excel: =tinv(α, df), với α = 5% và df = $n_m$ – 1, $n_m$ là tổng số mẫu rút thử.

Thông thường, kết quả tính số ô mẫu không phải là một số nguyên. Trong trường hợp đó, số lượng yêu cầu của mẫu phải được điều chỉnh đến số nguyên gần nhất mà không phải là nhỏ hơn so với giá trị ước tính. Ví dụ, nếu ước tính số ô mẫu là 62.04 ô, sau đó là số điều chỉnh phải là 63 ô.

Đôi khi ô mẫu không thể tiếp cận được, bị mất, hoặc không thể đo lặp lại vì nhiều lý do. Ví dụ, ô mẫu có thể bị cuốn trôi bởi lũ lụt hoặc bị cháy hoặc nằm trên vách đá, trên địa hình quá dốc, trên sông suối. Vì vậy tăng số lượng ô mẫu lên vài % có thể giúp cho việc bảo đảm đủ ô số lượng ô theo yêu cầu khi mà một số ô mẫu xác định ban đầu có thể không thể điều tra được.

*Bước 6.* Điều tra hiện trường để đo tính sinh khối trong các ô mẫu đã xác định ở bước 5.

*Bước 7.* Tính toán lại sai số % (Δ%) cho mỗi cấp, khối, trạng thái sau khi đã rút mẫu, đo đếm tất cả các ô mẫu. Mức sai số đạt nếu sai số sau cùng thấp hơn sai số yêu cầu từ đầu, ví dụ thấp hơn 10%. Nếu sai số có sự sai khác và lớn hơn yêu cầu thì hoặc phân chia nhỏ hoặc gộp các trạng thái hoặc là bổ sung thêm số ô mẫu để đạt được sai số cho phép. Lặp lại bước 5 – 7 cho đến khi sai số của mỗi trạng thái đạt được theo yêu cầu, ví dụ là <10%.

Sai số % (Δ%) so với trung bình (hay còn gọi là độ chính xác của rút mẫu - Precision level) về sinh khối, carbon cho mỗi trạng thái i được tính theo công thức sau:

$$\Delta_i\,\% = \frac{S_{xi} * t_{i(0.05,\ ni-1)}}{\overline{X}_i}\,\% \quad (2.12)$$

$$S_{xi} = \frac{S_i}{\sqrt{n_i}} \quad (2.13)$$

Trong đó: Các giá trị này được tính lại sau khi đã rút mẫu đầy đủ theo dung lượng mẫu:

$S_{xi}$: Sai số của số trung bình của trạng thái i

$\overline{X}_i$ là trung bình sinh khối, carbon của trạng thái i

$S_i$ = Sai tiêu chuẩn của trạng thái i

$n_i$: Số ô mẫu đã điều tra theo dung lượng mẫu của trạng thái i

Nếu đến đây tất cả các trạng thái i đều có sai số % ($\Delta\%_i$) bé hơn sai số cho trước thì việc rút mẫu đã hoàn thành.

Ví dụ tính toán số ô mẫu cần thiết để ước tính carbon tích lũy trong cây rừng phần trên mặt đất (tấn/ha) theo các trạng thái rừng ở vùng dự án SNV-REDD$^+$ ở huyện Bảo Lâm, tỉnh Lâm Đồng. Sử dụng Dữ liệu 2 trong Phụ lục được thu thập trong các năm 2011 – 2012.

Từ dữ liệu này có được tổng diện tích khu vực khảo sát là A = 33,068 ha, chia ra theo ba trạng thái rừng: trung bình $A_1$ = 5,783 ha, phục hồi $A_2$ = 19,048 ha và gỗ lồ ô $A_3$ = 8,237 ha. Diện tích mỗi ô mẫu là AP = 0.1 ha. Đã tiến hành rút mẫu n = 97 ô, trong đó $n_1$ = 40 ô, $n_2$ = 17 ô và $n_3$ = 40 ô.

Từ đó tính được:

$$N = \frac{33{,}068}{0.1} = 330{,}680$$

$$N_1 = \frac{5{,}783}{0.1} = 57{,}830;\ N_2 = \frac{19{,}048}{0.1} = 190{,}482;\ N_3 = \frac{8{,}237}{0.1} = 82{,}368$$

Bảng 2.2 trình bày kết quả sử dụng chức năng Descriptive Statistics trong excel để tính các chỉ tiêu thống kê cho mỗi trạng thái rừng dựa vào dữ liệu carbon ô mẫu ở ba trạng thái.

**Bảng 2.2.** Các chỉ tiêu thống kê của carbon trên mặt đất theo các trạng thái rừng

| $A_1$ Trung bình | | $A_2$ Phục hồi | | $A_3$ Gỗ - lồ ô | |
|---|---|---|---|---|---|
| Mean | 135.7 | Mean | 46.8 | Mean | 72.0 |
| Standard Error | 7.7 | Standard Error | 10.1 | Standard Error | 8.6 |
| Standard Deviation | 48.8 | Standard Deviation | 41.5 | Standard Deviation | 54.2 |
| Count | 40 | Count | 17 | Count | 40 |
| Confidence Level(95.0%) | 15.6 | Confidence Level(95.0%) | 21.3 | Confidence Level(95.0%) | 17.3 |

Kết quả từ Bảng 2.2 có được:

$\overline{X}_1$= 135.7 tấn/ha; $\overline{X}_2$ = 46.8 tấn/ha; $\overline{X}_3$ = 72.0 tấn/ha

$S_1$ = 48.8; $S_2$ = 41.5; $S_3$ = 54.2.

Từ đây tính được trung bình chung: $\overline{X}$ = 68.6 tấn/ha, E = 10%×68.6 = 6.68 tấn/ha, t = tinv(0.05, 96) = 1.98.

Với tất cả dữ liệu đầu vào nói trên, tính được tổng số ô mẫu và số ô mẫu theo từng trạng thái:

$n_{ct}$ = 177 ô ; $n_{1ct}$ = 33 ô; $n_{2ct}$ = 92 ô và $n_{3ct}$ = 52 ô.

Như vậy tổng số ô mẫu còn thiếu là: 177 – 97 = 80 ô, trong đó trạng thái 1 (rừng trung bình) dư: 40 – 33 = 7 ô, trạng thái 2 (phục hồi) thiếu: 17 – 92 = - 75 ô, trạng thái 3 (rừng hỗn giao gỗ - lồ ô) còn thiếu: 40 – 52 = -12 ô.

Từ kết quả này, kiểm tra độ tin cậy (sai số của rút mẫu) như sau:

Từ kết quả ở Bảng 2.2 có được sai số của trung bình theo trạng thái (Standard Error):

$S_{x1}$ = 7.7; $S_{x2}$ = 10.1; $S_{x3}$ = 8.6

Các giá trị $t_i$: $t_1$ = 2.02; $t_2$ = 2.12 và $t_3$ = 2.02 (theo hàm $t_i$ =tinv(0.05, $n_i$ -1) trong excel.

Từ đây tính được sai số của từng trạng thái:

$\Delta_1$% = 12%; $\Delta_2$% = 46% và $\Delta_3$% = 24%.

Như vậy, nếu dừng lại ở số mẫu đã điều tra, với yêu cầu sai số < 10% thì chỉ có trạng thái 1 (rừng trung bình) là có số ô mẫu gần đủ, với sai số gần với 10%, hai trạng thái còn lại đều có sai số > 10%, do đó cần rút mẫu bổ sung.

Kết quả đánh giá cho thấy sai số của trạng thái 1 là 12%, có nghĩa là vẫn còn thiếu ô mẫu để đạt được sai số 10% như yêu cầu. Trong khi đó trên đây lại tính ra được số mẫu cần thiết phân chia cho trạng thái này cũng với sai số cho phép là 10% là 33 ô, và cũng đã điều tra 40 ô, có nghĩa là đã dư ô mẫu để đạt yêu cầu sai số. Việc kiểm tra lại sai số sau cùng cho từng trạng thái chỉ có tính chất tham khảo về sai số đạt được cho từng khối, do sai số lúc này được tính riêng lẻ cho từng cấp, trạng thái, không phân khối, như vậy đây là sai số của rút mẫu cho từng khối riêng lẻ. Trong khi đó việc tính toán số ô mẫu ở đây được phân phối theo cấp, trạng thái trên cơ sở phân bố tối ưu số ô mẫu cho từng khối, do đó số ô mẫu từng cấp, khối có thể nhỏ hơn số ô mẫu nếu tính riêng lẻ. Vì vậy, đối với rút mẫu có phân cấp, khối, sử dụng kết quả xác định số ô mẫu cần thiết theo từng cấp, khối là đạt được sai số cho trước.

Ngoài ra có thể xác định nhanh và ước tính gần đúng dung lượng mẫu có phân cấp, khối ban đầu theo (Lackmann, 2011) mà không cần rút mẫu thử trước như sau:

$$n_{ct} = \frac{t^2 * CV\%^2}{E\%^2 + \frac{t^2 * CV\%^2}{N}} \quad (2.14)$$

$$n_{ict} = n_{ct}\frac{N_i}{N} \quad (2.15)$$

Trong đó: i: Số thứ tự trạng thái, kiểu, khối rừng được phân loại

N = Số lượng ô mẫu tối đa trong vùng điều tra

$N_i$ = Số lượng ô mẫu tối đa của trạng thái i

$n_{ct}$ = Tổng số ô mẫu cần thiết trong vùng điều tra

$n_{ict}$ = Số ô mẫu cần thiết cho trạng thái i

E% = Sai số tương đối cho trước, thường là 10%

t = 2 (giá trị ban đầu với độ tin cậy ước lượng là 95%)

CV%: (Coefficient of Variation %), hệ số biến động. Chọn trước một CV% cao nhất đã biết trước của đối tượng điều tra.

Sử dụng dữ liệu trình bày trên để ước tính số ô mẫu cần thiết theo trạng thái ban đầu trên cơ sở công thức của (Lackmann, 2011), trong đó CV% được xác định trước là 67% (Hệ số biến động cao nhất trong khu vực khi ước tính carbon rừng). Kết quả:

$$n_{ct} = \frac{2^2 * 67\%^2}{10\%^2 + \frac{2^2 * 67\%^2}{330\,680}} = 179 \text{ ô}$$

$$n_{1ct} = 179\frac{57{,}830}{330\,680} = 31 \text{ ô}$$

$$n_{2ct} = 179\frac{190{,}482}{330\,680} = 103 \text{ ô}$$

$$n_{3ct} = 179\frac{82{,}368}{330\,680} = 45 \text{ ô}$$

Cách tính này giảm được chi phí và thời gian vì không rút mẫu thử trước, nó phụ thuộc vào việc chọn giá trị CV% dự báo trước. Vì vậy, dung lượng mẫu được ước tính trước rất phụ thuộc vào CV%. Áp dụng phương pháp này cần có đánh giá lại sai số cho mỗi trạng thái sau khi đã rút mẫu xong, theo bước 7 của phương pháp rút mẫu có phân khối đã giới thiệu trên đây.

Ngoài ra nếu không có giá trị CV% cho trước, có thể tiến hành thu thập mẫu thử với dung lượng khoảng 30 - 50 ô, phân bố ngẫu nhiên trên tất cả các trạng thái, cấp, khối được phân loại. Từ đây tính được hệ số biến động chung CV% của toàn khu rừng điều tra.

## 2.2 Phương pháp bố trí, rút mẫu trong điều tra, đánh giá để xử lý thống kê

Trên cơ sở xác định số mẫu cần thu thập, cần tiến hành bố trí, rút mẫu theo một phương pháp thống nhất. Tùy theo đối tượng nghiên cứu, nguồn lực và độ tin cậy cho phép mà lựa chọn một trong ba phương pháp rút mẫu chính sau đây (Nguyễn Hải Tuất, 1982; Lackmann, 2011; Jayaraman, 1999; Huy et al., 2013).

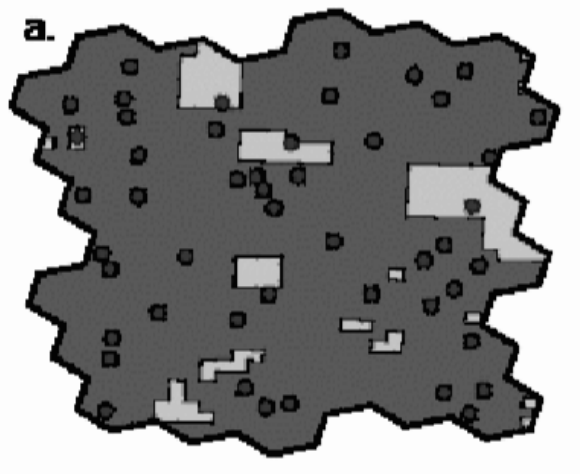
### 2.2.1 Rút mẫu ngẫu nhiên (Random sampling)

Đây là phương pháp rút mẫu các cá thể, ô mẫu, người,… để điều tra, nghiên cứu, phỏng vấn được lựa chọn ngẫu nhiên. Việc lựa chọn ngẫu nhiên có thể được tiến hành bằng nhiều cách tùy vào đối tượng nghiên cứu.

- Nếu đối tượng là các cây rừng trong một lô rừng, thì đánh số các cây và rút thăm ngẫu nhiên.

- Nếu đối tượng là người cần phỏng vấn, thì bốc thăm ngẫu nhiên trong danh sách đối tượng nghiên cứu.

- Nếu đối tượng là vị trí các ô mẫu trong một vùng, một khu rừng rộng lớn, thì bố trí ngẫu nhiên ô mẫu trong các phần mềm ArcGIS và chuyển tọa độ vào GPS cần được áp dụng.




**Hình 2.1**. Bố trí ô mẫu trong rút mẫu ngẫu nhiên. *Nguồn: (Lackmann, 2011)*

Hình 2.1 cho thấy các ô mẫu được bố trí ngẫu nhiên trên khu vực rừng điều tra. Rút mẫu ngẫu nhiên có ưu điểm là hoàn toàn khách quan và dùng để áp dụng thống kê để ước lượng, so sánh, đánh giá. Nhược điểm của nó là đối với điều tra tài nguyên rừng, đôi khi vị trí được xác định ngẫu nhiên ở nơi quá xa, ở vị trí địa hình phức tạp, nguy hiểm khó tiếp cận. Trường hợp này trong thống kê cũng cho phép dịch chuyển vị trí cá thể, ô mẫu đến nơi lân cận nhưng khá tương đồng. Tuy vậy, việc dịch chuyển đó sẽ làm mất tính ngẫu nhiên, do đó cần hạn chế. Đồng thời, như đã trình bày trong xác định dung lượng mẫu, thường nên tính toán số mẫu lớn hơn yêu cầu vài % để có thể bỏ bớt một số ô mẫu ở vị trí không thể tiếp cận được trong thực tế.

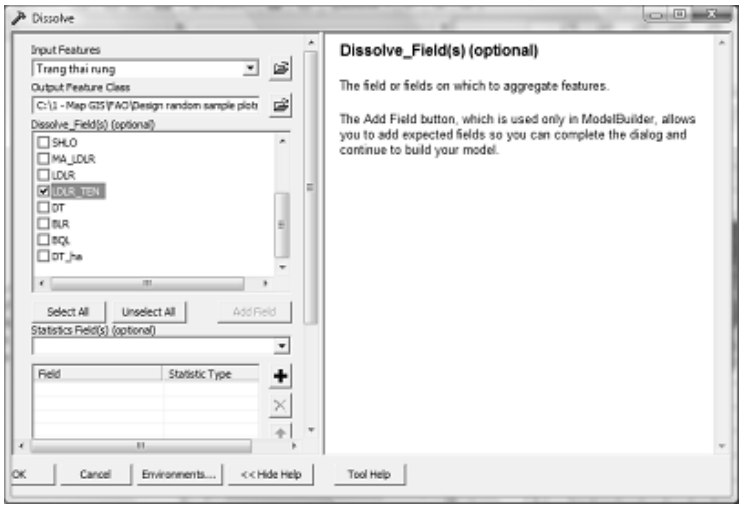
Dưới đây là giới thiệu cách bố trí các ô mẫu ngẫu nhiên trên bản đồ dựa vào phần mềm ArcGIS và chuyển tọa độ ô mẫu vào GPS để xác định vị trí điều tra trên hiện trường. Thiết kế ô

mẫu ngẫu nhiên cho từng trạng thái rừng có thể được tiến hành trong chức năng tạo lập điểm ngẫu nhiên "create random point" trong ArcGIS (Huy et al., 2013).

Số lượng ô mẫu phụ thuộc vào diện tích của trạng thái rừng, kích thước của ô mẫu và biến động của chỉ tiêu điều tra giám sát rừng. Việc xác định số ô mẫu cần thiết (đã được trình bày trong phần trên). Ngoài ra kích thước và hình dạng ô mẫu phải được thống nhất trong cả khu vực điều tra. Một bản đồ số được sử dụng để tạo các vị trí ô mẫu ngẫu nhiên theo phương pháp của Hawths được chạy trong phần mềm ArcGIS với công cụ "create random point". Các bước sau đây cần được tiến hành:

*Bước 1:* Gộp các mảnh/lô trạng thái trên bản đồ thành khối trạng thái để bố trí ô mẫu ngẫu nhiên:

Tạo thành các khối trạng thái đồng nhất, nghĩa là gộp các poligon có cùng trạng thái với nhau để thiết kế vị trí các ô mẫu cho từng khối trạng thái. Sử dụng chức năng Dissolve trong ArcGIS.




Chức năng Dissolve trong ArcGIS để gộp các lô theo khối trạng thái rừng:

Input Features: Chọn lớp dữ liệu chứa trạng thái rừng

Output Feature Class: Chọn địa chỉ và đặt tên file.

Dissolve Field: Chọn trường trạng thái rừng.

Tuy nhiên, trong trường hợp thiết kế ô mẫu cho từng lô trạng thái thì giữ nguyên lớp bản đồ phân chia lô trạng thái.

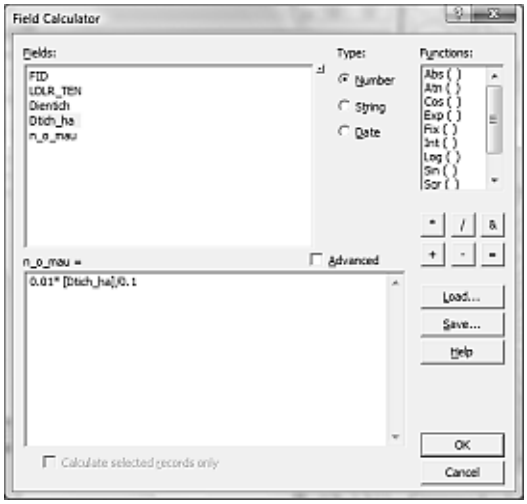
*Bước 2:* Cung cấp, xác định số ô mẫu ngẫu nhiên theo từng trạng thái rừng trên bản đồ:

Từ kết quả bản đồ đã Dissolve thành các khối trạng thái, lập một trường (field) định dạng số là số ô mẫu của từng khối trạng thái. Sau đó nhập vào trường này số ô mẫu cần thiết cho từng trạng thái của từng kiểu rừng, vùng sinh thái.

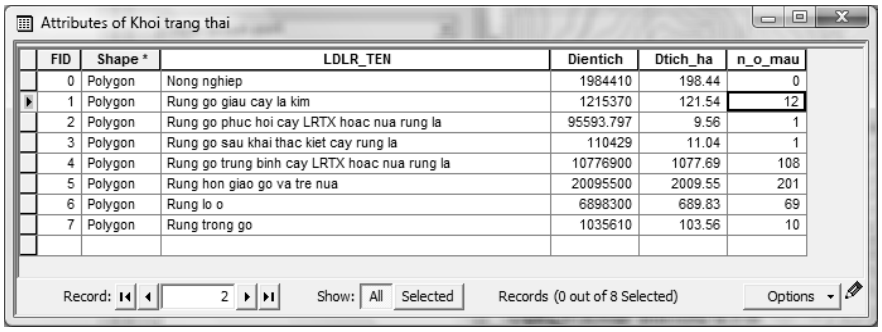
Có trường hợp số ô mẫu được xác định theo tỷ lệ % rút mẫu theo diện tích. Ví dụ tỷ lệ rút mẫu là 1% của diện tích trạng thái và diện tích ô mẫu là 0.1ha. Từ đây tính toán số ô mẫu cho từng khối trạng thái i là $n_i$ theo công thức sau:

$$ni = \frac{1\% * \text{Dien tich trang thai i (ha)}}{\text{Dien tich o mau (0.1 ha)}} \quad (2.16)$$

Lúc này tính toán số ô mẫu cho từng khối trạng thái trên bản đồ trong ArcGIS nhờ chức năng Field Calculator.



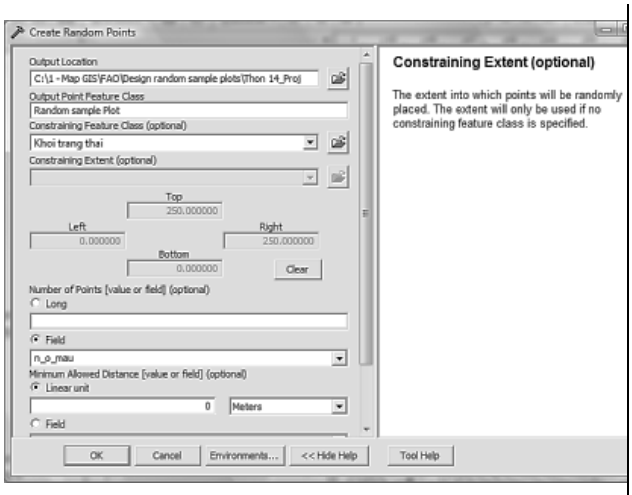

Số lượng ô mẫu tính theo công thức cho mỗi khối trạng thái trong chức năng Field Calculator trong ArcGIS




Bảng dữ liệu số lượng ô mẫu theo từng khối trạng thái trên bản đồ trong ArcGIS

*Bước 3:* Thiết kế hệ thống ô mẫu ngẫu nhiên trên bản đồ cho từng khối trạng thái trong ArcGIS:

Trong ArcGIS, sử dụng Creat Random Points để thiết kế hệ thống ô mẫu được phân bố ngẫu nhiên theo khối trạng thái trên bản đồ trạng thái rừng.




Thiết kế vị trí ô mẫu ngẫu nhiên dựa vào Creat Random Points:

Output location: Chỉ vị trí thư mục để xuất ra bản đồ ô mẫu.

Output Point Feature Class: Tên file bản đồ xuất ra.

Constrainning Feature Class: Chọn trường khối trạng thái

Field: Chọn trường dữ liệu chứa số liệu số ô mẫu cho từng khối trạng thái

Minimum Allowed Distance: Cự ly khoảng cách tối thiểu giữa 2 ô: có thể chọn khoảng cách tối thiểu như là một điều kiện (ví dụ với việc bố trí ô mẫu tròn có bán kính 17.84m thì khoảng cách tối

Nếu chọn giá trị của "Long" là cố định bao nhiêu thì có nghĩa mỗi khối trạng thái có số ô bằng nhau, điều này hiếm xảy ra vì số ô mẫu của mỗi khối trạng thái phụ thuộc và

thiểu giữa hai ô nên là > 18m để tránh trường hợp các ô chồng lấn lên nhau trên thực địa).

diện tích, sai tiêu chuẩn và biến động của nó.

Sau đó gán tọa độ cho các vị trí ô mẫu ngẫu nhiên trên bản đồ: Sử dụng chức năng của ArcGIS trên bản đồ các ô mẫu ngẫu nhiên đã lập: Add XY Coordinates. Sau đó tạo một trường đặt số hiệu cho ô mẫu: So_hieu_o = FID +1.

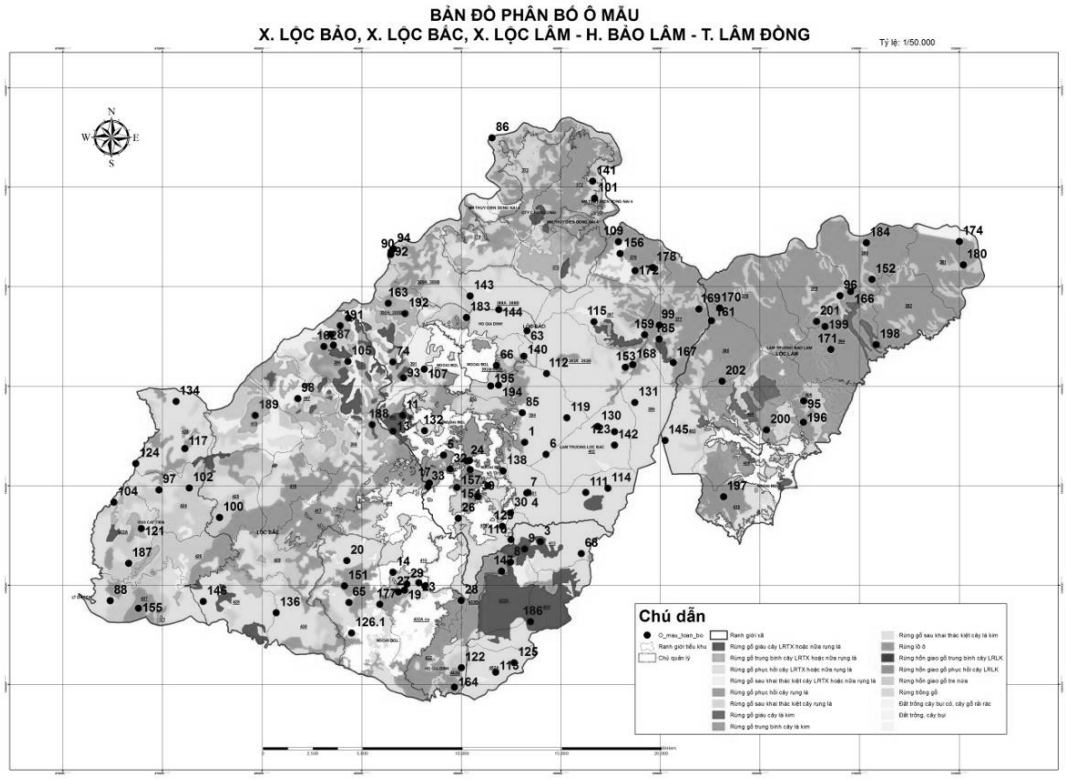
Attributes of O mau ngau nhien

| FID | Shape | CID | So_hieu_o | POINT_X | POINT_Y |
|---|---|---|---|---|---|
| 0 | Point | 1 | 1 | 524449.285074 | 1261684.32025 |
| 1 | Point | 1 | 2 | 525349.470495 | 1262866.43532 |
| 2 | Point | 1 | 3 | 524070.528416 | 1262757.47458 |
| 3 | Point | 1 | 4 | 524769.531447 | 1262624.19545 |
| 4 | Point | 1 | 5 | 524740.012971 | 1262802.18004 |
| 5 | Point | 1 | 6 | 524494.609521 | 1263223.52577 |
| 6 | Point | 1 | 7 | 525227.354617 | 1263174.75494 |
| 7 | Point | 1 | 8 | 524904.987619 | 1262817.95875 |
| 8 | Point | 1 | 9 | 524311.838026 | 1261822.54978 |
| 9 | Point | 1 | 10 | 524298.116124 | 1260735.51158 |
| 10 | Point | 1 | 11 | 524294.13666 | 1262184.81284 |
| 11 | Point | 1 | 12 | 524246.284969 | 1261993.1484 |
| 12 | Point | 2 | 13 | 521351.94811 | 1262390.3633 |
| 13 | Point | 3 | 14 | 524964.164553 | 1263121.98471 |
| 14 | Point | 4 | 15 | 523987.500389 | 1259208.83952 |
| 15 | Point | 4 | 16 | 523731.926787 | 1256459.66146 |
| 16 | Point | 4 | 17 | 524654.495667 | 1259795.67211 |
| 17 | Point | 4 | 18 | 525914.674794 | 1260698.2084 |
| 18 | Point | 4 | 19 | 525878.65148 | 1259910.30381 |
| 19 | Point | 4 | 20 | 523903.582455 | 1258682.29778 |
| 20 | Point | 4 | 21 | 526274.316065 | 1260783.22584 |
| 21 | Point | 4 | 22 | 525558.907285 | 1260212.03964 |
| 22 | Point | 4 | 23 | 524714.685258 | 1259701.7598 |
| 23 | Point | 4 | 24 | 525257.380468 | 1261738.67359 |
| 24 | Point | 4 | 25 | 526518.238395 | 1262489.20488 |
| 25 | Point | 4 | 26 | 524093.598079 | 1257622.08512 |
| 26 | Point | 4 | 27 | 524454.956944 | 1258073.2208 |
| 27 | Point | 4 | 28 | 523372.026925 | 1256838.52982 |

Record: 0 Show: All Selected Records (0 out of 402 Selected) Options

Kết quả thực hiện chức năng Add XY Coordinates: Gắn tọa độ ô mẫu trong ArcGIS.

Kết quả có được bản đồ thiết kế hệ thống ô mẫu ngẫu nhiên trên bản đồ như ở Hình 2.2.




**Hình 2.2.** Bản đồ thiết kế hệ thống ô mẫu ngẫu nhiên theo trạng thái rừng khu vực ba xã Lộc Bảo, Lộc Bắc và Lộc Lâm thuộc huyện Bảo Lâm, tỉnh Lâm Đồng. *Nguồn: Huy et al., (2013), dự án SNV-REDD+.*

*Bước 4:* Chuyển tọa độ điểm các ô mẫu ngẫu nhiên vào GPS để xác định trên thực địa:

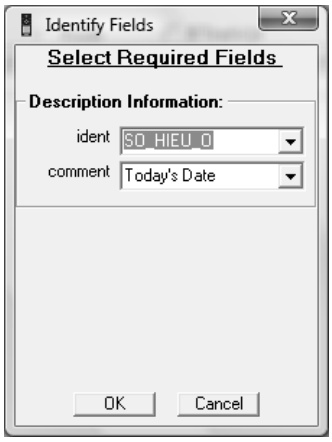
Tọa độ các ô mẫu cần được chuyển vào trong máy GPS để xác định chính xác vị trí ô mẫu trên thực địa. Công việc này có thể thực hiện thông qua chương trình DNR trên cơ sở kết nối máy tính với GPS.

Trong trường hợp sử dụng hệ tọa độ VN2000, vì có sự chênh lệch, do đó cần điều chỉnh tọa độ ô mẫu trước khi chuyển vào GPS. Cách làm như sau:

- Trong ArcGIS, xuất (export) file tọa độ ô mẫu sang dạng dbf

- Mở file dbf trong excel và tạo hai cột tọa độ X/Y mới, với X = x + 194 và Y = y – 112, trong đó x/y là tọa độ của ô mẫu, X/Y là tọa độ điều chỉnh để chuyển vào GPS với hệ chiếu VN2000

- Chuyển file này thành shape file trong ArcGIS




Mở DNR và cài đặt hệ tọa độ: File/Set Projection. Vào Load PRJ và chọn file tọa độ ô mẫu có đuôi *.prj có chứa thông tin hệ tọa độ (ví dụ Vn2000).

Mở file dữ liệu tọa độ trong DNR: File/Load from/File… Chọn file dạng shape và mở file tọa độ ô đã lưu. Chọn ident là trường So_hieu_o.

Đưa dữ liệu từ file tọa độ ô mẫu vào GPS: Waypoint/Upload.

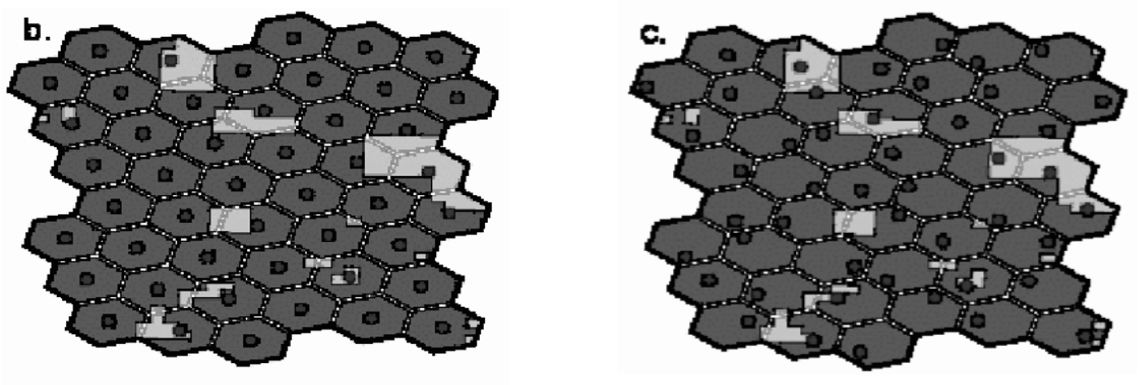
Kết quả hệ thống ô mẫu ngẫu nhiên trên bản đồ và tọa độ của nó đã được chuyển sang GPS, từ đây sử dụng chức năng dẫn đường (Go to) của GPS để đi đến đúng tọa độ của từng ô mẫu trên thực địa.

### 2.2.2 Rút mẫu hệ thống (Systematic sampling)

Rút mẫu hệ thống là thực hiện lấy mẫu tuân theo một quy luật đơn giản. Ví dụ có N mẫu đi từ 1 đến N, rút mẫu hệ thống mẫu ở vị trí thứ k và tuần tự như vậy (ví dụ k = 3, thì cứ 3 mẫu chọn một mẫu) (Jayaraman, 1999).

Đây là phương pháp rút mẫu khá truyền thống trong lĩnh vực điều tra rừng Việt Nam. Trên cơ sở số lượng ô mẫu đã được xác định, lập các tuyến song song và vuông góc với hệ thống đường đồng mức chính để có thể đi qua được các kiểu rừng, trạng thái khác nhau và cách đều trên bản đồ. Sau đó bố trí các ô mẫu cũng cách đều nhau trên từng tuyến. Nguyên tắc là cự ly giữa các tuyến và các ô trên tuyến càng xấp xỉ nhau càng tốt, vì như vậy hệ thống ô mẫu sẽ được bố trí khá đều trên thực địa.

Phương pháp rút mẫu hệ thống còn được chia ra là: rút mẫu hệ thống thẳng hàng (Systemmatic sampling aligned) và không thẳng hàng (Systematic sampling unaligned). Mẫu thẳng hàng thì các ô mẫu có thể ở trung tâm hoặc góc của một lưới. Mẫu không thẳng hàng thì mỗi ô mẫu được bố trí ngẫu nhiên trong một ô của lưới. Hai cách rút mẫu này được biểu diễn ở Hình 2.3.




**Hình 2.3.** Rút mẫu hệ thống thẳng hàng (trái) và không thẳng hàng (phải). *Nguồn: (Lackmann, 2011)*

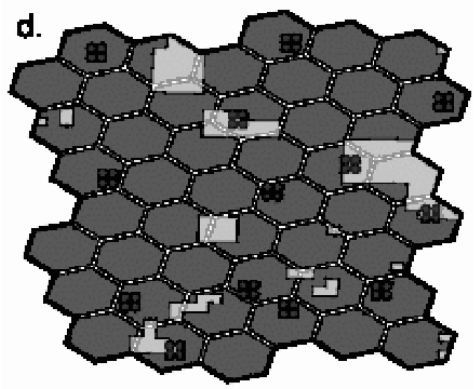
Bố trí ô mẫu hệ thống ngày nay cũng nên thiết kế trên các phần mềm GIS cho thuận tiện và chính xác, đồng thời có thể trích tọa độ ô mẫu để sử dụng GPS xác định vị trí ô trên thực địa.

Phương pháp rút mẫu hệ thống có ưu điểm là các ô được bố trí khá đồng đều trên thực địa, dễ quản lý thứ tự của hệ thống ô mẫu theo tuyến, ô. Nhược điểm lớn nhất của nó là việc chuyển theo tuyến đến từng ô mẫu, có nơi hoàn toàn không có khả năng thực hiện vì địa hình phức tạp, tuyến cắt qua sông suối lớn, đỉnh núi cao, vách đứng. Nhược điểm này cũng có thể khắc phục bằng cách sử dụng tọa độ ô để đi tránh nơi có địa hình phức tạp.

Ngoài ra trong thực tế người ta còn có thể phối hợp hai phương pháp rút mẫu hệ thống và ngẫu nhiên với nhau, gọi là rút mẫu hệ thống – ngẫu nhiên. Phương pháp này thì các tuyến được bố trí hệ thống (song song, cách đều, vuông góc với đường đồng mức chính), sau đó các ô được bốc thăm để bố trí ngẫu nhiên trên từng tuyến. Lý do là để tăng tính khách quan, ngẫu nhiên của phương pháp rút mẫu hệ thống. Vì rút mẫu hệ thống có tính chủ quan khi chọn vẽ hướng của các tuyến song song, do đó nếu các ô trên tuyến được bốc thăm ngẫu nhiên vị trí của nó là cơ hội để tăng tính khách quan và ngẫu nhiên của việc rút mẫu. Có nghĩa là tăng độ tin cậy, tránh định kiến (Bias).

### 2.2.3 Rút mẫu theo cụm (Cluster sampling)

Dựa trên phương pháp rút mẫu hệ thống, nhưng các ô mẫu được bố trí theo cụm như ở Hình 2.4. Tại vị trí ô mẫu xếp theo hệ thống, bố trí một cụm ô mẫu gần nhau. Việc bố trí ô mẫu theo cụm rất hiệu quả vì giảm thời gian di chuyển giữa các ô so với rút mẫu ngẫu nhiên và hệ thống. Tuy vậy vì trong mỗi cụm, các ô gần nhau nên có thể giảm phản ảnh sự biến động và thông tin sai khác trong trạng thái. Vì vậy, khoảng cách giữa các ô mẫu trong mỗi cụm cũng nên đủ lớn để tránh mối tương quan giữa các ô trong cụm. Thường nên ít nhất là 250 – 300m, phụ thuộc vào kích cỡ của trạng thái rừng điều tra (Lackmann, 2011).




**Hình 2.4.** Rút mẫu theo phương pháp phân bố cụm theo hệ thống. *Nguồn (Lackmann, 2011)*

### 2.2.4 Rút mẫu điển hình

Đây là phương pháp rút mẫu dựa vào việc chọn lựa đối tượng điều tra nghiên cứu đại diện, điển hình nhất. Phương pháp này được áp dụng trong các trường hợp sau:

- Đối tượng nghiên cứu không quá lớn để tổ chức rút mẫu ngẫu nhiên hay hệ thống nhưng cũng không quá nhỏ để có thể đo đếm, lấy số liệu toàn bộ,

- Đối tượng nghiên cứu là rộng lớn và nguồn lực có hạn để có thể tổ chức rút mẫu hệ thống, hoặc ngẫu nhiên.

- Phục vụ các nghiên cứu chuyên đề như cấu trúc rừng, rút mẫu chặt hạ cây để nghiên cứu thể tích, sinh khối.

Cả hai trường hợp trên đều cần chuyên gia có kinh nghiệm để lựa chọn đối tượng lấy mẫu đại diện. Ví dụ, trong nghiên cứu cấu trúc rừng của một kiểu rừng ở một vùng sinh thái, lúc này kích thước ô mẫu cần phải đủ lớn, thường ít nhất là 1ha (Phuong et at., 2012), hoặc có khi lên đến 100 ha rừng. Vì vậy, rút mẫu hệ thống hoặc ngẫu nhiên sẽ không đủ nguồn lực; trường hợp này dựa vào kinh nghiệm chuyên gia để lựa chọn vị trí mà lâm phần đó đại diện nhất cho cấu trúc rừng. Gọi là rút mẫu điển hình.

Việc lựa chọn đại diện để rút mẫu cũng cần có các tiêu chí định trước và tùy thuộc vào mục đích nghiên cứu. Ví dụ để nghiên cứu cấu trúc rừng mẫu rừng hỗn loại khác tuổi, người ta phải chọn nơi có cấu trúc chưa bị tác động (hoặc ít nhất có thể), có trữ lượng, sinh khối cao nhất, có phân bố các thế hệ hợp lý theo dạng giảm đều, có tổ thành loài mục đích ưu thế, đại diện cho lập địa…

Ưu điểm của phương pháp rút mẫu điển hình là giảm chí phí, thời gian và chọn lựa được mẫu đại diện cho mục đích nghiên cứu. Nhược điểm là phụ thuộc vào chủ quan, kinh nghiệm của nhà nghiên cứu, không xử lý được thống kê để tính toán biến động, sai số.

## 2.3 Nguyên tắc thiết kế thử nghiệm trong lâm nghiệp, quản lý tài nguyên môi trường rừng

Trong nghiên cứu lâm nghiệp thường có các thử nghiệm như, đánh giá các nguồn giống cây rừng khác nhau, ảnh hưởng của phân bón đến sinh trưởng cây con trong vườn ươm, hoặc của cả nguồn giống và phân bón. Trong trồng rừng thì có nghiên cứu ảnh hưởng của mật độ, hoặc tỉa thưa

hoặc cả hai đến tăng trưởng rừng. Trong gây trồng lâm sản ngoài gỗ dưới tán rừng thì thí nghiệm ảnh hưởng của một đến nhiều nhân tố sinh thái như: độ tàn che, độ cao so với mắt biển, độ ẩm không khí, kiểu rừng, loại đất, độ dốc,… đến năng suất. Tất cả những thử nghiệm, thí nghiệm, đánh giá nói trên cần được bố trí thí nghiệm theo các yếu tố thử nghiệm, nhân tố thí nghiệm và các bố trí thí nghiệm này cần theo các nguyên tắc để bảo đảm cho việc áp dụng xử lý, đánh giá thống kê khách quan. Theo (Jayaraman, 1999; Nguyễn Hải Tuất, Vũ Tiến Hinh, Ngô Kim Khôi, 2006) thì thiết kế thí nghiệm lâm nghiệp cần theo các nguyên tắc như là ngẫu nhiên, có lặp lại và khống chế các tác động ngoài nhân tố thí nghiệm.

**Các nguyên tắc trong bố trí thử nghiệm lâm nghiệp:**

*Dung lượng mẫu trong ô hoặc lô thí nghiệm:* Khi rút mẫu trong điều tra, đánh giá chúng ta có thể rút mẫu thử, sau đó tính toán biến động để xác định số mẫu và đánh giá sai số, nếu thiếu mẫu thì có thể rút bổ sung để bảo đảm sai số. Nhưng các bố trí thí nghiệm liên quan đến gây trồng cả trong vườn ươm lẫn rừng trồng, trồng dưới tán rừng, chúng ta chỉ có thể bố trí cố định số cây ngay từ đầu và không thể bổ sung (vì sẽ không đồng nhất). Trong thí nghiệm gây trồng, chúng ta chưa thể đánh giá biến động để tính số mẫu bảo đảm sai số. Do vậy, mẫu cần bao nhiêu trong thí nghiệm là đủ? Về nguyên tắc một lô thí nghiệm trong vườn ươm, rừng trồng, dưới tán rừng cần có đủ số cây để bảo đảm số trung bình của chỉ tiêu quan sát tiệm cận phân bố chuẩn, hay nói khác, không quá ít để số trung bình của ô mẫu không đại diện hoặc quá nhiều sẽ lãng phí. Trong thống kê, người ta có nhận xét rằng, với một mẫu có dung lượng lớn hơn 30 thì chỉ tiêu quan sát có nhiều khả năng tiệm cận chuẩn. Vì vậy, trong các loại thử nghiệm này, số cây nên ít nhất là 30 và lớn nhất khoảng 50 cây trong mỗi ô/lô thí nghiệm.

*Số lượng ô hoặc lô thí nghiệm theo các nhân tố nghiên cứu:* Trong thử nghiệm thông thường chúng ta đánh giá ít nhất một nhân tố ảnh hưởng hoặc nhiều hơn. Ví dụ, trong nghiên cứu sinh trưởng của lan kim truyến dưới tán rừng theo hai nhân tố độ tàn che của rừng và độ cao so với mặt biển. Lúc này mỗi nhân tố nghiên cứu cần xác định có bao nhiêu cấp (công thức thí nghiệm), ví dụ độ tàn che được chia làm 3 cấp: < 30%, 30 – 80% và > 80%, độ cao chia làm 4 cấp: < 500m, 500 – 800m, 800 – 1100m và > 1100m. Lúc này số ô thí nghiệm sẽ là 3 cấp độ tàn che * 4 cấp đai cao so với mặt biển = 12 ô thí nghiệm.

*Bố trí thí nghiệm có lặp lại:* Với ví dụ bố trí thí nghiệm trồng lan kim tuyến dưới tán rừng theo hai nhân tố độ tàn che và đai cao nói trên, thì có 12 ô thí nghiệm, mỗi ô thí nghiệm sẽ ứng với một tổ hợp công thức của hai nhân tố đó. Ví dụ ứng với tổ hợp độ tàn che < 30% và đai cao < 500m lúc này sẽ có một ô thí nghiệm. Thí nghiệm như vậy gọi là nghiên cứu hai nhân tố không lặp lại. Loại bố trí thí nghiệm từ hai nhân tố trở lên không có lặp lại thường có hạn chế về tính khách quan và có rủi ro không đánh giá được. Hạn chế tính khách quan là chỉ với một ô thí nghiệm, đôi khi chúng ta bố trí chúng trên một hoàn cảnh đặc biệt (ví dụ, đất tốt hoặc xấu hơn bình thường và đất không phải là nhân tố nghiên cứu) và kết quả đánh giá sẽ thiên lệch. Hạn chế của rủi ro là chỉ với một ô thí nghiệm, nếu do ảnh hưởng của vật nuôi phá hoại, thời tiết, sẽ không có dữ liệu để đánh giá.

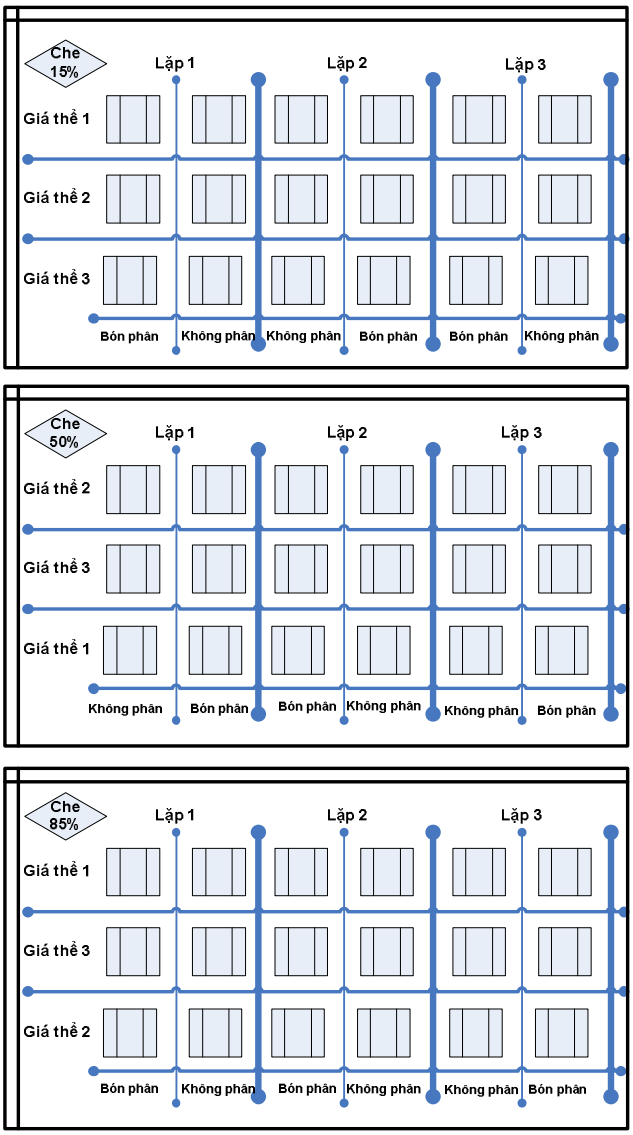
Vì vậy, trong bố trí thí nghiệm thường cố gắng có lần lặp lại, ít nhất là 2 lần và cao nhất cũng không quá 3 lần. Với ví dụ trồng lan kim tuyến dưới tán rừng theo 2 nhân tố độ tàn che và độ cao

nói trên, với số lần lặp lại là 3 lần, thì số ô thí nghiệm được tính là: 3 cấp độ tàn che * 4 cấp đai cao * 3 lần lặp lại = 36 ô thí nghiệm ứng với tổ hợp công thức sẽ có 3 ô thí nghiệm.

Đặc biệt là, nếu bố trí thí nghiệm chỉ một nhân tố, ví dụ ảnh hưởng của mật độ đến tăng trưởng rừng, trong đó mật độ có 3 cấp: < 1000 cây/ha, 1000 – 2000 cây/ha và 2000 – 3000 cây/ha; lúc này số ô thí nghiệm ở mỗi cấp mật độ ít nhất phải là 2 ô, như vậy, với 3 cấp mật độ thì cần có ít nhất 6 ô thí nghiệm. Các cấp mật độ có thể có số ô không bằng nhau, ví dụ có cấp là 3 ô, cấp có 4 ô, cấp có 2 ô, nhưng ít nhất phải là 2 ô (có lặp lại) với mỗi công thức thí nghiệm.

*Đồng nhất các yếu tố không nghiên cứu:* Khi bố trí thí nghiệm theo các nhân tố, chúng ta cần phải đồng nhất các nhân tố không nghiên cứu khác, để bảo đảm kết quả nghiên cứu là phản ảnh ảnh hưởng của các nhân tố thí nghiệm. Ví dụ, nghiên cứu ảnh hưởng của mật độ đến tăng trưởng rừng trồng, lúc này các ô thí nghiệm cần phải được đồng nhất các nhân tố khác có khả năng ảnh hưởng đến tăng trưởng như loại đất, vị trí địa hình, chế độ chăm sóc, bón phân. Có thể áp dụng bón phân nhưng phải đồng đều ở tất cả các công thức mật độ, nhưng loại đất thì nhất thiết phải như nhau ở các ô thí nghiệm.

*Bố trí ô thí nghiệm ngẫu nhiên:* Các ô thí nghiệm ứng với các công thức khác nhau cần được sắp xếp, bố trí một các ngẫu nhiên trên hiện trường, tránh tính hệ thống. Ví dụ các ô thí nghiệm cấp mật độ thưa < 1000 cây/ha không được luôn bố trí cạnh ô có cấp mật độ dày 2000 – 3000 cây/ha mà ngẫu nhiên với cấp mật độ khác, ví dụ với cấp 1000 – 2000 cây/ha. Lý do cần bố trí ngẫu nhiên là tránh sự tương tác có tính hệ thống, ví dụ ô mật độ dày có thể ảnh hưởng che ánh sáng đến ô lân cận. Vì vậy, để khách quan các ô thí nghiệm ở các công thức khác nhau, theo các nhân tố khác nhau và lần lặp lại cần bố trí một cách ngẫu nhiên trong khu vực thí nghiệm, không sắp xếp một cách hệ thống.




**Hình 2.5.** Sơ đồ bố trí thí nghiệm gieo ươm lan kim tuyến trong vườn ươm theo 3 nhân tố (che bóng (3 cấp), giá thể (3 loại) và bón phân (có hoặc không) với 3 lần lặp *(Nguyễn Thị Quỳnh, 2016)*

Tổng số có số ô thí nghiệm = 3 cấp che bóng * 3 loại giá thể * 2 cấp bón phân * 3 lần lặp = 48 ô thí nghiệm. Bố trí ngẫu nhiên: Mỗi cấp che bóng các ô thí nghiệm của các công thức giá thể được bố trí ngẫu nhiên, ứng với mỗi công thức giá thể, các ô thí nghiệm có hoặc không bón phân được bố trí ngẫu nhiên.

Chương 3

# TIN HỌC THỐNG KÊ MÔ TẢ VÀ KIỂM TRA PHÂN BỐ CHUẨN, DUNG LƯỢNG CỦA MẪU

## 3.1 Thông tin thống kê về đặc trưng của mẫu quan sát

Để có những thông số đặc trưng về một đối tượng quan sát như sinh trưởng của một lô rừng, sự đa dạng loài của lô rừng, ảnh hưởng của cháy rừng đến mật độ, chất lượng tái sinh, biến động trữ lượng, mật độ của một lô rừng trồng, trạng thái rừng... cần tiến thành thu thập dữ liệu theo một nhân tố chủ đạo và sau đó ước lượng, tính toán các đặc trưng cơ bản. Đây là các thông tin cơ bản về một đối tượng quan sát, theo một chỉ tiêu, nhân tố quan tâm. Gọi là mô tả đặc trưng của mẫu theo các chỉ tiêu thống kê.

Thống kê mô tả được Nguyễn Hải Tuất, Vũ Tiến Hinh, Ngô Kim Khôi (2006), Laar (2007) xem như là bước xử lý thống kê đầu tiên để chi ra các đặc điểm, đặc trưng của đối tượng quan sát, nghiên cứu. Các đặc trưng mẫu bao gồm tính các chỉ tiêu: Số trung bình, số trung vị (median), mode, phương sai, sai tiêu chuẩn, hệ số biến động, độ lệch, độ nhọn của dãy số liệu quan sát, phạm vi biến động của nó với một mức sai số cho phép đặt trước và các biểu đồ phân bố.

Ngoài ra đối với rút mẫu, cần quan tâm đến mẫu có đạt được phân bố chuẩn hay không. Việc này cần được làm rõ trong phân tích đặc trưng mẫu; đôi khi cũng cần xác định trước khi rút mẫu hoặc bố trí thí nghiệm. Vì mẫu đạt phân bố chuẩn thì các chỉ tiêu thống kê của nó là đại diện cho đối tượng quan sát và bảo đảm độ tin cậy.

**Sau đây là một số chỉ tiêu thống kê cơ bản mô tả đặc trưng mẫu:**

*Số trung bình (Average, mean):* Ký hiệu $\bar{X}$ đây là giá trị thống kê sử dụng phổ biến để đánh giá một mẫu và dùng để so sánh với các mẫu khác.

*Số trung vị (Median):* Là một số có vị trí ở giữa dãy số quan sát được xếp theo thứ tự từ nhỏ đến lớn. Nếu số quan sát là số chẵn, thì median được tính là trùng bình của hai giá trị nằm giữa dãy số.

*Mode*: Là giá trị có tần số xuất hiện nhiều nhất trong toàn bộ số liệu quan sát. Mode thường được sử dụng để chọn giá trị quan sát có tần số tập trung. Ví dụ chọn cây có đường kính phổ biến trong lâm phần. Giá trị này cũng dễ dàng tìm thấy nhờ biểu đồ phân bố tần số.

*Range*: Là khoảng biến động tuyệt đối giữa số lớn nhất (Max) và nhỏ nhất (Min) quan sát. Range = Max – Min

*Sai tiêu chuẩn (Standard Deviation):* Ký hiệu S, chỉ ra sự biến động của mẫu quan sát. S càng lớn thì dữ liệu quan sát của mẫu càng có sự biến động lớn, không tập trung. Công thức sau để tính S, với n là số mẫu quan sát, $X_i$ là giá trị quan sát, $\bar{X}$ là trung bình:

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2} \quad (3.1)$$

*Phương sai mẫu (Variance):* Ký hiệu $S^2$, là bình phương của sai tiêu chuẩn.

*Hệ số biến động (Coefficient of Variation):* Ký hiệu CV%, chỉ ra phần trăm biến động theo sai tiêu chuẩn so với số trung bình. Tính theo công thức sau:

$$CV\% = \frac{S}{\bar{X}}100 \quad (3.2)$$

*Sai số mẫu (Standard Error):* Ký hiệu $S_x$ tính theo công thức sau, với n là dung lượng mẫu quan sát:

$$S_x = \frac{S}{\sqrt{n}} \quad (3.3)$$

*Độ lệch (Skewness):* Chỉ ra phân bố tần số của mẫu quan sát lệch trái hay phải so với phân bố chuẩn. Độ lệch bằng 0 thì phân bố không bị lệch so với phân bố chuẩn, Stnd. Skewness (Standardized Skewness) là giá trị chuẩn hóa của độ lệch. Nếu -2 ≤ Stnd. Skewness ≤ +2 thì độ lệch của mẫu quan sát tiệm cận chuẩn với độ tin cậy 95%.

*Độ nhọn (Kurtosis):* Chỉ ra phân bố tần số của mẫu quan sát có đỉnh nhọn hay tù hơn so với phân bố chuẩn. Độ nhọn bằng 0 thì phân bố tần số có đỉnh tiệm cận chuẩn. Stnd. Kurtosis (Standardized Kurtosis) là giá trị chuẩn hóa của độ nhọn. Nếu -2 ≤ Stnd. Kurtosis ≤ +2 thì độ nhọn của mẫu quan sát tiệm cận chuẩn với độ tin cậy 95%.

*Ước lượng khoảng biến động của số trung bình*: Từ rút mẫu, tính được số trung bình và dựa vào biến động của nó để có thể ước lượng phạm vi biến động của trung bình ở một độ tin cậy cho trước. Công thức tính ước lượng khoảng số trung bình:

$$\mu = \bar{X} \pm t\frac{S}{\sqrt{n}} \quad (3.4)$$

Trong đó μ là biến động của ở một độ tin cậy cho trước, $\bar{x}$ là trung bình mẫu, S là độ chuẩn của mẫu, n là số mẫu quan sát, t biến số của hàm phân bố chuẩn được tính trên cơ sở mức ý nghĩa α

(thông thường là 0.05, ứng với độ tin cây 95% (Confidence Level)) và độ tự do df = n -1. Công thức tính t trong excel: t = tinv(0.05, n – 1).

*Biểu đồ phân bố tần số của chỉ tiêu quan sát được xếp theo cấp, nhóm,…*

Sử dụng Dữ liệu 1 trong Phụ lục để tính đặc trưng mẫu về trữ lượng (M) trong Statgraphics theo các bước sau:

Bước 1: Nhập dữ liệu từ excel vào Statgraphics:

File/Open/Open Data Source…

Trong hộp thoại đọc file excel, lựa chọn các thông số:

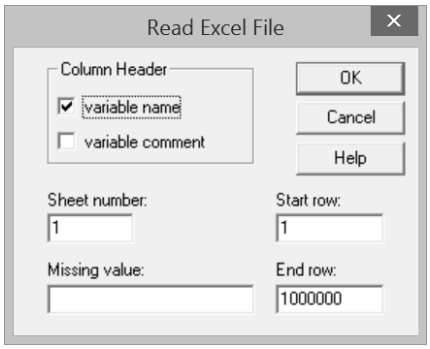
- Variable name: Có tên biến số

- Variable comment: Có chú giải về biến số

- Sheet number: Chọn số thứ tự sheetccuar excel chứa dữ liệu

- Start row: Dữ liệu bắt đầu từ hàng nào

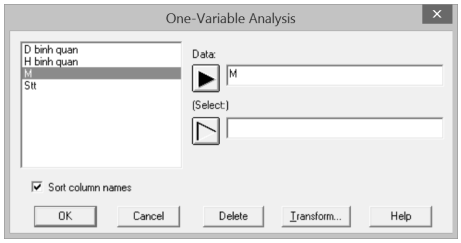
- End row: Dữ liệu kết thúc từ hàng nào.




*Bước 2*: Khai báo biến số tính toán, mô tả

Chọn biến số cần mô tả đưa vào Data.

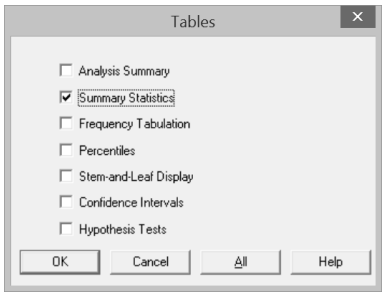
Ngoài ra còn có thể chọn Select: Có nghĩa là chọn một trường điều kiện, ví dụ Năm = 2015 (trong trường năm đo đạc, dữ liệu được đo nhiều năm, ở đây chỉ tính trong một năm nhất định)

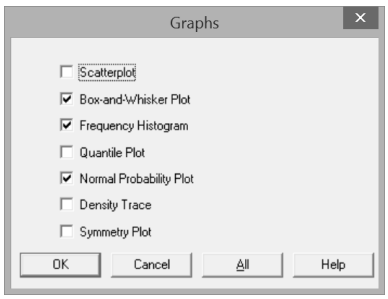



*Bước 3:* Chọn lựa các thông số, chỉ tiêu thống kê, biểu đồ để mô tả mẫu quan sát

Có hai hộp thoại được sử dụng để chọn xuất ra kết quả: Tables và Graphs.

Từ đây chọn các biểu đồ quan tâm



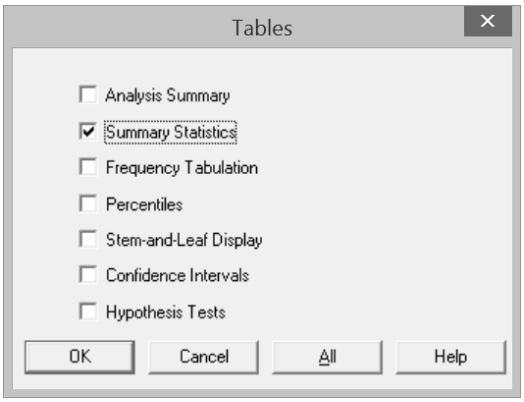
Sau đây là các kết quả và giải thích trong mô tả thống kê mẫu quan sát với ví dụ sử dụng Dữ liệu 1 trong Phụ lục: Khảo sát trữ lượng rừng của một trạng thái, sử dụng ô mẫu để đo tính trữ lượng (M, $m^3$/ha), từ đây mô tả thống kê các đặc trưng cơ bản về trữ lượng của trạng thái rừng này.

## 3.2 Tính toán các chỉ tiêu thống kê mô tả mẫu

Mô tả một mẫu quan sát thông qua các chỉ tiêu thống kê cơ bản như trình bày trên. Các chỉ tiêu này có thể tính toán nhanh chóng quan các chương trình excel, Statgraphics, SPSS, R. Sau đây là sử dụng Statgraphics.

Tóm tắt các chỉ tiêu thống kê mô tả (Summary Statistics):

Trong hộp thoại Tables, chọn Summary Statistics, có các chỉ tiêu thống kê sau:




Summary Statistics for M

| | |
|---|---|
| Count | 27 |
| Average | 76.1481 |
| Median | 78.0 |
| Mode | 78.0 |
| Variance | 572.67 |
| Standard deviation | 23.9305 |
| Coeff. of variation | 31.4263% |
| Standard error | 4.60543 |
| Minimum | 34.0 |
| Maximum | 124.0 |
| Range | 90.0 |
| Stnd. skewness | 0.249982 |
| Stnd. kurtosis | -0.415415 |

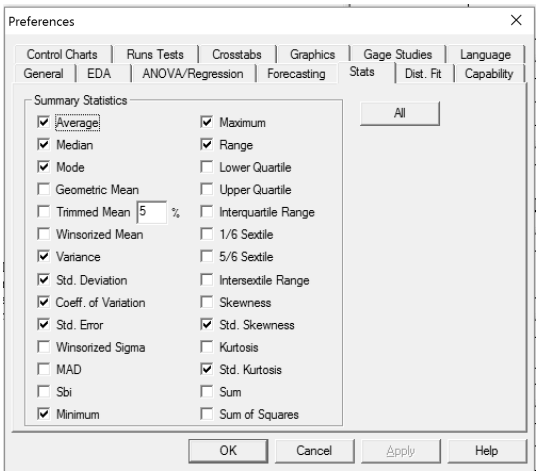
The StatAdvisor

This table shows summary statistics for M. It includes measures of central tendency, measures of variability, and measures of shape. Of particular interest here are the standardized skewness and standardized kurtosis, which can be used to determine whether the sample comes

Các chỉ tiêu thống kê xuất ra này có thể thay đổi, thêm vào hay bớt đi theo nhu cầu nghiên cứu.

Để thay đổi cài đặt chỉ tiêu thống kê xuất ra, tiến hành:

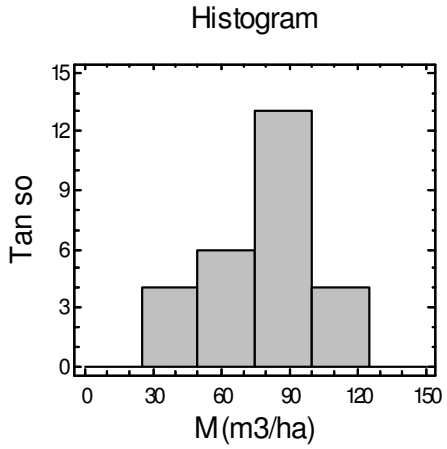
Edit/Preferences và lựa chọn trong hộp thoại



from a normal distribution. Values of these statistics outside the range of -2 to +2 indicate significant departures from normality, which would tend to invalidate any statistical test regarding the standard deviation. In this case, the standardized skewness value is within the range expected for data from a normal distribution. The standardized kurtosis value is within the range expected for data from a normal distribution.

Trong phần tư vấn thống kê của Statgraphics (The StatAdvisor) đã giải thích Stnd. Skewness và Stnd. Kurtosis nằm trong phạm vi -2 đến +2 nên phân bố của mẫu quan sát tiệm cận chuẩn.

Đặc trưng của mẫu cũng được biểu hiện qua phân bố tần số số ô mẫu theo cấp M trong Hình 3.1. Hình này cho thấy các phân bố ô mẫu có xu hướng tiệm cận phân bố chuẩn, với kiểu dạng phân bố khá đối xứng.




**Hình 3.1.** Phân bố số ô mẫu theo cấp trữ lượng (M, $m^3$/ha) thực hiện trong Statgraphics: Graphs/Frequency Hitogram

## 3.3 Kiểm tra phân bổ chuẩn của mẫu quan sát - bổ sung số liệu hoặc đổi biến số

Mẫu quan sát tiệm cận phân bố chuẩn rất quan trọng, nó cho biết giá trị trung bình và ước lượng biến động của nó có bảo đảm độ tin cậy hay không. Do vậy, phần mềm Statgraphics quan tâm áp dụng Stnd. Skewness và Stnd. Kurtosis để chỉ ra có đạt chuẩn hay không. Ngoài việc sử dụng Stnd. Skewness và Stnd. Kurtosis, còn có thể sử dụng các biểu đồ phân bố để đánh giá chuẩn hóa của mẫu như biểu đồ phân bố tần số, Quantile, Normal Probability, Q-Q, P-P (Wilk et al., 1968).

Nếu mẫu chưa chuẩn thì có hai giải pháp:

*Rút mẫu bổ sung*: Chỉ áp dụng trong điều tra, phỏng vấn; không thể áp dụng cho các bố trí thí nghiệm trồng cây, gieo ươm. Vì vậy, các thí nghiệm như vậy cần bố trí ngay từ đầu số mẫu lớn, ít nhất là 30.

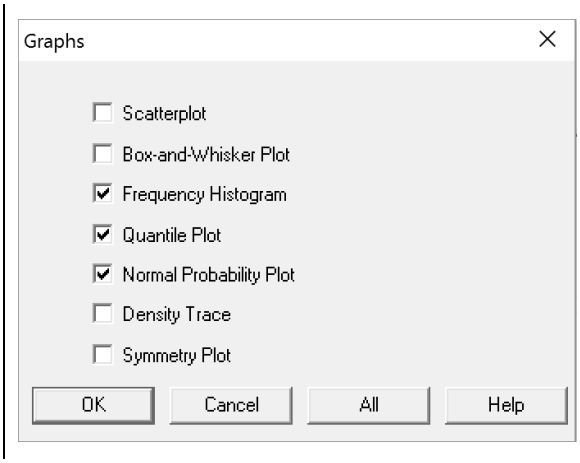
Việc bổ sung mẫu tính theo dưng lượng mẫu cần thiết để bảo đảm sai số cho trước và độ tin cậy (được trình bày trong mục trước "Tính toán dung lượng mẫu").

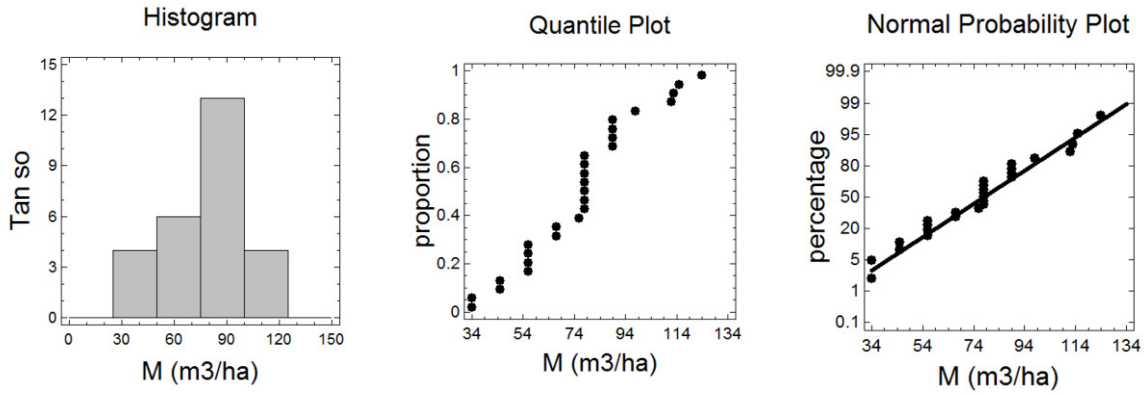
*Đổi biến số*: Bằng cách đổi biến số theo các dạng: logarit, sqrt, $x^2$, exp,... để biến mẫu có phân bố rời rạc thành liên tục, từ đây có thể tiệm cận được luật chuẩn. Thử với các cách đổi biến số khác nhau và kiểm tra có tuân theo phân bố chuẩn hay không.

Kiểm tra phân bố chuẩn của mẫu bằng biểu đồ trong Statgraphics:

Có ba loại biểu đồ thường dùng để đánh giá phân bố chuẩn được chọn trong hộp thoại Graphs:

Frequency Histogram, Quantile và Normal Probability Plot




**Hình 3.2.** Biểu đồ phân bố tần số Histogram, biểu đồ Quantile biểu diễn xác suất tích lũy theo biến số, biểu đồ Normal Probability biểu diễn xác suất % tích lũy theo biến số

Hình 3.2 cho thấy phân bố mẫu tiệm cận chuẩn khi có biểu đồ Histogram tiếp cận hình chuông (phân bố đối xứng), phân bố xác suất nằm trên đường chéo, càng xa hoặc lệch so với đường chéo thì mẫu có phân bố càng lệch chuẩn.

Ví dụ trong Hình 3.2 cho thấy mẫu quan sát khá chuẩn và phù hợp với kết quả theo Stnd. Skewness và Stnd. Kurtosis trên đây.

Phần mềm SPSS có chức năng lập các biểu đồ phân bố P-P và Q-Q để kiểm tra phân bố chuẩn của mẫu như dưới đây:

Lập biểu đồ P-P và Q-Q trong SPSS:

Thực hiện theo menu: Analyze/ Descriptive Statistics/ Chọn P-P Plots hoặc Q-Q plots.

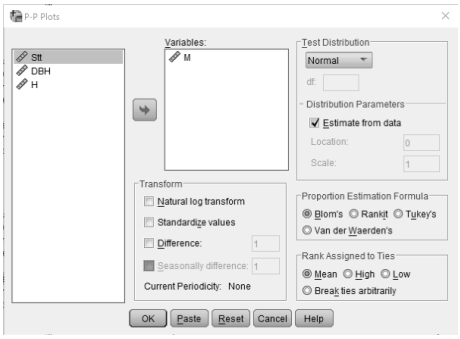
Trong hộp thoại P-P hoặc Q-Q Plots chọn biến số của mẫu cần kiểm tra luật chuẩn.

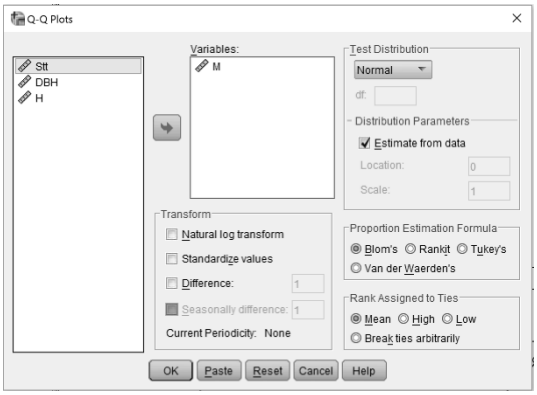
Ngoài ra còn có các lựa chọn khác như:

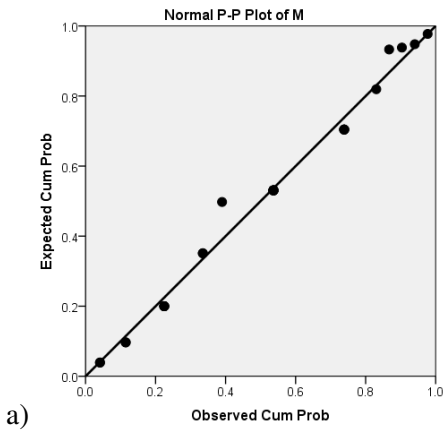
- Đổi biến số theo logarit neper để mẫu tiệm cận chuẩn: Chọn Natural log transform.

- Giá trị quan sát được chuẩn hóa: Chọn Standardized values.

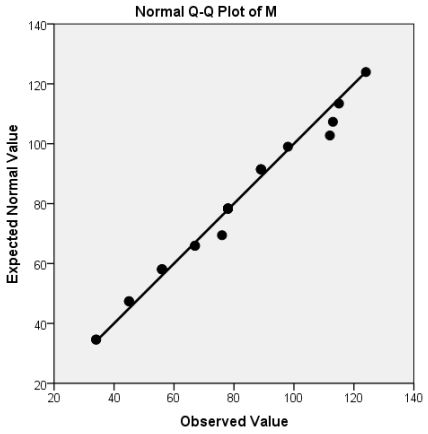
- ……




a)




b)

**Hình 3.3.** Biểu đồ P-P (a) và Q-Q (b)

Mẫu có phân bố chuẩn khi xác suất của giá trị quan sát và ước lượng nằm trên đường chéo từ tọa độ (0, 0) đến (1, 1) của biểu đồ P-P. Mẫu cũng có phân bố chuẩn khi giá trị quan sát và ước lượng phân bố trên đường chéo của biểu đồ Q-Q. Trong ví dụ rút mẫu ước tính trữ lượng rừng M, Hình 3.3 cho thấy dãy giá trị quan sát M có phân bố tiệm cận chuẩn, xác suất cũng như dữ liệu bám sát các đường chéo trên hai biểu đồ P-P và Q-Q.

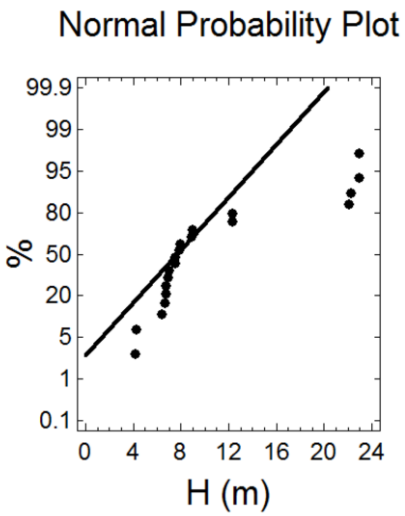
Một ví dụ khác là điều tra sinh trưởng chiều cao (H, m) của 20 cây Sao đen với số liệu như sau: 23.0; 23.0; 22.3; 22.1; 6.9; 7.0; 6.7; 6.4; 6.8; 6.8; 7.9; 8.0; 7.5; 7.5; 12.3; 12.3; 4.3; 4.2; 9.0; 8.9.

Kết quả tính đặc trưng mẫu và biểu đồ xác suất cho thấy việc rút mẫu với 20 cây để đánh giá sinh trưởng chiều cao (H) keo là chưa có độ tin cậy, vì mẫu chưa đủ (chưa chuẩn). Với Stnd.

Skewness = 2.34 > 2 và phân bố mẫu quan sát sai lệch quá lớn so với đường chéo chuẩn (trong Statgraphics).

Mô tả thống kê của mẫu đo chiều cao 20 cây sao đen thực hiện trong Statgraphics.

| | |
|---|---|
| Count | 20 |
| Average | 10.645 |
| Standard deviation | 6.44878 |
| Coeff. of variation | 60.5804% |
| Minimum | 4.2 |
| Maximum | 23.0 |
| Range | 18.8 |
| Stnd. skewness | 2.34108 |
| Stnd. kurtosis | 0.0990205 |




Biểu đồ xác suất của H cây Sao đen. Giá trị phân bố nằm rất lệch so với đường chéo cho thấy mẫu chưa chuẩn.

Với một mẫu điều tra chưa chuẩn, để chuẩn hóa có thể thông qua đổi biến số theo các dạng log(x), sqrt(x), 1/x.

Bổ sung thêm dữ liệu đo đạc để mẫu quan sát bảo đảm có phân bố chuẩn thường được thực hiện ở các nghiên cứu rút mẫu thông qua điều tra. Tính toán lại số cây đo đạc để bảo đảm độ tin cậy ước lượng 95% và sai số tương đối là 10% như sau:

Hệ số biến động: $CV\% = \frac{S}{\bar{X}}100 = \frac{6.449}{10.645}100 = 60.58\%$

Với độ tin cậy 95%: t = tiniv(0.05, 19) = 2.09 (thực hiện hàm tinv trong excel)

Số cây cần đo đạc để đạt sai số 10%: $n_{ct} = (t*CV\% / \Delta\%)^2 = \left(\frac{2.09*60.58\%}{10\%}\right)^2 = 160$ cây

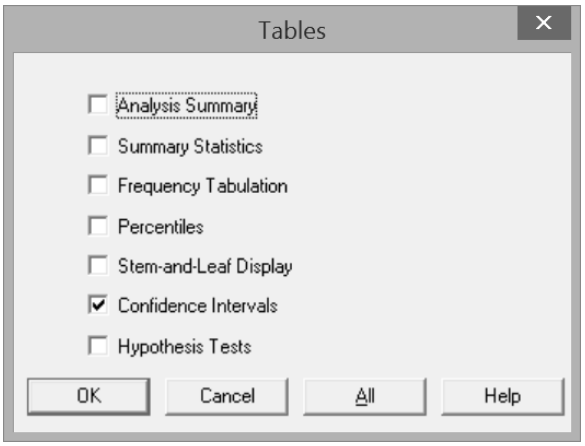
Như vậy khảo sát này chỉ mới đo tính được 20 cây, vậy số mẫu cần bổ sung để đạt chuẩn là 160 – 20 = 140 cây.

## 3.4 Ước lượng biến động của số trung bình với độ tin cậy cho trước

Số trung bình được ước lượng trong một khoảng biến động theo độ tin cậy cho trước. Độ tin cậy càng cao thì khoảng biến động càng rộng. Khoảng biến động theo một độ tin cậy (Confidence Intervals) trong Statgraphics được tiến hành như sau:

Ước lượng khoảng biến động của trung bình theo độ tin cậy cho trước:

Trong hộp thoại Tables, chọn Confidence Intervals




Kết quả như sau:

Confidence Intervals for M

95.0% confidence interval for mean: 76.1481 +/- 9.46662 [66.6815, 85.6148]

95.0% confidence interval for standard deviation: [18.8457, 32.7951]

Kết quả trên cho thấy biến động của trung bình trong khoảng ± 9.466$m^3$/ha, như vậy với độ tin cậy 95% thì trung bình trữ lượng M = từ 66.681 đến 85.615$m^3$/ha. Ngoài ra chương trình Statgraphics còn ước tính biến động của độ lệch chuẩn (Standard Deviation) tương ứng với biến động của trung bình.

Trong thực tế tùy theo đặc điểm, yêu cầu của cuộc điều tra đánh giá, thí nghiệm mà chọn mức độ tin cậy khác nhau, thông thường là 90% hoặc 95% hoặc 99% để ước khoảng biến động của số trung bình.

Chương 4

# TIN HỌC THỐNG KÊ SO SÁNH

## 4.1 So sánh trung bình một mẫu với một giá trị cho trước

Trong mô tả quan sát một mẫu, người ta có thể có yêu cầu đánh giá giá trị trung bình của mẫu với một giá trị cho trước, ví dụ từ đo đếm chiều cao của cây tái sinh trong rừng khộp, so sánh với một giá trị cho trước về chiều cao mong đợi để cây rừng vượt qua được lửa rừng, xem thật sự chiều cao tái sinh của lô rừng đó đã đạt yêu cầu hay chưa? Hoặc trong vườn ươm, so sánh trung bình chiều cao của cây con so với yêu cầu chiều cao xuất vườn.

Có thể có nhiều ví dụ cho việc áp dụng tiêu chuẩn thống kê này như là so sánh trung bình nồng độ $CO_2$ trong không khí với tiêu chuẩn an toàn; so sánh chỉ tiêu hàm lượng hóa chất có trong thực phẩm với nồng độ/hàm lượng cho phép,…

Để giải quyết vấn đề này, sử dụng kiểm định t một mẫu với điều kiện mẫu có phân bố chuẩn. Mẫu phân bố chuẩn để bảo đảm rằng giá trị trung bình là đủ đại diện cho đối tượng quan sát và biến số t của phân bố chuẩn có thể được áp dụng để đánh giá trung bình so với gía trị cho trước μ nào đó.

Giả thuyết $H_0$: $\bar{X} = \mu$: Trung bình giá trị quan sát bằng giá trị lý thuyết cho trước

Công thức t kiểm tra một mẫu với một giá trị cho trước (Laar và Akca, 2007):

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \quad (4.1)$$

Trong đó, $\bar{X}$ là giá trị trung bình của mẫu, μ là giá trị lý thuyết cho trước để so sánh với trung bình mẫu, S là sai tiêu chuẩn và n là số lượng mẫu quan sát.

Từ số liệu quan sát, với giả định mẫu có phân bố chuẩn, tính toán biến số t và kết luận:

- Nếu giá trị tuyệt đối t tính được: $|t| > t_{(0.05,\ df)}$ (t lý thuyết ở mức ý nghĩa α, thường là 5%) thì có thể kết luận có sự khác biệt có ý nghĩa thống kê giữa trung bình mẫu với giá trị cho trước đó (bác bỏ giả thuyết $H_0$). Trong trường hợp này nếu t tính < 0 thì có nghĩa trung bình của mẫu nhỏ thua có ý nghĩa so với giá trị cho trước, ngược lại nếu t tính > 0 thì trung bình của mẫu lớn hơn có

ý nghĩa so với giá trị cho trước. Đồng thời, để đơn giản, kết quả tính toán mức ý nghĩa thống kê α (thường là 5%) hay gọi là $P_{value}$ hay Significance alpha (Sig.), nếu Sig. < 0.05 thì kết luận có sự sai khác giữa trung bình mẫu với giá trị cho trước và nếu t < 0 thì mẫu có bình quân bé hơn gía trị so sánh và ngược lại nếu t > 0 thì trung bình lớn hẳn giá trị so sánh cho trước.

- Nếu |t| tính ≤ $t_{(0.05,\ df)}$ thì có thể kết luận ở mức ý nghĩa 5% trung bình mẫu quan sát xấp xỉ với giá trị cho trước (chấp nhận giả thuyết $H_0$). Hoặc $P_{value}$ hoặc α hoặc Sig. > 0.05.

Ví dụ: Rút mẫu đo tính chiều cao (H, m) cây tái sinh trong rừng khộp và kiểm tra xem trung bình H của cây tái sinh có lớn hơn 2m hay không; vì nếu H đã thực sự > 2m thì khu rừng này đã có lớp cây tái sinh có triển vọng thành cây gỗ, vượt qua được lửa rừng.

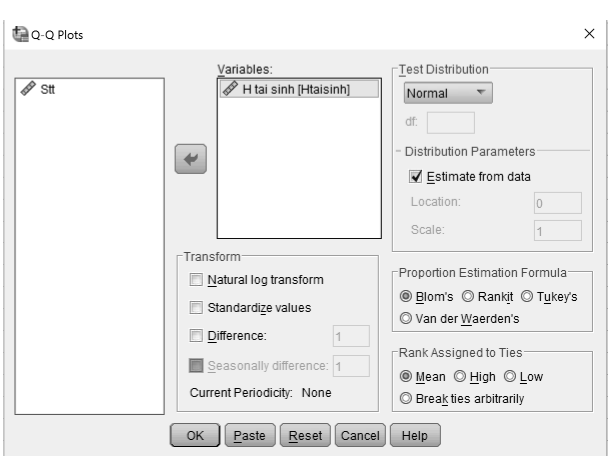
Sử dụng Dữ liệu 3 trong Phụ lục để so sánh chiều cao cây tái sinh rừng khộp với yêu cầu đạt chiều cao 2m, thực hiện trong chương trình SPSS.

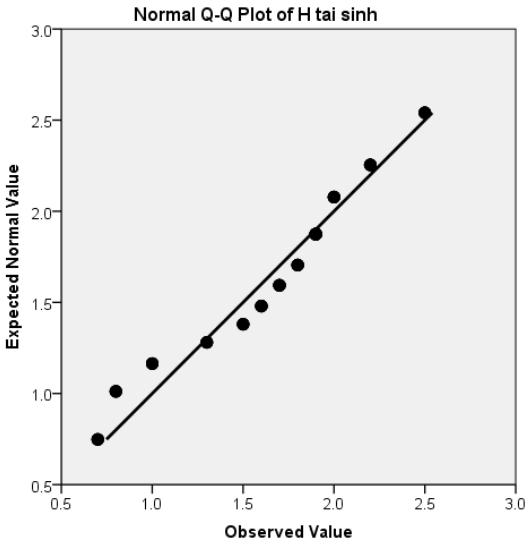
Đầu tiên kiểm tra mẫu có phân chuẩn hay không (hay số cây đo cao đã đủ để mẫu đạt chuẩn).

Sử dụng biểu đồ Q-Q để xem xét phân bố chuẩn của mẫu:

Trong SPSS, vào menu: Analyze/ Descriptive Statistics/Q-Q Plots.

Trong hộp thoại Q-Q Plots chọn biến số cần phân tích (H tái sinh)




**Hình 4.1.** Biểu đồ Q-Q của giá trị chuẩn và quan sát của H tái sinh cây con rừng khộp

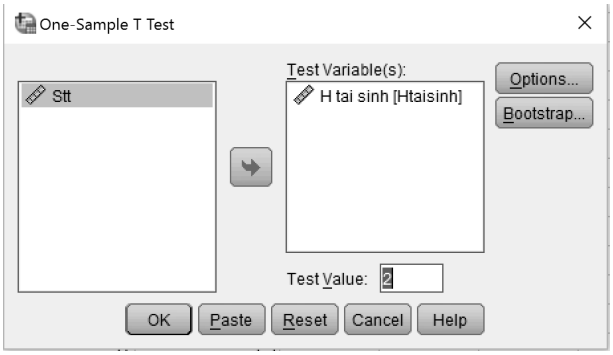
Kết quả sơ đồ Q-Q ở Hình 4.1 có giá trị ước lượng chuẩn và quan sát phân bố khá bám sát đường chéo, chấp nhận mẫu có phân bố tiệm cận chuẩn.

Trên cơ sở phân bố chuẩn của mẫu, sử dụng t test để so sánh trung bình mẫu với giá trị cho trước trong SPSS.

Sử dụng T test một mẫu trong SPSS:

Menu: Analyze/Compare Means/One-Sample T test.

Trong hộp thoại, chọn biến số phân tích (H tái sinh) và nhập giá trị cho trước để so sánh với trung bình: Test Value, trong ví dụ này là 2.




Kết quả T test của một mẫu so với giá trị cho trước test value = 2 trong SPSS:

One-Sample Statistics

| | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|
| H tai sinh | 61 | 1.644 | .4935 | .0632 |

One-Sample Test

| | Test Value = 2 | | | | | |
|---|---|---|---|---|---|---|
| | t | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference | |
| | | | | | Lower | Upper |
| H tai sinh | -5.630 | 60 | .000 | -.3557 | -.482 | -.229 |

Kết quả từ SPSS cho thấy chiều cao trung bình (Mean) của cây tái sinh rừng khộp của lô rừng nghiên cứu là 1.644m. Giá trị |t| = 5.63 với mức ý nghĩa Sig. (2-tailed) = 0.00 < < 0.05. Có nghĩa là chiều cao trung bình cây tái sinh khác biệt so với giá trị mong đợi 2 m so sánh ở độ tin cậy 95%. Trong khi đó t = - 5.63 < 0, vì vậy kết luận là trung bình H cây tái sinh nhỏ thua 2m rõ rệt, hay theo yêu cầu về sinh thái thì lô rừng này có cây tái sinh chưa vượt được lửa rừng.

## 4.2 So sánh hai mẫu quan sát - thí nghiệm

### 4.2.1 So sánh sự sai khác giữa trung bình hai mẫu quan sát độc lập

Trong các nghiên cứu, thí nghiệm thường người ta cần so sánh kết quả của 2 mẫu hoặc 2 công thức độc lập, ví dụ: Sản lượng rừng nơi có bón phân hay không bón phân, sinh trưởng cây con nơi có che bóng hay không che bóng, sinh trưởng và tái sinh của cây rừng nơi được chăm sóc và nơi không, sinh trưởng cây rừng nơi cháy và không cháy... Các nghiên cứu, thí nghiệm như vậy gọi

là so sánh hai mẫu độc lập. Việc áp dụng thống kê được tiến hành theo phương pháp so sánh 2 số trung bình của hai mẫu độc lập bằng các tiêu chuẩn t.

Giả thuyết $H_0$: $\bar{\mu}_1 = \bar{\mu}_2$, trong đó $\mu_1$ và $\mu_2$ là trung bình ước lượng tổng thể của hai mẫu

Công thức tính giá trị kiểm tra t (Nguyễn Hải Tuất, 1982; Ngôi Kim Khôi, 1998, Ngô Kim Khôi et al., 2002; Jayaraman, 1999):

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \quad (4.2)$$

Trong đó: $\bar{X}_1$, $\bar{X}_2$: Trung bình của mẫu 1 và 2.

$S_1^2$, $S_2^2$: Phương sai mẫu 1 và 2.

$n_1$, $n_2$: Dung lượng của mẫu 1 và 2.

Nếu giá trị tuyệt đối của t: |t| tính lớn hơn t lý thuyết với $\alpha$=0.05 và độ tự do df = $n_1+n_2-2$ (sử dụng excel để tính t lý thuyết: = tiniv(0.05, df)) thì bác bỏ giả thuyết $H_0$; hoặc $P_{value} < 0.05$, thì trung bình 2 mẫu sai khác có ý nghĩa. Ngược lại thì chấp nhận giả thuyết $H_0$, trung bình hai mẫu chưa có sự khác biệt có ý nghĩa.

Khi sử dụng tiêu chuẩn t để so sánh 2 mẫu độc lập, cần kiểm tra 2 điều kiện:

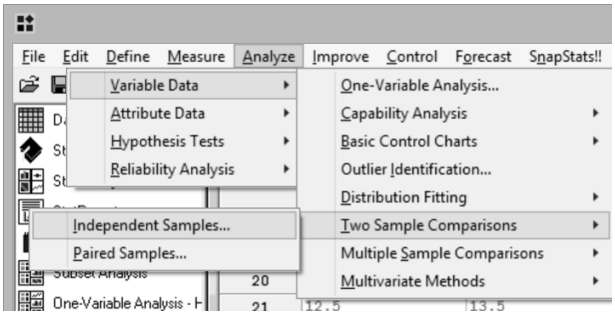
- Tần số phân bố của cả hai mẫu phải tuân theo phân bố chuẩn. Trong trường hợp mẫu chưa chuẩn, thì như đã giới thiệu trong mục thống kê mô tả, cần bổ sung thêm dữ liệu để mẫu đạt chuẩn, hoặc áp dụng tiểu chuẩn so sánh phi tham số *(trình bày ở các mục tiếp theo).*

- Phân biệt sai tiêu chuẩn hoặc phương sai của hai mẫu có bằng nhau hay không.

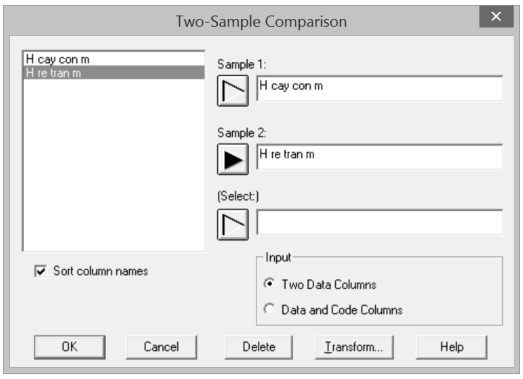
Ví dụ: Kiểm tra sai khác sinh trưởng chiều cao (H) của 2 phương pháp trồng thông 3 lá Pinus kesiya bằng cây con và rễ trần tại trạm thực nghiệm của Viện Nghiên cứu Lâm sinh ở Lang Hanh-Lâm Đồng: Mỗi công thức được rút mẫu độc lập theo ô tiêu chuẩn 1000$m^2$, trong ô đo đếm chiều cao tất cả các cây; số liệu trong Dữ liệu 4 ở phần Phụ lục.

Sử dụng Statgraphics để kiểm tra thống kê bằng tiêu chuẩn t trong trường hợp 2 mẫu độc lập:

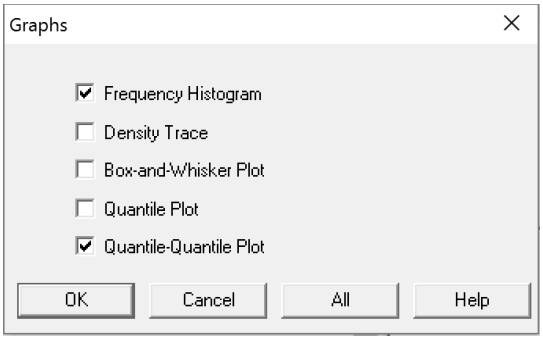
Sử dụng so sánh 2 mẫu độc lập trong Statgraphics: Analyze/ Variable Data/Two Sample Comparisions/ Independent Samples.



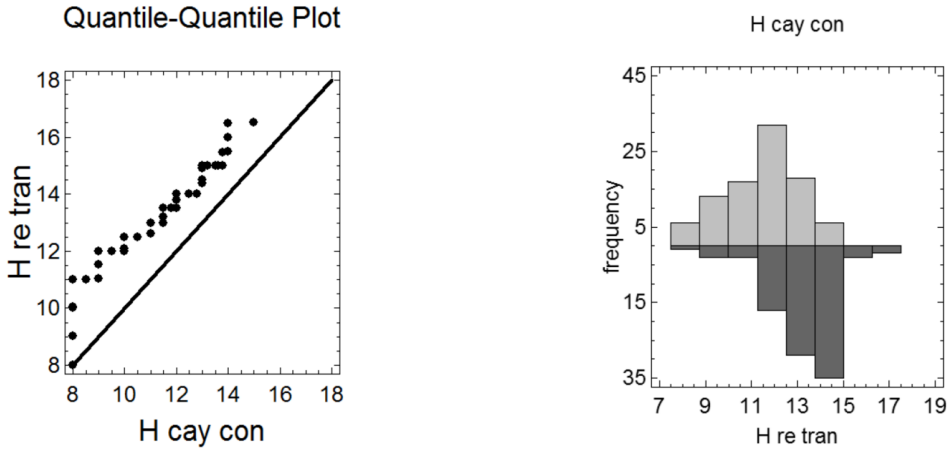
Trong hội thoại, chọn các biến số của hai mẫu so sánh




Kiểm tra hai mẫu có phân bố chuẩn hay không bằng biểu đồ: Trong hộp thoại Graphs chọn hai dạng biểu đồ để kiểm tra phân bố chuẩn của hai mẫu: Frequency Histogram và Quantile-Quantile Plot.




Kết quả hai biểu đồ kiểm tra phân bố chuẩn của hai mẫu thể hiện ở Hình 4.2. Từ biểu đồ Q-Q cho thấy xác suất phân bố của hai mẫu nằm lệch xa đường chéo và biểu đồ phân bố tần số cho thấy phân bố lệch. Vì vậy, cả hai mẫu đều có phân bố chưa chuẩn.




**Hình 4.2.** Biểu đồ Q-Q (trái) và phân bố tần số theo cấp H (phải) của hai mẫu thí nghiệm trồng thông 3 lá bằng cây con và rễ trần

Ngoài ra, để khẳng định có thể kiểm tra thông qua giá trị của Stnd. Skewness và Stnd. Kurtosis (kết quả của so sánh hai mẫu trong Summary Statistics). Kết quả cho thấy hoặc/và Stnd. Skewness và Stnd. Kurtosis nằm ngoài phạm vi [-2, +2]. Do đó khẳng định là hai mẫu chưa đạt chuẩn (Cách xác định các giá trị thống kê này sẽ trình bày trong bước tiếp theo khi so sánh hai mẫu). Trong trường hợp này không nên sử dụng tiêu chuẩn t để so sánh trung bình hai mẫu, vì kết quả có khả năng không phản ảnh đúng thực tế do mẫu được rút chưa chuẩn (chưa đủ lớn). Lúc này

có hai giải pháp: Rút bổ sung mẫu để đạt chuẩn và tiếp tục áp dụng tiêu chuẩn t hoặc áp dụng tiêu chuẩn thống kê phi tham số (không yêu cầu mẫu đạt chuẩn – sẽ giới thiệu ở phần tiếp theo trong giáo trình này). Tuy nhiêu, ở đây mẫu được thu thập khá lớn (>90 cây cho mỗi mẫu), do đó, tạm thời chấp nhận giả thuyết phân bố chuẩn của 2 mẫu để thử ứng dụng tiêu chuẩn t để so sánh.

Tiêu chuẩn F để kiểm tra sự bằng nhau của phương sai của hai mẫu với giả thuyết phương sai tổng thể của hai mẫu bằng nhau: Ho: $\sigma_1^2 = \sigma_2^2$. Công thức tính giá trị F (Jayaraman, 1999):

$$F = \frac{S_1^2}{S_2^2} \tag{4.3}$$

Trong đó $S_1^2$, $S_2^2$ là phương sai của mẫu 1 và 2. Kết quả F tính được sẽ so sánh với F lý thuyết ở mức P = 0.05 và độ tự do $df_1 = n_1 - 1$ và $df_2 = n_2 - 1$, với $n_1$ và $n_2$ là dung lượng mẫu của mẫu 1 và 2. Nếu $F < F_{(0.05, df1, df2)}$ hoặc $P_{value} > 0.05$ thì chấp nhận giả thuyết $H_0$ và kết luận là hai phương sai bằng nhau; ngược lại thì hai phương sai không bằng nhau. Giá trị F lý thuyết có thể xác định thông qua hàm =Finv(α, $df_1$, $df_2$) của Excel. Sau đây là tính F để kiểm tra sự bằng nhau của hai phương sai hai mẫu trong Statgraphics.

Kiểm tra phương sai của 2 mẫu bằng tiêu chuẩn F: Sử dụng hộp thoại Table và chọn: Comparision of Standard Deviations.

Cho ra kết quả:

Comparison of Standard Deviations

| | H cay con m | H re tran m |
|---|---|---|
| Standard deviation | 1.59993 | 1.46565 |
| Variance | 2.55976 | 2.14814 |
| Df | 91 | 92 |

Ratio of Variances = 1.19162

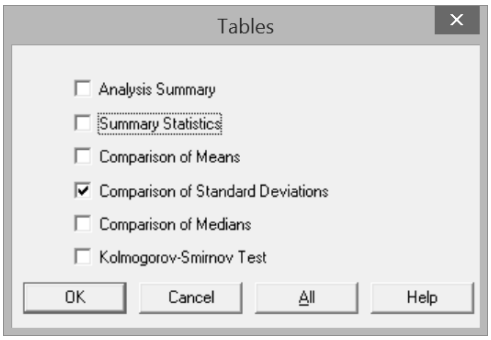
F-test to Compare Standard Deviations

Null hypothesis: sigma1 = sigma2

Alt. hypothesis: sigma1 NE sigma2

F = 1.19162  P-value = 0.403068

Do not reject the null hypothesis for alpha = 0.05.




Kết quả trên cho thấy F = 1.19 với $P_{value} = 0.40 > 0.05$, kết luận chưa thể bác bỏ giả thuyết $H_0$ về sự bằng nhau của hai phương sai của hai mẫu. Hay nói khác là chấp nhận giả thuyết $H_0$ (Null Hypothesis) là hai phương sai (sai tiêu chuẩn) của hai mẫu bằng nhau.

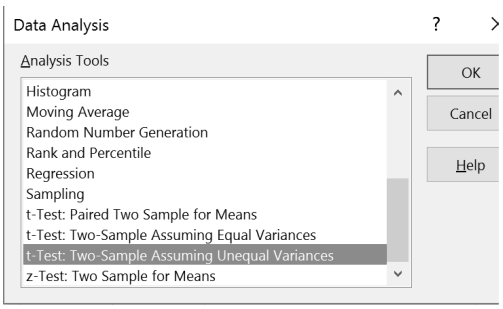
Trong trường hợp ngược lại nếu $P_{value} < 0,05$ thì phương sai 2 mẫu không bằng nhau, lúc này vẫn có thể áp dụng tiêu chuẩn t, nhưng trong trường hợp hai phương sai không bằng nhau. Excel đã cung cấp kiểm tra t cho cả hai trường hợp phương sai hai mẫu bằng nhau hoặc không bằng nhau.

Bao gồm: t-Test: Two-Sample Assumning Equal Variances và: t-Test: Two-Sample Assumning UnEqual Variances.

Dưới đây là ví dụ kiểm tra t với hai phương sai không bằng nhau khi sử dụng Excel

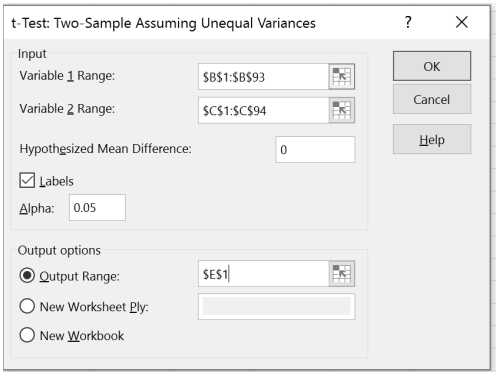
So sánh hai mẫu bằng tiêu chuẩn t trong Excel khi hai phương sai không bằng nhau: Data/Data Anlysis.

Trong hộp thoại chọn: t-Test: Two-Sample Assumning UnEqual Variances.




Trong hộp thoại: t-Test: Two-Sample Assumning UnEqual Variances. Cung cấp các thông tin:

- Variable 1 Range: Dãy địa chỉ dữ liệu của mẫu 1
- Variable 1 Range: Dãy địa chỉ dữ liệu của mẫu 2
- Hypothesized Mean Difference = 0
- Label: Kích chọn nếu có khai trong dãy dữ liệu
- Alpha: Mặc định là 0.05, có thể thay đổi theo mục đích nghiên cứu.
- Output Range: Nơi xuất ra kết quả phân tích t




**Bảng 4.1.** Kết quả so sánh trung bình hai mẫu độc lập bằng tiêu chuẩn t trong trường hợp phương sai hai mẫu không bằng nhau (thực hiện trong Excel)

t-Test: Two-Sample Assuming Unequal Variances

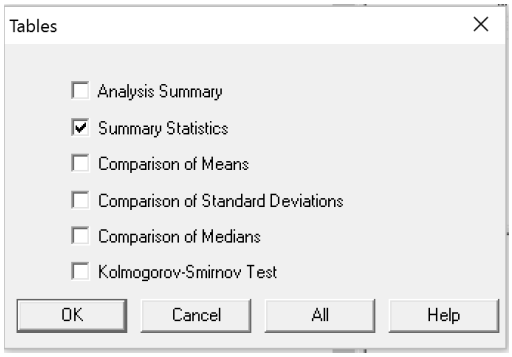
| Các chỉ tiêu thống kê | H cây con | H rễ trần |
|---|---|---|
| Mean (Trung bình, m) | 11.60435 | 13.40323 |
| Variance (Phương sai) | 2.559761 | 2.148142 |
| Observations (Số mẫu quan sát) | 92 | 93 |
| Hypothesized Mean Difference (Giả thuyết về sự khác nhau của hai trung bình | 0 | |
| Df (Độ tự do) | 181 | |
| t Stat | -7.97168 | |
| P(T<=t) one-tail | 8.52E-14 | |
| t Critical one-tail | 1.653316 | |
| P(T<=t) two-tail | 1.7E-13 | |
| t Critical two-tail | 1.973157 | |

Kết quả Bảng 4.1 cho thấy |t Stat| = 7.97 > t Critical two-tail = 1.97, hoặc P(T<=t) two-tail = 1.7E-13 < 0.05; có nghĩa là bác bỏ giả thuyết $H_0$ về sự bằng nhau của trung bình hai mẫu.

Trở lại với kiểm tra trung bình hai mẫu trong trường hợp hai phương sai bằng nhau (có thể áp dụng Excel như trường hợp hai phương sai không bằng nhau). Dưới đây tiếp tục giới thiệu áp dụng Statgraphics.

Mô tả hai mẫu trong Statgraphics.

Trong hộp thoại Tables, chọn: Summary Stattistics và Comparison of Means

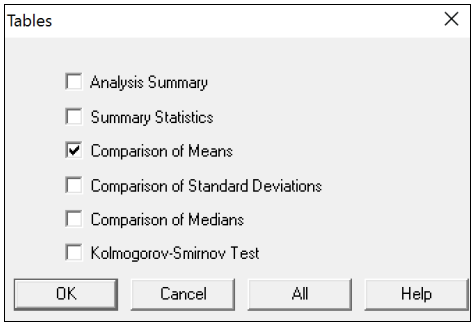



**Bảng 4.2.** Kết quả mô tả thống kê của hai mẫu trong Statgraphics

| Summary Statistics | H cay con | H re tran |
|---|---|---|
| Count | 92 | 93 |
| Average | 11.6043 | 13.4032 |
| Median | 12.0 | 13.5 |
| Mode | 12.5 | 14.0 |
| Variance | 2.55976 | 2.14814 |
| Standard deviation | 1.59993 | 1.46565 |
| Coeff. of variation | 13.7873% | 10.9351% |
| Standard error | 0.166804 | 0.151981 |
| Minimum | 8.0 | 8.0 |
| Maximum | 15.0 | 16.5 |
| Range | 7.0 | 8.5 |
| Stnd. skewness | -2.23744 | -3.38989 |
| Stnd. kurtosis | -0.398833 | 3.8466 |

Bảng 4.2 xuất ra các chỉ tiêu thống kê cho mỗi mẫu như đã giới thiệu trong mục thông kê mô tả. Kết quả này cũng khẳng định lại hai mẫu nghiên cứu chưa chuẩn vì Stnd. Skewness hoặc/và Stnd. Kurtosis nằm ngoài [-2, +2]. Do đó, kết quả so sánh trung bình hai mẫu trong trường hợp này chỉ là hướng dẫn ứng dụng Statgraphics.

So sánh trung bình hai mẫu trong Statgraphics (trường hợp hai mẫu chuẩn và phương sai bằng nhau): Trong hộp thoại Tables chọn Comparison of Means.




Kết quả như sau:

Comparison of Means:

95.0% confidence interval for mean of H cay con: 11.6043 +/- 0.331336 [11.273, 11.9357].

95.0% confidence interval for mean of H re tran: 13.4032 +/- 0.301848 [13.1014, 13.7051].

95.0% confidence interval for the difference between the means assuming equal variances: -1.79888 +/- 0.445016 [-2.24389, -1.35386]

t test to compare means:

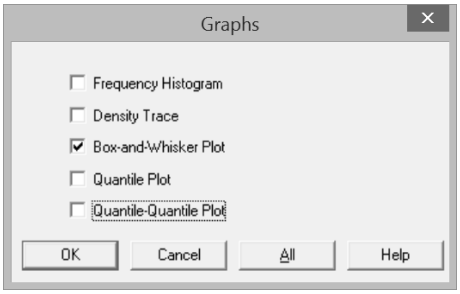
Null hypothesis: mean1 = mean2

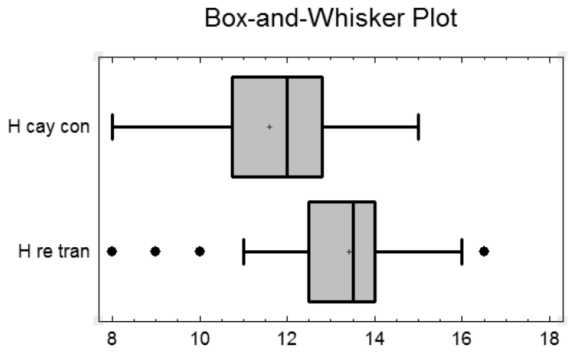
Alt. hypothesis: mean1 NE mean2 assuming equal variances: t = -7.97547 P-value = 1.79536E-7

Reject the null hypothesis for alpha = 0.05.

Kết quả trên cho thấy t = -7.97 và $P_{value}$ = 1.79E-7 < 0.05, có nghĩa là giả thuyết $H_0$ bị bác bỏ hay nói khác trung bình hai mẫu có sự sai khác ở độ tin cậy 95%.

So sánh sự sai khác của trung bình hai mẫu bằng biểu đồ Box and Whisker trong Statgraphics: Trong hộp thoại Graphs, chọn Box-and-Whisker Plot




**Hình 4.3.** Biểu đồ Box-Whisker biểu diễn trung bình và biến động H của hai mẫu quan sát trồng thông theo phương pháp cây con và rễ trần

Kết quả trên cho thấy sinh trưởng của *P. kesiya* trồng bằng 2 phương pháp khác nhau sai dị rõ. Chiều cao bình quân cây trồng bằng rễ trần hơn hẳn trồng bằng cây con qua biểu đồ, do vậy, trong trường hợp nghiên cứu này, phương pháp trồng thông 3 lá bằng rễ trần cần được ứng dụng trong thực tiễn.

### 4.2.2 So sánh sự sai khác hai trung bình của hai mẫu quan sát bắt cặp

Hai mẫu bắt cặp là trên cùng một đối tượng có một cặp dữ liệu được thu thập, có nghĩa là dữ liệu được sắp xếp và so sánh cặp đôi. Trong các nghiên cứu, có trường hợp cần so sánh kết quả từ hai phương pháp, thí nghiệm khác nhau trên cùng một đối tượng, ví dụ như bón phân theo độ tuổi khác nhau, như vậy sẽ có từng cặp dữ liệu sinh trưởng theo tuổi để so sánh sự ảnh hưởng của bón phân. Hoặc trong thẩm định các mô hình ước lượng, cần so sánh độ tin cậy của giá trị ước lượng qua mô hình với giá trị quan sát trên cùng một đối tượng. Ví dụ sử dụng mô hình quan hệ chiều cao (H) và đường kính của cây (D) để ước lượng H qua D mà không phải đo cao, để đánh giá độ tin cậy của mô hình, so sánh H ước lượng qua mô hình với H đo trực tiếp của từng cây, đây cũng là so sánh hai mẫu bắt cặp, tức là mỗi cây với D cụ thể sẽ có hai giá trị chiều cao cần so sánh với nhau.

Trường hợp này sử dụng tiêu chuẩn t bắt cặp (paired t test). Điều kiện để áp dụng tiêu chuẩn t này là sai lệch (d) của các cặp dữ liệu có phân bố chuẩn.

Có n cặp dữ liệu quan sát bao gồm: $(x_1, y_1)$, $(x_2, y_2)$…. $(x_n, y_n)$, với các giá trị quan sát $x_i$ có ước lượng trung bình tổng thể là $\mu_1$ và các giá trị quan sát $y_i$ có ước lượng trung bình tổng thể là $\mu_2$, lúc này giả thuyết $H_0$: $\mu_1 = \mu_2$. Trong đó sai lệch giữa từng cặp dữ liệu i là $d_i = x_i - y_i$ với i = 1, 2, … n và $d_i$ có phân bố chuẩn. Nói khác giả thuyết $H_0$: d = 0. Chỉ tiêu thống kê của t được tính theo công thức sau (Jayaraman, 1999; Nguyễn Hải Tuất et al., 2006):

$$t = \frac{\bar{d}}{\sqrt{\frac{S_d^2}{n}}} \tag{4.4}$$

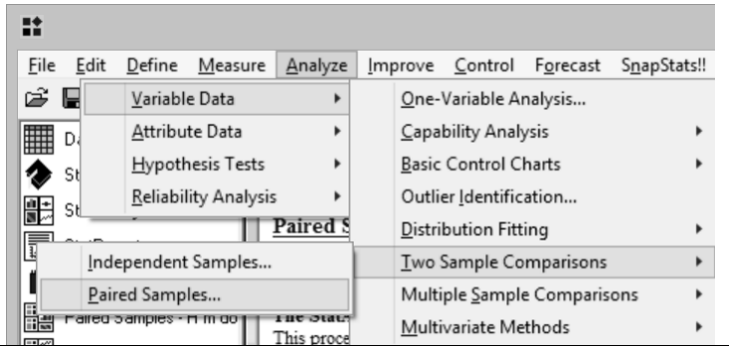
$$S_d^2 = \frac{1}{n-1}\left\{\sum d_i^2 - \frac{(\sum d_i)^2}{n}\right\} \tag{4.5}$$

Nếu $t > t_{(0.05,\ df = n-1)}$ hoặc $P_{value} < 0.05$ thì bác bỏ Ho, có nghĩa là trung bình của hai mẫu bắt cặp có sai khác có ý nghĩa ở độ tin cậy 95%. Ngược lại, nếu $t \leq t_{(0.05,\ df)}$ hoặc $P_{value} > 0.05$ thì chấp nhận Ho, hay chưa có sai khác giữa trung bình hai mẫu.

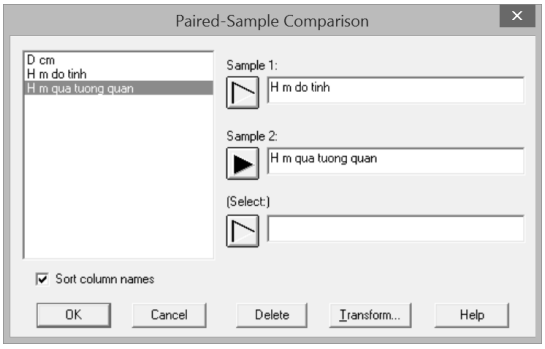
Ví dụ: Trong điều tra cây rừng, thường lập mô hình tương quan giữa chiều cao (H) theo đường kính (D) để từ đó ước tính H thông qua D để giảm chi phí khi đo cao cây. Tuy nhiên, để đánh giá độ tin cậy của mô hình tương quan, từ mỗi cây so sánh cặp dữ liệu gồm H đo cao trực tiếp và H ước tính qua mô hình tương quan. Đây là trường hợp so sánh 2 mẫu bắt cặp, tức là 2 giá trị bắt cặp trên một cây.

Sử dụng Statgraphics để so sánh bằng tiêu chuẩn t bắt cặp theo Dữ liệu 5 ở Phụ lục

Kiểm tra sai lệch 2 mẫu bắt cặp bằng tiêu chuẩn t: Variable Data/Two sample comparisions/Paired samples.

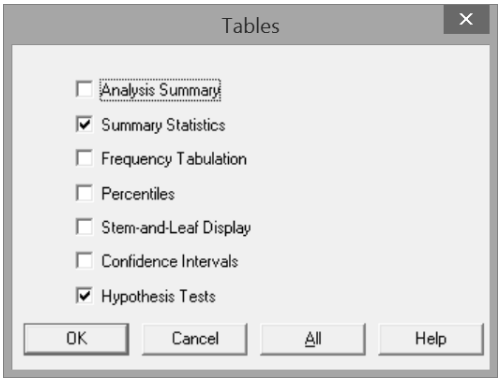


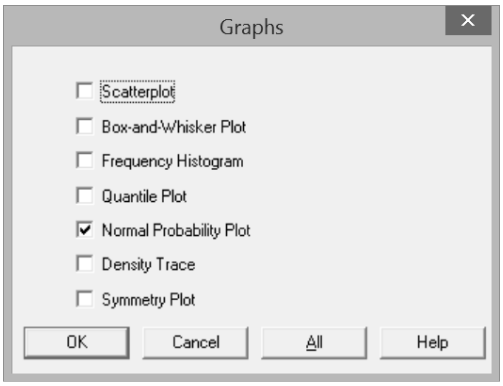

Trong hộp thoại chọn biến so sánh cho từng mẫu.



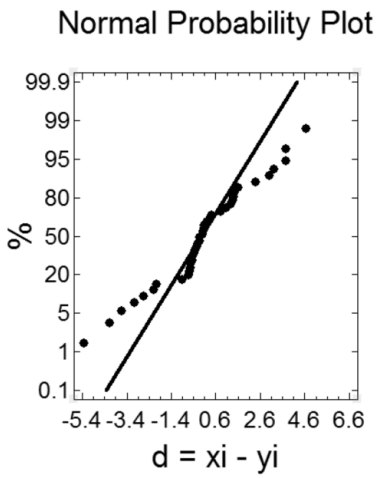

Kiểm tra sai lệch giữa hai mẫu có chuẩn hay không:

Trong hộp thoại Tables chọn Summary Statistics và trong Graphs chọn Normal Probability Plot




Kết quả có Stnd. Skewnes = -0.538061 và Stnd. Kurtosis = 0.81107 đều nằm trong phạm vi [-2, +2] và đồ thị xác suất phân bố chuẩn ở khá bám sát đường chéo (Hình 4.4); như vậy có thể chấp nhận các giá trị sai lệch (d) giữa hai mẫu tiệm cận chuẩn.




**Hình 4.4.** Biểu đồ phân bố xác suất chuẩn theo sai lệch giữa các dữ liệu bắt cặp (d)

Kiểm tra sự sai khác giữa các cặp quan sát trên cùng một mẫu: Trong hộp Table chọn Hypothesis.

*Kết quả:*

Hypothesis Tests for H m do truc tiep-H m qua tuong quan:

Sample mean = 0.0617335

Sample median = -0.0459924

Sample standard deviation = 2.11221

t-test

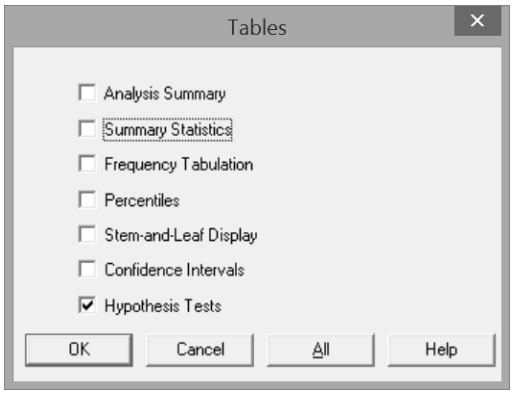
Null hypothesis: mean = 0.0

Alternative: not equal

Computed t statistic = 0.184848

P-Value = 0.854306

Do not reject the null hypothesis for alpha = 0.05.




Ở đây dùng tiêu chuẩn t bắt cặp để kiểm tra sai lệch (d) giữa H đo tính và H qua mô hình, với giả thuyết $H_0$ (Null Hypothesis) là trung bình sai lệch giữa 2 mẫu d = 0. Kết quả cho ra $P_{value}$ = 0.854 > 0.05, hoặc t = 0.18 < $t_{(0.05,\ df = n-1 = 39)}$ = 2.02; có nghĩa là chấp nhận giả thuyết $H_0$. Hay nói cách khác, trung bình sai lệch d là gần bằng 0, hay hai mẫu chưa có sự sai khác, hay H ước tính qua phương trình là bám sát với số liệu đo trực tiếp và có thể sử dụng phương trình vào thực tế.

### 4.2.3 Ước lượng biến động về tỷ lệ và so sánh tỷ lệ của hai mẫu

Tỷ lệ về một chỉ tiêu quan sát nào đó như là tỷ lệ sống, tỷ lệ cây tốt, tỷ lệ cây triển vọng,… thường được quan sát và cần chỉ ra khoảng biến động với một độ tin cậy cho trước (thường là 95%).

Gọi p là tỷ lệ quan sát, n là số mẫu đo đếm và q = 1-p; công thức ước lượng tỷ lệ của tổng thể mẫu P như sau (Nguyễn Hải Tuất, 1982; ayaraman, 1999):

$$P = p \pm t\sqrt{\frac{pq}{n}} \tag{4.6}$$

Với t lý thuyết ứng với mức ý nghĩa α (thường là 0.05) và độ tự do df = n-1. Giá trị t có thể xác định trong excel: = tinv(α, df).

Trong thực tế có thể cần so sánh tỷ lệ của một chỉ tiêu đánh giá nào đó giữa hai mẫu độc lập. Ví dụ so sánh tỷ lệ cây trồng sống hay tỷ lệ cây không bị sâu bệnh, tỷ lệ cây tốt,… giữa 2 lô thí nghiệm. Lúc này giả thuyết là tỷ lệ giữa hai mẫu là bằng nhau: Ho: $P_1 = P_2$, trong đó $P_1$ và $P_2$ là tỷ lệ tổng thể của hai mẫu.

Giá trị thống kê của t được tính theo công thức sau (Jayaraman, 1999):

$$t = \frac{p_1 - p_2}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \tag{4.7}$$

Trong đó $p_1$ và $p_2$ là tỷ lệ của mẫu 1 và 2; $q_1 = 1 - p_1$ và $q_2 = 1 - p_2$; $n_1$ và $n_2$ là dung lượng mẫu của mẫu 1 và 2.

Nếu giá trị $|t| > t_{(0.05,\ df = n1 + n2 - 2)}$ hoặc $P_{value} < 0.05$ thì bác bỏ Ho, có nghĩa là tỷ lệ của hai mẫu là khác nhau; ngược lại thì chấp nhận $H_0$ với kết luận là tỷ lệ hai mẫu là chưa có sự khác biệt rõ rệt.

Ví dụ đánh giá tỷ lệ sống của hai lô rừng trồng theo hai phương pháp khác nhau, mẫu 1 là bằng stump và mẫu 2 là bằng cây con có bầu. Kết quả có $p_1 = 0.65$, $p_2 = 0.85$; suy ra $q_1 = 0.35$ và $q_2 = 0.15$; $n_1 = 45$ và $n_2 = 57$. Tính được giá trị thống kê t:

$$t = \frac{0.65 - 0.85}{\sqrt{\frac{0.65x0.35}{45} + \frac{0.85x0.15}{57}}} = -2.34$$

Trong khi đó t lý thuyết: $t_{(0.05,\ df= 45+57-2 = 100)} = 1.98$ nhỏ thua |t|, kết luận là tỷ lệ sống của hai lô rừng là khác biệt có ý nghĩa ở độ tin cậy 95%. Trong đó, tỷ lệ sống của phương pháp trồng bằng cây con có bầu (85%) là cao hơn trồng bằng stump (65%); hay trong trường hợp này dùng cây con để trồng đạt tỷ lệ sống tốt hơn.

Ước lượng biến động tỷ lệ sống theo phương pháp trồng bằng cây con, trong đó với độ tin cậy 95% thì $t = tinv_{(0.05,\ 56)} = 2.0$:

$$P = p \pm t\sqrt{\frac{pq}{n}} = 0.85 \pm 2.0\sqrt{\frac{0.85x0.15}{57}} = 0.85 \pm 0.09$$

Chương 5

# TIN HỌC TRONG PHÂN TÍCH PHƯƠNG SAI (ANALYSIS OF VARIANCE - ANOVA)

Trong khi tiêu chuẩn t đề cập ở trên chỉ dừng lại so sánh giữa hai mẫu, hai công thức thí nghiệm. Trường hợp có nhiều hơn hai công thức thí nghiệm hoặc có nhiều nhân tố cần đánh giá, so sánh hơn; ví dụ thí nghiệm ba phương pháp trồng rừng (cây con có bầu, rễ trần, nuôi cấy mô) hoặc so sánh hai nhân tố là phương pháp trồng rừng (3 công thức) và nhân tố bón phân (2 công thức) đến sinh trưởng cây rừng trồng; lúc này tiêu chuẩn t hoàn toàn không có khả năng áp dụng để so sánh nhiều hơn hai công thức trong một nhân tố hoặc nhiều nhân tố, trong đó mỗi nhân tố có ít nhất 2 công thức thí nghiệm. Lúc này tiêu chuẩn thống kê phân tích phương sai cần được áp dụng.

Phân tích phương sai (Analysis of Variance – ANOVA) là một trong những phương pháp phân tích thống kê ứng dụng quan trọng, đặc biệt là trong các thí nghiệm để tìm ra các công thức và các nhân tố tác động đến hiệu quả, chất lượng của cây trồng, vật nuôi, gieo ươm, kiểm nghiệm xuất xứ cây trồng. ANOVA dùng để so sánh, đánh giá ảnh hưởng của nhiều công thức, nhân tố đến kết quả thí nghiệm, làm cơ sở cho việc lựa chọn công thức, phương pháp tối ưu trong lâm nghiệp.

Larson (2008) định nghĩa phân tích phương sai (ANOVA) là một kỹ thuật thống kê phân tích sự biến động của biến số phụ thuộc dưới các ảnh hưởng của các nhân tố. Thông thường chúng ta sử dụng ANOVA để kiểm tra sự giống nhau hay không giữa một vài giá trị trung bình thông qua so sánh phương sai giữa các nhóm/nhân tố tác động/ảnh hưởng trong mối quan hệ với biến động trong nội bộ của các nhóm/ nhân tố (sai số ngẫu nhiên).

ANOVA là một kỹ thuật cơ bản và là tiến trình biểu diễn tổng số các ảnh hưởng của các nhân tố, công thức thí nghiệm đến biến số phụ thuộc, quan sát theo mô hình sau đây (Yayaraman, 1999):

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij},\ i = 1, 2, ...., t;\ j = 1, 2, ...., ni \qquad (5.1)$$

Trong đó $y_{ij}$ là biến phụ thuộc, bị tác động của lần lặp/quan sát thứ j cùa nhân tố ảnh hưởng i, $\mu$ trung bình của tổng thể, $\alpha_i$ là ảnh hưởng của nhân tố i và $\varepsilon_{ij}$ là sai số ngẫu nhiên trong phạm vi của nhân tố i và lần quan sát j.

Giả thuyết $H_0$ và $H_1$ trong ANOVA được trình bày như sau.

Giả thuyết $H_0$:

$$Ho = \bar{\mu}_1 = \bar{\mu}_2 = \cdots = \bar{\mu}_t \quad (5.2)$$

Trong đó $\bar{\mu}_1, \bar{\mu}_2, \ldots, \bar{\mu}_n$ là trung bình giá trị quan sát ở các công thức khác nhau trong một nhân tố thí nghiệm; hoặc là trung bình theo các nhân tố khác nhau được so sánh, đánh giá.

Giả thuyết $H_1$ sẽ được chấp nhận (tức là có sự sai khác giữa các trung bình) nếu giả thuyết $H_0$ bị bác bỏ ($P_{value} < 0.05$). Lúc này giả thuyết $H_1$ cho biết có sự khác biệt ít nhất của hai số trung bình ở hai công thức hoặc ở hai nhân tố bất kỳ i và j trong các công thức, nhân tố so sánh:

$$H_1: \bar{\mu}_i \# \bar{\mu}_j \text{ hay } H_1: \bar{\mu}_i - \bar{\mu}_j \# 0 \quad (5.3)$$

Lúc này để biết được trung bình i nào khác biệt rõ với trung bình j (công thức, nhân tố nào khác biệt nhau), có thể tiếp tục sử dụng tiêu chuẩn t để so sánh bắt cặp như đã giới thiệu ở phần trên. Tuy nhiên, sử dụng t sẽ mất thời gian, dài dòng vì phải so sánh tất cả các cặp đôi, trong khi đó, có thể sử dụng rất đa dạng các tiêu chuẩn thống kê xếp nhóm các trung bình có sự sai biệt nhau; đó là các tiêu chuẩn: Duncan, LSD của Fisher (sai biệt có ý nghĩa ít nhất), HSD của Tukey, Scheffe, Bonferroni, Student-Newman-Keuls.

Có ba điều kiện trong bố trí thí nghiệm, thu thập số liệu để có thể sử dụng phân tích phương sai là:

- Các giá trị quan sát ở các mẫu, ô thí nghiệm của mỗi công thức cần đạt chuẩn.

- Phương sai của các mẫu, nhân tố thí nghiệm là bằng nhau.

- Mỗi công thức của nhân tố thí nghiệm đều được lặp lại ít nhất 2 lần.

Việc kiểm tra số liệu mỗi mẫu, công thức có đạt chuẩn hay không (đã trình bày trong mục thống kê mô tả ở trên). Trong đó sử dụng chỉ tiêu thống kê Stnd. Kurtosis và Skewness trong chương trình Statgraphics, đồng thời là các biểu đồ phân bố tần số, phân bố chuẩn như P-P, Q-Q để đánh giá dữ liệu của ô mẫu, công thức thí nghiệm đạt chuẩn hay không. Trong thực tế bố trí thí nghiệm trong phòng thí nghiệm, gây trồng, gieo ươm, để dữ liệu quan sát sau này có khả năng tiệm cận chuẩn, thì dung lượng mẫu ở mỗi ô/công thức thí nghiệm (số cây, số mẫu) cần đủ lớn, thông thường ít nhất là 30 và lớn nhất là 50. Với kinh nghiệm thống kê cho thấy các mẫu có dung lượng >30 có nhiều khả năng tiệm cận được luật chuẩn.

Sai tiêu chuẩn hoặc phương sai các mẫu hoặc nhân tố bằng nhau là điều kiện quan trọng để phân tích phương sai. Nếu phương sai của các mẫu, nhân tố có sự sai khác rõ rệt, khi áp dụng ANOVA sẽ cho ra kết quả đánh giá thống kê kém hiệu lực. Trong thực tế nhiều báo cáo nghiên cứu áp dụng ANOVA nhưng bỏ qua đi điều kiện này, đây là điều đáng tiếc và cho dù có dựa vào bố trí thí nghiệm tốt đến thế nào, nhưng nếu các công thức có phương sai không bằng nhau thì kết quả ANOVA có thể chỉ ra công thức, nhân tố tối ưu không đáng tin cậy. Vì thế, khi áp dụng ANOVA cần xem xét cẩn thận điều kiện phương sai bằng nhau. Cần kiểm tra giả thuyết $H_0$ về các phương sai các mẫu, các nhân tố bằng nhau. Có nhiều tiêu chuẩn để kiểm tra sự bằng nhau của các phương sai như Cochran, Barlett hoặc Leneve (Nguyễn Hải Tuất, 1982; Nguyễn Hải Tuất et al., 1996, 2005, 2006), kết quả kiểm tra nếu Pvalue < 0.05 thì bác bỏ $H_o$ có nghĩa là phương sai của các mẫu, nhân tố không bằng nhau, lúc này cần thay thế bằng tiêu chuẩn thống kê phi tham số, (sẽ trình bày

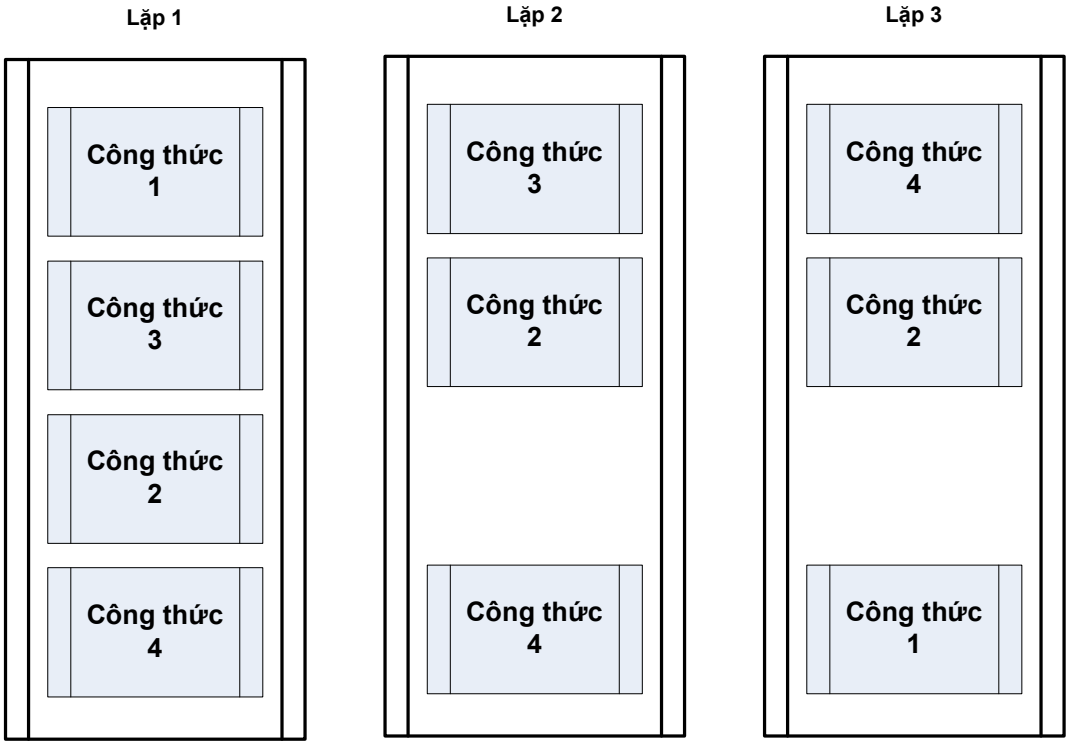
ở các mục tiêu theo). Trong đó kiểm tra Levene thường được sử dụng rộng rãi để kiểm tra giả thuyết $H_0$ là các phương sai là đồng nhất (Larson (2008).

Yêu cầu có lần lặp lại là điều kiện cơ bản để áp dụng ANOVA, có nghĩa là mỗi công thức, nhân tố thí nghiệm cần có ít nhất hai lần lặp lại. Điều kiện này cần được quan tâm trong bố trí thí nghiệm để sử dụng ANOVA.

## 5.1 Phân tích phương sai một nhân tố với bố trí thí nghiệm ngẫu nhiên hoàn toàn

Phân tích này nhằm nghiên cứu ảnh hưởng của một nhân tố ví dụ như: xuất xứ, mật độ trồng khác nhau, chế độ chăm sóc khác nhau,... đến sinh trưởng, sản lượng cây rừng. Trong nhân tố thí nghiệm sẽ có ít nhất hai công thức (tổng quát là a công thức), ví dụ a xuất xứ cây trồng và mỗi công thức theo điều kiện ANOVA cần được lặp lại ít nhất 2 lần (tổng quát có m lần lặp). Mỗi lần lặp lại của một công thức được bố trí thành một ô thí nghiệm, trong ô số cây, mẫu cần đủ lớn để tiệm cận chuẩn (30 – 50 cây, mẫu). Số lần lặp lại của các công thức có thể bằng hoặc không bằng nhau. Ngoài ra để bảo đảm tính khách quan, thì trong mỗi lần lặp, vị trí của các ô thí nghiệm của các công thức thí nghiệm cần được sắp xếp ngẫu nhiên (bốc thăm) – vì vậy còn gọi là bố trí thí nghiệm ngẫu nhiên hoàn toàn. Ngoài ra, vị trí ứng với một lần lặp lại các công thức cần đồng nhất các yếu tố khác không tham gia thí nghiệm.

Hình 5.1 minh họa các kiểu bố trí thí nghiệm để áp dụng ANOVA một nhân tố với bố trí thí nghiệm ngẫu nhiên có số lặp không bằng nhau.




**Hình 5.1.** Sơ đồ bố trí thí nghiệm để phân tích phương sai một nhân tố với bố trí ngẫu nhiên có số lần lặp không bằng nhau

*Ghi chú*: Diện tích mỗi lần lặp được chọn trên lập địa đồng nhất các yếu tố không thí nghiệm. Các công thức thí nghiệm của nhân tố nghiên cứu được bốc thăm để bố trí ngẫu nhiên trong mỗi lần lặp, số lần lặp của các công thức là không bằng nhau nhưng ít nhất là 2. Mỗi công thức ở một lần lặp là một ô thí nghiệm, ô có mẫu (số cây) đủ lớn để bảo đảm dữ liệu của ô đạt chuẩn.

Ví dụ: Khảo nghiệm 7 xuất xứ loài thông *Pinus caribeae* tại Trạm Thực nghiệm Lâm sinh Lang Hanh - Lâm Đồng. Thí nghiệm có 7 xuất xứ với 5 xuất xứ được trồng lặp lại 4 lần, còn 2 xuất xứ chỉ được lặp lại 2 lần vào năm 1991 (Bảng 5.1).

**Bảng 5.1.** Bảy xuất xứ P.caribeae được trồng thí nghiệm để phân tích phương sai một nhân tố với số lần lặp lại không bằng nhau

| Mã số xuất xứ | Tên loài theo xuất xứ | Số lần lặp (ô thí nghiệm) |
|---|---|---|
| 1 | P.alamicamba (NIC) | 4 |
| 2 | P.poptun (Guat) | 4 |
| 3 | *P.guanaja* (Nonduras) | 4 |
| 4 | *P.linures* (Nonduras) | 4 |
| 5 | *P.R482* (Australia) | 2 |
| 6 | *P.T473* (Australia) | 4 |
| 7 | P.little asaco (Bahamas) | 2 |

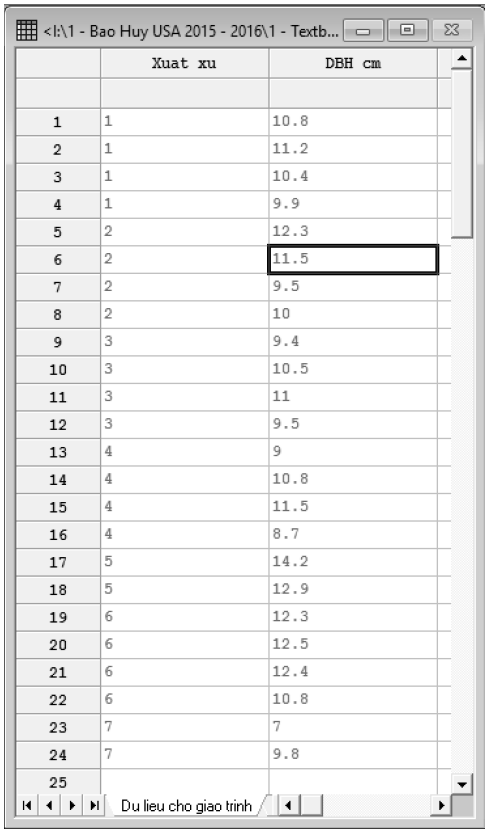
Tổng diện tích bố trí thí nghiệm là 1ha. Mỗi xuất xứ ứng với một lần lặp được trồng 25 cây, với cự ly 3x2m. Các điều kiện đất đai, vi khí hậu, địa hình, chăm sóc... đều được đồng nhất ở các lần lặp lại, nhân tố thay đổi để khảo sát chỉ còn lại là các xuất xứ khác nhau. Tại thời điểm điều tra (năm 1996), cây trồng trong các ô thí nghiệm có tuổi là 5. Tiến hành đo đếm toàn diện các chỉ tiêu đường kính ngang ngực (DBH, cm), chiều cao (H, m), đường kính tán ($D_t$, m), phẩm chất, tỉa cành, hình thân. Mỗi ô thí nghiệm tính giá trị trung bình của sinh trưởng và sử dụng hai chỉ tiêu DBH và H trung bình để đánh giá sinh trưởng của các xuất xứ thử nghiệm.

Dữ liệu trung bình DBH (cm) của các ô thí nghiệm theo 7 xuất xứ ở Dữ liệu 6 trong Phụ lục.

Sử dụng ANOVA để so sánh sự khác nhau về sinh trưởng DBH của 7 xuất xứ. Trong ba điều kiện để áp dụng ANOVA, thì ở đây chấp nhận 2 điều kiện: Mẫu chuẩn (vì mỗi ô thí nghiệm có số cây trồng đủ lớn, gần xấp xỉ 30 cây) và mỗi công thức được lặp lại ít nhất 2 lần. Lúc này cần kiểm tra sự bằng nhau của các phương sai ở các công thức thí nghiệm (xuất xứ).

Sử dụng phân tích phương sai một nhân tố để kiểm tra sự sai khác sinh trưởng DBH trung bình của 7 xuất xứ trong chương trình Statgraphics theo trình tự sau đây:

Nhập dữ liệu từ Excel vào Statgraphics: Trong đó có hai cột: Cột thứ nhất là mã số xuất xứ khác nhau, cột thứ hai là chỉ tiêu đánh giá (DBH, cm) trung bình của các ô thí nghiệm được lặp lại theo các xuất xứ khác nhau.



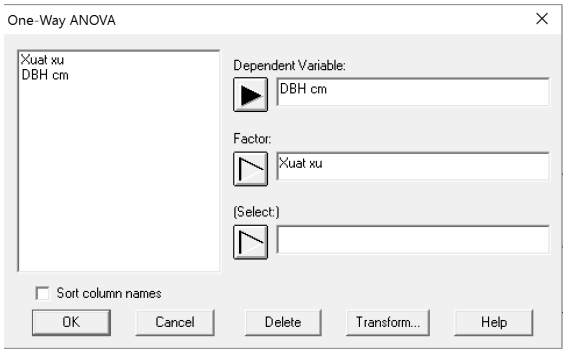

| | Xuat xu | DBH cm |
|---|---|---|
| 1 | 1 | 10.8 |
| 2 | 1 | 11.2 |
| 3 | 1 | 10.4 |
| 4 | 1 | 9.9 |
| 5 | 2 | 12.3 |
| 6 | 2 | 11.5 |
| 7 | 2 | 9.5 |
| 8 | 2 | 10 |
| 9 | 3 | 9.4 |
| 10 | 3 | 10.5 |
| 11 | 3 | 11 |
| 12 | 3 | 9.5 |
| 13 | 4 | 9 |
| 14 | 4 | 10.8 |
| 15 | 4 | 11.5 |
| 16 | 4 | 8.7 |
| 17 | 5 | 14.2 |
| 18 | 5 | 12.9 |
| 19 | 6 | 12.3 |
| 20 | 6 | 12.5 |
| 21 | 6 | 12.4 |
| 22 | 6 | 10.8 |
| 23 | 7 | 7 |
| 24 | 7 | 9.8 |
| 25 | | |

Tiến hành ANOVA một nhân tố trong Statgraphics: Analyze/Variable Data/One Variable Analysis.

Trong hộp thoại chọn:

- Dependent Variable: Biến số khảo sát (DBH)

- Factor: Nhân tố (xuất xứ)

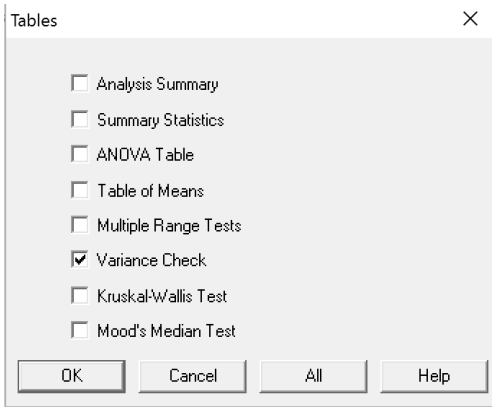



Kiểm tra sự bằng nhau của các phương sai ở các xuất xứ khác nhau theo tiêu chuẩn Leneve trong Statgraphics:

Trong nút Tables, chọn Variance Check.

Kết quả:

Variance Check



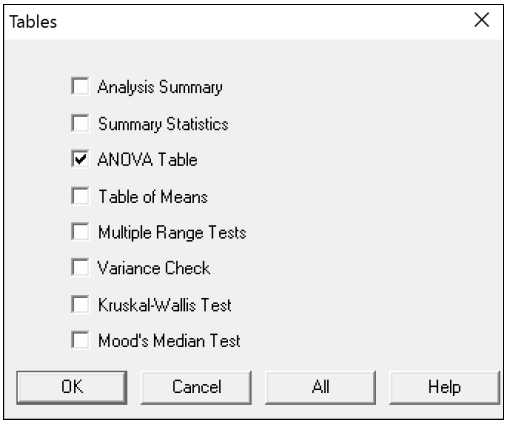

| | Test | P-Value |
|---|---|---|
| Levene's | 2.73661 | 0.0477133 |

Kết quả kiểm tra sự bằng nhau của các phương sai theo tiêu chuẩn Leneve cho thấy giá trị $P_{value} = 0.047 < 0.05$. Lúc này giả thuyết $H_0$ về các phương sai bằng nhau bị bác bỏ, hay nói khác là chấp nhận giả thuyết $H_1$ tức là có ít nhất hai phương sai ở hai công thức (xuất xứ) là có sự khác biệt rõ rệt. Trong trường hợp này sử dụng ANOVA để đánh giá sinh trưởng ở các xuất xứ khác nhau là chưa bảo đảm yêu cầu thống kê. Vì vậy, cần chuyển sang sử dụng các tiêu chuẩn thống kê so sánh phi tham số (được trình bày ở mục tiếp theo trong sách này).

Tuy nhiên, để giới thiệu cách áp dụng ANOVA một nhân tố, trong ví dụ này giả định rằng các phương sai là bằng nhau để tiếp tục so sánh đánh giá DBH ở các xuất xứ thông khác nhau.

ANOVA 1 nhân tố trong Statgraphics:

Trong Tables, chọn ANOVA Table để sử dụng ANOVA đánh giá sự sai khác trung bình DBH ở các xuất xứ khác nhau (Bảng 5.2).




**Bảng 5.2**. Kết quả phân tích phương sai (ANOVA) một nhân tố bố trí ngẫu nhiên hoàn toàn từ chương trình Statgraphics

ANOVA Table for DBH cm by Xuat xu

| Source | Sum of Squares | Df | Mean Square | F-Ratio | P-Value |
|---|---|---|---|---|---|
| Between groups | 37.0596 | 6 | 6.1766 | 5.22 | 0.0033 |
| Within groups | 20.1 | 17 | 1.18235 | | |
| Total (Corr.) | 57.1596 | 23 | | | |

Kết quả ANOVA cho thấy P-Value = 0.0033 < 0.05, có nghĩa là cần bác bỏ giả thuyết $H_0$ về các trung bình DBH ở các xuất xứ là bằng nhau. Hay chấp nhận giả thuyết $H_1$ với ít nhất có hai xuất xứ cho kết quả sinh trưởng bình quân DBH khác nhau.

Như vậy, kết quả đánh giá ANOVA chỉ cho biết có sự khác biệt DBH trung bình giữa các xuất xứ, chưa chỉ ra sự khác biệt từng xuất xứ với nhau và xuất xứ nào là tối ưu. Vì vậy cần tiếp tục sử dụng một trong các tiêu chuẩn xếp nhóm khác biệt giữa các xuất xứ khác nhau (Multiple Range Test), trong đó phổ biến hay được sử dụng là tiêu chuẩn Duncan hoặc tiêu chuẩn khác biệt có ý nghĩa tối thiểu LSD của Fisher.

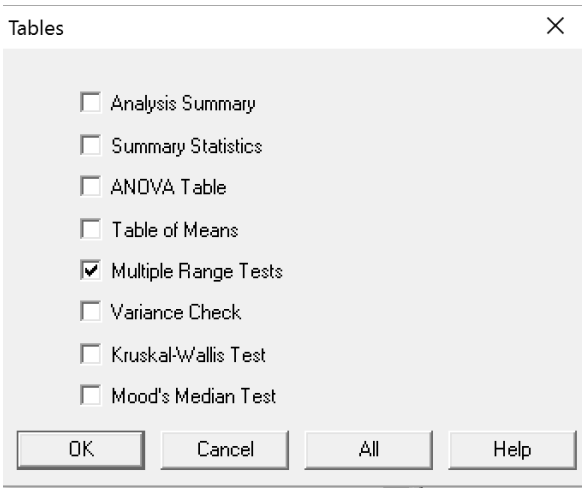
Sử dụng Statgraphics để xếp nhóm các xuất xứ có sự khác biệt:

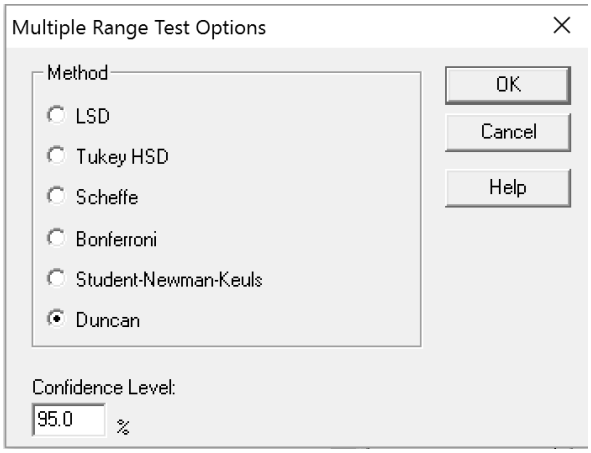
Trong nút Tables chọn Multiple Range Test.

Trong cửa sổ kết quả, kích chuột phải và chọn Pane Options; sau đó chọn tiêu chuẩn xếp nhóm khác biệt: Duncan, hoặc LSD,…

Đồng thời có thể chọn độ tin cậy: Confidenca Level, thông thường là 95%.

Kết quả cho ra xếp nhóm sự khác biệt giữa các xuất xứ theo tiêu chuẩn Duncan ở Bảng 5.3.




**Bảng 5.3.** Xếp nhóm đồng nhất và có sự khác biệt rõ rệt giữa các nhóm theo tiêu chuẩn Duncan ở độ tin cậy 95% cho 7 xuất xứ thông caribeae

| Xuất xứ<br>Xuat xu | Lần lặp<br>Count | Trung bình<br>Mean | Các nhóm đồng nhất<br>Homogeneous Groups |
|---|---|---|---|
| 7 | 2 | 8.4 | X |
| 4 | 4 | 10.0 | XX |
| 3 | 4 | 10.1 | XX |
| 1 | 4 | 10.575 | X |
| 2 | 4 | 10.825 | X |
| 6 | 4 | 12.0 | XX |
| 5 | 2 | 13.55 | X |

Trong bảng kết quả xếp nhóm của Duncan, các xuất xứ không có sự khác biệt có ý nghĩa về DBH trung bình (đồng nhất) sẽ được đánh dấu X và được xếp cùng trong một hàng theo chiều thẳng đứng trong cột "Các nhóm đồng nhất – Homogenous Groups"; đồng thời các xuất xứ được sắp xếp theo giá trị trung bình DBH từ thấp đến cao theo thứ tự từ trên xuống, xuất xứ có giá trị

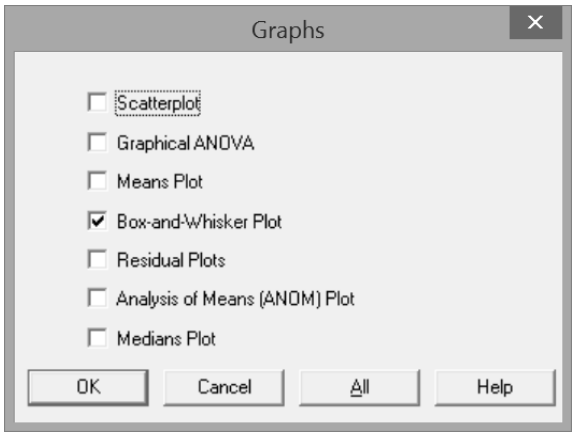
trung bình cao nhất nằm ở cuối cùng. Kết quả này cho thấy, các xuất xứ được xếp vào ba nhóm có sự khác biệt nhau: Nhóm có DBH thấp nhất gồm các xuất xứ: 7, 4 và 3; nhóm có DBH trung bình gồm các xuất xứ: 4, 3, 1, 2 và 6; và nhóm có DBH cao nhất gồm các xuất xứ: 6 và 5. Như vậy, trong ví dụ này giữa xuất xứ 5 và 6 chưa có sự khác biệt về sinh trưởng đường kính và đây là hai xuất xứ tốt nhất cần được lựa chọn cho trồng rừng.

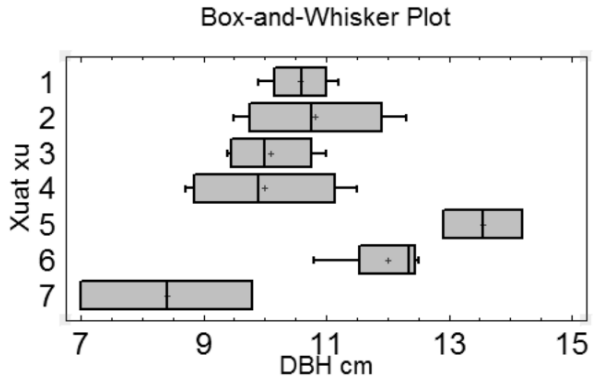
Ngoài ra để minh họa và so sánh sinh trưởng DBH ở các xuất xứ khác nhau, nên sử dụng sơ đồ hộp cùng với biến động của các trung bình (Box-and-Whisker Plot).

Vẽ đồ thị Box-and-Whisker Plot trong Statgraphics:

Trong nút đồ thị (Graphs), chọn Box-and-Whisker Plot.

Kết quả thể hiện ở Hình 5.2.



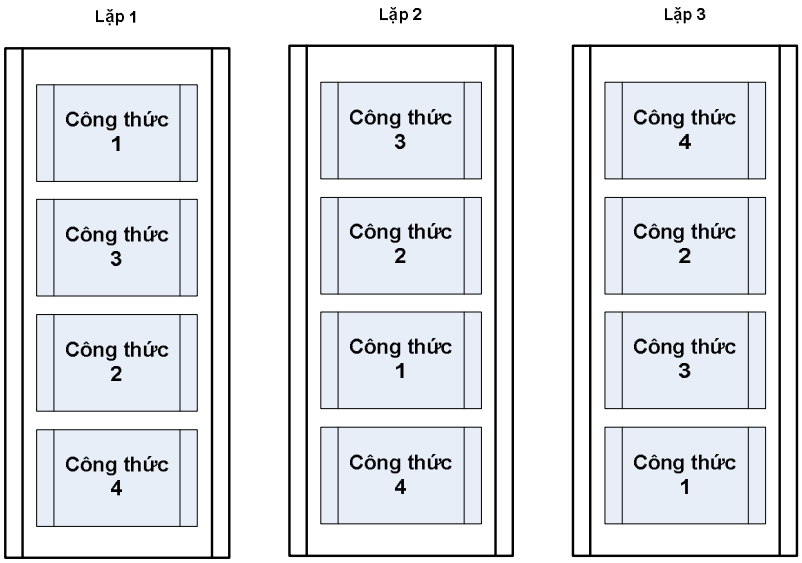




**Hình 5.2.** Sơ đồ Box-and-Whisker biến động trung bình DBH của bảy xuất xứ thông Caribeae khác nhau

Sơ đồ Box and Whisker chỉ ra vị trí giá trị trung bình ở dấu + và biến động của nó ở độ tin cậy 95% trong phạm vi box. Kết quả này cho thấy xuất xứ mã số 5 có trung bình cao nhất và giá trị ước lượng thấp nhất của nó cũng cao hơn giá trị cao nhất của xuất xứ 6 liền kề. Vì vậy, xuất xứ 5 cần được xem là xuất xứ tốt nhất.

## 5.2 Phân tích phương sai một nhân tố với bố trí thí nghiệm theo khối ngẫu nhiên đầy đủ (Randomized Complete Blocks) (RCB) hoặc phân tích phương sai hai nhân tố một lần lặp lại

Kiểu bố trí thí nghiệm RCB thường được sử dụng, nhân tố A chia làm a cấp và nhân tố B (hoặc Lần lặp) được chia b cấp (khối), mỗi tổ hợp 2 nhân tố chỉ có 1 lần lặp (1 ô thí nghiệm) (Hình 5.3).




**Hình 5.3.** Sơ đồ bố trí thí nghiệm để phân tích phương sai một nhân tố với bố trí thí nghiệm theo khối ngẫu nhiên đầy đủ

Ở Hình 5.3 mỗi công thức có số lần lặp bằng nhau. Diện tích mỗi lần lặp được chọn trên lập địa đồng nhất các yếu tố không thí nghiệm. Các công thức thí nghiệm của nhân tố nghiên cứu được bốc thăm để bố trí ngẫu nhiên trong mỗi lần lặp. Mỗi công thức ở một lần lặp là một ô thí nghiệm, ô có mẫu (số cây) đủ lớn để bảo đảm dữ liệu của ô đạt chuẩn. Lần lặp có thể là nhân tố thứ hai. Lúc này nó trở thành bố trí thí nghiệm hai nhân tố một lần lặp.

Ví dụ: Để khảo nghiệm 16 xuất xứ *Pinus kesiya* tại Trạm Thực nghiệm Lâm sinh Lang Hanh-Lâm Đồng: 16 xuất xứ thông ba lá *P.kesiya* đã được trồng khảo nghiệm tại Trạm Thực nghiệm Lang Hanh năm 1991. Việc bố trí thí nghiệm đã được tiến hành theo khối ngẫu nhiên đầy đủ RCB (Randomized Complete Blocks), bao gồm 16 công thức chỉ thị 16 xuất xứ và mỗi xuất xứ được trồng lặp lại 4 lần như nhau. Như vậy, tổng cộng có 64 ô thí nghiệm.

16 xuất xứ *P.kesiya* được mã số như sau:1: Bengliet; 2: Faplac; 3: Xuân Thọ; 4: Thác Prenn; 5: Lang Hanh; 6: Nong Kiating; 7: Doisupthep; 8: Doiinthranon; 9: Phu Kradung; 10: Nam nouv; 11: Cotomines; 12: Simao; 13: Watchan;14: Zo khua; 15: Aung ban; 16: Jingdury.

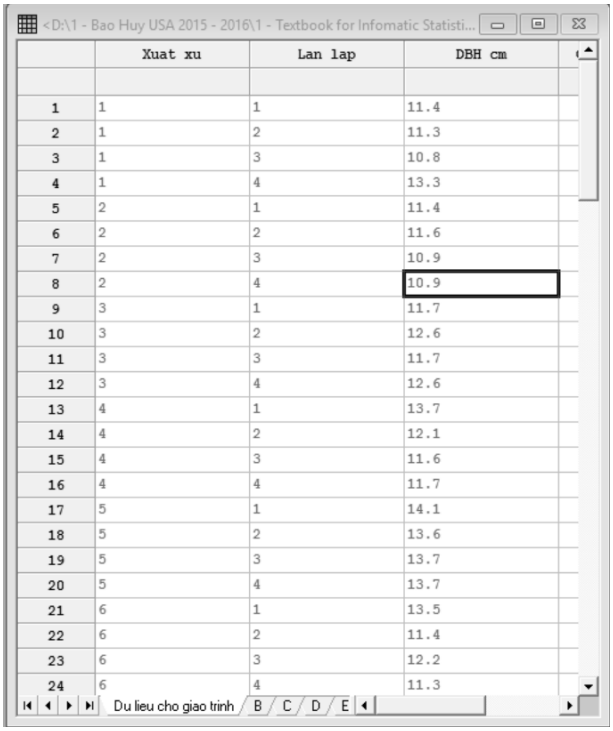
Mỗi công thức ứng với một lần lặp được trồng 25 cây, với cự ly 3x2m, tổng diện tích bố trí thí nghiệm là 1,5ha. Các điều kiện khí hậu, địa hình, chăm sóc... đều được đồng nhất, nhân tố thay đổi để khảo sát chỉ còn lại là các xuất xứ khác nhau. Tại thời điểm điều tra (1996), cây trồng trong các ô thí nghiệm có tuổi là 5. Tiến hành đo, đếm toàn diện các chỉ tiêu DBH (cm), H (m), đường kính tán (Dt, m), phẩm chất, tỉa cành, hình thân. Sử dụng 2 chỉ tiêu DBH và H để đánh giá sinh trưởng của các xuất xứ thử nghiệm.

Dữ liệu trung bình DBH (cm) của các ô thí nghiệm theo 16 xuất xứ với 4 lần lặp ở trong Dữ liệu 7 phần Phụ lục.

Sử dụng ANOVA để so sánh sự khác nhau về sinh trưởng DBH của 16 xuất xứ. Trong ba điều kiện để áp dụng ANOVA, thì ở đây chấp nhận 2 điều kiện: Mẫu chuẩn (vì mỗi ô thí nghiệm có số cây trồng đủ lớn, gần xấp xỉ 30 cây) và mỗi công thức được lặp lại ít nhất 2 lần. Lúc này cần kiểm tra sự bằng nhau của các phương sai ở các công thức thí nghiệm (xuất xứ).

Sử dụng ANOVA trong chương trình Statgraphics. Theo trình tự sau đây:

Nhập dữ liệu từ Excel vào Statgraphics: Gồm 3 trường dữ liệu: Nhân tố xuất xứ (mã số xuất xứ), mã số lần lặp và trường dữ liệu quan sát ứng với từng xuất xứ là lần lặp lại (DBH, cm)




| | Xuat xu | Lan lap | DBH cm |
|---|---|---|---|
| 1 | 1 | 1 | 11.4 |
| 2 | 1 | 2 | 11.3 |
| 3 | 1 | 3 | 10.8 |
| 4 | 1 | 4 | 13.3 |
| 5 | 2 | 1 | 11.4 |
| 6 | 2 | 2 | 11.6 |
| 7 | 2 | 3 | 10.9 |
| 8 | 2 | 4 | 10.9 |
| 9 | 3 | 1 | 11.7 |
| 10 | 3 | 2 | 12.6 |
| 11 | 3 | 3 | 11.7 |
| 12 | 3 | 4 | 12.6 |
| 13 | 4 | 1 | 13.7 |
| 14 | 4 | 2 | 12.1 |
| 15 | 4 | 3 | 11.6 |
| 16 | 4 | 4 | 11.7 |
| 17 | 5 | 1 | 14.1 |
| 18 | 5 | 2 | 13.6 |
| 19 | 5 | 3 | 13.7 |
| 20 | 5 | 4 | 13.7 |
| 21 | 6 | 1 | 13.5 |
| 22 | 6 | 2 | 11.4 |
| 23 | 6 | 3 | 12.2 |
| 24 | 6 | 4 | 11.3 |

Du lieu cho giao trinh / B / C / D / E

Kiểm tra sự bằng nhau về phương sai của các công thức thí nghiệm:

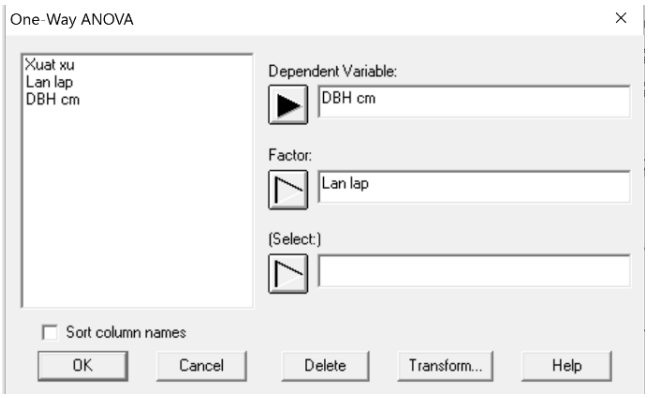
Tiến hành ANOVA một nhân tố trong Statgraphics: Analyze/ Variable Data/One Variable Analysis.

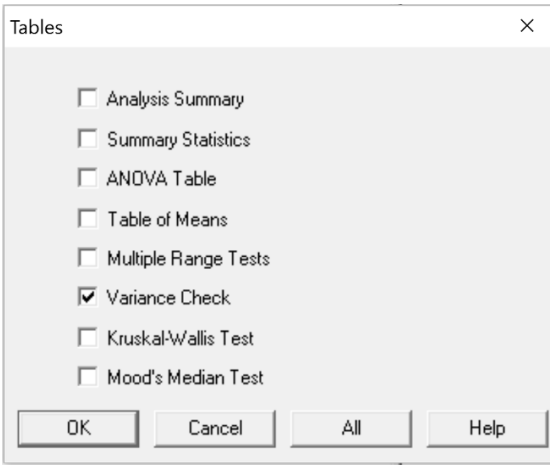
Trong hộp thoại chọn:

- Dependent Variable: Biến số khảo sát (DBH)

- Factor: Lần lặp

Trong nút Tables, chọn Variance Check để kiểm tra sự bằng nhau của các phương sai ở các công thức thí nghiệm theo tiêu chuẩn Levene. Kết quả:




Variance Check

| | Test | P-Value |
|---|---|---|
| Levene's | 0.731076 | 0.537507 |

Kết quả kiểm tra theo tiêu chuẩn Levene có P-Value = 0.537 > 0.05, có nghĩa là giả thuyết $H_0$ về sự bằng nhau giữa các phương sai ở các công thức thí nghiệm (theo xuất xử ở các lần lặp lại khác nhau) được chấp nhận. Như vậy, dữ liệu từ kết quả bố trí thí nghiệm này bảo đảm yêu cầu để áp dụng ANOVA nhằm đánh giá sự khác nhau về sinh trưởng của các xuất xứ thông ba lá.

Sử dụng Statgraphics với ANOVA nhiều nhân tố (ở đây có thể hiểu là một nhân tố là xuất xứ và nhân tố thứ hai là lần lặp lại). Nếu thay lần lặp lại là một nhân tố khác như loại đất hoặc bón phân, hoặc mật độ... thì lúc này là ANOVA hai nhân tố không có lần lặp (lặp lại chỉ một lần cho mỗi tổ hợp hai nhân tố).

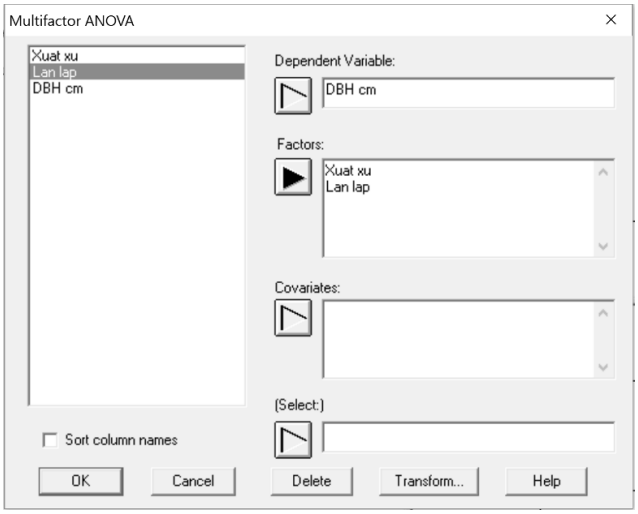
Sử dụng phân tích ANOVA nhiều nhân tố trong Statgraphics: Analysis of Variance/Multifactor ANOVA

Trong hộp thoại chọn:

- Dependent Variable: Biến giá trị đo tính, quan sát: DBH, cm để so sánh sinh trưởng thông 3 lá ở các xuất xứ khác nhau

- Factors: Chọn hai nhân tố: Xuất xứ và Lặp lại

Kết quả ANOVA thể hiện ở Bảng 5.4.




**Bảng 5.4.** Kết quả ANOVA hai nhân tố một lần lặp

Analysis of Variance for DBH cm - Type III Sums of Squares

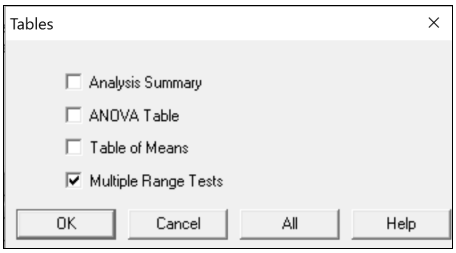
| Source | Sum of Squares | Df | Mean Square | F-Ratio | P-Value |
|---|---|---|---|---|---|
| MAIN EFFECTS | | | | | |
| A:Xuat xu | 81.775 | 15 | 5.45167 | 7.72 | 0.0000 |
| B:Lan lap | 3.47625 | 3 | 1.15875 | 1.64 | 0.1931 |
| RESIDUAL | 31.7587 | 45 | 0.70575 | | |
| TOTAL (CORRECTED) | 117.01 | 63 | | | |

Kết quả ở bảng ANOVA trên cho thấy, các lần lặp lại không có sự sai khác về sinh trưởng DBH, với kết quả ANOVA có P-Value = 0.193 > 0.05 (Chấp nhận giả thuyết $H_0$ về sự bằng nhau sinh trưởng DBH ở 4 lần lặp lại). Trong đó P-Value = 0.000 < 0.05 đối với các xuất xứ khác nhau, có nghĩa giả thuyết $H_0$ bị bác bỏ và chấp nhận giả thuyết $H_1$: Có sự sai khác về sinh trưởng DBH ở các xuất xứ (ít nhất là có sự khác biệt ở hai xuất xứ).

Lúc này chỉ cần xếp nhóm khác biệt giữa các xuất xứ khác nhau (Multiple Range Test), thực hiện trong Statgraphics như sau:

Xếp nhóm các xuất xứ có sự khác biệt theo trong Statgraphics:

Trong nút Tables chọn Multiple Range Test.

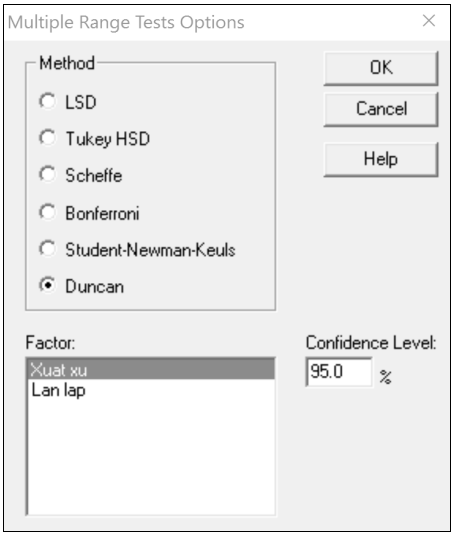



Trong cửa sổ kết quả, kích chuột phải và chọn Pane Options; sau đó chọn tiêu chuẩn xếp nhóm khác biệt: Duncan, hoặc LSD,…

Chọn nhân tố (Factor) để xếp nhóm: Ở đây lần lặp không có sự khác biệt, do đó, chỉ xếp nhóm các xuất xứ đồng nhất. Chọn: Xuat xu

Đồng thời có thể chọn độ tin cậy: Confidenca Level, thông thường là 95%.

Kết quả cho ra xếp nhóm sự khác biệt giữa các xuất xứ theo tiêu chuẩn Duncan ở Bảng 5.5.




**Bảng 5.5.** Xếp nhóm các xuất xứ đồng nhất về sinh trưởng DBH (giữa các nhóm có sự khác biệt rõ rệt) theo tiêu chuẩn Duncan ở độ tin cậy 95% cho 16 xuất xứ thông 3 lá khảo nghiệm

| *Xuat xu* | *Count* | *LS Mean* | *LS Sigma* | *Homogeneous Groups* |
|---|---|---|---|---|
| 14 | 4 | 9.35 | 0.420045 | X |
| 15 | 4 | 9.975 | 0.420045 | XX |
| 16 | 4 | 10.975 | 0.420045 | XX |
| 2 | 4 | 11.2 | 0.420045 | XXX |
| 12 | 4 | 11.55 | 0.420045 | XX |
| 1 | 4 | 11.7 | 0.420045 | XX |
| 10 | 4 | 11.75 | 0.420045 | XX |
| 6 | 4 | 12.1 | 0.420045 | XX |
| 3 | 4 | 12.15 | 0.420045 | XX |

| *Xuat xu* | *Count* | *LS Mean* | *LS Sigma* | *Homogeneous Groups* |
|---|---|---|---|---|
| 13 | 4 | 12.225 | 0.420045 | XX |
| 4 | 4 | 12.275 | 0.420045 | XX |
| 11 | 4 | 12.35 | 0.420045 | XX |
| 9 | 4 | 12.4 | 0.420045 | X |
| 7 | 4 | 12.525 | 0.420045 | X |
| 5 | 4 | 13.775 | 0.420045 | X |
| 8 | 4 | 13.9 | 0.420045 | X |

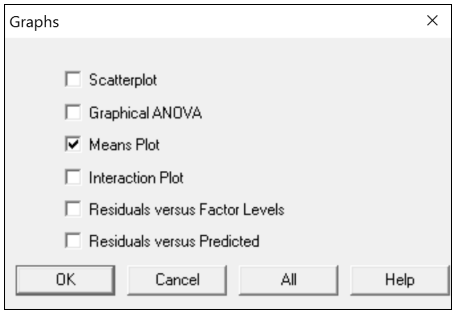
Trong bảng trên, kết quả xếp nhóm của Duncan, các xuất xứ không có sự khác biệt, có ý nghĩa về DBH trung bình (đồng nhất) sẽ được đánh dấu X và được xếp cùng một cột "Các nhóm đồng nhất – Homogenous Groups"; đồng thời các xuất xứ được sắp xếp theo giá trị trung bình DBH từ thấp đến cao theo thứ tự từ trên xuống, xuất xứ có giá trị trung bình cao nhất nằm ở cuối cùng. Kết quả này cho thấy, các xuất xứ được xếp vào bốn nhóm có sự khác biệt nhau: Nhóm có DBH cao nhất gồm hai xuất xứ: 5 và 8. Như vậy trong nghiên cứu này giữa xuất xứ 5 và 8 chưa có sự khác biệt về sinh trưởng đường kính và đây là hai xuất xứ tốt nhất cần được lựa chọn cho trồng rừng thông ba lá ở Lâm Đồng.

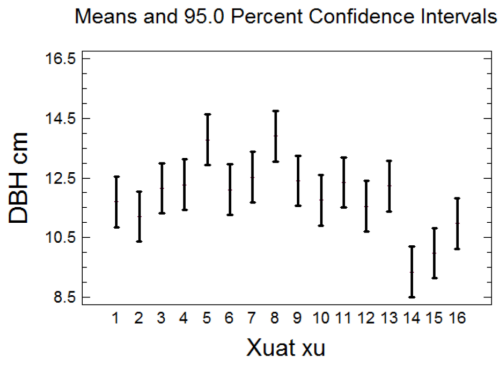
Ngoài ra để minh họa và so sánh sinh trưởng DBH của 16 xuất xứ khác nhau, nên sử dụng sơ đồ biến động trung bình (MeanPlot) trong Statgraphics.

Vẽ đồ thị biến động trung bình trong Statgraphics:

Trong nút đồ thị (Graphs), chọn Mean Plot.

Kết quả thể hiện ở Hình 5.4.




**Hình 5.4.** Đồ thị biến động DBH trung bình của 16 xuất xứ thông ba lá thí nghiệm ở độ tin cậy 95%

Kết quả đồ thị khẳng định xuất xứ Lang Hanh (5) và xuất xứ Doiinthranon (8) là các xuất xứ tốt nhất và sai khác rõ rệt với các xuất xứ khác.

## 5.3 Phân tích phương sai nhiều nhân tố với m lần lặp

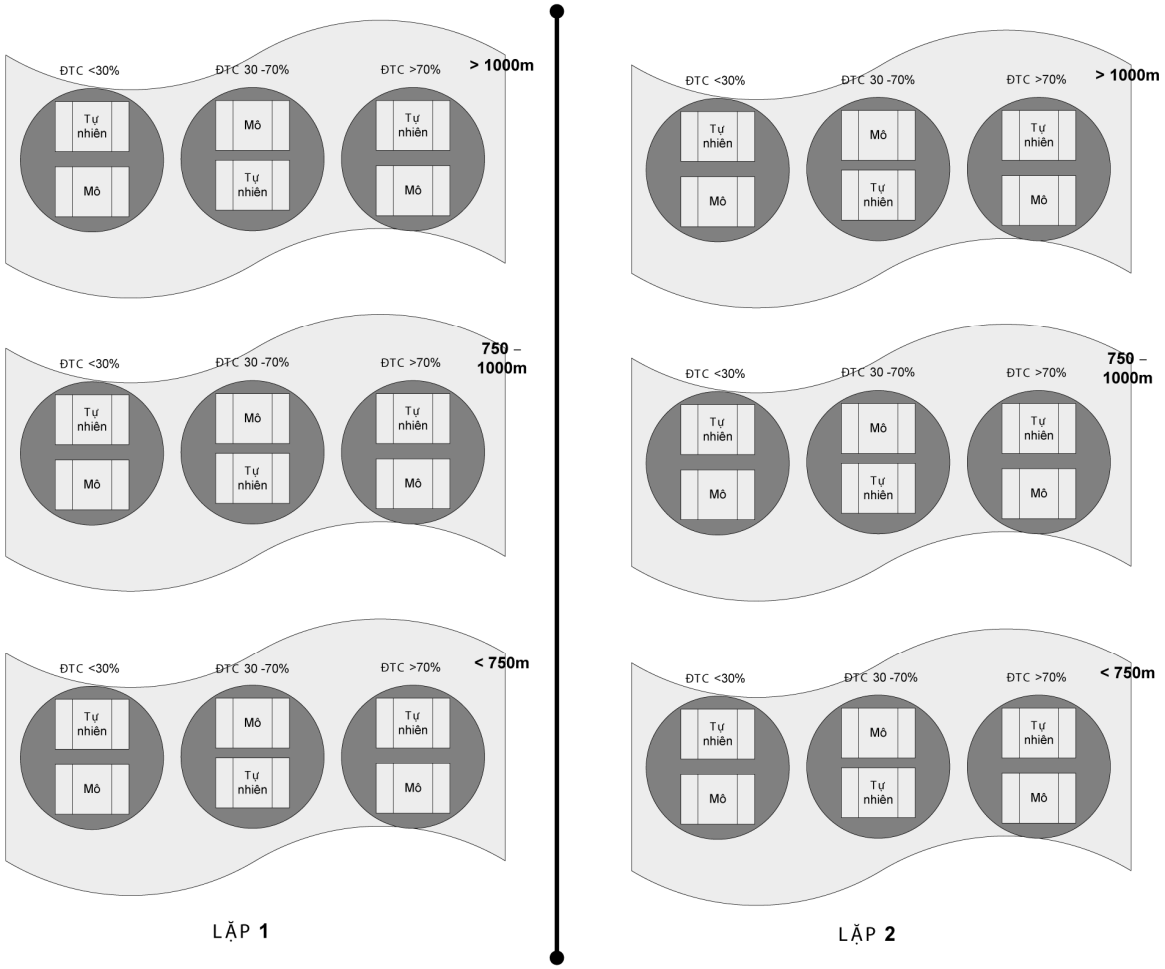
Trong thực tế, có những trường hợp cần nghiên cứu ảnh hưởng của nhiều nhân tố tổng hợp đến giá trị quan sát, khảo sát. Ví dụ, cần nghiên cứu ảnh hưởng của n nhân tố và áp dụng ANOVA với m lần lặp. Lúc này nhân tố A có a công thức, nhân tố B có b công thức và cuối cùng nhân tố thứ n là Z có z công thức, tạo thành a*b*…*z tổ hợp của n nhân tố; và mỗi tổ hợp các nhân tố cần được lặp lại m lần với (với $m \geq 2$ lần để có thể áp dụng ANOVA).

Về nghiên cứu liên quan đến sinh thái học và nhân tác trong lâm nghiệp như nghiên cứu sinh trưởng cây con trong vườn ươm, sinh trưởng cây rừng trong trồng rừng, mật độ phân bố, tái sinh loài,… thì chỉ tiêu khảo sát của các nghiên cứu này bị tác động không chỉ từng nhân tố sinh thái hoặc một nhân tác riêng lẻ. Do đó, đơn giản hóa các nhân tố ảnh hưởng để nghiên cứu thường cho ra kết quả phiến diện và ít có ý nghĩa trong thực tiễn. Ví dụ, nghiên cứu để nâng cao năng suất cây rừng, chỉ nghiên cứu một nhân tố là phân bón, các nhân tố có khả năng ảnh hưởng khác được đồng nhất như mật độ trồng, đất đai, địa hình, khí hậu... Do đó, cho dù kết quả có thể chỉ ra mức và loại phân bón nào là phù hợp để cho năng suất cao nhất, tuy nhiên có thể kết quả này sẽ không bằng việc thay đổi mật độ tối ưu. Có thể thay đổi mật độ sẽ cho ra năng suất cao hơn nhiều so với bón phân, trong khi đó nhân tố mật độ bị bỏ qua trong thí nghiệm này. Với ví dụ này, cần thêm nhân tố mật độ trong thử nghiệm tạo thành bố trí thí nghiệm hai nhân tố phân bón và mật độ và nghiên cứu ảnh hưởng tổng hợp của chúng đến năng suất cây rừng.

Còn nhiều ví dụ khác trong nghiên cứu ảnh hưởng của nhiều nhân tố như các nghiên cứu ảnh hưởng của các nhân tố độ tàn che, chế độ tưới, phân bón,... đến tỷ lệ sống, sinh trưởng cây con trong vườn ươm; hoặc ảnh hưởng của các nhân tố sinh thái như: đai cao, địa hình, đất đai đến mật độ phân bố, tái sinh của loài nghiên cứu. Vì vậy cần có tổ chức nghiên cứu, thực nghiệm đa nhân tố và áp dụng ANOVA để chỉ ra được sự ảnh hưởng tổng hợp và qua lại của nhiều nhân tố đến chỉ tiêu quan tâm.

Ví dụ một nghiên cứu của Nguyễn Thị Quỳnh (2016) về ảnh hưởng của các nhân tố sinh thái – nhân tác đến sinh trưởng - sinh khối của lan kim tuyến (*Anoectochilus formosanus* Hayata) trồng dưới tán rừng thường xanh ở Lâm Đồng. Lan kim tuyến là một loài lan quý hiếm, có trong sách đỏ Việt Nam và thế giới, ngoài hoa và lá đẹp, nó có giá trị kinh tế cao trên thị trường. Nghiên cứu này nhằm mục đích thử nghiệm để tìm kiếm khu vực có các nhân tố sinh thái thích hợp cho việc gây trồng loài lan quý hiếm này trong tự nhiên.

Bố trí thí nghiệm theo ba nhân tố: Nguồn gốc giống lan với 2 công thức: từ nuôi cấy mô hoặc từ tự nhiên; độ cao so với mặt biển có ba cấp: < 750m, 750 – 1000 m và > 1000m; và nhân tố độ tàn che của rừng với ba mức: < 30%, 30 – 70% và > 70%. Thí nghiệm được lặp lại 2 lần. Như vậy, có 2 nguồn gốc * 3 độ cao * 3 độ tàn che = 18 tổ hợp 3 nhân tố và có 18 tổ hợp * 2 lần lặp lại = 36 ô thí nghiệm lan được trồng dưới tán rừng. Bố trí thí nghiệm theo 3 nhân tố 2 lần lặp trong thực tế được minh họa ở Hình 5.5.



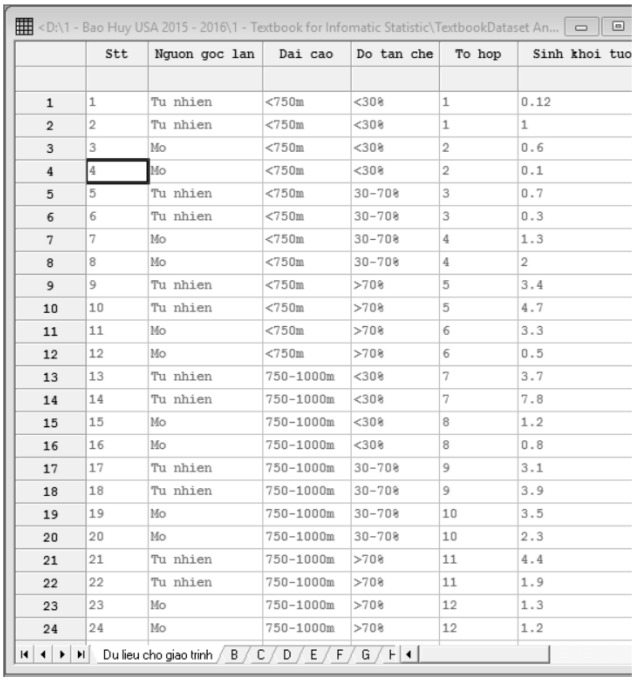

**Hình 5.5.** Sơ đồ bố trí thí nghiệm ba nhân tố: Nguồn gốc, độ cao, độ tàn che trong gây trồng lan kim tuyến với hai lần lặp dưới tán rừng thường xanh ở Lâm Đồng *(Nguyễn Thị Quỳnh, 2016)*

Mỗi ô thí nghiệm có diện tích 0.25m$^2$ (0.5 * 0.5m), mỗi ô trồng 35 cây lan kim tuyến, cây cách cây 5*5cm. Thời gian thử nghiệm là 6 tháng. Các chỉ tiêu theo dõi là số cây chết, số chồi mới, đường kính gốc, chiều cao theo từng tháng và chỉ tiêu quan trọng nhất là tổng sinh khối tươi (g) trên từng ô thí nghiệm ở lần đo cuối cùng. Dữ liệu sinh khối tươi (g) sau sáu tháng trồng ở 36 ô thử nghiệm theo 3 nhân tố 2 lần lặp (trình bày trong Dữ liệu 8 ở phần Phụ lục).

Sử dụng ANOVA đa nhân tố để so sánh sự khác nhau về sinh khối của lan kim tuyến theo 3 nhân tố nghiên cứu. Trong ba điều kiện để áp dụng ANOVA, thì ở đây chấp nhận 2 điều kiện: Mẫu chuẩn (vì mỗi ô thí nghiệm có số cây trồng đủ lớn, 35 cây) và mỗi công thức được lặp lại ít nhất 2 lần. Lúc này cần kiểm tra sự bằng nhau của các phương sai ở các công thức của từng nhân tố và tổ hợp các công thức.

Sử dụng ANOVA trong chương trình Statgraphics. Theo trình tự sau đây:

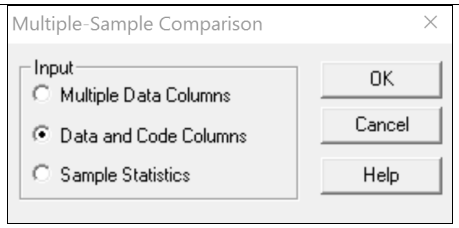
Nhập dữ liệu từ Excel vào Statgraphics: Gồm 5 trường dữ liệu: Nhân tố nguồn gốc lan (tự nhiên hoặc mô), nhân tố đai cao (3 cấp), nhân tố độ tàn che (3 cấp), tổ hợp (gồm 18 tổ hợp, mỗi tổ hợp lặp lại 2 lần) và trường dữ liệu quan sát, đánh giá là sinh khối tươi (g).



| | Stt | Nguon goc lan | Dai cao | Do tan che | To hop | Sinh khoi tuo |
|---|---|---|---|---|---|---|
| 1 | 1 | Tu nhien | <750m | <30% | 1 | 0.12 |
| 2 | 2 | Tu nhien | <750m | <30% | 1 | 1 |
| 3 | 3 | Mo | <750m | <30% | 2 | 0.6 |
| 4 | 4 | Mo | <750m | <30% | 2 | 0.1 |
| 5 | 5 | Tu nhien | <750m | 30-70% | 3 | 0.7 |
| 6 | 6 | Tu nhien | <750m | 30-70% | 3 | 0.3 |
| 7 | 7 | Mo | <750m | 30-70% | 4 | 1.3 |
| 8 | 8 | Mo | <750m | 30-70% | 4 | 2 |
| 9 | 9 | Tu nhien | <750m | >70% | 5 | 3.4 |
| 10 | 10 | Tu nhien | <750m | >70% | 5 | 4.7 |
| 11 | 11 | Mo | <750m | >70% | 6 | 3.3 |
| 12 | 12 | Mo | <750m | >70% | 6 | 0.5 |
| 13 | 13 | Tu nhien | 750-1000m | <30% | 7 | 3.7 |
| 14 | 14 | Tu nhien | 750-1000m | <30% | 7 | 7.8 |
| 15 | 15 | Mo | 750-1000m | <30% | 8 | 1.2 |
| 16 | 16 | Mo | 750-1000m | <30% | 8 | 0.8 |
| 17 | 17 | Tu nhien | 750-1000m | 30-70% | 9 | 3.1 |
| 18 | 18 | Tu nhien | 750-1000m | 30-70% | 9 | 3.9 |
| 19 | 19 | Mo | 750-1000m | 30-70% | 10 | 3.5 |
| 20 | 20 | Mo | 750-1000m | 30-70% | 10 | 2.3 |
| 21 | 21 | Tu nhien | 750-1000m | >70% | 11 | 4.4 |
| 22 | 22 | Tu nhien | 750-1000m | >70% | 11 | 1.9 |
| 23 | 23 | Mo | 750-1000m | >70% | 12 | 1.3 |
| 24 | 24 | Mo | 750-1000m | >70% | 12 | 1.2 |

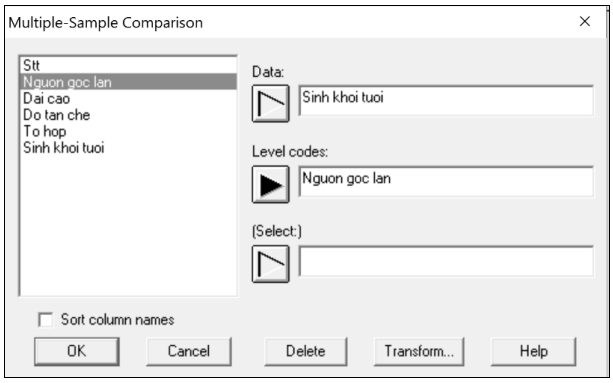
Kiểm tra lần lượt sự bằng nhau về phương sai của các công thức trong từng nhân tố và giữa các tổ hợp các nhân tố thí nghiệm:

Menu: Analyze/ Variable Data/ Multiple Sample Comparisions.




Chọn: Data and Code Columns (Dữ liệu và các nhân tố là code)

Trong hộp thoại lần lượt chọn:



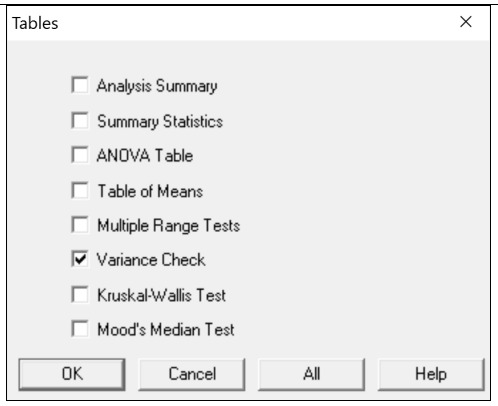

Data: Dữ liệu khảo sát "Sinh khoi tuoi".

Level Codes: Chọn nhân tố. Tức là các ô thí nghiệm của nhân tố cần kiểm tra sự bằng nhau các phương sai. Lần lượt chọn để kiểm tra sự bằng nhau về phương sai ở 3 nhân tố và cả tổ hợp.

Trong nút Tables, chọn Variance Check để kiểm tra sự bằng nhau của các phương sai theo từng nhân tố và các tổ hợp ba nhân tố.

Kết quả như sau theo tiêu chuẩn

Levene:




Nhân tố Nguồn gốc

| | Test | P-Value |
|---|---|---|
| Levene's | 6.02323 | 0.019395 |

Nhân tố Đai cao

| | Test | P-Value |
|---|---|---|
| Levene's | 2.14851 | 0.13269 |

Nhân tố Độ tàn che

| | Test | P-Value |
|---|---|---|
| Levene's | 0.027795 | 0.9726 |

Tổ hợp ba nhân tố

| | Test | P-Value |
|---|---|---|
| Levene's | 4.1859E31 | 0.0 |

Kết quả kiểm tra sự bằng nhau của các phương sai theo tiêu chuẩn Levene ở các công thức cho từng nhân tố và giữa các tổ hợp công thức ở trên cho thấy: giữa các công thức của hai nhân tố Đai cao và Độ tàn che, giá trị P-Value > 0.05 (chấp nhận giả thuyết $H_0$), có nghĩa dữ liệu của hai nhân tố này có phương sai là bằng nhau; trong khi đó nhân tố Nguồn gốc và Tổ hợp có P-Value < 0.05 (Bác bỏ giả thuyết $H_0$), hay nói khác giữa các công thức của nhân tố Nguồn gốc và Tổ hợp có sự sai khác về phương sai. Như vậy, nếu phân tích phương sai cả ba nhân tố có khả năng sẽ cho kết quả kém tin cậy, vì có một nhân tố với các công thức thí nghiệm cho phương sai không bằng nhau. Trường hợp này, tốt nhất là áp dụng ANOVA cho hai nhân tố có phương sai bằng nhau là Độ cao và Độ tàn che. Riêng nhân tố Nguồn gốc có phương sai không bằng nhau giữa các công thức, cần áp dụng tiêu chuẩn phi tham số (sẽ giới thiệu ở phần sau của sách này) (không yêu cầu dữ liệu có phân bố chuẩn và phương sai bằng nhau).

Sử dụng Statgraphics với ANOVA nhiều nhân tố (trường hợp này là hai nhân tố) để so sánh sinh khối tươi của lan kim tuyến.

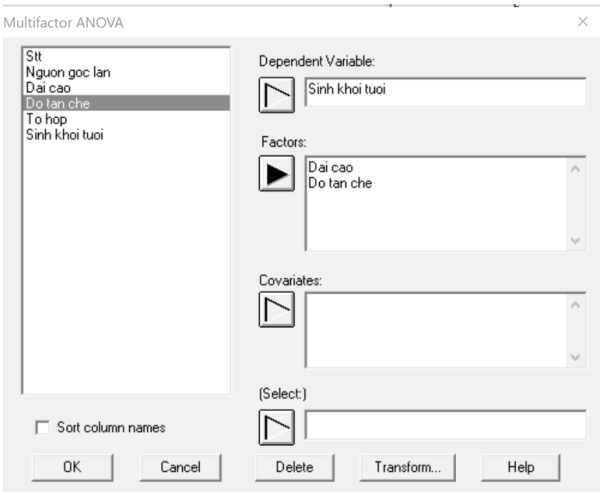
Trong Statgraphics: Analysis of Variance/Multifactor ANOVA

Trong hộp thoại chọn:

- Dependent Variable: Biến giá trị đo tính, quan sát: Sinh khoi tuoi

- Factors: Chọn hai nhân tố: Dai cao và Do tan che

Kết quả ANOVA thể hiện ở Bảng 5.6.



**Bảng 5.6.** Kết quả phân tích ANOVA nhiều nhân tố

Analysis of Variance for Sinh khoi tuoi - Type III Sums of Squares

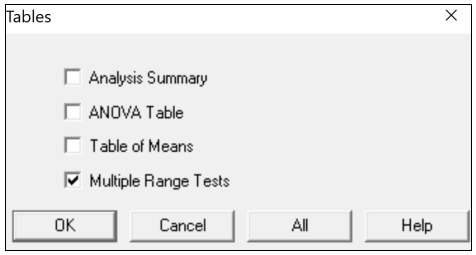
| Source | Sum of Squares | Df | Mean Square | F-Ratio | P-Value |
|---|---|---|---|---|---|
| MAIN EFFECTS | | | | | |
| A:Dai cao | 32.2319 | 2 | 16.116 | 7.16 | 0.0028 |
| B:Do tan che | 5.31794 | 2 | 2.65897 | 1.18 | 0.3202 |
| RESIDUAL | 69.7612 | 31 | 2.25036 | | |
| TOTAL (CORRECTED) | 107.311 | 35 | | | |

Kết quả ở bảng ANOVA trên cho thấy, nhân tố "độ tàn che" chưa có sự sai khác về sinh khối lan, do có P-Value = 0.3202 > 0.05 (chấp nhận giả thuyết $H_0$ về sự bằng nhau của sinh khối ở 3 cấp độ tàn che nghiên cứu). Trong đó P-Value = 0.0028 < 0.05 đối với nhân tố "đai cao", có nghĩa giả thuyết $H_0$ bị bác bỏ và chấp nhận giả thuyết $H_1$: Có sự sai khác về sinh khối lan ở các đai cao khác nhau (ít nhất là có sự khác biệt ở hai đai cao).

Lúc này chỉ cần xếp nhóm khác biệt sinh khối tươi của lan giữa các đai cao khác nhau (Multiple Range Test), thực hiện trong Statgraphics như sau:

Xếp nhóm các đai cao có sự khác biệt về sinh khối lan trong Statgraphics:

Trong nút Tables chọn Multiple Range Test.

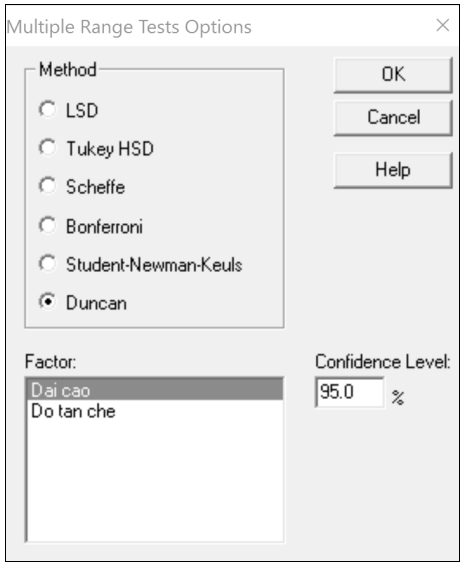



Trong cửa sổ kết quả, kích chuột phải và chọn Pane Options; sau đó chọn tiêu chuẩn xếp nhóm khác biệt: Duncan, hoặc LSD,…

Chọn nhân tố (Factor) để xếp nhóm: Ở đây Độ tàn che không cho sự khác biệt, do đó, chỉ xếp nhóm các đai cao đồng nhất về sinh khối với nhau. Chọn: Dai cao.

Đồng thời có thể chọn độ tin cậy: Confidenca Level, thông thường là 95%.

Kết quả cho ra xếp nhóm sự khác biệt giữa các xuất xứ theo tiêu chuẩn Duncan ở Bảng 5.7.



**Bảng 5.7.** Xếp nhóm các đai cao đồng nhất về sinh khối tươi của lan kim tuyến (giữa các nhóm có sự khác biệt rõ rệt) theo tiêu chuẩn Duncan ở độ tin cậy 95% cho 3 đai cao thử nghiệm

| Dai cao | Count | LS Mean | LS Sigma | Homogeneous Groups |
|---|---|---|---|---|
| > 1000m | 12 | 0.629167 | 0.433048 | X |
| < 750m | 12 | 1.50167 | 0.433048 | X |
| 750-1000m | 12 | 2.925 | 0.433048 | X |

Trong bảng trên, kết quả xếp nhóm của Duncan, cho ra hai nhóm có sự khác biệt về sinh khối tươi của lan: các đai cao không có sự khác biệt có ý nghĩa về sinh khối trung bình (đồng nhất) sẽ được đánh dấu X và được xếp cùng một cột "Các nhóm đồng nhất – Homogenous Groups"; đồng thời các đai cao được sắp xếp theo giá trị trung bình sinh khối từ thấp đến cao theo thứ tự từ trên xuống, đai cao có giá trị trung bình cao nhất nằm ở cuối cùng. Kết quả này cho thấy, các đai cao được xếp vào hai nhóm có sự khác biệt nhau: Nhóm có sinh khối cao nhất ở đai cao 700 – 1000m; nhóm có sinh khối thấp hơn rõ rệt bao gồm hai đai cao < 750m và > 1000m.

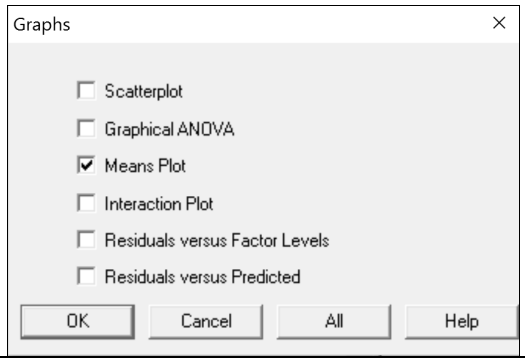
Với kết quả nghiên cứu này, thì đai cao 750 – 1000m là tối ưu cho việc gây trồng lan kim tuyến.

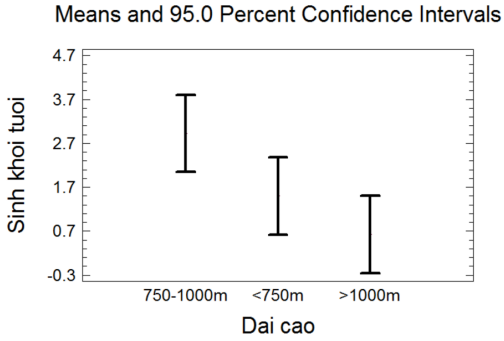
Ngoài ra, để minh họa và so sánh sinh trưởng DBH của 16 xuất xứ khác nhau, nên sử dụng sơ đồ biến động trung bình (MeanPlot) trong Statgraphics.

Vẽ đồ thị biến động trung bình trong Statgraphics:

Trong nút đồ thị (Graphs), chọn Mean Plot.

Kết quả thể hiện ở Hình 5.6.




**Hình 5.6.** Đồ thị trung bình và biến động sinh khối tươi của lan kim tuyến ở ba đai cao thí nghiệm với độ tin cậy 95%

Kết quả đồ thị khẳng định đai cao 750 – 1000m cho khối lượng sinh khối tối ưu.

Chương 6

# TIN HỌC TRONG ỨNG DỤNG TIÊU CHUẨN PHI THAM SỐ ĐỂ SO SÁNH CÁC MẪU QUAN SÁT ĐỘC LẬP HOẶC CÓ LIÊN HỆ

Trong sử dụng tiêu chuẩn tham số như tiêu chuẩn t hoặc phân tích phương sai (ANOVA) để so sánh các mẫu, các nhân tố ảnh hưởng đến chỉ tiêu quan sát, thì yêu cầu các mẫu, công thức đều phải thỏa mãn hai điều kiện là tuân theo phân bố chuẩn và phương sai bằng nhau. Trong thực tế các điều kiện rút mẫu, bố trí thí nghiệm có thể không bảo đảm yêu cầu của luật chuẩn, hoặc dung lượng mẫu nhỏ và phương sai có thể không bằng nhau. Trường hợp như vậy, không thể áp dụng các tiêu chuẩn thống kê tham số t và ANOVA và như vậy các tiêu chuẩn thống kê phi tham số cần được sử dụng. Hay nói khác, phương pháp so sánh nếu chưa rõ ràng luật phân bố với những tham số của nó được gọi là phương pháp phi tham số (Nguyễn Hải Tuất, Vũ Tiến Hinh, Ngô Kim Khôi, 2006; Ngô Kim Khôi, 1998; Ngô Kim Khôi và cộng sự, 2002).

## 6.1 Tiêu chuẩn phi tham số để so sánh các mẫu độc lập

Với các điều tra, thiết kế thí nghiệm với các mẫu, công thức độc lập trong một nhân tố có dữ liệu không theo phân bố chuẩn, hoặc dung lượng mẫu nhỏ, hoặc phương sai không bằng nhau; kiểm tra phi tham số của Kruskal-Wallis và Friedman được sử dụng hơn là ANOVA một nhân tố (Larson, 2008).

Khi sử dụng tiêu chuẩn phi tham số Kruskal Wallis và Friedman, tất cả dữ liệu quan sát của tất cả các mẫu được kết hợp chung trong một cột và được xếp hạng từ nhỏ đến lớn (thứ tự), sau đÓ tính giá trị trung bình thứ hạng cho từng mẫu (trung vị Median), công thức và sau cùng đem so sánh trung bình các thứ hạng của các mẫu, công thức so sánh với nhau.

Giả thuyết $H_0$: $x_1 = x_2 = \ldots. x_n$ hay giả thuyết $H_1$: $x_i \neq x_j$; với $x_{i,j}$ là các trung vị của mẫu i hoặc j trong n mẫu quan sát. Nếu P-value < 0.05 thì kết luận có sự khác biệt có ý nghĩa giữa các trung vị (ít nhất có hai trung vị khác biệt nhau) ở mức tin cậy 95%; hay nói khác giả thuyết bác bỏ giả thuyết và $H_0$ và chấp nhận giả thuyết $H_1$.

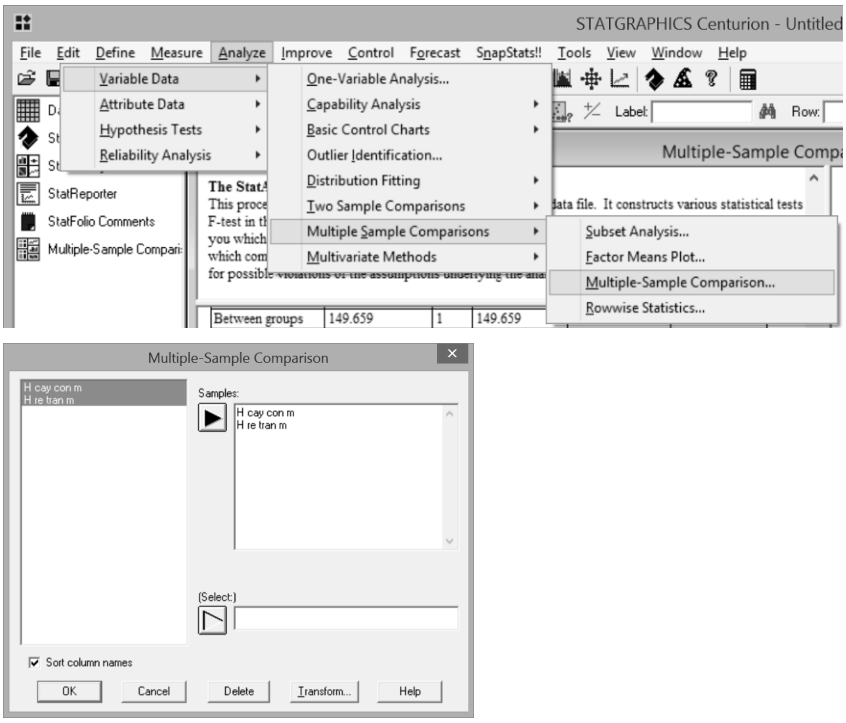
Tiêu chuẩn phi tham số Kruskal Wallis và Friedman có thể dùng để kiểm tra hai đến nhiều mẫu, công thức độc lập, và thay thế cho tiêu chuẩn t hoặc ANOVA một nhân tố.

Ví dụ, trong trường hợp so sánh hai mẫu độc lập theo hai phương pháp trồng thông 3 lá từ cây con hoặc rễ trần, với số liệu quan sát khá lớn (> 90 cây) nhưng cả hai mẫu đều chưa đạt chuẩn (Dữ liệu 4 trong Phụ lục). Do đó, nếu áp dụng t để so sánh sẽ chưa đủ độ tin cậy. Trong trường hợp này nên sử dụng tiêu chuẩn phi tham số Kruskal Wallis và Friedman để so sánh vì nó loại trừ được yêu

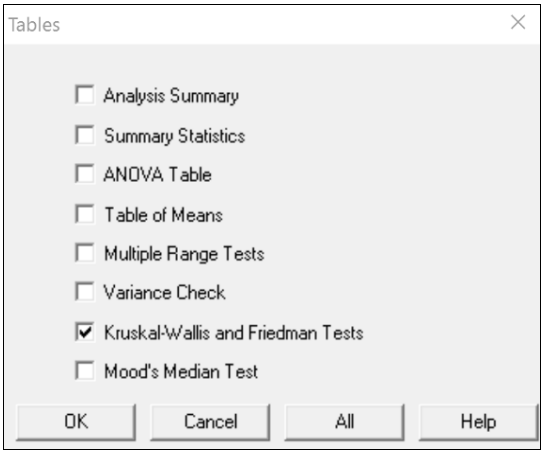
cầu hai mẫu có phân bố chuẩn.

Sử dụng phần mềm Statgraphics để kiểm tra thống kê theo Kruskal Wallis và Friedman với ví dụ trên theo "Dữ liệu 4" như sau:

Trong Statgraphics: Analysis/Variable Data/ Multiple Sample Comparisions/ Multiple-Sample Comnparirion. Trong hộp thoại chọn các mẫu so sánh.




Trong hộp thoại "Tables" chọn Kruskal-Wallis and Friedman Test.




**Bảng 6.1.** Kết quả kiểm tra sự sai khác hai trung bình thứ hạng của hai thí nghiệm trồng thông 3 lá bằng cây con và rễ trần theo Kruskal-Wasllis và Friedman

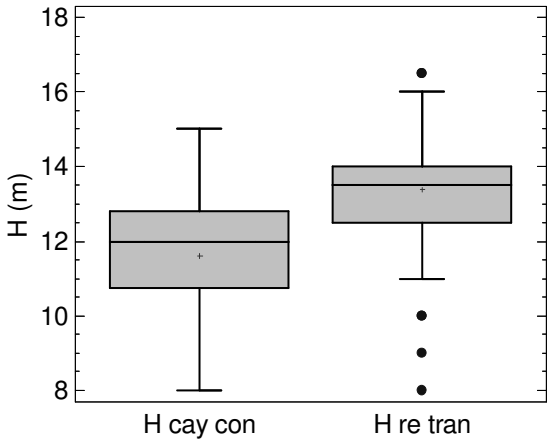
Kruskal-Wallis Test

| | Sample Size | Average Rank |
|---|---|---|
| H cay con | 92 | 63.9402 |
| H re tran | 93 | 121.747 |

Test statistic = 54.3713 P-Value = 0.0

Kết quả ở Bảng 6.1 kiểm tra theo Kruskal-Wallis và Friedman cho P-value = 0.0 < 0.05, có nghĩa là trung bình thứ hạng của hai mẫu trồng theo hai phương pháp khác nhau là có sự sai khác, có ý nghĩa ở độ tin cậy 95%. Trong đó trung bình thứ hạng (Average Rank) về chiều cao H của cây trồng bằng rễ trần cao hơn trồng bằng cây con; do vậy trong trường hợp này nên áp dụng phương pháp trồng bằng rễ trần cho thông 3 lá.

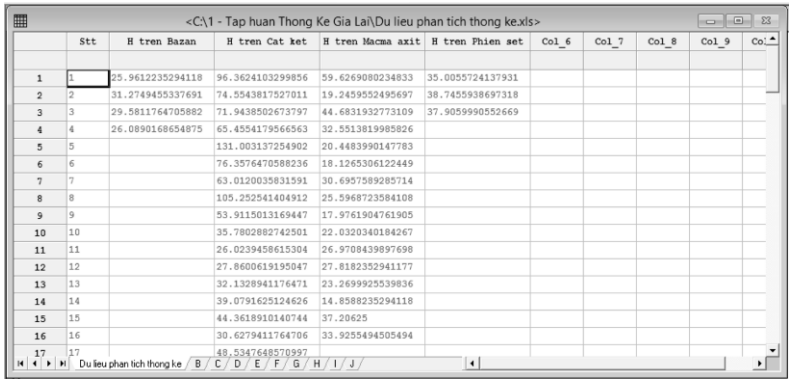
Ngoài ra theo Hình 6.1, trung bình H trồng theo phương pháp rễ trần cao hơn phương pháp cây con.




**Hình 6.1.** Đồ thị Box and Whisker biểu diễn biến động và trung bình H của hai phương pháp trồng thông 3 lá

Một ví dụ khác cho trường hợp kiểm tra trên hai mẫu độc lập theo tiêu chuẩn phi tham số theo Kruskal-Wallis và Friedman. Đó là so sánh tăng trưởng chiều cao trung bình cây tếch (TT_H, cm/năm) sau bốn năm đưa vào khi trồng làm giàu rừng khộp trên bốn loại đá mẹ (Dữ liệu 9; Bảo Huy, 2014). Thực hiện đánh giá trong Statgraphics theo trình tự tương tự trên, nhưng trước hết kiểm tra phân bố chuẩn của 4 mẫu và phương sai của chúng có bằng nhau hay không:

Nhập dữ liệu trung bình TT_H của từng ô thí nghiệm theo 4 loại đá mẹ, mỗi loại nằm trong một cột trong excel

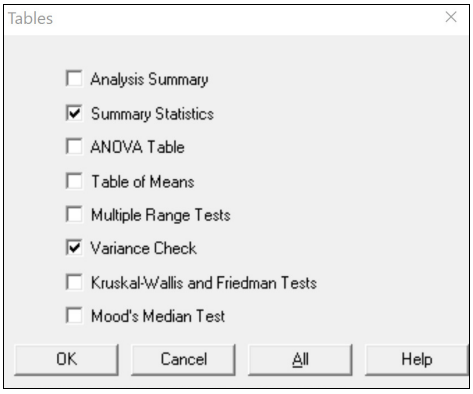



| | Stt | H tren Bazan | H tren Cat ket | H tren Macma axit | H tren Phien set | Col_6 | Col_7 | Col_8 | Col_9 | Co |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 25.9612235294118 | 96.3624103299856 | 59.6269080234833 | 35.0055724137931 | | | | | |
| 2 | 2 | 31.2749455337691 | 74.5543817527011 | 19.2459552495697 | 38.7455938697318 | | | | | |
| 3 | 3 | 29.5811764705882 | 71.9438502673797 | 44.6831932773109 | 37.9059990552669 | | | | | |
| 4 | 4 | 26.0890168654875 | 65.4554179566563 | 32.5513819985826 | | | | | | |
| 5 | 5 | | 131.003137254902 | 20.4483990147783 | | | | | | |
| 6 | 6 | | 76.3576470588236 | 18.1265306122449 | | | | | | |
| 7 | 7 | | 63.0120035831591 | 30.6957589285714 | | | | | | |
| 8 | 8 | | 105.252541404912 | 25.5968723584108 | | | | | | |
| 9 | 9 | | 53.9115013169447 | 17.9761904761905 | | | | | | |
| 10 | 10 | | 35.7802882742501 | 22.0320340184267 | | | | | | |
| 11 | 11 | | 26.0239458615304 | 26.9708439897698 | | | | | | |
| 12 | 12 | | 27.8600619195047 | 27.8182352941177 | | | | | | |
| 13 | 13 | | 32.1328941176471 | 23.2699925539836 | | | | | | |
| 14 | 14 | | 39.0791625124626 | 14.8588235294118 | | | | | | |
| 15 | 15 | | 44.3618910140744 | 37.20625 | | | | | | |
| 16 | 16 | | 30.6279411764706 | 33.9255494505494 | | | | | | |
| 17 | 17 | | 48.5347648570997 | | | | | | | |

Kiểm tra các mẫu có theo phân bố chuẩn và phương sai có bằng nhau hay không:

Trong nút Tables, chọn Summay Statistics và Varicance Check.

Kết quả như bảng sau:




**Bảng 6.2.** Tóm tắt thống kê và kiểm tra phương sai của 4 mẫu thử nghiệm trồng tếch trên 4 loại đá mẹ trong rừng khộp. Giá trị quan sát là tăng trưởng chiều cao trung bình của ô thí nghiệm. Tổng cộng có 64 ô thí nghiệm trên 4 loại đá mẹ

Summary Statistics

| | Count | Average | Standard deviation | Coeff. of variation | Minimum | Maximum | Range |
|---|---|---|---|---|---|---|---|
| TT_H tren Bazan | 4 | 28.2266 | 2.63492 | 9.3349% | 25.9612 | 31.2749 | 5.31372 |
| TT_H tren Cat ket | 41 | 48.4679 | 23.4149 | 48.3101% | 17.43 | 131.003 | 113.573 |
| TT_H tren Macma axit | 16 | 28.4396 | 11.5254 | 40.5259% | 14.8588 | 59.6269 | 44.7681 |
| TT_H tren Phien set | 3 | 37.2191 | 1.96236 | 5.27246% | 35.0056 | 38.7456 | 3.74002 |
| Total | 64 | 41.6684 | 21.6131 | 51.8693% | 14.8588 | 131.003 | 116.144 |

| | Stnd. skewness | Stnd. kurtosis |
|---|---|---|
| TT_H tren Bazan | 0.28028 | -1.64562 |
| TT_H tren Cat ket | 4.11222 | 4.25107 |
| TT_H tren Macma axit | 2.36311 | 2.01465 |
| TT_H tren Phien set | -0.977387 | |
| Total | 5.99291 | 7.47169 |

Variance Check

| | Test | P-Value |
|---|---|---|
| Levene's | 2.87066 | 0.043729 |

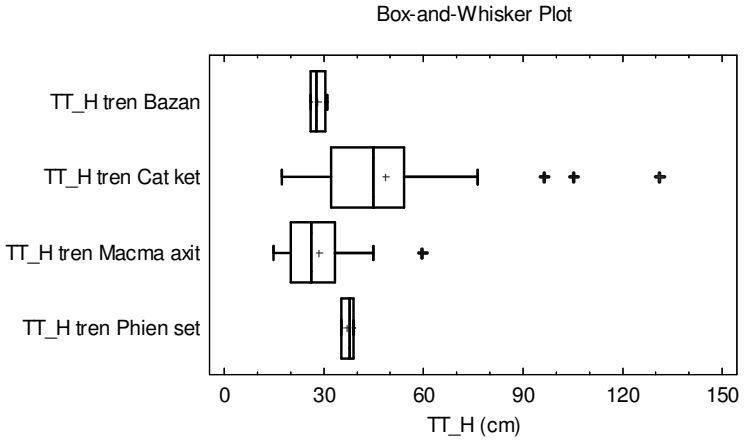
Kết quả cho thấy có ¾ mẫu là không đạt chuẩn với Stnd. Skewness hoặc Stnd. Kurtosis nằm ngoài phạm vị [-2, +2]. Kết quả kiểm tra phương sai theo tiêu chuẩn Levene cho P-Value = 0.043 < 0.05, có nghĩa là có sai khác rõ rệt giữa các phương sai của bốn mẫu thí nghiệm trên 4 loại đá mẹ. Vì vậy, với dữ liệu thí nghiệm này không thể áp dụng ANOVA một nhân tố để so sánh, đánh giá sai khác. Trường hợp này tiêu chuẩn phi tham số của Kruskal-Wallis được áp dụng, kết quả như sau:

**Bảng 6.3.** Kết quả kiểm tra 4 mẫu thí nghiệm trồng tếch làm giàu rừng khộp trên bốn loại đá mẹ theo tiêu chuẩn Kruskal-Wallis

| | Sample Size | Average Rank |
|---|---|---|
| TT_H tren Bazan | 4 | 19.0 |
| TT_H tren Cat ket | 41 | 39.0732 |
| TT_H tren Macma axit | 16 | 18.625 |
| TT_H tren Phien set | 3 | 34.6667 |

Test statistic = 16.1389 P-Value = 0.00106202.

Với P-Value = 0.001 < 0.05, có nghĩa là với độ tin cậy 95% cho thấy có sự khác biệt rõ rệt về tăng trưởng H trung bình của tếch trồng làm giàu rừng khộp trên 4 loại đá mẹ. Trong đó theo sắp xếp trung vị (Average Rank), thì tếch tăng trưởng tốt nhất trên đá mẹ Cát kết, sau đó đến đá Phiến sét và tiếp đến là đá Bazan, kém nhất trên Macma axit. Kết quả này cũng phù hợp với thứ tự tăng trưởng bình quân của tếch (TT_H, cm) trên 4 loại đá mẹ ở Hình 6.2.




**Hình 6.2.** Biểu đồ Box-Whisker về biến động tăng trưởng chiều cao tếch trên 4 loại đá mẹ

## 6.2 Tiêu chuẩn phi tham số kiểm tra, so sánh các mẫu liên hệ

Trong trường hợp có hai hay nhiều hơn các mẫu có liên hệ với nhau, và các mẫu này chưa đạt phân bố chuẩn, hoặc phương sai không bằng nhau, nên không thể sử dụng tiêu chuẩn t bắt cặp (với 2 mẫu) hoặc phân tích phương sai (trên 2 mẫu/công thức có liên hệ); thì tiêu chuẩn phi tham số là thích hợp để so sánh.

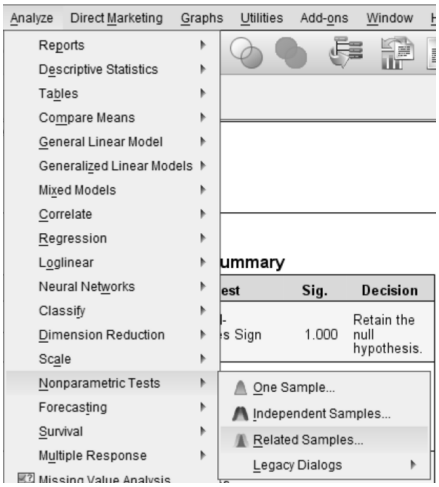
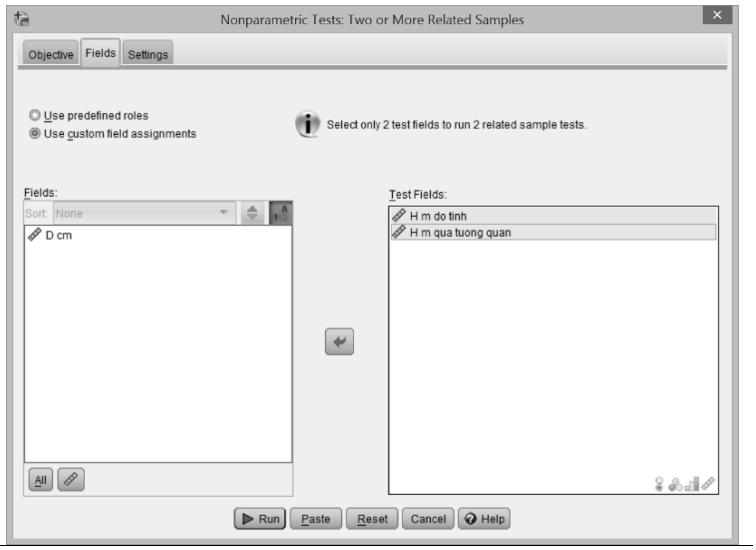
i) Trường hợp hai mẫu liên hệ, có thể sử dụng tiêu chuẩn phi tham số Wilcoxon, trong đó giá trị tuyệt đối chênh lệch |d| giữa các cặp dữ liệu được xếp hạng từ nhỏ đến lớn và tính giá trị trung vị Median cho hai nhóm cho chênh lệch $d < 0$ và $d > 0$, sau đó so sánh sự sai khác của hai trung vị của hai nhóm chênh lệch d:

Giả thuyết $H_0$: $x_1 = x_2$ hay giả thuyết $H_1$: $x_1 \# x_2$; trong đó $x_1$ và $x_2$ là trung vị (Median) của nhóm có chênh lệch âm và dương. Kiểm tra Wilcoxon, nếu P-Value < 0.05 thì bác bỏ giả thuyết $H_0$ và chấp nhận $H_1$, tức là hai mẫu bắt cặp có sự chênh lệch rõ rệt và ngược lại (IBM, 2011; Nguyễn Hải Tuất, 2982; Nguyễn Hải Tuất et al., 2006).

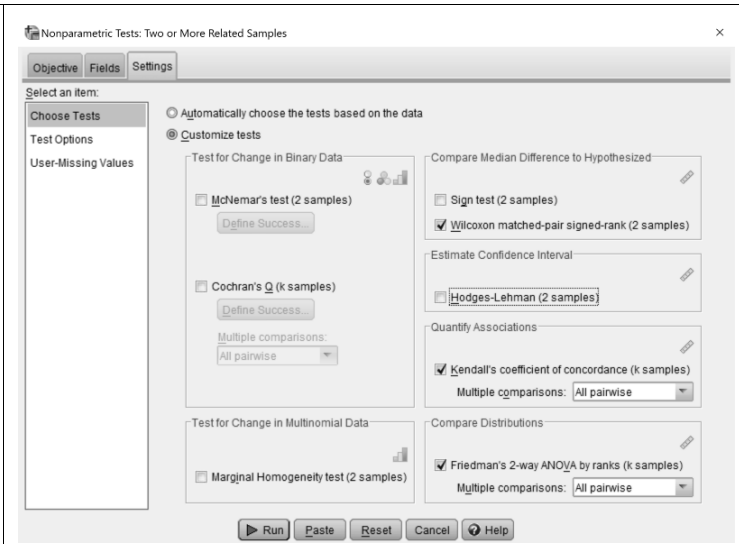
ii) Trường hợp có nhiều mẫu liên hệ (≥ 2 mẫu) thì tiêu chuẩn Friedman hoặc Kendall có thể được sử dụng để so sánh sự sai khác giữa các dãy phân bố dữ liệu của các mẫu:

Giả thuyết $H_0$: $F(x_1) = F(x_2) = ….. = F(x_n)$ hay giả thuyết $H_1$: $F(X_i) \# F(xj)$. Trong đó $F(x_1) = F(x_2) = ….. = F(x_n)$ là dãy phân bố tần số của các mẫu 1 đến n; $x_i$, $x_j$ là mẫu thứ i và j trong n mẫu. Kiểm tra theo một trong hai tiêu chuẩn của Friedman hoặc Kendall, nếu P-Value < 0.05 thì bác bỏ giả thuyết $H_0$ và chấp nhận $H_1$, có nghĩa là có ít nhất 2 trong n mẫu có sự sai khác rõ rệt về kiểu phân bố và ngược lại thì chấp nhận $H_0$, các mẫu có phân bố đồng nhất (IBM, 2011; Nguyễn Hải Tuất, 2982; Nguyễn Hải Tuất et al., 2006).

Sử dụng ví dụ kiểm tra sự sai khác giữa chiều cao được đo trực tiếp và thông qua mô hình tương quan (Dữ liệu 5) và áp dụng các tiêu chuẩn phi tham số Wilcoxon, Friwedman và Kendal để so sánh sự sai khác về trung vị của chênh lệch và phân bố trong SPSS như sau:

| | |
|---|---|
| Sử dụng tiêu chuẩn phi tham số để so sánh từ 2 đến nhiều mẫu liên hệ trong SPSS Analyze/Nonparametric Test/Related Samples |  |
| Trong hộp thoại với Tab: Field, đưa các biến so sánh vào Test Fields. |  |

Trong Tab: Settings chọn các test: Wilcoxon, Kendalls, Friedman, nếu so sánh hai mẫu theo trung vị và dãy phân bố; nếu trên hai mẫu thì không sử dụng test Wilcoxon, vì chỉ so sánh các dãy phân bố theo Friedman hoặc Kendalls.




**Bảng 6.4.** Kết quả so sánh hai mẫu liên hệ theo cả ba tiêu chuẩn Wilcoxon, Kendal và Friedman trong SPSS

**Hypothesis Test Summary**

| | Null Hypothesis | Test | Sig. | Decision |
|---|---|---|---|---|
| 1 | The median of differences between H m do truc tiep and H m qua tuong quan equals 0. | Related-Samples Wilcoxon Signed Rank Test | .936 | Retain the null hypothesis. |
| 3 | The distributions of H m do truc tiep and H m qua tuong quan are the same. | Related-Samples Kendall's Coefficient of Concordance | 1.000 | Retain the null hypothesis. |
| 2 | The distributions of H m do truc tiep and H m qua tuong quan are the same. | Related-Samples Friedman's Two-Way Analysis of Variance by Ranks | 1.000 | Retain the null hypothesis. |

Asymptotic significances are displayed. The significance level is .05.

Kết quả ở Bảng 6.4 cho thấy:

Kiểm tra theo Wilcoxon với 2 mẫu liên hệ theo trung vị của chênh lệch (Median) cho thấy Sig. (P-Value) = 0.936 > 0.05, có nghĩa là chưa thể bác bỏ giả thuyết $H_0$ với hai trung vị là bằng nhau. Như vậy, việc xác định H qua đo trực tiếp và ước tính qua tương quan là chưa có sai khác. Có thể sử dụng tương quan để giảm chi phí đo đếm chiều cao cây rừng.

Kiểm tra dãy phân bố của hai mẫu theo Kendall hoặc Friedman cho kết quả thấy Sig. (P-Value) = 1.0 > 0.05, như vậy, giả thuyết $H_0$ được chấp nhận, hay nói khác, chưa có sự sai khác giữa hai dãy số liệu đo H trực tiếp và ước tính qua phương trình. Có thể sử dụng phương trình để giảm chi phí điều tra.

Trong ví dụ trên, tiêu chuẩn Kendall và Friedman được sử dụng để so sánh dãy phân bố với 2 mẫu liên hệ; tuy nhiên, các tiêu chuẩn này được sử dụng tốt khi có trên 2 mẫu liên hệ; trong khi đó, tiêu chuẩn Wilcoxon chỉ giới hạn ứng dụng để so sánh hai mẫu liên hệ dựa vào trung vị của chênh lệch âm và dương.

Chương 7

# TIN HỌC ỨNG DỤNG TRONG MÔ HÌNH HÓA RỪNG (FOREST MODELLING)

## 7.1 Khái niệm chung về mô hình hóa rừng, mô hình quan hệ

Twery (2004) khi nói về mô hình hóa trong quản lý rừng đã chỉ ra rằng quản lý rừng truyền thống chỉ quan tâm đến cây cho gỗ. Thực tế nó bao gồm quản lý thực động vật, quản lý đất và quản lý con người với đa mục tiêu. Để quản lý được các mối quan hệ của các thành phần hệ sinh thái rừng với nhau và với hoạt động của con người, thì của mô hình hóa là chìa khóa cho quản lý rừng, đó là mô phỏng một cách chính xác động thái rừng với ảnh hưởng của đa nhân tố.

Các mô hình được sử dụng để quản lý rừng bao gồm một số dạng. Đầu tiên là các mô hình sinh trưởng và sản lượng rừng (Twery, 2004), trong truyền thống thì chủ yếu là các mô hình thể tích và sản lượng gỗ, ngày nay trong tình hình biến đổi khí hậu, cần mô hình ước tính sinh khối carbon cây rừng và lâm phần (Brown, 1997; Huy et at., 2016a,b,c). Một cách tổng quát, mô hình tương quan sinh học (allometry) là các mô hình tuyến tính hoặc phi tuyến tính mô tả quan hệ, tương quan giữa các biến số điều tra cây rừng, lâm phần, hệ sinh thái, xã hội (Picard et al., 2012). Basuki et al. (2009) đã sử dụng các mô hình hàm power được tuyến hóa bằng cách logarit để xây dựng mô hình sinh khối trên mặt đất cây rừng khộp ở vùng thấp của Indonesia. Trong khi đó Brown et al. (1989) và Chave et al. (2005) sử dụng hàm phi tuyến parabol. Pearson (2007) đã đề nghị sử dụng hàm mũ để thiết lập mô hình sinh khối cây rừng cho các loài cây và kiểu rừng ở Hoa Kỳ.

*Đóng góp của mô hình hóa bao gồm (Twery, 2004):*

*Mô hình hóa hệ sinh thái rừng*: Bao gồm mô hình hóa quá trình sinh trưởng và sản lượng rừng, mô hình sinh khối carbon rừng; mô hình quá trình tái sinh rừng; mô hình cây chết; mô hình nơi sống của động vật hoang dã (Habitat).

*Mô hình hóa mối quan hệ rừng và con người*: Bao gồm mô hình về khai thác rừng; mô hình về nghỉ dưỡng và sinh thái,...

Trong thực tế, người ta cần lập các mô hình tương quan hồi quy với dạng tổng quát $y = f(x_i)$ vì các mục đích:

- Để ước lượng một nhân tố khó đo đếm (gọi là biến phụ thuộc y) thông qua một hay nhiều biến dễ quan sát, đo đếm (gọi là biến độc lập $x_i$) và tất nhiên là phải có mối liên hệ giữa y và các

biến $x_i$.

- Để nghiên cứu tác động, ảnh hưởng của một hoặc nhiều nhân tố (biến $x_i$) đến một yếu tố cần quan tâm như sinh trưởng, sản lượng, chất lượng rừng, xói mòn đất, dòng chảy lưu vực (biến phụ thuộc y). Trên cơ sở đó, có giải pháp kỹ thuật thích hợp.

- Để dự báo một nhân tố trong tương lai (gọi là biến dự báo y) với một hay nhiều biến số độc lập, đầu vào ($x_i$)

Có thể thấy mô hình hóa quá trình sinh trưởng, tăng trưởng, các mối quan hệ của các nhân tố cây rừng, lâm phần, hệ sinh thái rừng, mô phỏng động thái rừng là một lĩnh vực được áp dụng rộng rãi và mang lại nhiều ý nghĩa, không chỉ cho khoa học thống kê mà còn đóng góp cho nhiều ứng dụng hữu ích trong quản lý tài nguyên rừng và môi trường.

Đồng Sĩ Hiền (1974) đã có ứng dụng mạnh mẽ mô hình quan hệ đa thức bậc cao, nhiều lớp để lập các phương trình đường sinh thân cây làm cơ sở lập biểu thể tích cây đứng cho rừng Việt Nam, khi mà tin học chưa được phát triển. Nhiều nhà nghiên cứu Việt Nam sau đó đã áp dụng và phát triển mạnh các ứng dụng của mô hình hóa, mô phỏng, lập mô hình quan hệ: Nguyễn Ngọc Lung (1989) đã chỉ ra một loạt các mô hình phi tuyến để mô phỏng quá trình sinh trưởng cây rừng; Bảo Huy (1993) đã áp dụng mô hình phi tuyến mũ exp để mô phỏng quá trính sinh trưởng rừng tự nhiên; Vũ Tiến Hinh (2012) đánh giá và phát triển thêm nhiều mô hình ước tính thể tích cây rừng tự nhiên Việt Nam; Vũ Tiến Hinh và Trần Văn Con (2012) đã vận dụng đa dạng các kiểu dạng mô hình hóa để dự đoán sản lượng rừng trồng và tự nhiên. Bảo Huy và cộng sự (1998), Bảo Huy và Đào Công Khanh (2008) đã mô hình hóa sinh trưởng và dự đoán sản lượng cho rừng trồng tếch, trắm trắng. Gần đây Bảo Huy (2013), Huy et al. (2016a,b,c) đã áp dụng mô hình hóa để thiết lập hệ thống mô hình ước tính sinh khối cây rừng tự nhiên cho nhiều kiểu rừng và vùng sinh thái của Việt Nam.

Vanclay (1994) đã chỉ ra ứng dụng rộng rãi mô hình trong nghiên cứu sinh trưởng và sản lượng rừng. Trong tình hình biến đổi khí hậu, để ước tính lượng $CO_2$ rừng hấp thụ nhằm giảm nhẹ biến đổi khí hậu, hàng loạt các mô hình ước tính sinh khối carbon cây rừng được xây dựng, đánh giá; đó là thiết lập mối quan hệ sinh khối cây rừng trên mặt đất (AGB) với một trong nhiều biến số độc lập như đường kính ngang ngực (DBH), chiều cao (H), khối lượng thể tích gỗ (WD), diện tích tán lá (CA), các dạng mô hình được sử dụng là power, logarit, hàm exp (Brown, 1997; Jenkins et al., 2003, 2004; IPCC, 2003; Basuki et at., 2009; Dietz et al., 2011; Johannes et al., 2011; Chave et al., 2005, 2014; Henry et al., 2015, Huy et al. 2016a,b,c).

Khái quát trên cho thấy, mô hình hóa các mối quan hệ cây rừng, lâm phần và hệ sinh thái rừng là một lĩnh vực khoa học rộng và mang lại nhiều ý nghĩa trong ứng dụng của quản lý rừng đa mục tiêu.

## 7.2 Các tiêu chuẩn, tiêu chí thống kê để so sánh, đánh giá, lựa chọn mô hình quan hệ

Quan hệ giữa đại lượng phụ thuộc y và các biến số độc lập, ảnh hưởng $x_i$ trong sinh học, sinh thái môi trường rừng thường có kiểu dạng quan hệ phức tạp, do vậy, việc lựa chọn được mô hình tối ưu để mô tả tốt nhất mối quan hệ $y = f(x_i)$ trong thực tế cần dựa vào nhiều tiêu chí thống kê khác nhau.

Phổ biến nhất dựa vào các chỉ tiêu thống kê:

Hệ số xác định $R^2$: Về tổng quát thì hàm tốt nhất khi $R^2$ đạt max và tồn tại ở mức sai $P < 0.05$. Tuy nhiên, có trường hợp $R^2$ đạt max nhưng chưa phải là hàm phù hợp nhất, do vậy, cần dựa thêm các chỉ tiêu thống kê khác.

Tồn tại của các tham số mô hình: Nếu là hàm có từ hai biến số độc lập trở lên, thì biến độc lập phải tồn tại qua kiểm tra theo tiêu chuẩn t ở mức $P < 0.05$.

Chỉ tiêu AIC (Akaike Information Criterion) - một chỉ tiêu đo lường độ tin cậy của dự báo qua mô hình, AIC càng bé thì mô hình càng có độ tin cậy cao hơn (Basuki et al., 2009; Picard et al., 2012; Bảo Huy, 2013; Huy et al. 2016a,b,c)

$$AIC = -2\ ln(L) + 2p \tag{7.1}$$

Trong đó L là Likelihood của mô hình, p là tổng số tham số của mô hình.

Trong so sánh các mô hình cùng biến y, chỉ tiêu AIC có tầm quan trọng hơn khi đánh giá so với hệ số xác định $R^2$. Có trường hợp hàm được chọn dựa vào AIC bé hơn cho dù $R^2$ của nó có thể bé hơn hàm so sánh; bởi vì AIC phản ảnh toàn diện độ tin cậy của giá trị ước lượng so với quan sát. Tuy nhiên, cần lưu ý khi sử dụng các chỉ tiêu thống kê $R^2$ và AIC để so sánh các dạng mô hình khác nhau, lúc này yêu cầu là các mô hình phải có cùng dạng biến số y; không thể áp dụng các chỉ tiêu này để so sánh một mô hình có biến là y và mô hình khác có biến y được đổi biến số ví dụ là log(y). Khi các mô hình có biến y khác dạng với nhau thì để so sánh cần áp dụng tiêu chuẩn Furnival (Furnival, 1961; Jayaraman, 1999) được giới thiệu trong phần tiếp theo.

Các loại sai số để đánh giá và so sánh các mô hình được chia làm hai nhóm sai số:

*Các sai số tuyệt đối:* (Mayer and Butler, 1993; Temesgen et al., 2014).

Bias là sai lệch trung bình giữa giá trị quan sát so với dự đoán qua mô hình, sai lệch âm có nghĩa là dự báo cao hơn thực tế và ngược lại. Khi lấy trung bình thì sai số âm dương sẽ bù trừ nhau, do đó, sai số này thường nhỏ. Vì vậy, sai số này thường được sử dụng để xét giá trị dự báo vượt trên hay nằm dưới quan sát nhờ dấu của nó.

$$Bias = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \tag{7.2}$$

RMSE (Root Mean Square Error): Sai số trung phương trung bình. Sai số được tính trên cơ sở sai lệch giữa quan sát và dự báo bình phương và luôn luôn dương. Vì vậy, RMSE dùng để đánh giá sai số trung bình không xét đến chiều hướng của sai lệch, như thế RMSE luôn lớn hơn Bias:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{7.3}$$

MAE (Mean Absolute Error): Sai số tuyệt đối trung bình. Sai số này khá tương đồng với RMSE, thay vì bình phương sai lệch giữa quan sát và dự báo thì sai số này lấy giá trị tuyệt đối, do đó MAE luôn dương và chỉ thị cho sai số trung bình không xét đến chiều hướng sai lệch.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (7.4)$$

*Các sai số tương đối:*

Trong trường hợp để thuận lợi hơn khi so sánh sai số của các mô hình, các sai số trên được tính theo phần trăm, bao gồm % sai lệch giữa quan sát và dự báo qua mô hình (Bias %), sai số trung bình % (RMSE %), và sai số tuyệt đối trung bình % (Mean Absolute Percent Error MAPE) (Mayer et al, 1993; Chave et al, 2005; Basuki et al., 2009; Swanson et al., 2011; Huy *et al.,* 2016a,b,c):

$$Bias\ \% = \frac{100}{n}\sum_{i=1}^{n}\frac{(y_i - \hat{y}_i)}{(y_i)} \quad (7.5)$$

$$RMSE\ \% = 100\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{y_i}\right)^2} \quad (7.6)$$

$$MAPE\ \% = \frac{100}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \quad (7.7)$$

Trong các công thức tính sai số nói trên, n là số mẫu; và $y_i$ và $\hat{y}_i$ là giá trị quan sát và ước tính qua mô hình.

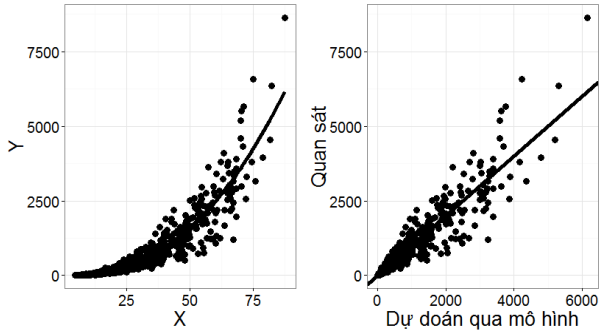
## 7.3 Các biểu đồ, đồ thị dùng để đánh giá, so sánh các mô hình

Ngoài các chỉ tiêu thống kê, sai số để đánh giá, lựa chọn mô hình, cũng cần xem xét độ tin cậy, sai số, sự bám sát của giá trị ước lượng so với quan sát thông qua các đồ thị trực quan.

Có nhiều loại đồ thị để đánh giá mô hình, trong đó các đồ thị sau thường được quan tâm (Mehtatalo, 2013; Huy et al., 2016b):

*Đồ thị biểu diễn mô hình so với giá trị quan sát*: Các giá trị quan sát càng bám sát mô hình thì mô hình càng tốt (Đồ trị trái của Hình 7.1).

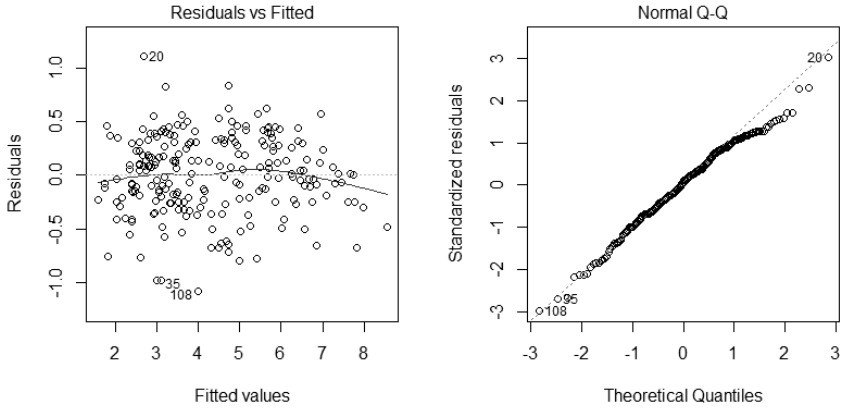
*Đồ thị biểu diễn quan hệ giữa giá trị quan sát với dự đoán qua mô hình*: Các đám mây điểm càng bám sát đường chéo có nghĩa giá trị quan sát và dự đoán càng gần trùng khớp (Đồ thị phải của Hình 7.1).




**Hình 7.1.** Trái: Đồ thị mô hình so với quan sát, Phải: Quan sát so với dự đoán qua mô hình

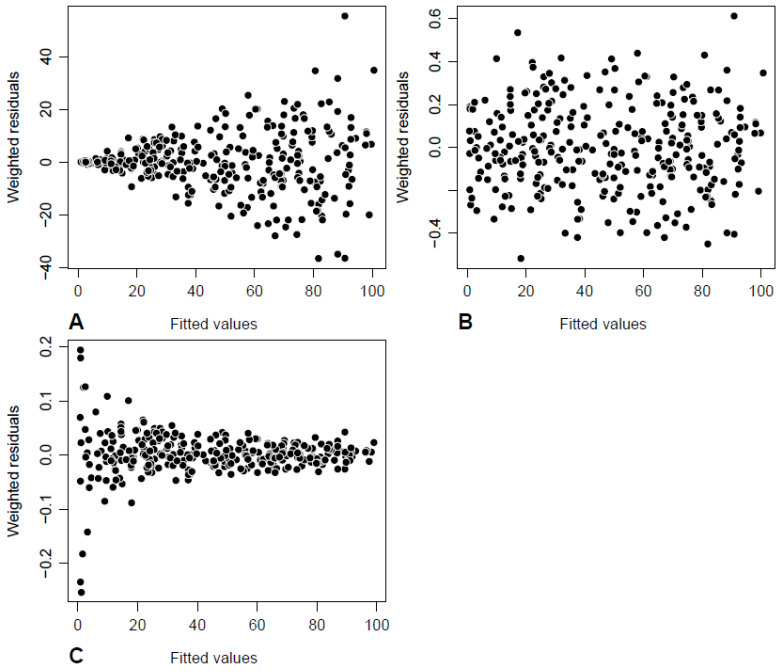
*Biểu đồ biến động sai số (Residuals) theo các giá trị dự báo qua mô hình (Fitted values):* Một mô hình tốt thì biến động sai số Residuals bám chung quanh đường y = 0 và rải đều theo đường này, biến động của Studentized Residuals tốt với độ tin cậy 95% thì trong phạm vi -2 đến +2 (Đồ thị trái của Hình 7.2).

*Biểu đồ phân bố Normal Q-Q*: Một mô hình tốt, bám sát dữ liệu quan sát thì biểu đồ phân bố đám mây điểm theo đường chéo, mô hình có ước lượng sai số lớn thì các điểm phân bố nằm xa đường chéo của đồ thị (Đồ thị phải của Hình 7.2).




**Hình 7.2.** Biến động sai số theo dự đoán (trái) và biểu đồ phân bố Q-Q (phải)

Picard et al. (2012) đã chỉ ra các loại biến động sai số residuals theo dự báo qua mô hình ở Hình 7.3, chọn mô hình thích hợp hay chọn biến trọng số đúng sẽ bảo đảm cho sai số có phân bố đều. Các loại sai số dạng A và C chỉ ra mô hình có giá trị dự báo bị sai lệch lớn, trong khi đó sai số ở dạng B chỉ ra mô hình phù hợp với dữ liệu và cho sai số phân bố đều ở các giá trị dự báo.




**Hình 7.3.** Các kiểu biến động sai số Residuals theo gía trị đự đoán qua mô hình: A: Biến động Residuals rộng ở các giá trị dự báo lớn; C: Biến động Residuals rộng ở các giá trị dự báo nhỏ; B: Biến động Residuals rải đều theo giá trị dự báo *(Picard et al., 2012)*

## 7.4 Mô hình tuyến tính

### 7.4.1 Mô hình tuyến tính đơn

Mô hình quan hệ giữa một biến ảnh hưởng, độc lập và một biến phụ thuộc, bị ảnh hưởng, dự báo trong đó $y_i$ và $x_i$ được đo đạc từ các mẫu i, i = 1, … n. Mô hình tuyến tính đơn biến cho cá thể. Mẫu thứ i được viết như sau (Laar và Akca, 2007; Mehtatalo, 2013):

$$y_i = b_0 + b_1 x_i + \varepsilon_i \tag{7.8}$$

Trong mô hình trên biến $y_i$ gồm hai bộ phận: Một là phụ thuộc vào biến $x_i$ theo mô hình tuyến tính và hai là sai số ngẫu nhiên của mô hình $\varepsilon_i$. $b_0$ và $b_1$ là hai tham số của mô hình tuyến tính đơn. Phương pháp ước lượng các tham số của mô hình chủ yếu là bình phương tối thiểu.

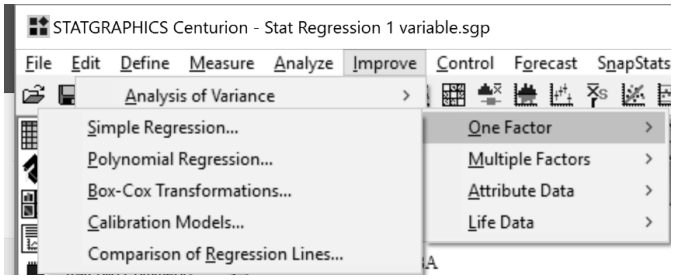
Việc thiết lập mô hình tuyến tính một biến số có thể áp dụng chương trình lập sẵn trong Statgraphics hoặc lập Codes để chạy trong R. Nếu áp dụng Statgraphics thì các chi tiết thống kê, đồ thị chỉ sử dụng theo chương trình lập sẵn, trong khi đó, nếu sử dụng R sẽ có điều kiện tính toán các chỉ tiêu thống kê, đồ thị khác nhau theo mong muốn.

Một ví dụ cho sử dụng mô hình tuyến tính đơn (một biến độc lập) đó là thiếp lập mối quan hệ giữa trữ lượng rừng (M, $m^3$/ha) theo tổng tiết diện ngang (BA, $m^2$/ha). Quan hệ M = f(BA) có khả năng biểu diễn tốt ở dạng tuyến tính M = a + b × BA. Từ 120 ô mẫu 1000$m^2$ rừng trồng tếch ở Tây Nguyên (Bảo Huy và cộng sự, 1998) đã tính toán được các biến số lâm phần tếch gồm M, BA, chiều cao và đường trung bình (Hg (m), Dg (cm)), mật độ (N, cây/ha),… trong Dữ liệu 10 ở phần phụ lục. Sử dụng bộ dữ liệu này để thiết lập mô hình M = a + b × BA theo hai chương trình thống kê là Statgraphics và mã nguồn mở R.

#### 7.4.1.1 Thiết lập mô hình tuyến tính một biến trong Statgraphics

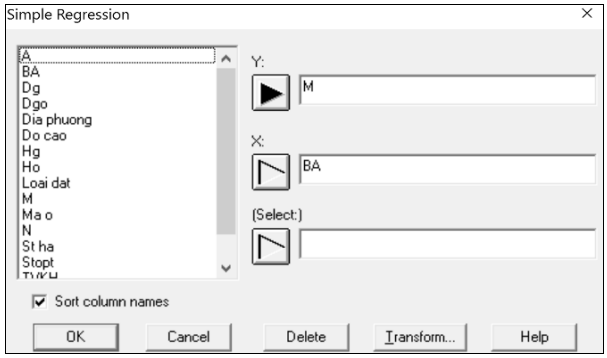
Thiết lập mô hình M = a + b × BA trong Statgraphics theo phương pháp bình phương tối thiểu từ Dữ liệu 10 với 120 ô mẫu của rừng trồng tếch ở Tây Nguyên như sau:

Thực hiện thiết lập mô hình tuyến tính 1 biến số trong Statgraphics: Improve / Regression Analysis / One Factor / Simple Regression.




Trong hộp thoại: Simple Regression:

Nhập biến M và vào Y và BA vào X.



**Kết quả thiết lập mô hình tuyến tính một biến trong Statgraphics**

**Simple Regression - M vs. BA**

Dependent variable: M
Independent variable: BA
Linear model: Y = a + b*X

**Coefficients**

| | *Least Squares* | *Standard* | *T* | |
|---|---|---|---|---|
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| Intercept | -23.1112 | 5.28506 | -4.37292 | 0.0000 |
| Slope | 9.48724 | 0.322611 | 29.4077 | 0.0000 |

**Analysis of Variance**

| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
|---|---|---|---|---|---|
| Model | 558148. | 1 | 558148. | 864.81 | 0.0000 |
| Residual | 76156.9 | 118 | 645.397 | | |
| Total (Corr.) | 634305. | 119 | | | |

Correlation Coefficient = 0.938049
R-squared = 87.9937 percent
R-squared (adjusted for d.f.) = 87.8919 percent
Standard Error of Est. = 25.4047
Mean absolute error = 17.7579
Durbin-Watson statistic = 0.852836 (P=0.0000)
Lag 1 residual autocorrelation = 0.552516

**The StatAdvisor**

The output shows the results of fitting a linear model to describe the relationship between M and BA. The equation of the fitted model is

**M = -23.1112 + 9.48724*BA**

Kết quả trên cho ra mô hình: M = -23.1112 + 9.48724*BA

Với hệ số xác định hiệu chỉnh $R^2$adj. (%) = 87.8919 %, ứng với P-Value = 0.0000 < 0.001 (trong bảng Analysis of Variance), chứng tỏ R tồn tại (bác bỏ giả thuyết $H_0$: $R^2 = 0$).

Các tham số của mô hình đều có P-Value = 0.000 < 0.001 (trong bảng các tham số Coefficients), cho thấy các tham số đều khác 0 (bác bỏ giả thuyết $H_0$: $b_i = 0$).

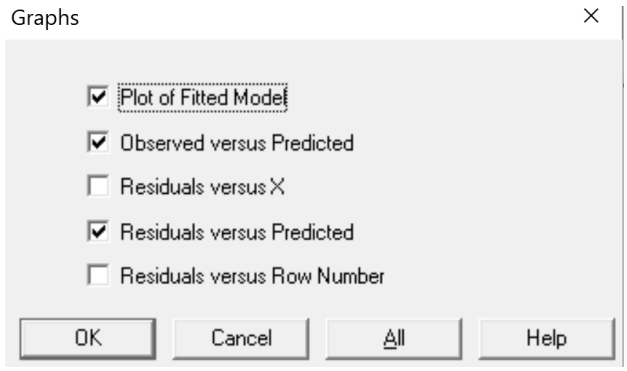
Sai số MAE (Mean Absolute Error - Sai số tuyệt đối trung bình) =17.7579 $m^3$/ha (đơn vị tính M là $m^3$/ha và BA là $m^2$/ha). Sai số này được tính từ chính dữ liệu lập mô hình.

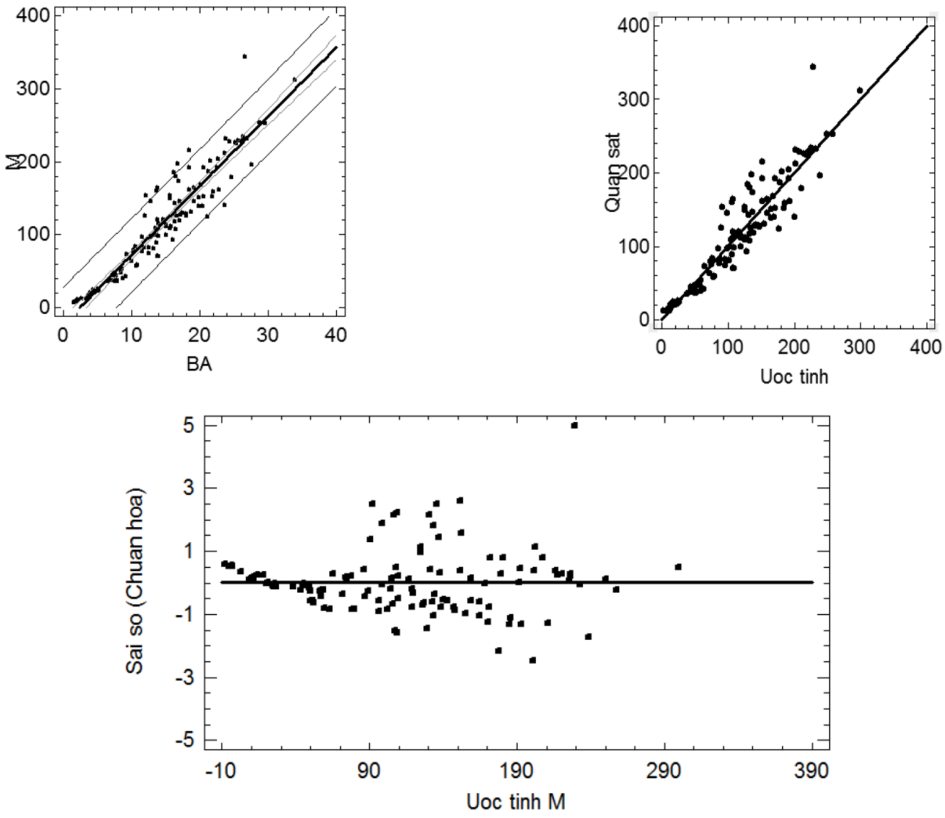
Xuất ra các đồ thị: Trong nút Graphs chọn xuất ra 3 loại đồ thị chính như sau:

Plot of Fitted Model: Đồ thị mô hình theo quan sát.

Observered versus Predicted: Đồ thì quan hệ giữa giá trị quan sdats và dự đoán qua mô hình.

Residuals versus Predicted: Đồ thị biến động sai số theo giá trị dự đoán quan mô hình.




**Hình 7.4.** Quan hệ giữa mô hình và dữ liệu quan sát (trên trái), dữ liệu quan sát với dự đoán (trên phải) và biến động sai số theo dự đoán (dưới) thực hiện trong Statgraphics (Mô hình: M = a + b×BA)

Các đồ thị chính biểu diễn quan hệ tuyến tính một biến và biến động sai số thực hiện trong Statgraphics biểu diễn ở Hình 7.4. Tổng quát mô hình và giá trị dự báo khá bám sát giá trị quan sát, sai số biến động chuẩn hóa khá rải đều theo giá trị quan sát và đa số nằm trong phạm vi -2 đến +2; vì vậy có thể kết luận mô hình tuyến tính mô tả khá tốt quan hệ giữa trữ lượng và tổng tiết diện ngang rừng trồng tếch.

Kết quả thiết lập mô hình quan hệ trong Statgraphics đã cho ra các kết quả chính, các đồ thị cơ bản để đánh giá mô hình. Tuy nhiên, đây là phần mềm lập sẵn, không thay đổi, vì vậy, nếu cần chỉ tiêu thống kê quan trọng như AIC để đánh giá mô hình, và các loại sai số khác như Bias, RMSE,

MAPE hoặc các đồ thị quan hệ Q-Q,... thì Statgraphics không cung cấp. Do vậy chương trình mã nguồn mở R là một lựa chọn tốt cho người sử dụng để có thể trực tiếp lập các bảng mã codes để thu được các kết quả đầu ra khác nhau, đầy đủ theo ý định nghiên cứu.

#### 7.4.1.2 Thiết lập mô hình tuyến tính một biến trong chương trình R

Như đã giới thiệu, chương trình R là một chương trình được thực hiện trên cơ sở người sử dụng viết các codes để chạy mô hình và đưa ra các kết quả, chỉ tiêu, đồ thị đầu ra theo ý muốn.

Sử dụng Dữ liệu 10 để lập quan hệ tuyến tính M = a + b × BA từ dữ liệu 120 ô mẫu của rừng trồng tếch.

Sau đây là các codes để lập và đánh giá mô hình tuyến tính một biến theo phương pháp bình phương tối thiểu trong R, theo chương trình "lm" (Chambers, 1992):

**Codes khởi động và xác định dữ liệu tính toán**

```
# Erase memory (Xóa bộ nhớ)
rm(list=ls())

# Clean plot window (Xóa dữ liệu cửa sổ cũ)
dev.off()

# Install.packages (Cài đặt các chương trình: ggplot2 là đồ thị)
library(ggplot2)

# Define the working directory (change \ with / using Edit>Find) (Xác định địa chỉ fiel dữ liệu)
setwd("E:/1 - Bao Huy 2017/Books All/Textbook for Infomatic Statistic/TextbookDataset Analysis/Dataset")

# Import data from *.txt file (Nhập dữ liệu dạng txt)
t <- read.table("Teak 120 plots.txt", header=T,sep="\t",stringsAsFactors = FALSE)
```

**Codes thiết lập mô hình tuyến tính một biến**

```
# Model development: (Lập mô hình theo lm)
Linear_model1 <- lm(M~BA, data=t)

# Outputs of the model (Giá trị dự đoán và sai số của mô hình):
t$Linear_model1.fit <- fitted.values(Linear_model1)
t$Linear_model1.res <- residuals(Linear_model1)

# Summary of linear model: (Tóm tắt kết quả mô hình)
summary(Linear_model1)
anova(Linear_model1)
```

Kết quả tính toán mô hình tuyến tính một biến trong R như sau:

```
summary(Linear_model1)

Call:
lm(formula = M ~ BA, data = t)

Residuals:
    Min      1Q  Median      3Q     Max
-60.836 -15.343  -1.533   9.258 113.897

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -23.1112     5.2851  -4.373 2.66e-05 ***
BA            9.4872     0.3226  29.408  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 25.4 on 118 degrees of freedom
Multiple R-squared:  0.8799,   Adjusted R-squared:  0.8789
F-statistic: 864.8 on 1 and 118 DF,  p-value: < 2.2e-16

> anova(Linear_model1)
Analysis of Variance Table

Response: M
           Df Sum Sq Mean Sq F value    Pr(>F)
BA          1 558148  558148  864.81 < 2.2e-16 ***
Residuals 118  76157     645
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Kết quả trên, mục Coeficients cung cấp các giá trị tham số mô hình và kiểm tra bằng tiêu chuẩn t với Pr được cung cấp. Với kết quả này có được mô hình:

$M = -23.1112 + 9.4872 \times BA$

Với $R^2adj = 0.8789$; các tham số đều có giá trị Pr < 0.0000, có nghĩa là tồn tại các tham số.

Bảng kết quả ANOVA cho thấy giá Pr kiểm tra sự tồn tại của R là 2.2e-16 < 0.000, có nghĩa là R # 0 rõ rệt, hay R tồn tại, hay tồn lại mối quan hệ giữa M và BA ở mức 87.89%.

**Codes trong R tính toán AIC và các sai số**

```
# Indicators for validation of the model (Cacsc chỉ tiêu đánh giá mô hình):
AIC(Linear_model1)

Bias <- mean(t$Linear_model1.res)
RMSE <- sqrt(mean((t$Linear_model1.res)^2))
MAPE <- 100*mean(abs(t$Linear_model1.res)/t$M)

Bias
RMSE
MAPE
```

Kết quả tính toán các chỉ tiêu đánh giá, sai số mô hình (ở đây tính Bias, RMSE và MAPE):

```
> AIC(Linear_model1)
[1] 1120.912

> Bias <- mean(t$Linear_model1.res)
> RMSE <- sqrt(mean((t$Linear_model1.res)^2))
> MAPE <- 100*mean(abs(t$Linear_model1.res)/t$M)

> Bias
[1] 8.082366e-16
> RMSE
[1] 25.19208
> MAPE
[1] 21.77369
```

Các chỉ tiêu đánh giá:

AIC = 1120.912;

Bias = 8.082366e-16; Bias gần bằng 0 có nghĩa là sai số âm và dương của giá trị dự báo là xấp xỉ nhau, hay nói khác mô hình không dự báo quá cao hay quá thấp so với quan sát.

RMSE = 25.19208 $m^3$/ha; Sai số trung phương cho thấy ước tính theo mô hình này sẽ cho sai số trung bình về M là 25 $m^3$/ha

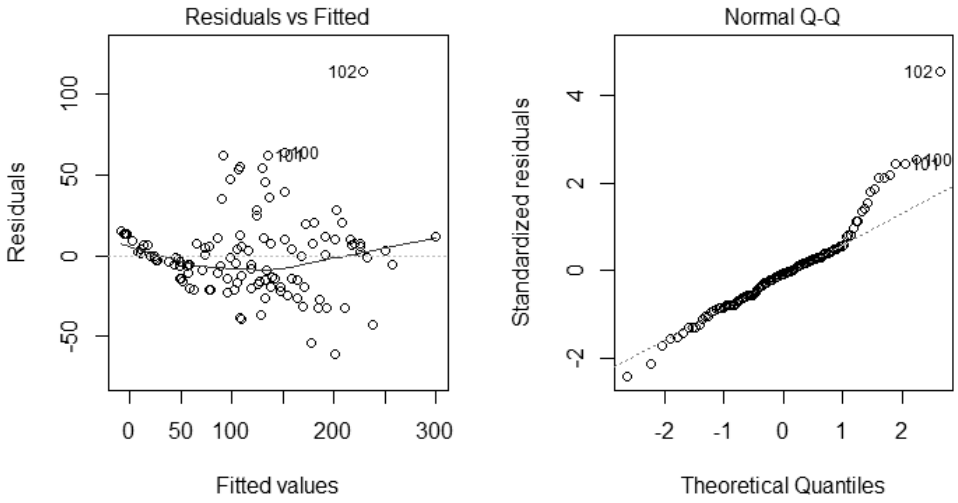
MAPE = 21.77 %; cho thấy sai số của mô hình ước tính M là khoảng 22%.

**Codes của R xuất ra các đồ thị đánh giá mô hình tuyến tính 1 biến:**

```
# Residuals and Normal Q-Q Plots: (Đồ thị sai số và Q-Q):
par(mfrow = c(1, 2))
plot(Linear_model1,1:2)

# Model and Observation: (Đồ thị mô hình và quan sát):
p <- ggplot(t, aes(x=BA, y=M))
p <- p + geom_point(cex=2)
p <- p + geom_line(cex = 1.5, aes(x=BA, y=Linear_model1.fit))
p <- p + xlab("BA (m2/ha)") + ylab("M (m3/ha") + theme_bw()+ theme_bw()
p = p + theme(axis.title.y = element_text(size = rel(1.5)))
p = p + theme(axis.title.x = element_text(size = rel(1.5)))
p <- p + theme(plot.title = element_text(size = rel(1.7)))
p = p + theme(axis.text.x = element_text(size=15))
p = p + theme(axis.text.y = element_text(size=15))
p

# Observated and Predicted Values: (Đồ thị Quan sát với dự báo):
p <- ggplot(t, aes(x=Linear_model1.fit, y=M))
p <- p + geom_point(cex=2)
p <- p + geom_abline(cex = 1.5, intercept = 0, slope = 1, col="black")
p <- p + xlab("Dự đoán (m3/ha)") + ylab("Quan sát (m3/ha)") + theme_bw()+ theme_bw()
```

```
p = p + theme(axis.title.y = element_text(size = rel(1.5)))
p = p + theme(axis.title.x = element_text(size = rel(1.5)))
p <- p + theme(plot.title = element_text(size = rel(1.7)))
p = p + theme(axis.text.x = element_text(size=15))
p = p + theme(axis.text.y = element_text(size=15))
p

# Resuduals and Predicted value (Đồ thị biến động sai số theo dự báo):
p <- ggplot(t, aes(x=Linear_model1.fit, y=Linear_model1.res))
p <- p + geom_point(cex = 2)
p <- p + geom_line(cex = 1.5, aes(x=Linear_model1.fit, y=0))
p = p + theme(axis.title.y = element_text(size = rel(1.5)))
p = p + theme(axis.title.x = element_text(size = rel(1.5)))
p <- p + theme(plot.title = element_text(size = rel(1.7)))
p = p + theme(axis.text.x = element_text(size=15))
p = p + theme(axis.text.y = element_text(size=15))
p <- p + xlab("Dự đoán (m3/ha)") + ylab("Sai số (m3/ha)") + theme_bw()+ theme_bw()
p
```

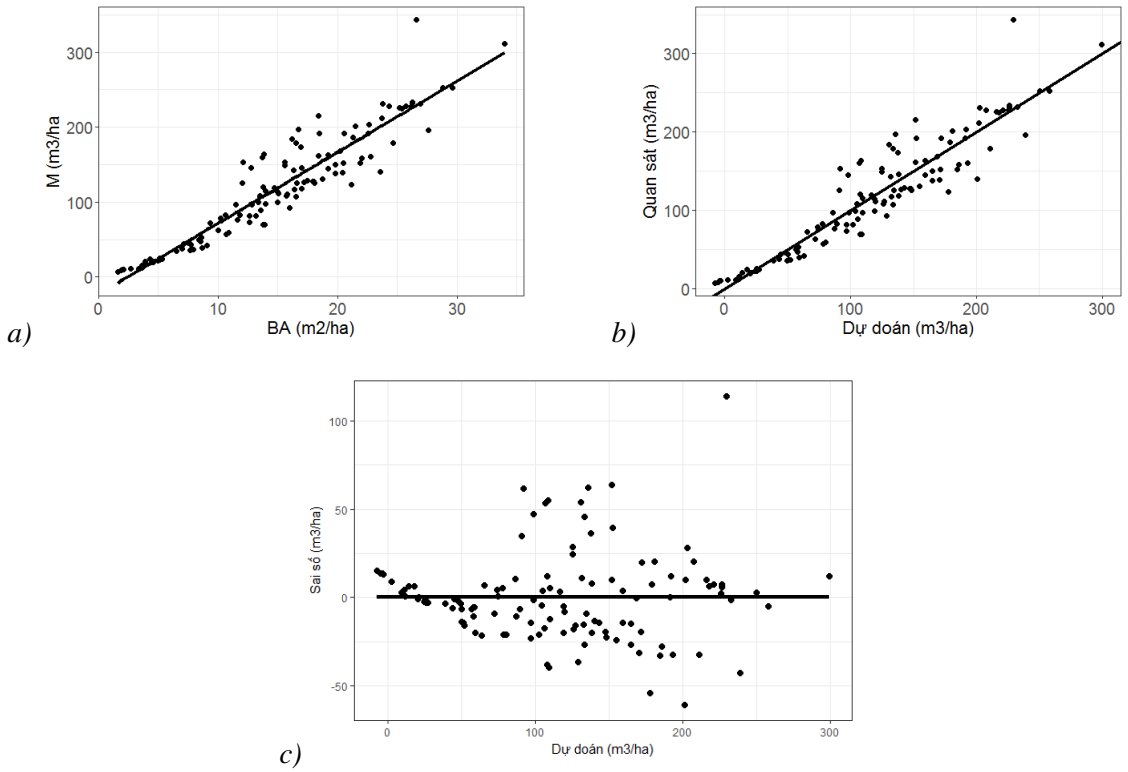
Kết quả cho ra các đồ thị đánh giá mô hình như sau:




**Hình 7.5.** Sai số theo dự báo (trái) và phân bố Q-Q (Mô hình: M = a + b×BA)

Đồ thị ở Hình 7.5 trên cho thấy sai số khá phân tán khi giá trị dự báo tăng lên và đồ thị Q-Q thì phân bố có phần nằm xa đường chéo, chứng tỏ mô hình chưa thực sự phù hợp.

Kết quả Hình 7.6 cho thấy giá trị dự báo và quan sát khá bám sát nhau, như vậy mô hình khá tốt, tuy nhiên đồ thị sai số chỉ ra biến động sai số tăng khi giá trị dự báo tăng lên, kết quả này phù hợp với đồ thị Q-Q, cho thấy mô hình tuyến tính mô tả chưa thực sự tốt quan hệ giữa M và BA của rừng trồng tếch. Vì vậy cần thử nghiệm mô phỏng theo các mô hình phi tuyến tính khác.




**Hình 7.6.** a) Đồ thị mô hình so với dữ liệu quan sát; b) Giá trị quan sát so với dự báo qua mô hình; c) Biến động sai số theo giá trị dự báo (Mô hình: M = a + b×BA)

### 7.4.2 Mô hình tuyến tính đa biến

Trong thực tế, một biến số thuộc y thông thường không chỉ bị ảnh hưởng, chi phối hoặc dự báo thông qua một biến số x mà có thể là nhiều biến số $x_i$. Lúc này cần xem xét thiết lập mối quan hệ theo mô hình tuyến tính đa biến.

Mô hình tuyến tính đa biến được biểu diễn như sau (Mehtatalo, 2013):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + \epsilon_i. \tag{7.9}$$

Trong mô hình trên với n dữ liệu quan sát có một biến phụ thuộc và p biến độc lập, ảnh hưởng. $Y_i$ là biến phụ thuộc với giá trị quan sát thứ i. $X_{ij}$ là giá trị quan sát thứ i của biến độc lập thứ j, với j đi từ 1 đến p. $\beta_j$ là các tham số được ước lượng qua mô hình tuyến tính nhiều lớp, và $\varepsilon_i$ sai số của biến phụ thuộc thứ i, có phân bố chuẩn.

#### 7.4.2.1 Lựa chọn các biến số độc lập ảnh hưởng trong mô hình tuyến tính đa biến

Trong thực tế, đôi khi chưa rõ biến phụ thuộc $Y_i$ bị ảnh hưởng bởi những biến độc lập $X_{ij}$ nào. Vì vậy sử dụng chỉ số Cp của Mallows (1973) sẽ hỗ trợ cho việc xác định số biến số độc lập ảnh hưởng.

Chỉ số Mallow' Cp (1973) được sử dụng để lựa chọn số biến số tham gia mô hình tốt nhất trong trường hợp có nhiều biến độc lập nhưng chưa rõ có ảnh hưởng đến $Y_i$ hay không. Chỉ số Cp càng bé và càng gần với số biến số độc lập p thì đó là các biến độc lập có ảnh hưởng rõ rệt đến biến

phụ thuộc; dựa vào đây để xác định p biến số $X_{ij}$ tham gia mô hình khi có quá nhiều biến số được giả định là có ảnh hưởng đến $Y_i$.

Nếu một mô hình có p biến số độc lập được lựa chọn từ một tập hợp K> p, chỉ tiêu thống kê Cp được tính toán:

$$C_p = \frac{SSE_p}{S^2} - N + 2P, \tag{7.10}$$

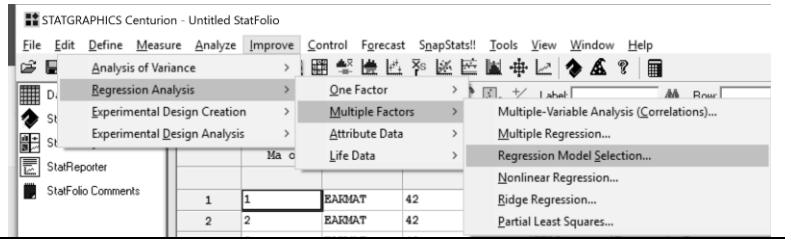
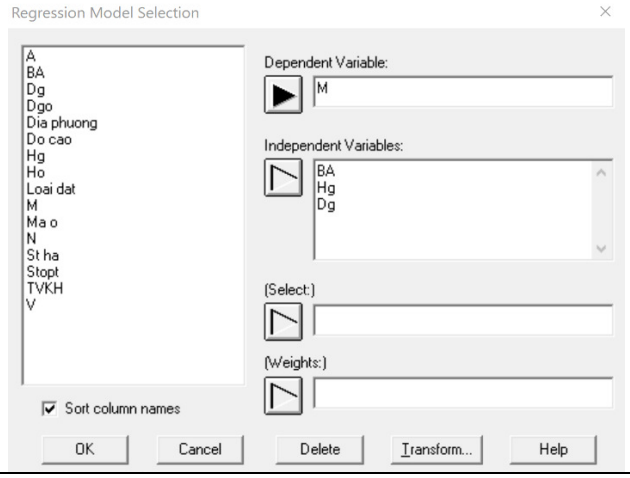
$$SSE_p = \sum_{i=1}^{N}(Y_i - Y_{pi})^2 \tag{7.11}$$

Trong đó: $Y_{pi}$ là giá trị dự đoán từ giá trị quan sát thứ i là của $Y_i$ của mô hình có p biến số độc lập; $S^2$ bình phương trung bình phần dư (residual mean square) sau khi mô hình quan hệ hoàn thành với K biến số độc lập và được ước tính từ sai số trung bình bình phương (mean square error – MSE); N là dung lượng mẫu quan sát.

Phần mềm Statgraphics hỗ trợ tốt cho việc áp dụng chỉ số Cp để tìm số biến số ảnh hưởng tối ưu.

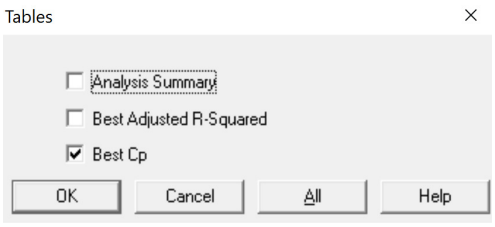
Cũng ví dụ với dữ liệu 120 ô mẫu rừng trồng tếch (trong Dữ liệu 10 ở Phụ lục), như mục trên trữ lượng M được lập quan hệ với biến độc lập là tổng tiết diện ngang BA. Tuy nhiên, có khả năng M còn bị chi phối bởi các nhân tố khác như chiều cao trung bình theo cây có tiết diện trung bình (Hg, m), mật độ (N, cây/ha), đường kính của cây có tiết diện ngang trung bình (Dg, cm),…

Áp dụng phần mềm Statgraphics để tìm số biến số độc lập như BA, Hg, Dg ảnh hưởng đến M rừng trồng tếch như sau:

| | |
|---|---|
| Sử dụng chức năng lựa chọn mô hình của Statgraphics: Improve / Regression Analysis / Multiple Factors / Regression Model Selection … |  |
| Trong hộp thoại Regression Model Selection:<br>Chọn biến độc lập (Dependent Variable): M<br>Chọn các biến độc lập cần xem xét khả năng ảnh hưởng đến M (Independent Variables): BA, Hg. Dg,… |  |

Tìm giá trị Cp tối ưu: Trong nút Tables, kích chọn: Best Cp.

Cũng có thể kích chọn Best Adjusted R-Squared nếu muốn xác định số biến ảnh hưởng dựa vào R cực đại.




**Kết quả tìm số biến số độc lập ảnh hưởng theo chỉ tiêu Cp trong Statgraphics:**

**Regression Model Selection - M**

Dependent variable: M

Independent variables:

A=BA

B=Hg

C=Dg

Number of complete cases: 120

Number of models fit: 8

**Models with Smallest Cp**

| | | *Adjusted* | | *Included* |
|---|---|---|---|---|
| *MSE* | *R-Squared* | *R-Squared* | *Cp* | *Variables* |
| 160.741 | 97.0604 | 96.9844 | **4.0** | **ABC** |
| 164.574 | 96.9644 | 96.9125 | 5.79012 | AB |
| 244.301 | 95.4938 | 95.4167 | 63.8215 | AC |
| 645.397 | 87.9937 | 87.8919 | 357.786 | A |
| 1725.59 | 67.8989 | 67.6268 | 1150.75 | B |
| 1740.33 | 67.8989 | 67.3502 | 1152.75 | BC |
| 2382.24 | 55.6831 | 55.3076 | 1632.8 | C |
| 5330.3 | 0.0 | 0.0 | 3828.13 | |

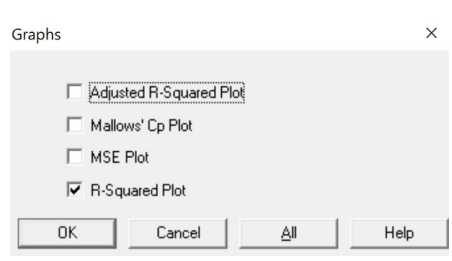
**The StatAdvisor**

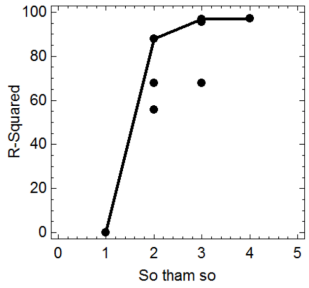
This table shows the models which give the smallest values of Mallows' Cp statistic. Cp is a measure of the Bias in a model, based on a comparison of total mean squared error to the true error variance. UnBiased models have an expected value of approximately p, where p is the number of coefficients in the fitted model (including the constant). You should look for models with Cp values close to p. The plot of Cp, available in the list of Graphical Options, contains a line equal to p to help you select the best models.

Kết quả trên cho thấy giá trị $C_p$ bé nhất là 4 và gần như bằng số tham số của các biến số (3 biến số BA, Hg và Dg cộng với tham số là hằng số = 4). Vì vậy, cả ba biến A=BA, B=Hg và C=Dg đều có ảnh hưởng đến M, do đó cần khảo sát để thiết lập mô hình M theo 3 biến số này.

Xuất ra đồ thị quan hệ R với số tham số của các biến số:

Trong nút Graphs, chọn R-Squared Plot




**Hình 7.7.** Quan hệ giữa R-Squared với số tham số gắn các biến số độc lập của mô hình

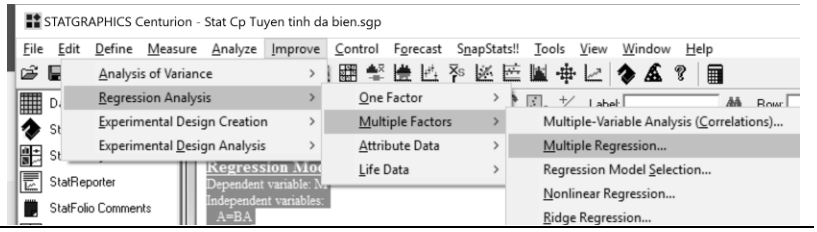
Hình 7.7 chỉ ra rằng $R^2$ đạt cực đại khi số tham số = 4; có nghĩa là có ba tham số gắn với ba biến số độc lập (BA, Hg và Dg) cùng với một tham số là hằng số ảnh hưởng đến M.

### 7.4.2.2 Thiết lập mô hình tuyến tính đa biến trong Statgraphics

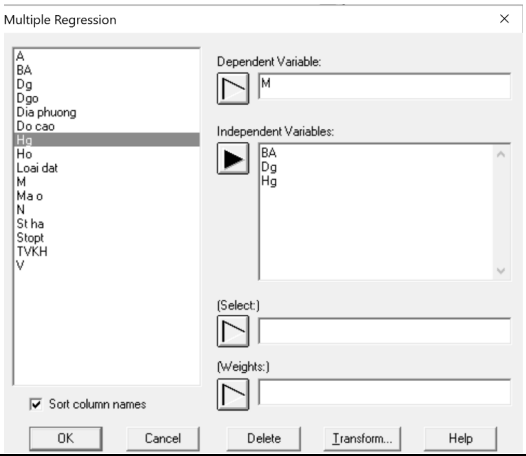
Trên cơ sở sử dụng chỉ tiêu $C_p$ của Mallow, xác định được số biến số ảnh hưởng đến biến độc lập, tiến hành thiết lập mô hình tuyến tính đa biến bằng phương pháp bình phương tối thiểu trong phần mềm Statgraphics.

Cũng với ví dụ trên về trữ lượng rừng trồng tếch cho thấy M bị ảnh hưởng bởi ba biến BA, Hg và Dg, thiết lập quan hệ $M = b_o + b_1BA + b_2Hg + b_3Dg$ trong Statgraphics như sau:

Thiết lập mô hình tuyến tính đa biến: Improve / Regression Analysis / Multiple Factors / Multiple Regression




Trong hộp thoại chọn các biến số độc lập (BA, Hg và Dg) và phụ thuộc (M)



**Kết quả thiết lập mô hình tuyến tính nhiều biến trong Statgraphics:**

**Multiple Regression - M**

Dependent variable: M

Independent variables:

BA

Dg

Hg

| | | *Standard* | *T* | |
|---|---|---|---|---|
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| CONSTANT | -62.7118 | 3.38776 | -18.5113 | 0.0000 |
| BA | 7.04942 | 0.207808 | 33.9227 | 0.0000 |
| Dg | 0.800355 | 0.411108 | 1.94682 | 0.0540 |
| Hg | 4.95232 | 0.629853 | 7.86267 | 0.0000 |

**Analysis of Variance**

| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
|---|---|---|---|---|---|
| Model | 615659. | 3 | 205220. | 1276.71 | 0.0000 |
| Residual | 18646.0 | 116 | 160.741 | | |
| Total (Corr.) | 634305. | 119 | | | |

R-squared = 97.0604 percent
R-squared (adjusted for d.f.) = 96.9844 percent
Standard Error of Est. = 12.6784
Mean absolute error = 10.0074
Durbin-Watson statistic = 1.34124 (P=0.0001)
Lag 1 residual autocorrelation = 0.327498

**The StatAdvisor**

The output shows the results of fitting a multiple linear regression model to describe the relationship between M and 3 independent variables. The equation of the fitted model is

M = -62.7118 + 7.04942*BA + 0.800355*Dg + 4.95232*Hg

Kết quả trên cho thấy biến số Dg có P -Value = 0.0540 > 0.05, có nghĩa là tham số gắn biến này xấp xỉ bằng 0 (chấp nhận giả thuyết $H_i$: $b_i = 0$); vì vậy nên loại biến số Dg ra khỏi mô hình. Lúc này lập mô hình chỉ còn hai biến số độc lập là BA và Hg

**Kết quả lập mô hình tuyến tính đa biến với biến số Dg loại khỏi mô hình:**

**Multiple Regression - M**

Dependent variable: M

Independent variables:

BA

Hg

| | | *Standard* | *T* | |
|---|---|---|---|---|
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| CONSTANT | -61.5752 | 3.37662 | -18.2357 | 0.0000 |
| BA | 7.02578 | 0.209912 | 33.4701 | 0.0000 |
| Hg | 6.00919 | 0.323173 | 18.5944 | 0.0000 |

**Analysis of Variance**

| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
|---|---|---|---|---|---|
| Model | 615050. | 2 | 307525. | 1868.61 | 0.0000 |
| Residual | 19255.2 | 117 | 164.574 | | |
| Total (Corr.) | 634305. | 119 | | | |

R-squared = 96.9644 percent
R-squared (adjusted for d.f.) = 96.9125 percent
Standard Error of Est. = 12.8287
Mean absolute error = 10.153
Durbin-Watson statistic = 1.40396 (P=0.0005)
Lag 1 residual autocorrelation = 0.295258

**The StatAdvisor**
The output shows the results of fitting a multiple linear regression model to describe the relationship between M and 2 independent variables. The equation of the fitted model is

M = -61.5752 + 7.02578*BA + 6.00919*Hg

Since the P-value in the ANOVA table is less than 0.05, there is a statistically significant relationship between the variables at the 95.0% confidence level.

The R-Squared statistic indicates that the model as fitted explains 96.9644% of the variability in M. The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 96.9125%. The standard error of the estimate shows the standard deviation of the residuals to be 12.8287. This value can be used to construct prediction limits for new observations by selecting the Reports option from the text menu. The mean absolute error (MAE) of 10.153 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the P-value is less than 0.05, there is an indication of possible serial correlation at the 95.0% confidence level. Plot the residuals versus row order to see if there is any pattern that can be seen.

In determining whether the model can be simplified, notice that the highest P-value on the independent variables is 0.0000, belonging to BA. Since the P-value is less than 0.05, that term is statistically significant at the 95.0% confidence level. Consequently, you probably don't want to remove any variables from the model.

Kết quả trên đã lập được mô hình: M = -61.5752 + 7.02578*BA + 6.00919*Hg

Mô hình có các chỉ tiêu thống kê và sai số sau tính được trong phần mềm Statgraphics:

Hệ số xác định: R-squared (adjusted for d.f.) = 96.9125 %, với P-Value = 0.000 < 0.0001. $R^2$ như vậy là rất cao và tồn tại rõ rệt, chứng tỏ mối quan hệ giữa M với hai biến số BA và Hg là rất chặt chẽ.

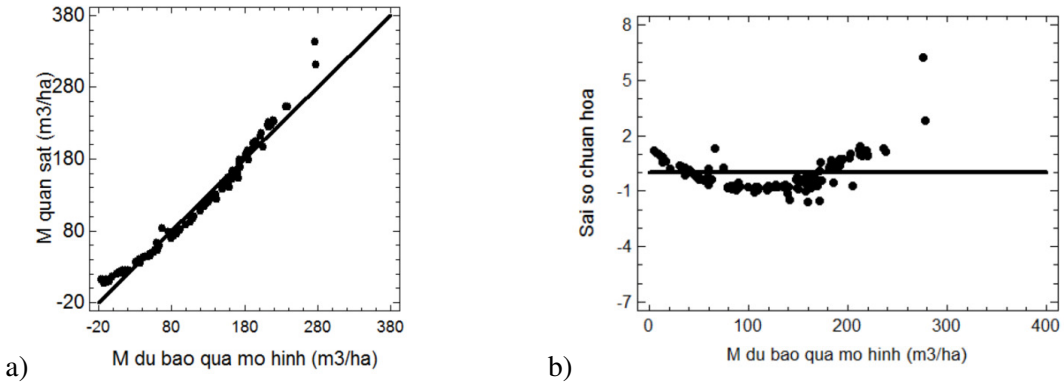
Các tham số gắn các biến độc lập (BA vÀ Hg) đều có P-Value = 0.0000 < 0.0001, do đó tồn tại hay nói khác là ảnh hưởng rõ rệt đến biến phụ thuộc M.

Sai số MAE của mô hình (Mean absolute error) = 10.153 $m^3$/ha.

So sánh kết quả này với mô hình M chỉ theo một biến số là BA ở phần trên cho thấy mô hình có thêm biến Hg đã nâng cao $R^2$ rất nhiều và sai số MAE cũng giảm rõ ràng. Điều này hoàn toàn phù hợp với thực tế, M không chỉ có quan hệ với BA mà còn với Hg, vì các lâm phần có cùng BA nhưng chiều cao trung bình lâm phần khác nhau thì M sẽ thay đổi khác nhau. Trong khi đó biến Dg bị loại khỏi mô hình là do BA đã phản ảnh Dg trong đó, vì vậy Dg là không cần đưa vào trong mô hình.

Chọn xuất ra hai dạng đồ thị ở Hình 7.8, cho thấy M dự đoán và quan sát khá bám sát nhau qua đường chéo, có nghĩa là mô hình khá tốt, tuy nhiên biến động sai số theo M không đều, sai số không rải đều theo dự báo và có một số điểm khi M lớn thì sai số chuẩn hóa > 2 rõ rệt. Điều này cho thấy mô hình tuyến tính có thể chưa mô tả tốt cho mối quan hệ này. Trong trường hợp này nên thử mô hình tổ hợp hoặc phi tuyến tính (sẽ trình bày trong mục tiếp theo).

Cũng cần nói thêm rằng trong thực tế rất ít quan hệ sinh học có dạng đường thẳng cho dù là nhiều biến. Quan hệ đường thẳng là quan hệ khá hiếm vì các mối quan hệ giữa các nhân tố cây, lâm phần, hệ sinh thái rừng luôn là phức tạp và khó có dạng "đường thẳng"; vì vậy mô hình đường thẳng cần áp dụng một cách thận trọng, chỉ chọn lựa khi mà các dạng phi tuyến đã được thử nghiệm nhưng không bằng nó.




**Hình 7.8.** Đồ thị: a) Quan hệ giữa giá trị M dự đoán và quan sát M; b) Biến động sai số chuẩn hóa theo M dự đoán qua mô hình: $M = b_o + b_1BA + b_2Hg + b_3Dg$

Tiếp tục thử nghiệm mô hình tuyến tính đa biến với biến độc lập được tổ hợp: BA×Hg theo dạng: $M = b_o + b_1(BA \times Hg)$ trong Statgraphics.

**Kết quả lập mô hình tuyến tính đa biến, tổ hợp biến:**
**Multiple Regression - M**
Dependent variable: M
Independent variables:

BA*Hg

| | | *Standard* | *T* | |
|---|---|---|---|---|
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| CONSTANT | 9.40253 | 0.918898 | 10.2324 | 0.0000 |
| BA*Hg | 0.52482 | 0.00372794 | 140.78 | 0.0000 |

**Analysis of Variance**

| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
|---|---|---|---|---|---|
| Model | 630551. | 1 | 630551. | 19819.07 | 0.0000 |
| Residual | 3754.21 | 118 | 31.8154 | | |
| Total (Corr.) | 634305. | 119 | | | |

R-squared = 99.4081 percent
R-squared (adjusted for d.f.) = 99.4031 percent
Standard Error of Est. = 5.64051
Mean absolute error = 4.19393
Durbin-Watson statistic = 1.16759 (P=0.0000)
Lag 1 residual autocorrelation = 0.410981

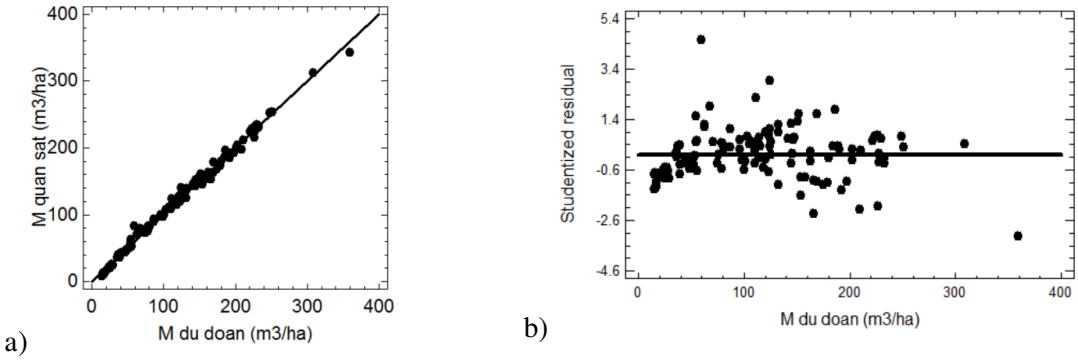
**The StatAdvisor**
The output shows the results of fitting a multiple linear regression model to describe the relationship between M and 1 independent variables. The equation of the fitted model is

M = 9.40253 + 0.52482*BA*Hg

Since the P-value in the ANOVA table is less than 0.05, there is a statistically significant relationship between the variables at the 95.0% confidence level.

The R-Squared statistic indicates that the model as fitted explains 99.4081% of the variability in M. The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 99.4031%. The standard error of the estimate shows the standard deviation of the residuals to be 5.64051. This value can be used to construct prediction limits for new observations by selecting the Reports option from the text menu. The mean absolute error (MAE) of 4.19393 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the P-value is less than 0.05, there is an indication of possible serial correlation at the 95.0% confidence level. Plot the residuals versus row order to see if there is any pattern that can be seen.

In determining whether the model can be simplified, notice that the highest P-value on the independent variables is 0.0000, belonging to BA*Hg. Since the P-value is less than 0.05, that term is statistically significant at the 95.0% confidence level. Consequently, you probably don't want to remove any variables from the model.




**Hình 7.9.** Đồ thị quan hệ $M = b_o + b_1(BA \times Hg)$. a) Quan hệ giữa M dự báo và quan sát; b) Biến động sai số theo M dự báo qua mô hình

Trong trường hợp này, mô hình tổ hợp biến $M = b_o + b_1(BA \times Hg)$ so với mô hình không tổ hợp biến $M = b_o + b_1 BA + b_2 Hg$ cho kết quả tốt hơn theo chỉ tiêu $R^2$ cao hơn và sai số MAE bé hơn nhiều. Đồng thời các đồ thị quan hệ giữa giá trị quan sát với dự đoán và biến động sai số ở Hình 7.9 là được cải thiện rõ rệt. Tuy nhiên, đây cũng là mô hình "đường thẳng", vì vậy tiếp tục thử nghiệm theo mô hình phi tuyến là cần thiết để mô phỏng tốt nhất mối quan hệ này.

#### 7.4.2.3 Thiết lập mô hình tuyến tính đa biến, tổ hợp biến theo chương trình "lm" trong R

Lập các codes để thiết lập mô hình tuyến tính đa biến, tổ hợp biến trong R sẽ cung cấp các kết quả đa dạng và tùy chọn cho người sử dụng

Sử dụng Dữ liệu 10 để minh họa việc lập quan hệ tuyến tính $M = b_o + b_1 \times BA + b_2 \times Hg$ từ dữ liệu 120 ô mẫu của rừng trồng tếch. Sau đây là các codes và kết quả lập và đánh giá mô hình tuyến tính đa biến, hoặc tổ hợp biến theo phương pháp bình phương tối thiểu trong R, theo chương trình "lm" (Chambers, 1992).

**Codes để lập mô hình đa biến theo chương trình "lm":**

```
# Erase memory
rm(list=ls())
# Clean plot window
dev.off()
# Cài đặt chương trình về đồ thị ggplot2:
library(ggplot2)

# Define the working directory (Thu mục)
setwd("E:/1 - Bao Huy 2017/Books All/Textbook for Infomatic Statistic/TextbookDataset Analysis/Dataset")

# Import data from a .txt file (Nhập dữ liệu):
t <- read.table("Teak 120 plots.txt", header=T,sep="\t",stringsAsFactors = FALSE)
library(ggplot2)
# Model development: M = bo + b1BA + b2Hg theo chương trình "lm":
Linear_model1 <- lm(M~BA+Hg, data=t)
# Outputs of the model
```

```
t$Linear_model1.fit <- fitted.values(Linear_model1)
t$Linear_model1.res <- residuals(Linear_model1)
# Summary of linear model: Tóm tắt kết quả mô hình:
summary(Linear_model1)
anova(Linear_model1)
```

Kết quả ước lượng mô hình tuyến tính đa biến trong R như sau:

```
> summary(Linear_model1)

Call:
lm(formula = M ~ BA + Hg, data = t)

Residuals:
    Min      1Q  Median      3Q     Max
-19.890  -9.968  -4.225   9.152  67.005

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -61.5752     3.3766  -18.24   <2e-16 ***
BA            7.0258     0.2099   33.47   <2e-16 ***
Hg            6.0092     0.3232   18.59   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.83 on 117 degrees of freedom
Multiple R-squared:  0.9696,      Adjusted R-squared:  0.9691
F-statistic:  1869 on 2 and 117 DF,  p-value: < 2.2e-16

> anova(Linear_model1)
Analysis of Variance Table

Response: M
           Df Sum Sq Mean Sq F value    Pr(>F)
BA          1 558148  558148 3391.46 < 2.2e-16 ***
Hg          1  56902   56902  345.75 < 2.2e-16 ***
Residuals 117  19255     165
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Mô hình: M = -61.5752 + 7.0258 BA + 6.0092 Hg

Với $R^2$ = 0.9691 với các p-value kiểm tra R và các tham số = 2.2e-16 < 0.0001;

Kết quả này hoàn toàn trùng hợp với kết quả chạy trong Statgraphics.

**Codes để tính AIC và các sai số của mô hình tuyến tính đa biến:**

```
# Outputs of the model: Tính giá trị dự báo và Bias của mô hình:
t$Linear_model1.fit <- fitted.values(Linear_model1)
t$Linear_model1.res <- residuals(Linear_model1)
# Indicators for validation of the model: AIC và các sai số:
AIC(Linear_model1)
Bias <- mean(t$Linear_model1.res)
RMSE <- sqrt(mean((t$Linear_model1.res)^2))
MAPE <- 100*mean(abs(t$Linear_model1.res)/t$M)
Bias
RMSE
MAPE
```

Kết quả tính AIC và sai số của mô hình tuyến tính đa biến:

```
> AIC(Linear_model1)
[1] 957.9107
>
> Bias <- mean(t$Linear_model1.res)
> RMSE <- sqrt(mean((t$Linear_model1.res)^2))
> MAPE <- 100*mean(abs(t$Linear_model1.res)/t$M)
>
> Bias
[1] 2.401725e-16
> RMSE
[1] 12.66728
> MAPE
[1] 22.74413
```

Như vậy mô hình có:

AIC = 957.9

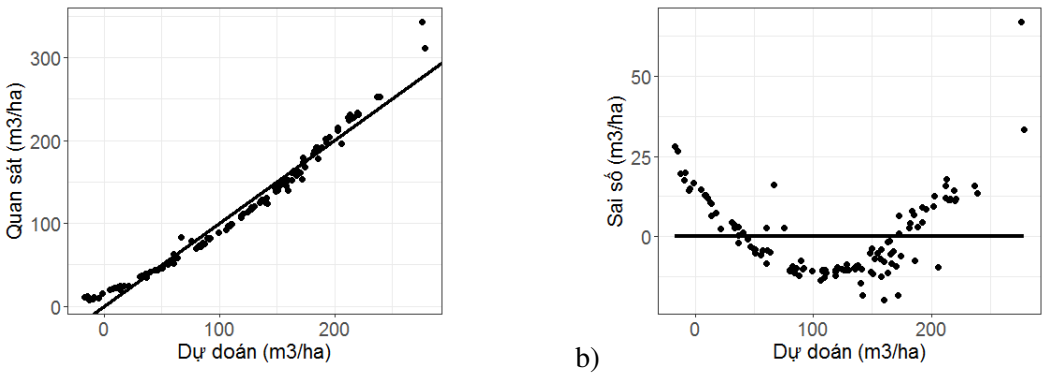
Bias = 2.4e-16

RMSE = 12.7$m^3$/ha

MAPE = 22.7%

**Codes để lập các đồ thị đánh giá mô hình tuyến tính đa biến:**

```
# Residuals and Normal Q-Q Plots: Đồ thị sai số và Q-Q
par(mfrow = c(1, 2))
plot(Linear_model1,1:2)
# Observated and Predicted Values: Đồ thị quan sát và dự báo qua mô hình:
p <- ggplot(t, aes(x=Linear_model1.fit, y=M))
p <- p + geom_point(cex=2)
p <- p + geom_abline(cex = 1.5, intercept = 0, slope = 1, col="black")
p <- p + xlab("Dự đoán (m3/ha)") + ylab("Quan sát (m3/ha)") + theme_bw()+ theme_bw()
p = p + theme(axis.title.y = element_text(size = rel(1.5)))
p = p + theme(axis.title.x = element_text(size = rel(1.5)))
p <- p + theme(plot.title = element_text(size = rel(1.7)))
p = p + theme(axis.text.x = element_text(size=15))
p = p + theme(axis.text.y = element_text(size=15))
p
# Resuduals and Predicted value: Đồ thị biến động sai số theo dự báo qua mô hình:
p <- ggplot(t, aes(x=Linear_model1.fit, y=Linear_model1.res))
p <- p + geom_point(cex = 2)
p <- p + geom_line(cex = 1.5, aes(x=Linear_model1.fit, y=0))
p <- p + xlab("Dự đoán (m3/ha)") + ylab("Sai số (m3/ha)") + theme_bw()+ theme_bw()
p = p + theme(axis.title.y = element_text(size = rel(1.5)))
p = p + theme(axis.title.x = element_text(size = rel(1.5)))
p <- p + theme(plot.title = element_text(size = rel(1.7)))
p = p + theme(axis.text.x = element_text(size=15))
p = p + theme(axis.text.y = element_text(size=15))
p
```




**Hình 7.10.** Đồ thị quan sát so với dự đoán qua mô hình: a) và biến động sai số theo giá trị dự đoán qua mô hình; b) Mô hình: $M = b_o + b_1 \times BA + b_2 \times Hg$

Kết quả ở Hình 7.10 cho thấy giá trị quan sát và dự báo khá bám sát nhau trên đường chéo, tuy nhiên sai số không phân bố đều quanh giá trị dự báo, điều này cho thấy mô hình dự báo cho sai số bị thiên lệch. Có nghĩa là cần tìm mô hình tốt hơn ví dụ như là phi tuyến để mô phỏng quan hệ M = f(BA, Hg) của rừng trồng tếch.

Tương tự như trên sử dụng chương trình "lm" để lập mô hình tuyến tính tổ hợp biến dạng $M = b_o + b_1 (BA \times Hg)$ (Chambers, 1992).

**Codes lập mô hình tuyến tính tổ hợp biến trong R**

```
# Combination of variables: Tổ hợp biến:
t$BAHg = t$BA*t$Hg
# Model development: Lập mô hình tuyến tính tổ hợp biến theo lm
Linear_model1 <- lm(M~BAHg, data=t)
# Summary of linear model: Tóm tắt kết quả mô hình
summary(Linear_model1)
anova(Linear_model1)
```

Kết quả lập mô hình tuyến tính tổ hợp biến theo "lm" trong R:

```
> summary(Linear_model1)

Call:
lm(formula = M ~ BAHg, data = t)

Residuals:
     Min       1Q   Median       3Q      Max
-16.6574  -3.3210  -0.0862   3.0580  23.5529

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 9.402527   0.918898   10.23   <2e-16 ***
BAHg        0.524820   0.003728  140.78   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.641 on 118 degrees of freedom
Multiple R-squared:  0.9941,   Adjusted R-squared:  0.994
F-statistic: 1.982e+04 on 1 and 118 DF,  p-value: < 2.2e-16

> anova(Linear_model1)
Analysis of Variance Table
```

```
Response: M
           Df Sum Sq Mean Sq F value    Pr(>F)
BAHg        1 630551  630551   19819 < 2.2e-16 ***
Residuals 118   3754      32
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Kết quả có mô hình: M = 9.403 + 0.525×(BA×Hg)

Với $R^2$ = 0.994, các p-value kiểm tra R và các tham số = 2.2e-16 < 0.0001

**Codes tính các chỉ tiêu thống kê của mô hình tổ hợp biến và sai số:**

```
# Outputs of the model: Ước tính và  Bias của mô hình:
t$Linear_model1.fit <- fitted.values(Linear_model1)
t$Linear_model1.res <- residuals(Linear_model1)
# Indicators for validation of the model: Các chỉ tiêu thống kê kiểm định mô hình:
AIC(Linear_model1)
Bias <- mean(t$Linear_model1.res)
RMSE <- sqrt(mean((t$Linear_model1.res)^2))
MAPE <- 100*mean(abs(t$Linear_model1.res)/t$M)
Bias
RMSE
MAPE
```

Kết quả:

```
> # Indicators for validation of the model
> AIC(Linear_model1)
[1] 759.7223
>
> Bias <- mean(t$Linear_model1.res)
> RMSE <- sqrt(mean((t$Linear_model1.res)^2))
> MAPE <- 100*mean(abs(t$Linear_model1.res)/t$M)
>
> Bias
[1] 5.600726e-16
> RMSE
[1] 5.59331
> MAPE
[1] 8.422975
```

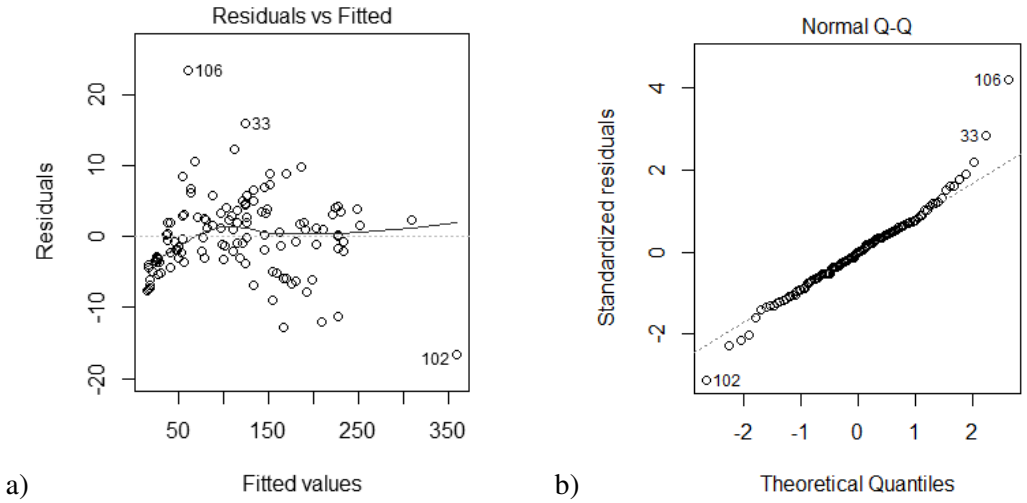
Từ kết quả trên, so sánh các chỉ tiêu thống kê của hai mô hình tuyến tính đa biến và tổ hợp biến ở Bảng 7.1 Từ đây cho thấy mô hình tổ hợp biến BA×Hg đều có các chỉ tiêu thống kê tốt hơn mô hình đa biến: $R^2$ lớn hơn, AIC và các sai số Bias, RMSE, MAPE đều nhỏ hơn.

**Bảng 7.1.** So sánh hai mô hình tuyến tính đa biến và tổ hợp biến

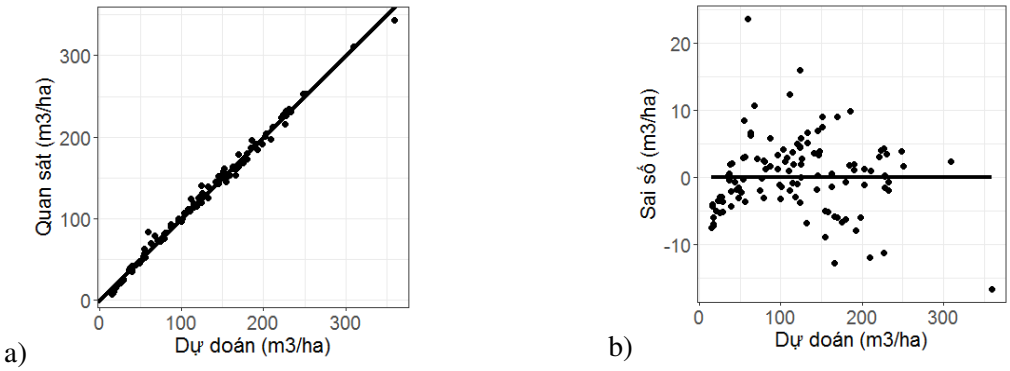
| Chỉ tiêu thống kê, sai số | Mô hình tuyến tính đa biến: $M = b_o + b_1BA + b_2Hg$ | Mô hình tuyến tính tổ hợp biến: $M = b_o + b_1(BA \times Hg)$ |
|---|---|---|
| $R^2$adjusted | 0.969 | 0.994 |
| AIC | 957.9 | 759.7 |
| Bias ($m^3$/ha) | 2.4e-16 | 5.6e-16 |
| RMSE ($m^3$/ha) | 12.7 | 5.59 |
| MAPE (%) | 22.8 | 8.42 |

**Codes vẽ các đồ thị đánh giá mô hình tuyến tính tổ hợp biến:**

```
# Residuals and Normal Q-Q Plots: Đồ thị sai số và Q-Q
par(mfrow = c(1, 2))
plot(Linear_model1,1:2)
# Observated and Predicted Values: Đồ thị quan sát và dự báo:
p <- ggplot(t, aes(x=Linear_model1.fit, y=M))
p <- p + geom_point(cex=2)
p <- p + geom_abline(cex = 1.5, intercept = 0, slope = 1, col="black")
p <- p + xlab("Dự đoán (m3/ha)") + ylab("Quan sát (m3/ha)") + theme_bw()+ theme_bw()
p = p + theme(axis.title.y = element_text(size = rel(1.5)))
p = p + theme(axis.title.x = element_text(size = rel(1.5)))
p <- p + theme(plot.title = element_text(size = rel(1.7)))
p = p + theme(axis.text.x = element_text(size=15))
p = p + theme(axis.text.y = element_text(size=15))
p
# Resuduals and Predicted value: Đồ thị sai số và dự báo:
p <- ggplot(t, aes(x=Linear_model1.fit, y=Linear_model1.res))
p <- p + geom_point(cex = 2)
p <- p + geom_line(cex = 1.5, aes(x=Linear_model1.fit, y=0))
p <- p + xlab("Dự đoán (m3/ha)") + ylab("Sai số (m3/ha)") + theme_bw()+ theme_bw()
p = p + theme(axis.title.y = element_text(size = rel(1.5)))
p = p + theme(axis.title.x = element_text(size = rel(1.5)))
p <- p + theme(plot.title = element_text(size = rel(1.7)))
p = p + theme(axis.text.x = element_text(size=15))
p = p + theme(axis.text.y = element_text(size=15))
p
```




**Hình 7.11.** Đồ thị sai số theo dự báo: a) và Q-Q; b) của mô hình tuyến tính tổ hợp biến ($M = b_o + b_1(BA \times Hg)$)




a) b)

**Hình 7.12.** Đồ thị quan sát với dự báo: a) và sai số theo dự báo; b) của mô hình tuyến tính tổ hợp biến ($M = b_o + b_1(BA \times Hg)$)

Kết quả các đồ thị ở Hình 7.11 và Hình 7.12 cho thấy sai số, Q-Q và quan hệ giữa quan sát với dự báo của mô hình tuyến tính tổ hợp biến là đều tốt hơn mô hình tuyến tính đa biến. Kết quả này phù hợp với so sánh các chỉ tiêu thống kê của hai dạng mô hình ở trên.

## 7.5 Mô hình phi tuyến tính

Mô hình phi tuyến tính có dạng tổng quát (Linton và Hardle, 1998; Mehtatalo, 2013; Picard et al., 2015):

$$y_i = f(x_{1i}, ....., x_{qi}, \beta_{1, ...}\beta_p) + \varepsilon_i \quad (7.12)$$

Trong đó $y_i$ là biến phụ thuộc theo mẫu i, f là dạng hàm phi tuyến, $x_{qi}$ là biến độc lập thứ q của mẫu i, $\beta_p$ là tham số thứ p của mô hình và $\varepsilon_i$ là sai số dự báo $y_i$ của mẫu thứ i. Dạng hàm f phi tuyến tính rất đa dạng, tùy thuộc rất lớn vào mối quan hệ giữa $y_i$ và $x_{qi}$ nghiên cứu.

Mối quan hệ giữa các nhân tố cây rừng, lâm phần, các thành phần của hệ sinh thái rừng thường rất phức tạp, khó diễn tả đầy đủ và chính xác theo mô hình tuyến tính. Vì vậy các mô hình phi tuyến từ một đến nhiều biến, hoặc tổ hợp biến, từ đơn giản ở dạng mũ đến phức tạp như mũ, exp nhiều tầng được áp dụng. Mô hình phi tuyến tính được sử dụng rất đa dạng trong mô hình hóa các mối quan hệ trong hệ sinh thái rừng, mô hình hóa quá trình sinh trưởng, sản lượng, sinh khối, carbon rừng.

Đối với quan hệ chiều cao (H) với đường kính ngang ngực (DBH), các quan hệ phi tuyến tính sau thường được sử dụng (Sola et al., 2014):

$$H = 1.3 + a \times DBH^b \quad (7.13)$$

$$H = 1.3 + a \times exp(-b/(DBH+c)) \text{ (Ratkowsky and Giles, 1990)} \quad (7.14)$$

$$H = 1.3 + a \times (1 - exp(-b \times DBH\text{^}c)) \text{ (Weibull in Yang et al., 1978)} \quad (7.15)$$

Đối với quan hệ chiều cao bình quân tầng trội ($H_o$) theo tuổi (A) để lập biểu cấp năng suất rừng trồng:

- Với rừng trồng tếch (*Tectona grandis* L.f.), Bảo Huy (1995), Bảo Huy và cộng sự (1998) đã sử dụng mô hình phi tuyến tính Schumacher:

$$H_o = a_i \times exp(-b_i A^{-m}) \quad (7.16)$$

- Với rừng trồng trám trắng (*Canarium album* Raeusch) Bảo Huy và Đào Công Khanh (2008) cũng sử dụng hàm Schumacher:

$$H_o = 1617.456 \times exp(-7.72465 \times A^{-0.15}) \quad (7.17)$$

Đối với mô hình sinh trưởng thể tích cây rừng (V) theo tuổi (A), một số mô hình phi tuyến mũ nhiều tầng thường được sử dụng như hàm Schumacher, Korf (Nguyễn Ngọc Lung, 1989). Ví dụ sinh trưởng thể tích loài bằng lăng (*Lagerstroemia calyculata* Kurz) theo hàm Korf (Bảo Huy, 1993):

$$V = 8.73218 \times exp(-25.27369 / A^{0.55211}) \quad (7.18)$$

Đối với phương trình thể tích cây rừng (V) theo các nhân tố điều tra như đường kính ngang ngực (DBH), chiều cao (H), một số dạng phi tuyến power sau được sử dụng; trong đó các biến số có thể là biến đơn hay tổ hợp (Vũ Tiến Hinh, 2012):

$$V = b_o \times DBH^{b1} \times H^{b2} \quad (7.19)$$

$$V = b_o \times (DBH^2 \times H)^{b1} \quad (7.20)$$

Đối với mô hình ước tính sinh khối – carbon cây rừng tự nhiên trên mặt đất (AGB) theo một đến nhiều biến số như DBH, H và khối lượng thể tích gỗ (WD) ở vùng nhiệt đới, Việt Nam, một số mô hình phi tuyến sau thường được sử dụng:

Brown (1997): (7.21)

$$AGB = exp(-2.134 + 2.530 \times \ln(DBH))$$

IPCC (2003): (7.22)

$$AGB = exp\left(-2.289 + 2.649 \times ln(DBH) - 0.021 \times \left(ln(DBH)\right)^2\right)$$

Chave et al. (2014): (7.23)

$$AGB = 0.0673 \times (WD \times DBH^2 \times H)^{0.976}$$

Huy et al., 2016b: AGB cho rừng lá rộng thường xanh vùng Nam Trung Bộ, trong đó CA là diện tích tán lá cây rừng:

$$AGB = 0.61345 \times (DBH^2 \times H \times WD)^{0.86983} \times CA^{0.18834} \quad (7.24)$$

Huy et al. 2016c thiết lập mô hình AGB cho rừng khộp Việt Nam:

$$AGB = 0.06203 \times DBH^{2.26430} \times H^{0.51415} \times WD^{0.79456} \quad (7.25)$$

Việc ước lượng các mô hình phi tuyến tính có thể áp dụng một trong các phương pháp: tuyến tính hóa và bình phương tối thiểu (chương trình "lm" trong R, Chambers, 1992) hoặc phi tuyến bình phương tối thiểu (chương trình "nls" trong R: Nonlinear Least Squares) (Bates và Watts, 1988) hoặc phương pháp phi tuyến của Marquardt (StatPoint-Inc., 2005) hoặc phi tuyến ảnh hưởng phức hợp hợp lý tối đa (chương trình "nlme" trong R: Nonlinear Mixed-Effects Models - Maximum Likelihood) (Davidian và Giltinan, 1995; Pinheiro et al., 2014).

### 7.5.1 Lựa chọn các biến số ảnh hưởng trong mô hình phi tuyến

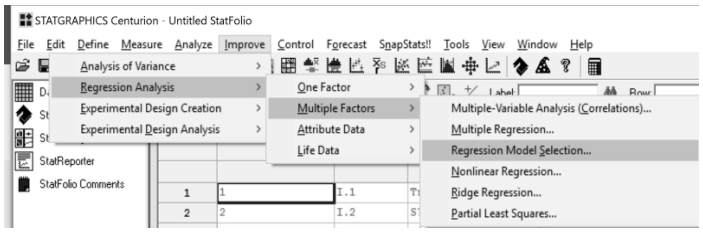
Cũng như mô hình tuyến tính đa biến, mối quan hệ phi tuyến cũng có dạng từ một biến số ảnh hưởng cho đến nhiều biến. Ví dụ đối với mô hình ước tính sinh khối cây rừng trên mặt đất (AGB), trong khi Chave et al. (2014) lập quan hệ với tối đa 3 biến số là DBH, H và WD, trong khi đó Huy et al. (2016b) đã phát hiện thêm biến CA đã làm tăng độ tin cậy của mô hình.

Chỉ tiêu Mallow's Cp (1973) được sử dụng để xác định số biến số độc lập $x_i$ ảnh hưởng có ý nghĩa tới biến phụ thuộc y; trong đó, khi Cp tiến đến giá trị p gần bằng nhất với số biến số độc lập thì đó là các biến tối ưu ảnh hưởng đến y.

Sử dụng Dữ liệu 110 cây mẫu xác định AGB và các nhân tố DBH, H, WD và CA của rừng lá rộng thường xanh vùng Nam Trung Bộ (Dữ liệu 11) để xác định số biến số tối ưu ảnh hưởng đến AGB cây rừng tự nhiên theo chỉ tiêu Cp của Mallow (1073). Thực hiện trong phần mềm Statgraphics như sau:

Trong Statgraphics, vào chức năng sử dụng Mallow's Cp:

Improve / Regression Analysis / Multiple Factors / Regression Selections

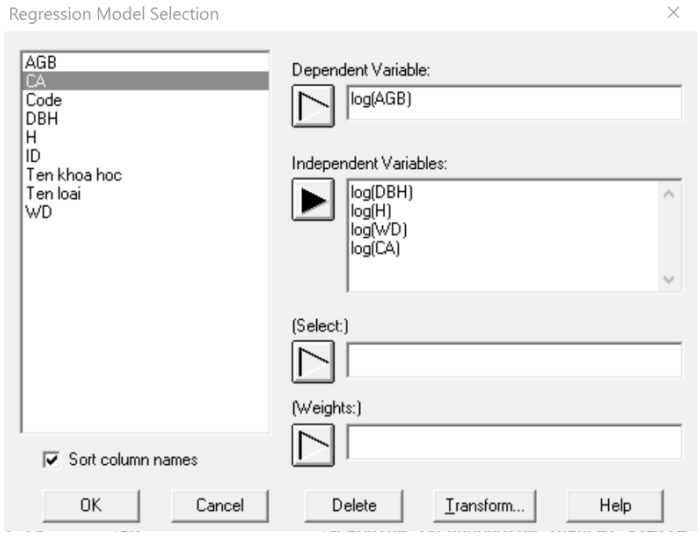



Trong hộp thoại Regression Model Selection:

Biến phụ thuộc (Dependent Variable): log(AGB)

Các biến độc lập (Independent Variables): Nhập vào 4 biến để tìm các biến tối ưu ảnh hưởng đến AGB, gồm log của DBH, H, WD và CA

Các biến phụ thuộc và độc lập đều được logarit do mô hình quan hệ sử dụng dạng hàm mũ, tuyến tính hóa để đánh giá mức độ quan hệ theo phương pháp bình phương tối thiểu đa biến



Kết quả xác định chỉ số Cp Mallow tối ưu trong Statgraphics:

**Regression Model Selection - log(AGB)**

Dependent variable: log(AGB)

Independent variables:

A=log(DBH)
B=log(H)
C=log(WD)
D=log(CA)

Number of complete cases: 110
Number of models fit: 16

**Models with Smallest Cp**

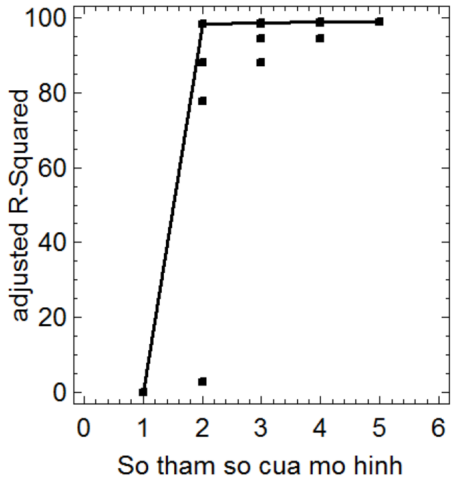
| | | *Adjusted* | | *Included* |
|---|---|---|---|---|
| *MSE* | *R-Squared* | *R-Squared* | *Cp* | *Variables* |
| 0.0465823 | 98.9146 | 98.8733 | 5.0 | ABCD |
| 0.0507418 | 98.8064 | 98.7727 | 13.4652 | ABC |
| 0.0568574 | 98.6626 | 98.6247 | 27.3815 | ABD |
| 0.05949 | 98.6007 | 98.561 | 33.372 | ACD |
| 0.0603782 | 98.5664 | 98.5396 | 34.6893 | AC |
| 0.0631657 | 98.5002 | 98.4721 | 41.0923 | AB |
| 0.0726672 | 98.2746 | 98.2423 | 62.9174 | AD |
| 0.0746345 | 98.2113 | 98.1947 | 67.0385 | A |
| 0.231369 | 94.5577 | 94.4036 | 424.489 | BCD |
| 0.232108 | 94.4887 | 94.3857 | 429.156 | BD |
| 0.489869 | 88.3684 | 88.151 | 1021.23 | BC |
| 0.494576 | 88.1469 | 88.0371 | 1040.66 | B |
| 0.922546 | 77.8901 | 77.6853 | 2032.9 | D |
| 4.01801 | 3.70342 | 2.81178 | 9209.68 | C |
| 4.13426 | 0.0 | 0.0 | 9565.94 | |

**The StatAdvisor**

This table shows the models which give the smallest values of Mallows' Cp statistic. Cp is a measure of the Bias in a model, based on a comparison of total mean squared error to the true error variance. UnBiased models have an expected value of approximately p, where p is the number of coefficients in the fitted model (including the constant). You should look for models with Cp values close to p. The plot of Cp, available in the list of Graphical Options, contains a line equal to p to help you select the best models.

Kết quả cho thấy Cp bé nhất = 5 và bằng số biến số là 4 (A, B, C, D) + 1 tham số là hằng số. Như vậy cả 4 biến độc lập DBH, H, WD và CA ảnh hưởng đến việc ước tính AGB qua mô hình power. Mô hình 4 biến cũng có $R^2_{adj.}$ cao nhất = 98.8733 so với các mô hình ít biến hơn (Hình 7.13).

Trên cơ sở này Huy et al. (2016b) đã lập mô hình ước tính AGB theo 4 biến số DBH, H, WD và CA cho cây rừng lá rộng thường xanh vùng Nam Trung Bộ; lần đầu tiên đưa biến CA vào mô hình AGB ở Việt Nam và đã chỉ ra rằng, nó có độ tin cậy cao nhất so với các mô hình có số biến số ít hơn hoặc so với mô hình chung cho rừng nhiệt đới của Brown (1997) và IPCC (2003) với một biến DBH hoặc của Chave et al. (2005, 2014) với ba biến DBH, H và WD.




**Hình 7.13.** Quan hệ $R^2_{adj.}$ với số tham số của mô hình quan hệ AGB = f(DBH, H, WD, CA). $R^2$ đạt max với số tham số của mô hình = 5 (Bốn biến số + một hằng số của mô hình)

### 7.5.2 Ước lượng mô hình phi tuyến theo phương pháp tuyến tính hóa

Phương pháp ước lượng mô hình phi tuyến tính phổ biến là tuyến tính hóa nó, thường là logarit hóa nếu mô hình dạng hàm mũ, từ đó ước lượng các tham số của mô hình tuyến tính từ một đến nhiều biến theo phương pháp bình phương tối thiểu.

Một số hàm phi tuyến và được tuyến tính hóa thông qua logarit phổ biến (Picard et al., 2012):

$y = b_o \times x_1^{b1} \times x_2^{b2} \ldots\ldots \times x_n^{bn} \times \varepsilon$, tuyến tính hóa: $\log(y) = \log(b_o) + b_1 \times \log(x_1) + b_2 \times \log(x_2) + \ldots\ldots. b_n \times \log(x_n) + \log(\varepsilon)$ (7.26)

$y = b_o \times \exp(x_1^{b1} \times x_2^{b2} \ldots\ldots \times x_n^{bn}) \times \varepsilon$, tuyến tính hóa: $\log(y) = \log(b_o) + b_1 \times x_1 + b_2 \times x_2 + \ldots\ldots. b_n \times x_n + \log(\varepsilon)$ (7.27)

Trong đó y là biến phụ thuộc, $x_i$ là biến độc lập thứ i đến n, $b_i$ là tham số theo biến số thứ i đến n của mô hình và $\varepsilon$ là sai số dự báo y.

Đối với mô hình tuyến tính hóa dạng logarit, khi chuyển về lại dạng nguyên thủy là phi tuyến để ước tính y, cần sử dụng một hệ số điều chỉnh (Corection Factor – CF) (Basuki et al., 2009; Bảo Huy, 2013). Sau khi tuyến tính hóa dạng logarit để lượng các tham số của mô hình theo phương pháp bình phương tối thiểu, cần tính giá trị CF; sau đó mô hình được chuyển về dạng nguyên thủy phi tuyến và khi dự đoán y cần nhân thêm hệ số CF. Ví dụ với mô hình power sau khi tính toán các tham số, dự đoán y sẽ là:

$$y = CF \times b_o \times x_1^{b1} \times x_2^{b2} \ldots\ldots \times x_n^{bn} \quad (7.28)$$

Trong đó CF được tính:

$$CF = exp(RSE^2/2) \quad (7.29)$$

CF luôn lớn hơn 1. Trong đó RSE (Residual standard error) là sai tiêu chuẩn của phần dư (sai lệch giữa quan sát và dự đoán). Khi RSE càng lớn thì CF càng lớn, có nghĩa mô hình càng có độ tin

cây thấp. Mô hình tốt khi CF càng tiến dần đến 1 (Chave et al., 2005; Basuki et al., 2009; Bảo Huy, 2013).

#### 7.5.2.1 Ước lượng mô hình phi tuyến được tuyến tính hóa trong Statgraphics

Phần mềm Statgraphics có khả năng hỗ trợ để ước lượng các mô hình được tuyến tính hóa (đổi biến số) từ một cho đến n biến số $x_i$.

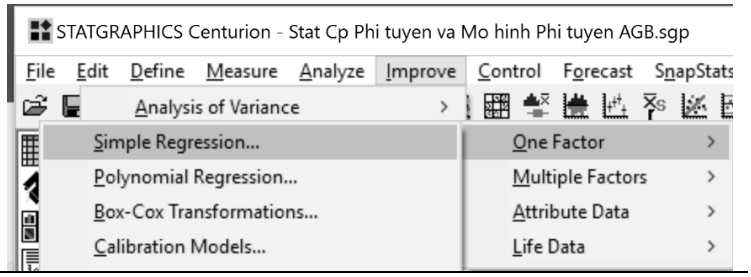
Sử dụng dữ liệu AGB theo các nhân tố điều tra của 110 cây rừng mẫu của rừng lá rộng thường xanh vùng Nam Trung Bộ (Dữ liệu 11, Huy et al., 2016b), ước lượng mô hình AGB theo một nhân tố DBH dạng hàm power được logarit hóa như sau:

$$AGB = b_o \times DBH^{b1} \times \varepsilon, \text{ logarit hóa: } \log(AGB) = \log(b_o) + b_1 \times \log(DBH) + \log(\varepsilon) \quad (7.30)$$

Tiến hành ước lượng mô hình phi tuyến một biến được logarit hóa trong chương trình Statgraphics theo phương pháp bình phương tối thiểu:

Thực hiện chương trình ước tính mô hình tuyến tính một biến trong Statgraphics:

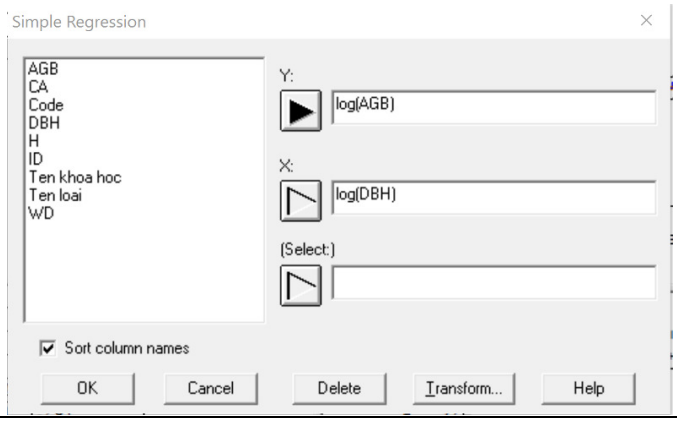
Improve / Simple Regression... / One Factor




Trong hộp thoại Simple Regression: Nhập và các biến Y và X và logarit:

Trong biến Y: log(AGB)

Trong biến X: log(DBH)




Kết quả ước lượng mô hình tuyến tính hóa một nhân tố: $\log(AGB) = \log(b_o) + b_1\log(DBH)$:

**Simple Regression - log(AGB) vs. log(DBH)**
Dependent variable: log(AGB)
Independent variable: log(DBH)
Linear model: Y = a + b*X

**Coefficients**

| | *Least Squares* | *Standard* | *T* | |
|---|---|---|---|---|
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| Intercept | -2.23646 | 0.0972079 | -23.007 | 0.0000 |
| Slope | 2.4715 | 0.032095 | 77.0057 | 0.0000 |

**Analysis of Variance**

| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
|---|---|---|---|---|---|
| Model | 442.574 | 1 | 442.574 | 5929.88 | 0.0000 |
| Residual | 8.06053 | 108 | 0.0746345 | | |
| Total (Corr.) | 450.634 | 109 | | | |

Correlation Coefficient = 0.991016
R-squared = 98.2113 percent
R-squared (adjusted for d.f.) = 98.1947 percent
Standard Error of Est. = 0.273193
Mean absolute error = 0.198873
Durbin-Watson statistic = 2.16837 (P=0.8102)
Lag 1 residual autocorrelation = -0.0938366

**The StatAdvisor**

The output shows the results of fitting a linear model to describe the relationship between log(AGB) and log(DBH). The equation of the fitted model is

log(AGB) = -2.23646 + 2.4715*log(DBH)

Since the P-value in the ANOVA table is less than 0.05, there is a statistically significant relationship between log(AGB) and log(DBH) at the 95.0% confidence level.

The R-Squared statistic indicates that the model as fitted explains 98.2113% of the variability in log(AGB). The correlation coefficient equals 0.991016, indicating a relatively strong relationship between the variables. The standard error of the estimate shows the standard deviation of the residuals to be 0.273193. This value can be used to construct prediction limits for new observations by selecting the Forecasts option from the text menu.

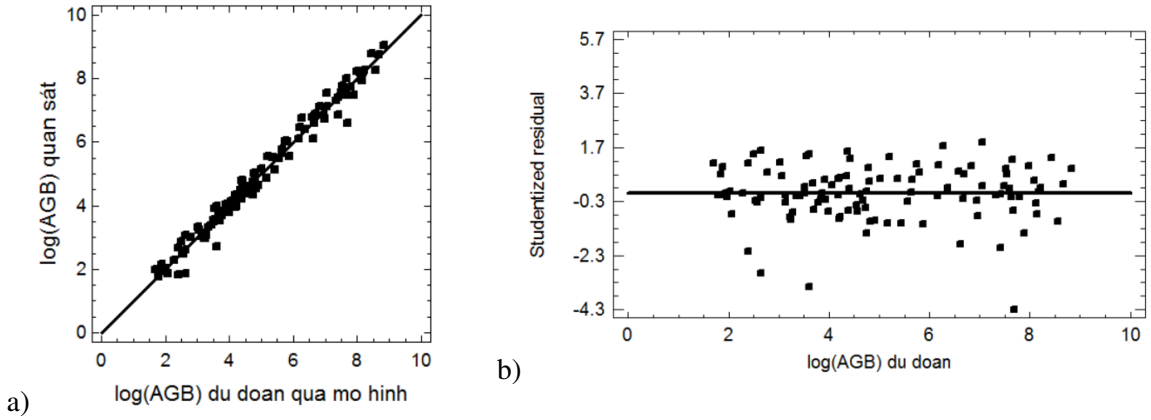
Kết quả mô hình tuyến tính hóa theo logarit một biến số:

log(AGB) = -2.23646 + 2.4715×log(DBH)

Với R-squared (adjusted) = 98.1947 % là cao và P-value kiểm tra R (trong bảng ANOVA) cũng như các tham số mô hình (trong bảng Coefficients) = 0.0000 < 0.001, có nghĩa là tồn tại R và các tham số.

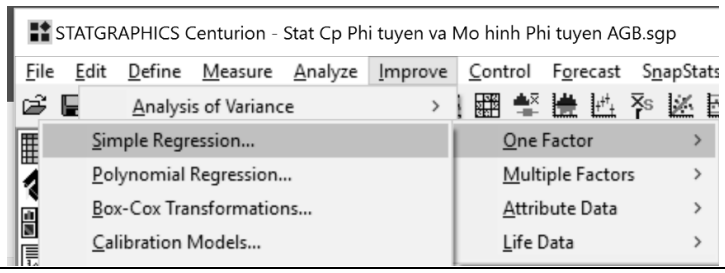
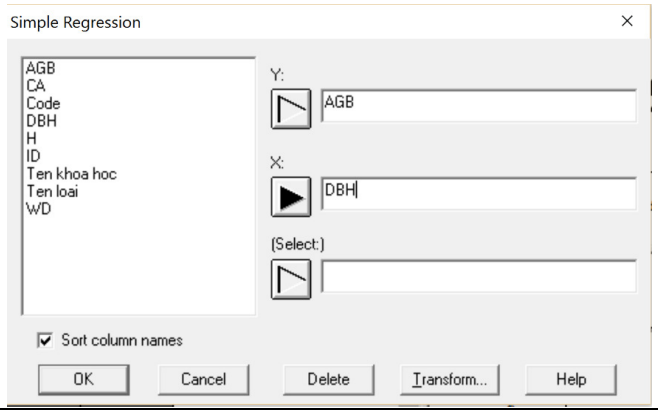
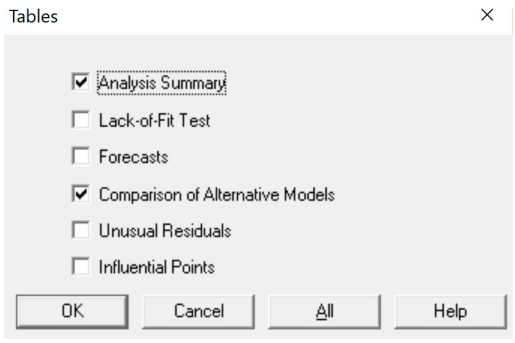
Statgraphics cũng tính sai số tuyệt đối trung bình MAE (Mean absolute error) = 0.198873, tuy nhiên lưu ý sai số này là của biến Y đã được tuyến tính hóa Y = log(y). Vì vậy đây không là sai số thực của sinh khối AGB. Tóm lại, các mô hình tuyến tính hóa dạng logarit, các sai số không thể sử dụng để đánh giá độ tin cậy vì đã bị logarit hóa và sai số của nó nhỏ hơn rất nhiều so với giá trị thực. Đây là hạn chế của phương pháp tuyến tính hóa trong đánh giá sai số của mô hình.

Hình 7.14 cho thấy quan hệ giữa log(AGB) quan sát so với dự đoán là chặt chẽ và biến động sai số rải đều quanh giá trị log(AGB) dự đoán và hầu hết nằm trong phạm vi -2 đến +2 (với độ tin cậy 95%); cho thấy mô hình mũ tuyến tính hóa logarit mô tả tốt quan hệ AGB theo biến số DBH của cây rừng lá rộng thường xanh ở Nam Trung Bộ.




**Hình 7.14.** log(AGB) = -2.23646 + 2.4715×log(DBH): a) log(AGB) quan sát so với log(AGB) dự đoán qua mô hình; b) biến động sai số chuẩn hóa theo giá trị dự đoán log(AGB)

Ngoài ra trong phần mềm Stagraphics có chương trình lập sẵn để hỗ trợ cho việc so sánh nhiều mô hình phi tuyến với một biến số độc lập tốt nhất dựa vào hệ số xác định $R^2$ đạt max. Cách tiến hành như sau

| | |
|---|---|
| Thực hiện lệnh lập mô hình một tuyến tính nhân tố:<br>Improve / Simple Regression / On Factor |  |
| Trong hộp thoại Simple Regression: Nhập và các biến Y và X và logarit:<br>Trong biến Y: AGB<br>Trong biến X: DBH |  |
| Trong nút Table của kết quả mô hình tuyến tính một biến số , chọn so sánh với các mô hình khác: Comnparision of Alternative Models |  |

Kết quả Statgraphics sẽ cho ra kết quả so sánh nhiều mô hình phi tuyến một biến độc lập và so với mô hình tuyến tính một biến vừa lập thông qua hệ số xác định $R^2$, mô hình có $R^2$ cao nhất được sắp lên trên cùng trong bảng so sánh các mô hình được trình bày ở Bảng 7.2.

**Bảng 7.2.** Kết quả so sánh các mô hình tuyến tính – phi tuyến một biến số dựa và hệ số xác định $R^2$. $R^2$ càng cao được xếp lên trên, và trên cùng là mô hình có $R^2$ cao nhất của mối quan hệ

Comparison of Alternative Models

| ***Model*** | ***Correlation*** | ***R-Squared*** |
|---|---|---|
| Multiplicative | 0.9910 | 98.21% |
| Square root-Y | 0.9801 | 96.05% |
| Logarithmic-Y square root-X | 0.9760 | 95.26% |
| Square root-Y squared-X | 0.9688 | 93.87% |
| Squared-X | 0.9571 | 91.60% |
| Double square root | 0.9560 | 91.38% |
| Exponential | 0.9374 | 87.87% |
| S-curve model | -0.9260 | 85.75% |
| Double reciprocal | 0.9058 | 82.06% |
| Square root-Y logarithmic-X | 0.9033 | 81.60% |
| Linear | 0.8913 | 79.45% |
| Logarithmic-Y squared-X | 0.8341 | 69.57% |
| Square root-X | 0.8294 | 68.79% |
| Double squared | 0.7904 | 62.47% |
| Reciprocal-Y logarithmic-X | -0.7495 | 56.18% |
| Logarithmic-X | 0.7462 | 55.69% |
| Square root-Y reciprocal-X | -0.7298 | 53.26% |
| Squared-Y | 0.6630 | 43.96% |
| Squared-Y square root-X | 0.5834 | 34.04% |
| Reciprocal-X | -0.5498 | 30.23% |
| Squared-Y logarithmic-X | 0.4972 | 24.72% |
| Reciprocal-Y squared-X | -0.4132 | 17.07% |
| Squared-Y reciprocal-X | -0.3353 | 11.24% |
| Reciprocal-Y | <no fit> | |
| Reciprocal-Y square root-X | <no fit> | |
| Logistic | <no fit> | |
| Log probit | <no fit> | |

Tư vấn của Statgraphics: The StatAdvisor

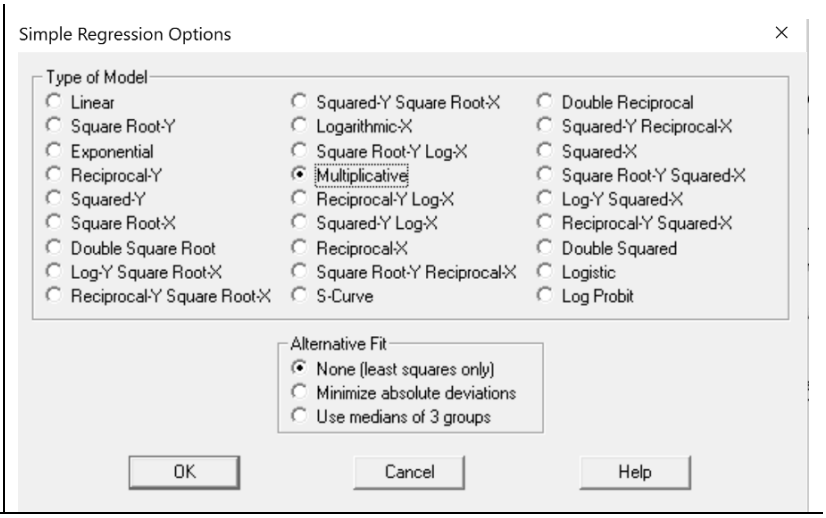
This table shows the results of fitting several curvilinear models to the data. Of the models fitted, the multiplicative model yields the highest R-Squared value with 98.2113%. This is 18.7611% higher than the currently selected linear model. To change models, select the Analysis Options dialog box.

Trong trường hợp quan hệ đang khảo sát là AGB = f(DBH) thì mô hình dạng Multiplicative (Power) có $R^2$ cao nhất. Kết quả này phù hợp với việc lựa chọn mô hình power để thiết lập mối quan hệ này ở phần trên.

Giả sử đang thử nghiệm với mô hình tuyến tính đơn, để tiếp tục thực hiện lập mô hình power như tư vấn của Statpraphics, tiến hành như sau:

Trong cửa sổ đồ thị, kích chuột phải và chọn trong menu: Analysis Options...

Trong hộp thoại: Simple Regression Options: Chọn hàm đã được tư vấn có $R^2$ cao nhất, trường hợp này là Multiplicative




Kết quả thiết lập hàm dạng Multiplicative (Power) được lựa chọn trên cơ sở với $R^2$ cao nhất của mô hình một biến số độc lập trong Statgraphics:

**Simple Regression - AGB vs. DBH**
Dependent variable: AGB
Independent variable: DBH
Multiplicative model: Y = a*X^b

**Coefficients**

| | *Least Squares* | *Standard* | *T* | |
|---|---|---|---|---|
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| Intercept | -2.23646 | 0.0972079 | -23.007 | 0.0000 |
| Slope | 2.4715 | 0.032095 | 77.0057 | 0.0000 |

NOTE: intercept = ln(a)

**Analysis of Variance**

| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
|---|---|---|---|---|---|
| Model | 442.574 | 1 | 442.574 | 5929.88 | 0.0000 |
| Residual | 8.06053 | 108 | 0.0746345 | | |
| Total (Corr.) | 450.634 | 109 | | | |

Correlation Coefficient = 0.991016
R-squared = 98.2113 percent

R-squared (adjusted for d.f.) = 98.1947 percent
Standard Error of Est. = 0.273193
Mean absolute error = 0.198873
Durbin-Watson statistic = 2.16837 (P=0.8102)
Lag 1 residual autocorrelation = -0.0938366

**The StatAdvisor**

The output shows the results of fitting a multiplicative model to describe the relationship between AGB and DBH. The equation of the fitted model is

AGB = exp(-2.23646 + 2.4715*ln(DBH))

or

ln(AGB) = -2.23646 + 2.4715*ln(DBH)

Since the P-value in the ANOVA table is less than 0.05, there is a statistically significant relationship between AGB and DBH at the 95.0% confidence level.

The R-Squared statistic indicates that the model as fitted explains 98.2113% of the variability in AGB. The correlation coefficient equals 0.991016, indicating a relatively strong relationship between the variables. The standard error of the estimate shows the standard deviation of the residuals to be 0.273193. This value can be used to construct prediction limits for new observations by selecting the Forecasts option from the text menu.

The mean absolute error (MAE) of 0.198873 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the P-value is greater than 0.05, there is no indication of serial autocorrelation in the residuals at the 95.0% confidence level.

Kết quả này hoàn toàn trùng hợp với thiết lập mô hình được logarits hóa ở phần trên. Tuy nhiên, ở phần trên mô hình power do người sử dụng quyết định, trong khi đó ở kết quả này mô hình được lựa chọn Multiplicative là nhờ sự so sánh của nhiều mô hình một biến số được lập trình sẵn trong Statgraphics. Vì vậy, trong nhiều trường hợp, khi chưa biết dạng quan hệ thích hợp với một biến số độc lập, chương trình này là một công cụ hỗ trợ hữu ích để phát hiện nhanh mô hình có $R^2$ cao nhất. Tuy nhiên, cũng cần lưu ý rằng không phải mô hình tối ưu bao giờ cũng có $R^2$ cao nhất, nó chỉ là một trong những chỉ tiêu thống kê để tham khảo đầu tiên, ngoài ra còn cần quan tâm đến các chỉ tiêu như AIC, các loại sai số như Bias, RMSE, MAPE,… và đặc biệt là các đồ thị biểu diễn quan hệ giữa giá trị quan sát với dự đoán và đồ thị biến động residials theo giá trị dự đoán khi so sánh các mô hình với nhau.

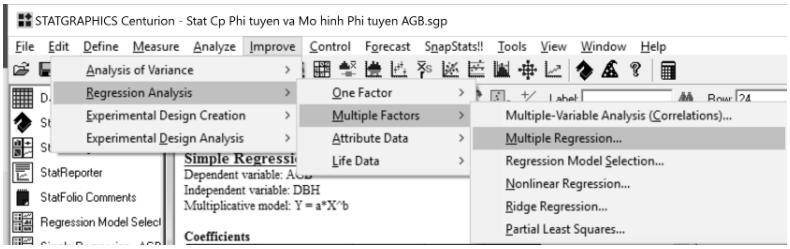
Ngoài ra, trong Statgraphics để thiết lập mô hình tuyến tính hóa qua logarit với nhiều biến số độc lập, thì cách tiến hành cũng tương tự, chỉ đưa thêm biến độc lập được lấy log. Ví dụ, tiếp tục sử dụng Dữ liệu 11 (Huy et al., 2016b), ước lượng mô hình AGB theo hai nhân tố DBH và H dạng hàm power được logarit hóa như sau:

$$AGB = b_o \times DBH^{b1} H^{b2} \times \varepsilon, \text{ logarit hóa: } log(AGB) = log(b_o) + b_1 \times log(DBH) + b_{2\times} log(H) + log(\varepsilon) \quad (7.31)$$

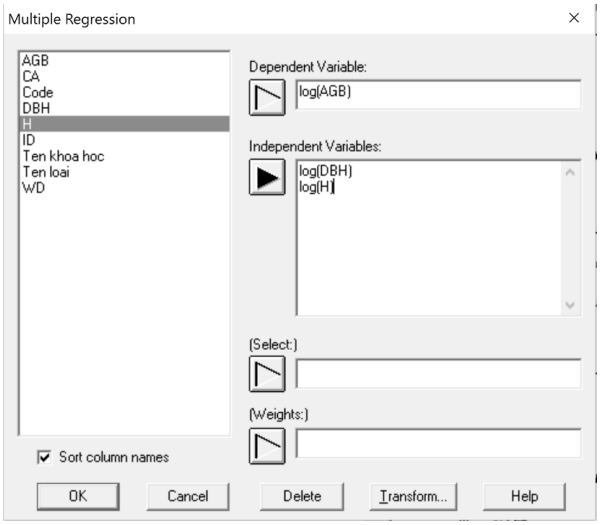
Tiến hành ước lượng mô hình phi tuyến hai biến được logarit hóa trong chương trình Statgraphics theo phương pháp bình phương tối thiểu:

Thực hiện chương trình tuyến tính đa biến trong Statgraphics:

Improve / Regression Analysis / Multiple Factors / Multiple Regression…




Nhập các biến phụ thuộc và độc lập và lấy log trong hộp thoại Multiple Regression




Kết quả thiết lập mô hình phi tuyến nhiều biến số được tuyến tính hóa theo hàm logarit và theo phương pháp bình phương tối thiểu trong Statgraphics:

**Multiple Regression - log(AGB)**

Dependent variable: log(AGB)

Independent variables:

log(DBH)

log(H)

| | | *Standard* | *T* | |
|---|---|---|---|---|
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| CONSTANT | -2.85846 | 0.163615 | -17.4707 | 0.0000 |
| log(DBH) | 2.1401 | 0.0787451 | 27.1776 | 0.0000 |
| log(H) | 0.579318 | 0.12761 | 4.53974 | 0.0000 |

**Analysis of Variance**

| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
|---|---|---|---|---|---|
| Model | 443.876 | 2 | 221.938 | 3513.58 | 0.0000 |
| Residual | 6.75873 | 107 | 0.0631657 | | |
| Total (Corr.) | 450.634 | 109 | | | |

R-squared = 98.5002 percent
R-squared (adjusted for d.f.) = 98.4721 percent
Standard Error of Est. = 0.251328
Mean absolute error = 0.191371
Durbin-Watson statistic = 1.96921 (P=0.4363)
Lag 1 residual autocorrelation = 0.00949601

**The StatAdvisor**
The output shows the results of fitting a multiple linear regression model to describe the relationship between log(AGB) and 2 independent variables. The equation of the fitted model is

log(AGB) = -2.85846 + 2.1401*log(DBH) + 0.579318*log(H)

Since the P-value in the ANOVA table is less than 0.05, there is a statistically significant relationship between the variables at the 95.0% confidence level.

The R-Squared statistic indicates that the model as fitted explains 98.5002% of the variability in log(AGB). The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 98.4721%. The standard error of the estimate shows the standard deviation of the residuals to be 0.251328. This value can be used to construct prediction limits for new observations by selecting the Reports option from the text menu. The mean absolute error (MAE) of 0.191371 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the P-value is greater than 0.05, there is no indication of serial autocorrelation in the residuals at the 95.0% confidence level.

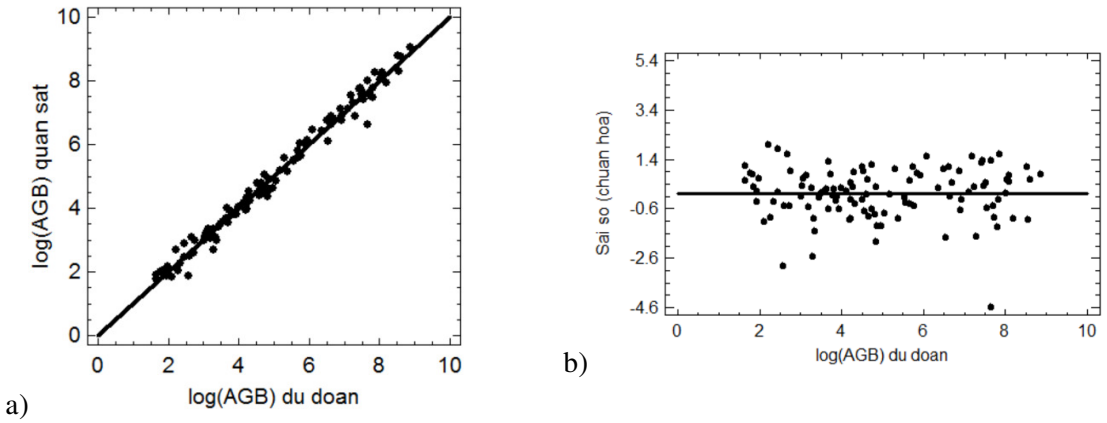
In determining whether the model can be simplified, notice that the highest P-value on the independent variables is 0.0000, belonging to log(H). Since the P-value is less than 0.05, that term is statistically significant at the 95.0% confidence level. Consequently, you probably don't want to remove any variables from the model.

Kết quả có được mô hình:

log(AGB) = -2.85846 + 2.1401*log(DBH) + 0.579318*log(H)

Với $R^2_{adj.}$ = 98.4721 % và P-Value = 0.0000 < 0.0001 khi kiểm tra tồn tại của R và các tham số của mô hình.

Hình 7.15 cho thấy giá trị log(AGB) dự báo bám sát quan sát và sai số được chuẩn hóa biến động rãi đều theo dự đoán và nằm hầu hết trong phạm vì -2 đến +2; tốt hơn so với mô hình một biến đã thực hiện ở phần trên.




**Hình 7.15.** Mô hình log(AGB) = -2.85846 + 2.1401*log(DBH) + 0.579318*log(H). a) Quan sát so với dự đoán; b) Biến động sai số theo dự đoán

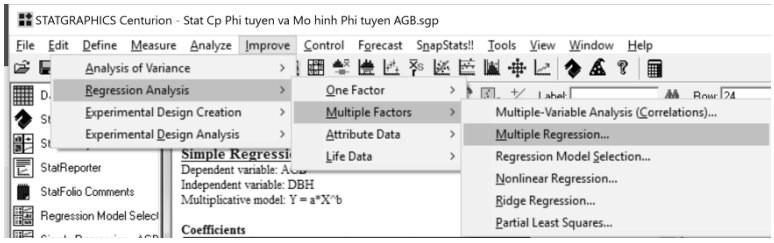
Đồng thời trong Statgraphics, còn có thể thiết lập mô hình tuyến tính hóa qua logarit với nhiều biến số độc lập được tổ hợp lại với nhau. Ví dụ, tiếp tục sử dụng Dữ liệu 11 (Huy et al., 2016b), ước lượng mô hình AGB theo bốn nhân tố DBH, H, WD (khối lượng thể tích gỗ, g/cm$^3$) và CA (diện tích tán lá, m$^2$) (Huy et al., 2016b) theo dạng hàm power có tổ hợp biến được logarit hóa như sau:

$$AGB = b_o \times (DBH^2 \times H \times WD)^{b1} \times CA^{b2} \times \varepsilon, \quad logarit\ hóa: \quad log(AGB) = log(b_o) + b_1 \times log(DBH^2 \times H \times WD) + b_2 \times log(CA) + log(\varepsilon) \qquad (7.32)$$

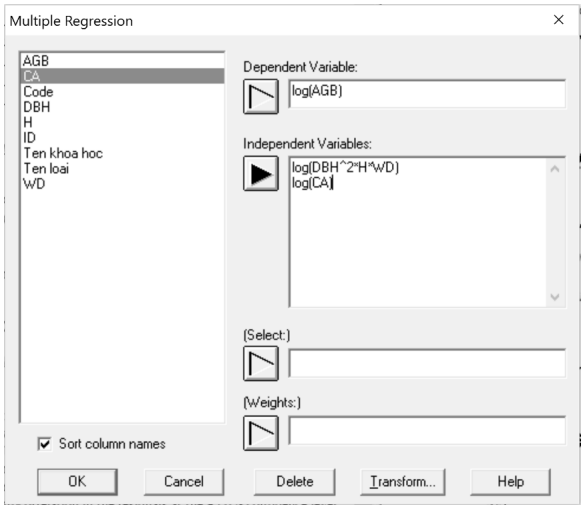
Tiến hành ước lượng mô hình phi tuyến nhiều biến, tổ hợp biến được logarit hóa trong chương trình Statgraphics theo phương pháp bình phương tối thiểu:

Thực hiện chương trình tuyến tính đa biến trong Statgraphics:

Improve / Regression Analysis / Multiple Factors / Multiple Regression ….




Trong hộp thoại Multiple Regression nhập các biến phụ thuộc, độc lập, tổ hợp biến và lấy log



Kết quả thiết lập mô hình phi tuyến nhiều biến số, tổ hợp biến được tuyến tính hóa theo hàm logarit và theo phương pháp bình phương tối thiểu trong Statgraphics:

**Multiple Regression - log(AGB)**

Dependent variable: log(AGB)
Independent variables:
log(DBH^2*H*WD)
log(CA)

| | | *Standard* | *T* | |
|---|---|---|---|---|
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| CONSTANT | -2.52301 | 0.0897843 | -28.1008 | 0.0000 |
| log(DBH^2*H*WD) | 0.87591 | 0.0197746 | 44.2947 | 0.0000 |
| log(CA) | 0.173596 | 0.0387972 | 4.47445 | 0.0000 |

**Analysis of Variance**

| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
|---|---|---|---|---|---|
| Model | 445.482 | 2 | 222.741 | 4625.43 | 0.0000 |
| Residual | 5.15266 | 107 | 0.0481557 | | |
| Total (Corr.) | 450.634 | 109 | | | |

R-squared = 98.8566 percent
R-squared (adjusted for d.f.) = 98.8352 percent
Standard Error of Est. = 0.219444
Mean absolute error = 0.161622
Durbin-Watson statistic = 1.77542 (P=0.1203)
Lag 1 residual autocorrelation = 0.106494

**The StatAdvisor**

The output shows the results of fitting a multiple linear regression model to describe the relationship between log(AGB) and 2 independent variables. The equation of the fitted model is

log(AGB) = -2.52301 + 0.87591*log(DBH^2*H*WD) + 0.173596*log(CA)

Since the P-value in the ANOVA table is less than 0.05, there is a statistically significant relationship between the variables at the 95.0% confidence level.

The R-Squared statistic indicates that the model as fitted explains 98.8566% of the variability in log(AGB). The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 98.8352%. The standard error of the estimate shows the standard deviation of the residuals to be 0.219444. This value can be used to construct prediction limits for new observations by selecting the Reports option from the text menu. The mean absolute error (MAE) of 0.161622 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the P-value is greater than 0.05, there is no indication of serial autocorrelation in the residuals at the 95.0% confidence level.

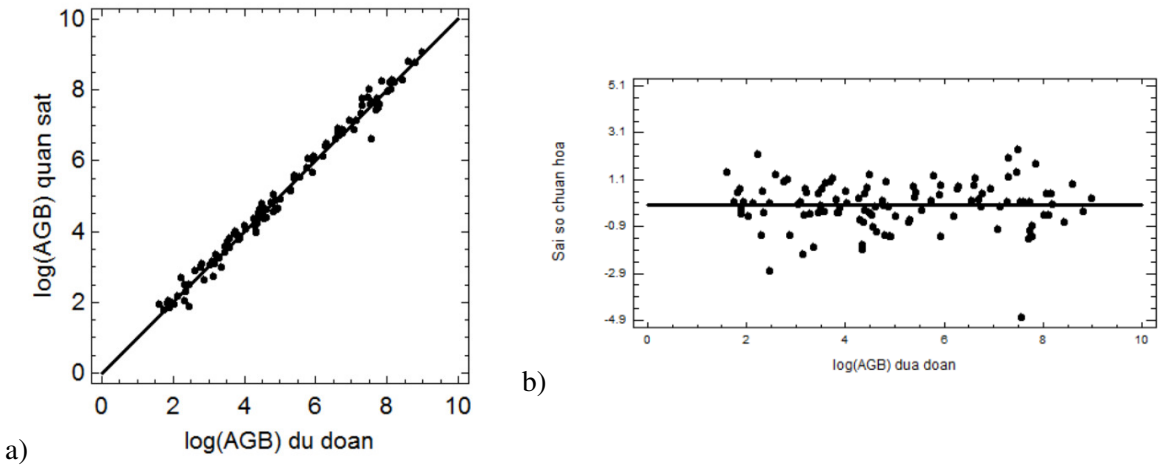
In determining whether the model can be simplified, notice that the highest P-value on the independent variables is 0.0000, belonging to log(CA). Since the P-value is less than 0.05, that term is statistically significant at the 95.0% confidence level. Consequently, you probably don't want to remove any variables from the model.

Kết quả có được mô hình:

$\log(AGB) = -2.52301 + 0.87591 \times \log(DBH^2 \times H \times WD) + 0.173596 \times \log(CA)$

Với $R^2_{adj.} = 98.8352\ \%$ và P-Value = 0.0000 < 0.0001 khi kiểm tra tồn tại của R và các tham số của mô hình.

Hình 7.16 cho thấy giá trị log(AGB) dự báo bám sát quan sát và sai số được chuẩn hóa biến động rải đều theo dự đoán và nằm hầu hết trong phạm vì -2 đến +2; tốt hơn so với mô hình hai biến đơn đã thực hiện ở phần trên.




**Hình 7.16.** Mô hình log(AGB) = -2.52301 + 0.87591*log($DBH^2$×H×WD) + 0.173596*log(CA).
a) Quan sát so với dự đoán; b) Biến động sai số theo dự đoán

### 7.6.2.2 Thiết lập mô hình phi tuyến được tuyến tính hóa theo logarit sử dụng codes "lm" trong chương trình R

Code "lm" (Chambers, 1992) thực hiện trong R giúp cho việc ước lượng các mô hình tuyến tính hoặc được tuyến tính hóa từ một đến nhiều biến. Chương trình này ước lượng các tham số của mô hình theo phương pháp bình phương tối thiểu. Đồng thời R còn cho phép chúng ta rất linh hoạt trong việc tính toán các chỉ tiêu thống kê và lập các đồ thị để đánh giá, so sánh các mô hình với nhau. Sau đây giới thiệu lần lượt sử dụng codes "lm" chạy trong R để ước lượng các mô hình phi tuyến được logarit hóa từ một đến nhiều biến độc lập hoặc tổ hợp biến và tính toán các chỉ tiêu sai số, đồ thị.

Tiếp tục sử dụng dữ liệu AGB theo các nhân tố điều tra (DBH, H, WD và CA) của 110 cây rừng mẫu của rừng lá rộng thường xanh vùng Nam Trung Bộ (Dữ liệu 11, Huy et al., 2016b), ước lượng mô hình AGB theo một đến nhiều nhân tố, tổ hợp nhân tố dạng hàm power được logarit hóa theo code "lm" trong R.

Tiến hành ước lượng mô hình phi tuyến power một biến được logarit hóa trong chương trình "lm" của R theo phương pháp bình phương tối thiểu:

$$AGB = b_o \times DBH^{b1} \times \varepsilon,\ logarit\ hóa:\ \log(AGB) = \log(b_o) + b_1 \times \log(DBH) + \log(\varepsilon) \quad (7.33)$$

**Chương trình "lm" trong R lập mô hình logarit: log(AGB) = a + b×log(DBH)**

```
# Erase memory (Xóa bộ nhớ):
rm(list=ls())

# Clean plot window (Xóa cửa sổ đồ thị cũ)
dev.off()

# Define the working directory (Thư mục)
setwd("E:/1 - Bao Huy 2017/Books All/Textbook for Infomatic Statistic/TextbookDataset Analysis/Dataset")

# Import data (File dữ liệu)
t_eq <- read.table("Du lieu 11 .txt", header=T,sep="\t",stringsAsFactors = FALSE)

# install.packages (Cài đặt các chương trình vẽ đồ thị)
library(ggplot2)
library(cowplot)
library(gridExtra)

# Mô hình tuyến tính logarit hóa (lm):
lmt1 <- lm(log(AGB)~log(DBH), data=t_eq)
summary(lmt1)
anova(lmt1)
# Outputs of the model
t_eq$lmt1.fit <- fitted.values(lmt1)
t_eq$lmt1.res <- residuals(lmt1)
# Calcul of model coefficients
a <- exp(coefficients(lmt1)[1])
b <- coefficients(lmt1)[2]
a
b
```

Kết quả:

```
Call:
lm(formula = log(AGB) ~ log(DBH), data = t_eq)

Residuals:
     Min       1Q   Median       3Q      Max
-1.07384 -0.13351 -0.00078  0.18722  0.50357

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.23646    0.09721  -23.01   <2e-16 ***
log(DBH)     2.47150    0.03210   77.01   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2732 on 108 degrees of freedom
Multiple R-squared:  0.9821,   Adjusted R-squared:  0.9819
F-statistic:  5930 on 1 and 108 DF,  p-value: < 2.2e-16

> anova(lmt1)
Analysis of Variance Table
```

```
Response: log(AGB)
            Df Sum Sq Mean Sq F value    Pr(>F)
log(DBH)     1 442.57  442.57  5929.9 < 2.2e-16 ***
Residuals 108   8.06    0.07
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> # Outputs of the model
> t_eq$lmt1.fit <- fitted.values(lmt1)
> t_eq$lmt1.res <- residuals(lmt1)
>
> # Calcul of model coefficients
> a <- exp(coefficients(lmt1)[1])
> b <- coefficients(lmt1)[2]
> a
(Intercept)
  0.1068359
> b
log(DBH)
2.471502
```

Kết quả trên lập được mô hình: AGB = 0.106836×DBH$^{2.471502}$ với các tham số tồn tại rõ rệt với Pr = <2e-16 < 0.0001

**Codes tính các chỉ tiêu thống kê, sai số của mô hình: AGB = 0.106836×DBH$^{2.471502}$**

```
# Calcul of correction factor: CF = exp(RSE^2/2):
summary(lmt1)$sigma^2
lmt1.CF <- exp(summary(lmt1)$sigma^2/2)
lmt1.CF

# Model back transformed: Y = exp(a)*X^b*CF (Đưa mô hình về dạng nguyên thủy phi tuyến)
# Calcul of fitted values and residuals
t_eq$backtr1.fit <- lmt1.CF * a * t_eq$DBH^b
t_eq$backtr1.res <- t_eq$AGB - t_eq$backtr1.fit

# Indicators for validation of the model (Các chỉ tiêu thống kê đánh giá mô hình) :
R2 <- 1- sum((t_eq$AGB - t_eq$backtr1.fit)^2)/sum((t_eq$AGB - mean(t_eq$AGB))^2)
R2
R2.adjusted <- 1 - (1-R2)*(length(t_eq$DBH)-1)/(length(t_eq$DBH)-3-1)
R2.adjusted

# Calcul of AIC = -2L + 2*p where p is numbers of parametters of model
ML.backtr1 <- -1/2*sum(t_eq$backtr1.res^2/var(t_eq$backtr1.res)+log(2*pi)
                +log(var(t_eq$backtr1.res)))
AIC.backtr1 <- -2*ML.backtr1 +2*2
AIC.backtr1

Bias <- mean(t_eq$backtr1.res)
RMSE <- sqrt(mean((t_eq$backtr1.res)^2))
MAPE <- 100*mean(abs(t_eq$backtr1.res)/t_eq$AGB)
Bias
RMSE
MAPE
```

Kết quả tính toán các chỉ tiêu thống kê của mô hình $AGB = 0.106836 \times DBH^{2.471502}$

```
> # Calcul of correction factor: CF = exp(RSE^2/2):
> summary(lmt1)$sigma^2
[1] 0.07463452
> lmt1.CF <- exp(summary(lmt1)$sigma^2/2)
> lmt1.CF
[1] 1.038022
>
> # Model back transformed: Y = exp(a)*X^b*CF
> # Calcul of fitted values and residuals
> t_eq$backtr1.fit <- lmt1.CF * a * t_eq$DBH^b
> t_eq$backtr1.res <- t_eq$AGB - t_eq$backtr1.fit
>
> # Indicators for validation of the model
> R2 <- 1- sum((t_eq$AGB - t_eq$backtr1.fit)^2)/sum((t_eq$AGB -
mean(t_eq$AGB))^2)
> R2
[1] 0.9353272
> R2.adjusted <- 1 - (1-R2)*(length(t_eq$DBH)-1)/(length(t_eq$DBH)-3-1)
> R2.adjusted
[1] 0.9334968
>
> # Calcul of AIC = -2L + 2*p where p is numbers of parametters of model
> ML.backtr1 <- -
1/2*sum(t_eq$backtr1.res^2/var(t_eq$backtr1.res)+log(2*pi)
+                          +log(var(t_eq$backtr1.res)))
>
> AIC.backtr1 <- -2*ML.backtr1 +2*2
> AIC.backtr1
[1] 1620.248
>
> Bias <- mean(t_eq$backtr1.res)
> RMSE <- sqrt(mean((t_eq$backtr1.res)^2))
> MAPE <- 100*mean(abs(t_eq$backtr1.res)/t_eq$AGB)
>
> Bias
[1] 5.553353
> RMSE
[1] 375.2613
> MAPE
[1] 22.61417
```

Kết quả trên chỉ ra các chỉ tiêu thống kê của mô hình: $AGB = 0.106836 \times DBH^{2.471502}$

CF = 1.038. CF gần bằng 1 cho thấy mô hình được ước lượng khá tốt

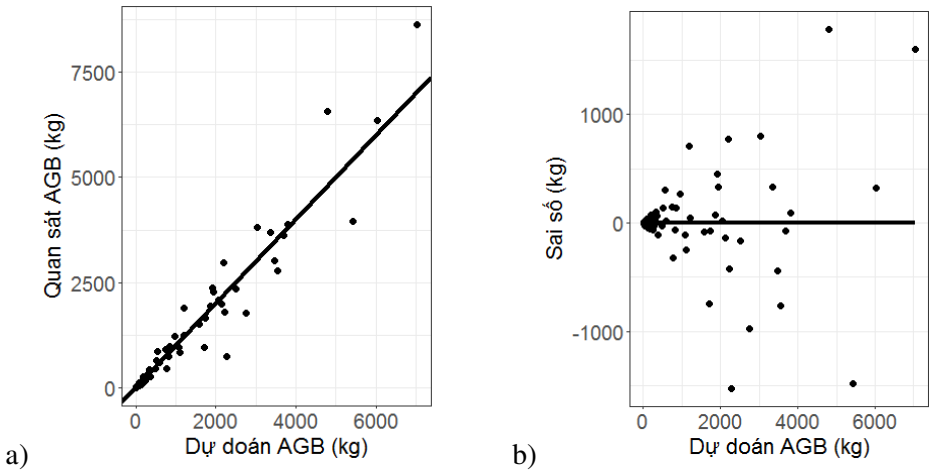
$R^2_{adjusted}$ = 0.9334968 và tồn tại với Pr = 2.2e-16 ; AIC = 1620.2

Các sai số: Bias = 5.55 kg ; RMSE = 375.3 kg và MAPE = 22.6%

**Codes vẽ các đồ thị của mô hình: $AGB = 0.106836 \times DBH^{2.471502}$**

```
# Observated and Predicted Values: (Đồ thị quan hệ giữa giá trị quan sát và dự đoán :
p1 <- ggplot(t_eq, aes(x=t_eq$backtr1.fit , y=AGB))
p1 <- p1 + geom_point(cex=2)
p1 <- p1 + geom_abline(cex = 1.5, intercept = 0, slope = 1, col="black")
p1 <- p1 + xlab("Dự đoán AGB (kg)") + ylab("Quan sát AGB (kg)") + theme_bw()+
theme_bw()
p1 = p1 + theme(axis.title.y = element_text(size = rel(1.5)))
p1 = p1 + theme(axis.title.x = element_text(size = rel(1.5)))
```

```
p1 <- p1 + theme(plot.title = element_text(size = rel(1.7)))
p1 = p1 + theme(axis.text.x = element_text(size=15))
p1 = p1 + theme(axis.text.y = element_text(size=15))
p1
# Resuduals and Predicted value (Đồ thị quan hệ giữa sai số voeis dự đoán) :
p2 <- ggplot(t_eq, aes(x=t_eq$backtr1.fit, y=t_eq$backtr1.res  ))
p2 <- p2 + geom_point(cex = 2)
p2 <- p2 + geom_line(cex = 1.5, aes(x=t_eq$backtr1.fit, y=0))
p2 <- p2 + xlab("Dự đoán AGB (kg)") + ylab("Sai số (kg)") + theme_bw()+ theme_bw()
p2 = p2 + theme(axis.title.y = element_text(size = rel(1.5)))
p2 = p2 + theme(axis.title.x = element_text(size = rel(1.5)))
p2 <- p2 + theme(plot.title = element_text(size = rel(1.7)))
p2 = p2 + theme(axis.text.x = element_text(size=15))
p2 = p2 + theme(axis.text.y = element_text(size=15))
p2
plot_grid(p1, p2, ncol = 2)
```




**Hình 7.17.** Mô hình: AGB = 0.106836×DBH$^{2.471502}$. a) Quan hệ giữa AGB quan sát và dự đoán; b) Biến động sai số theo giá trị AGB dự đoán qua mô hình

Kết quả ở Hình 7.17 cho thấy giá trị dự đoán và quan sát khá bám sát nhau, tuy nhiên, mô hình một biến số DBH có sai số rộng và phân tán khi giá trị dự đoán lớn (hay DBH tăng). Điều này có thể được cải thiện nếu áp dụng mô hình có trọng số (giới thiệu ở phần tiếp theo của giáo trình).

Để nâng cao độ tin cậy của ước lượng AGB, tiếp tục thử nghiệm tiến hành ước lượng mô hình phi tuyến power mở rộng gồm hai biến DBH và H được logarit hóa trong chương trình "lm" của R theo phương pháp bình phương tối thiểu:

$$AGB = b_o \times DBH^{b1} \times H^{b2} \times \varepsilon, \text{ logarit hóa: } log(AGB) = log(b_o) + b_1 \times log(DBH) + b_2 \times log(H) + log(\varepsilon) \quad (7.34)$$

**Codes lập mô hình tuyến tính hóa logarit nhiều biến: log(AGB) = log($b_o$) + $b_1$×log(DBH) +$b_2$×log(H)**

```
lmt1 <- lm(log(AGB)~log(DBH)+log(H), data=t_eq)
summary(lmt1)
anova(lmt1)
```

```
# Outputs of the model (Tính toán sai số và dự đoán)
t_eq$lmt1.fit <- fitted.values(lmt1)
t_eq$lmt1.res <- residuals(lmt1)
# Calcul of model coefficients (Tính toán các tham số của mô hình)
a <- exp(coefficients(lmt1)[1])
b <- coefficients(lmt1)[2]
c <- coefficients(lmt1)[3]
a
b
c
```

Kết quả tính toán mô hình:

```
> lmt1 <- lm(log(AGB)~log(DBH)+log(H), data=t_eq)
> summary(lmt1)

Call:
lm(formula = log(AGB) ~ log(DBH) + log(H), data = t_eq)

Residuals:
     Min       1Q   Median       3Q      Max
-1.04288 -0.13952  0.01563  0.18718  0.48347

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.85846    0.16361  -17.47  < 2e-16 ***
log(DBH)     2.14010    0.07875   27.18  < 2e-16 ***
log(H)       0.57932    0.12761    4.54 1.48e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2513 on 107 degrees of freedom
Multiple R-squared:  0.985,    Adjusted R-squared:  0.9847
F-statistic:  3514 on 2 and 107 DF,  p-value: < 2.2e-16

> anova(lmt1)
Analysis of Variance Table

Response: log(AGB)
           Df Sum Sq Mean Sq  F value    Pr(>F)
log(DBH)    1 442.57  442.57 7006.553 < 2.2e-16 ***
log(H)      1   1.30    1.30   20.609 1.482e-05 ***
Residuals 107   6.76    0.06
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> # Outputs of the model
> t_eq$lmt1.fit <- fitted.values(lmt1)
> t_eq$lmt1.res <- residuals(lmt1)
>
> # Calcul of model coefficients
> a <- exp(coefficients(lmt1)[1])
> b <- coefficients(lmt1)[2]
> c <- coefficients(lmt1)[3]
> a
(Intercept)
 0.05735691
> b
log(DBH)
2.140101
> c
   log(H)
0.5793182
```

Lập được mô hình: $AGB = 0.057356 \times DBH^{2.140101} \times H^{0.579318}$ với các tham số và $R^2$ tồn tại rõ rệt với Pr < 0.0001

**Codes tính toán các chỉ tiêu thống kê của mô hình: $AGB = 0.057356 \times DBH^{2.140101} \times H^{0.579318}$**

```
# Model back transformed: Y = exp(a)*X^b*CF
# Calcul of correction factor: CF = exp(RSE^2/2): Tính CF
summary(lmt1)$sigma^2
lmt1.CF <- exp(summary(lmt1)$sigma^2/2)
lmt1.CF

# Calcul of fitted values and residuals (Tính sai số và dự đoán)
t_eq$backtr1.fit <- lmt1.CF * a * t_eq$DBH^b*t_eq$H^c
t_eq$backtr1.res <- t_eq$AGB - t_eq$backtr1.fit

# Indicators for validation of the model (Các chỉ tiêu thống kê)
R2 <- 1- sum((t_eq$AGB - t_eq$backtr1.fit)^2)/sum((t_eq$AGB - mean(t_eq$AGB))^2)
R2

R2.adjusted <- 1 - (1-R2)*(length(t_eq$DBH)-1)/(length(t_eq$DBH)-4-1)
R2.adjusted

# Calcul of AIC = -2L + 2*p where p is numbers of parametters of model
ML.backtr1 <- -1/2*sum(t_eq$backtr1.res^2/var(t_eq$backtr1.res)+log(2*pi)
                  +log(var(t_eq$backtr1.res)))

AIC.backtr1 <- -2*ML.backtr1 +2*3
AIC.backtr1

Bias <- mean(t_eq$backtr1.res)
RMSE <- sqrt(mean((t_eq$backtr1.res)^2))
MAPE <- 100*mean(abs(t_eq$backtr1.res)/t_eq$AGB)

Bias
RMSE
MAPE
```

Kết quả có các chỉ tiêu thống kê của mô hình: $AGB = 0.057356 \times DBH^{2.140101} \times H^{0.579318}$

```
> # Model back transformed: Y = exp(a)*X^b*CF
> # Calcul of correction factor: CF = exp(RSE^2/2):
> summary(lmt1)$sigma^2
[1] 0.0631657
> lmt1.CF <- exp(summary(lmt1)$sigma^2/2)
> lmt1.CF
[1] 1.032087
>
> # Calcul of fitted values and residuals
> t_eq$backtr1.fit <- lmt1.CF * a * t_eq$DBH^b*t_eq$H^c
> t_eq$backtr1.res <- t_eq$AGB - t_eq$backtr1.fit
```

```
>
> # Indicators for validation of the model
> R2 <- 1- sum((t_eq$AGB - t_eq$backtr1.fit)^2)/sum((t_eq$AGB -
mean(t_eq$AGB))^2)
> R2
[1] 0.9420377
>
> R2.adjusted <- 1 - (1-R2)*(length(t_eq$DBH)-1)/(length(t_eq$DBH)-4-1)
> R2.adjusted
[1] 0.9398296
>
> # Calcul of AIC = -2L + 2*p where p is numbers of parametters of model
> ML.backtr1 <- -
1/2*sum(t_eq$backtr1.res^2/var(t_eq$backtr1.res)+log(2*pi)
+                         +log(var(t_eq$backtr1.res)))
>
> AIC.backtr1 <- -2*ML.backtr1 +2*3
> AIC.backtr1
[1] 1610.195
>
> Bias <- mean(t_eq$backtr1.res)
> RMSE <- sqrt(mean((t_eq$backtr1.res)^2))
> MAPE <- 100*mean(abs(t_eq$backtr1.res)/t_eq$AGB)
>
> Bias
[1] 18.01414
> RMSE
[1] 355.2594
> MAPE
[1] 20.93618
```

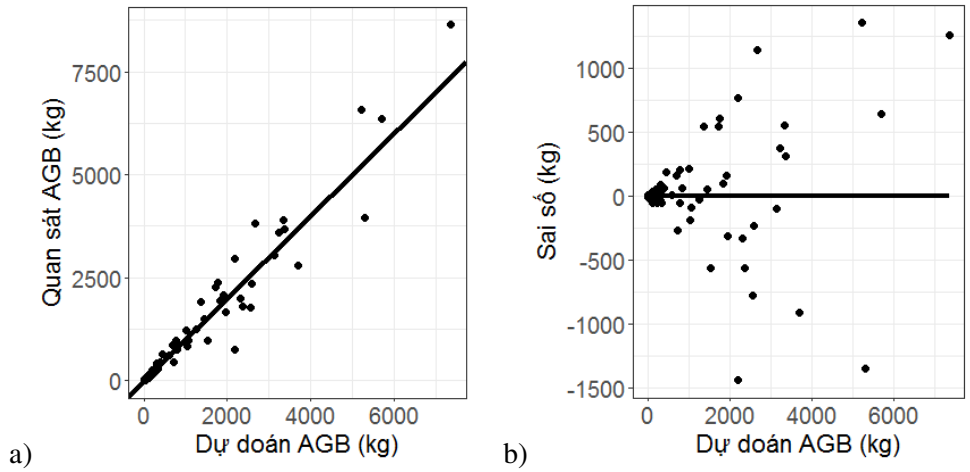
$R^2_{adj}$ = 0.939829; CF = 1.032; AIC = 1610.2

Các sai số: Bias = 18.0 kg; RMSE = 355.3 kg và MAPE = 20.9%

**Codes vẽ các đồ thị của mô hình: $AGB = 0.057356 \times DBH^{2.140101} \times H^{0.579318}$**

```
# Observated and Predicted Values:
p1 <- ggplot(t_eq, aes(x=t_eq$backtr1.fit , y=AGB))
p1 <- p1 + geom_point(cex=2)
p1 <- p1 + geom_abline(cex = 1.5, intercept = 0, slope = 1, col="black")
p1 <- p1 + xlab("Dự đoán AGB (kg)") + ylab("Quan sát AGB (kg)") + theme_bw()+
theme_bw()
p1 = p1 + theme(axis.title.y = element_text(size = rel(1.5)))
p1 = p1 + theme(axis.title.x = element_text(size = rel(1.5)))
p1 <- p1 + theme(plot.title = element_text(size = rel(1.7)))
p1 = p1 + theme(axis.text.x = element_text(size=15))
p1 = p1 + theme(axis.text.y = element_text(size=15))
p1
# Resuduals and Predicted value
p2 <- ggplot(t_eq, aes(x=t_eq$backtr1.fit, y=t_eq$backtr1.res  ))
p2 <- p2 + geom_point(cex = 2)
p2 <- p2 + geom_line(cex = 1.5, aes(x=t_eq$backtr1.fit, y=0))
p2 <- p2 + xlab("Dự đoán AGB (kg)") + ylab("Sai số (kg)") + theme_bw()+ theme_bw()
p2 = p2 + theme(axis.title.y = element_text(size = rel(1.5)))
```

```
p2 = p2 + theme(axis.title.x = element_text(size = rel(1.5)))
p2 <- p2 + theme(plot.title = element_text(size = rel(1.7)))
p2 = p2 + theme(axis.text.x = element_text(size=15))
p2 = p2 + theme(axis.text.y = element_text(size=15))
p2
plot_grid(p1, p2, ncol = 2)
```




**Hình 7.18.** Mô hình: AGB = $0.057356 \times DBH^{2.140101} \times H^{0.579318}$. a) Quan hệ giữa AGB quan sát và dự đoán; b) Biến động sai số theo giá trị AGB dự đoán qua mô hình

Kết quả trên đồ thị cho thấy, khi tăng thêm một biến số H, mô hình ước tính AGB bám sát hơn giá trị quan sát trên đường chéo (trái) và biến động sai số có giảm, tuy nhiên, cũng rất phân tán khi AGB lớn vì mô hình thiếu trọng số.

Tiếp tục thử nghiệm tiến hành ước lượng mô hình phi tuyến power mở rộng gồm bốn biến DBH, H, WD và CA, trong đó ba biến DBH, H và WD tạo thành một tổ hợp biến đại diện cho AGB ($DBH^2HWD = (DBH/100)^2 \times H \times WD \times 1000$ kg) và được logarit hóa trong chương trình "lm" của R theo phương pháp bình phương tối thiểu:

$$AGB = b_o \times (DBH^2HWD)^{b1} \times CA^{b2} \times \varepsilon, \; logarit\ hóa: \; log(AGB) = log(b_o) + b_1 \times log(DBH^2HWD) + b_2 \times log(CA) + log(\varepsilon) \quad (7.35)$$

**Codes lm thiết lập mô hình logarit hóa mô hình đa biến/tổ hợp biến $log(AGB) = log(b_o) + b_1 \times log(DBH2HWD) + b_2 \times log(CA)$**

```
# Combination of variable: DBH2HWD approximation of AGB (Tạo tổ hợp biến) :
t_eq$DBH2HWD = (t_eq$DBH/100)^2*t_eq$H*t_eq$WD*1000
# Code lm (Lập mô hình) :
lmt1 <- lm(log(AGB)~log(DBH2HWD)+log(CA), data=t_eq)
summary(lmt1)
anova(lmt1)
# Outputs of the model (Tính dự đoán và sai số mô hình):
t_eq$lmt1.fit <- fitted.values(lmt1)
t_eq$lmt1.res <- residuals(lmt1)
# Calcul of model coefficients (Tính các tham số mô hình):
```

```
a <- exp(coefficients(lmt1)[1])
b <- coefficients(lmt1)[2]
c <- coefficients(lmt1)[3]
a
b
c
```

Kết quả thu được mô hình trong R

```
> # Combination of variable: DBH2HWD approximation of AGB
> t_eq$DBH2HWD = (t_eq$DBH/100)^2*t_eq$H*t_eq$WD*1000
>
> lmt1 <- lm(log(AGB)~log(DBH2HWD)+log(CA), data=t_eq)
> summary(lmt1)

Call:
lm(formula = log(AGB) ~ log(DBH2HWD) + log(CA), data = t_eq)

Residuals:
     Min       1Q   Median       3Q      Max
-0.94397 -0.09699  0.02360  0.14820  0.49285

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.50521    0.06068  -8.325 3.01e-13 ***
log(DBH2HWD)  0.87581    0.01976  44.334  < 2e-16 ***
log(CA)       0.17352    0.03869   4.485 1.84e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2194 on 107 degrees of freedom
Multiple R-squared:  0.9886,   Adjusted R-squared:  0.9884
F-statistic:  4629 on 2 and 107 DF,  p-value: < 2.2e-16

> anova(lmt1)
Analysis of Variance Table

Response: log(AGB)
              Df Sum Sq Mean Sq  F value    Pr(>F)
log(DBH2HWD)   1 444.52  444.52 9237.988 < 2.2e-16 ***
log(CA)        1   0.97    0.97   20.119 1.837e-05 ***
Residuals    107   5.15    0.05
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> # Outputs of the model
> t_eq$lmt1.fit <- fitted.values(lmt1)
> t_eq$lmt1.res <- residuals(lmt1)
>
> # Calcul of model coefficients
> a <- exp(coefficients(lmt1)[1])
> b <- coefficients(lmt1)[2]
> c <- coefficients(lmt1)[3]
> a
(Intercept)
  0.6033778
> b
log(DBH2HWD)
   0.8758139
> c
  log(CA)
0.1735203
```

Mô hình ước lượng: $AGB = 0.603378 \times (DBH^2HWD)^{0.875814} \times CA^{0.173520}$ với các tham số và $R^2$ tồn tại rõ rệt với Pr < 0.0001

**Codes tính toán các chỉ tiêu thống kê của mô hình:**
**$AGB = 0.603378 \times (DBH^2HWD)^{0.875814} \times CA^{0.173520}$**

```
# Model back transformed: Y = exp(a)*X^b*CF
# Calcul of correction factor: CF = exp(RSE^2/2): (Tình CF)
summary(lmt1)$sigma^2
lmt1.CF <- exp(summary(lmt1)$sigma^2/2)
lmt1.CF

# Calcul of fitted values and residuals (Tính giá trị dự báo và sai số)
t_eq$backtr1.fit <- lmt1.CF * a * t_eq$DBH2HWD^b*t_eq$CA^c
t_eq$backtr1.res <- t_eq$AGB - t_eq$backtr1.fit

# Indicators for validation of the model (Tính các chỉ tiêu thống kê, sai số)
R2 <- 1- sum((t_eq$AGB - t_eq$backtr1.fit)^2)/sum((t_eq$AGB - mean(t_eq$AGB))^2)
R2

R2.adjusted <- 1 - (1-R2)*(length(t_eq$DBH)-1)/(length(t_eq$DBH)-4-1)
R2.adjusted

# Calcul of AIC = -2L + 2*p where p is numbers of parametters of model
ML.backtr1 <- -1/2*sum(t_eq$backtr1.res^2/var(t_eq$backtr1.res)+log(2*pi)
                  +log(var(t_eq$backtr1.res)))

AIC.backtr1 <- -2*ML.backtr1 +2*3
AIC.backtr1

Bias <- mean(t_eq$backtr1.res)
RMSE <- sqrt(mean((t_eq$backtr1.res)^2))
MAPE <- 100*mean(abs(t_eq$backtr1.res)/t_eq$AGB)

Bias
RMSE
MAPE
```

Kết quả thu được các chỉ tiêu thống kê từ R:

```
> # Model back transformed: Y = exp(a)*X^b*CF
> # Calcul of correction factor: CF = exp(RSE^2/2):
> summary(lmt1)$sigma^2
[1] 0.04811844
> lmt1.CF <- exp(summary(lmt1)$sigma^2/2)
> lmt1.CF
[1] 1.024351
>
> # Calcul of fitted values and residuals
> t_eq$backtr1.fit <- lmt1.CF * a * t_eq$DBH2HWD^b*t_eq$CA^c
> t_eq$backtr1.res <- t_eq$AGB - t_eq$backtr1.fit
```

```
>
> # Indicators for validation of the model
> R2 <- 1- sum((t_eq$AGB - t_eq$backtr1.fit)^2)/sum((t_eq$AGB -
mean(t_eq$AGB))^2)
> R2
[1] 0.961727
>
> R2.adjusted <- 1 - (1-R2)*(length(t_eq$DBH)-1)/(length(t_eq$DBH)-4-1)
> R2.adjusted
[1] 0.9602689
>
> # Calcul of AIC = -2L + 2*p where p is numbers of parametters of model
> ML.backtr1 <- -
1/2*sum(t_eq$backtr1.res^2/var(t_eq$backtr1.res)+log(2*pi)
+                         +log(var(t_eq$backtr1.res)))
>
> AIC.backtr1 <- -2*ML.backtr1 +2*3
> AIC.backtr1
[1] 1564.542
>
> Bias <- mean(t_eq$backtr1.res)
> RMSE <- sqrt(mean((t_eq$backtr1.res)^2))
> MAPE <- 100*mean(abs(t_eq$backtr1.res)/t_eq$AGB)
>
> Bias
[1] 3.226089
> RMSE
[1] 288.6817
> MAPE
[1] 17.26489
```

Kết quả mô hình: $AGB = 0.603378 \times (DBH^2HWD)^{0.875814} \times CA^{0.173520}$ có các chỉ tiêu thống kê, sai số như sau :

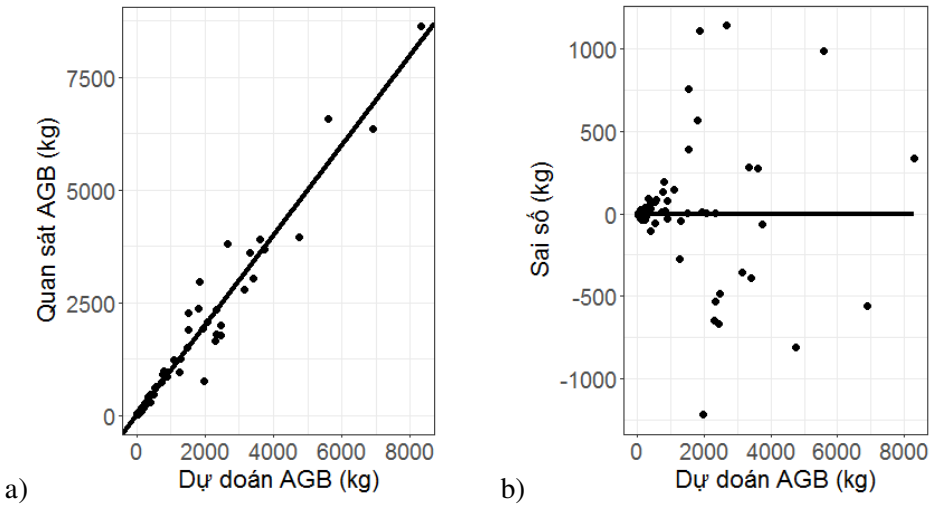
CF = 1.024; $R^2_{adj} = 0.960269$; AIC = 1564.5

Các sai số: Bias = 3.2 kg; RMSE = 288.7 kg và MAPE = 17.3%

**Codes vẽ đồ thị của mô hình $AGB = 0.603378 \times (DBH^2HWD)^{0.875814} \times CA^{0.173520}$**

```
# Observated and Predicted Values:
p1 <- ggplot(t_eq, aes(x=t_eq$backtr1.fit , y=AGB))
p1 <- p1 + geom_point(cex=2)
p1 <- p1 + geom_abline(cex = 1.5, intercept = 0, slope = 1, col="black")
p1 <- p1 + xlab("Dự đoán AGB (kg)") + ylab("Quan sát AGB (kg)") + theme_bw()+
theme_bw()
p1 = p1 + theme(axis.title.y = element_text(size = rel(1.5)))
p1 = p1 + theme(axis.title.x = element_text(size = rel(1.5)))
p1 <- p1 + theme(plot.title = element_text(size = rel(1.7)))
p1 = p1 + theme(axis.text.x = element_text(size=15))
p1 = p1 + theme(axis.text.y = element_text(size=15))
p1
# Resuduals and Predicted value
p2 <- ggplot(t_eq, aes(x=t_eq$backtr1.fit, y=t_eq$backtr1.res  ))
p2 <- p2 + geom_point(cex = 2)
p2 <- p2 + geom_line(cex = 1.5, aes(x=t_eq$backtr1.fit, y=0))
p2 <- p2 + xlab("Dự đoán AGB (kg)") + ylab("Sai số (kg)") + theme_bw()+ theme_bw()
p2 = p2 + theme(axis.title.y = element_text(size = rel(1.5)))
p2 = p2 + theme(axis.title.x = element_text(size = rel(1.5)))
p2 <- p2 + theme(plot.title = element_text(size = rel(1.7)))
```

```
p2 = p2 + theme(axis.text.x = element_text(size=15))
p2 = p2 + theme(axis.text.y = element_text(size=15))
p2

plot_grid(p1, p2, ncol = 2)
```




**Hình 7.19.** Mô hình: AGB = 0.603378×(DBH2HWD)$^{0.875814}$×CA$^{0.173520}$. a) Quan hệ giữa AGB quan sát và dự đoán. b) Biến động sai số theo giá trị AGB dự đoán qua mô hình

Từ đồ thị trên cho thấy mô hình ước tính AGB với bốn biến số DBH, H, WD và CA có giá trị dự đoán bám sát quan sát; tuy nhiên, với mô hình không có trọng số thì sai số vẫn biến động mạnh khi AGB tăng lên. Vì vậy, cần sử dụng mô hình có trọng số.

Thử so sánh độ tin cậy sai số của ba mô hình ước tính AGB có biến số đầu vào khác nhau ở Bảng 7.3.

**Bảng 7.3.** So sánh độ tin cậy, sai số của các mô hình ước tính AGB với biến số đầu vào khác nhau dạng tuyến tính hóa logarit ước lượng theo chương trình lm trong R

| Chỉ tiêu thống kê so sánh | log(AGB = log(a) +b×log(DBH) | log(AGB) = log(a) + b×log(DBH) + c×log(H) | log(AGB) = log(a) + b×log(DBH$^2$HWD) + c×log(CA) |
|---|---|---|---|
| $R^2_{adj}$ | 0.933496 | 0.939829 | 0.960269 |
| CF | 1.038 | 1.032 | 1.024 |
| AIC | 1620.2 | 1610.2 | 1564.5 |
| Bias (kg) | 5.6 | 18.0 | 3.2 |
| RMSE (kg) | 375.3 | 355.3 | 288.7 |
| MAPE (%) | 22.6 | 20.9 | 17.3 |

Kết quả trên cho thấy, khi gia tăng số biến số đầu vào từ một biến DBH lên hai biến (thêm H) và bốn biến (DBH, H, WD, CA) thì mô hình có độ tin cậy càng cao và sai số giảm. Mô hình tốt nhất có bốn biến số đầu vào ở dạng tổ hợp biến với $R^2_{adj}$ cao nhất, CF bé nhất và gần bằng 1, AIC bé nhất, các sai số Bias, RMSE, MAPE bé nhất. Sử dụng mô hình 4 biến số giảm sai số tương đối MAPE 5% so với mô hình một biến số.

### 7.5.3 Ước lượng mô hình theo phương pháp phi tuyến tính

Ngoài việc ước lượng các mô hình phi tuyến tính bằng cách tuyến tính hóa mô hình theo hàm logarit và áp dụng phương pháp bình phương tối thiểu, thì có hàng loạt các phương pháp ước lượng trực tiếp mô hình phi tuyến tính mà không phải tuyến tính hóa như là phương pháp phi tuyến Marquardt (StatPoint-Inc., 2005), phi tuyến bình phương tối thiểu (chương trình "nls" trong R: Nonlinear Least Squares) (Bates và Watts, 1988), phi tuyến ảnh hưởng phức hợp hợp lý tối đa (chương trình "nlme" trong R: Nonlinear Mixed-Effects - Maximum Likelihood) (Davidian và Giltinan, 1995; Pinheiro et al., 2014). Các phương pháp phi tuyến ước lượng trực tiếp các tham số mô hình mà không phải thông qua hàm trung gian để tuyến tính hóa, do đó, về mặt nguyên tắc chung nó đã làm giảm được sai số ước lượng trung gian, cụ thể là không cần sử dụng hệ số điều chỉnh CF như mô hình tuyến tính hóa, do đó, tăng độ tin cậy của mô hình; ngoài ra còn có kỹ thuật trọng số để giảm biến động sai số của mô hình khi ước lượng bằng phương pháp phi tuyến tính.

Sau đây lần lượt giới thiệu các phương pháp ước lượng các hàm phi tuyến khác nhau và cách thức đánh giá để lựa chọn phương pháp thích hợp trong từng trường hợp cụ thể.

#### 7.5.3.1 Phương pháp phi tuyến tính Marquardt

Phương pháp phi tuyến Marquardt (StatPoint-Inc., 2005) ước lượng trực tiếp các tham số của mô hình phi tuyến với thông tin đầu vào là các tham số khởi đầu của mô hình. Phương pháp này được áp dụng thuận lợi trong phần mềm Statgraphics.

Để giới thiệu phương pháp này, sử dụng Dữ liệu 12 của Bảo Huy và Đào Công Khanh (2008) trong lập biểu sản lượng rừng trồng trám trắng (*Canarium album* Raeusch) tại các tỉnh Lạng Sơn, Bắc Giang, Quảng Ninh. Có 73 lâm phần trám trắng được thu thập dữ liệu sinh trưởng của cây bình quân lâm phần như chiều cao bình quân tầng trội ($H_o$, m) (là chiều cao bình quân của 20% cây cao nhất trên 0.1 ha rừng), đường kính bình quân (Dbq, cm), chiều cao bình quân (Hbq, m), thể tích bình quân (V, $m^3$). Để phân cấp năng suất rừng trồng trám trắng đã lập mô hình sinh trưởng $H_o/A$ theo hàm Schumacher:

$$Ho = bo \times \exp(-b_1 \times A^{-m}) + \varepsilon \quad (7.36)$$

Tiến hành thiết lập mô hình trên theo Marquardt trong phần mềm Statgraphics như sau:

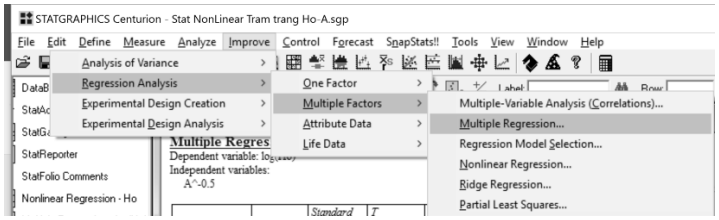
Tìm tham số đầu vào cho mô hình Schumacher:

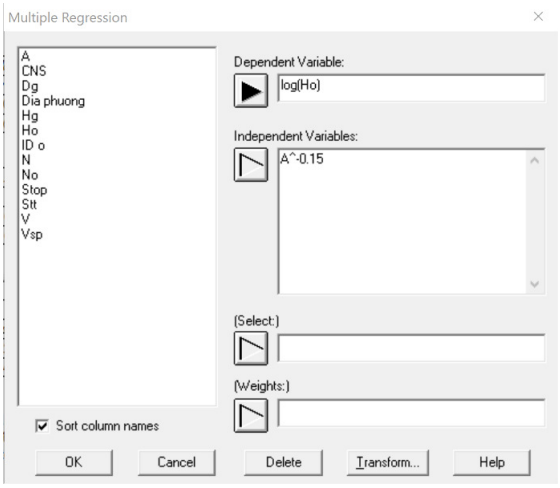
Các tham số $b_o$, $b_1$ và m cần được xác định trước thông qua mô hình logarit hóa:

$\log(H_o) = \log(b_o) - b_1 \times A^{-m}$

Lập mô hình tuyến tính hóa theo phương pháp bình phương tối thiểu trong Statgraphics:

Improve/ Regression



| | |
|---|---|
| Analysis/ Multiple Factors / Multiple Regression | |
| Nhập các biến số:<br>Dependent Variable: log(Ho)<br>Indepedent Variables: $A^{-m}$<br>Trong đó thử nghiệm thay m từ 0.1 – 2.0, để chọn m tối ưu khi mô hình có $R^2$ max và các sai số mô hình bé nhất.<br>Trong ví dụ này m tối ưu là 0.15 |  |

Kết quả mô hình Schumacher dạng logarit:

**Multiple Regression - log(Ho)**
Dependent variable: log(Ho)
Independent variables:
A^-0.15

| | | *Standard* | *T* | |
|---|---|---|---|---|
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| CONSTANT | 6.75039 | 0.499005 | 13.5277 | 0.0000 |
| A^-0.15 | -6.65422 | 0.668616 | -9.95224 | 0.0000 |

**Analysis of Variance**

| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
|---|---|---|---|---|---|
| Model | 4.54562 | 1 | 4.54562 | 99.05 | 0.0000 |
| Residual | 2.98308 | 65 | 0.0458935 | | |
| Total (Corr.) | 7.5287 | 66 | | | |

R-squared = 60.3772 percent
R-squared (adjusted for d.f.) = 59.7676 percent
Standard Error of Est. = 0.214228
Mean absolute error = 0.176139
Durbin-Watson statistic = 1.87811 (P=0.2824)
Lag 1 residual autocorrelation = 0.0394685

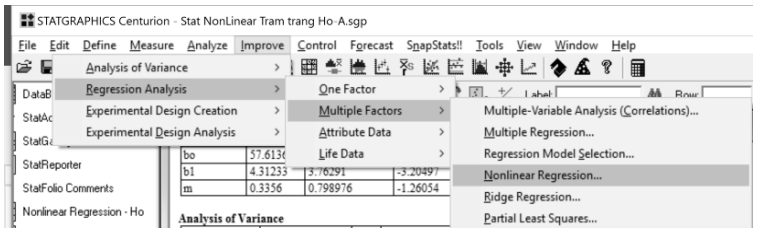
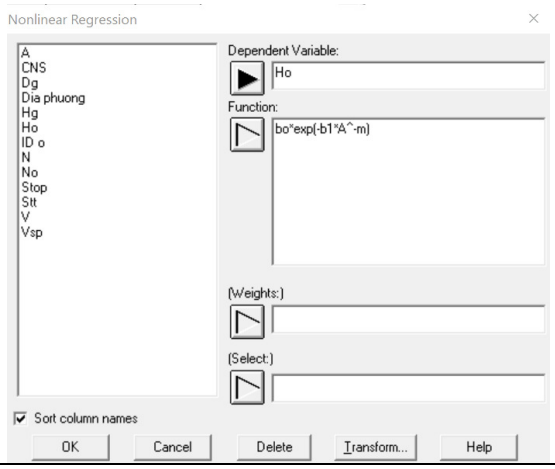
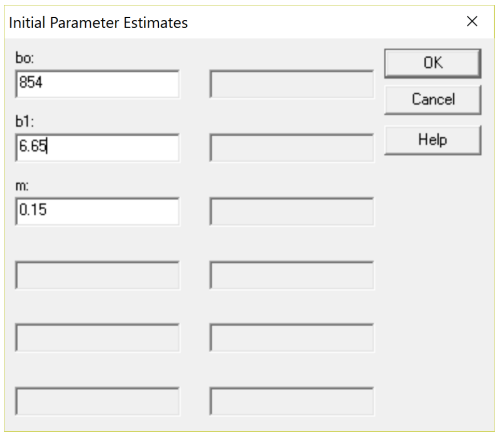
**The StatAdvisor**
The output shows the results of fitting a multiple linear regression model to describe the relationship between log(Ho) and 1 independent variables. The equation of the fitted model is

$\log(H_o) = 6.75039 - 6.65422*A^{-0.15}$

Như vậy tham số ban đầu của mô hình sẽ là:

$b_o$ = exp(6.75) = 854 ; $b_1$ = 6.65 và m = 0.15

Sử dụng các tham số đầu vào để tiếp tục ước lượng mô hình Schumacher theo phương pháp phi tuyến tính Marquardt như sau:

| | |
|---|---|
| Thiết lập mô hình phi tuyến Schumacher theo Marquardt trong Statgraphics: Improve / Regression Analysis / Multiple Factors / Nonlinear Regression ... |  |
| Trong hộp thoại Nonlinear Regression nhập biến số phụ thuộc và dạng mô hình Schumacher:<br>Dependent variable: $H_o$<br>Function: $b_o$*exp(-b1* A^-m) |  |
| Trong hộp thoại ước tính tham số khởi đầu: (Initial Parameter Estimates), nhập các các giá trị tham số $b_o$m $b_1$ và m tìm được từ mô hình logarit ở phần trên |  |

Kết quả ước lượng hàm Schumacher $H_o$/A theo phương pháp phi tuyến Marquardt trong Statgraphics như sau:

**Nonlinear Regression - $H_o$**
Dependent variable: $H_o$
Independent variables:
A

Function to be estimated: $b_o$*exp(-$b_1$*A^-m)
Initial parameter estimates:

bo = 854.0
b1 = 6.65
m = 0.15

Estimation method: Marquardt
Estimation stopped due to convergence of residual sum of squares.
Number of iterations: 5
Number of function calls: 25

**Estimation Results**

| | | | *Asymptotic* | *95.0%* |
|---|---|---|---|---|
| | | *Asymptotic* | *Confidence* | *Interval* |
| *Parameter* | *Estimate* | *Standard Error* | *Lower* | *Upper* |
| bo | 913.185 | 24003.5 | -47039.5 | 48865.8 |
| b1 | 6.73639 | 26.222 | -45.6481 | 59.1209 |
| m | 0.151546 | 0.789574 | -1.42581 | 1.7289 |

**Analysis of Variance**

| *Source* | *Sum of Squares* | *Df* | *Mean Square* |
|---|---|---|---|
| Model | 2840.22 | 3 | 946.739 |
| Residual | 122.83 | 64 | 1.91921 |
| Total | 2963.05 | 67 | |
| Total (Corr.) | 279.064 | 66 | |

R-Squared = 55.9852 percent
R-Squared (adjusted for d.f.) = 54.6097 percent
Standard Error of Est. = 1.38536
Mean absolute error = 1.10229
Durbin-Watson statistic = 1.7863
Lag 1 residual autocorrelation = 0.0814487

**Residual Analysis**

| | *Estimation* | *Validation* |
|---|---|---|
| n | 67 | |
| MSE | 1.91921 | |
| MAE | 1.10229 | |
| MAPE | 18.2573 | |
| ME | 0.00420461 | |
| MPE | -4.3777 | |

**The StatAdvisor**
The output shows the results of fitting a nonlinear regression model to describe the relationship between $H_o$ and 1 independent variables. The equation of the fitted model is

$H_o$ = 913.185*exp(-6.73639*A^-0.151546)

In performing the fit, the estimation process terminated successully after 5 iterations, at which point the estimated coefficients appeared to converge to the current estimates.

The R-Squared statistic indicates that the model as fitted explains 55.9852% of the variability in Ho. The adjusted R-Squared statistic, which is more suitable for comparing models with different

numbers of independent variables, is 54.6097%. The standard error of the estimate shows the standard deviation of the residuals to be 1.38536. This value can be used to construct prediction limits for new observations by selecting the Forecasts option from the text menu. The mean absolute error (MAE) of 1.10229 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file.

The output also shows aymptotic 95.0% confidence intervals for each of the unknown parameters. These intervals are approximate and most accurate for large sample sizes. You can determine whether or not an estimate is statistically significant by examining each interval to see whether it contains the value 0. Intervals covering 0 correspond to coefficients which may well be removed form the model without hurting the fit substantially.

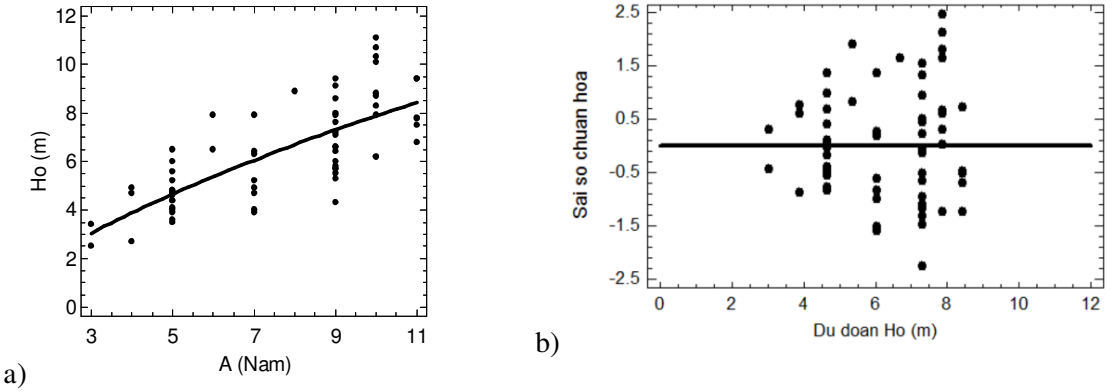
---

Kết quả:

$H_o$ = 913.185*exp(-6.73639*A^-0.151546)

Với $R^2_{adj}$ = 54.6097%; một số sai số chính Bias (ME – Mean eror) = 0.004m; MAE = 1.10m và MAPE = 18.26%.

Phương pháp Marquardt ước lượng trực tiếp mô hình từ giá trị quan sát, không bị thay đổi như logarit để áp dụng phương pháp tuyến tính, vì vậy mô hình ước lượng bám sát hơn giá trị quan sát thực và cho sai số bé hơn; cho dù $R^2$ của phương pháp phi tuyến có thể bé hơn mô hình lập theo phương pháp tuyến tính, vấn đề lựa chọn phương pháp ước lượng mô hình sẽ được trình bày ở phần sau của giáo trình này.




a) b)

**Hình 7.20.** Mô hình $H_o$ = 913.185*exp(-6.73639*A^-0.151546). a) Mô hình theo giá trị quan sát; b) biến động sai số theo dự đoán $H_o$.

Hình 7.20 cho thấy mô hình ước lượng đi qua trung tâm giá trị quan sát (trái), tuy nhiên, sai số có sự phân hóa cao khi giá trị dự đoán tăng lên (A tăng lên) (b). Điều này cho thấy mô hình có sai số lớn ở các tuổi cao, do đó cần có giải pháp cải thiện sai số của mô hình thông qua sử dụng trọng số theo biến số A: Weight = $1/A^a$ với a biến động từ -20 đến +20 (Picarrd et al. 2012). Lập mô hình có trọng số sẽ được trình bày ở phần tiếp theo của giáo trình này.

### 7.5.3.2 Phương pháp phi tuyến bình phương tối thiểu (Code nls trong R)

Phương pháp phi tuyến tính bình phương tối thiểu (nls - Nonlinear Least Squares) ước lượng trực tiếp các tham số của mô hình phi tuyến với thông tin đầu vào là các tham số khởi đầu của mô hình. Phương pháp này được áp dụng thuận lợi trong phần mềm mã nguồn mở R theo chương trình nls (Bates et al., 1988).

Sử dụng Dữ liệu 12 của Bảo Huy và Đào Công Khanh (2008) trong lập biểu sản lượng rừng trồng trám trắng (*Canarium album* Raeusch) tại các tỉnh Lạng Sơn, Bắc Giang, Quảng Ninh. Từ dữ liệu 73 lâm phần trám trắng được thu thập, minh họa phương pháp để thiết lập mô hình ước tính thể tích bình quân (V, $m^3$) theo đường kính (Dg, cm) và chiều cao bình quân lâm phần (Hg, m): V = f(Dg, Hg) theo mô hình power tổ hợp biến:

$$V = a \times (Dg^2 Hg)^b + \varepsilon \quad (7.37)$$

Trong đó $Dg^2Hg$ là tổ hợp biến đại diện cho thể tích cây ($m^3$) $Dg^2Hg = (Dg/100)^2 \times Hg$

Tiến hành thiết lập mô hình trên theo codes nls trong R như sau:

**Codes nls thiết lập mô hình phi tuyến tính trong R:**

$$V = a \times (Dg^2 Hg)^b$$

```
# Erase memory (Xóa bộ nhớ)
rm(list=ls())
# Clean plot window (Xóa các cửa sổ đồ thị cũ)
dev.off()
# Define the working directory (Đường dẫn đến thư mục chưa file dữ liệu)
setwd("E:/1 - Bao Huy 2017/Books All/Textbook for Infomatic Statistic/TextbookDataset Analysis/Dataset")
# Import data (Nhập dữ liệu) :
t_eq <- read.table("Du lieu 12 Tram trang.txt", header=T,sep="\t",stringsAsFactors = FALSE)
# install.packages for plots (cài đặt các chương trình đồ thị) :
library(ggplot2)
library(cowplot)
library(gridExtra)

#############################################################
#    Lập mô hình V = a*(Dg^2*Hg)^b theo code nls
#############################################################
# Combination of variables (Tạo tổ hợp biến đại diện thể tích) :
t_eq$Dg2Hg = (t_eq$Dg/100)^2*t_eq$Hg

# Initial parameters (Tham số khởi đầu của mô hình) :
start <- coefficients(lm(log(V)~log(Dg2Hg), data=t_eq))
names(start) <- c("a","b")
# Áp dụng chương trinh lập mô hình phi tuyến nls :
nls_least_square <- nls(V~a*Dg2Hg^b, data=t_eq, start=start)
# Tóm tắt kết quả mô hình :
summary(nls_least_square)
# Tính giá trị dự đoán và sai số của mô hình :
t_eq$nls_least_square.fit <- fitted.values(nls_least_square)
t_eq$nls_least_square.res <- residuals(nls_least_square)
```

Kết quả mô hình:

```
> summary(nls_least_square)

Formula: V ~ a * Dg2Hg^b

Parameters:
  Estimate Std. Error t value Pr(>|t|)
a  0.33842    0.01051   32.19   <2e-16 ***
b  0.88971    0.01145   77.73   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.001051 on 71 degrees of freedom

Number of iterations to convergence: 5
Achieved convergence tolerance: 9.293e-07

> t_eq$nls_least_square.fit <- fitted.values(nls_least_square)
> t_eq$nls_least_square.res <- residuals(nls_least_square)
```

Mô hình được thiết lập: $V = 0.33842 \times (Dg^2Hg)^{0.88971}$

Các tham số của mô hình có các Pr < 2e-16, chứng tỏ các tham số tồn tại rõ rệt.

Codes tính các chỉ tiêu thống kê, và sai số của mô hình phi tuyến lập theo codes nls:

```
# calcul of AIC, R2 and errors:
AIC = AIC(nls_least_square)
AIC
R2 <- 1- sum((t_eq$V - t_eq$nls_least_square.fit)^2)/sum((t_eq$V - mean(t_eq$V))^2)
R2.adjusted <- 1 - (1-R2)*(length(t_eq$V)-1)/(length(t_eq$V)-3-1)
R2.adjusted

Bias = mean(t_eq$nls_least_square.res)
Bias
RMSE =  sqrt(mean(t_eq$nls_least_square.res^2))
RMSE
MAPE = 100*mean(abs(t_eq$nls_least_square.res)/t_eq$V)
MAPE
```

Kết quả tính các chỉ tiêu thống kê sai số của mô hình chạy theo nls:

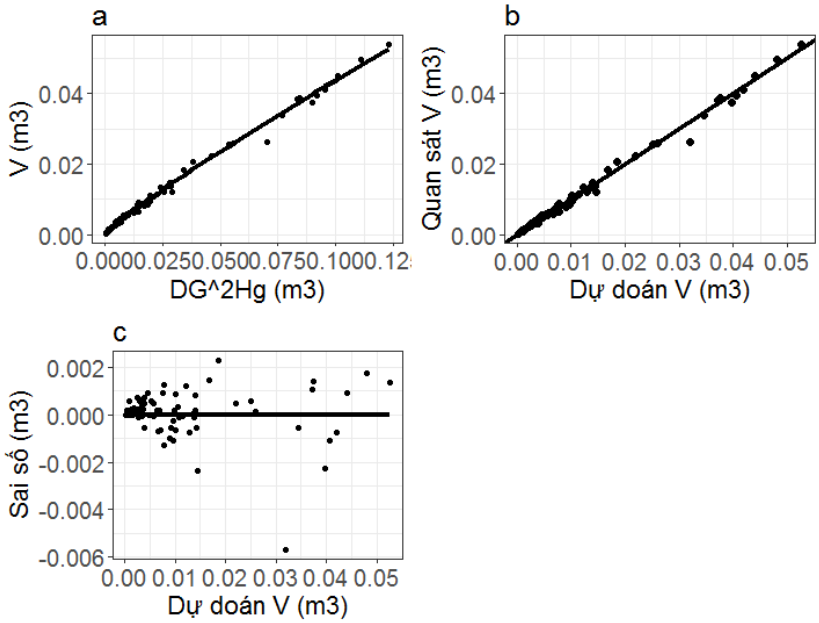
```
> AIC = AIC(nls_least_square)
> AIC
[1] -790.0709
> R2 <- 1- sum((t_eq$V - t_eq$nls_least_square.fit)^2)/sum((t_eq$V -
mean(t_eq$V))^2)
> R2.adjusted <- 1 - (1-R2)*(length(t_eq$V)-1)/(length(t_eq$V)-3-1)
> R2.adjusted
[1] 0.9933147
>
> Bias = mean(t_eq$nls_least_square.res)
> Bias
[1] 6.956402e-05
> RMSE =  sqrt(mean(t_eq$nls_least_square.res^2))
> RMSE
[1] 0.001036942
> MAPE = 100*mean(abs(t_eq$nls_least_square.res)/t_eq$V)
> MAPE
[1] 7.850645
```

AIC = -790.1; $R^2_{adj}$ = 0.993314 (Quan hệ là chặt chẽ)

Bias = 6.95e-05$m^3$; RMSE = 0.001036$m^3$ và MAPE = 7.85%. Các sai số bé cho thấy dạng mô hình tổ hợp biến mô phỏng tốt quan hệ V với Dg và Hg của loài trám trắng.

Codes thiết lập các đồ thị đánh giá mô hình:

```
# Plots:
# Fitted model (Mô hình theo quan sát)
p1 <- ggplot(t_eq)
p1 <- p1 + geom_point(aes(x=Dg2Hg, y=V))
p1 <- p1 + geom_line(cex = 1.5, aes(x=Dg2Hg, y=nls_least_square.fit))
p1 <- p1 + xlab("DG^2Hg (m3)") + ylab("V (m3)") + theme_bw()
p1 <- p1 + labs(title = "a")
p1 = p1 + theme(axis.title.y = element_text(size = rel(1.5)))
p1 = p1 + theme(axis.title.x = element_text(size = rel(1.5)))
p1 <- p1 + theme(plot.title = element_text(size = rel(1.7)))
p1 = p1 + theme(axis.text.x = element_text(size=15))
p1 = p1 + theme(axis.text.y = element_text(size=15))
p1

# Observated and Predicted Values: (Đồ thị Quan sát với dự báo):
p2 <- ggplot(t_eq, aes(x=t_eq$nls_least_square.fit, y=V))
p2 <- p2 + geom_point(cex=2)
p2 <- p2 + geom_abline(cex = 1.5, intercept = 0, slope = 1, col="black")
p2 <- p2 + xlab("Dự đoán (m3)") + ylab("Quan sát (m3)") + theme_bw()+ theme_bw()
p2 = p2 + theme(axis.title.y = element_text(size = rel(1.5)))
p2 = p2 + theme(axis.title.x = element_text(size = rel(1.5)))
p2 <- p2 + theme(plot.title = element_text(size = rel(1.7)))
p2 = p2 + theme(axis.text.x = element_text(size=15))
p2 = p2 + theme(axis.text.y = element_text(size=15))
p2 <- p2 + labs(title = "b")
p2

# Residuals vs predicted (Quan hệ sai số với dự đoán)
p3 <- ggplot(t_eq, aes(x=nls_least_square.fit, y=nls_least_square.res))
p3 <- p3 + geom_point()
p3 <- p3 + geom_line(cex = 1.5, aes(x=nls_least_square.fit, y=0))
p3 <- p3 + xlab("Fitted Values (m3)") + ylab("Residuals (m3)") + theme_bw()
p3 <- p3 + labs(title = "")
p3 = p3 + theme(axis.title.y = element_text(size = rel(1.5)))
p3 = p3 + theme(axis.title.x = element_text(size = rel(1.5)))
p3 <- p3 + theme(plot.title = element_text(size = rel(1.7)))
p3 = p3 + theme(axis.text.x = element_text(size=15))
p3 = p3 + theme(axis.text.y = element_text(size=15))
p3 <- p3 + labs(title = "c")
p3
plot_grid(p1, p2, p3, ncol = 2)
```

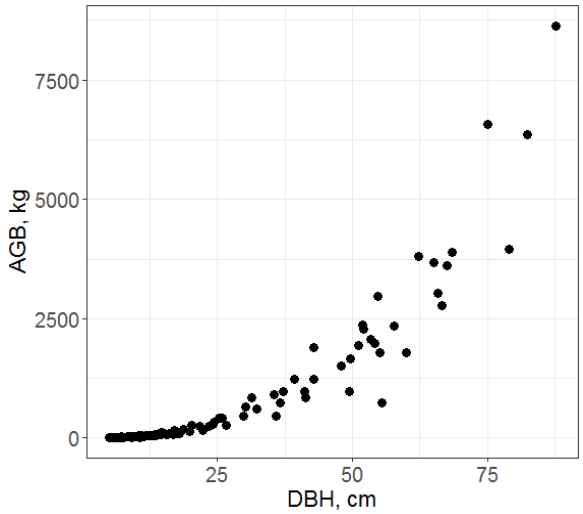



**Hình 7.21.** Các đồ thị của mô hình V = 0.33842×$(Dg^2Hg)^{0.88971}$. a) Mô hình với quan sát; b) Quan sát với dự đoán qua mô hình và c) Sai số theo dự đoán

Từ Hình 7.21 cho thấy, mô hình bám khá sát dữ liệu quan sát, tuy nhiên, biến động sai số cũng khá lớn ở các giá trị dự đoán lớn, cho thấy cần sử dụng trong số để lập mô hình nhằm cải thiện sai số. (Sử dụng Weight được trình bày ở phần tiếp theo của giáo trình này).

### 7.5.4 Mô hình phi tuyến có hay không có trọng số (Weight)

Qua thiết lập một số mô hình phi tuyến nói trên, cho thấy một số mối quan hệ ví dụ như $H_o/A$ hay AGB/DBH, dữ liệu biến phụ thuộc $H_o$ hay AGB sẽ càng phân tán (biến động lớn) khi biến độc lập như A hoặc DBH tăng lên (Hình 7.22); điều này làm cho mô hình có sai số lớn ở các giá trị dự báo cao. Để cải thiện điều này, mô hình phi tuyến có trọng số cần được áp dụng (Poso et al., 1999, Picard et al., 2012).



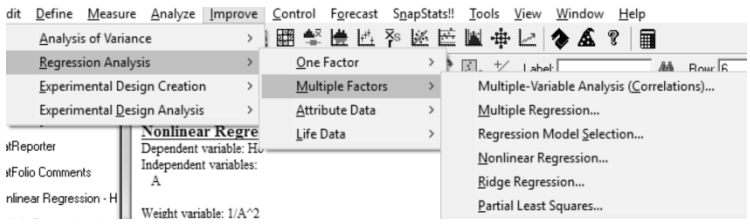
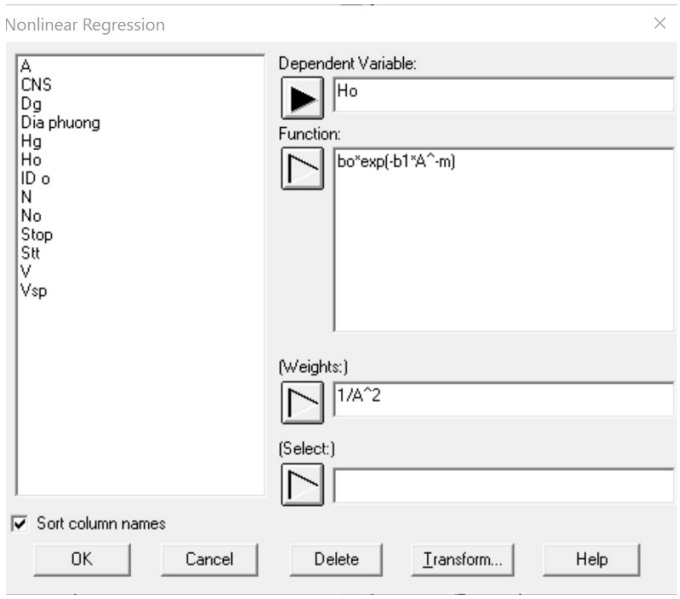

**Hình 7.22.** Biến động AGB tăng khi DBH tăng. Mô hình cần áp dụng có trọng số Weight = $1/DBH^a$

Biến số trọng số được tính: Weight = $1/X^a$ trong đó X là biến số độc lập ảnh hưởng rõ rệt nhất đến biến phụ thuộc và khi nó tăng lên thì biến phụ thuộc có sự phân hóa cao; a thường biến động từ -20 đến +20 (Picarrd et al., 2012); thay đổi a để mô hình có sai số phân bố đều theo các giá trị dự đoán trên đồ thị.

#### 7.5.4.1 Mô hình theo phương pháp phi tuyến theo Marquardt có trọng số thực hiện trong Statgraphics

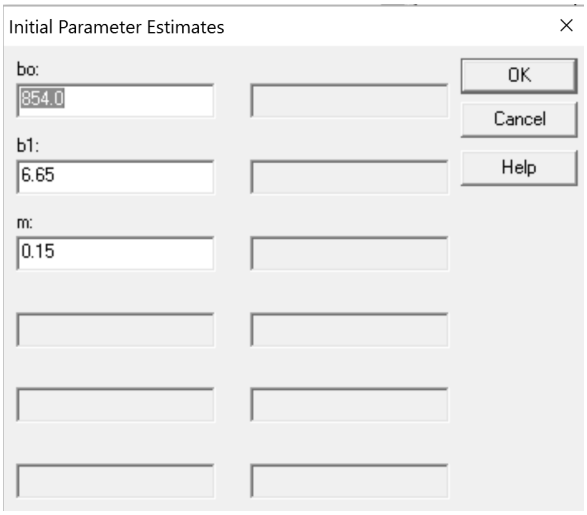
Trong lập mô hình $H_o$ theo A của rừng trồng trám trắng (Dữ liệu 12) theo hàm Schumacher như phần trên theo phương pháp phi tuyến tính Marquardt cho thấy sai số biến động lớn khi A tăng lên; vì vậy cần lập mô hình có trọng số.

Tiến hành lập mô hình $H_o$/A hàm Schumacher theo phương pháp phi tuyến Marquarrdt có trọng số từ dữ liệu 12: $Ho = bo \times \exp(-b_1 \times A^{-m}) + \varepsilon$

| | |
|---|---|
| Thực hiện thiết lập mô hình phi tuyến theo Marquardt trong Statgraphics:<br>Imp[rove / Regression Analysis / Multiple Factors / Nonlinear Regression |  |
| Trong hộp thoại Nonlinear Regression:<br>Depenedent Variable: $H_o$<br>Function: Nhập hàm Schumacher<br>Weights: Biến trọng số: $1/A^a$, trong đó A là biến số độc lập ảnh hưởng dự phân hóa $H_o$ và a thay đổi từ -20 đến +20 để chọn mô hình có các chỉ tiêu thống kê tốt nhất như $R^2$ cao, các sai số bé nhất và đặc biệt là sai số có phân bố đều theo giá trị dự đoán trên đồ thị residuals. Trong trường này chon a = 2 |  |

Trong hộp thoại Initial Parameters Estimates (Dự kiến tham số ban đầu):

Nhập các tham số ban đầu được xác định từ phương pháp logarit tuyến tính hóa (đã giới thiệu ở phần trên)




Kết quả thiết lập mô hình phi tuyến tính Marquardt có trọng số trong Statgraphics:

**Nonlinear Regression - $H_o$**
Dependent variable: $H_o$
Independent variables:
A

Weight variable: 1/A^2
Function to be estimated: $b_o$*exp(-$b_1$*A^-m)
Initial parameter estimates:
bo = 854.0
b1 = 6.65
m = 0.15

Estimation method: Marquardt
Estimation stopped due to convergence of residual sum of squares.
Number of iterations: 2
Number of function calls: 10

**Estimation Results**

| | | | *Asymptotic* | *95.0%* |
|---|---|---|---|---|
| | | *Asymptotic* | *Confidence* | *Interval* |
| *Parameter* | *Estimate* | *Standard Error* | *Lower* | *Upper* |
| bo | 860.434 | 15576.3 | -30256.9 | 31977.7 |
| b1 | 6.63987 | 17.9138 | -29.1472 | 42.427 |
| m | 0.150432 | 0.526199 | -0.900775 | 1.20164 |

**Analysis of Variance**

| *Source* | *Sum of Squares* | *Df* | *Mean Square* |
|---|---|---|---|
| Model | 49.7752 | 3 | 16.5917 |
| Residual | 2.15216 | 64 | 0.0336275 |
| Total | 51.9274 | 67 | |
| Total (Corr.) | 5.97288 | 66 | |

R-Squared = 63.9678 percent
R-Squared (adjusted for d.f.) = 62.8418 percent
Standard Error of Est. = 0.183378
Mean absolute error = 0.885531
Durbin-Watson statistic = 1.79717
Lag 1 residual autocorrelation = 0.0765417

**Residual Analysis**

| | *Estimation* | *Validation* |
|---|---|---|
| n | 67 | |
| MSE | 0.0336275 | |
| MAE | 0.885531 | |
| MAPE | 17.7424 | |
| ME | -0.00049614 | |
| MPE | -4.37533 | |

**The StatAdvisor**
The output shows the results of fitting a nonlinear regression model to describe the relationship between $H_o$ and 1 independent variables. The equation of the fitted model is

Ho = 860.434*exp(-6.63987*A^-0.150432)

In performing the fit, the estimation process terminated successully after 2 iterations, at which point the estimated coefficients appeared to converge to the current estimates.

The R-Squared statistic indicates that the model as fitted explains 63.9678% of the variability in Ho. The adjusted R-Squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 62.8418%. The standard error of the estimate shows the standard deviation of the residuals to be 0.183378. This value can be used to construct prediction limits for new observations by selecting the Forecasts option from the text menu. The mean absolute error (MAE) of 0.885531 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file.

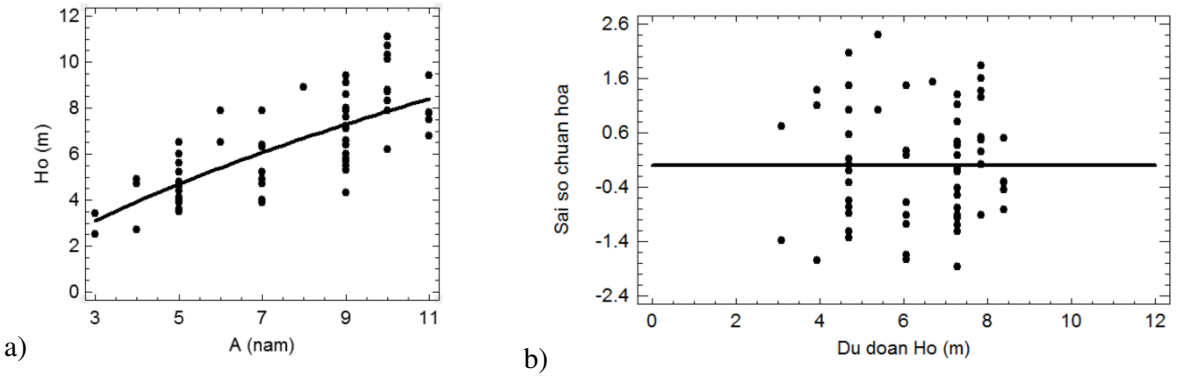
The output also shows aymptotic 95.0% confidence intervals for each of the unknown parameters. These intervals are approximate and most accurate for large sample sizes. You can determine whether or not an estimate is statistically significant by examining each interval to see whether it contains the value 0. Intervals covering 0 correspond to coefficients which may well be removed form the model without hurting the fit substantially.

Mô hình: $H_o$ = 860.434*exp(-6.63987*A^-0.150432)

Với Weight được lựa chọn = $1/A^2$ $R^2_{adj}$ = 62.8418%.

Các sai số chính: Bias (ME) = -0.00049614 m; MAPE = 17.74%

Các chỉ tiêu thống kê này đề tốt hơn mô hình không có trọng số đã lập ở phần trên, chứng tỏ khi dùng trọng số đối với dữ liệu bị phân hóa mạnh đã cải thiện được sai số của mô hình. Hình 7.23 cho thấy, mô hình có trọng số đã giúp cho sai số có phân bố đều theo các giá trị dự báo, cải thiện được tình trạng sai số phân tán mạnh khi giá trị dự đoán tăng trong mô hình không có trọng số.




a) b)

**Hình 7.23.** Mô hình: $H_o$ = 860.434*exp(-6.63987*A^-0.150432) với Weight = $1/A^2$.
a) Mô hình với quan sát; b) sai số theo dự đoán

#### 7.5.4.2 So sánh mô hình theo phương pháp phi tuyến bình phương tối thiểu (nls) có hay không có trọng số thực hiện R

Mô hình phi tuyến tính có hay không có trọng số có thể được thiết lập theo chương trình phi tuyến bình phương tối thiểu nls (nonlinear least square) (Bate et al., 1988) trong phần mềm mã nguồn mở R.

Dưới đây minh họa cách thiết lập các mô hình ước tính sinh khối cây rừng trên mặt đất (AGB, kg) theo các biến số DBH, H, WD được tổ hợp biến là $DBH^2HWD$ dạng hàm power (Huy et al. 2016b) trên cơ sở dữ liệu 11 của 110 cây mẫu ở vùng Nam Trung Bộ, so sánh mô hình có hay không có trọng số thực hiện theo code nls trong R.

$$AGB = a \times (DBH^2HWD)^b + \varepsilon \quad (7.38)$$

Codes nhập dữ liệu đầu vào:

```
# Erase memory
rm(list=ls())
# Clean plot window
dev.off()
# Define the working directory (change \ with / using Edit>Find)
setwd("E:/1 - Bao Huy 2017/Books All/Textbook for Infomatic Statistic/TextbookDataset Analysis/Dataset")

# Import data
t_eq <- read.table("Du lieu 11 AGB QNam .txt", header=T,sep="\t",stringsAsFactors = FALSE)
# install.packages for plots
library(ggplot2)
library(cowplot)
library(gridExtra)
```

**Codes lập mô hình phi tuyến không có trọng số theo nls trong R:**

```
# Tổ hợp biến đại diện sinh khối AGB:
t_eq$DBH2HWD = (t_eq$DBH/100)^2*t_eq$H*t_eq$WD*1000

# Initial parameters:
start <- coefficients(lm(log(AGB)~log(DBH2HWD), data=t_eq))
names(start) <- c("a","b")
start[1] = exp(start[1])

nls_least_square <- nls(AGB~a*DBH2HWD^b, data=t_eq,
              start=start)

# Tóm tắt kết quả mô hình
summary(nls_least_square)
# Dự đoán và sai số mô hình:
t_eq$nls_least_square.fit <- fitted.values(nls_least_square)
t_eq$nls_least_square.res <- residuals(nls_least_square)

# calcul of AIC, R2 and errors:
AIC = AIC(nls_least_square)
AIC
R2 <- 1- sum((t_eq$AGB - t_eq$nls_least_square.fit)^2)/sum((t_eq$AGB -
mean(t_eq$AGB))^2)
R2.adjusted <- 1 - (1-R2)*(length(t_eq$AGB)-1)/(length(t_eq$AGB)-3-1)
R2.adjusted

Bias = mean(t_eq$nls_least_square.res)
Bias
RMSE =  sqrt(mean(t_eq$nls_least_square.res^2))
RMSE
MAPE = 100*mean(abs(t_eq$nls_least_square.res)/t_eq$AGB)
MAPE

# Plots:
# Fitted model (Mô hình theo quan sát)
p1 <- ggplot(t_eq)
p1 <- p1 + geom_point(aes(x=DBH2HWD, y=AGB))
p1 <- p1 + geom_line(cex = 1.5, aes(x=DBH2HWD, y=nls_least_square.fit))
p1 <- p1 + xlab("DBH2HWD (kg)") + ylab("AGB (kg)") + theme_bw()
p1 <- p1 + labs(title = "a")
p1 = p1 + theme(axis.title.y = element_text(size = rel(1.5)))
p1 = p1 + theme(axis.title.x = element_text(size = rel(1.5)))
p1 <- p1 + theme(plot.title = element_text(size = rel(1.7)))
p1 = p1 + theme(axis.text.x = element_text(size=15))
p1 = p1 + theme(axis.text.y = element_text(size=15))
```

```
p1

# Observated and Predicted Values: (Đồ thị Quan sát với dự báo):
p2 <- ggplot(t_eq, aes(x=t_eq$nls_least_square.fit, y=AGB))
p2 <- p2 + geom_point(cex=2)
p2 <- p2 + geom_abline(cex = 1.5, intercept = 0, slope = 1, col="black")
p2 <- p2 + xlab("Dự đoán (kg)") + ylab("Quan sát AGB (kg)") + theme_bw()+ theme_bw()
p2 = p2 + theme(axis.title.y = element_text(size = rel(1.5)))
p2 = p2 + theme(axis.title.x = element_text(size = rel(1.5)))
p2 <- p2 + theme(plot.title = element_text(size = rel(1.7)))
p2 = p2 + theme(axis.text.x = element_text(size=15))
p2 = p2 + theme(axis.text.y = element_text(size=15))
p2 <- p2 + labs(title = "b")
p2

# Residuals vs predicted (Quan hệ sai số với dự đoán)
p3 <- ggplot(t_eq, aes(x=nls_least_square.fit, y=nls_least_square.res))
p3 <- p3 + geom_point()
p3 <- p3 + geom_line(cex = 1.5, aes(x=nls_least_square.fit, y=0))
p3 <- p3 + xlab("Dự đoán AGB (kg)") + ylab("Sai số (kg)") + theme_bw()
p3 <- p3 + labs(title = "")
p3 = p3 + theme(axis.title.y = element_text(size = rel(1.5)))
p3 = p3 + theme(axis.title.x = element_text(size = rel(1.5)))
p3 <- p3 + theme(plot.title = element_text(size = rel(1.7)))
p3 = p3 + theme(axis.text.x = element_text(size=15))
p3 = p3 + theme(axis.text.y = element_text(size=15))
p3 <- p3 + labs(title = "c")
p3 = p3 + ylim(-1500,1500)
p3

plot_grid(p1, p2, p3, ncol = 1)
```

**Codes lập mô hình phi tuyến có trọng số theo nls trong R:**

```
# Tổ hợp biến đại diện sinh khối AGB:
t_eq$DBH2HWD = (t_eq$DBH/100)^2*t_eq$H*t_eq$WD*1000

# Initial parameters:
start <- coefficients(lm(log(AGB)~log(DBH2HWD), data=t_eq))
names(start) <- c("a","b")
start[1] = exp(start[1])

nls_least_square <- nls(AGB~a*DBH2HWD^b, data=t_eq,
                        start=start, weights = 1/DBH2HWD^0.9)

# Tóm tắt kết quả mô hình
```

```
summary(nls_least_square)
# Dự đoán và sai số mô hình:
t_eq$nls_least_square.fit <- fitted.values(nls_least_square)
t_eq$nls_least_square.res <- residuals(nls_least_square)
t_eq$nls_least_square.res_weight <- residuals(nls_least_square)/t_eq$DBH2HWD^0.9

# calcul of AIC, R2 and errors:
AIC = AIC(nls_least_square)
AIC
R2 <- 1- sum((t_eq$AGB - t_eq$nls_least_square.fit)^2)/sum((t_eq$AGB - mean(t_eq$AGB))^2)
R2.adjusted <- 1 - (1-R2)*(length(t_eq$AGB)-1)/(length(t_eq$AGB)-3-1)
R2.adjusted

Bias = mean(t_eq$nls_least_square.res_weight)
Bias
RMSE =  sqrt(mean(t_eq$nls_least_square.res_weight^2))
RMSE
MAPE = 100*mean(abs(t_eq$nls_least_square.res_weight)/t_eq$AGB)
MAPE

# Plots:
# Fitted model (Mô hình theo quan sát)
p4 <- ggplot(t_eq)
p4 <- p4 + geom_point(aes(x=DBH2HWD, y=AGB))
p4 <- p4 + geom_line(cex = 1.5, aes(x=DBH2HWD, y=nls_least_square.fit))
p4 <- p4 + xlab("DBH2HWD (kg)") + ylab("AGB (kg)") + theme_bw()
p4 <- p4 + labs(title = "a")
p4 = p4 + theme(axis.title.y = element_text(size = rel(1.5)))
p4 = p4 + theme(axis.title.x = element_text(size = rel(1.5)))
p4 <- p4 + theme(plot.title = element_text(size = rel(1.7)))
p4 = p4 + theme(axis.text.x = element_text(size=15))
p4 = p4 + theme(axis.text.y = element_text(size=15))
p4

# Observated and Predicted Values: (Đồ thị Quan sát với dự báo):
p5 <- ggplot(t_eq, aes(x=t_eq$nls_least_square.fit, y=AGB))
p5 <- p5 + geom_point(cex=2)
p5 <- p5 + geom_abline(cex = 1.5, intercept = 0, slope = 1, col="black")
p5 <- p5 + xlab("Dự đoán AGB (kg)") + ylab("Quan sát AGB (kg)") + theme_bw()+ theme_bw()
p5 = p5 + theme(axis.title.y = element_text(size = rel(1.5)))
p5 = p5 + theme(axis.title.x = element_text(size = rel(1.5)))
p5 <- p5 + theme(plot.title = element_text(size = rel(1.7)))
p5 = p5 + theme(axis.text.x = element_text(size=15))
p5 = p5 + theme(axis.text.y = element_text(size=15))
p5 <- p5 + labs(title = "b")
```

```
p5

p6 <- ggplot(t_eq, aes(x=nls_least_square.fit, y=t_eq$nls_least_square.res_weight))
p6 <- p6 + geom_point()
p6 <- p6 + geom_line(cex = 1.5, aes(x=nls_least_square.fit, y=0))
p6 <- p6 + xlab("Dự đoán AGB (kg)") + ylab("Sai số có trọng số (kg)") + theme_bw()
p6 <- p6 + labs(title = "")
p6 = p6 + theme(axis.title.y = element_text(size = rel(1.5)))
p6 = p6 + theme(axis.title.x = element_text(size = rel(1.5)))
p6 <- p6 + theme(plot.title = element_text(size = rel(1.7)))
p6 = p6 + theme(axis.text.x = element_text(size=15))
p6 = p6 + theme(axis.text.y = element_text(size=15))
p6 <- p6 + labs(title = "c")
p6 = p6 + ylim(-0.6,0.6)
p6

plot_grid(p4, p5, p6, ncol = 1)
```

Kết quả thiết lập mô hình có hay không có trọng số theo chương trình nls trong phần mềm R được tổng hợp trong Bảng 7.4. Từ bảng này cho thấy mô hình phi tuyến có trọng số thực hiện theo nls trong R có các sai số bé hơn rất nhiều mô hình không có trọng số.

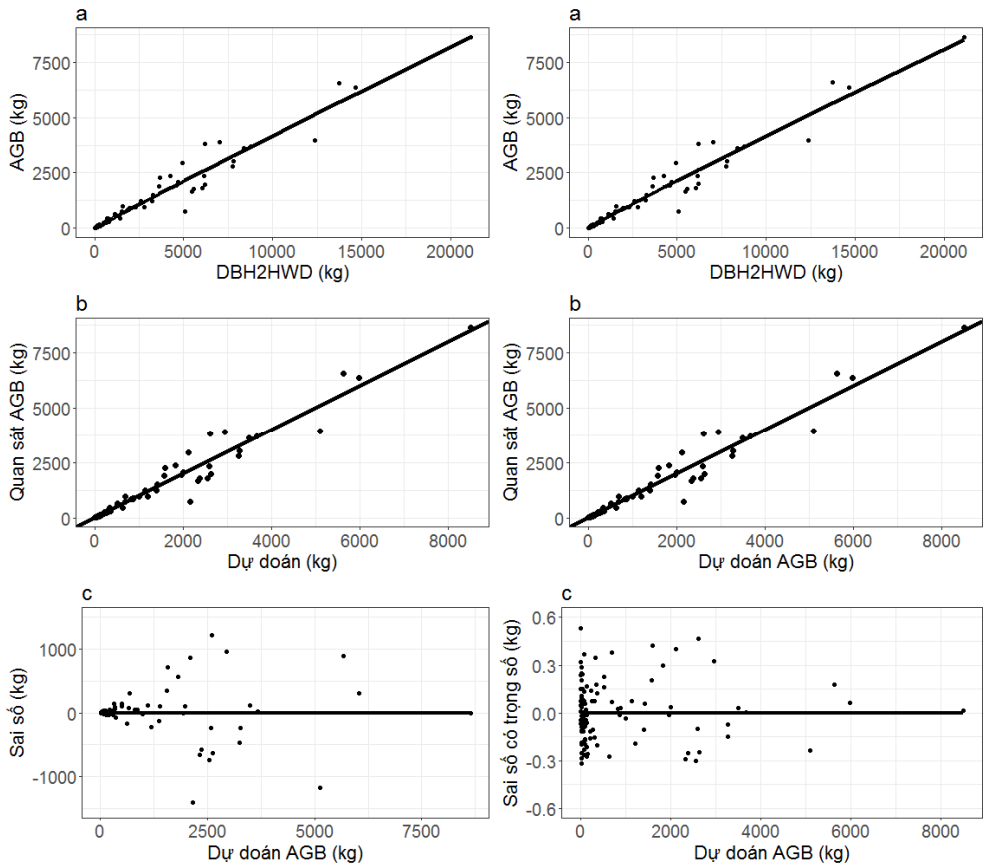
Vì vậy, đối với các dữ liệu quan sát bị phân tán khi biến độc lập tăng lên cần áp dụng phương pháp phi tuyến có trọng số, phương pháp giúp cho việc cải thiện sai số rất có ý nghĩa.

**Bảng 7.4.** Mô hình phi tuyến và các chỉ tiêu thống kê của mô hình có hay không có trọng số theo phương pháp phi tuyến bình phương tối thiểu (nls) trong R

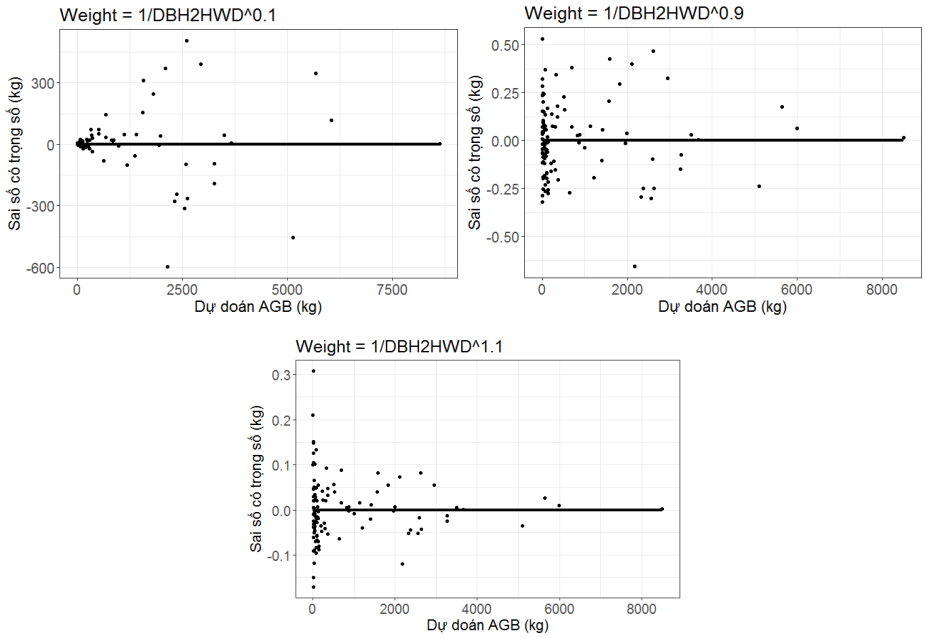
| Chỉ tiêu thống kê, sai số | Không trọng số | Có trọng số Weight = $1/DBH^2HWD^{0.9}$ |
|---|---|---|
| Mô hình | $AGB = 0.51083 \times (DBH^2HWD)^{0.97776}$ | $AGB = 0.60670 \times (DBH^2HWD)^{0.95883}$ |
| $R^2_{adj.}$ | 0.954 | 0.954 |
| Bias (kg) | 6.83 | -0.00065 |
| RMSE (kg) | 311.9 | 0.2 |
| MAPE (%) | 18.8 | 0.41 |

Hình 7.24 cũng cho thấy mô hình phi tuyến có trọng số thì sai số rất bé và phân bố đều theo giá trị dự báo, ngược lại khi không có trọng số sai số lớn và phân tán mạnh ở các giá trị AGB lớn.

Kết quả này khẳng định đối với các quan hệ nếu biến y bị phân tán mạnh khi x tăng lên, như trường hợp các mô hình ước tính sinh khối cây rừng (AGB) theo một đến nhiều biến như DBH, H, WD và CA hoặc mô hình quan hệ chiều cao bình quân tầng trội theo tuổi $H_o/A$ hoặc quan hệ H/DBH; thì mô hình có trọng số có ý nghĩa quan trọng trong cải thiện độ tin cậy như đã giới thiệu ở trên.




**Hình 7.24.** Mô hình phi tuyến AGB = a×DBH$^2$HWD. A) Mô hình theo quan sát; b) Quan sát so với dự đoán và c) phân bố sai số theo dự đoán. Cột trái: Mô hình không có trọng số; Cột phải: Mô hình với trọng số tối ưu




**Hình 7.25.** Mô hình phi tuyến AGB = a×DBH$^2$HWD với trọng số khác nhau. Weight = 1/DBH$^2$HWD$^{0.9}$ cho sai số bé nhất và phân bố đều theo giá trị AGB dự đoán quan mô hình

Hình 7.25 chỉ ra khi thay đổi trọng số khác nhau thì phân bố sai số cũng thay đổi, đây là cơ sở để lựa chọn một trọng số thích hợp cho mô hình. Phân bố sai số rải đều theo giá trị dự báo với Weight = 1/DBH2HWD$^{0.9}$ là tốt nhất trong ví dụ này.

### 7.5.5 Phương pháp phi tuyến ảnh hưởng tổng hợp (Nonlinear Mixed-Effects - nlme) Maximum Likelihood có trọng số để ước lượng mô hình phi tuyến tính

Ngoài phương pháp phi tuyến bình phương tối thiểu có trọng số sử dụng codes nls trong phần mềm mã nguồn mở R, còn có phương pháp phi tuyến ảnh hưởng phức hợp hợp lý tối đa (nlme: nonlinear mixed effects - Maximum Likelihood) có trọng số (weights), sau đây viết tắt là phương pháp phi tuyến Maximum Likelihood có trọng số (Davidian et al., 1995; Pinheiro et al., 2014). Trong một số quan hệ phức tạp như, mô hình bị ảnh hưởng của nhiều nhân tố khác ngoài các biến số độc lập, thì mô hình theo phương pháp nlme Maximum Likelohood có trọng số tỏ ra có hiệu quả để nâng cao độ tin cậy, vì vậy, cần có thử nghiệm để áp dụng phương pháp này so với nls thông thường và hay áp dụng.

Để minh họa áp dụng phương pháp phi tuyến Maximum Likelihood có trọng số, sử dụng mô hình power. Phần mềm mã nguồn mở R được áp dụng theo chương trình nlme (Bates *et al.,* 2010; Pinheiro *et al.,* 2014) và chẩn đoán qua sơ đồ sử dụng ggplot2 (Wickham *et al.,* 2013). Sử dụng Dữ liệu 11 để thiết lập mô hình ước tính khối cây rừng theo các biến số đầu vào khác nhau với kiểu dạng mô hình tổng quát như sau (Huy *et al.,* 2016a):

$$Y_i = \alpha \times X_i^{\beta} + \varepsilon_i \quad (7.39)$$

$$\varepsilon_i \sim iid\ N(0, \sigma^2) \quad (7.40)$$

Trong đó $Y_i$ là AGB (kg) ứng với cây thứ *i*; $\alpha$ và $\beta$ là tham số của mô hình; $X_i$ là các biến số DBH (cm), H (m), WD (g/cm$^3$), CA (m$^2$) hoặc tổ hợp biến DBH$^2$H, DBH$^2$HWD*;* và $\varepsilon_i$ là sai số ngẫu nhiên ứng với cây thứ *i.*

Phân tích ban đầu cho thấy, biến động của sai số có xu hướng gia tăng khi gia tăng $X_i$ trong các mô hình. Vì vậy, một hàm phương sai theo trọng số đã được áp dụng để điều chỉnh các tham số của mô hình nhằm giảm biến động sai số này. Hàm phương sai có dạng như sau (Huy *et al.,* 2016a):

$$Var(\varepsilon_i) = \widehat{\sigma^2}(\gamma_i)^{2k} \quad (7.41)$$

Trong đó $\varepsilon_i$ là sai số ngẫu nhiên; $\widehat{\sigma^2}$ là sai số bình phương; $\nu_i$ là biến trọng số (DBH, DBH$^2$HWD) tương ứng với cây thứ i; và k là hệ số của hàm phương sai.

Từ dữ liệu 11, minh họa thiết lập mô hình power theo codes nlme có trọng số Weight = 1/DBH$^2$HWWD trong R như sau:

$$AGB = a \times (DBH^2HWD)^b + \varepsilon \quad (7.42)$$

**Codes theo phương pháp nlme Maximum Likelihood có trọng số trong R cho mô hình: $AGB = a \times (DBH^2HWD)^b$ với Weight = 1/DBH$^2$HWWD:**

```
# Erase memory (Xóa bộ nhớ cũ)
rm(list=ls())

# Clean plot window (Xóa các cữa sổ cũ)
dev.off()

# Define the working directory (Thư mục chứa dữ liệu)
setwd("E:/1 - Bao Huy 2017/Books All/Textbook for Infomatic Statistic/TextbookDataset Analysis/Dataset")

# Import data (Nhập dữ liệu)
t <- read.table("Du lieu 11 AGB QNam .txt", header=T,sep="\t",stringsAsFactors = FALSE)

# Install.packages ggplot2 and nlme (Cài đặt chương trình nlme, ggplots)
library(ggplot2)
library(nlme)
library(cowplot)

# Combination of variables: (Tổ hợp biến)
t$DBH2HWD = (t$DBH/100)^2*t$H*t$WD*1000

# Model nlme (Mô hình theo nlme có trọng số)
start <- coefficients(lm(log(AGB)~log(DBH2HWD), data=t))
names(start) <- c("a","b")
start[1]<-exp(start[1])

Max_like <- nlme(AGB~a*(DBH2HWD)^b, data=cbind(t,g="a"), fixed=a+b~1,
                 start=start, groups=~g, weights=varPower(form=~DBH2HWD))

# Outputs of the model (Kết quả mô hình)
summary(Max_like)

k <- summary(Max_like)$modelStruct$varStruct[1]
k

t$Max_like.fit <- fitted.values(Max_like)
t$Max_like.res <- residuals(Max_like)
t$Max_like.res.weigh <- residuals(Max_like)/t$DBH2HWD^k

# Calcul of errors (Các sai số, chỉ tiêu thống kê của mô hình)
Bias = mean(t$Max_like.res.weigh)
Bias
```

```
RMSE = sqrt(mean((t$Max_like.res.weigh)^2))
RMSE
MAPE = 100*mean(abs(t$Max_like.res.weigh/t$AGB))
MAPE

AIC(Max_like)

R2 <- 1- sum((t$AGB - t$Max_like.fit)^2)/sum((t$AGB - mean(t$AGB))^2)
R2
R2.adjusted <- 1 - (1-R2)*(length(t$DBH)-1)/(length(t$DBH)-3-1)
R2.adjusted

# Plots (Các đồ thị)
# Fitted model (Mô hình theo quan sát)
p4 <- ggplot(t)
p4 <- p4 + geom_point(aes(x=DBH2HWD, y=AGB))
p4 <- p4 + geom_line(cex = 1.5, aes(x=DBH2HWD, y=Max_like.fit))
p4 <- p4 + xlab("DBH2HWD (kg)") + ylab("AGB (kg)") + theme_bw()
p4 <- p4 + labs(title = "a")
p4 = p4 + theme(axis.title.y = element_text(size = rel(1.5)))
p4 = p4 + theme(axis.title.x = element_text(size = rel(1.5)))
p4 <- p4 + theme(plot.title = element_text(size = rel(1.7)))
p4 = p4 + theme(axis.text.x = element_text(size=15))
p4 = p4 + theme(axis.text.y = element_text(size=15))
p4

# Observated and Predicted Values: (Quan sát và dự đoán):
p5 <- ggplot(t, aes(x=t$Max_like.fit, y=AGB))
p5 <- p5 + geom_point(cex=2)
p5 <- p5 + geom_abline(cex = 1.5, intercept = 0, slope = 1, col="black")
p5 <- p5 + xlab("Dự doán AGB (kg)") + ylab("Quan sát AGB (kg)") + theme_bw()+ theme_bw()
p5 = p5 + theme(axis.title.y = element_text(size = rel(1.5)))
p5 = p5 + theme(axis.title.x = element_text(size = rel(1.5)))
p5 <- p5 + theme(plot.title = element_text(size = rel(1.7)))
p5 = p5 + theme(axis.text.x = element_text(size=15))
p5 = p5 + theme(axis.text.y = element_text(size=15))
p5 <- p5 + labs(title = "b")
p5

# Residuals and predicted (Sai số theo dự đoán)
p6 <- ggplot(t, aes(x=Max_like.fit, y=t$Max_like.res.weigh))
p6 <- p6 + geom_point()
p6 <- p6 + geom_line(cex = 1.5, aes(x=Max_like.fit, y=0))
p6 <- p6 + xlab("Dự doán AGB (kg)") + ylab("Sai số có trọng số (kg)") + theme_bw()
p6 <- p6 + labs(title = "")
```

```
p6 = p6 + theme(axis.title.y = element_text(size = rel(1.5)))
p6 = p6 + theme(axis.title.x = element_text(size = rel(1.5)))
p6 <- p6 + theme(plot.title = element_text(size = rel(1.7)))
p6 = p6 + theme(axis.text.x = element_text(size=15))
p6 = p6 + theme(axis.text.y = element_text(size=15))
p6 <- p6 + labs(title = "c")
p6 = p6 + ylim(-0.6,0.6)
p6
plot_grid(p4, p5, p6, ncol = 1)
# The end #
```

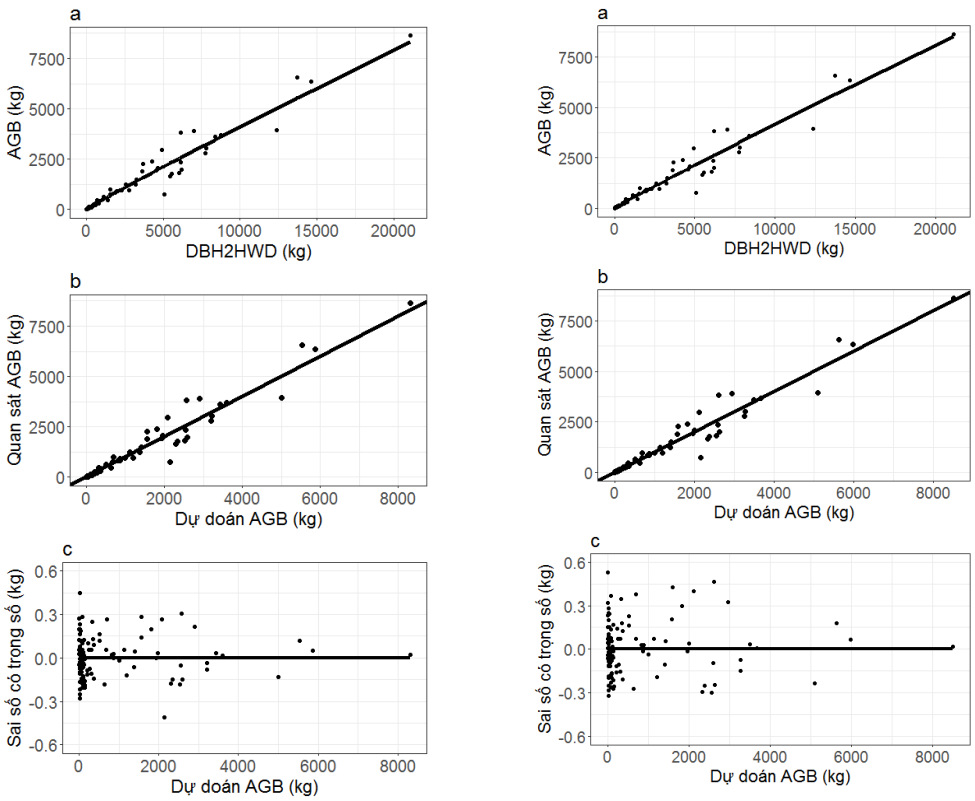
Kết quả thiết lập mô hình phi tuyến theo phương pháp nlme Maximum Likelihood có trọng số ở Bảng 7.5 và so sánh với mô hình theo phương pháp phi tuyến bình phương tối thiểu nls có trọng số được dò tìm. Kết quả ở bảng này cho thấy, đối với các mô hình có ảnh hưởng phức tạp, phương pháp nlme đã tỏ ra có độ tin cậy hơn, với AIC nhỏ hơn rõ rệt (một điều quan tâm là giá trị AIC trong mô hình theo nls có giá trị khá cao) và các sai số như Bias, RMSE và MAPE cũng được cải thiện hơn. Trong khi đó $R^2$ của mô hình theo nlme có xu hướng bé hơn, do đó, trong thống kê truyền thống thường chỉ sử dụng $R^2$ như là chỉ tiêu thống kê duy nhất để so sánh các mô hình là chưa phù hợp và đôi khi sẽ không mang lại kết quả đúng.

**Bảng 7.5.** Mô hình phi tuyến và các chỉ tiêu thống kê theo phương pháp Maximum Like Lihood (nlme) và phi tuyến bình phương tối thiểu (nls) có trọng số trong R

| Chỉ tiêu thống kê, sai số | Theo nlme với Weight = $1/DBH^2HWD^k$ | Theo nls với Weight = $1/DBH^2HWD^{0.9}$ |
|---|---|---|
| Mô hình | $AGB = 0.62596 \times (DBH^2HWD)^{0.95345}$ | $AGB = 0.60670 \times (DBH^2HWD)^{0.95883}$ |
| k | 0.95186 | - |
| AIC | 1098.9 | 62438.6 |
| $R^2_{adj.}$ | 0.953 | 0.954 |
| Bias (kg) | 1.225314e-07 | -0.00065 |
| RMSE (kg) | 0.14 | 0.20 |
| MAPE (%) | 0.35 | 0.41 |

*Ghi chú: k là hệ số của hàm phương sai*

Hình 7.26 cũng cho thấy ước lượng mô hình phi tuyến theo phương pháp nlme Maximum Likelihood so với phương pháp nls có trọng số thì sai số bé hơn và phân bố hẹp và đều hơn theo giá trị dự báo. Kết quả này cho thấy, phương pháp ước lượng mô hình phi tuyến tính Maximum Likelihood có trọng số Weight = $1/X^k$ (với k là hệ số của hàm phương sai) sử dụng codes nlme tỏ ra có hiệu quả cao nhất so với các phương pháp ước lượng là phi tuyến đối với các mối quan hệ phi tuyến phức tạp, bị chi phối bởi nhiều nhân tố bên ngoài. Vì vậy, xu thế áp dụng codes nlme trong R là hiện đại và cần được xem xét áp dụng trong thiết lập các mô hình phi tuyến trong lâm nghiệp.




**Hình 7.26.** Mô hình phi tuyến AGB = a×DBH$^2$HWD. a) Mô hình theo quan sát; b) Quan sát so với dự đoán và c) Phân bố sai số theo dự đoán. Cột trái: Mô hình theo phương pháp nlme có trọng số Weight = 1/(DBH$^2$HWD)$^k$; Cột phải: Mô hình theo phương pháp nls với trọng số Weight = 1/(DBH$^2$HWWD)$^{0.9}$

## 7.6 Chỉ Số Furnival's Index để lựa chọn dạng phương trình khác nhau hoặc phương pháp ước lượng mô hình phi tuyến: Tuyến tính hóa hay phi tuyến Maximum Likelihood

Các mô hình có thể khác nhau về biến phụ thuộc y, một mô hình không đổi biến số vẫn là y, mô hình khác thì đổi biến ví dụ như là log(y), 1/y, sqrt(y). Khi các mô hình khác nhau về đổi biến số thì cũng áp dụng phương pháp ước lượng khác nhau. Phổ biến là mô hình phi tuyến có thể ước lượng các tham số của mô hình theo một trong hai phương pháp chính là, tuyến tính hóa bình phương tối thiểu (sử dụng codes lm trong R) hoặc phi tuyến bao gồm: phi tuyến bình phương tối thiểu (sử dụng codes nls trong R) hoặc phi tuyến Marquardt hoặc phi tuyến Maximum Likelihood (code nlme trong R).

Để so sánh các mô hình khác nhau về biến y (y và log(y), 1.y,...) được ước lượng theo phương pháp khác nhau cơ bản như là, tuyến tính hoặc phi tuyến tính; lúc này cần sử dụng chỉ số Furnival's Index (FI) (1961) (Jayaraman, 1999)). Chỉ số Furnival's Index dùng để so sánh các mô hình không giống nhau về biến số phụ thuộc (ví dụ y và ln(y)), vì lúc này các chỉ tiêu thống kê như $R^2$, AIC, các loại sai số sẽ không thể dùng để so sánh do khác nhau về giá trị của biến phụ thuộc. Các mô hình theo phương pháp ước lượng áp dụng có chỉ số Furnival Index (FI) thấp hơn là tốt hơn.

Công thức tính Furnival's Index (FI) như sau:

$$FI = RMSE * \frac{1}{Geometric\ mean\ (y')} \tag{7.43}$$

Trong đó: RMSE: Root Mean Squared Error: Sai số trung phương; y' là đạo hàm bậc nhất của biến phụ thuộc y và bằng 1, nếu là biến phụ thuộc được đổi biến số là ln(y) thì sẽ bằng 1/y. Geomtric mean: Trung bình hình học.

Công thức tính trung bình hình học (Geometric mean) (Huy et al., 2016b):

$$\text{Geometric Mean} = \left(\prod_{i=1}^{n} x_i\right)^{1/n} = \sqrt[n]{x_1 x_2 \cdots x_n}. \quad (7.44)$$

Tính FI cho từng dạng mô hình và phương pháp lập mô hình có thể được thực hiện trong R như sau: Để minh họa, sử dụng codes nêu trên để tính FI cho các mô hình ước tính AGB theo một biến số DBH hoặc tổ hợp biến số $DBH^2HWD + CA$ theo hai dạng mô hình tuyến tính hóa logarit và power ứng với hai phương pháp tuyến tính bình phương tối thiểu (lm trong R) và phi tuyến Maximum likelihood có trọng số (nlme trong R), dữ liệu 11 từ sinh khối 110 cây mẫu ở Nam Trung Bộ được áp dụng.

**Codes của chương trình R để tính chỉ số Furnival's Index cho hàm mũ tuyến tính logarit theo lm: log(AGB) = a + b log(DBH)**

```
# Erase memory
rm(list=ls())
# Clean plot window
dev.off()
# install.packages("ggplot2")
library(ggplot2)
library(nlme)
# Define the working directory
setwd("E:/1 - Bao Huy 2017/Books All/Textbook for Infomatic Statistic/TextbookDataset Analysis/Dataset")
# Import data
t <- read.table("Du lieu 11 AGB QNam .txt", header=T,sep="\t",stringsAsFactors = FALSE)

# Modelling
lmt1 <- lm(log(AGB)~I(log(DBH)), data=t)

# Outputs of the model
summary(lmt1)
anova(lmt1)
t$lmt1.fit <- fitted.values(lmt1)
t$lmt1.res <- residuals(lmt1)

# Furnival Index (FI)= RMSE*(1/Geometric Mean of ln(y)'), ln(y)' = 1/y
# Geometric Mean (gm):
gm = exp(mean(log(1/t$AGB)))
RMSE = sqrt(mean(t$lmt1.res^2))
FI = RMSE*(1/gm)
FI

# The end
```

**Codes của chương trình R để tính chỉ số Furnival's Index cho hàm mũ tuyến tính logarit theo lm: log(AGB) = a + b*log(DBH2HWD) + c*log(CA)**

```
# Erase memory
rm(list=ls())
# Clean plot window
dev.off()
# install.packages("ggplot2")
library(ggplot2)
library(nlme)
# Define the working directory
setwd("E:/1 - Bao Huy 2017/Books All/Textbook for Infomatic Statistic/TextbookDataset Analysis/Dataset")
# Import data
t <- read.table("Du lieu 11 AGB QNam .txt", header=T,sep="\t",stringsAsFactors = FALSE)

# Modelling
# Combination of variable:
t$DBH2HWD = (t$DBH/100)^2*t$H*t$WD*1000
lmt1 <- lm(log(AGB)~log(DBH2HWD)+log(CA), data=t)

# Outputs of the model
summary(lmt1)
anova(lmt1)
t$lmt1.fit <- fitted.values(lmt1)
t$lmt1.res <- residuals(lmt1)

# Furnival Index (FI)= RMSE*(1/Geometric Mean of ln(y)'), ln(y)' = 1/y
# Geometric Mean (gm):
gm = exp(mean(log(1/t$AGB)))
RMSE = sqrt(mean(t$lmt1.res^2))
FI = RMSE*(1/gm)
FI
# The end
```

**Codes của chương trình R để tính chỉ số Furnival's Index cho hàm mũ theo phương pháp phi tuyến Maximum Likelihood nlme có trọng số: AGB = aDBH^b**

```
# Erase memory
rm(list=ls())
# Clean plot window
dev.off()
# install.packages("ggplot2")
library(ggplot2)
library(nlme)
# Define the working directory
```

```
setwd("E:/1 - Bao Huy 2017/Books All/Textbook for Infomatic Statistic/TextbookDataset Analysis/Dataset")
# Import data
t <- read.table("Du lieu 11 AGB QNam .txt", header=T,sep="\t",stringsAsFactors = FALSE)

# Modelling
start <- coefficients(lm(log(AGB)~log(DBH), data=t))
names(start) <- c("a","b")
start[1]<-exp(start[1])
Max_like <- nlme(AGB~a*DBH^b, data=cbind(t,g="a"), fixed=a+b~1,
                 start=start, groups=~g, weights=varPower(form=~DBH))

# Outputs of the model
summary(Max_like)
k <- summary(Max_like)$modelStruct$varStruct[1]
k
t$Max_like.fit <- fitted.values(Max_like)
t$Max_like.res <- residuals(Max_like)
t$Max_like.res.weigh <- residuals(Max_like)/t$DBH^k

# Calcul of R2
R2 <- 1- sum((t$AGB - t$Max_like.fit)^2)/sum((t$AGB - mean(t$AGB))^2)
R2
R2.adjusted <- 1 - (1-R2)*(length(t$DBH)-1)/(length(t$DBH)-3-1)
R2.adjusted

# Furnival Index FI:
RMSE = sqrt(mean((t$Max_like.res.weigh)^2))
FI = RMSE
FI

# The End
```

**Codes của chương trình R để tính chỉ số Furnival's Index cho hàm mũ theo phương pháp phi tuyến Maximum Likelihood nlme có trọng số: AGB = a*DBH2HWD^b*CA^c**

```
# Erase memory
rm(list=ls())
# Clean plot window
dev.off()
# install.packages("ggplot2")
library(ggplot2)
library(nlme)
```

```
# Define the working directory
setwd("E:/1 - Bao Huy 2017/Books All/Textbook for Infomatic Statistic/TextbookDataset Analysis/Dataset")
# Import data
t <- read.table("Du lieu 11 AGB QNam .txt", header=T,sep="\t",stringsAsFactors = FALSE)

# Modelling
# Combination of variables
t$DBH2HWD = (t$DBH/100)^2*t$H*t$WD*1000
start <- coefficients(lm(log(AGB)~log(DBH2HWD)+log(CA), data=t))
names(start) <- c("a","b","c")
start[1]<-exp(start[1])
Max_like <- nlme(AGB~a*DBH2HWD^b*CA^c, data=cbind(t,g="a"), fixed=a+b+c~1,
                 start=start, groups=~g, weights=varPower(form=~DBH))

# Outputs of the model
summary(Max_like)
k <- summary(Max_like)$modelStruct$varStruct[1]
k
t$Max_like.fit <- fitted.values(Max_like)
t$Max_like.res <- residuals(Max_like)
t$Max_like.res.weigh <- residuals(Max_like)/t$DBH^k

# Calcul of R2
R2 <- 1- sum((t$AGB - t$Max_like.fit)^2)/sum((t$AGB - mean(t$AGB))^2)
R2
R2.adjusted <- 1 - (1-R2)*(length(t$DBH)-1)/(length(t$DBH)-4-1)
R2.adjusted

# Furnival Index FI:
RMSE = sqrt(mean((t$Max_like.res.weigh)^2))
FI = RMSE
FI
# The end
```

Trong trường hợp mô hình sinh khối, kết quả ở Bảng 7.6 đã chỉ ra lập mô hình theo phương pháp phi tuyến Maximum Likelihood có trọng số là tốt hơn nhiều so với phương pháp tuyến tính hóa logarit bình phương tối thiểu. Tất cả các hàm với các biến số đầu vào khác nhau theo phương pháp Maximum Likelihood có trọng số đều có giá trị FI nhỏ hơn rất nhiều, vì vậy, phương pháp này là tốt nhất để lập các mô hình sinh khối dạng hàm power với một đến nhiều biến số đầu, tổ hợp biến. Trong khi đó $R^2$ không thể sử dụng để chọn mô hình có độ tin cậy tốt, vì kết quả này cho thấy $R^2$ của các mô hình theo dạng logarrit bình phương tối thiểu có $R^2$ cao hơn mô hình theo Maximum Likelihood, vì vậy, khẳng định $R^2$ cũng như AIC hoặc các sai số đều không thể dùng để sáng trong trường hợp mô hình có biến y khác nhau và ước lượng theo hai phương pháp khác nhau là tuyến tính và phi tuyến.

**Bảng 7.6.** Các hàm thử nghiệm theo các nhóm biến số đầu vào và sử dụng chỉ số Furnival (FI) để so sánh hai phương pháp logarrit và phi tuyến tính Maximum Likelihood có trọng số

| Biến đầu vào | Mô hình sinh khối | Theo phương pháp tuyến tính hóa logarit bình phương tối thiểu | | Theo phương pháp phi tuyến có trọng số Maximum Likelihood | | |
|---|---|---|---|---|---|---|
| | | Adj. $R^2$ | FI | Biến trọng số | Adj. $R^2$ | FI |
| DBH | $AGB = a \times DBH^b$ | 0.982 | 39.2 | $1/DBH^k$ | 0.934 | 0.024 |
| DBH, H, WD, and CA | $AGB = a \times DBH^2 HWD^b \times CA^c$ | 0.988 | 31.3 | $1/DBH^k$ | 0.960 | 0.018 |

*Ghi chú: FI: Chỉ số Furnival's Index. Tổ hợp biến: $DBH^2HWD$ (kg) = $(DBH/100)^2 \times H \times WD \times 1000$ là đại diện cho sinh khối. k là hệ số của hàm biến động. Nguồn: Huy et al., 2016b*

## 7.7 Mô hình thay đổi tham số dưới ảnh hưởng của các nhân tố ngẫu nhiên (random effect)

Trong thực tế, thiết lập mô hình quan hệ biến phụ thuộc không chỉ bị ảnh hưởng bởi các biến độc lập, mà còn bị chi phối bởi các nhân tố môi trường khác. Ví dụ quan hệ H/DBH hay quan hệ $H_o$/A hoặc AGB/DBH thì các các biến phụ thuộc không chỉ bị ảnh hưởng bởi một hoặc nhiều trong các biến độc lập trong mô hình, mà còn bị ảnh hưởng bởi các nhân tố sinh thái, môi trường rừng khác (random effect), như mật độ (N), tổng tiết diện ngang (BA), lập địa, khí hậu, đất đai,... Trong trường hợp này, để nâng cao độ tin cậy của mô hình, các tham số của mô hình cần có sự điều chỉnh theo từng nhân tố môi trường; phương pháp ước lượng mô hình như vậy gọi là phi tuyến tính Maximum Likelihood có trọng số có xét đến ảnh hưởng ngẫu nhiên của các nhân tố môi trường.

Phương pháp ước lượng mô hình có xét đến ảnh hưởng ngẫu nhiên (random effect) của các nhân tố môi trường bằng cách, đánh giá sự thay đổi các tham số khi các nhân tố ảnh hưởng thay đổi được thực hiện theo chương trình nlme trong phần mềm mã nguồn mở R (Bates *et al.,* 2010; Pinheiro *et al.,* 2014) và đánh giá qua sơ đồ sử dụng codes ggplot2 (Wickham *et al.,* 2013).

Ví dụ kiểu dạng mô hình phi tuyến tổng quát dạng power để ước tính AGB có xét ảnh hưởng của các nhân tố môi trường (Huy *et al.,* 2016c):

$$Y_{ij} = (\alpha + a_i) \times X_{ij}^{(\beta + b_i)} + \varepsilon_{ij} \quad (7.45)$$

$$\varepsilon_{ij} \sim iid\ N(0, \sigma^2) \quad (7.46)$$

$$a_i \sim iid\ N(0, \sigma_a^{\ 2}) \quad (7.47)$$

$$b_i \sim iid\ N(0, \sigma_b^{\ 2}) \quad (7.48)$$

Trong đó $Y_{ij}$ là AGB (kg) ứng với cây thứ *j* từ cấp *i* của nhân tố ảnh hưởng; $\alpha$ và $\beta$ là tham số của mô hình; $a_i$ và $b_i$ là thay đổi của tham số theo cấp *i;* $X_{ij}$ là các biến số DBH (cm), H (m) hoặc tổ hợp biến $DBH^2H$, $DBH^2HWD$ ứng với cây thứ *j* trong cấp i*;* và $\varepsilon_{ij}$ là sai số ngẫu nhiên ứng với cây thứ *j* và cấp nhân tố *i.*

Do phương sai của sai số mô hình có xu hướng tăng khi DBH tăng; do vậy, một hàm phương sai theo trọng số đã được áp dụng để điều chỉnh các tham số của mô hình nhằm giảm biến động sai số khi DBH tăng lên. Hàm phương sai có dạng như sau (Huy *et al.,* 2016a):

$$Var(\varepsilon_{ij}) = \widehat{\sigma^2}(v_{ij})^{2k} \quad (7.49)$$

Trong đó $\varepsilon_{ij}$ là sai số ngẫu nhiên; $\widehat{\sigma^2}$ là sai số bình phương; $v_{ij}$ là biến trọng số (DBH, hoặc $DBH^2H$ hoặc $DBH^2HWD$) tương ứng với cây thứ j và cấp nhân tố ảnh hưởng i; và k là hệ số của hàm phương sai.

Để minh họa cho thiết lập mô hình với ảnh hưởng của các nhân tố ngẫu nhiên, sử dụng Dữ liệu 13 với 968 cây mẫu được xác định khối cây rừng trên mặt đất (AGB, kg) cùng với các biến số DBH (cm), H (m), WD ($g/cm^3$) và các nhân tố môi trường như vùng sinh thái (có 5 vùng được thu thập số liệu), sinh học như họ thực vật cấp khối lượng thể tích gỗ (WD, $g/cm^3$). Thử nghiệm thiết lập mô hình ước tính AGB theo tổ hợp biến $DBH^2H$ dạng power và xét ảnh hưởng của vùng sinh thái đến các tham số của mô hình

Áp dụng chương trình nlme có trọng số và không có hay có xét ảnh hưởng của nhân tố môi trường (vùng sinh thái) trong R như sau:

**Codes lập mô hình $AGB = a*(DBH^2H)^b$ theo nlme trong R có trọng số không xét ảnh hưởng của yếu tố môi trường**

```
# Erase memory
rm(list=ls())
# Clean plot window
dev.off()
# Define the working directory
setwd("E:/1 - Bao Huy 2017/Books All/Textbook for Infomatic Statistic/TextbookDataset Analysis/Dataset")
# Import country data
t_all <- read.table("Du lieu 13 AGB Viet Nam.txt", header=T,sep="\t",stringsAsFactors = FALSE)
# install.packages
library(ggplot2)
library(nlme)
library(cowplot)
library(gridExtra)

# Develop Model: (Lập mô hình theo nlme không xét ảnh hưởng của vùng sinh thái)
# Combination of varialbe of DBH2H:
```

```
t_all$DBH2H = (t_all$DBH/100)^2*t_all$H
# Modelling (Mô hình hóa với trọng số weight = 1/DBH2H, không có random efect)
  start <- coefficients(lm(log(AGB)~log(DBH2H), data=t_all))
  names(start) <- c("a","b")
  start[1]<-exp(start[1])
    Max_like <- nlme(AGB~a*DBH2H^b, data=cbind(t_all,g="a"), fixed=a+b~1,
                start=start, groups=~g, weights=varPower(form=~DBH2H))

 # Output Model:
summary(Max_like)
k <- summary(Max_like)$modelStruct$varStruct[1]
k
# Estimated values and Predicted:
t_all$Max_like.fit <- fitted.values(Max_like)
t_all$Max_like.res <- residuals(Max_like)
t_all$Max_like.res.weigh <- residuals(Max_like)/t_all$DBH2H^k

# Calcul of AIC, R2
AIC(Max_like)
R2 <- 1- sum((t_all$AGB - t_all$Max_like.fit)^2)/sum((t_all$AGB - mean(t_all$AGB))^2)
R2
R2.adjusted <- 1 - (1-R2)*(length(t_all$DBH)-1)/(length(t_all$DBH)-3-1)
R2.adjusted

# Plots:
# Fitted and Obseved
p1 <- ggplot(t_all)
p1 <- p1 + geom_point(aes(x=DBH2H, y=AGB))
p1 <- p1 + geom_line(cex = 1.5, aes(x=DBH2H, y=Max_like.fit))
p1 <- p1 + xlab("DBH2H (m3)") + ylab("AGB (kg)") + theme_bw()
p1 <- p1 + labs(title = "a")
p1 = p1 + theme(axis.title.y = element_text(size = rel(1.5)))
p1 = p1 + theme(axis.title.x = element_text(size = rel(1.5)))
p1 <- p1 + theme(plot.title = element_text(size = rel(1.7)))
p1 = p1 + theme(axis.text.x = element_text(size=15))
p1 = p1 + theme(axis.text.y = element_text(size=15))
p1

# Weighted Residuals vs predictied
p2 <- ggplot(t_all, aes(x=Max_like.fit, y=Max_like.res.weigh))
p2 <- p2 + geom_point()
p2 <- p2 + stat_smooth(cex = 1.5, method = "auto", se = FALSE, colour="black")
p2 <- p2 + xlab("Dự đoán AGB (kg)") + ylab("Sai số có trọng số (kg)") + theme_bw()
p2 <- p2 + labs(title = "b")
p2 = p2 + theme(axis.title.y = element_text(size = rel(1.5)))
p2 = p2 + theme(axis.title.x = element_text(size = rel(1.5)))
p2 <- p2 + theme(plot.title = element_text(size = rel(1.7)))
```

```
p2 = p2 + theme(axis.text.x = element_text(size=15))
p2 = p2 + theme(axis.text.y = element_text(size=15))
p2

plot_grid(p1, p2, ncol = 2)
```

Kết quả lập mô hình theo nlme có trọng số không xét ảnh hưởng vùng sinh thái:

```
> # Develop Model:
> # Combination of varialbe of DBH2H:
> t_all$DBH2H = (t_all$DBH/100)^2*t_all$H
>
> # Moelling
>   start <- coefficients(lm(log(AGB)~log(DBH2H), data=t_all))
>   names(start) <- c("a","b")
>   start[1]<-exp(start[1])
>
>   Max_like <- nlme(AGB~a*DBH2H^b, data=cbind(t_all,g="a"),
fixed=a+b~1,
+                    start=start, groups=~g,
weights=varPower(form=~DBH2H))
>
> # Output Model:
> summary(Max_like)
Nonlinear mixed-effects model fit by maximum likelihood
  Model: AGB ~ a * DBH2H^b
 Data: cbind(t_all, g = "a")
       AIC      BIC    logLik
  10445.05 10479.17 -5215.523

Random effects:
 Formula: list(a ~ 1, b ~ 1)
 Level: g
 Structure: General positive-definite, Log-Cholesky parametrization
         StdDev       Corr
a        3.506221e-03 a
b        1.749664e-06 -0.007
Residual 8.347181e+01

Variance function:
 Structure: Power of variance covariate
 Formula: ~DBH2H
 Parameter estimates:
    power
0.9351279
Fixed effects: a + b ~ 1
      Value Std.Error  DF  t-value p-value
a 263.98561  2.777799 966  95.0341       0
b   0.93646  0.005568 966 168.1851       0
 Correlation:
  a
b 0.253

Standardized Within-Group Residuals:
       Min         Q1        Med         Q3        Max
-2.1723427 -0.7485622 -0.1440077  0.6302289  5.1209316

Number of Observations: 968
Number of Groups: 1
> k <- summary(Max_like)$modelStruct$varStruct[1]
> k
[1] 0.9351279
```

```
>
> # Estimated values and Predicted:
> t_all$Max_like.fit <- fitted.values(Max_like)
> t_all$Max_like.res <- residuals(Max_like)
> t_all$Max_like.res.weigh <- residuals(Max_like)/t_all$DBH2H^k
>
> # Calcul of AIC, R2
> AIC(Max_like)
[1] 10445.05
> R2 <- 1- sum((t_all$AGB - t_all$Max_like.fit)^2)/sum((t_all$AGB -
mean(t_all$AGB))^2)
> R2
[1] 0.8973651
> R2.adjusted <- 1 - (1-R2)*(length(t_all$DBH)-1)/(length(t_all$DBH)-3-
1)
> R2.adjusted
[1] 0.8970457
```

**Codes lập mô hình $AGB = a*(DBH^2H)^b$ theo nlme trong R có trọng số xét ảnh hưởng của vùng sinh thái**

```
# Erase memory
rm(list=ls())
# Clean plot window
dev.off()
# Define the working directory
setwd("E:/1 - Bao Huy 2017/Books All/Textbook for Infomatic Statistic/TextbookDataset Analysis/Dataset")
# Import country data
t_all <- read.table("Du lieu 13 AGB Viet Nam.txt", header=T,sep="\t",stringsAsFactors = FALSE)
# install.packages
library(ggplot2)
library(nlme)
library(cowplot)
library(gridExtra)

# Develop Model: (Lập mô hình)
# Combination of varialbe of DBH2H: (Tổ hợp biến DBH2H)
t_all$DBH2H = (t_all$DBH/100)^2*t_all$H

# Modelling (Mô hình phi tuyến có trọng số và random efect là vừng sinh thái (region) theo nlme)
start <- coefficients(lm(log(AGB)~log(DBH2H), data=t_all))
  names(start) <- c("a","b")
  start[1]<-exp(start[1])

  Max_like2 <- nlme(AGB~a*DBH2H^b, data=t_all, fixed=a+b~1, random=a~1,
                    start=start, groups=~Region, weights=varPower(form=~DBH2H))
```

```
# Output of Model
summary(Max_like2)
k <- summary(Max_like2)$modelStruct$varStruct[1]
k

# Fittied and predicted values of the model
t_all$Max_like2.fit <- fitted.values(Max_like2)
t_all$Max_like2.res <- residuals(Max_like2)
t_all$Max_like2.res.weigh <- residuals(Max_like2)/t_all$DBH2H^k

# Parameters and random parameters
fixef(Max_like2)
ranef(Max_like2)
coef(Max_like2)

# Calcul of AIC, R2
AIC(Max_like2)

R2 <- 1- sum((t_all$AGB - t_all$Max_like2.fit)^2)/sum((t_all$AGB - mean(t_all$AGB))^2)
R2
R2.adjusted <- 1 - (1-R2)*(length(t_all$DBH2H)-1)/(length(t_all$DBH2H)-3-1)
R2.adjusted

# Plots for Regions
p1 <- ggplot(t_all)
p1 <- ggplot(t_all, aes(x=DBH2H, y=AGB, pch=Region))
p1 <- p1 + geom_point(cex=2.5)
p1 <- p1 + geom_line(cex = 1.5, aes(x=DBH2H, y=Max_like2.fit, linetype=Region))
p1 <- p1 + xlab("DBH2H (m3)") + ylab("AGB (kg)") + theme_bw()
p1 <- p1 + theme(legend.title=element_blank())
p1 = p1 + theme(axis.title.y = element_text(size = rel(1.5)))
p1 = p1 + theme(axis.title.x = element_text(size = rel(1.5)))
p1 <- p1 + theme(plot.title = element_text(size = rel(1.7)))
p1 = p1 + theme(axis.text.x = element_text(size=15))
p1 = p1 + theme(axis.text.y = element_text(size=15))
p1

# The end
```

Kết quả lập mô hình có trọng số và xét ảnh hưởng của vùng sinh thái:

```
> # Develop Model:
> # Combination of varialbe of DBH2H:
> t_all$DBH2H = (t_all$DBH/100)^2*t_all$H
>
> # Modelling
> start <- coefficients(lm(log(AGB)~log(DBH2H), data=t_all))
```

```
>   names(start) <- c("a","b")
>   start[1]<-exp(start[1])
>
>   Max_like2 <- nlme(AGB~a*DBH2H^b, data=t_all, fixed=a+b~1,
random=a~1,
+                     start=start, groups=~Region,
weights=varPower(form=~DBH2H))
>
> # Output of Model
> summary(Max_like2)
Nonlinear mixed-effects model fit by maximum likelihood
  Model: AGB ~ a * DBH2H^b
 Data: t_all
       AIC      BIC    logLik
  10386.39 10410.77 -5188.195

Random effects:
 Formula: a ~ 1 | Region
                a Residual
StdDev: 23.62485 80.34504

Variance function:
 Structure: Power of variance covariate
 Formula: ~DBH2H
 Parameter estimates:
    power
0.9294789
Fixed effects: a + b ~ 1
      Value Std.Error  DF   t-value p-value
a 264.54585 10.960436 962  24.13643       0
b   0.95104  0.005603 962 169.74914       0
 Correlation:
  a
b 0.052

Standardized Within-Group Residuals:
       Min         Q1        Med         Q3        Max
-2.6445682 -0.7396476 -0.1005597  0.5979435  4.8975852

Number of Observations: 968
Number of Groups: 5
> k <- summary(Max_like2)$modelStruct$varStruct[1]
> k
[1] 0.9294789
>
> # Fittied and predicted values of the model
> t_all$Max_like2.fit <- fitted.values(Max_like2)
> t_all$Max_like2.res <- residuals(Max_like2)
> t_all$Max_like2.res.weigh <- residuals(Max_like2)/t_all$DBH2H^k
>
> # Parameters and random parameters
> fixef(Max_like2)
          a           b
264.5458471   0.9510392
> ranef(Max_like2)
             a
CH   39.665076
NCC -11.337119
NE   -7.840356
SCC   7.534166
SE  -28.021766
> coef(Max_like2)
             a         b
CH  304.2109 0.9510392
NCC 253.2087 0.9510392
NE  256.7055 0.9510392
```

```
SCC 272.0800 0.9510392
SE  236.5241 0.9510392
>
> # calcul of AIC, R2
> AIC(Max_like2)
[1] 10386.39
>
> R2 <- 1- sum((t_all$AGB - t_all$Max_like2.fit)^2)/sum((t_all$AGB -
mean(t_all$AGB))^2)
> R2
[1] 0.9043419
> R2.adjusted <- 1 - (1-R2)*(length(t_all$DBH2H)-
1)/(length(t_all$DBH2H)-3-1)
> R2.adjusted
[1] 0.9040442
```

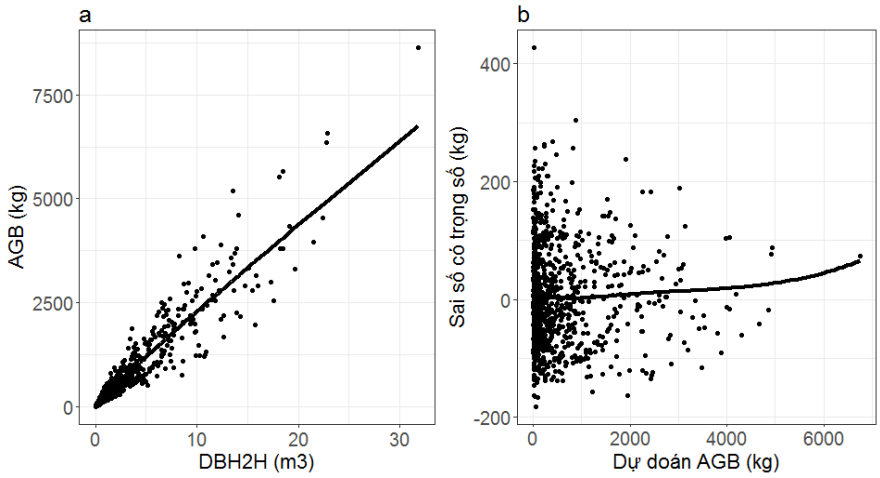
Kết quả thử nghiệm xét ảnh hưởng của vùng sinh thái khác nhau đến tham số a của mô hình AGB = a $DBH^2H^b$ cho thấy, a thay đổi rõ rệt ở các vùng sinh thái, hay nói khác vùng sinh thái đã ảnh hưởng đến ước tính AGB qua mô hình.

Kết quả thiết lập mô hình power ước tính AGB theo tổ hợp biến $DBH^2H$ theo phương pháp nlme Maximum Likelihood có trọng số và có hay không có ảnh hưởng của vùng sinh thái (được trình bày trong Bảng 7.7). Kết quả cho thấy, khi có xét vùng sinh thái, AIC giảm rõ rệt và $R^2$ tăng nhẹ, có nghĩa, khi đưa nhân tố vùng sinh thái vào mô hình sẽ nâng cao độ tin cậy, giảm sai số khi ước tính AGB cho từng vùng sinh thái.

**Bảng 7.7.** So sánh sự khác nhau của các mô hình *AGB = f(DBH, H)* có hay không xét đến ảnh hưởng của vùng sinh thái

| Dạng mô hình | Nhân tố ảnh hưởng<br>Random effect | Trong số<br>Weight variable | AIC | Adj. $R^2$ |
|---|---|---|---|---|
| $AGB = a \times (DBH^2H)^b$ | Không | $1/(DBH^2H)^k$ | 10445 | 0.897 |
| $AGB = a \times (DBH^2H)^b$ | Vùng sinh thái (Region) | $1/(DBH^2H)^k$ | 10386 | 0.904 |

Hình 7.27 chỉ ra mô hình và biến động của sai số của mô hình khi không xét vùng sinh thái, sai số khi rải đều nhưng lớn, biến động -200 đến 200 kg.

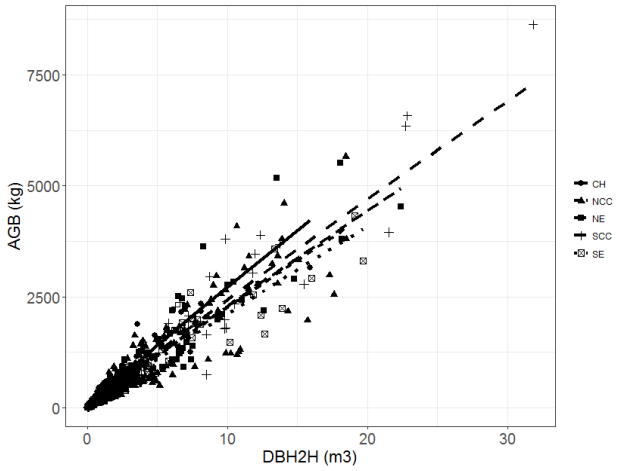



**Hình 7.27.** Mô hình AGB = a×$DBH^2H^b$ không có ảnh hưởng vùng sinh thái: a) Mô hình với giá trị quan sát; b) Sai số có trọng số theo dự đoán

Bảng 7.8 trình bày kết quả các tham số của mô hình thay đổi theo vùng sinh thái và Hình 7.28 là đồ thị của mô hình theo 5 vùng sinh thái.

**Bảng 7.8.** Các tham số và kích thước mẫu của mô hình $AGB=a\times(DBH^2H)^b$ có hay không xét ảnh hưởng của vùng sinh thái

| Nhân tố ảnh hưởng Random Effect | Vùng sinh thái | Tham số | | Số cây mẫu |
|---|---|---|---|---|
| | | a | b | |
| *Không* | Chung | 264.54585 | 0.95104 | 968 |
| *Vùng sinh thái* | Tây Nguyên | 304.2109 | 0.95104 | 222 |
| | Bắc Trung Bộ | 253.2087 | 0.95104 | 311 |
| | Đông Bắc | 256.7055 | 0.95104 | 215 |
| | Nam Trung Bộ | 272.0800 | 0.95104 | 110 |
| | Đông Nam Bộ | 236.5241 | 0.95104 | 110 |




**Hình 7.28.** $AGB=a\times(DBH2H)^b$ theo 5 vùng sinh thái: CH: Tây Nguyên, NCC: Bắc Trung Bộ, NE: Đông Bắc, SCC: Nam Trung Bộ và SE: Đông Nam

## 7.8 Phương pháp so sánh và thẩm định chéo các mô hình (Cross validation)

### 7.8.1 Thẩm định chéo để lựa chọn và đánh giá sai số, độ tin cậy của các mô hình

Trong thiết lập và sử dụng các mô hình, việc lựa chọn mô hình tối ưu, với độ tin cậy cao, hoặc cung cấp thông tin sai số của mô hình đã thiết lập một cách khách quan và chính xác là một nội dung vô cùng quan trọng trong khoa học mô hình hóa. Từ đây đã hình thành một lĩnh vực thống kê chuyên đề là "Thẩm định chéo – Cross Validation". Moore (2017) đã chỉ ra các phương pháp thẩm định chéo (Cross-Validation) các mô hình nhằm phát hiện và phòng tránh việc thiết lập và lựa chọn các mô hình có sai lệch lớn so với thực tế (overfitting). Các chỉ tiêu dùng để so sánh nhằm lựa chọn các mô hình chủ yếu là AIC, $R^2_{adj}$ và các sai số thường được áp dụng khi thẩm định chéo là Bias%, RMSE%, MAPE% (Mayer et al, 1993; Zhang, 1997; Chave et al, 2005; Basuki et al, 2009;

Temesgen et al. 2014; Huy et al, 2016a,b,c). Ngoài ra, Picard và Cook (1984) cũng chỉ ra rằng, thẩm định chéo ngoài việc xác định sai số, tránh cho mô hình dự đoán sai lệch với thực tế thì nó còn giúp cho việc lựa chọn các biến số thích hợp cho mô hình; ví dụ, khi ước tính AGB ở một vùng sinh thái cụ thể thì cần biến số nào trong các biến số như DBH, H, WD và CA để mô hình có độ tin cậy tốt nhất?.

Các phần mềm thống kê chuyên nghiệp như SPSS, Statgraphics chỉ cung cấp các sai số của mô hình so với dữ liệu lập mô hình mà lại không cung cấp công cụ để tính toán sai số theo phương pháp thẩm định chéo, trong khi đó, phần mềm mã nguồn mở R là cơ hội tốt cho việc áp dụng thẩm định chéo các mô hình một cách linh hoạt.

### 7.8.2 Phương pháp truyền thống – Sử dụng dữ liệu độc lập để so sánh và thẩm định sai số mô hình

Trong khoa học mô hình hóa truyền thống, việc đánh giá sai số của các mô hình thường tiến hành bằng cách sử dụng một bộ dữ liệu độc lập để đánh giá sai số của mô hình đã thiết lập, hoặc có thể phân chia ngẫu nhiên bộ dữ liệu thành hai phần, phần một dùng để lập các mô hình và một phần khác dùng đánh giá sai số của mô hình.

Các sai số của các mô hình được tính toán bao gồm % sai lệch giữa quan sát và dự báo qua mô hình (Bias %), sai số trung phương trung bình % (Root Mean Square Error - RMSE %), và sai số tuyệt đối trung bình % (Mean Absolute Percent Error - MAPE) (Mayer *et al.,* 1993; Chave *et al*., 2005; Basuki *et al.,* 2009; Swanson *et al.,* 2011; Huy *et al.,* 2016a,b):

$$Bias\ \% = \frac{100}{n}\sum_{i=1}^{n}\frac{(y_i - \hat{y}_i)}{(y_i)} \tag{7.50}$$

$$RMSE\ \% = 100\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{y_i}\right)^2} \tag{7.51}$$

$$MAPE\ \% = \frac{100}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \tag{7.52}$$

Trong đó n là số cây mẫu độc lập dùng đánh giá; và $y_i$ và $\hat{y}_i$ là giá trị quan sát và ước tính qua mô hình.

Thử nghiệm áp dụng phương pháp thẩm định sai số mô hình: $AGB = a \times DBH^2HWD^b$. Từ bộ dữ liệu 110 cây mẫu sinh khối ở vùng Nam Trung Bộ (Dữ liệu 11). Ước lượng mô hình bằng phương pháp phi tuyến Maximum Likelihood nlme có trọng số Weight = $1/DBH2HWD^k$ trong R. Phân chia dữ liệu ngẫu nhiên làm hai phần: 80% cho lập mô hình và 20% cho đánh giá sai số của các mô hình trong R, thực hiện việc phân chia ngẫu nhiên đến 200 lần bằng codes trong R như sau:

**Codes phân chia bộ dữ liệu ngẫu nhiên 200 lần thành 2 phần: 80% cho lập mô hình (t_eq) và 20% cho đánh giá mô hình (t_va):**

```
# Erase memory
rm(list=ls())
# Import data from a .txt file
setwd("E:/1 - Bao Huy 2017/Books All/Textbook for Infomatic Statistic/TextbookDataset Analysis/Dataset")
t <- read.table("Du lieu 11 AGB QNam .txt", header=T,sep="\t",stringsAsFactors = FALSE)

# Combination of Variables
t$DBH2H <- (t$DBH/100)^2*t$H
t$DBH2HWD <- t$DBH2H*t$WD*1000



# Selection of random sample tress (80% for equation development and 20% validation) (Phân chia ngẫu nhiên thành hai bộ dữ liwwuj 80%/20% số cây ngẫu nhiên 200 lần:
for(i in 1:200){
  t_va <- t[sample(nrow(t), length(t$AGB)/5), ]
  t_eq <- t[!t$ID %in% t_va$ID, ]
}

str(t_va)
View(t_va)
length(t_va$AGB)

str(t_eq)
View(t_eq)
length(t_eq$AGB)

# Sub_Table for Equation development (t_eq) and validation (t_va) (Ghi lại hai file dữ liệu dùng lập và đánh giá mô hình:
write.table(t_eq, file="t_eq.txt", sep="\t",dec=".", row.names= FALSE)
write.table(t_va, file="t_va.txt", sep="\t",dec=".", row.names= FALSE)
```

Trên cơ sở phân chia thành hai bộ dữ liệu, sử dụng bộ dữ liệu 80% (t_eq) để lập mô hình theo chương trình nlme có trọng số trong R như sau.

**Codes lập mô hình $AGB = a* DBH^2HWD^b$ theo nlme có trọng số với 80% dữ liệu và các chỉ tiêu thống kê:**

```
# Erase memory
rm(list=ls())
# Clean plot window
dev.off()
# Define the working directory
setwd("E:/1 - Bao Huy 2017/Books All/Textbook for Infomatic Statistic/TextbookDataset Analysis/Dataset")
```

```
# Import data
t <- read.table("t_eq.txt", header=T,sep="\t",stringsAsFactors = FALSE)
# Install.packages ggplot2 and nlme
library(ggplot2)
library(nlme)
library(cowplot)

# Model nlme with weight
start <- coefficients(lm(log(AGB)~log(DBH2HWD), data=t))
names(start) <- c("a","b")
start[1]<-exp(start[1])
Max_like <- nlme(AGB~a*(DBH2HWD)^b, data=cbind(t,g="a"), fixed=a+b~1,
                 start=start, groups=~g, weights=varPower(form=~DBH2HWD))

# Outputs of the model
summary(Max_like)
k <- summary(Max_like)$modelStruct$varStruct[1]
k
t$Max_like.fit <- fitted.values(Max_like)
t$Max_like.res <- residuals(Max_like)
t$Max_like.res.weigh <- residuals(Max_like)/t$DBH2HWD^k

# Calcul of AIC, R^2
AIC(Max_like)
R2 <- 1- sum((t$AGB - t$Max_like.fit)^2)/sum((t$AGB - mean(t$AGB))^2)
R2
R2.adjusted <- 1 - (1-R2)*(length(t$DBH)-1)/(length(t$DBH)-3-1)
R2.adjusted
# The end
```

**Kết quả lập mô hình với 80% dữ liệu:**

```
> # Model nlme
> start <- coefficients(lm(log(AGB)~log(DBH2HWD), data=t))
> names(start) <- c("a","b")
> start[1]<-exp(start[1])
>
> Max_like <- nlme(AGB~a*(DBH2HWD)^b, data=cbind(t,g="a"), fixed=a+b~1,
+                  start=start, groups=~g,
weights=varPower(form=~DBH2HWD))
>
> # Outputs of the model
> summary(Max_like)
Nonlinear mixed-effects model fit by maximum likelihood
  Model: AGB ~ a * (DBH2HWD)^b
 Data: cbind(t, g = "a")
       AIC      BIC    logLik
  886.9901 904.3314 -436.495

Random effects:
 Formula: list(a ~ 1, b ~ 1)
 Level: g
 Structure: General positive-definite, Log-Cholesky parametrization
```

```
         StdDev       Corr
a        9.864677e-06 a
b        1.703723e-06 -0.992
Residual 1.557526e-01

Variance function:
 Structure: Power of variance covariate
 Formula: ~DBH2HWD
 Parameter estimates:
    power
0.9331245
Fixed effects: a + b ~ 1

      Value  Std.Error DF  t-value p-value
a 0.6323179 0.04391993 86 14.39706       0
b 0.9516968 0.01094225 86 86.97450       0
 Correlation:
  a
b -0.939

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-3.08744021 -0.56655763 -0.04305784  0.42117832  3.02727329

Number of Observations: 88
Number of Groups: 1
>
> k <- summary(Max_like)$modelStruct$varStruct[1]
> k
[1] 0.9331245
>
> t$Max_like.fit <- fitted.values(Max_like)
> t$Max_like.res <- residuals(Max_like)
> t$Max_like.res.weigh <- residuals(Max_like)/t$DBH2HWD^k
>
> # calcul of AIC, R^2
> AIC(Max_like)
[1] 886.9901
>
> R2 <- 1- sum((t$AGB - t$Max_like.fit)^2)/sum((t$AGB - mean(t$AGB))^2)
> R2
[1] 0.958051
> R2.adjusted <- 1 - (1-R2)*(length(t$DBH)-1)/(length(t$DBH)-3-1)
> R2.adjusted
[1] 0.9565529
```

Kết quả có mô hình: $AGB = 0.63231 \times (DBH^2HWD)^{0.95169}$

Với n = 88, $R^2_{adj.} = 0.957$, các tham số tồn tại với p-value = 0; AIC = 887.0

Đem mô hình vừa thiết lập áp dụng với 20% dữ liệu độc lập để tính các sai số của mô hình. Thực hiện trong R như sau:

**Code tính các sai số từ 20% dữ liệu ngẫu nhiên độc lập**

```
# Erase memory
rm(list=ls())
# Clean plot window
dev.off()
# Define the working directory (change \ with / using Edit>Find)
setwd("E:/1 - Bao Huy 2017/Books All/Textbook for Infomatic Statistic/TextbookDataset Analysis/Dataset")
# Import country data from t_EBLF
t <- read.table("t_va.txt", header=T,sep="\t",stringsAsFactors = FALSE)
```

```
# Prediction of Model (Dự đoán AGB qua mô hình)
AGBpre = 0.63231*t$DBH2HWD^0.95169

# Bias and RSME and MAPE%
Bias <- 100*mean((t$AGB - AGBpre)/t$AGB)
RMSE <- 100*sqrt(mean(((t$AGB - AGBpre)/t$AGB)^2))
MAPE <- 100*mean(abs(t$AGB - AGBpre)/t$AGB)
Bias
RMSE
MAPE

# Plot Model vs validation data
p3 <- ggplot(t)
p3 <- p3 + geom_point(pch=19, cex=4,aes(x=DBH2HWD, y=AGB))
p3 <- p3 + geom_line(cex=1.5,aes(x=DBH2HWD, y= AGBpre))
p3 <- p3 + xlab("DBH^2HWD (kg)") + ylab("AGB (kg)") + theme_bw()
p3 <- p3 + theme(legend.title=element_blank())
p3 = p3 + theme(axis.title.y = element_text(size = rel(1.5)))
p3 = p3 + theme(axis.title.x = element_text(size = rel(1.5)))
p3 <- p3 + theme(plot.title = element_text(size = rel(1.7)))
p3 = p3 + theme(axis.text.x = element_text(size=15))
p3 = p3 + theme(axis.text.y = element_text(size=15))
p3

# The end
```

Kết quả tính sai số mô hình: $AGB = 0.63231 \times (DBH^2HWD)^{0.95169}$ từ 20% dữ liệu ngẫu nhiên độc lập:

```
> # Prediction of Model
> AGBpre = 0.63231*t$DBH2HWD^0.95169
>
> # Bias and RSME and MAPE%
> Bias <- 100*mean((t$AGB - AGBpre)/t$AGB)
> RMSE <- 100*sqrt(mean(((t$AGB - AGBpre)/t$AGB)^2))
> MAPE <- 100*mean(abs(t$AGB - AGBpre)/t$AGB)
>
> Bias
[1] -5.990722
> RMSE
[1] 25.95349
> MAPE
[1] 21.90881
```

Mô hình: $AGB = 0.63231 \times (DBH^2HWD)^{0.95169}$ có sai số từ 20% dữ liệu độc lập:

Bias = - 5.99 kg, RMSE% = 25.95% và MAPE = 21.91%

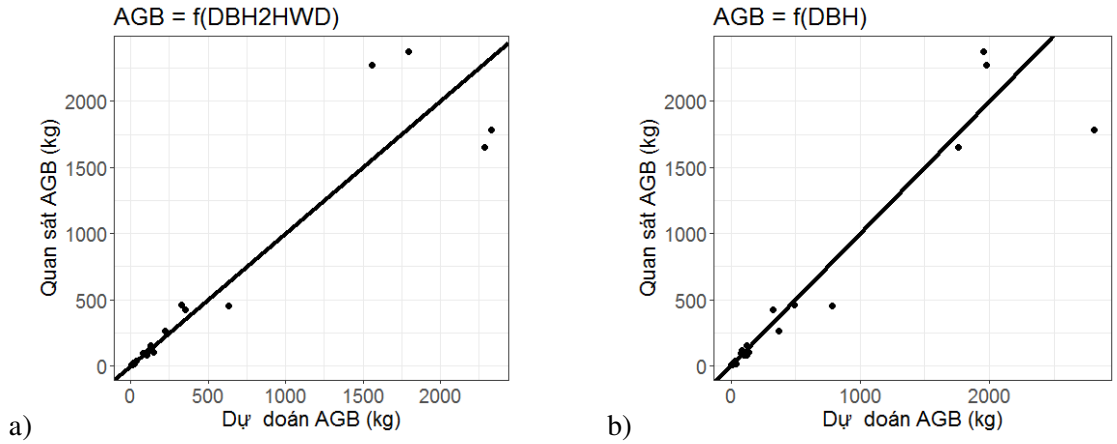
Tiếp tục áp dụng phương pháp lập và thẩm định sai số tương tự như trên với mô hình ước tính AGB chỉ với một biến số đơn giản DBH, từ đó so với mô hình tổ hợp 3 biến số ($DBH^2HWD$) trong Bảng 7.9.

Các mô hình dự đoán so với 20% dữ liệu rút ngẫu nhiên dùng đánh giá sai số thể hiện ở Hình 7.29. Từ hình này cho thấy, mô hình với ba biến số đầu vào ($DBH^2HWD$) đã cải thiện được sai số so với mô hình một biến DBH. Điều này cũng phù hợp với kết quả so sánh ở Bảng 7.9. Mô hình tổ hợp ba biến $DBH^2HWD$ có AIC và các sai số bé hơn mô hình chỉ một biến số DBH. Hay nói khác, khi tăng biến số đầu vào H và WD đã cải thiện đáng kể độ tin cậy của mô hình. RMSE của mô hình 3 biến số đã giảm gần 20% so với mô hình một biến số.

**Bảng 7.9.** So sánh và thẩm định sai số của hai mô hình AGB = f(DBH) và AGB = f($DBH^2HWWD$) theo phương pháp sử dụng dữ liệu độc lập

| Chỉ tiêu thống kê, sai số | Mô hình | |
|---|---|---|
| | $AGB = 0.10658 \times DBH^{2.48596}$ | $AGB = 0.63231 \times (DBH^2HWD)^{0.95169}$ |
| $R^2_{adj.}$ | 0.937 | 0.957 |
| AIC | 895.0 | 887.0 |
| Bias (kg) | -14.7 | -5.99 |
| RMSE % | 44.7 | 25.95 |
| MAPE % | 30.8 | 21.91 |

*Ghi chú: Mô hình và $R^2$, AIC được tính từ 80% dữ liệu độc lập; các sai số Bias, RMSE, MAPE được tính từ 20% dữ liệu rút ngẫu nhiên và độc lập với dữ liệu lập mô hình.*




**Hình 7.29.** Đồ thị quan hệ giữa giá trị AGB dự đoán qua mô hình với AGB quan sát của 20% dữ liệu ngẫu nhiên thẩm định độc lập. a) $AGB = 0.63231 \times (DBH^2HWD)^{0.95169}$; b) $AGB = 0.10658 \times DBH^{2.48596}$

Phương pháp thẩm định sai số truyền thống sử dụng dữ liệu độc lập để so sánh và thẩm định sai số mô hình có hạn chế rất lớn là sai số được xác định một lần cho một bộ dữ liệu độc lập nhất định, vì vậy, sai số có thể khác đi nếu áp dụng theo một bộ dữ liệu độc lập khác. Vì vậy, nó thường không cung cấp chính xác sai số trong mọi trường hợp ứng dụng. Do đó, các phương pháp thẩm định chéo các mô hình cần được xem xét áp dụng để cung cấp sai số ổn định.

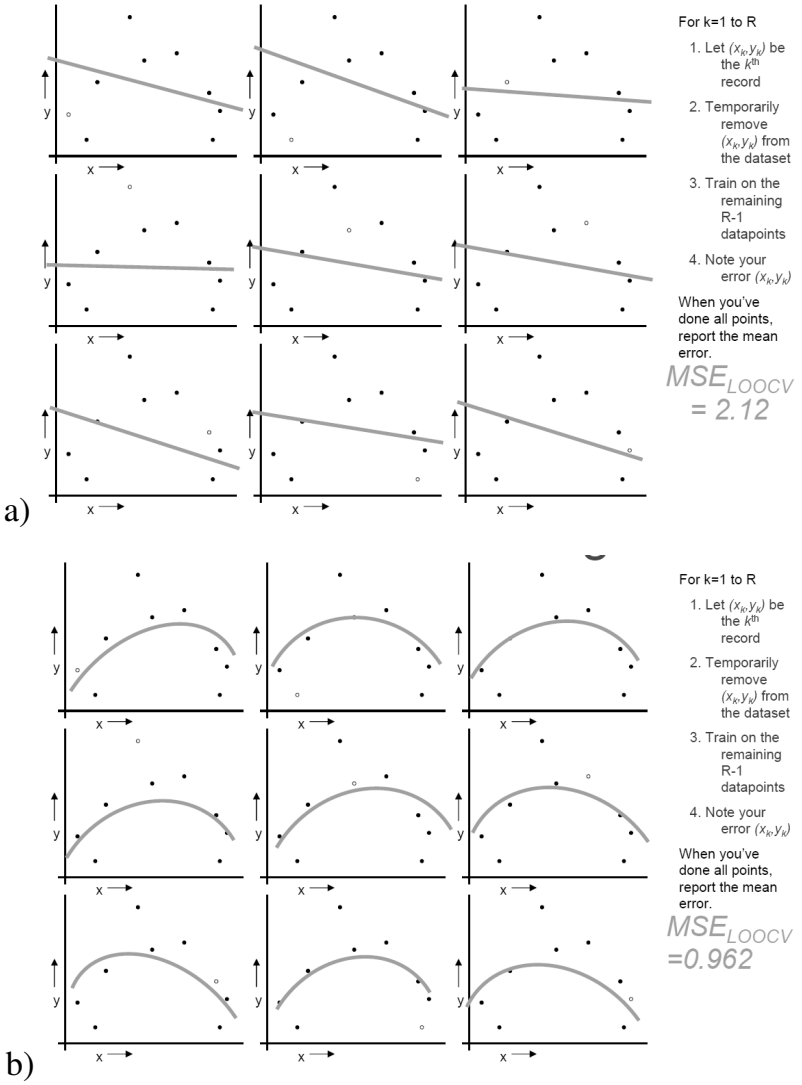
### 7.8.3 Phương pháp thẩm định chéo sai số - Leave-One-Out Cross Validation (LOOCV)

Giả sử có n dữ liệu, thì với phương pháp LOOCV sẽ sử dụng n-1 dữ liệu lập mô hình và 1 dữ liệu độc lập dùng để đánh giá sai số, lặp lại như vậy với n lần lập mô hình và đánh giá sai số, với

sai số mỗi lần được tính từ một dữ liệu độc lập không tham gia lập mô hình, sau đó lấy trung bình (Moore, 2017). Trên cơ sở sai số sẽ lựa chọn được mô hình có sai số bé nhất và cung cấp sai số chính xác.

Phương pháp này thì mọi dữ liệu đều tham gia lập mô hình hoặc dùng để tính sai số. Tuy vậy, sai số của mỗi lần được tính từ một dữ liệu cá thể lẻ, trong khi đó thực tế sai số khi áp dụng mô hình thường cho một quần thể với nhiều dữ liệu.

Ví dụ theo Moore (2017) ở Hình 7.30, có 9 dữ liệu, mỗi lần loại một dữ liệu khi lập mô hình và sử dụng dữ liệu loại ra đó để tính sai số theo mô hình vừa lập, lặp lại như vậy 9 lần và tính sai số trung bình. Ví dụ này cùng với 9 dữ liệu, thử lập hai loại mô hình là tuyến tính và parabol bậc 2 và áp dụng LOOCV để so sánh sai số MSE (Sai số trung phương). Kết quả cho thấy mô hình tuyến tính có MSE = 2.120 trong khi đó mô hình Quadratic có MSE= 0.962. Như vậy mô hình Quadratic là phù hợp hơn và sai số của nó là 0.962 cho mọi trường hợp dữ liệu.




**Hình 7.30.** Thẩm định chéo hai mô hình tuyến tính và Quadratic theo phương pháp LOOCV với 9 dữ liệu, sử dụng sai số trung phương MSE trung binh để so sánh. a) Thẩm định mô hình dạng tuyến tính; b) Thẩm định mô hình dạng hàm bậc 2 Quadratic cho. *(Nguồn: Moore, 2017)*

Cách tính các sai số tương đối khi áp dụng LOOCV:

$$Bias\ (\%) = \frac{100}{n}\sum_{i=1}^{L}\frac{y_i - \hat{y}_i}{y_i} \quad (7.53)$$

$$RMSE\ (\%) = 100\sqrt{\frac{1}{n}\sum_{i=1}^{L}\left(\frac{y_i - \hat{y}_i}{y_i}\right)^2} \quad (7.54)$$

$$MAPE\ (\%) = \frac{100}{n}\sum_{i=1}^{L}\frac{|y_i - \hat{y}_i|}{y_i} \quad (7.55)$$

Trong đó, L là số lần lặp lại tính sai số, mỗi lần sử dụng một dữ liệu độc lập để tính sai số mô hình (L=n dữ liệu); $y_i$ và $\hat{y}_i$ là giá trị quan sát và dự đoán qua mô hình.

Trên cơ sở Dữ liệu 11 của 110 cây mẫu sinh khối vùng Nam Trung Bộ, minh họa áp dụng phương pháp LOOCV để so sánh và đánh giá sai số của hai mô hình ước tính AGB: $AGB = a \times DBH^b$ và $AGB = a \times DBH^2 HWD^b$.

Sử dụng phần mềm mã nguồn mở R để lập mô hình theo nlme và tính toán các chỉ tiêu thống kê so sánh và sai số các mô hình theo phương pháp thẩm định chéo LOOCV như sau:

**Codes trong R để thiết lập mô hình theo nlme có trọng số và thẩm định chéo LOOCV mô hình: $AGB = a \times DBH^2 HWD^b$**

```
# Erase memory
rm(list=ls())

# Clean plot window
dev.off()

# Define the working directory
setwd("E:/1 - Bao Huy 2017/Books All/Textbook for Infomatic Statistic/TextbookDataset Analysis/Dataset")

# Import data
t <- read.table("Du lieu 11 AGB QNam .txt", header=T,sep="\t",stringsAsFactors = FALSE)
# Combination fo variables:
t$DBH2HWD = (t$DBH/100)^2*t$H*t$WD*1000

length(t$DBH)

# Using ggplot2 and nlme - install.packages("ggplot2") nlme
library(ggplot2)
```

```
library(nlme)
library(cowplot)

# Randomly shuffle the data
t <- t[sample(nrow(t)),]

# Create equally size folds = 1
folds <- cut(seq(1,nrow(t)),breaks=length(t$DBH),labels=FALSE)

AIC = rep(0, length(t$DBH))
R2adj = rep(0, length(t$DBH))
Bias = rep(0, length(t$DBH))
RMSE = rep(0, length(t$DBH))
MAPE = rep(0, length(t$DBH))

# Perform LOOCV cross validation: AGB = a*DBH2HWD^b
for(i in 1:length(t$DBH)){
  # Segement the data by fold using the which() function
  testIndexes <- which(folds==i,arr.ind=TRUE)
  n_va <- t[testIndexes, ]
  t_eq <- t[-testIndexes, ]

  # Modelling AGB = a*DBH2HWD^b
  start <- coefficients(lm(log(AGB)~log(DBH2HWD), data=t_eq))
  names(start) <- c("a","b")
  start[1]<-exp(start[1])

  Max_like <- nlme(AGB~a*DBH2HWD^b, data=cbind(t_eq,g="a"), fixed=a+b~1,
                   start=start, groups=~g, weights=varPower(form=~DBH2HWD))

  # Outputs of the model
  k <- summary(Max_like)$modelStruct$varStruct[1]
  k
  t_eq$Max_like.fit <- fitted.values(Max_like)
  t_eq$Max_like.res <- residuals(Max_like)
  t_eq$Max_like.res.weigh <- residuals(Max_like)/t_eq$DBH2HWD^k

  # Calcul of AIC, R2,
  AIC[i] = AIC(Max_like)
  R2 <- 1- sum((t_eq$AGB - t_eq$Max_like.fit)^2)/sum((t_eq$AGB - mean(t_eq$AGB))^2)
  R2.adjusted <- 1 - (1-R2)*(length(t_eq$DBH)-1)/(length(t_eq$DBH)-3-1)
  R2adj[i] = R2.adjusted
  # Prediction and Errors
  n_va$Pred <- predict(Max_like, newdata=cbind(n_va,g="a"))
```

```
  Bias[i] <- 100*mean((n_va$AGB - n_va$Pred)/n_va$AGB)
  RMSE[i] <- 100*sqrt(mean(((n_va$AGB - n_va$Pred)/n_va$AGB)^2))
  MAPE[i] <- 100*mean(abs(n_va$AGB - n_va$Pred)/n_va$AGB)
}

i
# Model fitting statistics:
mean(AIC)
mean(R2adj)

mean(Bias)
mean(RMSE)
mean(MAPE)
hist(Bias)
hist(RMSE)
hist(MAPE)
hist(Bias, main = paste("", ""), xlab = "Bias của mô hình AGB = a*DBH2HWWD^b", ylab = "Tần số",cex.lab=2, cex.axis=2, cex.main=2, cex.sub=2)

# The end
```

Kết quả tính toán các chỉ tiêu thống kê và các sai số của mô hình: $AGB = a \times DBH^b$ theo phương pháp LOOCV:

```
> # Model fitting statistics:
> mean(AIC)
[1] 1109.031
> mean(R2adj)
[1] 0.9335493
>
> mean(Bias)
[1] -7.856362
> mean(RMSE)
[1] 22.88872
> mean(MAPE)
[1] 22.88872
> hist(Bias)
> hist(RMSE)
> hist(MAPE)
> hist(Bias, main = paste("", ""), xlab = "Bias của mô hình AGB=
a*DBH^b", ylab = "Tần số")
```

Kết quả tính toán các chỉ tiêu thống kê và các sai số của mô hình: $AGB = a \times DBH2HWD^b$ theo phương pháp LOOCV:

```
> # Model fitting statistics:
> mean(AIC)
[1] 1088.982
> mean(R2adj)
[1] 0.9531261
>
> mean(Bias)
```

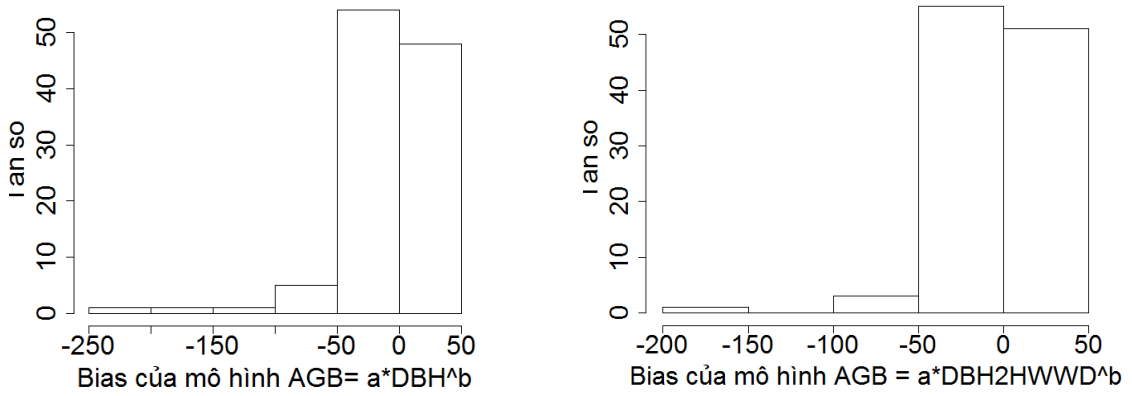
```
[1] -5.882556
> mean(RMSE)
[1] 19.68456
> mean(MAPE)
[1] 19.68456
> hist(Bias)
> hist(RMSE)
> hist(MAPE)
> hist(Bias, main = paste("", ""), xlab = "Bias của mô hình AGB =
a*DBH2HWWD^b", ylab = "Tần số")
>
> # The end
```

Các kết quả minh họa độc lập và thẩm định các mô hình AGB theo phương pháp LOOCV được tổng hợp trong Bảng 7.10. Kết quả này cho thấy mô hình tổ hợp ba biến $DBH^2HWD$ có độ tin cậy cao hơn (AIC bé hơn và $R^2$ cao hơn) và các sai số đều nhỏ hơn so với mô hình AGB chỉ với một biến số DBH.

**Bảng 7.10.** So sánh và thẩm định chéo LOOCV hai mô hình AGB = f(DBH) và AGB = f($DBH^2HWWD$)

| Chỉ tiêu thống kê, sai số | Mô hình | |
|---|---|---|
| | $AGB = a \times DBH^b$ | $AGB = a \times (DBH^2HWD)^b$ |
| $R^2_{adj.}$ | 0.934 | 0.953 |
| AIC | 1109 | 1089 |
| Bias (kg) | -7.86 | -5.88 |
| RMSE % | 22.89 | 19.68 |
| MAPE % | 22.89 | 19.68 |

*Ghi chú: Mô hình và $R^2$, AIC được tính từ n-1 dữ liệu độc lập; các sai số Bias, RMSE, MAPE được tính trung bình n lần từ một dữ liệu rút độc lập.*




**Hình 7.31.** Phân bố tần số Bias của hai mô hình AGB theo phương pháp LOOCV

Hình 7.31 cho thấy phân bố Bias của hai mô hình được thẩm định theo phương pháp LOOCV có xu hướng lệch phải và chưa tiệm cận chuẩn. Đây là nhược điểm của phương pháp thẩm định chéo LOOCV, do chỉ tính sai số của từng cá thể, trong khi đó, thực tế để sai số tiệm cận chuẩn thì mỗi lần rút mẫu đánh giá cần có số mẫu đủ lớn. Điều này cũng là hạn chế của phương pháp

LOOCV trong ứng dụng, vì trong thực tế sai số không tính cho từng cá thể mà cho một ô mẫu, hoặc lâm phần.

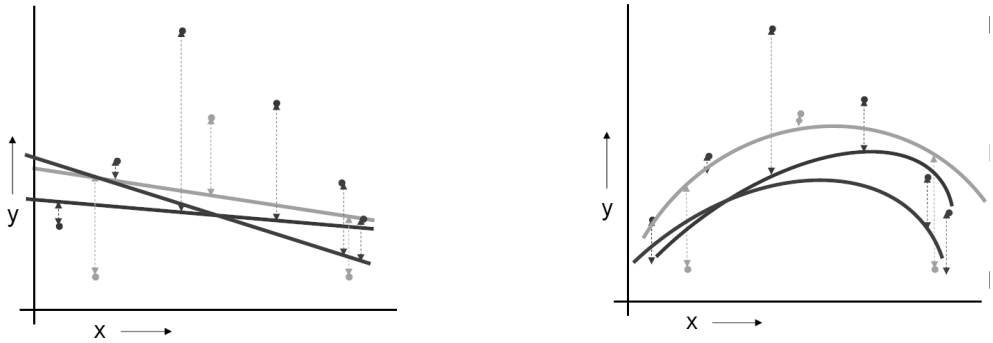
### 7.8.4 Phương pháp thẩm định chéo sai số k-fold Cross Validation

Phương pháp này phân chia n dữ liệu thành k phần bằng nhau (k-fold), một phần dữ liệu không tham gia lập mô hình dùng để đánh giá sai số, trong khi đó k-1 phần dữ liệu dùng lập mô hình. Tiến hành lập lại như vậy k lần, mỗi lần lấy một phần dữ liệu khác nhau để thẩm định mô hình và tính sai số trung bình từ k lần lặp (Moore, 2017).

Khi k = n dữ liệu quan sát thì phương pháp k-fold sẽ trở thành phương pháp LOOCV, vì vậy, thông thường để dữ liệu đánh giá > 1, trong k-fold có k > n, thường k = 5, 10.

Phương pháp này thì mọi dữ liệu đều tham gia lập mô hình hoặc dùng để tính sai số. Sai số của mỗi lần đánh giá được tính từ một bộ dữ liệu độc lập, cách tiếp cận như vậy gần gũi khi áp dụng cho một quần thể với nhiều dữ liệu. Tuy nhiên, số lần lặp lại để đánh giá thường không đủ lớn (k = 10) nên sai số có thể chưa ổn định và chưa tiệm cận chuẩn.

Hình 7.32 là ví dụ áp dụng k-fold để đánh gia sai số MSE của hai mô hình tuyến tính và Quadratic. Có 9 dữ liệu, được chia thành 3 phần (k=3), một phần có 3 dữ liệu. Một lần sử dụng 2 phần dữ liệu (6 dữ liệu) để lập mô hình, một phần dữ liệu độc lập (3 dữ liệu) để tính sai số. Lặp lại như vậy k lần (3 lần trong ví dụ này), sau đó tính sai số trung bình cho mỗi loại mô hình. Kết quả cho thấy dạng tuyến tính có MSE = 2.05 và hàm quadratic có MSE = 1.11. Như vậy dạng hàm Quadratic là tốt hơn với sai số bé hơn và sai số này là đúng cho mọi trường hợp của dữ liệu quan sát.




**Hình 7.32.** Thẩm định chéo hai mô hình tuyến tính và Quadratic theo phương pháp k-fold với 9 dữ liệu, tạo thành 3 fold (k=3), sử dụng sai số trung phương MSE trung bình để so sánh. a) Thẩm định mô hình dạng tuyến tính; b) Thẩm định mô hình dạng hàm bậc 2 Quadratic. (*Nguồn: Moore, 2017)*

Cách tính các sai số tương đối theo phương pháp k-fold như sau:

$$Bias\ (\%) = \frac{1}{k}\sum_{k=1}^{k} \frac{100}{n}\sum_{i=1}^{n} \frac{y_i - \hat{y}_i}{y_i} \tag{7.56}$$

$$RMSE\ (\%) = \frac{1}{k}\sum_{k=1}^{k} 100 \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{y_i}\right)^2} \quad (7.57)$$

$$MAPE\ (\%) = \frac{1}{k}\sum_{k=1}^{k} \frac{100}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \quad (7.58)$$

Trong đó, k là số phần dữ liệu bằng nhau được phân chia (k-fold), thường k = 5-10; n là số dữ liệu đánh giá của mỗi lần và $y_i$ và $\hat{y}_i$ là giá trị quan sát và dự đoán qua mô hình.

Trên cơ sở Dữ liệu 11 của 110 cây mẫu sinh khối vùng Nam Trung Bộ, minh họa áp dụng phương pháp k-fold để so sánh và đánh giá sai số của hai mô hình ước tính AGB: $AGB = a \times DBH^b$ và $AGB = a \times DBH^2 HWD^b$.

Sử dụng phần mềm mã nguồn mở R để lập mô hình theo nlme có trọng số và tính toán các chỉ tiêu thống kê so sánh và sai số các mô hình theo phương pháp thẩm định chéo k-fold với k = 10 như sau:

**Codes trong R để thiết lập mô hình theo nlme có trọng số và thẩm định chéo k-fold (k=10) mô hình: $AGB = a \times DBH^b$**

```
# Erase memory
rm(list=ls())
# Clean plot window
dev.off()
# Define the working directory
setwd("E:/1 - Bao Huy 2017/Books All/Textbook for Infomatic Statistic/TextbookDataset Analysis/Dataset")
# Import data
t <- read.table("Du lieu 11 AGB QNam .txt", header=T,sep="\t",stringsAsFactors = FALSE)
# Combination fo variables:
t$DBH2HWD = (t$DBH/100)^2*t$H*t$WD*1000
length(t$DBH)
# Using ggplot2 and nlme - install.packages("ggplot2") nlme
library(ggplot2)
library(nlme)
library(cowplot)
# Randomly shuffle the data
t <- t[sample(nrow(t)),]
# Create 10 equally size folds
folds <- cut(seq(1,nrow(t)),breaks=10,labels=FALSE)
AIC = rep(0, 10)
R2adj = rep(0, 10)
Bias = rep(0, 10)
RMSE = rep(0, 10)
```

```
MAPE = rep(0, 10)
# Perform 10 fold cross validation: Model AGB = a*DBH^b
for(i in 1:10){
  # Segement the data by fold using the which() function
  testIndexes <- which(folds==i,arr.ind=TRUE)
  n_va <- t[testIndexes, ]
  t_eq <- t[-testIndexes, ]
  # Modelling AGB = a*DBH^b
  start <- coefficients(lm(log(AGB)~log(DBH), data=t_eq))
  names(start) <- c("a","b")
  start[1]<-exp(start[1])
  Max_like <- nlme(AGB~a*DBH^b, data=cbind(t_eq,g="a"), fixed=a+b~1,
                   start=start, groups=~g, weights=varPower(form=~DBH))
  # Outputs of the model
  k <- summary(Max_like)$modelStruct$varStruct[1]
  k
  t_eq$Max_like.fit <- fitted.values(Max_like)
  t_eq$Max_like.res <- residuals(Max_like)
  t_eq$Max_like.res.weigh <- residuals(Max_like)/t_eq$DBH^k

  # Calcul of AIC, R2,
  AIC[i] = AIC(Max_like)
  R2 <- 1- sum((t_eq$AGB - t_eq$Max_like.fit)^2)/sum((t_eq$AGB - mean(t_eq$AGB))^2)
  R2.adjusted <- 1 - (1-R2)*(length(t_eq$DBH)-1)/(length(t_eq$DBH)-3-1)
  R2adj[i] = R2.adjusted
  # Prediction and Errors
  n_va$Pred <- predict(Max_like, newdata=cbind(n_va,g="a"))
  Bias[i] <- 100*mean((n_va$AGB - n_va$Pred)/n_va$AGB)
  RMSE[i] <- 100*sqrt(mean(((n_va$AGB - n_va$Pred)/n_va$AGB)^2))
  MAPE[i] <- 100*mean(abs(n_va$AGB - n_va$Pred)/n_va$AGB)
}
i
# Model fitting statistics:
mean(AIC)
mean(R2adj)
mean(Bias)
mean(RMSE)
mean(MAPE)
hist(Bias)
hist(RMSE)
hist(MAPE)
hist(Bias, main = paste("","" ), xlab = "Bias (%)  của mô hình AGB = a*DBH^b", ylab = "Tần
số",
```

```
    cex.lab=2, cex.axis=2, cex.main=2, cex.sub=2)
# Plots:
# Prediction and Validation data
p <- ggplot(n_va)
p <- ggplot(n_va, aes(x=DBH, y=AGB))
p <- p + geom_point(pch=19,cex=2)
p <- p + geom_line(cex = 1.5, aes(x=DBH, y=Pred))
p <- p + xlab("DBH (cm)") + ylab("AGB (kg)") + theme_bw()
p <- p + labs(title = "AGB = a*DBH^b")
p = p + theme(axis.title.y = element_text(size = rel(1.7)))
p = p + theme(axis.title.x = element_text(size = rel(1.7)))
p <- p + theme(legend.title=element_blank())
p <- p + theme(plot.title = element_text(size = rel(2)))
p = p + theme(axis.text.x = element_text(size=20))
p = p + theme(axis.text.y = element_text(size=20))
p
# The end
```

Kết quả tính toán các chỉ tiêu thống kê và các sai số của mô hình: $AGB = a \times DBH^b$ theo phương pháp k-fold:

```
> # Model fitting statistics:
> mean(AIC)
[1] 1008.058
> mean(R2adj)
[1] 0.9329573
>
> mean(Bias)
[1] -7.896921
> mean(RMSE)
[1] 33.06169
> mean(MAPE)
[1] 22.9158
```

Kết quả tính toán các chỉ tiêu thống kê và các sai số của mô hình: $AGB = a \times DBH2HWD^b$ theo phương pháp k-fold:

```
> # Model fitting statistics:
> mean(AIC)
[1] 989.9389
> mean(R2adj)
[1] 0.9520122
>
> mean(Bias)
[1] -6.062977
> mean(RMSE)
[1] 27.18281
> mean(MAPE)
[1] 19.72642
```

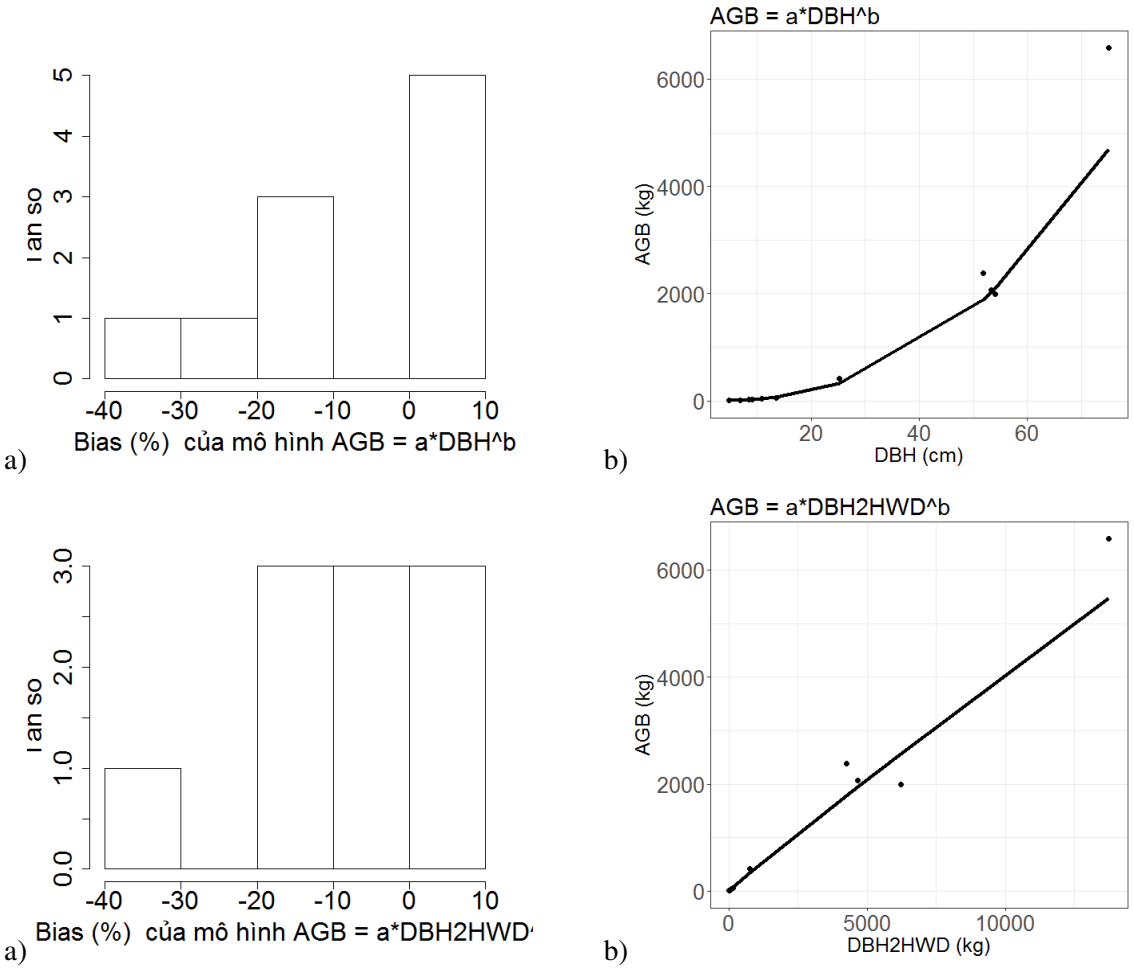
Các kết quả minh họa lập và thẩm định các mô hình AGB theo phương pháp k-fold được tổng hợp trong Bảng 7.11. Kết quả này cho thấy, mô hình tổ hợp ba biến $DBH^2HWD$ có độ tin cậy cao

hơn (AIC bé hơn và $R^2$ cao hơn) và các sai số đều nhỏ hơn so với mô hình AGB chỉ với một biến số DBH. Kết quả áp dụng phương pháp thẩm định chéo k-fold là đồng nhất với LOOCV ở trên.

**Bảng 7.11.** So sánh và thẩm định chéo k-fold (k=10) hai mô hình AGB = f(DBH) và AGB = f($DBH^2HWWD$)

| Chỉ tiêu thống kê, sai số | Mô hình | |
|---|---|---|
| | $AGB = a \times DBH^b$ | $AGB = a \times (DBH^2HWD)^b$ |
| $R^2_{adj.}$ | 0.933 | 0.952 |
| AIC | 1008 | 990 |
| Bias (kg) | -7.89 | -6.06 |
| RMSE % | 33.06 | 27.18 |
| MAPE % | 22.91 | 19.73 |

*Ghi chú: Mô hình và $R^2$, AIC được tính từ k-1 phần dữ liệu độc lập; các sai số Bias, RMSE, MAPE được tính trung bình k lần.*




**Hình 7.33.** Phân bố tần số Bias: a) và giá trị dự đoán qua mô hình so với dữ liệu đánh giá độc lập; b) phương pháp k-fold (k=10) của hai mô hình

Hình 7.31 cho thấy, phân bố Bias của hai mô hình được thẩm định theo phương pháp k-fold có xu hướng lệch phải và chưa tiệm cận chuẩn. Đây là nhược điểm của phương pháp thẩm định chéo k-fold, do số lần lại chưa đủ lớn k = 10.

### 7.8.5 Phương pháp thẩm định chéo sai số mô hình Monte Carlo Cross Validation

Phương pháp Monte Carlo dùng để thẩm định chéo các mô hình được mô tả như sau: Phân chia dữ liệu ngẫu nhiên làm 2 phần, một phần dùng để lập mô hình (thường từ 70 – 80% dữ liệu) và một phần dùng để đánh giá sai số (thường từ 20 – 30% dữ liệu); có trường hợp như Zhang (1997) chia dữ liệu làm hai phần bằng nhau. Mỗi lần như vậy tính toán các chỉ tiêu thống kê đánh giá, so sánh các mô hình như AIC, $R^2$ và các sai số khác nhau như Bias%, RMSE%, MAPE%. Tiến hành lặp lại như vậy R lần để thẩm định các mô hình và đánh giá sai số, cuối cùng, giá trị thống kê so sánh các mô hình và sai số được tính trung bình từ R lần; thường là R = 200 lần (theo Temesgen et al. (2014) và Huy et al. (2016a,b,c)); trong khi đó Zhang (1997) lặp lại lên đến 500 lần. Nguyên tắc lựa chọn số lần lặp lại R là dựa trên cơ sở sai số của mô hình ổn định và có phân bố tần số tiệm cận chuẩn.

Các sai số tương đối áp dụng theo phương pháp thẩm định chéo Monte Carlo với R lần lặp lại ngẫu nhiên như sau (Swanson et al., 2011; Huy et al. 2016a,b,c):

$$Bias\ (\%) = \frac{1}{R}\sum_{r=1}^{R}\frac{100}{n}\sum_{i=1}^{n}\frac{y_i-\hat{y}_i}{y_i} \tag{7.59}$$

$$RMSE\ (\%) = \frac{1}{R}\sum_{r=1}^{R} 100\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i-\hat{y}_i}{y_i}\right)^2} \tag{7.60}$$

$$MAPE\ (\%) = \frac{1}{R}\sum_{r=1}^{R}\frac{100}{n}\sum_{i=1}^{n}\frac{|y_i-\hat{y}_i|}{y_i} \tag{7.61}$$

Trong đó, R là số lần phân chia dữ liệu ngẫu nhiên thành hai phần: lập mô hình (thường là 70-80%) và đánh giá mô hình (thường là 20-30%); n là số dữ liệu đánh giá của mỗi lần rút mẫu (thường là 20-30% mẫu rút ngẫu nhiên) và $y_i$ và $\hat{y}_i$ là giá trị quan sát và dự đoán qua mô hình.

Phương pháp Monte Carlo cho phép sử dụng dữ liệu ngẫu nhiên để lập và tính sai số của mô hình rất khách quan. Ngoài ra với số lần lặp lại R đủ lớn thì hầu như tất cả các dữ liệu đều có thể tham gia lập và đánh giá mô hình và có lặp lại vì việc phân chia mẫu để lập và đánh giá mô hình là có hoàn lại. Với R đủ lớn thì phân bố các chỉ tiêu thống kê so sánh và các sai số sẽ tiệm cận chuẩn với sai số trung binh là ổn định. Có thể nói, đây là phương pháp cung cấp sai số khách quan và chính xác cho mọi trường hợp của dữ liệu quan sát.

Để minh họa cho áp dụng phương pháp Monte Carlo, thiết lập và so sánh, đánh giá sai số của hai mô hình ước tính AGB: $AGB = a\times DBH^b$ và $AGB = a\times DBH^2HWD^b$ trên cơ sở Dữ liệu 11 của 110 cây mẫu sinh khối vùng Nam Trung Bộ.

Sử dụng phần mềm mã nguồn mở R để lập mô hình theo nlme có trọng số và tính toán các chỉ tiêu thống kê so sánh và sai số các mô hình theo phương pháp thẩm định chéo Monte Carlo. Trong đó dữ liệu được phân chia ngẫu nhiên với 80% cho lập mô hình và 20% để thẩm định, lặp lại R = 200 lần.

**Codes trong R để thiết lập mô hình theo nlme có trọng số và thẩm định chéo theo Monte Carlo mô hình: $AGB = a \times (DBH^2HWD)^b$ với 80% dữ liệu lập mô hình, 20% đánh giá sai số, lặp R = 200 lần**

```
# Erase memory
rm(list=ls())
# Clean plot window
dev.off()
# Define the working directory
setwd("E:/1 - Bao Huy 2017/Books All/Textbook for Infomatic Statistic/TextbookDataset Analysis/Dataset")
# Import data
t <- read.table("Du lieu 11 AGB QNam .txt", header=T,sep="\t",stringsAsFactors = FALSE)
# Combination fo variables:
t$DBH2HWD = (t$DBH/100)^2*t$H*t$WD*1000
# install.packages("ggplot2")
library(ggplot2)
library(nlme)
library(cowplot)
library(gridExtra)
# Monte Carlo cross validation 200 times, 80% for training, 20% for error
AIC = rep(0, 200)
R2adj = rep(0, 200)
Bias = rep(0, 200)
RMSE = rep(0, 200)
MAPE = rep(0, 200)
for(i in 1:200){
  n_va <- t[sample(nrow(t), length(t$AGB)/5), ]
  t_eq <- t[!t$ID %in% n_va$ID, ]
   start <- coefficients(lm(log(AGB)~log(DBH2HWD), data=t_eq))
  names(start) <- c("a","b")
  start[1]<-exp(start[1])
   Max_like <- nlme(AGB~a*DBH2HWD^b, data=cbind(t_eq,g="a"), fixed=a+b~1,
              start=start, groups=~g, weights=varPower(form=~DBH2HWD))
  # Outputs of the model
  k <- summary(Max_like)$modelStruct$varStruct[1]
  k
  t_eq$Max_like.fit <- fitted.values(Max_like)
  t_eq$Max_like.res <- residuals(Max_like)
  t_eq$Max_like.res.weigh <- residuals(Max_like)/t_eq$DBH2HWD^k
```

```
  # Calcul of AIC, R2,
  AIC[i] = AIC(Max_like)
  R2 <- 1- sum((t_eq$AGB - t_eq$Max_like.fit)^2)/sum((t_eq$AGB - mean(t_eq$AGB))^2)
  R2.adjusted <- 1 - (1-R2)*(length(t_eq$DBH)-1)/(length(t_eq$DBH)-3-1)
  R2adj[i] = R2.adjusted
  # Prediction and Errors
  n_va$Pred <- predict(Max_like, newdata=cbind(n_va,g="a"))

  Bias[i] <- 100*mean((n_va$AGB - n_va$Pred)/n_va$AGB)
  RMSE[i] <- 100*sqrt(mean(((n_va$AGB - n_va$Pred)/n_va$AGB)^2))
  MAPE[i] <- 100*mean(abs(n_va$AGB - n_va$Pred)/n_va$AGB)
}
i
# Model fitting statistics:
mean(AIC)
mean(R2adj)
mean(Bias)
mean(RMSE)
mean(MAPE)
hist(Bias)
hist(RMSE)
hist(MAPE)
hist(Bias, main = paste("","" ), xlab = "Bias (%)  của mô hình AGB = a*DBH2HWD^b", ylab =
"Tần số",
     cex.lab=2, cex.axis=2, cex.main=2, cex.sub=2)

# Plots:
# Prediction and Validation data
p <- ggplot(n_va)
p <- ggplot(n_va, aes(x=DBH2HWD, y=AGB))
p <- p + geom_point(pch=19,cex=2)
p <- p + geom_line(cex = 1.5, aes(x=DBH2HWD, y=Pred))
p <- p + xlab("DBH2HWD (kg)") + ylab("AGB (kg)") + theme_bw()
p <- p + labs(title = "AGB = a*DBH2HWD^b")
p = p + theme(axis.title.y = element_text(size = rel(1.7)))
p = p + theme(axis.title.x = element_text(size = rel(1.7)))
p <- p + theme(legend.title=element_blank())
p <- p + theme(plot.title = element_text(size = rel(2)))
p = p + theme(axis.text.x = element_text(size=20))
p = p + theme(axis.text.y = element_text(size=20))
p
```

Kết quả tính toán các chỉ tiêu thống kê và các sai số của mô hình: $AGB = a \times DBH2HWD^b$ theo phương pháp Monte Carlo:

```
> # Model fitting statistics:
> mean(AIC)
[1] 880.5662
> mean(R2adj)
[1] 0.951661
>
> mean(Bias)
[1] -6.018152
> mean(RMSE)
[1] 27.38712
> mean(MAPE)
[1] 19.66181
```

Kết quả tính toán các chỉ tiêu thống kê và các sai số của mô hình: $AGB = a \times DBH^b$ theo phương pháp Monte Carlo:

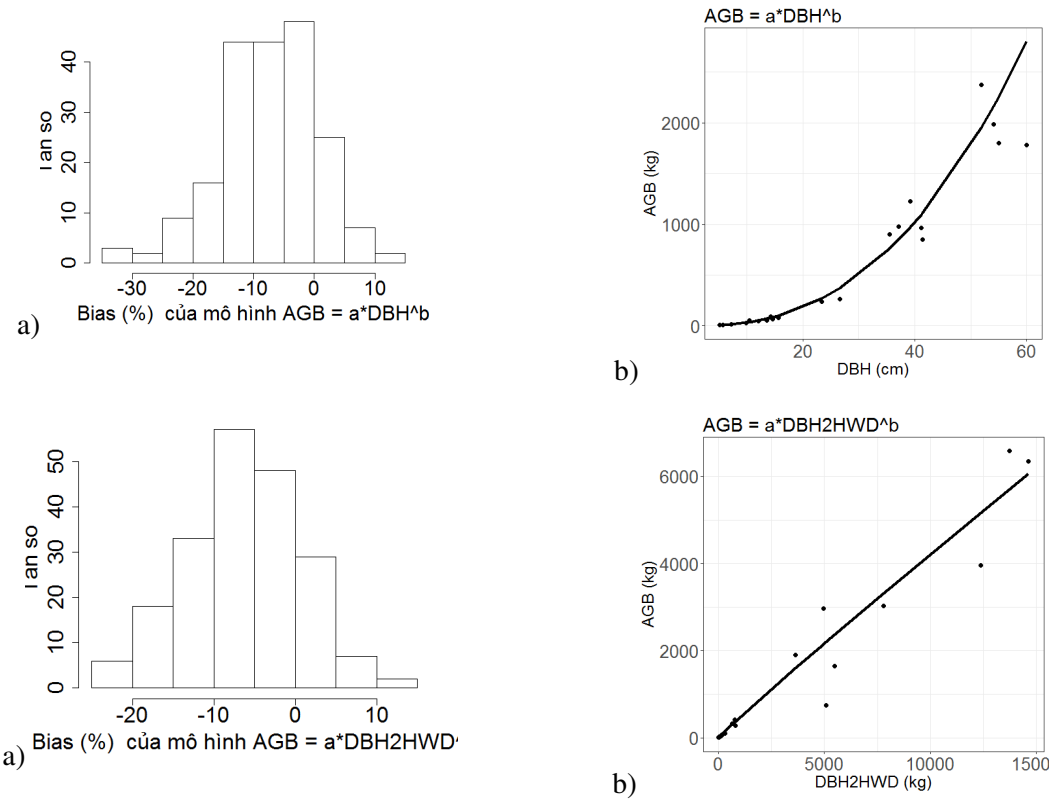
```
> # Model fitting statistics:
> mean(AIC)
[1] 897.5198
> mean(R2adj)
[1] 0.9335265
>
> mean(Bias)
[1] -7.424382
> mean(RMSE)
[1] 34.07336
> mean(MAPE)
[1] 22.9274
```

Các kết quả minh họa lập và thẩm định các mô hình AGB theo phương pháp Monte Carlo được tổng hợp trong Bảng 7.12. Kết quả này cho thấy, mô hình tổ hợp ba biến $DBH^2HWD$ có độ tin cậy cao hơn (AIC bé hơn và $R^2$ cao hơn) và các sai số đều nhỏ hơn so với mô hình AGB chỉ với một biến số DBH. Kết quả áp dụng phương pháp thẩm định chéo Monte Carlo là đồng nhất với LOOCV và k-fold ở trên.

**Bảng 7.12.** So sánh và thẩm định chéo theo phương pháp Monte Carlo cho hai mô hình AGB = f(DBH) và AGB = f($DBH^2HWWD$)

| Chỉ tiêu thống kê, sai số | Mô hình | |
|---|---|---|
| | $AGB = a \times DBH^b$ | $AGB = a \times (DBH^2HWD)^b$ |
| $R^2_{adj.}$ | 0.934 | 0.952 |
| AIC | 898 | 881 |
| Bias (kg) | -7.42 | -6.02 |
| RMSE % | 34.07 | 27.39 |
| MAPE % | 22.92 | 19.67 |

*Ghi chú: Mô hình và $R^2$, AIC được tính từ 80% dữ liệu rút ngẫu nhiên; các sai số Bias, RMSE, MAPE được tính từ 20% dữ liệu đánh giá được rút ngẫu nhiên, độc lập và trung bình từ 200 lần lại.*




**Hình 7.34.** Phân bố tần số Bias: a) và giá trị dự đoán qua mô hình so với dữ liệu đánh giá độc lập; b) phương pháp Monte Carlo của hai mô hình

Hình 7.34 cho thấy, phân bố Bias của hai mô hình được thẩm định theo phương pháp Monte Carlo với 200 lần lặp lại tiệm cận chuẩn; đặc biệt là mô hình có ba biến số tổ hợp $DBH^2HWD$. Vì vậy, phương pháp Monte Carlo có thể xem là đã cung cấp sai số ổn định và khách quan của mô hình so với các phương pháp thẩm định chéo khác.

**Bảng 7.13.** Tổng hợp kết quả thẩm định chéo mô hình $AGB = a\times(DBH^2HWD)^b$ theo các phương pháp khác nhau

| | Phương pháp thẩm định chéo mô hình | | | |
|---|---|---|---|---|
| Chỉ tiêu thống kê, sai số | Dữ liệu độc lập | LOOCV | k-fold | Monte Carlo |
| $R^2_{adj}$ | 0.957 | 0.953 | 0.952 | 0.952 |
| AIC | 887.0 | 1089 | 990 | 881 |
| Bias (%) | -5.99 | -5.88 | -6.06 | -6.02 |
| RMSE (%) | 25.95 | 19.68 | 27.18 | 27.39 |
| MAPE (%) | 21.91 | 19.68 | 19.73 | 19.67 |

Trong ví dụ minh họa cho mô hình sinh khối, với kết quả tổng hợp đánh giá sai số mô hình tốt nhất $AGB = a\times(DBH^2HWD)^b$ theo bốn phương pháp khác nhau ở Bảng 7.13 cho thấy, nếu lấy kết quả theo Monte Carlo làm chuẩn (vì có sai số ổn định và phân bố chuẩn), thì sai số cung cấp theo

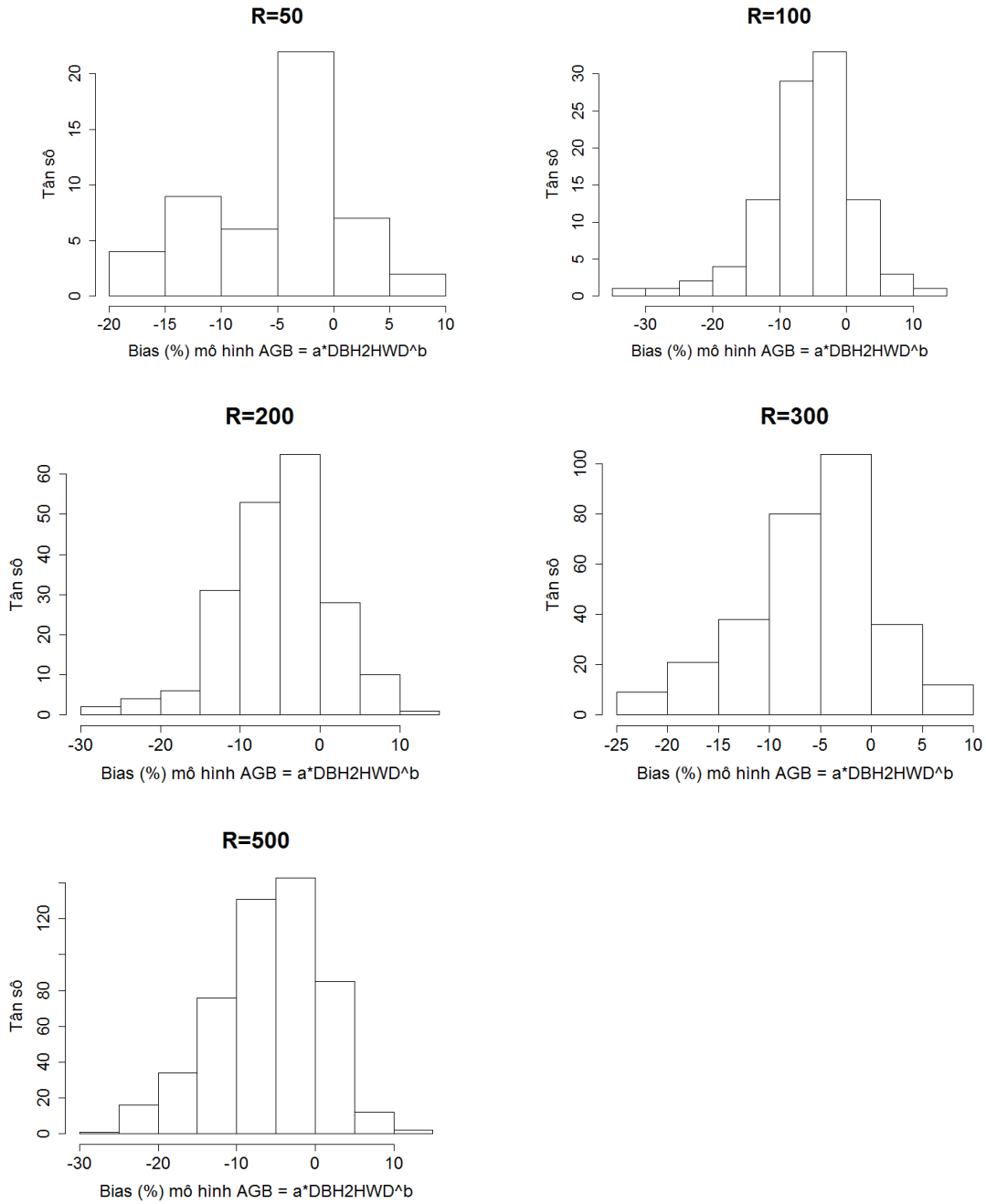
phương pháp k-fold là khá tương đồng, tuy nhiên, k-fold cho sai số chưa có phân bố chuẩn. Trong khi đó, hai phương pháp dùng dữ liệu độc lập, hoặc LOOCV có sai lệch sai số RMSE và MAPE và có phân bố sai số sai lệch. Vì vậy, phương pháp Monte Carlo dùng thẩm định chéo các mô hình sẽ cung cấp sai số ổn định, khách quan khi số lần lặp tăng lên đủ lớn, trong ví dụ này là 200 lần.

Thử nghiệm thay đổi số lần lặp lại theo phương pháp Monte Carlo trong thẩm định chéo mô hình $AGB = a \times (DBH^2 HWD)^b$ được lập theo phương pháp phi tuyến Maximum Likelihood có trọng số; số lần lặp R thay đổi từ 50, 100, 200, 300 và 500 trên cơ sở Dữ liệu 11.

**Bảng 7.14.** So sánh các chỉ tiêu thống kê, sai số thẩm định chéo mô hình $AGB = a \times (DBH^2 HWD)^b$ theo Monte Carlo với số lần lặp lại R khác nhau

| Chỉ tiêu thống kê, sai số | Số lần lặp R theo phương pháp thẩm định chéo Monte Carlo | | | | |
|---|---|---|---|---|---|
| | 50 | 100 | 200 | 300 | 500 |
| $R^2_{adj}$ | 0.951 | 0.952 | 0.952 | 0.952 | 0.951 |
| AIC | 880 | 884 | 880 | 879 | 881 |
| Bias (%) | -5.20 | -5.61 | -5.17 | -5.65 | -5.94 |
| RMSE (%) | 28.62 | 28.43 | 27.18 | 28.52 | 28.01 |
| MAPE(%) | 19.58 | 20.18 | 19.76 | 19.91 | 19.66 |

*Ghi chú: $R^2$, AIC được tính từ 80% dữ liệu rút ngẫu nhiên; các sai số Bias, RMSE, MAPE được tính từ 20% dữ liệu đánh giá được rút ngẫu nhiên, độc lập và tính trung bình từ R lần lại.*




**Hình 7.35.** Phân bố Bias của mô hình AGB = a×(DBH$^2$HWD)$^b$ theo phương pháp Monte Carlo với số lần lặp khác nhau

Kết quả thử ở Bảng 7.14 với số lần lặp R khác nhau cho thấy, với R = 50 trở lên thì các chỉ tiêu thống kê của mô hình (AIC, $R^2_{adj.}$) và các sai số Bias, RMSE, MAPE đã ổn định, không có sự khác biệt khi R tăng đến 500 lần. Tuy nhiên, xét thêm phân bố của Bias ở Hình 7.35 thì, với R=50, 100 phân bố có nhiều đỉnh, khi R ≥ 200 lần, dạng phân bố của Bias đã tiệm cận chuẩn. Vì vậy, có thể nói, trong trường hợp mô hình AGB ở vùng sinh thái này, sử dụng thẩm định chéo Monte Carlo với R = 200 lần là hợp lý, cung cấp sai số ổn định và có phân bố chuẩn, phù hợp với nghiên cứu

của Temesgen et al., (2014) và Huy et al., (2016a,b,c). Không nhất thiết lặp lại số quá lớn (R = 500 lần) như Zhang (1997) đề nghị.

Trong trường hợp ước lượng mô hình theo phương pháp phi tuyến Maximum Likehood có trọng số và có xét đến các nhân tố môi trường ảnh hưởng (random effect) (theo chương trình nlme trong R), codes trong R để tiến hành áp dụng thẩm định chéo theo Monte Carlo có sự điều chỉnh; được viết lại như sau:

Codes ước lượng mô hình $AGB = a \times (DBH^2HWD)^b$ theo phương pháp phi tuyến Maximum Likehood có trọng số và có xét đến các nhân tố môi trường ảnh hưởng (random effect) và thẩm định chéo theo Monte Carlo. Với 80% dữ liệu lập mô hình, 20% thẩm định, lặp lại R = 200 lần:

```
# Erase memory
rm(list=ls())

# Clean plot window
dev.off()

# Define the working directory
setwd("E:/1 - Bao Huy 2017/Books All/Textbook for Infomatic Statistic/TextbookDataset Analysis/Dataset")

# Import data
t <- read.table("Du lieu 13 AGB Viet Nam.txt", header=T,sep="\t",stringsAsFactors = FALSE)
# Combination fo variables:
t$DBH2HWD = (t$DBH/100)^2*t$H*t$WD*1000
length(t$AGB)

# install.packages("ggplot2")
library(ggplot2)
library(nlme)
library(cowplot)
library(gridExtra)

# Cross Validation Monte Carlo: 80/20%, R =200 times
AIC = rep(0, 200)
R2adj = rep(0, 200)
Bias = rep(0, 200)
RMSE = rep(0, 200)
MAPE = rep(0, 200)

for(i in 1:200){
  n_va <- t[sample(nrow(t), length(t$AGB)/5), ]
  t_eq <- t[!t$ID %in% n_va$ID, ]

  # Develop Model:
  start <- coefficients(lm(log(AGB)~log(DBH2HWD), data=t_eq))
  names(start) <- c("a","b")
```

```
  start[1]<-exp(start[1])

  Max_like2 <- nlme(AGB~a*DBH2HWD^b, data=t_eq, fixed=a+b~1, random=a~1,
                start=start, groups=~Region, weights=varPower(form=~DBH2HWD))

  # Outputs of the model
  k <- summary(Max_like2)$modelStruct$varStruct[1]
  k
  t_eq$Max_like2.fit <- fitted.values(Max_like2)
  t_eq$Max_like2.res <- residuals(Max_like2)
  t_eq$Max_like2.res.weigh <- residuals(Max_like2)/t_eq$DBH2HWD^k

  # Calcul of AIC, R2,
  AIC[i] = AIC(Max_like2)
  R2 <- 1- sum((t_eq$AGB - t_eq$Max_like2.fit)^2)/sum((t_eq$AGB - mean(t_eq$AGB))^2)
  R2.adjusted <- 1 - (1-R2)*(length(t_eq$DBH)-1)/(length(t_eq$DBH)-3-1)
  R2adj[i] = R2.adjusted

  # Prediction of the model for validation
  n_va$Pred <- predict(Max_like2, newdata=n_va)

  # Calcul of RMSE, Bias, MAPE%:
  Bias[i] = 100*mean((n_va$AGB - n_va$Pred)/n_va$AGB)
  RMSE[i] = 100*sqrt(mean(((n_va$AGB - n_va$Pred)/n_va$AGB)^2))
  MAPE[i] = 100*mean(abs(n_va$AGB - n_va$Pred)/n_va$AGB)
}

i
# Model fitting statistics:
mean(AIC)
mean(R2adj)

mean(Bias)
mean(RMSE)
mean(MAPE)

hist(Bias)
hist(RMSE)
hist(MAPE)
hist(Bias, main = paste("","" ), xlab = "Bias (%) c???a mô hình AGB = a*DBH2HWD^b", ylab
= "T???n s???",
     cex.lab=2, cex.axis=2, cex.main=2, cex.sub=2)

# Plots:
# Prediction and Validation data
p <- ggplot(n_va)
p <- ggplot(n_va, aes(x=DBH2HWD, y=AGB))
```

```
p <- p + geom_point(pch=19,cex=2)
p <- p + geom_line(cex = 1.5, aes(x=DBH2HWD, y=Pred))
p <- p + xlab("DBH2HWD (kg)") + ylab("AGB (kg)") + theme_bw()
p <- p + labs(title = "AGB = a*DBH2HWD^b")
p = p + theme(axis.title.y = element_text(size = rel(1.7)))
p = p + theme(axis.title.x = element_text(size = rel(1.7)))
p <- p + theme(legend.title=element_blank())
p <- p + theme(plot.title = element_text(size = rel(2)))
p = p + theme(axis.text.x = element_text(size=20))
p = p + theme(axis.text.y = element_text(size=20))
p
```

Kết quả các chỉ tiêu thống kê và sai số của mô hình $AGB = a \times (DBH^2HWD)^b$ có xét ảnh hưởng vùng sinh thái (random effect) được thẩm định chéo theo Monte Carlo, lặp lại R = 200 lần

```
> mean(AIC)
[1] 7999.769
> mean(R2adj)
[1] 0.9423663
> mean(Bias)
[1] -5.411498
> mean(RMSE)
[1] 27.55405
> mean(MAPE)
[1] 19.19206
```

Khi áp dụng các phương pháp thẩm định chéo các mô hình, sau khi lựa chọn được dạng mô hình tối ưu và để tối ưu hóa sử dụng dữ liệu, mô hình cuối cùng được thiết lập dựa vào toàn bộ dữ liệu, tức là không sử dụng một phần dữ liệu để lập mô hình như khi đánh giá, thẩm định. Đây là ưu điểm của các phương pháp thẩm định chéo, vì tối ưu hóa được việc sử dụng dữ liệu trong xây dựng các mô hình, đặc biệt là các mô hình với dữ liệu khó thu thập, giá thành cao như sinh khối – carbon, sinh trưởng, tăng trưởng cây rừng, lâm phần.

Tiến hành lập mô hình sinh khối với toàn bộ dữ liệu và các chỉ tiêu thống kê của mô hình (AIC, $R^2_{adj,}$); sai số của các mô hình được lấy từ kết quả áp dụng phương pháp Monte Carlo với R = 200 lần lặp (Bảng 7.15).

**Bảng 7.15.** Kết quả ước lượng các mô hình sinh khối từ toàn bộ dữ liệu và sai số từ thẩm định chéo theo phương pháp Monte Carlo

| Mô hình | Trọng số (Weight) | AIC | $R^2_{adj}$ | MAPE (%) |
|---|---|---|---|---|
| $AGB = 0.10959 \times DBH^{2.47432}$ | $1/DBH^k$ | 1119 | 0.934 | 22.1 |
| $AGB = 0.59164 \times (DBH^2HWD)^{0.98655}$ | $1/DBH^k$ | 1090 | 0.954 | 19.7 |

*Ghi chú: Mô hình và các chỉ tiêu AIC, $R^2_{adj}$ được thiết lập từ toàn bộ dữ liệu, sai số MAPE được lấy từ kết quả của phương pháp Monte Carlo với R = 200 lần; k là hệ số của hàm phương sai; P-value của các tham số <0.0001.*

Chương 8

# TIN HỌC THỐNG KÊ CHUYÊN ĐỀ

## 8.1 Mô phỏng quy luật phân bố - cấu trúc rừng, cấu trúc quần thể

Mô phỏng định lượng cấu trúc rừng là một lĩnh vực quan trọng trong khoa học lâm sinh. Nó giúp cho việc điều tra quần thể, mô tả phân bố, thực hiện các biến pháp kỹ thuật lâm sinh. Đặt nền móng cho định lượng cấu trúc rừng Việt Nam là Nguyễn Văn Trương (1983), Đồng Sĩ Hiền (1974).

Cấu trúc rừng định lượng bao gồm ba kiểu dạng chính: Phân bố số cây theo cấp kính (N/DBH), phân bố số cây theo cấp chiều cao hoặc theo tầng tán (N/H) và mạng hình tọa độ cây rừng trên mặt đất rừng, còn gọi là phân bố trên mặt bằng.

Tin học thống kê hỗ trợ đắc lực cho việc mô tả, định lượng, dự đoán cấu trúc rừng. Sử dụng thống kê chỉ ra được quy luật biến đổi của cá thể trong quần thể. Quy luật này còn có thể mô phỏng mẫu theo các quy luật phân bố lý thuyết làm cơ sở cho việc ước tính rộng ra cho hệ sinh thái rừng.

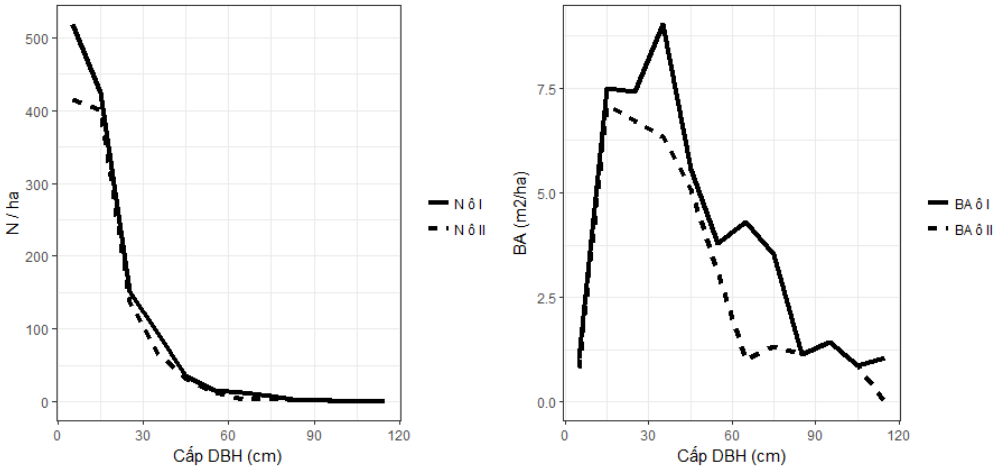
### 8.1.1 Sắp xếp và vẽ biểu đồ phân bố tần số xuất hiện theo cấp, cỡ, hạng

Phân bố tần số là một biểu diễn sự biến đổi của số lượng cá thể, tỷ lệ cá thể theo một biến số quan sát được phân cấp. Nó cho chúng ta dữ liệu định lượng về phân bố cá thể trong quần thể và quy luật biến đổi của nó. Điều này giúp cho việc xử lý lâm sinh, sinh thái theo hướng tuân theo các quy luật ổn định và bền vững của hệ sinh thái rừng.

Trong nghiên cứu xã hội, người ta cần nghiên cứu tần số phân bố số người theo cấp tuổi để biết sự phân bổ con người theo các thế hệ để có chiến lược quản lý nguồn nhân lực. Trong quản lý tài nguyên thiên nhiên, thường cần nghiên cứu sự phân bố số lượng cá thể loài theo cấp tuổi, cấp kích thước để biết được quy luật biến đổi cá thể theo thế hệ, theo kích thước, chất lượng,... là cơ sở quản lý, bảo tồn và định hướng khai thác sử dụng bền vững. Trong lâm nghiệp, thường cần sắp xếp phân bố số cây theo cỡ kính (N/D), số cây theo cỡ chiều cao (N/H), số cây theo cấp thể tích (N/V), số cây theo loài cây theo các tầng rừng, thế hệ để tổ chức quản lý điều chế rừng.

Một vài dạng phân bố thường được nghiên cứu và áp dụng, đó là phân bố số cây trên ha (N cây/ha) theo cấp đường kính ngang ngực (DBH, cm), N theo cấp chiều cao (H, m); số cá thể của động thực vật theo cấp tuổi A (năm). Hình 8.1 dưới đây giới thiệu phân bố tần số N/DBH và phân

bố tổng tiết diện ngang lâm phần BA ($m^2$/ha) BA/DBH của kiểu rừng tự nhiên lá rộng thường xanh Việt Nam.




**Hình 8.1.** Phân bố số cây N (cây/ha) theo cấp kính DBH (cm)(Phải) và phân bố tổng tiết diện ngang BA ($m^2$/ha) theo cấp DBH (Trái) của rừng lá rộng thường xanh vùng Nam Trung Bộ. *(Huy et al., 2016b)*

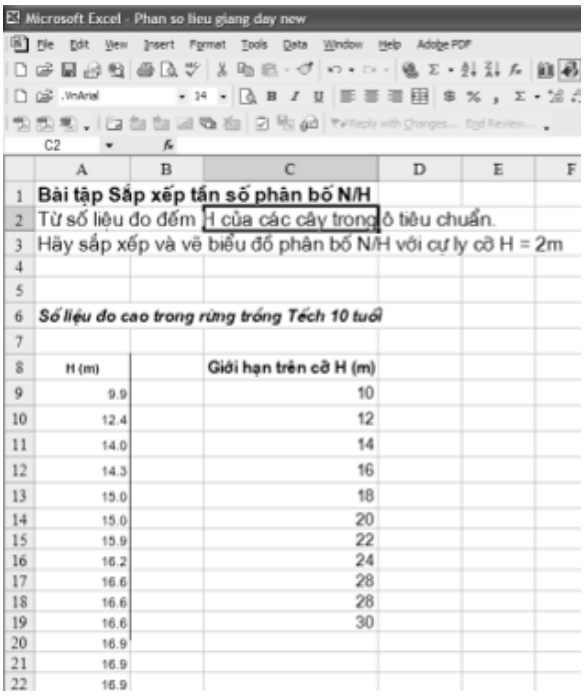
Trong nghiên cứu quy luật phân bố tần số, cá thể, người ta có thể thay đổi phạm vi mỗi cấp để tiếp cận được quy luật của quần thể và có thể mô phỏng bằng một hàm toán học.

Để có được tần số phân bố, sử dụng chức năng sắp xếp bảng phân bố tần số theo một nhân tố theo từng cấp, hạng,... và vẽ đồ thị phân bố trong excel.

Ví dụ, cũng từ số liệu quan sát rừng trồng tếch 10 tuổi, tiến hành sắp xếp phân bố thực nghiệm N/H và vẽ biểu đồ (cấp H là 2m) trong excel như sau:

Nhập số liệu chiều cao vào bảng tính theo cột (H (m))

Lập một cột giới hạn trên cỡ H. Vd: cỡ 2m.



Thực hiện sắp xếp tần số: Data/Data Analysis/Histogram

Xuất hiện hộp thoại Histogram, xác định:

+ Input range: Khai báo khối dữ liệu

+ Bin range: Khai báo khối chứa cự ly tổ.

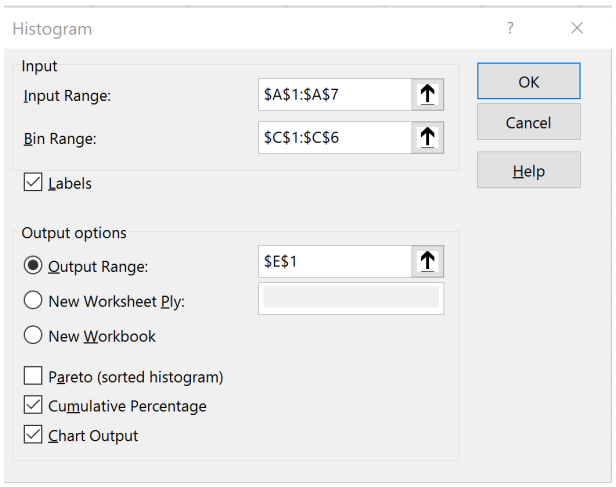
+ Label: Chọn nếu có tiêu đề

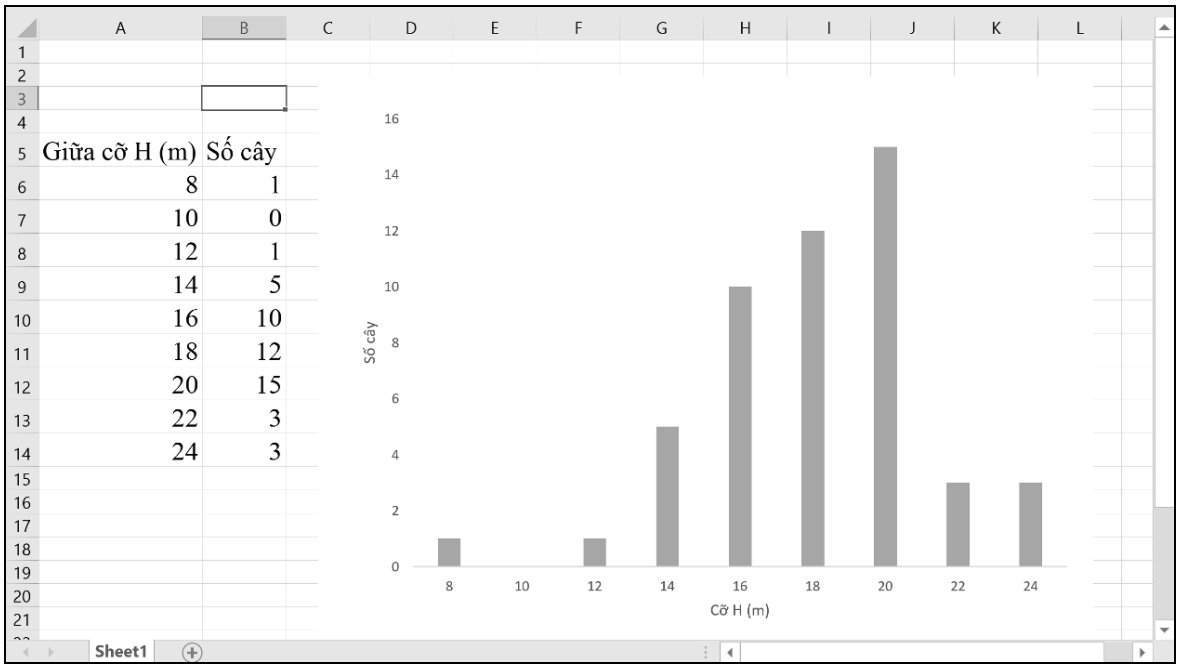
+ Output range: Khai địa chỉ ô trên trái nơi đưa ra kết quả.

+ Cumulative percentage: Tính phần trăm tần số tích lũy. (Đánh dấu).

+ Chart output: Vẽ biểu đồ. (Đánh dấu chọn).

+ OK.




**Hình 8.2.** Kết qủa sắp xếp phân bố tần số N/H trong excel

Kết quả sắp xếp tần số cho được một dãy dữ liệu theo cấp và biểu đồ phân bố. Nó phản ảnh cụ thể hơn đặc trưng mẫu và cho thấy hình ảnh của kiểu dạng phân bố theo cấp, thế hệ; từ đó giúp cho việc phân tích quần thể và đưa ra quyết định quản lý, sử dụng bền vững. Ví dụ trong biểu đồ trên, số cây phân hóa khá mạnh theo cấp chiều cao, một số cây sinh trưởng kém ở cấp chiều cao nhỏ 8 – 12m, một số cây vượt tán có cấp H trên 22m; giải pháp đề nghị ở đây là tỉa thưa, loại bỏ bớt cây sinh trưởng kém có H < 12m và có thể tỉa thưa một số cây lớn với H > 22m để lợi dụng trung gian, lúc này cá thể sẽ có kích thước tập trung trong phạm vi 14 – 22m và có đủ không gian dinh dưỡng để phát triển.

### 8.1.2 Kiểm tra thuần nhất K mẫu quan sát đứt quãng ứng dụng trong kiểm tra sự thuần nhất của các dãy phân bố N/DBH ở các ô tiêu chuẩn, số cá thể theo tuổi

Trong thực tế, để điều tra đánh giá, nghiên cứu một đối tượng rừng, một trạng thái, điều tra đa dạng sinh học; chúng ta thường rút mẫu theo ô tiêu chuẩn, quan sát các bầy đàn, sau đó, gộp chung để tính toán. Tuy nhiên, có khả năng một số ô mẫu, vị trí quan sát không nằm trong một trạng thái hoặc cùng một đàn, do đó, nếu gộp chung để tính toán trung bình, mô phỏng cấu trúc quần thể chung sẽ gặp sai số và biến động rất lớn, đồng thời không phản ảnh đúng đối tượng, trạng thái nghiên cứu. Do vậy, cần có sự kiểm tra sự thuần nhất của các số liệu này ở các ô mẫu, dãy số liệu bầy đàn trước khi gộp chung để tính toán như một tổng thể hoặc tách nhóm thành các trạng thái, đàn khác nhau.

Trong trường hợp này cần ứng dụng tiêu chuẩn thống kê $\chi^2$ để kiểm tra K mẫu quan sát đứt quãng, cụ thể là, ứng dụng để kiểm tra các dãy phân bố N/DBH ở các ô mẫu hoặc dãy phân bố số cá thể theo tuổi/thế hệ để đánh giá chúng có cùng một bầy đàn, một trạng thái hay không (Laar và Akca, 2007).

Ví dụ ở Bảng 8.1 là các dãy phân bố N/DBH của i ô mẫu, được sắp xếp lần lượt i = 1,2, ... k (k ô) và được phân bố theo j cấp kính, được sắp xếp theo thứ tự j = 1, 2, ... m (m cấp kính).

**Bảng 8.1.** Phân bố N/DBH của i ô mẫu theo các cỡ kính j

| | Tần số $f_{ij}$ (i=1 .... k) (N/ha) | | | | | |
|---|---|---|---|---|---|---|
| Cỡ DBH (j= 1 ... m) | ô 1 | ô 2 | ô 3 | ô 4 = k | Tổng | fj |
| 8 | 12 | 11 | 9 | 8 | 40 | f1 |
| 12 | 134 | 125 | 150 | 145 | 554 | f2 |
| 16 | 97 | 80 | 97 | 88 | 362 | |
| 20 | 56 | 56 | 41 | 54 | 207 | |
| 24 | 45 | 34 | 31 | 31 | 141 | |
| 28 | 33 | 31 | 25 | 31 | 120 | |
| 32 | 21 | 21 | 21 | 21 | 84 | |
| 36 | 15 | 11 | 15 | 15 | 56 | fj |
| 40 | 11 | 11 | 11 | 9 | 42 | |
| 44 | 9 | 7 | 15 | 5 | 36 | |
| 48 | 6 | 6 | 4 | 6 | 22 | |
| 52 | 3 | 3 | 3 | 1 | 10 | |
| 56 | 1 | 1 | 1 | 1 | 4 | fm |
| Tổng | 443 | 397 | 423 | 415 | 1678 | |
| ni | n1 | n2 | ni | nk | n | |

Gọi $f_{ij}$ là tần số (số cây) theo ô i và cấp kính j; $n_i$ là tổng tần số của ô i và $f_j$ là tổng tần số cấp kính j.

Giả thuyết $H_0$: F1 = F2 = ... = Fk (Cho mọi i và j). Hay nói khác là các dãy phân bố N/DBH là thuần nhất ở các ô mẫu khác nhau, hay cùng một trạng thái rừng; ngược lại chúng ở các trạng thái khác nhau và không thuần nhất.

Kiểm tra sự thuần nhất bằng tiêu chuẩn $\chi^2$ như sau:

$$\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{m}\frac{(fij - fj.ni/n)^2}{fj.ni/n} \quad (8.1)$$

So sánh với $\chi^2_{(0.05;\ df = (m-1)(k-1)}$. Nếu $\chi^2_t < \chi^2_{(0.05;\ df = (m-1)(k-1)}$ thì chấp nhận giải thuyết $H_0$, có nghĩa là các dãy phân bố N/DBH đồng nhất với nhau hay các ô mẫu thu thập nằm trong cùng một trạng thái, hệ thống và có thể gộp chung để ước lượng đặc trưng mẫu, mô phỏng phân bố N/DBH chung.

Trong ví dụ trên $\chi^2_t = 20.42 < \chi^2_{(0.05;\ df=36)} = 51.00$ (xác định theo hàm = Chiinv(0.05, 36) trong excel). Do đó, có thể xem các dãy N/DBH của 4 ô mẫu là thuần nhất hoặc nằm trong cùng một trạng thái, có thể gộp để mô phỏng N/DBH chung.

Việc ước lượng được số cá thể voi rừng trong điều tra giám sát đa dạng sinh học là việc làm khó, trong lập dự án bảo tồn voi (Bảo Huy và cộng sự, 2009) đã ứng dụng phương pháp kiểm tra sự thuần nhất của các dãy phân bố rời rạc bằng tiêu chuẩn $\chi^2$ để xây dựng phương pháp kiểm tra sự đồng nhất giữa các nhóm/đàn voi; từ đó dự báo được số lượng đàn/nhóm và cá thể voi hoang dã ở Đăk Lăk. Cách tiến hành như sau:

Trên các tuyến, điểm habitat tiến hành xác định tần số xuất hiện cá thể voi rừng theo cấp tuổi. Việc phân chia cấp tuổi được dựa vào các giai đoạn sinh học, sinh lý, khả năng thuần dưỡng, săn bắt của voi ở Bảng 8.2.

**Bảng 8.2.** Phân chia cấp tuổi voi

| Cấp tuổi | Đặc điểm sinh học, sinh lý |
|---|---|
| < 5 | Non, có thể bắt để thuần dưỡng |
| 5 - 15 | Trẻ, chưa có khả năng sinh sản và săn bắt |
| 16 - 45 | Trung niên, khả năng sinh sản tốt, tham gia săn bắt tốt |
| 46 - 55 | Già, giảm khả năng sinh sản và săn bắt |
| > 55 | Yếu, không còn khả năng sinh sản và săn bắt |

*Nguồn: Bảo Huy và cộng sự, Dự án bảo tồn Voi Đăk Lăk, FREM, ĐHTN, 2009*

Kết quả trên 49 tuyến, điểm habitat ở 5 khu vực phân bố voi rừng, đã đo, đếm và thống kê được 324 tần số xuất hiện cá thể voi rừng theo 5 cấp tuổi. Vấn đề đặt ra là đàn voi di chuyển rộng, thường xuyên, do đó, có nhiều khả năng trùng lặp số liệu điều tra tần số xuất hiện ở các tuyến điểm khác nhau. Do vậy, đã áp dụng thống kê xác suất sinh học để kiểm tra sự thuần nhất của dãy phân bố tần số cá thể theo cấp tuổi của các nhóm/đàn voi đã điều tra. Các nhóm đàn đồng nhất sẽ được

gộp lại và xem như một đàn độc lập, từ đây ước tính được số đàn/nhóm voi và cá thể voi rừng. Trong trường hợp này, sử dụng tiêu chuẩn thống kê $\chi^2$ để kiểm tra k mẫu quan sát đứt quãng, cụ thể là ứng dụng để kiểm tra các dãy phân bố tần số cá thể voi rừng theo cấp tuổi ở các tuyến, điểm habitat.

Dãy phân bố tần số cá thể voi theo cấp tuổi của các tuyến, habitat được sắp xếp lần lượt i = 1,2, .... k (Số tuyến, điểm habitat); và phân bố số cá thể theo cấp tuổi được sắp xếp theo thứ tự j = 1, 2, .... m (Số cấp tuổi) (Bảng 8.3).

**Bảng 8.3.** Bảng sắp xếp phân bố tần số cá thể voi fij của i tuyến/habitat theo các cấp tuổi j

| | Tần số fij (i=1 .... k) | | | | |
|---|---|---|---|---|---|
| Cấp tuổi (j= 1 ... m) | *Tuyến 1* | Tuyến 2 | Habitat 3 | Tuyến k | Tổng |
| < 5 | f11 | f21 | f31 | fk1 | f1 |
| 5 - 15 | f12 | f22 | f32 | fk2 | f2 |
| 16 - 45 | | | | | |
| 46 - 55 | | | | | |
| > 55 | f1m | f2m | f3m | fkm | fm |
| Tổng | n1 | n2 | ni | nk | n |

Gọi $f_{ij}$ là tần số voi rừng theo tuyến/habitat i và cấp tuổi j; $n_i$ là tổng tần số của tuyến/habitat i và $f_j$ là tổng tần số cấp tuổi j.

Giả thuyết $H_0$: $F_1 = F_2 = ... = F_k$ (Cho mọi i và j). Hay nói khác là các dãy phân bố tần số cá thể đàn voi theo cấp tuổi là đồng nhất ở các tuyến/habitat khác nhau.

Kiểm tra sự đồng nhất bằng tiêu chuẩn $\chi^2$: So sánh $\chi^2_t$ tính toán với $\chi^2_{(0.05;\ df = (m-1)(k-1))}$. Nếu $\chi^2_t < \chi^2_{(0.05;\ df = (m-1)(k-1))}$ thì chấp nhận giả thuyết $H_0$, có nghĩa là phân bố tần số voi theo cấp tuổi ở các tuyến/habitat đó nằm trong một đàn, cho dù chúng xuất hiện ở nhiều nơi, nhiều lần; từ đây tính tần số cá thể bình quân theo cấp tuổi của đàn đó hoặc chọn dãy tần số cao nhất để dự báo số cá thể tối đa theo cấp tuổi. Nếu ngược lại $\chi^2_t > \chi^2_{(0.05;\ df = (m-1)(k-1))}$ thì các dãy phân bố cá thể voi theo cấp tuổi ở các tuyến/habiat là độc lập, hay nói khác, chúng từ các nhóm/đàn khác nhau; từ đây thống kê được số đàn, cá thể theo cấp tuổi độc lập. Đã kiểm tra 49 tuyến, điểm habitat và cho thấy có thể gộp lại thành 6 nhóm/đàn voi có sự đồng nhất và có 4 nhóm/đàn voi độc lập.

Trên cơ sở kết quả kiểm tra sự đồng nhất của các nhóm/đàn voi, đã xác định được 10 nhóm/đàn voi ở Đăk Lăk. Cuối cùng kiểm tra 10 nhóm/đàn voi này xem chúng có thực sự độc lập nhau không để xác định được số cá thể voi ở Đăk Lăk (Bảng 8.4).

**Bảng 8.4**. Phân bố cá thể voi theo tuổi của 10 đàn voi hoang dã ở Đăk Lăk

| Cấp tuổi | Số cá thể voi theo ký hiệu tuyến - Habitat | | | | | | | | | | Tổng số cá thể voi |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | VQ GY D / ĐA3 | VQ GY D / TA1 | VQ GY D / ĐA4 | VQ GY D / ĐA6 | VQ G YD / ĐA7 2 | VQ G YD / ĐA4 2 | VQ G YD / Tr4 - TA3 | Cty LN EaH Mo / TA3 | Cty LN Ya Lop / TB1 | Cty LN Chu Pa / TA1 | |
| <5 | 0 | 3 | 1 | | 0 | 1 | | 2 | 0 | 0 | 7 |
| 5 - 15 | 1 | 1 | 1 | | 1 | 3 | 11 | 5 | 2 | 0 | 25 |
| 16 - 45 | 2 | 3 | 8 | | 2 | 7 | 4 | 9 | 2 | 2 | 39 |
| 46 - 55 | 0 | 1 | | 1 | 1 | 0 | | 1 | 0 | 1 | 5 |
| >55 | 0 | 2 | | 1 | 0 | 0 | | 3 | 0 | 1 | 7 |
| Tổng số cá thể voi | 3 | 10 | 10 | 2 | 4 | 11 | 15 | 20 | 4 | 4 | 83 |

Kết quả kiểm tra phân bố cá thể voi theo cấp tuổi của 10 đàn voi ở Đăk Lăk, cho thấy: $\chi^2_t = 390.9 > \chi^2_{(0.05, 36)} = 60.0$, điều này khẳng định 10 nhóm đàn này là độc lập. Như vậy đã ước lượng được ở Đăk Lăk trung bình có 83 cá thể voi hoang dã đang sinh sống theo 10 đàn và di chuyển trong pham vi Vườn Quốc Gia Yok Đôn, hai huyện Ea Soup và Ea H'Leo.

### 8.1.3 Mô hình hoá cấu trúc phân bố dạng giảm theo hàm Meyer

Cấu trúc phân bố số cây theo cấp kính (N/DBH) dạng giảm của rừng tự nhiên thường được mô phỏng theo các dạng hàm Meyer (Nguyễn Văn Trương, 1983; Đồng Sĩ Hiền, 1984; Nguyễn Hải Tuất và cộng sự, 2006 và nhiều tác giả khác).

Laar *et al.* (2007) cho thấy, kiểu phân bố giảm trong các kiểu rừng tự nhiên. Đặc điểm của phân bố đường kính của các loài thông của rừng tự nhiên ở Pháp theo dạng giảm J đã được Liocourt (1898) nghiên cứu và định lượng. Tác giả đã giả thuyết rằng, phân bố số cây $N_i$ theo cấp kính $d_i$ dạng $N_i = e^{b0+b1\times di}$, diễn tả theo một cách khác thì $N_i = k \times e^{-a \cdot di}$. Số cây $N_i$ và $n_{i+1}$ có thể được xác định từ phân tích mô hình. Trước đây, Meyer *et al.* (1943, 1951) (dẫn theo Laar *et al.,* 2007) đã mở rộng luật Licourt và giới thiệu khái niệm phân bố đường kính cân bằng, được biết như là một phân bố tạo ra sự bền vững sản lượng rừng. Moser (1976) (dẫn theo Laar *et al.*, 2007) đã trình bày một phương pháp thay thế để mô tả phân bố đường kính dạng J theo hàm Schumacher.

Hàm Meyer viết theo dạng phổ biến (Nguyễn Hải Tuất và cộng sự, 2006; Bảo Huy, 1993) như sau:

$$y = \alpha . e^{-\beta x} \tag{8.2}$$

Trong đó: y là số cây, cá thể; x là cấp, cỡ và α, β là hai tham số của mô hình. Hàm này thích hợp cho mô tả mô phỏng phân bố số cây theo cỡ kính (N/DBH) rừng có phân bố dạng giảm, hoặc mô phỏng sự giảm của số loài theo tầng, theo cỡ kính, tuổi.

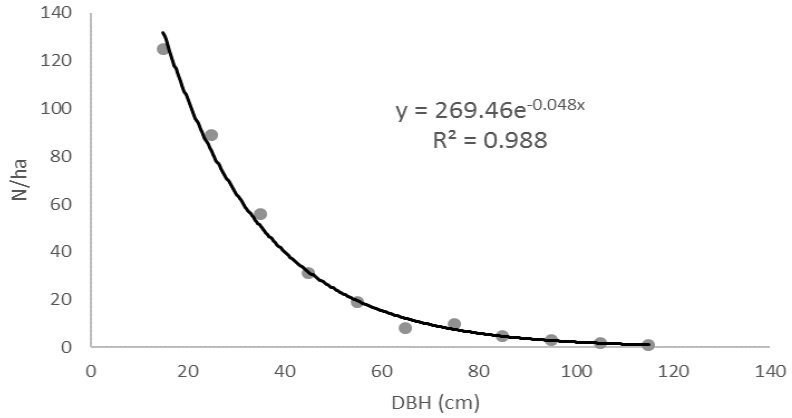
Sau đây giới thiệu cách thức lượng hóa cấu trúc N/DBH theo các hàm Meyer trong excel. Rất đơn giản là trong excel có chương trình lập sẵn để xác định mô hình quan hệ Meyer ngay trên đồ thị.

Nhập dữ liệu tần số N/DBH trên cơ sở sắp xếp tần số theo chương trình Histogam theo Bảng 8.5.

**Bảng 8.5.** Bảng dữ liệu tần số phân bố N/DBH trong Excel

| | A | B |
|---|---|---|
| 1 | DBH (cm) | N (c/ha) |
| 2 | 15 | 125 |
| 3 | 25 | 89 |
| 4 | 35 | 56 |
| 5 | 45 | 31 |
| 6 | 55 | 19 |
| 7 | 65 | 8 |
| 8 | 75 | 10 |
| 9 | 85 | 5 |
| 10 | 95 | 3 |
| 11 | 105 | 2 |
| 12 | 115 | 1 |

Tiến hành vẽ biểu đồ điểm N/DBH, chọn vào điểm kích chuột phải để vào: Add Trendline …, chọn hàm Exponential (dạng Meyer). Kết quả có được biểu đồ quan hệ N/DBH và mô hình Meyer hiển thị trên đồ thị cùng với hệ số xác định $R^2$ (Hình 8.3).




**Hình 8.3.** Mô hình Meyer mô tả phân bố N/DBH trên đồ thị của Excel

Phân bố Meyer còn có thể sử dụng để xem xét phân bố số lượng cá thể của một loài theo các giai đoạn tuổi. Kiểu dạng cấu trúc số cây theo tuổi (N/A) rừng nhiệt đới nhìn chung có dạng giảm, tuổi càng cao thì số cá thể càng ít, bảo đảm cho sự kế tục các thế hệ cây rừng và ổn định quần thể thực vật rừng theo thời gian và có thể mô phỏng theo phân bố dạng J. Với đặc trưng cấu trúc dạng giảm theo thế hệ, tuổi như vậy nên phương thức khai thác chính của rừng tự nhiên là chặt chọn theo cấp kính. Khai thác lớp cây thành thục và nuôi dưỡng rừng trong một luân kỳ để rừng phục hồi trạng thái ban đầu và tiếp tục khai thác lần 2. Việc xác định được cấu trúc N/A của lâm phần và N/A theo từng loài/nhóm loài chính sẽ rất thuận tiện cho việc xác định kỹ thuật lâm sinh như tuổi, đường kính khai thác, luân kỳ. Mô hình hoá cấu trúc N/A thường được biểu diễn tốt bằng hàm Meyer với hệ số tương quan $R^2$ rất cao (Bảo Huy, 1993). Tuy nhiên, trong thực tế rừng tự nhiên việc xác định A là rất khó khăn, do đó, thông thường được thay bằng đường kính, và kiểu cấu trúc phổ biến được nghiên cứu là số cây theo cỡ kính N/DBH, như đã giới thiệu trên để phục vụ cho điều tra, xác định chỉ tiêu kỹ thuật nuôi dưỡng, khai thác rừng.

Ngoài ra, hàm Meyer còn có thể sử dụng để mô tả phân bố số loài theo cấp kính, thế hệ rừng; (Bảo Huy, 1993) đã mô tả cấu trúc N loài/DBH của kiểu rừng nửa rụng ưu hợp bằng lăng – căm xe ở Đăk Lăk có kiểu dạng phân bố là dạng giảm liên tục, có nghĩa khi lên tầng cao, cấp kính lớn, số loài chiếm tỷ lệ thấp, đây là các loài ưu thế sinh thái. Với kiểu rừng này, số loài trên ha là 70 loài thân gỗ, và với cỡ kính thành thục từ 55cm trở lên thì số loài còn khoảng 5 loài. Kiểu dạng cấu trúc này được mô phỏng tốt bằng dạng hàm Meyer.

### 8.1.4 Mô phỏng cấu trúc phân bố theo phân bố khoảng cách - hình học

Phân bố khoảng cách và hình học đã được Nguyễn Hải Tuất (1975) giới thiệu và Bảo Huy (1993) áp dụng để mô tả tốt cấu trúc N/DBH của rừng nửa rụng lá ưu thế bằng lăng dạng giảm (phân bố hình học) và có đỉnh ở cấp kính thứ hai (phân bố khoảng cách).

*i)* Phân bố khoảng cách dùng mô phỏng phân bố dạng có đỉnh ở cỡ thứ hai được trình bày như sau (Nguyễn Hải Tuất, 1975):

$$P(x) = \begin{cases} \Upsilon & x=0 \\ (1-\alpha).(1-\Upsilon).\alpha^{x-1} & x>=1 \end{cases} \qquad (8.3)$$

Trong đó x là mã số các cỡ từ nhỏ đến lớn 0, 1, 2, 3...r ; $\Upsilon$ và $\alpha$ là hai tham số của phân bố.

| | |
|---|---|
| $\Upsilon < (1-\Upsilon)(1-\alpha)$ | Phân bố có đỉnh tại x=1 |
| $\Upsilon = 1 - \alpha$ | Phân bố giảm có thể thay thế bằng phân bố hình học |
| $\Upsilon > (1-\Upsilon)(1-\alpha)$ | Phân bố giảm. |

Ước lượng hai tham số bằng phương pháp cực đại hợp lý:

$$\Upsilon = N_0/N \qquad (8.4)$$

$$\alpha = 1 - \frac{\sum_{i=1}^{r} Ni}{\sum_{i=1}^{r} Ni.xi} \qquad (8.5)$$

Trong đó: $N_o$: Số cá thể ở cỡ kính nhỏ nhất, $N_i$: Số cá thể ở cỡ i, N: Tổng số cá thể, r: số cá thể.

Trình tự tính phân bố lý thuyết N/DBH của rừng nửa rụng lá ưu thế bằng lăng có dạng một đỉnh theo phân bố khoảng cách và kiểm tra sự phù hợp bằng tiêu chuẩn $\chi^2$ trong Excel (Bảo Huy, 1993) (Bảng 8.6 và Hình 8.4):

- Cột A: Mã số x
- Cột B: Giá trị giữa cỡ DBH ($x_i$)
- Cột C: Số cây theo cỡ kính ($N_i$).
- Cột D: $N_i.x_i$.
- Tính 2 tham số:

  $\Upsilon$ = C2/Sum(C2:C12)

  $\alpha$ = 1- Sum(C3:C12)/sum(D3:D12)
- Cột E: Xác suất từng cỡ kính P(xi): Ô E2: Pxo=$\Upsilon$; ô E3: Px1 = (1-$\Upsilon$)(1-$\alpha$)$\alpha$^(a3-1); copy cho các ô dưới.
- Cột F: Tần số lý thuyết: $Nlt_i$: Ô F2: =$C$13*E2; copy cho các ô dưới
- Cột G: Tính $\chi^2$ từng cỡ và tổng. Ô G2: = (F2-C2)^2/F2, copy cho các ô dưới, cộng tổng. Nếu các cỡ kính có $Nlt_i$ <5 thì gộp lại với nhau.
- Ô G14: Tra $\chi^2$ bảng ($\alpha$=0,05 ; K = l-r-1 = 8-2-1=5): =Chiinv(0.05,5) (Với l là số cỡ kính, r số tham số)

Tiêu chuẩn $\chi^2$ theo công thức sau:

$$\chi 2 = \sum_{i=1}^{l} \frac{(y_{lti} - y_i)^2}{y_{lti}} \quad (8.6)$$

Trong đó $y_{lti}$: Số cá thể lý thuyết ở cỡ i, $y_i$: Số cá thể quan sát ở cỡ i, l: Số cỡ, cấp.

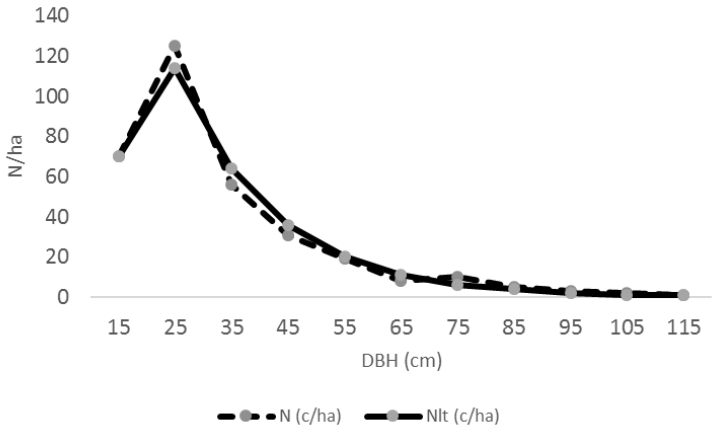
Kết quả $\chi^2 = 7.70 < \chi^2 (0.05, 5) = 11.07$; kết luận phân bố khoảng cách mô phỏng tốt phân bố thực nghiệm N/DBH có đỉnh ở cỡ kính thứ hai.

**Bảng 8.6.** Mô phỏng phân bố N/DBH rừng nửa rụng lá ưu thế bằng lăng theo phân bố khoảng cách trong Excel

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | x | Cỡ DBH (cm) | N (c/ha) | Nixi | Px | Nlt (c/ha) | $\chi 2$ |
| 2 | 0 | 15 | 70 | 0 | 0,212121 | 70 | 0,00 |
| 3 | 1 | 25 | 125 | 125 | 0,345444 | 114 | 1,06 |
| 4 | 2 | 35 | 56 | 112 | 0,193985 | 64 | 1,00 |
| 5 | 3 | 45 | 31 | 93 | 0,108932 | 36 | 0,68 |

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | x | Cỡ DBH (cm) | N (c/ha) | Nixi | Px | Nlt (c/ha) | $\chi 2$ |
| 6 | 4 | 55 | 19 | 76 | 0,061171 | 20 | 0,07 |
| 7 | 5 | 65 | 8 | 40 | 0,034351 | 11 | 0,98 |
| 8 | 6 | 75 | 10 | 60 | 0,01929 | 6 | 2,08 |
| 9 | 7 | 85 | 5 | 35 | 0,010832 | 4 | 1,82 |
| 10 | 8 | 95 | 3 | 24 | 0,006083 | 2 | |
| 11 | 9 | 105 | 2 | 18 | 0,003416 | 1 | |
| 12 | 10 | 115 | 1 | 10 | 0,001918 | 1 | |
| 13 | Tổng | | 330 | 593 | 0,997543 | 329 | 7,70 |
| 14 | | | $\Upsilon$ = | 0,212121 | | $\chi^2$ bảng= | 11,07 |
| 15 | | | $\alpha$ = | 0,561551 | | K=l-r-1 = 8-2-1=5 | |

*Nguồn: Bảo Huy (1993)*




**Hình 8.4.** Phân bố DBH quan sát và phân bố lý thuyết khoảng cách

*ii)* Phân bố hình học mô phỏng phân bố dạng giảm (Nguyễn Hải Tuất, 1975):

$$P(x) = \alpha^x . (1-\alpha) x=0, 1, 2, 3...r \quad (8.7)$$

Trong đó x là mã số cỡ, α là tham số của phân bố.

Ước lượng α bằng phương pháp cực đại hợp lý:

$$\alpha = \frac{\bar{x}}{\bar{x}+1} \quad (8.8)$$

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{r} N_i \,.\, x_i \qquad (8.9)$$

Trong đó N: Tổng số cá thể, r số cấp i, $N_i$: Số cá thể cấp kính i, $x_i$: Mã số cấp i = 0, 1, 2, …r.

Phân bố hình học dùng mô tả các phân bố thực nghiệm dạng giảm như phân bố N/DBH của rừng nửa rụng lá ưu thế bằng lăng (Bảo Huy, 1993), trình tự tính trong Excel: (Hình 8.5 và Bảng 8.7).

- Cột A: Mã số x
- Cột B: Giá trị giữa cỡ DBH.
- Cột C: Số cây theo cỡ kính $N_i$.
- Cột D: $N_i.x_i$.

Tính tham số α:

$\bar{x}$ = D13/C13

α = $\bar{x}$/(+1)

- Cột E: Xác suất từng cỡ kính P(xi): Ô E2: Pxo = (1-α)α^A3; copy cho các ô dưới.
- Cột F: Tần số lý thuyết: $Nlt_i$: Ô F2: =$C$13*E2; copy cho các ô dưới
- Cột G: Tính $\chi^2$ từng cỡ và tổng. Ô G2: = (F2-C2)^2/F2, copy cho các ô dưới, cộng tổng.
- Ô G14: Tra $\chi^2$ bảng (α=0,05 ; K = l-r-1 = 8-1-1=6): =Chiinv(0.05,6). (Với l là số cỡ kính, r số tham số)

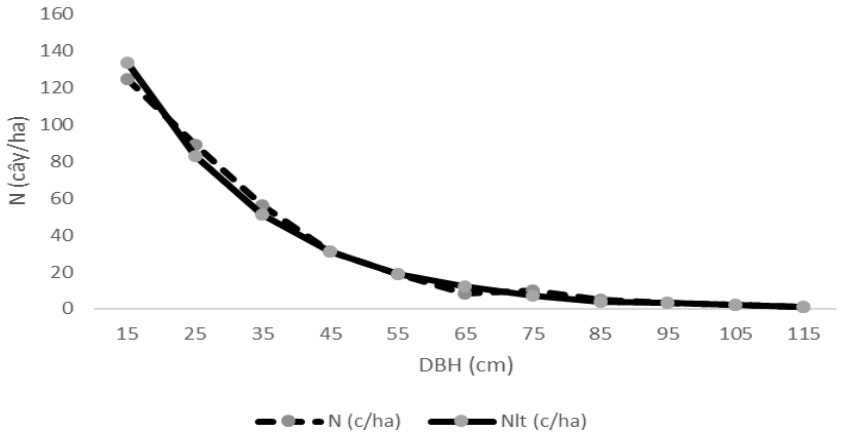
Kết quả $\chi^2 = 4.06 < \chi^2_{(0.05,\ 6)} = 15.59$. Kết luận: Phân bố hình học mô phỏng tốt, phân bố thực nghiệm N/DBH dạng giảm rừng nửa rụng lá ưu thế bằng lăng.

**Bảng 8.7.** Mô phỏng phân bố N/DBH rừng nửa rụng lá ưu thế bằng lăng theo phân bố hình học

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | x | Cỡ DBH (cm) | N (c/ha) | Nixi | Px | Nlt (c/ha) | χ2 |
| 2 | 0 | 15 | 125 | 0 | 0,38521 | 134 | 0,66 |
| 3 | 1 | 25 | 89 | 89 | 0,236823 | 83 | 0,49 |
| 4 | 2 | 35 | 56 | 112 | 0,145597 | 51 | 0,53 |
| 5 | 3 | 45 | 31 | 93 | 0,089511 | 31 | 0,00 |
| 6 | 4 | 55 | 19 | 76 | 0,055031 | 19 | 0,00 |
| 7 | 5 | 65 | 8 | 40 | 0,033832 | 12 | 1,23 |
| 8 | 6 | 75 | 10 | 60 | 0,0208 | 7 | 1,03 |

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | x | Cỡ DBH (cm) | N (c/ha) | Nixi | Px | Nlt (c/ha) | $\chi 2$ |
| 9 | 7 | 85 | 5 | 35 | 0,012788 | 4 | 0,12 |
| 10 | 8 | 95 | 3 | 24 | 0,007862 | 3 | |
| 11 | 9 | 105 | 2 | 18 | 0,004833 | 2 | |
| 12 | 10 | 115 | 1 | 10 | 0,002971 | 1 | |
| 13 | Tổng | | 349 | 557 | 0,995258 | 347 | 4,06 |
| | | | $\bar{x}$ = | 1,595989 | | $\chi^2$ (0.05, 6)= | 12,59 |
| | | | α = | 0,61479 | | (K=8-1-1=6) | |

*Nguồn: Bảo Huy (1993)*




**Hình 8.5.** Phân bố DBH quan sát và phân bố lý thuyết hình học

### 8.1.5 Mô phỏng phân bố, cấu trúc theo phân bố Weibull

Phân bố Weibull là một dạng phân bố tổng quát, có thể mô tả cho các kiểu dạng phân bố khác nhau như : giảm, có đỉnh, tiệm cận chuẩn, lệch phải, trái,... Với sự khái quát và đa dạng như vậy nên Weibull có thể được sử dụng để mô phỏng các kiểu dạng cấu trúc rừng khác nhau.

Phân bố Weibull là phân bố xác suất của biến ngẫu nhiên liên tục với miền giá trị $x \in (0,+\infty)$ (Laar, 2007; Nguyễn Hải Tuất và cộng sự 2006).

Hàm mật độ của phân bố Weibull được viết như sau:

$$f(x) = \alpha\lambda(x - x_{min})^{\alpha-1}.exp(-\lambda(x - x_{min})^{\alpha} \quad (8.10)$$

Hàm phân bố:

$$F(x) = 1 - exp(-\lambda(x - x_{min})^{\alpha} \quad (8.11)$$

Trong đó $x_{min}$: Trị số quan sát nhỏ nhất, x: Các giá trị quan sát, nếu xếp theo tổ thì x là giá trị giữa mỗi tổ.

Khi: $\alpha <= 1$: Phân bố giảm.

$1 < \alpha < 3$: Phân bố lệch trái

$\alpha = 3$: Phân bố đối xứng.

$\alpha > 3$: Phân bố lệch phải.

Ước lượng 2 tham số $\alpha$ và $\lambda$:

Tham số $\alpha$ thường được thăm dò trong một khoảng thích hợp dựa trên các đặc trưng mẫu, cho chạy $\alpha$ để tính $\lambda$. Sau đó kiểm tra sự phù hợp của phân bố lý thuyết bằng tiêu chuẩn $\chi^2$, chọn cặp tham số có $\chi^2$ bé nhất và so sánh với $\chi^2_{(0.05,\ df = 1-r-1)}$ (l: số cỡ, r số tham số).

Tham số $\lambda$ được ước lượng bằng phương pháp cực đại hợp lý:

$$\lambda = \frac{N}{\sum_{i=1}^{r}(x_i - x_{min})^{\alpha}} \quad (8.12)$$

Trong đó: N: Tổng dung lượng quan sát; $N_i$: Tần số tổ i; $x_{min}$: Trị số quan sát nhỏ nhất, x: Các giá trị quan sát, nếu xếp theo tổ thì x là giá trị giữa mỗi tổ.

Tính xác suất cho từng tổ:

+ Tổ 1: $P(x_1)=F(x_1) = 1 - \exp(-\lambda(x_1 + A - x_{min})^{\alpha})$

+ Tổ 2: $P(x_2)=F(x_2) - F(x_1) = \exp(-\lambda(x_1 + A - x_{min})^{\alpha}) - \exp(-\lambda(x_2 + A - x_{min})^{\alpha})$

+ Tổ 3: $P(x_3)=F(x_3) - F(x_2) = \exp(-\lambda(x_2 + A - x_{min})^{\alpha}) - \exp(-\lambda(x_3 + A - x_{min})^{\alpha})$

.....

+ Tổ r: $P(x_r)=F(x_r) - F(x_{r-1}) = \exp(-\lambda(x_{r-1} + A - x_{min})^{\alpha}) - \exp(-\lambda(x_r + A - x_{min})^{\alpha})$

Với A: giá trị 1/2 cự ly tổ.

Tần số lý thuyết $N_{lt}$ cho từng tổ: $Nlt_i = N.P(x_i)$.

Cuối cùng kiểm tra sự phù hợp bằng tiêu chuẩn $\chi^2$.

Tiến hành mô phỏng phân bố N/DBH rừng nửa rụng lá ưu thế bằng lăng có dạng giảm theo phân bố Weibull ở Bảng 8.8 và Hình 8.6 (Bảo Huy, 1993).

**Bảng 8.8.** Mô phỏng phân bố N/DBH rừng nửa rụng lá ưu thế bằng lăng theo phân bố Weibull

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | Cỡ DBH (cm) | N (c/ha) | Alpha | N(x-xmin)^α | λ | P(x) | Nlt (c/ha) | $\chi^2$ |
| 2 | 15 | 125 | 1 | 625,0 | 0,047710 | 0,379420 | 132 | 0,42 |
| 3 | 25 | 89 | | 1335,0 | | 0,235460 | 82 | 0,57 |
| 4 | 35 | 56 | | 1400,0 | | 0,146121 | 51 | 0,49 |
| 5 | 45 | 31 | | 1085,0 | | 0,090680 | 32 | 0,01 |
| 6 | 55 | 19 | | 855,0 | | 0,056274 | 20 | 0,02 |

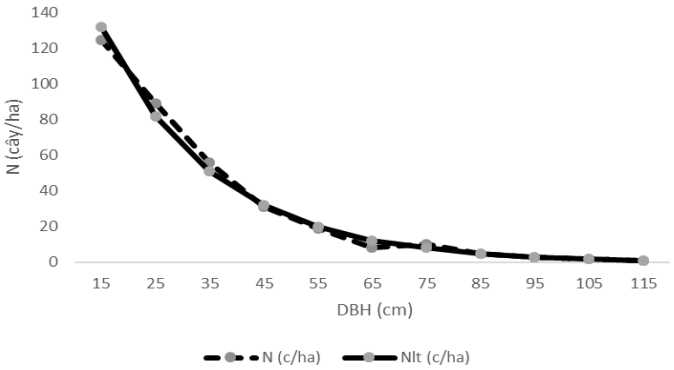
| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | Cỡ DBH (cm) | N (c/ha) | Alpha | N(x-xmin)^α | λ | P(x) | Nlt (c/ha) | $\chi^2$ |
| 7 | 65 | 8 | | 440,0 | | 0,034922 | 12 | 1,44 |
| 8 | 75 | 10 | | 650,0 | | 0,021672 | 8 | 0,78 |
| 9 | 85 | 5 | | 375,0 | | 0,013449 | 5 | 0,02 |
| 10 | 95 | 3 | | 255,0 | | 0,008346 | 3 | 0,00 |
| 11 | 105 | 2 | | 190,0 | | 0,005179 | 2 | |
| 12 | 115 | 1 | | 105,0 | | 0,003214 | 1 | |
| 13 | Tổng | 349 | | 7315,0 | | 1,0 | 347 | 3,76 |
| 14 | | | | | | | $\chi^2$ (0.05, 7) = | 14,07 |
| 15 | | | | | | | K=l-r-1= 9-1-1=7 | |

*Nguồn: Bảo Huy (1993)*

- Cột A: Giá trị giữa cỡ kính 15, 25, ... 115 với cự ly cỡ 10 cm.
- Cột B: Số cây từng cỡ Ni.
- Ô C2: Đưa tham số α thăm dò.
- Cột D: Giá trị: $Ni(x_i - 10)^{\alpha}$. Với $x_{min}$=10. Tính tại ô D2: =B2*(A2-10)^$C$2, sau đó copy cho các ô dưới.
- Ô E2: Tính tham số λ: = B13/Sum(D2:D12).
- Cột F: Tính xác suất P(x) từng tổ: Tính theo công thức địa chỉ ô.
- Cột G: Nlt từng tổ: Ô G2: =$B$13*F2, sau đó copy xuống và tính tổng.
- Cột H: Tính $\chi^2$ từng tổ và tổng $\chi^2$=3.76
- Ô H14: Tra $\chi^2$(0.05, df) =Chiinv(0.05,7)=14.07 (df=l-r-1; l: số cỡ kính kiểm tra, r: số tham số)

Kết quả có $\chi^2 = 3.76 < \chi^2_{(0.05, 7)} = 14.07$: Phân bố Weibull mô phỏng tốt phân bố thực nghiệm N/DBH rừng bằng lăng.

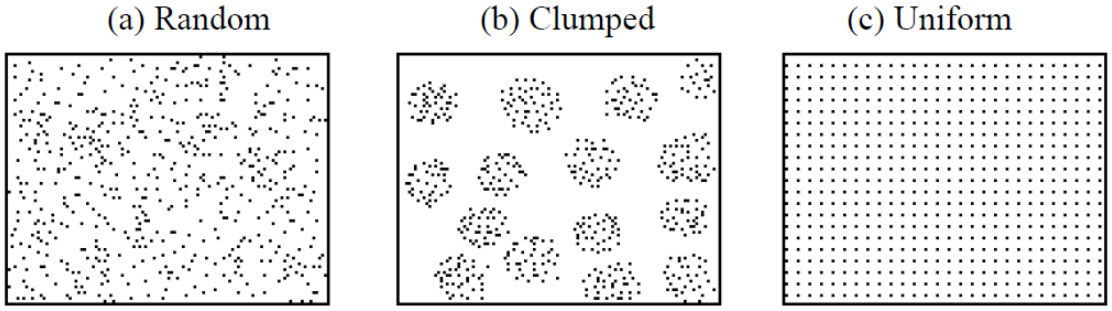
Chú ý: Để chọn được α tối ưu, lần lượt thay giá trị ở ô C2, bảng tính sẽ tự động tính lại, sau đó chọn một α tối ưu với $\chi^2$ bé nhất.




**Hình 8.6.** Phân bố DBH quan sát và phân bố lý thuyết Weibull

### 8.1.6 Xác định kiểu phân bố cây rừng trên mặt đất rừng

Cấu trúc mặt bằng thể hiện sự phân bố và sử dụng không gian dinh dưỡng trên mặt đất rừng, kiểu dạng phân bố thường được chia thành ba kiểu: *ngẫu nhiên, cụm hoặc đều* Hình 8.7 (Jayaraman, 1999). Trong đó, kiểu phân bố cụm thể hiện rừng chưa lợi dụng tốt không gian trên mặt đất. Cho nên, chặt nuôi dưỡng phải bảo đảm sao cho phân bố cây trên mặt đất rừng đồng đều hơn, tạo ra phân bố cách đều hoặc ngẫu nhiên, tránh để rừng ở trạng thái phân bố cụm, ảnh hưởng xấu đến quá trình tái sinh, sinh trưởng và phục hồi rừng. Tóm lại, nghiên cứu phân bố cây trên mặt đất nhằm phục vụ cho việc đề xuất giải pháp kỹ thuật trong chặt nuôi dưỡng, tỉa thưa, khai thác để điều tiết mật độ trên bề mặt đất rừng.




**Hình 8.7.** Ba kiểu phân bố cây trên mặt đất rừng: a) Ngẫu nhiên (Random); b) Cụm (Clumped) và c) Đều (uniform) *(Jayaraman, 1999).*

Phương pháp xác định kiểu phân bố cây rừng trên mặt đất rừng là áp dụng đánh giá, phân bố khoảng cách từ một cây chọn ngẫu nhiên đến cây gần nhất, với dung lượng mẫu n > 30 (số khoảng cách đo) dựa vào theo tiêu chuẩn U của Klark và Evans (Nguyễn Hải Tuất, 1990; Bảo Huy, 1993, 1997):

$$U = \frac{\bar{x}\sqrt{\lambda} - 0.5)\sqrt{n}}{0.26136} \quad (8.13)$$

trong đó: $\bar{x}$: Khoảng cách bình quân giữa các cây (Lấy tổng khoảng cách chia cho số lần đo là n).

$\lambda$: Số cây trên một $m^2$ diện tích đất rừng.

Kết quả tính U nếu:

$|U| \leq 1.96$ Cây rừng phân bố ngẫu nhiên trên mặt đất rừng;

$U > 1.96$ Cây rừng phân bố cách đều trên mặt đất rừng;

$U < -1.96$ Cây rừng phân bố cụm trên mặt đất rừng.

Ví dụ minh học đánh giá phân bố cây trên mặt đất rừng nửa rụng lá ưu thế bằng lăng có cấu trúc ổn định, ít bị tác động ở Tây Nguyên (Bảo Huy, 1993), với mẫu quan sát n = 285 (số khoảng cách đo) trên 5 ô điều tra 1ha và dùng tiêu chuẩn U của Klark và Evans để đánh giá phân bố trên bề mặt đất rừng. Kết quả tính toán theo tiêu chuẩn U như sau: Tham số: $\lambda = 0,095$ (Tổng số cây điều tra trong các ô/tổng diện tích các ô tiêu chuẩn); với: n = 285; cự ly bình quân giữa 285 cây quan sát ngẫu nhiên là $\bar{x} = 2,158$ m; giá trị U = 10.67.

*Kết luận:* Cây phân bố trên mặt bằng của lâm phần là phân bố cách đều.

Tuy nhiên, với đánh giá trên mới chỉ xem xét phân bố trên mặt bằng của tất cả các cây ở các thế hệ, cấp kính khác nhau; điều này chưa thể làm cơ sở để điều tiết cấu trúc mặt bằng. Cự ly tối ưu giữa các cây cần được thiết lập cho từng thế hệ hoặc cấp kính. Do vậy, trước khi đi vào xây dựng cấu trúc mặt bằng, cần thiết lập mô hình N/DBH định hướng (chuẩn, ổn định) để làm cơ sở xác định mật độ tối ưu cho từng cấp kính và chung lâm phần. Sau đó tìm khoảng cách tối ưu giữa các cây theo từng cấp kính.

Phương pháp tìm $\bar{x}$: Cự ly bình quân tối ưu để cây rừng có phân bố đồng đều trên mặt đất rừng chung và theo cấp kính (Bảo Huy, 1993): Từ công thức tính U, giả định rừng có phân bố đều thì U = 2, và rừng đạt mật độ chuẩn $N_{opt}$ (số cây tối ưu/ha tính được từ mô hình N/DBH định hướng đã xây dựng), suy ra cự ly $\bar{x}$ giữa cây rừng tối ưu:

$$\bar{x} = \frac{\frac{2 \times 0.26136}{\sqrt{N_{opt}}} + 0.5}{\sqrt{\lambda}} \quad (8.14)$$

Với $\lambda = N_{opt}/10.000$

Ví dụ đối với rừng nửa rụng lá bằng lăng, có $N_{opt} = 889$ cây/ha (số cây mẫu từ phân bố N/D đã xây dựng), suy ra: $\lambda = N_{opt} / 10.000 = 0.0889$, từ đây tính được cự ly bình quân tối ưu giữa 2 cây gần nhất trong lâm phần: $\bar{x} = 1.73$m (Bảo Huy, 1993).

Để đạt được hiệu quả cao hơn trong điều tiết cấu trúc mặt bằng, cần điều tiết cự ly này theo từng cỡ kính có nghĩa là, nếu trong một cỡ kính đã đạt được số cây tối ưu $N_{opti}$ nào đó theo cấu trúc N/DBH mẫu, nhưng phân bố chưa đều thì rừng cũng có năng suất thấp. Với lý do đó cần tiếp tục xác định cự ly bình quân tối ưu cho từng cỡ kính ( $\overline{x_i}$ ).

Phương pháp xác định như trên, nhưng chỉ khác là mật độ tối ưu ở đây được tính theo từng cỡ kính ($N_{opti}$) theo mô hình N/DBH định hướng, suy ra tham số $\lambda_i = N_{opti}/10.000$. Từ đây tính được cự

ly bình quân tối ưu ($\overline{x_i}$) cho từng cỡ kính i như ví dụ ở rừng nửa rụng lá bằng lăng (Bảo Huy, 1993).

**Bảng 8.9.** Mô hình cự ly bình quân tối ưu giữa các cây rừng theo cỡ kính

| Cấp kính DBH (cm) | $N_{opt}$ (c/ha) | λ | Cự ly bình quân tối ưu U $\bar{x}_i$ (m) | |
|---|---|---|---|---|
| 15 | 377 | 0.037667 | 2.72 | 2 |
| 20 | 208 | 0.020814 | 3.72 | 2 |
| 25 | 115 | 0.011502 | 5.12 | 2 |
| 30 | 64 | 0.006356 | 7.09 | 2 |
| 35 | 35 | 0.003512 | 9.93 | 2 |
| 40 | 19 | 0.001941 | 14.04 | 2 |
| 45 | 11 | 0.001072 | 20.14 | 2 |
| 50 | 6 | 0.000593 | 29.36 | 2 |
| 55 | 3 | 0.000327 | 43.59 | 2 |
| 60 | 2 | 0.000181 | 66.05 | 2 |
| 65 | 1 | 0.000100 | 102.27 | 2 |
| Tổng | 889 | | | |

*Nguồn: Bảo Huy (1993)*

Qua bảng trên cho thấy đối với cấp kính càng lớn thì cự ly tối ưu để rừng có phân bố đều càng lớn. Trong thực tế, để áp dụng cần đo khoảng cách từ một cây đến cây gần nhất nằm trong cùng một cấp kính, từ đó có thể có giải pháp lựa chọn trong tỉa thưa, nuôi dưỡng, khai thác chọn rừng, với định hướng đưa rừng về $N_{opti}$ và có cự ly bảo đảm cây rừng có phân bố đều, lợi dụng tốt nhất không gian dinh dưỡng ở tất cả các cấp kính, thế hệ.

## 8.2 Xác định mối quan hệ sinh thái loài trong rừng mưa nhiệt đới

Rừng hỗn loài nhiệt đới bao gồm nhiều loài cây cùng tồn tại, thời gian cùng tồn tại của một số loài trong đó phụ thuộc vào mức độ phù hợp hay đối kháng giữa chúng với nhau trong quá trình lợi dụng những yếu tố môi trường. Có thể phân ra làm 3 trường hợp:

*Liên kết dương:* Là trường hợp những loài cây có thể cùng tồn tại suốt quá trình sinh trưởng, giữa chúng không có sự cạnh tranh về ánh sáng, về các chất dinh dưỡng trong đất và không làm hại nhau thông qua các chất hoặc sinh vật trung gian khác.

*Liên kết âm:* Là trường hợp những loài cây không thể tồn tại lâu dài bên cạnh nhau được, do có những đối kháng quyết liệt trong quá trình lợi dụng các yếu tố môi trường (ánh sáng, chất dinh dưỡng trong đất, nước…), có khi loại trừ lẫn nhau thông qua nhiều yếu tố như: độc tố lá cây, các tinh dầu hoặc sinh vật trung gian..

*Quan hệ ngẫu nhiên:* Là trường hợp những loài cây tồn tại tương đối độc lập với nhau.

Việc nghiên cứu mối quan hệ giữa các loài là nhằm mục đích định hướng trong việc lựa chọn tổ thành loài cây hỗn giao trong phục hồi các hệ sinh thái rừng, trồng rừng hỗn giao, làm giàu rừng. Trong đó, theo hướng phục hồi tổ thành loài như tự nhiên với mối quan hệ ngẫu nhiên và hỗ trợ nhau (quan hệ dương).

Tuy nhiên, nghiên cứu đầy đủ mối quan hệ giữa các loài cây trong rừng tự nhiên là một vấn đề phức tạp, đòi hỏi căn cứ trên nhiều yếu tố sinh học và sinh thái. Trong thống kê sinh học, phương pháp dự báo được sử dụng để xác định mối quan hệ giữa các loài, làm cơ sở cho việc định hướng lựa chọn mô hình trồng rừng hỗn giao, điều chỉnh tổ thành trong công tác lâm sinh.

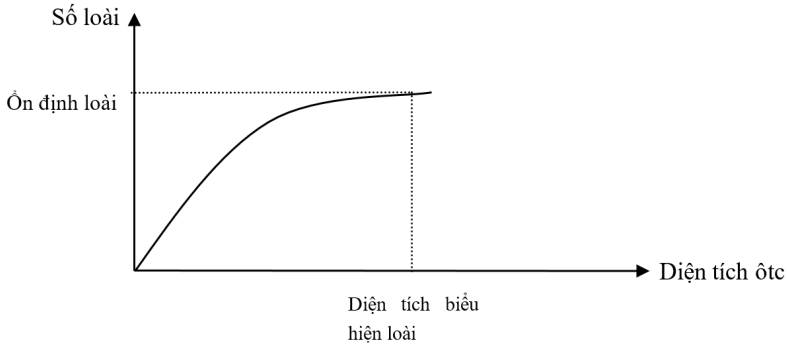
Phương pháp xác định mối quan hệ sinh thái loài gồm có ba bước chính (Bảo Huy, 1997): i) Xác định diện tích biểu hiện loài; ii) Xác định số ô mẫu cần thiết và thu thập dữ liệu loài và iii) Dự báo mối quan hệ giữa các loài

*i) Xác định diện tích biểu hiện loài:*

Để nghiên cứu mối quan hệ sinh thái giữa các loài, cần phải rút mẫu theo ô tiêu chuẩn (hoặc theo phương pháp K cây gần nhất) để tính toán xác suất xuất hiện các loài, vấn đề đặt ra là kích thước ô tiêu chuẩn bao nhiêu để bảo đảm đại diện, đó chính là xác định diện tích biểu hiện loài.

Trong trường hợp này là xác định một diện tích ô mẫu nhỏ nhất, nhưng bảo đảm xuất hiện các loài ưu thế sinh thái.

Tiến hành bằng cách thu thập số loài theo ô tiêu chuẩn diện tích thay đổi, diện tích ô bắt đầu là $100m^2$ và tăng dần đến khoảng 1 – 2ha. Khi diện tích ô tăng lên thì số loài tăng theo, đến một giới hạn diện tích ô nào đó thì số loài sẽ bão hòa. Diện tích đó gọi là diện tích biểu hiện loài. Có thể biểu thị việc xác định diện tích biểu hiện loài bằng đồ thị sau: (Bảo Huy, 1997; Jayaaman, 1999).




**Hình 8.8.** Quan hệ số loài theo diện tích ô mẫu

Để xác định diện tích biểu hiện loài, ở đó diện tích ô ứng với số loài tối đa trong kiểu rừng nghiên cứu, tiến hành mô phỏng quan hệ: $N_{số\ loài} = f(S = \text{diện tích ô mẫu})$, dạng quan hệ sau có thể được sử dụng:

$$N_{số\ loài} = a.e^{-b.S^{-m}} \quad (8.15)$$

Lim $N_{số\ loài} = a.e^{-b.S^{-m}}$ = a

Khi S → + ∞

Theo lý thuyết, khi diện tích ô mở ra vô hạn thì số loài sẽ đạt tới giới hạn là a loài. Ví dụ, đã điều tra thử nghiệm 53 ô có diện tích từ 100m$^2$ đến 10.000m$^2$ ở rừng khộp vùng Ea Soup, trên mỗi ô xác định số loài thuộc tầng cây gỗ (có đường kính ngang ngực lớn 10cm) xuất hiện. Tiến hành mô phỏng quy luật biến đổi số loài ($N_{số\ loài}$) theo diện tích ô (S) bằng một dạng hàm mũ cơ số e. Kết quả đã ước lượng các tham số:

$$N_{số\ loài} = 16.810.e^{-6.9.S^{-0.246}} \quad (8.16)$$

Với n = 53 R =0.907 Fr=722.58 α<0.01

Phương trình đạt hệ số tương quan cao, chứng tỏ có mối liên hệ chặt chẽ giữa $N_{số\ loài}$ và S, và dạng hàm này mô tả tốt chiều hướng biến thiên. Khảo sát hàm này cho thấy, khi tăng diện tích lên vô hạn thì số loài xuất hiện tiệm cận với giá trị của tham số a = 16.810. Như vậy, có nghĩa là đối với rừng khộp tại Ea Soup, số lượng loài thuộc tầng cây gỗ không nhiều, chỉ đạt đến 17 loài.

Tuy nhiên trong thực tế không thể lập ô có diện tích "vô hạn", vì vậy, trên cơ sở mô hình này, xác định số loài ưu thế sinh thái để thế vào mô hình và xác định được diện tích o mẫu biểu hiện loài.

Ví dụ như rừng khộp vùng Ea Soup, có các loài ưu thế sinh thái chính như Cà chắc (*Shorea obtusa*); Cẩm liên (*Shorea siamensis*), Dầu đồng (*Dipteocarpus tuberculatus*), Dầu trà beng (*Dipterocarpus obtusifolius*), Chiêu liêu (*Terminalia mycrocarpa*) và một số loài thuộc loài khác có tỷ lệ thấp hơn trong tổ thành. Như vậy số loài phổ biến trên một đơn vị diện tích rừng khộp chỉ khoảng 5-6 loài. Từ phương trình, thế giá trị $N_{số\ loài}$ = 6 vào suy được diện tích biểu hiện, đây cũng chính là diện tích cần có của một ô tiêu chuẩn trong rút mẫu điều tra nghiên cứu quan hệ sinh thái loài của khu rừng này. Diện tích biểu hiện trong trường hợp này là S = 2.500m$^2$ (50 × 50m).

*ii) Xác định số ô mẫu cần thiết và thu thập dữ liệu loài:*

Trên cơ sở đã xác định được diện tích ô biểu hiện loài ưu thế sinh thái; tiếp tục xác định dung lượng mẫu (số ô mẫu) theo công thức:

$$N_{ct} \geq \frac{t^2.V\%^2}{\Delta\%^2} \quad (8.17)$$

Trong đó: t = 1,96 khi độ tin cậy là 95% ; V%: Hệ số biến động về số loài, được tính theo công thức:

$$V\% = \frac{S}{\overline{X}} \times 100 \quad (8.18)$$

$$S = \sqrt{\frac{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)}{n-1}} \quad (8.19)$$

S: Sai tiêu chuẩn mẫu; n: Số ô rút mẫu thử (thường chọn n ≥ 30); x: Số loài trên mỗi ô

Δ%: Sai số cho phép từ 5% - 10%.

Sau khi xác định số lượng ô tiêu chuẩn cần thiết tiến hành xác định cự ly giữa các tuyến và cự ly giữa các ô trên tuyến để bảo đảm các ô mẫu được rải đều trên diện tích khảo sát. Tiến hành thu thập dữ liệu trên ô có diện tích biểu hiện loài ưu thế sinh thái, trong đó tập trung tấn số xuất hiện và xác định tên loài.

Từ số liệu quan sát, xác định số loài ưu thế để nghiên cứu mối quan hệ giữa chúng. Trên quan điểm sinh thái, loài ưu thế được chọn thường phải có chỉ số quan trọng (Important Value) IV% > 5% hoặc tần suất suất hiện loài F% > 5%.

$$IV\% = \frac{N\% + G\% + F\%}{3} \quad (8.20)$$

$$F\% = \frac{Số\ ô\ loài}{Số\ ô\ xuất\ hiện\ tất\ cả\ các\ loài} \quad (8.21)$$

Trong đó N% là % mật độ loài, G% là % tổng tiết diện ngang (BA) của loài và F% là % tần suất xuất hiện loài.

Ví dụ từ 32 ô tiêu chuẩn được rút mẫu ngẫu nhiên trong rừng thường xanh khu vực Đăk RLấp, Đăk Nông thống kê được tần suất xuất hiện của các loài ưu thế theo F% ở Bảng 8.10.

**Bảng 8.10.** Tần suất xuất hiện F% các loài ở rừng lá rộng thường xanh, Đăk Nông

| Stt | Loài | | Tần số số xuất hiện | Tần suất (F%) |
|---|---|---|---|---|
| 1 | Dẻ | *Lithocarpus sp* | 30 | 13.0 |
| 2 | Bằng lăng | *Lagerstroemia calyculata* | 27 | 11.7 |
| 3 | Xương cá | *Canthium didynum* | 23 | 10.0 |
| 4 | Xoan Mộc | *Toona sureni* | 19 | 8.2 |
| 5 | Bời lời | *Litsea glutinosa* | 18 | 7.8 |
| 6 | Bồ hòn | *Sapindus mukorossi* | 16 | 6.9 |
| 7 | Chò xót | *Schima superba* | 15 | 6.5 |
| 8 | Vạng trứng | *Endospermum chinnese* | 14 | 6.1 |
| 9 | Trâm | *Eugenia sp.* | 14 | 6.1 |
| 10 | Bứa | *Garcinia loureiri* | 11 | 4.8 |

| | | | | |
|---|---|---|---|---|
| 11 | Phay sừng | *Duabanga sonneratioides* | 8 | 3.5 |
| 12 | Cám | *Parinari anamense* | 6 | 2.6 |
| 13 | Dâu da đất | *Baccaurea sapida* | 6 | 2.6 |
| 14 | Thừng mực | *Wrightia annamensis* | 6 | 2.6 |
| 15 | Máu chó | *Knema conferta* | 4 | 1.7 |
| 16 | Chua khét | *Dysoxylum acutangulum* | 4 | 1.7 |
| 17 | Trám | *Canarium copaliferum* | 3 | 1.3 |
| 18 | Gạo | *Gossampinus malabaria* | 2 | 0.9 |
| 19 | Sầu đâu | *Azadirachta indica* | 2 | 0.9 |
| 20 | Chò chỉ | *Parashorea chinensis* | 2 | 0.9 |
| 21 | Gòn | *Bombax anceps* | 1 | 0.4 |

Từ bảng trên cho thấy có 9 loài có tần suất F% > 5%. Trong rừng hỗn loài, các loài có tần suất > 5% được xem là loài đóng vai trò quan trọng trong hình thành sinh thái rừng, do đó, chọn 9 loài này để xem xét quan hệ giữa chúng với nhau.

*iii) Dự báo mối quan hệ sinh thái giữa các loài:*

Từ số ô mẫu cần thiết có diện tích biểu hiện loài đã được rút mẫu ngẫu nhiên, thu thập số liệu xuất hiện loài; tiến hành kiểm tra quan hệ cho từng cặp loài theo tiêu chuẩn ρ và $\aleph^2$.

Sử dụng các tiêu chuẩn thống kê sau để đánh giá quan hệ theo từng cặp loài:

ρ: Hệ số tương quan giữa 2 loài A và B:

$$\rho = \frac{P(AB) - P(A).P(B)}{\sqrt{P(A).(1-P(A)).P(B).(1-P(B))}} \quad (8.22)$$

Trong đó khi:

ρ = 0 : 2 loài A và B độc lập nhau.

0 < ρ ≤ 1: loài A và B liên kết dương.

-1 ≤ ρ < 0: loài A và B liên kết âm (bài xích nhau).

Xác suất xuất hiện loài:

P(AB): Xác suất xuất hiện đồng thời của 2 loài A và B

P(A): Xác suất xuất hiện loài A.

P(B): Xác suất xuất hiện loài B.

$$P(AB) = \frac{nAB}{n} \quad P(A) = \frac{nA + nAB}{n} \quad P(B) = \frac{nB + nAB}{n}$$

Với: nA: Số ô tiêu chuẩn chỉ xuất hiện loài A.

nB: Số ô tiêu chuẩn chỉ xuất hiện loài B.

nAB: Số ô tiêu chuẩn xuất hiện đồng thời 2 loài A và B.

n: Tổng số ô quan sát ngẫu nhiên.

ρ nói lên chiều hướng liên hệ và mức độ liên hệ giữa 2 loài. ρ < 0: 2 loài liên kết âm và |ρ| càng lớn thì mức độ bài xích nhau càng mạnh, ngược lại ρ > 0: 2 loài liên kết dương và |ρ| càng lớn thì mức độ hỗ trợ nhau càng cao.

Trong trường hợp |ρ| xấp xỉ = 0, thì chưa thể biết giữa 2 loài có thực sự quan hệ với nhau hay không? Lúc này cần sử dụng thêm phương pháp kiểm tra tính độc lập bằng mẫu biểu 2x2 theo tiêu chuẩn $\chi^2$.

Công thức kiểm tra mối quan hệ giữa 2 loài A và B được thực hiện bằng tiêu chuẩn $\aleph^2$:

$$\aleph^2 = \frac{(|ad - bc| - 0.5).^2 n}{(a+b).(c+d).(a+c).(b+d)} \quad (8.23)$$

Trong đó: a = nAB; b = nB; c = nA; d: số ô không chứa cả 2 loài a và B.

$\aleph^2$t tính được ở công thức trên được so sánh với $\aleph^2_{0.05}$ ứng với bậc tự do K=1 là $\aleph^2_{0.05,\ K=1}$ = 3.84.

Nếu $\aleph^2 t \leq \aleph^2_{0.05} = 3.84$ thì mối quan hệ giữa 2 loài là ngẫu nhiên.

Nếu $\aleph^2 t > \aleph^2_{0.05} = 3.84$ thì giữa 2 loài có quan hệ với nhau và nếu ρ < 0 thì quan hệ âm, ρ > thì quan hệ dương.

Tóm lại để xem xét mối quan hệ theo từng cặp loài, sử dụng đồng thời 2 tiêu chuẩn ρ và $\aleph^2$:

$\aleph^2$: Để kiểm tra mối quan hệ từng cặp loài.

ρ: Trong trường hợp kiểm tra bằng $\aleph^2$ cho thấy có quan hệ, thì ρ sẽ cho biết chiều hướng mối quan hệ đó theo dấu của ρ (- hay +) và mức độ quan hệ qua giá trị |ρ|.

Bảng 8.11 giới thiệu một kết quả kiểm tra quan hệ sinh thái giữa các cặp loài ưu thế sinh thái (F% > 5%), trong đó tập trung cho loài nghiên cứu là Xoan mộc, ở rừng lá rộng thường xanh vùng Tây Nguyên (Bảo Huy, 1997).

**Bảng 8.11.** Kiểm tra quan hệ theo từng cặp loài ưu thế sinh thái rừng lá rộng thường xanh Đăk R'lấp, Đăk Nông

| Stt | Loài A | Loài B | nA(c) | nB(b) | nAB(a) | nAB-(d) | P(A) | P(B) | P(AB) | ρ | $\aleph^2$ | Quan hệ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Xoan Mộc | Bằng Lăng | 5 | 13 | 14 | 0 | 0.594 | 0.844 | 0.438 | -0.356 | 3.99 | Quan hệ âm |
| 2 | Xoan Mộc | Dẻ | 0 | 11 | 19 | 2 | 0.594 | 0.938 | 0.594 | 0.312 | 3.04 | Ngẫu nhiên |
| 3 | Xoan Mộc | Bời Lời | 7 | 6 | 12 | 7 | 0.594 | 0.563 | 0.375 | 0.168 | 0.89 | Ngẫu nhiên |
| 4 | Xoan Mộc | Vạng Trứng | 10 | 5 | 9 | 8 | 0.594 | 0.438 | 0.281 | 0.088 | 0.24 | Ngẫu nhiên |
| 5 | Xoan Mộc | Trâm | 10 | 5 | 9 | 8 | 0.594 | 0.438 | 0.281 | 0.088 | 0.24 | Ngẫu nhiên |

| Stt | Loài A | Loài B | nA(c) | nB(b) | nAB(a) | nAB-(d) | P(A) | P(B) | P(AB) | ρ | $\aleph^2$ | Quan hệ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | Xoan Mộc | Xương cá | 5 | 9 | 14 | 4 | 0.594 | 0.719 | 0.438 | 0.049 | 0.07 | Ngẫu nhiên |
| 7 | Xoan Mộc | Bồ hòn | 10 | 7 | 9 | 6 | 0.594 | 0.500 | 0.281 | -0.064 | 0.12 | Ngẫu nhiên |
| 8 | Xoan Mộc | Chò xót | 12 | 8 | 7 | 5 | 0.594 | 0.469 | 0.219 | -0.243 | 1.86 | Ngẫu nhiên |
| 9 | Bằng Lăng | Dẻ | 2 | 5 | 25 | 0 | 0.844 | 0.938 | 0.781 | -0.111 | 0.36 | Ngẫu nhiên |
| 10 | Bằng Lăng | Bời Lời | 13 | 4 | 14 | 2 | 0.844 | 0.563 | 0.438 | -0.206 | 0.40 | Ngẫu nhiên |
| 11 | Bằng Lăng | Vạng Trứng | 16 | 3 | 11 | 2 | 0.844 | 0.438 | 0.344 | -0.141 | 0.61 | Ngẫu nhiên |
| 12 | Bằng Lăng | Trâm | 14 | 1 | 13 | 4 | 0.844 | 0.438 | 0.406 | 0.206 | 1.32 | Ngẫu nhiên |
| 13 | Bằng Lăng | Xương cá | 9 | 5 | 18 | 0 | 0.844 | 0.719 | 0.563 | -0.269 | 2.27 | Ngẫu nhiên |
| 14 | Bằng Lăng | Bồ hòn | 13 | 2 | 14 | 3 | 0.844 | 0.500 | 0.438 | 0.086 | 0.22 | Ngẫu nhiên |
| 15 | Bằng Lăng | Chò xót | 13 | 1 | 14 | 4 | 0.844 | 0.469 | 0.438 | 0.232 | 1.68 | Ngẫu nhiên |
| 16 | Dẻ | Bời Lời | 14 | 2 | 16 | 0 | 0.938 | 0.563 | 0.500 | -0.228 | 1.60 | Ngẫu nhiên |
| 17 | Dẻ | Vạng Trứng | 18 | 2 | 12 | 0 | 0.938 | 0.438 | 0.375 | -0.293 | 2.67 | Ngẫu nhiên |
| 18 | Dẻ | Trâm | 17 | 1 | 13 | 1 | 0.938 | 0.438 | 0.406 | -0.033 | 0.03 | Ngẫu nhiên |
| 19 | Dẻ | Xương cá | 7 | 0 | 23 | 2 | 0.938 | 0.719 | 0.719 | 0.413 | 5.33 | Quan hệ dương |
| 20 | Dẻ | Bồ hòn | 14 | 0 | 16 | 2 | 0.938 | 0.500 | 0.500 | 0.258 | 2.07 | Ngẫu nhiên |
| 21 | Dẻ | Chò xót | 16 | 1 | 14 | 1 | 0.938 | 0.469 | 0.438 | -0.016 | 0.00 | Ngẫu nhiên |
| 22 | Bời lời | Vạng Trứng | 11 | 7 | 7 | 7 | 0.563 | 0.438 | 0.219 | -0.111 | 0.38 | Ngẫu nhiên |
| 23 | Bời lời | Trâm | 7 | 3 | 11 | 11 | 0.563 | 0.438 | 0.344 | 0.397 | 4.99 | Quan hệ dương |
| 24 | Bời lời | Xương cá | 5 | 10 | 13 | 4 | 0.563 | 0.719 | 0.406 | 0.009 | 0.00 | Ngẫu nhiên |
| 25 | Bời lời | Bồ hòn | 11 | 9 | 7 | 5 | 0.563 | 0.500 | 0.219 | -0.252 | 2.00 | Ngẫu nhiên |
| 26 | Bời lời | Chò xót | 13 | 10 | 5 | 4 | 0.563 | 0.469 | 0.156 | -0.434 | 5.97 | Quan hệ âm |
| 27 | Vạng trứng | Trâm | 9 | 9 | 5 | 9 | 0.438 | 0.438 | 0.156 | -0.143 | 0.64 | Ngẫu nhiên |
| 28 | Vạng trứng | Xương cá | 5 | 14 | 9 | 4 | 0.438 | 0.719 | 0.281 | -0.149 | 0.69 | Ngẫu nhiên |
| 29 | Vạng trứng | Bồ hòn | 5 | 7 | 9 | 11 | 0.438 | 0.500 | 0.281 | 0.252 | 2.00 | Ngẫu nhiên |
| 30 | Vạng trứng | Chò xót | 7 | 8 | 7 | 10 | 0.438 | 0.469 | 0.219 | 0.055 | 0.09 | Ngẫu nhiên |
| 31 | Trâm | Xương cá | 3 | 12 | 11 | 6 | 0.438 | 0.719 | 0.344 | 0.131 | 0.53 | Ngẫu nhiên |
| 32 | Trâm | Bồ hòn | 6 | 8 | 8 | 10 | 0.438 | 0.500 | 0.250 | 0.126 | 0.49 | Ngẫu nhiên |
| 33 | Trâm | Chò xót | 11 | 12 | 3 | 6 | 0.438 | 0.469 | 0.094 | -0.450 | 6.42 | Quan hệ âm |
| 34 | Xương cá | Bồ hòn | 9 | 2 | 14 | 7 | 0.719 | 0.500 | 0.438 | 0.348 | 3.82 | Ngẫu nhiên |
| 35 | Xương cá | Chò xót | 16 | 8 | 7 | 1 | 0.719 | 0.469 | 0.219 | -0.527 | 8.80 | Quan hệ âm |
| 36 | Bồ hòn | Chò xót | 9 | 8 | 7 | 8 | 0.500 | 0.469 | 0.219 | -0.063 | 0.12 | Ngẫu nhiên |

*Nguồn: Bảo Huy (1997)*

Từ kết quả này có thể xác định được:

*Các loài có quan hệ dương:* $\aleph^2 t > \aleph^2_{0.05} = 3.84$ và $\rho > 0$: Các loài này nên được lựa chọn để trồng hỗn giao, hoặc làm giàu rừng.

*Các loài có quan hệ âm*: $\aleph^2 t > \aleph^2_{0.05} = 3.84$ và $\rho < 0$: Các loài này không nên được lựa chọn để trồng hỗn giao, hoặc làm giàu rừng; và cần loài trừ bớt sự cạnh tranh giữa chúng.

*Các loài có quan hệ ngẫu nhiên*: $\aleph^2 t \leq \aleph^2_{0.05} = 3.84$: Các loài này có thể tồn tại khá độc lập, do vậy lựa chọn chúng hỗn giao hay loại trừ cũng không ảnh hưởng đến quan hệ sinh thái loài.

## 8.3 Mô hình quan hệ với các nhân tố định tính

Trong thực tế thiết lập các mô hình quan hệ, các biến độc lập ảnh hưởng đến biến phụ thuộc không phải bao giờ cũng có giá trị định lượng (số) để lập mô hình toán, mà còn có các nhân tố định tính như đơn vị đất, đá mẹ, vị trí địa hình, vùng sinh thái,… Ví dụ nghiên cứu quan hệ sinh trưởng cây rừng với tuổi và các nhân tố sinh thái như loại đất, điều kiện khí hậu, hoặc mô hình ước tính sinh khối cây rừng theo các nhân tố điều tra cây và yếu tố sinh thái, môi trường rừng. Các nhân tố định tính này cần được mã hóa một cách thích hợp để xem xét ảnh hưởng của nó đến biến phụ thuộc và đưa vào mô hình để tăng độ cậy.

Mô hình hồi quy đa biến dạng tuyến tính hoặc tuyến tính hóa hoặc phi tuyến, có trọng số như đã trình bày ở các phần trên sẽ là một công cụ mạnh giúp cho việc phát hiện các nhân tố định lượng và định tính ảnh hưởng đến biến số y khảo sát.

Đối với dạng mô hình cần mã hóa biến định tính, các bước tiến hành như sau:

i) Thiết kế thí nghiệm hoặc thu thập dữ liệu về biến số phụ thuộc y và cùng với nó là các nhân tố $x_i$ dự kiến có ảnh hưởng (có thể định tính hay định lượng).

ii) Mã hóa các biến định tính theo một phương pháp thích hợp.

iii) Xác định biến số $x_i$ có ảnh hưởng đến y ở mức độ tin cậy 95%.

iv) Thử nghiệm các mô hình tuyến tính hoặc phi tuyến nhiều lớp biểu diễn mối quan hệ $y = f(x_i)$.

v) Phân tích kết quả mô hình hồi quy đa biến để đánh giá chiều hướng tác động của các biến số đã được mã hóa đến biến phụ thuộc để đưa ra giải pháp.

Minh họa cho cho nội dung này là mô hình hóa để xác định các nhân tố định tính và định lượng như: sinh thái, lập địa, trạng thái rừng ảnh hưởng đến sinh trưởng, tăng trưởng và mức thích nghi của cây tếch được trồng làm giàu rừng khộp ở tỉnh Đăk Lăk (Bảo Huy, 2014).

*Bước 1*: Thiết kế thí nghiệm và thu thập số liệu: Bố trí 64 ô thí nghiệm trên nhiều tổ hợp sinh thái khác nhau của rừng khộp. Cây tếch ở các 64 ô thí nghiệm sau khi trồng trên 4 năm được thu thập số liệu sinh trưởng, tăng trưởng tếch và các nhân tố sinh thái trên có ô thử nghiệm như đá mẹ, đơn vị đất, tầng dày đất, đá nổi, kết von, độ tàn che, mật độ cây rừng, ngập nước,... vị trí, địa hình, độ dốc,... lý hóa tính đất. Mỗi ô xác định mức thích nghi (biến y) (Dữ liệu 14).

*Bước 2:* Mã hóa biến định tính: Các nhân tố định tính như đá mẹ, đơn vị đất, ngập nước... Cần được mã hóa để tạo thành biến số định lượng.

Có hai phương án mã hóa. Cách thức mã hóa khác nhau sẽ dẫn đến việc lựa chọn mô hình hồi quy có mức độ phức tạp khác nhau (Bảng 8.12):

*Mã hóa hệ thống (cơ giới):* Các nhân tố định tính được mã hóa hệ thống (cơ giới): 1, 2, 3,... Ví dụ mã hóa nhân tố vị trí địa hình theo thứ tự: Bằng = 1; chân = 2; sườn = 3 và đỉnh = 4.

*Mã hóa theo chiều biến thiên:* Các nhân tố định tính được mã hóa theo chiều biến thiên của nhân tố phụ thuộc. Sắp xếp nhân tố định tính theo chiều biến thiên của biến phụ thuộc (tăng hoặc giảm), sau đó các nhân tố định tính được mã hóa theo cùng một vector như vậy.

**Bảng 8.12.** Hai phương án mã hóa biến định tính khác nhau và việc chọn lựa mô hình hồi quy khác nhau

| Kiểu dạng hàm mô phỏng | Phương pháp mã hóa biến định tính | |
|---|---|---|
| | Hệ thống (Cơ giới)<br>(Mã hóa đơn giản) | Theo chiều biến thiên, vector của biến phụ thuộc (Mã hóa phức tạp) |
| Tuyến tính hoặc phi tuyến nhưng biến phụ thuộc theo một chiều (tăng hoặc giảm) (Xây dựng hàm đơn giản) | Không thực hiện được hoặc sai quy luật | Thực hiện được |
| Phi tuyến với biến phụ thuộc dạng tăng giảm phức tạp, (Xây dựng hàm phức tạp) | Thực hiện được | Thực hiện được nhưng không cần thiết |

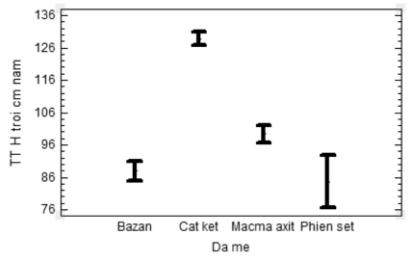
Trong minh họa này, do sự ảnh hưởng của các biến sinh thái, lập địa, lâm phần, đất đai đến sinh trưởng và tăng trưởng tếch trong rừng khộp, có nghĩa là tăng trưởng tếch không theo một chiều của sự thay đổi các nhân tố sinh thái, định tính. Vì vậy, phương án được lựa chọn là mã hóa các nhân tố sinh thái theo chiều biến thiên (cũng chiều) với sinh trưởng tếch, lúc này mô hình có biến y sẽ quan hệ một chiều với các biến $x_i$.

Để mã hóa theo chiều biến thiên (thuận hay nghịch) của mức thích nghi cây tếch theo nhân tố định tính như đơn vị đất, đá mẹ; hoặc biến thiên của mức thích nghi không theo cùng chiều với các cấp của nhân tố ví dụ như kết von < 30, 30-50, 50-70 và > 70%, mức thích nghi sẽ phù hợp với một cấp nào đó rồi giảm xuống. Do vậy, nếu không mã hóa theo chiều biến thiên của mức thích nghi thì có thể xác định sai hoặc kém tin cậy sự ảnh hưởng của các nhân tố sinh thái định tính và phân cấp.

Sử dụng tăng trưởng chiều cao cây tếch trội (TT H troi) để xem xét chiều biến thiên theo các nhân tố, cấp; vì tăng trưởng cây trội là chỉ tiêu phản ảnh mức thích nghi của tếch, từ đó mã hóa theo chiều biến thiên tăng trưởng cây trội. Thực hiện trong Statgraphics theo tiêu chuẩn Kruskal Wallis để kiểm tra có sự ảnh hưởng của nhân tố đó với tăng trưởng cây trội tếch, sau đó sử dụng trắc nghiệm Duncan để xem các công thức, cấp, yếu tố nào là đồng nhất hoặc khác biệt để gộp nhóm và mã hóa theo chiều biến thiên.

Ví dụ đối với nhân tố đá mẹ bao gồm 4 loại: Bazan, Cát kết, Macma axit và Phiến sét. Kết quả kiểm tra ở Bảng 8.13 cho thấy theo Kruskal-Wallis Test thì P < 0,05, có nghĩa đá mẹ khác nhau ảnh hưởng đến tăng trưởng tếch; và Duncan cho thấy đá phiến sét và bazan là cùng một nhóm và có tăng trưởng tếch thấp nhất, tiếp theo là bazan, cát kết ảnh hưởng độc lập và cát kết cho tăng trưởng cao nhất.

**Bảng 8.13.** Kết quả kiểm tra ảnh hưởng của nhân tố đá mẹ đến tăng trưởng cây tếch trội




TT H troi theo loại đá mẹ

**Kruskal-Wallis Test for TT H troi cm nam by Da me**

| *Da me* | *Sample Size* | *Average Rank* |
|---|---|---|
| Bazan | 427 | 685.779 |
| Cat ket | 835 | 1111.5 |
| Macma axit | 489 | 776.475 |
| Phien set | 58 | 629.603 |

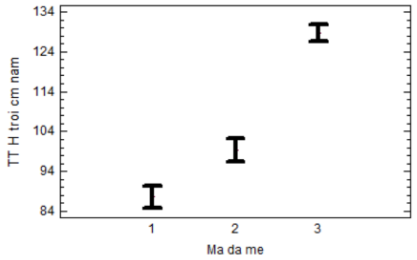
Test statistic = 251.429 P-Value = 0.0

**Method: 95.0 percent Duncan**

| *Da me* | *Count* | *Mean* | *Homogeneous Groups* |
|---|---|---|---|
| Phien set | 58 | 84.7038 | X |
| Bazan | 427 | 87.983 | X |
| Macma axit | 489 | 99.3378 | X |
| Cat ket | 835 | 128.794 | X |

Mã hóa theo chiều biến thiên (thuận hoặc nghịch) của tăng trưởng tếch và gộp nhóm đồng nhất: Phiến sét và Bazan = 1 (vì Phiến sét và Bazan không có ảnh hưởng khác biệt), Macma Axit = 2 và Cát Kết = 3. Sau mã hóa nhân tố đá mẹ, tăng trưởng cây trội tếch biến thiên cùng chiều với giá trị mã hóa và có sự sai khác rõ rệt giữa các yếu tố đã được mã hóa qua kiểm tra lại theo Kruskal-Wallis và Duncan ở độ tin cậy 95%, kết quả ở Bảng 8.14.

**Bảng 8.14.** Kết quả kiểm tra ảnh hưởng của mã hóa nhân tố đá mẹ đến tăng trưởng cây tếch trội




TT H troi theo mã loại đá mẹ

**Kruskal-Wallis Test for TT H troi cm nam by Ma da me**

| *Ma da me* | *Sample Size* | *Average Rank* |
|---|---|---|
| 1 | 485 | 679.061 |
| 2 | 489 | 776.475 |
| 3 | 835 | 1111.5 |

Test statistic = 250.838 P-Value = 0.0

**Method: 95.0 percent Duncan**

| *Ma da me* | *Count* | *Mean* | *Homogeneous Groups* |
|---|---|---|---|
| 1 | 485 | 87.5908 | X |
| 2 | 489 | 99.3378 | X |
| 3 | 835 | 128.794 | X |

Tương tự như vậy, mã hóa cho toàn bộ các nhân tố, cấp nhân tố sinh thái, lập địa, trạng thái rừng khộp theo chiều biến thiên của tăng trưởng tếch ở Bảng 8.15.

**Bảng 8.15.** Mã hóa các nhân tố sinh thái, lập địa, trạng thái rừng theo chiều biến thiên của tăng trưởng cây tầng trội tếch

| Stt | Nhân tố | Mã hóa | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| 1. | Độ cao so với mặt biển (m) | 300 -400 | 100-200 | 200-300 | | | |
| 2. | Vị trí địa hình | Khe | Bằng | Sườn + Đỉnh | | | |
| 3. | Cấp độ dốc (độ) | <3 | 15-20 | 3-8 | 8-15 | | |
| 4. | Đá mẹ | Phiến sét + Bazan | Macma Axit | Cát kết | | | |
| 5. | Đơn vị đất | Dat phu sa co gioi nhe, | Dat do chua, rat | Dat den co gioi | Dat xam tang rat | Dat xam soi san | Dat nau tang |

| Stt | Nhân tố | Mã hóa | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| | | dong nuoc<br>Dat xoi mon manh, tro soi san | ngheo kiem<br>Dat xam co gioi nhe<br>Dat co tang set chat, co tang ket von<br>Dat nau co gioi nhe<br>Dat den tang mong | nhe, soi san sau<br>Dat co tang set chat, nhan tac, it chua | mong<br>Dat xam tang mong | nong | mong |
| 6. | Cấp dày đất | < 30cm, > 50cm | 30 – 50cm | | | | |
| 7. | Ngập nước | Không | Có | | | | |
| 8. | Cấp đá nổi | 30 – 50% và > 50% | < 10% | 10-30% | | | |
| 9. | Cấp kết von | < 10% | 10-30% | > 50% | 30-50% | | |
| 10. | Cấp đá lẫn | < 10%, 10-30% và 30-50% | > 70% | 50-70% | | | |
| 11. | Sổ đất và mộc hoa | Có | Không | | | | |
| 12. | Cỏ lào | Không | Có | | | | |
| 13. | Cấp độ tàn che | 10 – 30% | <10% và > 50% | 30-50% | | | |
| 14. | Loài cây ưu thế rừng khộp | Dầu trà beng | Cà chít | Dầu đồng, Chiêu liêu đen, Căm xe | Cẩm liên | | |
| 15. | Cấp mật độ /ha rừng khộp | > 500 cây | < 100 cây và 100-300 cây | 300-500 cây | | | |
| 16. | Cấp BA /ha Prodan, Bi | > 15$m^2$ | < 5$m^2$ | 5-10$m^2$ | 10-15$m^2$ | | |
| 17. | Cấp trữ lượng M/ha | 100-150 $m^3$ | < 50 $m^3$ và > 150 $m^3$ | 50-100 $m^3$ | | | |

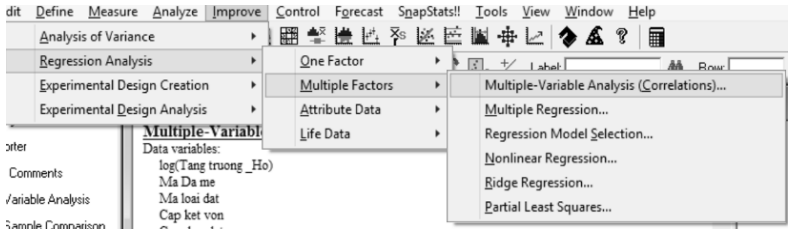
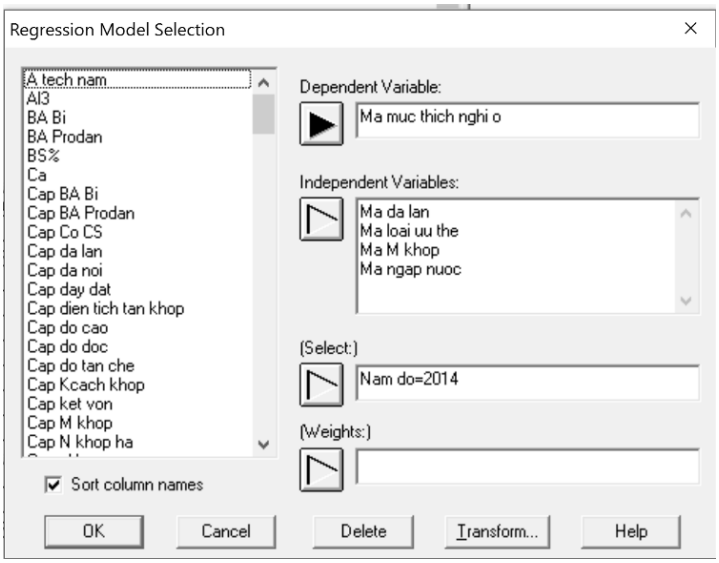
| Stt | Nhân tố | Mã hóa | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| 18. | Cấp tổng diện tích tán | > 15.000m$^2$, 10.000-15.000m$^2$ và < 1.000m$^2$ | 1.000-5.000m$^2$ và 5.000-10.000m$^2$ | | | | |
| 19. | Mức thích nghi tếch | Rất tốt | Tốt | Trung bình | Kém | | |

*Nguồn: Bảo Huy (2014)*

*Bước 3*: Xác định các biến số $x_i$ có ảnh hưởng đến y: Kết quả phân tích này cũng chỉ ra được các biến số có quan hệ với nhau và ảnh hưởng đến y.

Biến y trong minh họa này là "Mức thích nghi của tếch" trong rừng khộp và các biến $x_i$ thăm dò ảnh hưởng đến y là các nhân tố dễ quan sát trên hiện trường để xác định vùng thích nghi để trồng tếch: Các nhân tố định tính, định lượng được mã hóa: "Ngập nước", "Tỷ lệ đá lẫn", "Loài cây rừng khộp ưu thế" và "Cấp trữ lượng rừng khộp".

Sử dụng chức năng xác định các nhân tố ảnh hưởng trong Statgraphics trên cơ sở Dữ liệu 14 như sau:

| | |
|---|---|
| Phân tích mối quan hệ giữa các biến số trong Stat: Improve/Regression Analysis/Mutiple Factors/Multiple-Variable Analysis |  |
| Trong hộp thoại lựa chọn mô hình (Regression Model Selection), nhập biến y Dependent Variable và các $x_i$ Independent Variables thăm dò. |  |

Kết quả xác định các nhân tố định tính và định lượng ảnh hưởng đến mức thích nghi của tếch trong làm giàu rừng khộp trong Statgraphics:

**Regression Model Selection - Ma muc thich nghi o (Nam do=2014)**

Dependent variable: Ma muc thich nghi o

Independent variables:

A=Ma da lan
B=Ma loai uu the
C=Ma M khop
D=Ma ngap nuoc

Selection variable: Nam do=2014

Number of complete cases: 64

Number of models fit: 16

**Models with Largest Adjusted R-Squared**

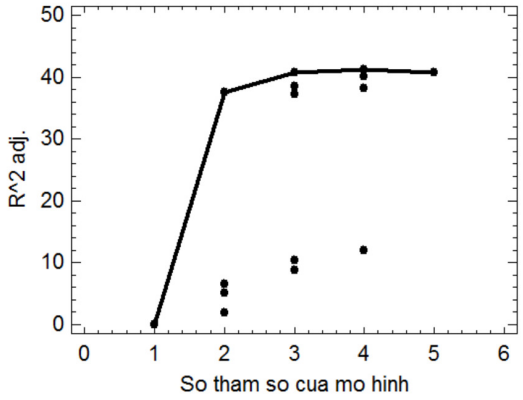
| | | *Adjusted* | | *Included* |
|---|---|---|---|---|
| *MSE* | *R-Squared* | *R-Squared* | *Cp* | *Variables* |
| 0.456631 | 44.0861 | 41.2904 | 3.50413 | ABD |
| 0.460436 | 44.5598 | 40.8011 | 5.0 | ABCD |
| 0.460833 | 42.6309 | 40.75 | 3.05268 | AD |
| 0.465473 | 43.0033 | 40.1534 | 4.65643 | ACD |
| 0.478036 | 40.4894 | 38.5383 | 5.3317 | AB |
| 0.480323 | 41.1849 | 38.2441 | 6.59158 | ABC |
| 0.485409 | 38.5809 | 37.5903 | 5.36278 | A |
| 0.488796 | 39.1499 | 37.1548 | 6.75724 | AC |
| 0.684842 | 16.1418 | 11.9488 | 33.2427 | BCD |
| 0.697021 | 13.228 | 10.383 | 34.3435 | BD |
| 0.709923 | 11.6218 | 8.72412 | 36.0529 | BC |
| 0.727154 | 7.99271 | 6.50873 | 37.915 | B |
| 0.737634 | 6.66667 | 5.16129 | 39.3262 | D |
| 0.762526 | 3.51715 | 1.96098 | 42.6779 | C |
| 0.777778 | 0.0 | 0.0 | 44.4209 | |

**The StatAdvisor**

This table shows the models which give the largest adjusted R-Squared values. The adjusted R-Squared statistic measures the proportion of the variability in Ma muc thich nghi o which is explained by the model. Larger values of adjusted R-Squared correspond to smaller values of the mean squared error (MSE). Up to 5 models in each subset of between 0 and 4 variables are shown. The best model contains 3 variables, Ma da lan, Ma loai uu the, and Ma ngap nuoc.

Kết quả trên cho thấy, ba biến ảnh hưởng rõ rệt đến mức thích nghi tếch là ABD: "Ma da lan", "Ma loai uu the", "Ma ngap nuoc", với $R^2_{adj.}$ cao nhất và chỉ tiêu Mallow's Cp (1973) = 3.5 gần bằng 4 (3 biến số ABD và một tham số là hằng số). Hình 8.9 minh họa cho $R^2_{adj.}$ đạt cực đại khi số

tham số của mô hình (số biến $x_i$ + 1) bằng 4. Vì vậy sử dụng ba biến số định tính đã được mã hóa theo chiều biến thiên với mức thích nghi tếch để lập mô hình dự báo theo các nhân tố dễ quan sát trên hiện trường.




**Hình 8.9.** Quan hệ $R^2_{adj.}$ với số tham số của quan hệ mức thích nghi tếch với các nhân tố sinh thái, trạng thái rừng khộp

*Bước 4*: Thử nghiệm và lựa chọn mô hình tối ưu quan hệ giữa y và $x_i$: Trên cơ sở xác định được các biến $x_i$ ảnh hưởng đến y, tiến hành lập mô hình quan hệ $y = f(x_i)$ theo các dạng mô hình và phương pháp khác nhau.

Trong ví dụ này đã thử nghiệm nhiều dạng hàm và kết quả cho thấy mô hình phi tuyến power là tốt nhất và ước lượng theo phương pháp phi tuyến tính của Marquardt có trọng số, thực hiện trong phần mềm Statgraphics với Dữ liệu 14 như sau:

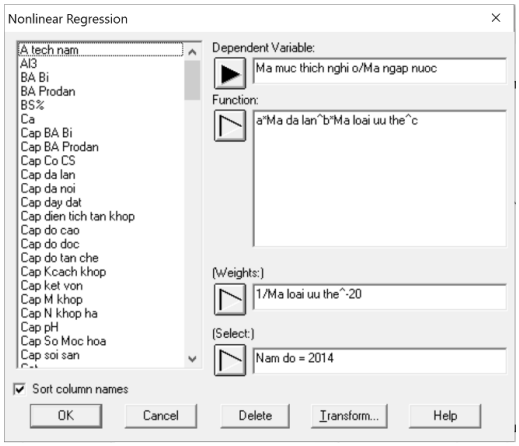
Trong hộp thoại Mô hình phi tuyến (Nonmlinear Regression):

Biến phụ thuộc (Dependenr Variable): Tổ hợp biến: Ma muc thich nghi o/Ma ngap nuoc

Function: Nhập hàm mũ:

a*Ma da lan^b*Ma loai uu the^c

Trọng số (Weight): 1/Ma loai uu the^-20




Kết quả thiết lập mô hình phi tuyến có trọng số theo Marquardt trong Statgraphics quan hệ mức thích nghi tếch với các nhân tố sinh thái định tính:

**Nonlinear Regression - Ma muc thich nghi o/Ma ngap nuoc (Nam do = 2014)**

Dependent variable: Ma muc thich nghi o/Ma ngap nuoc

Independent variables:

Madalan

Maloaiuuthe

Weight variable: 1/Ma loai uu the^-20

Selection variable: Nam do = 2014
Function to be estimated: a*Ma da lan^b*Ma loai uu the^c
Initial parameter estimates:
a = 5.0
b = 0.1
c = 0.1

Estimation method: Marquardt
Estimation stopped due to convergence of residual sum of squares.
Number of iterations: 5
Number of function calls: 22

**Estimation Results**

| | | | *Asymptotic* | *95.0%* |
|---|---|---|---|---|
| | | *Asymptotic* | *Confidence* | *Interval* |
| *Parameter* | *Estimate* | *Standard Error* | *Lower* | *Upper* |
| a | 5.52415 | 4.89944 | -4.27291 | 15.3212 |
| b | -0.626903 | 0.0942846 | -0.815437 | -0.438368 |
| c | -0.440852 | 0.64385 | -1.72831 | 0.846608 |

**Analysis of Variance**

| *Source* | *Sum of Squares* | *Df* | *Mean Square* |
|---|---|---|---|
| Model | 3.60495E13 | 3 | 1.20165E13 |
| Residual | 2.85081E12 | 61 | 4.67346E10 |
| Total | 3.89003E13 | 64 | |
| Total (Corr.) | 5.91703E12 | 63 | |

R-Squared = 51.8202 percent
R-Squared (adjusted for d.f.) = 50.2405 percent
Standard Error of Est. = 216182.
Mean absolute error = 0.603025
Durbin-Watson statistic = 1.81439
Lag 1 residual autocorrelation = 0.0531014

**Residual Analysis**

| | *Estimation* | *Validation* |
|---|---|---|
| n | 64 | 64 |
| MSE | 4.67346E10 | 0.505104 |
| MAE | 0.603025 | 0.603025 |
| MAPE | 29.9875 | 29.9875 |
| ME | -0.000174705 | -0.000174705 |
| MPE | -10.064 | -10.064 |

**The StatAdvisor**
The output shows the results of fitting a nonlinear regression model to describe the relationship between Ma muc thich nghi o/Ma ngap nuoc and 2 independent variables. The equation of the fitted model is

Ma muc thich nghi o/Ma ngap nuoc = 5.52415*Ma da lan^-0.626903*Ma loai uu the^-0.440852

In performing the fit, the estimation process terminated successully after 5 iterations, at which point the estimated coefficients appeared to converge to the current estimates.

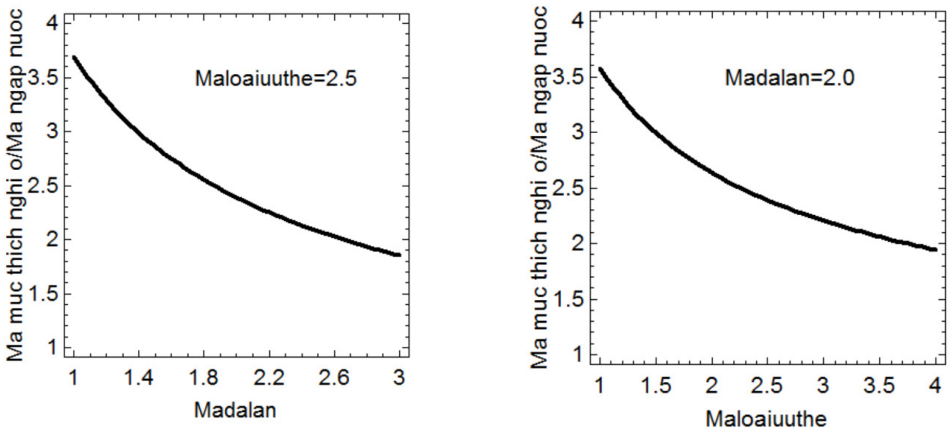
The R-Squared statistic indicates that the model as fitted explains 51.8202% of the variability in Ma muc thich nghi o/Ma ngap nuoc. The adjusted R-Squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 50.2405%. The standard error of the estimate shows the standard deviation of the residuals to be 216182. This value can be used to construct prediction limits for new observations by selecting the Forecasts option from the text menu. The mean absolute error (MAE) of 0.603025 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file.

The output also shows aymptotic 95.0% confidence intervals for each of the unknown parameters. These intervals are approximate and most accurate for large sample sizes. You can determine whether or not an estimate is statistically significant by examining each interval to see whether it contains the value 0. Intervals covering 0 correspond to coefficients which may well be removed form the model without hurting the fit substantially.

Kết quả mô hình được tóm tắt trong Bảng 8.16. Mô hình có sai số tuyệt đối trung bình MAE = 0.6, chưa đến một mức thích nghi, vì vậy có độ tin cậy tốt với sai số tương đối MAPE = 30%. Biến thiên của mức thích nghi tếch theo các nhân tố sinh thái qua mô hình thể hiện ở Hình 8.10.

**Bảng 8.16.** Mô hình quan hệ giữa mức thích nghi tếch với 3 nhân tố sinh thái có thể quan trắc trực tiếp trên hiện trường

| Mô hình | n ô | $R^2$ adj. % | Biến Weight | MAE | MAPE % |
|---|---|---|---|---|---|
| Ma muc thich nghi o/Ma ngap nuoc = 5,52415*Ma_da_lan^-0,626904*Ma loai uu the^-0,440856 | 64 | 50,24 | 1/Ma loai uu the^-20 | 0,60 | 29,98 |




**Hình 8.10.** Biến thiên "Mức thích nghi" tếch theo các nhân tố sinh thái định tính dễ xác định trên thực tế qua mô hình

**Bước 5:** Phân tích kết quả ứng dụng của mô hình: Mô hình được thiết lập thể hiện quan hệ giữa biến số phụ thuộc và các biến độc lập. Sự thay đổi các biến độc lập sẽ chỉ ra sự biến thiên của y. Do đó phân tích, lập ma trận ảnh hưởng của các biến độc lập đến phụ thuộc là cần thiết để có thể sử dụng đơn giản, từ đó hướng dẫn sử dụng.

Tiếp tục với minh họa lập mô hình nói trên, thế các giá trị mã hóa vào mô hình lập được bảng xác định mức thích nghi theo 3 nhân tố (Bảng 8.17). Trên hiện trường quan sát, xác định 3 nhân tố như sau:

*Ngập nước hay không?:* Quan sát và sử dụng thực vật chỉ thị là 2 loài sổ đất và mộc hoa, đa số các trường hợp khi xuất hiện 2 loài này thì đất sẽ ngập nhẹ đến úng trong mùa mưa.

*Loài cây ưu thế của rừng khộp*: Quan sát để xác định loài có số cây cao nhất trong lâm phần, việc này xác định khá thuận lợi vì ở các điều kiện lập địa khác nhau rừng khộp có loài ưu thế khác nhau rõ rệt, thường là 1 cho đến 2 loài ưu thế.

*Tỷ lệ đá lẫn:* Dùng khoan đất (hoặc đào) đến độ sâu 30-50cm, sau đó ước lượng tỷ lệ % sỏi lẫn có trong đất.

**Bảng 8.17.** Mức thích nghi của tếch theo 3 nhân tố sinh thái xác định nhanh trên hiện trường

| Mã / Loài ưu thế rừng khộp | Mã / Ngập nước | Mã / Cấp tỷ lệ đá lẫn | | |
|---|---|---|---|---|
| | | 1 / <50% | 2 / >70% | 3 / 50-70% |
| 1 / Dầu trà beng | 1 / Không | 4 | 4 | 3 |
| | 2 / Có | 4 | 4 | 4 |
| 2 / Cà chít | 1 / Không | 4 | 3 | 2 |
| | 2 / Có | 4 | 4 | 4 |
| 3 / Dầu đồng, Chiêu liêu đen, Căm xe | 1 / Không | 3 | 2 | 1 |
| | 2 / Có | 4 | 4 | 3 |
| 4 / Cẩm liên | 1 / Không | 3 | 2 | 1 |
| | 2 / Có | 4 | 4 | 3 |

# DỮ LIỆU THỰC HÀNH TIN HỌC THỐNG KÊ TRONG LÂM NGHIỆP

**Dữ liệu 1:** Số liệu trung bình các chỉ tiêu điều tra rừng trên ô mẫu rút thử trong một trạng thái rừng

| Stt | $\overline{DBH}$ | $\overline{H}$ | M |
|---|---|---|---|
| 1 | 15 | 17 | 34 |
| 2 | 16 | 18 | 34 |
| 3 | 17 | 19 | 45 |
| 4 | 21 | 23 | 45 |
| 5 | 21 | 23 | 56 |
| 6 | 22 | 24 | 56 |
| 7 | 23 | 25 | 56 |
| 8 | 21 | 23 | 56 |
| 9 | 22 | 24 | 67 |
| 10 | 21 | 23 | 67 |
| 11 | 15 | 17 | 76 |
| 12 | 25 | 27 | 78 |
| 13 | 25 | 27 | 78 |
| 14 | 23 | 25 | 78 |
| 15 | 24 | 26 | 78 |
| 16 | 25 | 27 | 78 |
| 17 | 26 | 28 | 78 |
| 18 | 21 | 23 | 78 |
| 19 | 23 | 25 | 89 |
| 20 | 24 | 26 | 89 |
| 21 | 34 | 36 | 89 |
| 22 | 32 | 34 | 89 |
| 23 | 31 | 33 | 98 |
| 24 | 29 | 31 | 112 |
| 25 | 29 | 31 | 113 |
| 26 | 31 | 33 | 115 |
| 27 | 35 | 37 | 124 |

*Nguồn: Tác giả. Ghi chú: $\overline{DBH}$: đường kính ngang ngực trung bình của ô (cm), $\overline{H}$: chiều cao trung bình của ô (m), M: trữ lượng quy ra ha của ô mẫu ($m^3$/ha).*

**Dữ liệu 2:** Carbon tích lũy trong cây rừng trên mặt đất (tấn/ha) ở các ô mẫu (0.1 ha) theo các trạng thái rừng

| Trung bình | Phục hồi | Gỗ - lồ ô |
|---|---|---|
| 147.3 | 43.3 | 189.0 |
| 109.9 | 33.4 | 59.7 |
| 161.7 | 20.9 | 12.8 |
| 107.6 | 32.7 | 16.0 |
| 226.0 | 40.6 | 20.7 |
| 172.5 | 10.1 | 38.8 |
| 126.7 | 56.0 | 14.2 |
| 159.3 | 7.0 | 29.8 |
| 126.4 | 23.2 | 11.4 |
| 190.7 | 99.4 | 93.2 |
| 135.0 | 172.6 | 33.4 |
| 75.5 | 51.4 | 89.4 |
| 97.1 | 86.4 | 22.1 |
| 147.8 | 46.2 | 145.5 |
| 204.1 | 5.3 | 215.2 |
| 126.1 | 15.8 | 141.7 |
| 161.1 | 51.2 | 156.8 |
| 58.0 | | 163.9 |
| 94.7 | | 92.5 |
| 314.6 | | 27.2 |
| 209.0 | | 143.0 |
| 98.2 | | 76.8 |
| 116.9 | | 55.6 |
| 165.8 | | 33.7 |

| Trung bình | Phục hồi | Gỗ - lồ ô |
|---|---|---|
| 122.9 | | 45.4 |
| 109.7 | | 74.5 |
| 93.4 | | 83.0 |
| 92.9 | | 50.9 |
| 124.1 | | 46.9 |
| 169.4 | | 110.8 |
| 172.1 | | 100.9 |
| 150.4 | | 124.8 |
| 85.6 | | 50.6 |
| 112.2 | | 38.8 |
| 95.3 | | 82.0 |
| 85.8 | | 107.0 |
| 122.9 | | 11.4 |
| 102.1 | | 42.0 |
| 88.6 | | 11.1 |
| 166.7 | | 16.5 |

*Nguồn: Dự án SNV-REDD+ ở huyện Bảo Lâm, tỉnh Lâm Đồng, 2013*

**Dữ liệu 3:** Số liệu đo cao (H, m) của cây tái sinh rừng khộp

| Stt | H tái sinh (m) | Stt | H tái sinh (m) | Stt | H tái sinh (m) |
|---|---|---|---|---|---|
| 1 | 1.5 | 21 | 1.0 | 41 | 1.7 |
| 2 | 1.3 | 22 | 0.7 | 42 | 2.2 |
| 3 | 0.8 | 23 | 1.9 | 43 | 1.9 |
| 4 | 1.9 | 24 | 1.8 | 44 | 1.8 |
| 5 | 1.7 | 25 | 1.6 | 45 | 1.6 |
| 6 | 2.2 | 26 | 2.0 | 46 | 2.0 |
| 7 | 2.5 | 27 | 1.9 | 47 | 1.5 |
| 8 | 1.0 | 28 | 1.7 | 48 | 1.3 |
| 9 | 0.7 | 29 | 2.2 | 49 | 0.8 |
| 10 | 1.9 | 30 | 2.5 | 50 | 1.9 |
| 11 | 1.8 | 31 | 1.0 | 51 | 1.7 |
| 12 | 1.6 | 32 | 0.7 | 52 | 2.2 |
| 13 | 2.0 | 33 | 1.9 | 53 | 2.5 |
| 14 | 1.5 | 34 | 1.8 | 54 | 1.0 |
| 15 | 1.3 | 35 | 1.6 | 55 | 0.7 |
| 16 | 0.8 | 36 | 2.0 | 56 | 1.9 |
| 17 | 1.9 | 37 | 1.5 | 57 | 1.8 |
| 18 | 1.7 | 38 | 1.3 | 58 | 1.6 |
| 19 | 2.2 | 39 | 0.8 | 59 | 2.0 |
| 20 | 2.5 | 40 | 1.9 | 60 | 1.9 |
| | | | | 61 | 1.7 |

*Nguồn: Tác giả*

**Dữ liệu 4:** Số liệu đo cao của hai mẫu theo hai phương pháp trồng thông 3 lá bằng cây con và rễ trần

| Stt | H cây con (m) | H rễ trần (m) | Stt | H cây con (m) | H rễ trần (m) | Stt | H cây con (m) | H rễ trần (m) |
|---|---|---|---|---|---|---|---|---|
| 1 | 13.6 | 13 | 32 | 11 | 14 | 63 | 9 | 14 |
| 2 | 14 | 13.5 | 33 | 11 | 13.5 | 64 | 12 | 13.5 |
| 3 | 13.8 | 12 | 34 | 12.5 | 12.5 | 65 | 8 | 14 |
| 4 | 13 | 13 | 35 | 11 | 12.5 | 66 | 9 | 13 |
| 5 | 11.8 | 12.5 | 36 | 11.5 | 15 | 67 | 9.5 | 13 |
| 6 | 12.8 | 13 | 37 | 12 | 14.5 | 68 | 8 | 15 |
| 7 | 13.5 | 12 | 38 | 10.5 | 13.5 | 69 | 11 | 12 |
| 8 | 13.2 | 12.5 | 39 | 12.5 | 11.5 | 70 | 9 | 13.5 |
| 9 | 11.5 | 12 | 40 | 12.5 | 15.5 | 71 | 12 | 11 |
| 10 | 11.5 | 12 | 41 | 12.5 | 15 | 72 | 13 | 12.5 |
| 11 | 10 | 13 | 42 | 11 | 14 | 73 | 11 | 14 |
| 12 | 13.5 | 13 | 43 | 10 | 16 | 74 | 12 | 14 |
| 13 | 14 | 12.5 | 44 | 11 | 15 | 75 | 12 | 14 |
| 14 | 13.5 | 14 | 45 | 11 | 16.5 | 76 | 13 | 8 |
| 15 | 10 | 14.5 | 46 | 10.5 | 15.5 | 77 | 12.5 | 15 |
| 16 | 12.8 | 15 | 47 | 12.5 | 11 | 78 | 13 | 15 |
| 17 | 14 | 16.5 | 48 | 13 | 13 | 79 | 12 | 14.5 |
| 18 | 12.5 | 10 | 49 | 12.5 | 11 | 80 | 12 | 15 |
| 19 | 13.5 | 12 | 50 | 12.5 | 12 | 81 | 11.5 | 15 |
| 20 | 11.8 | 15 | 51 | 10 | 13.5 | 82 | 12 | 14 |
| 21 | 12.5 | 13.5 | 52 | 11 | 15 | 83 | 11 | 14.5 |
| 22 | 12.5 | 13.5 | 53 | 12.5 | 13.5 | 84 | 13 | 12 |
| 23 | 10.5 | 14 | 54 | 13 | 13.5 | 85 | 12 | 14 |
| 24 | 13 | 14 | 55 | 8 | 13 | 86 | 11.5 | 14 |
| 25 | 8.5 | 14 | 56 | 10 | 13.5 | 87 | 11 | 14 |
| 26 | 10 | 13.5 | 57 | 9.5 | 14 | 88 | 12 | 12 |
| 27 | 13.8 | 13 | 58 | 10 | 15 | 89 | 13 | 13.5 |
| 28 | 15 | 13.5 | 59 | 8 | 13 | 90 | 11 | 15 |
| 29 | 13 | 13 | 60 | 8 | 13 | 91 | 12 | 14 |
| 30 | 11 | 13.5 | 61 | 10.5 | 12.5 | 92 | 12.5 | 10 |
| 31 | 12.5 | 14 | 62 | 9 | 13 | 93 | | 9 |

*Nguồn: Tác giả*

**Dữ liệu 5:** Các cặp dữ liệu chiều cao (H) qua tương quan và đo trực tiếp theo đường kính cây

| Stt | D (cm) | H (m) đo trực tiếp | H (m) qua tương quan |
|---|---|---|---|
| 1 | 31.3 | 22.0 | 24.2 |
| 2 | 32.0 | 21.8 | 24.8 |
| 3 | 30.6 | 21.5 | 23.6 |
| 4 | 27.9 | 21.6 | 21.2 |
| 5 | 10.2 | 6.4 | 6.6 |
| 6 | 10.2 | 6.5 | 6.6 |
| 7 | 9.6 | 5.9 | 6.2 |
| 8 | 9.5 | 5.7 | 6.1 |
| 9 | 9.5 | 6.1 | 6.1 |
| 10 | 10.2 | 6.0 | 6.6 |
| 11 | 16.4 | 7.3 | 11.5 |
| 12 | 15.9 | 7.4 | 11.0 |
| 13 | 10.0 | 6.5 | 6.5 |
| 14 | 10.1 | 6.6 | 6.5 |
| 15 | 15.7 | 11.7 | 10.9 |
| 16 | 15.5 | 11.8 | 10.7 |
| 17 | 6.7 | 3.5 | 4.1 |
| 18 | 6.8 | 3.6 | 4.1 |
| 19 | 11.9 | 8.0 | 7.9 |
| 20 | 11.9 | 8.1 | 7.9 |
| 21 | 15.5 | 12.1 | 10.7 |
| 22 | 15.4 | 12.1 | 10.6 |
| 23 | 10.1 | 6.1 | 6.5 |
| 24 | 10.1 | 6.4 | 6.5 |
| 25 | 17.7 | 12.2 | 12.5 |
| 26 | 18.1 | 12.4 | 12.8 |
| 27 | 17.8 | 15.6 | 12.6 |
| 28 | 18.3 | 15.4 | 13.0 |
| 29 | 17.4 | 16.0 | 12.3 |
| 30 | 16.5 | 16.2 | 11.5 |
| 31 | 17.0 | 15.7 | 11.9 |
| 32 | 17.2 | 15.3 | 12.1 |
| 33 | 14.4 | 9.3 | 9.8 |
| 34 | 14.7 | 9.2 | 10.1 |
| 35 | 16.1 | 12.8 | 11.2 |
| 36 | 14.6 | 11.4 | 10.0 |
| 37 | 29.3 | 17.1 | 22.4 |
| 38 | 25.9 | 16.8 | 19.4 |
| 39 | 12.6 | 9.7 | 8.4 |
| 40 | 13.9 | 9.8 | 9.5 |

*Nguồn: Tác giả*

**Dữ liệu 6:** Giá trị trung bình $\overline{DBH}$ (cm) của các ô thí nghiệm theo 7 xuất xứ thông Pinus caribeae khác nhau

| Mã số xuất xứ | $\overline{DBH}$ (cm) | Mã số xuất xứ | $\overline{DBH}$ (cm) |
|---|---|---|---|
| 1 | 10.8 | 4 | 9.0 |
| 1 | 11.2 | 4 | 10.8 |
| 1 | 10.4 | 4 | 11.5 |
| 1 | 9.9 | 4 | 8.7 |
| 2 | 12.3 | 5 | 14.2 |
| 2 | 11.5 | 5 | 12.9 |
| 2 | 9.5 | 6 | 12.3 |
| 2 | 10.0 | 6 | 12.5 |
| 3 | 9.4 | 6 | 12.4 |
| 3 | 10.5 | 6 | 10.8 |
| 3 | 11.0 | 7 | 7.0 |
| 3 | 9.5 | 7 | 9.8 |

*Nguồn: Tác giả*

**Dữ liệu 7:** Giá trị trung bình $\overline{DBH}$ (cm) của các ô thí nghiệm theo 16 xuất xứ thông Pinus kesiya với 4 lần lặp lại.

| Mã số xuất xứ | Lần lặp | DBH cm | Mã số xuất xứ | Lần lặp | DBH cm |
|---|---|---|---|---|---|
| 1 | 1 | 11.4 | 9 | 1 | 13.8 |
| 1 | 2 | 11.3 | 9 | 2 | 11.8 |
| 1 | 3 | 10.8 | 9 | 3 | 11.9 |
| 1 | 4 | 13.3 | 9 | 4 | 12.1 |
| 2 | 1 | 11.4 | 10 | 1 | 11.3 |
| 2 | 2 | 11.6 | 10 | 2 | 11.8 |
| 2 | 3 | 10.9 | 10 | 3 | 12.1 |
| 2 | 4 | 10.9 | 10 | 4 | 11.8 |
| 3 | 1 | 11.7 | 11 | 1 | 12.6 |
| 3 | 2 | 12.6 | 11 | 2 | 12.6 |
| 3 | 3 | 11.7 | 11 | 3 | 13.3 |
| 3 | 4 | 12.6 | 11 | 4 | 10.9 |
| 4 | 1 | 13.7 | 12 | 1 | 11.3 |
| 4 | 2 | 12.1 | 12 | 2 | 12.4 |
| 4 | 3 | 11.6 | 12 | 3 | 10.5 |
| 4 | 4 | 11.7 | 12 | 4 | 12.0 |
| 5 | 1 | 14.1 | 13 | 1 | 12.7 |
| 5 | 2 | 13.6 | 13 | 2 | 13.4 |
| 5 | 3 | 13.7 | 13 | 3 | 12.1 |
| 5 | 4 | 13.7 | 13 | 4 | 10.7 |
| 6 | 1 | 13.5 | 14 | 1 | 10.1 |
| 6 | 2 | 11.4 | 14 | 2 | 9.5 |
| 6 | 3 | 12.2 | 14 | 3 | 9.8 |
| 6 | 4 | 11.3 | 14 | 4 | 8.0 |
| 7 | 1 | 13.8 | 15 | 1 | 10.5 |
| 7 | 2 | 12.3 | 15 | 2 | 9.4 |
| 7 | 3 | 12.6 | 15 | 3 | 9.1 |
| 7 | 4 | 11.4 | 15 | 4 | 10.9 |
| 8 | 1 | 14.1 | 16 | 1 | 10.2 |
| 8 | 2 | 13.3 | 16 | 2 | 11.0 |
| 8 | 3 | 15.2 | 16 | 3 | 10.8 |
| 8 | 4 | 13.0 | 16 | 4 | 11.9 |

*Nguồn: Tác giả*

**Dữ liệu 8:** Dữ liệu thí nghiệm trồng lan kim tuyến dưới tán rừng theo 3 nhân tố: Nguồn gốc, độ cao và độ tàn che, với hai lần lặp lại, tạo thành 18 tổ hợp ba nhân tố với 36 ô thí nghiệm. Chỉ tiêu quan sát, đánh giá là tổng sinh khối tươi (g) / ô thí nghiệm sau 6 tháng gây trồng.

| Stt | Nguồn gốc lan | Đai cao | Độ tàn che | Tổ hợp | Sinh khối tươi (g) |
|---|---|---|---|---|---|
| 1 | Tu nhien | <750m | <30% | 1 | 0.1 |
| 2 | Tu nhien | <750m | <30% | 1 | 1.0 |
| 3 | Mo | <750m | <30% | 2 | 0.6 |
| 4 | Mo | <750m | <30% | 2 | 0.1 |
| 5 | Tu nhien | <750m | 30-70% | 3 | 0.7 |
| 6 | Tu nhien | <750m | 30-70% | 3 | 0.3 |
| 7 | Mo | <750m | 30-70% | 4 | 1.3 |
| 8 | Mo | <750m | 30-70% | 4 | 2.0 |
| 9 | Tu nhien | <750m | >70% | 5 | 3.4 |
| 10 | Tu nhien | <750m | >70% | 5 | 4.7 |
| 11 | Mo | <750m | >70% | 6 | 3.3 |
| 12 | Mo | <750m | >70% | 6 | 0.5 |
| 13 | Tu nhien | 750-1000m | <30% | 7 | 3.7 |
| 14 | Tu nhien | 750-1000m | <30% | 7 | 7.8 |
| 15 | Mo | 750-1000m | <30% | 8 | 1.2 |
| 16 | Mo | 750-1000m | <30% | 8 | 0.8 |
| 17 | Tu nhien | 750-1000m | 30-70% | 9 | 3.1 |
| 18 | Tu nhien | 750-1000m | 30-70% | 9 | 3.9 |
| 19 | Mo | 750-1000m | 30-70% | 10 | 3.5 |
| 20 | Mo | 750-1000m | 30-70% | 10 | 2.3 |
| 21 | Tu nhien | 750-1000m | >70% | 11 | 4.4 |
| 22 | Tu nhien | 750-1000m | >70% | 11 | 1.9 |
| 23 | Mo | 750-1000m | >70% | 12 | 1.3 |
| 24 | Mo | 750-1000m | >70% | 12 | 1.2 |
| 25 | Tu nhien | >1000m | <30% | 13 | 0.0 |
| 26 | Tu nhien | >1000m | <30% | 13 | 0.0 |
| 27 | Mo | >1000m | <30% | 14 | 0.0 |
| 28 | Mo | >1000m | <30% | 14 | 0.0 |
| 29 | Tu nhien | >1000m | 30-70% | 15 | 0.5 |
| 30 | Tu nhien | >1000m | 30-70% | 15 | 1.1 |
| 31 | Mo | >1000m | 30-70% | 16 | 0.1 |
| 32 | Mo | >1000m | 30-70% | 16 | 0.2 |
| 33 | Tu nhien | >1000m | >70% | 17 | 2.5 |
| 34 | Tu nhien | >1000m | >70% | 17 | 1.8 |
| 35 | Mo | >1000m | >70% | 18 | 0.5 |
| 36 | Mo | >1000m | >70% | 18 | 0.9 |

*Nguồn: Nguyễn Thị Quỳnh (2016)*

**Dữ liệu 9:** Tăng trưởng trung bình chiều cao cây tếch ở các ô thí nghiệm trồng làm giàu rừng khộp trên bốn loại đá mẹ ở Đăk Lăk

| Stt | TT_H tren Bazan | TT_H tren Cat ket | TT_H tren Macma axit | TT_H tren Phien set |
|---|---|---|---|---|
| 1 | 26.0 | 96.4 | 59.6 | 35.0 |
| 2 | 31.3 | 74.6 | 19.2 | 38.7 |
| 3 | 29.6 | 71.9 | 44.7 | 37.9 |
| 4 | 26.1 | 65.5 | 32.6 | |
| 5 | | 131.0 | 20.4 | |
| 6 | | 76.4 | 18.1 | |
| 7 | | 63.0 | 30.7 | |
| 8 | | 105.3 | 25.6 | |
| 9 | | 53.9 | 18.0 | |
| 10 | | 35.8 | 22.0 | |
| 11 | | 26.0 | 27.0 | |
| 12 | | 27.9 | 27.8 | |
| 13 | | 32.1 | 23.3 | |
| 14 | | 39.1 | 14.9 | |
| 15 | | 44.4 | 37.2 | |
| 16 | | 30.6 | 33.9 | |
| 17 | | 48.5 | | |
| 18 | | 51.9 | | |
| 19 | | 33.6 | | |
| 20 | | 44.8 | | |
| 21 | | 54.3 | | |
| 22 | | 53.1 | | |
| 23 | | 32.1 | | |
| 24 | | 43.6 | | |
| 25 | | 62.1 | | |
| 26 | | 49.9 | | |
| 27 | | 51.5 | | |
| 28 | | 50.8 | | |
| 29 | | 48.8 | | |
| 30 | | 29.4 | | |
| 31 | | 42.6 | | |
| 32 | | 22.0 | | |
| 33 | | 21.6 | | |
| 34 | | 25.8 | | |
| 35 | | 24.4 | | |
| 36 | | 17.4 | | |
| 37 | | 31.3 | | |
| 38 | | 36.7 | | |
| 39 | | 50.7 | | |
| 40 | | 50.9 | | |
| 41 | | 35.6 | | |

*TT_H: Tăng trưởng chiều cao trung bình trên ô thí nghiệm, cm. Nguồn: Bảo Huy (2014)*

**Dữ liệu 10:** Dữ liệu các giá trị sinh trưởng của 120 lâm phần rừng trồng tếch ở Tây Nguyên

| Ma o | Dia phuong | A | TVKH | Loai dat | Do cao | BA | Dg | Hg | Dgo | Ho | St ha | Stopt | N | V | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | EAKMAT | 42 | IIA3 | Fe nau do Bazan | 400 | 13.9 | 31.3 | 22.0 | 38.3 | 23.0 | 5178.0 | | 180 | 0.908 | 163.5 |
| 2 | EAKMAT | 42 | IIA3 | Fe nau do Bazan | 400 | 13.7 | 32.0 | 21.8 | 39.8 | 23.0 | 5370.1 | | 170 | 0.942 | 160.2 |
| 3 | EAKMAT | 42 | IIA3 | Fe nau do Bazan | 400 | 16.2 | 30.6 | 21.5 | 42.9 | 22.3 | 4233.3 | | 220 | 0.837 | 184.2 |
| 4 | EAKMAT | 42 | IIA3 | Fe nau do Bazan | 400 | 12.8 | 27.9 | 21.6 | 34.6 | 22.1 | 3174.4 | | 210 | 0.693 | 145.4 |
| 5 | BJVAM | 5 | IIA2 | Fe nau do Bazan | 340 | 5.3 | 10.2 | 6.4 | 12.3 | 6.9 | 7283.3 | | 650 | 0.037 | 24.2 |
| 6 | BJVAM | 5 | IIA2 | Fe nau do Bazan | 340 | 5.1 | 10.2 | 6.5 | 12.7 | 7.0 | 8135.3 | | 620 | 0.038 | 23.4 |
| 7 | BJVAM | 4 | IIA2 | Fe nau do Bazan | 340 | 5.0 | 9.6 | 5.9 | 12.3 | 6.7 | 8136.1 | | 690 | 0.032 | 22.0 |
| 8 | BJVAM | 4 | IIA2 | Fe nau do Bazan | 340 | 4.6 | 9.5 | 5.7 | 12.4 | 6.4 | 6725.1 | | 650 | 0.030 | 19.7 |
| 9 | BJVAM | 4 | IIA2 | Fe nau do Bazan | 340 | 4.6 | 9.5 | 6.1 | 11.9 | 6.8 | 8042.2 | | 650 | 0.032 | 20.8 |
| 10 | BJVAM | 4 | IIA2 | Fe nau do Bazan | 340 | 5.1 | 10.2 | 6.0 | 12.1 | 6.8 | 7313.2 | | 630 | 0.036 | 22.6 |
| 11 | BJVAM | 5 | IIA2 | Fe nau do Bazan | 340 | 13.9 | 16.4 | 7.3 | 19.8 | 7.9 | 23799.8 | | 660 | 0.105 | 69.5 |
| 12 | BJVAM | 5 | IIA2 | Fe nau do Bazan | 340 | 13.8 | 15.9 | 7.4 | 18.8 | 8.0 | 24379.4 | | 690 | 0.101 | 69.5 |
| 13 | NNUNG | 5 | IIA5 | Fe nau do Bazan | 600 | 7.7 | 10.0 | 6.5 | 13.0 | 7.5 | 10021.4 | | 980 | 0.037 | 35.9 |
| 14 | NNUNG | 5 | IIA5 | Fe nau do Bazan | 600 | 7.8 | 10.1 | 6.6 | 12.9 | 7.5 | 11119.5 | | 980 | 0.038 | 36.9 |
| 15 | NNUNG | 8 | IIA5 | Fe nau do Bazan | 600 | 18.8 | 15.7 | 11.7 | 19.0 | 12.3 | 14012.7 | | 970 | 0.135 | 130.7 |
| 16 | NNUNG | 8 | IIA5 | Fe nau do Bazan | 600 | 18.0 | 15.5 | 11.8 | 19.1 | 12.3 | 15099.8 | | 960 | 0.131 | 125.7 |
| 17 | NNUNG | 3 | IIA5 | Fe nau do Bazan | 600 | 3.4 | 6.7 | 3.5 | 9.7 | 4.3 | 6398.8 | | 960 | 0.012 | 11.3 |
| 18 | NNUNG | 3 | IIA5 | Fe nau do Bazan | 600 | 3.6 | 6.8 | 3.6 | 9.7 | 4.2 | 7319.9 | | 1000 | 0.012 | 12.3 |
| 19 | NNUNG | 7 | IIA5 | Fe nau do Bazan | 600 | 10.7 | 11.9 | 8.0 | 16.0 | 9.0 | 12473.9 | | 960 | 0.060 | 57.1 |
| 20 | NNUNG | 7 | IIA5 | Fe nau do Bazan | 600 | 10.9 | 11.9 | 8.1 | 16.3 | 8.9 | 13723.2 | | 970 | 0.061 | 58.7 |
| 21 | NNUNG | 9 | IIA5 | Fe nau do Bazan | 600 | 18.0 | 15.5 | 12.1 | 18.8 | 12.5 | 13296.0 | | 950 | 0.135 | 127.9 |
| 22 | NNUNG | 9 | IIA5 | Fe nau do Bazan | 600 | 18.0 | 15.4 | 12.1 | 18.8 | 12.5 | 15140.3 | | 970 | 0.132 | 127.9 |
| 23 | NNUNG | 5 | IIA5 | Fe nau do Bazan | 600 | 8.7 | 10.1 | 6.1 | 13.3 | 7.1 | 12259.2 | | 1090 | 0.036 | 39.4 |
| 24 | NNUNG | 5 | IIA5 | Fe nau do Bazan | 600 | 9.1 | 10.1 | 6.4 | 12.8 | 7.2 | 15036.9 | | 1140 | 0.037 | 41.8 |
| 25 | DLAP | 10 | IIA5 | Fe nau do Bazan | 500 | 22.0 | 17.7 | 12.2 | 21.6 | 12.8 | 6291.1 | | 890 | 0.178 | 158.2 |
| 26 | DLAP | 10 | IIA5 | Fe nau do Bazan | 500 | 24.7 | 18.1 | 12.4 | 21.7 | 13.0 | 12847.9 | | 960 | 0.186 | 178.4 |
| 27 | DLAP | 10 | IIA5 | Fe nau do Bazan | 500 | 26.2 | 17.8 | 15.6 | 22.6 | 16.3 | 11490.4 | | 1050 | 0.217 | 227.6 |
| 28 | DLAP | 10 | IIA5 | Fe nau do Bazan | 500 | 27.0 | 18.3 | 15.4 | 21.7 | 15.8 | 10857.7 | | 1020 | 0.227 | 231.6 |
| 29 | KANA | 12 | IIA3 | Fe nau do Bazan | 480 | 26.3 | 17.4 | 16.0 | 22.0 | 16.8 | 19487.5 | | 1110 | 0.211 | 233.8 |
| 30 | KANA | 12 | IIA3 | Fe nau do Bazan | 480 | 22.7 | 16.5 | 16.2 | 20.4 | 17.0 | 19220.2 | | 1060 | 0.192 | 203.8 |
| 31 | KANA | 12 | IIA3 | Fe nau do Bazan | 480 | 21.3 | 17.0 | 15.7 | 21.7 | 16.6 | 29048.7 | | 940 | 0.198 | 186.5 |
| 32 | KANA | 12 | IIA3 | Fe nau do Bazan | 480 | 17.0 | 17.2 | 15.3 | 21.7 | 16.0 | 12682.3 | | 730 | 0.200 | 145.7 |
| 33 | KANA | 9 | IIA3 | Fe nau do Bazan | 480 | 23.6 | 14.4 | 9.3 | 18.6 | 10.5 | 19342.9 | | 1440 | 0.097 | 140.0 |
| 34 | KANA | 9 | IIA3 | Fe nau do Bazan | 480 | 21.1 | 14.7 | 9.2 | 18.4 | 10.1 | 13407.3 | | 1240 | 0.100 | 123.7 |
| 35 | KANA | 10 | IIA3 | Fe nau do Bazan | 480 | 15.0 | 16.1 | 12.8 | 20.4 | 12.8 | 11732.4 | | 740 | 0.150 | 111.3 |
| 36 | KANA | 10 | IIA3 | Fe nau do Bazan | 480 | 11.8 | 14.6 | 11.4 | 19.3 | 13.3 | 14896.3 | | 710 | 0.116 | 82.4 |
| 37 | KANA | 14 | IIA3 | Fe nau do Bazan | 480 | 24.3 | 29.3 | 17.1 | 40.1 | 17.9 | 15136.2 | | 360 | 0.633 | 227.7 |
| 38 | KANA | 14 | IIA3 | Fe nau do Bazan | 480 | 20.6 | 25.9 | 16.8 | 31.5 | 17.7 | 11417.0 | | 390 | 0.492 | 191.8 |
| 39 | KANA | 11 | IIA3 | Fe nau do Bazan | 480 | 7.1 | 12.6 | 9.7 | 16.5 | 10.9 | 9368.4 | | 570 | 0.077 | 43.8 |
| 40 | KANA | 11 | IIA3 | Fe nau do Bazan | 480 | 8.6 | 13.9 | 9.8 | 18.6 | 11.1 | 8491.0 | | 570 | 0.093 | 53.2 |

| Ma o | Dia phuong | A | TVKH | Loai dat | Do cao | BA | Dg | Hg | Dgo | Ho | St ha | Stopt | N | V | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 41 | KANA | 10 | IIA3 | Fe nau do Bazan | 480 | 9.3 | 15.0 | 13.3 | 19.1 | 14.6 | 8211.5 | | 530 | 0.137 | 72.5 |
| 42 | KANA | 10 | IIA3 | Fe nau do Bazan | 480 | 10.2 | 14.9 | 10.9 | 19.0 | 12.2 | 11992.4 | | 590 | 0.133 | 78.5 |
| 43 | DLAP | 11 | IIA5 | Fe nau do Bazan | 500 | 22.8 | 18.9 | 11.9 | 22.8 | 12.7 | 10407.0 | 13.8 | 820 | 0.196 | 160.7 |
| 44 | NNUNG | 6 | IIA5 | Fe nau do Bazan | 600 | 8.5 | 11.1 | 8.8 | 13.5 | 9.6 | 5405.0 | 7.2 | 870 | 0.054 | 47.2 |
| 45 | NNUNG | 9 | IIA5 | Fe nau do Bazan | 600 | 7.5 | 9.7 | 9.7 | 13.6 | 11.4 | 6988.0 | 7.1 | 1030 | 0.044 | 45.7 |
| 46 | NNUNG | 10 | IIA5 | Fe nau do Bazan | 600 | 12.6 | 13.6 | 10.8 | 16.5 | 12.2 | 8948.0 | 11.1 | 870 | 0.094 | 82.1 |
| 47 | NNUNG | 4 | IIA5 | Fe nau do Bazan | 600 | 10.0 | 13.4 | 8.6 | 14.7 | 9.0 | 7547.0 | 11.1 | 810 | 0.078 | 63.0 |
| 48 | KANA | 13 | IIA3 | Fe nau do Bazan | 480 | 13.4 | 16.8 | 13.0 | 20.4 | 14.9 | 7142.0 | 13.7 | 600 | 0.166 | 99.5 |
| 49 | KANA | 15 | IIA3 | Fe nau do Bazan | 480 | 22.6 | 28.7 | 15.2 | 37.6 | 16.0 | 7247.0 | 25.1 | 350 | 0.547 | 191.6 |
| 50 | KANA | 11 | IIA3 | Fe nau do Bazan | 480 | 12.8 | 15.6 | 13.2 | 19.2 | 14.3 | 5056.0 | 10.2 | 670 | 0.145 | 96.9 |
| 51 | KANA | 10 | IIA3 | Fe nau do Bazan | 480 | 19.2 | 16.7 | 13.1 | 21.0 | 14.2 | 14027.0 | 17.1 | 880 | 0.165 | 145.0 |
| 52 | KANA | 13 | IIA3 | Fe nau do Bazan | 480 | 11.5 | 17.5 | 15.0 | 21.7 | 16.4 | 6319.0 | 17.0 | 480 | 0.201 | 96.7 |
| 53 | KANA | 9 | IIA3 | Fe nau do Bazan | 480 | 7.6 | 13.7 | 9.8 | 16.4 | 10.8 | 2348.0 | 5.4 | 510 | 0.089 | 45.5 |
| 54 | BJVAM | 6 | IIA2 | Fe nau do Bazan | 340 | 16.0 | 17.7 | 9.2 | 21.6 | 10.8 | 7834.0 | 14.3 | 650 | 0.142 | 92.5 |
| 55 | BJVAM | 5 | IIA2 | Fe nau do Bazan | 340 | 7.9 | 12.5 | 6.6 | 15.9 | 7.3 | 4877.0 | 9.5 | 640 | 0.057 | 36.3 |
| 57 | DLAP | 13 | IIA5 | Fe nau do Bazan | 500 | 18.4 | 17.8 | 15.9 | 23.3 | 16.9 | 4687.0 | 8.0 | 740 | 0.218 | 161.6 |
| 58 | DLAP | 13 | IIA5 | Fe nau do Bazan | 500 | 16.3 | 17.0 | 15.8 | 22.3 | 17.1 | 5947.0 | 12.5 | 720 | 0.198 | 142.7 |
| 59 | DLAP | 13 | IIA5 | Fe nau do Bazan | 500 | 20.5 | 15.9 | 12.9 | 20.7 | 13.7 | 5374.0 | 7.4 | 1030 | 0.148 | 152.0 |
| 60 | DLAP | 13 | IIA5 | Fe nau do Bazan | 500 | 19.8 | 15.8 | 13.2 | 20.3 | 13.9 | 5258.0 | 7.1 | 1010 | 0.148 | 149.9 |
| 61 | DLAP | 13 | IIA5 | Fe nau do Bazan | 500 | 13.8 | 18.9 | 15.8 | 24.1 | 16.6 | 5658.0 | 16.0 | 490 | 0.245 | 120.0 |
| 62 | DLAP | 13 | IIA5 | Fe nau do Bazan | 500 | 26.3 | 19.1 | 16.2 | 24.0 | 17.2 | 4849.0 | 7.0 | 910 | 0.255 | 232.3 |
| 63 | DLAP | 13 | IIA5 | Fe nau do Bazan | 500 | 20.2 | 16.5 | 15.6 | 24.1 | 17.2 | 5280.0 | 9.4 | 910 | 0.185 | 168.1 |
| 64 | DLAP | 13 | IIA5 | Fe nau do Bazan | 500 | 25.2 | 20.2 | 16.5 | 27.1 | 17.6 | 8449.0 | 13.7 | 780 | 0.290 | 226.1 |
| 65 | DLAP | 13 | IIA5 | Fe nau do Bazan | 500 | 25.7 | 19.7 | 16.2 | 25.5 | 16.8 | 8044.0 | 12.4 | 840 | 0.272 | 228.1 |
| 66 | KANA | 14 | IIA3 | Fe nau do Bazan | 480 | 25.4 | 20.7 | 15.9 | 26.2 | 17.2 | 6687.0 | 12.3 | 760 | 0.295 | 224.4 |
| 67 | KANA | 11 | IIA3 | Fe nau do Bazan | 480 | 17.5 | 16.4 | 12.7 | 20.5 | 13.7 | 3532.0 | 5.6 | 830 | 0.155 | 128.8 |
| 68 | KANA | 11 | IIA3 | Fe nau do Bazan | 480 | 16.4 | 17.5 | 12.3 | 22.7 | 13.0 | 6611.0 | 12.4 | 680 | 0.172 | 117.2 |
| 69 | KANA | 11 | IIA3 | Fe nau do Bazan | 480 | 14.0 | 16.9 | 14.8 | 21.2 | 19.0 | 3942.0 | 8.4 | 620 | 0.186 | 115.2 |
| 70 | KANA | 10 | IIA3 | Fe nau do Bazan | 480 | 15.8 | 15.9 | 11.9 | 19.7 | 13.2 | 3398.0 | 6.1 | 800 | 0.139 | 111.0 |
| 71 | KANA | 16 | IIA3 | Fe nau do Bazan | 480 | 18.5 | 27.6 | 19.4 | 33.3 | 20.9 | 15292.0 | 60.6 | 310 | 0.619 | 191.8 |
| 72 | KANA | 16 | IIA3 | Fe nau do Bazan | 480 | 15.6 | 26.2 | 18.2 | 32.3 | 19.1 | 8036.0 | 31.0 | 290 | 0.529 | 153.3 |
| 73 | KANA | 16 | IIA3 | Fe nau do Bazan | 480 | 16.9 | 30.6 | 19.2 | 41.6 | 20.8 | 11627.0 | 50.5 | 230 | 0.754 | 173.4 |
| 74 | KANA | 12 | IIA3 | Fe nau do Bazan | 480 | 15.0 | 15.9 | 11.0 | 20.2 | 12.2 | 2329.0 | 3.4 | 760 | 0.131 | 99.4 |
| 75 | KANA | 12 | IIA3 | Fe nau do Bazan | 480 | 8.5 | 14.5 | 10.4 | 17.5 | 11.4 | 1524.0 | 4.0 | 500 | 0.104 | 52.2 |
| 76 | KANA | 12 | IIA3 | Fe nau do Bazan | 480 | 11.6 | 15.3 | 11.0 | 19.6 | 12.3 | 1811.0 | 3.1 | 630 | 0.121 | 76.3 |
| 77 | KANA | 14 | IIA3 | Fe nau do Bazan | 480 | 34.0 | 23.1 | 16.8 | 28.9 | 17.5 | 15174.0 | 19.8 | 810 | 0.385 | 311.5 |
| 78 | KANA | 14 | IIA3 | Fe nau do Bazan | 480 | 28.8 | 21.1 | 16.0 | 26.9 | 17.8 | 6783.0 | 10.8 | 820 | 0.308 | 252.9 |
| 79 | KANA | 10 | IIA3 | Fe nau do Bazan | 480 | 17.0 | 15.8 | 11.8 | 19.5 | 13.1 | 3838.0 | 6.4 | 870 | 0.136 | 118.4 |
| 80 | KANA | 10 | IIA3 | Fe nau do Bazan | 480 | 15.7 | 15.6 | 11.7 | 19.1 | 12.8 | 2967.0 | 4.7 | 820 | 0.132 | 108.1 |
| 81 | KANA | 13 | IIA3 | Fe nau do Bazan | 480 | 16.6 | 19.0 | 13.3 | 24.0 | 14.3 | 9331.0 | 20.1 | 580 | 0.216 | 125.2 |
| 82 | KANA | 13 | IIA3 | Fe nau do Bazan | 480 | 13.5 | 18.0 | 14.3 | 23.4 | 15.8 | 7069.0 | 18.1 | 530 | 0.205 | 108.7 |
| 83 | KANA | 13 | IIA3 | Fe nau do Bazan | 480 | 10.3 | 17.2 | 12.8 | 22.3 | 14.1 | 4137.0 | 12.6 | 440 | 0.172 | 75.5 |

| Ma o | Dia phuong | A | TVKH | Loai dat | Do cao | BA | Dg | Hg | Dgo | Ho | St ha | Stopt | N | V | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 84 | BJVAM | 7 | IIA2 | Fe nau do Bazan | 340 | 7.0 | 16.7 | 8.2 | 22.0 | 9.1 | 1890.0 | 6.3 | 320 | 0.117 | 37.4 |
| 85 | BJVAM | 7 | IIA2 | Fe nau do Bazan | 340 | 6.5 | 14.5 | 8.8 | 17.5 | 9.7 | 2486.0 | 6.9 | 380 | 0.093 | 35.2 |
| 86 | BJVAM | 7 | IIA2 | Fe nau do Bazan | 340 | 8.4 | 16.7 | 9.7 | 21.2 | 10.4 | 2667.0 | 8.7 | 380 | 0.132 | 50.0 |
| 87 | BJVAM | 5 | IIA2 | Fe nau do Bazan | 340 | 1.9 | 9.2 | 6.5 | 11.1 | 7.2 | 1040.0 | 3.6 | 280 | 0.030 | 8.5 |
| 88 | BJVAM | 6 | IIA2 | Fe nau do Bazan | 340 | 3.9 | 11.9 | 8.0 | 16.9 | 9.3 | 3743.0 | 11.9 | 350 | 0.058 | 20.4 |
| 89 | BJVAM | 6 | IIA2 | Fe nau do Bazan | 340 | 4.3 | 12.5 | 8.8 | 17.5 | 10.4 | 2596.0 | 7.7 | 350 | 0.069 | 24.1 |
| 90 | BJVAM | 5 | IIA2 | Fe nau do Bazan | 340 | 2.1 | 10.4 | 7.0 | 13.7 | 8.0 | 2134.0 | 10.6 | 250 | 0.041 | 10.2 |
| 91 | BJVAM | 5 | IIA2 | Fe nau do Bazan | 340 | 2.0 | 9.7 | 7.1 | 12.1 | 7.7 | 2004.0 | 8.1 | 270 | 0.036 | 9.7 |
| 92 | BJVAM | 5 | IIA2 | Fe nau do Bazan | 340 | 1.6 | 9.2 | 6.3 | 12.0 | 7.3 | 1728.0 | 8.6 | 240 | 0.030 | 7.2 |
| 93 | CUMGA | 11 | IIA3 | Fe nau do Bazan | 325 | 27.6 | 22.7 | 12.2 | 26.9 | 12.9 | 7435.0 | 11.1 | 680 | 0.288 | 196.0 |
| 94 | CUMGA | 11 | IIA3 | Fe nau do Bazan | 325 | 20.4 | 17.2 | 11.5 | 21.3 | 12.4 | 7415.0 | 8.6 | 880 | 0.158 | 139.2 |
| 95 | CUMGA | 11 | IIA3 | Fe nau do Bazan | 325 | 19.8 | 18.2 | 11.9 | 23.9 | 12.1 | 5382.0 | 7.2 | 760 | 0.182 | 138.2 |
| 96 | CUMGA | 11 | IIA3 | Fe nau do Bazan | 325 | 14.0 | 16.7 | 11.8 | 22.2 | 12.7 | 7709.0 | 12.5 | 640 | 0.152 | 97.3 |
| 97 | CUMGA | 11 | IIA3 | Fe nau do Bazan | 325 | 21.9 | 17.5 | 11.8 | 22.5 | 12.2 | 6237.0 | 7.2 | 910 | 0.167 | 152.0 |
| 98 | CUMGA | 9 | IIA3 | Fe nau do Bazan | 325 | 13.6 | 15.5 | 10.9 | 21.0 | 12.0 | 6553.0 | 8.8 | 720 | 0.123 | 88.9 |
| 99 | CUMGA | 9 | IIA3 | Fe nau do Bazan | 325 | 16.5 | 16.3 | 10.8 | 21.4 | 11.1 | 6473.0 | 8.9 | 790 | 0.136 | 107.1 |
| 100 | CUMGA | 9 | IIA3 | Fe nau do Bazan | 325 | 13.2 | 15.4 | 10.0 | 19.5 | 10.1 | 4985.0 | 7.1 | 710 | 0.114 | 81.2 |
| 101 | EAKMAT | 44 | IIA3 | Fe nau do Bazan | 400 | 18.4 | 34.2 | 22.5 | 46.9 | 22.9 | 5677.0 | 28.4 | 200 | 1.077 | 215.4 |
| 102 | EAKMAT | 44 | IIA3 | Fe nau do Bazan | 400 | 16.7 | 35.3 | 22.8 | 44.6 | 23.2 | 4471.0 | 26.3 | 170 | 1.161 | 197.3 |
| 103 | EAKMAT | 44 | IIA3 | Fe nau do Bazan | 400 | 26.6 | 43.4 | 25.1 | 57.8 | 25.3 | 5529.0 | 30.7 | 180 | 1.906 | 343.1 |
| GT1 | KANA | 16 | IIA3 | Fe nau do Bazan | 480 | 16.5 | 27.9 | 19.7 | 34.2 | 20.6 | 7814.0 | 35.3 | 280 | 0.640 | 179.3 |
| GT2 | CUMGA | 11 | IIA3 | Fe nau do Bazan | 325 | 14.7 | 17.0 | 14.4 | 20.4 | 14.8 | 5702.0 | 10.8 | 650 | 0.184 | 119.6 |
| GT3 | BJVAM | 8 | IIA2 | Fe nau do Bazan | 340 | 7.7 | 14.3 | 8.6 | 18.7 | 10.3 | 5763.0 | 18.1 | 490 | 0.089 | 43.4 |
| GT4 | NNUNG | 5 | IIA5 | Fe nau do Bazan | 600 | 10.6 | 13.3 | 9.0 | 14.4 | 10.9 | 5816.0 | 7.1 | 1050 | 0.079 | 83.0 |
| GT5 | DLAP | 13 | IIA5 | Fe nau do Bazan | 500 | 21.5 | 18.1 | 17.1 | 22.4 | 17.7 | 9804.0 | 12.7 | 840 | 0.240 | 201.3 |
| I | DLAP | 13 | IIA5 | Fe nau do Bazan | 500 | 29.6 | 19.2 | 15.4 | 23.1 | 15.9 | 2940.0 | 4.1 | 1020 | 0.248 | 252.5 |
| II | DLAP | 14 | IIA5 | Fe nau do Bazan | 500 | 19.2 | 19.0 | 15.2 | 25.5 | 16.5 | 4572.0 | 6.2 | 680 | 0.240 | 163.2 |
| III | KANA | 17 | IIA3 | Fe nau do Bazan | 480 | 23.7 | 29.0 | 16.2 | 35.4 | 17.3 | 7206.0 | 23.6 | 360 | 0.588 | 211.8 |
| IV | DLAP | 12 | IIA5 | Fe nau do Bazan | 500 | 15.0 | 16.7 | 13.4 | 20.8 | 15.3 | 7036.0 | 13.8 | 680 | 0.168 | 114.1 |
| IX | EAKMAT | 45 | IIA3 | Fe nau do Bazan | 400 | 12.1 | 31.0 | 24.7 | 35.9 | 25.5 | 5757.0 | 43.4 | 160 | 0.959 | 153.5 |
| V | BJVAM | 6 | IIA2 | Fe nau do Bazan | 340 | 2.7 | 10.1 | 5.7 | 12.9 | 6.5 | 1603.0 | 5.3 | 340 | 0.034 | 11.5 |
| VI | BJVAM | 6 | IIA2 | Fe nau do Bazan | 340 | 5.1 | 11.2 | 7.2 | 13.5 | 7.7 | 1478.0 | 3.5 | 520 | 0.048 | 25.0 |
| VII | BJVAM | 7 | IIA2 | Fe nau do Bazan | 340 | 3.6 | 9.4 | 5.8 | 15.1 | 6.8 | 1843.0 | 8.4 | 520 | 0.030 | 15.4 |
| VIII | CUMGA | 12 | IIA3 | Fe nau do Bazan | 325 | 23.8 | 22.9 | 17.9 | 28.2 | 18.3 | 8184.0 | 15.1 | 580 | 0.398 | 231.0 |
| X | EAKMAT | 45 | IIA3 | Fe nau do Bazan | 400 | 12.0 | 31.0 | 19.5 | 33.5 | 20.1 | 5504.0 | 34.4 | 160 | 0.784 | 125.4 |
| XI | EAKMAT | 44 | IIA3 | Fe nau do Bazan | 400 | 15.6 | 31.5 | 17.7 | 34.8 | 18.3 | 6933.0 | 37.1 | 200 | 0.747 | 149.3 |
| XII | KTUM | 10 | IIA3 | Fe vang xam Mac ma acid | 500 | 12.6 | 10.8 | 9.3 | 15.9 | 10.8 | 9178.0 | 9.1 | 1380 | 0.053 | 73.6 |
| XIII | KTUM | 12 | IIA3 | Fe vang xam Mac ma acid | 500 | 17.2 | 14.1 | 12.8 | 18.4 | 13.3 | 10730.0 | 10.8 | 1100 | 0.115 | 126.9 |

*Nguồn: Bảo Huy và cộng sự 1998*

**Dữ liệu 11:** Dữ liệu sinh khối cây rừng trên mặt đất và các biến số điều tra cây rừng ở vùng sinh thái nam trung bộ, tỉnh Quảng Nam

| ID | Mã cây | Tên loài | Tên khoa học | DBH, cm | H, m | WD, g/cm$^3$ | CA, m$^2$ | AGB, kg |
|---|---|---|---|---|---|---|---|---|
| 1 | I.1 | Trám | *Canarium littorale* Blume | 14.6 | 12.0 | 0.588 | 16.62 | 67.7 |
| 2 | I.2 | Sổ | *Dillenia indica var. aurea* (Sm.) Kuntze | 9.6 | 9.3 | 0.552 | 7.07 | 27.9 |
| 3 | I.3 | Côm | *Elaeocarpus kontumensis* Gagnep. | 12.1 | 11.5 | 0.582 | 9.08 | 47.8 |
| 4 | I.4 | Lộc vừng | *Barringtonia racenmosa (*L.) Spreng. | 11.4 | 9.3 | 0.541 | 9.08 | 39.9 |
| 5 | I.5 | Trâm | *Syzygium levinei* (Merr.) Merr. | 13.6 | 13.5 | 0.603 | 13.85 | 53.4 |
| 6 | I.6 | Nhọc | *Polyalthia nemoralis* Aug.DC. | 6.5 | 6.2 | 0.593 | 7.07 | 14.7 |
| 7 | I.7 | Sổ | *Dillenia indica var. aurea* (Sm.) Kuntze | 13.4 | 14.0 | 0.559 | 13.20 | 76.0 |
| 8 | I.8 | Sổ | *Dillenia indica var. aurea* (Sm.) Kuntze | 9.3 | 11.6 | 0.476 | 5.73 | 21.9 |
| 9 | I.9 | Nhọc | *Polyalthia nemoralis* Aug.DC. | 13.5 | 15.1 | 0.614 | 10.75 | 77.7 |
| 10 | I.10 | Trôm | *Sterculia parviflora* Roxb. | 12.0 | 13.9 | 0.533 | 13.20 | 57.1 |
| 11 | I.11 | Bùi tía | *Ilex annamensis* Tardieu | 6.9 | 7.6 | 0.581 | 2.84 | 11.8 |
| 12 | I.12 | Sổ | *Dillenia indica var. aurea* (Sm.) Kuntze | 11.0 | 11.5 | 0.536 | 6.16 | 34.0 |
| 13 | I.13 | Thị rừng | *Diospyros decandra* Lour. | 14.5 | 17.4 | 0.664 | 13.20 | 120.3 |
| 14 | I.14 | sp | *Scaphium lychnophorum* (Hance) Pierre | 6.2 | 9.5 | 0.591 | 2.54 | 9.7 |
| 15 | I.15 | Dành dành | *Gardenia philastrei* Pierre ex Pit. | 10.6 | 12.9 | 0.566 | 9.62 | 54.3 |
| 16 | I.16 | Trám | *Canarium littorale* Blume | 5.6 | 12.2 | 0.620 | 1.33 | 7.7 |
| 17 | I.17 | Nhọc | *Polyalthia nemoralis* Aug.DC. | 10.2 | 10.9 | 0.586 | 7.07 | 33.2 |
| 18 | I.18 | sp | *Scaphium lychnophorum* (Hance) Pierre | 7.2 | 10.4 | 0.638 | 5.31 | 13.5 |
| 19 | I.19 | sp | *Scaphium lychnophorum* (Hance) Pierre | 10.2 | 12.1 | 0.568 | 3.80 | 35.4 |
| 20 | I.20 | sp | *Scaphium lychnophorum* (Hance) Pierre | 8.5 | 10.3 | 0.619 | 13.85 | 25.2 |
| 21 | I.21 | Dẻ | *Lithocarpus annamensis* (Hickel & A.Camus) Barnett | 11.7 | 14.3 | 0.585 | 0.79 | 44.2 |
| 22 | I.22 | Trâm | *Syzygium levinei* (Merr.) Merr. | 7.0 | 9.4 | 0.536 | 2.54 | 12.0 |
| 23 | I.23 | Lộc vừng | *Barringtonia racenmosa (*L.) Spreng. | 9.1 | 7.2 | 0.494 | 6.16 | 19.8 |
| 24 | I.24 | Giổi | *Aglaia roxburghiana* (Wight & Arn.) Miq. | 6.8 | 8.0 | 0.659 | 6.16 | 18.0 |
| 25 | I.25 | sp | *Scaphium lychnophorum* (Hance) Pierre | 8.8 | 11.0 | 0.561 | 2.54 | 21.1 |
| 26 | I.26 | Trâm | *Syzygium levinei* (Merr.) Merr. | 16.4 | 18.4 | 0.567 | 15.21 | 100.4 |
| 27 | I.27 | Xoan đào | *Prunus ceylanica* (Wight.) Miq. | 23.4 | 18.0 | 0.589 | 7.07 | 238.4 |
| 28 | I.28 | sp | *Scaphium lychnophorum* (Hance) Pierre | 21.7 | 23.3 | 0.589 | 10.18 | 248.5 |
| 29 | I.29 | Gáo | *Nauclea orientalis* (L.) L. | 16.9 | 17.0 | 0.430 | 9.08 | 77.5 |
| 30 | I.30 | Máu chó | *Knema pierrei* Warb. | 15.7 | 14.2 | 0.595 | 14.52 | 80.7 |
| 31 | I.31 | sp | *Scaphium lychnophorum* (Hance) Pierre | 18.8 | 20.5 | 0.591 | 18.10 | 175.0 |
| 32 | I.32 | Bình Linh | *Vitex sp.* | 16.0 | 14.3 | 0.524 | 20.43 | 100.1 |
| 33 | I.33 | Trâm | *Syzygium levinei* (Merr.) Merr. | 24.2 | 22.9 | 0.595 | 30.19 | 283.5 |
| 34 | I.34 | Máu chó | *Knema pierrei Warb.* | 17.2 | 16.3 | 0.591 | 26.42 | 136.1 |
| 35 | I.35 | Ngát vàng | *Gironniera subaequalis Planch.* | 19.9 | 13.5 | 0.506 | 15.21 | 128.5 |
| 36 | I.36 | Chè rừng | *Camellia fleuryi* (A.Chev.) Sealy | 32.3 | 21.7 | 0.505 | 32.17 | 610.0 |
| 37 | I.37 | Sòi | *Sapium baccatum Roxb.* | 35.9 | 20.2 | 0.544 | 8.04 | 450.4 |
| 38 | I.38 | Vàng nghệ | *Garcinia hanburyi* Hook.f. | 25.9 | 16.9 | 0.691 | 30.19 | 411.0 |
| 39 | I.39 | Bứa | *Garcinia oliveri* Pierre | 30.2 | 17.5 | 0.712 | 40.72 | 644.3 |

| ID | Mã cây | Tên loài | Tên khoa học | DBH, cm | H, m | WD, g/cm$^3$ | CA, m$^2$ | AGB, kg |
|---|---|---|---|---|---|---|---|---|
| 40 | I.40 | Chò | *Shorea farinosa* C.E.C.Fisch. | 31.3 | 31.2 | 0.637 | 38.48 | 853.7 |
| 41 | I.41 | Bưởi bung | *Maclurodendron oligophlebium* (Merr.) T.G. Hartley | 49.5 | 22.8 | 0.495 | 46.57 | 968.2 |
| 42 | I.42 | Bưởi bung | *Maclurodendron oligophlebium* (Merr.) T.G. Hartley | 36.6 | 22.2 | 0.518 | 40.72 | 742.6 |
| 43 | I.43 | Chò | *Shorea farinosa* C.E.C.Fisch. | 42.8 | 31.5 | 0.630 | 36.32 | 1903.4 |
| 44 | I.44 | Dẻ | *Lithocarpus annamensis* (Hickel & A.Camus) Barnett | 54.7 | 29.2 | 0.566 | 24.63 | 2960.7 |
| 45 | I.45 | Giổi | *Magnolia braianensis* (Gagnep.) Figlar | 54.1 | 33.5 | 0.634 | 40.72 | 1988.4 |
| 46 | I.46 | Trôm | *Sterculia parviflora* Roxb. | 49.7 | 34.4 | 0.644 | 51.53 | 1651.8 |
| 47 | I.47 | Côm | *Elaeocarpus kontumensis* Gagnep. | 55.5 | 27.6 | 0.598 | 30.19 | 747.1 |
| 48 | I.48 | Trám | *Canarium littorale* Blume | 55.0 | 32.7 | 0.610 | 34.21 | 1800.5 |
| 49 | I.49 | Bời lời lá bầu dục | *Litsea elliptica* Blume | 57.8 | 31.5 | 0.582 | 32.17 | 2345.4 |
| 50 | I.50 | Trám | *Canarium littorale* Blume | 65.0 | 32.5 | 0.640 | 78.54 | 3687.3 |
| 51 | I.51 | Còng | *Calophyllum dryobalanoides* Pierre | 68.4 | 26.4 | 0.567 | 201.06 | 3894.5 |
| 52 | I.52 | Xoan | *Melia azedarach* L. | 66.5 | 35.0 | 0.502 | 52.81 | 2785.9 |
| 53 | I.53 | Trám | *Canarium littorale* Blume | 82.4 | 33.5 | 0.644 | 201.06 | 6349.5 |
| 54 | I.54 | Sòi | *Sapium baccatum* Roxb. | 79.0 | 34.5 | 0.575 | 55.42 | 3954.1 |
| 55 | I.55 | Chò | *Shorea farinosa* C.E.C.Fisch. | 87.7 | 41.4 | 0.663 | 91.61 | 8633.0 |
| 56 | II.1 | An tức hương | *Styrax benzoin Dryand.* | 12.8 | 14.1 | 0.557 | 4.91 | 63.6 |
| 57 | II.2 | Chè rừng | *Camellia fleuryi* (A.Chev.) Sealy | 8.8 | 8.7 | 0.613 | 7.07 | 22.7 |
| 58 | II.3 | Trám | *Canarium littorale Blume* | 13.5 | 13.1 | 0.645 | 12.57 | 51.4 |
| 59 | II.4 | Trâm | *Syzygium levinei* (Merr.) Merr. | 14.3 | 12.6 | 0.620 | 15.21 | 91.6 |
| 60 | II.5 | Côm | *Elaeocarpus kontumensis Gagnep.* | 10.6 | 6.7 | 0.572 | 8.04 | 14.9 |
| 61 | II.6 | Chè rừng | *Camellia fleuryi* (A.Chev.) Sealy | 12.4 | 10.2 | 0.647 | 7.07 | 44.9 |
| 62 | II.7 | Chè rừng | *Camellia fleuryi* (A.Chev.) Sealy | 8.4 | 12.3 | 0.624 | 3.14 | 28.1 |
| 63 | II.8 | Nhọc | *Polyalthia nemoralis* Aug.DC. | 11.8 | 11.2 | 0.640 | 5.31 | 47.9 |
| 64 | II.9 | Chò | *Shorea farinosa* C.E.C.Fisch. | 10.4 | 15.2 | 0.553 | 4.52 | 50.6 |
| 65 | II.10 | Sơn | *Madhuca alpina* (A.Chev. ex Lecomte) A.Chev. | 14.8 | 19.6 | 0.613 | 10.18 | 118.0 |
| 66 | II.11 | Dâu da | *Baccaurea ramiflora Lour.* | 10.9 | 14.2 | 0.603 | 5.73 | 43.5 |
| 67 | II.12 | Ngâu rừng | *Aglaia elaeagnoidea* (A.Juss.) Benth. | 9.2 | 12.6 | 0.485 | 11.34 | 19.8 |
| 68 | II.13 | Nhọc | *Polyalthia nemoralis* Aug.DC. | 5.5 | 7.1 | 0.538 | 3.14 | 7.0 |
| 69 | II.14 | Thị | *Diospyros pilosula (*A.DC.) Wall. ex Hiern | 5.7 | 6.4 | 0.621 | 2.54 | 6.4 |
| 70 | II.15 | Thị | *Diospyros pilosula* (A.DC.) Wall. ex Hiern | 7.6 | 9.1 | 0.641 | 3.14 | 19.9 |
| 71 | II.16 | Giổi | *Magnolia braianensis* (Gagnep.) Figlar | 7.2 | 8.1 | 0.516 | 4.91 | 6.5 |
| 72 | II.17 | Ngát | *Gironniera subaequalis Planch.* | 13.4 | 14.1 | 0.568 | 13.85 | 64.3 |
| 73 | II.18 | Máu chó | Knema pierrei Warb. | 5.4 | 4.7 | 0.609 | 4.52 | 6.8 |
| 74 | II.19 | Bời lời | *Litsea baviensis var. venulosa* H. Liu | 5.2 | 7.4 | 0.515 | 7.07 | 7.6 |
| 75 | II.20 | Trâm | *Syzygium levinei* (Merr.) Merr. | 7.2 | 9.6 | 0.614 | 7.07 | 21.5 |
| 76 | II.21 | Máu chó | *Knema pierrei Warb.* | 5.4 | 6.8 | 0.596 | 9.08 | 6.9 |
| 77 | II.22 | Sơn | *Madhuca alpina* (A.Chev. ex Lecomte) A.Chev. | 5.3 | 9.2 | 0.633 | 3.14 | 8.6 |
| 78 | II.23 | Ngát | *Gironniera subaequalis* Planch. | 6.5 | 5.2 | 0.549 | 3.80 | 6.2 |
| 79 | II.24 | Nhãn rừng | *Lepisanthes rubiginosa (*Roxb.) Leenh. | 4.9 | 8.5 | 0.649 | 3.14 | 7.3 |
| 80 | II.25 | Nhọc | *Polyalthia nemoralis* Aug.DC. | 5.1 | 5.8 | 0.577 | 8.04 | 5.9 |
| 81 | II.26 | Chò | *Shorea farinosa* C.E.C.Fisch. | 9.9 | 11.1 | 0.600 | 6.16 | 30.0 |

| ID | Mã cây | Tên loài | Tên khoa học | DBH, cm | H, m | WD, g/cm$^3$ | CA, m$^2$ | AGB, kg |
|---|---|---|---|---|---|---|---|---|
| 82 | II.27 | Dẻ | *Lithocarpus annamensis* (Hickel & A.Camus) Barnett | 14.7 | 16.5 | 0.549 | 11.95 | 86.2 |
| 83 | II.28 | Dẻ | *Lithocarpus annamensis* (Hickel & A.Camus) Barnett | 20.2 | 20.0 | 0.626 | 16.62 | 260.0 |
| 84 | II.29 | Lộc vừng | *Barringtonia racenmosa (*L.) Spreng. | 16.9 | 10.4 | 0.556 | 20.43 | 117.9 |
| 85 | II.30 | Chò | *Shorea farinosa* C.E.C.Fisch. | 24.4 | 18.5 | 0.595 | 28.27 | 329.0 |
| 86 | II.31 | Trâm | *Syzygium levinei* (Merr.) Merr. | 17.1 | 14.0 | 0.650 | 12.57 | 153.8 |
| 87 | II.32 | sp | *Scaphium lychnophorum* (Hance) Pierre | 17.4 | 16.1 | 0.593 | 8.04 | 93.8 |
| 88 | II.33 | Bưởi bung | *Maclurodendron oligophlebium* (Merr.) T.G. Hartley | 18.0 | 17.0 | 0.560 | 11.34 | 103.4 |
| 89 | II.34 | Dẻ | *Lithocarpus annamensis* (Hickel & A.Camus) Barnett | 15.5 | 17.3 | 0.517 | 5.73 | 83.5 |
| 90 | II.35 | Nhãn rừng | *Lepisanthes rubiginosa (*Roxb.) Leenh. | 22.3 | 15.9 | 0.561 | 14.52 | 170.7 |
| 91 | II.36 | Giổi | *Aglaia roxburghiana* (Wight & Arn.) Miq. | 16.8 | 15.0 | 0.511 | 3.80 | 99.1 |
| 92 | II.37 | Bứa | *Garcinia oliveri Pierre* | 25.3 | 22.0 | 0.543 | 15.21 | 423.2 |
| 93 | II.38 | Giổi | *Aglaia roxburghiana (*Wight & Arn.) Miq. | 26.6 | 13.5 | 0.501 | 17.35 | 261.4 |
| 94 | II.39 | Chiêu liêu xanh | *Terminalia calamansanay Rolfe* | 35.5 | 27.3 | 0.574 | 15.90 | 903.3 |
| 95 | II.40 | Dẻ | *Lithocarpus annamensis (*Hickel & A.Camus) Barnett | 29.8 | 14.1 | 0.570 | 55.42 | 457.8 |
| 96 | II.41 | Lòng máng lá nhỏ | *Pterospermum diversifolium Blume* | 37.1 | 20.3 | 0.556 | 59.45 | 980.3 |
| 97 | II.42 | Giổi | *Magnolia braianensis* (Gagnep.) Figlar | 39.2 | 26.1 | 0.646 | 28.27 | 1224.3 |
| 98 | II.43 | Săng máu | *Horsfieldia amygdalina* (Wall.) Warb. | 41.1 | 23.9 | 0.565 | 18.10 | 968.6 |
| 99 | II.44 | Re Hương | *Cinnamomum subavenium Miq.* | 42.9 | 27.8 | 0.626 | 27.34 | 1243.2 |
| 100 | II.45 | Trâm | *Syzygium levinei* (Merr.) Merr. | 52.1 | 23.2 | 0.583 | 35.26 | 2275.1 |
| 101 | II.46 | Ngát | *Gironniera subaequalis* Planch. | 41.4 | 22.5 | 0.481 | 34.21 | 848.8 |
| 102 | II.47 | Dẻ | *Lithocarpus annamensis (*Hickel & A.Camus) Barnett | 51.9 | 24.5 | 0.645 | 45.36 | 2376.7 |
| 103 | II.48 | Thị | *Diospyros pilosula* (A.DC.) Wall. ex Hiern | 48.0 | 23.2 | 0.611 | 58.09 | 1503.1 |
| 104 | II.49 | Vạng trứng | *Endospermum chinense Benth.* | 60.0 | 27.2 | 0.570 | 66.48 | 1782.9 |
| 105 | II.50 | Sơn huyết | *Melanorrhea curtisii* Oliv. | 62.3 | 25.4 | 0.626 | 66.48 | 3816.3 |
| 106 | II.51 | Trám | *Canarium littorale* Blume | 51.2 | 27.5 | 0.634 | 45.36 | 1933.3 |
| 107 | II.52 | Sơn | *Madhuca alpina (*A.Chev. ex Lecomte) A.Chev. | 53.3 | 25.4 | 0.646 | 62.21 | 2074.1 |
| 108 | II.53 | Giổi | *Aglaia roxburghiana* (Wight & Arn.) Miq. | 65.9 | 27.2 | 0.660 | 84.95 | 3032.0 |
| 109 | II.54 | Vàng nghệ | *Garcinia hanburyi* Hook.f. | 67.5 | 26.3 | 0.698 | 50.27 | 3608.0 |
| 110 | II.55 | Chò | *Shorea farinosa* C.E.C.Fisch. | 75.1 | 40.5 | 0.602 | 81.71 | 6575.9 |

*Nguồn: Huy et al. (2016b)*

**Dữ liệu 12:** Giá trị sinh trưởng bình quân rừng trồng trám trắng ở các tính Lạng Sơn, Bắc Giang và Quảng Ninh

| Stt | Dia phuong | ID o | A (Năm) | Ho (m) | N (cây/ha) | Dg (cm) | Hg (m) | V ($m^3$) |
|---|---|---|---|---|---|---|---|---|
| 1 | Cao Loc | 01 (06) | 7 | 7.9 | 209 | 6.7 | 6.5 | 0.012232 |
| 2 | Cao Loc | 02 (07) | 7 | 4.9 | 191 | 3.8 | 4.7 | 0.003414 |
| 3 | Chi Lang | 1 | 11 | 9.4 | 620 | 5.7 | 6.2 | 0.010826 |
| 4 | Chi Lang | 2 | 11 | 7.5 | 460 | 3.1 | 4.6 | 0.002606 |
| 5 | Chi Lang | 3 | 10 | 8.3 | 420 | 4.5 | 6.5 | 0.006530 |
| 6 | Chi Lang | 4 | 10 | 6.2 | 475 | 4.9 | 4.2 | 0.005618 |
| 7 | Chi Lang | 5 | 11 | 6.8 | 450 | 5.7 | 6.0 | 0.011050 |
| 8 | Chi Lang | 6 | 11 | 7.8 | 410 | 3.6 | 4.4 | 0.003337 |
| 9 | Chi Lang | 7 | 10 | 10.7 | 487 | 7.3 | 5.1 | 0.013604 |
| 10 | Chi Lang | 8 | 9 | 8.0 | 600 | 10.7 | 6.7 | 0.033916 |
| 11 | Chi Lang | 9 | 9 | 9.4 | 450 | 11.9 | 8.7 | 0.053910 |
| 12 | Chi Lang | 10 | 9 | 4.3 |  | 3.0 | 3.1 | 0.002097 |
| 13 | Chi Lang | 11 | 9 | 6.6 | 450 | 5.4 | 4.8 | 0.008501 |
| 14 | Dinh Lap | 1 | 9 | 5.5 | 380 | 3.7 | 4.6 | 0.003944 |
| 15 | Dinh Lap | 2 | 9 | 6.6 | 440 | 5.6 | 5.4 | 0.008004 |
| 16 | Dinh Lap | 3 | 9 | 6.4 | 400 | 6.1 | 6.0 | 0.011446 |
| 17 | Dinh Lap | 4 | 9 | 7.6 | 390 | 3.3 | 5.0 | 0.003450 |
| 18 | Dinh Lap | 5 | 9 | 7.2 | 420 | 4.7 | 5.7 | 0.007081 |
| 19 | Dinh Lap | 6 | 9 | 8.6 | 420 | 6.4 | 6.8 | 0.014201 |
| 20 | Luc Nam HG | 1 | 5 | 3.5 | 327 | 3.1 | 3.4 | 0.002222 |
| 21 | Luc Nam HG | 2 | 5 | 5.2 | 300 | 4.3 | 4.5 | 0.004756 |
| 22 | Luc Nam HG | 3 | 5 | 6.0 | 388 | 5.8 | 5.2 | 0.008695 |
| 23 | Luc Nam Thuan | 1 | 6 | 6.5 | 282 | 6.0 | 5.8 | 0.010743 |
| 24 | Luc Nam Thuan | 2 | 6 | 7.9 | 400 | 6.7 | 6.1 | 0.013859 |
| 25 | Luc Nam Thuan | 3 | 5 | 4.7 | 300 | 4.7 | 4.7 | 0.006295 |
| 26 | Luc Nam Thuan | 4 | 5 | 6.5 | 309 | 5.7 | 5.7 | 0.009483 |
| 27 | Luc Nam Thuan | 5 | 5 | 4.0 | 390 | 3.6 | 3.3 | 0.002855 |
| 28 | Luc Nam Thuan | 6 | 3 | 2.5 | 291 | 1.4 | 2.1 | 0.000343 |
| 29 | Luc Nam Thuan | 7 | 3 | 3.4 | 300 | 2.4 | 2.9 | 0.001217 |
| 30 | Luc Ngan HG | 1 | 9 | 6.0 | 318 | 3.9 | 4.3 | 0.004338 |
| 31 | Luc Ngan HG | 2 | 9 | 5.3 | 290 | 4.7 | 4.3 | 0.005973 |
| 32 | Luc Ngan HG | 3 | 9 |  | 336 | 8.4 | 7.6 | 0.025633 |
| 33 | Luc Ngan HG | 4 | 9 | 5.7 | 300 | 3.9 | 4.3 | 0.004590 |
| 34 | Luc Ngan HG | 5 | 9 | 5.7 | 463 | 5.1 | 5.5 | 0.009001 |
| 35 | Luc Ngan HG | 6 | 9 | 5.8 | 450 | 3.4 | 4.4 | 0.003680 |
| 36 | Luc ngan (tram thuan) | 1 | 11 |  | 318 | 11.0 | 7.9 | 0.041159 |
| 37 | Luc ngan (tram thuan) | 2 | 11 | 7.8 | 318 | 8.5 | 7.7 | 0.026015 |

| Stt | Dia phuong | ID o | A (Năm) | Ho (m) | N (cây/ha) | Dg (cm) | Hg (m) | V ($m^3$) |
|---|---|---|---|---|---|---|---|---|
| 38 | Luc ngan (tram thuan) | 3 | 11 | 9.4 | 327 | 10.6 | 8.2 | 0.039464 |
| 39 | Luc ngan (tram thuan) | 4 | 11 | | 354 | 10.2 | 10.7 | 0.049755 |
| 40 | Luc ngan (tram thuan) | 5 | 11 | | 318 | 9.1 | 8.5 | 0.026236 |
| 41 | Luc ngan (tram thuan) | 6 | 10 | 10.3 | 300 | 10.6 | 8.0 | 0.037439 |
| 42 | Luc ngan (tram thuan) | 7 | 4 | 2.7 | 375 | 1.5 | 2.3 | 0.000586 |
| 43 | Luc ngan (tram thuan) | 8 | 10 | 8.7 | 355 | 10.2 | 8.1 | 0.038855 |
| 44 | Luc ngan (tram thuan) | 9 | 10 | 8.8 | 425 | 10.7 | 7.3 | 0.038261 |
| 45 | Luc ngan (tram thuan) | 10 | 10 | 11.1 | 345 | 10.6 | 9.0 | 0.044960 |
| 46 | Luc ngan (tram thuan) | 11 | 5 | 3.6 | 354 | 2.1 | 2.7 | 0.001444 |
| 47 | Loc Binh | 1 | 7 | 6.3 | 400 | 5.5 | 6.2 | 0.010031 |
| 48 | Son Dong HG | 1 | 5 | 5.6 | 280 | 3.3 | 3.8 | 0.002695 |
| 49 | Son Dong HG | 2 | 5 | 4.4 | 290 | 2.7 | 3.5 | 0.001670 |
| 50 | Son Dong HG | 3 | 4 | 4.9 | 290 | 3.3 | 3.4 | 0.002458 |
| 51 | Son Dong HG | 4 | 7 | | 221 | 6.8 | 6.2 | 0.013807 |
| 52 | Son Dong HG | 5 | 7 | 6.4 | 229 | 5.7 | 6.0 | 0.009518 |
| 53 | Son Dong HG | 6 | 7 | 4.7 | 228 | 5.6 | 4.6 | 0.006530 |
| 54 | Son Dong 1 HG | 20 | 5 | 4.7 | 460 | 5.0 | 4.7 | 0.006658 |
| 55 | Son Dong 1 HG | 21 | 5 | | 350 | 6.0 | 5.1 | 0.008579 |
| 56 | Son Dong 1 HG | 23 | 5 | 4.4 | 340 | 3.8 | 3.9 | 0.003737 |
| 57 | Son Dong 1 Thuan | 1 | 5 | 4.7 | 463 | 4.5 | 3.9 | 0.005463 |
| 58 | Son Dong 1 Thuan | 2 | 5 | 4.8 | 388 | 2.7 | 2.9 | 0.001658 |
| 59 | Son Dong 1 Thuan | 3 | 5 | 4.8 | 300 | 3.2 | 3.5 | 0.002280 |
| 60 | Son Dong 1 Thuan | 4 | 4 | 4.7 | 388 | 3.6 | 3.3 | 0.003356 |
| 61 | Son Dong 1 Thuan | 5 | 7 | 4.0 | 413 | 2.0 | 3.0 | 0.000855 |
| 62 | Son Dong 1 Thuan | 6 | 7 | 3.9 | 309 | 2.2 | 3.2 | 0.001116 |
| 63 | Son Dong 1 Thuan | 8 | 7 | 5.2 | 413 | 2.9 | 3.3 | 0.001917 |
| 64 | Son Dong II HG | 1 | 5 | 4.6 | 400 | 3.8 | 3.9 | 0.003982 |
| 65 | Son Dong II HG | 2 | 5 | 4.1 | 388 | 2.9 | 3.2 | 0.001935 |
| 66 | Son Dong II HG | 3 | 9 | 7.1 | 450 | 6.2 | 6.2 | 0.013402 |
| 67 | Son Dong II HG | 4 | 9 | 9.1 | 463 | 7.8 | 7.6 | 0.022441 |
| 68 | Son Dong II HG | 5 | 5 | 3.9 | 373 | 2.9 | 2.8 | 0.001774 |
| 69 | Son Dong II HG | 6 | 9 | 7.9 | 425 | 6.5 | 6.6 | 0.014834 |
| 70 | Son Dong II Thuan | 8 | 8 | 8.9 | 388 | 7.6 | 6.6 | 0.020793 |
| 71 | Son Dong II Thuan | 10 | 10 | 7.9 | 562 | 6.4 | 6.2 | 0.012132 |
| 72 | Son Dong II Thuan | 11 | 10 | 10.1 | 400 | 7.2 | 6.6 | 0.018270 |
| 73 | Son Dong II Thuan | 12 | 10 | 6.2 | 388 | 5.2 | 4.5 | 0.005998 |

*Nguồn: Bảo Huy và Đào Công Khanh (2008)*

Dữ liệu 13: Bộ dữ liệu sinh khối cây rừng trên mặt đất (AGB) rừng lá rộng thường xanh ở 5 vùng sinh thái, theo họ thực vật và cấp khối lượng thể tích gỗ (WD) ở Việt Nam

| ID | Vùng sinh thái | Họ thực vật | Tên khoa học loài | DBH, cm | H, m | WD, g/cm$^3$ | AGB, kg | WD class |
|---|---|---|---|---|---|---|---|---|
| 1 | NE | Ulmaceae | *Gironniera subaequalis* | 33.5 | 24.5 | 0.485 | 513.0 | WD2 |
| 2 | NE | Others | *Elaeocarpus tonkinensis* | 8.7 | 12.0 | 0.560 | 23.4 | WD2 |
| 3 | NE | Others | *Adinandra bockiana* | 25.5 | 15.5 | 0.415 | 197.2 | WD2 |
| 4 | NE | Fagaceae | *Castanopsis sp.* | 39.5 | 23.6 | 0.723 | 1262.2 | WD3 |
| 5 | NE | Others | *Canarium sp.* | 12.9 | 12.0 | 0.488 | 51.7 | WD2 |
| 6 | NE | Fagaceae | *Castanopsis sp.* | 44.7 | 24.0 | 0.613 | 1129.9 | WD3 |
| 7 | NE | Others | *Adinandra bockiana* | 23.9 | 14.8 | 0.390 | 158.1 | WD1 |
| 8 | NE | Others | *Elaeocarpus tonkinensis* | 30.4 | 20.1 | 0.628 | 419.9 | WD3 |
| 9 | NE | Others | *Garcinia multiflora* | 12.6 | 11.4 | 0.535 | 58.6 | WD2 |
| 10 | NE | Others | *Schefflera heptaphilla* | 28.2 | 19.4 | 0.378 | 261.4 | WD1 |
| 11 | NE | Lauraceae | *Cinnamomum balansae* | 12.1 | 12.2 | 0.510 | 42.9 | WD2 |
| 12 | NE | Others | *Manglietia sp.* | 19.3 | 13.7 | 0.405 | 125.6 | WD2 |
| 13 | NE | Fagaceae | *Castanopsis sp.* | 49.7 | 20.6 | 0.664 | 1548.4 | WD3 |
| 14 | NE | Others | *Rhaphiolepis indica* | 6.0 | 9.2 | 0.756 | 14.4 | WD3 |
| 15 | NE | Others | *Choerospondias axillaris* | 20.2 | 21.0 | 0.509 | 137.2 | WD2 |
| 16 | NE | Others | *Elaeocarpus tonkinensis* | 9.7 | 12.0 | 0.523 | 32.0 | WD2 |
| 17 | NE | Leguminosae | *Archidendron tonkinensis* | 7.8 | 11.2 | 0.357 | 10.8 | WD1 |
| 18 | NE | Others | *Elaeocarpus tonkinensis* | 7.6 | 8.2 | 0.560 | 14.3 | WD2 |
| 19 | NE | Leguminosae | *Cassia javanica* | 12.4 | 14.0 | 0.517 | 60.3 | WD2 |
| 20 | NE | Lauraceae | *Cinnamomum parthenoxylon* | 24.8 | 18.8 | 0.372 | 340.2 | WD1 |
| 21 | NE | Others | *Canarium sp.* | 17.2 | 14.3 | 0.442 | 86.6 | WD2 |
| 22 | NE | Myrtaceae | *Syzygium chanlos* | 6.2 | 6.0 | 0.570 | 10.8 | WD2 |
| 23 | NE | Others | *Trevesia palmata* | 12.6 | 9.6 | 0.410 | 21.8 | WD2 |
| 24 | NE | Others | *Prunus arborea* | 9.7 | 12.2 | 0.473 | 30.0 | WD2 |
| 25 | NE | Fagaceae | *Castanopsis sp.* | 8.6 | 9.6 | 0.473 | 23.6 | WD2 |
| 26 | NE | Fagaceae | *Castanopsis sp.* | 8.1 | 9.8 | 0.465 | 22.0 | WD2 |
| 27 | NE | Others | *Elaeocarpus tonkinensis* | 11.8 | 12.3 | 0.571 | 47.2 | WD2 |
| 28 | NE | Others | *Engelhardtia roxburghiana* | 39.4 | 22.9 | 0.528 | 681.7 | WD2 |
| 29 | NE | Euphorbiaceae | *Macaranga denticulata* | 17.7 | 20.5 | 0.448 | 130.2 | WD2 |
| 30 | NE | Euphorbiaceae | *Mallotus paniculatus* | 10.2 | 16.0 | 0.371 | 36.8 | WD1 |
| 31 | NE | Others | *Elaeocarpus tonkinensis* | 6.7 | 8.6 | 0.465 | 10.7 | WD2 |
| 32 | NE | Others | *Styrax tonkinensis* | 35.3 | 26.5 | 0.420 | 579.9 | WD2 |
| 33 | NE | Others | *Choerospondias axillaris* | 22.4 | 19.3 | 0.464 | 122.6 | WD2 |
| 34 | NE | Lauraceae | *Phoebe tovoyana* | 8.7 | 12.5 | 0.521 | 23.9 | WD2 |
| 35 | NE | Euphorbiaceae | *Macaranga denticulata* | 19.7 | 23.0 | 0.441 | 164.8 | WD2 |
| 36 | NE | Others | *Adinandra bockiana* | 31.0 | 17.2 | 0.570 | 447.9 | WD2 |
| 37 | NE | Fagaceae | *Lithocarpus sp.* | 35.1 | 21.6 | 0.770 | 1078.0 | WD3 |
| 38 | NE | Fagaceae | *Castanopsis sp.* | 27.4 | 18.5 | 0.681 | 477.2 | WD3 |
| 39 | NE | Fagaceae | *Castanopsis sp.* | 39.6 | 24.6 | 0.662 | 1013.4 | WD3 |
| 40 | NE | Lauraceae | *Cinnamomum parthenoxylon* | 30.1 | 25.4 | 0.479 | 575.3 | WD2 |
| 41 | NE | Others | *Styrax tonkinensis* | 30.2 | 21.0 | 0.368 | 339.6 | WD1 |
| 42 | NE | Others | *Elaeocarpus floribundus* | 24.5 | 18.0 | 0.616 | 402.4 | WD3 |
| 43 | NE | Others | *Choerospondias axillaris* | 35.6 | 24.0 | 0.590 | 672.3 | WD2 |
| 44 | NE | Fagaceae | *Castanopsis sp.* | 50.0 | 26.0 | 0.839 | 2504.1 | WD3 |
| 45 | NE | Others | *Rhaphiolepis indica* | 7.9 | 9.5 | 0.945 | 25.4 | WD3 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 46 | NE | Euphorbiaceae | *Endospermum chinense* | 49.2 | 27.0 | 0.428 | 1215.2 | WD2 |
| 47 | NE | Fagaceae | *Castanopsis sp.* | 52.9 | 25.0 | 0.669 | 2296.8 | WD3 |
| 48 | NE | Others | *Engelhardtia roxburghiana* | 74.6 | 32.7 | 0.524 | 3806.1 | WD2 |
| 49 | NE | Lauraceae | *Cinnamomum parthenoxylon* | 55.0 | 31.3 | 0.526 | 2058.2 | WD2 |
| 50 | NE | Others | *Schefflera heptaphilla* | 54.5 | 25.0 | 0.407 | 1090.5 | WD2 |
| 51 | NE | Others | *Elaeocarpus floribundus* | 65.5 | 28.5 | 0.451 | 2789.7 | WD2 |
| 52 | NE | Lauraceae | *Phoebe tovoyana* | 6.8 | 9.0 | 0.366 | 6.9 | WD1 |
| 53 | NE | Others | *Prunus arborea* | 12.5 | 11.0 | 0.391 | 35.2 | WD1 |
| 54 | NE | Others | *Manglietia sp.* | 6.8 | 8.2 | 0.292 | 9.2 | WD1 |
| 55 | NE | Meliaceae | *Aglaia spectabilis* | 7.5 | 8.5 | 0.433 | 11.5 | WD2 |
| 56 | NE | Others | *Gmelina arborea* | 10.5 | 10.0 | 0.432 | 21.4 | WD2 |
| 57 | NE | Others | *Garcinia multiflora* | 29.1 | 20.4 | 0.745 | 450.7 | WD3 |
| 58 | NE | Others | *Canarium parvum* | 16.2 | 19.4 | 0.581 | 153.8 | WD2 |
| 59 | NE | Lauraceae | *Cinnamomum sp.* | 24.5 | 21.8 | 0.665 | 368.4 | WD3 |
| 60 | NE | Others | *Garcinia oblongifolia* | 16.6 | 16.5 | 0.721 | 158.8 | WD3 |
| 61 | NE | Lauraceae | *Cinnamomum sp.* | 33.4 | 19.7 | 0.659 | 792.9 | WD3 |
| 62 | NE | Others | *Pterospermum heterophyllum* | 30.9 | 19.2 | 0.673 | 587.9 | WD3 |
| 63 | NE | Lauraceae | *Cinnamomum sp.* | 18.0 | 14.3 | 0.695 | 158.4 | WD3 |
| 64 | NE | Fagaceae | *Castanopsis sp.* | 65.0 | 23.7 | 0.801 | 2771.9 | WD3 |
| 65 | NE | Others | *Wendlandia paniculata* | 14.5 | 12.8 | 0.648 | 57.2 | WD3 |
| 66 | NE | Lauraceae | *Cinnamomum sp.* | 13.7 | 14.4 | 0.686 | 83.0 | WD3 |
| 67 | NE | Dipterocarpaceae | *Hopea mollissima* | 15.1 | 14.1 | 0.928 | 161.5 | WD3 |
| 68 | NE | Lauraceae | *Litsea glutinosa* | 23.9 | 15.7 | 0.674 | 331.1 | WD3 |
| 69 | NE | Others | *Prunus arborea* | 36.6 | 24.2 | 0.548 | 666.4 | WD2 |
| 70 | NE | Ulmaceae | *Gironniera subaequalis* | 6.8 | 7.8 | 0.507 | 10.4 | WD2 |
| 71 | NE | Lauraceae | *Phoebe tovoyana* | 28.8 | 14.5 | 0.499 | 329.1 | WD2 |
| 72 | NE | Others | *Huodendron biaristatum* | 24.8 | 17.5 | 0.719 | 294.2 | WD3 |
| 73 | NE | Fagaceae | *Castanopsis sp.* | 6.4 | 9.1 | 0.608 | 11.9 | WD3 |
| 74 | NE | Myrtaceae | *Syzygium chanlos* | 10.5 | 12.6 | 0.962 | 57.6 | WD3 |
| 75 | NE | Leguminosae | *Archidendron chevalieri* | 28.7 | 18.2 | 0.544 | 360.0 | WD2 |
| 76 | NE | Myrtaceae | *Syzygium chanlos* | 30.6 | 18.6 | 0.787 | 532.2 | WD3 |
| 77 | NE | Leguminosae | *Albizia lebbeck* | 19.9 | 15.6 | 0.770 | 235.5 | WD3 |
| 78 | NE | Dipterocarpaceae | *Hopea mollissima* | 57.3 | 25.2 | 0.960 | 3625.9 | WD3 |
| 79 | NE | Others | *Canarium parvum* | 10.7 | 13.3 | 0.594 | 36.6 | WD2 |
| 80 | NE | Others | *Pterospermum heterophyllum* | 25.5 | 22.7 | 0.693 | 425.8 | WD3 |
| 81 | NE | Leguminosae | *Archidendron chevalieri* | 44.3 | 21.6 | 0.547 | 1007.7 | WD2 |
| 82 | NE | Lauraceae | *Cinnamomum sp.* | 19.7 | 19.2 | 0.743 | 286.9 | WD3 |
| 83 | NE | Leguminosae | *Archidendron chevalieri* | 27.8 | 17.2 | 0.598 | 355.3 | WD2 |
| 84 | NE | Leguminosae | *Archidendron chevalieri* | 15.9 | 13.0 | 0.455 | 74.7 | WD2 |
| 85 | NE | Lauraceae | *Cinnamomum sp.* | 14.9 | 14.2 | 0.748 | 119.9 | WD3 |
| 86 | NE | Lauraceae | *Cinnamomum sp.* | 35.0 | 23.2 | 0.571 | 717.3 | WD2 |
| 87 | NE | Myrtaceae | *Syzygium chanlos* | 37.4 | 24.5 | 0.762 | 958.2 | WD3 |
| 88 | NE | Fagaceae | *Castanopsis sp.* | 50.1 | 26.7 | 0.544 | 1273.8 | WD2 |
| 89 | NE | Others | *Semecarpus perniciosa* | 27.0 | 19.8 | 0.550 | 316.7 | WD2 |
| 90 | NE | Others | *Adinandra bockiana* | 29.7 | 24.6 | 0.738 | 767.4 | WD3 |
| 91 | NE | Others | *Adinandra glischroloma* | 6.3 | 7.2 | 0.598 | 11.7 | WD2 |
| 92 | NE | Others | *Symplocos laurina* | 8.9 | 10.9 | 0.833 | 33.6 | WD3 |
| 93 | NE | Fagaceae | *Castanopsis sp.* | 39.6 | 24.7 | 0.794 | 1265.6 | WD3 |
| 94 | NE | Fagaceae | *Castanopsis sp.* | 67.1 | 24.5 | 0.719 | 2432.3 | WD3 |
| 95 | NE | Others | *Pterospermum heterophyllum* | 11.8 | 15.0 | 0.637 | 52.5 | WD3 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 96 | NE | Leguminosae | *Archidendron chevalieri* | 37.1 | 18.5 | 0.519 | 640.6 | WD2 |
| 97 | NE | Leguminosae | *Archidendron chevalieri* | 40.4 | 19.5 | 0.477 | 865.5 | WD2 |
| 98 | NE | Lauraceae | *Actinodaphne pilosa* | 8.3 | 10.2 | 0.328 | 15.1 | WD1 |
| 99 | NE | Others | *Styrax tonkinensis* | 7.8 | 10.3 | 0.326 | 13.4 | WD1 |
| 100 | NE | Others | *Engelhardtia roxburghiana* | 56.6 | 17.8 | 0.690 | 1750.0 | WD3 |
| 101 | NE | Fagaceae | *Castanopsis sp.* | 70.1 | 27.5 | 0.760 | 5178.5 | WD3 |
| 102 | NE | Lauraceae | *Phoebe tovoyana* | 13.2 | 14.2 | 0.553 | 88.8 | WD2 |
| 103 | NE | Others | *Garcinia oblongifolia* | 11.7 | 11.0 | 0.647 | 34.6 | WD3 |
| 104 | NE | Lauraceae | *Cinnamomum parthenoxylon* | 49.8 | 23.7 | 0.524 | 1671.8 | WD2 |
| 105 | NE | Fagaceae | *Castanopsis sp.* | 54.0 | 23.2 | 0.912 | 2461.5 | WD3 |
| 106 | NE | Myrtaceae | *Syzygium chanlos* | 46.0 | 19.5 | 0.920 | 1002.7 | WD3 |
| 107 | NE | Leguminosae | *Archidendron chevalieri* | 28.3 | 13.3 | 0.482 | 232.6 | WD2 |
| 108 | NE | Dipterocarpaceae | *Hopea mollissima* | 21.2 | 19.2 | 0.960 | 344.4 | WD3 |
| 109 | NE | Dipterocarpaceae | *Hopea mollissima* | 6.8 | 10.2 | 0.870 | 22.8 | WD3 |
| 110 | NE | Dipterocarpaceae | *Hopea mollissima* | 28.7 | 19.8 | 0.885 | 682.6 | WD3 |
| 111 | SE | Others | *Lagerstroemia calyculata* | 8.4 | 12.5 | 0.517 | 23.0 | WD2 |
| 112 | SE | Dipterocarpaceae | *Dipterocarpus alatus* | 21.0 | 17.0 | 0.495 | 137.8 | WD2 |
| 113 | SE | Dipterocarpaceae | *Dipterocarpus alatus* | 22.6 | 18.0 | 0.477 | 168.2 | WD2 |
| 114 | SE | Others | *Adina polycephala* | 13.4 | 14.6 | 0.320 | 36.2 | WD1 |
| 115 | SE | Dipterocarpaceae | *Dipterocarpus alatus* | 25.3 | 19.5 | 0.487 | 229.4 | WD2 |
| 116 | SE | Dipterocarpaceae | *Dipterocarpus alatus* | 20.7 | 19.0 | 0.505 | 157.7 | WD2 |
| 117 | SE | Others | *Irvingia malayana* | 15.9 | 13.8 | 0.675 | 124.9 | WD3 |
| 118 | SE | Others | *Microcos paniculata* | 10.4 | 13.3 | 0.546 | 34.1 | WD2 |
| 119 | SE | Dipterocarpaceae | *Dipterocarpus alatus* | 11.1 | 10.0 | 0.471 | 35.6 | WD2 |
| 120 | SE | Dipterocarpaceae | *Dipterocarpus alatus* | 35.9 | 26.8 | 0.524 | 663.8 | WD2 |
| 121 | SE | Dipterocarpaceae | *Dipterocarpus alatus* | 32.9 | 21.5 | 0.556 | 560.2 | WD2 |
| 122 | SE | Dipterocarpaceae | *Dipterocarpus alatus* | 35.7 | 20.0 | 0.543 | 574.0 | WD2 |
| 123 | SE | Others | *Adina polycephala* | 34.9 | 30.0 | 0.340 | 536.2 | WD1 |
| 124 | SE | Leguminosae | *Antheroporum pierrei* | 48.2 | 32.3 | 0.681 | 2022.2 | WD3 |
| 125 | SE | Others | *Lagerstroemia calyculata* | 34.9 | 22.3 | 0.521 | 384.7 | WD2 |
| 126 | SE | Myrtaceae | *Syzygium sp.* | 58.2 | 30.0 | 0.491 | 1482.7 | WD2 |
| 127 | SE | Leguminosae | *Xylia xylocarpa* | 27.4 | 25.5 | 0.632 | 524.1 | WD3 |
| 128 | SE | Others | *Mangifera minutifolia* | 29.5 | 19.3 | 0.470 | 374.8 | WD2 |
| 129 | SE | Myrtaceae | *Syzygium jambos* | 36.0 | 20.4 | 0.530 | 620.8 | WD2 |
| 130 | SE | Others | *Diospyros maritima* | 17.5 | 17.8 | 0.629 | 187.9 | WD3 |
| 131 | SE | Leguminosae | *Xylia xylocarpa* | 22.5 | 20.9 | 0.620 | 312.7 | WD3 |
| 132 | SE | Dipterocarpaceae | *Dipterocarpus alatus* | 66.9 | 35.8 | 0.581 | 2911.9 | WD2 |
| 133 | SE | Others | *Lagerstroemia calyculata* | 12.9 | 14.0 | 0.559 | 53.0 | WD2 |
| 134 | SE | Others | *Cephalanthus tetrandra* | 60.1 | 34.3 | 0.357 | 2091.3 | WD1 |
| 135 | SE | Others | *Vitex ajugiflora* | 20.5 | 19.1 | 0.490 | 117.0 | WD2 |
| 136 | SE | Myrtaceae | *Syzygium jambos* | 65.0 | 28.0 | 0.590 | 2547.9 | WD2 |
| 137 | SE | Others | *NA* | 70.9 | 38.0 | 0.666 | 4330.3 | WD3 |
| 138 | SE | Others | *Lagerstroemia calyculata* | 41.5 | 28.3 | 0.555 | 786.7 | WD2 |
| 139 | SE | Others | *Terminalia triptera* | 36.3 | 32.4 | 0.588 | 803.1 | WD2 |
| 140 | SE | Others | *NA* | 55.1 | 25.8 | 0.575 | 1984.8 | WD2 |
| 141 | SE | Dipterocarpaceae | *Dipterocarpus alatus* | 50.1 | 32.2 | 0.594 | 1887.3 | WD2 |
| 142 | SE | Others | *Microcos paniculata* | 14.7 | 12.9 | 0.569 | 60.9 | WD2 |
| 143 | SE | Dipterocarpaceae | *Hopea odorata* | 39.2 | 23.6 | 0.554 | 800.9 | WD2 |
| 144 | SE | Others | *Lagerstroemia calyculata* | 24.0 | 20.8 | 0.553 | 176.8 | WD2 |
| 145 | SE | Leguminosae | *Xylia xylocarpa* | 20.4 | 22.1 | 0.632 | 278.8 | WD3 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 146 | SE | Dipterocarpaceae | *Hopea odorata* | 22.6 | 20.9 | 0.540 | 254.0 | WD2 |
| 147 | SE | Others | *Diospyros maritima* | 14.1 | 12.0 | 0.577 | 77.5 | WD2 |
| 148 | SE | Others | *Lagerstroemia calyculata* | 36.6 | 26.5 | 0.558 | 552.4 | WD2 |
| 149 | SE | Others | *Lagerstroemia calyculata* | 10.4 | 14.6 | 0.521 | 35.5 | WD2 |
| 150 | SE | Myrtaceae | *Syzygium jambos* | 48.4 | 26.5 | 0.562 | 1754.0 | WD2 |
| 151 | SE | Dipterocarpaceae | *Dipterocarpus alatus* | 52.5 | 27.0 | 0.519 | 1580.1 | WD2 |
| 152 | SE | Others | *Terminalia chebula* | 26.0 | 24.1 | 0.672 | 365.5 | WD3 |
| 153 | SE | Leguminosae | *Bauhinia sp.* | 18.8 | 15.3 | 0.532 | 147.2 | WD2 |
| 154 | SE | Others | *Microcos paniculata* | 20.4 | 18.6 | 0.562 | 162.6 | WD2 |
| 155 | SE | Others | *Spathodea campanulata* | 10.2 | 10.5 | 0.570 | 32.9 | WD2 |
| 156 | SE | Others | *Microcos paniculata* | 9.9 | 13.9 | 0.546 | 40.2 | WD2 |
| 157 | SE | Others | *Microcos paniculata* | 6.8 | 9.1 | 0.538 | 11.2 | WD2 |
| 158 | SE | Others | *Vitex ajugiflora* | 21.3 | 19.6 | 0.518 | 130.2 | WD2 |
| 159 | SE | Others | *Diospyros maritima* | 12.6 | 10.0 | 0.558 | 63.3 | WD2 |
| 160 | SE | Others | *Terminalia chebula* | 26.0 | 24.3 | 0.651 | 241.7 | WD3 |
| 161 | SE | Others | *Dillenia scabrella* | 11.5 | 11.5 | 0.564 | 42.0 | WD2 |
| 162 | SE | Leguminosae | *Xylia xylocarpa* | 33.9 | 27.0 | 0.631 | 738.7 | WD3 |
| 163 | SE | Dipterocarpaceae | *Dipterocarpus alatus* | 13.5 | 11.7 | 0.505 | 49.1 | WD2 |
| 164 | SE | Others | *Adina polycephala* | 23.0 | 22.5 | 0.390 | 162.4 | WD1 |
| 165 | SE | Others | *Mangifera minutifolia* | 21.4 | 15.0 | 0.495 | 163.1 | WD2 |
| 166 | SE | Others | *Mischocarpus pentapetalus* | 15.4 | 16.2 | 0.512 | 83.1 | WD2 |
| 167 | SE | Leguminosae | *Dalbergia oliveri* | 9.2 | 8.5 | 0.631 | 18.8 | WD3 |
| 168 | SE | Others | *Lagerstroemia calyculata* | 13.7 | 17.3 | 0.529 | 70.2 | WD2 |
| 169 | SE | Dipterocarpaceae | *Dipterocarpus alatus* | 42.5 | 26.5 | 0.576 | 1153.0 | WD2 |
| 170 | SE | Dipterocarpaceae | *Anisoptera costata* | 41.2 | 26.8 | 0.590 | 912.5 | WD2 |
| 171 | SE | Leguminosae | *Xylia xylocarpa* | 51.7 | 27.5 | 0.702 | 2590.5 | WD3 |
| 172 | SE | Dipterocarpaceae | *Dipterocarpus alatus* | 72.5 | 37.5 | 0.587 | 3313.8 | WD2 |
| 173 | SE | Leguminosae | *Peltophorum pterocarpum* | 63.7 | 31.2 | 0.395 | 1664.4 | WD1 |
| 174 | SE | Others | *Adina polycephala* | 35.9 | 23.6 | 0.390 | 603.3 | WD1 |
| 175 | SE | Dipterocarpaceae | *Dipterocarpus alatus* | 66.5 | 31.5 | 0.525 | 2245.9 | WD2 |
| 176 | SE | Leguminosae | *Antheroporum pierrei* | 39.8 | 27.3 | 0.648 | 952.3 | WD3 |
| 177 | SE | Others | *Careya arborea* | 15.0 | 11.3 | 0.491 | 64.4 | WD2 |
| 178 | SE | Others | *Lagerstroemia crispa* | 9.8 | 11.9 | 0.623 | 30.7 | WD3 |
| 179 | SE | Leguminosae | *Dalbergia oliveri* | 8.1 | 6.8 | 0.629 | 13.1 | WD3 |
| 180 | SE | Leguminosae | *Dalbergia oliveri* | 11.1 | 10.5 | 0.660 | 31.7 | WD3 |
| 181 | SE | Others | *Terminalia chebula* | 15.2 | 18.0 | 0.594 | 127.7 | WD2 |
| 182 | SE | Others | *Cratoxylum formosum* | 15.4 | 16.9 | 0.599 | 90.6 | WD2 |
| 183 | SE | Dipterocarpaceae | *Dipterocarpus obtusifolius* | 30.6 | 20.2 | 0.654 | 602.2 | WD3 |
| 184 | SE | Others | *Lagerstroemia calyculata* | 12.3 | 13.1 | 0.531 | 39.9 | WD2 |
| 185 | SE | Others | *Diospyros decandra* | 11.6 | 15.1 | 0.693 | 67.7 | WD3 |
| 186 | SE | Others | *NA* | 19.4 | 19.5 | 0.620 | 206.7 | WD3 |
| 187 | SE | Others | *Lagerstroemia calyculata* | 16.0 | 16.1 | 0.552 | 62.7 | WD2 |
| 188 | SE | Others | *Vitex ajugiflora* | 24.9 | 21.5 | 0.567 | 287.5 | WD2 |
| 189 | SE | Others | *Cephalanthus tetrandra* | 24.4 | 20.5 | 0.395 | 217.4 | WD1 |
| 190 | SE | Others | *Lagerstroemia crispa* | 22.3 | 21.0 | 0.623 | 309.0 | WD3 |
| 191 | SE | Others | *Adina pilulifera* | 47.8 | 25.4 | 0.571 | 1048.8 | WD2 |
| 192 | SE | Others | *Adina pilulifera* | 30.9 | 23.2 | 0.591 | 503.3 | WD2 |
| 193 | SE | Others | *Lagerstroemia calyculata* | 31.5 | 17.6 | 0.528 | 277.0 | WD2 |
| 194 | SE | Dipterocarpaceae | *Dipterocarpus obtusifolius* | 28.6 | 21.5 | 0.650 | 514.4 | WD3 |
| 195 | SE | Dipterocarpaceae | *Shorea siamensis* | 55.7 | 21.8 | 0.586 | 1924.3 | WD2 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 196 | SE | Dipterocarpaceae | *Shorea roxburghii* | 68.5 | 28.5 | 0.570 | 3582.1 | WD2 |
| 197 | SE | Dipterocarpaceae | *Shorea roxburghii* | 45.2 | 21.8 | 0.598 | 1159.8 | WD2 |
| 198 | SE | Dipterocarpaceae | *Dipterocarpus alatus* | 47.6 | 29.0 | 0.595 | 1245.0 | WD2 |
| 199 | SE | Dipterocarpaceae | *Shorea roxburghii* | 30.6 | 22.0 | 0.571 | 521.8 | WD2 |
| 200 | SE | Others | *Irvingia malayana* | 45.6 | 20.4 | 0.700 | 889.3 | WD3 |
| 201 | SE | Dipterocarpaceae | *Dipterocarpus obtusifolius* | 39.0 | 26.7 | 0.667 | 1093.0 | WD3 |
| 202 | SE | Others | *Lagerstroemia calyculata* | 30.4 | 21.4 | 0.524 | 311.6 | WD2 |
| 203 | SE | Others | *Lagerstroemia crispa* | 32.9 | 18.7 | 0.641 | 522.8 | WD3 |
| 204 | SE | Others | *Terminalia chebula* | 36.5 | 23.5 | 0.638 | 583.5 | WD3 |
| 205 | SE | Leguminosae | *Xylia xylocarpa* | 28.8 | 20.8 | 0.669 | 402.7 | WD3 |
| 206 | SE | Leguminosae | *Xylia xylocarpa* | 26.1 | 23.0 | 0.655 | 489.9 | WD3 |
| 207 | SE | Others | *Terminalia chebula* | 25.3 | 20.5 | 0.651 | 296.2 | WD3 |
| 208 | SE | Others | *Terminalia chebula* | 9.9 | 15.5 | 0.627 | 39.5 | WD3 |
| 209 | SE | Dipterocarpaceae | *Shorea roxburghii* | 27.8 | 19.0 | 0.599 | 409.4 | WD2 |
| 210 | SE | Leguminosae | *Pterocarpus macrocarpus* | 16.9 | 15.6 | 0.581 | 103.7 | WD2 |
| 211 | SE | Others | *Vitex ajugiflora* | 9.4 | 13.0 | 0.540 | 25.6 | WD2 |
| 212 | SE | Others | *Irvingia malayana* | 7.5 | 10.1 | 0.674 | 20.3 | WD3 |
| 213 | SE | Others | *Lagerstroemia calyculata* | 8.9 | 12.3 | 0.518 | 18.5 | WD2 |
| 214 | SE | Leguminosae | *Pterocarpus macrocarpus* | 11.9 | 13.8 | 0.604 | 52.5 | WD3 |
| 215 | SE | Others | *Microcos paniculata* | 7.2 | 10.0 | 0.535 | 18.0 | WD2 |
| 216 | SE | Others | *Careya arborea* | 11.7 | 10.7 | 0.508 | 31.1 | WD2 |
| 217 | SE | Leguminosae | *Pterocarpus macrocarpus* | 12.9 | 15.5 | 0.538 | 57.0 | WD2 |
| 218 | SE | Leguminosae | *Xylia xylocarpa* | 15.6 | 16.3 | 0.661 | 111.3 | WD3 |
| 219 | SE | Leguminosae | *Xylia xylocarpa* | 52.3 | 25.7 | 0.714 | 1690.0 | WD3 |
| 220 | SE | Others | *Mangifera minutifolia* | 22.3 | 16.7 | 0.464 | 193.7 | WD2 |
| 221 | NCC | Others | *Manglietia dandyi* | 25.0 | 14.5 | 0.429 | 121.7 | WD2 |
| 222 | NCC | Fagaceae | *Lithocarpus pseudosundaicus* | 5.6 | 8.4 | 0.465 | 9.0 | WD2 |
| 223 | NCC | Others | *Ficus sp.* | 38.8 | 16.2 | 0.453 | 390.9 | WD2 |
| 224 | NCC | Leguminosae | *Erythrophleum fordii* | 23.0 | 19.4 | 0.687 | 389.8 | WD3 |
| 225 | NCC | Fagaceae | *Castanopsis tesselata* | 32.3 | 20.5 | 0.380 | 321.2 | WD1 |
| 226 | NCC | Others | *Canarium tramdenum* | 40.3 | 21.5 | 0.588 | 858.7 | WD2 |
| 227 | NCC | Leguminosae | *Archidendron balansae* | 35.0 | 16.0 | 0.431 | 237.8 | WD2 |
| 228 | NCC | Leguminosae | *Archidendron balansae* | 53.3 | 21.4 | 0.493 | 741.0 | WD2 |
| 229 | NCC | Lauraceae | *Litsea sp.* | 27.6 | 15.4 | 0.406 | 207.6 | WD2 |
| 230 | NCC | Others | *Polyalthia sp.* | 45.5 | 22.0 | 0.392 | 736.1 | WD1 |
| 231 | NCC | Leguminosae | *Aidia pycnantha* | 7.0 | 7.6 | 0.609 | 16.3 | WD3 |
| 232 | NCC | Others | *Canarium tramdenum* | 35.7 | 16.6 | 0.643 | 625.5 | WD3 |
| 233 | NCC | Euphorbiaceae | *Sapium sebiferum* | 25.5 | 17.1 | 0.342 | 207.5 | WD1 |
| 234 | NCC | Lauraceae | *Cinnadenia paniculata* | 28.2 | 19.2 | 0.592 | 333.5 | WD2 |
| 235 | NCC | Leguminosae | *Pithecolobium acumiratum* | 23.1 | 15.2 | 0.367 | 136.5 | WD1 |
| 236 | NCC | Leguminosae | *Ormosia balansae* | 8.1 | 9.7 | 0.451 | 11.5 | WD2 |
| 237 | NCC | Dipterocarpaceae | *Vatica odorata* | 5.0 | 7.2 | 0.702 | 9.7 | WD3 |
| 238 | NCC | Fagaceae | *Castanopsis tesselata* | 11.8 | 11.5 | 0.384 | 38.7 | WD1 |
| 239 | NCC | Dipterocarpaceae | *Vatica odorata* | 6.3 | 9.0 | 0.672 | 15.1 | WD3 |
| 240 | NCC | Others | *Helicia cochinchinensis* | 22.0 | 13.8 | 0.529 | 137.6 | WD2 |
| 241 | NCC | Others | *Engelhardtia roxburghiana* | 29.4 | 19.8 | 0.545 | 350.1 | WD2 |
| 242 | NCC | Others | *Camelia sp.* | 9.6 | 11.4 | 0.434 | 23.4 | WD2 |
| 243 | NCC | Meliaceae | *Dysoxylum binectariferum* | 22.3 | 15.7 | 0.406 | 140.3 | WD2 |
| 244 | NCC | Leguminosae | *Ormosia balansae* | 10.3 | 12.3 | 0.439 | 35.2 | WD2 |
| 245 | NCC | Ulmaceae | *Gironniera subaequalis* | 19.7 | 12.6 | 0.437 | 108.7 | WD2 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 246 | NCC | Others | *Prunus arborea* | 24.0 | 15.9 | 0.427 | 188.1 | WD2 |
| 247 | NCC | Euphorbiaceae | *Endospermum chinense* | 22.9 | 16.4 | 0.383 | 136.5 | WD1 |
| 248 | NCC | Others | *Elaeocarpus griffithii* | 19.3 | 11.6 | 0.494 | 113.2 | WD2 |
| 249 | NCC | Leguminosae | *Pithecolobium acumiratum* | 30.6 | 15.7 | 0.448 | 294.6 | WD2 |
| 250 | NCC | Fagaceae | *Castanopsis chinensis* | 22.1 | 14.3 | 0.388 | 167.3 | WD1 |
| 251 | NCC | Lauraceae | *Litsea sp.* | 20.6 | 14.5 | 0.529 | 147.3 | WD2 |
| 252 | NCC | Leguminosae | *Pithecolobium acumiratum* | 22.6 | 14.6 | 0.437 | 109.8 | WD2 |
| 253 | NCC | Others | *Ficus sp.* | 35.2 | 18.1 | 0.417 | 351.2 | WD2 |
| 254 | NCC | Fagaceae | *Castanopsis chinensis* | 19.3 | 14.5 | 0.527 | 143.1 | WD2 |
| 255 | NCC | Ulmaceae | *Gironniera subaequalis* | 7.9 | 8.5 | 0.384 | 13.9 | WD1 |
| 256 | NCC | Others | *Knema conferta* | 11.6 | 9.9 | 0.419 | 34.0 | WD2 |
| 257 | NCC | Leguminosae | *Pithecolobium acumiratum* | 7.9 | 10.0 | 0.450 | 11.4 | WD2 |
| 258 | NCC | Others | *Dracontomelon duperreanum* | 18.7 | 12.1 | 0.462 | 84.5 | WD2 |
| 259 | NCC | Ulmaceae | *Gironniera subaequalis* | 18.3 | 12.7 | 0.411 | 72.0 | WD2 |
| 260 | NCC | Fagaceae | *Castanopsis tesselata* | 25.8 | 14.8 | 0.390 | 221.7 | WD1 |
| 261 | NCC | Lauraceae | *Cinnadenia paniculata* | 14.5 | 13.3 | 0.424 | 51.8 | WD2 |
| 262 | NCC | Ulmaceae | *Gironniera subaequalis* | 12.1 | 10.6 | 0.500 | 37.1 | WD2 |
| 263 | NCC | Leguminosae | *Erythrophleum fordii* | 12.6 | 11.9 | 0.594 | 68.8 | WD2 |
| 264 | NCC | Dipterocarpaceae | *Vatica odorata* | 6.0 | 8.2 | 0.703 | 12.5 | WD3 |
| 265 | NCC | Others | *Symplocos sp.* | 10.1 | 10.0 | 0.366 | 21.4 | WD1 |
| 266 | NCC | Dipterocarpaceae | *Vatica odorata* | 7.1 | 7.8 | 0.640 | 15.4 | WD3 |
| 267 | NCC | Euphorbiaceae | *Mallotus macrostachyus* | 9.5 | 10.6 | 0.450 | 22.7 | WD2 |
| 268 | NCC | Leguminosae | *Pithecolobium acumiratum* | 10.3 | 13.5 | 0.482 | 36.6 | WD2 |
| 269 | NCC | Myrtaceae | *Syzygium jambos* | 18.2 | 14.8 | 0.577 | 116.3 | WD2 |
| 270 | NCC | Others | *Goniothalamus macrocalyx* | 5.9 | 7.6 | 0.602 | 15.2 | WD3 |
| 271 | NCC | Dipterocarpaceae | *Vatica odorata* | 52.9 | 24.8 | 0.760 | 2213.1 | WD3 |
| 272 | NCC | Dipterocarpaceae | *Vatica odorata* | 70.0 | 28.7 | 0.648 | 4605.2 | WD3 |
| 273 | NCC | Leguminosae | *Archidendron eberhardtia* | 28.7 | 22.8 | 0.526 | 458.8 | WD2 |
| 274 | NCC | Lauraceae | *Cryptocarya lenticellata* | 10.7 | 8.2 | 0.487 | 26.1 | WD2 |
| 275 | NCC | Others | *Engelhardtia roxburghiana* | 39.2 | 22.6 | 0.504 | 784.2 | WD2 |
| 276 | NCC | Others | *Elaeocarpus griffithii* | 10.4 | 7.1 | 0.529 | 35.8 | WD2 |
| 277 | NCC | Others | *Knema conferta* | 11.2 | 9.5 | 0.575 | 36.9 | WD2 |
| 278 | NCC | Others | *Engelhardtia roxburghiana* | 38.6 | 22.1 | 0.546 | 504.8 | WD2 |
| 279 | NCC | Ulmaceae | *Gironniera subaequalis* | 27.9 | 13.5 | 0.426 | 221.4 | WD2 |
| 280 | NCC | Others | *Symplocos laurina* | 22.6 | 17.8 | 0.469 | 248.9 | WD2 |
| 281 | NCC | Fagaceae | *Castanopsis hystrix* | 59.0 | 25.4 | 0.468 | 2436.3 | WD2 |
| 282 | NCC | Others | *Elaeocarpus griffithii* | 34.2 | 19.3 | 0.466 | 490.4 | WD2 |
| 283 | NCC | Lauraceae | *Actinodaphne ellipticibacca* | 17.0 | 15.8 | 0.380 | 77.0 | WD1 |
| 284 | NCC | Others | *NA* | 14.8 | 17.5 | 0.432 | 78.4 | WD2 |
| 285 | NCC | Others | *Canarium tramdenum* | 17.2 | 17.3 | 0.452 | 125.3 | WD2 |
| 286 | NCC | Others | *Engelhardtia roxburghiana* | 16.5 | 18.8 | 0.444 | 109.0 | WD2 |
| 287 | NCC | Leguminosae | *Peltophorum pterocarpum* | 56.5 | 28.2 | 0.553 | 2759.2 | WD2 |
| 288 | NCC | Others | *Rubus parvifolius* | 15.7 | 14.8 | 0.608 | 86.1 | WD3 |
| 289 | NCC | Others | *Alangium ridleyi* | 16.7 | 14.8 | 0.618 | 147.9 | WD3 |
| 290 | NCC | Fagaceae | *Castanopsis hystrix* | 37.4 | 27.3 | 0.463 | 966.5 | WD2 |
| 291 | NCC | Others | *Eberhardtia tonkinensis* | 9.4 | 12.2 | 0.375 | 18.3 | WD1 |
| 292 | NCC | Others | *Engelhardtia roxburghiana* | 33.0 | 20.3 | 0.573 | 730.5 | WD2 |
| 293 | NCC | Others | *NA* | 19.6 | 15.4 | 0.591 | 215.4 | WD2 |
| 294 | NCC | Others | *Garcinia oblongifolia* | 7.3 | 7.8 | 0.316 | 14.2 | WD1 |
| 295 | NCC | Others | *Engelhardtia roxburghiana* | 7.3 | 11.0 | 0.526 | 11.7 | WD2 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 296 | NCC | Others | *Garcinia oblongifolia* | 24.1 | 13.7 | 0.508 | 216.1 | WD2 |
| 297 | NCC | Ulmaceae | *Gironniera subaequalis* | 13.2 | 12.2 | 0.449 | 50.7 | WD2 |
| 298 | NCC | Others | *Elaeocarpus griffithii* | 56.2 | 22.6 | 0.567 | 2313.2 | WD2 |
| 299 | NCC | Dipterocarpaceae | *Vatica odorata* | 59.4 | 26.1 | 0.648 | 2972.5 | WD3 |
| 300 | NCC | Dipterocarpaceae | *Vatica odorata* | 41.0 | 23.6 | 0.742 | 1513.3 | WD3 |
| 301 | NCC | Dipterocarpaceae | *Vatica odorata* | 23.1 | 16.1 | 0.666 | 312.0 | WD3 |
| 302 | NCC | Dipterocarpaceae | *Vatica odorata* | 31.4 | 24.9 | 0.731 | 871.5 | WD3 |
| 303 | NCC | Others | *Eberhardtia tonkinensis* | 18.2 | 13.3 | 0.376 | 89.8 | WD1 |
| 304 | NCC | Ulmaceae | *Gironniera subaequalis* | 18.6 | 15.8 | 0.425 | 131.7 | WD2 |
| 305 | NCC | Fagaceae | *Castanopsis chinensis* | 14.6 | 14.9 | 0.546 | 123.6 | WD2 |
| 306 | NCC | Others | *Canarium tramdenum* | 14.5 | 12.4 | 0.626 | 85.5 | WD3 |
| 307 | NCC | Others | *Canarium tramdenum* | 33.8 | 25.9 | 0.453 | 639.1 | WD2 |
| 308 | NCC | Dipterocarpaceae | *Vatica odorata* | 49.1 | 27.6 | 0.727 | 1688.6 | WD3 |
| 309 | NCC | Others | *Engelhardtia roxburghiana* | 53.5 | 27.5 | 0.552 | 1878.9 | WD2 |
| 310 | NCC | Others | *Alangium ridleyi* | 29.0 | 18.2 | 0.630 | 792.3 | WD3 |
| 311 | NCC | Others | *NA* | 46.9 | 16.4 | 0.555 | 1195.4 | WD2 |
| 312 | NCC | Dipterocarpaceae | *Vatica odorata* | 38.3 | 25.6 | 0.707 | 1168.9 | WD3 |
| 313 | NCC | Dipterocarpaceae | *Vatica odorata* | 15.8 | 15.9 | 0.593 | 127.6 | WD2 |
| 314 | NCC | Dipterocarpaceae | *Vatica odorata* | 71.4 | 36.2 | 0.638 | 5659.5 | WD3 |
| 315 | NCC | Fagaceae | *Castanopsis hystrix* | 48.5 | 27.9 | 0.589 | 1829.7 | WD2 |
| 316 | NCC | Others | *Schima superba* | 22.4 | 23.7 | 0.487 | 215.5 | WD2 |
| 317 | NCC | Dipterocarpaceae | *Vatica odorata* | 54.0 | 29.8 | 0.750 | 2356.2 | WD3 |
| 318 | NCC | Dipterocarpaceae | *Vatica odorata* | 65.4 | 31.8 | 0.676 | 3421.6 | WD3 |
| 319 | NCC | Others | *Engelhardtia roxburghiana* | 39.1 | 19.3 | 0.538 | 511.4 | WD2 |
| 320 | NCC | Others | *Nephelium cuspidatum* | 63.3 | 26.6 | 0.748 | 4086.6 | WD3 |
| 321 | NCC | Ulmaceae | *Gironniera subaequalis* | 39.7 | 25.0 | 0.460 | 673.1 | WD2 |
| 322 | NCC | Fagaceae | *Castanopsis chinensis* | 18.4 | 17.0 | 0.407 | 135.5 | WD2 |
| 323 | NCC | Others | *Garcinia oblongifolia* | 41.8 | 25.8 | 0.480 | 709.8 | WD2 |
| 324 | NCC | Others | *Eberhardtia tonkinensis* | 25.3 | 23.6 | 0.340 | 186.9 | WD1 |
| 325 | NCC | Others | *Alangium ridleyi* | 32.9 | 17.5 | 0.647 | 926.5 | WD3 |
| 326 | NCC | Lauraceae | *Endiandra hainanensis* | 27.2 | 22.4 | 0.429 | 364.5 | WD2 |
| 327 | NCC | Lauraceae | *Actinodaphne pilosa* | 31.6 | 23.0 | 0.425 | 375.6 | WD2 |
| 328 | NCC | Leguminosae | *Archidendron eberhardtia* | 33.2 | 21.5 | 0.517 | 524.8 | WD2 |
| 329 | NCC | Others | *Garcinia oblongifolia* | 24.5 | 17.7 | 0.542 | 260.9 | WD2 |
| 330 | NCC | Others | *Alangium ridleyi* | 46.3 | 25.6 | 0.621 | 1614.9 | WD3 |
| 331 | NCC | Others | *NA* | 31.9 | 15.4 | 0.505 | 454.9 | WD2 |
| 332 | NCC | Dipterocarpaceae | *Vatica odorata* | 9.0 | 10.9 | 0.679 | 39.5 | WD3 |
| 333 | NCC | Leguminosae | *Archidendron eberhardtia* | 26.1 | 23.1 | 0.497 | 458.3 | WD2 |
| 334 | NCC | Leguminosae | *Antheroporum pierrei* | 54.3 | 33.6 | 0.474 | 2657.2 | WD2 |
| 335 | NCC | Euphorbiaceae | *Endospermum chinense* | 37.1 | 27.2 | 0.434 | 625.2 | WD2 |
| 336 | NCC | Leguminosae | *Antheroporum pierrei* | 55.1 | 25.5 | 0.468 | 1713.9 | WD2 |
| 337 | NCC | Leguminosae | *Antheroporum pierrei* | 8.1 | 9.4 | 0.431 | 14.8 | WD2 |
| 338 | NCC | Dipterocarpaceae | *Vatica odorata* | 36.3 | 24.8 | 0.703 | 1399.5 | WD3 |
| 339 | NCC | Fagaceae | *Castanopsis hystrix* | 56.9 | 29.2 | 0.574 | 2099.7 | WD2 |
| 340 | NCC | Others | *Garcinia oblongifolia* | 11.0 | 12.7 | 0.510 | 41.4 | WD2 |
| 341 | NCC | Dipterocarpaceae | *Vatica odorata* | 30.6 | 21.3 | 0.675 | 868.6 | WD3 |
| 342 | NCC | Others | *NA* | 39.6 | 17.8 | 0.548 | 656.2 | WD2 |
| 343 | NCC | Ulmaceae | *Gironniera subaequalis* | 33.6 | 17.5 | 0.432 | 466.8 | WD2 |
| 344 | NCC | Fagaceae | *Castanopsis chinensis* | 26.0 | 20.8 | 0.572 | 519.8 | WD2 |
| 345 | NCC | Others | *Alangium barbatum* | 20.9 | 19.8 | 0.391 | 134.3 | WD1 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 346 | NCC | Others | *Rubus parvifolius* | 18.1 | 14.1 | 0.599 | 123.5 | WD2 |
| 347 | NCC | Dipterocarpaceae | *Vatica odorata* | 11.6 | 10.9 | 0.652 | 69.0 | WD3 |
| 348 | NCC | Others | *Alangium ridleyi* | 45.8 | 23.3 | 0.622 | 1528.4 | WD3 |
| 349 | NCC | Lauraceae | *Endiandra hainanensis* | 30.1 | 24.9 | 0.519 | 520.9 | WD2 |
| 350 | NCC | Dipterocarpaceae | *Vatica odorata* | 38.1 | 23.3 | 0.729 | 1628.4 | WD3 |
| 351 | NCC | Leguminosae | *Erythrophleum fordii* | 11.0 | 16.1 | 0.640 | 62.3 | WD3 |
| 352 | NCC | Euphorbiaceae | *Endospermum chinense* | 38.1 | 28.4 | 0.386 | 712.5 | WD1 |
| 353 | NCC | Others | *Elaeocarpus griffithii* | 10.8 | 13.4 | 0.450 | 54.4 | WD2 |
| 354 | NCC | Lauraceae | *Cryptocarya lenticellata* | 9.8 | 8.9 | 0.588 | 23.2 | WD2 |
| 355 | NCC | Euphorbiaceae | *Aleurites montana* | 39.5 | 23.6 | 0.396 | 558.3 | WD1 |
| 356 | NCC | Others | *Alangium ridleyi* | 40.6 | 21.1 | 0.644 | 837.5 | WD3 |
| 357 | NCC | Others | *Schima superba* | 60.0 | 32.0 | 0.537 | 2676.2 | WD2 |
| 358 | NCC | Others | *Alangium ridleyi* | 19.6 | 13.7 | 0.658 | 249.8 | WD3 |
| 359 | NCC | Others | *Garcinia oblongifolia* | 15.3 | 15.0 | 0.483 | 86.2 | WD2 |
| 360 | NCC | Others | *Nephelium cuspidatum* | 21.2 | 20.4 | 0.630 | 273.6 | WD3 |
| 361 | NCC | Others | *Manglietia dandyi* | 22.2 | 15.2 | 0.370 | 142.1 | WD1 |
| 362 | NCC | Others | *Artocarpus rigidus* | 22.6 | 20.0 | 0.442 | 189.6 | WD2 |
| 363 | NCC | Leguminosae | *Antheroporum pierrei* | 65.3 | 32.6 | 0.578 | 3796.2 | WD2 |
| 364 | NCC | Dipterocarpaceae | *Vatica odorata* | 67.1 | 24.9 | 0.740 | 3148.1 | WD3 |
| 365 | NCC | Others | *Nephelium cuspidatum* | 66.1 | 26.4 | 0.647 | 2579.2 | WD3 |
| 366 | NCC | Others | *Elaeocarpus griffithii* | 66.0 | 32.9 | 0.527 | 2169.9 | WD2 |
| 367 | NCC | Others | *Oroxylum indicum* | 15.9 | 15.9 | 0.364 | 62.8 | WD1 |
| 368 | NCC | Others | *Knema conferta* | 7.9 | 11.2 | 0.367 | 12.1 | WD1 |
| 369 | NCC | Others | *NA* | 27.0 | 18.0 | 0.516 | 271.0 | WD2 |
| 370 | NCC | Fagaceae | *Castanopsis hystrix* | 59.4 | 26.4 | 0.511 | 2193.1 | WD2 |
| 371 | NCC | Others | *Elaeocarpus griffithii* | 18.6 | 15.3 | 0.363 | 121.6 | WD1 |
| 372 | SCC | Others | *Canarium littorale Bl.* | 14.6 | 12.0 | 0.588 | 67.7 | WD2 |
| 373 | SCC | Others | *Sterculia parviflora Roxb.* | 12.0 | 13.9 | 0.533 | 57.1 | WD2 |
| 374 | SCC | Others | *Ilex annamensis Tard* | 6.9 | 7.6 | 0.581 | 11.8 | WD2 |
| 375 | SCC | Others | *Dillenia indica L.* | 11.0 | 11.5 | 0.536 | 34.0 | WD2 |
| 376 | SCC | Others | *Diospyros decandra* | 14.5 | 17.4 | 0.664 | 120.3 | WD3 |
| 377 | SCC | Others | *Scaphium lychnophorum (Hance) Kosterm.* | 6.2 | 9.5 | 0.591 | 9.7 | WD2 |
| 378 | SCC | Others | *Gardenia philastrei Pierre ex Pit.* | 10.6 | 12.9 | 0.566 | 54.3 | WD2 |
| 379 | SCC | Others | *Canarium littorale Bl.* | 5.6 | 12.2 | 0.620 | 7.7 | WD3 |
| 380 | SCC | Others | *Polyalthia nemoralis A. Dc.* | 10.2 | 10.9 | 0.586 | 33.2 | WD2 |
| 381 | SCC | Others | *Scaphium lychnophorum (Hance) Kosterm.* | 7.2 | 10.4 | 0.638 | 13.5 | WD3 |
| 382 | SCC | Others | *Scaphium lychnophorum (Hance) Kosterm.* | 10.2 | 12.1 | 0.568 | 35.4 | WD2 |
| 383 | SCC | Others | *Dillenia indica L.* | 9.6 | 9.3 | 0.552 | 27.9 | WD2 |
| 384 | SCC | Others | *Scaphium lychnophorum (Hance) Kosterm.* | 8.5 | 10.3 | 0.619 | 25.2 | WD3 |
| 385 | SCC | Fagaceae | *Lithocarpus annamensis A. Camus.* | 11.7 | 14.3 | 0.585 | 44.2 | WD2 |
| 386 | SCC | Myrtaceae | *Syzygium levinei Merr. Et Perry.* | 7.0 | 9.4 | 0.536 | 12.0 | WD2 |
| 387 | SCC | Others | *Barringtonia racenmosa (L.) Spreng* | 9.1 | 7.2 | 0.494 | 19.8 | WD2 |
| 388 | SCC | Meliaceae | *Aglaia roxburghiana Miq.* | 6.8 | 8.0 | 0.659 | 18.0 | WD3 |
| 389 | SCC | Others | *Scaphium lychnophorum (Hance) Kosterm.* | 8.8 | 11.0 | 0.561 | 21.1 | WD2 |
| 390 | SCC | Myrtaceae | *Syzygium levinei Merr. Et Perry.* | 16.4 | 18.4 | 0.567 | 100.4 | WD2 |
| 391 | SCC | Others | *Prunus ceylanica (Wight.) Miq.* | 23.4 | 18.0 | 0.589 | 238.4 | WD2 |
| 392 | SCC | Others | *Scaphium lychnophorum (Hance) Kosterm.* | 21.7 | 23.3 | 0.589 | 248.5 | WD2 |
| 393 | SCC | Others | *Nauclea orientalis L.* | 16.9 | 17.0 | 0.430 | 77.5 | WD2 |
| 394 | SCC | Others | *Elaeocarpus kontumensis Gagn.* | 12.1 | 11.5 | 0.582 | 47.8 | WD2 |
| 395 | SCC | Others | *Knema pierre Warb.* | 15.7 | 14.2 | 0.595 | 80.7 | WD2 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 396 | SCC | Others | *Scaphium lychnophorum (Hance) Kosterm.* | 18.8 | 20.5 | 0.591 | 175.0 | WD2 |
| 397 | SCC | Others | *Vitex plerreana P. Dop.* | 16.0 | 14.3 | 0.524 | 100.1 | WD2 |
| 398 | SCC | Myrtaceae | *Syzygium levinei Merr. Et Perry.* | 24.2 | 22.9 | 0.595 | 283.5 | WD2 |
| 399 | SCC | Others | *Knema pierre Warb.* | 17.2 | 16.3 | 0.591 | 136.1 | WD2 |
| 400 | SCC | Ulmaceae | *Gironiera subaequalis Planch.* | 19.9 | 13.5 | 0.506 | 128.5 | WD2 |
| 401 | SCC | Others | *Camelia fleuryi (Pit.) Sealy* | 32.3 | 21.7 | 0.505 | 610.0 | WD2 |
| 402 | SCC | Euphorbiaceae | *Sapium baccatum Roxb.* | 35.9 | 20.2 | 0.544 | 450.4 | WD2 |
| 403 | SCC | Others | *Garcinia handburyi Hook.F* | 25.9 | 16.9 | 0.691 | 411.0 | WD3 |
| 404 | SCC | Others | *Garcinia oliveri Pierre.* | 30.2 | 17.5 | 0.712 | 644.3 | WD3 |
| 405 | SCC | Others | *Barringtonia racenmosa (L.) Spreng* | 11.4 | 9.3 | 0.541 | 39.9 | WD2 |
| 406 | SCC | Dipterocarpaceae | *Shorea farinosa* | 31.3 | 31.2 | 0.637 | 853.7 | WD3 |
| 407 | SCC | Others | *Acronychia oligophlebia Merr* | 49.5 | 22.8 | 0.495 | 968.2 | WD2 |
| 408 | SCC | Others | *Acronychia oligophlebia Merr* | 36.6 | 22.2 | 0.518 | 742.6 | WD2 |
| 409 | SCC | Dipterocarpaceae | *Shorea farinosa* | 42.8 | 31.5 | 0.630 | 1903.4 | WD3 |
| 410 | SCC | Fagaceae | *Lithocarpus annamensis A. Camus.* | 54.7 | 29.2 | 0.566 | 2960.7 | WD2 |
| 411 | SCC | Others | *Magnolia braianensis Gagnep.* | 54.1 | 33.5 | 0.634 | 1988.4 | WD3 |
| 412 | SCC | Others | *Sterculia parviflora Roxb.* | 49.7 | 34.4 | 0.644 | 1651.8 | WD3 |
| 413 | SCC | Others | *Elaeocarpus kontumensis Gagn.* | 55.5 | 27.6 | 0.598 | 747.1 | WD2 |
| 414 | SCC | Others | *Canarium littorale Bl.* | 55.0 | 32.7 | 0.610 | 1800.5 | WD3 |
| 415 | SCC | Lauraceae | *Litsea elliptica* | 57.8 | 31.5 | 0.582 | 2345.4 | WD2 |
| 416 | SCC | Myrtaceae | *Syzygium levinei Merr. Et Perry.* | 13.6 | 13.5 | 0.603 | 53.4 | WD3 |
| 417 | SCC | Others | *Canarium littorale Bl.* | 65.0 | 32.5 | 0.640 | 3687.3 | WD3 |
| 418 | SCC | Others | *Calophyllum dryobalanoides Pierre* | 68.4 | 26.4 | 0.567 | 3894.5 | WD2 |
| 419 | SCC | Meliaceae | *Melia azedarach L.* | 66.5 | 35.0 | 0.502 | 2785.9 | WD2 |
| 420 | SCC | Others | *Canarium littorale Bl.* | 82.4 | 33.5 | 0.644 | 6349.5 | WD3 |
| 421 | SCC | Euphorbiaceae | *Sapium baccatum Roxb.* | 79.0 | 34.5 | 0.575 | 3954.1 | WD2 |
| 422 | SCC | Dipterocarpaceae | *Shorea farinosa* | 87.7 | 41.4 | 0.663 | 8633.0 | WD3 |
| 423 | SCC | Others | *Polyalthia nemoralis A. Dc.* | 6.5 | 6.2 | 0.593 | 14.7 | WD2 |
| 424 | SCC | Others | *Dillenia indica L.* | 13.4 | 14.0 | 0.559 | 76.0 | WD2 |
| 425 | SCC | Others | *Dillenia indica L.* | 9.3 | 11.6 | 0.476 | 21.9 | WD2 |
| 426 | SCC | Others | *Polyalthia nemoralis A. Dc.* | 13.5 | 15.1 | 0.614 | 77.7 | WD3 |
| 427 | SCC | Others | *Styrax benjoin Dryand.* | 12.8 | 14.1 | 0.557 | 63.6 | WD2 |
| 428 | SCC | Others | *Madhuca alpina Chev.* | 14.8 | 19.6 | 0.613 | 118.0 | WD3 |
| 429 | SCC | Euphorbiaceae | *Baccaurea ramiflora Lour.* | 10.9 | 14.2 | 0.603 | 43.5 | WD3 |
| 430 | SCC | Meliaceae | *Aglaia elaeagnoidea Benth.* | 9.2 | 12.6 | 0.485 | 19.8 | WD2 |
| 431 | SCC | Others | *Polyalthia nemoralis A. Dc.* | 5.5 | 7.1 | 0.538 | 7.0 | WD2 |
| 432 | SCC | Others | *Diospyros pilosula Hiern.* | 5.7 | 6.4 | 0.621 | 6.4 | WD3 |
| 433 | SCC | Others | *Diospyros pilosula Hiern.* | 7.6 | 9.1 | 0.641 | 19.9 | WD3 |
| 434 | SCC | Others | *Magnolia braianensis Gagnep.* | 7.2 | 8.1 | 0.516 | 6.5 | WD2 |
| 435 | SCC | Ulmaceae | *Gironiera subaequalis Planch.* | 13.4 | 14.1 | 0.568 | 64.3 | WD2 |
| 436 | SCC | Others | *Knema pierre Warb.* | 5.4 | 4.7 | 0.609 | 6.8 | WD3 |
| 437 | SCC | Lauraceae | *Litsea baviensis var venulosa Liouho.* | 5.2 | 7.4 | 0.515 | 7.6 | WD2 |
| 438 | SCC | Others | *Camelia fleuryi (Pit.) Sealy* | 8.8 | 8.7 | 0.613 | 22.7 | WD3 |
| 439 | SCC | Myrtaceae | *Syzygium levinei Merr. Et Perry.* | 7.2 | 9.6 | 0.614 | 21.5 | WD3 |
| 440 | SCC | Others | *Knema pierre Warb.* | 5.4 | 6.8 | 0.596 | 6.9 | WD2 |
| 441 | SCC | Others | *Madhuca alpina Chev.* | 5.3 | 9.2 | 0.633 | 8.6 | WD3 |
| 442 | SCC | Ulmaceae | *Gironiera subaequalis Planch.* | 6.5 | 5.2 | 0.549 | 6.2 | WD2 |
| 443 | SCC | Others | *Lepisanthes rubiginosa Leenh.* | 4.9 | 8.5 | 0.649 | 7.3 | WD3 |
| 444 | SCC | Others | *Polyalthia nemoralis A. Dc.* | 5.1 | 5.8 | 0.577 | 5.9 | WD2 |
| 445 | SCC | Dipterocarpaceae | *Shorea farinosa* | 9.9 | 11.1 | 0.600 | 30.0 | WD3 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 446 | SCC | Fagaceae | *Lithocarpus annamensis A. Camus.* | 14.7 | 16.5 | 0.549 | 86.2 | WD2 |
| 447 | SCC | Fagaceae | *Lithocarpus annamensis A. Camus.* | 20.2 | 20.0 | 0.626 | 260.0 | WD3 |
| 448 | SCC | Others | *Barringtonia racenmosa (L.) Spreng* | 16.9 | 10.4 | 0.556 | 117.9 | WD2 |
| 449 | SCC | Others | *Canarium littorale Bl.* | 13.5 | 13.1 | 0.645 | 51.4 | WD3 |
| 450 | SCC | Dipterocarpaceae | *Shorea farinosa* | 24.4 | 18.5 | 0.595 | 329.0 | WD2 |
| 451 | SCC | Myrtaceae | *Syzygium levinei Merr. Et Perry.* | 17.1 | 14.0 | 0.650 | 153.8 | WD3 |
| 452 | SCC | Others | *Scaphium lychnophorum (Hance) Kosterm.* | 17.4 | 16.1 | 0.593 | 93.8 | WD2 |
| 453 | SCC | Others | *Acronychia oligophlebia Merr* | 18.0 | 17.0 | 0.560 | 103.4 | WD2 |
| 454 | SCC | Fagaceae | *Lithocarpus annamensis A. Camus.* | 15.5 | 17.3 | 0.517 | 83.5 | WD2 |
| 455 | SCC | Others | *Lepisanthes rubiginosa Leenh.* | 22.3 | 15.9 | 0.561 | 170.7 | WD2 |
| 456 | SCC | Meliaceae | *Aglaia roxburghiana Miq.* | 16.8 | 15.0 | 0.511 | 99.1 | WD2 |
| 457 | SCC | Others | *Garcinia oliveri Pierre.* | 25.3 | 22.0 | 0.543 | 423.2 | WD2 |
| 458 | SCC | Meliaceae | *Aglaia roxburghiana Miq.* | 26.6 | 13.5 | 0.501 | 261.4 | WD2 |
| 459 | SCC | Others | *Terminalia calamansanai Rolfe.* | 35.5 | 27.3 | 0.574 | 903.3 | WD2 |
| 460 | SCC | Myrtaceae | *Syzygium levinei Merr. Et Perry.* | 14.3 | 12.6 | 0.620 | 91.6 | WD3 |
| 461 | SCC | Fagaceae | *Lithocarpus annamensis A. Camus.* | 29.8 | 14.1 | 0.570 | 457.8 | WD2 |
| 462 | SCC | Others | *Pterospermum diversifolia Bl.* | 37.1 | 20.3 | 0.556 | 980.3 | WD2 |
| 463 | SCC | Others | *Magnolia braianensis Gagnep.* | 39.2 | 26.1 | 0.646 | 1224.3 | WD3 |
| 464 | SCC | Others | *Hosfieldia amygdalina (Wall.) Warb.* | 41.1 | 23.9 | 0.565 | 968.6 | WD2 |
| 465 | SCC | Lauraceae | *Cinnamomum subavenium Miq.* | 42.9 | 27.8 | 0.626 | 1243.2 | WD3 |
| 466 | SCC | Myrtaceae | *Syzygium levinei Merr. Et Perry.* | 52.1 | 23.2 | 0.583 | 2275.1 | WD2 |
| 467 | SCC | Ulmaceae | *Gironiera subaequalis Planch.* | 41.4 | 22.5 | 0.481 | 848.8 | WD2 |
| 468 | SCC | Fagaceae | *Lithocarpus annamensis A. Camus.* | 51.9 | 24.5 | 0.645 | 2376.7 | WD3 |
| 469 | SCC | Others | *Diospyros pilosula Hiern.* | 48.0 | 23.2 | 0.611 | 1503.1 | WD3 |
| 470 | SCC | Euphorbiaceae | *Endospermum sinensis Benth.* | 60.0 | 27.2 | 0.570 | 1782.9 | WD2 |
| 471 | SCC | Others | *Elaeocarpus kontumensis Gagn.* | 10.6 | 6.7 | 0.572 | 14.9 | WD2 |
| 472 | SCC | Others | *Melanorhea laccifera Pierre.* | 62.3 | 25.4 | 0.626 | 3801.3 | WD3 |
| 473 | SCC | Others | *Canarium littorale Bl.* | 51.2 | 27.5 | 0.634 | 1933.3 | WD3 |
| 474 | SCC | Others | *Madhuca alpina Chev.* | 53.3 | 25.4 | 0.646 | 2074.1 | WD3 |
| 475 | SCC | Meliaceae | *Aglaia roxburghiana Miq.* | 65.9 | 27.2 | 0.660 | 3032.0 | WD3 |
| 476 | SCC | Others | *Garcinia handburyi Hook.F* | 67.5 | 26.3 | 0.698 | 3466.9 | WD3 |
| 477 | SCC | Dipterocarpaceae | *Shorea farinosa* | 75.1 | 40.5 | 0.602 | 6575.9 | WD3 |
| 478 | SCC | Others | *Camelia fleuryi (Pit.) Sealy* | 12.4 | 10.2 | 0.647 | 44.9 | WD3 |
| 479 | SCC | Others | *Camelia fleuryi (Pit.) Sealy* | 8.4 | 12.3 | 0.624 | 28.1 | WD3 |
| 480 | SCC | Others | *Polyalthia nemoralis A. Dc.* | 11.8 | 11.2 | 0.640 | 47.9 | WD3 |
| 481 | SCC | Dipterocarpaceae | *Shorea farinosa* | 10.4 | 15.2 | 0.553 | 50.6 | WD2 |
| 482 | NE | Lauraceae | *Phoebe tovoyana* | 8.8 | 10.3 | 0.378 | 16.8 | WD1 |
| 483 | NE | Euphorbiaceae | *Sapium discolor* | 40.1 | 28.2 | 0.359 | 680.9 | WD1 |
| 484 | NE | Others | *Michelia mediocris* | 17.0 | 14.4 | 0.327 | 75.0 | WD1 |
| 485 | NE | Fagaceae | *Castanopsis sp.* | 19.1 | 17.0 | 0.377 | 114.9 | WD1 |
| 486 | NE | Others | *Terminalia bellirica* | 26.1 | 22.7 | 0.472 | 330.0 | WD2 |
| 487 | NE | Leguminosae | *Archidendron tonkinensis* | 49.4 | 28.0 | 0.497 | 1337.9 | WD2 |
| 488 | NE | Euphorbiaceae | *Endospermum chinense* | 19.7 | 14.7 | 0.327 | 82.0 | WD1 |
| 489 | NE | Dipterocarpaceae | *Vatica odorata* | 19.1 | 18.6 | 0.785 | 230.6 | WD3 |
| 490 | NE | Lauraceae | *Cryptocarya impressa* | 17.4 | 17.4 | 0.559 | 137.6 | WD2 |
| 491 | NE | Euphorbiaceae | *Endospermum chinense* | 14.5 | 14.0 | 0.300 | 46.3 | WD1 |
| 492 | NE | Lauraceae | *Litsea sp.* | 11.1 | 11.4 | 0.395 | 32.5 | WD1 |
| 493 | NE | Lauraceae | *Phoebe tovoyana* | 25.2 | 15.8 | 0.546 | 288.8 | WD2 |
| 494 | NE | Others | *Eberhardtia tonkinensis* | 12.7 | 13.0 | 0.373 | 38.2 | WD1 |
| 495 | NE | Dipterocarpaceae | *Vatica odorata* | 68.5 | 31.5 | 0.715 | 2908.8 | WD3 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 496 | NE | Lauraceae | *Phoebe tovoyana* | 51.3 | 26.1 | 0.362 | 911.9 | WD1 |
| 497 | NE | Others | *Dillenia turbiana* | 23.2 | 15.3 | 0.559 | 207.8 | WD2 |
| 498 | NE | Others | *Canarium sp.* | 63.7 | 31.1 | 0.470 | 2189.0 | WD2 |
| 499 | NE | Meliaceae | *Aglaia spectabilis* | 81.8 | 33.5 | 0.475 | 4531.9 | WD2 |
| 500 | NE | Lauraceae | *Cinnamomum glaucescens* | 25.2 | 17.8 | 0.335 | 266.7 | WD1 |
| 501 | NE | Lauraceae | *Phoebe tovoyana* | 9.2 | 13.0 | 0.445 | 30.7 | WD2 |
| 502 | NE | Others | *Tetradium glabrifolium* | 28.0 | 22.6 | 0.295 | 241.4 | WD1 |
| 503 | NE | Euphorbiaceae | *Macaranga denticulata* | 13.4 | 15.1 | 0.339 | 61.4 | WD1 |
| 504 | NE | Meliaceae | *Aglaia spectabilis* | 28.0 | 21.4 | 0.401 | 357.7 | WD2 |
| 505 | NE | Others | *Tetradium glabrifolium* | 31.2 | 22.8 | 0.341 | 357.4 | WD1 |
| 506 | NE | Leguminosae | *Archidendron tonkinensis* | 11.6 | 10.3 | 0.383 | 37.2 | WD1 |
| 507 | NE | Leguminosae | *Archidendron lucidum* | 9.2 | 11.4 | 0.404 | 20.1 | WD2 |
| 508 | NE | Leguminosae | *Archidendron tonkinensis* | 8.9 | 13.5 | 0.255 | 12.2 | WD1 |
| 509 | NE | Leguminosae | *Archidendron chevalieri* | 10.0 | 10.9 | 0.408 | 23.3 | WD2 |
| 510 | NE | Fagaceae | *Castanopsis sp.* | 12.7 | 14.5 | 0.398 | 45.2 | WD1 |
| 511 | NE | Others | *Schima wallichii* | 17.2 | 15.9 | 0.490 | 137.3 | WD2 |
| 512 | NE | Lauraceae | *Phoebe tovoyana* | 37.4 | 26.0 | 0.491 | 938.9 | WD2 |
| 513 | NE | Fagaceae | *Castanopsis sp.* | 32.8 | 23.0 | 0.594 | 552.3 | WD2 |
| 514 | NE | Others | *Sapindus delavayi* | 10.8 | 11.5 | 0.322 | 22.8 | WD1 |
| 515 | NE | Euphorbiaceae | *Macaranga denticulata* | 10.6 | 11.2 | 0.300 | 27.4 | WD1 |
| 516 | NE | Euphorbiaceae | *Macaranga denticulata* | 17.7 | 16.4 | 0.352 | 104.4 | WD1 |
| 517 | NE | Others | *Euodia sutchuenensis* | 45.5 | 23.1 | 0.378 | 652.7 | WD1 |
| 518 | NE | Ulmaceae | *Gironniera subaequalis* | 40.4 | 28.4 | 0.404 | 738.4 | WD2 |
| 519 | NE | Dipterocarpaceae | *Vatica odorata* | 35.2 | 26.1 | 0.757 | 891.4 | WD3 |
| 520 | NE | Others | *Elaeocarpus tonkinensis* | 25.8 | 18.0 | 0.463 | 459.4 | WD2 |
| 521 | NE | Others | *Huodendron biaristatum* | 26.3 | 21.5 | 0.406 | 263.3 | WD2 |
| 522 | NE | Dipterocarpaceae | *Vatica odorata* | 19.1 | 17.8 | 0.723 | 179.5 | WD3 |
| 523 | NE | Others | *Euodia sutchuenensis* | 30.3 | 23.4 | 0.259 | 290.5 | WD1 |
| 524 | NE | Others | *Elaeocarpus tonkinensis* | 33.8 | 27.0 | 0.586 | 722.0 | WD2 |
| 525 | NE | Leguminosae | *Archidendron tonkinensis* | 21.0 | 17.4 | 0.436 | 195.9 | WD2 |
| 526 | NE | Dipterocarpaceae | *Vatica odorata* | 22.3 | 20.2 | 0.742 | 356.3 | WD3 |
| 527 | NE | Leguminosae | *Archidendron chevalieri* | 43.6 | 27.8 | 0.508 | 1392.1 | WD2 |
| 528 | NE | Dipterocarpaceae | *Vatica odorata* | 14.0 | 12.4 | 0.738 | 87.0 | WD3 |
| 529 | NE | Dipterocarpaceae | *Vatica odorata* | 16.9 | 20.6 | 0.774 | 148.9 | WD3 |
| 530 | NE | Lauraceae | *Phoebe tovoyana* | 6.3 | 7.0 | 0.398 | 9.5 | WD1 |
| 531 | NE | Lauraceae | *Phoebe tovoyana* | 44.9 | 30.0 | 0.500 | 1078.1 | WD2 |
| 532 | NE | Euphorbiaceae | *Sapium discolor* | 23.1 | 17.7 | 0.325 | 124.6 | WD1 |
| 533 | NE | Ulmaceae | *Gironniera subaequalis* | 44.7 | 32.0 | 0.472 | 1081.8 | WD2 |
| 534 | NE | Others | *Ficus trivia* | 21.5 | 17.8 | 0.686 | 242.2 | WD3 |
| 535 | NE | Ulmaceae | *Gironniera subaequalis* | 21.6 | 16.3 | 0.397 | 136.6 | WD1 |
| 536 | NE | Dipterocarpaceae | *Vatica odorata* | 35.6 | 24.8 | 0.964 | 947.0 | WD3 |
| 537 | NE | Ulmaceae | *Gironniera subaequalis* | 39.6 | 30.2 | 0.467 | 890.1 | WD2 |
| 538 | NE | Others | *Elaeocarpus tonkinensis* | 57.1 | 29.5 | 0.652 | 2103.6 | WD3 |
| 539 | NE | Euphorbiaceae | *Sapium discolor* | 19.7 | 22.3 | 0.435 | 137.2 | WD2 |
| 540 | NE | Others | *Prunus arborea* | 8.7 | 10.2 | 0.469 | 17.7 | WD2 |
| 541 | NE | Others | *Symplocos laurina* | 25.7 | 20.2 | 0.781 | 366.1 | WD3 |
| 542 | NE | Meliaceae | *Aglaia spectabilis* | 45.5 | 25.8 | 0.441 | 918.1 | WD2 |
| 543 | NE | Others | *Canarium sp.* | 43.8 | 33.1 | 0.605 | 1728.7 | WD3 |
| 544 | NE | Lauraceae | *Litsea cubeba* | 22.9 | 20.2 | 0.411 | 161.3 | WD2 |
| 545 | NE | Euphorbiaceae | *Endospermum chinense* | 10.3 | 11.1 | 0.320 | 23.2 | WD1 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 546 | NE | Fagaceae | *Castanopsis sp.* | 12.6 | 13.7 | 0.431 | 55.6 | WD2 |
| 547 | NE | Others | *Nephelium melliferum* | 11.7 | 15.3 | 0.592 | 67.0 | WD2 |
| 548 | NE | Lauraceae | *Phoebe tovoyana* | 46.0 | 35.5 | 0.540 | 1398.2 | WD2 |
| 549 | NE | Dipterocarpaceae | *Vatica odorata* | 43.7 | 31.8 | 0.902 | 2196.3 | WD3 |
| 550 | NE | Leguminosae | *Albizia lebbeck* | 55.1 | 30.7 | 0.698 | 1985.3 | WD3 |
| 551 | NE | Fagaceae | *Castanopsis sp.* | 10.7 | 12.8 | 0.488 | 38.2 | WD2 |
| 552 | NE | Others | *Symplocos laurina* | 9.5 | 11.6 | 0.558 | 23.0 | WD2 |
| 553 | NE | Others | *Garcinia oblongifolia* | 11.1 | 12.7 | 0.656 | 39.9 | WD3 |
| 554 | NE | Others | *Symplocos laurina* | 14.7 | 14.5 | 0.516 | 77.1 | WD2 |
| 555 | NE | Others | *Symplocos laurina* | 14.2 | 14.6 | 0.565 | 84.8 | WD2 |
| 556 | NE | Euphorbiaceae | *Sapium discolor* | 18.4 | 18.0 | 0.385 | 122.6 | WD1 |
| 557 | NE | Dipterocarpaceae | *Shorea roxburghii* | 35.5 | 23.3 | 0.508 | 653.9 | WD2 |
| 558 | NE | Dipterocarpaceae | *Vatica odorata* | 19.4 | 16.3 | 0.830 | 209.9 | WD3 |
| 559 | NE | Others | *Pometia pinnata* | 47.0 | 32.2 | 0.576 | 1473.3 | WD2 |
| 560 | NE | Others | *Holarrhena pubescens* | 34.1 | 25.1 | 0.437 | 518.4 | WD2 |
| 561 | NE | Meliaceae | *Aglaia spectabilis* | 34.7 | 16.0 | 0.512 | 475.1 | WD2 |
| 562 | NE | Dipterocarpaceae | *Vatica odorata* | 8.5 | 11.1 | 0.817 | 30.0 | WD3 |
| 563 | NE | Others | *Adinandra bockiana* | 10.8 | 11.9 | 0.711 | 47.2 | WD3 |
| 564 | NE | Fagaceae | *Castanopsis sp.* | 6.7 | 8.8 | 0.766 | 16.5 | WD3 |
| 565 | NE | Others | *Garcinia oblongifolia* | 21.5 | 19.0 | 0.708 | 209.0 | WD3 |
| 566 | NE | Meliaceae | *Aglaia spectabilis* | 5.4 | 6.1 | 0.693 | 7.2 | WD3 |
| 567 | NE | Others | *Engelhardtia roxburghiana* | 9.9 | 12.6 | 0.409 | 26.4 | WD2 |
| 568 | NE | Others | *Ficus vasculosa* | 9.4 | 10.5 | 0.361 | 19.5 | WD1 |
| 569 | NE | Others | *Schefflera heptaphilla* | 5.7 | 8.9 | 0.509 | 7.0 | WD2 |
| 570 | NE | Myrtaceae | *Syzygium wightianum* | 31.7 | 27.2 | 0.797 | 863.3 | WD3 |
| 571 | NE | Others | *Prunus arborea* | 12.5 | 11.6 | 0.445 | 38.1 | WD2 |
| 572 | NE | Others | *Prunus arborea* | 33.4 | 21.3 | 0.342 | 312.0 | WD1 |
| 573 | NE | Others | *Elaeocarpus tonkinensis* | 25.3 | 19.7 | 0.560 | 313.1 | WD2 |
| 574 | NE | Others | *Choerospondias axillaris* | 29.0 | 25.2 | 0.494 | 444.5 | WD2 |
| 575 | NE | Euphorbiaceae | *Sapium discolor* | 28.2 | 22.2 | 0.400 | 358.0 | WD1 |
| 576 | NE | Others | *Ficus vasculosa* | 30.5 | 21.9 | 0.479 | 402.6 | WD2 |
| 577 | NE | Dipterocarpaceae | *Vatica odorata* | 34.0 | 21.7 | 0.833 | 919.1 | WD3 |
| 578 | NE | Others | *Nephelium melliferum* | 61.8 | 27.4 | 0.882 | 2832.0 | WD3 |
| 579 | NE | Ulmaceae | *Gironniera subaequalis* | 38.8 | 18.3 | 0.531 | 664.2 | WD2 |
| 580 | NE | Ulmaceae | *Gironniera subaequalis* | 39.4 | 27.4 | 0.689 | 1238.5 | WD3 |
| 581 | NE | Leguminosae | *Archidendron chevalieri* | 49.9 | 24.2 | 0.439 | 1003.5 | WD2 |
| 582 | NE | Fagaceae | *Castanopsis sp.* | 70.3 | 36.6 | 0.889 | 5521.8 | WD3 |
| 583 | NE | Others | *Garuga pinnata* | 16.4 | 17.4 | 0.601 | 129.2 | WD3 |
| 584 | NE | Dipterocarpaceae | *Vatica odorata* | 5.9 | 8.2 | 0.840 | 11.4 | WD3 |
| 585 | NE | Others | *Kitabalia macrophylla* | 9.8 | 6.5 | 0.556 | 18.7 | WD2 |
| 586 | NE | Myrtaceae | *Syzygium zeylanicum* | 13.2 | 11.1 | 0.838 | 88.1 | WD3 |
| 587 | NCC | Others | *Zygocaeus truncatus* | 48.0 | 20.5 | 0.475 | 701.1 | WD2 |
| 588 | NCC | Dipterocarpaceae | *Vatica odorata* | 19.7 | 17.0 | 0.703 | 246.8 | WD3 |
| 589 | NCC | Fagaceae | *Castanopsis chinensis* | 25.9 | 16.5 | 0.546 | 466.7 | WD2 |
| 590 | NCC | Leguminosae | *Erythrophleum fordii* | 19.4 | 14.5 | 0.701 | 184.1 | WD3 |
| 591 | NCC | Others | *Adinandra integerrima* | 36.0 | 23.5 | 0.469 | 607.6 | WD2 |
| 592 | NCC | Euphorbiaceae | *Endospermum chinense* | 27.1 | 22.0 | 0.369 | 208.5 | WD1 |
| 593 | NCC | Others | *Prunus arborea* | 25.2 | 19.5 | 0.576 | 270.8 | WD2 |
| 594 | NCC | Euphorbiaceae | *Sapium discolor* | 38.1 | 24.0 | 0.415 | 692.2 | WD2 |
| 595 | NCC | Euphorbiaceae | *Sapium discolor* | 21.5 | 21.3 | 0.403 | 175.6 | WD2 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 596 | NCC | Lauraceae | *Cryptocarya sp.* | 35.6 | 28.0 | 0.512 | 911.2 | WD2 |
| 597 | NCC | Dipterocarpaceae | *Vatica chevalieri* | 35.6 | 30.0 | 0.686 | 997.0 | WD3 |
| 598 | NCC | Leguminosae | *Peltophorum pterocarpum* | 10.6 | 8.4 | 0.647 | 31.3 | WD3 |
| 599 | NCC | Ulmaceae | *Gironniera subaequalis* | 38.1 | 23.4 | 0.449 | 492.7 | WD2 |
| 600 | NCC | Others | *Engelhardtia roxburghiana* | 26.4 | 20.0 | 0.431 | 271.4 | WD2 |
| 601 | NCC | Meliaceae | *Aglaia macrocarpa* | 46.5 | 26.0 | 0.373 | 943.2 | WD1 |
| 602 | NCC | Fagaceae | *Castanopsis chinensis* | 39.9 | 23.9 | 0.541 | 1435.6 | WD2 |
| 603 | NCC | Meliaceae | *Aglaia macrocarpa* | 29.0 | 18.0 | 0.500 | 430.5 | WD2 |
| 604 | NCC | Others | *NA* | 28.4 | 21.5 | 0.397 | 268.6 | WD1 |
| 605 | NCC | Myrtaceae | *Syzygium wightianum* | 27.0 | 18.5 | 0.437 | 257.6 | WD2 |
| 606 | NCC | Euphorbiaceae | *Endospermum chinense* | 62.0 | 28.1 | 0.401 | 1259.0 | WD2 |
| 607 | NCC | Fagaceae | *Castanopsis pierrei* | 35.2 | 20.0 | 0.468 | 449.1 | WD2 |
| 608 | NCC | Fagaceae | *Castanopsis acuminatissima* | 18.8 | 10.0 | 0.543 | 140.0 | WD2 |
| 609 | NCC | Euphorbiaceae | *Endospermum chinense* | 33.5 | 18.0 | 0.398 | 411.7 | WD1 |
| 610 | NCC | Fagaceae | *Castanopsis pierrei* | 40.7 | 19.0 | 0.473 | 727.0 | WD2 |
| 611 | NCC | Leguminosae | *Ormosia balansae* | 35.1 | 23.0 | 0.507 | 862.8 | WD2 |
| 612 | NCC | Others | *Manglietia conifera* | 29.6 | 15.0 | 0.383 | 203.5 | WD1 |
| 613 | NCC | Euphorbiaceae | *Endospermum chinense* | 35.2 | 19.7 | 0.391 | 289.7 | WD1 |
| 614 | NCC | Euphorbiaceae | *Endospermum chinense* | 59.2 | 29.2 | 0.422 | 1224.4 | WD2 |
| 615 | NCC | Euphorbiaceae | *Endospermum chinense* | 26.5 | 19.7 | 0.404 | 268.1 | WD2 |
| 616 | NCC | Euphorbiaceae | *Endospermum chinense* | 60.5 | 29.8 | 0.392 | 1314.5 | WD1 |
| 617 | NCC | Others | *Pometia pinnata* | 25.3 | 22.5 | 0.389 | 281.2 | WD1 |
| 618 | NCC | Fagaceae | *Castanopsis pierrei* | 34.7 | 19.5 | 0.518 | 385.6 | WD2 |
| 619 | NCC | Euphorbiaceae | *Endospermum chinense* | 36.7 | 20.0 | 0.350 | 400.1 | WD1 |
| 620 | NCC | Others | *Prunus arborea* | 20.5 | 14.5 | 0.544 | 97.1 | WD2 |
| 621 | NCC | Lauraceae | *Cinnamomum parthenoxylon* | 27.0 | 21.0 | 0.364 | 262.4 | WD1 |
| 622 | NCC | Leguminosae | *Erythrophleum fordii* | 20.6 | 13.5 | 0.602 | 234.7 | WD3 |
| 623 | NCC | Euphorbiaceae | *Endospermum chinense* | 55.2 | 25.2 | 0.353 | 926.0 | WD1 |
| 624 | NCC | Euphorbiaceae | *Endospermum chinense* | 67.2 | 23.7 | 0.368 | 1199.1 | WD1 |
| 625 | NCC | Euphorbiaceae | *Endospermum chinense* | 60.2 | 23.8 | 0.375 | 1084.1 | WD1 |
| 626 | NCC | Euphorbiaceae | *Endospermum chinense* | 45.6 | 24.0 | 0.382 | 782.7 | WD1 |
| 627 | NCC | Euphorbiaceae | *Endospermum chinense* | 45.5 | 19.5 | 0.361 | 609.8 | WD1 |
| 628 | NCC | Euphorbiaceae | *Endospermum chinense* | 35.8 | 21.5 | 0.351 | 405.5 | WD1 |
| 629 | NCC | Others | *Alstonia scholaris* | 46.5 | 18.5 | 0.428 | 586.4 | WD2 |
| 630 | NCC | Fagaceae | *Castanopsis pierrei* | 24.8 | 17.0 | 0.434 | 314.5 | WD2 |
| 631 | NCC | Leguminosae | *Ormosia balansae* | 22.6 | 18.7 | 0.472 | 355.2 | WD2 |
| 632 | NCC | Euphorbiaceae | *Endospermum chinense* | 45.4 | 22.5 | 0.383 | 557.4 | WD1 |
| 633 | NCC | Euphorbiaceae | *Endospermum chinense* | 46.3 | 23.0 | 0.353 | 596.7 | WD1 |
| 634 | NCC | Fagaceae | *Castanopsis cerebrina* | 47.9 | 22.5 | 0.429 | 501.8 | WD2 |
| 635 | NCC | Euphorbiaceae | *Endospermum chinense* | 12.0 | 9.0 | 0.323 | 26.6 | WD1 |
| 636 | NCC | Euphorbiaceae | *Endospermum chinense* | 22.0 | 17.0 | 0.385 | 162.3 | WD1 |
| 637 | NCC | Leguminosae | *Sindora siamensis* | 60.5 | 31.5 | 0.578 | 3409.6 | WD2 |
| 638 | NCC | Myrtaceae | *Syzygium sp.* | 29.0 | 17.0 | 0.699 | 387.4 | WD3 |
| 639 | NCC | Others | *Canarium sp.* | 31.5 | 19.7 | 0.750 | 526.8 | WD3 |
| 640 | NCC | Others | *Polyathia lauii* | 46.8 | 24.5 | 0.386 | 847.7 | WD1 |
| 641 | NCC | Others | *Tarrietia javanica* | 47.4 | 27.5 | 0.616 | 1462.0 | WD3 |
| 642 | NCC | Dipterocarpaceae | *Vatica odorata* | 26.0 | 25.5 | 0.925 | 635.9 | WD3 |
| 643 | NCC | Dipterocarpaceae | *Vatica odorata* | 42.7 | 30.9 | 0.873 | 1806.2 | WD3 |
| 644 | NCC | Others | *Tarrietia javanica* | 70.4 | 34.9 | 0.623 | 2986.3 | WD3 |
| 645 | NCC | Others | *Canarium sp.* | 37.4 | 25.1 | 0.708 | 792.6 | WD3 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 646 | NCC | Others | *Tarrietia javanica* | 59.6 | 34.2 | 0.714 | 2780.3 | WD3 |
| 647 | NCC | Others | *Alangium ridleyi* | 41.4 | 22.7 | 0.733 | 1324.8 | WD3 |
| 648 | NCC | Myrtaceae | *Syzygium wightianum* | 54.9 | 28.5 | 0.639 | 2031.4 | WD3 |
| 649 | NCC | Others | *Xerospermum macrophyllum* | 32.5 | 26.5 | 0.748 | 551.8 | WD3 |
| 650 | NCC | Others | *Microcos paniculata* | 26.9 | 20.4 | 0.637 | 429.9 | WD3 |
| 651 | NCC | Others | *Xerospermum macrophyllum* | 25.1 | 15.9 | 0.864 | 323.8 | WD3 |
| 652 | NCC | Euphorbiaceae | *Koilodepas longifolium* | 13.1 | 17.8 | 0.861 | 104.9 | WD3 |
| 653 | NCC | Lauraceae | *Machilus sp.* | 22.4 | 22.1 | 0.538 | 165.6 | WD2 |
| 654 | NCC | Others | *Artocarpus lakoocha* | 11.1 | 14.7 | 0.416 | 26.4 | WD2 |
| 655 | NCC | Myrtaceae | *Syzygium sp.* | 6.1 | 8.0 | 0.660 | 9.9 | WD3 |
| 656 | NCC | Lauraceae | *Cinnamomum sp.* | 7.5 | 9.0 | 0.526 | 14.9 | WD2 |
| 657 | NCC | Others | *Carallia brachiata* | 10.6 | 10.2 | 0.411 | 21.9 | WD2 |
| 658 | NCC | Ulmaceae | *Gironniera subaequalis* | 19.7 | 14.2 | 0.453 | 126.5 | WD2 |
| 659 | NCC | Ulmaceae | *Gironniera subaequalis* | 25.1 | 21.0 | 0.559 | 232.7 | WD2 |
| 660 | NCC | Dipterocarpaceae | *Vatica odorata* | 5.0 | 9.0 | 0.643 | 8.8 | WD3 |
| 661 | NCC | Euphorbiaceae | *Macaranga denticulata* | 5.9 | 6.9 | 0.780 | 9.8 | WD3 |
| 662 | NCC | Lauraceae | *Machilus odoratissima* | 14.4 | 17.5 | 0.722 | 103.6 | WD3 |
| 663 | NCC | Meliaceae | *Aphanamixis grandiflora* | 6.4 | 9.8 | 0.776 | 15.3 | WD3 |
| 664 | NCC | Others | *Symplocos sp.* | 6.7 | 7.7 | 0.658 | 10.8 | WD3 |
| 665 | NCC | Others | *Diospyros sylvatica* | 24.6 | 19.2 | 0.624 | 322.1 | WD3 |
| 666 | NCC | Others | *Glycosmis citrifolia* | 17.2 | 15.4 | 0.671 | 139.0 | WD3 |
| 667 | NCC | Euphorbiaceae | *Macaranga denticulata* | 9.1 | 10.8 | 0.823 | 41.2 | WD3 |
| 668 | NCC | Leguminosae | *Ormosia sp.* | 7.0 | 9.6 | 0.456 | 10.7 | WD2 |
| 669 | NCC | Fagaceae | *Castanopsis sp.* | 16.8 | 18.3 | 0.567 | 153.3 | WD2 |
| 670 | NCC | Myrtaceae | *Syzygium sp.* | 40.8 | 20.5 | 0.721 | 990.6 | WD3 |
| 671 | NCC | Others | *Alangium ridleyi* | 5.4 | 9.3 | 0.548 | 9.3 | WD2 |
| 672 | NCC | Myrtaceae | *Syzygium sp.* | 11.1 | 12.5 | 0.927 | 65.7 | WD3 |
| 673 | NCC | Lauraceae | *Litsea sp.* | 12.5 | 14.3 | 0.665 | 66.1 | WD3 |
| 674 | NCC | Others | *Xerospermum macrophyllum* | 21.0 | 16.5 | 0.752 | 178.1 | WD3 |
| 675 | NCC | Others | *Diospyros apiculata* | 9.1 | 8.6 | 0.608 | 27.5 | WD3 |
| 676 | NCC | Euphorbiaceae | *Endospermum chinense* | 14.5 | 15.6 | 0.396 | 59.0 | WD1 |
| 677 | NCC | Myrtaceae | *Syzygium sp.* | 15.1 | 14.3 | 0.926 | 142.2 | WD3 |
| 678 | NCC | Others | *Alangium ridleyi* | 18.6 | 18.6 | 0.736 | 199.4 | WD3 |
| 679 | NCC | Others | *Alangium ridleyi* | 21.3 | 14.6 | 0.765 | 240.4 | WD3 |
| 680 | NCC | Others | *Michelia mediocris* | 9.6 | 12.5 | 0.547 | 21.5 | WD2 |
| 681 | NCC | Others | *Tarrietia javanica* | 36.1 | 24.7 | 0.595 | 921.7 | WD2 |
| 682 | NCC | Others | *Diospyros apiculata* | 7.3 | 6.9 | 0.646 | 15.3 | WD3 |
| 683 | NCC | Others | *Diospyros sylvatica* | 7.9 | 9.0 | 0.744 | 14.3 | WD3 |
| 684 | NCC | Others | *Xerospermum macrophyllum* | 9.3 | 10.2 | 0.650 | 36.4 | WD3 |
| 685 | NCC | Euphorbiaceae | *Koilodepas longifolium* | 9.7 | 11.3 | 0.792 | 43.9 | WD3 |
| 686 | NCC | Meliaceae | *Aphanamixis grandiflora* | 11.1 | 12.2 | 0.553 | 41.8 | WD2 |
| 687 | NCC | Ulmaceae | *Gironniera subaequalis* | 15.2 | 11.8 | 0.468 | 36.6 | WD2 |
| 688 | NCC | Leguminosae | *Sindora siamensis* | 54.8 | 31.8 | 0.677 | 2557.3 | WD3 |
| 689 | NCC | Others | *Canarium tramdenum* | 67.5 | 33.1 | 0.508 | 3333.8 | WD2 |
| 690 | NCC | Leguminosae | *Ormosia sp.* | 65.2 | 32.0 | 0.537 | 2801.1 | WD2 |
| 691 | NCC | Others | *Xerospermum macrophyllum* | 55.1 | 22.8 | 0.942 | 1759.2 | WD3 |
| 692 | NCC | Lauraceae | *Litsea umbellata* | 44.5 | 22.1 | 0.437 | 650.4 | WD2 |
| 693 | NCC | Others | *Engelhardtia roxburghiana* | 32.6 | 23.3 | 0.511 | 509.8 | WD2 |
| 694 | NCC | Others | *Canarium tramdenum* | 49.8 | 26.5 | 0.528 | 1366.3 | WD2 |
| 695 | NCC | Others | *Tarrietia javanica* | 31.0 | 16.4 | 0.762 | 438.3 | WD3 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 696 | NCC | Others | *Tarrietia javanica* | 55.0 | 25.4 | 0.639 | 1935.7 | WD3 |
| 697 | NCC | Leguminosae | *Erythrophleum fordii* | 34.5 | 24.1 | 0.860 | 1108.0 | WD3 |
| 698 | NCC | Fagaceae | *Castanopsis sp.* | 25.1 | 21.2 | 0.597 | 316.1 | WD2 |
| 699 | NCC | Dipterocarpaceae | *Vatica odorata* | 28.1 | 21.2 | 0.838 | 623.8 | WD3 |
| 700 | NCC | Ulmaceae | *Gironniera subaequalis* | 39.3 | 23.1 | 0.441 | 735.7 | WD2 |
| 701 | NCC | Others | *Ficus sp.* | 33.0 | 21.3 | 0.480 | 364.6 | WD2 |
| 702 | NCC | Meliaceae | *Aphanamixis grandiflora* | 74.5 | 33.3 | 0.716 | 3805.0 | WD3 |
| 703 | NCC | Others | *Glenniea philippinensis* | 63.0 | 33.2 | 0.730 | 3240.3 | WD3 |
| 704 | NCC | Ulmaceae | *Gironniera subaequalis* | 39.8 | 26.3 | 0.581 | 1385.8 | WD2 |
| 705 | NCC | Others | *Xerospermum macrophyllum* | 29.5 | 22.8 | 0.847 | 698.9 | WD3 |
| 706 | NCC | Others | *Xerospermum macrophyllum* | 52.7 | 31.2 | 0.863 | 2336.3 | WD3 |
| 707 | NCC | Others | *Alangium ridleyi* | 12.3 | 13.6 | 0.652 | 63.2 | WD3 |
| 708 | NCC | Others | *Alangium ridleyi* | 10.0 | 12.4 | 0.647 | 35.0 | WD3 |
| 709 | NCC | Others | *Canarium sp.* | 9.2 | 14.4 | 0.762 | 37.1 | WD3 |
| 710 | NCC | Others | *Symplocos sp.* | 14.5 | 19.0 | 0.573 | 75.2 | WD2 |
| 711 | NCC | Others | *Alangium ridleyi* | 6.6 | 9.3 | 0.552 | 12.9 | WD2 |
| 712 | NCC | Others | *Knema sp.* | 11.5 | 11.9 | 0.499 | 36.9 | WD2 |
| 713 | NCC | Others | *Symplocos sp.* | 7.6 | 9.0 | 0.679 | 13.8 | WD3 |
| 714 | NCC | Others | *Schefflera heptaphilla* | 26.6 | 17.7 | 0.402 | 221.7 | WD2 |
| 715 | NCC | Leguminosae | *Peltophorum pterocarpum* | 22.6 | 20.4 | 0.572 | 246.2 | WD2 |
| 716 | NCC | Leguminosae | *Erythrophleum fordii* | 7.8 | 9.4 | 0.584 | 16.3 | WD2 |
| 717 | NCC | Fagaceae | *Castanopsis sp.* | 10.4 | 13.0 | 0.605 | 41.8 | WD3 |
| 718 | NCC | Myrtaceae | *Syzygium sp.* | 6.4 | 7.5 | 0.521 | 9.0 | WD2 |
| 719 | NCC | Others | *Artocarpus rigidus* | 13.4 | 16.1 | 0.448 | 38.6 | WD2 |
| 720 | NCC | Others | *Xerospermum macrophyllum* | 5.9 | 8.5 | 0.752 | 11.2 | WD3 |
| 721 | NCC | Others | *Horsfieldia amygdalina* | 17.5 | 12.8 | 0.465 | 77.9 | WD2 |
| 722 | NCC | Others | *Knema sp.* | 9.5 | 9.3 | 0.497 | 23.2 | WD2 |
| 723 | NCC | Others | *Tarrietia javanica* | 10.4 | 12.2 | 0.549 | 29.8 | WD2 |
| 724 | NCC | Others | *Vernonia arborea* | 18.6 | 21.5 | 0.661 | 207.0 | WD3 |
| 725 | NCC | Lauraceae | *Machilus odoratissima* | 45.6 | 29.5 | 0.442 | 971.9 | WD2 |
| 726 | NCC | Lauraceae | *Cinnamomum obtusifolium* | 21.7 | 16.0 | 0.534 | 198.2 | WD2 |
| 727 | NCC | Lauraceae | *Cryptocarya lenticellata* | 9.7 | 10.5 | 0.496 | 31.9 | WD2 |
| 728 | NCC | Lauraceae | *Litsea sp.* | 17.1 | 15.9 | 0.518 | 105.6 | WD2 |
| 729 | NCC | Lauraceae | *Litsea sp.* | 7.0 | 10.3 | 0.451 | 10.3 | WD2 |
| 730 | NCC | Others | *Horsfieldia amygdalina* | 5.6 | 6.9 | 0.484 | 4.9 | WD2 |
| 731 | NCC | Others | *Garcinia oblongifolia* | 5.2 | 8.4 | 0.667 | 7.1 | WD3 |
| 732 | NCC | Myrtaceae | *Syzygium sp.* | 14.1 | 14.8 | 0.869 | 118.0 | WD3 |
| 733 | NCC | Others | *Diospyros apiculata* | 6.3 | 6.5 | 0.568 | 9.3 | WD2 |
| 734 | NCC | Others | *Artocarpus rigidus* | 15.6 | 18.5 | 0.746 | 115.7 | WD3 |
| 735 | NCC | Others | *Artocarpus masticata* | 8.6 | 11.7 | 0.446 | 16.4 | WD2 |
| 736 | NCC | Lauraceae | *Machilus sp.* | 72.2 | 33.8 | 0.443 | 2550.6 | WD2 |
| 737 | NCC | Lauraceae | *Cinnamomum sp.* | 10.3 | 12.9 | 0.473 | 26.1 | WD2 |
| 738 | NCC | Others | *Knema sp.* | 22.7 | 19.4 | 0.669 | 258.6 | WD3 |
| 739 | NCC | Myrtaceae | *Syzygium sp.* | 20.3 | 21.5 | 0.668 | 274.4 | WD3 |
| 740 | NCC | Lauraceae | *Cryptocarya sp.* | 12.0 | 12.0 | 0.443 | 35.0 | WD2 |
| 741 | NCC | Others | *Glycosmis citrifolia* | 19.0 | 16.2 | 0.462 | 85.6 | WD2 |
| 742 | NCC | Others | *Garcinia cowa* | 7.8 | 9.1 | 0.674 | 22.9 | WD3 |
| 743 | NCC | Others | *Polyathia lauii* | 58.2 | 29.2 | 0.435 | 1232.8 | WD2 |
| 744 | NCC | Euphorbiaceae | *Endospermum chinense* | 68.5 | 33.5 | 0.444 | 1971.2 | WD2 |
| 745 | NCC | Leguminosae | *Ormosia sp.* | 36.8 | 23.9 | 0.537 | 809.3 | WD2 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 746 | NCC | Dipterocarpaceae | *Vatica odorata* | 45.5 | 21.4 | 0.946 | 1191.6 | WD3 |
| 747 | CH | Fagaceae | *Castanopsis sp.* | 4.7 | 5.3 | 0.622 | 3.9 | WD3 |
| 748 | CH | Euphorbiaceae | *Mallotus cochinchinensis* | 5.0 | 7.9 | 0.339 | 5.0 | WD1 |
| 749 | CH | Others | *Garcinia oblongifolia* | 5.0 | 6.3 | 0.614 | 5.3 | WD3 |
| 750 | CH | Myrtaceae | *Syzygium zeylanicum* | 5.1 | 7.1 | 0.597 | 9.6 | WD2 |
| 751 | CH | Others | *NA* | 5.2 | 4.7 | 0.375 | 2.9 | WD1 |
| 752 | CH | Fagaceae | *Castanopsis sp.* | 5.3 | 6.9 | 0.600 | 9.5 | WD2 |
| 753 | CH | Lauraceae | *Phoebe lanceolata* | 5.6 | 8.8 | 0.395 | 5.0 | WD1 |
| 754 | CH | Others | *Symplocos sp.* | 5.6 | 4.5 | 0.480 | 5.9 | WD2 |
| 755 | CH | Lauraceae | *Litsea glutinosa* | 5.6 | 5.5 | 0.491 | 7.3 | WD2 |
| 756 | CH | Others | *Cratoxylum prunifolium* | 5.7 | 8.7 | 0.416 | 11.3 | WD2 |
| 757 | CH | Others | *NA* | 5.8 | 5.3 | 0.422 | 6.2 | WD2 |
| 758 | CH | Others | *Garcinia oblongifolia* | 5.9 | 7.4 | 0.574 | 7.0 | WD2 |
| 759 | CH | Others | *Michelia mediocris* | 6.1 | 5.5 | 0.397 | 6.3 | WD1 |
| 760 | CH | Others | *Hydnocarpus kurzii (King) Warb* | 6.4 | 7.4 | 0.547 | 11.6 | WD2 |
| 761 | CH | Lauraceae | *Cinnamomum iners* | 6.4 | 9.8 | 0.640 | 11.2 | WD3 |
| 762 | CH | Others | *Alstonia scholaris* | 6.4 | 6.1 | 0.425 | 6.1 | WD2 |
| 763 | CH | Others | *Trema orientalis* | 6.5 | 8.0 | 0.352 | 10.1 | WD1 |
| 764 | CH | Euphorbiaceae | *Mallotus cochinchinensis* | 6.5 | 7.5 | 0.341 | 7.3 | WD1 |
| 765 | CH | Lauraceae | *Litsea glutinosa* | 6.5 | 11.5 | 0.681 | 14.8 | WD3 |
| 766 | CH | Ulmaceae | *Gironniera nervosa* | 6.5 | 9.7 | 0.413 | 10.0 | WD2 |
| 767 | CH | Others | *Prunus arborea* | 6.5 | 8.1 | 0.483 | 9.5 | WD2 |
| 768 | CH | Others | *NA* | 6.5 | 8.8 | 0.219 | 7.1 | WD1 |
| 769 | CH | Others | *Diospyros ehretioides* | 6.6 | 6.5 | 0.632 | 9.8 | WD3 |
| 770 | CH | Myrtaceae | *Syzygium hancei* | 6.7 | 3.9 | 0.555 | 10.2 | WD2 |
| 771 | CH | Others | *Grewia paniculata* | 6.8 | 6.3 | 0.459 | 11.1 | WD2 |
| 772 | CH | Meliaceae | *Walsura pinnata Hassk.* | 6.8 | 6.9 | 0.622 | 15.4 | WD3 |
| 773 | CH | Euphorbiaceae | *Macaranga indica* | 7.0 | 9.5 | 0.341 | 11.0 | WD1 |
| 774 | CH | Others | *NA* | 7.0 | 7.8 | 0.695 | 14.7 | WD3 |
| 775 | CH | Others | *Tetradium sp* | 7.0 | 8.2 | 0.530 | 16.6 | WD2 |
| 776 | CH | Fagaceae | *Castanopsis sp.* | 7.0 | 7.4 | 0.569 | 14.5 | WD2 |
| 777 | CH | Myrtaceae | *Syzygium sp.* | 7.0 | 8.5 | 0.326 | 13.8 | WD1 |
| 778 | CH | Others | *Calophyllum sp.* | 7.0 | 7.1 | 0.590 | 18.0 | WD2 |
| 779 | CH | Others | *Lagerstroemia floribunda* | 7.1 | 10.4 | 0.237 | 6.3 | WD1 |
| 780 | CH | Others | *Barringtonia acutangula* | 7.2 | 7.6 | 0.435 | 21.8 | WD2 |
| 781 | CH | Fagaceae | *Castanopsis sp.* | 7.2 | 8.5 | 0.554 | 17.1 | WD2 |
| 782 | CH | Others | *Garcinia oblongifolia* | 7.3 | 9.4 | 0.596 | 16.0 | WD2 |
| 783 | CH | Myrtaceae | *Syzygium sp.* | 7.3 | 10.0 | 0.682 | 44.5 | WD3 |
| 784 | CH | Others | *Garcinia oblongifolia* | 7.5 | 9.0 | 0.672 | 16.4 | WD3 |
| 785 | CH | Myrtaceae | *Syzygium sp.* | 7.5 | 8.7 | 0.660 | 18.0 | WD3 |
| 786 | CH | Others | *Lagerstroemia floribunda* | 7.6 | 10.2 | 0.603 | 22.5 | WD3 |
| 787 | CH | Others | *NA* | 7.6 | 9.4 | 0.573 | 17.8 | WD2 |
| 788 | CH | Euphorbiaceae | *Aporosa microcalyx* | 7.6 | 6.3 | 0.584 | 19.7 | WD2 |
| 789 | CH | Lauraceae | *Cinnamomum iners* | 7.7 | 9.4 | 0.522 | 18.4 | WD2 |
| 790 | CH | Fagaceae | *Castanopsis sp.* | 7.7 | 9.4 | 0.669 | 19.8 | WD3 |
| 791 | CH | Lauraceae | *Cinnamomum iners* | 7.8 | 12.5 | 0.507 | 25.5 | WD2 |
| 792 | CH | Others | *Symplocos sp.* | 7.9 | 9.9 | 0.528 | 21.2 | WD2 |
| 793 | CH | Ulmaceae | *Gironniera nervosa* | 7.9 | 10.5 | 0.502 | 19.2 | WD2 |
| 794 | CH | Others | *Alphonsea sp.* | 7.9 | 8.4 | 0.560 | 16.3 | WD2 |
| 795 | CH | Myrtaceae | *Syzygium sp.* | 8.0 | 9.8 | 0.546 | 32.1 | WD2 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 796 | CH | Lauraceae | *Cinnamomum iners* | 8.0 | 7.8 | 0.286 | 12.8 | WD1 |
| 797 | CH | Fagaceae | *Castanopsis sp.* | 8.1 | 6.0 | 0.658 | 13.4 | WD3 |
| 798 | CH | Lauraceae | *Cinnamomum parthenoxylon* | 8.1 | 8.9 | 0.556 | 21.0 | WD2 |
| 799 | CH | Others | *NA* | 8.1 | 13.2 | 0.573 | 28.2 | WD2 |
| 800 | CH | Others | *Knema poilanei* | 8.2 | 11.9 | 0.344 | 15.5 | WD1 |
| 801 | CH | Others | *NA* | 8.2 | 8.3 | 0.706 | 22.4 | WD3 |
| 802 | CH | Others | *Pterospermum grewiifolium Pierre* | 8.3 | 7.0 | 0.182 | 7.6 | WD1 |
| 803 | CH | Others | *NA* | 8.3 | 7.3 | 0.580 | 19.4 | WD2 |
| 804 | CH | Others | *Grewia paniculata* | 8.3 | 9.2 | 0.730 | 29.1 | WD3 |
| 805 | CH | Euphorbiaceae | *Mallotus cochinchinensis* | 8.5 | 9.0 | 0.304 | 18.3 | WD1 |
| 806 | CH | Others | *Schima superba* | 8.5 | 10.5 | 0.303 | 25.3 | WD1 |
| 807 | CH | Meliaceae | *Melia sp.* | 8.5 | 12.4 | 0.203 | 18.8 | WD1 |
| 808 | CH | Ulmaceae | *Gironniera nervosa* | 8.6 | 8.0 | 0.165 | 8.3 | WD1 |
| 809 | CH | Others | *Lagerstroemia floribunda* | 8.6 | 9.9 | 0.410 | 18.1 | WD2 |
| 810 | CH | Dipterocarpaceae | *Dipterocarpus duperreanus* | 8.6 | 9.5 | 0.489 | 33.4 | WD2 |
| 811 | CH | Dipterocarpaceae | *Dipterocarpus duperreanus* | 8.7 | 6.5 | 0.291 | 14.0 | WD1 |
| 812 | CH | Others | *Styrax annamensis* | 8.7 | 11.3 | 0.455 | 20.7 | WD2 |
| 813 | CH | Fagaceae | *Castanopsis sp.* | 8.8 | 13.1 | 0.458 | 36.0 | WD2 |
| 814 | CH | Others | *Lagerstroemia floribunda* | 8.8 | 10.5 | 0.611 | 27.9 | WD3 |
| 815 | CH | Euphorbiaceae | *Mallotus cochinchinensis* | 8.8 | 9.6 | 0.318 | 16.8 | WD1 |
| 816 | CH | Lauraceae | *Cinnamomum iners* | 8.9 | 6.5 | 0.362 | 13.5 | WD1 |
| 817 | CH | Others | *NA* | 8.9 | 8.8 | 0.598 | 31.5 | WD2 |
| 818 | CH | Others | *NA* | 8.9 | 10.3 | 0.545 | 23.8 | WD2 |
| 819 | CH | Others | *Vitex pubescens* | 9.0 | 12.4 | 0.294 | 21.8 | WD1 |
| 820 | CH | Fagaceae | *Lithocarpus annamensis* | 9.0 | 11.5 | 0.561 | 37.1 | WD2 |
| 821 | CH | Myrtaceae | *Syzygium sp.* | 9.0 | 11.6 | 0.597 | 57.0 | WD2 |
| 822 | CH | Others | *Trema orientalis* | 9.3 | 7.3 | 0.517 | 24.9 | WD2 |
| 823 | CH | Meliaceae | *Aglaia annamensis* | 9.3 | 11.2 | 0.877 | 42.2 | WD3 |
| 824 | CH | Meliaceae | *Walsura pinnata Hassk.* | 9.4 | 8.0 | 0.590 | 40.3 | WD2 |
| 825 | CH | Others | *Trema orientalis* | 9.4 | 9.1 | 0.307 | 21.2 | WD1 |
| 826 | CH | Others | *NA* | 9.5 | 9.7 | 0.393 | 21.9 | WD1 |
| 827 | CH | Meliaceae | *Xylocarpus granatum J.Koenig* | 9.7 | 10.7 | 0.467 | 34.1 | WD2 |
| 828 | CH | Fagaceae | *Castanopsis sp.* | 9.7 | 7.7 | 0.456 | 25.0 | WD2 |
| 829 | CH | Others | *Streblus ilicifolius* | 9.8 | 7.5 | 0.727 | 41.9 | WD3 |
| 830 | CH | Euphorbiaceae | *Mallotus cochinchinensis* | 9.8 | 12.5 | 0.353 | 44.6 | WD1 |
| 831 | CH | Myrtaceae | *Syzygium sp.* | 9.9 | 11.0 | 0.369 | 36.2 | WD1 |
| 832 | CH | Others | *Pterospermum heterophyllum* | 9.9 | 13.9 | 0.610 | 35.4 | WD3 |
| 833 | CH | Others | *NA* | 10.0 | 11.6 | 0.598 | 34.4 | WD2 |
| 834 | CH | Meliaceae | *Chukrasia tabularis A.Juss* | 10.0 | 8.3 | 0.487 | 26.9 | WD2 |
| 835 | CH | Fagaceae | *Quercus glauca Thunb* | 10.1 | 12.5 | 0.692 | 52.9 | WD3 |
| 836 | CH | Meliaceae | *Sandoricum sp.* | 10.1 | 12.8 | 0.493 | 28.3 | WD2 |
| 837 | CH | Others | *Trevesia palmata* | 10.2 | 10.2 | 0.476 | 26.3 | WD2 |
| 838 | CH | Myrtaceae | *Syzygium sp.* | 10.2 | 8.9 | 0.581 | 24.4 | WD2 |
| 839 | CH | Meliaceae | *Dysoxylum binectariferum* | 10.2 | 9.5 | 0.725 | 41.4 | WD3 |
| 840 | CH | Lauraceae | *Litsea glutinosa* | 10.5 | 8.6 | 0.482 | 21.4 | WD2 |
| 841 | CH | Fagaceae | *Lithocarpus annamensis* | 10.6 | 10.8 | 0.717 | 65.3 | WD3 |
| 842 | CH | Others | *Donella Sp.* | 10.7 | 14.1 | 0.744 | 60.9 | WD3 |
| 843 | CH | Others | *Trema orientalis* | 10.8 | 9.7 | 0.319 | 30.3 | WD1 |
| 844 | CH | Ulmaceae | *Gironniera subaequalis* | 10.8 | 10.6 | 0.461 | 28.3 | WD2 |
| 845 | CH | Others | *NA* | 11.0 | 8.3 | 0.451 | 27.9 | WD2 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 846 | CH | Others | *Garcinia sp1* | 11.0 | 12.0 | 0.735 | 52.7 | WD3 |
| 847 | CH | Fagaceae | *Castanopsis sp.* | 11.1 | 12.6 | 0.569 | 69.4 | WD2 |
| 848 | CH | Others | *NA* | 11.2 | 11.6 | 0.540 | 57.3 | WD2 |
| 849 | CH | Myrtaceae | *Syzygium sp.* | 11.2 | 8.4 | 0.564 | 34.7 | WD2 |
| 850 | CH | Meliaceae | *Sandoricum sp.* | 11.3 | 13.1 | 0.618 | 40.9 | WD3 |
| 851 | CH | Others | *Semecarpus sp.* | 11.4 | 10.9 | 0.458 | 32.2 | WD2 |
| 852 | CH | Lauraceae | *Cinnamomum iners* | 11.5 | 13.2 | 0.486 | 62.0 | WD2 |
| 853 | CH | Others | *Morinda sp.* | 11.6 | 11.0 | 0.562 | 47.4 | WD2 |
| 854 | CH | Meliaceae | *Sandoricum sp.* | 11.6 | 9.0 | 0.629 | 53.6 | WD3 |
| 855 | CH | Others | *Lagerstroemia floribunda* | 11.9 | 12.8 | 0.579 | 75.3 | WD2 |
| 856 | CH | Myrtaceae | *Syzygium zeylanicum* | 12.1 | 12.7 | 0.492 | 70.9 | WD2 |
| 857 | CH | Others | *Trema orientalis* | 12.2 | 8.8 | 0.277 | 36.6 | WD1 |
| 858 | CH | Others | *NA* | 12.3 | 13.2 | 0.513 | 18.2 | WD2 |
| 859 | CH | Lauraceae | *Litsea glutinosa* | 12.3 | 7.1 | 0.721 | 61.9 | WD3 |
| 860 | CH | Others | *Calophyllum calaba* | 12.6 | 15.3 | 0.279 | 42.6 | WD1 |
| 861 | CH | Euphorbiaceae | *Macaranga indica* | 13.0 | 13.8 | 0.349 | 49.3 | WD1 |
| 862 | CH | Others | *Trevesia palmata* | 13.0 | 15.2 | 0.354 | 54.4 | WD1 |
| 863 | CH | Lauraceae | *Cinnamomum iners* | 13.1 | 9.4 | 0.346 | 54.6 | WD1 |
| 864 | CH | Myrtaceae | *Syzygium sp.* | 13.7 | 10.0 | 0.641 | 80.2 | WD3 |
| 865 | CH | Fagaceae | *Castanopsis sp.* | 14.0 | 18.7 | 0.609 | 37.9 | WD3 |
| 866 | CH | Others | *NA* | 14.0 | 12.6 | 0.265 | 52.0 | WD1 |
| 867 | CH | Others | *Ixora coccinea* | 14.0 | 14.8 | 0.197 | 52.2 | WD1 |
| 868 | CH | Others | *Trema orientalis* | 14.1 | 10.3 | 0.292 | 46.0 | WD1 |
| 869 | CH | Others | *Donella Sp.* | 14.6 | 11.3 | 0.712 | 115.2 | WD3 |
| 870 | CH | Others | *Lagerstroemia floribunda* | 15.0 | 14.4 | 0.267 | 45.2 | WD1 |
| 871 | CH | Fagaceae | *Castanopsis sp.* | 15.1 | 14.1 | 0.556 | 153.5 | WD2 |
| 872 | CH | Fagaceae | *Castanopsis sp.* | 15.2 | 17.0 | 0.224 | 99.9 | WD1 |
| 873 | CH | Others | *Trema orientalis* | 15.4 | 9.8 | 0.299 | 50.1 | WD1 |
| 874 | CH | Others | *Manglietia sp.* | 15.4 | 19.6 | 0.616 | 134.0 | WD3 |
| 875 | CH | Fagaceae | *Castanopsis sp.* | 15.5 | 10.7 | 0.457 | 67.1 | WD2 |
| 876 | CH | Others | *Careya sphaerica* | 15.8 | 15.3 | 0.621 | 99.2 | WD3 |
| 877 | CH | Others | *Acronychia pedunculata* | 15.8 | 12.9 | 0.576 | 140.2 | WD2 |
| 878 | CH | Others | *Celtis philippinensis Blanco var. wightii (Planch.) Soepadmo* | 15.9 | 19.2 | 0.683 | 138.5 | WD3 |
| 879 | CH | Meliaceae | *Melia sp.* | 16.0 | 13.0 | 0.329 | 79.3 | WD1 |
| 880 | CH | Myrtaceae | *Syzygium levinei* | 16.2 | 10.5 | 0.732 | 58.1 | WD3 |
| 881 | CH | Lauraceae | *Cinnamomum iners* | 16.5 | 11.8 | 0.390 | 59.0 | WD1 |
| 882 | CH | Fagaceae | *Castanopsis sp.* | 16.5 | 20.8 | 0.575 | 261.8 | WD2 |
| 883 | CH | Fagaceae | *Lithocarpus annamensis* | 16.5 | 15.0 | 0.736 | 210.3 | WD3 |
| 884 | CH | Others | *Schefflera octophylla* | 16.6 | 10.2 | 0.404 | 55.8 | WD2 |
| 885 | CH | Others | *Canarium sp1* | 16.9 | 14.5 | 0.645 | 129.4 | WD3 |
| 886 | CH | Fagaceae | *Castanopsis sp.* | 17.0 | 13.9 | 0.463 | 201.4 | WD2 |
| 887 | CH | Fagaceae | *Castanopsis sp.* | 17.0 | 14.8 | 0.330 | 182.1 | WD1 |
| 888 | CH | Others | *Lagerstroemia floribunda* | 17.1 | 16.1 | 0.596 | 175.8 | WD2 |
| 889 | CH | Others | *Polyalthia cerasoides* | 17.6 | 17.2 | 0.751 | 180.4 | WD3 |
| 890 | CH | Myrtaceae | *Syzygium sp.* | 17.8 | 15.5 | 0.507 | 128.1 | WD2 |
| 891 | CH | Others | *Elaeocarpus sp.* | 18.0 | 14.6 | 0.542 | 184.7 | WD2 |
| 892 | CH | Myrtaceae | *Syzygium sp.* | 18.0 | 16.5 | 0.508 | 222.8 | WD2 |
| 893 | CH | Others | *Diospyros maritima* | 18.1 | 12.1 | 0.715 | 172.6 | WD3 |
| 894 | CH | Dipterocarpaceae | *Dipterocarpus duperreanus* | 18.3 | 15.3 | 0.323 | 133.4 | WD1 |
| 895 | CH | Leguminosae | *Dialium cochinchinense* | 18.3 | 4.3 | 0.564 | 65.4 | WD2 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 896 | CH | Lauraceae | *Cinnamomum parthenoxylon* | 18.4 | 16.7 | 0.443 | 118.3 | WD2 |
| 897 | CH | Others | *Alstonia scholaris* | 18.5 | 15.6 | 0.410 | 88.9 | WD2 |
| 898 | CH | Lauraceae | *Cinnamomum parthenoxylon* | 18.9 | 13.1 | 0.520 | 167.4 | WD2 |
| 899 | CH | Myrtaceae | *Syzygium zeylanicum* | 19.3 | 17.5 | 0.465 | 202.2 | WD2 |
| 900 | CH | Others | *Nephelium lappaceum* | 19.6 | 15.0 | 0.882 | 276.6 | WD3 |
| 901 | CH | Others | *Canarium album* | 19.6 | 16.5 | 0.767 | 265.5 | WD3 |
| 902 | CH | Meliaceae | *Melia sp.* | 20.0 | 14.8 | 0.235 | 126.2 | WD1 |
| 903 | CH | Lauraceae | *Phoebe lanceolata* | 20.1 | 13.2 | 0.595 | 150.4 | WD2 |
| 904 | CH | Others | *NA* | 20.3 | 10.6 | 0.608 | 117.4 | WD3 |
| 905 | CH | Others | *Alstonia scholaris* | 21.7 | 15.5 | 0.340 | 102.8 | WD1 |
| 906 | CH | Others | *Polyalthia cerasoides* | 22.0 | 21.4 | 0.756 | 410.1 | WD3 |
| 907 | CH | Others | *Elaeocarpus sp.* | 22.1 | 18.0 | 0.492 | 200.9 | WD2 |
| 908 | CH | Others | *Schima superba* | 22.5 | 15.6 | 0.726 | 263.7 | WD3 |
| 909 | CH | Others | *NA* | 22.8 | 13.5 | 0.823 | 348.7 | WD3 |
| 910 | CH | Others | *Hydnocarpus kurzii (King) Warb* | 23.0 | 16.5 | 0.320 | 230.6 | WD1 |
| 911 | CH | Others | *Lagerstroemia speciosa* | 23.2 | 18.2 | 0.723 | 402.3 | WD3 |
| 912 | CH | Lauraceae | *Machilus parviflora* | 23.5 | 23.4 | 0.586 | 359.7 | WD2 |
| 913 | CH | Others | *Garcinia oblongifolia* | 23.7 | 13.4 | 0.732 | 244.2 | WD3 |
| 914 | CH | Euphorbiaceae | *Aporosa microcalyx* | 23.8 | 26.8 | 0.706 | 528.0 | WD3 |
| 915 | CH | Fagaceae | *Castanopsis sp.* | 24.0 | 20.4 | 0.409 | 429.5 | WD2 |
| 916 | CH | Myrtaceae | *Syzygium sp.* | 24.0 | 14.6 | 0.501 | 448.5 | WD2 |
| 917 | CH | Others | *Ternstroemia kwangtungensis* | 24.1 | 12.1 | 0.628 | 285.5 | WD3 |
| 918 | CH | Fagaceae | *Castanopsis sp.* | 24.4 | 16.0 | 0.484 | 472.6 | WD2 |
| 919 | CH | Myrtaceae | *Syzygium zeylanicum* | 24.4 | 22.0 | 0.549 | 319.3 | WD2 |
| 920 | CH | Others | *Michelia balansae* | 24.5 | 25.2 | 0.452 | 341.1 | WD2 |
| 921 | CH | Myrtaceae | *Syzygium hancei* | 24.8 | 14.2 | 0.664 | 461.9 | WD3 |
| 922 | CH | Others | *Calophyllum dongnaiense* | 25.1 | 27.0 | 0.642 | 456.5 | WD3 |
| 923 | CH | Myrtaceae | *Syzygium sp.* | 25.4 | 21.0 | 0.369 | 435.1 | WD1 |
| 924 | CH | Others | *Diospyros sp.* | 25.5 | 16.0 | 0.765 | 420.1 | WD3 |
| 925 | CH | Fagaceae | *Castanopsis sp.* | 26.0 | 17.5 | 0.345 | 554.3 | WD1 |
| 926 | CH | Others | *Engelhardtia roxburghiana* | 26.0 | 13.5 | 0.357 | 274.4 | WD1 |
| 927 | CH | Others | *Irvingia malayana* | 26.5 | 17.5 | 0.846 | 413.4 | WD3 |
| 928 | CH | Others | *Lagerstroemia floribunda* | 26.8 | 20.5 | 0.434 | 369.5 | WD2 |
| 929 | CH | Myrtaceae | *Syzygium zeylanicum* | 26.8 | 14.8 | 0.641 | 501.4 | WD3 |
| 930 | CH | Fagaceae | *Castanopsis sp.* | 27.1 | 18.8 | 0.517 | 600.1 | WD2 |
| 931 | CH | Euphorbiaceae | *Endospermum chinense* | 27.5 | 15.4 | 0.280 | 245.6 | WD1 |
| 932 | CH | Others | *Schefflera octophylla* | 28.3 | 21.0 | 0.407 | 285.9 | WD2 |
| 933 | CH | Others | *Garcinia sp1* | 28.7 | 21.9 | 0.874 | 514.0 | WD3 |
| 934 | CH | Lauraceae | *Litsea glutinosa* | 30.5 | 15.8 | 0.565 | 533.2 | WD2 |
| 935 | CH | Others | *Schima superba* | 30.5 | 24.2 | 0.320 | 678.6 | WD1 |
| 936 | CH | Others | *Manglietia sp.* | 30.8 | 23.6 | 0.464 | 526.1 | WD2 |
| 937 | CH | Ulmaceae | *Gironniera nervosa* | 31.1 | 15.0 | 0.475 | 435.2 | WD2 |
| 938 | CH | Myrtaceae | *Syzygium zeylanicum* | 31.6 | 21.5 | 0.462 | 437.7 | WD2 |
| 939 | CH | Others | *Camellia vietnamensis* | 32.4 | 23.9 | 0.746 | 944.3 | WD3 |
| 940 | CH | Myrtaceae | *Syzygium sp.* | 32.5 | 23.0 | 0.404 | 844.7 | WD2 |
| 941 | CH | Others | *Schima superba* | 33.0 | 27.3 | 0.212 | 611.5 | WD1 |
| 942 | CH | Myrtaceae | *Syzygium levinei* | 33.1 | 19.6 | 0.586 | 672.3 | WD2 |
| 943 | CH | Myrtaceae | *Syzygium sp.* | 33.8 | 23.0 | 0.581 | 913.1 | WD2 |
| 944 | CH | Fagaceae | *Castanopsis sp.* | 34.0 | 18.3 | 0.336 | 716.7 | WD1 |
| 945 | CH | Others | *Vitex quinata* | 34.4 | 11.3 | 0.622 | 425.1 | WD3 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 946 | CH | Myrtaceae | *Syzygium sp.* | 34.5 | 19.5 | 0.602 | 648.0 | WD3 |
| 947 | CH | Others | *Manglietia conifera* | 35.1 | 24.0 | 0.631 | 864.9 | WD3 |
| 948 | CH | Lauraceae | *Litsea glutinosa* | 36.0 | 19.5 | 0.294 | 762.4 | WD1 |
| 949 | CH | Fagaceae | *Castanopsis sp.* | 36.4 | 18.5 | 0.426 | 733.6 | WD2 |
| 950 | CH | Others | *Hydnocarpus kurzii (King) Warb* | 37.0 | 21.8 | 0.223 | 660.0 | WD1 |
| 951 | CH | Others | *Elaeocarpus sp.* | 37.4 | 24.1 | 0.617 | 844.7 | WD3 |
| 952 | CH | Lauraceae | *Litsea glutinosa* | 38.6 | 22.0 | 0.527 | 865.3 | WD2 |
| 953 | CH | Others | *Wrightia pubescens* | 38.7 | 17.3 | 0.451 | 494.3 | WD2 |
| 954 | CH | Fagaceae | *Castanopsis sp.* | 39.8 | 22.2 | 0.541 | 1147.1 | WD2 |
| 955 | CH | Fagaceae | *Castanopsis sp.* | 40.4 | 22.0 | 0.607 | 1882.4 | WD3 |
| 956 | CH | Others | *Calophyllum sp.* | 41.0 | 23.8 | 0.298 | 1385.5 | WD1 |
| 957 | CH | Others | *Spondias pinnata* | 42.6 | 19.1 | 0.642 | 1003.2 | WD3 |
| 958 | CH | Others | *Schima superba* | 45.0 | 25.0 | 0.319 | 1497.7 | WD1 |
| 959 | CH | Others | *Schima superba* | 46.0 | 25.0 | 0.356 | 1359.9 | WD1 |
| 960 | CH | Others | *Schima superba* | 47.8 | 21.0 | 0.550 | 1631.1 | WD2 |
| 961 | CH | Myrtaceae | *Syzygium sp.* | 49.0 | 22.8 | 0.425 | 1600.5 | WD2 |
| 962 | CH | Others | *Schima superba* | 52.5 | 26.2 | 0.310 | 1580.7 | WD1 |
| 963 | CH | Fagaceae | *Castanopsis sp.* | 53.5 | 23.5 | 0.447 | 2159.2 | WD2 |
| 964 | CH | Others | *Schima superba* | 55.4 | 26.5 | 0.410 | 2337.0 | WD2 |
| 965 | CH | Others | *Schima superba* | 56.0 | 24.7 | 0.272 | 1860.0 | WD1 |
| 966 | CH | Others | *Elaeocarpus sp.* | 56.7 | 22.9 | 0.274 | 1255.8 | WD1 |
| 967 | CH | Others | *Tetrameles nudiflora* | 60.8 | 22.0 | 0.356 | 2179.5 | WD1 |
| 968 | CH | Others | *Lagerstroemia floribunda* | 76.0 | 27.5 | 0.456 | 3149.3 | WD2 |

*Ghi chú: Ký hiệu vùng sinh thái: NE: Đồng Bắc, NCC: Bắc miền Trung, SCC: Nam miền Trung, CH: Tây Nguyên và SE: Đông Nam Bộ. Nguồn: Huy et al. (2016a)*

**Dữ liệu 14:** Dữ liệu mức thích nghi của tếch làm giàu rừng khộp theo các nhân tố sinh thái, trạng thái rừng ở 64 ô thực nghiệm

| Mã ô TN | Mã ô sinh thái | Năm đo | Ngập nước | Mã ngập nước | Loài ưu thể rừng khộp | Mã loài ưu thế | Cấp M rừng khộp | Mã cấp M | Cấp đá lẫn | Mã đá lẫn | Mức thích nghi | Mã mức thích nghi |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VN1 | VN1.1 | 2014 | Khong | 1 | Ca chit | 2 | <50 $m^3$ | 2 | 10 - 30% | 1 | Thich nghi TB | 3 |
| VN1 | VN1.2 | 2014 | Khong | 1 | Ca chit | 2 | <50 $m^3$ | 2 | 50 - 70% | 3 | Thich nghi TB | 3 |
| VN2 | VN2 | 2014 | Khong | 1 | Dau tra beng | 1 | <50 $m^3$ | 2 | <10% | 1 | Thichnghikem | 4 |
| VN4 | VN4 | 2014 | Khong | 1 | Ca chit | 2 | <50 $m^3$ | 2 | 30 - 50% | 1 | Thichnghikem | 4 |
| VN5 | VN5.1 | 2014 | Khong | 1 | Ca chit | 2 | <50 $m^3$ | 2 | <10% | 1 | Thichnghikem | 4 |
| VN5 | VN5.2 | 2014 | Khong | 1 | Ca chit | 2 | <50 $m^3$ | 2 | <10% | 1 | Thichnghikem | 4 |
| VN6 | VN6 | 2014 | Khong | 1 | Ca chit | 2 | 50 - 100 $m^3$ | 3 | <10% | 1 | Thichnghikem | 4 |
| VN7 | VN7.1 | 2014 | Co | 2 | Dau tra beng | 1 | <50 $m^3$ | 2 | <10% | 1 | Thichnghikem | 4 |
| VN7 | VN7.2 | 2014 | Khong | 1 | Chieu lieu den | 3 | <50 $m^3$ | 2 | <10% | 1 | Thichnghikem | 4 |
| VN8 | VN8 | 2014 | Khong | 1 | Ca chit | 2 | <50 $m^3$ | 2 | <10% | 1 | Thich nghi TB | 3 |
| VN9 | VN9.1 | 2014 | Khong | 1 | Ca chit | 2 | 50 - 100 $m^3$ | 3 | 10 - 30% | 1 | Thichnghikem | 4 |
| VN9 | VN9.2 | 2014 | Khong | 1 | Ca chit | 2 | <50 $m^3$ | 2 | <10% | 1 | Thich nghi TB | 3 |
| VN10 | VN10.1 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | <10% | 1 | Thichnghikem | 4 |
| VN10 | VN10.2 | 2014 | Co | 2 | Dau dong | 3 | <50 $m^3$ | 2 | <10% | 1 | Thichnghikem | 4 |
| VN11 | VN11 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | <10% | 1 | Thich nghi TB | 3 |
| YD1 | YD1.1 | 2014 | Co | 2 | Dau dong | 3 | <50 $m^3$ | 2 | <10% | 1 | Thichnghikem | 4 |
| YD1 | YD1.2 | 2014 | Co | 2 | Dau dong | 3 | <50 $m^3$ | 2 | <10% | 1 | Thichnghikem | 4 |
| YD2 | YD2 | 2014 | Co | 2 | Dau dong | 3 | 50 - 100 $m^3$ | 3 | <10% | 1 | Thichnghikem | 4 |
| BD1 | BD1 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | 50 - 70% | 3 | Rat thich nghi | 1 |
| BD2 | BD2 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | 50 - 70% | 3 | Thich nghi | 2 |
| BD3 | BD3 | 2014 | Khong | 1 | Dau dong | 3 | 50 - 100 $m^3$ | 3 | > 70% | 2 | Thich nghi | 2 |
| BD4 | BD4.1 | 2014 | Khong | 1 | Dau dong | 3 | 50 - 100 $m^3$ | 3 | > 70% | 2 | Thich nghi | 2 |
| BD4 | BD4.2 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | 50 - 70% | 3 | Rat thich nghi | 1 |
| BD5 | BD5 | 2014 | Khong | 1 | Dau dong | 3 | 50 - 100 $m^3$ | 3 | 50 - 70% | 3 | Rat thich nghi | 1 |
| BD6 | BD6.1 | 2014 | Khong | 1 | Cam lien | 4 | 50 - 100 $m^3$ | 3 | 50 - 70% | 3 | Thich nghi | 2 |
| BD6 | BD6.2 | 2014 | Khong | 1 | Cam lien | 4 | <50 $m^3$ | 2 | 50 - 70% | 3 | Rat thich nghi | 1 |
| EW1 | EW1.1 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | 30 - 50% | 1 | Thichnghikem | 4 |
| EW1 | EW1.2 | 2014 | Khong | 1 | Cam xe | 3 | <50 $m^3$ | 2 | 30 - 50% | 1 | Thich nghi | 2 |
| EW2 | EW2 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | 10 - 30% | 1 | Thichnghikem | 4 |
| EW3 | EW3.1 | 2014 | Khong | 1 | Ca chit | 2 | <50 $m^3$ | 2 | <10% | 1 | Thichnghikem | 4 |
| EW3 | EW3.2 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | 10 - 30% | 1 | Thichnghikem | 4 |
| EW4 | EW4 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | 50 - 70% | 3 | Thichnghikem | 4 |
| EW5 | EW5 | 2014 | Khong | 1 | Dau tra beng | 1 | <50 $m^3$ | 2 | 10 - 30% | 1 | Thichnghikem | 4 |
| EW6 | EW6 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | 30 - 50% | 1 | Thichnghikem | 4 |
| EW7 | EW7.1 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | 30 - 50% | 1 | Thichnghikem | 4 |
| EW7 | EW7.2 | 2014 | Co | 2 | Dau dong | 3 | <50 $m^3$ | 2 | 50 - 70% | 3 | Thichnghikem | 4 |
| EW8 | EW8.1 | 2014 | Khong | 1 | Dau dong | 3 | 50 - 100 $m^3$ | 3 | <10% | 1 | Thichnghikem | 4 |
| EW8 | EW8.2 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | <10% | 1 | Thichnghikem | 4 |

| Mã ô TN | Mã ô sinh thái | Năm đo | Ngập nước | Mã ngập nước | Loài ưu thế rừng khộp | Mã loài ưu thế | Cấp M rừng khộp | Mã cấp M | Cấp đá lẫn | Mã đá lẫn | Mức thích nghi | Mã mức thích nghi |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| EW9 | EW9 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | <10% | 1 | Thichnghikem | 4 |
| EW10 | EW10 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | <10% | 1 | Thichnghikem | 4 |
| EW11 | EW11.1 | 2014 | Khong | 1 | Cam lien | 4 | <50 $m^3$ | 2 | <10% | 1 | Thich nghi TB | 3 |
| EW11 | EW11.2 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | 10 - 30% | 1 | Thichnghikem | 4 |
| EW12 | EW12 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | 30 - 50% | 1 | Thich nghi TB | 3 |
| EW13 | EW13 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | 10 - 30% | 1 | Thich nghi TB | 3 |
| BN1 | BN1.1 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | <10% | 1 | Thichnghikem | 4 |
| BN1 | BN1.2 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | <10% | 1 | Thichnghikem | 4 |
| BN1 | BN1.3 | 2014 | Khong | 1 | Dau dong | 3 | 50 - 100 $m^3$ | 3 | 30 - 50% | 1 | Thich nghi TB | 3 |
| BN1 | BN1.4 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | 10 - 30% | 1 | Thichnghikem | 4 |
| BN2 | BN2.1 | 2014 | Khong | 1 | Dau dong | 3 | 50 - 100 $m^3$ | 3 | 10 - 30% | 1 | Thichnghikem | 4 |
| BN2 | BN2.2 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | 30 - 50% | 1 | Thich nghi TB | 3 |
| BN3 | BN3.1 | 2014 | Khong | 1 | Cam lien | 4 | 100 - 150 $m^3$ | 1 | 10 - 30% | 1 | Thichnghikem | 4 |
| BN3 | BN3.2 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | 10 - 30% | 1 | Thich nghi TB | 3 |
| BN4 | BN4 | 2014 | Khong | 1 | Dau dong | 3 | 50 - 100 $m^3$ | 3 | 50 - 70% | 3 | Thich nghi TB | 3 |
| BN5 | BN5 | 2014 | Khong | 1 | Dau dong | 3 | >150 $m^3$ | 2 | <10% | 1 | Thich nghi TB | 3 |
| BN6 | BN6.1 | 2014 | Khong | 1 | Dau dong | 3 | <50 $m^3$ | 2 | > 70% | 2 | Thichnghikem | 4 |
| BN6 | BN6.2 | 2014 | Khong | 1 | Dau dong | 3 | 50 - 100 $m^3$ | 3 | 30 - 50% | 1 | Thich nghi TB | 3 |
| BN7 | BN7.1 | 2014 | Khong | 1 | Ca chit | 2 | 50 - 100 $m^3$ | 3 | <10% | 1 | Thichnghikem | 4 |
| BN7 | BN7.2 | 2014 | Khong | 1 | Chieu lieu den | 3 | <50 $m^3$ | 2 | 10 - 30% | 1 | Thich nghi TB | 3 |
| BN7 | BN7.3 | 2014 | Khong | 1 | Chieu lieu den | 3 | <50 $m^3$ | 2 | 10 - 30% | 1 | Thich nghi TB | 3 |
| BN8 | BN8 | 2014 | Khong | 1 | Cam xe | 3 | <50 $m^3$ | 2 | <10% | 1 | Thich nghi TB | 3 |
| TS1 | TS1 | 2014 | Co | 2 | Ca chit | 2 | <50 $m^3$ | 2 | 30 - 50% | 1 | Thich nghi TB | 3 |
| TS2 | TS2 | 2014 | Co | 2 | Ca chit | 2 | <50 $m^3$ | 2 | <10% | 1 | Thichnghikem | 4 |
| TS3 | TS3.1 | 2014 | Co | 2 | Cam lien | 4 | <50 $m^3$ | 2 | <10% | 1 | Thichnghikem | 4 |
| TS3 | TS3.2 | 2014 | Co | 2 | Ca chit | 2 | <50 $m^3$ | 2 | <10% | 1 | Thichnghikem | 4 |

*Nguồn: Bảo Huy (2014)*

# TÀI LIỆU THAM KHẢO

**Tiếng Việt**


[1] Vũ Tiến Hinh, 2012. *Phương pháp lập biểu thể tích cây đứng rừng tự nhiên Việt Nam*. NXB. Nông Nghiệp, Hà Nội.

[2] Vũ Tiến Hinh, Trần Văn Con, 2012. *Sản lượng rừng*. NXB. Nông nghiệp.

[3] Đồng Sĩ Hiền, 1974. *Lập biểu thể tích và độ thon cây đứng cho rừng Việt Nam*. NXB. Khoa học và Kỹ thuật, Hà Nội

[4] Bảo Huy, 1993. *Góp phần nghiên cứu đặc điểm cấu trúc rừng nửa rụng lá- rụng lá ưu thế bằng lăng ở Tây Nguyên*. Luận văn Tiến sĩ Lâm nghiệp. Viện Khoa học Lâm nghiệp Việt Nam.

[5] Bảo Huy, 1995. *Sinh trưởng và sản lượng rừng trồng tếch. Kỷ yếu Hội thảo quốc gia lần thứ nhất về trồng rừng tếch ở Việt Nam. Hội khoa học kỹ thuật Lâm nghiệp Việt Nam.* Nhà in Đại học Quốc gia Hà Nội.

[6] Bảo Huy, 1997. *Nghiên cứu đặc điểm sinh thái và sinh trưởng loài Xoan mộc (Toona sureni)*. Báo cáo khoa học. Sở NN & PTNT Đăk Lăk.

[7] Bảo Huy, Nguyễn Văn Hòa, Nguyễn Thị Kim Liên, 1998. *Nghiên cứu các cơ sở khoa học để kinh doanh rừng trồng tếch*. Báo cáo khoa học đề tài cấp Bộ trọng điểm. Bộ Giáo dục và Đào tạo.

[8] Bảo Huy, Đào Công Khanh, 2008. *Biểu sản lượng rừng trồng Trám trắng tại các tỉnh Lạng Sơn, Bắc Giang, Quảng Ninh*. Bộ Nông nghiệp và Phát triển nông thôn, Hà Nội.

[9] Bảo Huy, 2012. *Xác định lượng $CO_2$ hấp thụ của rừng lá rộng thường xanh vùng Tây Nguyên làm cơ sở tham gia chương trình REDD$^+$*. Báo cáo khoa học đề tài cấp Bộ trọng điểm. Bộ Giáo dục và Đào tạo.

[10] Bảo Huy, 2013. *Mô hình sinh trắc và Viễn thám - GIS để xác định $CO_2$ hấp thụ của rừng lá rộng thường xanh vùng Tây Nguyên*. Nxb. Khoa học và Kỹ thuật, Tp. Hồ Chí Minh.

[11] Bảo Huy, 2014. *Xác định lập địa, trạng thái thích hợp và kỹ thuật làm giàu rừng khộp bằng cây tếch*. Báo cáo kết quả đề tài nghiên cứu khoa học công nghệ. Sở Khoa học Công nghệ, tỉnh Đăk Lăk

[12] Ngô Kim Khôi, 1998. *Thống kê toán học trong lâm nghiệp*. NXB. Nông nghiệp, Hà Nội

[13] Ngô Kim Khôi, Nguyễn Hải Tuất, Nguyễn Văn Tuấn, 2002. *Tin học ứng dụng trong lâm nghiệp*. Nxb. Nông nghiệp, Hà Nội.

[14] Nguyễn Ngọc Lung, 1989. *Điều tra rừng thông Pinus kesiya Việt Nam làm cơ sở tổ chức kinh doanh*. Luận án Tiến sĩ khoa học. Học viện kỹ thuật lâm nghiệp Leningrad mang tên S.M. Kirov, Leningrad. (Bản dịch tiếng Việt).

[15] Nguyễn Thị Quỳnh, 2016. *Xác định nhân tố sinh thái ảnh hưởng đến sinh trưởng lan kim tuyến (Anoectochilus formosanus Hayata) trồng dưới tán rừng thường xanh ở tỉnh Lâm Đồng*. Luận văn Thạc Sĩ Lâm nghiệp, Đại học Tây Nguyên.

[16] Nguyễn Văn Trương, 1983. *Quy luật cấu trúc rừng gỗ hỗn loài*. NXB. Khoa học và Kỹ thuật, Hà Nội.

[17] Nguyễn Hải Tuất, 1982. *Thống kê toán học trong lâm nghiệp*. NXB. Nông nghiệp, Hà Nội.

[18] Nguyễn Hải Tuất, 1975. *Phân bố khoảng cách và ứng dụng của nó*. Thông tin khoa học kỹ thuật. Đại học Lâm nghiệp, số 4(1975).

[19] Nguyễn Hải Tuất, 1990. *Quá trình Poisson và ứng dụng trong nghiên cứu cấu trúc quần thể rừng*. Thông tin khoa học kỹ thuật, Đại học Lâm nghiệp, số 1(1990): 1-7.

[20] Nguyễn Hải Tuất, Ngô Kim Khôi, 1996. *Xử lý thống kê các kết quả nghiên cứu thực nghiệm trong nông – lâm nghiệp trên máy tính*. NXB. Nông Nghiệp, Hà Nội

[21] Nguyễn Hải Tuất, Nguyễn Trọng Bình, 2005. *Khai thác và sử dụng SPSS để xử lý số liệu nghiên cứu trong lâm nghiệp*. NXB. Nông Nghiệp, Hà Nội.

[22] Nguyễn Hải Tuất, Vũ Tiến Hinh và Ngô Kim Khôi, 2006. *Phân tích thống kê trong lâm nghiệp*. NXB. Nông Nghiệp, Hà Nội.

[23] Wikipedia, 2015. R - *Ngôn ngữ lập trình*. Retrieved from https://vi.wikipedia.org/wiki/R

## Tiếng Anh

[1] Basuki, T.M.; Van Lake, P.E.; Skidmore, A.K.; Hussin, Y.A. 2009. *Allometric equations for estimating the above-ground biomass in the tropical lowland Dipterocarp forests.* For. Ecol. and Manag. 257(2009): 1684-1694. DOI 10.1016/j.foreco.2009.01.027.

[2] Bates, D. M. and Watts, D. G. 1988. Nonlinear Regression Analysis and Its Applications, Wiley

[3] Bates, D.M., 2010. lme4: *Mixed-effects modeling with R*. Springer, 131 p.

[4] Brown S. 1997. *Estimating biomass and biomass change of tropical forests: A Primer.* FAO Forestry paper – 134. ISBN 92-5-103955-0. Available on-line: http://www.fao.org/docrep/w4095e/w4095e00.htm

[5] Brown, S., Gillespie A.J.R., and Lugo, A.E. 1989. Biomass estimation methods for tropical forests with applications to forest inventory data. For. Sci. 35:881-902.

[6] Chambers, J. M. 1992. *Linear models. Chapter 4 of Statistical Models* in S eds J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole.

[7] Chave J, Andalo C, Brown S, Cairns MA, Chambers JQ, Eamus D, Folster H, Fromard F, Higuchi N, Kira T, Lescure JP, Nelson BW, Ogawa H, Puig H, Rier B, Yamakura T. 2005. *Tree allometry and improved estimation of carbon stocks and balance in tropical forests.* Oecologia 145 (2005): 87-99. DOI 10.1007/s00442-005-0100-x.

[8] Chave, J., Mechain, M.R., Burquez, A., Chidumayo, E., Colgan, M.S., Delitti, W.B.C., Duque, A., Eid, T., Fearnside, P.M., Goodman, R.C., Henry, M., Yrizar, A.M., Mugasha, W.A., Mullerlandau, H.C., Mencuccini, M., Nelson, B.W., Ngomanda, A., Nogueira, E.M., Ortiz-Malavassi, E., Pelissier, R., Ploton, P., Ryan, C.M., Saldarriaga, J.G., and Vieilledent,

G. 2014. *Improved allometric models to estimate the aboveground biomass of tropical trees.* Global change biology, 20(2014): 3177-3190. Doi: 10.1111/gcb.12629.

[9] Davidian, M., and Giltinan, D.M. 1995. Nonlinear Mixed Effects Models for Repeated Measurement Data, Chapman and Hall.

[10] Dietz, J., Kuyah, S., 2011. *Guidelines for establishing regional allometric equations for biomass estimation through destructive sampling*. World Agroforestry Center (ICRAF). Protocol CBP 1.3. Available at http://reddcommunity.org/sites/default/files/field/publications/allometry2013.pdf.

[11] Freese, F. (1976). *Elementary Forest Sampling*. U.S. Department of Agriculture and Forest Service, USA

[12] Furnival, G.M. (1961). An index for comparing equations used in constructing volume tables. For. Sci. 7: 337-341.

[13] Henry, M., Jara, M.C., Réjou-Méchain, M., Piotto, D., Fuentes, J.M.M., Wayson, C., Guier, F.A., Lombis, H.C., López, E.C., Lara, R.C., Rojas, K.C., Pasquel, J.D.A., Montoya, A.D., Vega, J.F., Galo, A.J., López, O.R., Marklund, L.G., Milla, F., Cahidez, J.J.N., Malavassi, E.O., Pérez, J., Zea, C.R., García, L.R., Pons, R.R., Sanquetta, C., Scott, C., Westfall, J., Zapata-Cuartas, M., Saint-André, L. 2015. *Recommendations for the use of tree models to estimate national forest biomass and assess their uncertainty*. Annals of Forest Science, Issue 6, 72(2015): 769 – 777.

[14] Huy, B., Kralicek, K., Poudel, K.P., Phuong, V.T., Khoa, P.V., Hung, N.D., Temesgen, H. 2016a. *Allometric Equations for Estimating Tree Aboveground Biomass in Evergreen Broadleaf Forests of Viet Nam.* For. Ecol. and Mgmt. 382: 193-205.

[15] Huy, B., Poudel K.P., Temesgen, H. 2016b. Aboveground biomass equations for evergreen broadleaf forests in South Central Coastal ecoregion of Viet Nam: Selection of eco-regional or pantropical models. For. Ecol. and Mgmt. 376: 276-282.

[16] Huy, B., Poudel, K.P., Kralicek, K., Hung, N.D., Khoa, P.V., Phuong, V.T., Temesgen, H. 2016c. *Allometric Equations for Estimating Tree Aboveground Biomass in Tropical Dipterocarp Forests of Viet Nam.* Forests 2016, 7(180): 1-19.

[17] Huy, B., Sharma, B.D., Nguyen, Q.V. 2013. *Participatory Carbon Monitoring. SNV,* Ha Noi, Viet Nam.

[18] IBM. (2011). IBM SPSS Statistics 20 Brief Guide. USA.

[19] IPCC (Intergovernmental Panel on Climate Change). 2003. *IPCC Good Practice Guidance for Land Use, Land-Use Change and Forestry*. Prepared by the National Greenhouse Gas Inventories Programme, Penman, J., Gytarsky, M., Hiraishi, T., Krug, T., Kruger, D., Pipatti, R., Buendia, L., Miwa, K., Ngara, T., Tanabe, K., Wagner F., (eds). Published: IGES, Japan.

[20] Jayaraman, K. (1999). *A Statistical Manual for Forestry Research*. FAO, Bangkok. Thailand.

[21] Jenkins, J.C., Chojnacky, D.C., Heath, L.S. and Birdsey, R.A. 2003. *National-Scale Biomass Estimators for United States Tree Species.* Journal of Forest Science 49(1): 12-35.

[22] Jenkins, J.C., Chojnacky, D.C., Heath, L.S. and Birdsey, R.A. 2004. *Comprehensive Database of Diameter-based Biomass Regressions for North American Tree Species.* United States Department of Agriculture, 45 pp.

[23] Johannes, D., Shem, K. 2011. Guidelines for establishing regional allometric equations for biomass estimation through destructive sampling. CIFOR.

[24] Laar, A.V., Akca, A., 2007. *Forest Mansuration.* Springer, Netherland. ISBN-13 978-1-4020-5991-9 (e-book).

[25] Lackmann, S. (2011). *Lesson 8 - Good Practice in Designing a Forest Inventory*. Regional Workshop: "Capacity Development for Sustainable National Greenhouse Gas Inventories - AFOLU sector (CD-REDD II) Programme. Quito, Ecuador.

[26] Larson, M.G. 2008. *Analysis of Variance*. Circulation, 2008; 117:115-121. Doi: 10.1161/circulationaha.107.654335.

[27] Linton, O. and Härdle, W. 1998. *Nonparametric regression.* In: Kotz, S., Read, C.B., and Banks, D.L. (Eds.). Encyclopedia of statistical sciences. Update vol. 2, 470-485. Wiley. New York.

[28] Mallows, C.L., 1973. *Some Comments on CP*. Technometrics 15 (4): 661–675. doi:10.2307/1267380. JSTOR 1267380.

[29] Mayer DG, Butler DG. 1993. *Statistical validation*. Ecological Modelling, 68(1993): 21-32.

[30] Mehtatalo, L. 2013. *Forest Biometrics with Examples in R*. University of Eastern Finland. School of Computing, 398p.

[31] Moore, A.W. 2017. *Cross-validation for detecting and preventing overfitting.* School of Computer Science. Carnegie Mellon University. Available on-line: https://www.autonlab.org/_media/tutorials/overfit10.pdf on February 02, 2017.

[32] Pearson, T.R.H., Brown, S.L, Birdsey, R.A. 2007. *Measurement Guidelines for the Sequestration of Forest Carbon.* General Technical Report. NRS – 18. U.S. Department of Agriculture Forest Service, PA, USA.

[33] Phuong, V.T., Huy, B., Hung, N.D., Khoa, P.V., Trieu, D.T., Cuong, P.M. 2012. *Guidelines on Destructive Measurement for Forest Above Ground Biomass Estimation.* For Technical Staff Use. UN-REDD Programme, Hanoi, Viet Nam.

[34] Picard, N., Rutishauser E., Ploton P., Ngomanda A., and Henry, M., 2015. *Should tree biomass allometry be restricted to power models*? Forest Ecology and Management 353, 156 – 163.

[35] Picard, N., Saint-André L., Henry M. 2012. *Manual for building tree volume and biomass allometric equations: from field measurement to prediction*. Food and Agricultural Organization of the United Nations, Rome, and Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Montpellier, 215 pp.

[36] Picard, R., Cook, D. 1984. *Cross-Validation of Regression Models*. Journal of the American Statistical Association. 79 (387): 575–583. doi:10.2307/2288403. JSTOR 2288403.

[37] Pinheiro, J., Bates, D., Debroy, S., Sarkar, D. & Team, R. C. 2014. *nlme: Linear and nonlinear mixed effects models*. R package version 3.1-117.

[38] Poso, S., Wang, G., and Tuominen, S. 1999. Weighting alternative estimates when using multisource auxiliary data for forest inventory. Silva Fennica, 33(1): 41–50.

[39] R-Core-Team. (2015). *R: A language and environment for statistical computing. R Foundation for statistical computing*. Vienna, Austria. Retrieved from https://www.R-project.org/

[40] Sola, G., Inoguchi, A., Garcia-Perez, J., Donegan, E., Birigazzi, L., Henry, M. 2014. Allometric equations at national scale for tree biomass assessment in Viet Nam. Context, methodology and summary of the results, UN-REDD Programme, Ha Noi, Viet Nam.

[41] StatPoint-Inc. 2005. The User's Guide to Statgraphics Centurion XV. USA.

[42] Subedi, B.P., Pandey, S. S., Pandey, A., Rana, E. B., Bhattarai, S., Banskota, T. R., Charmakar, S., Tamrakar, R., 2010. *Forest Carbon Stock Measurement*: *Guidelines for measuring carbon stocks in community – managed forests.* Asia Network for Sustainable, Agriculture and Bioresources (ANSAB). Federation of Community Forest, Users, Nepal (FECOFUN). International Centre for Integrated, Mountain Development (ICIMOD). Kathmadu, Nepal. 69p.

[43] Swanson, D.A., Tayman, J., Bryan, T.M., 2011. MAPE-R: a rescaled measure of accuracy for cross-sectional subnational population forecastts. J Pop Research 28(2011):225-243. DOI 10.1007/s12546-011-9054-5.

[44] Temesgen, H., Zhang, C.H., Zhao, X.H. 2014. Modelling tree height-diameter relationships in multi-species and multi-layered forests: A large observational study from Northeast China. Journal of Forest Ecology and Management, 316(2014): 78-89

[45] Twery, M.J. 2004. *Modelling in Forest Management*. In: (eds) Wainwright, J and Mulligan, M. Environmental Modelling: Finding Simplicity in Complexity. John Wiley & Sons, Ltd ISBNs: 0-471-49617-0 (HB); 0-471-49618-9 (PB).

[46] Vanclay, J.K. 1994. Modelling forest growth and yield: applications to mixed tropical forests, CAB International, Wallingford, UK.

[47] Wickham, H. & Chang, W. 2013. Package 'ggplot2': an implementation of the Grammar of Graphics.

[48] Wilk, M.B., Gnanadesikan, R. 1968. *Probability plotting methods for the analysis for the* analysis of data. Biometrika, 55(1), 1-17. doi:10.1093/biomet/55.1.1

[49] Zhang, L. 1997. Cross-validation of Non-linear Growth Functions for Modelling Tree Height–Diameter Relationships. Annals of Botany 79(1997): 251–257.

**PGS.TS. BẢO HUY**

# TIN HỌC THỐNG KÊ TRONG LÂM NGHIỆP

## Sử dụng các chương trình R, Statgraphics, SPSS

*Chịu trách nhiệm xuất bản:*
GIÁM ĐỐC - TỔNG BIÊN TẬP
**PHẠM NGỌC KHÔI**

*Biên tập* : PHẠM THỊ MAI
*Thiết kế bìa* : HOÀNG VIỆT
*Trình bày* : PHẠM THỊ MAI
*Sửa bản in* : PHẠM THỊ MAI

**NHÀ XUẤT BẢN KHOA HỌC VÀ KỸ THUẬT**
*70 Trần Hưng Đạo – Quận Hoàn Kiếm – Hà Nội*
ĐT: (04) 3942 2443 Fax: (04) 3822 0658
Website: http://www.nxbkhkt.com.vn Email: nxbkhkt@hn.vnn.vn
**CHI NHÁNH NHÀ XUẤT BẢN KHOA HỌC VÀ KỸ THUẬT**
*28 Đồng Khởi, 12 Hồ Huấn Nghiệp – Quận 1 – TP. Hồ Chí Minh*
ĐT: (08) 3822 5062 Fax: (08) 3829 6628
Email: chinhanhnxbkhkt@yahoo.com.vn
(Sách thật có đóng dấu và dán tem ở bìa 3)

---

In 300 bản, khổ 19cm × 27cm tại Công ty cổ phần thương mại In Nhật Nam
Địa chỉ: 007 Lô I – KCN Tân Bình – P. Tây Thạnh – Q. Tân Phú – TP. Hồ Chí Minh
Số ĐKXB: 691 – 2017/CXBIPH/5 – 18/KHKT
Quyết định XB số: 23/QĐ-NXBKHKT, ngày 13/04/2017
Mã ISBN: 978-604-67-0853-7
In xong và nộp lưu chiểu quí I năm 2017

# TIN HỌC THỐNG KÊ TRONG LÂM NGHIỆP

PGS.TS. BẢO HUY

PGS.TS. BẢO HUY TIN HỌC THỐNG KÊ TRONG LÂM NGHIỆP

GIÁ: 225.000đ