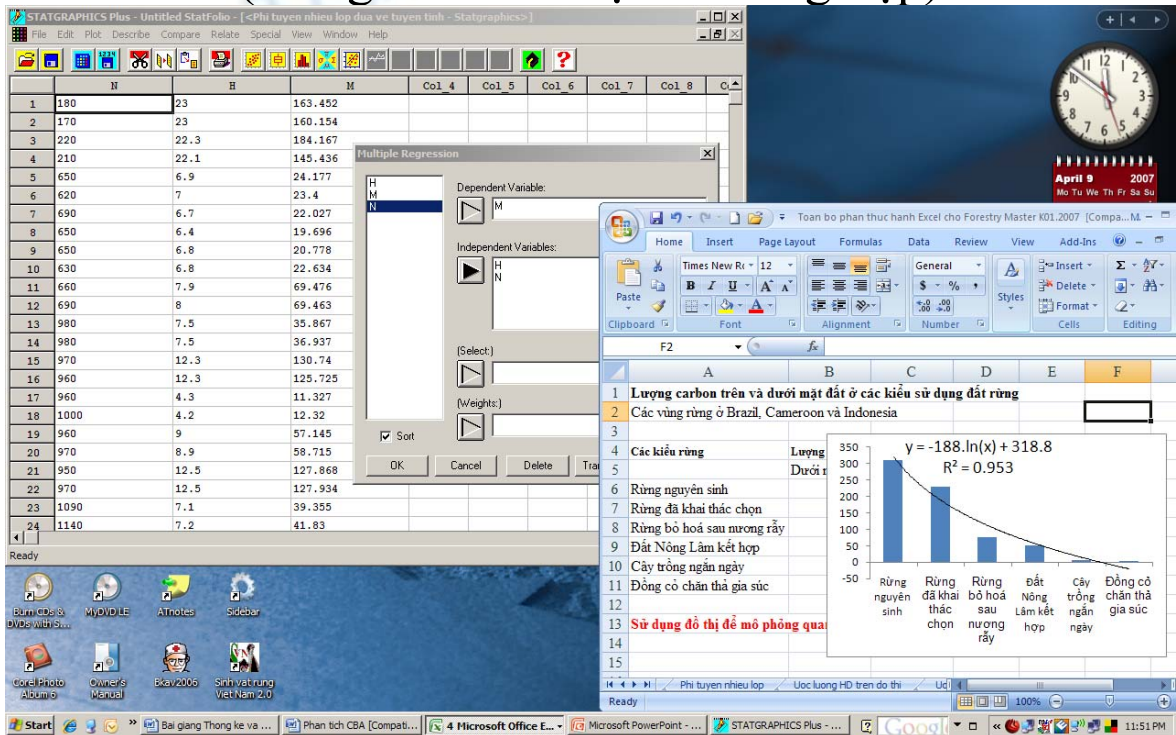


**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC TÂY NGUYÊN**

PGS.TS. BẢO HUY

**THỐNG KÊ VÀ TIN HỌC
TRONG LÂM NGHIỆP**
(Dùng cho Cao học Lâm nghiệp)



Tháng 3 năm 2008

Mục lục

1	TỔNG QUÁT VỀ CHỨC NĂNG XỬ LÝ THỐNG KÊ CỦA EXCEL VÀ STATGRAPHICS PLUS	4
1.1	Tổng quát về phần xử lý thống kê trong Excel	4
1.2	Tổng quát về phần mềm xử lý thống kê Statgraphics Plus	5
2	TÍNH TOÁN CÁC ĐẶC TRƯNG THỐNG KÊ CỦA MỘT MẪU QUAN SÁT VÀ ƯỚC LƯỢNG KHOẢNG	6
3	SO SÁNH TRUNG BÌNH 02 MẪU QUAN SÁT BẰNG TIÊU CHUẨN T CỦA STUDENT 10	
4	NGHIÊN CỨU MỐI QUAN HỆ SINH THÁI LOÀI TRONG RỪNG MƯA NHIỆT ĐỚI DỰA VÀO TIÊU CHUẨN χ^2 (Bảo Huy, 1997)	13
5	PHÂN TÍCH PHƯƠNG SAI	19
5.1.	Phân tích phương sai 1 nhân tố với các thí nghiệm ngẫu nhiên hoàn toàn	19
5.2.	Phân tích phương sai 2 nhân tố	22
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):	22
5.2.2.	Phân tích phương sai 2 nhân tố m lần lặp	28
6.	PHÂN TÍCH TƯƠNG QUAN - HỒI QUY	32
6.1.	Hồi quy tuyến tính 1 lớp	32
6.2.	Dạng phi tuyến đưa về tuyến tính 1 lớp	34
6.2.1.	Lập mô hình hàm mũ trong Excel:	34
6.2.2.	Lập mô hình hàm mũ và Schumacher trong Statgraphics Plus:	36
6.3.	Hồi quy tuyến tính nhiều lớp	41
6.4.	Hồi quy phi tuyến tính nhiều lớp, tổ hợp biến	44
6.4.1.	Lập mô hình phi tuyến nhiều lớp chuyển về tuyến tính nhiều lớp trong Excel	44
6.4.2.	Lập mô hình phi tuyến nhiều lớp chuyển về tuyến tính trong Statgraphics Plus	46
7.	ƯỚC LƯỢNG CÁC DẠNG HỒI QUY MỘT BIẾN TRÊN ĐỒ THỊ	50
8.	SẮP XẾP VÀ VẼ BIỂU ĐỒ PHÂN BỐ TẦN SỐ XUẤT HIỆN THEO CẤP, CỖ, HẠNG .	54
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	56
10.	MÔ HÌNH HOÁ QUY LUẬT PHÂN BỐ	57
10.1.	Mô hình hoá phân bố giảm theo hàm Mayer	58
10.2.	Mô phỏng phân bố thực nghiệm theo phân bố khoảng cách-hình học:	62
10.3.	Mô phỏng phân bố thực nghiệm theo phân bố Weibull:	64
11.	PHƯƠNG PHÁP NGHIÊN CỨU CẤU TRÚC MẶT BẰNG RỪNG (Bảo Huy, 1993) .	66
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) ...	68

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 Plus, 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 Plus 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 PLUS

1.1 Tổng quát về phần xử lý thống kê trong Excel

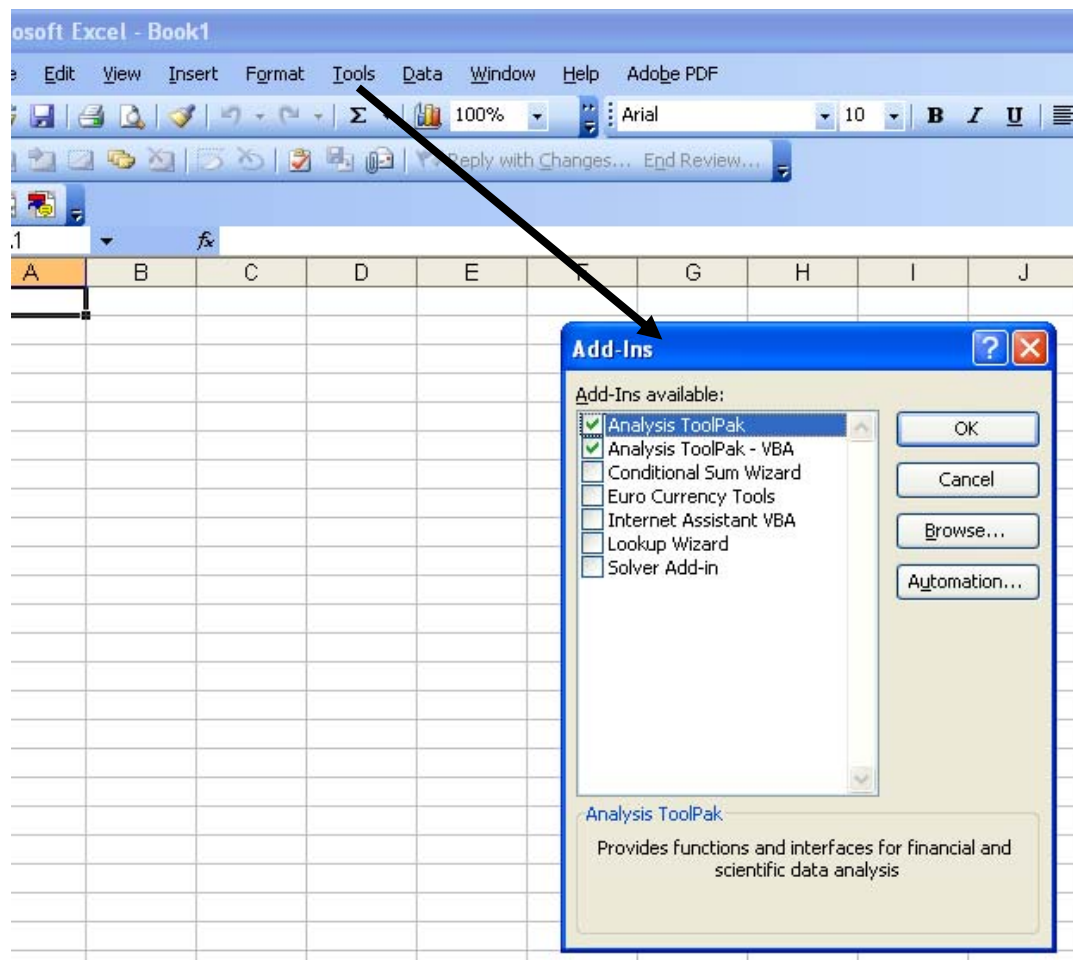
Excel thiết kế sẵn một số chương trình để xử lý số liệu và phân tích thống kê cơ bản ứng dụng trong nhiều lĩnh vực:

- 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, χ^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, AA OVA, 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 đồ thị.

.....

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

- o *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.*
- o *Khi chạy Excel lần đầu cần mở chế độ phân tích dữ liệu bằng cách: Menu Tools/Add-Ins và chọn Analysis Toolpak-OK.*



ả hư vậy 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 thống kê sẵn có ở một phần mềm phổ biến trong bất kỳ một máy vi tính cá nhân nào.

Trước hết cần lưu ý sử dụng các hàm, các tiêu chuẩn thống kê thông dụng trong Excel như sau:

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, χ^2 : Trong phân tích thống kê, khi áp dụng một tiêu chuẩn nào đó, cần thiết phải so sánh với giá trị tra bảng ở mức độ tin cậy nhất định để đánh giá và kết luận. Trong Excel đã lập và tính sẵn các hàm để tra các giá trị này.

- 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 χ^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 Plus

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.
- 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 biến ảnh hưởng đến một hậu quả (biến phụ thuộc).

The screenshot shows the STATGRAPHICS Plus interface. The main window displays regression statistics and an Analysis of Variance table. To the right, a data table titled '<Phi tuyen nhieu lop' lists values for 'N' from 1 to 18.

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	173850.0	2	86925.0	104.16	0.0000
Residual	30876.9	37	834.51		
Total (Corr.)	204727.0	39			

	N
1	180
2	170
3	220
4	210
5	650
6	620
7	690
8	650
9	650
10	630
11	660
12	690
13	980
14	980
15	970
16	960
17	960
18	1000
19	650

Để giao tiếp trong Statgraphics, số liệu đầu vào có thể được nhập trực tiếp trong file bảng tính và cơ sở dữ liệu của nó; 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 Plus để tính toán, thiết lập mô hình, Cơ sở dữ liệu lập trong Excel cần lưu dưới dạng phiên bản của Excel 5.0 trở về trước, vì Statgraphics Plus chưa nhận được các phiên bản sau này như Excel 97 – 2003, 2007.

2 TÍNH TOÁN CÁC ĐẶC TRƯNG THÔNG KÊ CỦA MỘT MẪU QUAN SÁT VÀ ƯỚC LƯỢNG KHOẢNG

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 thành thu thập dữ liệu theo một nhân tố chủ đạo và sau đó ước lượng, tính toán các đặc trưng cơ bản. Đây là các thông tin cơ bản về một đối tượng quan sát, theo một chỉ tiêu, nhân tố quan tâm.

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:

- 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: ấu 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 Tách

	A	B	C	D	E	F
1						
2	Số liệu đo đường kính cây tách trong lâm phần					
3	Tính các đặc trưng mẫu					
4						
5						
6	D1,3 (cm)					
7	21.0			<i>D_{1,3} (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

The screenshot shows the 'Descriptive Statistics' dialog box in Microsoft Excel. The input range is \$A\$1:\$A\$52. The 'Labels in first row' checkbox is checked. Under 'Output options', 'New Worksheet Ply' is selected. 'Summary statistics' and 'Confidence Level for Mean' (95%) are also checked. The 'Kth Largest' and 'Kth Smallest' options are set to 1.

Kết quả tính đặc trưng mẫu

D1,3 (cm)	
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.
Sk = 0 phân bố đối xứng.
Sk > 0 đỉnh đường cong lệch trái so với số trung bình.
Sk < 0 đỉnh đường cong lệch phải so với số trung bình.
Sk = 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. Ắt ếu Sk = 0 thì độ lệch tiệm cận chuẩn.

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

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 TRUNG BÌNH 02 MẪU QUAN SÁT BẰNG TIÊU CHUẨN T CỦA STUDENT

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.

ả ếu 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 lẫn 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 ả - H 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
1	H (cây con)	H (rễ trần)
2	13,6	13
3	14	13,5
	13,8	12
	13	13,5

	A	B
1	H (cây con)	H (rễ trần)
	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-â ext (trong Function name): Xuất hiện hội 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).*

The screenshot shows the Microsoft Excel interface with the 'Data Analysis' task pane open. The 't-Test: Two-Sample Assuming Equal Variances' option is selected. The background spreadsheet displays data for two groups: 'H (cây con)' in column A and 'H (rễ trần)' in column B. The 'Data Analysis' dialog box includes the following options: Histogram, Moving Average, Random Number Generation, Rank and Percentile, Regression, Sampling, t-Test: Paired Two-Sample for Means, t-Test: Two-Sample Assuming Equal Variances (selected), t-Test: Two-Sample Assuming Unequal Variances, and z-Test: Two Sample for Means. Buttons for OK, Cancel, and Help are visible on the right side of the dialog.

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)
 ở đây đưa cả tiêu đề.
- Hypothesized mean difference: 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: ở đây 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.

ở đây: $P(T \leq t)$ 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 χ^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.

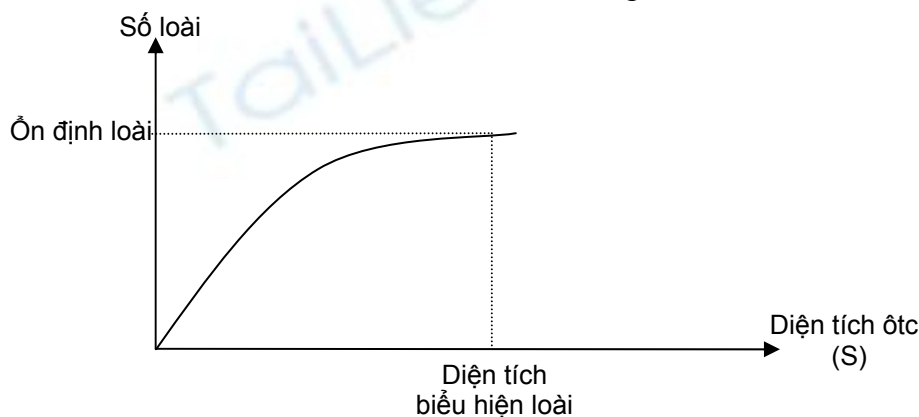
Nguyên tắc nghiên cứu trong trường hợp này là xác định một diện tích ô mẫu nhỏ nhất nhưng bảo đảm xuất hiện các loài ưu thế sinh thái.

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 $100m^2$ và tăng dần đến khoảng 1 – 2ha.

Số loài theo kích thước ô mẫu tăng dần

Stt	Diện tích ô tiêu chuẩn (m ²)	Số loài	Tên loài
1	100	5	a, d, hg
2	100	4	
3	100	6	
4	100		
5	200	6	
6	200	8	
7	200		
8	400	10	
9	400	11	
10	400		
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=diện tích ô tiêu chuẩn), dạng quan hệ sau có thể được sử dụng:

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

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

Ví dụ: Tiến hành điều tra thử nghiệm 53 ô có diện tích từ 100m² đến 10.000m² 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:

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

$$\text{Với } n=53 \quad R=0.907 \quad Fr=722.58 \quad \alpha < 0.01$$

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 $a = 16.810$. ả 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. ả 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 $a = 6$ 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 $S = 2.500m^2$. Vậy có thể chọn ô hình vuông với kích thước $50 \times 50m$.

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:

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

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

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

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%) %
	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 didynum	23	10.0
4	Xoan Mộc	Toona sureni	19	8.2
5	Bời lồi	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
14	Thừng mực	Wrightia annamensis	6	2.6
15	Máu chó	Knema conferta	4	1.7
16	Chua khét	Dysoxylum acutangulum	4	1.7
17	Trám	Canarium copaliferum	3	1.3
18	Gạo	Gossampinus 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 ρ và χ^2 .

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

ρ : Hệ số tương quan giữa 2 loài A và B.

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

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.

$$P(AB) = \frac{nAB}{n} \quad P(A) = \frac{nA + nAB}{n} \quad P(B) = \frac{nB + nAB}{n}$$

Với:

nA : số ô tiêu chuẩn chỉ xuất hiện loài A.

nB : số ô tiêu chuẩn chỉ xuất hiện loài B.

nAB : số ô tiêu chuẩn xuất hiện đồng thời 2 loài A và B.

n : tổng số ô quan sát ngẫu nhiên.

ρ nói lên chiều hướng liên hệ và mức độ liên hệ giữa 2 loài. $\rho < 0$: 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 2×2 :

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

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

Trong đó:

$a = nAB$; $b = nB$; $c = nA$; d : số ô không chứa cả 2 loài a và B.

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

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

ả ếu $\chi^2 > \chi^2_{0.05} = 3.84$ 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 ρ và χ^2 :

χ^2 : để kiểm tra mối quan hệ từng cặp loài.

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

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)	ρ	χ^2	Quan hệ
1	Xoan Mộc	Bằng Lăng	5	13	14	0	0.594	0.844	0.438	-0.356	3.99	Có quan hệ âm
2	Xoan Mộc	Dẻ	0	11	19	2	0.594	0.938	0.594	0.312	3.04	Ngẫu nhiên
3	Xoan Mộc	Bời Lồi	7	6	12	7	0.594	0.563	0.375	0.168	0.89	Ngẫu nhiên
4	Xoan Mộc	Vạng Trứng	10	5	9	8	0.594	0.438	0.281	0.088	0.24	Ngẫu nhiên
5	Xoan Mộc	Trâm	10	5	9	8	0.594	0.438	0.281	0.088	0.24	Ngẫu nhiên
6	Xoan Mộc	Xương cá	5	9	14	4	0.594	0.719	0.438	0.049	0.07	Ngẫu nhiên
7	Xoan Mộc	Bồ hòn	10	7	9	6	0.594	0.500	0.281	-0.064	0.12	Ngẫu nhiên
8	Xoan Mộc	Chò xốt	12	8	7	5	0.594	0.469	0.219	-0.243	1.86	Ngẫu nhiên
9	Bằng Lăng	Dẻ	2	5	25	0	0.844	0.938	0.781	-0.111	0.36	Ngẫu nhiên
10	Bằng Lăng	Bời Lồi	13	4	14	2	0.844	0.563	0.438	-0.206	0.40	Ngẫu nhiên
11	Bằng Lăng	Vạng Trứng	16	3	11	2	0.844	0.438	0.344	-0.141	0.61	Ngẫu nhiên
12	Bằng Lăng	Trâm	14	1	13	4	0.844	0.438	0.406	0.206	1.32	Ngẫu nhiên
13	Bằng Lăng	Xương cá	9	5	18	0	0.844	0.719	0.563	-0.269	2.27	Ngẫu nhiên
14	Bằng Lăng	Bồ hòn	13	2	14	3	0.844	0.500	0.438	0.086	0.22	Ngẫu nhiên
15	Bằng Lăng	Chò xốt	13	1	14	4	0.844	0.469	0.438	0.232	1.68	Ngẫu nhiên
16	Dẻ	Bời Lồi	14	2	16	0	0.938	0.563	0.500	-0.228	1.60	Ngẫu nhiên
17	Dẻ	Vạng Trứng	18	2	12	0	0.938	0.438	0.375	-0.293	2.67	Ngẫu nhiên
18	Dẻ	Trâm	17	1	13	1	0.938	0.438	0.406	-0.033	0.03	Ngẫu nhiên
19	Dẻ	Xương cá	7	0	23	2	0.938	0.719	0.719	0.413	5.33	Có quan hệ dương
20	Dẻ	Bồ hòn	14	0	16	2	0.938	0.500	0.500	0.258	2.07	Ngẫu nhiên
21	Dẻ	Chò xốt	16	1	14	1	0.938	0.469	0.438	-0.016	0.00	Ngẫu nhiên
22	Bời lồi	Vạng Trứng	11	7	7	7	0.563	0.438	0.219	-0.111	0.38	Ngẫu nhiên
23	Bời lồi	Trâm	7	3	11	11	0.563	0.438	0.344	0.397	4.99	Có quan hệ dương
24	Bời lồi	Xương cá	5	10	13	4	0.563	0.719	0.406	0.009	0.00	Ngẫu nhiên
25	Bời lồi	Bồ hòn	11	9	7	5	0.563	0.500	0.219	-0.252	2.00	Ngẫu nhiên
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:
 - Kiểm tra bằng đặc trưng mẫu, sơ đồ.
 - ả ều 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 onduras) “ 4 “
- 4: P.línures (ả onduras) “ 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 hư 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
	Xuất xứ	Lặp 1	Lặp 2	Lặp 3	Lặp 4
1					
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: ả hộp 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 colum (row).
- Output range: Đưa địa chỉ ô trên trái nơi xuất kết quả.
- Kích OK.