

PGS.TS. BÁO HUY

# TIN HỌC THỐNG KÊ TRONG LÂM NGHIỆP

SỬ DỤNG CÁC CHƯƠNG TRÌNH  
R, STATGRAPHICS, SPSS

(Giáo trình đại học và sau đại học)



NHÀ XUẤT BẢN KHOA HỌC VÀ KỸ THUẬT



# LỜI MỞ ĐẦU

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

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

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

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

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

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

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

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

Chương 4: Tin học thống kê so sánh;

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

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

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

Chương 8: Tin học thống kê chuyên đề.

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

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

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

*Thành phố Corvallis, Bang Oregon, USA,*

Ngày 18 tháng 04 năm 2016

**Bảo Huy**

# MỤC LỤC

Lời nói đầu ..... 5

## Chương 1

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

1.1 Phương pháp luận về tin học thống kê trong nghiên cứu khoa học lâm nghiệp .....	11
1.2 Cơ sở để kết luận thống kê lâm nghiệp .....	12
1.3 Giới thiệu các chương trình xử lý thống kê .....	15
1.3.1 Chương trình mã nguồn mở R .....	15
1.3.2 Chương trình Statgraphics .....	17
1.3.3 Chương trình SPSS .....	19
1.3.4 Chương trình thống kê trong Excel .....	21

## Chương 2

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

2.1 Tính toán dung lượng mẫu trong điều tra, đánh giá .....	23
2.1.1 Xác định dung lượng mẫu khi không có phân cấp, khối, loại.....	23
2.1.2 Xác định dung lượng mẫu theo phân cấp, khối, loại .....	25
2.2 Phương pháp bố trí, rút mẫu trong điều tra, đánh giá để xử lý thống kê .....	30
2.2.1 Rút mẫu ngẫu nhiên (Random sampling) .....	30
2.2.2 Rút mẫu hệ thống (Systematic sampling) .....	34
2.2.3 Rút mẫu theo cụm (Cluster sampling).....	35
2.2.4 Rút mẫu điển hình .....	36
2.3 Nguyên tắc thiết kế thử nghiệm trong lâm nghiệp, quản lý tài nguyên môi trường rừng.....	36

## Chương 3

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

3.1 Thông tin thống kê về đặc trưng của mẫu quan sát.....	40
3.2 Tính toán các chỉ tiêu thống kê mô tả mẫu .....	43
3.3 Kiểm tra phân bố chuẩn của mẫu quan sát - bổ sung số liệu hoặc đổi biến số .....	44
3.4 Ước lượng biến động của số trung bình với độ tin cậy cho trước .....	47

Chương 4  
**TIN HỌC THỐNG KÊ SO SÁNH**

4.1 So sánh trung bình một mẫu với một giá trị cho trước.....	49
4.2 So sánh hai mẫu quan sát - thí nghiệm.....	51
4.2.1 So sánh sự sai khác giữa trung bình hai mẫu quan sát độc lập.....	51
4.2.2 So sánh sự sai khác hai trung bình của hai mẫu quan sát bắt cặp.....	58
4.2.3 Ước lượng biến động về tỷ lệ và so sánh tỷ lệ của hai mẫu .....	61

Chương 5  
**TIN HỌC TRONG PHÂN TÍCH PHƯƠNG SAI  
(ANALYSIS OF VARIANCE - ANOVA)**

5.1 Phân tích phương sai một nhân tố với bô trí thí nghiệm ngẫu nhiên hoàn toàn.....	64
5.2 Phân tích phương sai một nhân tố với bô trí thí nghiệm theo khói ngẫu nhiên đầy đủ (Randomized Complete Blocks) (RCB) hoặc phân tích phương sai hai nhân tố một lần lặp lại .....	69
5.3 Phân tích phương sai nhiều nhân tố với m lần lặp .....	75

Chương 6

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

6.1 Tiêu chuẩn phi tham số để so sánh các mẫu độc lập.....	81
6.2 Tiêu chuẩn phi tham số kiểm tra, so sánh các mẫu liên hệ.....	85

Chương 7

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

7.1 Khái niệm chung về mô hình hóa rừng, mô hình quan hệ.....	88
7.2 Các tiêu chuẩn, tiêu chí thống kê để so sánh, đánh giá, lựa chọn mô hình quan hệ.....	89
7.3 Các biểu đồ, đồ thị dùng để đánh giá, so sánh các mô hình .....	91
7.4 Mô hình tuyến tính.....	93
7.4.1 Mô hình tuyến tính đơn.....	93
7.4.2 Mô hình tuyến tính đa biến .....	100
7.5 Mô hình phi tuyến tính.....	114
7.5.1 Lựa chọn các biến số ảnh hưởng trong mô hình phi tuyến .....	116
7.5.2 Ước lượng mô hình phi tuyến theo phương pháp tuyến tính hóa .....	118
7.5.3 Ước lượng mô hình theo phương pháp phi tuyến tính.....	142
7.5.4 Mô hình phi tuyến có hay không có trọng số (Weight) .....	150

7.5.5 Phương pháp phi tuyến ánh hưởng tổng hợp (Nonlinear Mixed-Effects - nlme) Maximum Likelihood có trọng số để ước lượng mô hình phi tuyến	160
7.6 Chỉ Số Furnival's Index để lựa chọn dạng phương trình khác nhau hoặc phương pháp ước lượng mô hình phi tuyến: Tuyến tính hóa hay phi tuyến Maximum Likelihood	164
7.7 Mô hình thay đổi tham số dưới ảnh hưởng của các nhân tố ngẫu nhiên (random effect)	169
7.8 Phương pháp so sánh và thẩm định chéo các mô hình (Cross validation) .....	177
7.8.1 Thẩm định chéo để lựa chọn và đánh giá sai số, độ tin cậy của các mô hình .....	177
7.8.2 Phương pháp truyền thống – Sử dụng dữ liệu độc lập để so sánh và thẩm định sai số mô hình.....	178
7.8.3 Phương pháp thẩm định chéo sai số - Leave-One-Out Cross Validation (LOOCV) .....	183
7.8.4 Phương pháp thẩm định chéo sai số k-fold Cross Validation.....	189
7.8.5 Phương pháp thẩm định chéo sai số mô hình Monte Carlo Cross Validation.....	194

## Chương 8

### **TIN HỌC THỐNG KÊ CHUYÊN ĐỀ**

8.1 Mô phỏng quy luật phân bố - cấu trúc rừng, cấu trúc quần thể .....	204
8.1.1 Sắp xếp và vẽ biểu đồ phân bố tần số xuất hiện theo cấp, cỡ, hang .....	204
8.1.2 Kiểm tra thuần nhất K mẫu quan sát đứt quãng ứng dụng trong kiểm tra sự thuần nhất của các dãy phân bố N/DBH ở các ô tiêu chuẩn, số cá thể theo tuổi .....	207
8.1.3 Mô hình hoá cấu trúc phân bố dạng giảm theo hàm Meyer .....	210
8.1.4 Mô phỏng cấu trúc phân bố theo phân bố khoảng cách - hình học.....	212
8.1.5 Mô phỏng phân bố, cấu trúc theo phân bố Weibull.....	216
8.1.6 Xác định kiểu phân bố cây rừng trên mặt đất rừng.....	219
8.2 Xác định mối quan hệ sinh thái loài trong rừng mưa nhiệt đới .....	221
8.3 Mô hình quan hệ với các nhân tố định tính .....	228
<b>DANH MỤC BẢNG DỮ LIỆU THỰC HÀNH .....</b>	238
<b>TÀI LIỆU THAM KHẢO .....</b>	278

PGS.TS. BẢO HUY

# TIN HỌC THỐNG KÊ TRONG LÂM NGHIỆP

Sử dụng các chương trình R, Statgraphics, SPSS

*Chịu trách nhiệm xuất bản:*  
**GIÁM ĐỐC - TỔNG BIÊN TẬP**  
**PHẠM NGỌC KHÔI**

*Biên tập* : PHẠM THỊ MAI  
*Thiết kế bìa* : HOÀNG VIỆT  
*Trình bày* : PHẠM THỊ MAI  
*Sửa bản in* : PHẠM THỊ MAI

## NHÀ XUẤT BẢN KHOA HỌC VÀ KỸ THUẬT

70 Trần Hưng Đạo – Quận Hoàn Kiếm – Hà Nội

ĐT: (04) 3942 2443 Fax: (04) 3822 0658

Website: <http://www.nxbkhkt.com.vn> Email: nxbkhkt@hn.vnn.vn

**CHI NHÁNH NHÀ XUẤT BẢN KHOA HỌC VÀ KỸ THUẬT**  
28 Đồng Khởi, 12 Hồ Huân Nghiệp – Quận 1 – TP. Hồ Chí Minh

ĐT: (08) 3822 5062 Fax: (08) 3829 6628

Email: chinhhanhnxbskhkt@yahoo.com.vn

(Sách thật có đóng dấu và dán tem ở bìa 3)

---

In 300 bản, khổ 19cm × 27cm tại Công ty cổ phần thương mại In Nhật Nam  
Địa chỉ: 007 Lô I – KCN Tân Bình – P. Tây Thạnh – Q. Tân Phú – TP. Hồ Chí Minh  
Số ĐKXB: 691 – 2017/CXBIPH/5 – 18/KHKT  
Quyết định XB số: 23/QĐ-NXBKHKT, ngày 13/04/2017  
Mã ISBN: 978-604-67-0853-7  
In xong và nộp lưu chiểu quý I năm 2017