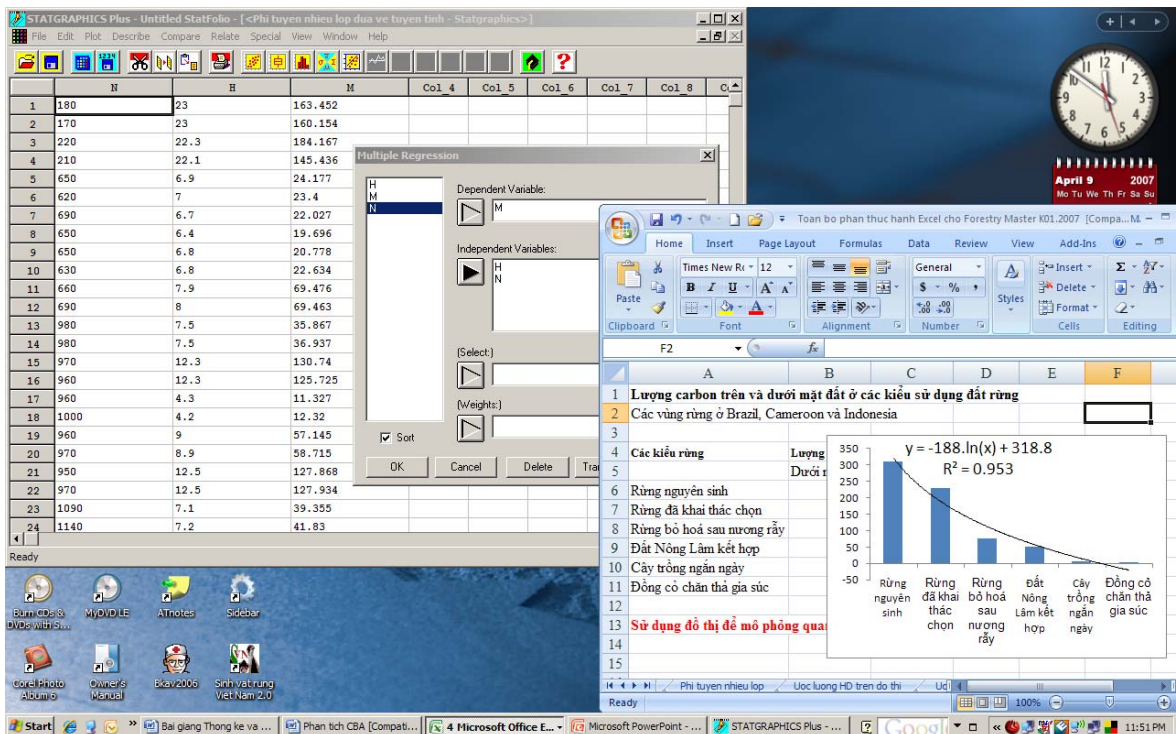


PGS.TS. BẢO HUY

# THỐNG KÊ TÍN HỌC TRONG LÂM NGHIỆP

Áp dụng phần mềm Statgraphics Centurion và MS. Excel  
(Dùng cho Cao học Lâm nghiệp)



Tháng 5 năm 2009



# Mục lục

|            |   |           |
|------------|---|-----------|
| <b>1</b>   | <b>TỔNG QUÁT VỀ CHỨC NĂNG XỬ LÝ THỐNG KÊ CỦA EXCEL VÀ STATGRAPHICS</b> .....  | <b>7</b>  |
| 1.1        | Tổng quát về phần xử lý thống kê trong Excel .....  | 7         |
| 1.2        | Tổng quát về phần mềm xử lý thống kê Statgraphics Centurion .....   | 8         |
| <b>2</b>   | <b>THỐNG KÊ MÔ TẢ</b> .....   | <b>10</b> |
| <b>3</b>   | <b>SO SÁNH 1 – 2 MẪU QUAN SÁT BẰNG TIÊU CHUẨN T</b> .....   | <b>13</b> |
| 3.1        | So sánh một mẫu với một giá trị cho trước – Kiểm tra T một mẫu .....  | 13        |
| 3.2        | So sánh sự sai khác giữa trung bình 2 mẫu – Kiểm tra T 2 mẫu .....  | 14        |
| <b>4</b>   | <b>NGHIÊN CỨU MỐI QUAN HỆ SINH THÁI LOÀI TRONG RỪNG MƯA NHIỆT ĐỚI DỰA VÀO TIÊU CHUẨN <math>\chi^2</math> (Bảo Huy, 1997)</b> .....              | <b>17</b> |
| <b>5</b>   | <b>PHÂN TÍCH PHƯƠNG SAI</b> .....   | <b>23</b> |
| 5.1.       | Phân tích phương sai 1 nhân tố với các thí nghiệm ngẫu nhiên hoàn toàn ..   | 23        |
| 5.2.       | Phân tích phương sai 2 nhân tố .....  | 27        |
| 5.2.1.     | Phân tích phương sai 2 nhân tố với 1 lần lặp lại: (Bố trí thí nghiệm theo khối ngẫu nhiên đầy đủ (Randomized Complete Blocks) (RCB): .....      | 27        |
| 5.2.2.     | Phân tích phương sai 2 nhân tố m lần lặp .....  | 32        |
| <b>6.</b>  | <b>PHÂN TÍCH TƯƠNG QUAN - HỒI QUY</b> .....   | <b>36</b> |
| 6.1.       | Hồi quy tuyến tính 1 lớp .....  | 36        |
| 6.2.       | Dạng phi tuyến đưa về tuyến tính 1 lớp .....  | 38        |
| 6.2.1.     | Lập mô hình hàm mũ trong Excel: .....   | 38        |
| 6.2.2.     | Lập mô hình hàm mũ và Schumacher trong Statgraphics: .....  | 40        |
| 6.3.       | Hồi quy tuyến tính nhiều lớp .....  | 46        |
| 6.4.       | Hồi quy phi tuyến tính nhiều lớp, tổ hợp biến .....   | 49        |
| 6.4.1.     | Lập mô hình phi tuyến nhiều lớp chuyển về tuyến tính nhiều lớp trong Excel ..   | 49        |
| 6.4.2.     | Lập mô hình phi tuyến nhiều lớp chuyển về tuyến tính trong Statgraphics .....   | 51        |
| <b>7.</b>  | <b>ƯỚC LƯỢNG CÁC DẠNG HỒI QUY MỘT BIẾN TRÊN ĐỒ THỊ</b> .....  | <b>55</b> |
| <b>8.</b>  | <b>SẮP XẾP VÀ VẼ BIỂU ĐỒ PHÂN BỐ TẦN SỐ XUẤT HIỆN THEO CẤP, CỖ, HẠNG</b> .....  | <b>59</b> |
| <b>9.</b>  | <b>KIỂM TRA THUẬN NHẤT K MẪU QUAN SÁT ĐÚT QUẢNG - ỨNG DỤNG: KIỂM TRA SỰ THUẬN NHẤT CỦA CÁC DÃY PHÂN BỐ N/D, N/H Ở CÁC Ô TIÊU CHUẨN</b> .....    | <b>61</b> |
| <b>10.</b> | <b>MÔ HÌNH HOÁ QUY LUẬT PHÂN BỐ</b> .....   | <b>62</b> |
| 10.1.      | Mô hình hoá phân bố giảm theo hàm Mayer .....   | 63        |
| 10.2.      | Mô phỏng phân bố thực nghiệm theo phân bố khoảng cách-hình học: .....   | 67        |
| 10.3.      | Mô phỏng phân bố thực nghiệm theo phân bố Weibull: .....  | 69        |
| <b>11.</b> | <b>PHƯƠNG PHÁP NGHIÊN CỨU CẤU TRÚC MẶT BẰNG RỪNG (Bảo Huy, 1993)</b> .....  | <b>71</b> |
| <b>12.</b> | <b>PHÂN TÍCH, PHÁT HIỆN CÁC NGUYÊN NHÂN, NHÂN TỐ ĐỊNH TÍNH, ĐỊNH LƯỢNG ẢNH HƯỞNG ĐẾN BIẾN PHỤ THUỘC (HẬU QUẢ, VẤN ĐỀ) (Bảo Huy, 2006)</b> ..... | <b>73</b> |



## LỜI NÓI ĐẦU

Tài liệu này được biên soạn phục vụ cho việc giảng dạy môn học **“Thống kê và Tin học trong lâm nghiệp”** cho lớp Cao học Lâm nghiệp ở trường Đại học Tây Nguyên. Môn học này giúp cho người học phân tích, xử lý số liệu thống kê trên máy vi tính trong quá trình học tập, làm đề tài nghiên cứu cũng như ứng dụng vào thực tiễn.

Có rất nhiều phần mềm ứng dụng để xử lý thống kê như SPSS, Statgraphics Plus, Excel... Microsoft Excel được mọi người biết đến khi nói đến công cụ bảng tính, tính toán..., nhưng những chức năng chuyên sâu về ứng dụng thống kê trong sinh học, nông lâm nghiệp, quản lý tài nguyên thiên nhiên, môi trường lại ít được đề cập đến. Mục đích của môn học này là khai thác chức năng xử lý thống kê hết sức phong phú và mạnh của phần mềm Excel để ứng dụng trong phân tích các kết quả thí nghiệm, đánh giá các kết quả điều tra khảo sát trong lâm nghiệp, nghiên cứu về quản lý tài nguyên thiên nhiên. Trong đó bao gồm các xử lý thống kê phổ biến như: Phân tích các đặc trưng mẫu, so sánh các mẫu thí nghiệm, phân tích phương sai, tương quan hồi quy, dự báo..... do đó phần mềm Excel được chọn lựa để giới thiệu.

Các phần mềm thống kê chuyên dụng và phổ biến trên thế giới là Statgraphics, SPSS, .... Đây là các phần mềm thống kê được ứng dụng rộng trong hầu hết các lĩnh vực nghiên cứu, phân tích dữ liệu của nhiều ngành khác nhau về xã hội, tự nhiên. Ứng dụng mạnh của các phần mềm này là phân tích các mô hình hồi quy đa biến dạng tuyến tính hay phi tuyến tính với các cách phân tích đa dạng như hồi quy lọc, hồi quy từng bước, tổ hợp biến, mã hóa tự động các biến định tính, ..... Do đó phần mềm Statgraphics cũng được giới thiệu để người học có thể tiếp cận với công cụ phân tích thống kê này.

Tài liệu này sẽ không đi sâu vào lý thuyết xác suất thống kê, mà thiên về hướng ứng dụng đơn giản, dễ hiểu, kèm theo các ví dụ để người đọc có thể thực hành các chức năng xử lý, phân tích dữ liệu bằng Excel, Statgraphics Plus một cách nhanh chóng, thuận tiện trong hoạt động quản lý và nghiên cứu lâm nghiệp, quản lý tài nguyên thiên nhiên, môi trường.



# 1 TỔNG QUÁT VỀ CHỨC NĂNG XỬ LÝ THỐNG KÊ CỦA EXCEL VÀ STATGRAPHICS

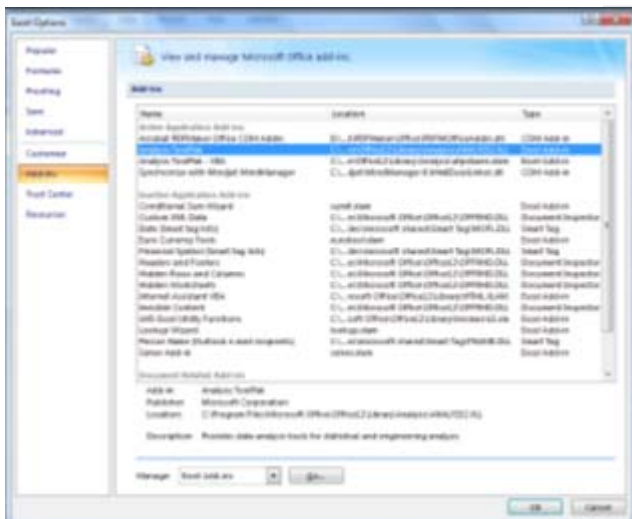
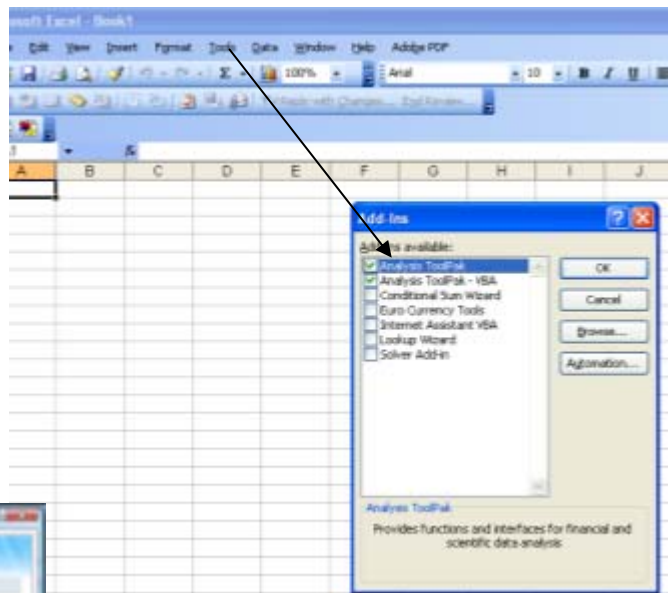
## 1.1 Tổng quát về phần xử lý thống kê trong Excel

Excel thiết kế sẵn một số chương trình để xử lý số liệu và phân tích thống kê cơ bản ứng dụng trong nhiều lĩnh vực:

- Chức năng xử lý số liệu, tạo bảng tổng hợp dữ liệu: Sắp xếp, tính toán nhanh các bảng tổng hợp từ số liệu thô,...
- Chức năng của các hàm: Cung cấp hàng loạt các hàm về kỹ thuật, thống kê, kinh tế tài chính, hàm tra các chỉ tiêu thống kê như t, F,  $\chi^2$
- Chức năng Data Analysis: Dùng để phân tích thống kê như phân tích các đặc trưng mẫu, tiêu chuẩn t để so sánh sự sai khác, phân tích phương sai, ước lượng các tương quan hồi quy
- Phân tích mô hình tương quan hoặc hồi quy để dự báo các thay đổi theo thời gian ngay trên đề thị.

*Lưu ý: Về việc cài đặt chương trình phân tích dữ liệu (Data Analysis) trong Excel:*

- Khi cài đặt phần mềm Excel phải thực hiện trong chế độ chọn lựa cài đặt, sau đó phải chọn mục: Add-Ins và Analysis Toolpak.
- Khi chạy Excel lần đầu cần mở chế độ phân tích dữ liệu bằng cách: Menu Tools/Add-Ins và chọn Analysis Toolpak-OK. (Đối với MS. Office 2003)



*Đối với MS. Office 2007, tiến hành mở chế độ phân tích thống kê như sau: Kích vào Microsoft Office Button sau đó chọn excel options, kích vào Add-ins, và chọn Analysis ToolPak trong hộp thoại - OK.*

Ảnh hưởng trong thực tế quản lý dữ liệu nông lâm nghiệp nói riêng, việc khai thác hết tiềm năng ứng dụng của Excel cũng mang lại hiệu quả tốt mà không nhất thiết phải tìm kiếm thêm một phần mềm chuyên dụng nào khác. Vấn đề đặt ra là xác định chiến lược ứng dụng và khai thác đúng và sâu các công cụ chức năng sẵn có ở một phần mềm phổ biến ở bất kỳ một vị tính cá nhân nào.

### **Một số hàm thông dụng trong thống kê:**

- Tính tổng: =Sum(dãy ds).
- Tổng bình phương: =Sumq(dãy ds).
- Trung bình: =Average(dãy ds).
- Lấy giá trị tuyệt đối: =Abs(ds).
- Trị lớn nhất, nhỏ nhất: =Max(dãy ds), Min(dãy ds).
- Các hàm lượng giác: =Cos(ds), =Sin(ds), =tan(ds).
- Hàm mũ, log: =Exp(ds), =Ln(ds), =Log(ds).
- Căn bậc 2: =Sqrt(ds)..
- Sai tiêu chuẩn mẫu chưa hiệu đính: =Stdevp(dãy ds); đã hiệu đính =Stdev(dãy ds).
- Phương sai mẫu chưa hiệu đính: =Varp(dãy ds); đã hiệu đính =Var(dãy ds).
- Giai thừa: =Fact(n).
- Số Pi: =Pi().

### **Tra các giá trị T, F, $\chi^2$ :**

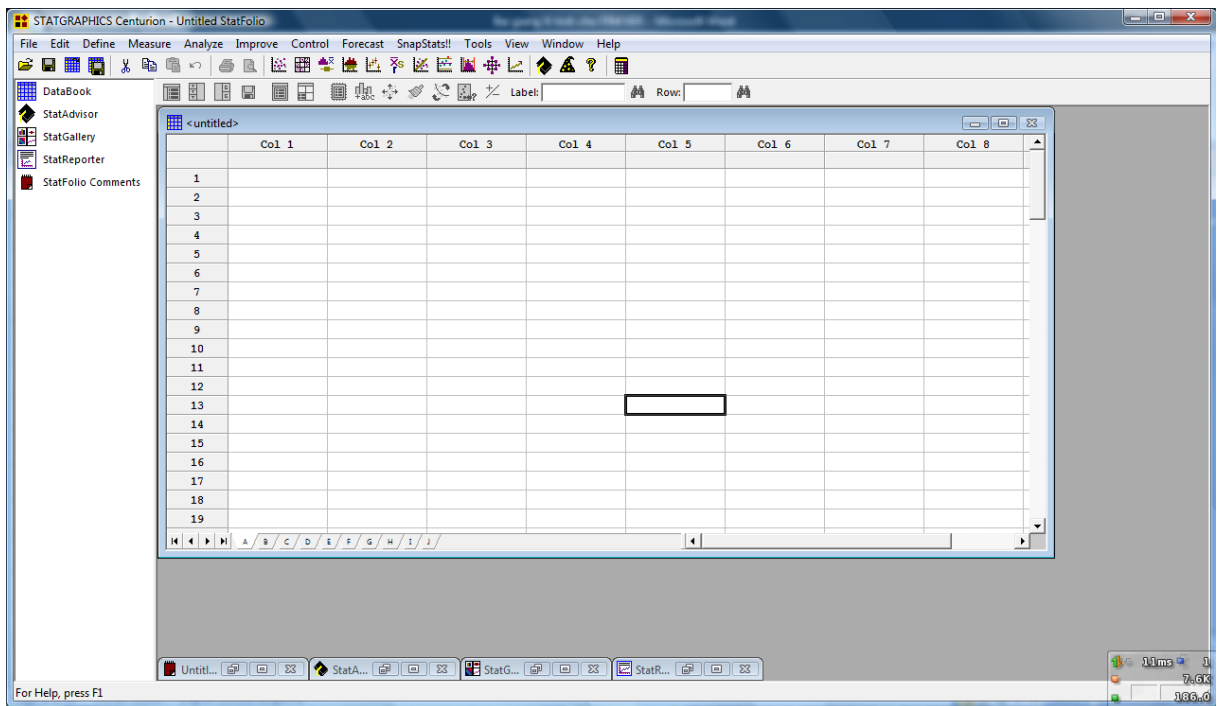
- Chọn 1 ô lấy giá trị tra.
- Kích nút fx trên thanh công cụ chuẩn. Trong hộp thoại Function Category, chọn Statistical.
- Trong mục Function name, chọn 1 trong các hàm:
  - Hàm Tinv:** để tra T.
  - Hàm Chiinv:** để tra  $\chi^2$ .
  - Hàm Finv:** để tra F.Bấm **Alt** ext.
- Trong hộp thoại tiếp theo: Function Wizard chọn:
  - Probability (fx): Gõ vào mức ý nghĩa  $\alpha=0.05$  ; 0.01 hay 0.001.
  - Degrees Freedom (fx): Gõ vào bậc tự do. Đối với tiêu chuẩn F cần đưa vào 2 độ tự do.
  - Finish.

## **1.2 Tổng quát về phần mềm xử lý thống kê Statgraphics Centurion**

Đây là một phần mềm chuyên dụng trong xử lý thống kê, bao gồm các chức năng:

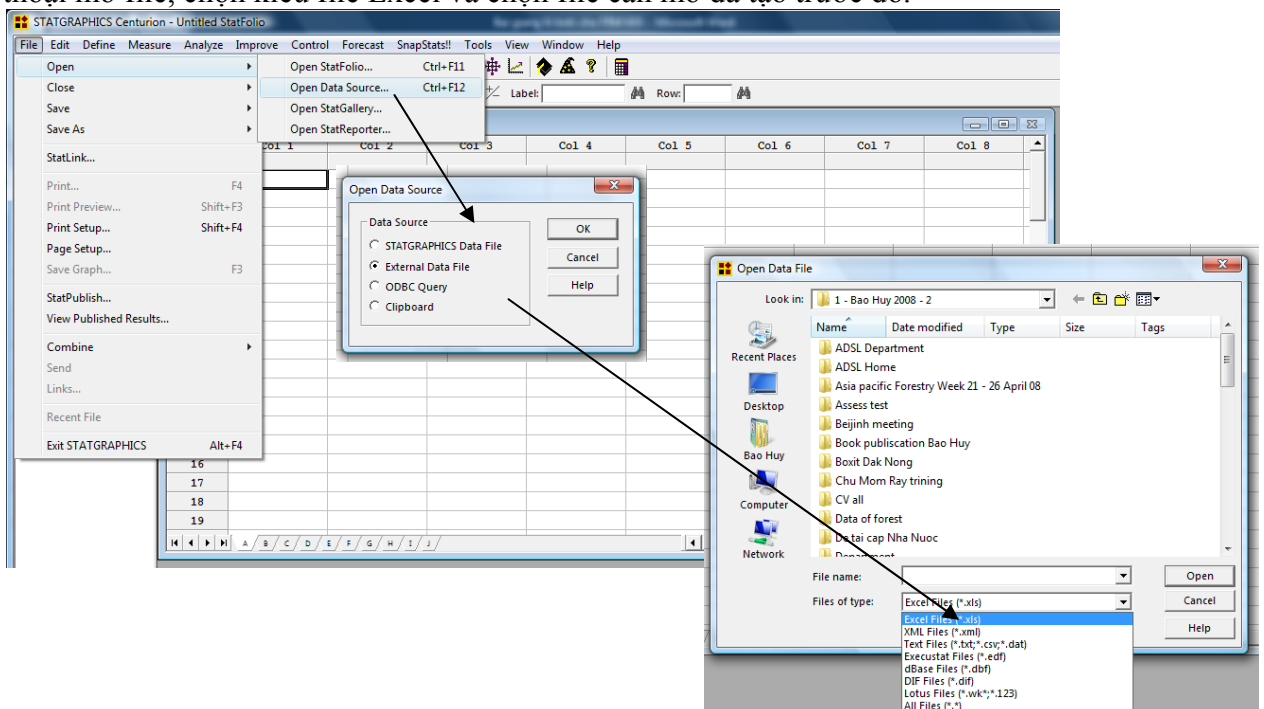
- Tạo lập cơ sở dữ liệu dưới dạng bảng tính
- Tính toán các đặc trưng mẫu, vẽ sơ đồ, đồ thị quan hệ
- So sánh hai hay nhiều mẫu bằng các tiêu chuẩn thống kê t, U, F và nhiều tiêu chuẩn phi tham số khác.
- Phân tích phương sai Aả OVA.
- Kiểm tra tính chuẩn của dữ liệu và đổi biến số.
- Thiết lập các mô hình hồi quy tuyến tính hay phi tuyến tính từ một cho đến nhiều lớp, tổ hợp biến. Với cách xử lý đa dạng để chọn lựa được các biến ảnh hưởng đến một hậu quả (biến phụ thuộc).





Giao tiếp trong Statgraphics Centurion, số liệu đầu vào có thể được nhập trực tiếp trong file bảng tính và cơ sở dữ liệu; 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 2007.

Sau khi nhập dữ liệu trong Excel 97-2003, đóng file của Excel và mở nó trong Statgraphics 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.



## 2 THÔNG KÊ MÔ TẢ

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, sự 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 hành 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.

Các đặc trưng mẫu bao gồm tính các chỉ tiêu: Số trung bình, số trung vị, phương sai, sai tiêu chuẩn, 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.

Ví dụ: Khảo sát các đặc trưng cơ bản về sinh trưởng đường kính của rừng trồng tẻch.

Số liệu đo  $D_{1,3}$  rừng trồng Tẻch 14 tuổi trong ô tiêu chuẩn  $500m^2$ .

Các đặc trưng mẫu có thể tính đồng thời trong Excel theo các bước:

- ả nhập số liệu theo cột hoặc hàng.
- Menu Tools/Data Analysis/Descriptive Statistics/OK. Có hộp thoại, trong đó cần xác định:
  - Input range: Khai báo khối dữ liệu.
  - Grouped by: Chọn dữ liệu nhập theo cột (Columns) hoặc hàng (Rows).
  - Label in first row: ả ếu đưa vào cả hàng tiêu đề thì đánh dấu.
  - Output range: Đánh vào địa chỉ ô trên trái nơi đưa ra kết quả.
  - Summary Statistics: Thông tin tóm lược các đặc trưng thống kê (đánh dấu).
  - Kích nút OK

**Bảng nhập dữ liệu đường kính  $D_{1,3}$  của Tẻch**

|    | A  | B    | C | D                  | E            | F |
|----|--|------|---|--------------------|--------------|---|
| 1  |  |      |   |                    |              |   |
| 2  | <b>Số liệu đo đường kính cây tẻch trong lâm phần</b> |      |   |                    |              |   |
| 3  | Tính các đặc trưng mẫu                               |      |   |                    |              |   |
| 4  |  |      |   |                    |              |   |
| 5  |  |      |   |                    |              |   |
| 6  | <b>D1,3 (cm)</b>                                     |      |   |                    |              |   |
| 7  |  | 21.0 |   | <i>D1,3 (cm)</i>   |              |   |
| 8  |  | 18.8 |   |                    |              |   |
| 9  |  | 16.9 |   | Mean               | 18.98627451  |   |
| 10 |  | 20.1 |   | Standard Error     | 0.442329126  |   |
| 11 |  | 20.7 |   | Median             | 19.1         |   |
| 12 |  | 18.1 |   | Mode               | 19.4         |   |
| 13 |  | 15.0 |   | Standard Deviation | 3.158861796  |   |
| 14 |  | 17.5 |   | Sample Variance    | 9.978407843  |   |
| 15 |  | 27.1 |   | Kurtosis           | 0.855133084  |   |
| 16 |  | 19.4 |   | Skewness           | -0.221834351 |   |
| 17 |  | 22.6 |   | Range              | 17.2         |   |
| 18 |  | 21.3 |   | Minimum            | 9.9          |   |
| 19 |  | 18.1 |   | Maximum            | 27.1         |   |

**Sử dụng chức năng phân tích đặc trưng mẫu của Excel**

Microsoft Excel - Trình bày giảng dạy Vi tính all

File Edit View Insert Format Tools Data Window Help Adobe PDF

100% VNbook-Antiqua 14 B I U

Reply with Changes... End Review...

A1 D1,3 (cm)

|    | A                | B    | C | D | E | F |
|----|------------------|------|---|---|---|---|
| 1  | <b>D1,3 (cm)</b> |      |   |   |   |   |
| 2  |                  | 21.0 |   |   |   |   |
| 3  |                  | 18.8 |   |   |   |   |
| 4  |                  | 16.9 |   |   |   |   |
| 5  |                  | 20.1 |   |   |   |   |
| 6  |                  | 20.7 |   |   |   |   |
| 7  |                  | 18.1 |   |   |   |   |
| 8  |                  | 15.0 |   |   |   |   |
| 9  |                  | 17.5 |   |   |   |   |
| 10 |                  | 27.1 |   |   |   |   |
| 11 |                  | 19.4 |   |   |   |   |
| 12 |                  | 22.6 |   |   |   |   |
| 13 |                  | 21.3 |   |   |   |   |
| 14 |                  | 18.1 |   |   |   |   |

**Descriptive Statistics**

Input

Input Range:

Grouped By:  Columns  Rows

Labels in first row

Output options

Output Range:

New Worksheet Ply:

New Workbook

Summary statistics

Confidence Level for Mean:  %

Kth Largest:

Kth Smallest:

OK Cancel Help

**Kết quả tính đặc trưng mẫu**

| <b>D1,3 (cm)</b>         |        |
|--------------------------|--------|
| Mean                     | 18,98  |
| Standard Error           | 0,442  |
| Median                   | 19,1   |
| Mode                     | 19,42  |
| Standard Deviation       | 3,16   |
| Sample Variance          | 9,986  |
| Kurtosis                 | 0,852  |
| Skewness                 | -0,227 |
| Range                    | 17,19  |
| Minimum                  | 9,868  |
| Maximum                  | 27,06  |
| Sum                      | 968    |
| Count                    | 51     |
| Confidence Level (95,0%) | 0,889  |

Giải thích:

- Mean: Số trung bình.
- Standard Error: Sai số của số trung bình mẫu.
- Median: Trung vị mẫu.
- Mode: Trị số ứng với tần số phân bố tập trung nhất.
- Standard deviation: Sai tiêu chuẩn mẫu.
- Sample variance: Phương sai mẫu.
- Kurtosis: Độ nhọn của phân bố
  - Ku = 0 phân bố thực nghiệm tiệm cận chuẩn.
  - Ku > 0 đường cong có dạng bẹt hơn so với phân bố chuẩn.
  - Ku < 0 đường cong có đỉnh nhọn hơn so với phân bố chuẩn.
  - Ku = Kurt(A2:A52) = 0.852. Đỉnh đường cong thấp hơn so với phân bố chuẩn.
- Skewness: Độ lệch của phân bố.
  - S<sub>k</sub> = 0 phân bố đối xứng.
  - S<sub>k</sub> > 0 đỉnh đường cong lệch trái so với số trung bình.
  - S<sub>k</sub> < 0 đỉnh đường cong lệch phải so với số trung bình.
  - S<sub>k</sub> = Skew(A2:A52) = -0.227. Đường cong hơi lệch phải.
- Minimum: Trị số quan sát bé nhất.
- Maximum: Trị số quan sát lớn nhất.
- Sum: Tổng các trị số quan sát.
- Count: Dung lượng mẫu.
- Confidence level (95%): Sai số tuyệt đối của ước lượng với độ tin cậy 95%.

Trong kết quả phân tích đặc trưng mẫu nói trên, ngoài các chỉ số phổ biến cần quan tâm như số trung bình, phương sai; thì hai giá trị quan trọng thuyết minh kiểu dạng phân bố của dữ liệu quan sát là Ku và Sk.

Khi Ku > 0 thì giá trị quan sát có xu hướng phân tán xa số trung bình, ngược lại Ku < 0 thì giá trị quan sát tập trung quanh số trung bình nhiều hơn. Khi Ku = 0 thì độ nhọn của số liệu quan sát tiệm cận chuẩn

Khi Sk > 0 thì số liệu quan sát có xu hướng nghiêng về các giá trị nhỏ hơn trung bình, nếu là số liệu sinh trưởng rừng, thì cây rừng đang ở giai đoạn non; ngược lại Sk < 0, giá trị quan sát thiên về các giá trị lớn hơn trung bình, nếu quan sát sinh trưởng rừng, thì đây là các khu rừng đã đi vào thành thực. ả ếu Sk = 0 thì độ lệch tiệm cận chuẩn.

Khi một mẫu có Ku = 0 và Sk = 0 thì nó có phân bố chuẩn.

ả ếu mẫu phân bố chưa chuẩn thì cần bổ sung mẫu theo công thức mẫu cần thiết nct:

$$nct \geq t^2 \cdot V\% / \Delta\%^2$$

Trong đó V% là hệ số biến động:  $V\% = \frac{S}{x_{bq}} 100$  và  $\Delta\%$  là sai số tương đối cho trước.

Giá trị Confidence Level (95%) cho phép ước lượng phạm vi biến động của số trung bình với độ tin cậy 95%:

$$P(\text{mean} - t.S/\sqrt{n} \leq \mu \leq \text{mean} + t.S/\sqrt{n}) = 0.95$$

trong đó  $t.S/\sqrt{n} = \text{Confidence Level (95\%)}$

Vì vậy giá trị biến động trung bình của tổng thể được ước lượng:

$$\mu = \text{mean} \pm \text{Confidence Level (95\%)}$$

Tùy theo yêu cầu của cuộc điều tra đánh giá, thí nghiệm mà chọn mức độ tin cậy khác nhau: 90%, 95%, 99%.

### 3 SO SÁNH 1 – 2 MẪU QUAN SÁT BẰNG TIÊU CHUẨN T

Kiểm tra mẫu bằng tiêu chuẩn t dựa vào giả thiết phân phối chuẩn của mẫu quan sát. Có hai loại kiểm tra t: kiểm tra t một mẫu (one-sample t-test), và t cho hai mẫu (two-sample t-test). Kiểm tra t một mẫu để đánh giá số trung bình của một mẫu có phải thật sự bằng một giá trị nào đó hay không?. Kiểm tra t hai mẫu thì để so sánh hai mẫu có cùng một luật phân phối, hay cụ thể hơn là hai mẫu có thật sự có cùng trị số trung bình hay không? Hay nói khác đi có sự sai khác giữa hai mẫu quan sát hay không?

#### 3.1 So sánh một mẫu với một giá trị cho trước – Kiểm tra T một mẫu

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?

Để giải quyết vấn đề này, sử dụng kiểm định t một mẫu. Theo lý thuyết thống kê công thức t kiểm tra một mẫu với một giá trị cho trước:

$$t = \frac{X_{bq} - \mu}{\frac{S}{\sqrt{n}}}$$

Trong đó,  $X_{bq}$  là giá trị trung bình của mẫu,  $\mu$  là trung bình theo giả thuyết, S là sai tiêu chuẩn và n là số lượng mẫu quan sát.

- ả ếu giá trị tuyệt |t| tính cao hơn giá trị t lý thuyết ở mức sai có ý nghĩa, 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. Và trong trường hợp này nếu t tính <0 thì có nghĩa trung bình của mẫu nhỏ thua có ý nghĩa so với trung bình lý thuyết, ngược lại nếu t tính > 0 thì trung bình của mẫu lớn hơn có ý nghĩa so với trung bình lý thuyết
- ả ếu |t| tính  $\leq t(0.05, df)$  thì có thể kết luận ở mức sai 5% trung bình mẫu quan sát xấp xỉ với trung bình lý thuyết.

Trong đó t lý thuyết được tính theo hàm =tinv(0.05, df), với độ tự do df = n-1.

#### Số liệu đo cao cây tái sinh rừng khộp trong Excel

| Stt | Chiều cao cây tái sinh (m) |
|-----|----------------------------|
| 1   | 1.5                        |
| 2   | 1.3                        |
| 3   | 0.8                        |
| 4   | 1.9                        |
| 5   | 1.7                        |
| 6   | 2.2                        |
| 7   | 2.5                        |
| 8   | 1.0                        |
| 9   | 0.7                        |
| 10  | 1.9                        |
| 11  | 1.8                        |

.....

|    |     |
|----|-----|
| 58 | 1.6 |
| 59 | 2.0 |
| 60 | 1.9 |
| 61 | 1.7 |

Để tính được giá trị t, cần tính toán đặc trưng mẫu để có các giá trị thông kê về  $\bar{X}$ , S.

### **Kết quả tính đặc trưng mẫu tái sinh rừng khộp**

*Chiều cao cây tái sinh (m)*

|                         |         |
|-------------------------|---------|
| Mean                    | 1.64    |
| Standard Error          | 0.06318 |
| Median                  | 1.7     |
| Mode                    | 1.9     |
| Standard Deviation      | 0.49347 |
| Sample Variance         | 0.24351 |
| Kurtosis                | -0.4499 |
| Skewness                | -0.4627 |
| Range                   | 1.8     |
| Minimum                 | 0.7     |
| Maximum                 | 2.5     |
| Sum                     | 100.3   |
| Count                   | 61      |
| Confidence Level(95.0%) | 0.12638 |

Từ đó tính giá trị thống kê t: So sánh trung bình chiều cao tái sinh với giá trị lý thuyết  $\mu = 2m$

$$t = \frac{1.64 - 2}{\frac{0.493}{\sqrt{61}}} = -5.63$$

Và t lý thuyết:  $t(0.05, df = n-1) = tinv(0.05, 60) = 2.00$

Kết quả cho thấy  $|t| = 5.63 > t(0.05, 60)$ . Kết luận: Có sự sai khác có ý nghĩa giữa trung bình chiều cao cây tái sinh rừng khộp với giá trị trung bình lý thuyết mong đợi là 2m. Và  $t < 0$  do đó có nghĩa là chiều cao trung bình cây tái sinh nhỏ thua có ý nghĩa khi so với chiều cao mong đợi là 2m; hay nói khác nếu với yêu cầu cao trên 2m thì mới thoát được ảnh hưởng của lửa rừng, thì lô rừng này cây tái sinh chưa đạt được.

### **3.2 So sánh sự sai khác giữa trung bình 2 mẫu – Kiểm tra T 2 mẫu**

Trong các thí nghiệm thường người ta cần so sánh kết quả của 2 công thức, ví dụ: Bón phân hay không bón, che bóng hay không che, sinh trưởng, tái sinh của cây rừng nơi được chăm sóc và nơi không, sinh trưởng cây rừng nơi cháy và không cháy.....Việc kiểm tra tiến hành theo 2 mẫu trên cơ sở so sánh 2 số trung bình bằng các tiêu chuẩn t.

**Công thức tính giá trị kiểm tra t:**

$$t = \frac{X_1 - X_2}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Với:  $X_1, X_2$ : 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 2 mẫu 1 và 2.

Giá trị t tính lớn hơn t bảng với  $\alpha=0.05$  và độ tự do  $K=n_1+n_2-2$  thì bác bỏ giả thuyết  $H_0$ , có nghĩa trung bình 2 mẫu sai khác có ý nghĩa, và người ta sẽ chọn mẫu có trung bình cao.

**Trước khi sử dụng tiêu chuẩn t, cần kiểm tra 2 điều kiện:**

- Hai mẫu có phân bố chuẩn.
- Phương sai của hai mẫu có bằng nhau hay không

□ **Hai mẫu có phân bố chuẩn:** Có thể vẽ biểu đồ phân bố tần số từng mẫu, hoặc dựa vào độ lệch và độ nhọn để xem xét có tiệm cận chuẩn hay không. Trong nhiều trường hợp thống kê về xã hội lần tự nhiên, người ta phải rút mẫu đủ lớn để bảo đảm tiêu chí này. Khi dung lượng mỗi mẫu  $>30$  thì có thể xem là tiệm cận chuẩn.

□ **Kiểm tra bằng nhau của 2 phương sai của 2 mẫu bằng tiêu chuẩn F.**

Trước khi chọn lựa tiêu chuẩn t để so sánh trung bình 2 mẫu, cần kiểm tra sự sai khác phương sai của chúng bằng tiêu chuẩn F.

Ví dụ: Kiểm tra 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 Lang Hanh-Lâm Đồng: Mỗi công thức được rút mẫu theo ô tiêu chuẩn  $1000m^2$ , đo đếm chiều cao:

- Dung lượng quan sát mỗi mẫu  $>90$  cây, nên chấp nhận giả thuyết phân bố ngẫu nhiên của từng mẫu tiệm cận chuẩn.
- Kiểm tra bằng nhau của 2 phương sai bằng tiêu chuẩn F:

**Bảng tóm tắt số liệu sinh trưởng H của hai mẫu**

|          | A                  | B                  |
|----------|--------------------|--------------------|
| <b>1</b> | <b>H (cây con)</b> | <b>H (rễ trần)</b> |
| 2        | 13,6               | 13                 |
| 3        | 14                 | 13,5               |
|          | 13,8               | 12                 |
|          | 13                 | 13,5               |
|          | 11                 | 15                 |
|          | 12                 | 14                 |
| 93       | 12,5               | 10                 |
| 94       |                    | 9                  |

Tính F: Một trong 2 cách:

*C1: Kích nút fx, có hộp thoại:* Chọn: Statistical (trong Function Category) và Ftest- $\alpha$  ext (trong Function name): Xuất hiện hộp thoại tiếp theo:

Array 1: Đưa vào dãy 1: A2:A93

Array 2: Đưa vào dãy 2: B2:B94  
Finish.

C2: Đưa đến ô kết quả: =Ftest(A2:A93,B2:b94) Enter.

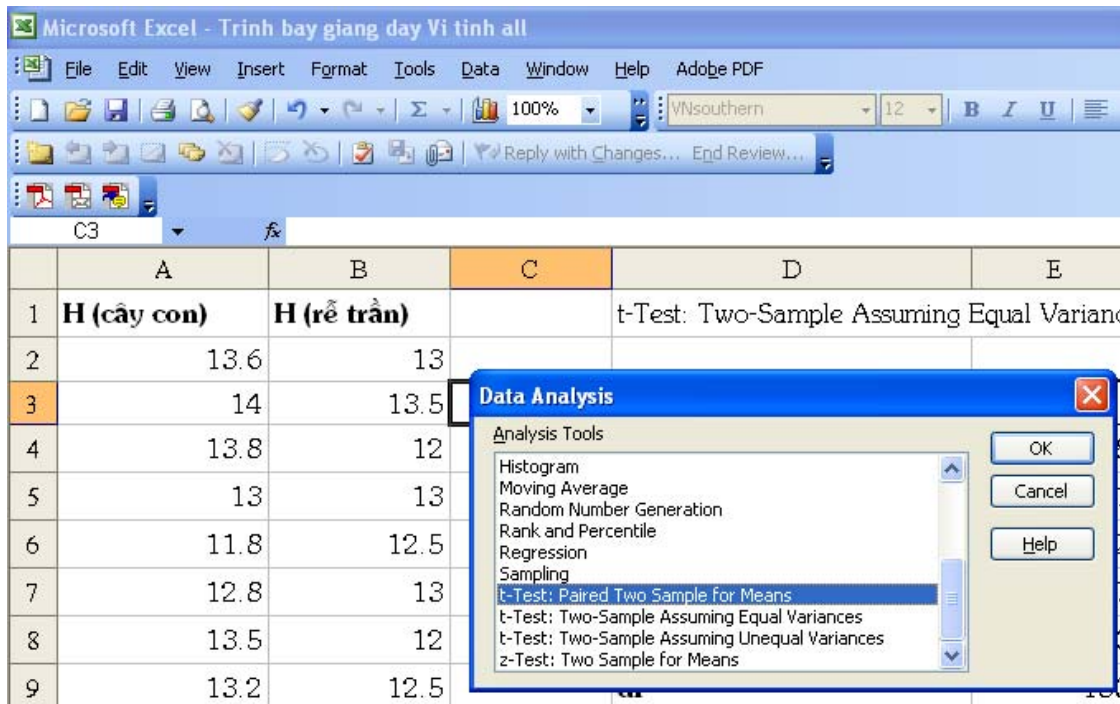
Nếu giá trị xác suất  $P > 0.05$ , kết luận hai phương sai bằng nhau, nếu ngược lại thì bác bỏ.

Kết quả ví dụ trên có  $P=0.40 > 0.05$ , kết luận phương sai hai mẫu bằng nhau (chưa có sai dị rõ).

• **Dùng tiêu chuẩn t để kiểm tra giả thuyết Ho theo trình tự:**

Trong menu Tools/Data Analysis: Chọn trong hộp thoại một trong hai trường hợp tùy theo phương sai hai mẫu có bằng nhau hay không qua kiểm tra bằng F ở bước trước

- *t-Test: Two sample assuming equal variance (Trường hợp phương sai bằng nhau).*
- *t-Test: Two sample assuming unequal variance (Trường hợp phương sai không bằng nhau).*



Trong Hộp thoại: Xác định:

- Variable 1 range: Khối dữ liệu mẫu 1 (A1:A93)
- Variable 2 range: Khối dữ liệu mẫu 2 (B1:B94)  
    ả ên đưa cả tiêu đề.
- Hypothesized mean diference: Đưa vào 0 (Có nghĩa giả thuyết  $H_0=0$ ). Có thể thay đổi giả thuyết này theo yêu cầu đánh giá thí nghiệm. Ví dụ nếu muốn kiểm tra xem hai trung bình của hai mẫu có thực sự sai khác nhau không, giả thuyết  $H_0$  sẽ là:  $H_0: \text{Mean1} = \text{Mean2}$  hay nói khác  $\text{Mean1} - \text{Mean2} = 0$ , lúc này giả thuyết  $H_0$  được đặt giá trị là 0.
- Label: ả êu có đưa hàng tiêu đề vào thì cần đánh dấu vào label
- Output range: Đưa địa chỉ ô trên trái nơi xuất kết quả.
- OK.



ả ếu:  $P(T \leq t)$  two tail (hai chiều)  $< 0.05$ , bác bỏ  $H_0$ , có nghĩa 2 mẫu sai dị rõ, ngược lại thì trung bình hai mẫu chưa có sai khác.

Hoặc  $|t \text{ Stat}| > t \text{ Critical two tail}$  (t hai chiều), bác bỏ  $H_0$ , hai mẫu sai dị rõ, ngược lại thì sai khác là ngẫu nhiên.

t-Test: Two-Sample Assuming Equal Variances

|                              | <i>H (cây con)</i> | <i>H (rễ trần)</i> |
|------------------------------|--------------------|--------------------|
| Mean                         | 11,60434783        | 13,40322581        |
| Variance                     | 2,559761108        | 2,148141655        |
| Observations                 | 92                 | 93                 |
| Pooled Variance              | 2,352826738        |                    |
| Hypothesized Mean Difference | 0                  |                    |
| df                           | 183                |                    |
| t Stat                       | -7,975469453       |                    |
| P(T<=t) one-tail             | 7,98781E-14        |                    |
| t Critical one-tail          | 1,653222625        |                    |
| P(T<=t) two-tail             | 1,59756E-13        |                    |
| t Critical two-tail          | 1,973012331        |                    |

Từ 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, do vậy phương pháp trồng thông 3 lá bằng rễ trần cần được ứng dụng trong thực tiễn.

#### 4 NGHIÊN CỨU MỐI QUAN HỆ SINH THÁI LOÀI TRONG RỪNG MƯA NHIỆT ĐỚI DỰA VÀO TIÊU CHUẨN $\chi^2$ (Bảo Huy, 1997)

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:

- Phục vụ việc “đơn giản hóa tổ thành”, xác định việc nên giữ lại và đào thải loài cây nào trong thiết kế nuôi dưỡng, khai thác rừng tự nhiên.
- Định hướng trong việc lựa chọn nhóm loài cây hỗn giao trong trồng rừng, làm giàu rừng.

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. 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 nghiên cứu gồm có các bước chính:

- Xác định diện tích biểu hiện loài
- 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 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.

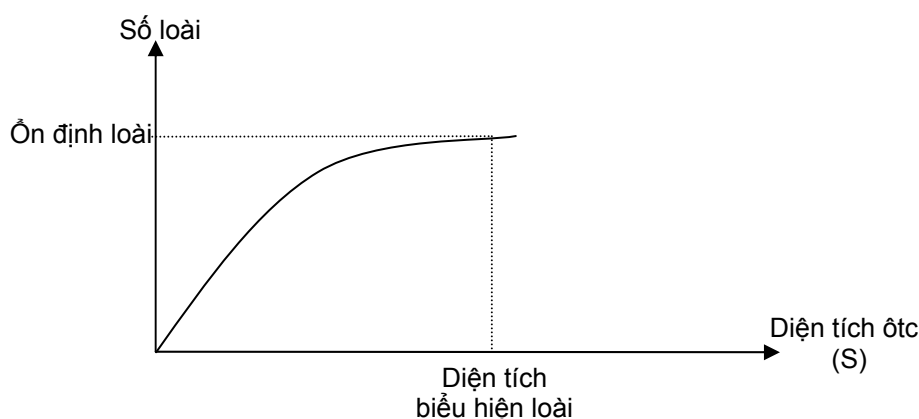
Nguyên tắc nghiên cứu trong trường hợp này là xác định một diện tích ô mẫu nhỏ nhất nhưng bảo đảm xuất hiện các loài ưu thế sinh thái.

Thu thập số loài theo ô tiêu chuẩn diện tích thay đổi (Theo từng loại: Cây gỗ, cây tái sinh), diện tích ô bắt đầu là 100m<sup>2</sup> và tăng dần đến khoảng 1 – 2ha.

**Số loài theo kích thước ô mẫu tăng dần**

| Stt | Diện tích ô tiêu chuẩn (m <sup>2</sup> ) | Số loài | Tên loài |
|-----|--|---------|----------|
| 1   | 100                                      | 5       | a, d, hg |
| 2   | 100                                      | 4       |          |
| 3   | 100                                      | 6       |          |
| 4   | ,,,                                      | ,,,     | ,,,      |
| 5   | 200                                      | 6       |          |
| 6   | 200                                      | 8       |          |
| 7   | ,,,                                      | ,,,     | ,,,      |
| 8   | 400                                      | 10      |          |
| 9   | 400                                      | 11      |          |
| 10  | ,,,                                      | ,,,     |          |
| 11  | 600                                      |         |          |
|     |  |         |          |
|     |  |         |          |
| n   | 2500                                     | 18      |          |

Có thể biểu thị việc xác định diện tích biểu hiện loài bằng đồ thị sau:



### Xác định diện tích biểu hiện loài

Mô phỏng quan hệ:  $\hat{a}$  (số loài) =  $f(S=\text{diện tích ô tiêu chuẩn})$ , dạng quan hệ sau có thể được sử dụng:

$$N = a.e^{-b.S^{-m}}$$

$$\lim_{\text{Khi } S \rightarrow +\infty} N = a.e^{-b.S^{-m}} = a$$

Ví dụ: Tiến hành điều tra thử nghiệm 53 ô có diện tích từ 100m<sup>2</sup> đến 10.000m<sup>2</sup> 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 ( $\hat{a}$ ) theo diện tích ô (S) bằng một dạng hàm mũ cơ số e. Kết quả đã ước lượng các tham số:

$$N = 16.810.e^{-6.900.S^{-0.246}}$$

Với n=53      R=0.907      Fr=722.58       $\alpha < 0.01$

Phương trình đạt hệ số tương quan cao chứng tỏ có mối liên hệ chặt chẽ giữa  $\hat{a}$  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  $a = 16.810$ .  $\hat{a}$  hư 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.

Đặc biệt một số loài có hệ số tổ thành gần như tuyệt đối. Các loài phổ biến của rừng khộp: Cà chấu (*Shorea obtusa*); Cẩm liên (*Pentacme siamensis*), Dầu đồng (*Dipteocarpus tuberculatus*), Dầu trà beng (*Dipterocarpus obtusifolius*), Chiêu liêu (*Terminalia myrocarpa*) và một số loài thuộc loài khác có tỷ lệ thấp hơn trong tổ thành.  $\hat{a}$  hư vậy số loài phổ biến trên một đơn vị diện tích rừng khộp chỉ khoảng 5-6 loài.

Từ phương trình, thế giá trị  $\hat{a} = 6$  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. Diện tích biểu hiện trong trường hợp này là  $S = 2.500\text{m}^2$ . Vậy có thể chọn ô hình vuông với kích thước 50x50m.

### ii) Dự báo mối quan hệ sinh thái giữa các loài

Trên cơ sở đã xác định được diện tích ô biểu hiện sinh thái loài; tiếp tục xác định dung lượng mẫu (số ô tiêu chuẩn) cho từng sinh cảnh theo công thức:

$$N_{ct} \geq \frac{t^2 \cdot V\%}{\Delta\%} \quad \text{Trong đó: } t = 1,96 \text{ khi độ tin cậy là } 95\%$$

V%: hệ số biến động về số loài, được tính theo công thức:

$$V\% = \frac{S}{\bar{X}} \times 100 \quad S = \sqrt{\frac{\left( \sum x^2 - \frac{(\sum x)^2}{n} \right)}{n-1}}$$

S: sai tiêu chuẩn mẫu

n: số ô rút mẫu thử (thường chọn  $n \geq 30$ )

x: số loài trên mỗi ô

$\Delta\%$ : sai số cho phép từ 5% - 10%.

Thường rút thử 30 ô để điều tra, nếu số liệu ghi nhận không đảm bảo dung lượng mẫu cần thiết theo công thức trên thì cần phải tiến hành điều tra bổ sung, ngược lại thì việc điều tra bổ sung không còn cần thiết.

Sau khi xác định số lượng ô tiêu chuẩn rút mẫu thử, 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, trong đó tập trung xác định tên loài xuất hiện

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 IV% > 5% hoặc tần suất F% > 5%.

Ví dụ: Từ 32 ô tiêu chuẩn được rút mẫu ngẫu nhiên trong rừng thường xanh khu vực Đắk RLấp, thống kê được tần suất xuất hiện của các loài chủ yếu:

#### Tần suất xuất hiện các loài

| Stt | Loài         |                          | Tần số xuất hiện | Tần suất (F%)<br>% |
|-----|--------------|--------------------------|------------------|--------------------|
|     | Tên Việt Nam | Tên Khoa Học             |                  |                    |
| 1   | Dẻ           | Lithocarpus sp           | 30               | 13.0               |
| 2   | Bằng lăng    | Lagerstroemia calyculata | 27               | 11.7               |
| 3   | Xương cá     | Canthium didyllum        | 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 xốt      | Schima superba           | 15               | 6.5                |
| 8   | Vạng trứng   | Endospermum chinense     | 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                |

| Stt | Loài         |                        | Tần số xuất hiện | Tần suất (F%)<br>% |
|-----|--------------|------------------------|------------------|--------------------|
|     | Tên Việt Nam | Tên Khoa Học           |                  |                    |
| 14  | Thường 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 malabarica | 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 biểu trên cho thấy trong các loài chủ yếu, có 9 loài có tần suất > 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.

Từ ô tiêu chuẩn có diện tích biểu hiện được rút mẫu ngẫu nhiên, tiến hành kiểm tra quan hệ cho từng cặp loài theo tiêu chuẩn  $\rho$  và  $\chi^2$ .

Sử dụng các tiêu chuẩn thống kê sau để đánh giá quan hệ theo từng cặp loài:

**$\rho$ : Hệ số tương quan giữa 2 loài A và B.**

$$\rho = \frac{P(AB) - P(A).P(B)}{\sqrt{P(A).(1 - P(A)).P(B).(1 - P(B))}}$$

Trong đó:

$\rho = 0$  : 2 loài A và B độc lập nhau.

$0 < \rho \leq 1$ : loài A và B liên kết dương.

$-1 \leq \rho < 0$ : 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.

$$P(AB) = \frac{n_{AB}}{n} \quad P(A) = \frac{n_A + n_{AB}}{n} \quad P(B) = \frac{n_B + n_{AB}}{n}$$

Với:

$n_A$ : số ô tiêu chuẩn chỉ xuất hiện loài A.

$n_B$ : số ô tiêu chuẩn chỉ xuất hiện loài B.

$n_{AB}$ : số ô tiêu chuẩn xuất hiện đồng thời 2 loài A và B.

$n$ : tổng số ô quan sát ngẫu nhiên.

$\rho$  nói lên chiều hướng liên hệ và mức độ liên hệ giữa 2 loài.  $\rho < 0$ : 2 loài liên kết âm và  $|\rho|$  càng lớn thì mức độ bài xích nhau càng mạnh, ngược lại  $\rho > 0$ : 2 loài liên kết dương và  $|\rho|$  càng lớn thì mức độ hỗ trợ nhau càng cao.

Trong trường hợp  $|\rho|$  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  $2 \times 2$ :

Việc kiểm tra mối quan hệ giữa 2 loài A và B được thực hiện bằng tiêu chuẩn  $\chi^2$ :

$$\chi^2 = \frac{(|ad - bc| - 0.5)^2 n}{(a+b).(c+d).(a+c).(b+d)}$$

Trong đó:

$a = n_{AB}$  ;  $b = n_B$ ;  $c = n_A$ ;  $d$ : số ô không chứa cả 2 loài a và B.

$\chi^2$  tính được ở công thức trên được so sánh với  $\chi^2_{0.05}$  ứng với bậc tự do  $K=1$   $\chi^2_{0.05, K=1} = 3.84$

Nếu  $\chi^2 \leq \chi^2_{0.05} = 3.84$  thì mối quan hệ giữa 2 loài là ngẫu nhiên.

Nếu  $\chi^2 > \chi^2_{0.05} = 3.84$  thì giữa 2 loài có quan hệ với nhau.

Tóm lại để xem xét mối quan hệ theo từng cặp loài, sử dụng đồng thời 2 tiêu chuẩn  $\rho$  và  $\chi^2$ :

$\chi^2$ : để kiểm tra mối quan hệ từng cặp loài.

$\rho$ : trong trường hợp kiểm tra bằng  $\chi^2$  cho thấy có quan hệ, thì  $\rho$  sẽ cho biết chiều hướng mối quan hệ đó theo dấu của  $\rho$  (- hay +) và mức độ quan hệ qua giá trị  $|\rho|$ .

### Kiểm tra quan hệ theo từng cặp loài

| Stt | Loài A          | Loài B           | nA(c) | nB(b) | nAB(a) | nAB-(d) | P(A)         | P(B)         | P(AB)        | $\rho$       | $\chi^2$    | Quan hệ                 |
|-----|-----------------|------------------|-------|-------|--------|---------|--------------|--------------|--------------|--------------|-------------|-------------------------|
| 1   | <b>Xoan Mọc</b> | <b>Bằng Lãng</b> | 5     | 13    | 14     | 0       | 0.594        | 0.844        | 0.438        | -0.356       | 3.99        | <b>Có quan hệ âm</b>    |
| 2   | Xoan Mọc        | Dê               | 0     | 11    | 19     | 2       | 0.594        | 0.938        | 0.594        | 0.312        | 3.04        | Ngẫu nhiên              |
| 3   | Xoan Mọc        | Bời Lờ           | 7     | 6     | 12     | 7       | 0.594        | 0.563        | 0.375        | 0.168        | 0.89        | Ngẫu nhiên              |
| 4   | Xoan Mọc        | Vạng Trúng       | 10    | 5     | 9      | 8       | 0.594        | 0.438        | 0.281        | 0.088        | 0.24        | Ngẫu nhiên              |
| 5   | Xoan Mọc        | Trâm             | 10    | 5     | 9      | 8       | 0.594        | 0.438        | 0.281        | 0.088        | 0.24        | Ngẫu nhiên              |
| 6   | Xoan Mọc        | Xương cá         | 5     | 9     | 14     | 4       | 0.594        | 0.719        | 0.438        | 0.049        | 0.07        | Ngẫu nhiên              |
| 7   | Xoan Mọc        | Bồ hòn           | 10    | 7     | 9      | 6       | 0.594        | 0.500        | 0.281        | -0.064       | 0.12        | Ngẫu nhiên              |
| 8   | Xoan Mọc        | Chò xốt          | 12    | 8     | 7      | 5       | 0.594        | 0.469        | 0.219        | -0.243       | 1.86        | Ngẫu nhiên              |
| 9   | Bằng Lãng       | Dê               | 2     | 5     | 25     | 0       | 0.844        | 0.938        | 0.781        | -0.111       | 0.36        | Ngẫu nhiên              |
| 10  | Bằng Lãng       | Bời Lờ           | 13    | 4     | 14     | 2       | 0.844        | 0.563        | 0.438        | -0.206       | 0.40        | Ngẫu nhiên              |
| 11  | Bằng Lãng       | Vạng Trúng       | 16    | 3     | 11     | 2       | 0.844        | 0.438        | 0.344        | -0.141       | 0.61        | Ngẫu nhiên              |
| 12  | Bằng Lãng       | Trâm             | 14    | 1     | 13     | 4       | 0.844        | 0.438        | 0.406        | 0.206        | 1.32        | Ngẫu nhiên              |
| 13  | Bằng Lãng       | Xương cá         | 9     | 5     | 18     | 0       | 0.844        | 0.719        | 0.563        | -0.269       | 2.27        | Ngẫu nhiên              |
| 14  | Bằng Lãng       | Bồ hòn           | 13    | 2     | 14     | 3       | 0.844        | 0.500        | 0.438        | 0.086        | 0.22        | Ngẫu nhiên              |
| 15  | Bằng Lãng       | Chò xốt          | 13    | 1     | 14     | 4       | 0.844        | 0.469        | 0.438        | 0.232        | 1.68        | Ngẫu nhiên              |
| 16  | Dê              | Bời Lờ           | 14    | 2     | 16     | 0       | 0.938        | 0.563        | 0.500        | -0.228       | 1.60        | Ngẫu nhiên              |
| 17  | Dê              | Vạng Trúng       | 18    | 2     | 12     | 0       | 0.938        | 0.438        | 0.375        | -0.293       | 2.67        | Ngẫu nhiên              |
| 18  | Dê              | Trâm             | 17    | 1     | 13     | 1       | 0.938        | 0.438        | 0.406        | -0.033       | 0.03        | Ngẫu nhiên              |
| 19  | <b>Dê</b>       | <b>Xương cá</b>  | 7     | 0     | 23     | 2       | <b>0.938</b> | <b>0.719</b> | <b>0.719</b> | <b>0.413</b> | <b>5.33</b> | <b>Có quan hệ dương</b> |
| 20  | Dê              | Bồ hòn           | 14    | 0     | 16     | 2       | 0.938        | 0.500        | 0.500        | 0.258        | 2.07        | Ngẫu nhiên              |
| 21  | Dê              | Chò xốt          | 16    | 1     | 14     | 1       | 0.938        | 0.469        | 0.438        | -0.016       | 0.00        | Ngẫu nhiên              |
| 22  | Bời lờ          | Vạng Trúng       | 11    | 7     | 7      | 7       | 0.563        | 0.438        | 0.219        | -0.111       | 0.38        | Ngẫu nhiên              |
| 23  | <b>Bời lờ</b>   | <b>Trâm</b>      | 7     | 3     | 11     | 11      | <b>0.563</b> | <b>0.438</b> | <b>0.344</b> | <b>0.397</b> | <b>4.99</b> | <b>Có quan hệ dương</b> |
| 24  | Bời lờ          | Xương cá         | 5     | 10    | 13     | 4       | 0.563        | 0.719        | 0.406        | 0.009        | 0.00        | Ngẫu nhiên              |
| 25  | Bời lờ          | Bồ hòn           | 11    | 9     | 7      | 5       | 0.563        | 0.500        | 0.219        | -0.252       | 2.00        | Ngẫu nhiên              |

| Stt | Loài A     | Loài B   | nA(c) | nB(b) | nAB(a) | nAB-(d) | P(A)  | P(B)  | P(AB) | $\rho$ | $\chi^2$ | Quan hệ          |
|-----|------------|----------|-------|-------|--------|---------|-------|-------|-------|--------|----------|------------------|
| 26  | Bời lồi    | Chò xốt  | 13    | 10    | 5      | 4       | 0.563 | 0.469 | 0.156 | -0.434 | 5.97     | Có 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 xốt  | 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 xốt  | 11    | 12    | 3      | 6       | 0.438 | 0.469 | 0.094 | -0.450 | 6.42     | Có quan hệ dương |
| 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 xốt  | 16    | 8     | 7      | 1       | 0.719 | 0.469 | 0.219 | -0.527 | 8.80     | Có quan hệ âm    |
| 36  | Bồ hòn     | Chò xốt  | 9     | 8     | 7      | 8       | 0.500 | 0.469 | 0.219 | -0.063 | 0.12     | Ngẫu nhiên       |

Từ kết quả này có thể xác định được:

- **Các loài có quan hệ dương:**  $\chi^2_t > \chi^2_{0.05} = 3.84$  và  $\rho > 0$ : Các loài này nên được lựa chọn để trồng hỗn giao, hoặc làm giàu rừng
- **Các loài có quan hệ âm:**  $\chi^2_t > \chi^2_{0.05} = 3.84$  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:  $\chi^2_t \leq \chi^2_{0.05} = 3.84$ : 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.

## 5 PHÂN TÍCH PHƯƠNG SAI

Phân tích phương sai là một trong những phương pháp phân tích thống kê quan trọng, đặc biệt là trong các thí nghiệm giống, thí nghiệm các nhân tố tác động đến hiệu quả, chất lượng của cây trồng, vật nuôi, gieo ươm, kiểm nghiệm xuất xứ cây trồng. Chủ yếu đánh giá ảnh hưởng của các công thức, nhân tố đến kết quả thí nghiệm, làm cơ sở cho việc lựa chọn công thức, phương pháp tối ưu trong nông lâm nghiệp.

**Điều kiện để phân tích phương sai là:**

- Các giá trị quan sát trong từng ô thí nghiệm có phân bố chuẩn:
  - o Kiểm tra bằng đặc trưng mẫu, sơ đồ.
  - o ầu dung lượng quan sát đủ lớn ( $n > 30$ ) thì chấp nhận giả thuyết phân bố chuẩn.
- Các phương sai của từng nhân tố bằng nhau: Kiểm tra bằng tiêu chuẩn Cochran (nếu số lần lặp lại bằng nhau), bằng tiêu chuẩn Bartlett (nếu số lần lặp của các công thức không bằng nhau).

### 5.1. Phân tích phương sai 1 nhân tố với các thí nghiệm ngẫu nhiên hoàn toàn

Phân tích này có một nhân tố như xuất xứ cây trồng, mật độ trồng khác nhau, chế độ chăm sóc khác nhau, .... Có nghĩa trong đó có a công thức, mỗi công thức được lặp lại m lần, số lần lặp của mỗi công thức có thể bằng hoặc không bằng nhau.

Trong trường hợp này có thể sử dụng chương trình phân tích phương sai một nhân tố để kiểm tra ảnh hưởng của các công thức đến kết quả thí nghiệm.

**Cách bố trí thí nghiệm trên hiện trường để phân tích phương sai 1 nhân tố**

| Các công thức của 1 nhân tố | Số lần lặp lại |     |     |     |
|-----------------------------|----------------|-----|-----|-----|
|                             | 1              | 2   | 3   | m   |
| 1                           | 11             | 12  | 13  | 1m  |
| 2                           | 21             | 22  |     |     |
| ....                        | ...            | ... | ... | ... |
| a                           | a1             | a2  |     | am  |

Ví dụ: Đánh giá kết quả khảo nghiệm xuất xứ *Pinus caribaea* tại Lang Hanh-Lâm Đồng. Theo dự kiến sẽ có 10 xuất xứ *P.caribaea* được trồng khảo nghiệm tại trạm thực nghiệm Lang Hanh năm 1991. Việc bố trí thí nghiệm ban đầu đã dự kiến tiến hành theo khối ngẫu nhiên đầy đủ RCB (Randomized Complete Blocks), bao gồm 10 công thức chỉ thị 10 xuất xứ và được lặp lại ở 4 khối.

Ả hưng trong quá trình triển khai trồng thực nghiệm, chỉ còn lại 7 xuất xứ và chỉ có 5 xuất xứ lặp lại đủ 4 lần, còn 2 xuất xứ chỉ được lặp lại 2 lần.

*7 xuất xứ P.caribaea được trồng thực tế, được đánh số và lặp lại như sau:*

- 1: Xuất xứ P.alamicamba (Ả IC) lặp lại 4 lần.
- 2: P.poptun (Guat) “ 4 “
- 3: P.guanaja (Ả ondures) “ 4 “
- 4: P.linures (Ả ondures) “ 4 “
- 5: P.R482 (Australia) “ 2 “
- 6: P.T473 (Australia) “ 4 “
- 8: P.little asaco (Bahamas) 2 “

- Mỗi xuất xứ ứng với 1 lần lặp được trồng 25 cây, với cự ly 3x2m, tổng diện tích bố trí thí nghiệm là 1ha.
- Các điều kiện đất đai, vi 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  $D_{1,3}$ , H,  $D_t$ , phẩm chất, tia cành, hình thân. Sử dụng 2 chỉ tiêu  $D_{1,3}$  và H để đánh giá sinh trưởng của các xuất xứ thử nghiệm.

Dùng phân tích phương sai để đánh giá sự sai khác về sinh trưởng ở các xuất xứ

Trước hết đã kiểm tra 2 điều kiện để phân tích phương sai:

- Điều kiện phân bố chuẩn: Các giá trị quan sát ở từng ô thí nghiệm qua kiểm tra bằng biểu đồ đều có dạng tiệm cận chuẩn nên chấp nhận giả thuyết phân bố chuẩn.
- Phương sai bằng nhau: Do dung lượng mẫu ở các xuất xứ không bằng nhau nên dùng tiêu chuẩn Bartlett để kiểm tra, kết quả tính được:

$$X^2 = 3,73 < X^2(0,05; 6) = 12,59$$

Do đó chấp nhận giả thuyết bằng nhau của các phương sai mẫu.

Ả hưng vậy 2 điều kiện trên là thỏa mãn để tiến hành phân tích phương sai.

Dùng phân tích phương sai 1 nhân tố để kiểm tra. Trong đó nhân tố là Xuất xứ với 7 công thức:



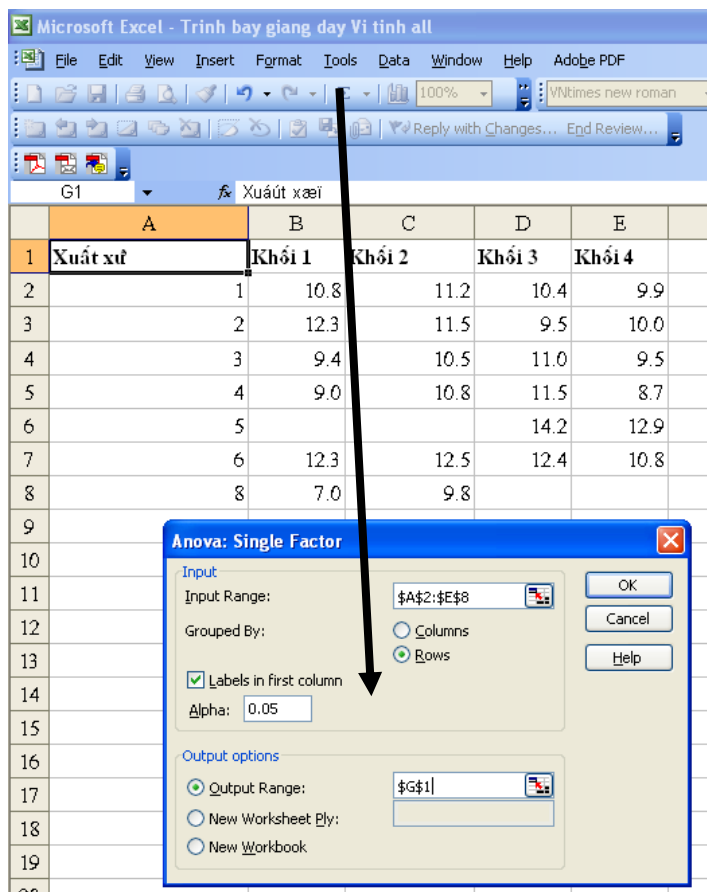
**Giá trị  $D_{1,3}$  (cm) bình quân ứng với từng ô thí nghiệm của các Xuất xứ theo khối (lần lặp lại)**

|   | A       | B     | C     | D     | E     |
|---|---------|-------|-------|-------|-------|
| 1 | Xuất xứ | Lặp 1 | Lặp 2 | Lặp 3 | Lặp 4 |
| 2 | 1       | 10.8  | 11.2  | 10.4  | 9.9   |
| 3 | 2       | 12.3  | 11.5  | 9.5   | 10.0  |
| 4 | 3       | 9.4   | 10.5  | 11.0  | 9.5   |
| 5 | 4       | 9.0   | 10.8  | 11.5  | 8.7   |
| 6 | 5       |       |       | 14.2  | 12.9  |
| 7 | 6       | 12.3  | 12.5  | 12.4  | 10.8  |
| 8 | 8       | 7.0   | 9.8   |       |       |

Phân tích phương sai 1 nhân tố:

Vào menu Tools/Data Analysis/Anova: Chọn Single Factor có được Hộp thoại:

- Input range: Nhập địa chỉ khối dữ liệu. Vd: A2:E8. (Có cột đầu chứa số hiệu công thức, nhưng bỏ hàng đầu).
- Grouped by: Chọn Columns hoặc Rows.
- Đánh dấu vào Label in first column (row).
- Output range: Nhập địa chỉ ô trên trái nơi xuất kết quả.
- Kích OK.



## Kết quả phân tích phương sai 1 nhân tố

Anova: Single Factor

### SUMMARY

| Groups | Count | Sum  | Average | Variance |
|--------|-------|------|---------|----------|
| 1      | 4     | 42.3 | 10.6    | 0.299523 |
| 2      | 4     | 43.2 | 10.8    | 1.703825 |
| 3      | 4     | 40.3 | 10.1    | 0.616404 |
| 4      | 4     | 40.0 | 10.0    | 1.780196 |
| 5      | 2     | 27.1 | 13.5    | 0.797116 |
| 6      | 4     | 48.1 | 12.0    | 0.673895 |
| 8      | 2     | 16.7 | 8.4     | 3.903367 |

### AẢ OVA

| Source of Variation | SS       | df | MS       | F        | P-value  | F crit   |
|---------------------|----------|----|----------|----------|----------|----------|
| Between Groups      | 37.53507 | 6  | 6.255846 | 5.338286 | 0.002925 | 2.698656 |
| Within Groups       | 19.92201 | 17 | 1.171883 |          |          |          |
| Total               | 57.45708 | 23 |          |          |          |          |

Từ bảng AẢ OVA nhận được: Đối với các xuất xứ khác nhau:  $F = 5,33 > F_{(0,05)} = 2,69$ . Kết luận: Các xuất xứ khác nhau có sự sai khác về sinh trưởng đường kính. ả ếu ngược lại thì kết luận rằng giữa các xuất xứ chưa có sự sai khác về sinh trưởng

Sinh trưởng bình quân đường kính các xuất xứ theo thứ tự từ cao đến thấp ở bảng sau:

#### Thứ tự sinh trưởng đường kính từ tốt đến xấu

| Xuất xứ | ni | D1,3 tb (cm) |
|---------|----|--------------|
| 5       | 2  | 13.5         |
| 6       | 4  | 12.0         |
| 2       | 4  | 10.8         |
| 1       | 4  | 10.6         |
| 3       | 4  | 10.1         |
| 4       | 4  | 10.0         |
| 8       | 2  | 8.4          |

Tiếp theo dùng tiêu chuẩn t để so sánh sinh trưởng đường kính lớn nhất của xuất xứ 5 với các xuất xứ có đường kính lần lượt nhỏ hơn:

- So sánh D bình quân giữa xuất xứ 5 với 6:  $t = 1,61 < t_{(0,05 ; k=17)} = 2,11$ .  
Kết luận: Sinh trưởng D giữa xuất xứ 5 và 6 chưa có sai khác rõ rệt.
- So sánh D bình quân giữa xuất xứ 5 với 2:  $t = 2,90 > t_{(0,05 ; k=17)} = 2,11$ .  
Kết luận: Sinh trưởng D giữa xuất xứ 5 và 2 có sai khác rõ rệt.

ả hư vậy, xét theo chỉ tiêu đường kính, xuất xứ tối ưu trong 7 xuất xứ khảo nghiệm là 5 và 6, hai xuất xứ này có chỉ tiêu D lớn nhất, chưa có sai dị với nhau và có sai khác rõ rệt với các xuất xứ còn lại. Đó là 2 xuất xứ: P.R482 (Australia) và P.T473 (Australia).

## 5.2. Phân tích phương sai 2 nhân tố

Trong các thí nghiệm người ta thường so sánh và phân tích tác động đồng thời 2 nhân tố lên kết quả thí nghiệm như: năng suất, sinh khối... Phân tích phương sai lúc này chia 2 trường hợp: Hai nhân tố với một lần lặp và Hai nhân tố với nhiều lần lặp lại.

### 5.2.1. Phân tích phương sai 2 nhân tố với 1 lần lặp lại: (Bố trí thí nghiệm theo khối ngẫu nhiên đầy đủ (Randomized Complete Blocks) (RCB):

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

**Bố trí thí nghiệm trên hiện trường**

| Nhân tố B1 | Nhân tố B2 | Nhân tố B3 | Nhân tố Bb |
|------------|------------|------------|------------|
| A1         | Aa         | A2         | A2         |
| A2         | A4         | A3         | A3         |
| A3         | A3         | Aa         | A1         |
| A4         | A2         | A4         | A4         |
| Aa         | A1         | A1         | Aa         |

ả nhân tố B được chia thành b khối, ở mỗi khối bố trí a công thức của nhân tố A một cách ngẫu nhiên.

Ví dụ: Đánh giá kết quả khảo nghiệm 16 xuất xứ Pinus kesiya tại Lang Hanh-Lâm Đồng: 16 xuất xứ 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 được lặp lại ở 4 cấp đất (khối)

16 xuất xứ P.kesiya được đánh số như sau:

- 1: Xuất xứ Bengliet.
- 2: Faplac.
- 3: Xuân Thọ.
- 4: Thác Prenn.
- 5: Lang Hanh.
- 6: ả ong Kiating.
- 7: Doisupthep.
- 8: Doiinthranon.
- 9: Phu Kradung.
- 10: ả am nouv.
- 11: Cotomines.
- 12: Simao.
- 13: Watchan.
- 14: Zo khoa.
- 15: Aung ban.
- 16: Jingdury.

- Mỗi công thức ứng với 1 lần lặp được trồng 25 cây, với cự ly 3x2m, tổng diện tích bố trí thí nghiệm là 1,5ha.

- Các khí hậu, địa hình, chăm sóc...đều được đồng nhất, nhân tố thay đổi để khảo sát chỉ còn lại là các xuất xứ và cấp đất 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  $D_{1,3}$ , H,  $D_t$ , phẩm chất, tia cành, hình thân. Sử dụng 2 chỉ tiêu  $D_{1,3}$  và H để đánh giá sinh trưởng của các xuất xứ thử nghiệm.

Dùng phân tích phương sai để đánh giá sự sai khác về sinh trưởng, cụ thể cho từng chỉ tiêu sinh trưởng như sau:

**Trước hết đã kiểm tra 2 điều kiện để phân tích phương sai:**

- Điều kiện phân bố chuẩn: Các giá trị quan sát ở từng ô thí nghiệm qua kiểm tra bằng biểu đồ đều có dạng tiệm cận chuẩn nên chấp nhận giả thuyết phân bố chuẩn.
- Phương sai bằng nhau: Dùng tiêu chuẩn Cochran, kết quả tính được:

$$G_{max} = 0,11 < G_{max}(0,05; 16; 3) = 0,28$$

Do đó chấp nhận giả thuyết bằng nhau của các phương sai mẫu.

ả hư vậy 2 điều kiện trên là thỏa mãn để tiến hành phân tích phương sai.

**Dùng phân tích phương sai 2 nhân tố 1 lần lặp để kiểm tra:**

Với nhân tố thứ nhất là 16 xuất xứ, nhân tố thứ 2 là cấp đất với 4 cấp. Ứng với 1 tổ hợp Xuất xứ - Cấp đất chỉ có 1 ô thí nghiệm (lặp lại 1 lần).

**Bảng dữ liệu phân tích phương sai 2 nhân tố 1 lần lặp**

Giá trị  $D_{1,3}$  (cm) bình quân ứng với từng ô thí nghiệm theo 2 nhân tố 1 lần lặp

|    | A       | B         | C         | D         | E         |
|----|---------|-----------|-----------|-----------|-----------|
| 1  | Xuất xứ | Cấp đất 1 | Cấp đất 2 | Cấp đất 3 | Cấp đất 4 |
| 2  | 1       | 11.4      | 11.3      | 10.8      | 13.3      |
| 3  | 2       | 11.4      | 11.6      | 10.9      | 10.9      |
| 4  | 3       | 11.7      | 12.6      | 11.7      | 12.6      |
| 5  | 4       | 13.7      | 12.1      | 11.6      | 11.7      |
| 6  | 5       | 14.1      | 13.6      | 13.7      | 13.7      |
| 7  | 6       | 13.5      | 11.4      | 12.2      | 11.3      |
| 8  | 7       | 13.8      | 12.3      | 12.6      | 11.4      |
| 9  | 8       | 14.1      | 13.3      | 15.2      | 13.0      |
| 10 | 9       | 13.8      | 11.8      | 11.9      | 12.1      |
| 11 | 10      | 11.3      | 11.8      | 12.1      | 11.8      |
| 12 | 11      | 12.6      | 12.6      | 13.3      | 10.9      |
| 13 | 12      | 11.3      | 12.4      | 10.5      | 12.0      |
| 14 | 13      | 12.7      | 13.4      | 12.1      | 10.7      |
| 15 | 14      | 10.1      | 9.5       | 9.8       | 8.0       |
| 16 | 15      | 10.5      | 9.4       | 9.1       | 10.9      |
| 17 | 16      | 10.2      | 11.0      | 10.8      | 11.9      |

Phân tích phương sai 2 nhân tố 1 lần lặp:

- Tools/Data Analysis/Anova: Two Factor Without Replication - OK.
- Hộp thoại:

Input range: Địa chỉ khối dữ liệu (ả ên quét cả hàng, cột đầu làm nhãn). Vd:

A1:E17

Đánh dấu vào Labels.  
 Output range: Địa chỉ ô trên trái nơi xuất kết quả  
 OK

The screenshot shows an Excel spreadsheet with the following data:

|    | A6   | B                | C                | D                | E                | F | G | H | I | J | K |
|----|--|------------------|------------------|------------------|------------------|---|---|---|---|---|---|
| 4  |  |                  |                  |                  |                  |   |   |   |   |   |   |
| 5  | <b>Giá trị D bình quân ở các ô thử nghiệm các xuất xứ P. kesiya trên 4 cấp đất</b> |                  |                  |                  |                  |   |   |   |   |   |   |
| 6  | <b>Xuất xứ</b>   | <b>Cấp đất 1</b> | <b>Cấp đất 2</b> | <b>Cấp đất 3</b> | <b>Cấp đất 4</b> |   |   |   |   |   |   |
| 7  | 1  | 11.4             | 11.3             | 10.8             | 13.3             |   |   |   |   |   |   |
| 8  | 2  | 11.4             | 11.6             | 10.9             | 10.9             |   |   |   |   |   |   |
| 9  | 3  | 11.7             | 12.6             | 11.7             | 12.6             |   |   |   |   |   |   |
| 10 | 4  | 13.7             | 12.1             | 11.6             | 11.7             |   |   |   |   |   |   |
| 11 | 5  | 14.1             | 13.6             | 13.7             | 13.7             |   |   |   |   |   |   |
| 12 | 6  | 13.5             | 11.4             | 12.2             | 11.3             |   |   |   |   |   |   |
| 13 | 7  | 13.8             | 12.3             | 12.6             | 11.4             |   |   |   |   |   |   |
| 14 | 8  | 14.1             | 13.3             | 15.2             | 13.0             |   |   |   |   |   |   |
| 15 | 9  | 13.8             | 11.8             | 11.9             | 12.1             |   |   |   |   |   |   |

The ANOVA dialog box 'Anova: Two-Factor Without Replication' is open, showing the following settings:

- Input Range: \$A\$6:\$E\$22
- Labels:
- Alpha: 0.05
- Output options:  Output Range: [empty],  New Worksheet Ply: [empty],  New Workbook

### Kết quả phân tích phương sai 2 nhân tố 1 lần lặp lại

Anova: Two-Factor Without Replication

| <i>SUMMARY</i> | <i>Count</i> | <i>Sum</i> | <i>Average</i> | <i>Variance</i> |
|----------------|--------------|------------|----------------|-----------------|
| 1              | 4            | 46.9       | 11.7           | 1.253512        |
| 2              | 4            | 44.8       | 11.2           | 0.156318        |
| 3              | 4            | 48.6       | 12.2           | 0.268337        |
| 4              | 4            | 49.1       | 12.3           | 0.933224        |
| 5              | 4            | 55.1       | 13.8           | 0.049285        |
| 6              | 4            | 48.5       | 12.1           | 1.064903        |
| 7              | 4            | 50.0       | 12.5           | 0.975826        |
| 8              | 4            | 55.7       | 13.9           | 0.926688        |
| 9              | 4            | 49.7       | 12.4           | 0.817143        |
| 10             | 4            | 47.0       | 11.7           | 0.107475        |
| 11             | 4            | 49.3       | 12.3           | 1.054463        |
| 12             | 4            | 46.1       | 11.5           | 0.664541        |
| 13             | 4            | 48.9       | 12.2           | 1.255351        |
| 14             | 4            | 37.4       | 9.3            | 0.85117         |
| 15             | 4            | 39.9       | 10.0           | 0.763403        |
| 16             | 4            | 43.9       | 11.0           | 0.514494        |

|           |    |       |      |          |
|-----------|----|-------|------|----------|
| Cấp đất 1 | 16 | 196.1 | 12.3 | 2.077919 |
| Cấp đất 2 | 16 | 190.2 | 11.9 | 1.470334 |
| Cấp đất 3 | 16 | 188.3 | 11.8 | 2.263297 |
| Cấp đất 4 | 16 | 186.3 | 11.6 | 1.767392 |

#### AẢ OVA

| Source of Variation | SS       | df | MS       | F        | P-value  | F crit   |
|---------------------|----------|----|----------|----------|----------|----------|
| Rows                | 82.11826 | 15 | 5.474551 | 7.804468 | 3.58E-08 | 1.894875 |
| Columns             | 3.402532 | 3  | 1.134177 | 1.616873 | 0.198718 | 2.811547 |
| Error               | 31.56586 | 45 | 0.701464 |          |          |          |
| Total               | 117.0867 | 63 |          |          |          |          |

Từ bảng AẢ OVA nhận được:

- Đối với các xuất xứ khác nhau (Hàng - Rows):  $F = 7,80 > F_{(0,05)} = 1,89$ . Kết luận: Các xuất xứ khác nhau có sự sai khác về sinh trưởng đường kính.
- Đối với các cấp đất (Cột – Columns):  $F = 1,62 < F_{(0,05)} = 2,81$ . Kết luận: Các cấp đất khác nhau chưa có ảnh hưởng đến sinh trưởng.

ả hư vậy 16 xuất xứ khi trồng ở Lang Hanh đã có sinh trưởng khác nhau, do việc cấp đất không ảnh hưởng rệt, nên để đánh giá chính xác hơn chỉ cần phân tích phương sai 1 nhân tố (xuất xứ).

#### Phân tích phương sai 1 nhân tố

Anova: Single Factor

##### SUMMARY

| Groups | Count | Sum  | Average | Variance |
|--------|-------|------|---------|----------|
| 1      | 4     | 46.9 | 11.7    | 1.253512 |
| 2      | 4     | 44.8 | 11.2    | 0.156318 |
| 3      | 4     | 48.6 | 12.2    | 0.268337 |
| 4      | 4     | 49.1 | 12.3    | 0.933224 |
| 5      | 4     | 55.1 | 13.8    | 0.049285 |
| 6      | 4     | 48.5 | 12.1    | 1.064903 |
| 7      | 4     | 50.0 | 12.5    | 0.975826 |
| 8      | 4     | 55.7 | 13.9    | 0.926688 |
| 9      | 4     | 49.7 | 12.4    | 0.817143 |
| 10     | 4     | 47.0 | 11.7    | 0.107475 |
| 11     | 4     | 49.3 | 12.3    | 1.054463 |
| 12     | 4     | 46.1 | 11.5    | 0.664541 |

|    |   |      |      |          |
|----|---|------|------|----------|
| 13 | 4 | 48.9 | 12.2 | 1.255351 |
| 14 | 4 | 37.4 | 9.3  | 0.85117  |
| 15 | 4 | 39.9 | 10.0 | 0.763403 |
| 16 | 4 | 43.9 | 11.0 | 0.514494 |

Aả OVA

| Source of Variation | SS       | df | MS       | F        | P-value  | F crit   |
|---------------------|----------|----|----------|----------|----------|----------|
| Between Groups      | 82.11826 | 15 | 5.474551 | 7.514741 | 3.59E-08 | 1.880174 |
| Within Groups       | 34.9684  | 48 | 0.728508 |          |          |          |
| Total               | 117.0867 | 63 |          |          |          |          |

Kết quả từ bảng Aả OVA cho thấy  $F = 7,51 > F_{(0,05)} = 1,88$ . Kết luận: Sinh trưởng đường kính của 16 xuất xứ là khác nhau khi trồng ở Lang Hanh.

Sinh trưởng bình quân đường kính các xuất xứ theo thứ tự từ cao đến thấp ở bảng sau:

**Thứ tự sinh trưởng đường kính từ tốt đến xấu**

| Xuất xứ | D1,3 tb(cm) |
|---------|-------------|
| 8       | 13.9        |
| 5       | 13.8        |
| 7       | 12.5        |
| 9       | 12.4        |
| 11      | 12.3        |
| 4       | 12.3        |
| 13      | 12.2        |
| 3       | 12.2        |
| 6       | 12.1        |
| 10      | 11.7        |
| 1       | 11.7        |
| 12      | 11.5        |
| 2       | 11.2        |
| 16      | 11.0        |
| 15      | 10.0        |
| 14      | 9.3         |

Kết quả dùng tiêu chuẩn t để so sánh sinh trưởng đường kính lớn nhất của xuất xứ 8 với các xuất xứ có đường kính lần lượt nhỏ hơn:

- So sánh D bình quân giữa xuất xứ 8 với 5:  $t = 0,26 < t_{(0,05 ; k=48)} = 2,01$ .  
Kết luận: Sinh trưởng D giữa xuất xứ 8 và 5 chưa có sai khác rõ rệt.
- So sánh D bình quân giữa xuất xứ 8 với 7:  $t = 2,34 > t_{(0,05 ; k=48)} = 2,01$ .  
Kết luận: Sinh trưởng D giữa xuất xứ 8 và 7 có sai khác rõ rệt.

Âm vậy, xét theo chỉ tiêu đường kính, xuất xứ tối ưu trong 16 xuất xứ khảo nghiệm là 8 và 5, hai xuất xứ này có chỉ tiêu D lớn nhất, chưa có sai dị với nhau và có sai khác rõ rệt với các xuất xứ còn lại. Đó là 2 xuất xứ: Doiinthranon và Lang Hanh.

### 5.2.2. Phân tích phương sai 2 nhân tố m lần lặp

Trường hợp phân tích phương sai sau 2 nhân tố m lần lặp: Nhân tố A có a công thức và nhân tố B có b công thức; và này mỗi tổ hợp nhân tố A và B được lặp lại m lần một cách ngẫu nhiên. Lúc này ngoài việc đánh giá ảnh hưởng của từng nhân tố A, B ta còn phải tính ảnh hưởng qua lại của chúng đến kết quả thí nghiệm.

Ví dụ: Âm nghiên cứu ảnh hưởng của hai nhân tố thí nghiệm là mật độ và bón phân đến năng suất của bông.

- Nhân tố A: Mật độ chia làm 3 cấp.
- Nhân tố B: Phân bón được chia làm 4 mức
- Mỗi tổ hợp được thí nghiệm lặp lại ngẫu nhiên 4 lần.

**Bố trí thí nghiệm 2 nhân tố m lần lặp**

| Lặp 1 |    | Lặp 2 |    | Lặp 3 |    | Lặp 4 |    |
|-------|----|-------|----|-------|----|-------|----|
| B1    | B2 | B1    | B2 | B1    | B2 | B1    | B2 |
| A1    |    | A2    |    | A1    |    | A1    |    |
| B3    | B4 | B3    | B4 | B4    | B3 | B3    | B4 |
| B3    | B1 | B1    | B2 | B1    | B2 | B1    | B2 |
| A2    |    | A1    |    | A2    |    | A2    |    |
| B2    | B4 | B4    | B3 | B3    | B4 | B4    | B3 |
| B1    | B2 | B1    | B2 | B1    | B2 | B1    | B2 |
| A3    |    | A3    |    | A3    |    | A3    |    |
| B4    | B3 | B4    | B3 | B4    | B3 | B3    | B4 |



**Bảng số liệu sản lượng bông theo tổ hợp 2 nhân tố và lặp lại 4 lần ở một tổ hợp  
(Đ/v: Tạ/ha)**

|           | <b>A</b>   | <b>B</b> | <b>C</b> | <b>D</b> |
|-----------|------------|----------|----------|----------|
| <b>1</b>  | <b>B\A</b> | A1       | A2       | A3       |
| <b>2</b>  | B1         | 16       | 17       | 18       |
| <b>3</b>  |            | 14       | 15       | 18       |
| <b>4</b>  |            | 21       | 17       | 19       |
| <b>5</b>  |            | 16       | 19       | 17       |
| <b>6</b>  |            | B2       | 19       | 19       |
| <b>7</b>  | 20         |          | 18       | 23       |
| <b>8</b>  | 23         |          | 18       | 21       |
| <b>9</b>  | 19         |          | 20       | 21       |
| <b>10</b> | B3         |          | 19       | 21       |
| <b>11</b> |            | 21       | 21       | 18       |
| <b>12</b> |            | 22       | 22       | 21       |
| <b>13</b> |            | 20       | 23       | 21       |
| <b>14</b> |            | B4       | 20       | 20       |
| <b>15</b> | 24         |          | 20       | 22       |
| <b>16</b> | 21         |          | 22       | 21       |
| <b>17</b> | 17         |          | 19       | 23       |

***Phân tích phương sai 2 nhân tố m lần lặp:***

- Tools/Data Analysis/Anova: Two Factor With Replication- OK.
- Hộp thoại: Xác định:

Input range: ả hộp khối dữ liệu kể cả hàng cột tiêu đề. Vd: A1:D17.

Rows per sample: ả hộp số lần lặp. Vd: 4.

Output range: ả hộp địa chỉ ô trên trái nơi xuất kết quả.

OK.

Microsoft Excel - Trình bày giảng dạy Vi tính all

File Edit View Insert Format Tools Data Window Help Adobe PDF

100% WNTimes new roman 12 B I U

Reply with Changes... End Review...

F1 BVA

|    | A   | B  | C  | D  | E | F | G | H | I | J |
|----|-----|----|----|----|---|---|---|---|---|---|
| 1  | BVA | A1 | A2 | A3 |   |   |   |   |   |   |
| 2  | B1  | 16 | 17 | 18 |   |   |   |   |   |   |
| 3  |     | 14 | 15 | 18 |   |   |   |   |   |   |
| 4  |     | 21 | 17 | 19 |   |   |   |   |   |   |
| 5  |     | 16 | 19 | 17 |   |   |   |   |   |   |
| 6  | B2  | 19 | 19 | 20 |   |   |   |   |   |   |
| 7  |     | 20 | 18 | 23 |   |   |   |   |   |   |
| 8  |     | 23 | 18 | 21 |   |   |   |   |   |   |
| 9  |     | 19 | 20 | 21 |   |   |   |   |   |   |
| 10 | B3  | 19 | 21 | 22 |   |   |   |   |   |   |
| 11 |     | 21 | 21 | 18 |   |   |   |   |   |   |

**Anova: Two-Factor With Replication**

Input  
 Input Range: \$A\$1:\$D\$17  
 Rows per sample: 4  
 Alpha: 0,05

Output options  
 Output Range: \$F\$1  
 New Worksheet Ply:  
 New Workbook

OK Cancel Help

## Kết quả phân tích phương sai 2 nhân tố m lần lặp

Anova: Two-Factor With Replication

SUMMARY 1 2 3 Total

1

|          |          |          |          |          |
|----------|----------|----------|----------|----------|
| Count    | 4        | 4        | 4        | 12       |
| Sum      | 67       | 68       | 72       | 207      |
| Average  | 16,75    | 17       | 18       | 17,25    |
| Variance | 8,916667 | 2,666667 | 0,666667 | 3,659091 |

2

|          |          |          |          |          |
|----------|----------|----------|----------|----------|
| Count    | 4        | 4        | 4        | 12       |
| Sum      | 81       | 75       | 85       | 241      |
| Average  | 20,25    | 18,75    | 21,25    | 20,08333 |
| Variance | 3,583333 | 0,916667 | 1,583333 | 2,810606 |

3

|          |          |          |      |          |
|----------|----------|----------|------|----------|
| Count    | 4        | 4        | 4    | 12       |
| Sum      | 82       | 87       | 82   | 251      |
| Average  | 20,5     | 21,75    | 20,5 | 20,91667 |
| Variance | 1,666667 | 0,916667 | 3    | 1,901515 |

4

|       |    |    |    |     |
|-------|----|----|----|-----|
| Count | 4  | 4  | 4  | 12  |
| Sum   | 82 | 81 | 91 | 254 |

|          |          |          |          |          |
|----------|----------|----------|----------|----------|
| Average  | 20,5     | 20,25    | 22,75    | 21,16667 |
| Variance | 8,333333 | 1,583333 | 2,916667 | 4,878788 |

*Total*

|          |      |          |          |
|----------|------|----------|----------|
| Count    | 16   | 16       | 16       |
| Sum      | 312  | 311      | 330      |
| Average  | 19,5 | 19,4375  | 20,625   |
| Variance | 7,2  | 4,529167 | 4,783333 |

AẢ OVA

| <i>Source of Variation</i> | <i>SS</i> | <i>df</i> | <i>MS</i> | <i>F</i> | <i>P-value</i> | <i>F crit</i> |
|----------------------------|-----------|-----------|-----------|----------|----------------|---------------|
| Sample                     | 116,2292  | 3         | 38,74306  | 12,65079 | 8,45E-06       | 2,866265      |
| Columns                    | 14,29167  | 2         | 7,145833  | 2,333333 | 0,111468       | 3,259444      |
| Interaction                | 21,20833  | 6         | 3,534722  | 1,154195 | 0,352014       | 2,363748      |
| Within                     | 110,25    | 36        | 3,0625    |          |                |               |
| Total                      | 261,9792  | 47        |           |          |                |               |

- Bảng Summary: Cho kết quả tính toán từng tổ hợp nhân tố A/B và chung cho từng nhân tố B, nhân tố A, gồm các chỉ tiêu: Dung lượng (Count), Tổng (Sum), Trung bình (Average), Phương sai (Variance).
- Bảng AẢ OVA:

Cột đầu tiên là các nguồn biến động:

- Sample: Biến động do nhân tố B tạo nên (do được xếp theo hàng).
- Columns: Biến động do nhân tố A tạo nên (do được xếp theo cột).
- Interaction: Tác động qua lại.
- Within: Biến động ngẫu nhiên.
- Total: Biến động chính của n giá trị quan sát.

Từ kết quả này cho thấy:

$F_B = 12.65 > F_{0.05} = 2.87$ . Kl: Phân bón có tác động rõ rệt đến năng suất bông.

$F_A = 2.33 < F_{0.05} = 3.26$ . Kl: Mật độ ảnh hưởng không rõ đến năng suất bông.

$F_{AB} = 1.15 < F_{0.05} = 3.36$ . Kl: Đồng thời thay đổi mật độ và phân bón ảnh hưởng không rõ đến năng suất.

Lúc này chỉ còn việc lựa chọn công thức bón phân tối ưu. Qua số trung bình năng suất theo từng công thức bón phân cho thấy công thức 4 có năng suất cao nhất là 21.16 tạ/ha. Có thể dùng tiêu chuẩn t để kiểm tra lại xem công thức 4 có sai khác với công thức nào còn lại để lựa chọn công thức có hiệu quả nhất.

Lưu ý trong phân tích phương sai 2 nhân tố m lần lặp, nếu kiểm tra Interaction (tương tác giữa 2 nhân tố) có ảnh hưởng đến chỉ tiêu quan sát, thì lúc này cần chọn tổ hợp công thức có giá trị bình quân tốt nhất; đồng thời có thể so sánh nó với tổ hợp đứng thứ 2, 3 ... để tìm tổ hợp công thức tối ưu.

## 6. PHÂN TÍCH TƯƠNG QUAN - HỒI QUY

Trong thực tế người ta cần lập các mô hình tương quan hồi quy 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  $x$ ) và tất nhiên là phải có mối liên hệ giữa  $y$  và  $x$ . Từ đây có thể lập các biểu điều tra phục vụ cho việc giảm nhẹ các quan sát đo đếm một số nhân tố phức tạp
- Để dự báo một nhân tố trong tương lai (gọi là biến dự báo  $y$ ) với một số biến độc lập, đầu vào (gọi là biến độc lập  $x$ )
- Để nghiên cứu tác động, ảnh hưởng của một hoặc nhiều nhân tố đến một yếu tố cần quan tâm như sinh trưởng, sản lượng, chất lượng rừng, xói mòn đất, dòng chảy lưu vực. Trên cơ sở đó có giải pháp kỹ thuật thích hợp hoặc các biện pháp thích hợp

Sử dụng chương trình Excel hoặc Statgraphics Plus để thiết lập các mô hình tương quan/hồi quy tuyến tính từ một cho đến nhiều biến số độc lập. Trong chương trình này, các tham số được ước lượng bằng phương pháp bình phương tối thiểu. Riêng các dạng phi tuyến khi ứng dụng chương trình này cần đổi biến số để quy về dạng tuyến tính.

### 6.1. Hồi quy tuyến tính 1 lớp

Hồi quy tuyến tính một lớp có nghĩa là có một biến số độc lập  $x$  được nghiên cứu ảnh hưởng đến biến phụ thuộc  $y$ , dạng quan hệ được xác định là đường thẳng. Có nghĩa là khi  $x$  tăng hoặc giảm thì  $y$  cũng tăng hoặc giảm đều theo dạng đường thẳng. Dạng phương trình tổng quát:  $Y = A + B.X$ .

Vd: Lập mô hình tương quan giữa chiều cao dưới cành ( $H_{dc}$ ) với chiều cao cả cây ( $H$ ) rừng Tách dạng đường thẳng:  $H_{dc} = A + B.H$ . Vì  $H_{dc}$  là chỉ tiêu khó đo đếm hơn  $H$ , nên dùng quan hệ này để xác định  $H_{dc}$  thông qua  $H$ .

- **Nhập số liệu theo bảng:**

**Các cặp số liệu  $H_{dc}$  -  $H$**

|           | <b>A</b>                      | <b>B</b>                 |
|-----------|-------------------------------|--------------------------|
| <b>1</b>  | <b><math>H_{dc}(m)</math></b> | <b><math>H(m)</math></b> |
| <b>2</b>  | 22,0                          | 23,0                     |
| <b>3</b>  | 21,8                          | 23,0                     |
| <b>4</b>  | 21,5                          | 22,3                     |
| .....     | .....                         | .....                    |
| <b>40</b> | 9,7                           | 10,9                     |
| <b>41</b> | 9,8                           | 11,1                     |

- **Ước lượng tương quan hồi quy đường thẳng:**

- Tools/Data Analysis/Regression. OK.
- Hộp thoại:  
Input Y range: ả nhập địa chỉ cột biến Y (Có thể nhập cả nhãn). Vd: A1:A41.  
Input X range: ả nhập địa chỉ cột biến X (Có thể nhập cả nhãn). Vd: B1:B41.  
Label: Đánh dấu nếu đã nhập cả hàng đầu làm nhãn.  
Output range: ả nhập địa chỉ ô trên trái nơi xuất kết quả.  
OK.

Excel interface showing a regression analysis. The spreadsheet contains data for Hdc (m) and H (m) from rows 5 to 17. A 'Regression' dialog box is open, showing the following settings:

- Input Y Range: \$A\$5:\$A\$45
- Input X Range: \$B\$5:\$B\$45
- Labels
- Constant is Zero
- Confidence Level: 95 %
- Output options:
  - Output Range: \$E\$5
  - New Worksheet Ply:
  - New Workbook
- Residuals:
  - Residuals
  - Standardized Residuals
  - Residual Plots
  - Line Fit Plots
- Normal Probability:
  - Normal Probability Plots

## Kết quả ước lượng hồi quy tuyến tính 1 lớp

### SUMMARY OUTPUT

| Regression Statistics |             |
|-----------------------|-------------|
| Multiple R            | 0,998189546 |
| R Square              | 0,99638237  |
| Adjusted R Square     | 0,996287169 |
| Standard Error        | 0,318271114 |
| Observations          | 40          |

Ả OVA

|            | <i>df</i> | <i>SS</i>   | <i>MS</i> | <i>F</i> | <i>Significance F</i> |
|------------|-----------|-------------|-----------|----------|-----------------------|
| Regression | 1         | 1060,180842 | 1060,181  | 10466,12 | 5,24804E-48           |
| Residual   | 38        | 3,84926708  | 0,101297  |          |                       |
| Total      | 39        | 1064,030109 |           |          |                       |

|           | <i>Coefficients</i> | <i>Standard Error</i> | <i>t Stat</i> | <i>P-value</i> | <i>Lower 95%</i> | <i>Upper 95%</i> |
|-----------|---------------------|-----------------------|---------------|----------------|------------------|------------------|
| Intercept | -0,715306008        | 0,127254043           | -5,62109      | 1,88E-06       | -0,972918358     | -0,457693658     |
| Hgo(m)    | 0,994341123         | 0,009719471           | 102,304       | 5,25E-48       | 0,974665081      | 1,014017165      |

Phương trình tương quan:

$$H_{dc} = - 0.715 + 0.994.H$$

Với  $\hat{a} = 40$      $R = 0.998$      $Fr = 10466.12$     với  $\alpha < 0.0000$

Từ phương trình hồi quy, có thể xác định  $H_{dc}$  gián tiếp qua  $H$ .

## 6.2. Dạng phi tuyến đưa về tuyến tính 1 lớp

Trong thực tế biến  $y$  có thể không có dạng quan hệ đường thẳng với  $x$ , do đó cần sử dụng mô hình phi tuyến. Trường hợp các hàm phi tuyến, để ước lượng cần biến đổi thành dạng tuyến tính để ước lượng trong các phần mềm Excel, Statgraphics Plus.

Một số hàm phi tuyến phổ biến như:

$$y = a.x^b \text{ tuyến tính hóa: } \ln(y) = \ln(a) + b.\ln(x)$$

$$y = a.e^{bx} \text{ tuyến tính hóa: } \ln(y) = \ln(a) + b.x$$

### 6.2.1. Lập mô hình hàm mũ trong Excel:

Ví dụ: Lập mô hình tương quan H/D rừng trồng Tách dạng hàm mũ:

$$H = a.D^b$$

- Tuyến tính hóa: Logarit neper 2 vế:**

$$\ln(H) = \ln(a) + b.\ln(D)$$

Đặt  $Y = \ln(H)$      $X = \ln(D)$      $A = \ln(a)$      $B = b$ .

Vậy  $Y = A + B.X$

- Nhập số liệu và đổi biến số:**

- o Cột A: Số liệu D.
- o Cột B: Số liệu H.
- o Cột C:  $\ln(D)$ . Tại ô C2: =Ln(A2), copy cho cả cột.
- o Cột D:  $\ln(H)$ . Tại ô D2: =Ln(B2), copy cho cả cột.

**Số liệu H/D và đổi biến số**

|           | A            | B           | C            | D            |
|-----------|--------------|-------------|--------------|--------------|
| <b>1</b>  | <b>D(cm)</b> | <b>H(m)</b> | <b>Ln(D)</b> | <b>Ln(H)</b> |
| <b>2</b>  | 31,3         | 22,0        | 3,443863     | 3,091042     |
| <b>3</b>  | 32,0         | 21,8        | 3,466237     | 3,08191      |
| ...       | ...          | ....        | .....        | .....        |
| ....      | ....         | ....        | .....        | .....        |
| <b>40</b> | 12,6         | 9,7         | 2,536373     | 2,270804     |
| <b>41</b> | 13,9         | 9,8         | 2,629481     | 2,277972     |

**Ước lượng tương quan hồi quy đường thẳng trong Excel:**

- o Tools/Data Analysis/Regression. OK.
- o Hộp thoại:  
Input Y range:  $\hat{a}$  hộp địa chỉ cột biến Y (Có thể nhập cả nhãn). Vd: D1:D41.  
Input X range:  $\hat{a}$  hộp địa chỉ cột biến X (Có thể nhập cả nhãn). Vd: C1:C41.  
Label: Đánh dấu nếu đã nhập cả hàng đầu làm nhãn.  
Output range:  $\hat{a}$  hộp địa chỉ ô trên trái nơi xuất kết quả.  
Kích OK.

Microsoft Excel interface showing a regression analysis setup. The worksheet contains data for diameter (D) and height (H) of trees, with natural logarithms of H (ln(H)) calculated. A Regression dialog box is open, showing the input ranges for Y (\$D\$7:\$D\$47) and X (\$C\$7:\$C\$47), and the output range (\$F\$7).

|    | A   | B            | C            | D            | E | F | G | H | I | J | K |
|----|---|--------------|--------------|--------------|---|---|---|---|---|---|---|
| 1  | <b>Bài tập: Thiết lập mô hình quan hệ phi tuyến tính</b>                    |              |              |              |   |   |   |   |   |   |   |
| 2  | Từ số liệu đo D/H một số cây, thiết lập mô hình H/D dạng hàm mũ phức vụ ước |              |              |              |   |   |   |   |   |   |   |
| 3  | lượng nhanh H   |              |              |              |   |   |   |   |   |   |   |
| 4  | $H = a D^b$   |              |              |              |   |   |   |   |   |   |   |
| 5  |   |              |              |              |   |   |   |   |   |   |   |
| 6  | <b>Số liệu đo D/H để lập mô hình</b>  |              |              |              |   |   |   |   |   |   |   |
| 7  | <b>D (cm)</b>   | <b>H (m)</b> | <b>ln(D)</b> | <b>ln(H)</b> |   |   |   |   |   |   |   |
| 8  | 31.3  | 22.0         | 3.44362      | 3.09104      |   |   |   |   |   |   |   |
| 9  | 32.0  | 21.8         | 3.46574      | 3.08191      |   |   |   |   |   |   |   |
| 10 | 30.6  | 21.5         | 3.421        | 3.06805      |   |   |   |   |   |   |   |
| 11 | 27.9  | 21.6         | 3.32863      | 3.07269      |   |   |   |   |   |   |   |
| 12 | 10.2  | 6.4          | 2.32239      | 1.8563       |   |   |   |   |   |   |   |
| 13 | 10.2  | 6.5          | 2.32239      | 1.8718       |   |   |   |   |   |   |   |
| 14 | 9.6   | 5.9          | 2.26176      | 1.77495      |   |   |   |   |   |   |   |
| 15 | 9.5   | 5.7          | 2.25129      | 1.74047      |   |   |   |   |   |   |   |
| 16 | 9.5   | 6.1          | 2.25129      | 1.80829      |   |   |   |   |   |   |   |
| 17 | 10.2  | 6.0          | 2.32239      | 1.79176      |   |   |   |   |   |   |   |
| 18 | 16.4  | 7.3          | 2.79728      | 1.98787      |   |   |   |   |   |   |   |
| 19 | 15.9  | 7.4          | 2.76632      | 2.00148      |   |   |   |   |   |   |   |
| 20 | 10.0  | 6.5          | 2.30259      | 1.8718       |   |   |   |   |   |   |   |

### Kết quả ước lượng hồi quy tuyến tính

#### SUMMARY OUTPUT

| <i>Regression Statistics</i> |          |
|------------------------------|----------|
| Multiple R                   | 0.940546 |
| R Square                     | 0.884627 |
| Adjusted R Square            |          |
| Standard Error               | 0.881591 |
| Observations                 | 40       |

| <i>ANOVA</i> |           |           |           |          |                       |
|--------------|-----------|-----------|-----------|----------|-----------------------|
|              | <i>df</i> | <i>SS</i> | <i>MS</i> | <i>F</i> | <i>Significance F</i> |
| Regression   | 1         | 8.141566  | 8.14156   | 291.366  | 2.062E-19             |
| Residual     | 38        | 1.061823  | 0.02794   |          |                       |
| Total        | 39        | 9.203388  |           |          |                       |

|           | <i>Coefficients</i> | <i>Standard Error</i> | <i>t Stat</i> | <i>P-value</i> | <i>Lower 95%</i> | <i>Upper 95%</i> |
|-----------|---------------------|-----------------------|---------------|----------------|------------------|------------------|
| Intercept | -0.80087            | 0.18412               | -4.34972      | 9.88E-05       | -1.173600        | -0.428137967     |
| ln(D)     | 1.157937            | 0.067837              | 17.0694       | 2.06E-19       | 1.020609         | 1.295265915      |



Phương trình tương quan:

$$\ln(H) = -0.800 + 1.157 \cdot \ln(D)$$

Với  $\hat{a} = 40$     $R = 0.940$     $Fr = 291.36$    với  $\alpha < 0.0000$

Đưa về dạng nguyên thủy: Tính  $a = \exp(A) = \exp(-0.800) = 0.449$

Vậy:       $H = 0.449 \cdot D^{1.157}$

Từ mô hình này có thể ước lượng H thông qua D mà không phải đo đếm

### 6.2.2. *Lập mô hình hàm mũ và Schumacher trong Statgraphics:*

Trong Statgraphics Plus, việc tính toán mô hình đơn giản hơn vì không cần tạo thêm các cột đổi biến số, biến số được đổi trực tiếp trong hộp thoại khi thiết lập mô hình.

#### i) **Lập mô hình hàm mũ trong Statgraphics**

Trong Statgraphics, việc ước lượng mô hình phi tuyến tính đơn giản hơn vì không cần tạo thêm các cột đổi biến số, biến số được đổi trực tiếp trong hộp thoại khi thiết lập mô hình.

Đầu tiên nhập dữ liệu trong Excel với hai cột x và y, ví dụ là D và H như sau

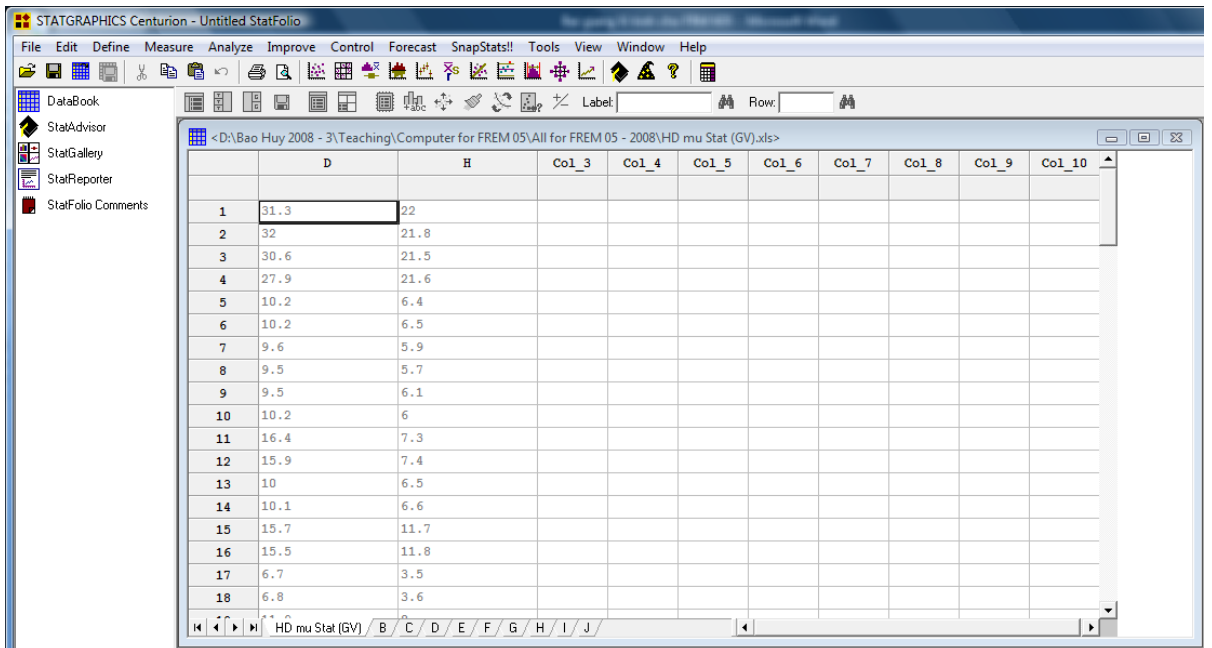
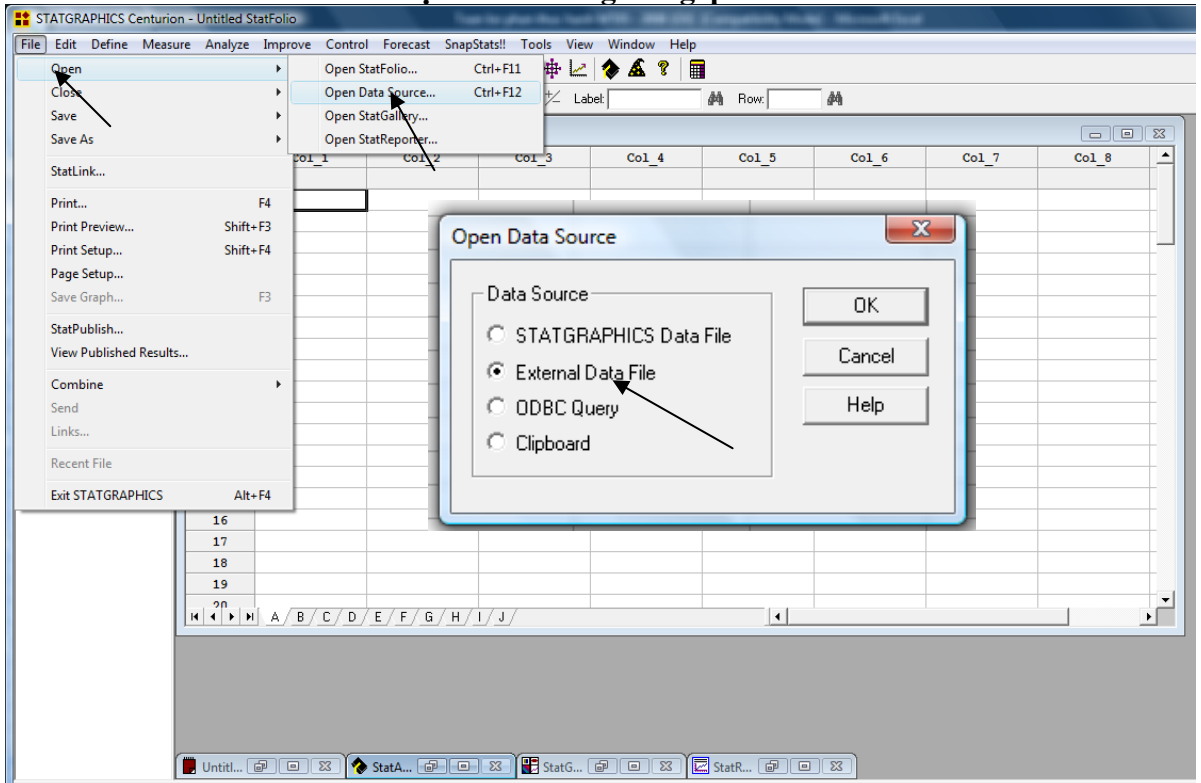
|    | D (cm) | H (m) |
|----|--------|-------|
| 1  |        |       |
| 2  | 31.3   | 22.0  |
| 3  | 32.0   | 21.8  |
| 4  | 30.6   | 21.5  |
| 5  | 27.9   | 21.6  |
| 6  | 10.2   | 6.4   |
| 7  | 10.2   | 6.5   |
| 8  | 9.6    | 5.9   |
| 9  | 9.5    | 5.7   |
| 10 | 9.5    | 6.1   |
| 11 | 10.2   | 6.0   |
| 12 | 16.4   | 7.3   |
| 13 | 15.9   | 7.4   |
| 14 | 10.0   | 6.5   |
| 15 | 10.1   | 6.6   |
| 16 | 15.7   | 11.7  |
| 17 | 15.5   | 11.8  |
| 18 | 6.7    | 3.5   |
| 19 | 6.8    | 3.6   |
| 20 | 11.9   | 8.0   |
| 21 | 11.9   | 8.1   |
| 22 | 15.5   | 12.1  |
| 23 | 15.4   | 12.1  |
| 24 | 10.1   | 6.1   |
| 25 | 10.1   | 6.4   |
| 26 | 17.7   | 12.2  |
| 27 | 18.1   | 12.4  |

File dữ liệu Excel cần được lưu với version của Microsoft Excel 97-2003 về trước, vì Statgraphics chưa nhận được kiểu file MS. Office 2007

Sau đó mở file dữ liệu này trong Statgraphics Centurion: File/Open/Open Data Source/External Data file - OK

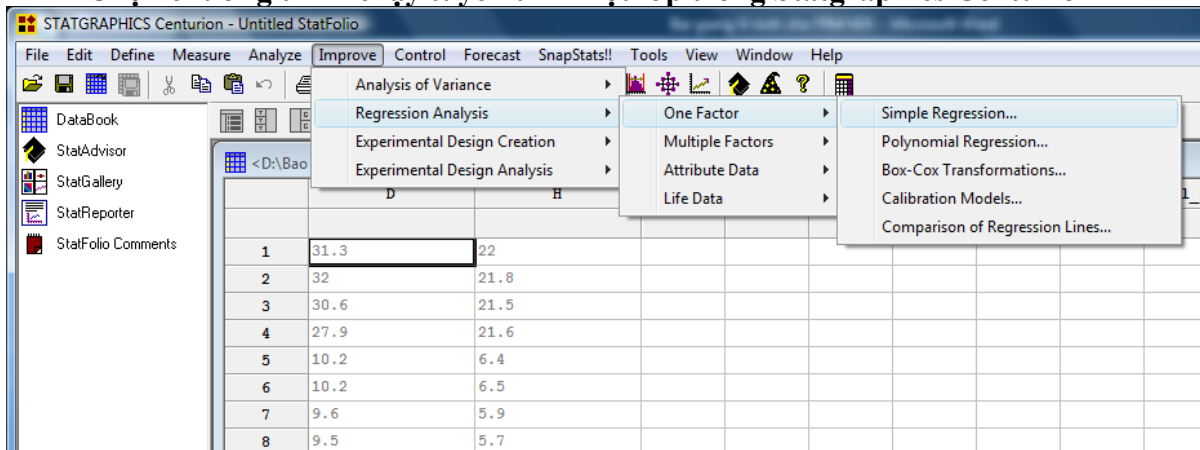


## Mở file dữ liệu Excel trong Statgraphics Centurion

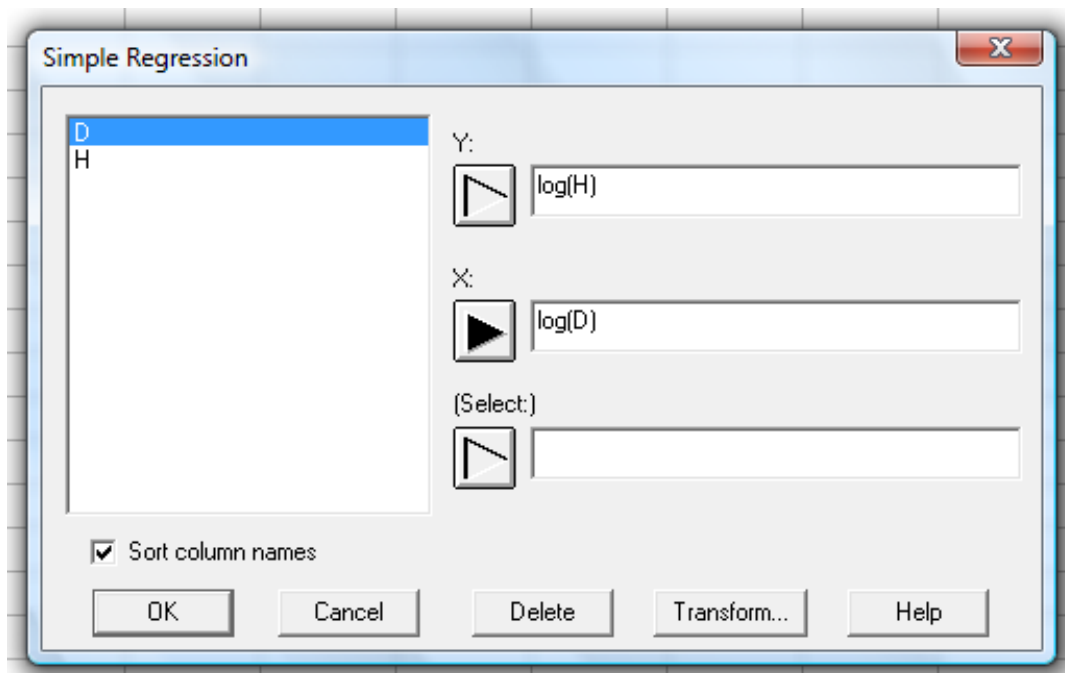


Chạy phần xử lý hàm tương quan một lớp: Improve/Regression Analysis/One Factor/Simple Regression

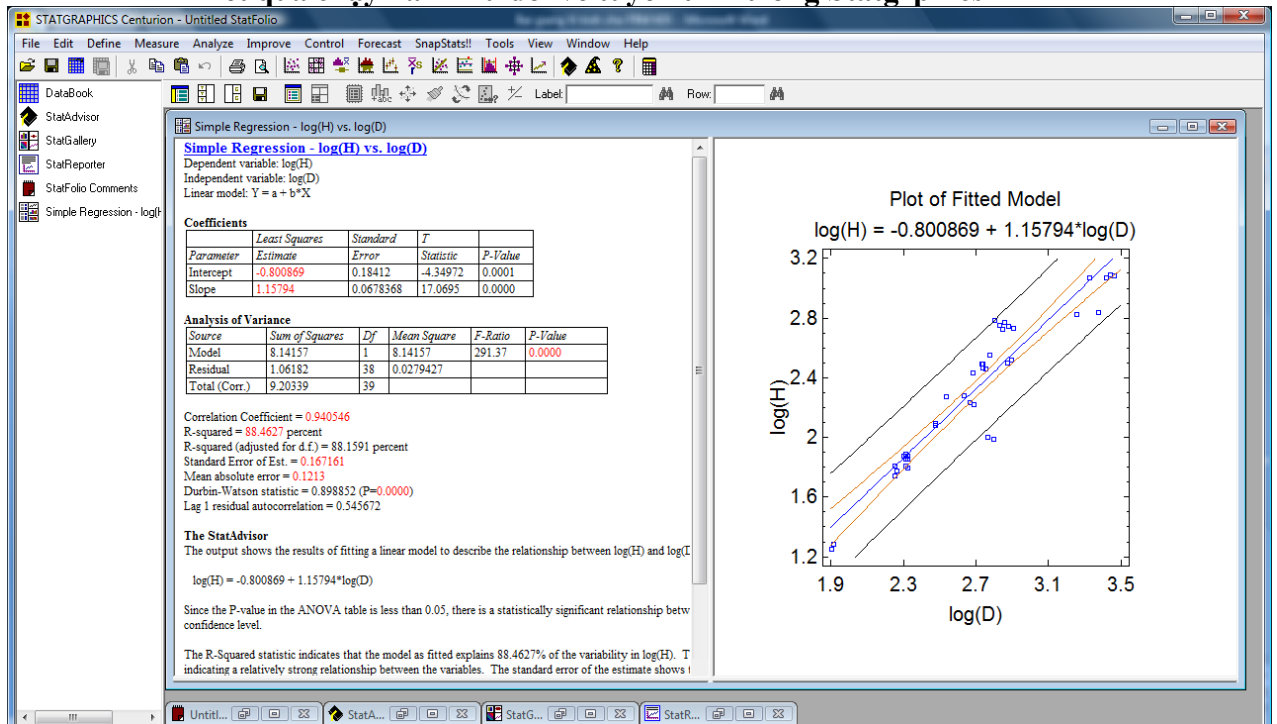
## Chọn chương trình chạy tuyến tính một lớp trong Statgraphics Centurion



Trong hộp thoại chọn biến y và x và đổi biến số ngay trong hộp thoại: log(H) và log(D).  
Kích OK để có kết quả. (Lưu ý ký hiệu log trong Statgraphics là logarit neper)



## Kết quả chạy hàm mũ đổi về tuyến tính trong Statgraphics



### Simple Regression - log(H) vs. log(D)

Dependent variable: log(H)

Independent variable: log(D)

Linear model:  $Y = a + b \cdot X$

#### Coefficients

|                  | <i>Least Squares</i> | <i>Standard</i> | <i>T</i>         |                |
|------------------|----------------------|-----------------|------------------|----------------|
| <i>Parameter</i> | <i>Estimate</i>      | <i>Error</i>    | <i>Statistic</i> | <i>P-Value</i> |
| Intercept        | -0.800869            | 0.18412         | -4.34972         | 0.0001         |
| Slope            | 1.15794              | 0.0678368       | 17.0695          | 0.0000         |

#### Analysis of Variance

| <i>Source</i> | <i>Sum of Squares</i> | <i>Df</i> | <i>Mean Square</i> | <i>F-Ratio</i> | <i>P-Value</i> |
|---------------|-----------------------|-----------|--------------------|----------------|----------------|
| Model         | 8.14157               | 1         | 8.14157            | 291.37         | 0.0000         |
| Residual      | 1.06182               | 38        | 0.0279427          |                |                |
| Total (Corr.) | 9.20339               | 39        |                    |                |                |

Correlation Coefficient = 0.940546

R-squared = 88.4627 percent

R-squared (adjusted for d.f.) = 88.1591 percent

Standard Error of Est. = 0.167161

Mean absolute error = 0.1213

Durbin-Watson statistic = 0.898852 (P=0.0000)

Lag 1 residual autocorrelation = 0.545672

#### The StatAdvisor

The output shows the results of fitting a linear model to describe the relationship between log(H) and log(D). The equation of the fitted model is

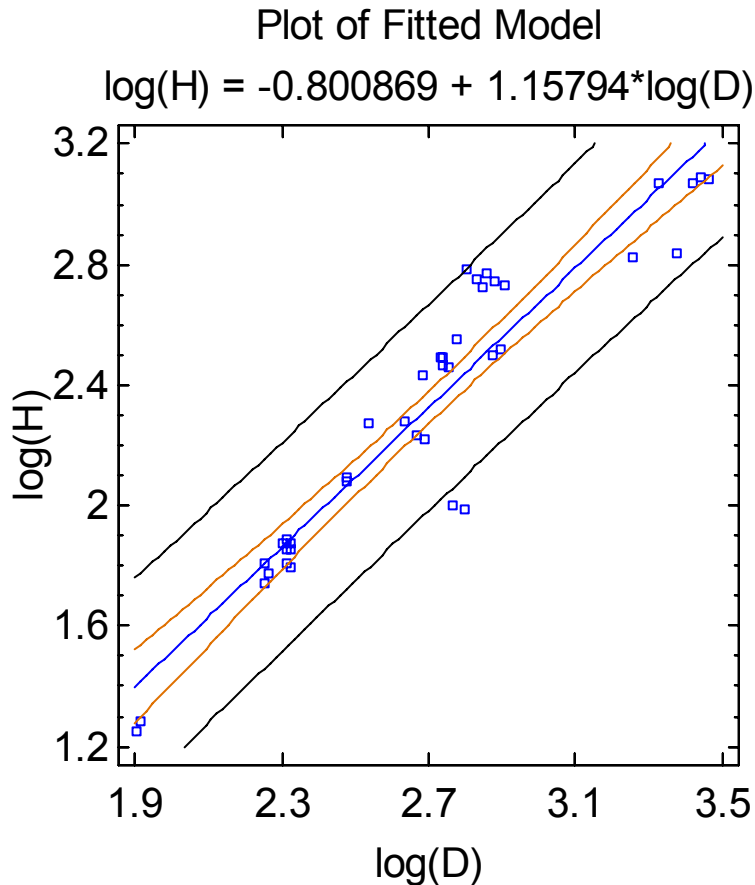
$$\log(H) = -0.800869 + 1.15794 \cdot \log(D)$$

Since the P-value in the ANOVA table is less than 0.05, there is a statistically significant relationship between log(H) and log(D) at the 95.0% confidence level.

The R-Squared statistic indicates that the model as fitted explains 88.4627% of the variability in log(H). The correlation coefficient equals 0.940546, indicating a relatively strong relationship between the variables. The standard error of the

estimate shows the standard deviation of the residuals to be 0.167161. 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 0.1213 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.



Kết quả cho ra hàm trực tiếp viết dưới dạng tuyến tính đã đổi biến số  
 Các kết quả kiểm tra hệ số tương quan R và các biến số được hiểu giống như trong Excel

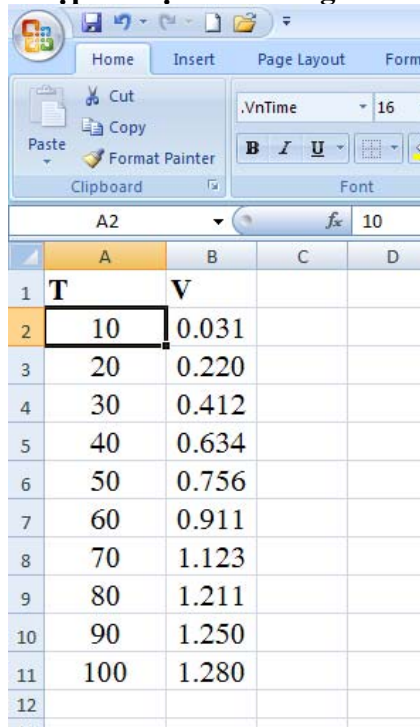
**ii) Lập mô hình hàm Schumacher trong Statgraphics**

Trong lâm nghiệp hàm Schumacher được sử dụng phổ biến để mô phỏng quá trình sinh trưởng cây rừng và lâm phần. Dạng hàm Schumacher:

$y = a.e^{-b \cdot x^{-m}}$ , trong đó a, b và m là tham số; y là giá trị sinh trưởng D, H, V, G, M và x là tuổi (T). Khi ước lượng hàm này, tham số m thường chạy từ 0.1; 0.2; ..... 1.9; 2.0. Từ đó chọn hàm tối ưu với m cho hệ số tương quan R cao nhất.  
 Để ước lượng hàm này, tuyến tính hóa:  $\ln(y) = \ln(a) - b \cdot x^{-m}$ .

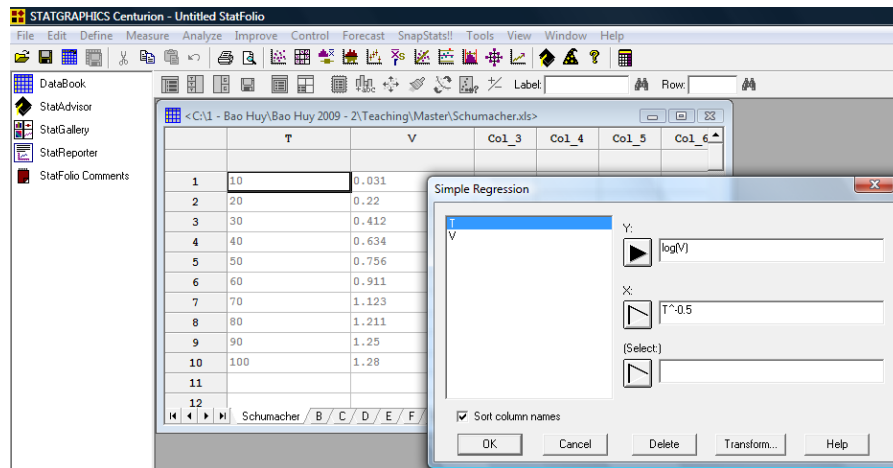
Ví dụ sử dụng Statgraphics để ước lượng sinh trưởng V theo tuổi (T) của loài bằng lăng theo hàm Schumacher.

## Nhập dữ liệu V/T trong Excel



|    | A   | B     | C | D |
|----|-----|-------|---|---|
| 1  | T   | V     |   |   |
| 2  | 10  | 0.031 |   |   |
| 3  | 20  | 0.220 |   |   |
| 4  | 30  | 0.412 |   |   |
| 5  | 40  | 0.634 |   |   |
| 6  | 50  | 0.756 |   |   |
| 7  | 60  | 0.911 |   |   |
| 8  | 70  | 1.123 |   |   |
| 9  | 80  | 1.211 |   |   |
| 10 | 90  | 1.250 |   |   |
| 11 | 100 | 1.280 |   |   |
| 12 |     |       |   |   |

## Đổi biến số với m chạy trong Statgraphics Improve/Regression Analysis/One Factor/Simple Regression



STATGRAPHICS Centurion - Untitled StatFolio

File Edit Define Measure Analyze Improve Control Forecast SnapStats! Tools View Window Help

<C:\1 - Bao Huy\Bao Huy 2009 - 2\Teaching\Master\Schumacher.xls>

|    | T   | V     | Col_3 | Col_4 | Col_5 | Col_6_6 |
|----|-----|-------|-------|-------|-------|---------|
| 1  | 10  | 0.031 |       |       |       |         |
| 2  | 20  | 0.22  |       |       |       |         |
| 3  | 30  | 0.412 |       |       |       |         |
| 4  | 40  | 0.634 |       |       |       |         |
| 5  | 50  | 0.756 |       |       |       |         |
| 6  | 60  | 0.911 |       |       |       |         |
| 7  | 70  | 1.123 |       |       |       |         |
| 8  | 80  | 1.211 |       |       |       |         |
| 9  | 90  | 1.25  |       |       |       |         |
| 10 | 100 | 1.28  |       |       |       |         |
| 11 |     |       |       |       |       |         |
| 12 |     |       |       |       |       |         |

Simple Regression

Y: log(V)

X:  $T^{0.5}$

(Select)

Sort column names

OK Cancel Delete Transform... Help

## Kết quả ước lượng hàm Schumacher

### Simple Regression - $\log(V)$ vs. $T^{-0.5}$

Dependent variable:  $\log(V)$

Independent variable:  $T^{-0.5}$

Linear model:  $Y = a + b \cdot X$

#### Coefficients

|                  | <i>Least Squares</i> | <i>Standard</i> | <i>T</i>         |                |
|------------------|----------------------|-----------------|------------------|----------------|
| <i>Parameter</i> | <i>Estimate</i>      | <i>Error</i>    | <i>Statistic</i> | <i>P-Value</i> |
| Intercept        | 2.12522              | 0.111291        | 19.0961          | 0.0000         |
| Slope            | -17.117              | 0.650282        | -26.3224         | 0.0000         |

#### Analysis of Variance

| <i>Source</i> | <i>Sum of Squares</i> | <i>Df</i> | <i>Mean Square</i> | <i>F-Ratio</i> | <i>P-Value</i> |
|---------------|-----------------------|-----------|--------------------|----------------|----------------|
| Model         | 11.9519               | 1         | 11.9519            | 692.87         | 0.0000         |
| Residual      | 0.137999              | 8         | 0.0172499          |                |                |
| Total (Corr.) | 12.0899               | 9         |                    |                |                |

Correlation Coefficient = -0.994276

R-squared = 98.8586 percent

R-squared (adjusted for d.f.) = 98.7159 percent

Standard Error of Est. = 0.131339

Mean absolute error = 0.0958475

Durbin-Watson statistic = 1.26469 (P=0.0470)

Lag 1 residual autocorrelation = 0.141506

#### The StatAdvisor

The output shows the results of fitting a linear model to describe the relationship between  $\log(V)$  and  $T^{-0.5}$ . The equation of the fitted model is

$$\log(V) = 2.12522 - 17.117 \cdot T^{-0.5}$$

Since the P-value in the ANOVA table is less than 0.05, there is a statistically significant relationship between  $\log(V)$  and  $T^{-0.5}$  at the 95.0% confidence level.

The R-Squared statistic indicates that the model as fitted explains 98.8586% of the variability in  $\log(V)$ . The correlation coefficient equals -0.994276, indicating a relatively strong relationship between the variables. The standard error of the estimate shows the standard deviation of the residuals to be 0.131339. 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 0.0958475 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.

$$\log(V) = 2.12522 - 17.117 \cdot T^{-0.5}$$

Với  $R = -0.994$  và  $P < 0.000$

Từ đây suy ra hàm nguyên thủy:  $a = \exp(2.12522) = 8.374$ ;  $b = 17.117$

$$V = 8.374 \cdot \exp(-17.117 \cdot T^{-0.5})$$

Kết quả trên là với  $m = 0.5$ ; lần lượt thay  $m$  khác nhau để tìm hàm tối ưu với  $R$  max.

### 6.3. Hồi quy tuyến tính nhiều lớp

Trong thực tế biến phụ thuộc  $Y$  bị chi phối bởi nhiều biến số độc lập  $X_i$ . Ví dụ như trữ lượng rừng được đóng góp bởi nhiều nhân tố như mật độ, tiết diện ngang, chiều cao, cấp đất; hoặc biến đổi dòng chảy, mức độ xung yếu của lưu vực bị chi phối bởi nhiều nhân tố như lượng mưa, độ dốc, địa hình, loại đất, che phủ thảm thực vật; hoặc tái sinh tự nhiên phụ thuộc vào nhiều nhân tố như ánh sáng, hạt giống, lửa rừng, đất, thực bì, ..... Tuy nhiên biến nào là chủ đạo thì chúng ta chưa biết, do vậy với phương pháp mô hình hóa với nhiều thử nghiệm khác

nhau giúp chúng ta dự báo được nhân tố ảnh hưởng quan trọng, trên cơ sở đó sẽ điều tiết các nhân tố này đáp ứng được mục tiêu mong đợi

Trong trường hợp này để ước lượng biến phụ thuộc Y người ta cần lập mô hình hồi quy nhiều biến số để có thể phản ánh chính xác giá trị ước lượng, dự báo Y.

Dạng phương trình tổng quát:

$$Y = a_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n$$

Ví dụ: Thiết lập mô hình dự đoán trữ lượng rừng (M) Tách theo 2 biến số mật độ (â) và chiều cao bình quân (H) theo dạng hàm mũ:

$$M = a + b_1 \hat{a} + b_2 H$$

Đây là dạng tuyến tính 2 lớp  $Y = a + b_1X_1 + b_2X_2$

Dùng phương pháp bình phương tối thiểu ước lượng phương trình

- **Nhập số liệu**

**Bảng số liệu M/N/H**

|           | A              | B           | C               |
|-----------|----------------|-------------|-----------------|
| <b>1</b>  | <b>N(c/ha)</b> | <b>H(m)</b> | <b>M(m3/ha)</b> |
| <b>2</b>  | 180            | 23,0        | 163,452         |
| <b>3</b>  | 170            | 23,0        | 160,154         |
| <b>4</b>  | 220            | 22,3        | 184,167         |
| ...       | ....           | ....        | .....           |
| ...       | ....           | ....        | .....           |
| <b>40</b> | 570            | 10,9        | 43,846          |
| <b>41</b> | 570            | 11,1        | 53,212          |

- **Ước lượng tương quan tuyến tính nhiều lớp:**

- Tools/Data Analysis/Regression.OK.

- Hộp thoại:

Input Y range: ả hập địa chỉ cột biến Y (Có thể nhập cả nhãn). Vd: C1:C41.

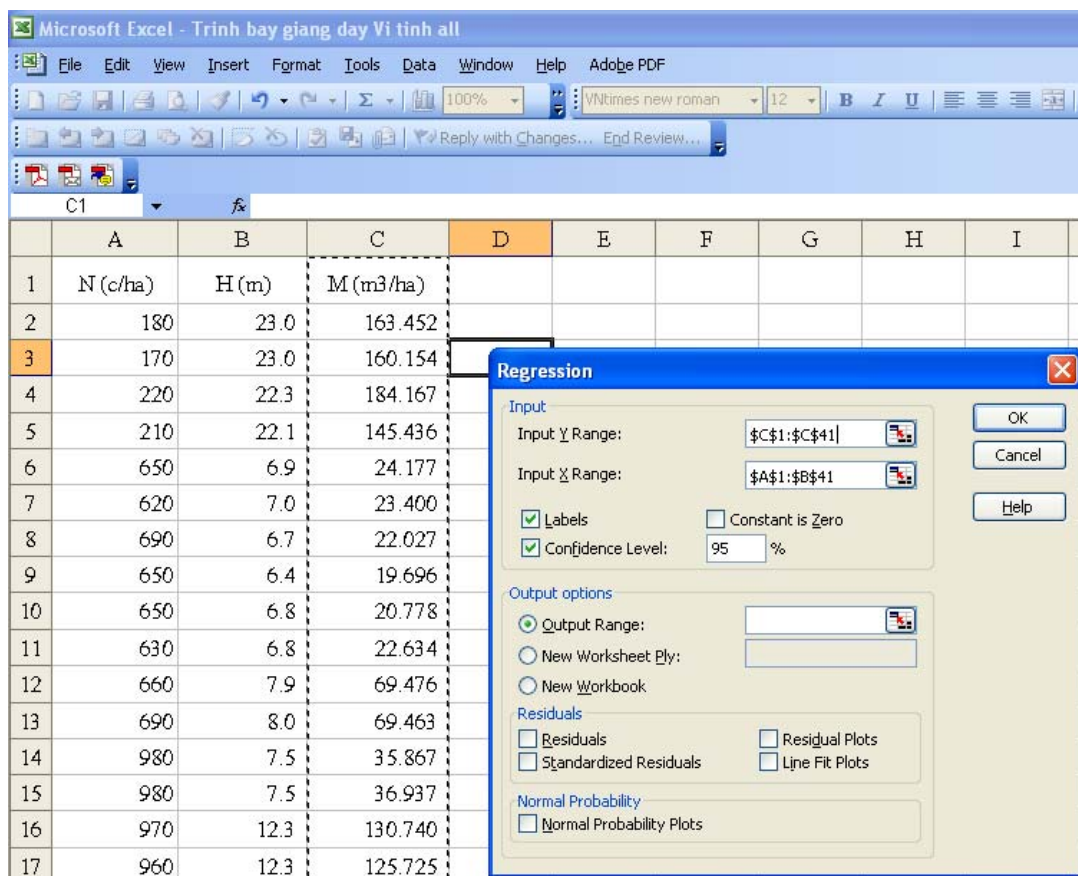
Input X range: ả hập địa chỉ khối các biến X (Có thể nhập cả nhãn). Vd:

A1:B41.

Label: Đánh dấu nếu đã nhập cả hàng đầu làm nhãn.

Output range: ả hập địa chỉ ô trên trái nơi xuất kết quả.

OK.



## Kết quả ước lượng mô hình hồi quy tuyến tính 2 lớp

### SUMMARY OUTPUT

#### Regression Statistics

|                   |           |
|-------------------|-----------|
| Multiple R        | 0.9256776 |
| R Square          | 0.856879  |
| Adjusted R Square | 0.8491427 |
| Standard Error    | 28.140919 |
| Observations      | 40        |

#### ẢŦ OVA

|            | <i>df</i> | <i>SS</i> | <i>MS</i> | <i>F</i> | <i>Significance F</i> |
|------------|-----------|-----------|-----------|----------|-----------------------|
| Regression | 2         | 175426.2  | 87713.1   | 110.7613 | 2.40166E-16           |
| Residual   | 37        | 29300.72  | 791.9113  |          |                       |
| Total      | 39        | 204726.9  |           |          |                       |

|           | <i>Coefficients</i> | <i>Standard Error</i> | <i>t Stat</i> | <i>P-value</i> | <i>Lower 95%</i> | <i>Upper 95%</i> |
|-----------|---------------------|-----------------------|---------------|----------------|------------------|------------------|
| Intercept | -154.77144          | 22.13662              | -6.99165      | 2.91E-08       | -199.6244851     | -109.918392      |
| Ả (c/ha)  | 0.1095484           | 0.016994              | 6.446152      | 1.57E-07       | 0.075114494      | 0.143982284      |
| H (m)     | 14.52156            | 0.97677               | 14.86692      | 3.49E-17       | 12.54243676      | 16.50068344      |



Phương trình tương quan hồi quy:

$$M = - 154.771 + 0.109 N + 14.521 H$$

Với  $\hat{a} = 40$      $R = 0.926$      $Fr = 110.76$     với  $\alpha < 0.00$

$tb_1 = 6.44$      $tb_2 = 14.86$     với  $\alpha < 0.00$

Từ mô hình này có thể ước lượng, dự báo trữ lượng rừng thông qua 2 biến số mật độ (á /ha) và chiều cao bình quân (H).

### 6.4. Hồi quy phi tuyến tính nhiều lớp, tổ hợp biến

Trong trường hợp nhiều biến số xi ảnh hưởng đến y không theo dạng tuyến tính mà có dạng quan hệ phi tuyến, trường hợp này cần đổi biến số để trở về dạng tuyến tính, hoặc lập mô hình tổ hợp biến.

Một số dạng phi tuyến nhiều lớp phổ biến và cách quy về tuyến tính hoặc tổ hợp biến:

$y = a \cdot x_1^{b_1} \cdot x_2^{b_2} \dots x_n^{b_n}$  tuyến tính hóa:  $\ln(y) = \ln(a) + b_1 \ln(x_1) + b_2 \ln(x_2) + \dots + b_n \ln(x_n)$

$y = a \cdot e^{b_1 x_1 + b_2 x_2 + \dots + b_n x_n}$  tuyến tính hóa:  $\ln(y) = \ln(a) + b_1 x_1 + b_2 x_2 + \dots + b_n x_n$

.....

#### 6.4.1. Lập mô hình phi tuyến nhiều lớp chuyển về tuyến tính nhiều lớp trong Excel

Hàm phi tuyến nhiều lớp trong Excel được ước lượng bằng cách tuyến tính hóa, có nghĩa là cần tạo thêm các trường/cột biến số mới.

Ví dụ lập quan hệ giữa  $M = f(\hat{a}, H)$  dạng hàm mũ:  $M = a \cdot \hat{a}^{b_1} H^{b_2}$ .

Tuyến tính hóa bằng cách lấy log 2 vế:  $\ln(M) = \ln(a) + b_1 \ln(\hat{a}) + b_2 \ln(H)$

á hư vậy phải tạo thêm 3 cột log để đổi biến số

**Bảng dữ liệu đổi biến số để lập mô hình phi tuyến nhiều lớp**

| N (c/ha) | H (m) | M (m3/ha) | Ln(M)    | Ln(N)    | Ln(H)    |
|----------|-------|-----------|----------|----------|----------|
| 180      | 23.0  | 163.452   | 5.096519 | 5.192957 | 3.135494 |
| 170      | 23.0  | 160.154   | 5.076136 | 5.135798 | 3.135494 |
| 220      | 22.3  | 184.167   | 5.215843 | 5.393628 | 3.104587 |
| 210      | 22.1  | 145.436   | 4.979736 | 5.347108 | 3.095578 |
| 650      | 6.9   | 24.177    | 3.185402 | 6.476972 | 1.931521 |
| 620      | 7.0   | 23.400    | 3.152736 | 6.429719 | 1.945910 |
| 690      | 6.7   | 22.027    | 3.092269 | 6.536692 | 1.902108 |
| 650      | 6.4   | 19.696    | 2.980416 | 6.476972 | 1.856298 |
| 650      | 6.8   | 20.778    | 3.033895 | 6.476972 | 1.916923 |
| 630      | 6.8   | 22.634    | 3.119453 | 6.445720 | 1.916923 |
| 660      | 7.9   | 69.476    | 4.240981 | 6.492240 | 2.066863 |
| 690      | 8.0   | 69.463    | 4.240794 | 6.536692 | 2.079442 |
| 980      | 7.5   | 35.867    | 3.579818 | 6.887553 | 2.014903 |
| 980      | 7.5   | 36.937    | 3.609214 | 6.887553 | 2.014903 |
| 970      | 12.3  | 130.740   | 4.873211 | 6.877296 | 2.509599 |
| 960      | 12.3  | 125.725   | 4.834097 | 6.866933 | 2.509599 |
| 960      | 4.3   | 11.327    | 2.427189 | 6.866933 | 1.458615 |
| 1000     | 4.2   | 12.320    | 2.511224 | 6.907755 | 1.435085 |
| 960      | 9.0   | 57.145    | 4.045592 | 6.866933 | 2.197225 |
| 970      | 8.9   | 58.715    | 4.072695 | 6.877296 | 2.186051 |
| 950      | 12.5  | 127.868   | 4.850998 | 6.856462 | 2.525729 |

| N (c/ha) | H (m) | M (m3/ha) | Ln(M)    | Ln(N)    | Ln(H)    |
|----------|-------|-----------|----------|----------|----------|
| 970      | 12.5  | 127.934   | 4.851515 | 6.877296 | 2.525729 |
| 1090     | 7.1   | 39.355    | 3.672623 | 6.993933 | 1.960095 |
| 1140     | 7.2   | 41.830    | 3.733614 | 7.038784 | 1.974081 |
| 890      | 12.8  | 158.216   | 5.063961 | 6.791221 | 2.549445 |
| 960      | 13.0  | 178.428   | 5.184185 | 6.866933 | 2.564949 |
| 1050     | 16.3  | 227.649   | 5.427805 | 6.956545 | 2.791165 |
| 1020     | 15.8  | 231.610   | 5.445055 | 6.927558 | 2.760010 |
| 1110     | 16.8  | 233.835   | 5.454616 | 7.012115 | 2.821379 |
| 1060     | 17.0  | 203.820   | 5.317237 | 6.966024 | 2.833213 |
| 940      | 16.6  | 186.508   | 5.228474 | 6.845880 | 2.809403 |
| 730      | 16.0  | 145.706   | 4.981591 | 6.593045 | 2.772589 |
| 1440     | 10.5  | 139.988   | 4.941557 | 7.272398 | 2.351375 |
| 1240     | 10.1  | 123.668   | 4.817601 | 7.122867 | 2.312535 |
| 740      | 12.8  | 111.255   | 4.711825 | 6.606650 | 2.549445 |
| 710      | 13.3  | 82.429    | 4.411937 | 6.565265 | 2.587764 |
| 360      | 17.9  | 227.702   | 5.428038 | 5.886104 | 2.884801 |
| 390      | 17.7  | 191.751   | 5.256198 | 5.966147 | 2.873565 |
| 570      | 10.9  | 43.846    | 3.780683 | 6.345636 | 2.388763 |
| 570      | 11.1  | 53.212    | 3.974284 | 6.345636 | 2.406945 |

Từ đây chạy phần Regression trong Excel để thiết lập mô hình như là mô hình tuyến tính nhiều lớp đã trình bày ở mục trên.

The screenshot shows the following data in the Excel worksheet:

| Số liệu N - H - M rừng Tách | N (c/ha) | H (m) | M (m3/ha) | ln(M)   | ln(N)   | ln(H)   |
|-----------------------------|----------|-------|-----------|---------|---------|---------|
| 8                           | 180      | 23.0  | 163.452   | 5.09652 | 5.19296 | 3.13549 |
| 9                           | 170      | 23.0  | 160.154   | 5.07614 | 5.1358  | 3.13549 |
| 10                          | 220      | 22.3  | 184.167   | 5.21584 | 5.39363 | 3.10459 |
| 11                          | 210      | 22.1  | 145.436   | 4.97974 | 5.34711 | 3.09558 |
| 12                          | 650      | 6.9   | 24.177    | 3.1854  | 6.47697 | 1.93152 |
| 13                          | 620      | 7.0   | 23.400    | 3.15274 | 6.42972 | 1.94591 |
| 14                          | 690      | 6.7   | 22.027    | 3.09227 | 6.53669 | 1.90211 |
| 15                          | 650      | 6.4   | 19.696    | 2.98042 | 6.47697 | 1.8563  |
| 16                          | 650      | 6.8   | 20.778    | 3.03389 | 6.47697 | 1.91692 |
| 17                          | 630      | 6.8   | 22.634    | 3.11945 | 6.44572 | 1.91692 |
| 18                          | 660      | 7.9   | 69.476    | 4.24098 | 6.49224 | 2.06686 |
| 19                          | 690      | 8.0   | 69.463    | 4.24079 | 6.53669 | 2.07944 |
| 20                          | 980      | 7.5   | 35.867    | 3.57982 | 6.88755 | 2.0149  |
| 21                          | 980      | 7.5   | 36.937    | 3.60921 | 6.88755 | 2.0149  |

The Regression dialog box is configured as follows:

- Input Y Range: \$D\$7:\$D\$47
- Input X Range: \$E\$7:\$F\$47
- Labels:  Labels
- Confidence Level: 95%
- Output Range: \$H\$7
- Residuals:  Residuals,  Standardized Residuals,  Residual Plots,  Line Fit Plots
- Normal Probability:  Normal Probability Plots

## Kết quả ước lượng mô hình phi tuyến đối về tuyến tính 2 lớp

### SUMMARY OUTPUT

| <i>Regression Statistics</i> |          |
|------------------------------|----------|
| Multiple R                   | 0.967567 |
| R Square                     | 0.936185 |
| Adjusted R Square            | 0.932736 |
| Standard Error               | 0.235045 |
| Observations                 | 40       |

### ANOVA

|            | <i>df</i> | <i>SS</i> | <i>MS</i> | <i>F</i>    | <i>Significance F</i> |  |
|------------|-----------|-----------|-----------|-------------|-----------------------|--|
| Regression | 2         | 29.98797  | 14.99398  | 271.4024864 | 7.78065E-23           |  |
| Residual   | 37        | 2.044113  | 0.055246  |             |                       |  |
| Total      | 39        | 32.03208  |           |             |                       |  |

|           | <i>Coefficients</i> | <i>Standard Error</i> | <i>t Stat</i> | <i>P-value</i> | <i>Lower 95%</i> | <i>Upper 95%</i> |
|-----------|---------------------|-----------------------|---------------|----------------|------------------|------------------|
| Intercept | -5.143280           | 0.681084              | -7.551614     | 0.000000       | -6.523287        | -3.763274        |
| Ln(N)     | 0.641785            | 0.081858              | 7.840245      | 0.000000       | 0.475926         | 0.807645         |
| Ln(H)     | 2.205407            | 0.095991              | 22.975155     | 0.000000       | 2.010911         | 2.399903         |

Mô hình có dạng:  $\ln(M) = -5.143 + 0.642\ln(\hat{a}) + 2.205\ln(H)$

Với  $\hat{a} = 40$      $R = 0.968$      $F_r = 271.40$     với  $\alpha < 0.00$

$tb_1 = 7.84$      $tb_2 = 22.98$     với  $\alpha < 0.00$

Từ mô hình này có thể dự báo M thông qua  $\hat{a}$  và H

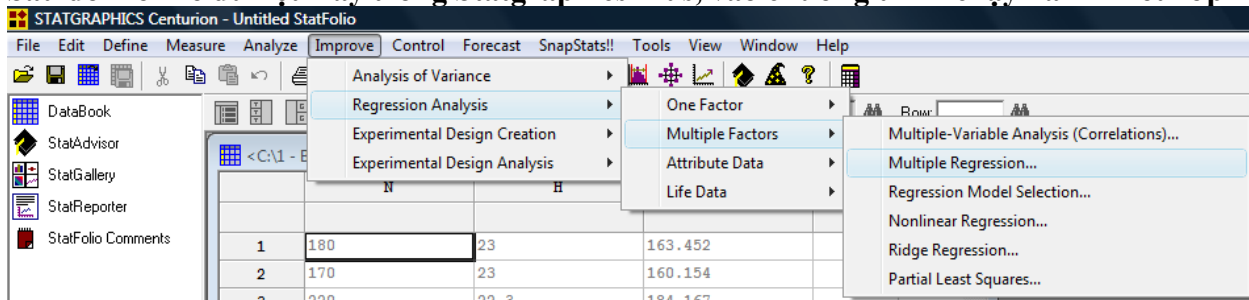
#### 6.4.2. Lập mô hình phi tuyến nhiều lớp chuyển về tuyến tính trong Statgraphics

Trong Statgraphics, việc tính toán mô hình phi tuyến nhiều lớp đơn giản hơn vì không cần tạo thêm các cột đối biến số, biến số được đối trực tiếp trong hộp thoại khi thiết lập mô hình.

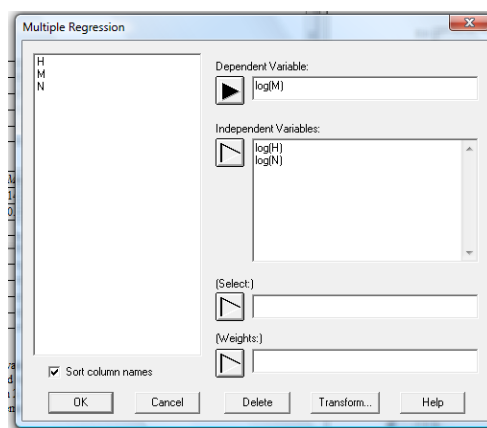
**Đầu tiên nhập dữ liệu trong Excel với biến y và các biến xi, mỗi biến một cột dữ liệu**

|   | A1  |      | <i>f<sub>x</sub></i> | N |
|---|-----|------|----------------------|---|
|   | A   | B    | C                    | D |
| 1 | N   | H    | M                    |   |
| 2 | 180 | 23.0 | 163.452              |   |
| 3 | 170 | 23.0 | 160.154              |   |
| 4 | 220 | 22.3 | 184.167              |   |
| 5 | 210 | 22.1 | 145.436              |   |
| 6 | 650 | 6.9  | 24.177               |   |
| 7 | 620 | 7.0  | 23.400               |   |
| 8 | 600 | 6.7  | 22.027               |   |

Sau đó mở file dữ liệu này trong Statgraphics Plus, vào chương trình chạy hàm nhiều lớp



Đổi biến số ngay trong hộp thoại



Kết quả chạy hàm phi tuyến nhiều lớp quy về tuyến tính

**Multiple Regression - log(M)**

Dependent variable: log(M)

Independent variables:

log(H)

log(â )

|            |          | Standard  | T         |         |
|------------|----------|-----------|-----------|---------|
| Parameter  | Estimate | Error     | Statistic | P-Value |
| COã STAã T | -5.14328 | 0.681084  | -7.55161  | 0.0000  |
| log(H)     | 2.20541  | 0.095991  | 22.9752   | 0.0000  |
| log(â )    | 0.641785 | 0.0818578 | 7.84025   | 0.0000  |

**Analysis of Variance**

| Source        | Sum of Squares | Df | Mean Square | F-Ratio | P-Value |
|---------------|----------------|----|-------------|---------|---------|
| Model         | 29.988         | 2  | 14.994      | 271.40  | 0.0000  |
| Residual      | 2.04411        | 37 | 0.0552463   |         |         |
| Total (Corr.) | 32.0321        | 39 |             |         |         |

**R-squared = 93.6185 percent**

R-squared (adjusted for d.f.) = 93.2736 percent

Standard Error of Est. = 0.235045

Mean absolute error = 0.16285

Durbin-Watson statistic = 1.47918 (P=0.0243)

Lag 1 residual autocorrelation = 0.243443

### The StatAdvisor

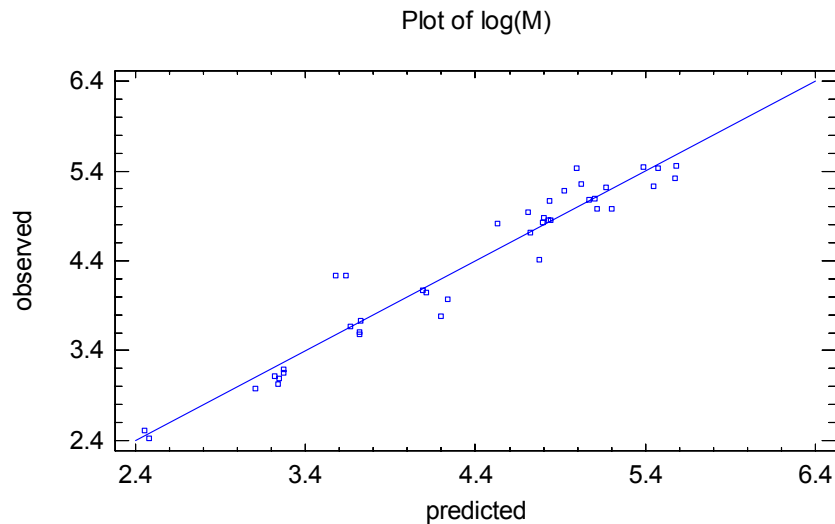
The output shows the results of fitting a multiple linear regression model to describe the relationship between  $\log(M)$  and 2 independent variables. The equation of the fitted model is

$$\log(M) = -5.14328 + 2.20541 \cdot \log(H) + 0.641785 \cdot \log(N)$$

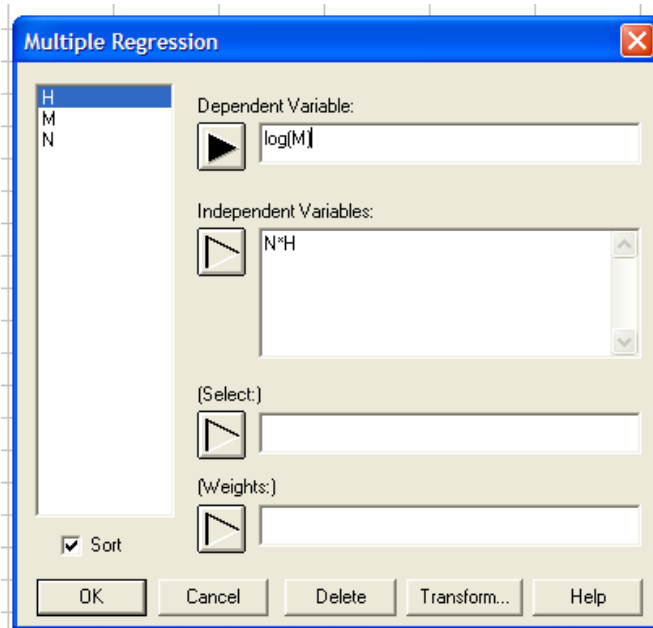
Since the P-value in the A $\hat{A}$  OVA table is less than 0.05, there is a statistically significant relationship between the variables at the 95.0% confidence level.

The R-Squared statistic indicates that the model as fitted explains 93.6185% of the variability in  $\log(M)$ . The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 93.2736%. The standard error of the estimate shows the standard deviation of the residuals to be 0.235045. 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 0.16285 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  $\log(\hat{A})$ . 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.



Trong Statgraphics Plus còn cho phép tạo tổ hợp biến ngay trong hộp thoại, ví dụ có thể lập hàm dạng:  $\ln(M) = a + b_1 \cdot N \cdot H$ , trong đó  $\hat{A} \cdot \ln(H)$  là tổ hợp biến. Trong hộp thoại tạo tổ hợp biến như sau



### Kết quả có hàm theo quan hệ nhiều biến dưới dạng tổ hợp biến

#### Multiple Regression - log(M)

Dependent variable: log(M)

Independent variables:

$\hat{\alpha} * H$

|  |                 | <i>Standard</i> | <i>T</i>         |                |
|--|-----------------|-----------------|------------------|----------------|
| <i>Parameter</i>                       | <i>Estimate</i> | <i>Error</i>    | <i>Statistic</i> | <i>P-Value</i> |
| CO $\hat{\alpha}$ STA $\hat{\alpha}$ T | 3.17609         | 0.248379        | 12.7873          | 0.0000         |
| $\hat{\alpha} * H$                     | 0.000133068     | 0.0000252748    | 5.26485          | 0.0000         |

#### Analysis of Variance

| <i>Source</i> | <i>Sum of Squares</i> | <i>Df</i> | <i>Mean Square</i> | <i>F-Ratio</i> | <i>P-Value</i> |
|---------------|-----------------------|-----------|--------------------|----------------|----------------|
| Model         | 13.5104               | 1         | 13.5104            | 27.72          | 0.0000         |
| Residual      | 18.5217               | 38        | 0.487412           |                |                |
| Total (Corr.) | 32.0321               | 39        |                    |                |                |

**R-squared = 42.1778 percent**

R-squared (adjusted for d.f.) = 40.6561 percent

Standard Error of Est. = 0.698149

Mean absolute error = 0.515141

Durbin-Watson statistic = 0.780029 (P=0.0000)

Lag 1 residual autocorrelation = 0.559301

#### The StatAdvisor

The output shows the results of fitting a multiple linear regression model to describe the relationship between log(M) and 1 independent variables. The equation of the fitted model is

$$\log(M) = 3.17609 + 0.000133068 * N * H$$

Since the P-value in the A $\hat{\alpha}$  OVA 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 42.1778% of the variability in log(M). The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 40.6561%. The standard error of the estimate shows the standard deviation of the residuals to be 0.698149. 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 0.515141 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  $\hat{\alpha} \cdot H$ . 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.

## 7. ƯỚC LƯỢNG CÁC DẠNG HỒI QUY MỘT BIẾN TRÊN ĐỒ THỊ

Trong thực tế trực quan các mối quan hệ, người ta thường dùng đồ thị để biểu diễn, và để dễ dàng trong việc xem xét các dự báo, Excel hỗ trợ chương trình xác định mô hình hồi quy một biến ngay trên đồ thị. Excel lập sẵn 5 dạng hàm phổ biến trong phần này.

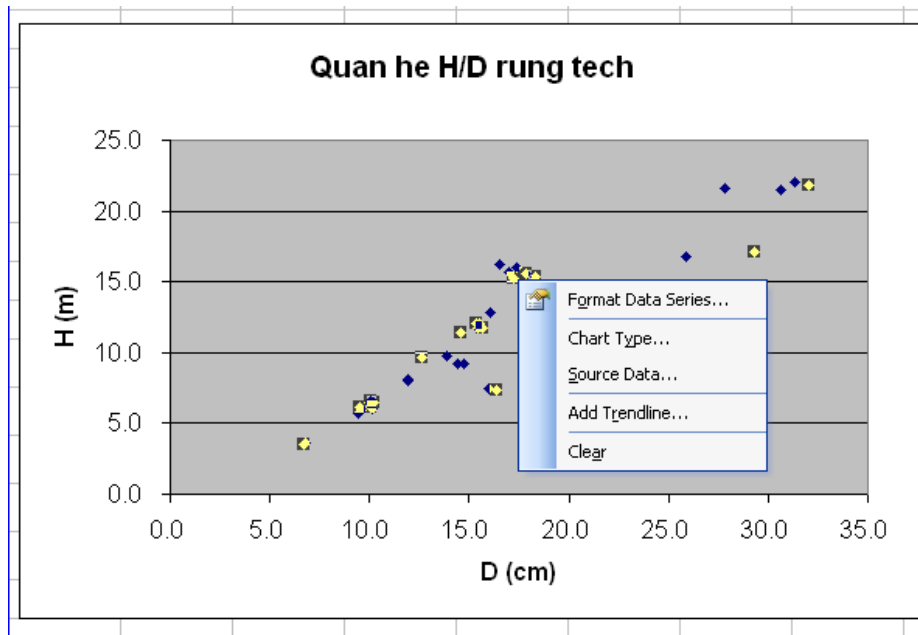
Ví dụ: Lập mô hình hồi quy H/D cho rừng trồng Tách ngay trên đồ thị quan hệ

- Nhập số liệu:**

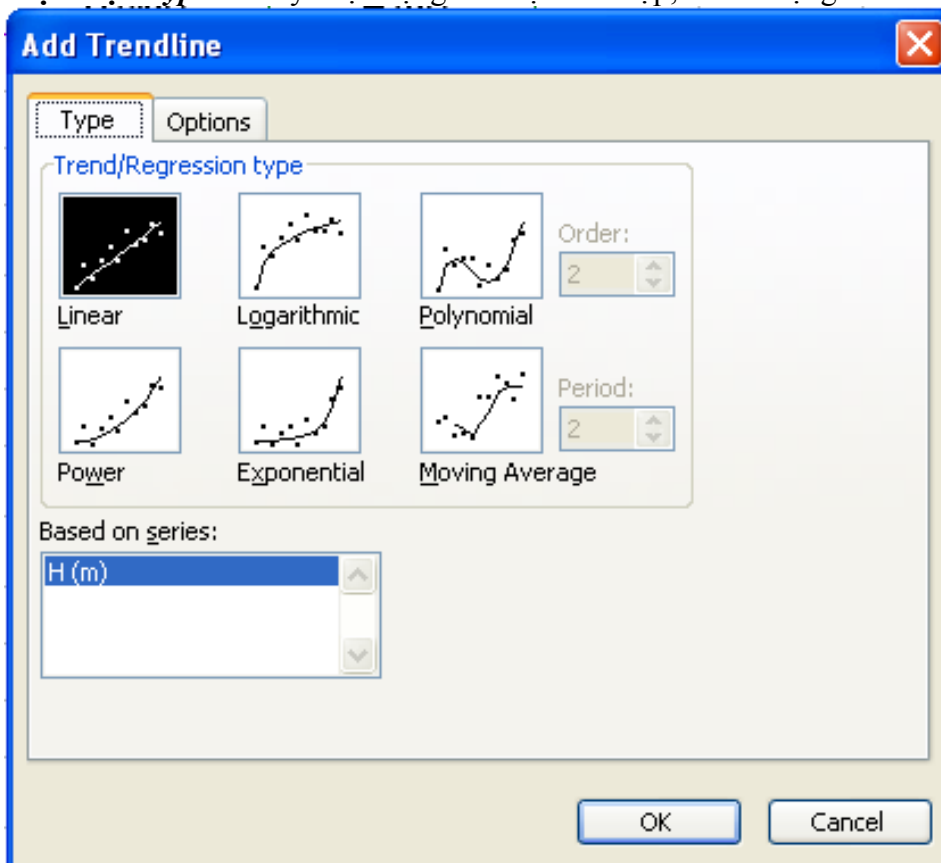
**Số liệu về quan hệ H/D**

|           | <b>A</b>     | <b>B</b>    |
|-----------|--------------|-------------|
| <b>1</b>  | <b>D(cm)</b> | <b>H(m)</b> |
| <b>2</b>  | 6,7          | 3,5         |
| <b>3</b>  | 6,8          | 3,6         |
| <b>4</b>  | 9,5          | 5,7         |
| <b>5</b>  | 9,5          | 6,1         |
| ...       | ...          | ...         |
| <b>40</b> | 31,3         | 22,0        |
| <b>41</b> | 32,0         | 21,8        |

- Vẽ đồ thị:** Tiến hành các bước vẽ đồ thị quan hệ H/D. (ở đây vẽ dạng đám mây điểm).
- Tính toán mô hình quan hệ dựa vào đồ thị:**
  - Kích hoạt đồ thị: Kích chuột trái.
  - Chọn đám mây điểm trên đồ thị: Kích chuột phải vào đám mây điểm này.
  - Chọn Add Trendline



**Chọn mục Type:** Ở đây chọn dạng liên hệ thích hợp, có các dạng sau:



Linear:  $y = mx + b$

Logarithmic:  $y = c \ln x + b$

Polynomial:  $y = b + c_1x + c_2x^2 + \dots + c_6x^6$

Có thể chọn 1 đến 6 bậc trong ô Order: Xác định số bậc.

Power:  $y = cx^b$

Exponential:  $y = c.e^{bx}$



**Chọn mục Option:** Xác định:

Forecast: Foward: Xác định độ dài dự đoán tiếp theo.

Backward: Xác định độ dài dự đoán lùi.

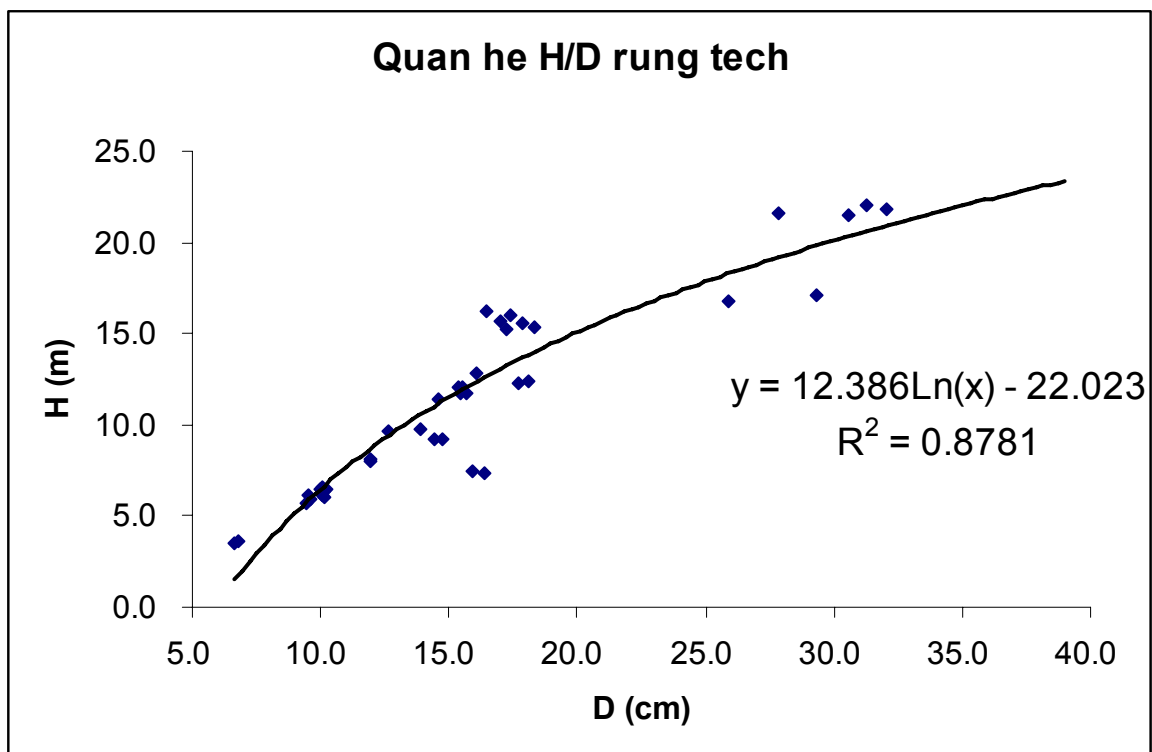
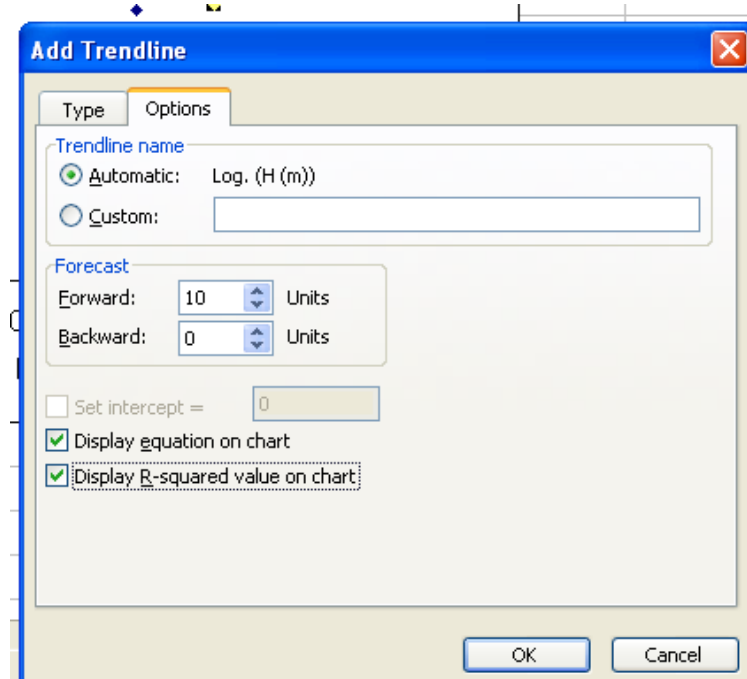
Set intercept (0): ấu ều đánh dấu thì tham số  $b=0$  trong các hàm đường thẳng

Display Equation on Chart: Đánh dấu để đưa hàm lên đồ thị.

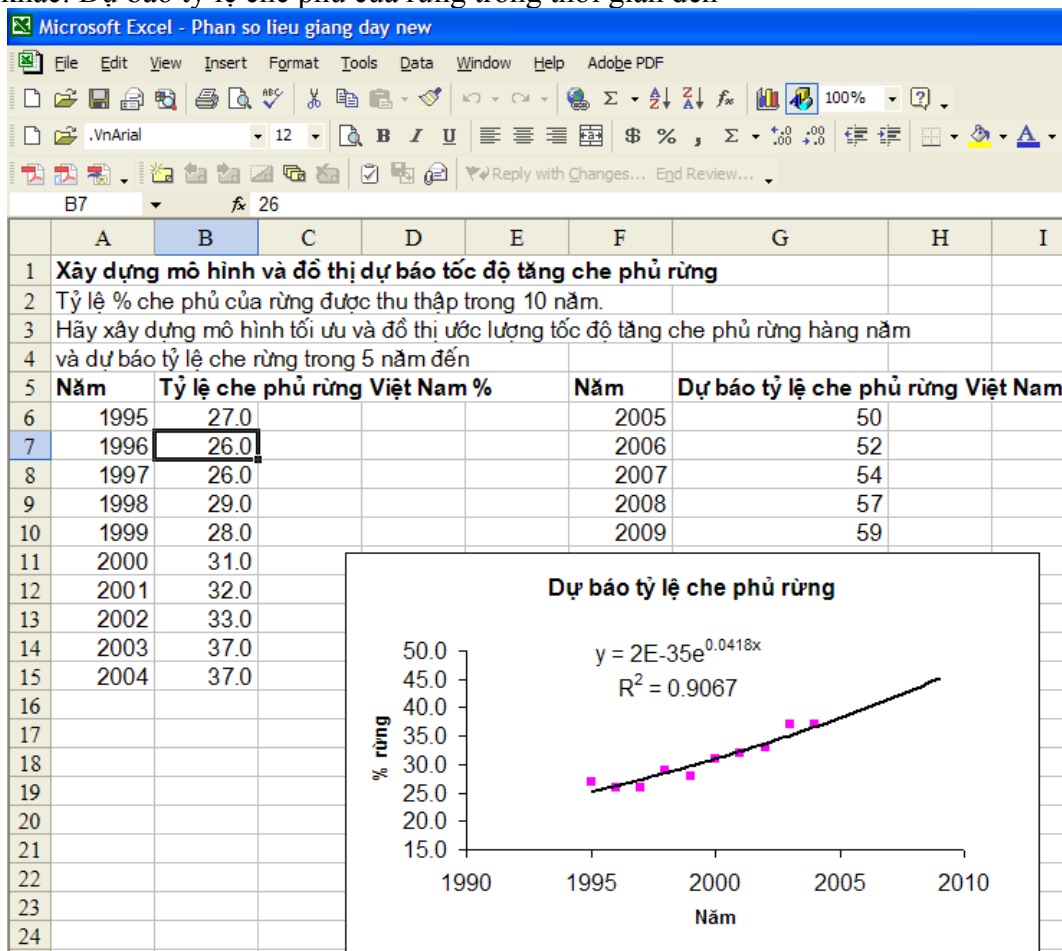
Display R-squared Value on Chart: Đánh dấu nếu muốn tính hệ số tương quan

Cuối cùng là OK.

**Kết quả như sau:**



Ví dụ khác: Dự báo tỷ lệ che phủ của rừng trong thời gian đến

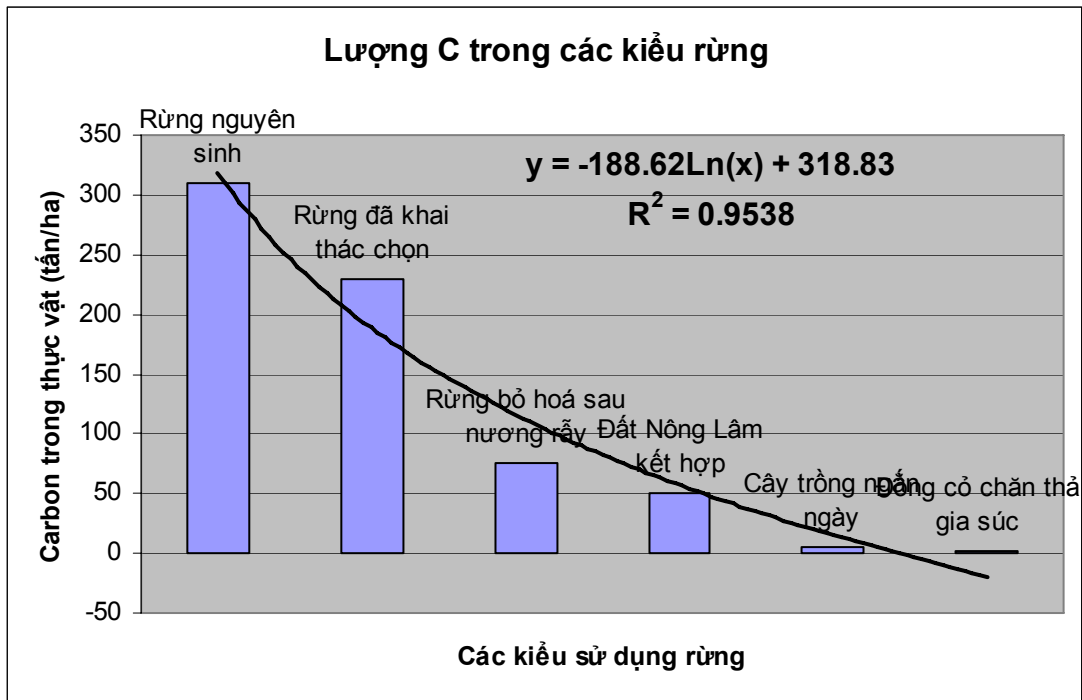


Ví dụ khác: Lượng carbon được lưu trữ trong các kiểu rừng khác nhau được mô phỏng bằng dạng hàm phi tuyến trên đồ thị. Trong đó không cần mã hóa biến số x (kiểu rừng), lúc này sử dụng sơ đồ cột để vẽ và chạy phương trình thích hợp. Lúc này máy tính đã tự động mã hóa các kiểu rừng là 1, 2, 3, 4 .....

#### Lượng carbon trên và dưới mặt đất ở các kiểu sử dụng đất rừng

Các vùng rừng ở Brazil, Cameroon và Indonesia

| Các kiểu rừng             | Lượng carbon (tấn/ha) |                |
|---------------------------|-----------------------|----------------|
|                           | Dưới mặt đất          | Trong thực vật |
| Rừng nguyên sinh          | 48                    | 310            |
| Rừng đã khai thác chọn    | 48                    | 230            |
| Rừng bỏ hoá sau nương rẫy | 48                    | 75             |
| Đất ở ông Lâm kết hợp     | 45                    | 50             |
| Cây trồng ngắn ngày       | 25                    | 5              |
| Đồng cỏ chăn thả gia súc  | 20                    | 2              |



## 8. SẮP XẾP VÀ VẼ BIỂU ĐỒ PHÂN BỐ TẦN SỐ XUẤT HIỆN THEO CẤP, CỖ, HẠNG

Đây là chức năng sắp xếp bảng phân bố tần số theo một nhân tố theo từng cấp, hạng, ... và vẽ đồ thị phân bố.

Trong nghiên cứu xã hội, người ta cần nghiên cứu tần số phân bố số người theo cấp tuổi để biết sự phân bố 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 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 ( $\hat{a}/D$ ), số cây theo cỡ chiều cao ( $\hat{a}/H$ ), số cây theo cấp thể tích ( $\hat{a}/V$ ), số cây theo loài cây theo các tầng rừng, thế hệ.

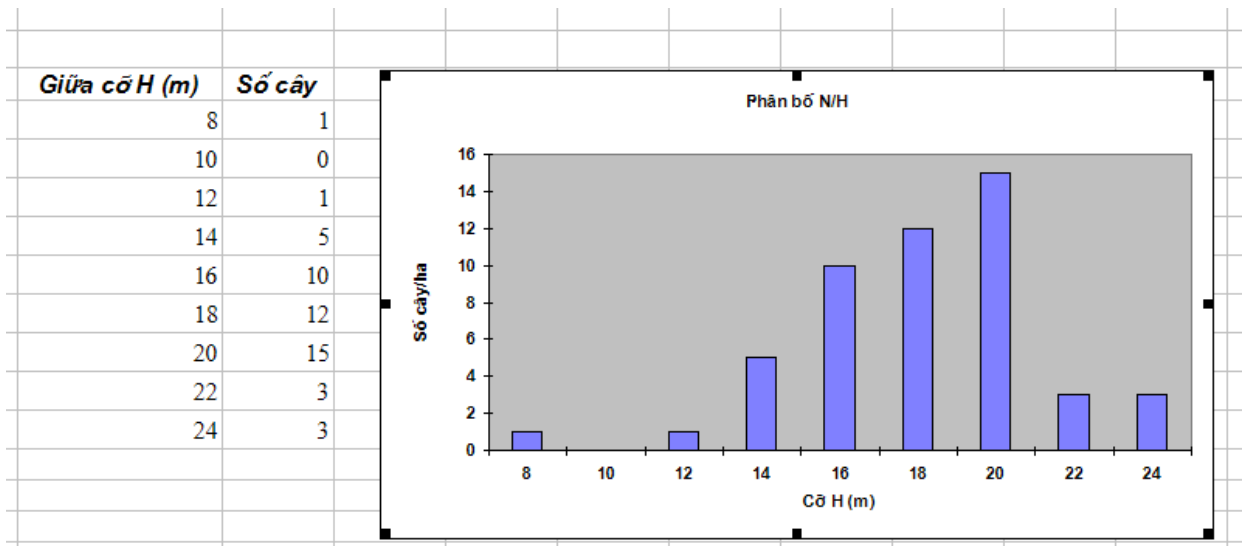
Ví dụ cũng từ số liệu quan sát rừng trồng Tách 10 tuổi, tiến hành sắp xếp phân bố thực nghiệm  $\hat{a}/H$  và vẽ biểu đồ (cấp H là 2m):

- $\hat{a}$  áp số liệu vào bảng tính theo cột. Vd: A9:A52.
- Lập một cột giới hạn trên cỡ kính. Vd: cỡ 2m.

## Bảng tóm tắt dữ liệu

|    | A   | B | C                      | D | E | F |
|----|---|---|------------------------|---|---|---|
| 1  | <b>Bài tập Sắp xếp tần số phân bố N/H</b>                 |   |                        |   |   |   |
| 2  | Từ số liệu đo đếm H của các cây trong ô tiêu chuẩn.       |   |                        |   |   |   |
| 3  | Hãy sắp xếp và vẽ biểu đồ phân bố N/H với cự ly cỡ H = 2m |   |                        |   |   |   |
| 4  |   |   |                        |   |   |   |
| 5  |   |   |                        |   |   |   |
| 6  | <b>Số liệu đo cao trong rừng trồng Tách 10 tuổi</b>       |   |                        |   |   |   |
| 7  |   |   |                        |   |   |   |
| 8  | H (m)   |   | Giới hạn trên cỡ H (m) |   |   |   |
| 9  | 9.9   |   | 10                     |   |   |   |
| 10 | 12.4  |   | 12                     |   |   |   |
| 11 | 14.0  |   | 14                     |   |   |   |
| 12 | 14.3  |   | 16                     |   |   |   |
| 13 | 15.0  |   | 18                     |   |   |   |
| 14 | 15.0  |   | 20                     |   |   |   |
| 15 | 15.9  |   | 22                     |   |   |   |
| 16 | 16.2  |   | 24                     |   |   |   |
| 17 | 16.6  |   | 28                     |   |   |   |
| 18 | 16.6  |   | 28                     |   |   |   |
| 19 | 16.6  |   | 30                     |   |   |   |
| 20 | 16.9  |   |                        |   |   |   |

- Menu Tools/Data Analysis/Histogram/OK. Xuất hiện hộp thoại, xác định:
  - + Input range: Khai báo khối dữ liệu
  - + Bin range: Khai báo khối chứa cự ly tổ (với giá trị lớn nhất của mỗi tổ)
  - + 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.



Kết quả sắp xếp tần số cho được một dãy dữ liệu theo cấp và biểu đồ phân bố. Nó phản ánh cụ thể hơn đặc trưng mẫu và cho thấy hình ảnh của kiểu dạng phân bố theo cấp, thể 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 tía thưa loại bỏ bớt cây sinh trưởng kém có H < 12m và có thể tía 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.

Công cụ sắp xếp tần số, tần suất phân bố là một công cụ mạnh và hữu ích trong lâm nghiệp, vì thông thường từ số liệu đo đếm, điều tra, cần phải sắp xếp theo tổ, nhóm, cấp. Ví dụ sắp xếp  $\hat{a}/D$ ,  $\hat{a}/H$ , số cây theo cấp tuổi rừng trồng, .....

## 9. KIỂM TRA THUẬN NHẤT K MẪU QUAN SÁT ĐỨT QUẢNG - ỨNG DỤNG: KIỂM TRA SỰ THUẬN NHẤT CỦA CÁC DÃY PHÂN BỐ N/D, N/H Ở CÁC Ô TIÊU CHUẨN

Trong thực tế, để điều tra đánh giá, nghiên cứu một đối tượng rừng, một trạng thái; chúng ta thường rút mẫu theo ô tiêu chuẩn, sau đó gộp chung để tính toán. Tuy nhiên có khả năng một số ô mẫu **không** nằm trong một trạng thái đó, do đó nếu gộp chung để tính toán trung bình, mô phỏng cấu trúc chung sẽ gặp sai số rất lớn và không phản ánh đúng đối tượng, trạng thái nghiên cứu. Do vậy cần có sự kiểm tra sự thuận nhất của các số liệu này ở các ô mẫu trước khi gộp chung để tính toán như một tổng thể.

Trong trường hợp này cần ứng dụng tiêu chuẩn thống kê  $\chi^2$  để kiểm tra K mẫu quan sát đứt quảng, cụ thể là ứng dụng để kiểm tra các dãy phân bố  $\hat{a}/D$  ở các ô mẫu để đánh giá chúng có cùng một tổng thể, một trạng thái hay không.

Trong đó:

Dãy phân bố  $\hat{a}$  của các ô được sắp xếp lần lượt  $i = 1, 2, \dots, k$  (k ô)

và được phân bố theo các cỡ kính được sắp xếp theo thứ tự  $j = 1, 2, \dots, m$  (m cỡ kính)

### Phân bố N của các ô i mẫu theo các cỡ kính j

| Cỡ D1.3 (j= 1 ... m) | Tần số f <sub>ij</sub> (i=1 .... k) (N/ha) |                |                |                |                | Tổng           | f <sub>j</sub> |
|----------------------|--|----------------|----------------|----------------|----------------|----------------|----------------|
|                      | ô 1  | ô 2            | ô 3            | ô 4 = k        | Tổng           |                |                |
| 8                    | 12   | 11             | 9              | 8              | 40             | f <sub>1</sub> |                |
| 12                   | 134  | 125            | 150            | 145            | 554            | f <sub>2</sub> |                |
| 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             | f <sub>j</sub> |                |
| 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              | f <sub>m</sub> |                |
| <b>Tổng</b>          | <b>443</b>                                 | <b>397</b>     | <b>423</b>     | <b>415</b>     | <b>1678</b>    |                |                |
|                      | n <sub>i</sub>                             | n <sub>1</sub> | n <sub>2</sub> | n <sub>i</sub> | n <sub>k</sub> | n              |                |

Gọi f<sub>ij</sub> là tần số (số cây) theo ô i và cấp kính j; f<sub>i</sub> là tổng tần số của ô i và f<sub>j</sub> là tổng tần số cấp kính j.

Giả thuyết H<sub>0</sub>: F<sub>1</sub> = F<sub>2</sub> = ... = F<sub>k</sub> (Cho mọi i và j). Hay nói khác là các dãy phân bố  $\hat{A}/D$  là thuần nhất ở các ô mẫu khác nhau, hay cùng 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:

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(f_{ij} - f_j \cdot n_i / n)^2}{f_j \cdot n_i / n}$$

So sánh với  $\chi^2_{(0.05; df = (m-1)(k-1))}$ , nếu  $\chi^2_t < \chi^2_{(0.05; df = (m-1)(k-1))} >$  Chấp nhận giả thuyết H<sub>0</sub>, có nghĩa là 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  $\hat{A}/\hat{D}$  chung .....

Trong ví dụ trên  $\chi^2_t = 20.42 < \chi^2_{(0.05; df=36)} = 51.00$ . Do đó có thể xem các dãy  $\hat{A}/D$  của 4 ô mẫu là thuần nhất hoặc nằm trong cùng một trạng thái, có thể gộp để mô phỏng  $\hat{A}/D$  chung.

## 10. MÔ HÌNH HOÁ QUY LUẬT PHÂN BỐ

Trong nghiên cứu các lâm phần, người ta thường khái quát quy luật phân bố số cây theo cỡ kính, chiều cao để làm cơ sở cho việc điều tra rừng và xác định các giải pháp lâm sinh thích hợp để dẫn dắt rừng. Hoặc nghiên cứu phân bố số cá thể theo tuổi, thế hệ; phân bố số loài theo tầng thứ, phân bố vi sinh vật đất theo các lớp đất, .... để hiểu rõ quy luật sinh học, sinh thái học làm cơ sở quản lý tài nguyên thiên nhiên bền vững.

## 10.1. Mô hình hoá phân bố giảm theo hàm Mayer

Hàm Mayer có dạng:  $y = \alpha \cdot e^{-\beta \cdot x}$ . Kiểu dạng này thích hợp cho mô tả mô phỏng phân bố số cây theo cỡ kính (đường kính) rừng chặt chọn có dạng giảm, hoặc mô phỏng sự giảm của số loài theo tầng, theo cỡ kính, ...

Trong Excel có chương trình lập sẵn tính quan hệ tuyến tính, vì vậy để xác định các tham số cần tuyến tính hóa.

Ví dụ mô phỏng phân bố đường kính theo dạng Mayer:  $\hat{a} = \alpha \cdot e^{-\beta \cdot D}$

□ **Tuyến tính hoá:**

$$\ln(\hat{a}) = \ln \alpha - \beta D$$

Đặt  $Y = \ln(\hat{a})$ ;  $A = \ln(\alpha)$ ;  $B = -\beta$ ;  $X = D$

Có hàm tuyến tính:  $Y = A + BX$

Trình tự tính toán trong Excel:

□ **Nhập số liệu:**

- Cột A là giá trị giữa cỡ kính (D).
- Cột B là tần số thực nghiệm ( $\hat{a}$ ).
- Cột C:  $\ln(\hat{a})$ . Đầu tiên tính cho ô đầu tiên (C2):  $=\ln(b2)$ , sau đó copy xuống cho các ô còn lại.

**Bảng dữ liệu đôi biến số để ước lượng hàm Mayer**

|    | A           | B        | C        |
|----|-------------|----------|----------|
| 1  | C D1,3 (cm) | N (c/ha) | Ln(N)    |
| 2  | 15          | 125      | 4,828314 |
| 3  | 25          | 89       | 4,488636 |
| 4  | 35          | 56       | 4,025352 |
| 5  | 45          | 31       | 3,433987 |
| 6  | 55          | 19       | 2,944439 |
| 7  | 65          | 8        | 2,079442 |
| 8  | 75          | 10       | 2,302585 |
| 9  | 85          | 5        | 1,609438 |
| 10 | 95          | 3        | 1,098612 |
| 11 | 105         | 2        | 0,693147 |
| 12 | 115         | 1        | 0        |

□ **Ước lượng hàm tuyến tính 1 lớp:** Menu Tools/Analysis/Regression/OK- Hộp thoại:

- Input X Range: Khai khối dữ liệu X. (A2:A12)
- Input Y range: Khai khối dữ liệu Y. (C2:C12)
- Output range: nhập địa chỉ ô trên ra nơi ra kết quả (E1).
- OK.

## Kết quả ước lượng hàm tuyến tính

E F G H

### SUMMARY OUTPUT

| <i>Regression Statistics</i> |             |
|------------------------------|-------------|
| Multiple R                   | 0,993975878 |
| R Square                     | 0,987988046 |
| Adjusted R Square            | 0,986653385 |
| Standard Error               | 0,183611852 |
| Observations                 | 11          |

### Ả OVA

|            | <i>df</i> | <i>SS</i>   | <i>MS</i>   | <i>F</i>    | <i>Significance F</i> |
|------------|-----------|-------------|-------------|-------------|-----------------------|
| Regression | 1         | 24,95640165 | 24,95640165 | 740,2536312 | 5,9318E-10            |
| Residual   | 9         | 0,303419808 | 0,033713312 |             |                       |
| Total      | 10        | 25,25982146 |             |             |                       |

|              | <i>Coefficients</i> | <i>Standard Error</i> | <i>t Stat</i> | <i>P-value</i> | <i>Lower 95%</i> | <i>Upper 95%</i> |
|--------------|---------------------|-----------------------|---------------|----------------|------------------|------------------|
| Intercept    | 5,596409498         | 0,126545736           | 44,2244021    | 7,72246E-12    | 5,310142937      | 5,882676058      |
| X Variable 1 | -0,047631542        | 0,00175067            | -27,20760245  | 5,9318E-10     | -0,051591836     | -0,043671248     |

a= 5,596      Alpha=                      269,457

b=-0,048      Beta=                                      0,048

- Quy về tham số nguyên thủy của Mayer:  
Tham số  $\alpha = e^a = e^{5,596} = 269,457$ ;  $\beta = -(-0,048) = 0,048$ .
- **Kết quả phương trình:  $N = 269,457 \cdot \exp(-0,048 \cdot D)$ .**
- Kiểm tra sự phù hợp của phương trình bằng tiêu chuẩn  $\chi^2$ :

Công thức tính  $\chi^2 = (\hat{a} \cdot l_t - \hat{a} \cdot t)^2 / \hat{a} \cdot l_t$

+ Cột E: Giá trị  $\chi^2$ : Tại ô E2:  $= (D2 - B2)^2 / D2$ . Copy cho các ô dưới. Chú ý các ô có  $\hat{a} \cdot l_t < 5$  cần gộp lại. Tổng cột E có  $\chi^2 = 4,00$ .

+ Tra  $\chi^2$  bảng với  $\alpha = 0,05$  và  $K = 1 - r - 1 = 9 - 2 - 1 = 6$ :  $= \text{Chiinv}(0,05; 6) = 12,59$

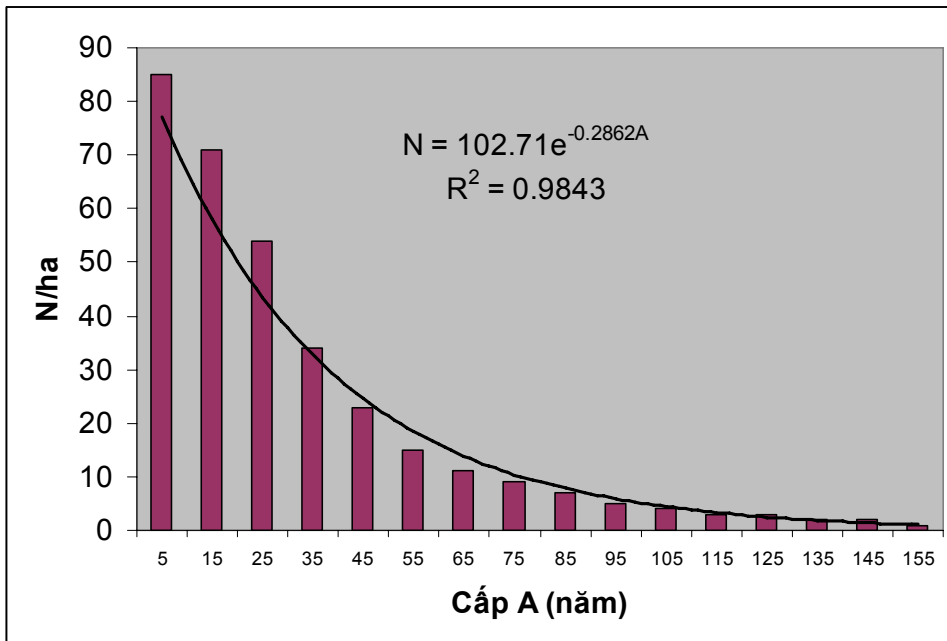
+  $\chi^2$  tính  $< \chi^2$  bảng. Kl: Hàm Mayer mô phỏng tốt cấu trúc  $\hat{a} \cdot -D$  thực nghiệm.



**Kết quả kiểm tra sự phù hợp của hàm Mayer bằng tiêu chuẩn  $\chi^2$**

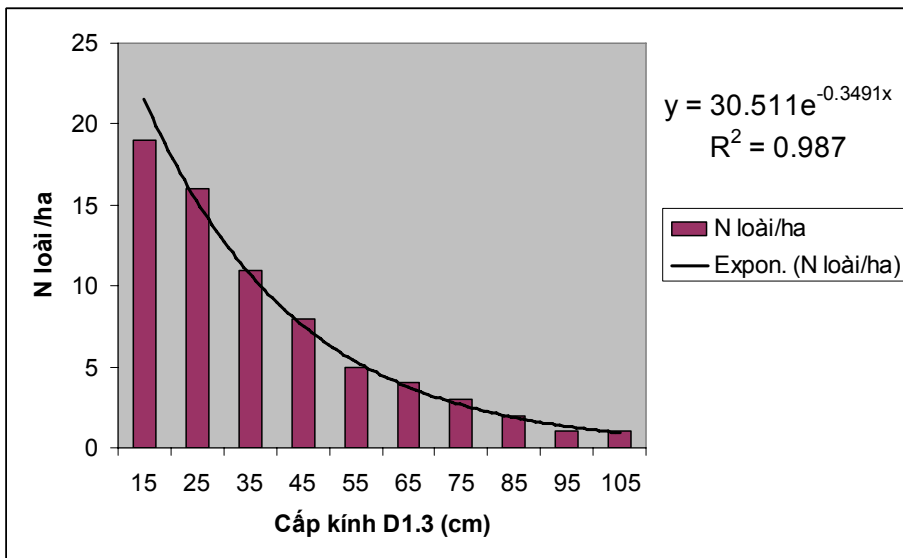
|    | A            | B          | C        | D          | E           |
|----|--------------|------------|----------|------------|-------------|
| 1  | Cỡ D1,3 (cm) | N (c/ha)   | Ln(N)    | Nlt (c/ha) | X2          |
| 2  | 15           | 125        | 4,828314 | 131        | 0,29        |
| 3  | 25           | 89         | 4,488636 | 81         | 0,76        |
| 4  | 35           | 56         | 4,025352 | 50         | 0,67        |
| 5  | 45           | 31         | 3,433987 | 31         | 0,00        |
| 6  | 55           | 19         | 2,944439 | 19         | 0,00        |
| 7  | 65           | 8          | 2,079442 | 12         | 1,28        |
| 8  | 75           | 10         | 2,302585 | 7          | 0,94        |
| 9  | 85           | 5          | 1,609438 | 5          | 0,06        |
| 10 | 95           | 3          | 1,098612 | 3          | 0,00        |
| 11 | 105          | 2          | 0,693147 | 2          |             |
| 12 | 115          | 1          | 0        | 1          |             |
| 13 | <b>Tổng</b>  | <b>349</b> |          | <b>342</b> | <b>4,00</b> |

Phân bố Mayer 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 ( $\hat{a}/A$ ) 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. 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 thực 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  $\hat{a}/A$  của lâm phần và  $\hat{a}/A$  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,.... Tuy nhiên trong thực tế 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  $\hat{a}/D$  để phục vụ cho điều tra, xác định chỉ tiêu kỹ thuật nuôi dưỡng, khai thác rừng. Mô hình hoá cấu trúc  $\hat{a}/A$  thường được biểu diễn tốt bằng hàm Mayer với hệ số tương quan  $R^2$  rất cao.



Ví dụ mô hình cấu trúc N/A rừng hỗn loài khác tuổi theo hàm Mayer

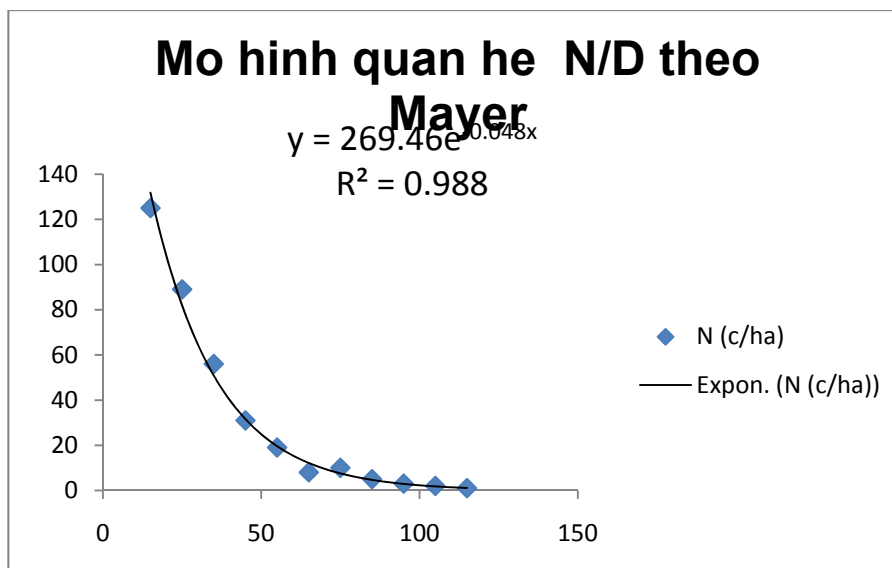
Rừng mưa nhiệt đới có khu hệ thực vật đa dạng với thành loài phong phú, phân bố ở nhiều thế hệ, cấp tuổi khác nhau. Trên 01 ha rừng có thể phát hiện trên 60 loài cây thân gỗ, ngoài ra rừng mưa rất phong phú các loài dây leo, song mây, rêu, dương xỉ, phong lan. Các loài cây nói chung là ưa sáng, cố gắng vươn lên cạnh tranh ánh sáng, tuy vậy cũng có loài chịu được ở tầng dưới và hình thành sự phân bố loài theo tầng, theo cấp tuổi, cấp kính khá rõ rệt. Trong thực tế việc xác định tuổi cây rừng là khó khăn, do đó thường nghiên cứu cấu trúc số loài theo cấp kính (số loài/D)



Cấu trúc số loài theo cấp kính rừng nửa rụng lá ưu hợp bằng lăng – cẩm xe ở Đắk Lắk

Cấu trúc s số loài/D của kiểu rừng nửa rụng lá ưu hợp bằng lăng – cẩm xe ở Đắk Lắk 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 thực từ 55cm trở lên thì số loài còn khoảng 5 loài. Kiểu dạng cấu trúc này cũng có thể mô phỏng tốt bằng dạng hàm Mayer.

Mô hình Mayer ngoài việc đổi biến số và tính toán, trong Excel có thể ước lượng nhanh ngay trên đồ thị



## 10.2. Mô phỏng phân bố thực nghiệm theo phân bố khoảng cách-hình học:

i) Dạng phân bố khoảng cách:

$$P(x) = \begin{cases} \gamma & x=0 \\ (1-\gamma) \cdot (1-\gamma) \cdot \alpha^{x-1} & x \geq 1 \end{cases}$$

Với x là mã số các cỡ kính từ nhỏ đến lớn 0,1,2,3....

Khi:  $\gamma < (1-\gamma)(1-\alpha)$  Phân bố có đỉnh tại x=1.

$\gamma = 1 - \alpha$  Phân bố giảm có thể thay thế bằng phân bố hình học.

$\gamma > (1-\gamma)(1-\alpha)$  Phân bố giảm.

Ước lượng 2 tham số bằng phương pháp cực đại hợp lý:

$$\gamma = N_0/N$$

$$\alpha = 1 - \frac{\sum_{i=1}^r N_i}{\sum_{i=1}^r N_i \cdot x_i}$$

Trình tự tính trong Excel: Vd: Mô phỏng phân bố N/D có dạng 1 đỉnh:

- Cột A: Mã số x
- Cột B: Giá trị giữa cỡ D.
- Cột C: Số cây theo cỡ kính. Tổng tại ô C13=SUM(C2:C12)
- Cột D:  $\alpha \cdot i \cdot x_i$ . Tại ô D2:=A2\*C2; copy cho các ô dưới. Tổng tại ô D13
- Tính 2 tham số:
  - $\gamma = C2/\text{SUM}(C2:C12)$
  - $\alpha = 1 - \text{SUM}(C3:C12)/\text{SUM}(D2:D12)$
- Cột E: Xác suất từng cỡ kính P(x<sub>i</sub>): Ô E2: P<sub>x0</sub>= $\gamma$ ; ô E3: P<sub>x1</sub> = (1- $\gamma$ )(1- $\alpha$ ) $\alpha^{(a3-1)}$ ; copy cho các ô dưới.
- Cột F: Tần số lý thuyết:  $\hat{a} \cdot l_{ti}$ : Ô F2: =C\$13\*E2; copy cho các ô dưới

- Cột G: Tính  $\chi^2$  từng cỡ và tổng.  $\hat{O} G2: = (f2-c2)^2/f2$ , copy cho các ô dưới, cộng tổng.
- $\hat{O} G14$ : Tra  $\chi^2$  bảng ( $\alpha=0,05$  ;  $K = 8-2-1=5$ ): =Chiinv(0.05,5)

Kết quả  $\chi^2$  tính <  $\chi^2$  bảng . K1: Phân bố Khoảng cách mô phỏng tốt phân bố thực nghiệm  $\hat{a}/D$ .

**Kết quả mô phỏng phân bố N/D theo phân bố khoảng cách**

|    | A           | B            | C          | D          | E               | F          | G           |
|----|-------------|--------------|------------|------------|-----------------|------------|-------------|
| 1  | x           | Cỡ D1,3 (cm) | N (c/ha)   | Nixi       | Px              | Nlt (c/ha) | X2          |
| 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        |
| 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 | <b>Tổng</b> |              | <b>330</b> | <b>593</b> | <b>0,997543</b> | <b>329</b> | <b>7,70</b> |
| 14 |             |              | Gamma=     | 0,212121   |                 | X2 bảng=   | 11,07       |
| 15 |             |              | Alpha=     | 0,561551   |                 | K=8-2-1=5  |             |

**ii) Phân bố hình học:**

$$P(x) = \alpha^x \cdot (1-\alpha) \quad x=0,1,2,3...r$$

Ước lượng  $\alpha$  bằng phương pháp cực đại hợp lý:

$$\alpha = \frac{x}{x+1}$$

$$x = \frac{1}{N} \sum_{i=1}^r Ni \cdot xi$$

Phân bố hình học dùng mô tả các phân bố thực nghiệm dạng giảm

Trình tự tính trong Excel: Vd: Mô phỏng phân bố N/D có dạng giảm:

- Cột A: Mã số x
- Cột B: Giá trị giữa cỡ D.
- Cột C: Số cây theo cỡ kính. Tổng tại ô C13=sum(c2:c12)
- Cột D:  $\hat{a} \cdot i \cdot xi$ . Tại ô D2:=A2\*C2; copy cho các ô dưới. Tổng tại ô D13
- Tính tham số  $\alpha$ :  
 $x = D13/c13$   
 $\alpha = x/(x+1)$
- Cột E: Xác suất từng cỡ kính P(xi):  $\hat{O} E2: Px0 = (1-\alpha)\alpha^x$ ; copy cho các ô dưới.
- Cột F: Tần số lý thuyết:  $\hat{a} \cdot lti$ :  $\hat{O} F2: =\$C\$13*E2$ ; copy cho các ô dưới

- Cột G: Tính  $\chi^2$  từng cỡ và tổng.  $\hat{O} G2: = (f2-c2)^2/f2$ , copy cho các ô dưới, cộng tổng.
- Ô G14: Tra  $\chi^2$  bảng ( $\alpha=0,05$  ;  $K = 8-1-1=6$ ): =Chiinv(0.05,6)

Kết quả  $\chi^2$  tính <  $\chi^2$  bảng . K1: Phân bố hình học mô phỏng tốt phân bố thực nghiệm ả /D.

### Kết quả mô phỏng phân bố N/D theo phân bố hình học

|    | A           | B            | C          | D          | E               | F          | G           |
|----|-------------|--------------|------------|------------|-----------------|------------|-------------|
| 1  | x           | Cỡ D1,3 (cm) | N (c/ha)   | Nixi       | Px              | Nlt (c/ha) | X2          |
| 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        |
| 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 | <b>Tổng</b> |              | <b>349</b> | <b>557</b> | <b>0,995258</b> | <b>347</b> | <b>4,06</b> |
|    |             |              | xbq=       | 1,595989   |                 | X2 bảng=   | 12,59       |
|    |             |              | Alpha=     | 0,61479    |                 | K=8-1-1=6  |             |

### 10.3. Mô phỏng phân bố thực nghiệm theo phân bố Weibull:

Phân bố Weibull là phân bố xác suất của biến ngẫu nhiên liên tục với miền giá trị  $x \in (0, +\infty)$ .

Hàm mật độ:

$$f(x) = \alpha \cdot \lambda (x - x_{\min})^{\alpha-1} \cdot \exp(-\lambda (x - x_{\min})^\alpha)$$

Hàm phân bố:

$$F(x) = 1 - \exp(-\lambda (x - x_{\min})^\alpha)$$

Với  $x_{\min}$ : 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:

$\alpha \leq 1$ : 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 nhỏ thua  $\chi^2$  bảng.

Tham số  $\lambda$  được ước lượng bằng phương pháp cực đại hợp lý:

$$\lambda = \hat{a} / \sum_{i=1}^r \hat{a}_i \cdot (x_i - x_{\min})^\alpha$$

$\hat{a}$  : Tổng dung lượng quan sát.

$\hat{a}_i$ : Tần số tổ  $i$ .

□ **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 Nlt cho từng tổ:**

$$\hat{a} \cdot l_i = \hat{a} \cdot P(x_i).$$

\* **Kiểm tra sự phù hợp bằng tiêu chuẩn  $\chi^2$ .**

#### Kết quả mô phỏng phân bố N/D theo hàm Weibull

|           | A                   | B               | C            | D                      | E            | F           | G                 | H           |
|-----------|---------------------|-----------------|--------------|------------------------|--------------|-------------|-------------------|-------------|
| <b>1</b>  | <b>Cỡ D1,3 (cm)</b> | <b>N (c/ha)</b> | <b>Alpha</b> | <b>N(x-xmin)^alpha</b> | <b>Lamda</b> | <b>P(x)</b> | <b>Nlt (c/ha)</b> | <b>X2</b>   |
| 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        |
| 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                 |             |
| <b>13</b> | <b>Tổng</b>         | <b>349</b>      |              | <b>7315,0</b>          |              | <b>1,0</b>  | <b>347</b>        | <b>3,76</b> |
| 14        |                     |                 |              |                        |              |             | X2 bảng=          | 14,07       |
| 15        |                     |                 |              |                        |              |             | K=9-1-1=7         |             |

- Cột A: Giá trị giữa cỡ kính 15, 25,...115 với cự ly cỡ 10 cm.
- Cột B: Số cây từng cỡ  $\hat{a}_i$ . Ô B13: tổng  $\hat{a} = \text{Sum}(b2:b12)$
- Ô C2: Đưa tham số  $\alpha$  thăm dò.
- Cột D: Giá trị:  $\hat{a}_i(x_i - 10)^\alpha$ . Với  $x_{\min}=10$ . Tính tại ô d2:  $=B2*(A2-10)^\alpha$ , sau đó copy cho các ô dưới. Ô D13 tính tổng  $=\text{Sum}(d2:d12)$ .

- Ô E2: Tính tham số  $\lambda$ : = B13/Sum(d2:d12).
- Cột F: Tính xác suất P(x) từng ô: Tính theo công thức địa chỉ ô.
- Cột G: ả lt từng ô: Ô G2: =\$B\$13\*F2, sau đó copy xuống và tính tổng.
- Cột H: Tính  $\chi^2$  từng ô và tổng  $\chi^2=3.76$
- Ô H14: Tra  $\chi^2(0.05,7)=\text{Chiinv}(0.05,7)=14.07$
- KL: Phân bố Weibull mô phỏng tốt phân bố thực nghiệm.

Chú ý: Để chọn được  $\alpha$  tối ưu, lần lượt thay giá trị ở ô C2, bảng tính sẽ tự động tính lại, sau đó chọn một  $\alpha$  với  $\chi^2$  bé nhất.

## 11. PHƯƠNG PHÁP NGHIÊN CỨU CẤU TRÚC MẶT BẰNG RỪNG (Bảo Huy, 1993)

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**; 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, tía thưa, khai thác để điều tiết mật độ trên bề mặt đất rừng.

Phương pháp áp dụng là phân bố khoảng cách từ một cây chọn ngẫu nhiên đến cây gần nhất, với dung lượng mẫu  $n > 30$  (số khoảng cách đo) tính theo tiêu chuẩn U

$$U = \frac{(\bar{x} \sqrt{\lambda} - 0,5) \times \sqrt{n}}{0,26136}$$

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

Ấu :  $|U| \leq 1,96$  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: Đánh giá phân bố cây trên mặt đất rừng:

Với mẫu quan sát  $n = 285$  (số khoảng cách đo) trên 5 ô điều tra nên 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; Tính toán được giá trị :  $U = 10,667$

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 ả /D 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 x: cự ly bình quân tối ưu để cây rừng có phân bố đồng đều trên mặt đất rừng chung và theo cấp kính.

Từ công thức tính U, giả định rừng có phân bố đều thì U=2, và rừng đạt mật độ chuẩn ẩ opt (số cây tối ưu/ha tính được từ mô hình ẩ /D định hướng đã xây dựng), suy ra cự ly giữa cây rừng tối ưu:

$$x = \frac{\left( \frac{2 \times 0,26136}{\sqrt{N_{opt}}} + 0,5 \right)}{\sqrt{\lambda}}$$

Với  $\lambda = \text{ẩ opt}/10.000$

Ví dụ: Với ẩ opt = 889 cây/ha (số cây mẫu từ phân bố ẩ /D đã xây dựng)

Suy ra :  $\lambda = \text{ẩ opt} / 10.000 = 0.0889$

Từ đây tính được cự ly bình quân tối ưu giữa 2 cây gần nhất trong lâm phần :  $x = 1.73\text{m}$

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 ẩ opti nào đó theo cấu trúc ẩ /D, 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 (xi), phương pháp xác định như trên, nhưng chỉ khác là mật độ tối ưu ở đây được tính theo từng cỡ kính (ẩ opti) theo mô hình ẩ /D định hướng, suy ra tham số  $\lambda_i = \text{ẩ opti}/10.000$ . Từ đây ta tính được cự ly bình quân tối ưu (xi) cho từng cỡ kính như ví dụ ở bảng sau

**Mô hình cự ly bình quân tối ưu giữa các cây rừng theo cỡ kính**

| Cấp kính $D_{1.3}$<br>(cm) | $N_{opt}$ (c/ha) | $\lambda$ | Cự ly<br>bình quân<br>tối ưu<br>$X_i$ (m) | U |
|----------------------------|------------------|-----------|---|---|
| 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 |
| <b>Tổng</b>                | <b>889</b>       |           |   |   |

Qua bảng trên cho thấy đối với cấp kính càng lớn thì cự ly tối ưu để rừng có phân bố đều giữa 2 cây 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 tía thưa, nuôi dưỡng, khai



thác chọn rừng, với định hướng đưa rừng về ả opti và có cự ly bảo đảm cây rừng có phân bố đều, lợi dụng tốt nhất không gian dinh dưỡng.

## 12. PHÂN TÍCH, PHÁT HIỆN CÁC NGUYÊN NHÂN, NHÂN TỐ ĐỊNH TÍNH, ĐỊNH LƯỢNG ẢNH HƯỞNG ĐẾN BIẾN PHỤ THUỘC (HẬU QUẢ, VẤN ĐỀ) (Bảo Huy, 2006)

Trong thực tế chúng ta cần phát hiện các nhân tố chủ đạo ảnh hưởng đến một vấn đề, hậu quả. Ví dụ các nhân tố nào ảnh hưởng đến mức độ xung yếu của lưu vực, từ đây giúp cho việc quy hoạch lưu vực; hoặc tìm kiếm các nhân tố chủ đạo ảnh hưởng đến sinh trưởng sản lượng của một loài cây trồng, làm cơ sở quy hoạch, chọn vùng trồng thích hợp, ....

Mô hình hồi quy đa biến dạng tuyến tính hoặc phi tuyến hoặc tổ hợp biến sẽ là một công cụ mạnh giúp cho việc phát hiện các nhân tố ảnh hưởng rõ rệt cả về tự nhiên lẫn nhân tố xã hội.

Trong trường hợp nhiều biến số xi ảnh hưởng đến y không theo dạng tuyến tính mà có dạng quan hệ phi tuyến, trường hợp này cần đổi biến số để trở về dạng tuyến tính, hoặc lập mô hình tổ hợp biến.

Một số dạng tuyến tính, phi tuyến nhiều lớp phổ biến và cách quy về tuyến tính hoặc tổ hợp biến:

$$y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

$$y = a \cdot x_1^{b_1} \cdot x_2^{b_2} \dots x_n^{b_n} \text{ tuyến tính hóa: } \ln(y) = \ln(a) + b_1 \ln(x_1) + b_2 \ln(x_2) + \dots + b_n \ln(x_n)$$

$$y = a \cdot e^{b_1x_1 + b_2x_2 + \dots + b_nx_n} \text{ tuyến tính hóa: } \ln(y) = \ln(a) + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

Hoặc dạng tổ hợp biến và đổi biến số kết hợp:

$$\ln(y) = a + b_1 \cdot \log(x_1 \cdot x_2) + b_2 \exp(x_3/x_4) + \dots$$

Trong Statgraphics, việc tính toán mô hình kiểu này rất đơn giản vì không cần tạo thêm các cột đổi biến số, biến số được đổi trực tiếp trong hộp thoại khi thiết lập mô hình.

Ví dụ nghiên cứu phát hiện các nhân tố sinh thái, nhân tác ảnh hưởng đến sự thay đổi mật độ tái sinh rừng khộp thông qua mã hóa và thiết lập mô hình hồi quy đa biến:

Cách tiến hành bao gồm:

**Các bước tiến hành như sau:**

- i) Thu thập dữ liệu về biến số phụ thuộc y là mật độ tái sinh (Nts), cùng với nó là các nhân tố sinh thái, nhân tác ảnh hưởng (có thể định tính hay định lượng)
- ii) Mã hóa các biến định tính
- iii) Kiểm tra dạng chuẩn của mỗi biến số, nếu chưa chuẩn phải đổi biến số để đưa về chuẩn ( $\log(x)$ ,  $1/x$ ,  $\sqrt{x}$ ,  $\exp(x)$ , ...)
- iv) Chọn biến số xi có ảnh hưởng đến y
- v) Chạy mô hình tuyến tính nhiều lớp hoặc được đổi biến số, khi cần thiết phải tổ hợp biến nếu các biến xi có quan hệ với nhau. Kiểm tra mô hình: Hệ số xác định  $R^2$  có  $P < 0.05$  và các tham số gắn biến số qua kiểm tra theo t phải có  $P < 0.05$ . Nếu một biến số chưa bảo đảm  $P < 0.05$  thì phải loại khỏi mô hình hoặc đổi biến số, hoặc tổ hợp với biến số khác.
- vi) Phân tích kết quả mô hình hồi quy đa biến để đánh giá chiều hướng tác động của các biến số sinh thái, nhân tác đến tài nguyên, loài làm cơ sở quy hoạch, phối trí không gian cảnh quan.

**Thu thập số liệu và mã hóa các biến số định tính:**

Các nhân tố tác động, ảnh hưởng có thể là nhân tố định lượng được như: lượng mưa, độ dốc, độ cao, .... hoặc là định tính như vị trí địa hình, mức độ tác động, mức độ lửa rừng, nhân tác, .... Trong trường hợp nhân tố là định tính thì cần phải mã hóa để có thể thiết lập mô hình hồi quy.

Có hai phương án mã hóa:

- i. **Mã hóa hệ thống:** Các mức độ, cấp của của nhân tố được mã hóa hệ thống 1, 2, 3, .... Ví dụ mã hóa nhân tố vị trí địa hình: Bằng = 1; chân = 2; sườn = 3 và đỉnh = 4
- ii. **Mã hóa theo chiều biến thiên:** Các mức độ, cấp được mã hóa theo chiều biến thiên của nhân tố phụ thuộc. Sắp xếp nhân tố phụ thuộc theo một chiều nào đó (tăng hoặc giảm), sau đó các nhân tố được mã hóa theo cùng một vector như vậy.

Ví dụ: Mã hóa biến số theo chiều biến thiên sinh trưởng

|                                   |      |      |      |      |      |
|-----------------------------------|------|------|------|------|------|
| <b>Sinh trưởng cây rừng H (m)</b> | 10   | 15   | 20   | 25   | 30   |
| <b>Vị trí địa hình</b>            | Đỉnh | Đỉnh | Chân | Sườn | Bằng |
| <b>Mã số vị trí</b>               | 1    | 2    | 3    | 4    | 5    |

Và việc mã hóa khác nhau sẽ dẫn đến việc lựa chọn mô hình hồi quy có mức độ phức tạp khác nhau

Hai phương án mã hóa biến định tính khác nhau sẽ dẫn đến việc chọn lựa mô hình hồi quy khác nhau

| <b>Kiểu dạng hàm mô phỏng</b>   | <b>Phương pháp mã hóa biến định tính</b> |   |
|---|--|---|
|   | <b>Hệ thống (Mã hóa đơn giản)</b>        | <b>Theo chiều biến thiên, vector của biến phụ thuộc (Mã hóa phức tạp)</b> |
| Tuyến tính hoặc phi tuyến nhưng theo 1 chiều (tăng hoặc giảm)<br><b>(Xây dựng hàm đơn giản)</b> | Không thực hiện được hoặc sai quy luật   | <b>Thực hiện được</b>   |
| Phi tuyến dạng tăng giảm phức tạp, hoặc tổ hợp biến<br><b>(Xây dựng hàm phức tạp)</b>           | <b>Thực hiện được</b>                    | Thực hiện được nhưng không cần thiết                                      |

Bảng mã hoá các nhân tố liên quan đến ả ts rừng khộp

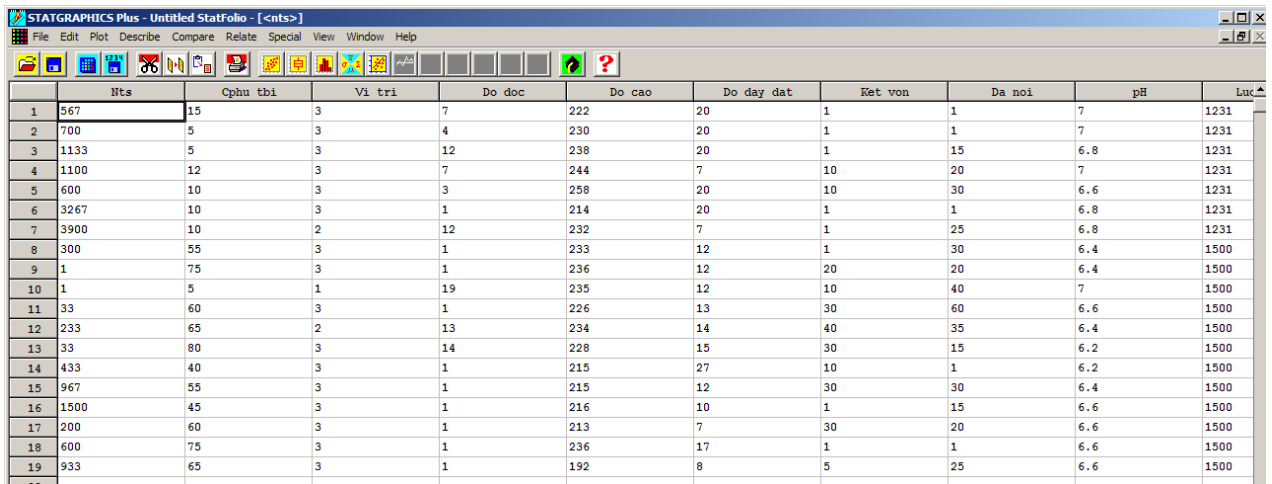
| <b>Nhân tố</b>         | <b>Ký hiệu</b>   | <b>Phân cấp và mã hoá các biến số theo cấp</b> |                  |                  |
|------------------------|------------------|--|------------------|------------------|
|                        |                  | <b>Cấp 1 = 1</b>                               | <b>Cấp 2 = 2</b> | <b>Cấp 3 = 3</b> |
| Độ che phủ thực bì (%) | Cphu tbi         | Không phân cấp, lấy theo đo đếm                |                  |                  |
| Vị trí địa hình        | Vi tri           | Đỉnh   | Sườn             | Chân             |
| Độ dốc (Độ)            | Do doc           | Không phân cấp, lấy theo đo đếm                |                  |                  |
| Độ cao (m)             | Do cao           | Không phân cấp, lấy theo đo đếm                |                  |                  |
| Độ dày đất (Cm)        | Do day dat       | Không phân cấp, lấy theo đo đếm                |                  |                  |
| Kết von (%)            | Ket von          | Không phân cấp, lấy theo đo đếm                |                  |                  |
| Đá nôi (%)             | Da noi           | Không phân cấp, lấy theo đo đếm                |                  |                  |
| pH đất                 | pH               | Không phân cấp, lấy theo đo đếm                |                  |                  |
| Lượng mưa (mm)         | Luong mua        | Không phân cấp, lấy theo đo đếm                |                  |                  |
| Khai thác              | Muc do khai thac | Sau nương rẫy                                  | Chặt chọn        | Không có         |
| Lửa rừng               | Muc do lua rung  | Hàng năm                                       | Thỉnh thoảng     | Không có         |

Biểu tổng hợp mã hóa các nhân tố và ảnh hưởng khộp

| Nts  | Cphu tbi | Vi tri | Do doc | Do cao | Do day dat | Ket von | Da noi | pH  | Luong mua | Muc do khai thac | Muc do Lua rung |
|------|----------|--------|--------|--------|------------|---------|--------|-----|-----------|------------------|-----------------|
| 567  | 15       | 3      | 7      | 222    | 20         | 1       | 1      | 7   | 1231      | 1                | 2               |
| 700  | 5        | 3      | 4      | 230    | 20         | 1       | 1      | 7   | 1231      | 2                | 2               |
| 1133 | 5        | 3      | 12     | 238    | 20         | 1       | 15     | 6.8 | 1231      | 2                | 2               |
| 1100 | 12       | 3      | 7      | 244    | 7          | 10      | 20     | 7   | 1231      | 2                | 2               |
| 600  | 10       | 3      | 3      | 258    | 20         | 10      | 30     | 6.6 | 1231      | 3                | 3               |
| 3267 | 10       | 3      | 1      | 214    | 20         | 1       | 1      | 6.8 | 1231      | 2                | 2               |
| 3900 | 10       | 2      | 12     | 232    | 7          | 1       | 25     | 6.8 | 1231      | 2                | 2               |
| 300  | 55       | 3      | 1      | 233    | 12         | 1       | 30     | 6.4 | 1500      | 3                | 2               |
| 1    | 75       | 3      | 1      | 236    | 12         | 20      | 20     | 6.4 | 1500      | 3                | 2               |
| 1    | 5        | 1      | 19     | 235    | 12         | 10      | 40     | 7   | 1500      | 3                | 3               |
| 33   | 60       | 3      | 1      | 226    | 13         | 30      | 60     | 6.6 | 1500      | 3                | 3               |
| 233  | 65       | 2      | 13     | 234    | 14         | 40      | 35     | 6.4 | 1500      | 3                | 1               |
| 33   | 80       | 3      | 14     | 228    | 15         | 30      | 15     | 6.2 | 1500      | 3                | 1               |
| 433  | 40       | 3      | 1      | 215    | 27         | 10      | 1      | 6.2 | 1500      | 3                | 3               |
| 967  | 55       | 3      | 1      | 215    | 12         | 30      | 30     | 6.4 | 1500      | 3                | 2               |
| 1500 | 45       | 3      | 1      | 216    | 10         | 1       | 15     | 6.6 | 1500      | 3                | 2               |
| 200  | 60       | 3      | 1      | 213    | 7          | 30      | 20     | 6.6 | 1500      | 3                | 2               |
| 600  | 75       | 3      | 1      | 236    | 17         | 1       | 1      | 6.6 | 1500      | 3                | 2               |
| 933  | 65       | 3      | 1      | 192    | 8          | 5       | 25     | 6.6 | 1500      | 3                | 2               |

Từ file dữ liệu mã hóa trong Excel, tiến hành nhập vào Statgraphics Plus để phát hiện các nhân tố ảnh hưởng và xây dựng mô hình hồi quy dự báo ảnh hưởng.  
 Dữ liệu được nhập vào Statgraphics Plus như sau: File/Open/Open Data file .... Chọn thư mục và kiểu file Excel để mở file trong Statgraphics Plus.

Cơ sở dữ liệu từ Excel được nhập vào trong Statgraphics Plus



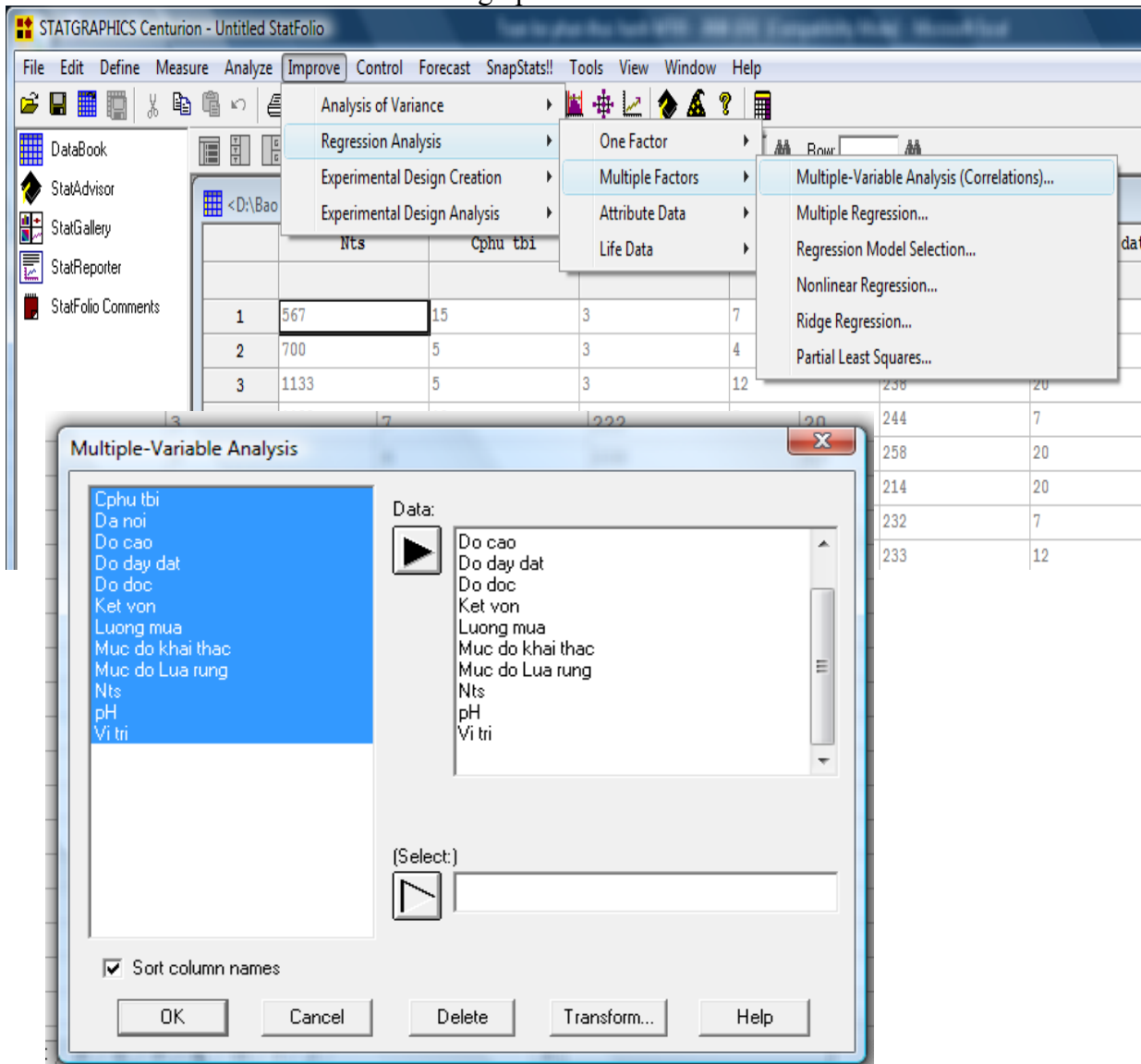
**Kiểm tra dạng chuẩn của mỗi biến số:**

Mỗi biến số tham gia vào mô hình nhiều biến phải bảo đảm phân bố chuẩn, một biến số có phân bố chuẩn khi đủ mẫu quan sát trong môi tương tác với các biến số khác, trong một số trường hợp tuy mẫu đủ lớn, nhưng khi tương tác vẫn chưa chuẩn, do đó cần đổi biến số để chuẩn hóa.

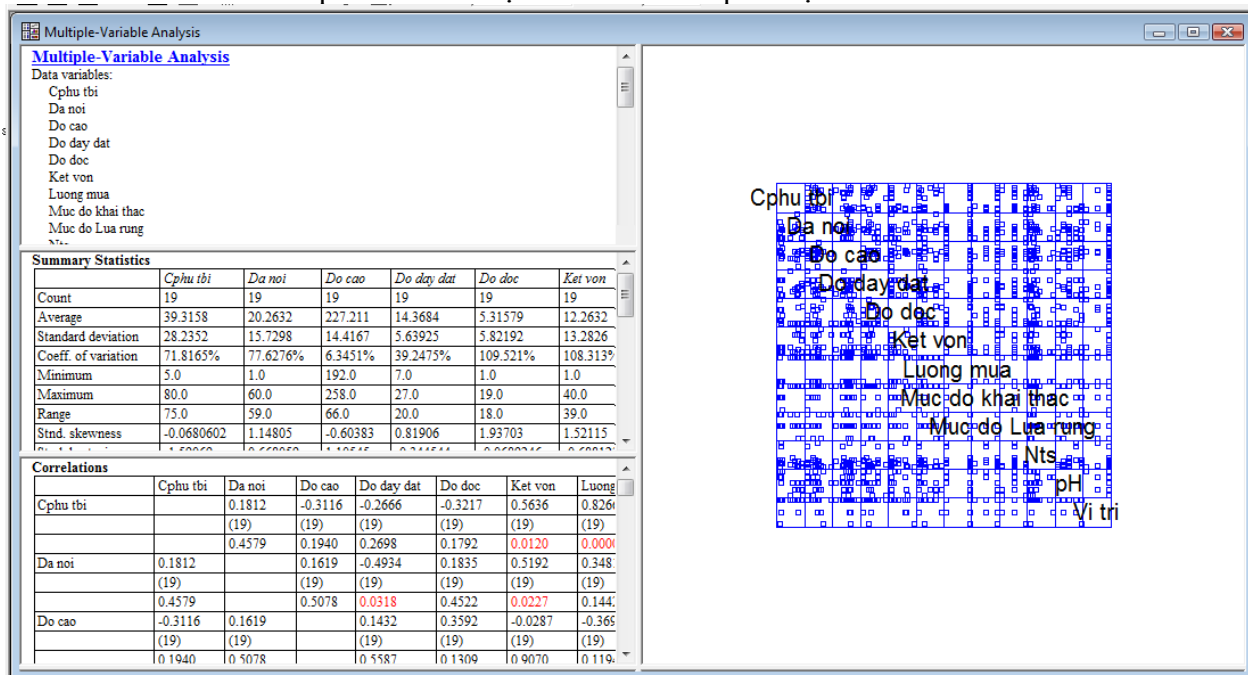
Kiểm tra dạng chuẩn của các biến số trong Statgraphics và định hướng đổi biến số:

Improve/Regression Analysis/Multiple Factors/Multiple Variable Analysis. Sau đó đưa tất cả biến y và xi vào hộp thoại data.

Chọn chương trình kiểm tra luật chuẩn và định hướng đổi biến số để chuẩn hóa trong Statgraphics Centurion



## Kết quả kiểm tra luật chuẩn và mối quan hệ các biến số



### - Kết quả kiểm tra phân bố chuẩn của các biến số:

#### - Summary Statistics

|                     | <i>Cphu tbi</i> | <i>Da noi</i> | <i>Do cao</i> | <i>Do day dat</i> | <i>Do doc</i> | <i>Ket von</i> | <i>Luong mua</i> | <i>Muc do khai thac</i> |
|---------------------|-----------------|---------------|---------------|-------------------|---------------|----------------|------------------|-------------------------|
| Count               | 19              | 19            | 19            | 19                | 19            | 19             | 19               | 19                      |
| Average             | 39.3158         | 20.2632       | 227.211       | 14.3684           | 5.31579       | 12.2632        | 1400.89          | 2.63158                 |
| Standard deviation  | 28.2352         | 15.7298       | 14.4167       | 5.63925           | 5.82192       | 13.2826        | 133.315          | 0.597265                |
| Coeff. of variation | 71.8165%        | 77.6276%      | 6.3451%       | 39.2475%          | 109.521%      | 108.313%       | 9.51641%         | 22.6961%                |
| Minimum             | 5.0             | 1.0           | 192.0         | 7.0               | 1.0           | 1.0            | 1231.0           | 1.0                     |
| Maximum             | 80.0            | 60.0          | 258.0         | 27.0              | 19.0          | 40.0           | 1500.0           | 3.0                     |
| Range               | 75.0            | 59.0          | 66.0          | 20.0              | 18.0          | 39.0           | 269.0            | 2.0                     |
| Std. skewness       | -0.0680602      | 1.14805       | -0.60383      | 0.81906           | 1.93703       | 1.52115        | -1.05608         | <b>-2.56858</b>         |
| Std. kurtosis       | -1.59069        | 0.668059      | 1.10545       | -0.344544         | -0.0689246    | -0.688123      | -1.65147         | 1.22788                 |

|                     | <i>Muc do Lua rung</i> | <i>Nts</i>     | <i>pH</i> | <i>Vi tri</i>   |
|---------------------|------------------------|----------------|-----------|-----------------|
| Count               | 19                     | 19             | 19        | 19              |
| Average             | 2.10526                | 868.474        | 6.63158   | 2.78947         |
| Standard deviation  | 0.567131               | 1054.29        | 0.260454  | 0.535303        |
| Coeff. of variation | 26.9387%               | 121.395%       | 3.92748%  | 19.1901%        |
| Minimum             | 1.0                    | 1.0            | 6.2       | 1.0             |
| Maximum             | 3.0                    | 3900.0         | 7.0       | 3.0             |
| Range               | 2.0                    | 3899.0         | 0.8       | 2.0             |
| Std. skewness       | 0.0906087              | <b>3.63749</b> | 0.0232827 | <b>-4.72906</b> |
| Std. kurtosis       | 0.52516                | <b>3.5476</b>  | -0.823423 | <b>6.1244</b>   |

#### The StatAdvisor

This table shows summary statistics for each of the selected data variables. 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 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 many of the statistical procedures normally applied to this data. In this case, the following variables show standardized skewness values outside the expected range:

Muc do khai thac

â ts

Vi tri

The following variables show standardized kurtosis values outside the expected range:

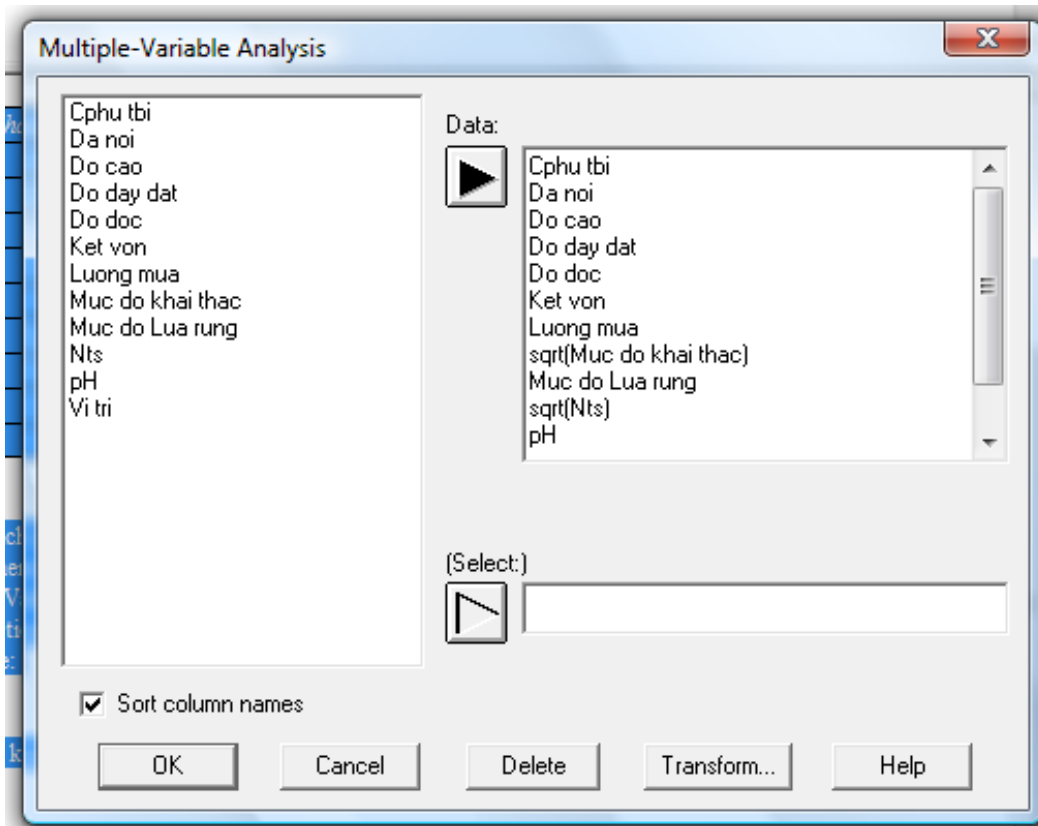
â ts

Vi tri

To make the variables more normal, you might try a transformation such as LOG(Y), SQRT(Y), or 1/Y.

Kết quả cho thấy có 3 biến số có *Standardized Sk hoặc Ku* không bảo đảm có phân bố chuẩn là: *Nts, Muc do khai thác và Vi tri*. Và 3 biến này cần đổi biến số ở các dạng LOG(Y), SQRT(Y), or 1/Y để chuẩn hóa.

### Đổi biến số để chuẩn hóa



#### Summary Statistics

|                     | <i>Cphu tbi</i> | <i>Da noi</i> | <i>Do cao</i> | <i>Do day dat</i> | <i>Do doc</i> | <i>Ket von</i> | <i>Luong mua</i> |
|---------------------|-----------------|---------------|---------------|-------------------|---------------|----------------|------------------|
| Count               | 19              | 19            | 19            | 19                | 19            | 19             | 19               |
| Average             | 39.3158         | 20.2632       | 227.211       | 14.3684           | 5.31579       | 12.2632        | 1400.89          |
| Standard deviation  | 28.2352         | 15.7298       | 14.4167       | 5.63925           | 5.82192       | 13.2826        | 133.315          |
| Coeff. of variation | 71.8165%        | 77.6276%      | 6.3451%       | 39.2475%          | 109.521%      | 108.313%       | 9.51641%         |
| Minimum             | 5.0             | 1.0           | 192.0         | 7.0               | 1.0           | 1.0            | 1231.0           |
| Maximum             | 80.0            | 60.0          | 258.0         | 27.0              | 19.0          | 40.0           | 1500.0           |
| Range               | 75.0            | 59.0          | 66.0          | 20.0              | 18.0          | 39.0           | 269.0            |
| Std. skewness       | -0.0680602      | 1.14805       | -0.60383      | 0.81906           | 1.93703       | 1.52115        | -1.05608         |
| Std. kurtosis       | -1.59069        | 0.668059      | 1.10545       | -0.344544         | -0.0689246    | -0.688123      | -1.65147         |

|                     | <i>sqrt(Muc do khai thác)</i> | <i>Muc do Lua rung</i> | <i>sqrt(Nts)</i> | <i>pH</i> | <i>log(Vi tri)</i> |
|---------------------|-------------------------------|------------------------|------------------|-----------|--------------------|
| Count               | 19                            | 19                     | 19               | 19        | 19                 |
| Average             | 1.60988                       | 2.10526                | 24.5836          | 6.63158   | 0.99811            |
| Standard deviation  | 0.205131                      | 0.567131               | 16.697           | 0.260454  | 0.273236           |
| Coeff. of variation | 12.742%                       | 26.9387%               | 67.9193%         | 3.92748%  | 27.3753%           |
| Minimum             | 1.0                           | 1.0                    | 1.0              | 6.2       | 0.0                |
| Maximum             | 1.73205                       | 3.0                    | 62.45            | 7.0       | 1.09861            |
| Range               | 0.732051                      | 2.0                    | 61.45            | 0.8       | 1.09861            |
| Std. skewness       | <b>-3.07989</b>               | 0.0906087              | 1.22414          | 0.0232827 | <b>-5.60515</b>    |
| Std. kurtosis       | <b>2.6152</b>                 | 0.52516                | 0.490076         | -0.823423 | <b>9.35136</b>     |

#### The StatAdvisor

This table shows summary statistics for each of the selected data variables. 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 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 many of the statistical procedures normally applied to this data. In this case, the following variables show standardized skewness values outside the expected range:

sqrt(Muc do khai thac)  
log(Vi tri)

The following variables show standardized kurtosis values outside the expected range:

sqrt(Muc do khai thac)  
log(Vi tri)

To make the variables more normal, you might try a transformation such as LOG(Y), SQRT(Y), or 1/Y.

Ví dụ sau khi thử đổi biến số thì biến sqrt(ẩ ts) bảo đảm luật chuẩn, trong khi đó thì 2 biến Muc do khai thac và Vi trí vẫn chưa thỏa mãn; nếu tiếp tục đổi biến số mà cũng không bảo đảm thì có 2 phương án: i) Đổi biến số theo kiểu khác; ii) Thu thập thêm dữ liệu để bảo đảm chuẩn;

### Chọn biến số xi 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 (ẩ ts)

#### Correlations

|                        | Cphu tbi        | Da noi          | Do cao          | Do day dat      | Do doc          | Ket von         | Luong mua       | sqrt(Muc do khai thac) |
|------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|------------------------|
| Cphu tbi               |                 | 0.1812<br>(19)  | -0.3116<br>(19) | -0.2666<br>(19) | -0.3217<br>(19) | 0.5636<br>(19)  | 0.8266<br>(19)  | 0.6420<br>(19)         |
|                        |                 | 0.4579          | 0.1940          | 0.2698          | 0.1792          | <b>0.0120</b>   | <b>0.0000</b>   | <b>0.0030</b>          |
| Da noi                 | 0.1812<br>(19)  |                 | 0.1619<br>(19)  | -0.4934<br>(19) | 0.1835<br>(19)  | 0.5192<br>(19)  | 0.3481<br>(19)  | 0.4579<br>(19)         |
|                        | 0.4579          |                 | 0.5078          | <b>0.0318</b>   | 0.4522          | <b>0.0227</b>   | 0.1442          | <b>0.0486</b>          |
| Do cao                 | -0.3116<br>(19) | 0.1619<br>(19)  |                 | 0.1432<br>(19)  | 0.3592<br>(19)  | -0.0287<br>(19) | -0.3695<br>(19) | -0.0594<br>(19)        |
|                        | 0.1940          | 0.5078          |                 | 0.5587          | 0.1309          | 0.9070          | 0.1194          | 0.8092                 |
| Do day dat             | -0.2666<br>(19) | -0.4934<br>(19) | 0.1432<br>(19)  |                 | -0.0680<br>(19) | -0.2313<br>(19) | -0.2668<br>(19) | -0.2309<br>(19)        |
|                        | 0.2698          | <b>0.0318</b>   | 0.5587          |                 | 0.7820          | 0.3407          | 0.2695          | 0.3415                 |
| Do doc                 | -0.3217<br>(19) | 0.1835<br>(19)  | 0.3592<br>(19)  | -0.0680<br>(19) |                 | 0.1117<br>(19)  | -0.1692<br>(19) | -0.1966<br>(19)        |
|                        | 0.1792          | 0.4522          | 0.1309          | 0.7820          |                 | 0.6490          | 0.4885          | 0.4197                 |
| Ket von                | 0.5636<br>(19)  | 0.5192<br>(19)  | -0.0287<br>(19) | -0.2313<br>(19) | 0.1117<br>(19)  |                 | 0.5135<br>(19)  | 0.4748<br>(19)         |
|                        | <b>0.0120</b>   | <b>0.0227</b>   | 0.9070          | 0.3407          | 0.6490          |                 | <b>0.0245</b>   | <b>0.0400</b>          |
| Luong mua              | 0.8266<br>(19)  | 0.3481<br>(19)  | -0.3695<br>(19) | -0.2668<br>(19) | -0.1692<br>(19) | 0.5135<br>(19)  |                 | 0.8012<br>(19)         |
|                        | <b>0.0000</b>   | 0.1442          | 0.1194          | 0.2695          | 0.4885          | <b>0.0245</b>   |                 | <b>0.0000</b>          |
| sqrt(Muc do khai thac) | 0.6420<br>(19)  | 0.4579<br>(19)  | -0.0594<br>(19) | -0.2309<br>(19) | -0.1966<br>(19) | 0.4748<br>(19)  | 0.8012<br>(19)  |                        |
|                        | <b>0.0030</b>   | <b>0.0486</b>   | 0.8092          | 0.3415          | 0.4197          | <b>0.0400</b>   | <b>0.0000</b>   |                        |
| Muc do Lua rung        | -0.3769<br>(19) | 0.2521<br>(19)  | 0.1194<br>(19)  | 0.2478<br>(19)  | -0.2294<br>(19) | -0.2546<br>(19) | -0.0520<br>(19) | 0.1167<br>(19)         |
|                        | 0.1117          | 0.2979          | 0.6262          | 0.3064          | 0.3449          | 0.2928          | 0.8325          | 0.6343                 |
| sqrt(ẩ ts)             | -0.4810<br>(19) | -0.3686<br>(19) | -0.1715<br>(19) | -0.0247<br>(19) | -0.1215<br>(19) | -0.5421<br>(19) | -0.5983<br>(19) | -0.4547<br>(19)        |
|                        | <b>0.0371</b>   | 0.1204          | 0.4826          | 0.9199          | 0.6203          | <b>0.0165</b>   | <b>0.0068</b>   | 0.0505                 |
| pH                     | -0.7690<br>(19) | -0.1160<br>(19) | 0.1786<br>(19)  | -0.0916<br>(19) | 0.2715<br>(19)  | -0.5164<br>(19) | -0.6796<br>(19) | -0.6910<br>(19)        |
|                        | <b>0.0001</b>   | 0.6361          | 0.4643          | 0.7093          | 0.2608          | <b>0.0236</b>   | <b>0.0014</b>   | <b>0.0011</b>          |
| log(Vi tri)            | 0.2821<br>(19)  | -0.3823<br>(19) | -0.1869<br>(19) | 0.2069<br>(19)  | -0.7285<br>(19) | -0.0642<br>(19) | -0.1223<br>(19) | -0.1035<br>(19)        |
|                        | 0.2420          | 0.1062          | 0.4436          | 0.3953          | <b>0.0004</b>   | 0.7940          | 0.6180          | 0.6733                 |

|          | Muc do Lua rung | sqrt(ẩ ts)      | pH              | log(Vi tri)    |
|----------|-----------------|-----------------|-----------------|----------------|
| Cphu tbi | -0.3769<br>(19) | -0.4810<br>(19) | -0.7690<br>(19) | 0.2821<br>(19) |
|          | 0.1117          | <b>0.0371</b>   | <b>0.0001</b>   | 0.2420         |
| Da noi   | 0.2521          | -0.3686         | -0.1160         | -0.3823        |

|                        |         |         |         |         |
|------------------------|---------|---------|---------|---------|
|                        | (19)    | (19)    | (19)    | (19)    |
|                        | 0.2979  | 0.1204  | 0.6361  | 0.1062  |
| Do cao                 | 0.1194  | -0.1715 | 0.1786  | -0.1869 |
|                        | (19)    | (19)    | (19)    | (19)    |
|                        | 0.6262  | 0.4826  | 0.4643  | 0.4436  |
| Do day dat             | 0.2478  | -0.0247 | -0.0916 | 0.2069  |
|                        | (19)    | (19)    | (19)    | (19)    |
|                        | 0.3064  | 0.9199  | 0.7093  | 0.3953  |
| Do doc                 | -0.2294 | -0.1215 | 0.2715  | -0.7285 |
|                        | (19)    | (19)    | (19)    | (19)    |
|                        | 0.3449  | 0.6203  | 0.2608  | 0.0004  |
| Ket von                | -0.2546 | -0.5421 | -0.5164 | -0.0642 |
|                        | (19)    | (19)    | (19)    | (19)    |
|                        | 0.2928  | 0.0165  | 0.0236  | 0.7940  |
| Luong mua              | -0.0520 | -0.5983 | -0.6796 | -0.1223 |
|                        | (19)    | (19)    | (19)    | (19)    |
|                        | 0.8325  | 0.0068  | 0.0014  | 0.6180  |
| sqrt(Muc do khai thac) | 0.1167  | -0.4547 | -0.6910 | -0.1035 |
|                        | (19)    | (19)    | (19)    | (19)    |
|                        | 0.6343  | 0.0505  | 0.0011  | 0.6733  |
| Muc do Lua rung        |         | -0.1064 | 0.2019  | -0.1764 |
|                        |         | (19)    | (19)    | (19)    |
|                        |         | 0.6648  | 0.4071  | 0.4699  |
| sqrt(â ts)             | -0.1064 |         | 0.3337  | 0.1746  |
|                        | (19)    |         | (19)    | (19)    |
|                        | 0.6648  |         | 0.1627  | 0.4748  |
| pH                     | 0.2019  | 0.3337  |         | -0.2960 |
|                        | (19)    | (19)    |         | (19)    |
|                        | 0.4071  | 0.1627  |         | 0.2186  |
| log(Vi tri)            | -0.1764 | 0.1746  | -0.2960 |         |
|                        | (19)    | (19)    | (19)    |         |
|                        | 0.4699  | 0.4748  | 0.2186  |         |

**Correlation  
(Sample Size)  
P-Value**

**The StatAdvisor**

This table shows Pearson product moment correlations between each pair of variables. These correlation coefficients range between -1 and +1 and measure the strength of the linear relationship between the variables. Also shown in parentheses is the number of pairs of data values used to compute each coefficient. The third number in each location of the table is a P-value which tests the statistical significance of the estimated correlations. P-values below 0.05 indicate statistically significant non-zero correlations at the 95.0% confidence level. The following pairs of variables have P-values below 0.05:

- Cphu tbi and Ket von**
- Cphu tbi and Luong mua**
- Cphu tbi and sqrt(Muc do khai thac)**
- Cphu tbi and sqrt(Nts)**
- Cphu tbi and pH**
- Da noi and Do day dat**
- Da noi and Ket von**
- Da noi and sqrt(Muc do khai thac)**
- Do doc and log(Vi tri)**
- Ket von and Luong mua**
- Ket von and sqrt(Muc do khai thac)**
- Ket von and sqrt(Nts)**
- Ket von and pH**
- Luong mua and sqrt(Muc do khai thac)**
- Luong mua and sqrt(Nts)**
- Luong mua and pH**
- sqrt(Muc do khai thac) and pH**

Từ kết quả này cho thấy â ts bị chi phối bởi 3 nhân tố chính là: Cphu tbi, Kvon, Luong mua. Từ đây thiết lập mô hình quan hệ â ts với 3 biến này để lượng hóa sự ảnh hưởng.

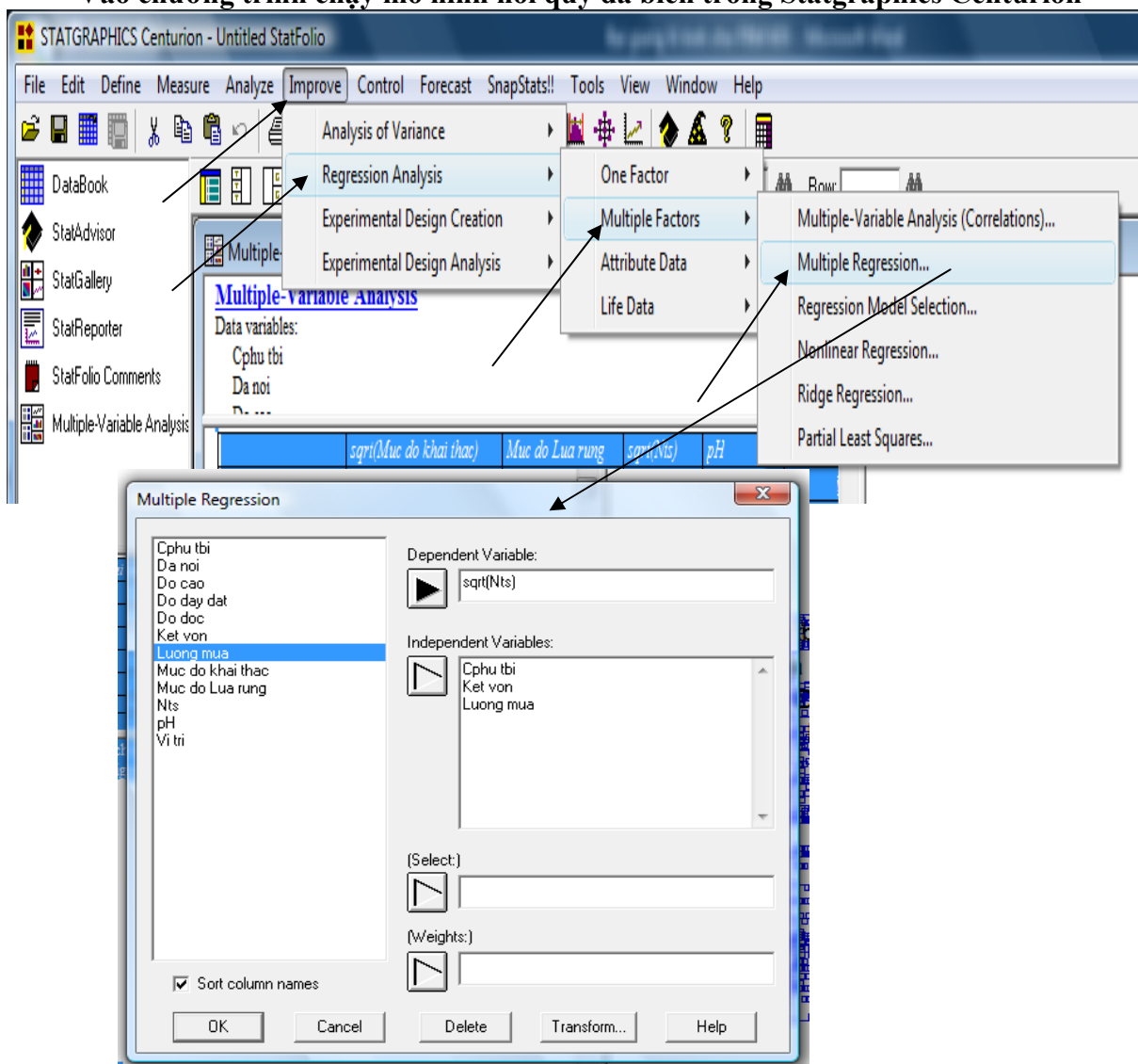


## Xây dựng mô hình đa biến

Các nhân tố sinh thái, nhân tác được đưa vào nghiên cứu ảnh hưởng đến chỉ tiêu ẩ ts: Độ che phủ của thực bì (%), Vị trí địa hình, độ dốc (độ), độ cao so với mặt biển (m), độ dày đất (cm), kết von (%), đá nổi (%), pH đất, lượng mưa (mm), mức độ khai thác, mức độ lửa rừng. Các nhân tố này được kiểm tra có hay không quan hệ ẩ ts bằng mô hình hồi quy đa biến. Giữa các nhân tố sinh thái, nhân tác có sự ảnh hưởng qua lại rất phức tạp và chúng cũng có ảnh hưởng tổng hợp đến các nhân tố ẩ ts, do đó khi kiểm tra mối quan hệ ở mô hình hồi quy đa biến, chấp nhận mức sai là  $<5\% - 10\%$  ( $P < 0.05 - 0.10$ ).

Xây dựng mô hình: *Improve/Regression Analysis/Multiple Factors/Multiple Regression* – Sau đó chọn các biến y, xi vào trong hộp thoại. Lưu ý đổi biến số để chuẩn hóa như đã xác định ở bước trên.

### Vào chương trình chạy mô hình hồi quy đa biến trong Statgraphics Centurion



## Multiple Regression - sqrt(Nts)

Dependent variable: sqrt(â ts)

Independent variables:

- Cphu tbi
- Ket von
- Luong mua

|                  |            | Standard  | T         |               |
|------------------|------------|-----------|-----------|---------------|
| Parameter        | Estimate   | Error     | Statistic | P-Value       |
| COÂ STAÂ T       | 127.22     | 53.9381   | 2.35863   | 0.0323        |
| <b>Cphu tbi</b>  | 0.118008   | 0.21119   | 0.558777  | <b>0.5846</b> |
| <b>Ket von</b>   | -0.4484    | 0.29441   | -1.52305  | <b>0.1485</b> |
| <b>Luong mua</b> | -0.0726513 | 0.0430591 | -1.68725  | <b>0.1122</b> |

### Analysis of Variance

| Source        | Sum of Squares | Df | Mean Square | F-Ratio | P-Value       |
|---------------|----------------|----|-------------|---------|---------------|
| Model         | 2230.26        | 3  | 743.419     | 4.00    | <b>0.0281</b> |
| Residual      | 2787.98        | 15 | 185.866     |         |               |
| Total (Corr.) | 5018.24        | 18 |             |         |               |

R-squared = 44.443 percent

R-squared (adjusted for d.f.) = 33.3316 percent

Standard Error of Est. = 13.6333

Mean absolute error = 10.1868

Durbin-Watson statistic = 1.17117 (P=0.0106)

Lag 1 residual autocorrelation = 0.363982

### The StatAdvisor

The output shows the results of fitting a multiple linear regression model to describe the relationship between sqrt(â ts) and 3 independent variables. The equation of the fitted model is

$$\text{sqrt(Nts)} = 127.22 + 0.118008 * \text{Cphu tbi} - 0.4484 * \text{Ket von} - 0.0726513 * \text{Luong mua}$$

Since the P-value in the AÂ OVA 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 44.443% of the variability in sqrt(â ts). The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 33.3316%. The standard error of the estimate shows the standard deviation of the residuals to be 13.6333. 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 10.1868 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 0.5846, belonging to Cphu tbi. Since the P-value is greater or equal to 0.05, that term is not statistically significant at the 95.0% or higher confidence level. Consequently, you should consider removing Cphu tbi from the model.

Kết quả cho thấy cả 3 biến số đều có Pvalue>0.05; do đó chưa tham gia được vào mô hình; lúc này cần đổi biến số (log, exp, sqrt, 1/xi, ...) hoặc tổ hợp biến để bảo đảm sự tồn tại của biến số đó. ầu một biến nào chưa tìm được cách đổi biến số thích hợp hoặc tổ hợp biến thì cần loại khỏi mô hình, tuy nhiên thực tế biến này có ảnh hưởng đến y, nhưng chưa được phát hiện dạng biến số thích hợp.

**Kết quả thử nghiệm đổi biến số, tổ hợp biến, loại biến số**

**Multiple Regression - sqrt(Nts)**

Dependent variable: sqrt(Nts)

Independent variables:

log(Luong mua\*Ket von)

|                               |                 | Standard       | T               |               |
|-------------------------------|-----------------|----------------|-----------------|---------------|
| Parameter                     | Estimate        | Error          | Statistic       | P-Value       |
| COẢ STAẢ T                    | 83.901          | 18.0012        | 4.66085         | 0.0002        |
| <b>log(Luong mua*Ket von)</b> | <b>-6.68159</b> | <b>1.99815</b> | <b>-3.34389</b> | <b>0.0038</b> |

**Analysis of Variance**

| Source        | Sum of Squares | Df | Mean Square | F-Ratio | P-Value       |
|---------------|----------------|----|-------------|---------|---------------|
| Model         | 1991.09        | 1  | 1991.09     | 11.18   | <b>0.0038</b> |
| Residual      | 3027.15        | 17 | 178.068     |         |               |
| Total (Corr.) | 5018.24        | 18 |             |         |               |

**R-squared = 39.677 percent**

R-squared (adjusted for d.f.) = **36.1286 percent**

Standard Error of Est. = **13.3442**

Mean absolute error = **10.4431**

Durbin-Watson statistic = 1.34835 (P=**0.0522**)

Lag 1 residual autocorrelation = 0.293351

**The StatAdvisor**

The output shows the results of fitting a multiple linear regression model to describe the relationship between sqrt(Nts) and 1 independent variables. The equation of the fitted model is

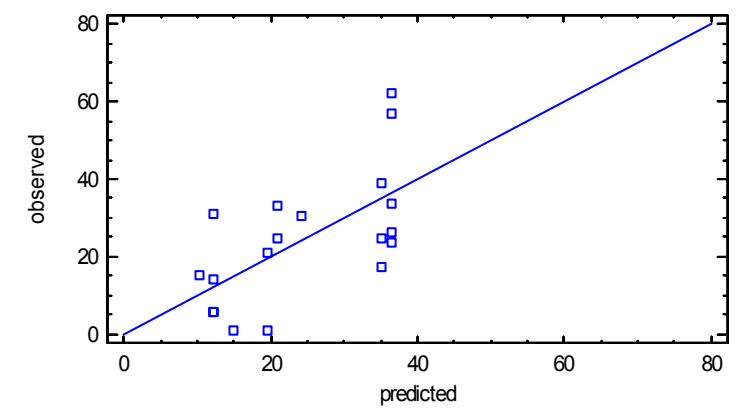
**sqrt(Nts) = 83.901 - 6.68159\*log(Luong mua\*Ket von)**

Since the P-value in the AẢ OVA 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 39.677% of the variability in sqrt(Nts). The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 36.1286%. The standard error of the estimate shows the standard deviation of the residuals to be 13.3442. 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 10.4431 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 0.0038, belonging to log(Luong mua\*Ket von). 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.

Plot of sqrt(Nts)



Kết quả thiết lập được mô hình:

$$\text{sqrt}(Nts) = 83.901 - 6.68159 * \log(\text{Luong mua} * \text{Ket von})$$

Với R-squared = 39.677 percent; Pvalue < 0.05

Các tham số đều tồn tại với Pvalue = 0.0038 < 0.05

### Phân tích kết quả mô hình

Với việc thử nghiệm với nhiều nhân tố ảnh hưởng, kết quả cho thấy mật độ tái sinh cây họ dầu giảm khi ở vùng có lượng mưa tăng, có nghĩa có nhiều loài cây thường xanh sẽ xen vào trong quần thể, lúc này rừng chuyển dần sang dạng chuyên tiếp; và kết von gia tăng (liên quan đến lửa rừng, xói mòn, độ chua tăng) làm khả năng tái sinh cây họ dầu suy giảm. Kết quả này sẽ hỗ trợ cho giải pháp lâm sinh xúc tiến tái sinh rừng khộp nghèo kiệt.

Một ví dụ khác về nghiên cứu phát hiện các nhân tố chủ đạo ảnh hưởng đến mức độ xói mòn đất và dòng chảy mặt trong một lưu vực ở Vườn quốc gia Bi Đúp ở tỉnh Lâm Đồng. Trên thực địa thu nhập nhân tố phụ thuộc y là xói mòn, dòng chảy mặt và các nhân tố ảnh hưởng như thảm thực vật rừng, đất đai, địa hình, lượng mưa, nhân tác. Sau đó mã hóa các nhân tố định tính theo phương pháp hệ thống.

### Mẫu phiếu nghiên cứu các nhân tố tác động đến môi trường rừng (Xói mòn đất, dòng chảy trong lưu vực)

Điểm khảo sát: Lô: ..... Địa điểm: .....

Tọa độ UTM: X: ..... Y: .....

Ngày khảo sát: ..... Người khảo sát: .....

| Các nhân tố   | Giá trị, hoặc mô tả | Giải thích   |
|---|---------------------|--|
| <b>Nhân tố phụ thuộc, bị tác động (y)</b><br>- Xói mòn đất<br>- Dòng chảy mặt |                     | Mã hóa từ 1 – 5:<br>1: An toàn<br>2: Ít nguy cơ<br>3: Nguy cơ trung bình<br>4: Nguy cơ<br>5: Nguy cơ cao |
| <b>Các nhân tố ảnh hưởng (xi)</b>   |                     |  |
| <b>1. Nhóm nhân tố thảm thực vật</b>  |                     |  |
| - Kiểu rừng   |                     | TX: 1, HG: 2; Thông: 3; Le tre: 4  |
| - Trạng thái  |                     | Giàu (TT): 1; TB (TN): 2; Nghèo (Sào): 3, Non: 4; Tcô: 5   |
| - Ưu hợp  |                     | Tên 2-3 loài cây gỗ ưu thế. Mã số ngẫu nhiên: 1,2,3,4,   |
| - Độ tàn che (1/10)   |                     |  |
| - Tổng G (m <sup>2</sup> /ha)   |                     | Dùng thước Bitterlich  |
| - Cấu trúc tầng tán   |                     | Số tầng rừng (1- 5 tầng)   |
| - Mức độ đồng đều của cây rừng trên mặt đất                                   |                     | 1: Đồng đều<br>2: Ngẫu nhiên<br>3: Cụm   |
| - Loài le tre   |                     | Mã số ngẫu nhiên   |
| - % che phủ của le tre  |                     |  |
| - Loài thảm thực bì   |                     | Tên 1-2 loài chính<br>Mã số ngẫu nhiên   |
| - % che phủ của thực bì   |                     |  |
| <b>2. Nhóm nhân tố địa hình</b>   |                     |  |
| - Độ cao so với mặt biển (m)  |                     | Dùng GPS   |
| - Vị trí  |                     | 0: Khe suối  |

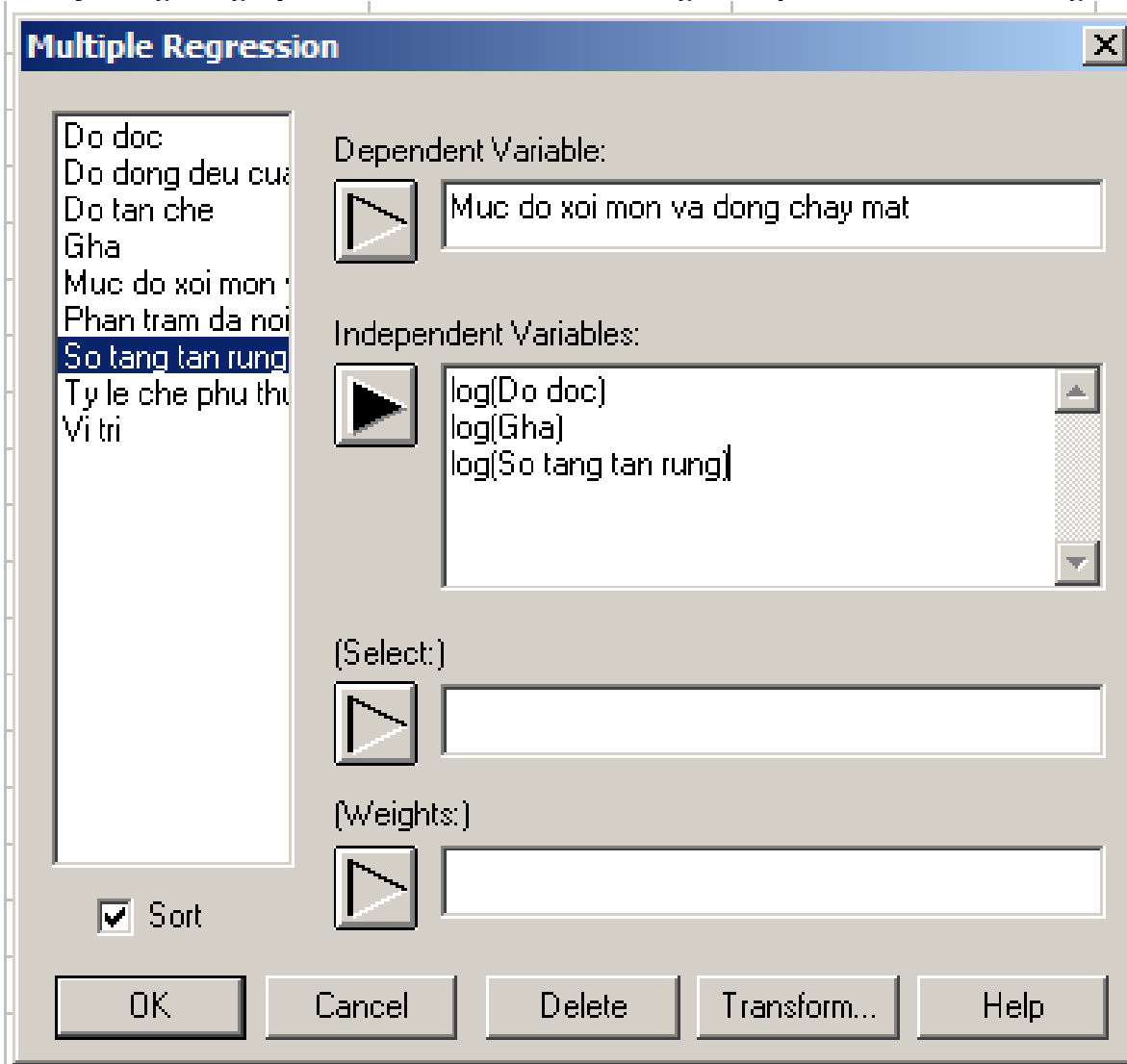
| Các nhân tố                              | Giá trị, hoặc mô tả  | Giải thích   |
|--|--|--|
|  |  | 1: Chân<br>2: Sườn<br>3: Đỉnh  |
| - Độ dốc (o)                             | Cấp: 1: <10; 2: 10 – 20; 3: 20 – 30,<br>4: 30 – 40; 5: >40 | Dùng thước Sunnto  |
| - Chiều dài dốc (m)                      |  | Thước dây 30m  |
| - Hướng phơi                             |  | Địa bàn  |
| <b>3 Nhóm nhân tố đất đai</b>            |  |  |
| - Loại đất                               |  |  |
| - Màu sắc đất                            |  | Mã số ngẫu nhiên   |
| - Độ dày tầng đất (cm)                   |  |  |
| - Độ xốp đất                             |  | 1: Tơi xốp<br>2: Chặt<br>3: Bí chặt  |
| - Độ ẩm đất (%)                          |  | Dụng cụ đo   |
| - pH đất                                 |  | nt   |
| - Nhiệt độ đất (oC)                      |  | nt   |
| - % kết von                              |  |  |
| - % đá nổi                               |  |  |
| - % ụ đất do giun đất tạo nên            |  |  |
| <b>4. Nhóm nhân tố khí hậu thủy văn</b>  |  |  |
| - Lượng mưa trung bình năm (mm)          |  | Số liệu thứ cấp  |
| - Độ ẩm không khí (%)                    |  | Dụng cụ đo   |
| - Nhiệt độ không khí (oC)                |  | nt   |
| - Lux                                    |  | nt   |
| - Cự ly đến sông suối gần nhất (m)       |  | Bản đồ địa hình + GPS  |
| <b>5. Nhóm nhân tố nhân tác</b>          |  |  |
| Mức độ lửa rừng                          |  | 0: Không có<br>1: Ít có<br>2: Vài năm<br>3: Hàng năm   |
| Mức động tác động đến thảm thực vật rừng |  | 0: Nguyên sinh<br>1: Khai thác chọn<br>2: Bỏ hóa sau nương rẫy<br>3: Chặt trắng để trồng cây nông nghiệp |

**Bảng tổng hợp số liệu mã hóa các nhân tố liên quan đến xói mòn, dòng chảy mặt trong lưu vực Bi Đúp – Núi Bà – Lâm Đồng**

| Mức độ xói mòn và dòng chảy mặt | Độ tan che | Ghè | Số tầng tan rưng | Độ đồng đều của cây rưng | Tỷ lệ che phủ thực bì | Vị trí | Độ dốc | Phân trạm địa nơi |
|---------------------------------|------------|-----|------------------|--------------------------|-----------------------|--------|--------|-------------------|
| 2                               | 0.7        | 20  | 5                | 1                        | 60                    | 2      | 45     | 0                 |
| 2                               | 0.6        | 10  | 5                | 2                        | 7                     | 0      | 12     | 80                |
| 3                               | 0.8        | 12  | 4                | 2                        | 7                     | 2      | 23     | 0                 |
| 2                               | 0.3        | 10  | 1                |                          | 16                    | 1      | 9      | 0                 |
| 1                               | 0.0        | 0   | 1                | 1                        | 95                    | 0      | 9      | 0                 |
| 1                               | 0.8        | 33  | 5                | 1                        | 5                     | 3      | 50     | 2                 |
| 2                               | 0.8        | 34  | 5                | 1                        | 5                     | 3      | 80     | 0                 |
| 2                               | 0.6        | 34  | 5                | 2                        | 9                     | 2      | 14     | 0                 |
| 2                               | 0.7        | 23  | 5                | 1                        | 70                    | 2      | 30     | 0                 |
| 2                               | 0.7        | 22  | 4                | 1                        | 6                     | 2      | 30     | 0                 |
| 2                               | 0.7        | 25  | 5                | 1                        | 5                     | 0      | 20     | 20                |
| 2                               | 0.8        | 32  | 5                | 1                        | 60                    | 2      | 30     | 0                 |
| 1                               | 0.8        | 31  | 5                | 1                        | 6                     | 3      | 40     | 0                 |
| 1                               | 0.8        | 30  | 4                | 2                        | 0                     | 2      | 15     | 0                 |
| 1                               | 0.7        | 31  | 5                | 2                        | 7                     | 1      | 20     | 0                 |
| 1                               | 0.7        | 20  | 5                | 2                        | 2                     | 2      | 20     | 0                 |
| 1                               | 0.7        | 34  | 5                | 2                        | 5                     | 2      | 30     | 0                 |
| 1                               | 0.7        | 32  | 5                | 2                        | 5                     | 2      | 70     | 0                 |
| 1                               | 0.7        | 10  | 5                | 2                        | 7                     | 1      | 40     | 2                 |
| 1                               | 0.6        | 15  | 4                | 2                        | 5                     | 2      | 46     | 0                 |
| 1                               | 0.7        | 33  | 3                | 2                        | 0                     | 2      | 15     | 0                 |
| 1                               | 0.3        | 26  | 5                | 2                        | 90                    | 2      | 18     | 0                 |
| 1                               | 0.3        | 20  | 4                | 3                        | 90                    | 2      | 18     | 0                 |
| 1                               | 0.6        | 8   | 5                | 2                        | 0                     | 0      | 15     | 0                 |
| 1                               | 0.5        | 27  | 5                | 2                        | 20                    | 2      | 15     | 0                 |
| 1                               | 0.7        | 32  | 5                | 2                        | 50                    | 0      | 5      | 0                 |
| 1                               | 0.5        | 22  | 4                | 2                        | 0                     | 3      | 0      | 0                 |
| 2                               | 0.6        | 28  | 4                | 1                        | 0                     | 2      | 10     | 0                 |
| 2                               | 0.3        | 4   | 5                | 3                        | 30                    | 2      | 8      | 0                 |
| 2                               | 0.5        | 12  | 4                | 3                        | 30                    | 2      | 15     | 0                 |
| 2                               | 0.0        | 0   | 2                | 2                        | 90                    | 2      | 20     | 0                 |
| 1                               | 0.2        | 18  | 2                | 3                        | 60                    | 3      | 0      | 0                 |
| 3                               | 0.1        | 5   | 1                |                          | 70                    | 2      | 5      | 0                 |
| 1                               | 0.3        | 35  | 5                | 3                        | 70                    | 1      | 5      | 0                 |
| 1                               | 0.0        | 0   | 0                |                          | 70                    | 1      | 3      | 0                 |
| 1                               | 0.6        | 20  | 4                | 2                        | 40                    | 2      | 12     | 0                 |
| 1                               | 0.7        | 27  | 4                | 2                        | 50                    | 3      | 0      | 0                 |
| 1                               | 0.7        | 26  | 3                | 2                        | 0                     | 3      | 0      | 0                 |
| 1                               | 0.3        | 26  | 5                | 2                        | 0                     | 2      | 9      | 0                 |
| 1                               | 0.6        | 22  | 5                | 2                        | 50                    | 3      |        | 0                 |
| 1                               | 0.6        | 26  | 4                | 2                        | 50                    | 2      | 2      | 0                 |
| 1                               | 0.7        | 21  | 5                | 2                        | 20                    | 0      | 14     | 0                 |
| 1                               | 0.6        | 33  | 5                | 2                        | 50                    | 3      | 0      | 0                 |
| 3                               | 0.0        | 0   | 0                |                          | 80                    | 2      | 15     | 0                 |
| 3                               | 0.7        | 35  | 4                | 2                        | 30                    | 2      | 40     | 15                |
| 2                               | 0.7        | 24  | 4                | 2                        | 30                    | 2      | 15     | 0                 |
| 2                               | 0.8        | 45  | 5                | 2                        | 5                     | 2      | 20     | 0                 |
| 1                               | 0.6        | 36  | 3                | 2                        | 0                     | 3      | 5      | 0                 |
| 1                               | 0.7        | 27  | 5                | 2                        | 10                    | 3      | 0      | 0                 |
| 3                               | 0.0        | 0   | 0                |                          | 90                    | 2      | 22     | 15                |
| 1                               | 0.5        | 37  | 1                | 2                        | 40                    | 2      | 5      | 0                 |
| 1                               | 0.8        | 24  | 4                | 1                        | 0                     | 2      | 18     | 0                 |

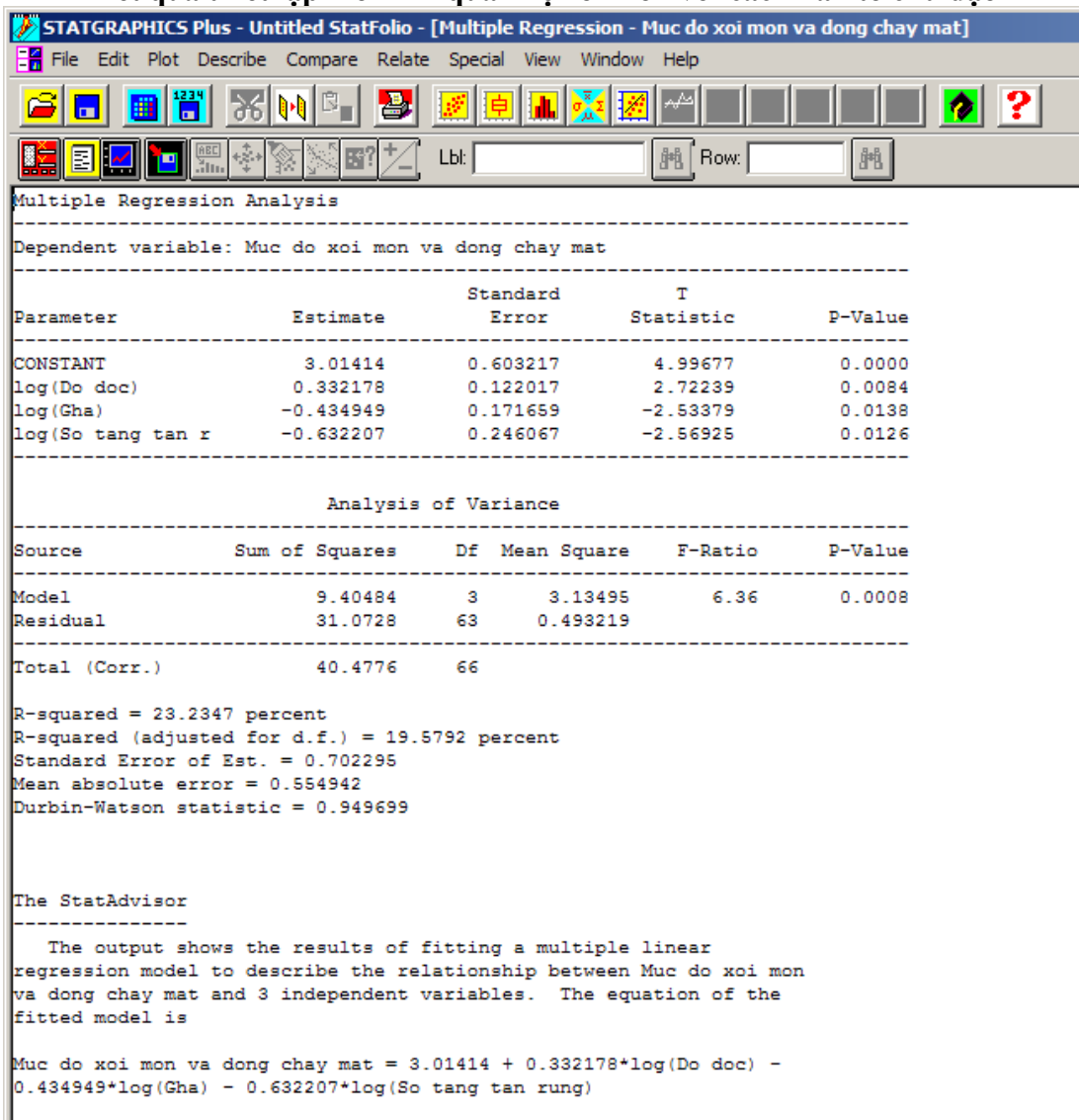
| Muc do xoi mon va dong chay mat | Do tan che | Gha | So tang tan rung | Do dong deu cua cay rung | Ty le che phu thuc bi | Vi tri | Do doc | Phan tram da noi |
|---------------------------------|------------|-----|------------------|--------------------------|-----------------------|--------|--------|------------------|
| 2                               | 0.7        | 37  | 4                | 1                        | 0                     | 3      | 10     | 0                |
| 2                               | 0.8        | 32  | 5                | 1                        | 5                     | 3      | 12     | 0                |
| 2                               | 0.7        | 36  | 5                | 2                        | 10                    | 2      | 40     | 0                |
| 2                               | 0.7        | 26  | 4                | 1                        | 25                    | 2      | 25     | 0                |
| 3                               | 0.8        | 23  | 4                | 1                        | 30                    | 2      | 18     | 0                |
| 2                               | 0.0        | 0   | 1                |                          | 90                    | 1      | 0      | 0                |
| 2                               | 0.5        | 22  | 4                | 2                        | 30                    | 3      | 17     | 0                |
| 2                               | 0.8        | 13  | 3                | 1                        | 0                     | 3      | 62     | 0                |
| 1                               | 0.4        | 19  | 3                | 2                        | 60                    | 1      | 22     | 0                |
| 3                               | 0.3        | 21  | 3                | 2                        | 40                    | 2      | 20     | 0                |
| 3                               | 0.9        | 14  | 4                | 2                        | 20                    | 2      | 54     | 20               |
| 2                               | 0.0        | 0   | 2                |                          | 80                    | 1      | 0      | 0                |
| 2                               | 0.9        | 22  | 3                | 1                        | 30                    | 1      | 12     | 0                |
| 2                               | 0.9        | 19  | 4                | 1                        | 30                    | 1      | 22     | 0                |
| 2                               | 0.8        | 16  | 4                | 2                        | 80                    | 1      | 6      | 0                |
| 2                               | 0.2        | 18  | 2                | 3                        | 95                    | 2      | 15     | 0                |
| 2                               | 0.9        | 25  | 4                | 1                        | 35                    | 2      | 25     | 0                |
| 2                               | 0.8        | 21  | 4                | 1                        | 30                    | 2      | 19     | 0                |
| 4                               | 0.6        | 4   | 4                | 3                        | 30                    | 0      | 16     | 0                |
| 3                               | 0.0        | 0   | 1                | 1                        | 85                    | 1      | 0      | 0                |
| 4                               | 0.3        | 21  | 2                | 3                        | 80                    | 2      | 25     | 2                |
| 3                               | 0.3        | 20  | 4                | 3                        | 90                    | 1      | 0      | 10               |
| 3                               | 0.8        | 38  | 4                | 1                        | 25                    | 1      | 40     | 0                |
| 3                               | 0.0        | 0   | 1                |                          | 90                    | 1      | 0      | 0                |
| 3                               | 0.3        | 22  | 2                | 2                        | 90                    | 3      | 45     | 5                |
| 3                               | 0.5        | 30  | 3                | 4                        | 20                    | 2      | 37     | 1                |
| 2                               | 0.5        | 13  | 2                | 1                        | 95                    | 1      | 0      | 0                |
| 2                               | 0.3        | 17  | 4                | 1                        | 60                    | 1      | 0      | 0                |
| 2                               | 0.9        | 13  | 3                | 1                        | 0                     | 3      | 0      | 0                |
| 2                               | 0.8        | 32  | 4                | 1                        | 15                    | 2      | 45     | 0                |
| 2                               | 0.6        | 26  | 3                | 1                        | 35                    | 1      | 40     | 0                |
| 2                               | 0.6        | 20  | 4                | 2                        | 40                    | 0      | 25     | 3                |
| 2                               | 0.1        | 1   | 2                | 3                        | 95                    | 0      | 0      | 0                |
| 2                               | 0.6        | 32  | 5                | 2                        | 30                    | 2      | 42     | 0                |
| 2                               | 0.8        | 28  | 4                | 1                        | 30                    | 3      | 0      | 0                |
| 1                               | 0.6        | 48  | 3                | 2                        | 40                    | 3      | 0      | 0                |
| 1                               | 0.9        | 18  | 4                | 1                        | 20                    | 1      | 0      | 0                |
| 1                               | 0.9        | 30  | 4                | 1                        | 10                    | 3      | 0      | 0                |
| 1                               | 0.8        | 28  | 4                | 1                        | 15                    | 2      | 32     | 0                |
| 1                               | 0.6        | 45  | 4                | 2                        | 45                    | 2      | 35     | 2                |
| 1                               | 0.3        | 18  | 3                | 2                        | 78                    | 3      | 0      | 0                |

Xử lý trong Statgraphics để tìm nhân tố ảnh hưởng chủ đạo đến xói mòn và dòng chảy.





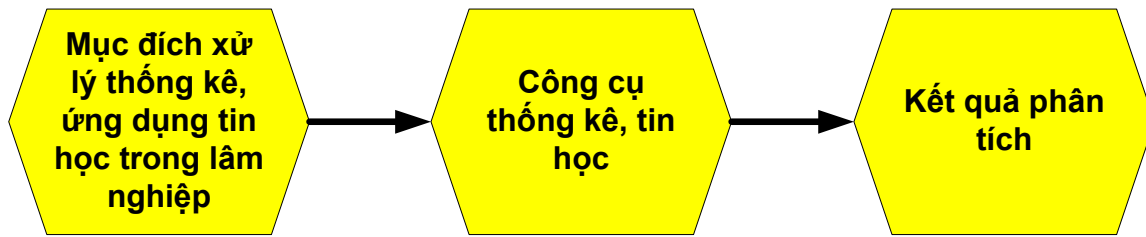
## Kết quả thiết lập mô hình quan hệ xói mòn với các nhân tố chủ đạo



**Muc do xoi mon va dong chay mat = 3.01414 + 0.332178\*log(Do doc) - 0.434949\*log(Gha) - 0.632207\*log(So tang tan rung)**

Với R = 0.5 và các giá trị P kiểm tra các nhân tố ảnh hưởng đều < 0.1

Mô hình này phản ánh mức độ xói mòn gia tăng theo cấp độ dốc, do đó phân cấp độ dốc là một lớp nhân tố trong phân cấp xung yếu phòng hộ đầu nguồn; bên cạnh đó nếu thảm phủ rừng có G và số tầng tán gia tăng thì làm giảm nguy cơ xói mòn, giảm dòng chảy mặt. Vì vậy lớp thảm phủ rừng cần xác định các nhân tố G và số tầng tán làm chủ đạo trong phân cấp xung yếu đầu nguồn.



|  |   |  |
|--|---|--|
| Mô tả một mẫu: Trạng thái rừng, lô rừng, ô thí nghiệm, vườn ươm  | Phân tích các đặc trưng của một mẫu quan sát  | Phản ánh đối tượng:<br>- Số trung bình<br>- Phương sai, biến động<br>- Kiểu phân bố: Lẻch (Sk), nhọn (Ku)<br>- Ước lượng khoảng biến động theo độ tin cậy ?% |
| So sánh các thí nghiệm: Trong phòng, hiện trường:<br>- So sánh 2 công thức, 2 mẫu<br>- So sánh tác động 1 – 2 nhân tố        | - Tiêu chuẩn t<br>- ANOVA 1 -2 nhân tố  | Chọn lựa được công thức tối ưu: Loại, biện pháp kỹ thuật, giống, ...   |
| Phát hiện quan hệ sinh thái loài   | Phân tích xác suất theo q và X <sup>2</sup>   | Các loại quan hệ: Dương, âm, ngẫu nhiên phục vụ trồng và làm giàu rừng hỗn giao  |
| Lập các mô hình tuyến tính, phi tuyến 1 - nhiều biến   | Phân tích tương quan hồi quy  | Các mô hình quan hệ phục vụ:<br>- Điều tra rừng<br>- Phát hiện các nhân tố ảnh hưởng đến 1 nhân tố phụ thuộc   |
| Mô phỏng quy luật cấu trúc 3 chiều   | - Sắp xếp tần số<br>- X <sup>2</sup> kiểm tra thuần nhất<br>- Tiêu chuẩn U để đánh giá phân bố mặt bằng<br>- Mô phỏng cấu trúc: Mayer, Weibull, ... | - Phát hiện và mô tả quy luật phân bố<br>- Ứng dụng trong điều chế rừng  |
| Phát hiện các yếu tố ảnh hưởng đến một vấn đề: Các nhân tố sinh thái, nhân tác ảnh hưởng đến sinh trưởng, đất, nước, lưu vực | - Mã hóa biến định tính<br>- Phân tích hồi quy đa biến, lọc biến, tổ hợp biến   | - Phát hiện các nhân tố chủ đạo về sinh thái, sinh trưởng, lưu vực<br>- Quy hoạch, đề xuất biện pháp kỹ thuật  |

### Tóm tắt ứng dụng thống kê và tin học trong lâm nghiệp