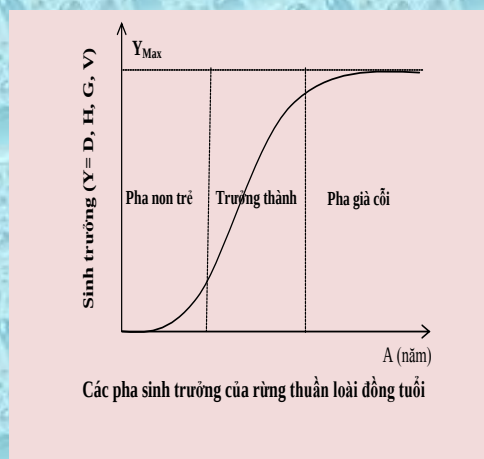


TRƯỜNG ĐẠI HỌC LÂM NGHIỆP

PGS. TS. Nguyễn Trọng Bình & PGS. TS. Nguyễn Văn Thêm

oOo

Hướng dẫn sử dụng IBM SPSS Statistics
để xử lý thông tin trong lâm học
(Sách tham khảo)



NHÀ XUẤT BẢN NÔNG NGHIỆP

Hà Nội, 2013

MỤC LỤC

Lời nói đầu.....	i
Chương 1. NHẬP DỮ LIỆU VÀO BẢNG TÍNH.....	1
1.1. Các thang đo	1
1.2. Cách nhập dữ liệu, mã hóa biến và tên biến	5
1.3. Chọn lọc và biến đổi dữ liệu.....	7
1.4. In ấn và copy số liệu	8
Chương 2. THỐNG KÊ MÔ TẢ VÀ KIỂM ĐỊNH PHÂN BỐ.....	10
2.1. Sử dụng thủ tục Descriptives.....	10
2.2. Sử dụng thủ tục Explore	10
2.3. Sử dụng thủ tục Frequencies	13
2.4. Kiểm định phân bố Normal và cân bằng phương sai	15
2.5. Sử dụng thủ tục Reports	17
2.6. Tính đặc trưng thống kê của bảng phân phối tần số	20
2.7. Lập bảng phân bố N – D, N – H.....	22
Một số hàm mô tả phân bố N – D, N – H.....	23
Phân bố Weibull	23
Phân bố Meyer.....	25
Phân bố khoảng cách	28
Chương 3. KIỂM ĐỊNH CÁC GIẢ THUYẾT THỐNG KÊ	30
3.1. Kiểm định phi tham số	30
Kiểm định 2 mẫu độc lập bằng thống kê Wilcoxon -Mann- Whitney	30
Kiểm định K mẫu ($K \geq 3$) độc lập bằng tiêu chuẩn Kruskal – Wallis.....	31
3.2. Kiểm định mẫu liên hệ hay mẫu bắt cặp	33
Kiểm định 2 mẫu liên hệ bằng thống kê Wilcoxon	33
Kiểm định K mẫu liên hệ ($K \geq 3$) bằng thống kê Friedman	34
3.3. Kiểm định 2 mẫu bằng thống kê T.....	35
Kiểm định T cho hai mẫu độc lập	35
Kiểm định T cho hai mẫu bắt cặp.....	36
Kiểm định T hai mẫu với việc sử dụng một điểm giới hạn	38
3.4. Kiểm định lô (Runs)	40
Chương 4. PHÂN TÍCH PHƯƠNG SAI (ANOVA).....	42
4.1. Khái quát	42
4.2. Các mô hình phân tích phương sai	43
ANOVA 1 yếu tố: Kiểu thí nghiệm ngẫu nhiên hoàn toàn	47
ANOVA 1 yếu tố: Kiểu thí nghiệm theo khối đầy đủ ngẫu nhiên	48
ANOVA 1 yếu tố: Kiểu thí nghiệm bình phương la tinh	50
ANOVA 2 yếu tố: Kiểu thí nghiệm ngẫu nhiên hoàn toàn	51
ANOVA 2 yếu tố: Kiểu thí nghiệm theo khối đầy đủ ngẫu nhiên	52
ANOVA 2 yếu tố: Kiểu thí nghiệm bình phương la tinh	53
ANOVA 2 yếu tố: Kiểu thí nghiệm có lô phụ	54
4.3. Thực hành phân tích phương sai	55
ANOVA 1 yếu tố: Kiểu thí nghiệm ngẫu nhiên hoàn toàn	55
ANOVA 1 yếu tố: Kiểu thí nghiệm theo khối đầy đủ ngẫu nhiên	61
ANOVA 1 yếu tố: Kiểu thí nghiệm bình phương la tinh	63
ANOVA 2 yếu tố: Kiểu thí nghiệm theo khối đầy đủ ngẫu nhiên	67
ANOVA 2 yếu tố: Kiểu thí nghiệm ngẫu nhiên hoàn toàn	69
ANOVA 2 yếu tố: Kiểu thí nghiệm bình phương la tinh	72

ANOVA 2 yếu tố: Kiểu thí nghiệm có lô phụ	74
ANOVA 3 yếu tố: Kiểu thí nghiệm theo khối đầy đủ ngẫu nhiên	77
ANOVA nhiều yếu tố: Mô hình với nhiều biến độc lập.....	80
4.4. Phân tích ANOVA cho nhiều biến phụ thuộc	82
Phân tích MANOVA 1 yếu tố theo khối đầy đủ ngẫu nhiên.....	83
Phân tích MANOVA 2 yếu tố tổng quát theo khối đầy đủ ngẫu nhiên.....	85
4.5. Phân tích hiệp phương sai (ANOCOVA)	87
ANOCOVA 1 yếu tố: Kiểu thí nghiệm hoàn toàn ngẫu nhiên.....	87
ANOCOVA 1 yếu tố: Kiểu thí nghiệm theo khối ngẫu nhiên	92
ANOCOVA 1 yếu tố: Kiểu thí nghiệm bình phương la tính.....	95
ANOCOVA 2 yếu tố	97
Phân tích đa hiệp phương sai (ANOMUCOVA)	99
ANOMUCOVA: Thí nghiệm 1 yếu tố được bố trí ngẫu nhiên hoàn toàn	100
ANOMUCOVA: Thí nghiệm 1 yếu tố được bố trí theo khối ngẫu nhiên.....	102
Chương 5. PHÂN TÍCH HỒI QUY TƯƠNG QUAN	105
5.1. Hồi quy tuyến tính đơn.....	105
5.2. Hồi quy đa biến	116
5.3. Thực hành phân tích hồi quy tương quan giữa 2 biến	130
5.4. Thực hành phân tích hồi quy đa tuyến tính	138
Nội dung phân tích hồi quy tuyến tính đa biến tổ	138
Thực hành phân tích hồi quy đa tuyến tính	143
Tương quan và hồi quy đa phi tuyến tính.....	146
5.5. Một số hàm mô tả quan hệ giữa các nhân tố điều tra lâm phần	147
Chương 6. QUAN HỆ GIỮA CÁC BIẾN SỐ CÓ THANG ĐO KHÁC NHAU	156
6.1. Kiểm định χ^2 (Chi – square).....	156
Kiểm định về tính độc lập của các đại lượng	156
Kiểm định ngang bằng về tỉ lệ	159
6.2. Quan hệ giữa hai biến định tính không có thứ bậc.....	161
6.3. Quan hệ giữa hai biến định tính có thứ bậc.....	168
6.4. Đo đặc tính thống nhất đối với một bảng R×R ô.....	172
6.5. Kiểm định trong nội bộ các lớp của 1 bảng nhiều chiều	172
Chương 7. SO SÁNH CÁC SỐ TRUNG BÌNH	177
7.1. Những thống kê mô tả trong một bảng chéo nhiều yếu tố	177
7.2. Phân tích ANOVA 1 yếu tố cùng với phép kiểm định tuyến tính.....	181
Chương 8. PHÂN TÍCH NHÓM.....	184
8.1. Giới thiệu chung	184
8.2. Phân tích mô hình 2 nhóm	188
8.3. Phân tích mô hình nhiều nhóm	194
Phân tích mô hình 4 nhóm.....	196
Phân tích mô hình 5 nhóm.....	203
8.4. Chọn những biến hữu ích nhất để phân tích tách biệt các nhóm.....	209
TÀI LIỆU THAM KHẢO	213
PHỤ LỤC	214

Chương I

NHẬP DỮ LIỆU VÀO BẢNG TÍNH SPSS

1.1. CÁC THANG ĐO

Để dễ dàng phân tích, so sánh và xử lý bằng phương pháp toán học, các dữ liệu về rừng phải được đo đạc, sắp xếp và trình bày theo một hệ thống nhất định. Các số liệu ghi lại thành Bảng được gọi là Bảng phân phối (phân bố). Để dễ dàng tính toán, phân tích và trình bày kết quả nghiên cứu, trước hết chúng ta xem xét những thuật ngữ quan trọng sau đây:

1. Các quan sát. Thuật ngữ này biểu thị các trường hợp¹ hay chủ thể. Một trường hợp bao gồm nhiều thông tin của một đại lượng nghiên cứu. Ví dụ: Đường kính thân cây ở vị trí 1.3 m ($D_{1.3}$, cm), chiều cao thân cây (H, m), trữ lượng rừng (M, m^3/ha), tình trạng sinh lực của cây...- đó là những đại lượng quan sát trong khi mô tả thảm thực vật rừng.

2. Các biến² hay tiêu thức. Thuật ngữ biểu thị các thông tin hay đặc điểm của từng trường hợp hay chủ thể. Ví dụ: Đường kính thân cây ở vị trí 1.3 m ($D_{1.3}$, cm), chiều cao thân cây (H, m), trữ lượng rừng (M, m^3/ha), tình trạng sinh lực của cây... Mỗi biến được đặt một tên biến³ cụ thể. Để mô tả tên đầy đủ của một biến, người ta dùng nhãn biến⁴. Mỗi giá trị của một biến cũng có thể được đặt một nhãn giá trị⁵. Ví dụ: Độ che phủ của thảm cỏ là một biến số. Giả sử biến này có tên là ĐCP với nhãn đầy đủ là “Độ che phủ của thảm cỏ”. Mỗi biểu hiện của biến ĐCP có thể được chỉ định bằng một giá trị biến như sau: 1 = Soc, 2 = Cop₃, 3 = Cop₂, 4 = Cop₁, 5 = Sol, 6 = Sp và 7 = Un. Việc quy định đặt tên biến, nhãn biến và nhãn giá trị như trên là tạo thuận lợi cho việc xử lý và tổng hợp số liệu bằng toán học.

3. Các thang đo. Những giá trị của biến được xác định bằng một thang đo nào đó. Có 4 loại thang đo được dùng phổ biến trong các nghiên cứu khoa học: định danh, thứ bậc, khoảng và tỉ lệ.

a. Thang đo định danh⁶. Thang đo định danh là thang đo được đánh số hay gán chuỗi dạng ngắn gọn cho những biểu hiện của một biến số. Vì thế người ta gọi biến số này là biến số định danh⁷. Những trị số của biến định danh chỉ thể hiện các nhóm phân loại không có thứ bậc hơn kém⁸. Nếu biến định danh được đo đạc bằng những con số thì giữa những con số này không có quan hệ hơn kém. Vì thế, ta không thể xử lý chúng bằng các phép tính đại số. Ví dụ: (1) Hai biểu hiện của biến độ bất gặp một loài cây nào đó thường được ghi bằng số như sau: 1 = có cây quan tâm và 0 = không có cây quan tâm; (2) hai biểu hiện của biến bệnh cây được đánh số là 1 = có bệnh và 2 = không có bệnh.

b. Thang đo thứ bậc⁹. Đó là một thang đo định danh, trong đó các trị số của biến số có thứ bậc hơn kém. Ví dụ 1: Khi điều tra một loại bệnh cây nào đó ta có thể ghi như sau: 1 = bệnh rất nặng, 2 = bệnh nặng, 3 = bệnh trung bình, 4 = bệnh nhẹ và 5 = không có bệnh. Ví dụ 2: Khi nghiên cứu quan hệ giữa năng suất rừng trồng với cấp đất, thì các biểu hiện của biến cấp đất có thể được ghi bằng các trị số như: 1 = cấp đất I, 2 = cấp đất II, 3 = cấp đất III, 4 = cấp đất IV và 5 = cấp đất V. Cần lưu ý rằng các con số có trị số cao hơn không có nghĩa là

¹ Case

² Variables

³ Variable name

⁴ Variable label

⁵ Value label

⁶ Nominal

⁷ Nominal variable

⁸ Unordered categories

⁹ Ordinal

chúng luôn ở bậc cao hơn và ngược lại. Việc gán các con số như vậy chỉ mang ý nghĩa quy ước nhằm tạo thuận lợi cho việc xử lý số liệu. Những biến số được đo bằng thang đo thứ bậc được gọi là biến định danh có thứ bậc. Trong lâm học, chúng ta thường gặp rất nhiều biến định danh có thứ bậc như cấp đất, cấp bệnh cây, cấp cháy rừng, cấp độ dày tầng đất, cấp lập địa, cấp sinh trưởng cây rừng... Trong các phép phân tích của SPSS và Stagraphics Plus, người ta gọi các biến định danh có thứ bậc là các biến phân loại hay biến lập nhóm¹.

c. Thang đo khoảng². Đó là thang đo thứ bậc có khoảng cách đều nhau. Thang đo này sử dụng cho những số liệu định lượng. Ví dụ: Đường kính thân cây được phân chia thành cấp 1, 2...k (cm); chiều cao thân cây được phân chia thành 2 m một cấp...

d. Thang đo tỉ lệ³. Đó là thang đo khoảng với một điểm Zero lấy làm gốc để so sánh. Thang đo này cũng sử dụng cho những số liệu định lượng. Ví dụ: Đo sự khác biệt giữa hai kiểu rừng bằng việc sử dụng trữ lượng gỗ thân cây để so sánh. Khác với thang đo khoảng, điểm zero của thang đo tỉ lệ là một số có thật.

Ngoài cách phân biệt các thang đo trên đây, trong lâm học người ta còn phân biệt các biến thành biến định tính và biến định lượng. Những đặc tính có thể đếm hay cân đo được theo một đơn vị vật lý nào đó được gọi là đặc tính định lượng (hay dấu hiệu về lượng). Ngược lại, những đặc tính không đo đếm được thì người ta gọi là đặc tính định tính (hay dấu hiệu về chất).

4. Đặc tính định tính. Trong lâm học, những dấu hiệu định tính có rất nhiều. Chúng đem lại ý nghĩa rất lớn nhằm thuyết minh cho vấn đề nghiên cứu. Chẳng hạn: Khi mô tả một quần hệ thực vật, một quần xã thực vật hoặc một ô tiêu chuẩn, chúng ta phải thu thập những dấu hiệu mà không thể đo đạc được như: (1) Màu sắc của lá, hoa và quả, (2) sự có mặt hay vắng mặt của loài cây quan tâm trên khu vực nghiên cứu, (3) tình trạng sinh lực của cá thể, (4) cây có bệnh hay không có bệnh, (5) cây đã chết hay còn sống... Ngay cả một số đặc tính định lượng, vì lý do khó đo đếm hay khó đến được để đo đếm, chúng ta cũng phải xem chúng như những đặc tính định tính. Ví dụ: Lượng hoa quả trên một cây gỗ, độ rậm rạp của tán lá, số lượng hay độ che phủ của thảm cỏ trên một diện tích...

Để dễ dàng đưa những dữ liệu định tính vào xử lý hay so sánh, chúng ta có thể ghi nhận bên cạnh chúng số lần hay tần số bất gặp. Tổng tần số của các sự kiện gặp được phải bằng tổng số cá thể (hay tổng số đơn vị) được quan sát. Ví dụ: Nếu a, b và c là tần số của 3 sự kiện nào đó và N là tổng số cá thể quan sát, thì $a + b + c = N$; còn a/N , b/N và c/N được gọi là tần suất của các nhóm. Tần suất của mỗi nhóm đối tượng có thể biểu thị bằng đơn vị phần 10 hoặc phần trăm. Trong trường hợp tần suất được biểu thị bằng đơn vị phần mười thì tổng các tần suất của các nhóm phải bằng 1, nghĩa là $a/N + b/N + c/N = 1$. Ngược lại, nếu chúng được biểu thị bằng đơn vị phần trăm thì tổng các tần suất của các nhóm phải bằng 100%.

Tùy theo mục đích nghiên cứu, những đặc tính định tính được ghi vào Bảng phân phối theo hai cách là Bảng phân phối một chiều (**Bảng 1.1**) và Bảng phân phối hai chiều (**Bảng 1.2**). Từ đó, nhà lâm học có thể tiến hành các xử lý và so sánh kết quả.

Bảng 1.1. Phân phối kết quả điều trị bệnh A trên 200 cây gỗ

	Tần số	Tần suất	Tỉ lệ %
	(1)	(2)	(3)
Bệnh khỏi	105	0,525	52,5
Bệnh giảm	55	0,275	27,5
Tử vong	40	0,200	20,0
Tổng	200	1.000	100.0

¹ Ordered categorical Variable, Controlling Variable, Grouping Variable

² Interval

³ Ratio

Bảng 1.2. Phân phối kết quả điều trị của 3 loại thuốc trên 205 cây

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

5. Đặc tính định lượng. Người ta phân biệt đặc tính định lượng theo hai loại: rời rạc và liên tục.

a. Đặc tính định lượng rời rạc. Đó là những đại lượng của một đặc tính được biểu thị bằng những con số nguyên. Thí dụ: Số cây trên một ô mẫu có kích thước nhất định, số cây gỗ mang bệnh, số cây tái sinh trên một đơn vị diện tích... Trong Bảng thống kê, chúng ta ghi trong một cột (hoặc một hàng) các trị số 0, 1, 2...n của đặc tính, còn cột bên cạnh (hoặc hàng dưới) ghi nhận các tần số, tần suất và tỉ lệ phần trăm tương ứng với chúng (**Bảng 1.3**).

Bảng 1.3. Phân bố cây tái sinh Dầu song nang trên mặt đất qua điều tra 576 ô dạng bản 1 m²

Số cây/ô	Tần số	Tần suất	Tỉ lệ %
(1)	(2)	(3)	(4)
0	350	0,6076	60,76
1	160	0,2778	27,78
2	28	0,0486	4,86
3	16	0,0278	2,78
4	14	0,0243	2,43
≥ 5	8	0,0139	1,39
Tổng	576	1,0000	100,00

Bảng 1.4. Phân bố cây tái sinh Dầu song nang trên mặt đất qua điều tra 576 ô dạng bản 1 m²

Số cây/ô	Tần số đôn	Tần suất đôn	Tỉ lệ đôn
(1)	(2)	(3)	(4)
0	350	0,6076	60,76
1	510	0,8854	88,54
2	538	0,9340	93,40
3	554	0,9618	96,18
4	568	0,9861	98,61
≥ 5	576	1,0000	100,00

Chúng ta cũng có thể lập Bảng 1.4 như sau:

Số cây/ô	Tần số đôn	Tần suất đôn	Tỉ lệ đôn
(1)	(2)	(3)	(4)
0	576	1,0000	100,00

1	226	0,3924	39,24
2	66	0,1146	11,46
3	38	0,0660	6,60
4	22	0,0382	3,82
≥ 5	8	0,0139	1,39

Cần lưu ý rằng, trong một vài trường hợp, người ta còn tính các tần số dồn, tần suất dồn và tỉ lệ dồn bằng cách cộng dồn tần số (hay tần suất, tỉ lệ) của nhóm này với những nhóm trên hoặc dưới (**Bảng 1.4**). Sự khác nhau của hai cách trình bày này là ở chỗ, cách thứ nhất cho ta biết có bao nhiêu ô tương ứng với 0, 1...n cây cụ thể, còn cách thứ 2 cho biết có bao nhiêu ô ít hơn (hay nhiều hơn) n cây/ô.

b. Đặc tính định lượng liên tục. Những đặc tính định lượng mà chúng có thể lấy bất kỳ giá trị nào nằm trong một khoảng nào đó được gọi là đặc tính định lượng liên tục. Ví dụ: Những đặc tính như chiều cao và đường kính thân cây, trữ lượng và diện ngang thân cây... Tuy nhiên, trong thực tế, do những giới hạn của các công cụ đo đạc, nên chúng ta vẫn chỉ nhận được các giá trị không liên tục. Ví dụ: Vì thước đo chiều cao Blume -Leiss có độ chính xác là 0,5 m, nên trong khoảng chiều cao cây gỗ từ 20 đến 22 m chúng ta sẽ nhận được 3 giá trị là 20,5; 21 và 21,5 m.

Khi trình bày các số liệu định lượng liên tục dưới dạng Bảng, người ta phải thực hiện việc phân nhóm (chia tổ hay cấp) theo một cự ly (hay khoảng cách) nhất định và tính tần số, tần suất và tỉ lệ của các nhóm. Ví dụ: Theo Brooks và Carruther, số nhóm (m) và cự ly giữa các nhóm (k) tính theo công thức

$$m \geq 5 \log N, \quad (1.1)$$

$$k = \frac{(X_{\max} - X_{\min})}{m} \quad (1.2)$$

trong đó N là dung lượng quan sát, $X_{\max}$ và $X_{\min}$ là những trị số lớn nhất và nhỏ nhất của dãy số liệu quan sát. Hoppman tính cự ly giữa các nhóm hoặc tổ theo công thức

$$d = \frac{(X_{\max} - X_{\min})}{2 \sqrt[3]{N}} \quad (1.3)$$

trong đó N là dung lượng quan sát. Cần nhận thấy rằng, mặc dù số tổ và cự ly giữa các tổ được tính theo công thức **1.1**, **1.2** và **1.3** là khách quan, nhưng thực tế có thể nhận được kết quả kém ý nghĩa. Ví dụ: Cự ly giữa các lớp chiều cao của quần xã thực vật được tính theo công thức **1.2** và **1.3** có thể là 0,2 – 0,5m, nhưng thực tế ta dùng dụng cụ đo cao có độ chính xác là 1,0 m. Do đó, ta không thể tính được chính xác tần số tương ứng từng cấp. Vì vậy, tùy theo đối tượng và mục tiêu nghiên cứu, trong thực hành người ta thấy số tổ được phân chia nằm trong khoảng từ 8 - 12 là phù hợp, còn cự ly giữa các tổ có thể tính dựa trên độ chính xác của dụng cụ đo đạc. Ví dụ: Khi phân cấp chiều cao ở rừng trồng đồng tuổi thuần loài, người ta lấy cự ly giữa các lớp chiều cao là 1 - 2 m, còn ở rừng tự nhiên từ 2 - 4 m. Đối với đường kính thân cây ở rừng trồng đồng tuổi thuần loài, người ta thường lấy cự ly tổ là 1 cm khi đa số cây gỗ có $D_{1.3} \leq 10$ cm, 2 cm khi đa số cây gỗ có $D_{1.3} \leq 20$ cm. Ở rừng tự nhiên khác tuổi tuyệt đối, cự ly giữa các cấp kính biến động từ 4 - 12 cm. Khi tổng hợp số liệu thành Bảng số liệu, chúng ta có thể ghi vào một cột nào đó trị trung tâm (hoặc ranh giới) của mỗi nhóm và các cột bên cạnh tương ứng là tần số, tần suất và tỉ lệ (cũng có thể gồm cả tần số dồn và tần suất dồn...) của mỗi nhóm. Ví dụ: Số liệu về phân bố số cây theo cấp đường kính (N - D) và cấp chiều cao (N - H) của một quần thụ có thể được ghi như sau (**Bảng 1.4** và **1.5**).

Bảng 1.4. Phân bố số cây theo cấp đường kính

D, cm	10	14	18	22	26	30	34	38	42	≥ 46	Σ
N, c/ha	87	50	46	28	18	11	7	5	5	8	265
N%	32,8	18,9	17,4	10,6	6,8	4,2	2,6	1,9	1,9	3,0	100
N% _{LT}	32,8	51,7	69,1	79,6	86,4	90,6	93,2	95,1	97,0	100	

Bảng 1.5. Phân bố số cây theo cấp chiều cao

H, m	11	15	19	23	27	≥ 28	Σ
N, cây/ha	36	64	132	84	34	15	365
N%	9,9	17,5	36,2	23,0	9,3	4,1	100
N% _{LT}	9,9	27,4	63,6	86,6	95,9	100	

Từ **Bảng 1.4** có thể thấy, số cây thuộc cấp kính 10 và 14 cm tương ứng là 32,8 và 18,9%, còn những cây có $D < 14$ cm là 57,7%.

1.2. CÁCH NHẬP DỮ LIỆU, MÃ HOÁ BIẾN VÀ TÊN BIẾN

1. Nhập dữ liệu vào Bảng tính. Sau khi mở một Bảng tính mới của SPSS¹ (**Hình 1.1**), bạn nhấp chuột vào ô trên cùng của một cột nào đó thuộc Bảng tính và nhập vào số liệu. Sau khi số liệu được nhập vào ô đầu tiên và nhấp Enter, kí tự đầu cột có tên là Var sẽ chuyển thành Var0001. Đây là tên mặc định của biến thứ nhất. Sau khi vào xong số liệu ở cột 1, bạn chuyển qua cột 2, 3...; tại cột 2, 3...bạn cũng vào số liệu như thông thường. Trang này được gọi là trang nhập dữ liệu và quan sát dữ liệu (Data View). Lưu ý rằng, để nhập dữ liệu vào SPSS nhanh và thuận tiện hơn, bạn cũng có thể nhập dữ liệu vào một Bảng tính Excel; sau đó Copy dữ liệu này và dán vào Bảng tính SPSS.

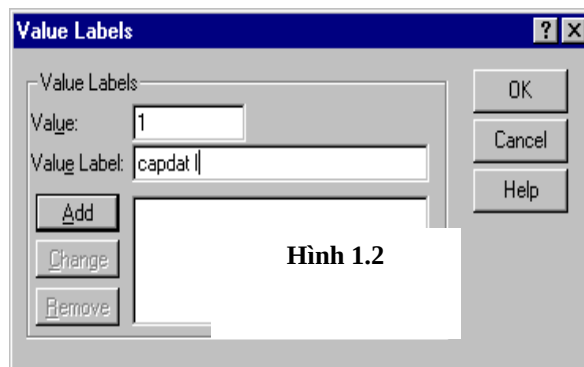
2. Mã hóa biến và đặt tên đầy đủ cho biến. Khi chưa mã hóa biến thì tên mặc định của mỗi biến là Var0001, Var0002, Var0003... Muốn điều chỉnh các dữ liệu và đổi tên biến mặc định bằng tên biến do bạn tự đặt, bạn hãy thực hiện như sau đây:

+ **Điều chỉnh các dữ liệu và đổi tên biến.** Trước hết, bạn hãy nhấp vào trang Variable View nằm ở bên cạnh trang Data View. Sau đó bạn điều chỉnh các dữ liệu như số chữ số thập phân (Decimals), độ rộng cột (Width), kiểu biến số (dạng số - Numeric và dạng chuỗi kí tự - String)... Số chữ số thập phân được điều chỉnh bằng cách nhấp chuột vào Cell Decimals và điều chỉnh theo mũi tên lên hoặc xuống. Khi bạn vào số liệu đường kính (D) và chiều cao (H) thân cây ở hai cột liên tiếp, chương trình tự động gán cho hai biến này hai tên mặc định là Var0001, Var0002. Để dễ dàng nhận biết hai biến Var0001 và Var0002 trên Bảng tính, bạn có thể đổi tên biến Var0001 bằng D và biến Var0002 bằng H như sau: (1) chọn trang Variable view, (2) nhấp chuột vào Cell đầu tiên của dòng đầu ở cột Name và đánh chữ D; sau đó Enter và đánh chữ H vào Cell thứ nhất của dòng hai. Khi trở lại trang Data view, bạn sẽ thấy hai biến D và H xuất hiện tương ứng trên đầu cột 1 và 2. Bạn có thể áp dụng thao tác này cho những biến khác. Lưu ý rằng, mỗi dòng ở trang Variable view tương ứng với một cột ở trang Data view.

+ **Mã hóa tên biến bằng tên đầy đủ.** Những tên biến ngắn gọn (ví dụ chữ $D_{1,3}$ là tên ngắn gọn của biến số đường kính thân cây ngang ngực – $D_{1,3}$, cm) xuất hiện trên đầu mỗi cột

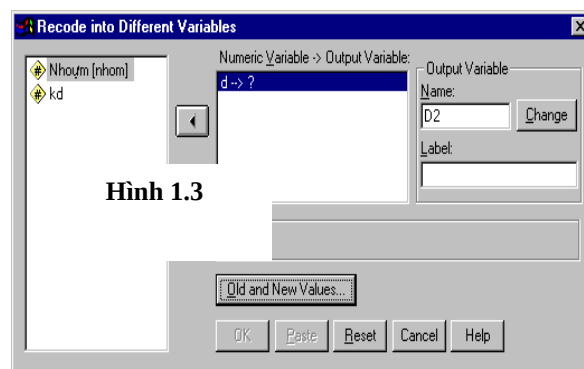
¹ SPSS = Statistical Products for the Social Services

chỉ có ý nghĩa giúp bạn quan sát và phân biệt vị trí của các biến số tương ứng với từng cột của Bảng tính. Những tên biến ngắn gọn này không được ghi ra trang kết quả. Để chương trình ghi các tên biến này ra trang kết quả tính toán bằng một tên đầy đủ, bạn cần mã hóa lại tên biến. Để đạt mục đích này, bạn nhấp chuột vào cột Label của trang Variable view; tiếp theo ghi tên đầy đủ cho biến số tại Cell tương ứng với cột nhập dữ liệu ở trang Data view. Lưu ý rằng, để dễ quan sát tên đầy đủ của biến trong Bảng tính SPSS, bạn có thể ghi tên biến không có dấu tiếng Việt. Nhưng nếu muốn sau này khỏi phải đánh lại tên biến đầy đủ thì bạn có thể ghi tên biến đầy đủ bằng tiếng Việt.



Hình 1.2

+ **Đặt nhãn cho một số giá trị.** Một số biến định tính có thể được mã hóa bằng những kí tự hoặc số. Ví dụ: Bạn có thể mã hóa biến cấp đất I bằng số 1; biến cấp đất II bằng số 2... Muốn mã hóa các giá trị này bằng một số hay kí tự nào đó, bạn nhấp vào cột Values của trang Variable view. Khi đó một cửa sổ Value labels xuất hiện (Hình 1.2). Để mã hóa số 1 cho cấp đất I, trước hết bạn đặt con chuột vào ô Value và đánh số 1. Tiếp theo, bạn đặt con chuột vào ô Value labels và đánh chữ “Cấp đất I”. Sau cùng bạn nhấp nút Add để máy ghi nhận các thay đổi này. Trình tự đặt nhãn cho các giá trị khác cũng được thực hiện tương tự như trên. Lưu ý rằng, sau mỗi lần đặt nhãn cho một số thì bạn phải nhấp nút Add. Nếu cần thay đổi nhãn cho một số nào đó thì bạn nhấp nút Change hay Remove. Nếu có giá trị khuyết thiếu (tức một mẫu nào đó thiếu số liệu) thì bạn nhấp nút Define missing variable và ghi số -1 hay 9 (số mặc định).



Hình 1.3

+ **Mã hóa lại biến số.** Mã hóa lại một số biến là một thủ tục rất hữu ích cho những trường hợp mà bạn mong muốn ghép nhóm các dữ liệu theo một số đặc trưng nào đó. Chẳng hạn: (1) Bạn muốn ghép tập hợp những cây có $D < 20$ cm vào nhóm cây nuôi dưỡng – kí hiệu bằng số 1, còn những cây có $D \geq 20$ vào nhóm khai thác – kí hiệu bằng số 2; (2) Bạn muốn nhóm những đặc trưng của quần thể theo từng loài, nhóm loài, nhóm gỗ, nhóm cấp sinh trưởng của Kraft, nhóm lập địa, nhóm cây theo cấp kính và cấp chiều cao...

Thực hành 1. Giả sử bạn muốn tập hợp số cây theo những nhóm $D_{1.3}$. Để đạt được mục đích ấy, bạn hãy thực hiện theo thủ tục sau đây :

+ Chọn Transform $\Rightarrow$ Recode $\Rightarrow$ Into Different Variables (đến đây chuyển biến D vào ô Numeric Variables $\rightarrow$ output variable (Hình 1.3);

+ Tiếp theo bạn ghi tên biến mới (ví dụ D2) vào ô Name;

+ Sau đó bạn chọn tiếp Old And New Values. Đến đây, nếu bạn chọn:

× Range Lowest through và ghi số 20 vào ô trống bên cạnh, rồi ghi số 1 vào ô value của khoảng New value, sau đó chọn Add;

× Range through highest và ghi số 20 vào ô bên cạnh, rồi ghi số 2 vào ô value của khoảng New value, sau đó chọn Add;

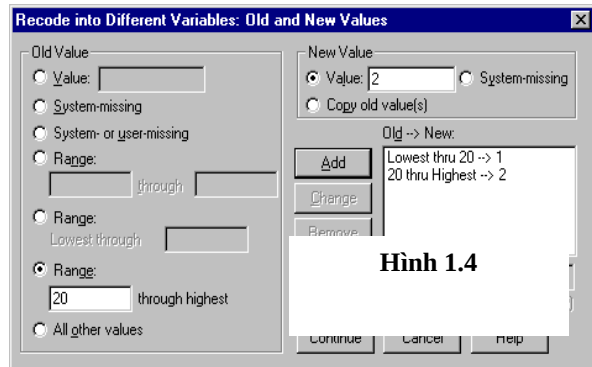
Cuối cùng bạn chọn Continue thì chương trình sẽ mã hóa những cây có $D < 20$ bằng số 1 và những cây có $D \geq 20$ bằng số 2 (Hình 1.4).

Sau khi mã hóa, bạn có thể gọi chương trình để tính toán các đặc trưng quần thụ theo các biến ghép nhóm.

Thực hành 2. Lập một Bảng phân phối đường kính thân cây. Giả sử một quần thụ có dãy $D_{1,3}$ phân bố từ $7 \leq D_{1,3} \leq 25$ cm. Bạn mong muốn lập một Bảng phân bố đường kính thân cây với mỗi cấp cách nhau 2 cm; trong đó cấp kính thứ nhất có trị số giữa cấp là 8 cm, còn cấp kính cuối cùng có trị số giữa cấp là 20 cm. Để đạt được mục đích đặt ra, bạn hãy thực hiện như sau :

× Chọn Range $\Rightarrow$ Lowest through. Kế đến ghi số 9 vào ô trắng bên cạnh và số 8 vào ô value của khoảng New value. Sau khi chọn Add thì chương trình sẽ mã hóa tất cả những cây có $D < 9$ cm vào cấp kính 8 cm.

× Chọn Range $\square$ (ghi số 9) Through $\square$ (ghi số 11), rồi ghi số 10 vào ô value của khoảng New value; sau đó chọn Add. Thủ tục này nhằm ghép những cây có giá trị $9 < D < 11$ vào cấp kính 10. Việc mã hóa cho những cấp kính tiếp theo (12, 14 ...18) cũng được thao tác tương tự như trên.



Hình 1.4

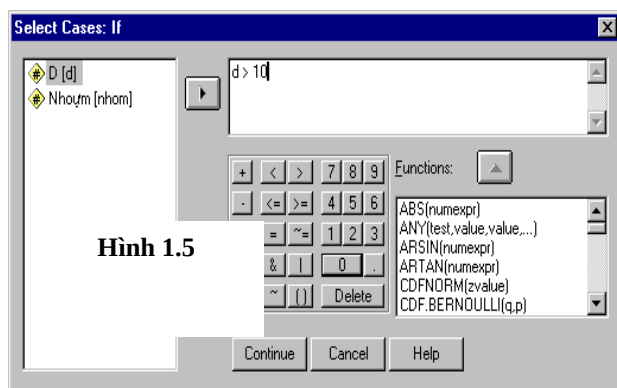
× Cuối cùng, chọn Range $\Rightarrow$ through highest rồi ghi số 19 vào ô trắng bên cạnh và số 20 vào ô value của khoảng New value; sau đó chọn Add. Thao tác này yêu cầu chương trình mã hóa tất cả những cây có $D > 19$ cm vào cấp kính 20 cm.

Sau khi hoàn thành việc mã hóa, bạn chọn Continue để chương trình thực hiện mã hóa từng cây vào những cấp kính đã định. Những trị số mã hóa sẽ được ghi lại thành một biến. Với cách làm tương tự như trên, bạn có thể mã hóa lại cho bất kỳ thông tin gì mà bạn mong muốn.

1.3. CHỌN LỌC VÀ BIẾN ĐỔI DỮ LIỆU

1. Chọn một hay nhiều tập con từ một tập lớn. Trước khi thực hiện thủ tục này, bạn cần lưu ý các toán tử sau đây: $>$, $<$, $\equiv$ (không bằng), $\neq$, $\leq$, $\geq$, $\&$ (và), $|$ (hay là, hoặc là). Giả thiết bạn chỉ mong muốn tính toán trên một bộ phận (tập con) nào đó của một tập dữ liệu lớn. Để đạt được mục đích này, bạn có thể thực hiện như sau :

+ Trước hết mở Data $\rightarrow$ Select Cases $\rightarrow$ If condition is satisfied $\rightarrow$ If. Đến đây bạn chuyển tên của biến cần lọc vào hộp thoại trống bên phải bằng cách nhấp nút mũi tên bên cạnh.



Hình 1.5

+ Tiếp theo bạn chọn các toán tử từ ô vuông bên dưới; sau đó ghi số cần lọc rồi nhấp Continue và nhấp OK (Hình 1.5).

Thực hành lọc số liệu. 1. Nếu bạn ghi $D \geq 10$ & $D \leq 30$ thì chương trình chỉ chọn những biến D nằm trong khoảng từ 10 đến 30 cm để đưa vào tính toán. Lưu ý rằng, bạn không thể chọn cùng lúc $D = 10$ & $D = 12$. Bởi vì toán tử & chỉ đúng khi hai biến khác nhau.

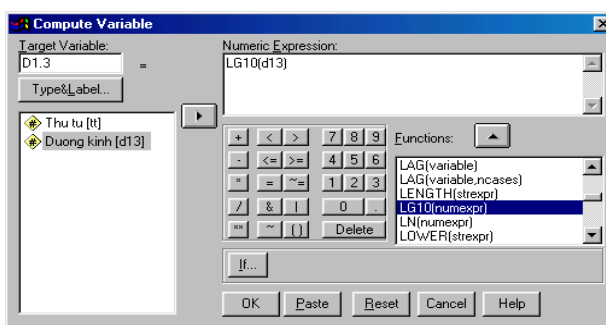
2. Nếu bạn ghi $D > 10$ & $H > 8$ thì chương trình chỉ chọn những biến thỏa điều kiện $D > 10$ cm nhưng $H > 8$ m.

3. Nếu bạn ghi $D \cong 10$ thì chương trình chỉ chọn những biến D khác 10.

4. Nếu bạn ghi $D = 8 \mid D = 10$ thì chương trình sẽ chọn những biến $D = 8$ và $D = 10$.

5. Nếu bạn chọn $D = 12$ & $H = 12.5$ & Nhóm gỗ = II thì chương trình sẽ chọn những biến $D = 12$ nhưng H phải là 12.5 và thuộc nhóm gỗ = II.

2. Biến đổi dữ liệu. Thủ tục này được đặt ra khi bạn cần phải biến đổi số liệu trước khi thực hiện những tính toán nào đó. Nhưng muốn biến đổi số liệu và tạo ra một biến mới trước khi tính toán, bạn cần lưu ý đến những kí hiệu biểu thị cho các phép tính như cộng (+), trừ (-), nhân ($\times$), chia (/), lũy thừa bậc k ($\times k$), căn bậc 2 (SQRT), logarit tự nhiên (Ln), logarit cơ số 10 (Lg10)...Việc biến đổi số liệu được thực hiện theo trình tự sau đây :



Hình 1.6. Cửa sổ biến đổi số liệu

+ Chọn Transform → Compute

→ Reset → Target Variable (đến đây

bạn ghi biến đích) → Numeric Expression (biểu thức số).

+ Tiếp theo bạn chọn hàm số từ ô Functions và nhấp mũi tên bên trên để đưa nó vào ô Numeric Expression.

+ Khi hàm số xuất hiện trong ô Numeric Expression, bạn đưa biến số cần biến đổi vào hàm (nơi có dấu ?) bằng cách chọn biến số trong ô Type & Label và nhấp mũi tên bên phải.

+ Sau cùng thì bạn chọn OK (Hình 1.6). Đến đây kết quả biến đổi số liệu sẽ được ghi vào một cột bên cạnh của Bảng tính SPSS. Nếu cần tính toán những thống kê nào đó (thống kê mô tả, tương quan...) thì bạn đưa dữ liệu vừa biến đổi vào ô biến phụ thuộc. Lưu ý rằng những thao tác trên đây cũng có thể được thực hiện trong lúc bạn đang xử lý số liệu.

1.4. IN ÁN VÀ COPY SỐ LIỆU

Sau khi tính toán xong, bạn có thể in các kết quả tính toán bằng cách chọn các Bảng và chọn Print từ Menu File. Nếu muốn copy các kết quả từ Bảng SPSS 10.0 sang Microsoft Word để tiện cho việc chỉnh sửa, thì bạn có thể thực hiện như sau :

- Chọn Bảng tính cần Copy bằng cách nhấp chuột vào Bảng tính hoặc vào tiêu đề của Bảng nằm ở bên trái cửa sổ kết quả.

- Mở Edit và chọn Copy.

- Mở Microsoft Word rồi chọn Paste.

- Sau cùng bạn chỉnh sửa Bảng kết quả theo mục đích của mình.

Cần lưu ý rằng, khi nhập số liệu vào Bảng tính SPSS thì bạn nên chọn Font VNI Times hoặc VNI Time và nhập các kí tự theo chữ tiếng Việt. Điều này có thể gây ra cho bạn chút ít khó khăn khi quan sát những tiêu đề của Bảng kết quả. Nhưng chúng sẽ giúp bạn giảm thiểu việc chỉnh sửa Font chữ tiếng Việt trong Bảng kết quả. Nói chung, những kết quả nhận được từ SPSS 10.0 được ghi lại thành các Bảng rất đẹp và rất khoa học. Tuy vậy, bạn vẫn cần phải chỉnh sửa chúng ít nhiều. Thủ tục chỉnh sửa các Bảng số liệu cũng được thực hiện theo cách thông thường như trong Microsoft Word và Microsoft Excel.

Chương II

THỐNG KÊ MÔ TẢ, KIỂM ĐỊNH PHÂN BỐ CHUẨN VÀ KIỂM ĐỊNH NGANG BẰNG PHƯƠNG SAI

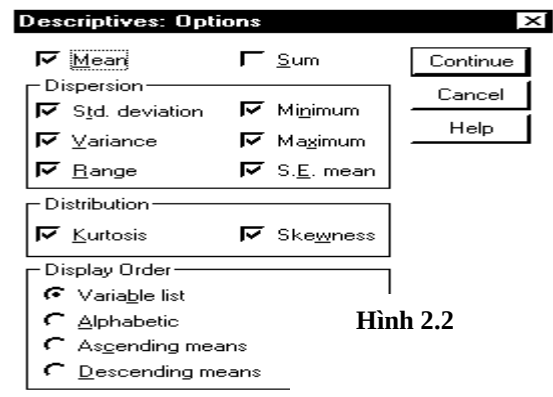
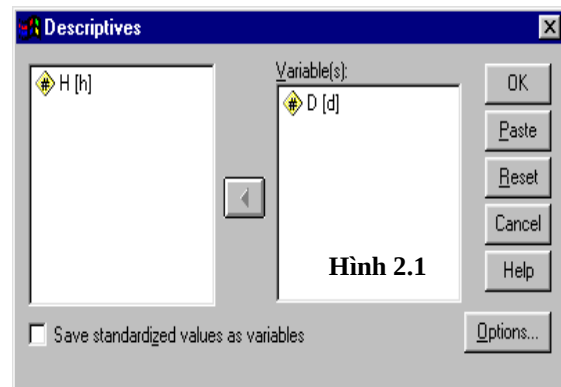
2.1. SỬ DỤNG THỦ TỤC DESCRIPTIVES ĐỂ THU ĐƯỢC THỐNG KÊ MÔ TẢ

Thủ tục thống kê mô tả (Descriptive Statistics) giúp bạn xác định những đặc trưng thống kê như giá trị trung bình, phương sai, phạm vi biến động, trị lớn nhất và nhỏ nhất, độ lệch, độ nhọn, khoảng tin cậy của trung bình mẫu... của một hoặc nhiều mẫu ngẫu nhiên. Giả thiết bạn cần biết những đặc trưng thống kê của đại lượng đường kính ngang ngực ($D_{1.3}$, cm) trong một quần thụ. Để nhận được những thống kê mà bạn quan tâm, trong thực hành bạn có thể thực hiện các bước như sau :

Bước 1. Nhập số liệu. Số liệu của một hoặc nhiều mẫu ngẫu nhiên có thể được nhập trực tiếp vào Bảng tính SPSS hoặc nhập vào Bảng tính Excel rồi Copy sang Bảng tính SPSS.

Bước 2. Gọi chương trình tính toán. Từ Menu chính, bạn mở Analyze $\Rightarrow$ Descriptive Statistic $\Rightarrow$ Descriptives. Khi cửa sổ Descriptives xuất hiện, bạn hãy chuyển tất cả các biến cần tính vào ô Variables (Hình 2.1).

Bước 3. Từ cửa sổ Descriptives, bạn chọn Options. Khi cửa sổ Options xuất hiện, bạn hãy chọn những thống kê cần quan tâm (Hình 2.2). Sau cùng thì nhấp OK để chương trình tính toán. Đến đây chương trình sẽ hiện thị những thống kê mô tả theo chỉ định của bạn (Bảng 2.1).



2.2. SỬ DỤNG THỦ TỤC EXPLORE ĐỂ THU ĐƯỢC CÁC THỐNG KÊ MÔ TẢ VÀ KIỂM ĐỊNH PHÂN BỐ CHUẨN

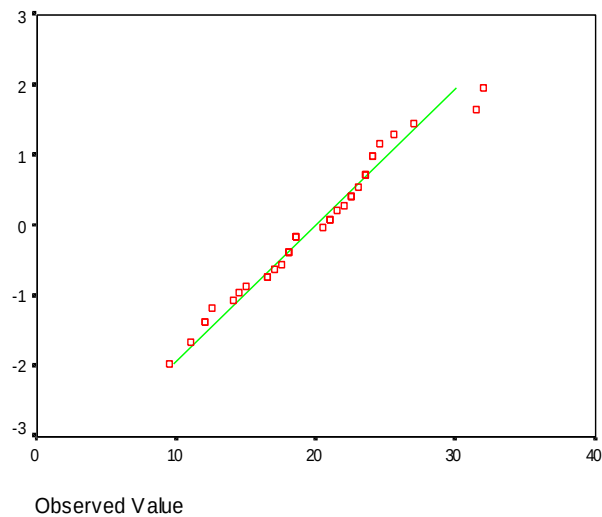
Thủ tục Explore cho phép bạn không chỉ xác định các thống kê mô tả (trung bình, phương sai, độ lệch, độ nhọn...), khoảng tin cậy C% (95, 99%...) cho số trung bình, trung bình bị tía bớt 5% dữ liệu (5% Trimmed Mean), trung vị (Median) và khoảng tứ phân vị (Interquartile Range) (Bảng 2.2) mà còn các ước lượng số trung bình M – Estimators (Bảng 2.3), các phân vị (Percentiles, Bảng 2.4), các giá trị cực trị (Extreme Values, Bảng 2.5), kiểm định tính chuẩn (Tests of Normality, Bảng 2.6) và đồ thị kiểm định phân bố chuẩn (Normal Q-Q plot, Hình 2.3)... Để nhận được các thống kê trên đây, bạn có thể thực hiện theo trình tự sau đây :

Bước 1. Từ Menu chính, mở Analyze $\Rightarrow$ Descriptive Statistic $\Rightarrow$ Explore. Đến đây bạn chọn biến cần xử lý và chuyển nó vào ô Dependent list (Hình 2.4).

Bảng 2.1. Các thống kê mô tả (Descriptive Statistics)

		D1	D2	D3	D4	Valid N (listwise)
N	Statistic	40	40	40	40	40
Range	Statistic	22.5	21.0	25.5	23.0	
Minimum	Statistic	9.5	11.0	6.0	7.5	
Maximum	Statistic	32.0	32.0	31.5	30.5	
Mean	Statistic	19.888	20.000	20.450	20.875	
	Std. Error	.810	.889	.898	1.012	
Std. Deviation	Statistic	5.121	5.624	5.677	6.401	
Variance	Statistic	26.224	31.628	32.228	40.971	
Skewness	Statistic	.115	.146	-.467	-.482	
	Std. Error	.374	.374	.374	.374	
Kurtosis	Statistic	.098	-.904	-.187	-.498	
	Std. Error	.733	.733	.733	.733	

Normal Q-Q Plot of D1



Bước 2. Từ **Hình 2.3.** Đồ thị kiểm định phân phối chuẩn, chọn các ô Descriptives, Mean Estimators, Outliers, và chọn Continue rồi chọn OK để chương trình tính toán và hiện kết quả.

Bảng 2.2. Thống kê mô tả (Descriptives)

		Statistic	Std. Error
D1	Mean	19.888	.810
	95% Confidence Interval for Mean	Lower Bound	
		Upper Bound	
	5% Trimmed Mean	21.525	
	Median	19.764	
	Variance	20.750	
	Std. Deviation	26.224	
	Minimum	5.121	
	Maximum	9.5	
	Range	32.0	
	Interquartile Range	22.5	
	Skewness	6.875	
	Kurtosis	.115	.374
		.098	.733

Giải thích và nhận xét kết quả. Khi phân bố của các đại lượng không tuân theo phân bố chuẩn thì việc ước lượng số trung bình được thực hiện theo thủ tục M – Estimators. Thủ tục này cung cấp 4 ước lượng là Huber, Tukeys's Beweight, Hampel và Andrew. Các số trung bình của Huber và số trung bình bị tĩa bớt 5% là các ước lượng gần đúng cho trung bình tổng thể. Trung bình bị tĩa bớt 5% được tính bằng cách : Trước hết sắp xếp các số liệu từ nhỏ nhất đến lớn nhất; tiếp theo bỏ bớt 5% số liệu ở mỗi phía của dãy số; sau đó tính số trung bình trên những số liệu còn lại.

+ **Bảng 2.5** đưa ra 10 giá trị cực trị (Extreme Values), bao gồm 5 số bé nhất và 5 số lớn nhất của dãy số liệu.

+ **Bảng 2.6** báo cáo các kiểm định về tính chuẩn (Test of normality) của dãy số liệu bằng kiểm định Kolmogorov-Smirnov và Shapiro-Wilk. Khi $P < 0.000$ (giá trị làm tròn của số 0.0005) thì giả thuyết về phân bố chuẩn bị bác bỏ. Ngược lại, khi $P > 0.0005$ thì giả thuyết về phân bố chuẩn được chấp nhận.

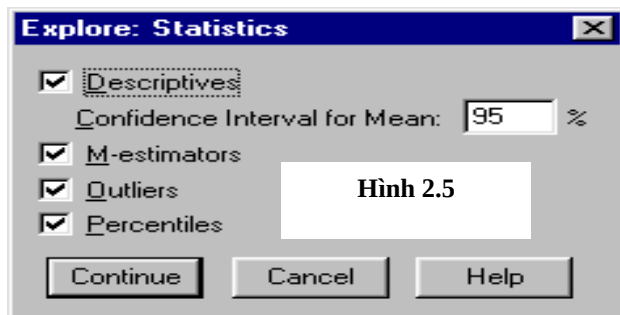
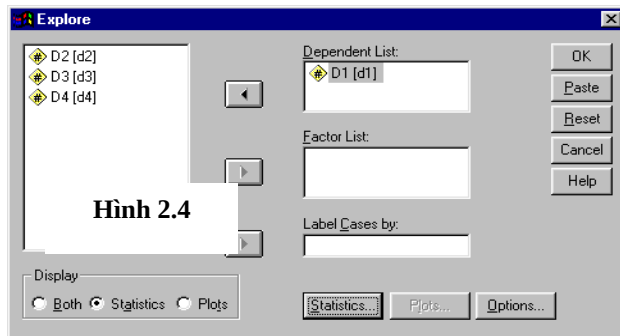
+ Cùng với các thống kê được tính toán, chương trình còn hiển thị 5 đồ thị là Histogram, Stem and leaf plot (đồ thị cành và lá), đồ thị xác suất chuẩn (Normal Q-Q plot), đồ thị xác suất chuẩn đã khử xu hướng (Detrended normal q – q plot) và đồ thị hộp (Boxplot). Các đồ thị này cũng giúp bạn nhận định sơ bộ về dãy số liệu có tuân theo luật phân bố chuẩn hay không. Đồ thị xác suất chuẩn (Normal Q-Q plot) ở **Hình 2.3** được vẽ với trục hoành là trị số quan sát, còn trục tung là trị số chuẩn hoá kỳ vọng. Khi các giao điểm giữa các trị số quan sát và các trị số chuẩn hoá kỳ vọng cùng nằm trên một đường thẳng thì phân bố của dãy số liệu là tiệm cận chuẩn.

Bảng 2.3.. Các ước lượng số trung bình (M - Estimators)

Huber's M-Estimator	Tukey's Biweight	Hampel's M-Estimator	Andrews' Wave
20.126 ^a	20.038 ^b	19.977 ^c	20.010 ^d

Bảng 2.4. Các phân vị (Percentiles)

Percentiles	Weighted Average	Tukey's Hinges
	D1	D1
5	11.050	
10	12.050	
25	16.625	16.750
50	20.750	20.750
75	23.500	23.500
90	25.400	
95	31.275	



+ Khi ước lượng khoảng tin cậy cho số trung bình và phương sai, bạn cũng có thể thay đổi mức tin cậy khác nhau (ví dụ: 90, 95, 99%); trong đó mặc định là mức tin cậy 95%. Việc chọn mức tin cậy nào là tùy thuộc vào yêu cầu của cuộc nghiên cứu.

+ Bạn cũng có thể sử dụng thủ tục Explore để tính toán các thống kê cho từng nhóm đối tượng. Ví dụ, bạn có thể tính toán các thống kê cho toàn bộ quần thụ và cho từng nhóm gỗ hay cấp sinh trưởng cây rừng. Để đạt được mục đích đặt ra, bạn hãy mã hoá các nhóm đối tượng bằng các con số hay các kí tự. Sau đó bạn chuyển các biến số cần tính vào ô Dependent variable, còn biến lập nhóm được đưa vào Factor List. Cuối cùng bạn chọn các thống kê mô tả rồi chọn OK. Đến đây kết quả cần tính toán sẽ hiện thị ra màn Hình.

Bảng 2.5. Các giá trị cực trị (Extreme Values)

			Case Number	Value
D1	Highest	1	20	32.0
		2	13	31.5
		3	14	27.0
		4	9	25.5
		5	28	24.5
	Lowest	1	10	9.5
		2	11	11.0
		3	16	12.0
		4	12	12.0
		5	21	12.5

Bảng 2.6. Kiểm định phân phối chuẩn (Tests of Normality)

		D1
Kolmogorov-Smirnov	Statistic	.086
	df	40
	Sig.	.200
Shapiro-Wilk	Statistic	.970
	df	40
	Sig.	.466

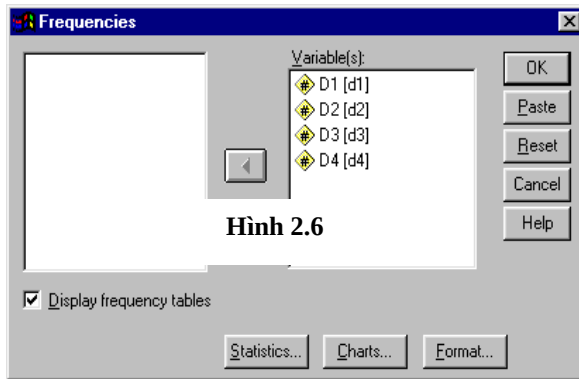
2.3. SỬ DỤNG THỦ TỤC FREQUENCIES ĐỂ THU ĐƯỢC CÁC THỐNG KÊ MÔ TẢ VÀ KIỂM ĐỊNH PHÂN BỐ CHUẨN

Thủ tục Frequencies cũng cho phép bạn xác định những thống kê mô tả khuynh hướng trung tâm (trung bình, trung vị, mode...), các đặc trưng phân tán (phương sai, sai tiêu chuẩn, phạm vi biến động...), các đặc trưng phân bố (độ lệch, độ nhọn), các tần số, các phân vị (Percentiles), giản đồ cột cùng với một đường cong chuẩn... Để thu được các thống kê mô tả, bạn có thể thực hiện theo trình tự sau đây :

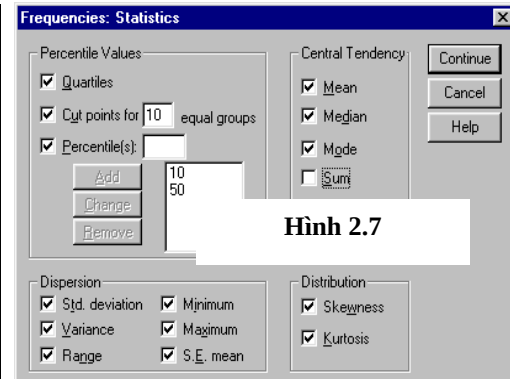
Bước 1. Từ Menu chính, mở Analyze $\Rightarrow$ Descriptive Statistic $\Rightarrow$ Frequencies. Đến đây bạn chọn những biến số cần xử lý và chuyển chúng vào ô Variables (Hình 2.6).

Bước 2. Từ cửa sổ Frequencies, bạn vào Statistics và đánh dấu chọn các ô Percentiles, Central tendency, Dispersions, Distributions (Hình 2.7). Sau đó bạn vào Charts và đánh dấu chọn Histograms with normal curve. Nếu cần so sánh các nghiệm thức thì bạn vào Formate và đánh dấu chọn mục Compare Variables trong ô Multiple Variables. Cuối cùng bạn chọn

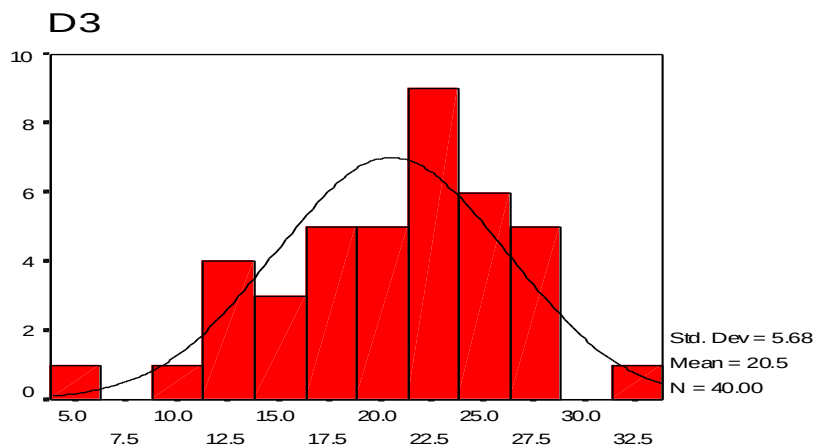
Continue rồi chọn OK để chương trình tính toán và hiển thị kết quả. Dưới đây chỉ trích dẫn giản đồ Hình cột cùng với một đường cong phân bố chuẩn (Hình 2.8) và các phân vị (Bảng 2.7).



Hình 2.6



Hình 2.7



Hình 2.8. Giản đồ hình cột cùng với đường cong chuẩn

		D1	D2	D3	D4
N	Valid	40	40	40	40
	Missing	0	0	0	0
Percentiles	10	12.050	12.000	11.550	11.050
	20	15.300	15.000	15.100	14.600
	30	17.650	16.300	18.000	18.000
	40	18.200	17.700	19.400	19.900
	50	20.750	20.000	21.500	21.500
	60	21.800	22.300	22.500	22.800
	70	22.850	22.850	23.850	24.500
	80	23.500	26.300	25.800	27.700
	90	25.400	27.400	27.450	28.900

Lưu ý rằng, phân vị (đơn vị là phần trăm) là một số mà dưới nó bao gồm bao nhiêu phần trăm số mẫu nhỏ hơn nó. Phân vị được dùng cho những dãy số liệu lớn. Nó đặc biệt hữu ích khi dãy số liệu phân bố không tuân theo luật phân bố chuẩn. Trong thực hành, người ta thường dùng các phân vị 10, 25, 50, 75 và 90% giá trị quan sát; trong đó phân vị 25% được gọi là tứ phân vị (Percentile) thứ nhất, 50% - tứ phân vị thứ hai, còn 75% - tứ phân vị thứ 3.

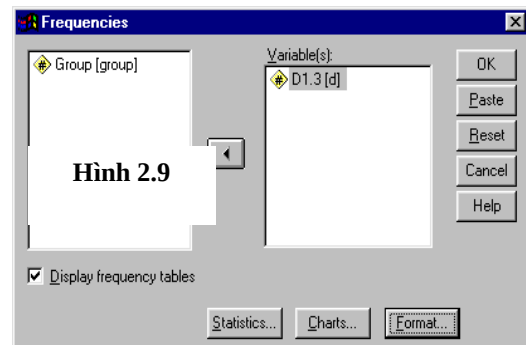
2.4. KIỂM ĐỊNH PHÂN BỐ CHUẨN VÀ SỰ BẰNG NHAU CỦA PHƯƠNG SAI

2.4.1. Kiểm định phân bố chuẩn

Để kiểm định giả thuyết H_0 cho rằng dãy số liệu có phân bố chuẩn, SPSS đưa ra ba kiểm định là Kolmogorov – Smirnov, Shapiro–Wilk và Lilliefors. Kiểm định Kolmogorov–Smirnov và Lilliefors được dùng trong trường hợp số trung bình và phương sai của tổng thể là chưa biết trước. Để kiểm định về phân bố chuẩn, bạn hãy thực hiện theo các bước sau đây :

Bước 1. Từ Menu chính, mở Analyze $\Rightarrow$ Descriptive Statistic $\Rightarrow$ Frequencies. Đến đây bạn chọn những biến số cần xử lý và chuyển chúng vào ô Variables (**Hình 2.9**).

Bước 2. Từ cửa sổ Frequencies, bạn chọn Statistics và đánh dấu mục Skewness và Kurtosis trong ô Distribution. Sau đó vào Charts và đánh dấu chọn Histograms và With normal curve. Nếu bạn cần so sánh các biến số (nghiệm thức) thì vào Formate rồi đánh dấu chọn ô Compare variables. Cuối cùng bạn chọn Continue rồi chọn OK để chương trình tính toán và hiển thị kết quả.



Bảng 2.8. Các đặc trưng phân bố

N	Valid	40
	Missing	0
Skewness		.115
Std. Error of Skewness		.374
Kurtosis		.098
Std. Error of Kurtosis		.733

Giải thích kết quả. **Bảng 2.8** dẫn ra những thống kê mô tả về các đặc trưng phân bố – đó là độ lệch (Skewness - kí hiệu Sk) và độ nhọn (Kurtosis - kí hiệu Ex) cùng với các sai số chuẩn của chúng – tương ứng kí hiệu là S_{Sk} và S_{Ex} .

+ Các kiểm định về phân bố chuẩn (Test of normality) của dãy số liệu. Để đạt được mục đích đặt ra, bạn có thể dựa vào một số thống kê sau đây :

1. Kiểm định Kolmogorov-Smirnov và Shapiro-Wilk (**Bảng 2.6**). Khi $P < 0.000$ ($P < 0.0005$ - mặc định) thì giả thuyết về phân bố chuẩn bị bác bỏ. Ngược lại, khi $P > 0.0005$ (mặc định) thì giả thuyết về phân bố chuẩn được chấp nhận.

2. Tỷ số giữa độ lệch (Sk) với sai số chuẩn của nó (S_{Sk}) và giữa độ nhọn (Ex) với sai số chuẩn của nó (S_{Ex}). Đối với phân bố chuẩn, Sk và Ex = 0. Khi tỷ số Sk/S_{Sk} và Ex/S_{Ex} nằm trong khoảng ± 2 thì dãy số liệu có phân bố tiệm cận chuẩn. Kiểm định phân bố chuẩn cũng có thể dựa vào độ lệch và độ nhọn chuẩn hoá. Khi P – value > 0.0005 (mặc định) thì dãy số liệu có phân bố tiệm cận chuẩn.

3. Thống kê phù hợp Chi – Square (Chi-Square goodness-of-fit statistic). Khi P-value $< \alpha$ (0.0005) thì giả thuyết H_0 bị bác bỏ. Ngược lại, khi P-value $> \alpha$ (0.0005) thì giả thuyết H_0 được chấp nhận.

4. Dựa vào đồ thị xác suất chuẩn (Normal Q-Q plot). Đồ thị xác suất chuẩn ở **Hình 2.3** được vẽ với trục hoành là trị số quan sát, còn trục tung là trị số kỳ vọng (chuẩn hoá) từ một phân bố chuẩn. Khi các giao điểm giữa các trị số quan sát và các trị số chuẩn hoá kỳ vọng cùng nằm trên một đường thẳng thì phân bố của dãy số liệu là tiệm cận chuẩn.

5. Lược đồ tần suất (Histogram) có gắn một đường cong chuẩn.

2.4.2. Kiểm định ngang bằng phương sai

Chương trình SPSS xây dựng kiểm định sự ngang bằng phương sai mẫu bằng kiểm định Levene's Test. Giả thiết cần kiểm định ngang bằng phương sai giữa các biến $D_{1,3}$ của hai hay nhiều quần thể (nghiệm thức). Để đạt mục đích đó, bạn có thể thực hiện theo thủ tục sau đây :

1. Dùng thủ tục Compare Means : Independent Samples T- Test. Thủ tục này cho phép kiểm định ngang bằng phương sai giữa hai mẫu độc lập. Cách gọi chương trình tính toán như sau :

Bước 1. Trước hết, bạn hãy nhập số liệu như sau : Cột 1 nhập thứ tự đại lượng quan sát (có hay không cũng được); cột 2 nhập số liệu của các biến số theo thứ tự từ biến 1 đến biến 2; cột 3 nhập các yếu tố phân loại (ở đây là các quần thể khác nhau) theo thứ tự từ 1 đến 2.

Bước 2. Gọi chương trình tính toán. Mở Analyze → Compare Means → Independent Samples T- test (đến đây hãy đưa biến phụ thuộc (cột 2) vào ô Test Variables, còn biến lập nhóm (cột 3) vào Grouping Variable) (Hình 2.10).

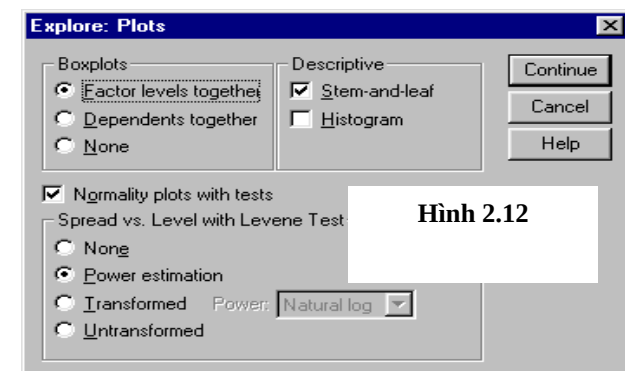
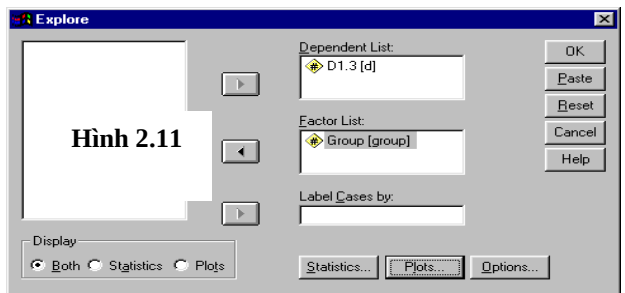
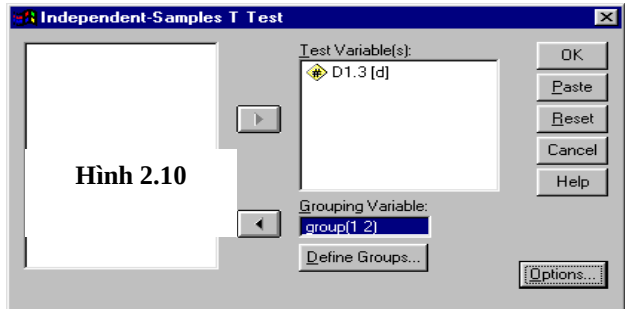
Bước 3. Từ cửa sổ Independent Samples T- test, chọn Define Groups (ghi số 1 trong Group 1, số 2 trong Group 2). Kế đến chọn Options và đánh dấu những thủ tục cần tính. Sau cùng bạn chọn OK để chương trình xử lý và báo cáo kết quả.

2. Dùng thủ tục Descriptives Statistic : Explore. Thủ tục này cho phép so sánh sự ngang bằng phương sai của hơn hai mẫu. Sau khi nhập số liệu giống như thủ tục Compare Means : Independent Samples T- Test, bạn gọi chương trình tính toán như sau :

Bước 1. Từ Menu chính, mở Analyze ⇒ Descriptive Statistic ⇒ Explore. Đến đây bạn chọn những biến số cần kiểm định và chuyển chúng vào ô Dependent List; kế đến chọn các biến số phân loại và chuyển chúng vào ô Factor List (Hình 2.11).

Bước 2. Từ cửa sổ Explore, bạn chọn Both; kế đến chọn Statistics và đánh dấu các thống kê cần tính. Tiếp theo chọn Plots và đánh dấu chọn Normality plots with tests, Power estimation (Hình 2.12). Cuối cùng bạn chọn Continue rồi chọn OK để chương trình tính toán và hiển thị kết quả.

Giải thích kết quả. Khi sử dụng thủ tục Compare Means: Independent Samples T- Test để kiểm định ngang bằng phương sai giữa hai mẫu, kết quả sẽ hiển thị một Bảng như sau (Bảng 2.9):



Bảng 2.9. Kiểm định ngang bằng phương sai (Test for Equality of Variances)

VAR1		Levene's Test for Equality of Variances	Sig.
		F	
	Equal variances assumed	1.105	.296
	Equal variances not assumed		

Bạn hãy quan sát các giá trị F và P – value (Sig). Nếu $P > 0.05$ (mặc định) thì phương sai hai mẫu bằng nhau ở mức tin cậy 95% và ngược lại. Khi phương sai hai mẫu bằng nhau thì kết quả phân tích sẽ sử dụng phương sai gộp (Equal Variances Assumed). Ngược lại, khi phương sai hai mẫu khác nhau thì kết quả phân tích sẽ sử dụng phương sai không gộp (Equal Variances Not Assumed). Ở ví dụ trên, vì $P > 0.05$ nên phương sai của hai mẫu là bằng nhau.

Bảng 2.10. Kiểm định ngang bằng phương sai
(Test of Homogeneity of Variance)

		Levene Statistic	df1	df2	Sig.
D1.3	Based on Mean	.452	1	78	.503
	Based on Median	.480	1	78	.490
	Based on Median and with adjusted df	.480	1	77.4	.490
	Based on trimmed mean	.447	1	78	.506

Khi sử dụng thủ tục Descriptives Statistic: Explore để kiểm định ngang bằng phương sai giữa K mẫu ($K > 2$), kết quả sẽ hiển thị một Bảng kiểm định ngang bằng phương sai (Bảng 2.10). Bạn hãy quan sát thống kê Levene Statistic và mức ý nghĩa P (Sig). Nếu $P > 0.05$ (mặc định) thì phương sai của K mẫu bằng nhau ở mức tin cậy 95% và ngược lại. Ở ví dụ trên, vì $P > 0.05$ nên phương sai của các mẫu là bằng nhau.

2.5. SỬ DỤNG THỦ TỤC REPORTS ĐỂ TÍNH CÁC THỐNG KÊ MÔ TẢ THEO NHÓM ĐỐI TƯỢNG

Thủ tục Reports cho phép bạn tóm tắt các thông tin về từng đối tượng (trường hợp) hoặc một nhóm đối tượng. Ví dụ: Khi điều tra một khu tiêu chuẩn, bạn thống kê tên loài cây gỗ cùng với các chỉ tiêu mô tả cho vai trò của loài ($D_{1,3}$, cm; H, m; V, m^3 ; G, m^2 ...). Từ số liệu này, bạn có thể yêu cầu chương trình tính toán và tổng hợp các đặc trưng thống kê về D, H, V, G... của từng loài cây, từng nhóm loài cây (phân theo một tiêu chuẩn nào đó, ví dụ nhóm gỗ) và toàn bộ khu tiêu chuẩn. Để đạt được mục đích đặt ra, bạn có thể thực hiện theo thủ tục sau đây:

Bước 1. Thu thập dữ liệu. Sau khi thiết lập một khu tiêu chuẩn, bạn hãy tiến hành thu thập tất cả các số liệu về loài cây cùng với các đặc trưng D, H, V, G... theo loài cây.

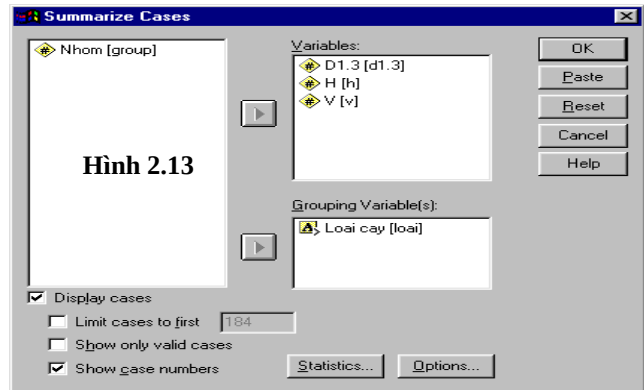
Bảng 2.11. Bảng thống kê loài cây trên khu tiêu chuẩn

Thứ tự	Loài cây	$D_{1,3}$, cm	H, m	V, m^3	Nhóm loài/gỗ
1()	(2)	(3)	(4)	(5)	(6)
1	Dầu rái	12.5	10.5	0.1187	4
2	Sao đen	13.0	11.5	0.1526	4
3	Vên vên	12.5	11.0	0.1349	4
4...	Thị	18.5	16.5	0.4197	8

Bước 2. Nhập số liệu vào Bảng tính. Số liệu của một khu tiêu chuẩn có thể được nhập trực tiếp vào Bảng tính SPSS 10.0 hoặc vào Bảng tính Excel. Cột 1 ghi thứ tự cây; cột 2 ghi biến lập nhóm theo loài cây; cột 3, 4... ghi các đặc trưng D, H, G, V của từng loài...; cột cuối cùng ghi lại biến lập nhóm theo nhóm gỗ (**Bảng 2.11**).

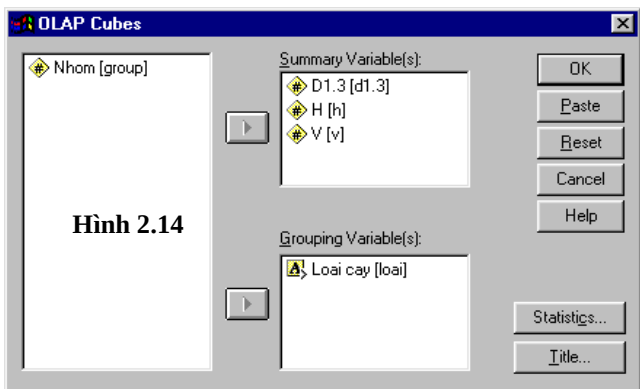
Bước 3. Gọi chương trình tính toán theo trình tự sau đây :

+ **Tóm tắt từng trường hợp và tất cả các trường hợp (Summarize Cases).** Từ cửa sổ chính, bạn vào lệnh Analyze ⇒ Reports ⇒ Case Summaries. Đến đây cửa sổ Summarize Cases xuất hiện như Hình 2.13. Tại cửa sổ Summarize Cases, bạn chọn các đặc trưng D, H, V, G... rồi chuyển chúng sang ô Variables. Nếu cần biết các đặc trưng thống kê theo loài cây, bạn chọn biến “Loài cây” rồi chuyển nó sang ô Grouping Variables. Tương tự, nếu cần biết các đặc trưng thống kê theo nhóm loài hay nhóm gỗ, bạn chọn biến “Nhóm loài cây hay nhóm gỗ” rồi chuyển nó sang ô Grouping Variables. Kế đến, chọn Statistics; tại đây bạn có thể chọn các thống kê mô tả theo mục đích của bạn. Nếu muốn hiển thị các đặc trưng thống kê mô tả cho từng loài kèm theo thứ tự từng trường hợp trong danh sách của biểu đồ đếm thì bạn chọn Display Cases (**Bảng 2.12**). Ngược lại, nếu chỉ muốn hiển thị các đặc trưng thống kê cho từng loài mà không kèm theo thứ tự từng trường hợp thì bạn tắt dấu chọn Display Cases (**Bảng 2.13**). Tại đây bạn cũng có thể chọn một tập hợp tính toán cho một cỡ mẫu nào đó bằng cách ghi cỡ mẫu vào ô Limit Cases to first... Cuối cùng bạn chọn OK để chương trình xử lý và hiển thị kết quả.



+ **Tóm tắt theo nhóm đối tượng (OLAP CUBES...).** Thủ tục này cũng tính toán các đặc trưng thống kê mô tả cho từng loài hay nhóm loài. Kết quả tính toán hiển thị một Bảng tóm tắt mà theo đó bạn có thể chọn bất kỳ trường hợp nào để xem xét. Cách thực hiện như sau :

Từ Menu chính, bạn chọn Analyze ⇒ Reports ⇒ OLAP CUBES. Đến đây cửa sổ OLAP CUBES xuất hiện như Hình 2.14. Tại cửa sổ OLAP CUBES, bạn chọn các đặc trưng D, H, V và G rồi chuyển chúng sang ô Summary Variables. Nếu cần biết các đặc trưng thống kê mô tả theo loài cây, thì bạn chọn biến “Loài cây” rồi chuyển nó sang ô Grouping Variables. Nếu cần biết các đặc trưng thống kê theo nhóm loài hay nhóm gỗ, thì bạn chọn biến “Nhóm loài cây hay nhóm gỗ” rồi chuyển nó sang ô Grouping Variables. Kế đến, bạn chọn Statistics; tại đây bạn có thể chọn các thống kê mô tả theo mục đích của bạn. Cuối cùng chọn OK để chương trình xử lý và hiển thị kết quả (**Bảng 2.14 và 2.15**). **Bảng 2.14** có được nhờ chọn Pivot ⇒ Reset Pivots to Defaults.



Bảng 2.15 có được nhờ chọn Pivot ⇒ Move Layers to Row. Để có được Bảng này, bạn cần thực hiện như sau :

Tại Bảng OLAP CUBES, bạn đưa con trỏ chuột vào Bảng kết quả và nhấp đúp nút chuột trái. Thao tác này làm xuất hiện một Bảng Pivot. Trên Bảng Pivot có một nút chọn. Khi dò mũi chuột vào nút chọn này và nhấn nút chuột trái, bạn sẽ thấy xuất hiện một cửa sổ. Tại đây bạn có thể chọn những chủ đề theo yêu cầu của mình. Thao tác này giúp bạn nhận được tất cả các đặc trưng thống kê cho từng loài cây và toàn bộ các loài cây. Tại đây, nếu bạn muốn chuyển hàng thành cột hoặc ngược lại thì chọn Pivot trên Menu chính. Sau đó chọn Pivot $\Rightarrow$ Transpose Rows to Columns. Nếu bạn muốn có được các Bảng có cấu trúc khác thì hãy chọn Pivot trên thanh Menu chính.

Bảng 2.12. Tóm tắt từng trường hợp (Case Summaries)

Loài cây			Case Number	D1.3	H	V
Cà chắt	1		5	21.0	19.5	.6751
	2		21	12.0	10.5	.1187
	3		53	6.5	5.0	.0166
	4		69	8.5	7.0	.0397
	...	...				
	Total	N		11	11	11
		Mean		14.718	13.218	.381573
		Minimum		6.5	5.0	.0166
		Maximum		30.0	28.5	2.0135
		Variance		45.154	45.154	.331

+ **Tóm tắt các đặc trưng thống kê theo hàng và theo cột.** Để đạt được mục đích đặt ra, bạn hãy chọn Analyze $\Rightarrow$ Reports Summaries in Row và Analyze $\Rightarrow$ Reports Summaries in Columns.

+ Tóm tắt các đặc trưng thống kê theo nhiều lớp. Giả sử bạn cần biết các đặc trưng thống kê theo nhóm gỗ (lớp 1), rồi trong mỗi nhóm gỗ chỉ ra loài cây (lớp 2), thì bạn có thể thực hiện như sau :

Trước hết, tại cửa sổ OLAP CUBES, bạn chọn các đặc trưng D, H, V và G rồi chuyển chúng sang ô Summary Variables. Kế đến chọn biến “Nhóm gỗ” rồi chuyển nó sang ô Grouping Variables. Tiếp theo bạn lại chọn biến loài cây và chuyển nó sang ô Grouping Variables. Cuối cùng, bạn chọn Statistics rồi chọn các thống kê mô tả và nhấn OK để chương trình xử lý và hiển thị kết quả.

Bảng 2.13. Tóm tắt từng trường hợp (Case Summaries)

Loài cây		D1.3	H	V
Daurai	N	24	24	24
	Mean	14.954	13.454	.333667
	Std. Error of Mean	1.093	1.093	6.65256E-02
	Minimum	7.5	6.0	.0265
	Maximum	23.0	21.5	.8928
	Variance	28.650	28.650	.106
...	...	...	...	...
Total	N	184	184	184
	Mean	15.684	14.184	.395579
	Std. Error of Mean	.424	.424	2.99134E-02

Minimum	4.5	3.0	.0048
Maximum	32.0	30.5	2.4517
Variance	33.093	33.093	.165

Bảng 2.14. Tóm tắt các đặc trưng thống kê của các loài cây (OLAP CUBES)

	Sum	N	Mean	Std. Deviation	% of Total Sum	% of Total N
D1.3	2885.9	184	15.684	5.753	100.0%	100.0%
H	2609.9	184	14.184	5.753	100.0%	100.0%
V	72.7866	184	.395579	.405765	100.0%	100.0%

Bảng 2.15. Tóm tắt các đặc trưng thống kê theo loài cây (OLAP Cubes)

Loai cay		Sum	N	Mean	S _{td}	% of Total Sum	% of Total N
Daurai	D1.3	358.9	24	14.954	5.353	12.4%	13.0%
	H	322.9	24	13.454	5.353	12.4%	13.0%
	V	8.0080	24	.333667	.325907	11.0%	13.0%
Saoden	D1.3	376.9	22	17.132	6.814	13.1%	12.0%
	H	343.9	22	15.632	6.814	13.2%	12.0%
	V	11.8614	22	.539155	.585477	16.3%	12.0%
Total	D1.3	2885.9	184	15.684	5.753	100.0%	100.0%
	H	2609.9	184	14.184	5.753	100.0%	100.0%
	V	72.7866	184	.395579	.405765	100.0%	100.0%

2.6. TÍNH CÁC ĐẶC TRƯNG THỐNG KÊ CỦA BẢNG PHÂN PHỐI TẦN SỐ

Khi số liệu đã được tập hợp thành một Bảng phân phối tần số (**Bảng 2.16**) thì thủ tục tính các đặc trưng thống kê như sau :

Bảng 2.16. Phân phối thu nhập (Xi, ngàn đồng) của xí nghiệp A

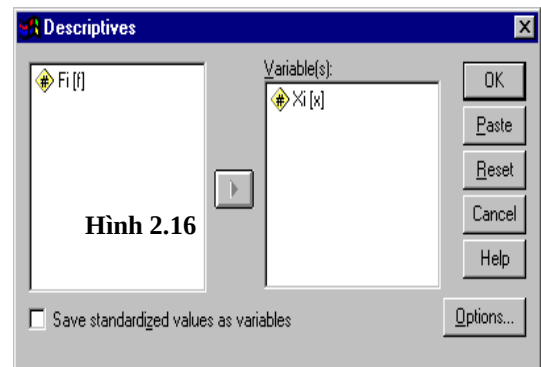
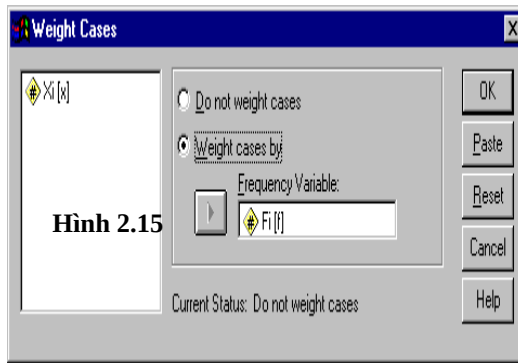
TT	Xi	Fi
(1)	(2)	(3)
1	510	81
2	530	12
3	550	20
4	570	56
5	590	18
6	610	16
7	630	10
Tổng		213

Bước 1. Nhập số liệu vào Bảng tính. Cột 1 ghi thứ tự nhóm (cấp). Cột 2 ghi trị số giữa cấp của từng nhóm (ở ví dụ này là thu nhập của người lao động). Cột 3 ghi tần số của mỗi nhóm (số người thuộc các nhóm thu nhập khác nhau).

Bước 2. Gọi chương trình tính toán. Để tính các đặc trưng thống kê mô tả, bạn thực hiện như sau:

+ Mở Data – Weight Cases – Weight Cases by. Đến đây bạn đưa biến tần số (cột 3) vào ô Frequency variable (Hình 2.15). Sau đó chọn OK.

Từ Menu chính, bạn hãy mở Analyze ⇒ Descriptive Statistic



⇒ Descriptives. Khi cửa sổ Descriptives xuất hiện, bạn hãy chuyển số liệu thu nhập (cột 2) vào ô Variables (Hình 2.16).

Bước 3. Tính các thống kê mô tả. Từ cửa sổ Descriptives, bạn chọn Options. Khi cửa sổ Options xuất hiện, bạn hãy chọn những thống kê mô tả cần quan tâm (Hình 2.17). Sau cùng thì nhấp OK để chương trình tính toán. Đến đây chương trình sẽ hiện thị những thống kê mô tả theo chỉ định của bạn (Bảng 2.17).



Giải thích kết quả. Bạn hãy căn cứ vào kết quả tính toán và đưa ra những nhận xét về :

+ Những thống kê mô tả khuynh hướng trung tâm của dãy số liệu (trung bình, Mode, Mediance...);

+ Những thống kê mô tả khuynh hướng phân tán của dãy số liệu (sai tiêu chuẩn, độ lệch, độ nhọn...).

Lưu ý rằng, nếu bạn gọi lệnh Analyze ⇒ Descriptive Statistic ⇒ Frequencies và chọn các thống kê mô tả thì kết quả hiển thị như Bảng 2.18. Tương tự như trên, bạn có thể chọn thủ tục Explore.

Bảng 2.17. Các thống kê mô tả (tóm tắt)

		Xi	Valid N (listwise)
N	Statistic	213	213
Mean	Statistic	550.56	
	Std. Error	2.61	
Std. Deviation	Statistic	38.06	
Variance	Statistic	1448.738	
Skewness	Statistic	.378	
	Std. Error	.167	
Kurtosis	Statistic	-1.030	
	Std. Error	.332	

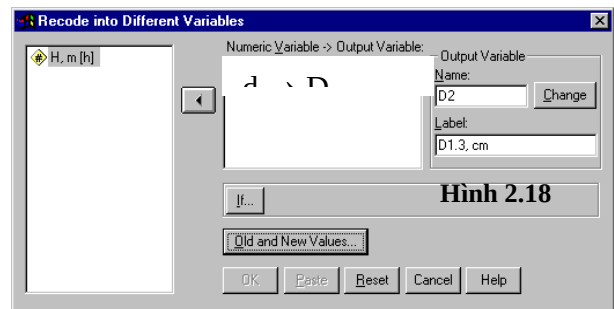
Bảng 2.18. Các thống kê mô tả

Missing	0		Kurtosis		-1.030
N	Valid	213	Std. Error of Kurtosis		.332
Mean		550.56	Range		120
Std. Error of Mean		2.61	Minimum		510
Median		551.84	Maximum		630
Mode		510	Sum		117270
Std. Deviation		38.06	Percentiles	10	.
Variance		1448.74		25	515.48
Skewness		.378		50	551.84
Std. Error of Skewness		.167		75	580.14

2.7. LẬP BẢNG PHÂN BỐ ĐƯỜNG KÍNH VÀ CHIỀU CAO THÂN CÂY

2.7.1. Lập Bảng phân bố đường kính thân cây (N – D). Giả sử một quần thụ có 312 cây; $D_{1.3}$ phân bố trong khoảng $6 \leq D_{1.3} \leq 33$ cm; trị trung bình là 20.4 cm. Để lập Bảng phân bố N – D, bạn cần biết cự ly mỗi cấp kính và số cấp kính. Số cấp kính (m) có thể xác định theo công thức: $m \geq 5 \times \log N$ với N là dung lượng quan sát. Cự ly một cấp kính (K) có thể tính theo công thức: $K = (X_{\max} - X_{\min})/m$. Ở ví dụ này, $m = 5 \times \log(312) = 12$; $K = (33 - 6)/(5 \times \log(312)) = 2,2$ cm (lấy tròn bằng 2 cm). Theo đó trị số trung tâm (Midpoint) của cấp kính thứ 1, 2, 3... tương ứng là 8, 10, 12 cm... Trị số trung tâm của cấp kính cuối cùng là 32 cm. Để tính tần số của các cấp kính và những đặc trưng thống kê mô tả cho quần thụ, bạn hãy thực hiện như sau :

Bước 1. Nhập số liệu $D_{1.3}$ của 312 cây vào Bảng tính. Sau đó chọn Transform $\Rightarrow$ Recode $\Rightarrow$ Into Different Variables. Đến đây bạn chuyển biến $D_{1.3}$ vào ô Numeric Variables $\rightarrow$ Output Variable. Kế đến bạn ghi tên biến mới (ví dụ D_2) và mã biến mới (ví dụ $D_{1.3}, \text{cm}$) tương ứng vào ô Name và Label; rồi chọn Change (Hình 2.18).

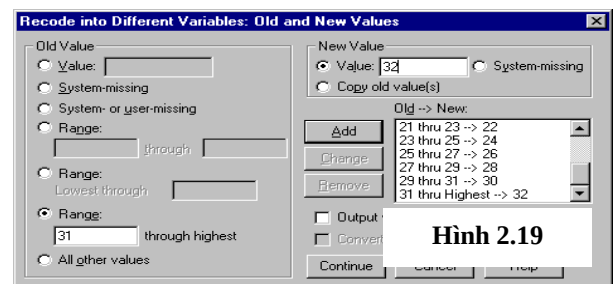


Hình 2.18

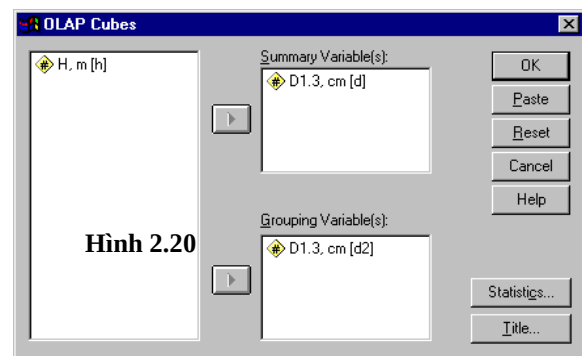
Bước 2. Chọn tiếp Old and New Values. Sau đó bạn hãy thực hiện liên tiếp những thủ tục sau đây :

2a. Trong ô Old value, bạn chọn Range $\Rightarrow$ Lowest through. Kế đến ghi số 9 vào ô trống bên cạnh, còn số 8 vào ô value của khoảng New value. Tiếp theo chọn Add. Thủ tục này yêu cầu chương trình mã hoá tất cả những cây có $D < 9$ cm bằng số 8 (nghĩa là cấp kính 8 cm).

2b. Chọn Range $\square$ (ghi số 9) Through



Hình 2.19



Hình 2.20

□ (ghi số 11), rồi ghi số 10 vào ô value của khoang New value. Sau đó chọn Add. Thủ tục này yêu cầu chương trình mã hoá những cây có giá trị $9 < D < 11$ bằng số 10 (cấp kính 10 cm).

2c. Hãy lập lại bước 2b để mã hoá cho những cấp kính tiếp theo (12, 14 ...30).

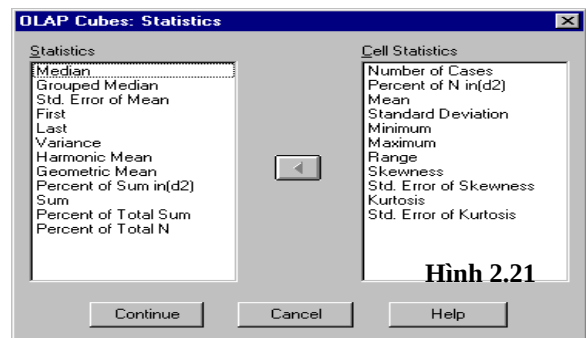
2d. Để mã hoá cho cấp kính cuối cùng (32 cm), bạn chọn Range $\Rightarrow$ through highest. Kế đến ghi số 31 vào ô trống bên cạnh, còn số 32 đưa vào ô value của khoang New value. Sau đó chọn Add. Thao tác này yêu cầu chương trình mã hoá tất cả những cây có $D > 31$ cm vào cấp kính 32 cm (Hình 2.19).

Bước 3. Sau khi hoàn thành bước 2, bạn chọn Continue rồi chọn OK. Đến đây chương trình sẽ mã hoá từng cây vào các cấp kính tương ứng và ghi kết quả mã hoá vào một cột bên cạnh của Bảng SPSS.

Bước 4. Lập Bảng phân bố N – D. Để thu được Bảng phân bố N - D (trong đó cột 1, 2 và 3...ghi tương ứng trị trung tâm, tần số và tần suất của mỗi cấp), bạn hãy sử dụng thủ tục Reports như sau :

- **Mở Analyze** $\Rightarrow$ Reports $\Rightarrow$ OLEP Cubes. Đến đây bạn chuyển biến D_{1.3} (số liệu gốc) vào ô Summary variables, còn biến D₂ (biến mã hoá cấp kính) vào ô Grouping variables (Hình 2.20).

- **Chọn Statistics.** Khi cửa sổ Statistics xuất hiện, bạn có thể chọn tần số (Number of Cases), tần suất (Percent of N in D₂), trị trung bình (Mean), sai tiêu chuẩn, độ lệch (Skewness), độ nhọn (Kurtosis)...(Hình 2.21). Sau cùng bạn chọn Continue rồi chọn OK. Kết quả sẽ hiển thị trên màn Hình những thống kê mô tả cho dãy phân bố N – D. Bạn hãy quan sát kết quả và chỉnh sửa nội dung của nó sao cho tương tự như Bảng 2.19.



Hình 2.21

Bảng 2.19. Đặc trưng phân bố N – D của quần
thụ

D1.3, cm	N	% N	Mean	Sd.	Min	Max	Range	Skewness	Kurtosis
8.0	14	4.5%							
10.0	9	2.9%							
...	...	...							
32.0	4	1.3%							
Total	312	100.0%	20.370	5.738	6.0	33.0	27.0	-.307	-.384

2. Lập Bảng phân bố chiều cao quần thụ (N – H). Giả sử một quần thụ có H phân bố trong khoảng $5 \leq H \leq 32$ m; trị trung bình là 19,2 m. Từ số liệu này ta thấy, cự ly một cấp kính là $K = (32 - 5)/(5 \times \log(312)) = 2.2$ m hay làm tròn thành 2 m. Theo đó trị số trung tâm (Midpoint) của các cấp chiều cao tương ứng là 6, 8, 10, 12 m... Trị số trung tâm của cấp chiều cao cuối cùng là 30 m. Để tính tần số của các cấp chiều cao và những đặc trưng thống kê mô tả chiều cao quần thụ, bạn hãy thực hiện những thủ tục mã hoá và gọi chương trình xử lý tương tự như đã làm với thủ tục lập Bảng phân bố N – D.

Một số hàm phân bố N/D và N/H

Đối với rừng trồng thuần loài đồng tuổi, đường cong mô tả sự suy giảm mật độ quần thụ (N, cây/ha) theo tuổi (A, năm) thường có dạng đường cong hypecbol. Vì thế,

phân bố N/A có thể được mô tả bằng các hàm dưới đây. Giá trị N_0 là mật độ trồng rừng ban đầu hoặc số cây ở tuổi (A , năm) bắt đầu nghiên cứu; b và b_2 là tỉ lệ hay tốc độ suy giảm số cây theo luật hàm số mũ ở giai đoạn tuổi ban đầu (Rừng non); tham số k là số cây ở A cao nhất.

Dạng hàm	Hàm phân bố	
$N_A = N_0 \times A^{-b}$	Lũy thừa (Power)	(2.1)
$N_A = N_0 \times \exp(-b \times A)$	Mũ (Exponent)	(2.2)
$N_A = N_0 \times \exp(-b \times A) + k$	Mũ biến đổi	(2.3)
$N_A = N_0 \times \exp(b \times A^{-c})$	Mũ luật lũy thừa	(2.4)
$N_A = N_0 \times D^{-b} \times \exp(-c \times A)$	Mũ tổng quát	(2.5)
$N_A \times \frac{(N_0 \times b) \exp(-b \times A)}{1 - \exp(-b \times A)}$	Mũ nghịch đảo	(2.6)
$N_A = N_0 \times \exp(-a \times (1 - \exp(-b \times A)))$	Thêm (2025)	(2.7a)
$N_A = N_0 \times \exp(-a \times (1 - \exp(-b \times A^c)))$	Thêm (2025)	(2.7b)

Nguồn: (Gompertz, 1825; Gupta & Kundu, 1999; Nguyễn Văn Thêm et al., 2026)

Phân bố số cây (N , cây/ha) theo cấp chiều cao (H , m) (Kí hiệu: N/H) của rừng trồng thuần loài đồng tuổi có dạng 1 đỉnh lồi. Vì thế, hàm phân bố N/H thể được mô tả bằng các hàm 2.8÷2.20. Ở hàm 2.8÷2.20, $f_{(H)}$ là hàm mật độ xác suất, μ và σ tương ứng là trung bình và độ lệch chuẩn của biến $H(m)$. Ở hàm 2.10, α là Hình dạng hay độ lệch của đường cong hàm mật độ, β là tham số tỉ lệ. Ở hàm 2.11, a và b là tham số của phân bố khoảng cách, $H = 0, 1, 2, \dots, k$. Ở hàm 2.12÷2.15, F_H là xác suất tích lũy số cây theo cấp H , a là tham số vị trí, b là tốc độ suy giảm số cây theo cấp H , c là tham số Hình dạng. Ở hàm 2.16, q là tham số vị trí, p là tham số tỉ lệ và r là tham số Hình dạng ($p, q > 0$; $r < 0$). Ở hàm 2.17, a và b tương ứng là H_{Min} và H_{Max} , α là tham số Hình dạng, β tham số tỉ lệ, $B_{(p, q)}$ là hàm phân bố Beta, p và q tương ứng là H_{Min} và H_{Max} . Ở hàm 2.18÷2.20, F_H là xác suất tích lũy số cây theo cấp H ; H và H_{Max} tương ứng là cấp H thứ i và cấp H_{Max} ; $b_0 \div b$ là các tham số.

Dạng hàm	Hàm phân bố	
$f_{(H)} = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{1}{2}(\frac{H - \mu}{\sigma})^2)$	Chuẩn	(2.8)
$f_{(H)} = \frac{1}{X\sigma\sqrt{2\pi}} \exp(-\frac{(\ln H - \mu)^2}{2\sigma^2})$	Lognormal	(2.9)
$f_{(H)} = \frac{(H/\beta)^{\alpha-1} \exp(-H/\beta)}{\beta\Gamma(\alpha)}$	Gamma	(2.10)
$f_{(H)} = a$ với $H = 0$ $f_{(H)} = (1 - a) \times (1 - b) \times b^{(H-1)}$ với $H > 1$	Khoảng cách	(2.11)
$F_{(H)} = (1 - \exp(-b \times H^c))$	Weibull 2 tham số	(2.12)
$F_{(H)} = 1 - \exp(-((H - a)/b))^c$	Weibull 3 tham số	(2.13)
$F_{(H)} = (1 - \exp(-b \times H))^c$	Gompertz 2 tham số (1825)	(2.14a)
$F_{(H)} = (1 - a \times \exp(-b \times H))^c$	Gompertz 3 tham số (1825)	(2.14b)
$F_{(H)} = 1 - (1 + (H/a)^b)^{-c}$	Burr (1942)	(2.15)
$F_{(H)} = 1 + \exp(-(H - q)/p))^{-r}$	Richards (1980)	(2.16)

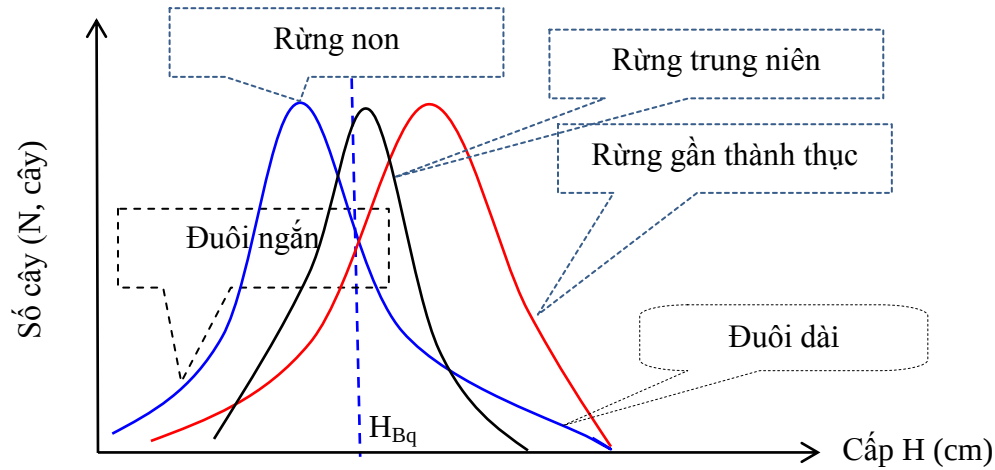
$$F_{(H)} = \frac{1}{B(p, q)} \frac{(H - a)^{p-1} (b - H)^{q-1}}{(b - a)^{p+q-1}} \quad \text{Beta} \quad (2.17)$$

$$F_{(H)} = \exp(b_0 + b_1 \times \ln(H) + b_2 \times (H/H_{\text{Max}})^{-b_3}) \quad \text{Thêm 2025(c)} \quad (2.18)$$

$$F_{(H)} = \exp(b_0 + b_1 \times \sqrt{H} + b_2 \times (H/H_{\text{Max}})^{-b_3}) \quad \text{Thêm 2025(d)} \quad (2.19)$$

$$F_{(H)} = \exp(b_0 + b_1 \times (H/H_{\text{Max}}) + b_2 \times (H/H_{\text{Max}})^{-b_3}) \quad \text{Thêm 2025(e)} \quad (2.20)$$

Nguồn: (Gompertz, 1825; Lê Thanh Quang, 2024; Nguyễn Hải Tuất & Nguyễn Trọng Bình, 2004)



Hình 2.21b. Phân bố N/H của rừng trồng thuần loài đồng tuổi.

Chương III

KIỂM ĐỊNH CÁC GIẢ THUYẾT THỐNG KÊ

3.1. PHƯƠNG PHÁP KIỂM ĐỊNH PHI THAM SỐ

3.1.1. Kiểm định 2 mẫu độc lập bằng thống kê Wilcoxon – Mann - Whitney

Điều kiện áp dụng: Hai biến độc lập và ngẫu nhiên; các giá trị sắp xếp theo thứ tự.

Nội dung kiểm định Wilcoxon-Mann - Whitney. Giả sử mẫu 1 có n_1 số liệu; mẫu 2 có n_2 số liệu. Để so sánh sự khác biệt giữa hai mẫu, người ta dựa vào hạng của các mẫu để kiểm tra. Để tính hạng của các mẫu, trước hết gộp số liệu của hai mẫu ($n_1 + n_2$) và sắp xếp chúng theo thứ tự từ nhỏ đến lớn. Tiếp theo tính tổng hạng cho từng mẫu, tương ứng kí hiệu R_1 và R_2 . Sau đó tính đại lượng W_1 và W_2 theo công thức :

$$W_1 = n_1 n_2 \frac{n_1(n_1 + 1)}{2} - R_1 \text{ và } W_2 = n_1 n_2 \frac{n_2(n_2 + 1)}{2} - R_2 \quad (3.1)$$

So sánh sự khác biệt giữa hai mẫu theo tiêu chuẩn U:

$$U = \frac{W - \frac{1}{2} n_1 n_2}{\sqrt{\frac{1}{12} n_1 n_2 (n_1 + n_2 + 1)}}; \text{ ở đây } W \text{ có thể là } W_1 \text{ hoặc } W_2. \quad (3.2)$$

Các giả định: H_0^+ : Hai mẫu không khác nhau hay $W_1 = W_2 = \frac{1}{2} n_1 n_2$; H_0^- : Hai mẫu khác nhau ($W_1 \neq W_2$). Sau đó người ta so sánh U với $U_{0.05}$. Nếu $U < U_{0.05}$ hay $P > 0.05$ thì hai mẫu không khác nhau. Ngược lại, $U > U_{0.05}$ hay $P < 0.05$ thì hai mẫu khác nhau.

Thực hành kiểm định Wilcoxon-Mann- Whitney. Nội dung của kiểm định Wilcoxon-Mann- Whitney được SPSS giải quyết theo các bước sau đây:

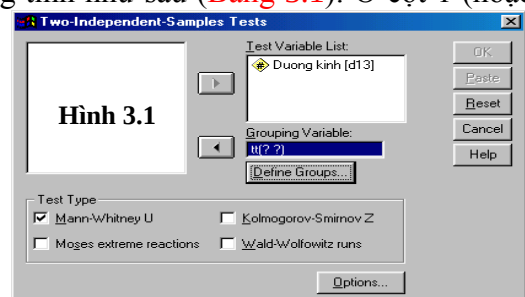
Bảng 3.1. Cách nhập dữ liệu trong kiểm định Wilcoxon -Mann- Whitney

Y	50	41	48	60	46	60	51	42	62	54	42	46	38	40	47	51	63	50	63	57	59	51
Xi	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2

Bước 1. Nhập số liệu. Số liệu nhập vào Bảng tính như sau (Bảng 3.1): Ở cột 1 (hoặc hàng 1) nhập lần lượt 2 biến phụ thuộc (Dependent Variable); trong đó biến 1 được nhập vào trước, còn biến 2 được nhập vào sau. Ở cột 2 (hoặc hàng 2), nhập biến độc lập (yếu tố thí nghiệm hoặc biến lập nhóm - Grouping Variable); trong đó mỗi biến độc lập biểu thị bằng một chuỗi số hay chuỗi kí tự giống nhau (Bảng 3.1).

Bước 2. Gọi chương trình tính toán. Để tính toán các nội dung, thực hiện theo các bước sau đây:

+ Chọn Analyze → Nonparametric Test → 2 Independent Samples → Test Variable List (đến đây đưa biến Data vào từ danh sách biến – cột 1) → Grouping variable (đưa biến lập nhóm vào – cột 2) → chọn Define Groups (trong ô Group1: ghi số 1, trong ô Group 2: ghi số 2).



+ Chọn Options và đánh dấu những thủ tục cần tính. Cuối cùng chọn OK (Hình 3.1). Đến đây kết quả hiển thị trên màn Hình như Bảng 3.2.

Bảng 3.2. Kết quả của kiểm định 2 mẫu độc lập

3.2a. Hạng trung bình và tổng hạng (Ranks) của các nghiệm thức

	VAR02	N	Mean Rank	Sum of Ranks
VAR01	1	12	10.79	129.50
	2	10	12.35	123.50
	Total	22		

3.2b. Các thống kê kiểm định (Test Statistics)

	VAR001
Mann-Whitney U	51.500
Wilcoxon W	129.500
Z	-.562
Asymp. Sig. (2-tailed)	.574
Exact Sig. [2×(1-tailed Sig.)]	.582

Giải thích và nhận xét kết quả. Bảng 3.2a cho biết tổng hạng (Ranks) và hạng trung bình của từng mẫu. Bảng 3.2b ghi lại các thống kê kiểm định (Test Statistics): (1) kiểm định Mann – Whitney U (đó là giá trị W_i của nhóm có tổng hạng nhỏ hơn) và Wilcoxon W (tổng hạng của nhóm có số mẫu lớn hơn hay $R_{i\max}$), (2) thống kê Z - đó là thống kê Mann – Whitney (hay kí hiệu U_i trong công thức 3.2), (3) xác suất 2 phía (Asymp. Sig. 2-tailed) cho biết mức ý nghĩa (xác suất) quan sát và mức ý nghĩa chính xác của kiểm định (Exact Sig. [2×(1-tailed Sig.)]). Nếu $P < 0.05$ (mặc định) thì giả thuyết H_0 bị bác bỏ, nghĩa là hai mẫu khác nhau một cách có ý nghĩa thống kê. Ngược lại, nếu $P > 0.05$ (mặc định) thì giả thuyết H_0 được chấp nhận, nghĩa là hai mẫu khác nhau không có ý nghĩa thống kê. Ở ví dụ này, vì $P = 0.574 > 0.05$ nên hai mẫu khác nhau không có ý nghĩa thống kê.

3.1.2. Kiểm định K mẫu ($K \geq 3$) độc lập bằng tiêu chuẩn Kruskal – Wallis

Để kiểm định sự khác nhau giữa K mẫu độc lập ($K \geq 3$), sử dụng tiêu chuẩn Kruskal – Wallis. Điều kiện áp dụng: Các biến độc lập với nhau và các giá trị có thể sắp xếp theo thứ tự. Giả thuyết H_0 : Các mẫu cùng tổng thể (phân bố giống nhau); H_0^- : Các mẫu khác tổng thể.

Nội dung kiểm định Kruskal – Wallis. Kiểm định Kruskal – Wallis dựa vào hạng của các mẫu để kiểm tra. Nội dung tiến hành như sau: Trước hết các mẫu được gộp lại và tìm tổng hạng cho từng mẫu. Kiểm định Kruskal – Wallis H được tính từ các tổng hạng này. Khi K mẫu có phân phối giống nhau thì $H \approx \chi^2$.

Thực hành kiểm định Kruskal – Wallis. Giả sử cần so sánh sự khác nhau về sinh khối của Thông ba lá 10 tuổi mọc trên ba cấp địa Hình khác nhau.

Bước 1. Nhập số liệu vào Bảng tính SPSS. Các số liệu của K mẫu độc lập được nhập như sau: Ở cột 1 (hoặc hàng 1) nhập biến phụ thuộc (ở đây là sinh khối của Thông ba lá); trong đó biến 1 được nhập vào trước, kế đến là biến 2 và 3. Cột 2 (hoặc hàng 2) nhập biến độc lập (ở đây là 3 cấp địa Hình – đó là yếu tố thí nghiệm hoặc biến lập nhóm - Grouping Variable); trong đó mỗi biến độc lập biểu thị bằng một chuỗi số hay chuỗi kí tự giống nhau (Bảng 3.3).

Bảng 3.3. Cách nhập số liệu vào Bảng tính SPSS trong kiểm định K mẫu ($K \geq 3$) độc lập bằng tiêu chuẩn Kruskal – Wallis

Dependent Variable (Sinh khối)	12	11	10	9	12	15	16	13	10	9	11	5	8	12	16	...
Gouping Variable (Địa Hình)	1	1	1	1	1	2	2	2	2	2	2	3	3	3	3	...

Bảng 3.4. Kiểm định k mẫu độc lập bằng tiêu chuẩn Kruskal – Wallis H

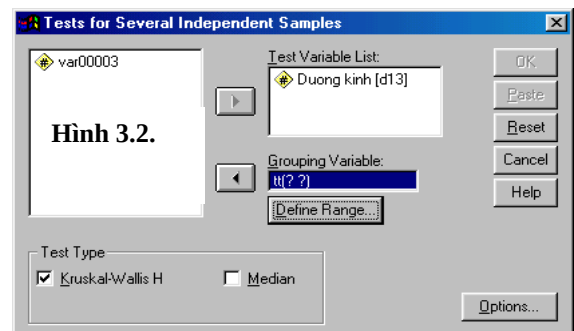
3.4a. Hạng trung bình và tổng hạng (Ranks) của các nghiệm thức

	VAR00004	N	Mean Rank
VAR00003	1.00	5	7.30
	2.00	6	9.33
	3.00	4	6.88
	Total	15	

3.4b. Các thống kê kiểm định (Test Statistics)

	Chi-Square	df	Asymp. Sig.
VAR00003	0.922	2	0.631

Bước 2. Gọi chương trình tính toán. Sau khi nhập xong dữ liệu, chọn Analyze → Nonparametric Tests → K Independent Samples → Test Variable List (đến đây đưa biến Data vào từ danh sách biến – cột 1) → Grouping variable (đưa biến lập nhóm vào – cột 2) → chọn Define Groups (trong ô Minimum thì ghi số 1, trong ô Maximum thì ghi số 3). Chọn Options và đánh dấu những thủ tục cần tính (**Hình 3.2**). Sau cùng chọn OK để chương trình xử lý và hiển thị kết quả (**Bảng 3.4**).



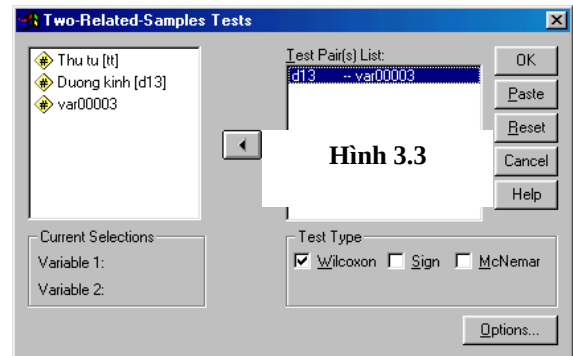
Giải thích và nhận xét kết quả. Theo trị số χ^2 tính toán hoặc χ^2 điều chỉnh (do nhiều quan sát có hạng bằng nhau) và xác suất P để kết luận. Nếu P (Asymp. Sig.) < 0.05 thì K mẫu khác nhau. Ngược lại, khi $P > 0.05$ thì K mẫu khác nhau không có ý nghĩa thống kê. Khi K mẫu khác nhau thì bạn có thể sử dụng thủ tục Multiple range tests để kiểm định phân hạng các nhóm.

3.2. KIỂM ĐỊNH MẪU LIÊN HỆ HAY MẪU BẮT CẬP

3.2.1. Kiểm định 2 mẫu liên hệ bằng thống kê Wilcoxon

Điều kiện áp dụng. Hai biến phụ thuộc và ngẫu nhiên; các giá trị có thể sắp xếp theo thứ tự. Giả thuyết H_0 : Hai mẫu cùng tổng thể (phân bố giống nhau); H_0^- - Hai mẫu khác tổng thể.

Nội dung của kiểm định Wilcoxon. Kiểm định Wilcoxon sử dụng chiều hướng chênh lệch giữa hai mẫu (kí hiệu bằng dấu – và dấu +). Vì thế, người ta còn gọi kiểm định Wilcoxon là kiểm định theo dấu. Để thực hiện kiểm định Wilcoxon, trước hết bạn hãy tính sai dị (chênh lệch) giữa hai cặp số liệu ($X_i - Y_i = di$). Kế đến bạn xếp hạng các sai dị này theo trị tuyệt đối của chúng. Tiếp theo bạn tính tổng hạng các sai dị dương (kí hiệu S^+) và sai dị âm (kí hiệu S^-) cho từng mẫu. Khi số cặp có sai dị khác zero lớn hơn 25 thì S^- hoặc S^+ sẽ có phân bố chuẩn. Do đó ta có thể dùng tiêu chuẩn U của phân bố chuẩn để kiểm định hai mẫu liên hệ. Để đạt được mục đích trên đây, bạn hãy thực hiện như sau (Hình 3.3):



Từ Menu chính, mở Analyze → Nonparametric Tests → 2 Related Samples → Test Pair lists (đến đây chọn cả mẫu 1 và mẫu 2 rồi nhấp nút ➤ để chuyển vào Test Pair Lists) → chọn Wilcoxon rồi chọn OK. Đến đây thống kê Wilcoxon Z và thống kê mô tả sẽ hiển thị trên màn Hình (Bảng 3.5).

Bảng 3.5. Kết quả của kiểm định 2 mẫu liên hệ bằng tiêu chuẩn Wilcoxon

3.5a. Các thống kê mô tả

	N	Mean	Std. Deviation	Minimum	Maximum
VAR00001	11	11.2727	2.2843	8.00	15.00
VAR00002	11	11.0364	2.3926	7.40	15.10

3.5b. Hạng trung bình và tổng hạng (Ranks) của các nghiệm thức

		N	Mean Rank	Sum of Ranks
VAR00002 - VAR00001	Negative Ranks	8	6.88	55.00
	Positive Ranks	3	3.67	11.00
	Ties	0		
	Total	11		

3.5c. Các thống kê kiểm định (Test Statistics)

	VAR00002 - VAR00001
Z	-1.970
Asymp. Sig. (2-tailed)	.049

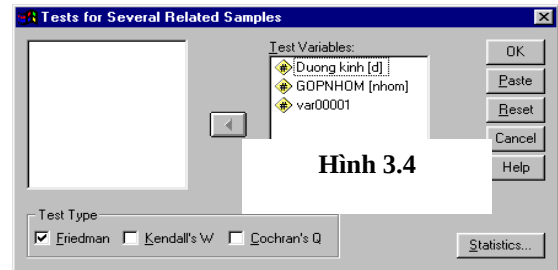
Giải thích và nhận xét kết quả. Nếu P (Asymp. Sig.) < 0.05 (mặc định) thì hai mẫu khác nhau. Ngược lại, nếu $P > 0.05$ thì hai mẫu bằng nhau. Ở ví dụ trên, vì $P = 0.049 < 0.05$ nên hai mẫu khác nhau với mức tin cậy 95%.

3.2.2. Kiểm định K mẫu liên hệ ($K \geq 3$) bằng thống kê Friedman

Giả thiết có ba điều tra viên cùng đo đặc đường kính, tiết diện ngang và thể tích thân cây của N cây rừng. Hãy cho biết kết quả đo đặc của ba điều tra viên này có giống nhau hay không? Để trả lời câu hỏi này, bạn có thể sử dụng kiểm định Friedman. Giả thuyết H_0 : Các

mẫu là như nhau hay cùng tổng thể; H^0 : Các mẫu khác nhau hay không cùng tổng thể. Cách nhập số liệu và gọi chương trình tính toán như sau:

Bạn hãy nhập số liệu của K mẫu (K nhân tố thí nghiệm) vào Bảng tính SPSS theo ba cột (hàng) liên tiếp nhau. Để gọi thủ tục tính toán, bạn mở Analyze → Nonparametric Tests → K Related Simple → Test Variable (chọn cả K mẫu và nhấp nút mũi tên bên cạnh để chuyển chúng vào Test Variable) → chọn Friedman Test → Statistics (đánh dấu mục cần tính) → OK (Hình 3.4). Đến đây kết quả của thống kê Friedman và các thống kê mô tả sẽ hiển thị trên màn Hình (Bảng 3.6).



Hình 3.4

Bảng 3.6. Kết quả kiểm định nhiều mẫu liên hệ bằng tiêu chuẩn Friedman

3.6a. Các thống kê mô tả

	N	Mean	Std. Deviation	Minimum	Maximum
D1	11	11.2727	2.2843	8.00	15.00
D2	11	11.0364	2.3926	7.40	15.10
D3	11	10.9909	2.3097	7.40	14.80

3.6b. Hạng trung bình và tổng hạng (Ranks) của các nghiệm thức

	VAR0001	VAR0002	VAR0003
Mean Rank	2.5	1.91	1.59

3.6c. Các thống kê kiểm định (Test Statistics)

N	Chi-Square	df	Asymp. Sig.
11	5.282	2	0.071

Giải thích và nhận xét kết quả. Từ số liệu của Bảng 3.6c, nếu P (Asymp. Sig.) > 0.05 (mặc định) thì các mẫu đem so sánh khác nhau không có ý nghĩa thống kê. Ngược lại, khi P < 0.05 (mặc định) thì các mẫu đem so sánh khác nhau có ý nghĩa thống kê.

3.3. KIỂM ĐỊNH HAI MẪU BẰNG THỐNG KÊ T (T- TEST)

3.3.1. Kiểm định T cho hai mẫu độc lập

Kiểm định T đòi hỏi hai mẫu đem so sánh là những biến độc lập và ngẫu nhiên; số liệu của hai mẫu có phân bố chuẩn và phương sai mẫu bằng nhau. Giả thuyết H_0 : Hai mẫu cùng tổng thể (phân bố giống nhau); H_0^- - Hai mẫu khác tổng thể.

Kiểm định T tính theo công thức:

+ Khi phương sai hai mẫu bằng nhau

$$T = \frac{X_1 - X_2}{\sqrt{S^2_p \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} ; \quad (3.3)$$

Trong công thức 3.3: X_i – trung bình của nhóm i ; n_i – dung lượng mẫu của nhóm i ; S_i^2 – phương sai mẫu của nhóm i ; S_p^2 – phương sai gộp (Pooled Variance):

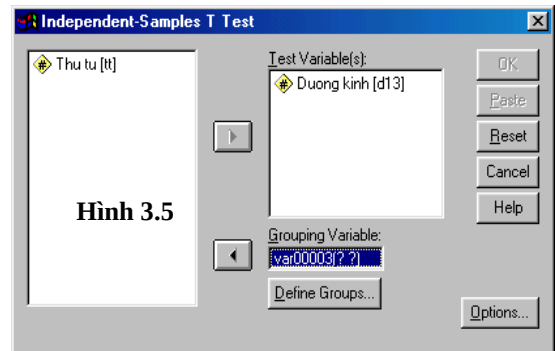
$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}$$

+ Khi phương sai hai mẫu không bằng nhau

$$T = \frac{X_1 - X_2}{\sqrt{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)}} \quad (3.4)$$

Để thực hiện kiểm định T, bạn nhập số liệu như sau : Cột 1 nhập kết quả thí nghiệm hay biến phụ thuộc (Dependent Variable), còn cột 2 là yếu tố thí nghiệm hay biến lập nhóm (Grouping Variable).

Gọi thủ tục tính toán như sau : Mở Analyze → Compare Means → Independent Samples T- test. Đến đây hãy đưa biến phụ thuộc (cột 1) vào ô Test Variables, còn biến lập nhóm (cột 2) vào ô Grouping Variable. Kế đến chọn Define Groups rồi ghi số 1 vào Group 1, còn số 2 vào Group 2. Tiếp theo chọn Options và đánh dấu những thủ tục cần tính. Sau cùng chọn OK (Hình 3.5).



Giải thích và nhận xét kết quả. Kết quả hiển thị trên màn Hình bao gồm những nội dung sau đây (Bảng 3.7):

Bảng 3.7. Kết quả của kiểm định 2 mẫu độc lập bằng tiêu chuẩn T

3.7a. Thống kê theo nghiệm thức

	VAR00003	N	Mean	Std. Deviation	Std. Error Mean
VAR00002	1.00	11	11.0364	2.3926	.7214
	2.00	11	10.9909	2.3097	.6964

3.7b. Kiểm định mẫu độc lập

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
Equal variances assumed	.028	.868	.045	20	.964	4.545E-02	1.0027	Lower	Upper
Equal variances not assumed			.045	19.975	.964	4.545E-02	1.0027	-2.0461	2.1370

× Phần thống kê mô tả gồm có X_{bq} , S^2 , S_x , Skewness, Kurtosis, Range...

× Phần kiểm định 2 mẫu độc lập bao gồm:

+ Kiểm định sự bằng nhau của phương sai mẫu bằng thống kê Levene's Test. Kết quả hiển thị thống kê F và P. Nếu $P > 0.05$ (mặc định) thì phương sai hai mẫu bằng nhau ở mức tin cậy 95%. Ngược lại, $P < 0.05$ (mặc định) thì phương sai hai mẫu khác nhau ở mức tin cậy 95%. Lưu ý rằng khi phương sai hai mẫu bằng nhau thì kết quả phân tích sẽ sử dụng phương sai gộp (Equal Variances Assumed). Ngược lại, khi phương sai hai mẫu khác nhau thì kết quả phân tích sẽ sử dụng phương sai riêng biệt (Equal Variances Not Assumed).

+ Kiểm định T về sự bằng nhau của trung bình mẫu (T- test for equality of means). Kết quả báo cáo thống kê T- Value, độ tự do (df) và xác suất hai phía (2- tail P), sự khác biệt

của hai số trung bình và khoảng tin cậy ở ngưỡng tin cậy 95%. Nếu $P > 0.05$ và khoảng tin cậy 95% của hiệu $X_{1bq} - X_{2bq}$ chứa số Zero thì hai mẫu khác nhau không có ý nghĩa thống kê. Ngược lại, khi $P < 0.05$ và khoảng tin cậy 95% của hiệu $X_{1bq} - X_{2bq}$ không chứa số Zero thì hai mẫu khác nhau một cách có ý nghĩa thống kê. Ở ví dụ trên, vì phương sai hai mẫu là bằng nhau, nên ta có thể sử dụng kết quả phân tích với phương sai gộp. Kết quả cho $P = 0.964$ và khoảng tin cậy 95% của hiệu $X_{1bq} - X_{2bq}$ chứa số Zero; do đó hai mẫu khác nhau không có ý nghĩa thống kê.

Lưu ý rằng, khi có K nghiệm thức ($K \geq 3$) nhưng chỉ cần so sánh từng cặp nghiệm thức với nhau thì bạn có thể thực hiện tương tự như so sánh 2 nghiệm thức. Nhưng khi kiểm định nhiều mẫu và cần phân hạng các nghiệm thức thì bạn phải áp dụng kiểm định Kruskal – Wallis hoặc dùng phương pháp phân tích phương sai (xem chương 4).

3.3.2. Kiểm định T cho hai mẫu bất cặp

Điều kiện áp dụng: Hai biến ngẫu nhiên và có phân bố chuẩn với phương sai bằng nhau. Giả thuyết H_0 : Hai mẫu cùng tổng thể (phân bố giống nhau); H_0^- - Hai mẫu khác tổng thể. Kiểm định T cho các mẫu bất cặp được tính theo công thức

$$T = \frac{D_{bq}}{\frac{S_D}{\sqrt{N}}} \quad (3.5)$$

Trong đó: D_{bq} – sự chênh lệch bình quân giữa hiệu số của các cặp quan sát; S_D – độ lệch chuẩn giữa hiệu số của các cặp quan sát; N – dung lượng quan sát.

Thực hành kiểm định T cho các mẫu bất cặp

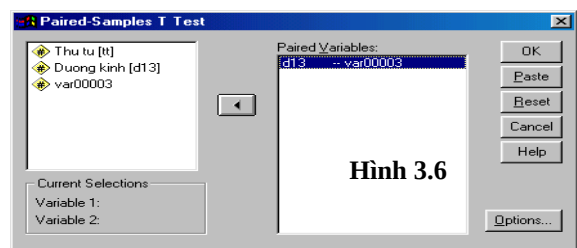
a. Trường hợp có hai mẫu. Giả sử hai điều tra viên A và B cùng đo đường kính thân cây của 15 cây gỗ. Kết quả được ghi lại ở **Bảng 3.8**. Hãy so sánh kết quả đo đặc của hai điều tra viên ? Để giải quyết vấn đề đặc ra, bạn có thể áp dụng kiểm định T cho hai mẫu bất cặp như sau :

Bảng 3.8. Số liệu đo đếm đường kính của 15 cây

A	18.5	18.5	18	17	17	17.5	18	18	14.5	14	19	20	22	22.5	27
B	18.1	18.1	17.6	16.6	16.6	17.1	17.6	17.8	13.9	13.4	18.4	19.8	21.4	22	26.5

Bước 1. Nhập số liệu vào Bảng tính SPSS. Cột 1 (hoặc hàng 1) nhập vào biến 1, còn cột 2 (hoặc hàng 1) nhập vào biến 2 (**Hình 3.6**).

Bước 2. Gọi chương trình tính toán. Trước hết, mở Analyze → Compare Mean → Paired simple T- test. Đến đây nhấn Shift và chọn cả hai biến, sau đó nhấp mũi tên bên cạnh để chuyển cả hai biến vào Paired Variable. Tiếp theo chọn Options và đánh dấu các mục cần tính rồi chọn OK.



Kết quả sẽ hiển thị trên màn Hình các thống kê mô tả, hệ số tương quan, sự khác biệt giữa các số bình quân, sai tiêu chuẩn, sai số chuẩn, khoảng tin cậy giữa hai số bình quân (**Bảng 3.9**).

Bảng 3.9. Kết quả của kiểm định 2 mẫu liên hệ bằng thống kê T**3.9a. Thống kê mô tả hai mẫu bất cặp**

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	D1	18.767	15	3.212	.829
	D2	18.327	15	3.215	.830

3.9b. Tương quan giữa hai mẫu bất cặp

		N	Correlation	Sig.
Pair 1	D1 & D2	15	.999	.000

3.9c. Kiểm định T của hai mẫu bất cặp

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	D1 - D2	.440	.130	3.352E-02	.368	.512	13.125	14	.000

Giải thích và nhận xét kết quả. Căn cứ vào hệ số R để xác định mức độ chặt chẽ của mối quan hệ giữa hai mẫu. Nếu $P > 0.05$ hoặc khoảng tin cậy giữa hiệu hai trung bình mẫu chứa số Zero thì hai mẫu khác nhau không có ý nghĩa thống kê. Ngược lại, nếu $P < 0.05$ hoặc khoảng tin cậy giữa hiệu hai trung bình mẫu không chứa số Zero thì hai mẫu khác nhau có ý nghĩa thống kê. Ở ví dụ này, vì $P = 0.000$ và khoảng tin cậy giữa hiệu hai trung bình mẫu không chứa số Zero nên hai mẫu là khác nhau.

b. Trường hợp có nhiều mẫu. Khi có K mẫu (nghiệm thức) cần so sánh từng cặp riêng rẽ với nhau thì cách thực hiện cũng tương tự như so sánh hai mẫu. Ví dụ: Có 3 nghiệm thức cần so sánh từng cặp với nhau. Trước hết, bạn so sánh nghiệm thức 1 với nghiệm thức thứ 2, nghiệm thức 1 với nghiệm thức 3; tiếp theo so sánh nghiệm thức 2 với nghiệm thức 3...Lưu ý rằng, kiểm định T chỉ cho biết sự khác biệt giữa hai nghiệm thức ở mức xác suất sai lầm được chỉ định (thường là 0.05 hoặc 0.01). Nhưng khi so sánh nhiều nghiệm thức với nhau thì bạn không thể sử dụng kiểm định T mà phải sử dụng kiểm định nhiều cặp (Pairwise multiple comparisons). Bởi vì xác suất liên quan đến thống kê T chỉ giả thiết có 1 kiểm định được thực hiện. Ngược lại, khi có k nghiệm thức ($k \geq 3$) cần kiểm định từng cặp thì xác suất tìm ra một cặp sai khác có ý nghĩa thống kê sẽ tăng lên rất nhanh theo số cặp cần so sánh. Giả thiết có k cặp số trung bình cần so sánh ở mức ý nghĩa 5% thì xác suất khác nhau rơi vào miền bác bỏ là $1 - 0,95^k$. Ví dụ, khi có 6 cặp nghiệm thức cần so sánh từng cặp ở mức ý nghĩa 5% thì xác suất khác nhau rơi vào miền bác bỏ là $0,265 = 1 - 0,95^6$.

3.3.3. Kiểm định T hai mẫu với việc sử dụng một điểm giới hạn (Cut Point) để phân chia các nhóm

Giả sử bạn cần phân chia một tập hợp lớn số liệu (N quan sát) thành hai nhóm; đồng thời kiểm định sự khác nhau giữa trung bình của hai nhóm. Ví dụ: Bạn cần phân biệt hai nhóm cây có cấp sinh trưởng khác nhau, đồng thời kiểm định sự khác biệt về D, H, G...giữa hai nhóm quan sát này. Điều này có thể được giải quyết dễ dàng bằng kiểm định T (có thể sử dụng cho

cả mẫu độc lập lẫn mẫu liên hệ). Nhưng khi kiểm định cho k yếu tố, bạn cần chú ý đến phép đa kiểm định. Phép đa kiểm định đòi hỏi phải điều chỉnh mức ý nghĩa (xác suất) của k yếu tố. Ví dụ: Khi thực hiện cùng một lúc 5 kiểm định (5 yếu tố) và mong muốn mức ý nghĩa chung cho 5 khoảng tin cậy của số trung bình là 95%, thì bạn phải chia mức ý nghĩa $\alpha = 0.05$ cho 5 yếu tố, nghĩa là $0.05/5 = 0.01$. Từ đó bạn phải yêu cầu một khoảng tin cậy 99% (nghĩa là $1 - 0.01$) để có được một khoảng tin cậy thích hợp cho cả 5 kiểm định ở mức tin cậy chung là 95%.

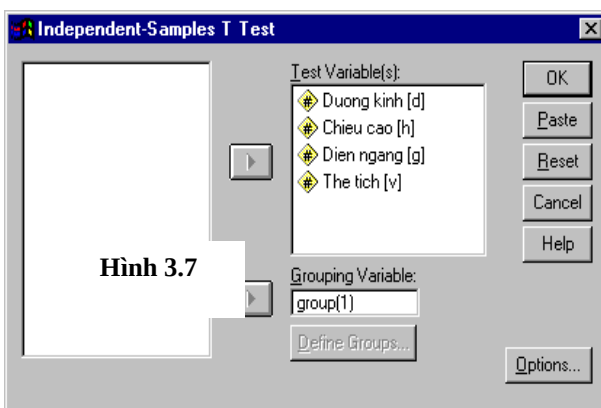
Giả thiết một tập hợp có N quan sát và mỗi quan sát có K tham số. Hãy chọn một điểm giới hạn để phân chia dãy quan sát thành hai nhóm. Điểm giới hạn có thể được chọn theo chủ định riêng của bạn. Ví dụ: Cùng một lúc bạn muốn kiểm định bốn biến số D, H, G, V của một quần thụ và dùng một điểm giới hạn để phân chia chúng thành hai nhóm. Để đạt được mục đích đặt ra, bạn có thể thực hiện theo thủ tục như sau:

Bước 1. Trước hết, bạn nhập bốn biến số D, H, G, V của từng cây vào 4 cột kế liên nhau. Kế đến bạn có thể chọn biến đường kính ($D_{1.3}$) và tính các hệ số đường kính thân cây bằng cách chia đường kính của từng cây riêng biệt cho đường kính trung bình của quần thụ, nghĩa là $D_{1.3}/D_{1.3bq}$. Sau đó bạn sử dụng hệ số $1 = D_{1.3}/D_{1.3bq}$ làm điểm giới hạn hai nhóm. Nhóm 1 bao gồm những cây có $D < D_{1.3}$, còn nhóm 2 có $D > D_{1.3}$. Các hệ số này được ghi vào cột thứ 5 của Bảng tính.

Bước 2. Xác định mức tin cậy chung cho 4 khoảng tin cậy của 4 số trung bình. Vì mong muốn mức tin cậy chung cho 4 khoảng tin cậy của 4 số trung bình là 95%, nên bạn phải chia mức ý nghĩa $\alpha = 0.05$ cho 4 yếu tố, nghĩa là $0.05/4 = 0.0125$. Từ đó bạn phải yêu cầu một khoảng tin cậy 99% (làm tròn từ $1 - 0.0125 = 0.9875$) để có được một khoảng tin cậy thích hợp cho cả 4 phép kiểm định ở mức tin cậy chung là 95%.

Bước 3. Gọi chương trình tính toán. Mở Analyze $\Rightarrow$ Compare means $\Rightarrow$ Independent Samples T Test. Khi cửa sổ Independent Samples T Test xuất hiện, bạn hãy chuyển cả 4 biến số cần kiểm định vào ô Test Variables; chuyển biến lập nhóm vào ô Grouping Variable. Kế đến mở Define Groups, chọn Cut Point = 1 rồi chọn continue (Hình 3.7). Điểm giới hạn bằng 1 sẽ phân chia dữ liệu thành hai nhóm: nhóm 1 là những cây có tỷ số $D_{1.3}/D_{bq} \leq 1$, còn nhóm hai có $D_{1.3}/D_{bq} > 1$. Cuối cùng chọn OK để chương trình tính toán và hiển thị kết quả (Bảng 3.10 và 3.11).

Giải thích và nhận xét kết quả. Bảng 3.10 cho biết các thống kê mô tả của hai nhóm như dung lượng mẫu (N), trung bình mẫu (Means), sai tiêu chuẩn mẫu (Std. Deviation), sai số chuẩn của trung bình mẫu (Std. Error Mean). Bảng 3.11 cho biết các thống kê kiểm định về ngang bằng phương sai của hai nhóm (Levene's Test for Equality of Variances), về ngang bằng hai số trung bình bằng T test, về sự khác biệt giữa hai số trung bình (Mean Difference), khoảng khác biệt của hai số trung bình mẫu ở mức tin cậy 95% (95% Confidence Interval of the Difference)... Ở ví dụ này, phương sai mẫu của các biến D, H là bằng nhau. Vì thế, bạn có thể sử dụng kết quả ứng với dòng phương sai gộp (Equal variances assumed). Đối với D, vì $P = 0.141 > 0.05$ (cột 7) nên hai mẫu khác nhau không rõ rệt. Đối với H, vì $P = 0.005 < 0.05$ (cột 7) nên hai mẫu có chiều cao khác nhau rõ rệt.



Hình 3.7

Bảng 3.10. Thống kê theo nhóm (Group Statistics)

	Grouping variable	N	Mean	Std. Deviation	Std. Error Mean
D _{1.3}	>= 1.00	5	7.580	.512	.229
	< 1.00	4	7.025	.479	.239
H	>= 1.00	5	8.100	.418	.187
	< 1.00	4	7.000	.408	.204
G	>= 1.00	5	0.0044	0.0005	0.0002
	< 1.00	4	0.0040	0.0008	0.0004
V	>= 1.00	5	0.0182	0.0033	0.0015
	< 1.00	4	0.0138	0.0024	0.0012

Bảng 3.11. Kiểm định mẫu độc lập (Independent Samples Test)

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
1	2	3	4	5	6	7	8	9	10	11
D_{1.3}	Equal variances assumed	.085	.779	1.662	7	.141	.555	.334	-.235	1.345
	Equal variances not assumed			1.676	6.758	.139	.555	.331	-.234	1.344
H	Equal variances assumed	.174	.689	3.960	7	.005	1.100	.278	.443	1.757
	Equal variances not assumed			3.973	6.642	.006	1.100	.277	.438	1.762
...	...	...								

3.4. KIỂM ĐỊNH LÔ (RUNS TEST)

Lô (Runs, kí hiệu là R) là sự xuất hiện liên tiếp của những quan sát giống nhau. Ví dụ: Khi điều tra 20 ô dạng bản (N = 20), ta nhận được 11 ô có cây tái sinh (kí hiệu là n₁), 9 ô không có cây tái sinh (kí hiệu là n₂). Những ô có cây và không có cây tái sinh lập thành 10 cụm hay 10 lô (R = 10) như sau : +0+++00++++000++00+0. Ở đây, lô 1 gồm 1 ô có cây, lô 2 gồm 1 ô không có cây, lô 3 gồm 3 ô có cây...Kiểm định “lô” là kiểm định giả thuyết về tính ngẫu nhiên trong sự phân bố của các lô mà các quan sát được ghi nhận bằng hai khả năng “có” và “không có”. Khi áp dụng kiểm định lô để nghiên cứu phân bố cây trên mặt đất, ta phải kiểm định giả thuyết H₀: phân bố số lô trống và lô có cây (R) là ngẫu nhiên. Để kiểm định giả thuyết trên, người ta dùng công thức kiểm định T:

$$T = \frac{R - \frac{2n_1 n_2}{n_1 + n_2} + 1}{\sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}}} \quad (3.6)$$

Xác suất $T_{\text{lý thuyết}}$ tìm theo Bảng phân bố T với mức ý nghĩa $P = 0,05$ hoặc $P = 0,01$. Giả sử cần xác định kiểu phân bố cây tái sinh trên mặt đất. Để đạt được mục đích này, bạn có thể vận dụng kiểm định lô. Cách tiến hành như sau:

Bước 1. Bố trí thí nghiệm. Khi vận dụng kiểm định lô, trước hết các ô dạng bản dùng để thống kê cây tái sinh được bố trí liên tục trên các tuyến (dải) song song cách đều hoặc không đều dưới tán rừng trong lô thí nghiệm (quần xã thực vật). Bề rộng của tuyến được chọn lựa tùy thuộc vào loài cây nghiên cứu, thông thường là 2 đến 4 mét, còn chiều dài của tuyến được bố trí tùy hứng. Tiếp theo chia các tuyến thành các phân đoạn đều nhau với chiều dài mỗi phân đoạn bằng bề rộng của tuyến; mỗi phân đoạn được xem là một ô dạng bản (Hình 3.8).

Tuyến I	+	+	-	+	-	-	-	2 m
II	-	+	+	-	+	-	+	2 m
III	+	+	-	-	+	-	+	+

Hình 3.8. Sơ đồ bố trí các ô dạng bản trong quần xã thực vật

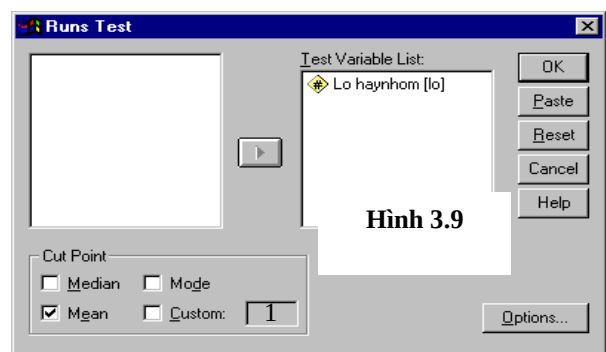
Kí hiệu: (+) và (-) - tương ứng là ô “có” và ô “không có” cây tái sinh của loài quan tâm

Bước 2. Thu thập thông tin trên ô dạng bản. Thông tin về cây tái sinh trong mỗi ô dạng bản (mỗi phân đoạn) gồm có hai dấu hiệu định tính là “có” - kí hiệu bằng dấu (+) hoặc số 1 và “không có” - kí hiệu bằng dấu (-) hoặc số 0. Giả sử điều tra 80 ô dạng bản, bạn nhận được $R = 36$, $n_1 = 44$, $n_2 = 36$.

Bước 3. Kiểm tra dạng phân bố cây tái sinh trên mặt đất theo những thủ tục sau đây:

+ Nhập số liệu vào Bảng tính. Cột 1 ghi thứ tự ô dạng bản, cột 2 ghi mã số ô bắt gặp hay không bắt gặp cây tái sinh; trong đó ô có cây ghi bằng số 1, còn ô không có cây ghi bằng số 0.

+ Gọi chương trình tính toán. Từ Menu chính, mở Analyze $\Rightarrow$ Nonparametric tests $\Rightarrow$ Runs. Khi cửa sổ Runs tests xuất hiện (Hình 3.9), bạn chuyển chuỗi số liệu vào ô Test variable list. Kế đến đánh dấu chọn mục Mean, rồi chọn tiếp Custom trong ô Cut Point. Tiếp theo bạn ghi số 1 vào ô bên cạnh mục Custom. Sau cùng bạn chọn OK để chương trình tính toán và hiển thị kết quả.



Hình 3.9

Giải thích và nhận xét kết quả. Kết quả hiển thị bao gồm 1 Bảng kiểm định lô như Bảng 3.12. Giá trị Test Value = 1 – đó là giá trị do bạn chỉ định thông qua chọn Custom với điểm chặn (Cut Point) = 1. Ở đây bạn cũng có thể chọn Cut point theo Mode, Median hay Mean. Giá trị Total Cases = 80 – đó là tổng số mẫu nghiên cứu ($N = 80$). Number of Runs =

36 – đó là số lô R. Thống kê $Z = -1.046$ chính là giá trị T được tính từ công thức 3.6. Kí hiệu Asymp. Sig. (2-tailed) hay P-value – đó là mức ý nghĩa của kiểm định. Nếu P-value > 0.05 (mức định) thì giả thuyết H_0 cho rằng phân bố số lô trống và lô có cây (R) là ngẫu nhiên sẽ được chấp nhận. Ngược lại, khi P-value < 0.05 (mức định) thì giả thuyết H_0 sẽ bị bác bỏ. Ở ví dụ này, vì $P = 0.296 > 0.05$ nên giả thuyết H_0 được chấp nhận.

Bảng 3.12. Kiểm định Lô (Runs Test)

	Lô hay nhóm
Test Value	1
Total Cases	80
Number of Runs	36
Z	-1.046
Asymp. Sig. (2-tailed)	.296

Theo Terenchiev (1964), nếu $T_{tt} < -2$ thì phân bố của loài cây quan tâm trên mặt đất có dạng cụm. Khi $T_{tt} = 0 \div \pm 2$ thì phân bố của loài cây quan tâm trên mặt đất có dạng phân bố ngẫu nhiên. Khi $T_{tt} > +2$ thì phân bố của loài cây quan tâm trên mặt đất có dạng phân bố đều.

Chương IV

PHÂN TÍCH PHƯƠNG SAI

4.1. KHÁI QUÁT

Phân tích phương sai (ANOVA) là một kỹ thuật rất hữu ích cho những vấn đề thống kê như kiểm định ảnh hưởng của một hoặc nhiều yếu tố không đo đạc được (biến độc lập) đến yếu tố phụ thuộc (biến phản hồi – Response variable, biến kết quả). Trong kỹ thuật phân tích phương sai, những biến không đo đạc được biểu thị cho những yếu tố (Factors) hoặc biến phân loại (Classification variables). Mặc dù phân tích phương sai là kỹ thuật dùng để phân tích những yếu tố không đo đạc được, nhưng nó vẫn có thể được sử dụng để phân tích những biến đo đạc được nếu những biến ấy có thể ghép thành nhóm hoặc khối (Block). Ví dụ, bạn có thể sử dụng phân tích phương sai để xem xét ảnh hưởng của độ dày tầng đất (được phân thành cấp có thứ bậc) đến năng suất rừng.

Tùy theo yếu tố tác động và kiểu bố trí thí nghiệm, bạn có thể xây dựng các mô hình phân tích phương sai khác nhau. Phân tích phương sai cũng có thể được thiết kế theo mô hình tuyến tính tổng quát (General Linear Model = GML). Để giải kiểu bài toán này, bạn có thể sử dụng thủ tục mô hình tuyến tính tổng quát trong SPSS. Mô hình tuyến tính tổng quát là một tập hợp các giả định thống kê mà qua đó phân tích tương quan, phân tích hồi quy và phân tích biến động đều dựa trên những giả định này. Mô hình tuyến tính tổng quát là một thành phần của họ các mô hình tuyến tính tổng quát được Nelder và Wedderburn (1972) giới thiệu lần đầu. Phân tích mô hình tuyến tính tổng quát là công cụ rất mềm dẻo, đặc biệt đối với các mô hình bao gồm cả yếu tố phân loại và yếu tố định lượng. Chẳng hạn, bạn có thể dùng nó để phân tích biến động (ANOVA), phân tích hiệp phương sai (ANOCOVA), phân tích đa hiệp phương sai (ANOMUCOVA)... Nói chung, thủ tục GML có thể giúp bạn chỉ rõ nhiều hơn một biến phụ thuộc, làm phù hợp mô hình mà bạn chọn cho mỗi biến phụ thuộc, xác định những ảnh hưởng chính và ảnh hưởng tổng hợp, tạo ra các Bảng chứa các giá trị dự đoán và các sai lệch từ việc phân tích, chọn kiểu sai số được kết hợp với mỗi tác động trong mô hình. Tóm lại, bạn có thể sử dụng mô hình GML để giải quyết các vấn đề sau đây:

- Khi bạn có hai yếu tố đan chéo nhau và hai yếu tố phân cấp. Chẳng hạn như thiết kế lô chính - phụ và thiết kế đo lặp.
- Khi bạn muốn đưa vào các tương phản (contrasts) trong phân tích biến động.
- Khi bạn muốn thu được thống kê cho nhiều biến phụ thuộc (MANOVA)...

Phân tích biến động (1 yếu tố, đa yếu tố và thiết kế phân cấp) giả định rằng số liệu tuân theo phân bố chuẩn với phương sai bằng nhau. Ba thủ tục phân tích biến động trên đây được tóm tắt như sau :

Thủ tục	Số biến	Kiểu số liệu	Mô tả
Phân tích ANOVA 1 yếu tố	2	Số	Phân tích ảnh hưởng của 1 yếu tố định tính hoặc định lượng đến 1 yếu tố phụ thuộc. Các thiết kế cân bằng hoặc không cân bằng.
Phân tích ANOVA nhiều yếu tố	Nhiều	Số	Phân tích ảnh hưởng của 1 hoặc nhiều yếu tố định tính hoặc định lượng đến 1 yếu tố phụ thuộc. Phân tích Covariance với 1 hoặc nhiều Covariance.
Phân tích ANOVA phân cấp (Nested designs)	Nhiều	Số	Phân tích ảnh hưởng của 1 hoặc nhiều yếu tố đến 1 yếu tố phụ thuộc khi số mẫu ở tất cả các tổ hợp của các mức yếu tố là bằng nhau.

Chương IV giới thiệu nội dung thực hành phân tích phương sai. Phần đầu giới thiệu khái quát về các kiểu bố trí thí nghiệm. Phần tiếp theo giới thiệu các thủ tục trong thực hành phân tích phương sai và hiệp phương sai theo mô hình tuyến tính tổng quát.

4.2. CÁC MÔ HÌNH PHÂN TÍCH PHƯƠNG SAI

Kết quả của phân tích phương sai báo cáo nhiều nội dung khác nhau; trong đó chúng ta thường lưu ý nhiều hơn đến những nội dung sau đây: (1) Các đặc trưng thống kê mô tả; (2) Kiểm định phân bố chuẩn và ngang bằng phương sai; (3) Phân tích các biến động (ANOVA); (4) So sánh giữa các nghiệm thức và kiểm định phân hạng các nghiệm thức; (5) Xác định các tương phản (Contrasts)...Nội dung 4 và 5 chỉ được thực hiện khi phân tích các biến động cho biết các trung bình mẫu khác nhau có ý nghĩa thống kê.

+ **So sánh giữa các nghiệm thức và kiểm định phân hạng.** So sánh sự khác nhau giữa các nghiệm thức được thực hiện theo hai cách sau đây: (1) So sánh từng cặp nghiệm thức với nhau (Pairwise comparisons), (2) So sánh nhiều cặp nghiệm thức với nhau (Pairwise multiple comparisons). Khi muốn phân hạng các nghiệm thức, người ta sử dụng các kiểm định phân hạng các nghiệm thức (Post Hoc range tests). Các kiểm định phân hạng các nghiệm thức cho biết những giá trị trung bình nào khác nhau có ý nghĩa thống kê. Chúng được thực hiện dựa trên các giá trị trung bình chưa được điều chỉnh. Các kiểm định này chỉ được áp dụng để so sánh ảnh hưởng giữa các yếu tố cố định đến một biến phụ thuộc. Khi có nhiều biến phụ thuộc, thì các kiểm định này sẽ thực hiện cho từng biến.

Khi so sánh sai khác giữa hai nghiệm thức với nhau, người ta có thể sử dụng kiểm định T. Kiểm định T cho biết sự khác biệt giữa hai nghiệm thức ở mức xác suất sai lầm được chỉ định trước (Thông thường là $P = \alpha = 0.05$ hoặc 0.01). Nhưng khi có nhiều nghiệm thức cần so sánh từng cặp với nhau, chúng ta phải sử dụng kiểm định nhiều cặp (Pairwise multiple comparisons). Bởi vì xác suất liên quan đến thống kê T chỉ giả định có 1 kiểm định được thực hiện. Ngược lại, khi có k nghiệm thức ($k > 2$) cần kiểm định từng cặp, thì xác suất tìm ra một cặp sai khác ở mức ý nghĩa P sẽ tăng lên theo số cặp cần so sánh. Giả thiết có k cặp số trung bình cần so sánh ở mức ý nghĩa $P = 0,05$, thì xác suất khác nhau rơi vào miền bác bỏ là $1 - 0,95^k$. Ví dụ: Khi có 6 cặp nghiệm thức cần so sánh từng cặp ở mức ý nghĩa $\alpha = 0,05$, thì xác suất khác nhau rơi vào miền bác bỏ là $1 - 0,95^6 = 0,265$. Giả sử có k nghiệm thức, số cặp so sánh là $Y = k(k-1)/2 = k!/(N-k!)$; trong đó $N =$ dung lượng quan sát.

Trong SPSS, khi có nhiều nghiệm thức nhưng chỉ cần so sánh từng cặp nghiệm thức riêng rẽ với nhau, người ta có thể sử dụng các kiểm định: Sự khác nhau có ý nghĩa nhỏ nhất (Least Significant Difference = LSD), Bonferroni và Sidak. Nhưng nếu có nhiều nghiệm thức cần so sánh không chỉ từng cặp với nhau mà còn xếp hạng các nghiệm thức, chúng ta cần sử dụng các kiểm định Tukey HSD, Scheffe, Duncan, Student - Newman - Keuls. Tất cả các kiểm định này đòi hỏi phương sai mẫu là bằng nhau.

Kiểm định Tukey sử dụng thống kê biên độ được Student hóa để xây dựng tất cả các so sánh cặp giữa các nhóm nghiệm thức và thiết lập tỉ lệ sai số cho những cặp so sánh. Khi có k cặp so sánh ở mức ý nghĩa $\alpha = 0,05$, thì mức ý nghĩa của 1 cặp so sánh là $0,05/k$. Ví dụ: Khi $k = 5$ và $\alpha = 0,05$, thì mức ý nghĩa cho 1 cặp so sánh là $0,05/5 = 0,01$. Vì thế, kiểm định Tukey được dùng để so sánh nhiều cặp trung bình ở mức ý nghĩa α . Kiểm định Bonferroni sử dụng thống kê Student's T để điều chỉnh mức ý nghĩa quan sát dựa trên số lần so sánh. Khi chỉ có ít cặp trung bình cần so sánh, thì kiểm định Bonferroni hữu dụng hơn. Các kiểm định Duncan, Student-Newman-Keuls (S-N-K) và Tukey's b là các kiểm định dựa trên hạng của các trung bình mẫu. Kiểm định LSD tương đương với kiểm định T. Hạn chế của kiểm định LSD là không có một kiểm định nào được điều chỉnh mức tin cậy cho nhiều so sánh.

Thực chất của phép so sánh giữa các nghiệm thức là xây dựng khoảng tin cậy của so sánh (C_i) và dựa vào khoảng tin cậy để kiểm định giả thuyết H_0 (Không có sự sai khác giữa hai nghiệm thức). Nếu khoảng tin cậy C_i chứa giá trị Zero, thì giả thuyết H_0 được chấp nhận ở mức xác suất sai lầm $P = \alpha$ hoặc mức tin cậy $1-P$. Ngược lại, nếu khoảng tin cậy C_i không chứa giá trị Zero, thì giả thuyết H_0 bị bác bỏ ở mức xác suất sai lầm P hoặc mức tin cậy $1-P$. Ba mức xác suất sai lầm thường được sử dụng là $\alpha = 0,05$; $\alpha = 0,01$ và $\alpha = 0,001$. Trong lĩnh vực lâm nghiệp, người ta thường chọn mức xác suất sai lầm $\alpha = 0,05$.

Dạng tổng quát của khoảng tin cậy của so sánh C_i như sau:

$$C_i = C_i^{(*)} \pm S_{C_i} \times T_p^{(a)}(k)$$

Trong đó:

C_i = hiệu số của hai trung bình tổng thể hay hai trung bình mẫu;

$C_i^{(*)}$ = ước lượng của C_i hay hiệu số của hai trung bình mẫu;

S_{C_i} = độ lệch chuẩn của $C_i^{(*)}$:

$$S_{C_i} = \sqrt{\frac{2MSE}{n}} \quad (\text{khi dung lượng hai mẫu bằng nhau});$$

$$S_{C_i} = \sqrt{\frac{(1/n_i + 1/n_j)MSE}{n}} \quad (\text{khi dung lượng hai mẫu khác nhau});$$

SE = sai số chuẩn;

$T_p^{(a)}(k)$ = đại lượng có phân phối Student hoặc phân phối F tùy theo kiểu so sánh đối chiếu;

α = mức xác suất sai lầm (còn kí hiệu là P – value);

a = số nghiệm thức (với các trắc nghiệm không xếp hạng các nghiệm thức) hoặc hạng (l) của cặp nghiệm thức đem so sánh (với các trắc nghiệm có xếp hạng các nghiệm thức);

k = bậc tự do của sai số trung bình bình phương (MSE).

Giá trị $T_p^{(a)}(k)$ được xác định tùy theo kiểu so sánh.

(a) Đối với những trắc nghiệm không xếp hạng các nghiệm thức

+ Trắc nghiệm LSD: $T_p^{(a)}(k) = T_p(k)$. Trắc nghiệm LSD không phụ thuộc vào số nghiệm thức cần so sánh, nghĩa là nó chỉ so sánh từng cặp nghiệm thức.

+ Trắc nghiệm Scheffe: $T_p^{(a)}(k) = \sqrt{(l-1)Fp(l-1; k)}$. Trắc nghiệm Scheffe phụ thuộc vào số nghiệm thức so sánh.

+ Trắc nghiệm Dunnett: $T_p^{(a)}(k)$. Trắc nghiệm này phụ thuộc vào số nghiệm thức so sánh và được tra từ Bảng lập sẵn dùng cho trắc nghiệm Dunnett.

(b) Đối với những trắc nghiệm có xếp hạng các nghiệm thức

+ Trắc nghiệm Tukey: $T_p^{(a)}(k) = \frac{q_p^{(a)}(k)}{\sqrt{2}}$; trong đó $q_p^{(a)}(k)$ là trị số tra Bảng lập sẵn dùng cho trắc nghiệm Tukey. Trắc nghiệm Tukey phụ thuộc vào số nghiệm thức so sánh.

+ Trắc nghiệm Duncan: $T_p^{(a)}(k) = \frac{q_p^{(a)}(k)}{\sqrt{2}}$; trong đó $q_p^{(a)}(k)$ là trị số tra Bảng lập sẵn dùng cho trắc nghiệm Duncan. Trắc nghiệm Duncan phụ thuộc vào hạng của cặp nghiệm thức so sánh.

+ Trắc nghiệm Newman - Keuls: $T_p^{(a)}(k) = \frac{q_p^{(a)}(k)}{\sqrt{2}}$; trong đó $q_p^{(a)}(k)$ là trị số tra Bảng lập sẵn dùng cho trắc nghiệm Newman – Keuls. Trắc nghiệm Newman - Keuls cũng phụ thuộc vào hạng của cặp nghiệm thức so sánh.

Ở những trắc nghiệm không xếp hạng các nghiệm thức (LSD, Scheffe và Dunnett), giá trị $S_{C_i} \times T_p(k)$ là sai khác nhỏ nhất có ý nghĩa so sánh (LSD). Quy tắc quyết định: (1) Nếu $C_i^{(*)} < S_{C_i} \times T_p(k)$, thì giá trị trung bình của 2 nghiệm thức khác nhau không rõ rệt ở mức ý nghĩa sai lầm P . (2) Nếu $C_i^{(*)} > S_{C_i} \times T_p(k)$, thì giá trị trung bình của 2 nghiệm thức khác nhau rõ rệt ở mức ý nghĩa sai lầm α .

Đối với những trắc nghiệm có xếp hạng các nghiệm thức (Tukey, Duncan, Newman - Keuls), trị số $S_{Ci} \times T_p^{(a)}(k)$ là hạng sai khác nhỏ nhất ở mức ý nghĩa sai lầm α . Khi $|C_I^{(*)}| \geq S_{Ci} \times T_p^{(a)}(k)$, thì khoảng tin cậy của C_i không chứa trị số Zero. Do đó, giả thuyết H_0 bị bác bỏ ở mức ý nghĩa sai lầm α . Ngược lại, khi $|C_I^{(*)}| < S_{Ci} \times T_p^{(a)}(k)$, thì khoảng tin cậy của C_i chứa giá trị Zero. Do đó, giả thuyết H_0 được chấp nhận ở mức ý nghĩa sai lầm α .

Lưu ý rằng tất cả các kiểm định Tukey HSD, LSD, Scheffe, Bonferroni, Duncan, Newman - Keuls và Sidak đòi hỏi phương sai mẫu là bằng nhau. Khi phương sai mẫu khác nhau, thì chúng ta sử dụng các kiểm định Games – Howell, Tamhane's T_2 , Dunnett's T_3 và Dunnett's C.

Nếu cần so sánh từng giá trị trung bình với một giá trị trung bình kiểm tra, người ta có thể sử dụng kiểm định Dunnett. Giá trị trung bình kiểm tra có thể là trị trung bình của nhóm thứ nhất hoặc nhóm cuối cùng. Chúng ta cũng có thể chọn kiểm định một bên (1-sided) hoặc hai bên (2-sided). Kiểm định hai bên áp dụng với các giả thuyết cho rằng trung bình ở một mức yếu tố nào đó (Ngoại trừ yếu tố kiểm tra) khác với hay không bằng yếu tố kiểm tra. Ở trường hợp này, sau khi chọn thống kê Dunnett, chúng ta chọn tiếp tiêu chuẩn kiểm tra là Last hoặc First và kiểm định hai phía (2-sided). Kiểm định 1 bên áp dụng đối với các giả thuyết cho rằng trung bình ở một mức yếu tố nào đó (Ngoại trừ yếu tố kiểm tra) nhỏ hơn hoặc lớn hơn yếu tố kiểm tra. Để đạt mục đích này, sau khi chọn thống kê Dunnett, chúng ta chọn tiếp tiêu chuẩn kiểm tra là Last hoặc First và kiểm định nhỏ hơn kiểm tra (<Control) hoặc lớn hơn kiểm tra (>Control).

(c) Kiểm định tương phản giữa các trung bình mẫu. Phân tích phương sai với kiểm định F chỉ đưa ra một kết luận chung chung cho một câu hỏi cũng chung chung. Đó là có tồn tại sự khác biệt giữa các trung bình mẫu hay không? Khi P – value < 0.05 (Mặc định), thì kết quả kiểm định chỉ cho biết các trung bình mẫu khác nhau ở mức xác suất P (0.05 hay 0.01). Muốn biết rõ những nghiệm thức nào khác nhau ở mức xác suất P, chúng ta có thể sử dụng kiểm định T cho từng cặp trung bình mẫu (Pairwise comparisons).

Để kiểm định nhiều trung bình mẫu và ghép chúng thành những nhóm thuần nhất (Những trung bình mẫu không khác nhau ở mức xác suất P), chúng ta sử dụng các so sánh đa phạm vi (Multiple comparisons) và kiểm định phân hạng các nghiệm thức (Multiple range tests). Trái lại, trong một số trường hợp khác, chúng ta không chỉ mong muốn những so sánh đơn giản như thế mà còn cần đến những so sánh sâu hơn như so sánh một hoặc một số nhóm này với một hoặc một số nhóm khác. Chẳng hạn: Có 4 thí nghiệm bón phân cho Sao đen: (1) Đối chứng (không bón phân), (2) Bón P, (3) Bón K, (4) Bón PK. Đối với kiểu thí nghiệm này, chúng ta có thể đặt ra yêu cầu: Hãy cho biết sinh trưởng của Sao đen trên những nghiệm thức bón lân có sai khác thực sự với những nghiệm thức không bón lân hay không? Để trả lời câu hỏi này, chúng ta phải so sánh giữa tập hợp trung bình mẫu của những công thức có bón lân (P và PK) với đối chứng và K trung bình mẫu của những công thức không bón lân (Đối chứng và K). Phương pháp kiểu kiểm định như trên đây được gọi là so sánh tương phản.

Về lý thuyết, một tương phản là một quan hệ tuyến tính của K số trung bình. Tương phản ở tổng thể có dạng $\psi = \sum^K a_i \mu_i$, còn ở mẫu có dạng $C = \sum^K a_i X_{bqi}$; trong đó $\sum a_i = 0$. Như vậy, các so sánh tương phản được sử dụng để kiểm định cho những khác biệt giữa các mức của một yếu tố. Chúng ta cũng có thể chỉ định một tương phản cho mỗi yếu tố trong mô hình.

Khi thí nghiệm với 1 yếu tố, so sánh tương phản bằng cách mở thủ tục Compare means: One – Way ANOVA hoặc General linear model: Univariate. Nhưng khi thí nghiệm với nhiều yếu tố, thì chúng ta chỉ cần đến thủ tục mô hình tuyến tính tổng quát (General linear model: Univariate).

(C₁) Đối với thủ tục One-Way ANOVA Contrasts. Khi chọn Polynomial (đa thức), chương trình sẽ chia tổng bình phương giữa các nhóm thành các thành phần xu hướng. Chúng ta có thể kiểm định khuynh hướng biên đổi của biến phụ thuộc theo các mức có thứ bậc của

biến yếu tố. Ví dụ: Chúng ta có thể kiểm định khuynh hướng tuyến tính (tăng hoặc giảm) đối với chiều cao thân cây theo thứ bậc cao dần của hàm lượng phân bón trong ruột bầu. Sau khi chọn Polynomial, rồi chọn tiếp Degree (bậc) thì chương trình sẽ chia tổng bình phương giữa các nhóm thành các thành phần xu hướng theo đa thức. Khi phân tích xu hướng biến đổi của các trung bình nghiệm thức, chúng ta có thể chọn lần lượt từ bậc 1 đến bậc 5 (Mặc định là lớn nhất). Để kiểm định các nghiệm thức bằng thống kê T, chúng ta cũng phải chỉ định các hệ số (Coefficients) cho những tương phản cụ thể. Để đạt được mục đích này, chúng ta đưa vào các hệ số cho mỗi nhóm (loại) của biến yếu tố (phân loại) và chọn Add sau mỗi lần nhập một hệ số. Theo mặc định, chúng ta có thể nhập vào 50 hệ số (nhóm). Khi đưa vào ô Coefficients một giá trị mới của một hệ số, thì nó sẽ được ghi vào dòng cuối của danh sách hệ số. Chúng ta có thể chỉ định cùng một lúc nhiều tập hợp (bộ) tương phản khác nhau (lớn nhất là 10) bằng cách nhấp nút Next sau mỗi tập hợp giá trị. Sử dụng nút Next và Previous để thay đổi qua lại giữa các bộ tương phản. Khi gõ nhầm một hệ số nào đó, chúng ta có thể hủy bỏ nó bằng cách chọn nó và nhấp nút Remove.

Lưu ý rằng, trật tự của các hệ số được đưa vào ô hệ số (Coefficients) là rất quan trọng, bởi vì chúng chỉ định rõ thứ tự các biến yếu tố. Hệ số thứ nhất trong danh sách các biến phân loại chỉ định cho giá trị của biến phân loại thuộc nhóm thấp nhất, còn hệ số cuối cùng thuộc về nhóm cao nhất. Chẳng hạn: Nếu có 6 biến phân loại và khi chỉ định thứ tự các hệ số -1, 0, 0, 0, 0.5 và 0.5, thì chương trình sẽ thực hiện so sánh đối chiếu giá trị trung bình của nhóm 1 với giá trị trung bình của tập hợp hai nhóm 5 và 6. Nếu bạn chỉ định các hệ số -1, -1, -1, -1, 2, 2, thì chương trình sẽ thực hiện so sánh đối chiếu giá trị trung bình của tập hợp 4 nhóm đầu với giá trị trung bình của tập hợp hai nhóm 5 và 6. Chúng ta có thể chỉ định các hệ số cho một bộ tương phản theo số nguyên hoặc theo số tỉ lệ, nhưng điều kiện là tổng các hệ số phải bằng zero. Thông thường, để có những bộ so sánh tương phản có ý nghĩa, chúng ta phải chủ động chỉ định chúng trước khi làm một cuộc thí nghiệm hoặc thu thập dữ liệu.

Kết quả của so sánh tương phản bao gồm: (1) Các hệ số của các bộ tương phản; (2) Thống kê F cho mỗi tập hợp (bộ) tương phản, (3) Khoảng tin cậy kiểu Bonferroni cho những khác biệt của mỗi tương phản dựa trên phân phối T. Lưu ý rằng, khi sử dụng thủ tục One-Way ANOVA và chọn contrast là đa thức (Polynomial), thì kết quả tương phản sẽ hiển thị trong Bảng kiểm định cùng lúc gồm cả mỗi quan hệ tuyến tính bậc 1 lẫn quan hệ phi tuyến tính (cao nhất đến bậc 5).

(C₂) Đối với thủ tục General linear model: Univariate. Những tương phản có hiệu lực là Deviation, Simple, Difference, Helmert, Repeated và Polynomial. Khi sử dụng thủ tục so sánh tương phản Deviation và Simple, chúng ta có thể chọn loại tóm tắt là Last hoặc First. Trong ô Contrast, nếu chọn None thì không có một so sánh tương phản nào được thực hiện. Nếu chọn Difference contrasts thì chương trình thực hiện những so sánh giữa giá trị trung bình của từng mẫu (trừ mẫu thứ 1) với giá trị trung bình của tổ hợp các mẫu trên. Nếu chọn Deviation contrasts thì chương trình thực hiện các so sánh giữa giá trị trung bình của từng mẫu (có thể bỏ mẫu đầu bằng cách chọn First hoặc bỏ mẫu cuối bằng cách chọn Last) với giá trị trung bình chung của các mẫu. Nếu chọn Simple contrasts thì chương trình thực hiện so sánh giữa trung bình của mỗi mẫu với giá trị trung bình mẫu được chỉ định. Mẫu chỉ định có thể là mẫu cuối (Last) hoặc mẫu đầu (First). Nếu chọn Repeated contrasts thì chương trình sẽ lần lượt so sánh giữa trung bình của mỗi mẫu (trừ mẫu cuối) với trung bình của mẫu kế tiếp. Nếu chọn Helmert contrasts thì chương trình sẽ so sánh giữa trung bình của mỗi mẫu (trừ mẫu 1) với trung bình của tổ hợp các mức dưới. Nếu chọn Polynomial contrasts thì chương trình sẽ tính mỗi quan hệ tuyến tính (bậc 1), bậc 2 và bậc 3 hoặc sự thẳng giăng giữa các trung bình mẫu có thứ bậc. Ví dụ: Mức tăng dần theo bậc của N, K hoặc P.

Tóm lại, về cơ bản trình tự các bước so sánh tương phản như sau: (1) Đặt giả thuyết H_0 : các tương phản = 0; H_1 : các tương phản khác 0; (2) Chỉ định các bộ tương phản bằng cách

nhập vào các hệ số (tổng các hệ số phải bằng zero); (3) Hoàn thành các thủ tục gọi chương trình tính toán; (4) giải thích và nhận xét kết quả.

4.2.1. MÔ HÌNH PHÂN TÍCH PHƯƠNG SAI MỘT YẾU TỐ

Phần này giới thiệu phân tích phương sai 1 yếu tố với ba kiểu bố trí thí nghiệm sau đây: (1) ngẫu nhiên hoàn toàn, (2) khối ngẫu nhiên đầy đủ, (3) la tinh bình phương.

A. PHÂN TÍCH PHƯƠNG SAI 1 YẾU TỐ: Kiểu thí nghiệm ngẫu nhiên hoàn toàn¹

Kiểu thí nghiệm ngẫu nhiên hoàn toàn (CRD) được thiết kế theo sơ đồ sau đây:

Số liệu quan sát	Thí nghiệm			
	1	2	...	r
1	Y_{11}	Y_{21}	...	Y_{r1}
2	...	...	...	...
...	...	...	...	...
j	Y_{1j}	...	...	Y_{rj}

Trong kiểu bố trí thí nghiệm ngẫu nhiên hoàn toàn, số liệu quan sát (j) có thể bằng nhau hoặc không bằng nhau và trật tự của các thí nghiệm không ảnh hưởng đến kết quả phân tích biến động. Ở kiểu thí nghiệm này có hai nguồn biến động: (1) nguồn biến động giữa các thí nghiệm, (2) nguồn biến động ngẫu nhiên của các số liệu được đo đạc ở mỗi thí nghiệm (sai số thí nghiệm). Giả thuyết H_0 : các thí nghiệm (yếu tố thí nghiệm) ảnh hưởng đồng đều đến kết quả thí nghiệm. Đối thuyết H_0 : các thí nghiệm ảnh hưởng khác nhau đến kết quả thí nghiệm. Khi nguồn biến động giữa các thí nghiệm lớn hơn biến động ngẫu nhiên thì giả thuyết H_0 bị bác bỏ.

Mô hình phân tích thống kê:

$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$; trong đó:

Y_{ij} = biến phụ thuộc;

μ = trung bình tổng thể;

α_i = tham số đặc trưng cho ảnh hưởng của yếu tố thí nghiệm, $i = 1, 2 \dots a$;

ε_{ij} = sai số ngẫu nhiên ($\varepsilon_{ij} \rightarrow N(0, \sigma^2)$).

Phân tích phương sai 1 yếu tố theo thí nghiệm kiểu CRD cũng có thể kiểm định bằng mô hình tuyến tính tổng quát (GML).

Mô hình tuyến tính tổng quát:

$Y = \beta_0 + \beta_1 X_1 + \varepsilon$;

X_1 = yếu tố thí nghiệm, ε = sai số ngẫu nhiên. Kết quả phân tích biến động theo mô hình tuyến tính tổng quát được báo cáo như sau :

Kiểm định ảnh hưởng giữa các yếu tố (Tests of Between-Subjects Effects)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power
Corr. Model (1)	416.144	2	208.072	6.235	.004	12.469	.872
Intercept (2)	70725.689	1	70725.68	2119.24	.000	2119.248	1.000
Dinh (3)	416.144	2	208.072	6.235	.004	12.469	.872
Error (4)	1401.667	42	33.373				
Total (5)	72543.500	45					

¹ Complete Randomized Design = CRD

Corr. Total (6)	1817.811	44				
-----------------	----------	----	--	--	--	--

Giải thích Bảng ANOVA. Cột có nhãn Source cho biết các nguồn biến động của biến phụ thuộc. Corrected Model (1) biểu thị tổng các biến thiên của biến phụ thuộc do mô hình hay do yếu tố nghiên cứu gây ra (3). Intercept là điểm chặn của đường hồi quy giữa trung bình các nghiệm thức. Giả thuyết H_0 : điểm chặn = 0. Nếu $P = 0.000$ (trị làm tròn của $P < 0.0005$) thì giả thuyết về điểm chặn của mô hình bằng không sẽ bị bác bỏ. Nghiệm thức (3) biểu thị biến động của biến phụ thuộc do nghiệm thức gây ra. Error (4) là biến thiên của biến phụ thuộc do sai số thí nghiệm hay do những yếu tố không giải thích được. Total (5) là tổng bình phương của biến phụ thuộc (Y_i^2). Corrected Total (6) là tổng chênh lệch bình phương của bộ phận tạo ra sự biến thiên của biến phụ thuộc do mô hình (1) cộng với bộ phận tạo ra biến thiên của biến phụ thuộc do sai số (4), nghĩa là $(6) = (1) + (4)$. Cột có nhãn Type III Sum of Squares cho biết tổng bình phương của các nguồn biến thiên theo kiểu III.

Cột có nhãn df (degree of freedom) chỉ độ tự do của các thành phần trong mô hình. Cột có nhãn Mean Square biểu thị trung bình bình phương của các nguồn biến động. Nó được tính bằng cách chia tổng bình phương của các nguồn biến thiên (cột 2) cho độ tự do tương ứng (cột 3). Cột có nhãn F (cột 5) là tiêu chuẩn kiểm định giả thuyết H_0 cho rằng các mức của yếu tố nghiên cứu (biến độc lập, biến phân loại) ảnh hưởng đồng đều đến kết quả thí nghiệm (biến phụ thuộc). $F = (\text{Means square/Error})$, nghĩa là trung bình bình phương của các nguồn biến động chia cho sai số tương ứng. Nếu giả thuyết H_0 là đúng thì $F = 1$. Nếu $F \gg 1$ thì bác bỏ H_0 và chấp nhận H_0^- , nghĩa là các mức của yếu tố nghiên cứu ảnh hưởng không đồng đều đến biến phụ thuộc.

Cột có nhãn (Sig) cho biết mức ý nghĩa của kiểm định hay xác suất chấp nhận giả thuyết H_0 . Trong các sách thống kê, mức ý nghĩa của kiểm định thường được kí hiệu bằng chữ P hay $P - \text{value}$. Nếu $P = 0.000$ thì giả thuyết H_0 bị bác bỏ.

Cột có nhãn Noncenter Parameter (tham số phi trung tâm) dùng để tính độ mạnh của phép kiểm định (Observed Power). Nếu ta sử dụng lặp lại một kiểm định với mức ý nghĩa α cố định, và khi giả thuyết H_0 đúng thì ta sẽ bị sai $\alpha\%$ và đúng $100 - \alpha\%$. Vì thế chúng ta cần quan tâm đến độ mạnh của phép kiểm định để phát hiện H_0 là sai. Độ mạnh của phép kiểm định đo xác suất bác bỏ giả thuyết H_0 nếu H_0 đúng. Khi xác suất bác bỏ giả thuyết lấy giá trị càng cao thì phép kiểm định càng hiệu quả.

B. PHÂN TÍCH PHƯƠNG SAI 1 YẾU TỐ: Kiểu thí nghiệm theo khối đầy đủ ngẫu nhiên¹

Trong kiểu bố trí thí nghiệm theo khối đầy đủ ngẫu nhiên (RCBD), số liệu trong mỗi nghiệm thức được đo đạc từ những điều kiện khác nhau, còn số liệu của các nghiệm thức được đo đạc trong mỗi khối có điều kiện như nhau. Khi bố trí thí nghiệm theo khối đầy đủ ngẫu nhiên, bạn có thể thiết kế theo sơ đồ sau đây:

Số lần lặp (Khối)	Nghiệm thức			
	1	2	...	k
1	Y_{11}	Y_{21}	...	Y_{k1}
...	...	...	...	...
r	Y_{1r}	...	...	Y_{kr}

Thí nghiệm RCBD có ba nguồn biến động: (1) nguồn biến động của các nghiệm thức, (2) nguồn biến động ngẫu nhiên của các số liệu được đo đạc ở mỗi nghiệm thức (sai số thí nghiệm), (3) nguồn biến động theo khối nghiệm thức.

¹ Randomized Complete Block Design = RCBD

Giả thuyết H_0 : Các nghiệm thức và các khối ảnh hưởng đồng đều đến kết quả thí nghiệm.

Đôi thuyết H_0 : Các nghiệm thức và các khối ảnh hưởng không đồng đều đến kết quả thí nghiệm. Khi nguồn biến động của các nghiệm thức và khối lớn hơn biến động ngẫu nhiên thì giả thuyết H_0 bị bác bỏ.

Mô hình phân tích thống kê:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij};$$

Y_{ij} = biến phụ thuộc; μ = trung bình tổng thể; α_i = tham số đặc trưng cho ảnh hưởng của yếu tố thí nghiệm, $i = 1, 2 \dots a$; β_j - tham số đặc trưng cho ảnh hưởng của khối, $j = 1, 2 \dots r$; ε_{ij} = sai số ngẫu nhiên ($\varepsilon_{ij} \rightarrow N(0, \sigma^2)$).

Khi sử dụng thủ tục GLM trong SPSS để phân tích biến động 1 yếu tố theo kiểu thí nghiệm RCBD, bạn có thể thiết kế mô hình theo hai kiểu.

+ Kiểu thứ nhất chứa đầy đủ yếu tố thí nghiệm và yếu tố khối:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

β_0 – điểm chặn (Intercept), X_1 – yếu tố thí nghiệm, X_2 – yếu tố khối, ε - sai số ngẫu nhiên.

+ Kiểu thứ hai chỉ bao gồm yếu tố thí nghiệm, còn yếu tố khối được gộp vào thành phần sai số:

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

β_0 = điểm chặn (Intercept), X_1 = yếu tố thí nghiệm, ε = sai số ngẫu nhiên.

Việc thiết kế mô hình theo kiểu nào là tùy thuộc vào mục tiêu nghiên cứu. Nếu thiết kế đầy đủ các yếu tố trong mô hình tuyến tính tổng quát thì kết cấu của Bảng phân tích biến động có dạng:

Kiểm định ảnh hưởng của các yếu tố thí nghiệm

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model (1)	522358.750	7	74622.679	18.553	.000
Intercept (2)	6446801.250	1	6446801.25	1602.851	.000
Nghiệm thức (3)	264955.000	4	66238.750	16.469	.000
Khối (4)	257403.750	3	85801.250	21.333	.000
Error (5)	48265.000	12	4022.083		
Total (6)	7017425.000	20			
Corrected Total (7)	570623.750	19			

a. R Squared = .915 (Adjusted R Squared = .866)

Giải thích Bảng ANOVA. Corrected Model biểu thị biến động do mô hình; trị số này bằng tổng các biến thiên do yếu tố nghiên cứu gây ra, nghĩa là biến động do nghiệm thức (3) cộng với biến động do khối (4). Error (5) – Biến thiên do sai số thí nghiệm. Corrected Total là tổng chênh lệch bình phương của bộ phận tạo ra sự biến thiên của biến phụ thuộc do mô hình cộng với bộ phận tạo ra biến thiên của biến phụ thuộc do sai số. Intercept là điểm chặn của đường hồi quy giữa trung bình các nghiệm thức. Nếu $P = 0.000$ thì giả thuyết về điểm chặn của mô hình bằng không sẽ bị bác bỏ. Những thành phần khác của Bảng ANOVA được giải thích tương tự như mục 4.2.1A.

C. PHÂN TÍCH PHƯƠNG SAI 1 YẾU TỐ: Kiểu thí nghiệm bình phương la tinh¹

Kiểu bình phương la tinh là kiểu thí nghiệm mà các nghiệm thức được bố trí ngẫu nhiên cả ở khối hàng lẫn khối cột; trong đó số nghiệm thức, số khối hàng và số khối cột đều bằng nhau và bằng số lần lặp lại của mỗi nghiệm thức. Ví dụ: Nghiên cứu ảnh hưởng của 6 cấp hàm lượng super photphat (0, 1, 2, 3, 4, 5% so với trọng lượng bầu) đến sinh trưởng chiều cao (H, cm) của Đậu song năng sau 8 tháng tuổi. Thí nghiệm được bố trí theo khối với 6 lần lặp lại. Tổng cộng có 36 lô thí nghiệm, được phân phối theo sơ đồ gồm 6 hàng và 6 cột như sau:

		Hàng					
		1	2	3	4	5	6
Cột	1	A	B	C	D	E	F
	2	B	A	E	F	C	D
	3	C	F	A	E	D	B
	4	D	C	B	A	F	E
	5	E	D	F	C	B	A
	6	F	E	D	B	A	C

Kí hiệu: ABCDEF – 6 nghiệm thức; trong đó: A = 0, B = 1, C = 2, D = 3, E = 4, F = 5% Super photphat so với trọng lượng bầu.

Trong bố trí thí nghiệm theo kiểu bình phương la tinh như sơ đồ trên đây có bốn nguồn biến động: (1) nguồn biến động của các nghiệm thức, (2) nguồn biến động ngẫu nhiên của các số liệu được lấy ở mỗi nghiệm thức (sai số thí nghiệm), (3) nguồn biến động của các số liệu được lấy ở mỗi nghiệm thức theo khối cột, (4) nguồn biến động của các số liệu được lấy ở mỗi nghiệm thức theo khối hàng.

Mô hình phân tích thống kê:

$$Y_{ij} = \mu + \alpha_i + \beta_j \times \text{Khối cột} + \beta_k \times \text{Khối hàng} + \varepsilon_{ij}$$

Trong đó: Y_{ij} = biến phụ thuộc; μ = trung bình tổng thể; α_i = tham số đặc trưng cho ảnh hưởng của yếu tố thí nghiệm, $i = 1, 2 \dots a$; β_j - tham số đặc trưng cho ảnh hưởng của khối cột, $i = 1, 2 \dots r$; β_k = tham số đặc trưng cho ảnh hưởng của khối hàng, $j = 1 \dots k$; ε_{ijk} = sai số ngẫu nhiên ($\varepsilon_{ijk} \rightarrow N(0, \sigma^2)$).

Giả thuyết H_0 : các nghiệm thức và các khối ảnh hưởng đồng đều đến kết quả thí nghiệm. Khi nguồn biến động của các nghiệm thức và khối lớn hơn biến động ngẫu nhiên thì giả thuyết H_0 bị bác bỏ.

Phân tích biến động 1 yếu tố được bố trí theo kiểu bình phương la tinh cũng có thể được thực hiện theo mô hình tuyến tính tổng quát.

Mô hình tuyến tính đầy đủ:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

trong đó: X_1 – yếu tố thí nghiệm, X_2 - yếu tố khối cột, X_3 – yếu tố khối hàng, ε - sai số ngẫu nhiên. Nếu mô hình chỉ bao gồm yếu tố thí nghiệm, còn yếu tố khối (hàng và cột) được gộp vào thành phần sai số thì mô hình có dạng:

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

β_0 = điểm chặn (Intercept), X_1 = yếu tố thí nghiệm, ε = sai số ngẫu nhiên.

¹ Latin – Square Design

Kiểm định ảnh hưởng của yếu tố thí nghiệm đến yếu tố phụ thuộc ^(a)
(Tests of Between-Subjects Effects)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	12024.097	15	801.606	24.514	.000
Intercept	26238.600	1	26238.600	802.413	.000
Nghiệm thức	10866.691	5	2173.338	66.464	.000
Cột	255.515	5	51.103	1.563	.216
Hàng	901.891	5	180.378	5.516	.002
Error	653.992	20	32.700		
Total	38916.690	36			
Corrected Total	12678.090	35			

(a) Thí nghiệm 1 nhân tổ kiểu bình phương la tinh.

4.2.2. MÔ HÌNH PHÂN TÍCH PHƯƠNG SAI HAI YẾU TỐ

Phần này giới thiệu phân tích phương sai 2 yếu tố với bốn kiểu bố trí thí nghiệm sau đây: (1) Ngẫu nhiên hoàn toàn, (2) Khối ngẫu nhiên đầy đủ, (3) Bình phương la tinh, (4) Lô phụ (Split plot experiment).

A. PHÂN TÍCH PHƯƠNG SAI 2 YẾU TỐ: Kiểu thí nghiệm hoàn toàn ngẫu nhiên

Mô hình phân tích thống kê có dạng: $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$

Trong đó μ = trung bình tổng thể được ước lượng từ trung bình mẫu Y_{bq} ; α_i ($i = 1 \dots a$) = tham số đặc trưng cho ảnh hưởng của yếu tố thí nghiệm A; β_j ($j = 1 \dots b$) = tham số đặc trưng cho ảnh hưởng của yếu tố thí nghiệm B; $(\alpha\beta)_{ij}$ = tương tác giữa yếu tố A với yếu tố B; ε_{ijk} = sai số ngẫu nhiên; k = số lần lặp ($k = 1 \dots m$). Giả thuyết đối với yếu tố A, H_a : $\alpha_i = 0$ với mọi i ; H_1 : ít nhất có một $\alpha_i \neq 0$. Đối với yếu tố B, H_b : $\beta_i = 0$ với mọi i ; H_1 : ít nhất có một $\beta_i \neq 0$. Đối với tổ hợp yếu tố A và B, H_{ab} : không có tương tác giữa yếu tố A với yếu tố B. Kiểu thí nghiệm hai yếu tố ngẫu nhiên hoàn toàn (RCD) có m lần lặp lại ở mỗi tổ hợp cấp của yếu tố A và yếu tố B được thiết kế theo sơ đồ như sau :

Yếu tố A	Yếu tố B				
	1	2	3	...	b
A1	X_{111}	X_{121}	X_{131}		X_{1b1}
	X_{112}	X_{122}	X_{132}		X_{1b2}
	X_{11m}	X_{12m}	X_{13m}		X_{1bm}
A2	...				
A3	X_{311}	X_{321}	X_{331}		X_{3b1}
	X_{312}	X_{322}	X_{332}		X_{3b2}
	X_{31m}	X_{32m}	X_{33m}		X_{3bm}

Khi sử dụng thủ tục GLM để phân tích phương sai 2 yếu tố theo kiểu thí nghiệm RCD thì mô hình tuyến tính tổng quát có dạng như sau: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$; trong đó: X_1 và X_2 – yếu tố thí nghiệm 1 và 2; ε - sai số thí nghiệm.

B. PHÂN TÍCH PHƯƠNG SAI 2 YẾU TỐ: Thiết kế theo khối đầy đủ ngẫu nhiên¹

Mô hình phân tích thống kê có dạng:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + k_k + (\alpha\beta)_{ij} + \varepsilon_{ijk};$$

Trong đó:

μ = trung bình tổng thể được ước lượng từ trung bình mẫu Y_{bq} ;

α_i ($i = 1 \dots a$) = tham số đặc trưng cho ảnh hưởng của yếu tố thí nghiệm A;

β_j ($j = 1 \dots b$) = tham số đặc trưng cho ảnh hưởng của yếu tố thí nghiệm B;

$(\alpha\beta)_{ij}$ = đặc trưng cho ảnh hưởng tương tác giữa yếu tố A với yếu tố B;

ε_{ijk} = sai số ngẫu nhiên;

k_k = ảnh hưởng của khối ($k = 1 \dots r$).

Giả thuyết đối với yếu tố A, $H_a: \alpha_i = 0$ với mọi i ; H_1 : ít nhất có 1 $\alpha_i \neq 0$. Đối với yếu tố B, $H_b: \beta_i = 0$ với mọi i ; H_1 : ít nhất có 1 $\beta_i \neq 0$. Đối với tổ hợp yếu tố A và B, H_{ab} : không có tương tác giữa yếu tố A với yếu tố B.

Kiểm thí nghiệm hai yếu tố được bố trí theo khối đầy đủ ngẫu nhiên (RCBD) với r lần lặp lại được thiết kế theo sơ đồ sau đây :

	Yếu tố A	Yếu tố B				
		1	2	3	...	b
Khối 1	A_1	Y_{111}	Y_{121}	Y_{131}	...	Y_{1b1}
	A_2	Y_{211}	Y_{221}	Y_{231}	...	Y_{2b1}
	A_3	Y_{311}	Y_{321}	X_{331}	...	X_{3b1}
Khối 2	A_1	Y_{112}	Y_{122}	Y_{132}	...	Y_{1b2}
	A_2	Y_{212}	Y_{222}	Y_{232}	...	Y_{2b2}
	A_3	Y_{312}	Y_{322}	X_{332}	...	X_{3b2}
Khối 3	A_1	Y_{113}	Y_{123}	Y_{133}	...	Y_{1b3}
	A_2	Y_{213}	Y_{223}	Y_{233}	...	Y_{2b3}
	A_3	Y_{313}	Y_{323}	X_{333}	...	X_{3b3}

/x/ Yếu tố A có a mức, B có b mức; $k = 1$ đến 3 – số lần lặp (khối)

Khi sử dụng thủ tục GLM để phân tích phương sai 2 yếu tố được bố trí theo khối đầy đủ ngẫu nhiên thì mô hình tuyến tính tổng quát có dạng: $Y = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 AB + \beta_5 \times \text{Khối} + \varepsilon$; trong đó A = yếu tố thí nghiệm 1, B = yếu tố thí nghiệm 2; AB = sự tương tác giữa hai yếu tố A với B; Khối = số lần lặp; ε = sai số thí nghiệm. Nếu gộp yếu tố khối vào thành phần sai số thì mô hình tuyến tính tổng quát có dạng: $Y = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 AB + \varepsilon$. Ở dạng đầy đủ các yếu tố, kết cấu của Bảng ANOVA có dạng như sau:

Kiểm định ảnh hưởng của đạm và lân đến sinh trưởng H của Dầu song nòng
(Theo mô hình tuyến tính tổng quát)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	10665.422	12	888.785	35.830	.000
Intercept	333852.800	1	333852.80	13458.79	.000
Đạm	3040.533	2	1520.267	61.287	.000

¹ Randomized Complete Block Design = RCBD

Lân	578.133	2	289.067	11.653	.000
Khối	257.422	4	64.356	2.594	.055
Đạm × Lân ^(*)	6789.333	4	1697.333	68.426	.000
Error	793.778	32	24.806		
Total	345312.000	45			
Corrected Total	11459.200	44			

a R Squared = .931 (Adjusted R Squared = .905); (*) Tương tác giữa hai yếu tố thí nghiệm

C. PHÂN TÍCH PHƯƠNG SAI 2 YẾU TỐ: Kiểu thí nghiệm bình phương la tinh

Giả thiết cần phân tích ảnh hưởng của hai yếu tố đến sinh trưởng chiều cao của Dầu song nà 8 tháng tuổi: (1) yếu tố X (kích thước bầu) với 2 mức độ là 15×25 cm và 20×30 cm; (2) yếu tố P (phân NPK) với 3 mức độ là 1%, 2% và 3% NPK so với trọng lượng bầu. Như vậy, thí nghiệm có 2×3 = 6 nghiệm thức. Các nghiệm thức được kí hiệu như sau: A = X₁P₁; B = X₁P₂; C = X₁P₃; D = X₂P₁; E = X₂P₂; F = X₂P₃. Sơ đồ thí nghiệm có thể được bố trí như sau:

		Khối hàng					
		1	2	3	4	5	6
Khối cột	1	A	B	C	D	E	F
	2	B	A	E	F	C	D
	3	C	F	A	E	D	B
	4	D	C	B	A	F	E
	5	E	D	F	C	B	A
	6	F	E	D	B	A	C

Thí nghiệm 2 yếu tố theo kiểu bình phương la tinh có 6 nguồn biến động: (1) nguồn biến động của các nghiệm thức, (2) nguồn biến động ngẫu nhiên của các số liệu được lấy ở mỗi nghiệm thức (sai số thí nghiệm), (3) nguồn biến động gây ra bởi yếu tố 1, (4) nguồn biến động gây ra bởi yếu tố 2, (5) nguồn biến động của các số liệu được lấy ở mỗi nghiệm thức theo khối cột, (6) nguồn biến động của các số liệu được lấy ở mỗi nghiệm thức theo khối hàng. Các giả thuyết: (1) Các yếu tố độc lập (1 và 2) ảnh hưởng đồng đều đến kết quả thí nghiệm; (2) Không có ảnh hưởng tương tác giữa các yếu tố độc lập đến kết quả thí nghiệm; (3) Các khối hàng và khối cột ảnh hưởng đồng đều đến kết quả thí nghiệm.

Mô hình phân tích thống kê :

$$Y_{ijch} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_c + \delta_h + \varepsilon_{ijch}$$

trong đó: μ = trung bình tổng thể; α_i ($i = 1 \dots a$) - tham số đặc trưng cho ảnh hưởng của yếu tố thứ nhất; β_j ($j = 1 \dots b$) = tham số đặc trưng cho ảnh hưởng của yếu tố thứ hai; $(\alpha\beta)_{ij}$ = tương tác giữa yếu tố 1 với yếu tố 2; γ_c = tham số đặc trưng cho ảnh hưởng của yếu tố cột; δ_h = tham số đặc trưng cho ảnh hưởng của yếu tố hàng; ε_{ijch} – sai số ngẫu nhiên.

Nếu sử dụng mô hình tuyến tính tổng quát để phân tích biến động 2 yếu tố được bố trí theo kiểu bình phương la tinh thì mô hình có dạng:

$$Y = \beta_0 + \beta_1(\text{khối hàng}) + \beta_2(\text{khối cột}) + \beta_3A + \beta_4B + \beta_5AB + \varepsilon$$

Trong đó A = yếu tố thí nghiệm 1, B = yếu tố thí nghiệm 2; khối hàng = số lần lặp theo hàng; khối cột = số lần lặp theo cột; ε = sai số thí nghiệm.

Nếu gộp những biến động theo khối hàng và theo khối cột vào thành phần sai số, thì mô hình GLM có dạng: $Y = \beta_0 + \beta_1A + \beta_2B + \beta_3AB + \varepsilon$; trong đó A = yếu tố thí nghiệm 1, B = yếu tố thí nghiệm 2; ε = sai số thí nghiệm. Theo mô hình GLM đầy đủ các yếu tố, kết cấu của Bảng ANOVA có dạng :

Kiểm định ảnh hưởng của các yếu tố đến sinh trưởng H của Dầu song nòng

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1851.784	15	123.45	40.24	.000
Intercept	170280.023	1	170280.02	55498.90	.000
Khối hàng	8.709	5	1.742	.568	.724
Khối cột	137.573	5	27.52	8.968	.000
Kích thước bầu	755.334	1	755.33	246.18	.000
NPK	712.500	2	356.25	116.11	.000
Ktbau $\times$ NPK	237.669	2	118.83	38.73	.000
Error	61.363	20	3.07		
Total	172193.170	36			
Corrected Total	1913.147	35			

a R Squared = .968 (Adjusted R Squared = .944)

D. PHÂN TÍCH PHƯƠNG SAI 2 YẾU TỐ: Kiểu thí nghiệm có lô phụ

Thí nghiệm 2 yếu tố (A và B) theo kiểu lô chính (yếu tố A) – lô phụ (yếu tố B) có 6 nguồn biến động: (1) nguồn biến động của yếu tố A, (2) nguồn biến động của yếu tố B, (3) sai số gây ra bởi yếu tố A, (4) sai số gây ra bởi yếu tố B, (5) nguồn biến động số liệu do lặp lại (khối), (6) nguồn biến động gây ra bởi sự tương tác giữa yếu tố A với yếu tố B. Các giả thuyết: (1) Từng yếu tố A và B ảnh hưởng đồng đều đến kết quả thí nghiệm; (2) Không có ảnh hưởng tương tác giữa yếu tố A với yếu tố B đến kết quả thí nghiệm; (3) Các khối ảnh hưởng đồng đều đến kết quả thí nghiệm.

Mô hình phân tích thống kê :

$$Y_{ijk} = \mu + \alpha_i + k_k + \varepsilon_{ik} + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ij}$$

Trong đó: μ = trung bình tổng thể được ước lượng từ trung bình mẫu Y_{bq} ; α_i ($i = 1 \dots a$) = tham số đặc trưng cho ảnh hưởng của yếu tố chính A; β_j ($j = 1 \dots b$) = tham số đặc trưng cho ảnh hưởng của yếu tố phụ B; $(\alpha\beta)_{ij}$ = tương tác giữa yếu tố A với yếu tố B; k_k = tham số đặc trưng cho ảnh hưởng của khối (lặp); ε_{ik} = sai số ngẫu nhiên của yếu tố A; ε_{ijk} = sai số ngẫu nhiên của yếu tố B.

Giả thiết cần xác định sinh trưởng chiều cao (H, cm) của 3 loài cây (Dầu rái, Sao đen, Vên vên) dưới ảnh hưởng của 3 phương thức lâm sinh. Mỗi thí nghiệm được lặp lại 3 lần. Ở đây chọn lô chính là 3 phương thức lâm sinh, còn lô phụ là loài cây. Sơ đồ bố trí thí nghiệm như sau :

Sơ đồ bố trí thí nghiệm hai yếu tố có lô phụ

Thí nghiệm	Số lần lặp lại						
	I			II	III		
	Dầu rái	Sao đen	Vên vên	...	Dầu rái	Sao đen	Vên vên
1	Y_{A11}	Y_{A21}	Y_{A31}	...	Y_{A13}	Y_{A23}	Y_{A33}

2	Y_{B11}	Y_{B21}	Y_{B31}	$\dots$	Y_{B13}	Y_{B23}	Y_{B33}
3	Y_{C11}	Y_{C21}	Y_{C31}	$\dots$	Y_{C13}	Y_{C23}	Y_{C33}

Nếu sử dụng mô hình tuyến tính tổng quát để phân tích biến động 2 yếu tố được bố trí theo kiểu lô chính – lô phụ thì mô hình có dạng:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \beta_4 \text{Khối} \times X_1 + \varepsilon;$$

Trong đó X_1 = yếu tố thí nghiệm 1 (bố trí ở lô chính); X_2 = yếu tố thí nghiệm 2 (bố trí ở lô phụ); $\text{Khối} \times X_1$ = sai số của lô chính; ε - sai số thí nghiệm.

Kết cấu của Bảng ANOVA tổng quát có dạng :

Kiểm định ảnh hưởng của các yếu tố đến sinh trưởng H
của Dầu rái, Sao đen, Vên vên

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3222.306	17	189.547	52.761	.000
Intercept	149640.028	1	149640.028	41652.379	.000
Khối	61.639	3	20.546	5.719	.006
Thuốc	476.722	2	238.361	66.348	.000
Thời gian	997.389	2	498.694	138.812	.000
Khối $\times$ Thuốc ^(*)	229.944	6	38.324	10.668	.000
Thuốc $\times$ Thời gian	1456.611	4	364.153	101.362	.000
Error	64.667	18	3.593		
Total	152927.000	36			
Corrected Total	3286.972	35			

a. R Squared = .980 (Adjusted R Squared = .962); (*) sai số của lô chính.

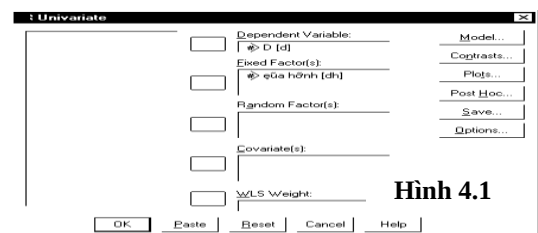
4.3. THỰC HÀNH PHÂN TÍCH PHƯƠNG SAI

4.3.1. PHÂN TÍCH PHƯƠNG SAI MỘT YẾU TỐ

A. PHÂN TÍCH PHƯƠNG SAI 1 YẾU TỐ: Kiểu thí nghiệm ngẫu nhiên hoàn toàn

Ví dụ 1. Giả sử cần nghiên cứu ảnh hưởng của 3 cấp địa Hình (chân, sườn, đỉnh) đến sinh trưởng $D_{1.3}$ (cm) của rừng Sao đen theo mô hình tuyến tính tổng quát: $Y = \beta_0 + \beta_1 X_1 + \varepsilon$; trong đó X_1 – yếu tố thí nghiệm (địa Hình), ε - sai số ngẫu nhiên. Để giải quyết vấn đề đặt ra, bạn hãy thực hiện theo các bước sau:

Bước 1. Thu thập và xử lý số liệu trung gian. Trước hết, bạn phân chia đối tượng (khu vực) nghiên cứu thành 3 cấp địa Hình. Tiếp đến thực hiện bố trí ô đo đếm. Trong mỗi ô đo đếm, bạn thống kê đầy đủ $D_{1.3}$ (cm) của từng cây. Để đảm bảo kết quả phân tích chính xác, mỗi cấp địa Hình có thể bố trí đo lặp lại r cây ($r > 30$ cây).



Hình 4.1

Bước 2. Nhập số liệu vào Bảng tính. Số liệu được nhập vào Bảng tính SPSS theo hai cột: Cột 1 nhập biến phụ thuộc ($D_{1.3} = Y_i$) theo thứ tự từ nghiệm thứ 1 đến nghiệm thứ 3; cột 2 nhập mã biến độc lập (X_i) hay mã biến lập nhóm (biến phân loại - Grouping Variable). Mỗi biến lập nhóm biểu thị cho một yếu tố thí nghiệm và được mã hóa bằng các số nguyên giống nhau hoặc kí tự giống nhau. Số lượng mã biến độc lập lặp lại giống nhau trong mỗi mức của yếu tố lập nhóm cho biết số lượng quan sát của mỗi nhóm (**Bảng 4.1**).

Bảng 4.1. Cách nhập dữ liệu trong phân tích phương sai một yếu tố
(Chi tiết số liệu xem ở phụ lục 1)

D _{1.3} , cm	27.5	25.5	44.5	40.5		48		35.5	34.5	49.5	45.5	37.5	...
Địa Hình	1	1	1	1	...	2	...	2	2	3	3	3	...

Bước 3. Kiểm định về phân bố chuẩn và ngang bằng phương sai mẫu. Các điều kiện này được kiểm định bằng cách chọn Option trong cửa sổ Univariate. Bên cạnh kiểm định sự bằng nhau của phương sai mẫu theo kiểm định Levene's Test, thủ tục này còn tính toán các tham số thống kê mô tả, lập đồ thị các trung bình mẫu...

Bước 4. So sánh các cặp giá trị trung bình mẫu và kiểm định phân hạng các mẫu. Giả thuyết $H_0: X_{bq1} = X_{bq2} = \dots = X_{bqk}$; $H_0': X_{bq1} \neq X_{bq2} \neq \dots \neq X_{bqk}$. Để kiểm tra giả thuyết H_0 , từ menu chính, bạn chọn Analyze $\rightarrow$ GLM: Univariate $\rightarrow$ Dependent list (đến đây hãy đưa biến phụ thuộc vào bằng cách chọn nó và nhấn nút mũi tên bên cạnh) $\rightarrow$ Fixed factor (đưa biến lập nhóm vào bằng cách chọn nó và nhấn nút mũi tên bên cạnh) (**Hình 4.1**). Để thực hiện các phân tích thống kê, từ cửa sổ Univariate, bạn hãy thực hiện như sau:

+ Chọn Model, sau đó đánh dấu mục Custom và chuyển biến số D từ ô Factor & Covariate sang ô Model.

+ Chọn Option. Tại cửa sổ Univariate: Option, chọn biến D từ ô Factor(s) and Factor Interactions để chuyển nó sang ô Display means for. Kế đến đánh dấu chọn Compare main effect và các thống kê cần quan tâm trong ô Display (hiển thị).

+ Chọn Post Hoc. Khi cửa sổ Univariate Post Hoc Multiple Comparisons for Observed means xuất hiện, bạn chọn D trong ô Factor và chuyển nó sang ô Post Hoc Tests for. Kế đến đánh dấu chọn Bonferroni và Tukey rồi chọn continue. Lưu ý rằng, khi có 2 nghiệm thức cần so sánh từng cặp với nhau, thì bạn chỉ cần chọn kiểm định LSD hoặc Bonferroni. Nhưng nếu số nghiệm thức lớn hơn 2, thì bạn cần đến không chỉ thủ tục so sánh từng cặp nghiệm thức với nhau mà còn cả thủ tục phân hạng các nghiệm thức. Để so sánh phân hạng các nghiệm thức, bạn hãy chọn kiểm định Tukey's Test. Kiểm định Tukey HSD được chọn khi có nhiều nghiệm thức (trên 2 nghiệm) và dung lượng của các cặp mẫu có thể khác nhau. Tại đây, bạn cũng có thể thay đổi mức tin cậy (Confidence Level) mà bạn muốn sử dụng cho các kiểm định phân hạng. Mức tin cậy mặc định là 0.95. Đối với kiểm định phân hạng Tukey HSD, mức tin cậy 0.95 là có hiệu quả.

Đến đây bạn hãy chọn OK để chương trình xử lý và hiển thị kết quả. Kết quả hiển thị bao gồm các nội dung cơ bản sau đây: (1) Thống kê mô tả (**Bảng 4.2**), (2) kiểm định sự bằng nhau của phương sai (**Bảng 4.3**), (3) Bảng ANOVA (**Bảng 4.4**), (4) so sánh giữa các cặp số trung bình nghiệm thức (**Bảng 4.5**) và phân hạng các nghiệm thức (**Bảng 4.6**).

Giải thích và nhận xét kết quả. **Bảng 4.3** ghi lại kết quả kiểm định sự bằng nhau của phương sai mẫu. Vì $P = 0.059 \approx 0.06$ nên phương sai các mẫu là bằng nhau. **Bảng 4.4** báo cáo kết quả phân tích phương sai (ANOVA). Hãy quan sát trị số P - value ở dòng có nhãn "Địa Hình". Vì $P = 0.004 < 0.01$ nên sinh trưởng của rừng Sao đen trên 3 cấp địa Hình khác nhau rõ rệt với mức tin cậy 99%.

Bảng 4.2. Thống kê mô tả (Descriptives)

Địa Hình	Mean	Std. Deviation	N
Đỉnh đồi	35.567	7.176	15
Sườn đồi	42.867	5.430	15
Chân đồi	40.500	4.375	15
Total	39.644	6.428	45

Bảng 4.3. Kiểm định sự bằng nhau của phương sai mẫu^(a)
(Test of Homogeneity of Variances D_{1,3} (cm))

F	df1	df2	Sig.
3.039	2	42	.059

a Design: Intercept+ Địa Hình

Bảng 4.4. Phân tích phương sai (ANOVA)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power
Corrected Model	416.144	2	208.072	6.235	.004	12.469	.872
Intercept	70725.689	1	70725.689	2119.248	.000	2119.248	1.000
Địa Hình	416.144	2	208.072	6.235	.004	12.469	.872
Error	1401.667	42	33.373				
Total	72543.500	45					
Corrected Total	1817.811	44					

a Computed using alpha = .05; b R Squared = .229 (Adjusted R Squared = .192)

Bảng 4.5. So sánh các cặp trung bình nghiệm thức (Pairwise Comparisons)

		Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
(I) Địa Hình	(J) Địa Hình				Lower Bound	Upper Bound
Đỉnh đồi	Sườn đồi	-7.300×	2.109	.001a	-11.557	-3.043
	Chân đồi	-4.933×	2.109	.024	-9.190	-.676
Sườn đồi	Đỉnh đồi	7.300×	2.109	.001	3.043	11.557
	Chân đồi	2.367	2.109	.268	-1.890	6.624
Chân đồi	Đỉnh đồi	4.933×	2.109	.024	.676	9.190
	Sườn đồi	-2.367	2.109	.268	-6.624	1.890

× The mean difference is significant at the .05 level.

a Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Bảng 4.6. Kiểm định phân hạng các nghiệm thức

Tukey HSD	Địa Hình	N	Subset	
			1	2
	Đỉnh đồi	15	35.567	
	Chân đồi	15	40.500	40.500
	Sườn đồi	15		42.867
	Sig.		.061	.506

b Alpha = .05.

Bảng 4.5 trình bày kiểm định so sánh giữa các cặp số trung bình nghiệm thức. Dấu hoa thị (×) ở bên cạnh các con số cho biết hai cặp mẫu ấy khác nhau rõ rệt. **Bảng 4.6** dẫn kết

quả kiểm định phân hạng các nghiệm thức. Những nghiệm thức không có khác biệt (cùng hạng) về mặt thống kê được ghi lại cùng một cột. Ở ví dụ này, địa Hình thực sự có ảnh hưởng đến sinh trưởng của rừng Sao đen. Kết quả kiểm định cho thấy rừng Sao đen mọc tốt ở địa Hình từ chân đồi đến sườn đồi, còn ở đỉnh đồi thì sinh trưởng kém nhất.

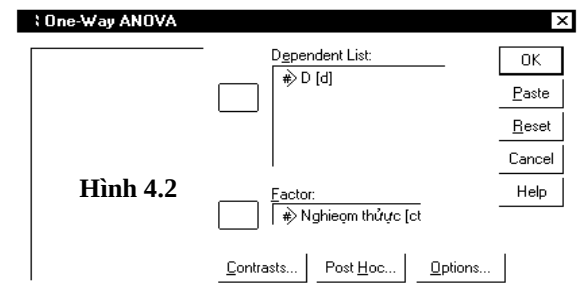
Ví dụ 2. Để xác định phản ứng sinh trưởng D của Sao đen với thành phần chất khoáng, người ta sử dụng 4 nghiệm thức sau đây: (1) Đối chứng, (2) bón P, (3) bón K, (4) bón PK. Thí nghiệm được bố trí ngẫu nhiên hoàn toàn. Hãy so sánh các nghiệm thức, đồng thời cho biết sự khác biệt về sinh trưởng của Sao đen giữa những nghiệm thức có bón P với những nghiệm thức không bón P (đối chứng và chỉ bón K)? Để giải quyết vấn đề đặt ra, bạn hãy thực hiện theo các bước sau đây:

Bước 1. Nhập số liệu vào Bảng tính. Số liệu nhập vào Bảng tính SPSS bao gồm hai cột (hoặc 2 hàng). Cột 1 nhập biến phụ thuộc (D, cm) theo thứ tự từ nghiệm thức 1 đến nghiệm thức 4. Cột 2 nhập mã biến độc lập hay lập nhóm (phân bón). Mỗi biến lập nhóm biểu thị cho một yếu tố thí nghiệm (Bảng 4.7).

Bảng 4.7. Cách nhập dữ liệu trong phân tích phương sai một yếu tố

D	2.45	2.25	4.38	4.35	3.42	3.27
Đối chứng	1	1	1	1	1	1
D	6.71	5.44	4.92	5.23	6.74	4.74
P	2	2	2	2	2	2
D	3.22	4.14	2.32	4.42	3.28	4
K	3	3	3	3	3	3
D	6.34	5.44	5.22	8.00	6.96	6.96
PK	4	4	4	4	4	4

Bước 2. So sánh các cặp số trung bình mẫu. Giả định $H_0: X_{bq1} = X_{bq2} = \dots = X_{bq4}$; $H_0^-: X_{bq1} \neq X_{bq2} \neq \dots \neq X_{bq4}$. Để kiểm tra giả định này, từ menu chính, bạn chọn Analyze → Compare means: One – way ANOVA → Dependent List (đến đây đưa biến D vào bằng cách chọn nó và nhấn nút mũi tên bên cạnh) → factor (đưa vào biến nghiệm thức bằng cách chọn nó và nhấn nút mũi tên bên cạnh) (Hình 4.2).



Hình 4.2

Bước 3. Phân tích thống kê. Từ cửa sổ One –way ANOVA, bạn hãy thực hiện như sau:

+ Chọn Post Hoc Test. Thủ tục này so sánh đa phạm vi cho các trung bình mẫu (Post Hoc Multiple Comparisons for Observed Means). Để đạt được mục đích này, bạn có thể chọn cả hai kiểm định Bonferroni và Tukey's Test. Sau đó chọn mức ý nghĩa của các kiểm định (mặc định là 0,05).

+ Chọn Option. Thủ tục này tính các thống kê mô tả và kiểm định phân hạng giữa các nghiệm thức. Để đạt được mục đích này, từ cửa sổ One –way ANOVA, bạn chọn Descriptive, Homogeneity – of – variance và Mean plot.

Cuối cùng chọn OK để chương trình tính toán và hiển thị kết quả. Kết quả hiển thị bao gồm các nội dung cơ bản sau đây: (1) Thống kê mô tả (Bảng 4.8), (2) kiểm định sự bằng nhau

của phương sai (Bảng 4.9), (3) Bảng ANOVA (Bảng 4.10), (4) so sánh giữa các cặp số trung bình nghiệm thức (Bảng 4.11) và phân hạng các nghiệm thức (Bảng 4.12).

Bảng 4.8. Thống kê mô tả (Descriptives)

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
Đối chứng	6	3.3533	.9047	.3693	2.4039	4.3027
P	6	5.6300	.8821	.3601	4.7043	6.5557
K	6	3.5633	.7747	.3163	2.7503	4.3764
PK	6	6.4867	1.0450	.4266	5.3900	7.5833
Total	24	4.7583	1.6062	.3279	4.0801	5.4366

Bảng 4.9. Kiểm định sự bằng nhau của phương sai
(Test of Homogeneity of Variances)

Levene Statistic	df1	df2	Sig.
.201	3	20	.894

Bảng 4.10. Phân tích phương sai (ANOVA)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	42.894	3	14.298	17.390	.000
Within Groups	16.444	20	.822		
Total	59.338	23			

Bảng 4.11. So sánh giữa các cặp trung bình (Multiple Comparisons)

		Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
(I) NT	(J) NT				Lower Bound	Upper Bound
Đối chứng	P	-2.2767×	.5235	.002	-3.7420	-.8114
	K	-.2100	.5235	.978	-1.6753	1.2553
	PK	-3.1333×	.5235	.000	-4.5986	-1.6680
P	Đối chứng	2.2767×	.5235	.002	.8114	3.7420
	K	2.0667×	.5235	.004	.6014	3.5320
	PK	-.8567	.5235	.382	-2.3220	.6086
K	Đối chứng	.2100	.5235	.978	-1.2553	1.6753
	P	-2.0667×	.5235	.004	-3.5320	-.6014
	PK	-2.9233×	.5235	.000	-4.3886	-1.4580
PK	Đối chứng	3.1333×	.5235	.000	1.6680	4.5986
	P	.8567	.5235	.382	-.6086	2.3220
	K	2.9233×	.5235	.000	1.4580	4.3886

× The mean difference is significant at the .05 level.

Giải thích và nhận xét kết quả. Bảng 4.10 báo cáo kết quả phân tích phương sai. Từ đó cho thấy các nghiệm thức khác nhau rất có ý nghĩa ($P < 0.001$). Bảng 4.11 và 4.12 báo cáo tương ứng kết quả so sánh đa phạm vi (Multiple Comparisons) và kiểm định phân hạng bằng thống kê Tukey HSD. Phân tích Bảng 4.11 và 4.12 cho thấy sinh trưởng của Sao đen trên từng công thức có bốn P vượt trội hơn những công thức không bón P.

Bảng 4.12. Kiểm định phân hạng các nghiệm thức

Tukey HSD	Nghiệm thức	N	Subset for alpha = .05	
			1	2
	Đối chứng	6	3.3533	
	K	6	3.5633	

	P	6		5.6300
	PK	6		6.4867
	Sig.		.978	.382

Muốn phân tích rõ hơn sự khác biệt về sinh trưởng của Sao đen giữa hai tổ hợp phân bón (có bón P và không bón P), bạn cần phải thực hiện so sánh tương phản giữa hai tổ hợp phân bón với nhau. Để so sánh tương phản giữa tổ hợp phân có chứa lân (P và PK) với tổ hợp phân chỉ có K, hoặc giữa tổ hợp phân có chứa lân (P và PK) với tổ hợp phân không có P (đối chứng và K), bạn có thể xây dựng tương ứng các bộ tương phản như sau: 0 -1 2 -1 và -1 1 -1 1. Muốn so sánh các bộ tương phản trên đây bằng thủ tục Compare means: One – Way ANOVA, sau khi đã chọn xong các thủ tục Post Hoc và Option như trên, bạn thực hiện tiếp như sau:

Tại cửa sổ One – Way ANOVA, bạn mở contrasts. Khi cửa sổ One – Way ANOVA contrasts xuất hiện, bạn chọn Polynomial; kế đến nhập lần lượt các bộ hệ số tương phản vào ô coefficients. Sau đó chọn tiếp Quadratic trong ô Degree. Lưu ý rằng khi chọn Quadratic thì bạn chỉ yêu cầu chương trình báo cáo kết quả phân tích quan hệ bậc 1 và bậc 2 cho các trung bình thuộc các tổ hợp trên. Nếu muốn thử tương phản ở các bậc cao hơn, bạn có thể thử đến bậc 5 (mặc định). Cuối cùng chọn Continue và OK. Đến đây chương trình sẽ hiển thị kết quả của các tương phản theo chỉ định của bạn (Bảng 4.13 và 4.14).

Lưu ý rằng, trong một số kiểu thí nghiệm chúng ta có thể cho yếu tố thí nghiệm tăng giảm theo bậc. Chẳng hạn, hàm lượng phân P tăng dần từ 0, 1, 2, ..., 10%; hay mật độ gieo ươm tăng từ 10×10, 15×15, 20×20, 25×25 cm...Nếu bố trí các mức yếu tố thí nghiệm tăng giảm dần theo bậc như trên thì thủ tục mở chương trình so sánh tương phản trong mô hình tuyến tính tổng quát (GLM) như sau:

+ Hoàn thành các bước từ 1 đến 4 như ở ví dụ 1.

+ Tại cửa sổ Univariate, mở tiếp contrasts. Khi cửa sổ Univariate: Contrasts xuất hiện, bạn chọn Polynomial trong ô contrast. Kế đến chọn yếu tố rồi nhấp nút change và nút continue. Sau cùng thì chọn OK để hiển thị kết quả. Khi chọn contrast là Polynomial thì kết quả tương phản bao gồm hai Bảng. Bảng 1 là kết quả tương phản (Contrast Results (K Matrix)), còn Bảng 2 là kết quả kiểm định (Test Results).

Bảng 4.13. Các hệ số tương phản (Contrast Coefficients)

Contrast	Nghiem thức			
	Đối chứng	P	K	PK
1	0	-1	2	-1
2	-1	1	-1	1

Bảng 4.14. Kiểm định tương phản (Contrast Tests)

	Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
Assume equal variances	1	-4.9900	.9067	-5.503	20	.000
	2	5.2000	.7404	7.024	20	.000
Does not assume equal variances	1	-4.9900	.8437	-5.914	12.061	.000
	2	5.2000	.7404	7.024	19.123	.000

Giải thích và nhận xét kết quả kiểm định tương phản. Bảng 4.13 báo cáo các bộ hệ số tương phản mà bạn đã yêu cầu. Bảng 4.14 hiển thị kết quả của các tương phản. Ở Bảng 4.9, kiểm định sự bằng nhau của phương sai đã cho thấy các mẫu có phương sai bằng nhau. Vì thế, ở đây bạn chỉ cần lưu ý đến kết quả của phần giả thiết phương sai bằng nhau (Assume

equal variances). Vì cả hai tổ hợp mẫu kiểm định tương phản đều có $P = 0.000 (< 0.01)$, nên sinh trưởng trung bình về D của Sao đen trên những công thức có bón P khác rõ rệt với những công thức không bón P ở mức tin cậy 99%.

B. PHÂN TÍCH PHƯƠNG SAI 1 YẾU TỐ: Kiểu khối đầy đủ ngẫu nhiên

Giả sử cần xác định ảnh hưởng của 5 cấp độ tàn che khác nhau đến sinh trưởng chiều cao (H, cm) của cây con Dầu song năng 8 tháng tuổi trong giai đoạn vườn ươm. Thí nghiệm được bố trí theo kiểu khối đầy đủ ngẫu nhiên với 5 nghiệm thức, mỗi nghiệm thức lặp lại 3 lần (mỗi lần là 1 khối). Hãy phân tích kết quả thí nghiệm theo mô hình tuyến tính tổng quát: $Y = \beta_0 + \beta_1 X_1 + \beta_2 \times \text{Khối} + \varepsilon$; trong đó X_1 – yếu tố thí nghiệm, ε – sai số ngẫu nhiên. Để giải quyết vấn đề đặt ra, bạn có thể thực hiện theo các bước sau đây:

Bước 1. Nhập số liệu. Số liệu được nhập vào Bảng tính SPSS theo 3 cột (**Bảng 4.15**): cột 1 nhập thứ tự đại lượng quan sát (có hay không cũng được), cột 2 nhập kết quả thí nghiệm (biến phụ thuộc Y); cột 3 nhập mã biến độc lập (biến lập nhóm theo yếu tố thí nghiệm); cột 4 nhập số lần lặp lại của mỗi nghiệm thức (biến lập nhóm theo số lần lặp).

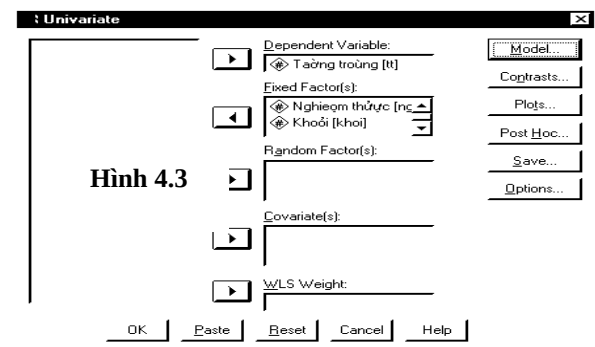
Bảng 4.15. Nhập số liệu trong phân tích phương sai 1 yếu tố bố trí theo khối
(Chi tiết số liệu xem ở phụ biểu 2)

Thứ tự	Biến phụ thuộc (H, cm)	Nghiệm thức	Khối (số lần lặp)
1	73.2	1	1
2	81.0	2	1
3	90.4	3	1
4	76.6	4	1
5...	82.8...	1...	2...

Bước 2. Kiểm tra giả thuyết về sự bằng nhau của phương sai mẫu, về sự khác biệt giữa các trung bình mẫu, so sánh sự khác nhau giữa các cặp số trung bình mẫu và kiểm định phân hạng các nghiệm thức được thực hiện theo các thủ tục sau đây:

+ Từ Menu chính, mở Analyze → General Linear Model → Univariable → Dependent List (đến đây đưa biến phụ thuộc vào bằng cách chọn nó và nhấn nút mũi tên bên **cạnh**) → factor (đưa vào các biến nghiệm thức và khối bằng cách chọn từng biến hoặc cả hai biến và nhấn nút mũi tên bên **cạnh**) (**Hình 4.3**).

+ Xác định ảnh hưởng của từng yếu tố nghiên cứu (ảnh hưởng chính – Main Effects) và ảnh hưởng của khối. Lưu ý rằng, khi bố trí thí nghiệm 1 yếu tố với K nghiệm thức, mỗi nghiệm thức lặp lại L khối thì bạn có thể vận dụng mô hình phân tích phương sai hai yếu tố để phân tích. Trong trường hợp này, mô hình phân tích thống kê không bao gồm sự tương tác giữa các yếu tố thí nghiệm với yếu tố khối. Do đó, sau khi chọn Model, tiếp theo tại cửa sổ Univariate Model bạn chỉ chọn yếu tố



nghiệm thức trong ô Factors & Covariates. Sau đó chọn Main Effects từ ô Build Terms và nhấn nút mũi tên bên cạnh để chuyển nghiệm thức sang ô Model.

Để phân tích thống kê, từ cửa sổ Univariate, bạn hãy thực hiện như sau:

+ Chọn Options. Thủ tục này tính các thống kê mô tả và kiểm tra sự bằng nhau của phương sai mẫu... Nếu chọn Parameter Estimates tại cửa sổ Display thì chương trình sẽ hiển thị các tham số của phương trình hồi quy. Nếu chọn Observed power thì chương trình sẽ hiển thị tham số Noncenter Parameter – đó là tham số phi trung tâm, dùng để tính độ mạnh của phép kiểm định (Observed Power). Khi Observed Power > 1 hay > 0.9995 thì những ảnh hưởng của yếu tố nghiên cứu đến biến phụ thuộc được thấy trong mẫu sẽ được phát hiện đến 95% trường hợp. Tại đây, bạn cũng có thể thay đổi mức ý nghĩa (Significance Level) mà bạn muốn sử dụng cho các kiểm định phân hạng. Mức ý nghĩa mặc định là 0.05.

+ Chọn Post Hoc Test. Thủ tục này so sánh đa phạm vi cho các trung bình mẫu (Post Hoc Multiple Comparisons for Observed Means). Để đạt được mục đích này, bạn chọn Post Hoc Test, sau đó đánh dấu ô Bonferroni và Tukey's Test.

Cuối cùng thì bạn chọn OK để chương trình tính toán và hiển thị kết quả. Kết quả hiển thị trên màn Hình bao gồm những thống kê sau đây: thống kê mô tả (Descriptive Statistics), kiểm định ngang bằng phương sai (Levene's Test of Equality of Error Variances), kiểm định ảnh hưởng giữa các yếu tố (Test of Between Subjects Effects), ước lượng trung bình nghiệm thức và trung bình khối, so sánh từng cặp đối với các nghiệm thức và khối, phân hạng các nghiệm thức (Homogeneous Subsets)... Kết quả phân tích biến động (ANOVA) được ghi lại ở **Bảng 4.16**.

Bảng 4.16. Kiểm định ảnh hưởng của các yếu tố thí nghiệm

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model (1)	522358.750	7	74622.679	18.553	.000
Intercept (2)	6446801.250	1	6446801.250	1602.851	.000
Nghiệm thức	264955.000	4	66238.750	16.469	.000
Khối	257403.750	3	85801.250	21.333	.000
Error (3)	48265.000	12	4022.083		
Total	7017425.000	20			
Corrected Total	570623.750	19			

a R Squared = .915 (Adjusted R Squared = .866)

Giải thích và nhận xét kết quả. Dòng có nhãn Corrected Model biểu thị những biến thiên do mô hình. Trị số này bằng tổng các biến thiên do yếu tố nghiên cứu cộng với biến thiên do khối gây ra. Giả thuyết H_0 : Các yếu tố nghiên cứu ảnh hưởng đồng đều đến kết quả thí nghiệm. Khi $P - \text{value} < 0.000$ (trị làm tròn của Sig. = 0.0005) thì các yếu tố nghiên cứu (biến lập nhóm hay biến độc lập X_i) có ảnh hưởng không đồng đều (hay rõ rệt) đến kết quả thí nghiệm (biến phụ thuộc Y). Ngược lại, khi $P - \text{value} > \alpha$ (0.05 hoặc 0.01) thì các yếu tố nghiên cứu có ảnh hưởng đồng đều đến kết quả thí nghiệm.

+ Kiểm định ảnh hưởng chính (yếu tố nghiệm thức và khối). Giả thuyết H_0 : Từng yếu tố nghiệm thức và khối ảnh hưởng đồng đều đến kết quả thí nghiệm. Khi $P - \text{value} < 0.05$ thì giả thuyết trên bị bác bỏ. Ngược lại, nếu $P > 0.05$ thì giả thuyết trên được chấp nhận.

+ Error: Cho biết những sai số không thể giải thích được bởi mô hình.

+ Total: Tổng bình phương của biến phụ thuộc (Y_i^2) = (1) + (2) + (3).

+ Corrected Total. Tham số này cho biết tổng chênh lệch bình phương của bộ phận tạo ra sự biến thiên của biến phụ thuộc do mô hình (1) cộng với bộ phận tạo ra sự biến thiên của biến phụ thuộc do sai số (3).

+ Intercept. Tham số này cho biết điểm chặn của đường hồi quy giữa trung bình các nghiệm thức. Nếu $P < 0.05$ thì giả thuyết về điểm chặn của mô hình bằng không sẽ bị bác bỏ.

+ Ghi chú (a) trong Bảng kiểm định ảnh hưởng của các yếu tố thí nghiệm (kí hiệu ở Bảng trên là a R Squared = .915; Adjusted R Squared = .866) cho biết 91,5% biến thiên của yếu tố phụ thuộc được giải thích bởi mô hình.

C. PHÂN TÍCH PHƯƠNG SAI 1 YẾU TỐ: Thiết kế theo kiểu bình phương la tinh

Giả sử cần nghiên cứu ảnh hưởng của 6 cấp hàm lượng Super photphat (0, 1, 2, 3, 4, 5% so với trọng lượng bầu) đến sinh trưởng chiều cao (H, cm) của Dầu song nà 8 tháng tuổi trong điều kiện vườn ươm. Thí nghiệm được bố trí theo khối với 6 lần lặp lại. Tổng cộng có 36 lô thí nghiệm. Các lô thí nghiệm được phân phối theo sơ đồ gồm 6 hàng và 6 cột như **Hình 4.4**. Hãy phân tích kết quả thí nghiệm theo mô hình tuyến tính tổng quát: $Y = \beta_0 + \beta_1 X_1 + \beta_2 \times \text{Khối cột} + \beta_3 \times \text{Khối hàng} + \varepsilon$; trong đó X_1 – yếu tố thí nghiệm, ε – sai số ngẫu nhiên. Để giải quyết vấn đề đặt ra trên đây, bạn thực hiện các bước sau đây:

		Hàng					
		1	2	3	4	5	6
Cột	1	A	B	C	D	E	F
	2	B	A	E	F	C	D
	3	C	F	A	E	D	B
	4	D	C	B	A	F	E
	5	E	D	F	C	B	A
	6	F	E	D	B	A	C

Hình 4.4. Sơ đồ bố trí thí nghiệm 1 yếu tố kiểu bình phương la tinh
A = 0, B = 1, C = 2, D = 3, E = 4, F = 5% Super photphat so với trọng lượng bầu

Bước 1. Nhập số liệu vào Bảng tính SPSS hoặc Bảng tính Excel như sau (**Bảng 4.17**): Cột 1 nhập biến phụ thuộc (H, cm), cột 2 nhập mã số biểu thị thứ tự nghiệm thức (khi nhập phải chú ý đến sơ đồ bố trí các nghiệm thức trong mỗi cột và hàng), cột 3 nhập mã số biểu thị thứ tự hàng (khối hàng), cột 4 nhập mã số biểu thị thứ tự cột (khối cột). Để tránh mất số liệu, bạn hãy lưu số liệu này thành một file.

Bảng 4.17. Ảnh hưởng của Super photphat đến sinh trưởng H của Dầu song nà 8 tháng tuổi
(Chi tiết số liệu xem ở phụ lục 3)

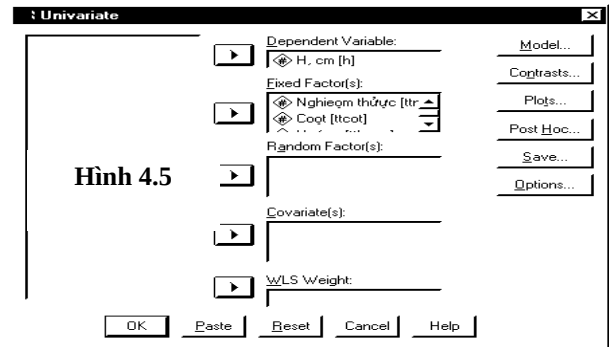
H, cm	Nghiệm thức	Thứ tự hàng	Thứ tự cột
67	1	1	1
46	2	2	1
47	3	3	1
28	4	4	1
15	5	5	1
7	6	6	1
...	...	...	...
19	4	2	6
57	2	3	6

17	5	4	6
87	1	5	6
45	3	6	6

Bước 2. Gọi chương trình tính toán. Trình tự kiểm tra giả thuyết về sự khác biệt giữa các trung bình mẫu (nghiệm thức), so sánh sự khác nhau giữa các cặp số trung bình mẫu và kiểm định phân hạng các nghiệm thức được thực hiện theo những thủ tục sau đây:

+ Từ Menu chính, mở Analyze → General Linear Model → Univariate → Dependent List (đến đây đưa vào biến phụ thuộc H) → factor (đưa vào các biến nghiệm thức, cột và hàng) (Hình 4.5).

+ Kiểm định ảnh hưởng của yếu tố nghiệm cứu (ảnh hưởng chính – Main Effects) đến biến phụ thuộc. Trước hết, hãy chọn Model. Kế đến tại cửa sổ Univariate



Model, bạn đánh dấu chọn Custom. Tiếp theo chọn yếu tố nghiệm thức (Super lân), cột và hàng (lần lặp hay khối) trong ô Factors & Covariates; đồng thời chọn Main Effects từ ô Build Terms và nhấn nút mũi tên bên cạnh để chuyển chúng sang ô Model.

+ Chọn Options. Thủ tục này xác định các thống kê mô tả, kiểm định ảnh hưởng của từng yếu tố... Để đạt được mục đích này, bạn chọn các biến từ ô Factor(s) and factor interaction rồi chuyển chúng sang ô Display means for. Sau đó chọn tiếp Compare main effects, Descriptive statistics và Homogeneity tests...

+ Chọn Post Hoc Test. Thủ tục này so sánh đa phạm vi cho các trung bình mẫu (Post Hoc Multiple Comparisons for Observed Means). Để đạt được mục đích này, bạn chọn Post Hoc Test, sau đó đánh dấu ô Bonferroni và Tukey's Test. Cuối cùng chọn OK để chương trình tính toán và hiển thị kết quả.

Giải thích và nhận xét kết quả. Kết quả phân tích thí nghiệm hiển thị nhiều thống kê khác nhau, trong đó có một số nội dung đáng quan tâm sau đây (Bảng 4.18 - 4.22).

+ Bảng 4.18 báo cáo các thống kê mô tả; nội dung bao gồm trung bình nghiệm thức theo cột và hàng. Bảng 4.19 đưa ra các thống kê ước lượng đối với trung bình nghiệm thức và khoảng tin cậy của số trung bình nghiệm thức.

Bảng 4.18. Các thống kê mô tả (Descriptive Statistics)

Dependent Variable: H, cm					
Nghiệm thức	Cột	Hàng	Mean	Std. Deviation	N
1	1	1	44.700	.	1
	2	2	41.300	.	1
	3	3	43.300	.	1
	4	4	60.000	.	1
	5	6	70.000	.	1
	6	5	58.000	.	1
	Total		58.000	.	1
...	...	...	...	...	...
Total	Total	1	21.883	17.512	6
		2	22.000	16.144	6
		3	24.650	16.014	6
		4	26.900	21.071	6

	6	36.000	24.743	6
	5	30.550	21.800	6
	Total	26.997	19.032	36

Bảng 4.19. Các thống kê ước lượng (Estimates)

Nghiệm thức	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	52.883	2.335	48.014	57.753
...	...	...	...	...
6	3.317	2.335	-1.553	8.186

+ Bảng ANOVA (Bảng 4.20) trình bày kết quả kiểm định ảnh hưởng của yếu tố thí nghiệm đến yếu tố phụ thuộc. Khi P – value (hay Sig) < 0.05 thì giả thuyết cho rằng hàm lượng phân Super photphat khác nhau ảnh hưởng đồng đều đến sinh trưởng chiều cao của Dầu song nạng sẽ bị bác bỏ và ngược lại. Ở ví dụ này, vì P = 0.000 < 0.01 nên sinh trưởng của Dầu song nạng khác nhau rất rõ rệt ở mức tin cậy 99%. Theo kết quả so sánh từng cặp số trung bình, bạn có thể thấy những cặp mẫu nào khác nhau. Để biết được những cặp mẫu nào khác nhau, bạn hãy quan sát giá trị P – value (Sig) của Bảng 4.21. Khi P < 0.05 thì giả thuyết cho rằng hai số trung bình mẫu không khác nhau ở mức ý nghĩa 0.05 (mặc định) sẽ bị bác bỏ và ngược lại.

Bảng 4.20. Kiểm định ảnh hưởng của yếu tố thí nghiệm đến yếu tố phụ thuộc

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	12024.097	15	801.606	24.514	.000
Intercept	26238.600	1	26238.600	802.413	.000
Nghiệm thức	10866.691	5	2173.338	66.464	.000
Khối cột	255.515	5	51.103	1.563	.216
Khối hàng	901.891	5	180.378	5.516	.002
Error	653.992	20	32.700		
Total	38916.690	36			
Corrected Total	12678.090	35			

a R Squared = .948 (Adjusted R Squared = .910)

Bảng 4.21. So sánh từng cặp số trung bình (Pairwise Comparisons)

		Mean Difference (I-J)	Std. Error	Sig. ^a	95% Confidence Interval for Difference ^a	
(I) Ng nghiệm thức	(J) Ng nghiệm thức				Lower Bound	Upper Bound
1	2	9.217×	3.301	.011	2.330	16.103
	3	23.883×	3.301	.000	16.997	30.770
	4	29.983×	3.301	.000	23.097	36.870
	5	42.667×	3.301	.000	35.780	49.553
	6	49.567×	3.301	.000	42.680	56.453
2	1	-9.217×	3.301	.011	-16.103	-2.330
	3	14.667×	3.301	.000	7.780	21.553
	4	20.767×	3.301	.000	13.880	27.653
	5	33.450×	3.301	.000	26.563	40.337
	6	40.350×	3.301	.000	33.463	47.237
...	...	...	...	...	...	...

+ **Bảng 4.22** trình bày kết quả kiểm định phân hạng các nghiệm thức. Những nghiệm thức không khác biệt thực sự (hay cùng hạng) được ghi lại trong cùng một cột. Trong **Bảng 4.22**, dòng có nhãn Sig cho biết các nghiệm thức không khác biệt thực sự với mức ý nghĩa P. Ở ví dụ này đã sử dụng mức tin cậy mặc định cho kiểm định đa phạm vi là 95%.

Bảng 4.22. Kiểm định phân hạng các nghiệm thức

Tukey HSD	Nghiệm thức	N	Subset		
			1	2	3
	6	6	3.317		
	5	6	10.217		
	4	6		22.900	
	3	6		29.000	
	2	6			43.667
	1	6			52.883
	Sig.		.331	.460	.100

4.3.2. PHÂN TÍCH PHƯƠNG SAI NHIỀU YẾU TỐ TỔNG QUÁT

Phân tích phương sai hai hay nhiều yếu tố tổng quát kiểm định ảnh hưởng của hai hay nhiều yếu tố định tính hoặc định lượng đến biến phụ thuộc. Khi có hai yếu tố nghiên cứu định lượng thì mô hình phân tích phương sai nhiều yếu tố tổng quát có dạng :

+ Nếu ảnh hưởng của hai yếu tố thí nghiệm (phân loại) là cộng tuyến tính thì phương trình đáp ứng có dạng: $Y = \beta_0 + \beta_1X_1 + \beta_2X_2$.

+ Nếu ảnh hưởng của hai yếu tố có tương tác với nhau thì phương trình đáp ứng có thể tồn tại ở các dạng sau đây :

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2.$$

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2 + \beta_4X_1^2X_2.$$

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2 + \beta_4X_1X_2^2.$$

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2 + \beta_4X_1^2 + \beta_5X_2^2 \dots$$

Một cách tổng quát: $Y = \beta_0 + \beta_1X_1 + \beta_2X_1^2 + \beta_3X_2 + \beta_4X_2^2 + \beta_5X_1X_2 + \beta_6X_1^2X_2 + \beta_7X_1X_2^2 + \beta_8X_1^2X_2^2$. Để thấy dạng nào là phù hợp, chúng ta phải thực hiện các kiểm định.

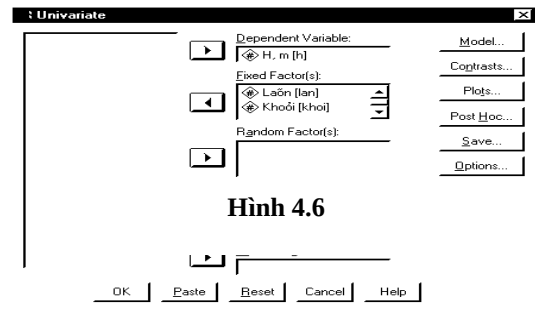
A. PHÂN TÍCH PHƯƠNG SAI HAI YẾU TỐ: **Bố trí theo khối hoàn toàn ngẫu nhiên**

Giả sử cần phải phân tích ảnh hưởng của 3 mức đạm (kí hiệu yếu tố A: 1, 2 và 3% đạm) và 3 mức lân (kí hiệu yếu tố B: 0.5, 1.0, 1.5%) đến sinh trưởng chiều cao (H, cm) của cây con Dầu song năng trong giai đoạn 6 tháng tuổi ở vườn ươm. Thí nghiệm bao gồm 9 nghiệm thức, mỗi nghiệm thức được lặp lại 5 lần; tổng cộng có 45 lô thí nghiệm. Mỗi lô thí nghiệm sử dụng 49 cây đồng đều về Hình thái và sức sống. Hãy phân tích và so sánh kết quả thí nghiệm theo mô hình tuyến tính tổng quát: $Y = \beta_0 + \beta_1A + \beta_2B + \beta_3AB + \beta_5 \times \text{Khối} + \epsilon$; trong đó A – yếu tố thí nghiệm 1, B – yếu tố thí nghiệm 2; Khối – số lần lặp; ϵ - sai số thí nghiệm.

Các giả thuyết: (1) Các mức đạm khác nhau (kí hiệu A) có ảnh hưởng đồng đều đến sinh trưởng H của Dầu song năng; (2) Các mức lân khác nhau (kí hiệu B) cũng có ảnh hưởng đồng đều đến sinh trưởng H của Dầu song năng; (3) Sự phối hợp của các mức đạm với các mức lân khác nhau có ảnh hưởng đồng đều đến sinh trưởng H của Dầu song năng. Để kiểm định các giả thuyết trên, bạn hãy thực hiện các bước sau đây:

Bước 1. Nhập số liệu vào Bảng tính SPSS như sau (**Bảng 4.23**): cột 1 hoặc hàng 1 nhập biến kết quả thí nghiệm (H, cm); cột 2 nhập số lần lặp của mỗi công thức (số khối); cột 3 hoặc hàng 3 nhập số lần lặp lại của yếu tố đậm (biến phân loại A); cột 4 hoặc hàng 4 nhập số lần lặp lại của yếu tố lân (biến phân loại B). Sau khi nhập xong dữ liệu, bạn nên ghi lại thành 1 file để có thể dùng về sau.

Bước 2. Kiểm định giả thuyết về sự bằng nhau của phương sai mẫu, về sự khác biệt giữa các trung bình mẫu, kiểm định phân hạng các nghiệm thức được thực hiện theo các thủ tục sau đây:



Hình 4.6

+ Từ Menu chính, mở Analyze → General Linear Model → Univariate → Dependent List (đến đây đưa biến phụ thuộc H vào) → factor (đưa vào các biến phân loại cho nghiệm thức đậm, lân và khối bằng cách chọn 1 hay nhiều biến và nhấn nút mũi tên bên cạnh để chuyển chúng sang ô Fixed factors, **Hình 4.6**).

Bảng 4.23. Cách nhập dữ liệu trong phân tích phương sai 2 yếu tố được bố trí ngẫu nhiên
(Chi tiết số liệu xem ở phụ lục 4)

H, cm	TT. khối	Đậm (A)	Lân (B)
60	1	1	1
56	2	1	1
58	3	1	1
62	4	1	1
60	5	1	1
70	1	1	2
76	2	1	2
...	...	...	...

+ Xác định ảnh hưởng của từng yếu tố nghiên cứu (ảnh hưởng chính – Main Effects) và ảnh hưởng tổng hợp (sự tương tác giữa các yếu tố – Interactions) của các yếu tố nghiên cứu đến biến phụ thuộc. Trước hết, bạn hãy chọn Model. Kế đến tại cửa sổ Univariate Model, bạn đánh dấu chọn Custom. Tiếp theo chọn yếu tố đậm, lân và số lần lặp (khối) trong ô Factors & Covariates; đồng thời chọn Main Effects từ ô Build Terms và nhấn nút mũi tên bên cạnh để chuyển chúng sang ô Model. Sau đó, bạn lại chọn cả hai yếu tố đậm và lân trong ô Factors & Covariates; kế đến chọn Interactions (tương tác) từ ô Build Terms và nhấn mũi tên bên cạnh để chuyển chúng sang ô Model.

+ Chọn Options. Thủ tục này xác định các thống kê mô tả, kiểm tra sự bằng nhau của phương sai mẫu, so sánh các ảnh hưởng chính... Để đạt được mục đích đặt ra, bạn chọn các biến từ ô Factor(s) and factor interaction rồi chuyển chúng sang ô Display means for. Sau đó chọn tiếp Compare main effects, descriptive statistics và homogeneity tests...

+ Chọn Post Hoc Test. Thủ tục này so sánh đa phạm vi cho các trung bình mẫu (Post Hoc Multiple Comparisons for Observed Means). Để đạt được mục đích này, bạn chọn Post Hoc Test, sau đó đánh dấu ô Bonferroni và Tukey's Test. Cuối cùng chọn OK để chương trình tính toán và hiển thị kết quả.

Giải thích và nhận xét kết quả. Kết quả hiển thị trên màn Hình bao gồm các Bảng sau đây: các thống kê mô tả (Descriptive Statistics), kiểm định ngang bằng phương sai (Levene's Test of Equality of Error Variances), kiểm định ảnh hưởng giữa các yếu tố (Test of Between Subjects Effects), ước lượng trung bình nghiệm thức và trung bình khối, so sánh từng cặp (Multiple Comparisons) đối với các nghiệm thức và khối, phân hạng các nghiệm thức (Homogeneous Subsets)...Dưới đây chỉ trích dẫn Bảng phân tích biến động (**Bảng 4.24**) và Bảng phân hạng các nghiệm thức (**Bảng 4.25**) để các bạn tham khảo.

Bảng 4.24. Kiểm định ảnh hưởng của đạm và lân đến H của Dầu song nòng

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	10665.422	12	888.785	35.830	.000
Intercept	333852.800	1	333852.800	13458.791	.000
Đạm	3040.533	2	1520.267	61.287	.000
Lân	578.133	2	289.067	11.653	.000
Khối	257.422	4	64.356	2.594	.055
Đạm × Lân	6789.333	4	1697.333	68.426	.000
Error	793.778	32	24.806		
Total	345312.000	45			
Corrected Total	11459.200	44			

a. R Squared = .931 (Adjusted R Squared = .905)

Bảng 4.25. Phân hạng các nghiệm thức về H theo yếu tố đạm và lân

Đạm		N	Subset		
			1	2	3
Tukey HSD	1	15	76.13		
	3	15		86.00	
	2	15			96.27
	Sig.		1.000	1.000	1.000

Lân		N	Subset	
			1	2
Tukey HSD	2	15	83.47	
	3	15	83.73	
	1	15		91.20
	Sig.		.988	1.000

Bạn hãy quan sát Bảng ANOVA và đưa ra nhận định về ảnh hưởng của các yếu tố thí nghiệm đến biến phụ thuộc. Khi $P - \text{value} < 0.05$ thì các yếu tố nghiên cứu (biến lập nhóm hay biến độc lập X_i) có ảnh hưởng rõ rệt đến kết quả thí nghiệm (biến phụ thuộc Y). Ngược lại, khi $P - \text{value} > 0.05$ thì các yếu tố nghiên cứu có ảnh hưởng không rõ rệt đến kết quả thí nghiệm. Từ **Bảng 4.24** cho thấy, từng yếu tố và tổ hợp hai yếu tố có ảnh hưởng rõ rệt đến kết quả thí nghiệm. Những giải thích chi tiết cho các thành phần khác được thực hiện tương tự như phân tích phương sai một yếu tố.

B. PHÂN TÍCH PHƯƠNG SAI 2 YẾU TỐ: Kiểu thí nghiệm hoàn toàn ngẫu nhiên

Giả sử cần thán dò sinh trưởng chiều cao (H, cm) của cây con Dầu song nòng trong giai đoạn 6 tháng tuổi dưới ảnh hưởng của 3 mức đạm (kí hiệu yếu tố A: 1, 2 và 3% đạm); mỗi mức đạm được sủi sung thêm 3 mức lân (kí hiệu yếu tố B: 0.5, 1.0, 1.5%). Mỗi lô thí nghiệm sử dụng 15 cây 1 tuần tuổi đồng đều về Hình thái và sức sống. Hãy phân tích và so sánh kết quả thí nghiệm theo mô hình GLM: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$; trong đó: X_1 và X_2 – yếu tố thí nghiệm 1 và 2; ε - sai số thí nghiệm

Các giả thuyết: (1) Các mức đạm khác nhau (kí hiệu A) có ảnh hưởng đồng đều đến sinh trưởng H của Dầu song nòng; (2) Các mức lân khác nhau (kí hiệu B) có ảnh hưởng đồng đều đến sinh trưởng H của Dầu song nòng; (3) Sự phối hợp của các mức đạm và lân khác nhau cũng có ảnh hưởng đồng đều đến sinh trưởng H của Dầu song nòng. Để kiểm định các giả thuyết trên, bạn hãy thực hiện các bước sau:

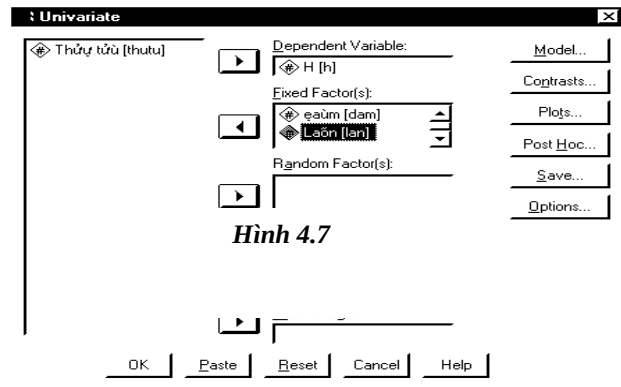
Bước 1. Nhập số liệu vào Bảng tính SPSS như sau (**Bảng 4.26**): cột 1 hoặc hàng 1 nhập biến kết quả thí nghiệm (H, cm); cột 2 nhập số trường hợp thu mẫu (lần lặp) của mỗi nghiệm thức; cột 3 hoặc hàng 3 nhập mã số biểu thị số lần lặp lại của yếu tố đạm (biến phân loại A); cột 4 hoặc hàng 4 nhập mã số biểu thị số lần lặp lại của yếu tố lân (biến phân loại B). Sau khi nhập xong dữ liệu, bạn nên ghi lại thành 1 file để có thể dùng về sau.

Bảng 4.26. Nhập dữ liệu trong phân tích phương sai 2 yếu tố bố trí ngẫu nhiên
(Chi tiết số liệu xem ở phụ lục 5)

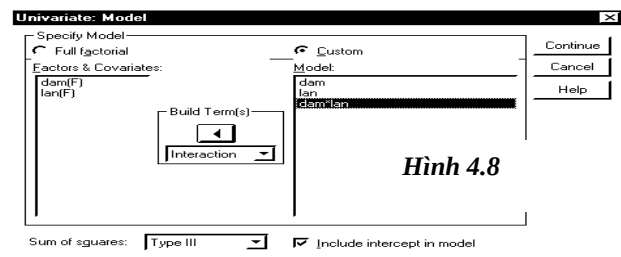
H, cm	TT. cây	Đạm (A)	Lân (B)
55	1	1	1
51	2	1	1
...	...	...	...
65	15	3	3

Bước 2. Kiểm tra giả thuyết về sự bằng nhau của phương sai mẫu, về sự khác biệt giữa các trung bình mẫu, so sánh sự khác nhau giữa các cặp số trung bình mẫu và kiểm định phân hạng các nghiệm thức được thực hiện theo các thủ tục sau:

+ Từ Menu chính, bạn mở Analyze → General Linear Model → Univariable → Dependent List (đến đây đưa biến phụ thuộc H vào) → factor (đến đây đưa vào các biến phân loại cho nghiệm thức đạm, lân bằng cách chọn 1 hay nhiều biến và nhấn nút mũi tên bên cạnh để chuyển chúng sang ô Fixed factors (**Hình 4.7**).



Hình 4.7



Hình 4.8

+ Xác định ảnh hưởng của từng yếu tố nghiên cứu (ảnh hưởng chính – Main Effects) và ảnh hưởng tổng hợp (sự tương tác giữa các yếu tố – Interactions) của các yếu tố nghiên cứu đến biến phụ thuộc. Trước hết, bạn hãy chọn Model. Kế đến tại cửa sổ Univariate Model, bạn đánh dấu chọn Custom. Tiếp theo chọn yếu tố đậm và lân trong ô Factors & Covariates; đồng thời chọn Main Effects từ ô Build Terms và nhấn nút mũi tên bên cạnh để chuyển các nghiệm thức sang ô Model. Sau đó, bạn lại chọn cả hai yếu tố đậm và lân trong ô Factors & Covariates; kế đến chọn Interactions (tương tác) từ ô Build Terms và nhấn mũi tên bên cạnh để chuyển chúng sang ô Model (**Hình 4.8**).

+ Chọn Options. Thủ tục này xác định các thống kê mô tả, kiểm tra sự bằng nhau của phương sai... Để đạt được mục đích này, bạn chọn các biến từ ô Factor(s) and factor interaction rồi chuyển chúng sang ô Display means for. Sau đó chọn tiếp Compare main effects, Descriptive statistics và Homogeneity tests...

+ Chọn Post Hoc Test. Thủ tục này so sánh đa phạm vi cho các trung bình mẫu (Post Hoc Multiple Comparisons for Observed Means). Để đạt được mục đích này, bạn chọn Post Hoc Test, sau đó đánh dấu ô Bonferroni và Tukey's Test. Cuối cùng chọn OK để chương trình tính toán và hiển thị kết quả.

Kết quả hiển thị trên màn Hình bao gồm các Bảng sau đây: các thống kê mô tả (Descriptive Statistics), kiểm định ngang bằng phương sai (Levene's Test of Equality of Error Variances), kiểm định ảnh hưởng giữa các yếu tố (Test of Between Subjects Effects), ước lượng trung bình nghiệm thức, so sánh từng cặp (Multiple Comparisons) đối với các nghiệm thức, phân hạng các nghiệm thức (Homogeneous Subsets). Dưới đây chỉ trích dẫn Bảng thống kê mô tả (**Bảng 4.27**), kiểm định ngang bằng phương sai (**Bảng 4.28**), phân tích biến động (**Bảng 4.29**), so sánh cặp đối với các nghiệm thức đậm, lân (**Bảng 4.30**) và Bảng phân hạng các nghiệm thức (**Bảng 4.31**) để các bạn tham khảo.

Bảng 4.27. Thống kê mô tả (Descriptive Statistics)

Đậm	Lân	Mean	Std. Deviation	N
1	1	55.93	3.81	15
	2	57.43	2.69	15
	3	57.36	2.56	15
	Total	56.91	3.08	45
....				
3	1	65.53	3.81	15
	2	67.03	2.69	15
	3	66.96	2.56	15
	Total	66.51	3.08	45
Total	1	59.13	5.90	45
	2	60.63	5.28	45
	3	60.56	5.21	45
	Total	60.11	5.48	135

Bảng 4.28. Kiểm định ngang bằng phương sai

F	df1	df2	Sig.
0.521	8	126	0.839

a. Design: Intercept + DAM + LAN + DAM × LAN

Giải thích và nhận xét kết quả. **Bảng 4.27** chỉ ra các trị trung bình và sai tiêu chuẩn về H của các nghiệm thức (đậm, lân). **Bảng 4.28 kiểm** định ngang bằng phương sai giữa các

nghiệm thức. **Bảng 4.29** cho biết kết quả phân tích biến động (ANOVA). Ở đây việc kiểm định được thiết kế theo mô hình tuyến tính đa yếu tố: Intercept (điểm chặn) + đạm + lân + đạm×lân. Vì P – value ứng với các dòng đạm và lân đều nhỏ hơn 0.05, nên các yếu tố đạm và lân có ảnh hưởng không đồng đều (rõ rệt) đến kết quả thí nghiệm (biến phụ thuộc Y). Ngược lại, ở dòng Đạm×Lân có $P = 1.000 > 0.05$, nên ảnh hưởng phối hợp của đạm với lân đến kết quả thí nghiệm là không rõ rệt. Nếu thiết kế lại mô hình chỉ bao gồm (Intercept + đạm + lân) thì kết quả như **Bảng 4.30**. Dòng có nhân Intercept được đưa ra để kiểm định giả thuyết H_0 : Trung bình các nghiệm thức phân bố trên cùng một đường thẳng. Giả định này bị bác bỏ, vì $P < 0.05$. Vì thế, ta có thể tìm phương trình đáp ứng ở dạng bậc hai. **Bảng 4.31** trình bày kết quả phân hạng các nghiệm thức theo yếu tố lân và đạm.

Bảng 4.29. Kiểm định ảnh hưởng của đạm và lân đến H của Dầu song nàg

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2829.161	8	353.645	37.509	.000
Intercept	487765.601	1	487765.601	51734.98	.000
Đạm	2764.800	2	1382.400	146.625	.000
Lân	64.361	2	32.181	3.413	.036
Đạm × Lân	.000	4	.000	.000	1.000
Error	1187.948	126	9.428		
Total	491782.710	135			
Corrected Total	4017.109	134			

a R Squared = .704 (Adjusted R Squared = .686)

Bảng 4.30. Kiểm định ảnh hưởng của đạm và lân đến H của Dầu song nàg

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2829.161	4	707.290	77.400	.000
Intercept	487765.601	1	487765.601	53377.360	.000
Đạm	2764.800	2	1382.400	151.279	.000
Lân	64.361	2	32.181	3.522	.032
Error	1187.948	130	9.138		
Total	491782.710	135			
Corrected Total	4017.109	134			

a R Squared = .704 (Adjusted R Squared = .695)

Bảng 4.31. Phân hạng các nghiệm thức theo H (trên ví dụ yếu tố đạm)

	Đạm	N	Subset	
			1	2
Tukey HSD ^{a,b}	1	45	56.91	
	2	45	56.91	
	3	45		66.51
	Sig.		1.000	1.000

a Uses Harmonic Mean Sample Size = 45.000; b Alpha = .05.

C. PHÂN TÍCH PHƯƠNG SAI 2 YẾU TỐ: Kiểu thí nghiệm bình phương la tinh

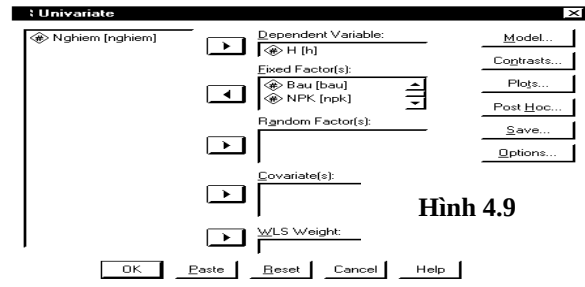
Giả thiết cần đánh giá ảnh hưởng của hai yếu tố (kích thước bầu và phân tổng hợp NPK) đến sinh trưởng chiều cao của Dầu song nàg 8 tháng tuổi. Yếu tố kích thước bầu (kí hiệu X) với 2 mức độ: 15×25 cm và 20×30 cm. Yếu tố phân NPK (kí hiệu P) với 3 mức độ: 1%, 2% và 3% so với trọng lượng bầu. Như vậy, ta có $2 \times 3 = 6$ nghiệm thức. Mỗi nghiệm thức được lặp lại 6 lần. Các nghiệm thức được kí hiệu như sau: A = X_1P_1 ; B = X_1P_2 ; C = X_1P_3 , D =

X_2P_1 ; $E = X_2P_2$; $F = X_2P_3$. Hãy phân tích và so sánh kết quả thí nghiệm theo mô hình tuyến tính tổng quát : $Y = \beta_0 + \beta_1\text{Hàng} + \beta_2\text{Cột} + \beta_3X + \beta_4P + \beta_5XP + \varepsilon$; trong đó X – yếu tố kích thước bầu, P – yếu tố phân NPK; ε - sai số thí nghiệm. Để đạt được mục đích đặt ra, bạn hãy thực hiện như sau:

Bước 1. Nhập số liệu vào Bảng tính như sau (Bảng 4.32): cột 1 hoặc hàng 1 nhập biến phụ thuộc (H , cm); cột 2 nhập số thứ tự nghiệm thức; cột 3 và 4 nhập tương ứng mã khối theo hàng và theo cột; cột 5 nhập mã số kích thước bầu theo khối hàng; cột 6 nhập mã số phân tổng hợp NPK theo khối cột. Sau khi nhập xong dữ liệu, bạn nên ghi lại thành 1 file để có thể dùng về sau.

Bước 2. Phân tích thống kê. Kiểm tra giả thuyết về sự bằng nhau của phương sai mẫu, về sự khác biệt giữa các trung bình mẫu, so sánh sự khác nhau giữa các cặp số trung bình mẫu và kiểm định phân hạng các nghiệm thức được thực hiện theo các thủ tục sau đây:

+ Mở Analyze → General Linear Model → Univariable → Dependent List (đến đây đưa vào biến phụ thuộc H) → factor (đến đây đưa vào các biến phân loại: hàng, cột, kích thước bầu và NPK bằng cách chọn 1 hay nhiều biến và nhấn nút mũi tên bên cạnh để chuyển chúng sang ô Fixed factors (Hình 4.9).

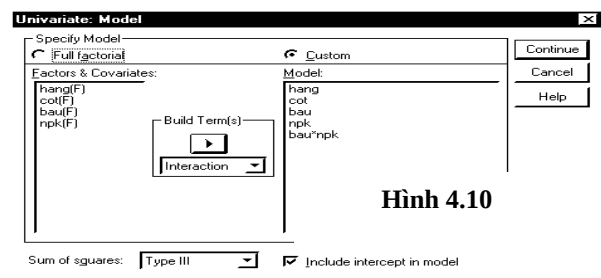


Hình 4.9

Bảng 4.32. Nhập dữ liệu trong phân tích phương sai 2 yếu tố được bố trí theo kiểu bình phương la tinh
(Chi tiết số liệu xem ở phụ lục 6)

H, cm	Nghiệm thức	Khối hàng	Khối cột	Kích thước bầu	Phân NPK
(1)	(2)	(3)	(4)	(5)	(6)
67.5	3	1	1	1	3
68.8	4	1	2	2	1
71.3	6	1	3	2	3
...	...	...	...	...	...

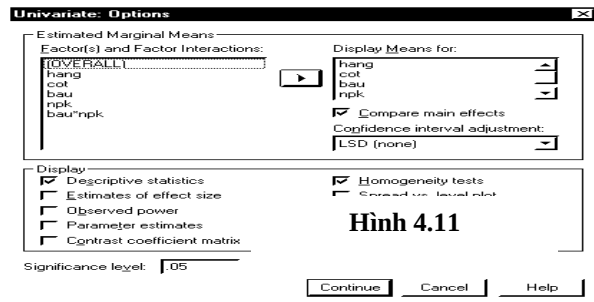
+ Để xác định ảnh hưởng của từng yếu tố nghiên cứu (ảnh hưởng chính – Main Effects), trước hết bạn hãy chọn Model. Kế đến tại cửa sổ Univariate Model bạn đánh dấu chọn Custom. Tiếp theo chọn các yếu tố (hàng, cột, kích thước bầu, NPK) trong ô Factors & Covariates; đồng thời chọn Main Effects từ ô Build Terms và nhấn nút mũi tên bên cạnh để chuyển các nghiệm thức sang ô Model. Để xác định ảnh hưởng tổng hợp (sự tương tác giữa các yếu tố – Interactions) của các yếu tố nghiên cứu đến biến phụ thuộc, bạn chọn cả hai yếu tố kích thước bầu và NPK trong ô Factors & Covariates; kế đến chọn Interactions (tương tác) từ ô Build Terms và nhấn mũi tên bên cạnh để chuyển chúng sang ô Model (Hình 4.10).



Hình 4.10

+ Chọn Options. Thủ tục này xác định các thống kê mô tả, kiểm tra sự bằng nhau của phương sai... Để đạt được mục đích đặt ra, bạn hãy chọn tất cả các yếu tố trong ô Factor(s) and Factor Interactions và chuyển chúng sang ô Display Means for; đồng thời đánh dấu chọn các mục thống kê **mô tả như Hình 4.11**.

+ Chọn Post Hoc Test. Thủ tục này so sánh đa phạm vi cho các trung bình mẫu (Post Hoc Multiple Comparisons for Observed Means). Để đạt được mục đích đặt ra, bạn chọn Post Hoc Test, sau đó đánh dấu ô Bonferroni và Tukey's Test. Cuối cùng chọn OK để chương trình tính toán và hiển thị kết quả.



Hình 4.11

Kết quả hiển thị trên màn Hình bao gồm các Bảng sau đây: các thống kê mô tả (Descriptive Statistics), kiểm định ngang bằng phương sai (Levene's Test of Equality of Error Variances), kiểm định ảnh hưởng giữa các yếu tố (Test of Between Subjects Effects), ước lượng trung bình nghiệm thức, so sánh từng cặp nghiệm thức (Multiple Comparisons), phân hạng các nghiệm thức (Homogeneous Subsets). Dưới đây chỉ trích dẫn Bảng phân tích biến động (Bảng 4.33) và Bảng phân hạng các nghiệm thức (Bảng 4.34) theo yếu tố kích thước bầu và NPK để các bạn tham khảo.

Bảng 4.33. Kiểm định ảnh hưởng giữa các yếu tố

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1851.784	15	123.452	40.236	.000
Intercept	170280.023	1	170280.023	55498.948	.000
Hang	8.709	5	1.742	.568	.724
Cot	137.573	5	27.515	8.968	.000
Ktbau	755.334	1	755.334	246.184	.000
NPK	712.500	2	356.250	116.112	.000
Ktbau × NPK	237.669	2	118.834	38.731	.000
Error	61.363	20	3.068		
Total	172193.170	36			
Corrected Total	1913.147	35			

Bảng 4.34. Kiểm định phân hạng các nghiệm thức

	NPK	N	Subset	
			1	2
Tukey HSD	1.00	12	62.5250	
	3.00	12		71.2750
	2.00	12		72.5250
	Sig.		1.000	.213

b Alpha = .05.

D. PHÂN TÍCH PHƯƠNG SAI 2 YẾU TỐ: Kiểu thí nghiệm có lô phụ

Giả thiết cần xác định sinh trưởng chiều cao (H, cm) của 3 loài cây (Dầu rái, Sao đen, Vên vên) dưới ảnh hưởng của 3 phương thức lâm sinh. Mỗi thí nghiệm được lặp lại 4 lần. Ở đây chọn lô chính là 3 phương thức lâm sinh, còn lô phụ là 3 loài cây. Hãy phân tích và so sánh kết quả thí nghiệm theo mô hình tuyến tính tổng quát: $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2 +$

$\beta_4 \text{Khối} \times X_1 + \varepsilon$; trong đó X_1 – yếu tố thí nghiệm 1 (phương thức lâm sinh - bố trí ở lô chính), X_2 – yếu tố thí nghiệm 2 (loài cây - bố trí ở lô phụ); $\beta_4 \text{Khối} \times X_1$ – sai số của lô chính; ε - sai số thí nghiệm. Để đạt được mục đích đặt ra, bạn hãy thực hiện các bước sau :

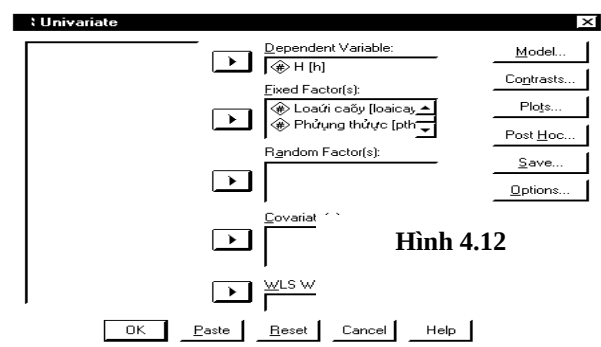
Bước 1. Nhập số liệu vào Bảng tính SPSS như sau (**Bảng 4.35**): cột 1 hoặc hàng 1 nhập biến phụ thuộc (H, cm); cột 2 nhập mã số biểu thị phương thức lâm sinh (lô chính); cột 3 nhập loài cây (lô phụ); cột 4 nhập mã số biểu thị khối (lần lặp). Sau khi nhập xong dữ liệu, bạn nên ghi lại thành 1 file để có thể dùng về sau.

Bảng 4.35. Cách nhập dữ liệu trong phân tích phương sai 2 yếu tố được bố trí theo lô chính - lô phụ
(Chi tiết số liệu xem ở phụ lục 7)

H, cm	Phương thức	Loài cây	Khối
54	1	1	1
58	1	1	2
63	1	1	3
..			
82	1	3	1
83	1	3	2
...			

Bước 2. Phân tích thống kê. Kiểm tra giả thuyết về sự bằng nhau của phương sai mẫu, về sự khác biệt giữa các trung bình mẫu, so sánh sự khác nhau giữa các cặp số trung bình mẫu và kiểm định phân hạng các nghiệm thức được thực hiện theo các thủ tục sau đây:

+ Mở Analyze → General Linear Model → Univariate → Dependent List (đến đây đưa vào biến phụ thuộc H) → Factor (đến đây đưa vào các biến phân loại theo thứ tự: khối, loài cây và phương thức lâm sinh bằng cách chọn 1 hay nhiều biến và nhấn nút mũi tên bên cạnh để chuyển chúng sang ô Fixed factors) (**Hình 4.12**).



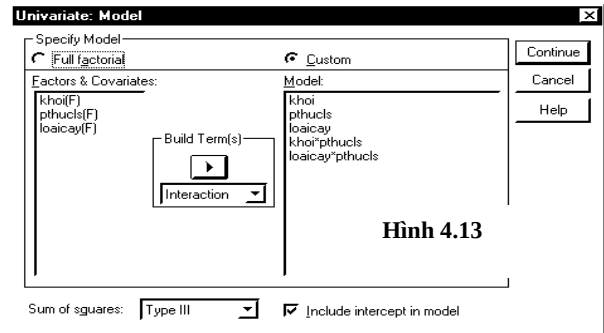
+ Để xác định ảnh hưởng của từng yếu tố nghiên cứu (ảnh hưởng chính – Main Effects) đến biến phụ thuộc, bạn hãy chọn Model. Kế đến tại cửa sổ Univariate Model bạn đánh dấu chọn Custom. Tiếp theo hãy chọn các yếu tố (khối, phương thức lâm sinh, loài cây) trong ô Factors & Covariates; đồng thời chọn Main Effects từ ô Build Terms và nhấn nút mũi tên bên cạnh để chuyển chúng sang ô Model. Để tính sai số cho lô chính, bạn chọn cả hai yếu tố khối và phương thức lâm sinh trong ô Factors & Covariates; kế đến chọn Interactions (tương tác) từ ô Build Terms và nhấn mũi tên bên cạnh để chuyển chúng sang ô Model. Để xác định ảnh hưởng tổng hợp (sự tương tác giữa các yếu tố – Interactions) của các yếu tố nghiên cứu đến biến phụ thuộc, bạn chọn cả hai yếu tố phương thức lâm sinh và loài cây trong ô Factors & Covariates; kế đến chọn Interactions (tương tác) từ ô Build Terms và nhấn mũi tên bên cạnh để chuyển chúng sang ô Model (**Hình 4.13**).

+ Chọn Options. Thủ tục này xác định các thống kê mô tả, kiểm tra sự bằng nhau của phương sai... Để đạt được mục đích đặt ra, bạn hãy chọn tất cả các yếu tố trong ô Factor(s)

and Factor Interactions và chuyển chúng sang ô Display Means for; đồng thời đánh dấu chọn các mục thống kê mô tả.

+ Chọn Post Hoc Test. Thủ tục này so sánh đa phạm vi cho các trung bình mẫu (Post Hoc Multiple Comparisons for Observed Means). Để đạt được mục đích đặt ra, bạn chọn Post Hoc Test, sau đó đánh dấu ô Bonferroni và Tukey's Test. Cuối cùng chọn OK để chương trình tính toán và hiển thị kết quả.

Kết quả hiển thị trên màn Hình bao gồm các Bảng sau đây: Các thống kê mô tả (Descriptive Statistics), kiểm định ngang bằng phương sai (Levene's Test of Equality of Error Variances), kiểm định ảnh hưởng giữa các yếu tố (Test of Between Subjects Effects), ước lượng trung bình nghiệm thức, so sánh từng cặp nghiệm thức (Multiple Comparisons), phân hạng các nghiệm thức (Homogeneous Subsets). Ở đây chỉ trích dẫn Bảng phân tích biến động (Bảng 4.36), Bảng so sánh từng cặp phương thức lâm sinh (Bảng 4.37), Bảng so sánh các loài cây (Bảng 4.38), Bảng kiểm định phân hạng các phương thức lâm sinh (Bảng 4.39) và kiểm định phân hạng sinh trưởng giữa ba loài cây (Bảng 4.40) để các bạn tham khảo.



Bảng 4.36. Kiểm định ảnh hưởng giữa các yếu tố

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3222.306	17	189.547	52.761	.000
Intercept	149640.028	1	149640.028	41652.37	.000
Khối	61.639	3	20.546	5.719	.006
Pthuc	476.722	2	238.361	66.348	.000
Loaicay	997.389	2	498.694	138.812	.000
Khối × Pthuc	229.944	6	38.324	10.668	.000
Pthuc × Loaicay	1456.611	4	364.153	101.362	.000
Error	64.667	18	3.593		
Total	152927.000	36			
Corrected Total	3286.972	35			

a. R Squared = .980 (Adjusted R Squared = .962)

Bảng 4.37. So sánh từng cặp phương thức lâm sinh

(I) Phương thức	(J) Phương thức	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
1	2	1.667×	.774	.045	0.041	3.292
	3	8.417×	.774	.000	6.791	10.042
2	1	-1.667×	.774	.045	-3.292	-0.041
	3	6.750×	.774	.000	5.124	8.376
3	1	-8.417×	.774	.000	-10.042	-6.791
	2	-6.750×	.774	.000	-8.376	-5.124

× The mean difference is significant at the .05 level.

Bảng 4.38. So sánh sự khác nhau về sinh trưởng giữa các loài cây

(I) Loai cay	(J) Loai cay	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Dầu rái (1)	2	-4.250×	.774	.000	-5.876	-2.624
	3	-12.667×	.774	.000	-14.292	-11.041
Sao đen (2)	1	4.250×	.774	.000	2.624	5.876
	3	-8.417×	.774	.000	-10.042	-6.791
Vên vên (3)	1	12.667×	.774	.000	11.041	14.292
	2	8.417×	.774	.000	6.791	10.042

Bảng 4.39. Kiểm định phân hạng các phương thức lâm sinh

	Phương thức	N	Subset	
			1	2
	Tukey HSD			
	3	12	59.42	
	2	12		66.17
	1	12		67.83
	Sig.		1.000	.107

Bảng 4.40. Kiểm định phân hạng sinh trưởng giữa ba loài cây

	Loai cay	N	Subset		
			1	2	3
	Tukey HSD				
	Dầu rái	12	58.83		
	Sao đen	12		63.08	
	Vên vên	12			71.50
	Sig.		1.000	1.000	1.000

E. PHÂN TÍCH PHƯƠNG SAI 3 YẾU TỐ: Kiểu thí nghiệm theo khối ngẫu nhiên

Trong phần này chỉ giới thiệu thí nghiệm 3 yếu tố (A, B và C) được bố trí theo khối ngẫu nhiên. Ở trường hợp này thí nghiệm có 11 nguồn biến động: (1) nguồn biến động của yếu tố “A”, (2) nguồn biến động của yếu tố “B”, (3) nguồn biến động của yếu tố “C”, (4) sai số gây ra bởi yếu tố “A”, (5) sai số gây ra bởi yếu tố “B”, (6) sai số gây ra bởi yếu tố “C”, (7) nguồn biến động số liệu do lặp lại (khối), (8) nguồn biến động gây ra bởi sự tương tác giữa yếu tố “A” với yếu tố “B”, (9) nguồn biến động gây ra bởi sự tương tác giữa yếu tố “A” với yếu tố “C”, (10) nguồn biến động gây ra bởi sự tương tác giữa yếu tố “B” với yếu tố “C”, (11) nguồn biến động gây ra bởi sự tương tác giữa 3 yếu tố “A×B×C”.

Khi sử dụng mô hình tuyến tính tổng quát để phân tích phương sai 3 yếu tố được bố trí theo khối đầy đủ ngẫu nhiên thì bạn có thể xây dựng mô hình tuyến tính tổng quát như sau: $Y = \beta_0 + \beta_1A + \beta_2B + \beta_3C + \beta_4AB + \beta_5AC + \beta_6BC + \beta_7ABC + \beta_8\text{Khối} + \varepsilon$; trong đó A – yếu tố thí nghiệm 1, B – yếu tố thí nghiệm 2; C – yếu tố thí nghiệm 3; AB – sự tương tác giữa hai yếu tố A với B; AC – sự tương tác giữa hai yếu tố A với C, BC – sự tương tác giữa hai yếu tố B với C, ABC – sự tương tác giữa 3 yếu tố (A, B và C); Khối – số lần lặp; ε - sai số thí nghiệm. Nếu gộp yếu tố khối vào thành phần sai số thì mô hình tuyến tính tổng quát có dạng: $Y = \beta_0 + \beta_1A + \beta_2B + \beta_3C + \beta_4AB + \beta_5AC + \beta_6BC + \beta_7ABC + \varepsilon$.

Các giả thuyết. Đối với yếu tố A, $H_0: \alpha_i = 0$ với mọi i; H_1 : ít nhất có 1 $\alpha_i \neq 0$. Đối với yếu tố B, $H_0: \beta_i = 0$ với mọi i; H_1 : ít nhất có 1 $\beta_i \neq 0$. Đối với yếu tố C, $H_0: \gamma_i = 0$; H_1 : ít

nhất có 1 $\gamma_i \neq 0$. Đối với tổ hợp các yếu tố (A, B, C), H_0 : không có tương tác giữa các mức yếu tố.

Thực hành. Giả sử cần thăm dò phản ứng sinh trưởng về D (mm) của Dầu song nài 3 tháng tuổi dưới ảnh hưởng của 3 yếu tố: ánh sáng (3 mức), phân bón (4 mức) và nước tưới (2 mức). Tổng số có 24 nghiệm thức: $A_0B_0C_0, A_0B_0C_1, \dots, A_2B_3C_1$. Thí nghiệm được bố trí lặp lại 2 lần; trong đó các nghiệm thức được phân phối hoàn toàn ngẫu nhiên trên khu thí nghiệm có hoàn cảnh giống nhau. Kết quả thí nghiệm được ghi lại ở **Bảng 4.41**. Phân tích phương sai ba yếu tố bằng chương trình SPSS được thực hiện theo các bước sau đây:

Bước 1. Nhập số liệu vào Bảng tính (xem phụ lục 8). Cột 1 nhập lần lượt biến phụ thuộc (D) cho từng nghiệm thức. Cột 2, 3 và 4 nhập tương ứng mã số biểu thị các yếu tố phân loại (ánh sáng, phân bón và nước tưới). Sau đó bạn thực hiện mã hóa các tên biến và ghi thành 1 file để có thể dùng về sau.

Bước 2. Kiểm định giả thuyết về sự bằng nhau của phương sai mẫu, về sự khác biệt giữa các trung bình mẫu, kiểm định phân hạng các nghiệm thức được thực hiện theo các thủ tục sau đây:

+ Từ Menu chính, mở Analyze $\rightarrow$ General Linear Model $\rightarrow$ Univariable $\rightarrow$ Dependent List (đến đây đưa vào biến phụ thuộc Do) $\rightarrow$ Fixed factor (đưa vào các biến phân loại cho nghiệm thức ánh sáng, phân bón, nước và khối bằng cách chọn 1 hay nhiều biến và nhấn nút mũi tên bên cạnh để chuyển chúng sang ô Fixed factors).

Bảng 4.41. Sinh trưởng D của Dầu song nài 3 tháng tuổi
(Chi tiết số liệu xem ở phụ lục 8)

Nghiệm thức			D, mm	
A ₀	B ₀	C ₀	2.49	2.49
		C ₁	2.55	2.55
	B ₁	C ₀	2.54	2.59
		C ₁	2.63	2.66
	B ₂	C ₀	2.86	2.8
		C ₁	2.77	2.66
	B ₃	C ₀	3.58	3.64
		C ₁	3.74	3.44
A ₁ ...A ₂	B ₀ ...B ₃	C ₀ ...C ₁	...	...

+ Xác định ảnh hưởng của từng yếu tố nghiên cứu (ảnh hưởng chính – Main Effects) và ảnh hưởng tổng hợp (sự tương tác giữa các yếu tố – Interactions) của các yếu tố nghiên cứu đến biến phụ thuộc. Trước hết, bạn hãy chọn Model. Kế đến tại cửa sổ Univariate Model, bạn đánh dấu chọn Custom. Tiếp theo chọn yếu tố ánh sáng, phân bón, nước và số lần lặp (khối) trong ô Factors & Covariates; đồng thời chọn Main Effects từ ô Build Terms và nhấn nút mũi tên bên cạnh để chuyển chúng sang ô Model. Sau đó, bạn lại chọn cả hai yếu tố “ánh sáng và phân bón” trong ô Factors & Covariates; kế đến chọn Interactions (tương tác) từ ô Build Terms và nhấn mũi tên bên cạnh để chuyển chúng sang ô Model. Thao tác này được lặp lại cho các cặp yếu tố “ánh sáng và nước”, “phân bón và nước” và “ánh sáng, phân bón và nước”.

+ Chọn Options. Thủ tục này xác định các thống kê mô tả, so sánh các ảnh hưởng chính và hiển thị các kết quả xử lý thống kê. Để đạt được mục đích đặt ra, bạn chọn các biến từ ô Factor(s) and factor interaction rồi chuyển chúng sang ô Display means for; sau đó chọn tiếp Compare main effects, descriptive statistics và homogeneity tests...

+ Chọn Post Hoc Test. Thủ tục này so sánh đa phạm vi cho các trung bình mẫu (Post Hoc Multiple Comparisons for Observed Means). Để đạt được mục đích đặt ra, bạn chọn Post Hoc Test, sau đó đánh dấu ô Bonferroni và Tukey's Test. Cuối cùng chọn OK để chương trình tính toán và hiển thị kết quả.

Kết quả hiển thị trên màn Hình bao gồm các Bảng sau đây: các thống kê mô tả (Descriptive Statistics), kiểm định ngang bằng phương sai (Levene's Test of Equality of Error Variances), kiểm định ảnh hưởng giữa các yếu tố (Test of Between Subjects Effects), ước lượng trung bình nghiệm thức và trung bình khối, so sánh từng cặp nghiệm thức (Multiple Comparisons), phân hạng các nghiệm thức (Homogeneous Subsets)...Dưới đây chỉ trích dẫn Bảng phân tích biến động (Bảng 4.42), Bảng so sánh từng cặp nghiệm thức (Bảng 4.43) và Bảng phân hạng các nghiệm thức (Bảng 4.44) để các bạn tham khảo.

Bảng 4.42. Kiểm định ảnh hưởng của các yếu tố ánh sáng, phân bón và nước đến Do của Dầu song nòng

Source	Type III Sum of Squares	df	Mean Square	F	Sig. ^a	Noncent. Parameter	Observed Power
Corrected Model	13.911b	24	.580	138.868	.000	3332.840	1.000
Intercept	519.7540	1	519.7540	124527.151	0.000	124527.151	1.000
KHOI	0.0029	1	0.0029	0.683	0.417	0.683	0.124
AS	7.7090	2	3.8550	923.548	0.000	1847.096	1.000
PHAN	4.9550	3	1.6520	395.689	0.000	1187.067	1.000
NUOC	0.0776	1	0.0776	18.593	0.000	18.593	0.985
AS × PHAN	1.0270	6	0.1710	41.014	0.000	246.081	1.000
AS × NUOC	0.0382	2	0.0191	4.578	0.021	9.156	0.719
PHAN × NUOC	0.0246	3	0.0082	1.964	0.148	5.891	0.439
AS × PHAN × NUOC	0.0763	6	0.0127	3.045	0.024	18.273	0.814
Error	0.0960	23	0.0042				
Total	533.7610	48					
Corrected Total	14.0070	47					

a. Computed using alpha = .05; b. R Squared = .993 (Adjusted R Squared = .986)

Bảng 4.43. So sánh từng cặp nghiệm thức (trên ví dụ yếu tố ánh sáng)

		Mean Difference (I-J)	Std. Error	Sig. ^a	95% Confidence Interval for Difference ^a	
(I) Ánh sáng	(J) Ánh sáng				Lower Bound	Upper Bound
1	2	-.291×	.023	.000	-.339	-.244
	3	-.958×	.023	.000	-1.005	-.910
2	1	.291×	.023	.000	.244	.339
	3	-.666×	.023	.000	-.714	-.619
3	1	.958×	.023	.000	.910	1.005
	2	.666×	.023	.000	.619	.714

× The mean difference is significant at the .05 level.

Giải thích và nhận xét kết quả. Bảng 4.42 báo cáo kết quả phân tích biến động. Khi $P - \text{value} < 0.05$ thì các yếu tố nghiên cứu (biến lập nhóm hay biến độc lập Xi) có ảnh hưởng rõ rệt đến kết quả thí nghiệm (biến phụ thuộc Y). Ngược lại, khi $P - \text{value} > 0.05$ thì các yếu tố nghiên cứu có ảnh hưởng không rõ rệt đến kết quả thí nghiệm. Ở ví dụ này, từng yếu tố thí nghiệm có ảnh hưởng rõ rệt đến sinh trưởng D của Dầu song nòng. Ngoài ra, trừ tổ hợp phân bón và nước, các tổ hợp nhân tố sinh thái khác (ánh sáng×nước, phân bón×ánh sáng, ánh sáng×nước×phân bón) có ảnh hưởng rõ rệt đến sinh trưởng D của Dầu song nòng.

Bảng 4.44. Phân hạng các nghiệm thức theo yếu tố ánh sáng

Tukey HSD ^{a, b}	Ánh sáng	N	Subset		
			1	2	3
	1	16	2.8744		
	2	16		3.1656	
	3	16			3.8319
	Sig.		1.000	1.000	1.000

a Uses Harmonic Mean Sample Size = 16.000; b Alpha = .05.

F. PHÂN TÍCH PHƯƠNG SAI NHIỀU YẾU TỐ: Mô hình hồi quy với nhiều biến độc lập

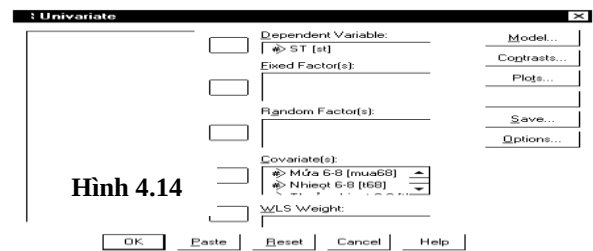
Giả sử cần xác định ảnh hưởng của lượng mưa, nhiệt độ và chế độ thủy nhiệt của tháng 6-8 (tương ứng kí hiệu M(6-8), T(6-8) và I(6-8)) đến sinh trưởng của cây rừng (ST). Số liệu về khí tượng và sinh trưởng của cây rừng được thu thập trong 39 năm liên tiếp. Từ số liệu này, tính các chỉ số sinh trưởng (Bảng 4.45) và sử dụng thủ tục GLM: Univariate để kiểm định theo mô hình thống kê sau đây: $ST = a + bM(6-8) + cT(6-8) + dI(6-8) + \varepsilon$. Để nhận được kết quả, bạn có thể thực hiện theo các thủ tục sau đây:

Bảng 4.45. Các chỉ số sinh trưởng của cây rừng và của yếu tố khí tượng

ST	M(68)	T(68)	TN(68)
1.02	0.71	1.08	0.64
1.02	0.95	0.96	0.98
1.00	1.27	1.04	1.23
1.25	0.81	1.01	0.8
0.91	1.12	0.89	1.24
0.98	0.99	1.09	0.9
1.04	0.95	0.96	0.99
1.00	1.12	1.00	1.12
0.98	0.77	1.08	0.7

Bước 1. Nhập dữ liệu vào Bảng tính. Tại cửa sổ Data Editor, bạn nhập biến ST (biến phụ thuộc) vào cột 1; nhập ba biến độc lập (M6-8, T6-8 và I6-8) tương ứng vào cột 2, 3 và 4. Sau đó mở trang Variable View để mã hóa tên cho các biến.

Bước 2. Gọi chương trình tính toán. Từ thanh Menu chính, bạn hãy chọn Analyze $\Rightarrow$ General Linear Model $\Rightarrow$ Univariate. Khi cửa sổ Univariate xuất hiện (Hình 4.14), bạn hãy thực hiện như sau:



Hình 4.14

+ Chọn biến ST và chuyển nó sang ô Dependent Variable;

+ Chọn toàn bộ các biến độc lập và chuyển chúng vào ô Covariate;

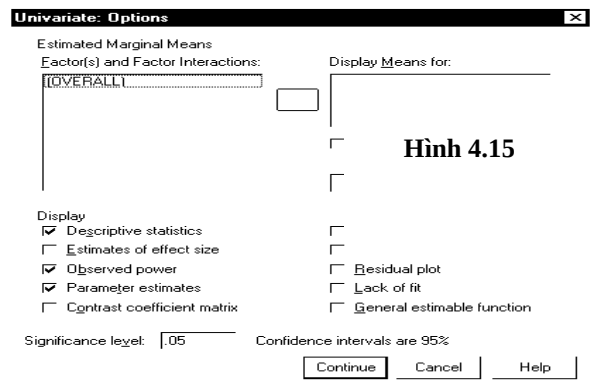
+ Chọn Model và đánh dấu ô Custom. Kế đó chọn tất cả các biến độc lập từ ô Factors & Covariate, rồi chọn tiếp Main Effect trong Build Term(s) và chuyển chúng sang ô Custom; chọn Continue;

+ Chọn Option và đánh dấu các mục Descriptive statistics, Observed power, Parameter estimates, Residual plot..., rồi chọn Continue (Hình 4.15);

+ Sau cùng chọn OK để chương trình tính toán và hiển thị kết quả.

Giải thích và nhận xét kết quả.

Bảng 4.46, 4.47 và 4.48 báo cáo kết quả tương ứng về kiểm định ảnh hưởng giữa các biến độc lập đến biến phụ thuộc và ước lượng các tham số của mô hình hồi quy. Trong **Bảng 4.47**, mỗi hàng chứa một thành phần của mô hình gây ra những biến thiên của biến phụ thuộc. Nội dung của **Bảng 4.47** như sau:



+ Hàng có nhãn Corrected Model cho biết tổng biến thiên gây ra do mô hình mà bạn đã chọn dưới dạng: $ST = a + bM(6-8) + cT(6-8) + dI(6-8) + \varepsilon$. Giả thuyết H_0 : Mô hình $ST = a + bM(6-8) + cT(6-8) + dI(6-8) + \varepsilon$ tồn tại. Nếu $P < \alpha$ (0.05 hay 0.01) thì giả thuyết H_0 bị bác bỏ. Ở ví dụ này, vì $P = 0.000 < 0.01$ nên mô hình $ST = a + bM(6-8) + cT(6-8) + dI(6-8) + \varepsilon$ không tồn tại.

+ Các thành phần $M(6-8)$, $T(6-8)$ và $I(6-8)$ biểu thị những ảnh hưởng chính trong mô hình. Kết quả nhận được là do bạn đã chọn Main effects từ ô Build terms. Giả thuyết H_0 : Từng yếu tố không có ảnh hưởng đến biến phụ thuộc. Nếu $P < \alpha$ (0.05 hay 0.01) thì giả thuyết H_0 bị bác bỏ. Ở ví dụ này chỉ có yếu tố $T(6-8)$ có ảnh hưởng rõ rệt đến sinh trưởng của rừng ($P = 0.041$).

+ Hàng Error biểu thị cho những sai số do mô hình không thể giải thích được.

+ Hàng Total là tổng bình phương của các trị quan sát của biến phụ thuộc.

+ Hàng Corrected Total là tổng bình phương các độ lệch giữa giá trị Y_i so với giá trị trung bình của biến phụ thuộc. Trị số này bằng Corrected model cộng với Error.

+ Mỗi giá trị của cột $F = (\text{Mean Square/Error})$ tương ứng.

+ Cột có nhãn Sig (hay $P - \text{value}$) cho biết mức ý nghĩa của các kiểm định. Giả thuyết H_0 : từng thành phần của mô hình bằng zero, hoặc các biến độc lập không ảnh hưởng rõ rệt đến biến phụ thuộc. Khi $P < \alpha$ (0.05 hay 0.01) thì giả thuyết H_0 bị bác bỏ. Ở ví dụ này chỉ có biến $T(6-8)$ có ảnh hưởng rõ rệt đến biến ST .

Bảng 4.46. Thống kê mô tả (Descriptive Statistics)

Mean	Std. Deviation	N
1.0044	.1212	39

Bảng 4.47. Kiểm định ảnh hưởng giữa các yếu tố

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power
Corrected Model	0.2710	3	0.0904	11.034	.000	33.101	.998
Intercept	0.0663	1	0.0663	8.092	.007	8.092	.790
M(6-8)	0.0206	1	0.0206	2.518	.122	2.518	.339
T(6-8)	0.0370	1	0.0370	4.513	.041	4.513	.542
I(6-8)	0.0183	1	0.0183	2.237	.144	2.237	.307
Error	0.2870	35	0.0082				
Total	39.899	39					
Corrected Total	0.5580	38					

a. Computed using alpha = .05; b. R Squared = .486 (Adjusted R Squared = .442)

Bảng 4.48. Ước lượng các tham số (Parameter Estimates)

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval		Noncent. parameter	Observed Power
					Lower Bound	Upper Bound		
Intercept	3.374	1.186	2.845	.007	.966	5.782	2.845	.790
Mưa 6-8	2.102	1.325	1.587	.122	-.587	4.791	1.587	.339
Nhiệt độ 6-8	-2.509	1.181	-2.124	.041	-4.907	-.111	2.124	.542
Thủynhiệt	-1.957	1.308	-1.496	.144	-4.613	.699	1.496	.307

a Computed using alpha = .05

+ Cột có nhãn Noncenter Parameter (tham số phi trung tâm) dùng để tính độ mạnh của phép kiểm định được quan sát (Observed Power). Nếu ta sử dụng lặp lại một kiểm định với mức ý nghĩa α cố định, và khi giả thuyết H_0 đúng, thì ta sẽ mắc một sai $\alpha\%$ lần và đúng $100 - \alpha\%$ lần. Vì thế chúng ta cần quan tâm đến độ mạnh của phép kiểm định để phát hiện H_0 là sai. Độ mạnh của phép kiểm định đo xác suất bác bỏ giả thuyết H_0 nếu H_0 đúng. Xác suất này càng cao thì phép kiểm định càng hiệu quả. Ở ví dụ này, độ mạnh của phép kiểm định được quan sát rất nhỏ (0.339, 0.542 và 0.307). Kết quả ấy cho biết khi sử dụng cùng một cỡ mẫu thì khả năng phát hiện những ảnh hưởng của các thành phần ở tổng thể thông qua mẫu là rất nhỏ. Nói khác đi, mô hình giả định trên đây là kém chính xác.

Phân tích số liệu của **Bảng 4.47 và 4.48** cho thấy chỉ có biến T(6-8) có ảnh hưởng rõ rệt đến biến ST. Nếu ta xây dựng lại mô hình chỉ bao gồm biến T(6-8) thì kết quả kiểm định ảnh hưởng của T(6-8) đến biến ST được **dẫn ra ở Bảng 4.49 và 4.50**. Từ số liệu của **Bảng 4.49**, ta có mô hình để dự đoán ST thông qua T(6-8) như sau: $ST = 2.150 - 1.140 \times T(6-8)$.

Bảng 4.49. Kiểm định ảnh hưởng giữa các yếu tố

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power
Corr. Model	.230	1	.230	25.924	.000	25.924	.999
Intercept	.806	1	.806	90.860	.000	90.860	1.000
T ₆₈	.230	1	.230	25.924	.000	25.924	.999
Error	.328	37	0.0089				
Total	39.899	39					
Corrected Total	.558	38					

a Computed using alpha = .05

b R Squared = .412 (Adjusted R Squared = .396)

Bảng 4.50. Ước lượng các tham số (Parameter Estimates)

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval		Noncent. Parameter	Observed Power
					Lower Bound	Upper Bound		
Intercept	2.150	.226	9.532	.000	1.693	2.608	9.532	1.000
T ₆₈	-1.145	.225	-5.092	.000	-1.601	-.689	5.092	.999

a Computed using alpha = .05

4.4. PHÂN TÍCH ẢNH HƯỞNG CỦA MỘT HOẶC NHIỀU YẾU TỐ ĐẾN NHIỀU BIẾN PHỤ THUỘC (MANOVA)

4.4.1. Khái quát

Ở các mục 4.2 và 4.3, chúng ta đã phân tích ảnh hưởng của các yếu tố độc lập (X_i) đến một biến phụ thuộc bằng thủ tục Univariate (Một biến Y). Phần này chúng ta phân tích ảnh hưởng của các yếu tố độc lập (X_i) đến nhiều biến phụ thuộc (Y_i) bằng thủ tục Multivariate. Ví dụ: Phân tích ảnh hưởng của các yếu tố đậm và lân đến sinh trưởng về đường kính và chiều cao thân cây. Bạn có thể sử dụng phân tích phương sai đa biến số (MANOVA = Multivariate Analysis of Variance) để phân tích phương sai một hoặc nhiều yếu tố theo mô hình tuyến tính tổng quát.

Để vào thủ tục MANOVA, từ Menu chính bạn thực hiện như sau: (1) Chọn Analyze $\Rightarrow$ General Linear Models $\Rightarrow$ Multivariate. Khi đó hộp thoại phân tích mô hình tuyến tính tổng quát xuất hiện (General Linear Models Analysis). (2) Hoàn thành hộp thoại và chọn OK. Hộp thoại chuyên về mô hình tuyến tính xuất hiện. (3) Hoàn thành hộp thoại và chọn OK. Bảng tóm tắt phân tích MANOVA xuất hiện.

Phân tích MANOVA theo mô hình tuyến tính tổng quát cũng bao gồm những thủ tục tương tự như phân tích ANOVA. Để vào một thủ tục, bạn dùng nút chuột chọn nó. Sau khi hoàn thành một thủ tục, hãy chọn Continue để chuyển sang hộp thoại khác.

4.4.2. THỰC HÀNH PHÂN TÍCH MANOVA

Thủ tục phân tích MANOVA được thực hiện tương tự như phân tích ANOVA. Điểm khác biệt cơ bản giữa Bảng MANOVA với Bảng ANOVA là ở chỗ Bảng MANOVA là sự kết hợp của nhiều Bảng ANOVA. Dưới đây dẫn ví dụ phân tích phương sai 1 yếu tố và 2 yếu tố được thiết kế theo khối đầy đủ ngẫu nhiên.

A. PHÂN TÍCH MANOVA 1 YẾU TỐ THEO KHỐI ĐẦY ĐỦ NGẪU NHIÊN

Giả sử cần xác định ảnh hưởng của 3 mức đậm khác nhau đến sinh trưởng về đường kính (D , mm) và chiều cao (H , cm) của cây con Dầu song nòng 3 tháng tuổi trong giai đoạn vườn ươm. Thí nghiệm được bố trí theo kiểu khối đầy đủ ngẫu nhiên với 3 nghiệm thức, mỗi nghiệm thức lặp lại 5 lần (mỗi lần là 1 khối). Hãy phân tích ảnh hưởng của các yếu tố thí nghiệm đến biến phụ thuộc theo mô hình tổng quát: $Y_i = \beta_0 + \beta_1 X_1 + \beta_2 \text{Khối} + \varepsilon$; trong đó Y_i – yếu tố phụ thuộc thứ i , X_1 – yếu tố thí nghiệm, ε - sai số ngẫu nhiên. Để đạt mục đích đặt ra, trong thực hành bạn thực hiện như sau:

Bảng 4.51. Cách nhập số liệu trong phân tích phương sai 1 yếu tố bố trí theo khối

Thứ tự	D (mm)	H (cm)	Nghiệm thức	Khối (số lần lặp)
1	12	10.5	1	1
2	13	11.5	1	2
3	12.5	11	1	3
4	14	12.5	1	4
5...	21	19.5	1...	5...

Bước 1. Cách vào số liệu để phân tích phương sai 1 yếu tố được bố trí theo khối (Bảng 4.51). Số liệu được nhập vào Bảng tính SPSS theo 5 cột. Cột 1 nhập thứ tự đại lượng quan sát (có hay không cũng được). Cột 2 và 3 nhập tương ứng hai biến D (mm) và H (cm). Cột 4 nhập mã biến độc lập (biến lập nhóm theo yếu tố thí nghiệm). Cột 5 nhập thứ tự số lần lặp lại của mỗi nghiệm thức (biến lập nhóm theo số lần lặp).

Bước 2. Kiểm định giả thuyết về sự bằng nhau của phương sai mẫu, về sự khác biệt giữa các trung bình mẫu, so sánh sự khác nhau giữa các cặp số trung bình mẫu và kiểm định phân hạng các nghiệm thức được thực hiện theo các thủ tục sau:

+ Mở Analyze → General Linear Model → Multivariable → Dependent List (đến đây đưa vào các biến phụ thuộc) → Fixed factor (đưa vào các biến nghiệm thức và khối) bằng cách chọn 1 hay nhiều biến và nhấn nút mũi tên bên cạnh (Hình 4.16).

+ Xác định ảnh hưởng của yếu tố nghiên cứu (ảnh hưởng chính – Main Effects) và ảnh hưởng của khối. Trước hết, bạn chọn Model. Kế đến tại cửa sổ Multivariate Model, bạn chọn yếu tố nghiệm thức và khối trong ô Factors & Covariates. Sau đó chọn Main Effects từ ô Build Terms và nhấn nút mũi tên bên cạnh để chuyển nghiệm thức sang ô Model.

+ Chọn Options. Thủ tục này tính các thống kê mô tả, so sánh ảnh hưởng chính, kiểm định thuần nhất các phương sai mẫu... Tại đây, bạn cũng có thể thay đổi mức ý nghĩa (Significance Level) mà bạn muốn sử dụng cho các kiểm định. Mức ý nghĩa mặc định là 0.05.

+ Chọn Post Hoc Test. Thủ tục này so sánh đa phạm vi cho các trung bình mẫu (Post Hoc Multiple Comparisons for Observed Means). Để đạt được mục đích đặt ra, bạn chọn Post Hoc Test, sau đó đánh dấu ô Bonferroni và Tukey's Test. Cuối cùng chọn OK để chương trình tính toán và hiển thị kết quả.

Kết quả hiển thị trên màn Hình bao gồm những thống kê sau đây: thống kê mô tả (Descriptive Statistics), kiểm định ngang bằng phương sai (Levene's Test of Equality of Error Variances), kiểm định ảnh hưởng giữa các yếu tố (Test of Between Subjects Effects), ước lượng trung bình nghiệm thức và trung bình khối, so sánh từng cặp đối với các nghiệm thức (Multiple Comparisons), kiểm định phân hạng các nghiệm thức (Homogeneous Subsets). Dưới đây chỉ trích dẫn kết quả của Bảng MANOVA để các bạn tham khảo (Bảng 4.52).

Giải thích kết quả của Bảng MANOVA. Trước hết, bạn có thể nhận thấy cấu trúc của Bảng MANOVA cũng tương tự như Bảng ANOVA. Điểm khác biệt giữa Bảng MANOVA với Bảng ANOVA là ở chỗ Bảng MANOVA là sự kết hợp của nhiều Bảng ANOVA. Vì thế, cách giải thích các nội dung của Bảng MANOVA cũng tương tự như Bảng ANOVA.

Bảng 4.52. Kiểm định ảnh hưởng của đạm đến sinh trưởng của Dầu song nà

Source	Dep. Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	D	203.560	6	33.927	3.585	.007
	H	205.713	6	34.286	3.606	.006
Intercept	D	10475.990	1	10475.990	1106.939	.000
	H	8926.721	1	8926.721	938.808	.000
Đạm	D	199.872	4	49.968	5.280	.002
	H	200.964	4	50.241	5.284	.002

Khối	D	3.688	2	1.844	.195	.824
	H	4.750	2	2.375	.250	.780
Error	D	359.629	38	9.464		
	H	361.326	38	9.509		
Total	D	11039.180	45			
	H	9493.760	45			
Corrected Total	D	563.190	44			
	H	567.039	44			

a R Squared = .361 (Adjusted R Squared = .261); b R Squared = .363 (Adjusted R Squared = .262)

B. PHÂN TÍCH MANOVA HAI YẾU TỐ TỔNG QUÁT: Kiểu thí nghiệm theo khối hoàn toàn ngẫu nhiên

Giả sử cần phải phân tích ảnh hưởng của 3 mức đạm (kí hiệu yếu tố A: 1, 2 và 3% đạm) và 3 mức lân (kí hiệu yếu tố B: 0.5, 1.0, 1.5%) đến sinh trưởng chiều cao (H, cm) và D (mm) của cây con Dầu song nòng 3 tháng tuổi trong giai đoạn vườn ươm. Thí nghiệm bao gồm 9 nghiệm thức, mỗi nghiệm thức được lặp lại 5 lần; tổng cộng có 45 lô thí nghiệm. Mỗi lô thí nghiệm sử dụng 49 cây đồng đều về Hình thái và sức sống. Hãy phân tích và so sánh kết quả thí nghiệm theo mô hình tuyến tính tổng quát: $Y_i = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 AB + \beta_5 \text{Khối} + \varepsilon$; trong đó Y_i – biến phụ thuộc thứ i , A – yếu tố thí nghiệm 1, B – yếu tố thí nghiệm 2; Khối – số lần lặp; ε - sai số thí nghiệm.

Các giả thuyết: (1) Các mức đạm khác nhau (kí hiệu A) có ảnh hưởng đồng đều đến sinh trưởng H và D của Dầu song nòng; (2) Các mức lân khác nhau (kí hiệu B) có ảnh hưởng đồng đều đến sinh trưởng H và D của Dầu song nòng; (3) Sự phối hợp của các mức đạm và lân khác nhau cũng có ảnh hưởng đồng đều đến sinh trưởng H và D của Dầu song nòng. Để kiểm định các giả thuyết trên, bạn hãy thực hiện các bước sau:

Bước 1. Nhập số liệu vào Bảng tính SPSS như sau (**Bảng 4.53**): cột 1 và 2 nhập 2 biến phụ thuộc (H, cm và D, mm); cột 3 nhập số lần lặp của mỗi nghiệm thức (số khối); cột 4 nhập mã số biểu thị cho yếu tố đạm (biến phân loại A); cột 5 nhập mã số biểu thị cho yếu tố lân (biến phân loại B). Sau khi nhập xong dữ liệu, bạn nên ghi lại thành 1 file để có thể dùng về sau.

Bước 2. Kiểm tra giả thuyết về sự bằng nhau của phương sai mẫu, về sự khác biệt giữa các trung bình mẫu, so sánh sự khác nhau giữa các cặp số trung bình mẫu và kiểm định phân hạng các nghiệm thức được thực hiện theo các thủ tục sau:

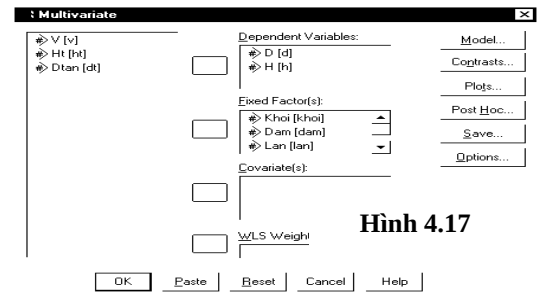
+ Mở Analyze → General Linear Model → Multivariable → Dependent List (đến đây đưa vào 2 biến phụ thuộc H và D) → Fixed factor (đưa vào các biến phân loại cho nghiệm thức đạm, lân và khối bằng cách chọn 1 hay nhiều biến và nhấn nút mũi tên bên cạnh để chuyển chúng sang ô Fixed factors (**Hình 4.17**)).

Bảng 4.53. Cách nhập dữ liệu trong phân tích phương sai 2 yếu tố được bố trí ngẫu nhiên

D, mm	H, cm	TT. khối	Đạm (A)	Lân (B)
12	10.7	1	1	1
13	11.8	2	1	1
12.5	11.2	3	1	1
14	12.7	4	1	1
21	19.7	5	1	1

10	8.7	1	1	2
16	14.7	2	1	2
18	16.8	3	1	2
...	...	...	...	...

+ Xác định ảnh hưởng của từng yếu tố nghiên cứu (ảnh hưởng chính – Main Effects) và ảnh hưởng tổng hợp (sự tương tác giữa các yếu tố – Interactions) của các yếu tố nghiên cứu đến biến phụ thuộc. Trước hết, bạn hãy chọn Model. Kế đến tại cửa sổ Multivariate Model bạn đánh dấu chọn Custom. Tiếp theo chọn yếu tố đậm, lân và số lần lặp (khối) trong ô Factors & Covariates; kế đến chọn Main Effects từ ô Build Terms và nhấn nút mũi tên bên cạnh để chuyển chúng sang ô Model. Sau đó, bạn lại chọn cả hai yếu tố đậm và lân trong ô Factors & Covariates; kế đến chọn Interactions (tương tác) từ ô Build Terms và nhấn mũi tên bên cạnh để chuyển chúng sang ô Model.



Hình 4.17

+ Chọn Options. Khi cửa sổ Option xuất hiện, bạn hãy chọn các thống kê mô tả, kiểm tra sự bằng nhau của phương sai và những tham số cần hiển thị.

+ Chọn Post Hoc Test. Thủ tục này so sánh đa phạm vi cho các trung bình mẫu (Post Hoc Multiple Comparisons for Observed Means). Để đạt được mục đích này, bạn chọn Post Hoc Test, sau đó đánh dấu ô Bonferroni và Tukey's Test. Cuối cùng chọn OK để chương trình tính toán và hiển thị kết quả.

Kết quả hiển thị trên màn Hình bao gồm các Bảng sau đây: các thống kê mô tả (Descriptive Statistics), kiểm định ngang bằng phương sai (Levene's Test of Equality of Error Variances), kiểm định ảnh hưởng giữa các yếu tố (Test of Between Subjects Effects), ước lượng trung bình nghiệm thức và trung bình khối, so sánh từng cặp (Multiple Comparisons) đối với các nghiệm thức và khối, phân hạng các nghiệm thức (Homogeneous Subsets). Ở đây chỉ trích dẫn Bảng phân tích biến động (Bảng 4.54) và Bảng so sánh từng cặp nghiệm thức (Bảng 4.55) để các bạn tham khảo.

Bảng 4.54. Kiểm định ảnh hưởng của yếu tố đậm và lân đến sinh trưởng D và H của Dầu song nòng 3 tháng tuổi

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	D	224.918a	12	18.743	1.773	.097
	H	227.967b	12	18.997	1.793	.093
Intercept	D	10475.990	1	10475.990	991.012	.000
	H	8926.721	1	8926.721	842.460	.000
Khối	D	199.872	4	49.968	4.727	.004
	H	200.964	4	50.241	4.741	.004
Đạm	D	3.688	2	1.844	.174	.841
	H	4.750	2	2.375	.224	.800
Lân	D	12.428	2	6.214	.588	.561
	H	13.350	2	6.675	.630	.539
Đạm × Lân	D	8.929	4	2.232	.211	.930
	H	8.904	4	2.226	.210	.931

Error	D	338.272	32	10.571		
	H	339.072	32	10.596		
Total	D	11039.180	45			
	H	9493.760	45			
Corrected Total	D	563.190	44			
	H	567.039	44			

a. R Squared = .399 (Adjusted R Squared = .174)

4.5. PHÂN TÍCH HIỆP PHƯƠNG SAI

Phân tích hiệp phương sai (Analysis of covariance = ANOCOVA) được áp dụng cho những đối tượng không đồng nhất, nhưng chúng vẫn ảnh hưởng đến kết quả thí nghiệm. Ví dụ: Khi phân tích ảnh hưởng của độ tàn che đến sinh trưởng của Dầu song nòng, nhưng cây con đưa vào thí nghiệm lại có tuổi khác nhau hoặc kích thước khác nhau. Dưới đây chúng ta xem xét một số mô hình phân tích hiệp phương sai thường được áp dụng.

Bảng 4.55. So sánh sự khác biệt giữa các mức đậm đến sinh trưởng D của Dầu song nòng 3 tháng tuổi

			Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
Tukey HSD	(I) Dam	(J) Dam				Lower Bound	Upper Bound
	1	2	.313	1.187	.962	-2.604	3.231
		3	-.387	1.187	.943	-3.304	2.531
	2	1	-.313	1.187	.962	-3.231	2.604
		3	-.700	1.187	.827	-3.617	2.217
	3	1	.387	1.187	.943	-2.531	3.304
		2	.700	1.187	.827	-2.217	3.617

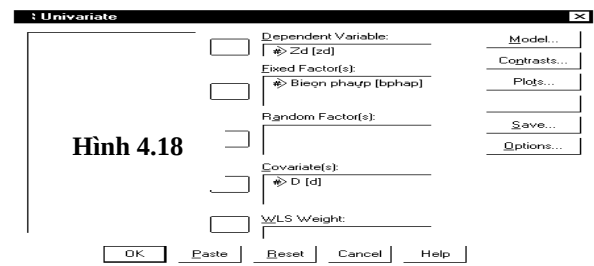
4.5.1. PHÂN TÍCH HIỆP PHƯƠNG SAI 1 YẾU TỐ

A. PHÂN TÍCH HIỆP PHƯƠNG SAI 1 YẾU TỐ: Kiểu thiết kế hoàn toàn ngẫu nhiên

1. Trường hợp có hai nghiệm thức

Giả sử nghiên cứu phản ứng tăng trưởng đường kính (ZD, mm) của cây con Dầu song nòng dưới ảnh hưởng của hai biện pháp lâm sinh khác nhau. Biện pháp thứ nhất (A) được thực hiện trên 11 cây, còn biện pháp thứ hai (B) được thực hiện trên 19 cây. Những cây con sử dụng trong thí nghiệm có kích thước đường kính ban đầu khác nhau (D, cm). Thí nghiệm được đánh giá sau 3 tháng.

Số liệu về D ban đầu và lượng tăng trưởng D (Zd, mm) của Dầu song nòng sau 3 tháng thí nghiệm được ghi lại ở **Bảng 4.56**. Để giải bài toán này bằng mô hình tuyến tính tổng quát, bạn hãy thực hiện theo các bước sau đây:



Hình 4.18

Bước 1. Nhập dữ liệu vào Bảng tính. Tại cửa sổ Data Editor, bạn nhập lần lượt tương ứng biến D và Zd của biện pháp A và B vào cột 1 và 2; cột 3 nhập mã số biểu thị cho hai biện pháp lâm sinh. Sau đó mở trang Variable View để mã hóa tên cho các biến.

Bước 2. Gọi chương trình tính toán. Từ thanh Menu chính, bạn hãy chọn Analyze $\Rightarrow$ General Linear Model $\Rightarrow$ Univariate. Khi cửa sổ Univariate xuất hiện (Hình 4.18), bạn thực hiện như sau :

+ Chọn biến Zd và chuyển nó sang ô Dependent Variable;

+ Chọn biến D và chuyển nó vào ô Covariate, kế đến chọn biến biện pháp lâm sinh và chuyển nó sang ô Fixed factor(s).

+ Chọn Model và đánh dấu ô Custom. Kế đến chọn tất cả các biến (Biện pháp lâm sinh và D) từ ô Factors & Covariate. Tiếp đến chọn Main Effect trong Build Term(s) và chuyển Biện pháp lâm sinh và D sang ô Custom. Sau đó chọn Continue.

+ Chọn Option và đánh dấu các mục Descriptive statistics, Observed power, Parameter estimates, Residual plot...(Hình 4.19). Sau đó chọn Continue.

+ Cuối cùng chọn OK để chương trình tính toán và hiển thị kết quả.

Bảng 4.56. Đường kính (D, cm) và lượng tăng Zd (mm) của Dầu song nòng sau hai tháng (Chi tiết số liệu xem ở phụ lục 9)

TT	D, cm	Zd, cm	Nghiệm	TT	D, cm	Zd, cm	Nghiệm
1	4.6	1.8	1	11	5.8	1.9	1
2	5.2	2.3	1	12	1.8	1.4	2
...	...	...	...	...	...	...	...
9	7.1	2.2	1	19	5.8	2.6	2
10	4.1	1.1	1	20	6.3	3.4	2

Giải thích và nhận xét kết quả. Bảng 4.57 và 4.58 báo cáo kết quả phân tích các đặc trưng thống kê mô tả và kiểm định ngang bằng phương sai mẫu. Từ Bảng 4.58 cho thấy, $P = 0.309 > 0.05$, nên phương sai hai mẫu là bằng nhau. Bảng 4.59 dẫn ra kết quả phân tích ảnh hưởng của các biện pháp lâm sinh và kích thước cây con ban đầu đến tăng trưởng Zd. Giả thuyết H_0 : Các biện pháp lâm sinh ảnh hưởng đồng đều đến tăng trưởng Zd. Vì $P = 0.086 > 0.05$, nên các biện pháp lâm sinh ảnh hưởng đồng đều đến Zd của các cây con sau 3 tháng tuổi.

Bảng 4.57. Thống kê mô tả (Descriptive Statistics)
Dependent Variable: Zd

Biện pháp	Mean	Std. Deviation	N
1	2.073	.647	11
2	2.179	.593	19
Total	2.140	.604	30

Bảng 4.58. Kiểm định ngang bằng phương sai

F	df1	df2	Sig.
1.076	1	28	.309

a. Design: Intercept + BIENPHAP + D

Bảng 4.59. Kiểm định ảnh hưởng giữa các yếu tố

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power
Corrected Model	5.510	2	2.755	14.640	.000	29.279	.997
Intercept	1.671	1	1.671	8.878	.006	8.878	.819
BIENPHAP	.598	1	.598	3.175	.086	3.175	.405
D	5.432	1	5.432	28.862	.000	28.862	.999
Error	5.082	27	.188				
Total	147.980	30					
Corrected Total	10.592	29					

a. Computed using alpha = .05; b. R Squared = .520 (Adjusted R Squared = .485)

Bảng 4.60. Các tham số ước lượng (Parameter Estimates)

	B	Std. Error	t	Sig.	95% Confidence Interval		Noncent. Parameter	Observed Power
					Lower Bound	Upper Bound		
Parameter Intercept	.934	.252	3.701	.001	.416	1.451	3.701	.946
[BPHAP=1]	-.300	.168	-1.782	.086	-.645	4.541E-02	1.782	.405
[BPHAP=2]	0	.	.	.	.	.	.	.
D	.271	.050	5.372	.000	.168	.375	5.372	.999

Lưu ý rằng, khi các nghiệm thức khác nhau rõ rệt, thì bước tiếp theo cần thực hiện so sánh các giá trị trung bình của các nghiệm thức. Vì thí nghiệm được thực hiện trên những cây con không đồng nhất về kích thước, nên trước khi so sánh các trung bình nghiệm thức, chúng ta cần phải điều chỉnh các giá trị trung bình của các nghiệm thức. Các giá trị trung bình đã được điều chỉnh giữa các nghiệm thức được gọi là các trung bình điều chỉnh. Để đạt được mục đích ấy, trước hết chúng ta phải đồng nhất những giá trị đường kính thân cây ban đầu của những cây con đưa vào nghiên cứu. Sau đó tính những giá trị trung bình điều chỉnh (Z_{dc}) - đó là những giá trị trung bình được tính bằng cách san phẳng sự khác biệt giữa các giá trị của đường kính thân cây con ban đầu.

Bảng 4.61. Các ước lượng đã được điều chỉnh (Estimates)

Biện pháp	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	1.950	.133	1.678	2.223
2	2.250	.100	2.044	2.456

a. Evaluated at covariates appeared in the model: D = 4.857.

Bảng 4.62. So sánh các nghiệm thức (Pairwise Comparisons)

Dependent Variable: Zd

		Mean Difference (I-J)	Std. Error	Sig.	95% Conf. Interval for Difference	
(I) Biện pháp	(J) Biện pháp				Lower Bound	Upper Bound
1	2	-.300	.168	.086	-0.645	0.045
2	1	.300	.168	.086	-0.045	0.645

Based on estimated marginal means

Để xác định được Zd_{dc} , trước hết chúng ta tính đường kính trung bình (Dc , cm) của các giá trị D_i ($i = 1 \div n$) ban đầu ở hai nghiệm thức, nghĩa là $Dc = \sum D_i/n$ với n là tổng dung lượng của hai mẫu. Sau đó giá trị Zd_{DC} của từng mẫu được tính theo công thức:

+ Đối với nghiệm thức A : $Zd_{dc}(A) = Zd_{bq}(A) + b(Dc - D_{Abq})$.

+ Đối với nghiệm thức B : $Zd_{dc}(B) = Zd_{bq}(B) + b(Dc - D_{Bbq})$.

Trong các công thức trên, b là hệ số hồi quy của mô hình tuyến tính bậc 1 giữa biến D với Zd . Ở ví dụ trên đây, hệ số b và Zd_{DC} của từng mẫu được ghi lại tương ứng ở **Bảng 4.60 và 4.61**. **Bảng 4.62** là kết quả so sánh trung bình điều chỉnh của nghiệm thức A với B.

2. Trường hợp có nhiều nghiệm thức

Giả sử cần nghiên cứu phản ứng sinh trưởng về chiều cao (Zh , cm) của cây con Cẩm lai dưới ảnh hưởng của 4 hỗn hợp ruột bầu khác nhau. Mỗi nghiệm thức có 10 cây. Những cây con dùng trong thí nghiệm có chiều cao ban đầu khác nhau (H , cm). Các nghiệm thức được bố trí ngẫu nhiên trong cùng điều kiện môi trường. Thí nghiệm được theo dõi trong 2 tháng. Số liệu về H ban đầu và Zh của Cẩm lai sau 2 tháng thí nghiệm được ghi lại ở **Bảng 4.63**. Để giải bài toán này bằng mô hình tuyến tính tổng quát, bạn hãy thực hiện như sau :

Bước 1. Nhập dữ liệu vào Bảng tính (Xem ở phụ lục 10). Tại cửa sổ Data Editor, bạn nhập lần lượt biến H và Zh của 4 nghiệm thức (Các hỗn hợp ruột bầu) tương ứng vào cột 1 và 2; cột 3 nhập mã số biểu thị cho 4 nghiệm thức. Sau đó mở trang Variable View để mã hóa cho các tên biến.

Bước 2. Gọi chương trình tính toán. Từ thanh Menu chính, bạn hãy chọn Analyze $\Rightarrow$ General Linear Model $\Rightarrow$ Univariate. Khi cửa sổ Univariate xuất hiện, bạn hãy thực hiện như sau :

+ Chọn biến Zh và chuyển nó sang ô Dependent Variable. Kế đến chọn biến H và chuyển nó vào ô Covariate. Sau đó chọn biến nghiệm thức và chuyển nó sang ô Fixed factor(s).

+ Chọn Model và đánh dấu ô Custom. Kế đến chọn tất cả các biến nghiệm thức và H từ ô Factors & Covariate, rồi chọn tiếp Main Effect trong Build Term(s), sau đó chuyển chúng sang ô Custom. Đến đây chọn Continue.

+ Chọn Option và đánh dấu các mục Descriptive statistics, Observed power, Parameter estimates, Residual plot, rồi chọn Continue. Sau cùng chọn OK để chương trình tính toán và hiển thị kết quả.

Bảng 4.63. Chiều cao ban đầu (H , cm) và lượng tăng trưởng chiều cao (Zh , cm) của Cẩm lai hai tháng tuổi

Hỗn hợp ruột bầu (nghiệm)	A		79	65	57	51	57	66	44	41	44	36
		Zh	1.96	1.77	1.62	1.76	1.88	1.5	1.6	1.49	1.77	1.27
	B	H	61	59	59	53	56	50	45	39	38	45
		Zh	1.4	1.79	1.61	1.47	1.69	1.48	1.4	1.42	1.29	1.26
	C	H	62	73	58	43	50	44	48	51	40	38
		Zh	1.22	1.39	1.28	1.28	1.45	1.22	1.31	1.57	1.21	1.06
	D	H	71	60	54	50	60	61	44	53	41	38
		Zh	1.15	1.28	1.4	1.37	1.19	1.18	1.2	0.96	1.13	1.12

Bảng 4.64. Thống kê mô tả (Descriptive Statistics)
Dependent Variable: Zh

Hỗn hợp ruột bầu	Mean	Std. Deviation	N
1	1.6620	.2067	10
2	1.4810	.1693	10
3	1.2990	.1426	10
4	1.1980	.1279	10
Total	1.4100	.2392	40

Bảng 4.65. Kiểm định ngang bằng phương sai

F	df1	df2	Sig.
.287	3	36	.835

a Design: Intercept+HONHOP+H

Số liệu của các **Bảng 4.64 – 4.69** báo cáo tóm tắt các kết quả phân tích thống kê. Nội dung và cách giải thích kết quả của các Bảng này cũng tương tự như trường hợp có 2 mẫu.

Bảng 4.66. Kiểm định ảnh hưởng của các yếu tố

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1.488	4	.372	17.519	.000
Intercept	1.596	1	1.596	75.170	.000
Honhop	1.220	3	.407	19.154	.000
H	.230	1	.230	10.811	.002
Error	.743	35	2.123E-02		
Total	81.755	40			
Corrected Total	2.231	39			

a R Squared = .667 (Adjusted R Squared = .629)

Bảng 4.67. Ước lượng các tham số (Parameter Estimates)

Parameter ^a	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept ^b	0.808	0.127	6.341	0	0.549	1.066
[PTHUC=1]	0.458	0.065	7.028	0	0.326	0.59
[PTHUC=2]	0.303	0.065	4.628	0	0.17	0.436
[PTHUC=3]	0.119	0.065	1.825	0.077	-0.01342	0.252
[PTHUC=4]	0	.	.	.	.	.
H	7.3E-03b	0.002	3.288	0.002	0.003	0.012

a This parameter is set to zero because it is redundant; b. Hệ số góc của mô hình tuyến tính bậc 1 giữa H với Zh

Bảng 4.68. Ước lượng trung bình điều chỉnh (Estimates)

Hỗn hợp ruột bầu	Mean ^(a)	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	1.648	.046	1.554	1.742
2	1.493	.046	1.399	1.587
3	1.309	.046	1.216	1.403
4	1.190	.046	1.096	1.284

a Evaluated at covariates appeared in the model: X = 52.100.

Bảng 4.69. So sánh các cặp nghiệm thức (Pairwise Comparisons)

(I) Hỗn hợp ruột bầu	(J) Hỗn hợp ruột bầu	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
1	2	.155×	.066	.024	0.022	.289
	3	.339×	.066	.000	.206	.472
	4	.458×	.065	.000	.326	.590
...	...	...	...	...	...	...
4	1	-.458×	.065	.000	-.590	-.326
	2	-.303×	.065	.000	-.436	-.170
	3	-.119	.065	.077	-.252	-0.134

B. PHÂN TÍCH HIỆP PHƯƠNG SAI 1 YẾU TỐ: Kiểu thiết kế theo khối ngẫu nhiên

Giả sử cần so sánh sinh trưởng chiều cao (H, mm) của 6 dòng Dó bầu sau 2 tháng thí nghiệm dưới ảnh hưởng của mật độ cấy (N, cây/m²) trong gieo ươm. Thí nghiệm được bố trí theo kiểu khối ngẫu nhiên hoàn toàn với 4 lần lặp lại. Kết quả thí nghiệm được ghi lại ở **Bảng 4.70**. Trình tự phân tích ANOCOVA bằng mô hình tuyến tính tổng quát như sau:

Bước 1. Nhập dữ liệu vào Bảng tính (xem ở phụ lục 11). Tại cửa sổ Data Editor, bạn nhập lần lượt biến N và H vào cột 1 và 2, còn cột 3 và 4 nhập tương ứng mã số biểu thị cho 4 khối và 6 dòng Dó bầu. Sau đó mở trang Variable view để mã hóa cho các tên biến.

Bước 2. Gọi chương trình tính toán. Từ thanh Menu chính, bạn hãy chọn Analyze ⇒ General Linear Model ⇒ Univariate. Khi cửa sổ Univariate xuất hiện, bạn hãy thực hiện như sau :

+ Chọn biến H và chuyển nó sang ô Dependent Variable. Kế đến chọn biến N và chuyển nó vào ô Covariate. Sau đó chọn đồng thời hai biến khối và dòng rồi chuyển chúng sang ô Fixed factor(s).

+ Chọn Model và đánh dấu ô Custom. Kế đến chọn tất cả các biến khối và dòng từ ô Factors & Covariate, rồi chọn tiếp Main Effect trong Build Term(s) và chuyển chúng sang ô Custom; chọn Continue.

+ Chọn Option và đánh dấu các mục Descriptive statistics, Observed power, Parameter estimates, Residual plot; rồi chọn Continue. Sau cùng chọn OK để chương trình tính toán và hiển thị kết quả.

Bảng 4.71 – 4.75 dẫn số liệu báo cáo tóm tắt các kết quả phân tích thống kê để các bạn tham khảo.

Bảng 4.70. Mật độ cấy và chiều cao (H, cm) của Dó bầu sau hai tháng

Dòng	Khu đất (Khối)							
	I		II		III		IV	
	N	H	N	H	N	H	N	H
A	28	202	22	165	27	191	19	134
B	23	145	26	201	28	203	24	180
C	27	188	24	185	27	185	28	220
D	24	201	28	231	30	238	30	261
E	30	202	26	178	26	198	29	226
F	30	228	25	221	27	207	24	204

Bảng 4.71. Thống kê mô tả (Descriptive Statistics)

Khối	Dòng	Mean	Std. Deviation	N
1	1	202.00	.	1
	2	145.00	.	1
	3	188.00	.	1
	4	201.00	.	1
	5	202.00	.	1
	6	228.00	.	1
	Total	194.33	27.46	6
...	...	...	...	

Bảng 4.72. Kiểm định ảnh hưởng của các yếu tố

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power ^a
Corrected Model	17317.214 ^b	9	1924.135	19.789	.000	178.097	1.000
Intercept	25.230	1	25.230	.259	.618	.259	.076
KHOI	1502.394	3	500.798	5.150	.013	15.451	.828
DONG	3227.275	5	645.455	6.638	.002	33.191	.972
N	7391.047	1	7391.047	76.012	.000	76.012	1.000
Error	1361.286	14	97.235				
Total	976280.000	24					
Corrected Total	18678.500	23					

a. Computed using alpha = .05; b. R Squared = .927 (Adjusted R Squared = .880)

Giải thích và nhận xét kết quả. Số liệu của **Bảng 4.71** là chiều cao trung bình và sai tiêu chuẩn về H của 6 dòng Dó bầu thuộc từng khối và của chung 4 khối. **Bảng 4.72** dẫn kết quả phân tích ANOCOVA. Bạn hãy chú ý đến giá trị P của dòng có nhãn “DONG”. Vì giá trị $P = 0.002 \ll 0.05$, nên sinh trưởng chiều cao của các dòng Dó bầu khác nhau rất có ý nghĩa. Ở dòng có nhãn N, giá trị $P = 0.000$; điều đó cho biết mật độ cấy có ảnh hưởng không đồng đều đến sinh trưởng chiều cao của các dòng Dó bầu. **Bảng 4.73** là ước lượng các tham số của mô hình.

Bảng 4.73. Ước lượng các tham số (Parameter Estimates)

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval		Noncent. Parameter	Observed Power
					Lower Bound	Upper Bound		
Intercept	11.256	24.627	.457	.655	-41.565	64.077	.457	.071
[KHOI=1]	-20.577	5.825	-3.533	.003	-33.070	-8.084	3.533	.907
[KHOI=4]	0 ^b	.	.	.	.	.	.	.
[DONG=1]	-21.855	7.345	-2.975	.010	-37.610	-6.101	2.975	.790
...	...							
[DONG=6]	0	.	.	.	.	.	.	.
N	8.058	.924	8.719	.000	6.076	10.040	8.719	1.000

a. Computed using alpha = .05; b. This parameter is set to zero because it is redundant.

Bảng 4.74. Ước lượng H cho các dòng Dó bầu (Estimates)

Dòng	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	191.802 _a	5.381	180.260	203.344
2	190.979 _a	5.031	180.189	201.770

3	193.157 _a	4.933	182.577	203.737
4	219.320 _a	5.165	208.242	230.399
5	189.585 _a	5.101	178.644	200.526
6	213.657 _a	4.933	203.077	224.237

a Evaluated at covariates appeared in the model: N = 26.33.

Bảng 4.74 là các trị chiều cao điều chỉnh của 6 dòng Dó bầu. **Bảng 4.75** là các so sánh đối chiếu H trung bình đã điều chỉnh của từng cặp nghiệm thức với nhau. Dấu (×) ở bên cạnh các con số của cột Mean Difference (I-J) cho biết các trung bình mẫu khác nhau có ý nghĩa về mặt thống kê.

C. PHÂN TÍCH HIỆP PHƯƠNG SAI 1 YẾU TỐ: Thiết kế theo kiểu khối la tinh

Giả sử cần so sánh tăng trưởng về chiều cao (Zh, cm) của 4 dòng Keo lai (A, B, C, D) sau 4 đợt thu hoạch khác nhau. Thí nghiệm được bố trí theo kiểu khối la tinh: Khối cột biểu thị 4 lần lặp lại, khối dòng biểu thị 4 đợt thu hoạch. Những cây ban đầu đưa vào thí nghiệm có mức độ bị nhiễm bệnh (X%) khác nhau. Kết quả thí nghiệm được ghi lại ở **Bảng 4.76**. Để giải bài toán này bằng mô hình tuyến tính tổng quát trong SPSS 10.0, bạn hãy thực hiện như sau :

Bước 1. Nhập dữ liệu vào Bảng tính (xem phụ lục 12). Tại cửa sổ Data Editor, bạn nhập lần lượt biến X và Zh vào cột 1 và 2; cột 3 và 4 nhập tương ứng mã số biểu thị cho 4 khối cột và 4 khối dòng; cột 5 nhập mã số biểu thị cho các dòng Keo lai. Sau đó mở trang Variable view để mã hóa cho các tên biến.

Bảng 4.75. So sánh sự khác biệt về H của từng cặp dòng Dó bầu

		Mean Difference (I-J)	Std. Error	Sig. ^a	95% Confidence Interval for Difference	
(I) Dòng	(J) Dòng				Lower Bound	Upper Bound
1	2	.822	7.068	.909	-14.336	15.981
	3	-1.355	7.345	.856	-17.110	14.399
	4	-27.519×	7.892	.004	-44.445	-10.592
	5	2.217	7.786	.780	-14.483	18.917
	6	-21.855×	7.345	.010	-37.610	-6.101
...	...	...	...	...	...	...
6	1	21.855×	7.345	.010	6.101	37.610
	2	22.678×	7.068	.006	7.519	37.836
	3	20.500×	6.973	.011	5.545	35.455
	4	-5.663	7.109	.439	-20.911	9.584
	5	24.072×	7.068	.004	8.914	39.231

× The mean difference is significant at the .05 level.

a Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Bước 2. Gọi chương trình tính toán. Từ thanh Menu chính, bạn hãy chọn Analyze ⇒ General Linear Model ⇒ Univariate. Khi cửa sổ Univariate xuất hiện, bạn hãy thực hiện theo trình tự như sau :

+ Chọn biến Zh và chuyển nó sang ô Dependent variable. Kế đến chọn biến X và chuyển nó vào ô Covariate. Sau đó chọn cả ba biến (khối dòng, khối cột và dòng Keo lai) rồi chuyển chúng sang ô Fixed factor(s).

+ Chọn Model và đánh dấu ô Custom. Kế đó chọn tất cả các biến từ ô Factors & Covariate, rồi chọn tiếp Main Effect trong Build Term(s) và chuyển chúng sang ô Custom, sau đó chọn Continue.

+ Chọn Option và đánh dấu các mục Descriptive statistics, Observed power, Parameter estimates, Residual plot..., rồi chọn Continue. Sau cùng chọn OK để chương trình tính toán và hiển thị kết quả. Kết quả phân tích thống kê dẫn ra ở **Bảng 4.77 – 4.80**).

Giải thích và nhận xét kết quả. **Bảng 4.77** dẫn kết quả phân tích ANOCOVA. Giả thuyết H_0 : Sinh trưởng H của các dòng Keo lai khác nhau không có ý nghĩa. Để kiểm định giả thuyết này, bạn hãy chú ý đến dòng có nhãn DONG. Vì giá trị $P = 0.059 \approx 0.06$, nên sinh trưởng chiều cao của các dòng Keo lai khác nhau có ý nghĩa thống kê. **Bảng 4.78** dẫn kết quả ước lượng các tham số của mô hình hồi quy. Giả thuyết H_0 : Hệ số hồi quy bằng zero. Từ **Bảng 4.78** cho thấy hệ số hồi quy ($b = -0.346$) tồn tại ($P = 0.005$). **Bảng 4.79** là các giá trị tăng trưởng về chiều cao của 4 dòng Keo lai đã được điều chỉnh. **Bảng 4.80** là các so sánh từng cặp nghiệm thức với nhau (Zh đã được điều chỉnh). Dấu (×) ở bên cạnh các con số của cột Mean Difference (I-J) cho biết các trung bình mẫu khác nhau có ý nghĩa thống kê.

Bảng 4.76. Mức độ bị bệnh và chiều cao (H, cm) của Keo lai sau 4 đợt thu hoạch

			Khối cột			
			I	II	III	IV
Khối dòng	1	Dòng	A	B	D	C
		X	29.2	34.7	30.2	48.2
		Zh	19.7	20.7	20.1	14.7
	2	Dòng	C	A	B	D
		X	4.3	19.3	10.1	14
		Zh	26.7	21.3	28.3	25.1
	3	Dòng	B	D	C	A
		X	14	7.2	6.3	1
		Zh	26	24.9	29	28.7
	4	Dòng	D	C	A	B
		X	8.9	6.7	6.4	5.6
		Zh	29.8	29	27.3	34.1

Bảng 4.77. Kiểm định ảnh hưởng giữa các yếu tố (ANOCOVA)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power
Corr. Model	350.277	10	35.028	17.689	.003	176.893	.997
Intercept	1368.219	1	1368.219	690.962	.000	690.962	1.000
KHOI	6.422	3	2.141	1.081	.437	3.243	.163
DONG	29.402	3	9.801	4.949	.059	14.848	.581
DOT	23.140	3	7.713	3.895	.089	11.686	.480
X	45.324	1	45.324	22.889	.005	22.889	.964
Error	9.901	5	1.980				
Total	10632.000	16					
Corrected Total	360.178	15					

a. Computed using alpha = .05; b. R Squared = .973 (Adjusted R Squared = .918)

Bảng 4.78. Ước lượng các tham số (Parameter Estimates)

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval		Noncent. Parameter	Observed Power
					Lower Bound	Upper Bound		
Intercept	33.234	1.316	25.263	.000	29.853	36.616	25.263	1.000
[KHOI=1]	-1.172	1.020	-1.149	.302	-3.794	1.450	1.149	.156
[KHOI=2]	-1.753	.995	-1.761	.138	-4.311	.805	1.761	.299
[KHOI=3]	-.841	1.035	-.813	.453	-3.502	1.820	.813	.102
[KHOI=4]	0	.	.	.	.	.	.	.
[DONG=1]	-1.430	1.010	-1.416	.216	-4.026	1.166	1.416	.212
[DONG=2]	2.330	.995	2.341	.066	-.228	4.888	2.341	.473
[DONG=3]	-.325	.999	-.325	.758	-2.894	2.245	.325	.058
[DONG=4]	0	.	.	.	.	.	.	.
[DOT=1]	-1.332	2.299	-.579	.588	-7.243	4.579	.579	.076
[DOT=2]	-2.962	1.059	-2.796	.038	-5.685	-.239	2.796	.613
[DOT=3]	-2.822	.995	-2.836	.036	-5.380	-.264	2.836	.625
[DOT=4]	0	.	.	.	.	.	.	.
X	-.346	.072	-4.784	.005	-.532	-.160	4.784	.964

a. Computed using alpha = .05

Bảng 4.79. Các ước lượng Zh đã tu chỉnh (Estimates)

Dòng	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	23.764 _a	.711	21.936	25.591
2	27.524	.706	25.710	29.337
3	24.869	.704	23.060	26.679
4	25.194	.707	23.376	27.012

Bảng 4.80. So sánh từng cặp trung bình nghiệm thức

(I) Dòng	(J) Dòng	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
1	2	-3.760×	1.007	.014	-6.348	-1.172
	3	-1.105	.998	.319	-3.671	1.460
	4	-1.430	1.010	.216	-4.026	1.166
2	1	3.760×	1.007	.014	1.172	6.348
	3	2.655×	.998	.045	0.08963	5.219
	4	2.330	.995	.066	-.228	4.888
3	1	1.105	.998	.319	-1.460	3.671
	2	-2.655×	.998	.045	-5.219	-0.089
	4	-.325	.999	.758	-2.894	2.245
4	1	1.430	1.010	.216	-1.166	4.026
	2	-2.330	.995	.066	-4.888	.228
	3	.325	.999	.758	-2.245	2.894

Bảng 4.81. Chiều cao (H, cm) của Dầu song nài 2 tháng tuổi

Nghiệm thức		Khối					
		I	II	III	IV	V	VI
To	X	183	176	291	254	225	249
	Zh	2.45	2.25	4.38	4.35	3.42	3.27
Tp	X	356	300	301	271	288	258
	Zh	6.71	5.44	4.92	5.23	6.74	4.74
Tk	X	224	258	244	217	192	236
	Zh	3.22	4.14	2.32	4.42	3.28	4

Tpk	X	329	283	308	326	318	318
	Zh	6.34	5.44	5.22	8	6.96	6.96

4.5.2. PHÂN TÍCH HIỆP PHƯƠNG SAI HAI YẾU TỐ

Giả sử cần so sánh lượng tăng trưởng về chiều cao (Zh, cm) của Vên vên sau 2 tháng thí nghiệm dưới ảnh hưởng của 4 hỗn hợp ruột bầu: Đối chứng (kí hiệu To), bón phân lân (kí hiệu Tp), bón phân kali (kí hiệu Tk) và bón phân PK (kí hiệu Tpk). Thí nghiệm được bố trí theo kiểu khối ngẫu nhiên với 6 lần lặp lại. Các nghiệm thức được bố trí với mật độ cây (X) khác nhau. Kết quả thí nghiệm được ghi lại ở **Bảng 4.81**.

Để giải bài toán này bằng mô hình tuyến tính tổng quát trong SPSS 10.0, bạn hãy thực hiện như sau :

Bước 1. Nhập dữ liệu vào Bảng tính (**Phụ lục 13**). Tại cửa sổ Data Editor, bạn nhập lần lượt biến X và Zh vào cột 1 và 2, còn cột 3 và 4 nhập tương ứng mã số biểu thị cho 6 khối và 4 nghiệm thức. Sau đó mở trang Variable view để mã hóa cho các tên biến.

Bước 2. Gọi chương trình tính toán. Từ thanh Menu chính, bạn hãy chọn Analyze ⇒ General Linear Model ⇒ Univariate. Khi cửa sổ Univariate xuất hiện, bạn hãy thực hiện theo thủ tục như sau :

+ Chọn biến Zh và chuyển nó sang ô Dependent variable. Kế đến chọn biến X và chuyển nó vào ô Covariate. Sau đó chọn cả hai biến (khối và nghiệm thức) rồi chuyển chúng sang ô Fixed factor(s).

+ Chọn Model và đánh dấu ô Custom. Kế đến chọn tất cả các biến từ ô Factors & Covariate, rồi chọn tiếp Main Effect trong Build Term(s) và chuyển chúng sang ô Custom, sau đó chọn Continue.

+ Chọn Option và đánh dấu các mục Descriptive statistics, Observed power, Parameter estimates, Residual plot, rồi chọn Continue. Sau cùng chọn OK để chương trình tính toán và hiển thị kết quả (**Bảng 4.82 – 4.86**).

Bảng 4.82. Các thống kê mô tả (Descriptive Statistics)

Dependent Variable: Zh

Khối	Nghiệm thức	Mean	Std. Deviation	N
1	Đối chứng	2.4500	.	1
	P	6.7100	.	1
	K	3.2200	.	1
	PK	6.3400	.	1
	Total	4.6800	2.1588	4
...	...	...	...	...

Bảng 4.83. Kiểm định ảnh hưởng giữa các yếu tố

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power
Corrected Model	56.182	9	6.242	27.696	.000	249.263	1.000
Intercept	.380	1	.380	1.684	.215	1.684	.227
KHOI	7.227	5	1.445	6.413	.003	32.063	.967
NGHIEM	2.035	3	.678	3.010	.066	9.029	.579
X	8.616	1	8.616	38.224	.000	38.224	1.000
Error	3.156	14	.225				
Total	602.739	24					

Corrected Total	59.338	23					
-----------------	--------	----	--	--	--	--	--

a Computed using alpha = .05; b R Squared = .947 (Adjusted R Squared = .913)

Bảng 4.84. Ước lượng các tham số (Parameter Estimates)

Dependent Variable: Zh

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval		Noncent. Parameter	Observed Power
					Lower Bound	Upper Bound		
Intercept	-.580	1.177	-.493	.630	-3.105	1.944	.493	.075
[KHOI=1]	-.238	.337	-.705	.492	-.960	.485	.705	.101
[KHOI=2]	-.176	.338	-.522	.610	-.902	.549	.522	.078
[NGHIEM=1]	-1.235	.412	-3.001	.010	-2.118	-.353	3.001	.797
...	...	...			...	...	...	
[NGHIEM=4]	0	.	.	.	.	.	.	.
X	0.226	.004	6.183	.000	0.0147	0.0304	6.183	1.000

a Computed using alpha = .05

b This parameter is set to zero because it is redundant.

Giải thích và nhận xét kết quả. Bảng 4.82 dẫn kết quả phân tích các đặc trưng thống kê mô tả về Zh của Vên vên 2 tháng tuổi ở các nghiệm thức khác nhau. Bảng 4.83 dẫn kết quả phân tích ANOCOVA để kiểm định ảnh hưởng giữa các yếu tố. Bạn hãy chú ý đến dòng có nhãn “NGHIEM”. Vì giá trị $P = 0.066 \approx 0.07$, nên tăng trưởng về chiều cao của Vên vên 2 tháng tuổi ở các nghiệm thức khác nhau ($P = 0.07$).

Bảng 4.85. Các ước lượng Zh đã tu chỉnh (Estimates)

Nghiệm thức	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Đối chứng	4.194 _a	.237	3.686	4.702
P	4.979	.221	4.506	5.452
K	4.430	.239	3.917	4.944
PK	5.429	.258	4.875	5.984

a Evaluated at covariates appeared in the model: X = 266.875.

Bảng 4.84 báo cáo kết quả ước lượng các tham số của mô hình hồi quy. Kết quả cho thấy hệ số hồi quy ($b = 0.0226$) tồn tại ($P = 0.0005$). Bảng 4.85 là các giá trị chiều cao trung bình đã được điều chỉnh của 4 nghiệm thức. Bảng 4.86 là các so sánh từng cặp nghiệm thức với nhau. Dấu (×) ở bên cạnh các con số của cột Mean Difference (I-J) cho biết các trung bình mẫu khác nhau ở mức ý nghĩa 0.05. Từ số liệu của Bảng 4.86 cho thấy, ở những nghiệm thức có bốn lần, lượng trưởng về chiều cao của Vên vên sau 2 tháng thí nghiệm đều vượt trội hơn so với các nghiệm thức khác.

Bảng 4.86. So sánh từng cặp trung bình nghiệm thức

(I) Nghiệm thức	(J) Nghiệm thức	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Đối chứng	P	-.785×	.365	.049	-1.568	-0.0022
	K	-.236	.274	.403	-.824	.352
	PK	-1.235×	.412	.010	-2.118	-.353
P	Đối chứng	.785×	.365	.049	0.0022	1.568

	K	.549	.368	.158	-.240	1.338
	PK	-.450	.282	.133	-1.055	.155
K	Đối chứng	.236	.274	.403	-.352	.824
	P	-.549	.368	.158	-1.338	.240
	PK	-.999×	.415	.030	-1.888	-.109
PK	Đối chứng	1.235×	.412	.010	.353	2.118
	P	.450	.282	.133	-.155	1.055
	K	.999×	.415	.030	.109	1.888

4.5.3. PHÂN TÍCH ĐA HIỆP PHƯƠNG SAI

Phân tích đa hiệp phương sai (Analysis of multiple covariance = ANOMUCOVA) được áp dụng trong trường hợp có nhiều yếu tố không đồng nhất cùng phối hợp với yếu tố thí nghiệm ảnh hưởng đến biến phụ thuộc. Vì thế, thủ tục phân tích ANOMUCOVA cũng được thực hiện tương tự như phân tích hiệp phương sai (ANOCOVA). Dưới đây chỉ giới thiệu phân tích ANOMUCOVA qua hai kiểu bố trí thí nghiệm: (1) thí nghiệm một yếu tố được bố trí ngẫu nhiên hoàn toàn, (2) thí nghiệm một yếu tố được bố trí ngẫu nhiên theo khối.

A. PHÂN TÍCH ĐA HIỆP PHƯƠNG SAI: Thí nghiệm 1 yếu tố được bố trí ngẫu nhiên hoàn toàn

Giả sử cần phân tích ảnh hưởng của 4 cấp kích thước bầu đến lượng tăng trưởng về đường kính (Zd, mm) của Dầu song nòng sau 3 tháng thí nghiệm. Những cây con Dầu song nòng đưa vào thí nghiệm có chiều cao (H, cm) và tuổi (A, tháng) ban đầu khác nhau. Kết quả thí nghiệm được ghi lại ở **Bảng 4.87**. Để giải bài toán này bằng mô hình tuyến tính tổng quát trong SPSS 10.0, bạn hãy thực hiện như sau:

Bước 1. Nhập dữ liệu vào Bảng tính tính (xem phụ lục 14). Tại cửa sổ Data Editor, bạn nhập lần lượt biến H và A vào cột 1 và 2, còn cột 3 và 4 nhập tương ứng Zd và mã số biểu thị cho 4 nghiệm thức. Sau đó mở trang Variable view để mã hóa cho các tên biến.

Bảng 4.87. Lượng tăng trưởng đường kính của Dầu song nòng 3 tháng tuổi

Thí nghiệm	1	H	78	90	94	71	99	80	83	75	62	67
		A	6	6	8	5	6	5	6	5	4	4
		Zd	1.4	1.79	1.72	1.47	1.26	1.28	1.34	1.55	1.57	1.26
	2	H	78	99	80	75	94	91	75	63	62	67
		A	7	8	6	5	6	4	5	4	5	4
		Zd	1.61	1.31	1.12	1.35	1.29	1.24	1.29	1.43	1.29	1.26
	3	H	78	83	79	70	85	83	71	66	67	67
		A	8	6	6	5	6	4	5	4	4	4
		Zd	1.67	1.41	1.73	1.23	1.29	1.22	1.39	1.59	1.56	1.36
	4	H	77	71	78	70	95	96	71	63	62	67
		A	6	6	6	4	6	5	4	4	5	4
		Zd	1.4	1.47	1.37	1.15	1.22	1.48	1.31	1.27	1.22	1.36

Bước 2. Gọi chương trình tính toán. Từ thanh Menu chính, bạn hãy chọn Analyze $\Rightarrow$ General Linear Model $\Rightarrow$ Univariate. Khi cửa sổ Univariate xuất hiện, bạn hãy thực hiện như sau :

+ Chọn biến Zd và chuyển nó sang ô Dependent variable. Kế đến chọn biến A và H rồi chuyển chúng vào ô Covariate. Tiếp theo chọn biến nghiệm thức và chuyển nó sang ô Fixed factor(s).

Bảng 4.88. Thống kê mô tả (Descriptive Statistics)

Dependent Variable: Zd

Nghiệm thức	Mean	Std. Deviation	N
1	1.4640	.1906	10
2	1.3190	.1291	10
3	1.4450	.1821	10
4	1.3250	.1105	10
Total	1.3882	.1651	40

+ Chọn Model và đánh dấu ô Custom. Kế đến chọn tất cả các biến từ ô Factors & Covariate, rồi chọn tiếp Main Effect trong Build Term(s) và chuyển chúng sang ô Custom, sau đó chọn Continue.

+ Chọn Option và đánh dấu các mục Descriptive statistics, Observed power, Parameter estimates, Residual plot..., rồi chọn Continue. Sau cùng chọn OK để chương trình tính toán và hiển thị kết quả.

Bảng 4.88 đến 4.92 dẫn một số Bảng báo cáo kết quả phân tích thống kê để các bạn tham khảo. Cách giải thích và nhận xét kết quả phân tích ANOMUCOVA của mục này cũng tương tự như phân tích ANOCOVA.

Bảng 4.89. Kiểm định ảnh hưởng giữa các yếu tố

Dependent Variable: Zd

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power
Corrected Model	.326	5	0.065	3.013	.023	15.063	.800
Intercept	1.343	1	1.343	61.998	.000	61.998	1.000
NGHIEM	.160	3	0.0533	2.461	.079	7.382	.561
H	0.044	1	0.0441	2.035	.163	2.035	.284
A	.148	1	0.1480	6.833	.013	6.833	.719
Error	.736	34	0.0217				
Total	78.152	40					
Corrected Total	1.063	39					

a. Computed using alpha = .05; b. R Squared = .307 (Adjusted R Squared = .205)

Bảng 4.90. Các tham số ước lượng (Parameter Estimates)

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval		Noncent. Parameter	Observed Power
					Lower Bound	Upper Bound		
Intercept	1.285	0.170	7.549	0.000	0.939	1.631	7.549	1.000
[NGHIEM=1]	0.125	0.067	1.868	0.070	-0.011	0.261	1.868	0.442
[NGHIEM=2]	-0.019	0.066	-0.291	0.773	-0.154	0.116	0.291	0.059
[NGHIEM=3]	0.106	0.066	1.610	0.117	-0.028	0.241	1.610	0.347
[NGHIEM=4]	0.000	.	.	.	.	.	.	.

H	-0.004	0.003	-1.427	0.163	-0.009	0.002	1.427	0.284
A	0.066	0.025	2.614	0.013	0.015	0.118	2.614	0.719

a Computed using alpha = .05; b This parameter is set to zero because it is redundant.

Bảng 4.91. Các ước lượng trung bình mẫu đã điều chỉnh (Estimates)

Nghiệm thức	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	1.460a	.047	1.365	1.556
2	1.316a	.047	1.221	1.411
3	1.442a	.047	1.346	1.537
4	1.335a	.047	1.240	1.431

a Evaluated at covariates appeared in the model: H = 77.05, A = 5.28.

Bảng 4.92. So sánh từng cặp nghiệm thức (Pairwise Comparisons)

(I) Nghiệm thức	(J) Nghiệm thức	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
1	2	.144×	.066	.036	0.0103	.278
	3	0.0186	.067	.783	-0.1170	.154
	4	0.1250	.067	.070	-0.0110	.261
2	1	-.144×	.066	.036	-0.2780	-0.0103
	3	-0.1260	.066	.066	-0.2600	0.0090
	4	-0.0193	.066	.773	-0.1540	0.1160
3	1	-0.0186	.067	.783	-0.1540	0.1170
	2	0.1260	.066	.066	-0.0090	0.2600
	4	0.1060	.066	.117	-0.0279	0.2410
4	1	-0.125	.067	.070	-0.261	0.0110
	2	1.93E-02	.066	.773	-0.116	0.1540
	3	-0.106	.066	.117	-0.241	0.0279

B. PHÂN TÍCH ĐA HIỆP PHƯƠNG SAI: Thí nghiệm 1 yếu tố được bố trí theo khối ngẫu nhiên

Giả sử cần phân tích ảnh hưởng của 6 thành phần hỗn hợp ruột bầu đến lượng tăng trưởng về chiều cao (Zh, cm) của Dầu song **nàng sau 3 tháng** thí nghiệm. Những cây con đưa vào thí nghiệm có chiều cao (H, cm) và độ dài rễ (L_R, cm) khác nhau. Kết quả thí nghiệm được ghi lại ở **Bảng 4.93**.

Bảng 4.93. Lượng tăng trưởng về chiều cao của Dầu song nàng sau 3 tháng

Khối	I	H	25.6	28.5	25.4	33	30.8	28
		Lr	14.9	12.8	13.3	14.7	4.6	7.5
		Zh	19	25.3	22.2	32.8	35.3	35.8
	II	H	25.4	25.9	28.3	32.4	35.3	24.2
		Lr	7.2	9.2	9.5	9.7	6.8	7.5
		Zh	32.4	28.3	32.2	35.7	43.7	35.2
	III	H	27.9	23.7	34.4	27.5	32.5	32.9
		Lr	18.6	14.4	22.2	17.6	10	7.9
		Zh	26.2	20.6	34.7	29.6	40	47.2

Bảng 4.94. Thống kê mô tả

Khối	Nghiệm thức	Mean	Std. Deviation	N
1	1	19.000	.	1
	2	25.300	.	1
	3	22.200	.	1
	4	32.800	.	1
	5	35.300	.	1
	6	35.800	.	1
	Total	28.400	7.185	6
...	...			

Để giải bài toán này bằng mô hình tuyến tính tổng quát trong SPSS, bạn hãy thực hiện các bước giống như trường hợp phân tích ANOCOVA với thí nghiệm 1 yếu tố được bố trí theo khối. Bạn có thể tham khảo cách vào số liệu như ở phụ lục 15. Ở đây chỉ lưu ý rằng, bạn phải chọn cùng lúc cả biến H và Lr rồi chuyển chúng vào ô Covariance. Kết quả phân tích thống kê ở ví dụ này được ghi lại ở **Bảng 4.94 – 4.98**. Bảng 4.95 ghi lại thống kê mô tả. Bảng 4.95 cho biết kết quả kiểm định ảnh hưởng giữa các yếu tố. Bảng 4.96 báo cáo các tham số ước lượng. Bảng 4.97 báo cáo các ước lượng trung bình mẫu đã điều chỉnh (Estimates). Bảng 4.98 ghi lại kết quả so sánh từng cặp nghiệm thức (Pairwise Comparisons).

Bảng 4.95. Kiểm định ảnh hưởng giữa các yếu tố

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power
Corrected Model	954.230	9	106.026	30.220	.000	271.979	1.000
Intercept	.249	1	.249	.071	.797	.071	.056
NGHIEM	101.040	5	20.208	5.760	.015	28.799	.863
KHOI	91.928	2	45.964	13.101	.003	26.202	.969
H	194.384	1	194.384	55.404	.000	55.404	1.000
Lr	29.597	1	29.597	8.436	.020	8.436	.721
Error	28.068	8	3.508				
Total	19427.100	18					
Corrected Total	982.298	17					

Bảng 4.96. Các tham số ước lượng (Parameter Estimates)

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval		Noncent. Parameter	Observed Power
					Lower Bound	Upper Bound		
Intercept	9.275	5.368	1.728	.122	-3.104	21.654	1.728	.331
[NGHIEM=1]	-6.978	2.104	-3.317	.011	-11.830	-2.127	3.317	.827
...								
H	1.315	.177	7.443	.000	.907	1.722	7.443	1.000
L	-.647	.223	-2.904	.020	-1.160	-.133	2.904	.721

Bảng 4.97. Các ước lượng trung bình mẫu đã điều chỉnh (Estimates)

Nghiệm thức	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	30.681	1.296	27.691	33.671
2	28.971	1.218	26.162	31.780
3	31.410	1.316	28.374	34.445
4	31.659	1.225	28.833	34.485
5	31.686	1.707	27.751	35.622
6	37.659	1.383	34.471	40.848

Bảng 4.98. So sánh từng cặp nghiệm thức (Pairwise Comparisons)
Dependent Variable: Zh

(I) Nghiệm thức	(J) Nghiệm thức	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
1	2	1.710	1.561	.305	-1.890	5.309
	3	-.729	1.631	.667	-4.489	3.032
	4	-.978	1.730	.587	-4.968	3.012
	5	-1.005	2.540	.703	-6.863	4.853
	6	-6.978×	2.104	.011	-11.830	-2.127
2	1	-1.710	1.561	.305	-5.309	1.890
	3	-2.438	1.712	.192	-6.386	1.510
	4	-2.688	1.765	.166	-6.758	1.382
	5	-2.715	2.370	.285	-8.180	2.749
	6	-8.688×	1.921	.002	-13.119	-4.258
...	...	...	...	...	...	...

CHƯƠNG V

PHÂN TÍCH TƯƠNG QUAN VÀ HỒI QUY

A. KHÁI QUÁT VỀ PHÂN TÍCH HỒI QUY

5.1. MÔ HÌNH HỒI QUY TUYẾN TÍNH ĐƠN

5.1.1. Mô hình hồi quy và hàm hồi quy

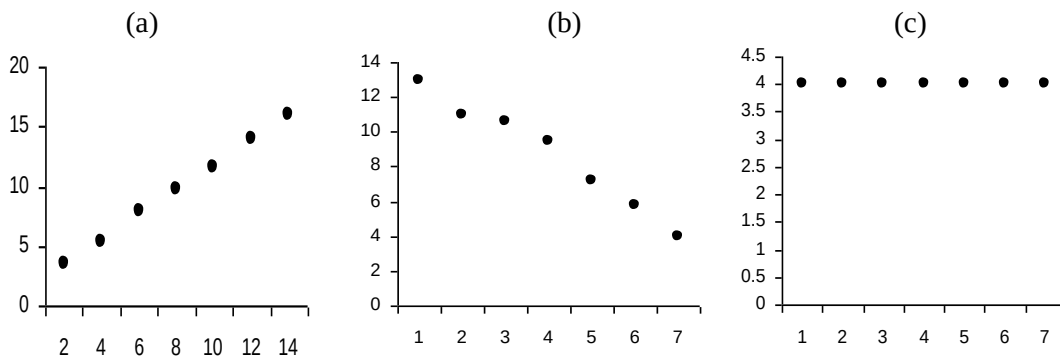
Hàm số mô tả quan hệ giữa y với x và sai số e được gọi là mô hình hồi quy. Mô hình hồi quy tuyến tính đơn có dạng:

$$Y = b_0 + b_1x + e \quad (5.1)$$

Trong mô hình hồi quy tuyến tính đơn, y là hàm số tuyến tính của x cộng với e ; b_0 và b_1 là các tham số của mô hình; e là biến ngẫu nhiên biểu thị sai số của mô hình. Sai số cho biết biến động của y không thể giải thích được bằng mối quan hệ tuyến tính giữa y với x . Phần sau chúng ta sẽ thảo luận các giả định của mô hình tuyến tính đơn. Một trong những giả định là trị số kỳ vọng của e bằng zero, nghĩa là $E(e) = 0$. Một giả định khác là $E(y) = b_0 + b_1x$. Nói khác đi, trị số trung bình của y là hàm tuyến tính của x .

Hàm số mô tả giá trị trung bình của y có quan hệ với x như thế nào được gọi là hàm hồi quy. Hàm hồi quy đối với hồi quy tuyến tính đơn có dạng:

$$E(y) = b_0 + b_1x \quad (5.2)$$



Hình 5.1. Mối quan hệ giữa y với x

Đồ thị của hàm hồi quy tuyến tính đơn là một đường thẳng; trong đó b_0 là điểm chặn (Intercept) của đường hồi quy, b_1 là độ dốc (Slope) của đường hồi quy, $E(y)$ là giá trị trung bình (kỳ vọng) của y tương ứng với một giá trị x nhất định. **Hình 5.1** là đồ thị của các hồi quy tuyến tính đơn. **Hình 5.1a** cho biết y có mối quan hệ dương với x ; nghĩa là y và x là đồng biến. **Hình 5.1b** cho biết y có mối quan hệ âm với x ; nghĩa là y và x là nghịch biến. **Hình 5.1c** cho biết y không có mối quan hệ với x .

5.1.2. Hàm hồi quy ước lượng

Khi biết các tham số b_0 và b_1 , chúng ta có thể sử dụng **hàm 5.2 để tính** giá trị trung bình y đối với một giá trị x nhất định. Nhưng trong thực tế chúng ta chưa biết các tham số b_0 và b_1 ; do đó chúng ta phải ước lượng các tham số này từ các quan sát mẫu. Các thống kê mẫu b_0 và b_1 được xem như là các ước lượng tương ứng của b_0 và b_1 . Khi thay thế b_0 và b_1 tương

ứng cho b_0 và b_1 , chúng ta nhận được hàm hồi quy ước lượng. Như vậy, hàm hồi quy ước lượng có dạng:

$$y = b_0 + b_1x \quad (5.3)$$

Ở **Hàm 5.3**, b_0 là điểm chặn, còn b_1 là độ dốc của đường hồi quy ước lượng, y là giá trị ước lượng trung bình (kỳ vọng) của y tương ứng với một giá trị x nhất định. Các tham số b_0 và b_1 được tính theo phương pháp bình phương nhỏ nhất.

5.1.3. Phương pháp bình phương nhỏ nhất

Phương pháp bình phương nhỏ nhất sử dụng số liệu mẫu để tính các tham số b_0 và b_1 của hàm hồi quy ước lượng. Bản chất của phương pháp này là làm nhỏ nhất tổng bình phương các sai lệch giữa giá trị của biến y_i và giá trị ước lượng của biến y_i .

$$\text{Tiêu chuẩn bình phương nhỏ : } \min S(y_i - y_i)^2 \quad (5.4)$$

Để $S(y_i - y_i)^2$ nhỏ nhất thì b_0 và b_1 được tính theo công thức

$$b_1 = \frac{\sum(x_i - x_{bq})(y_i - y_{bq})}{\sum(x_i - x_{bq})^2} \quad (5.5)$$

$$b_0 = y_{bq} - b_1x_{bq} \quad (5.6)$$

n = tổng số quan sát

x_i = giá trị của biến độc lập tương ứng với quan sát i

y_i = giá trị của biến phụ thuộc tương ứng với quan sát i

x_{bq} = giá trị trung bình của biến độc lập

y_{bq} = giá trị trung bình của biến phụ thuộc

5.1.4. Hệ số xác định

Hệ số xác định là số đo mức độ phù hợp của hàm hồi quy ước lượng. Hiệu số $y_i - y_i$ được gọi là sai lệch của quan sát i (Residual). Tổng bình phương các sai lệch hoặc sai số (kí hiệu SSE) là một đại lượng được tối thiểu hóa bằng phương pháp bình phương nhỏ nhất.

Tổng bình phương sai số (Error Sum of Square = SSE)

$$SSE = \sum(y_i - y_i)^2 \quad (5.7)$$

Giá trị SSE đo đặc sai số khi sử dụng hàm hồi quy ước lượng để ước lượng các giá trị của biến phụ thuộc y_i ở mẫu. Nếu sử dụng giá trị y_{bq} của mẫu để ước lượng y_i thì sai số ước lượng là $y_i - y_{bq}$. Đại lượng $\sum(y_i - y_{bq})^2$ được gọi là tổng bình phương toàn bộ (SST = Total Sum of Square).

Tổng bình phương do hồi quy

$$SSR = \sum(y - y_{bq})^2 \quad (5.8)$$

Tổng bình phương toàn bộ

$$SST = \sum(y_i - y_{bq})^2 \quad (5.9)$$

Chúng ta có thể Hình dung SST đo đặc sự phân bố của các y_i xung quanh y_{bq} tốt như thế nào, còn SSE đo đặc sự phân bố của các y_i xung quanh y tốt như thế nào (**Hình 5.2**).

Để đo đặc giá trị y trên đường hồi quy ước lượng cách xa y_{bq} bao nhiêu, người ta sử dụng đại lượng tổng bình phương do hồi quy (SSR = Regression Sum of Square).

Giữa SST, SSR và SSE có quan hệ với nhau như sau:
 $SST = SSR + SSE$ (5.10)

Nếu tất cả các điểm quan sát y_i đều nằm trên đường y thì hàm hồi quy ước lượng phù hợp hoàn toàn. Điều đó có nghĩa là $y_i - y_i = 0$ hay $SSE = 0$. Vì $SST = SSR + SSE$, nên để hàm hồi quy ước lượng phù hợp hoàn toàn thì $SSR/SST = 1$. Như vậy, mức phù hợp của hàm hồi quy ước lượng phụ thuộc vào SSE lớn hay nhỏ. $SSE_{\max}$ khi $SSR = 0$ và $SST = SSE$. Tỷ số SSR/SST được gọi là hệ số xác định (Coefficient of Determination), kí hiệu là r^2 . Hệ số r^2 lấy giá trị từ 0 đến 1.

Hệ số xác định r^2

$$r^2 = \frac{SSR}{SST} \quad (5.11)$$

Căn bậc hai của hệ số xác định được gọi là hệ số tương quan (kí hiệu r). Hệ số tương quan r lấy giá trị từ -1 đến +1. Khi $r = 0$ thì x và y không có quan hệ. Khi $r = -1$ thì x và y có quan hệ tuyến tính âm rất chặt chẽ. Điều đó có nghĩa là các điểm quan sát nằm trên đường thẳng có độ dốc âm. Khi $r = +1$ thì x và y có quan hệ tuyến tính dương rất chặt chẽ. Điều đó có nghĩa là các điểm quan sát nằm trên đường thẳng có độ dốc dương.

Hệ số tương quan mẫu (Sample Correlation Coefficient)

$$r_{xy} = (\text{dấu của } b_1) \sqrt{\frac{SSR}{SST}} = \sqrt{r^2} \quad (5.12)$$

Khi x và y có quan hệ tuyến tính thì cả r và r^2 đều đo đặc mức độ (cường độ) chặt chẽ của mối quan hệ giữa hai biến. Lưu ý rằng, hệ số r chỉ đo đặc mức độ chặt chẽ của mối quan hệ tuyến tính giữa hai biến; hệ số r^2 còn được dùng để đo đặc mức độ chặt chẽ của mối quan hệ tuyến tính và phi tuyến tính giữa hai hay nhiều biến độc lập. Vì thế, r^2 có phạm vi ứng dụng rất rộng.

Khi r_{xy} lấy giá trị nhỏ thì cần phải kiểm định ý nghĩa của hệ số r_{xy} .

Giả thuyết $H_0: \rho_{xy} = 0$; $H_a: \rho_{xy} \neq 0$

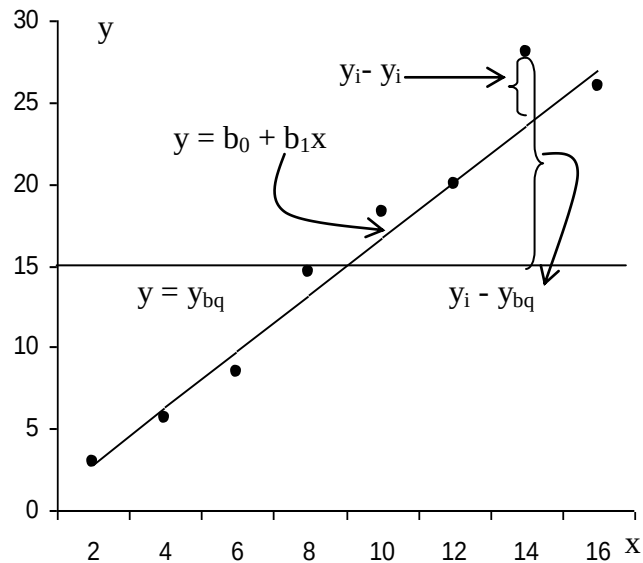
Thống kê kiểm định (Test Statistic)

$$t = r_{xy} \sqrt{\frac{n-2}{1-r_{xy}^2}} \quad (5.13)$$

Quy tắc quyết định (Rejection Rule)

H_0 : Nếu $t < -t_{\alpha/2(n-2)}$ hoặc $t > t_{\alpha/2(n-2)}$

ở đây $t_{\alpha/2}$ dựa trên phân bố t với $n-2$ bậc tự do; n = dung lượng quan sát.



Hình 5.2. Sai lệch giữa đường hồi quy ước lượng y và $y = y_{bq}$.

5.1.5. Các giả định của mô hình phù hợp

Mô hình hồi quy tuyến tính đơn giả định có dạng:

$$y = \beta_0 + \beta_1 x + e$$

Hàm hồi quy tuyến tính ước lượng có dạng:

$$y = b_0 + b_1 x$$

Chúng ta đã biết r^2 đo đặc mức độ phù hợp của hàm hồi quy ước lượng. Mặc dù vậy, khi r^2 rất lớn nhưng nó vẫn không thể phản ánh rõ ràng mức độ phù hợp của mô hình giả định. Vì thế, để biết mô hình giả định có phù hợp hay không, bước tiếp theo là kiểm định mức ý nghĩa của mối quan hệ. Trong phân tích hồi quy, kiểm định mức ý nghĩa dựa trên các giả định về sai số e .

Các kiểm định về sai số trong mô hình hồi quy $y = \beta_0 + \beta_1 x + e$:

1. Sai số e là một biến ngẫu nhiên có $E(e) = 0$.
2. Phương sai của e (kí hiệu s^2) bằng nhau đối với mọi x .
3. Các giá trị của e là độc lập.
4. Sai số e có phân bố $N(0, \sigma^2)$.

5.1.6. Kiểm định mức ý nghĩa

Trong hàm hồi quy tuyến tính đơn, $E(y) = \beta_0 + \beta_1 x$. Nếu $\beta_1 = 0$ thì $E(y) = \beta_0$. Trong trường hợp này y_{bq} không phụ thuộc vào x . Ngược lại, nếu $\beta_1 \neq 0$ thì x và y có quan hệ với nhau. Như vậy, thực chất của kiểm định mức ý nghĩa là kiểm định giả thuyết $H_0: \beta_1 = 0$. Để kiểm định giả thuyết H_0 , chúng ta cần thực hiện các tính toán sau đây:

1. Ước lượng phương sai tổng thể (d^2)

Phương sai tổng thể (d^2) được ước lượng gần đúng thông qua sai số trung bình bình phương (MSE).

Phương sai tổng thể (d^2) ước lượng

$$d^2 = s^2 = \text{MSE} = \frac{\text{SSE}}{n - 2} \quad (5.14)$$

Sai số chuẩn ước lượng (Standard Error of the Estimate)

$$s = \sqrt{\text{MSE}} = \sqrt{\frac{\text{SSE}}{n - 2}} = \sqrt{s^2} \quad (5.15)$$

2. Kiểm định t

Mục đích của kiểm định t là xác định hệ số b_1 có khác zero hay không. Tóm tắt của kiểm định t như sau:

Giả thuyết $H_0: b_1 = 0$, $H_a: b_1 \neq 0$.

Thống kê kiểm định t (t Test Statistic)

$$t = \frac{b_1}{s_{b1}} \quad (5.16)$$

Sai tiêu chuẩn ước lượng của hệ số b_1 (Estimated Standard Deviation of b_1)

$$s_{b1} = \frac{s}{\sqrt{\sum (x_i - \bar{x}_{bq})^2}} \quad (5.17)$$

Quy tắc quyết định

Ở mức ý nghĩa α ($\alpha = 0,05$ hoặc $0,01$), bác bỏ H_0 khi:

$$\frac{b_1}{s_{b1}} < -t_{\alpha/2(n-2)} \text{ hay } \frac{b_1}{s_{b1}} > t_{\alpha/2(n-2)}.$$

ở đây $\alpha/2$ dựa trên phân bố t với $n - 2$ độ tự do.

Sử dụng mức ý nghĩa (p - value)

Ở mức ý nghĩa α ($\alpha = 0,05$ hoặc $0,01$), bác bỏ H_0 khi p - value $< \alpha$.

Sử dụng khoảng tin cậy của b_1

Khoảng tin cậy của b_1 có dạng như sau: $b_1 \pm t_{\alpha/2} s_{b1}$; trong đó b_1 là ước lượng điểm, $t_{\alpha/2}$ là giá trị t lấy vùng $\alpha/2$ ở đuôi trên của phân bố t với $n - 2$ độ tự do, $t_{\alpha/2} s_{b1}$ là biên độ sai số. Khi kiểm định mức ý nghĩa bằng kiểm định t , giả thuyết đặt ra như sau:

$$H_0: b_1 = 0, H_a: b_1 \neq 0.$$

Quy tắc quyết định

Giả thuyết H_0 bị bác bỏ khi khoảng tin cậy của b_1 khác zero.

3. Kiểm định F

Kiểm định F (dựa trên phân bố xác suất F) cũng được sử dụng để kiểm định mức ý nghĩa trong hồi quy. Với hàm hồi quy tuyến tính 1 biến, kiểm định F cũng đưa ra cùng kết luận như kiểm định t . Nhưng khi hàm hồi quy đa tuyến tính (từ hai biến độc lập trở lên) thì chỉ có kiểm định F được sử dụng.

Thống kê kiểm định F (The F Test Statistic)

$$F = \frac{MSR}{MSE} \quad (5.18)$$

Trong đó MSR là hồi quy bình phương trung bình (Mean Square Regression).

$$MSR = \frac{SSR}{\text{số biến độc lập (p)}}$$

Quy tắc quyết định

+ Ở mức ý nghĩa α ($\alpha = 0,05$ hoặc $0,01$), bác bỏ H_0 khi $F > F_\alpha$; trong đó F_α dựa trên phân bố F với 1 độ tự do ở tử số và $n - 2$ độ tự do ở mẫu số.

+ Sử dụng mức ý nghĩa (p - value). Ở mức ý nghĩa α ($\alpha = 0,05$ hoặc $0,01$), bác bỏ H_0 khi p - value $< \alpha$.

Bảng phân tích biến động (ANOVA) có thể được sử dụng để tổng hợp kiểm định F (Bảng 5.1).

5.1.7. Sử dụng hàm hồi quy ước lượng để ước lượng và dự đoán

5.1.7.1. Ước lượng điểm

Sau khi xây dựng được hàm hồi quy giữa x với y , chúng ta có thể sử dụng nó để phát triển một ước lượng điểm của giá trị trung bình y đối với giá trị x nhất định hoặc dự đoán một giá trị nhất định của y tương ứng với một giá trị nhất định của x .

Bảng 5.1. Hình dạng chung của Bảng ANOVA đối với hồi quy tuyến tính đơn

Nguồn biến động (Source)	Tổng bình phương (Sum of Squares)	Độ tự do (Degrees of Freedom)	Trung bình bình phương (Mean Square)	F
Hồi quy (Regression)	SSR	1	$MSR = \frac{SSR}{1}$	$\frac{MSR}{MSE}$
Sai số (Error)	SSE	$n - 2$	$MSE = \frac{SSE}{n - 2}$	
Tổng (Total)	SST	$n - 1$	MST	

Ví dụ. Hàm hồi quy biểu thị trữ lượng rừng (y , m³/ha) phụ thuộc vào tuổi (x , năm) có dạng như sau: $y = 60 + 5x$. Giả sử chủ rừng mong muốn biết được ước lượng trữ lượng rừng trung bình của các quần thụ ở tuổi 10 năm. Từ $y = 60 + 5x$, khi thay $x = 10$, ta có $y = 60 + 5 \times 10 = 110$ m³/ha.

Giả sử chủ rừng lại có yêu cầu dự đoán trữ lượng rừng (y , m³/ha) đối với một quần thụ nhất định ở tuổi 10 năm. Trong trường hợp này, chủ rừng không quan tâm đến trữ lượng trung bình của tất cả các quần thụ, mà chỉ quan tâm đến việc dự đoán trữ lượng gỗ cho một quần thụ cụ thể. Để đạt mục đích ấy, chúng ta có thể sử dụng ước lượng điểm giống như ước lượng điểm đối với giá trị trung bình y . Từ $y = 60 + 5x$, thay $x = 10$, ta có $y = 60 + 5 \times 10 = 110$ m³/ha.

5.1.7.2. Ước lượng khoảng

Ước lượng điểm không cung cấp một thông tin nào về độ chính xác có liên quan đến một ước lượng. Vì thế, chúng ta cần phải phát triển một ước lượng khoảng. Có hai kiểu ước lượng khoảng: (1) ước lượng khoảng tin cậy, (2) ước lượng khoảng dự đoán. Kiểu thứ nhất của ước lượng khoảng là ước lượng khoảng tin cậy của giá trị trung bình (y_{bq}) đối với giá trị cụ thể của x . Ước lượng khoảng dự đoán được sử dụng khi chúng ta mong muốn một ước lượng khoảng của một giá trị y nhất định đối với một giá trị x nhất định.

Ước lượng khoảng tin cậy đối với giá trị trung bình y

Hàm hồi quy ước lượng chỉ cung cấp một ước lượng điểm của giá trị trung bình y đối với x nhất định. Khi mô tả thủ tục ước lượng khoảng, chúng ta sẽ sử dụng những kí hiệu sau đây:

x_p = giá trị cụ thể của x

$E(y_p)$ = giá trị trung bình hoặc kỳ vọng của y tương ứng với x_p .

$y = b_0 + b_1x_p$ = ước lượng điểm của $E(y_p)$ khi $x = x_p$.

Khi sử dụng các kí hiệu trên đây để ước lượng trữ lượng rừng ở tuổi 10, chúng ta có $x_p = 10$, $E(y_p)$ biểu thị cho giá trị gỗ trung bình chưa biết đối với tất cả các quần thụ khi $x_p = 10$. Ước lượng của $E(y_p) = y = 60 + 5 \times 10 = 110$ m³/ha.

Nói chung, chúng ta không thể mong đợi $y = E(y_p)$. Nếu muốn biết y_p là một giá trị trung bình thực của $E(y_p)$, chúng ta cần ước lượng phương sai của y_p (kí hiệu là $s^2_{y_p}$).

Phương sai ước lượng của y_p

$$s^2_{y_p} = s^2 \left\{ \frac{1}{n} + \frac{(x_p - x_{bq})^2}{\sum (x_i - x_{bq})^2} \right\} \quad (5.19)$$

Sai tiêu chuẩn ước lượng của y_p (Estimated Standard Deviation of y_p)

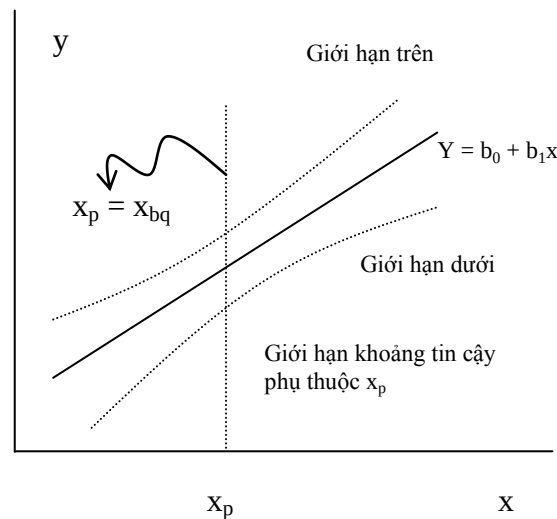
$$s_{y_p} = s \sqrt{\frac{1}{n} + \frac{(x_p - x_{bq})^2}{\sum (x_i - x_{bq})^2}} \quad (5.20)$$

Ước lượng khoảng tin cậy của $E(y_p)$ (Confidence Interval Estiamte of $E(y_p)$)

$$y_p \pm t_{\alpha/2(n-2)} s_{y_p} \quad (5.21)$$

Ở công thức 5.21, hệ số tin cậy là $1 - \alpha$, $t_{\alpha/2}$ dựa trên phân bố t với $n - 2$ độ tự do, $n =$ dung lượng quan sát.

Khi $\alpha = 0,05$ thì $1 - \alpha = 1 - 0,05 = 0,95$ hay 95%. Chúng ta có thể sử dụng công thức (5.21) để xây dựng khoảng tin cậy 95%. Ví dụ: $y_p = 100$, $\alpha/2 = 0,025$, độ tự do (df) = $10 - 2 = 8$, $s_{y_p} = 4,95$. Từ Bảng phân bố t, với $\alpha/2 = 0,025$, df = 8, $t_{\alpha/2} = 2,306$ và $t_{\alpha/2} s_{y_p} = 2,306(4,95) = 11,415$, thì ước lượng khoảng tin cậy 95% là $110 \pm 11,415$.



Hình 5.3. Khoảng tin cậy của số trung bình y

Lưu ý rằng, s_{y_p} nhỏ nhất khi $x_p = x_{bq}$ hay $s_{y_p} = s \sqrt{\frac{1}{n}}$. Như vậy, ước lượng số trung bình tốt nhất của y khi $x_p = x_{bq}$. Rõ ràng là khi x_p càng xa x_{bq} thì $x_p - x_{bq}$ càng lớn. Điều đó cho thấy, khoảng tin cậy của giá trị trung bình y càng rộng khi x_p càng xa x_{bq} . Hình 5.3 mô tả khoảng tin cậy đối với số trung bình y .

Ước lượng khoảng dự đoán đối với giá trị y nhất định

Để xây dựng khoảng ước lượng dự đoán, trước hết chúng ta phải xác định phương sai tương ứng với việc sử dụng y_p thay thế cho một ước lượng của giá trị y nhất định khi $x = x_p$. Phương sai này bao gồm hai thành phần: (1) phương sai của giá trị y nhất định theo số trung bình $E(y)$

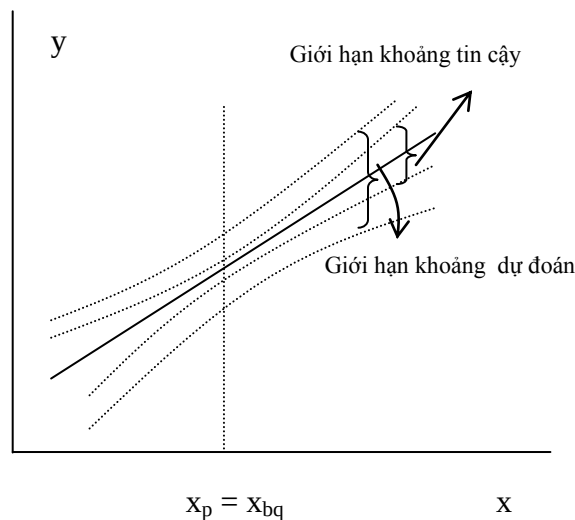
mà ước lượng của nó là s^2 ; (2) phương sai có quan hệ với việc sử dụng y_p để ước lượng $E(y_p)$ mà ước lượng của nó là $s^2 y_p$.

Phương sai của giá trị y_i riêng biệt (kí hiệu s^2_x):

$$\begin{aligned} s^2_x &= s^2 + s^2 y_p = s^2 + s^2 \left\{ \frac{1}{n} + \frac{(x_p - x_{bq})^2}{\sum (x_i - x_{bq})^2} \right\} \\ &= s^2 \left\{ 1 + \frac{1}{n} + \frac{(x_p - x_{bq})^2}{\sum (x_i - x_{bq})^2} \right\} \end{aligned} \quad (5.22)$$

Sai tiêu chuẩn ước lượng của giá trị y_i :

$$s^x = s \sqrt{1 + \frac{1}{n} + \frac{(x_p - x_{bq})^2}{\sum (x_i - x_{bq})^2}} \quad (5.23)$$



Hình 5.4. Khoảng tin cậy và khoảng dự đoán của y tương với giá trị x nhất định

Ước lượng khoảng dự đoán của y_p (Prediction Interval Estiamte of y_p)

$$y_p \pm t_{\alpha/2(n-2)} s^x \quad (5.24)$$

ở đây hệ số tin cậy là $1 - \alpha$, $t_{\alpha/2}$ dựa trên phân bố t với $n - 2$ độ tự do.

Cả hai ước lượng khoảng tin cậy và ước lượng khoảng dự đoán là tốt nhất khi $x_p = x_{bq}$. So với ước lượng khoảng tin cậy, ước lượng khoảng dự đoán của y_p có biên độ rộng hơn (**Hình 5.4**).

5.1.8. Phân tích các sai lệch

Các giả định của mô hình phù hợp

Sai lệch đối với quan sát i là hiệu số giữa giá trị quan sát của biến phụ thuộc y_i và giá trị ước lượng của biến phụ thuộc y_i . Nói khác đi, sai lệch thứ i là sai số thu được khi sử dụng hàm hồi quy ước lượng để dự đoán giá trị y_i .

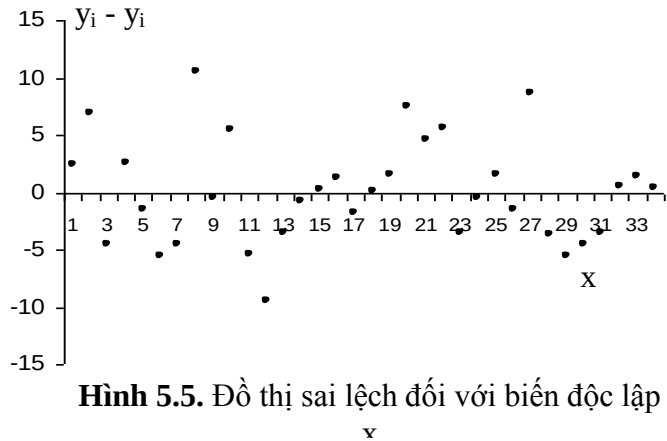
Sai lệch của quan sát i (Residual for Observation i)

$$y_i - \hat{y}_i \quad (5.25)$$

Mô hình hồi quy giả định

$$y = b_0 + b_1x + e \quad (5.26)$$

Trong mô hình hồi quy giả định, y là một hàm số tuyến tính của x với sai số e . Các giả định về e là cơ sở lý thuyết để xây dựng các kiểm định t và F nhằm xác định mối quan hệ giữa x với y có ý nghĩa hay không. Ngoài ra, các giả định về e cũng là cơ sở để xây dựng khoảng tin cậy và khoảng dự đoán cho một giá trị nhất định của y . Nếu các giả định về e bị vi phạm thì những kiểm định về mức ý nghĩa và việc xây dựng khoảng tin cậy và khoảng dự đoán cho một giá trị nhất định của y là không chính xác. Vì thế, các sai lệch cung cấp thông tin tốt nhất về sai số e . Điều đó có nghĩa là việc phân tích các sai lệch là một bước quan trọng trong việc xác định các giả định về e có phù hợp hay không. Thông thường các sai lệch được mô tả bằng đồ thị. Chúng ta sẽ thảo luận các loại đồ thị sai lệch sau đây: (1) đồ thị sai lệch đối với biến độc lập x ; (2) đồ thị sai lệch đối với biến dự đoán y ; (3) đồ thị các sai lệch chuẩn; (4) đồ thị xác suất chuẩn.



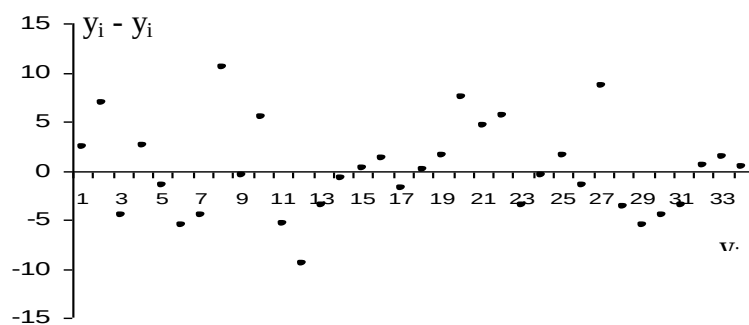
Hình 5.5. Đồ thị sai lệch đối với biến độc lập

Đồ thị sai lệch đối với biến độc lập x

Khi vẽ đồ thị sai lệch đối với biến độc lập x , trục hoành đặt giá trị x , còn trục tung đặt sai lệch $y_i - y_i$ (Hình 5.5). Mô hình phù hợp nhất khi các sai lệch $y_i - y_i$ nằm trên cùng một đường thẳng đi qua zero.

Đồ thị sai lệch đối với biến dự đoán y

Khi vẽ đồ thị sai lệch đối với biến y , trục hoành đặt giá trị y_i , còn trục tung là $y_i - y_i$ (Hình 5.6). Mô hình phù hợp nhất khi các sai lệch $y_i - y_i$ nằm trên cùng một đường thẳng đi qua zero.



Hình 5.6. Đồ thị sai lệch đối với biến y

Đồ thị các sai lệch chuẩn

Theo phương pháp bình phương nhỏ nhất, giá trị trung bình của các sai lệch bằng zero. Như vậy, khi chia mỗi sai lệch cho sai tiêu chuẩn của nó thì chúng ta nhận được sai lệch chuẩn. Như đã biết, sai tiêu chuẩn của sai lệch i ($s_{y_i - \hat{y}_i}$) phụ thuộc vào sai số chuẩn ước lượng s và giá trị của biến độc lập x_i .

Sai tiêu chuẩn của sai lệch thứ i (Standard Deviation of i th Residual)

$$s_{y_i - \hat{y}_i} = s \sqrt{1 - h_i} \quad (5.27)$$

ở đây $s_{y_i - \hat{y}_i}$ = sai tiêu chuẩn của sai lệch thứ i ;

s = sai số ước lượng chuẩn.

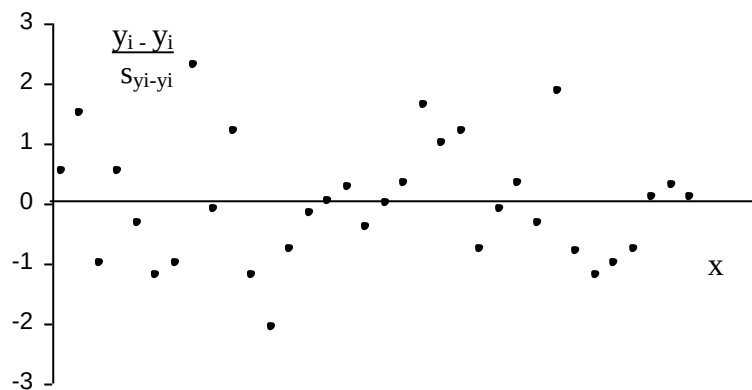
h_i = ảnh hưởng của quan sát i (Leverage of i th Observation)

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x}_{bq})^2}{\sum (x_i - \bar{x}_{bq})^2} \quad (5.28)$$

Sai lệch chuẩn của quan sát thứ i (Standardized Residual of i th Observation)

$$\frac{y_i - \hat{y}_i}{s_{y_i - \hat{y}_i}} \quad (5.29)$$

Khi vẽ đồ thị sai lệch chuẩn đối với biến độc lập x , trục hoành đặt giá trị x , còn trục tung là sai lệch chuẩn của quan sát thứ i $(y_i - \hat{y}_i)/s_{y_i - \hat{y}_i}$ (Hình 5.7). Các sai lệch chuẩn nằm trong giới hạn từ - 2 đến + 2. Những quan sát có sai lệch chuẩn nằm ngoài giới hạn từ - 2 đến + 2 là những quan sát có ảnh hưởng lớn đến hàm hồi quy ước lượng. Nếu sai số e phân bố chuẩn thì trị trung bình các sai lệch chuẩn bằng zero.



Hình 5.7. Đồ thị sai lệch chuẩn

Đồ thị xác suất chuẩn

Đồ thị xác suất chuẩn là một cách khác để xác định tính đúng đắn của giả định về phân bố chuẩn của sai số e . Để vẽ đồ thị xác suất chuẩn, chúng ta cần phải biết các điểm chuẩn (Normal Scores). Những giá trị này được các nhà thống kê tính toán và ghi lại trong các sách thống kê tùy theo số mẫu quan sát.

Khi vẽ đồ thị xác suất chuẩn, trục hoành đặt giá trị chuẩn, còn trục tung là các sai lệch chuẩn $(y_i - \hat{y}_i)/s_{y_i - \hat{y}_i}$. Nếu sai lệch chuẩn phù hợp với phân bố chuẩn thì các điểm tọa độ phải phân bố trên đường thẳng có độ dốc bằng 45° .

Một cách khác để vẽ đồ thị xác suất chuẩn là đặt y_i trên trục tung, còn trục hoành là phần trăm mẫu (Hình 5.8). Trong cả hai trường hợp ấy, nếu các điểm tọa độ phân tán xa đường thẳng có độ dốc bằng 45° thì các sai lệch phân bố không chuẩn. Điều đó có nghĩa là mô hình giả định không phù hợp.

Ngày nay những loại đồ thị sai lệch trên đây được xác định rất nhanh chóng bằng các phần mềm thống kê như Excel, SPSS, Statgraphics, Minitab...

5.1.9. Phân tích các sai lệch: Các số ngoại lai và các quan sát ảnh hưởng

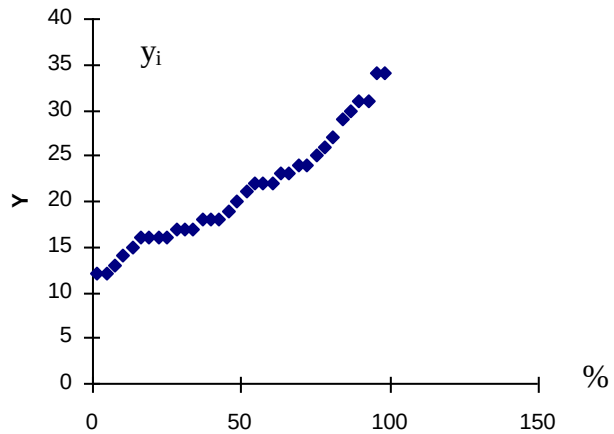
Phần này chúng ta xem xét việc phân tích các sai lệch được sử dụng để xác định các số ngoại lai (Outliers) và những quan sát có ảnh hưởng (Leverage) đến hàm hồi quy ước lượng như thế nào.

(1) Những quan sát ngoại lai

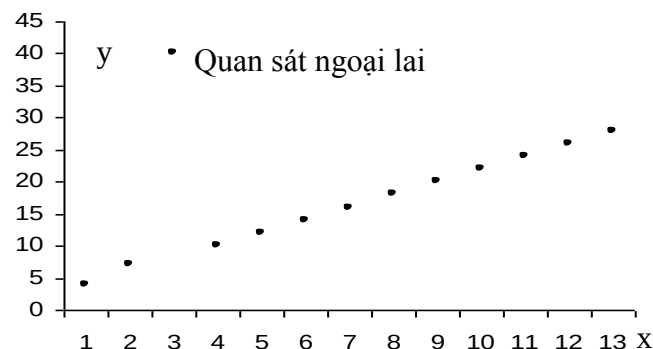
Những quan sát ngoại lai là những số không phù hợp với khuynh hướng biến đổi của các tập hợp quan sát khác hoặc nằm cách xa trị trung bình của các quan sát khác. Để loại bỏ các số ngoại lai, chúng ta có thể sử dụng đồ thị phân tán (Scatter diagram) hoặc các sai lệch chuẩn. Nếu một quan sát phân tán xa trị trung bình của các quan sát khác thì sai lệch chuẩn của nó sẽ lớn hơn về trị tuyệt đối. Nói chung, những giá trị quan sát có sai lệch chuẩn lớn hơn +2 và nhỏ hơn -2 là những số ngoại lai.

(2) Các quan sát ảnh hưởng

Đôi khi chúng ta gặp những quan sát ảnh hưởng lớn đến hàm hồi quy ước lượng. Hình 9.9 là đồ thị chỉ ra quan sát có ảnh hưởng. Khi loại bỏ trị số này thì các thông kê của hàm hồi quy ước lượng sẽ thay đổi. Những quan sát ảnh hưởng có thể là số ngoại lai. Số ngoại lai có thể do biến y hoặc biến x gây ra hoặc do cả hai gây ra.



Hình 5.8. Đồ thị xác suất chuẩn



Hình 5.9. Tập hợp số liệu với một quan sát ảnh hưởng

Vì những quan sát ảnh hưởng chỉ tác động mạnh đến hàm hồi quy ước lượng, nên chúng ta cần xem xét chúng thật cẩn thận. Trước hết, chúng ta kiểm tra lại cách ghi chép và đo đạc các biến số. Nếu sai sót do đo đạc thì phải thực hiện đo đạc lại. Đối với biến độc lập, những quan sát có ảnh hưởng lớn được gọi là ảnh hưởng cao (High Leverage = h_i). Chúng được xác định bằng cách so sánh với trị trung bình của tất cả các quan sát bình thường khác. Đối với biến x, ảnh hưởng lớn (Leverage) của quan sát i được xác định theo công thức:

$$h_i = \frac{1}{n} + \frac{(x_i - x_{bq})^2}{\sum (x_i - x_{bq})^2} \quad (5.30)$$

Từ **công thức 5.30** cho thấy, khi x càng xa x_{bq} thì h_i càng lớn. Thông thường nếu một quan sát có $h_i > 3(p + 1)/n$ thì quan sát ấy có ảnh hưởng lớn. Ở đây p là số biến độc lập, (đối với hồi quy tuyến tính đơn $p = 1$), n là dung lượng mẫu. Nhưng những quan sát gây ra bởi sự tương tác giữa các sai lệch lớn và ảnh hưởng lớn (h_i) thì rất khó bị loại bỏ. Để loại bỏ những trường hợp này, người ta sử dụng thủ tục chuẩn đoán (Diagnostic procedure). Khi một quan sát có ảnh hưởng lớn, thủ tục chuẩn đoán cho phép đưa cả hai (các sai lệch lớn và ảnh hưởng lớn) vào phân tích. Trong trường hợp này, chúng ta có thể sử dụng các số đo khoảng cách của Cook, Dfits và Mahalanobis.

Số đo khoảng cách của Cook (Cook's Distance Measure)

$$D_i = \frac{(y_i - y_{bq})^2}{(p - 1)s^2} \left[\frac{h_i}{(1 - h_i)^2} \right] \quad (5.31)$$

D_i = số đo khoảng cách của quan sát i

$y_i - y_{bq}$ = sai lệch của quan sát i

h_i = ảnh hưởng của quan sát i

p = số lượng biến độc lập

s = sai số ước lượng chuẩn.

Từ **công thức 5.31** cho thấy, một quan sát có ảnh hưởng lớn đến hàm hồi quy ước lượng nếu $y_i - y_{bq}$ và h_i đều lớn. Nói chung, khi $D_i > 1$ thì quan sát i là số ngoại lai. Dfits là số đo mức độ thay đổi các hệ số của hàm hồi quy ước lượng khi mỗi quan sát ảnh hưởng bị loại bỏ. Nói chung, các số đo quan sát ảnh hưởng (Cook, Dfits và Mahalanobis) có thể được tính toán dễ dàng bằng các phần mềm thống kê như SPSS, Statgraphics, Minitab...

5.2. PHÂN TÍCH TƯƠNG QUAN VÀ HỒI QUY ĐA BIẾN

5.2.1. Mô hình hồi quy đa biến

Phân tích hồi quy đa biến là nghiên cứu biến phụ thuộc y có quan hệ với 2 hay nhiều biến độc lập hay không. Trong trường hợp chung, chúng ta sử dụng p để chỉ số biến độc lập.

1. Mô hình hồi quy và hàm hồi quy đa biến

Mô hình hồi quy đa biến có dạng:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + e \quad (5.32)$$

$\beta_0, \beta_1, \beta_2 \dots \beta_p$ = các tham số;

e = sai số của mô hình.

Hàm hồi quy đa biến có dạng:

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (5.33)$$

Hàm hồi quy đa biến ước lượng có dạng:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p \quad (5.34)$$

2. Giải thích các hệ số của hồi quy đa biến

Trong hồi quy tuyến tính đơn, hệ số b_1 là ước lượng sự thay đổi của y theo một đơn vị thay đổi của x . Trong hồi quy đa tuyến tính, hệ số b_i là ước lượng sự thay đổi của y tương ứng với một đơn vị thay đổi của x_i khi tất cả các biến độc lập khác là hằng số (không đổi).

Ví dụ. Biến động chỉ số tăng trưởng đường kính của Thông ba lá (Y) có mối quan hệ với biến động của tổ hợp chỉ số nhiệt độ tháng 9 (T_9), chỉ số lượng mưa tháng 9 (M_9) và chỉ số giờ nắng tháng 9 (N_9) theo dạng: $Y = 8,117 + 0,058M_9 - 7,000T_9 - 1,172N_9$. Ở đây hệ số 0,058 là ước lượng sự gia tăng trung bình của chỉ số tăng trưởng đường kính Thông ba lá theo một đơn vị gia tăng chỉ số mưa khi chỉ số nhiệt độ và chỉ số giờ nắng tháng 9 là không thay đổi. Tương tự như thế, hệ số -7,0 là ước lượng sự suy giảm trung bình của chỉ số tăng trưởng đường kính Thông ba lá theo một đơn vị gia tăng chỉ số nhiệt độ khi chỉ số mưa và chỉ số giờ nắng tháng 9 là không thay đổi.

5.2.2. Hệ số xác định đa biến

Tương tự như hồi quy đơn, trong hồi quy đa biến SST, SSR và SSE có quan hệ như sau:

$$SST = SSR + SSE \quad (5.35)$$

SSE là tổng bình phương do sai số

$$SSE = \sum (y_i - \hat{y})^2 \quad (5.36)$$

SSR là tổng bình phương do hồi quy

$$SSR = \sum (y - y_{bq})^2 \quad (5.37)$$

SST là tổng bình phương toàn bộ

$$SST = \sum (y_i - y_{bq})^2 \quad (5.38)$$

Hệ số xác định đa biến

$$R^2 = \frac{SSR}{SST} \quad (5.39)$$

Hệ số tương quan đa biến của mẫu

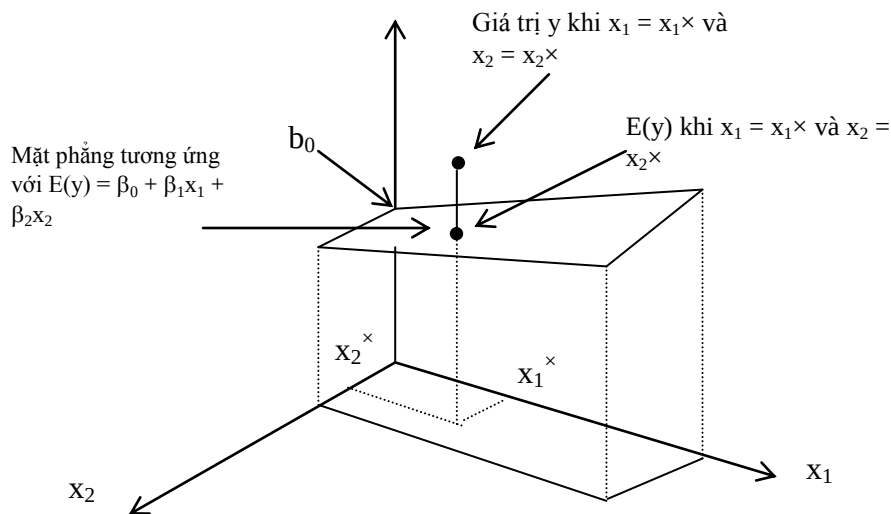
$$R = \sqrt{\frac{SSR}{SST}} = \sqrt{R^2} \quad (5.40)$$

Hệ số R^2 giải thích tỉ lệ biến thiên của biến phụ thuộc y có thể được giải thích bởi hàm hồi quy đa biến ước lượng. Giá trị $R^2 \times 100$ giải thích bao nhiêu phần trăm biến thiên của biến phụ thuộc y có thể được giải thích bởi hàm hồi quy đa biến ước lượng. Giá trị của R^2 luôn tăng lên theo số lượng biến độc lập (p) đưa vào mô hình. Hệ số tương quan đa biến giải thích cường độ quan hệ giữa y với các biến độc lập p .

Nhiều nhà thống kê còn sử dụng hệ số xác định điều chỉnh (kí hiệu R_a^2) để phân tích ảnh hưởng của số lượng biến độc lập đưa thêm vào mô hình đến biến động được giải thích bởi hàm hồi quy ước lượng. Với n là kích thước mẫu, p là số biến độc lập, hệ số R_a^2 được tính theo công thức:

$$R_a^2 = 1 - (1 - R^2) \frac{n - 1}{n - p - 1} \quad (5.41)$$

Từ công thức 5.41 cho thấy, khi p tăng lên thì R_a^2 giảm xuống.



Hình 5.10. Đồ thị hàm hồi quy để phân tích hồi quy đa biến với 2 biến phản hồi

5.2.3. Các giả định của mô hình hồi quy đa biến

Các giả định trong mô hình hồi quy đa tuyến tính

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + e \quad (5.42)$$

- (a) Sai số e là một biến ngẫu nhiên có $E(e) = 0$.
- (b) Phương sai của e (kí hiệu s^2) bằng nhau đối với tất cả các biến độc lập.
- (c) Các giá trị của e là độc lập.
- (d) Sai số e là một biến ngẫu nhiên có phân bố $N(0, s^2)$. Giá trị e biểu thị sai lệch giữa y_i với y_i , nghĩa là $e = y_i - \hat{y}_i$.

Trong phân tích hồi quy, thuật ngữ biến phản hồi (Response Variable) được dùng để thay thế cho biến độc lập. Ngoài ra, vì hồi quy đa biến tạo ra một mặt phẳng, nên đồ thị của nó được gọi là mặt phản hồi (Hình 5.10).

5.2.4. Kiểm định mức ý nghĩa

Trong hồi quy tuyến tính đơn, thống kê t và F được sử dụng để kiểm định sự tồn tại của mô hình hay của hệ số b_1 . Hai kiểm định t và F có ý nghĩa như nhau, nghĩa là cùng đưa ra những kết luận giống nhau. Nhưng trong hồi quy đa tuyến tính, thống kê t và F được sử dụng với ý nghĩa khác nhau.

- (a) Thống kê F được dùng để xác định mô hình hồi quy đa tuyến tính có tồn tại hay không. Nói khác đi, giữa y với các x_i có tồn tại mối quan hệ hay không.
- (b) Nếu kiểm định F chỉ ra mối quan hệ giữa y với các x_i tồn tại thực sự, thì tiếp theo kiểm định t sẽ được sử dụng để xác định từng biến độc lập có ý nghĩa hay không. Kiểm định t riêng biệt được xây dựng cho từng biến độc lập. Vì thế, mỗi kiểm định t là một kiểm định mức ý nghĩa riêng biệt.

1. Kiểm định F

Mô hình hồi quy đa biến có dạng

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + e$$

Các giả định đối với kiểm định F

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0.$$

$$H_a: \text{Một hoặc nhiều hệ số} \neq 0.$$

Nếu H_0 bị bác bỏ, chúng ta có thể kết luận mối quan hệ giữa y với tập hợp biến x_i là có ý nghĩa thống kê.

Khi xây dựng kiểm định F, chúng ta cần tính các thống kê sau đây:

Hồi quy bình phương trung bình

$$MSR = \frac{SSR}{p} \quad (5.43)$$

Sai số bình phương trung bình

$$s^2 = MSE = \frac{SSE}{n - p - 1} \quad (5.44)$$

Sai số chuẩn ước lượng

$$s = \sqrt{MSE} = \sqrt{\frac{SSE}{n - p - 1}} = \sqrt{s^2} \quad (5.45)$$

Thống kê kiểm định F

$$F = \frac{MSR}{MSE} \quad (5.46)$$

Quy tắc quyết định

Bác bỏ H_0 nếu $F > F_a$ ($a = 0,05$ hoặc $0,01$) hoặc $p\text{-value} < a$; trong đó F_a dựa trên phân bố F với p độ tự do ở tử số và $n - p - 1$ độ tự do ở mẫu số.

Trong phân tích hồi quy, kết quả của kiểm định F được chỉ ra ở Bảng ANOVA (Bảng 5.2).

Bảng 5.2. Bảng ANOVA đối với mô hình hồi quy đa biến

Nguồn biến động (Source)	Tổng bình phương (Sum of Squares)	Độ tự do (Degrees of Freedom)	Trung bình bình phương (Mean Square)	F
Hồi quy (Regression)	SSR	p	$MSR = \frac{SSR}{p}$	$\frac{MSR}{MSE}$
Sai số (Error)	SSE	$n - p - 1$	$MSE = \frac{SSE}{n - p - 1}$	
Tổng (Total)	SST	$n - 1$	MST	

2. Kiểm định t

Nếu kiểm định F chỉ ra mối quan hệ giữa y với các x_i tồn tại thực sự, thì tiếp theo kiểm định t sẽ được sử dụng để xác định từng biến độc lập có ý nghĩa hay không. Kiểm định t riêng biệt được xây dựng cho từng biến độc lập trong mô hình. Vì thế, mỗi kiểm định t là một kiểm định mức ý nghĩa riêng biệt.

Các giả định đối với kiểm định t

$$H_0: \beta_i = 0.$$

$$H_a: \beta_i \neq 0.$$

Thống kê kiểm định t

$$t = \frac{b_i}{s_{b_i}} \quad (5.47)$$

Sai tiêu chuẩn ước lượng của hệ số b_i

$$s_{b_i} = \frac{s}{\sqrt{\sum (x_{ij} - \bar{x}_{bj})^2}} \quad (5.48)$$

Quy tắc quyết định

Bác bỏ H_0 nếu $t > t_{a/2}$ hoặc $t < -t_{a/2}$ ($\alpha = 0,05$ hoặc $0,01$)

trong đó $t_{a/2}$ dựa trên phân bố t với $n - p - 1$ độ tự do.

Sử dụng p - value: Bác bỏ H_0 nếu p - value $< \alpha$.

5.2.5. Cộng tuyến tính

Chúng ta đã sử dụng biến độc lập x trong phân tích hồi quy để biểu thị cho một biến số nào đó được dùng để dự đoán hoặc giải thích giá trị của biến phụ thuộc y. Tuy vậy, điều đó không có nghĩa là các biến độc lập tự chúng là độc lập với nhau về mặt thống kê. Trên thực tế, hầu hết các biến số độc lập trong hồi quy đa biến có quan hệ với nhau ở một mức độ nào đó. Ví dụ: Trong mối quan hệ giữa chỉ số tăng trưởng đường kính thân cây (biến y) với các biến khí hậu (nhiệt độ, mưa, nắng...), chúng ta đã biết chỉ số nhiệt độ và chỉ số mưa, chỉ số nhiệt độ và chỉ số nắng là có quan hệ chặt chẽ với nhau.

Trong phân tích hồi quy đa biến, cộng tuyến tính biểu thị mối quan hệ giữa các biến độc lập với nhau. Nói chung, theo quy tắc kiểm định ảnh hưởng, cộng tuyến tính xuất hiện khi hệ số tương quan giữa hai biến độc lập lớn hơn hoặc bằng 0,7. Khi các biến độc lập có quan hệ chặt chẽ với nhau thì chúng ta không thể dễ dàng phân tích ảnh hưởng của từng biến độc lập riêng biệt đến biến phụ thuộc.

Hiện nay có nhiều phần mềm thống kê (SPSS, Minitab, Statgraphics...) cho phép thực hiện các thủ tục kiểm định cộng tuyến tính và chọn những biến độc lập hữu ích nhất để xây dựng hàm hồi quy ước lượng.

5.2.6. Sử dụng hàm hồi quy ước lượng để ước lượng và dự đoán

Các thủ tục ước lượng số trung bình y và dự đoán giá trị của y trong hồi quy đa biến cũng được thực hiện tương tự như trong phân tích hồi quy tuyến tính đơn.

Trong hồi quy đa biến, ước lượng điểm cho trị số kỳ vọng y đối với một giá trị nhất định của các biến độc lập cũng được thực hiện giống như ước lượng điểm của giá trị y nhất định. Trong cả hai trường hợp ấy, chúng ta đều sử dụng $y = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p$ như là một ước lượng điểm.

Để xây dựng ước lượng khoảng cho trị số kỳ vọng y và cho giá trị y riêng biệt đối với một giá trị nhất định của biến độc lập cũng được thực hiện giống như phân tích hồi quy tuyến tính một biến. Nói chung, các phần mềm thống kê (Excel, SPSS, Minitab, Statgraphics) đều cho phép tính nhanh các ước lượng khoảng cho trị số kỳ vọng y và cho giá trị y riêng biệt.

5.2.7. Các biến độc lập định tính

Trong nhiều trường hợp chúng ta phải làm việc với cả biến định lượng và biến định tính (những biến không thể đo đạc bằng các đơn vị vật lý nào đó). Ví dụ: Sinh trưởng của rừng có quan hệ chặt chẽ không chỉ với cường độ ánh sáng, nhiệt độ, lượng mưa, số giờ nắng, hàm lượng các chất khoáng trong đất, mà còn với sự có mặt của các loài cây khác, với loại địa Hình và đất, kiểu khí hậu, kiểu rừng... Cường độ ánh sáng, nhiệt độ, mưa, hàm lượng chất khoáng trong đất là những biến có thể đo đạc được. Nhưng sự có mặt của loài cây, loại đất, kiểu khí hậu, kiểu rừng là những biến không thể đo đạc được.

Trường hợp hai biến định tính

Giả sử cần xác định quan hệ giữa chiều cao của rừng Tách (y , m) với tuổi quần thụ (x_1 , năm) và loại đất (x_2). Loại đất bao gồm hai loại là đất đỏ nâu trên đá bazan (kí hiệu Bazan) và đất feralit đỏ vàng trên đá phiến sét (kí hiệu Fs).

Để chuyển các biến định tính thành kiểu số thích hợp trong mô hình, chúng ta có thể định nghĩa biến x_2 như sau:

$x_2 = 0$ nếu loại đất là Bazan

$x_2 = 1$ nếu loại đất là Fs

Trong phân tích hồi quy, biến số x_2 được gọi là biến giả hoặc biến chỉ định (Dummy or Indicator Variable). Từ ví dụ trên, bằng cách sử dụng biến giả, chúng ta có thể viết mô hình hồi quy đa biến như sau:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e$$

Bảng 5.3 ghi lại số liệu về sinh trưởng chiều cao của rừng Tách trong quan hệ với tuổi và loại đất, trong đó loại đất là một biến giả.

Bảng 5.3. Chiều cao của rừng Tách tương ứng với tuổi và loại đất

TT	H, m	Tuổi, năm (x_1)	Loại đất (x_2)	H	Tuổi, năm (x_1)	Loại đất (x_2)
1	3,7	2	0	2,2	2	1
2	7,5	4	0	6,2	4	1
3	11,3	6	0	9,8	6	1
4	13,3	8	0	12,1	8	1
5	14,7	10	0	13,8	10	1
6	16,4	12	0	15,2	12	1
7	17,7	14	0	16,7	14	1
8	19,4	16	0	18,4	16	1
9	20,6	18	0	19,3	18	1
10	21,5	20	0	20,4	20	1

Bằng thuật toán thống kê, chúng ta nhận được các kết quả phân tích tương quan và hồi quy như **Bảng 5.4 và 5.5**. Mô hình mối quan hệ có dạng:

$$H = 4,260 + 0,941x_1 - 1,200x_2$$

$$\text{với } R^2 = 0,937; s = \pm 1,37; t_{b1} = 17,6 (p = 0,000); t_{b2} = -1,96 (p = 0,06).$$

Ở mức ý nghĩa $\alpha = 0,05$, $p = 0,000$ với $F = 156,95$. Điều đó cho thấy giữa chiều cao với tuổi quần thụ và loại đất tồn tại mối quan hệ rất chặt chẽ. Kiểm định sự tồn tại của các hệ số b_1 và b_2 cho thấy $t_{b_1} = 17,6$ với $p = 0,000$; $t_{b_2} = -1,96$ với $p = 0,06$. Điều đó chứng tỏ x_1 tồn tại trong mô hình với ý nghĩa rất cao, còn x_2 cũng có thể được sử dụng trong mô hình nếu ta chấp nhận mức ý nghĩa $p = 0,10$. Giá trị $R^2 = 94,8\%$, $R_a^2 = 94,2\%$ chỉ ra mô hình hồi quy ước lượng được dùng để giải thích biến động của H . Mô hình $H = 4,260 + 0,941x_1 - 1,200x_2$ được sử dụng để ước lượng chiều cao quần thụ tương ứng với tuổi và loại đất khác nhau.

Bảng 5.4. Phân tích phương sai (ANOVA)

Nguồn	df	SS	MS	F	P
Hồi quy	2	591,50	295,8	156,95	0,000
Sai lệch	17	32,03	1,9		
Tổng	19	623,54			

Bảng 5.5. Phân tích hồi quy

	Hệ số	Sai số chuẩn	t	P-value	Cận dưới 95%	Cận trên 95%
b_0	4,260	0,731	5,830	0,000	2,718	5,802
b_1	0,941	0,053	17,609	0,000	0,828	1,054
b_2	-1,200	0,614	-1,955	0,067	-2,495	0,095

Giải thích các tham số

Hàm hồi quy đa biến bao gồm hai biến có dạng:

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \quad (5.49)$$

Để hiểu cách thức giải thích các tham số b_0 , b_1 và b_2 khi đưa vào mô hình một biến định tính, chúng ta xem xét trường hợp $x_2 = 0$. Nếu gọi $E(H/Bazan)$ là chiều cao trung bình của tập hợp các quần thụ Tách 2 - 20 tuổi trên đất Bazan, chúng ta có:

$$E(H/Fs) = b_0 + b_1 x_1 + b_2(0) = b_0 + b_1 x_1 \quad (5.50)$$

Tương tự như trên, khi thay $x_2 = 1$, chúng ta nhận được chiều cao trung bình của tập hợp các quần thụ Tách 2 - 20 tuổi trên đất Fs như sau:

$$\begin{aligned} E(H/Fs) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 \\ &= b_0 + b_1 x_1 + b_2(1) \\ &= (b_0 + b_2) + b_1 x_1 \end{aligned} \quad (5.51)$$

So sánh (5.50) và (5.51), chúng ta thấy chiều cao quần thụ kỳ vọng là một hàm tuyến tính của x_1 đối với cả hai loại đất. Độ dốc của cả hai hàm là b_1 , nhưng điểm chặn trên trục y có giá trị khác nhau. Nếu xét trường hợp H của quần thụ trên đất Bazan thì điểm chặn là b_0 . Ngược lại, nếu xét trường hợp H của quần thụ trên đất Fs thì điểm chặn là $b_0 + b_2$.

Như vậy, b_2 cho biết sự khác nhau giữa chiều cao kỳ vọng của các quần thụ Tách trên hai loại đất. Nếu $b_2 < 0$, $E(H/Bazan) > E(H/Fs)$. Khi $b_2 = 0$ thì sinh trưởng chiều cao của các quần thụ trên hai loại đất là không khác nhau. Từ ví dụ trên, chúng ta có:

$$\begin{aligned} &+ \text{khi } x_2 = 0 \\ &H = 4,260 + 0,941x_1; \end{aligned}$$

+ khi $x_2 = 1$
 $H = 3,060 + 0,941x_1$.

So sánh hai hàm này dễ dàng nhận thấy chiều cao trung bình của các quần thụ Tách mọc trên đất bazan lớn hơn chiều cao trung bình của các quần thụ Tách mọc trên đất feralít đỏ vàng phiến sét là 1,2 m ($4,260 - 3,060 = 1,2$ m). Hình 5.11 là đồ thị mô tả chiều cao của quần thụ Tách phụ thuộc vào tuổi và loại đất.

Tổ hợp nhiều biến định tính

Khi biến định tính được phân chia thành nhiều mức, chúng ta cần thận trọng trong việc định nghĩa và giải thích các biến giả. Về nguyên tắc, khi có k biến định tính, chúng ta cần mã hóa thành $k - 1$ biến giả; trong đó mỗi biến giả lấy giá trị 0 hoặc 1.

Ví dụ. Nghiên cứu sự khác biệt về chiều cao (H , m) của các quần thụ Tách 2 - 20 tuổi mọc trên 3 loại đất A, B và C. Nhà lâm nghiệp mong muốn sử dụng một hàm hồi quy chung để dự đoán chiều cao quần thụ. Giả sử chiều cao quần thụ phụ thuộc vào một số biến độc lập như tuổi, khí hậu, đất, biện pháp kinh doanh. Nhà lâm nghiệp nhận thấy rằng, trong một vùng khí hậu như nhau, chiều cao quần thụ phụ thuộc khá lớn vào loại đất. Bởi vì loại đất bao gồm ba mức A, B và C, do đó chúng ta cần có $3 - 1 = 2$ biến giả để chỉ các loại đất. Mỗi biến giả sẽ được định nghĩa như sau:

$$\begin{aligned} x_1 &= 1 \text{ nếu loại đất là B} \\ x_1 &= 0 \text{ nếu loại đất khác B (đất A và C)} \\ x_2 &= 1 \text{ nếu loại đất là C} \\ x_2 &= 0 \text{ nếu loại đất khác C (đất A và B)} \end{aligned}$$

Theo các định nghĩa như trên, chúng ta có các giá trị x_1 và x_2 như sau:

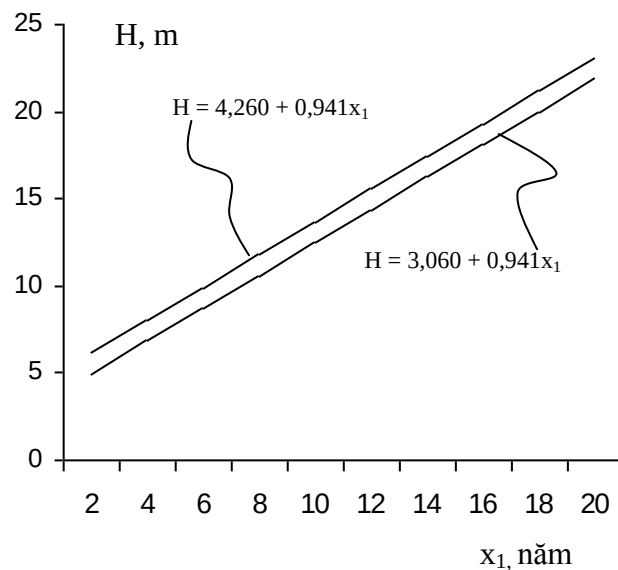
Loại đất	x_1	x_2
A	0	0
B	1	0
C	0	1

Như vậy, các quan sát ở loại đất A được mã hóa tương ứng $x_1 = 0$, $x_2 = 0$; các quan sát ở loại đất B được mã hóa tương ứng $x_1 = 1$, $x_2 = 0$; các quan sát ở loại đất C được mã hóa tương ứng $x_1 = 0$, $x_2 = 1$. Cách mã hóa như thế cho phép nhà lâm nghiệp so sánh từng cặp chiều cao của quần thụ trên loại đất B và C với chiều cao của quần thụ trên loại đất A.

Hàm hồi quy biểu thị quan hệ giữa chiều cao, $E(H)$, với các biến giả được viết dưới dạng sau đây:

$$E(H) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Để giải thích các tham số β_0 , β_1 và β_2 , chúng ta xem xét ba biến động của hàm hồi quy như sau:



Hình 5.11. Chiều cao lâm phần Tách trên hai loại đất

$$\begin{aligned} E(H/\text{loại đất A}) &= \beta_0 + \beta_1(0) + \beta_2(0) = b_0 \\ E(H/\text{loại đất B}) &= \beta_0 + \beta_1(1) + \beta_2(0) = b_0 + b_1 \\ E(H/\text{loại đất C}) &= \beta_0 + \beta_1(0) + \beta_2(1) = b_0 + b_2 \end{aligned}$$

Như vậy, β_0 là kỳ vọng hoặc trị số trung bình của H rừng Tách mọc trên loại đất A; $\beta_0 + b_1$ là trị số trung bình của H rừng Tách mọc trên loại đất B; $b_0 + b_2$ là trị số trung bình của H rừng Tách mọc trên loại đất C. Tham số b_1 là sự khác biệt về H trung bình của rừng Tách mọc trên loại đất A và B; còn b_2 là sự khác biệt về H trung bình của rừng Tách mọc trên loại đất A và C.

Chúng ta đã sử dụng hai biến giả, bởi vì đất bao gồm 3 loại khác nhau. Nhưng việc quy ước $x_1 = 0, x_2 = 0$ để chỉ loại đất A; $x_1 = 1, x_2 = 0$ để chỉ loại đất B và $x_1 = 0, x_2 = 1$ để chỉ loại đất C chỉ là quy ước tùy ý. Bởi vì, chúng ta vẫn có thể quy ước $x_1 = 1, x_2 = 0$ để chỉ loại đất A; $x_1 = 0, x_2 = 0$ để chỉ loại đất B và $x_1 = 0, x_2 = 1$ để chỉ loại đất C. Nếu quy ước như thế, chúng ta nhận được các hàm sau đây:

$$\begin{aligned} E(H/\text{loại đất A}) &= \beta_0 + \beta_1(1) + \beta_2(0) = b_0 + b_1 \\ E(H/\text{loại đất B}) &= \beta_0 + \beta_1(0) + \beta_2(0) = b_0 \\ E(H/\text{loại đất C}) &= \beta_0 + \beta_1(0) + \beta_2(1) = b_0 + b_2 \end{aligned}$$

Trong trường hợp này, b_1 giải thích sự khác nhau giữa chiều cao trung bình của rừng mọc trên loại đất A với rừng mọc trên loại đất B, còn b_2 giải thích sự khác nhau giữa chiều cao trung bình của rừng mọc trên loại đất C với rừng mọc trên loại đất B.

Khi $k = 4$ (A, B, C, D), ta có 3 biến giả như sau:

$$\begin{aligned} x_1 &= 1 \text{ nếu loại đất là B} \\ x_1 &= 0 \text{ nếu loại đất khác B (đất A, C, D)} \\ x_2 &= 1 \text{ nếu loại đất là C} \\ x_2 &= 0 \text{ nếu loại đất khác C (đất A, B, D)} \\ x_3 &= 1 \text{ nếu loại đất là D} \\ x_3 &= 0 \text{ nếu loại đất khác D (đất A, B, C)} \end{aligned}$$

Theo các định nghĩa như trên, chúng ta có các giá trị x_1 và x_2 như sau:

Loại đất	x_1	x_2	x_3
A	0	0	0
B	1	0	0
C	0	1	0
D	0	0	1

Từ đó chúng ta có các hàm sau đây:

$$\begin{aligned} E(H/\text{loại đất A}) &= \beta_0 + \beta_1(0) + \beta_2(0) + \beta_3(0) = b_0 \\ E(H/\text{loại đất B}) &= \beta_0 + \beta_1(1) + \beta_2(0) + \beta_3(0) = b_0 + b_1 \\ E(H/\text{loại đất C}) &= \beta_0 + \beta_1(0) + \beta_2(1) + \beta_3(0) = b_0 + b_2 \\ E(H/\text{loại đất D}) &= \beta_0 + \beta_1(0) + \beta_2(0) + \beta_3(1) = b_0 + b_3 \end{aligned}$$

Như vậy so sánh với chiều cao trung bình của rừng Tách trên loại đất A, chiều cao trung bình của rừng Tách trên loại đất B, C và D khác nhau tương ứng là b_1, b_2 và b_3 .

5.2.8. Phân tích các sai lệch

Sai lệch của quan sát thứ i được tính theo công thức:

$$y_i - \hat{y}_i \quad (5.52)$$

Sai tiêu chuẩn của sai lệch thứ i được tính theo công thức:

$$s_{y_i - y_i} = s\sqrt{1 - h_i} \quad (5.53)$$

Sai lệch chuẩn đối với quan sát thứ i được tính theo công thức:

$$\frac{y_i - \bar{y}_i}{s_{y_i - y_i}} \quad (5.54)$$

Trong các công thức trên, s là sai số chuẩn ước lượng; h_i là ảnh hưởng của quan sát i . Trong phân tích hồi quy đa biến, việc tính toán các h_i , $s_{y_i - y_i}$ và sai lệch chuẩn của quan sát i là khá phức tạp. Những tham số này có thể được xác định bằng các phần mềm thống kê như Excel, SPSS, Minitab, Statgraphics ...

Khi vẽ đồ thị sai lệch chuẩn, người ta đặt y trên trục hoành, còn sai lệch chuẩn trên trục tung. Các sai lệch chuẩn nằm trong giới hạn từ -2 đến $+2$. Những quan sát có sai lệch nằm ngoài giới hạn từ -2 đến $+2$ là những quan sát có ảnh hưởng lớn đến hàm hồi quy ước lượng.

Đồ thị xác suất chuẩn cũng được sử dụng để xác định phân bố chuẩn của e . Khi vẽ đồ thị xác suất chuẩn, trục hoành đặt giá trị chuẩn, còn trục tung là các sai lệch chuẩn $(y_i - \bar{y}_i)/s_{y_i - y_i}$. Nếu sai lệch chuẩn phù hợp với phân bố chuẩn thì các điểm tọa độ phải phân bố trên đường thẳng có độ dốc bằng 45° . Một cách khác để vẽ đồ thị xác suất chuẩn là đặt y_i trên trục tung, còn trục hoành là phần trăm mẫu. Trong cả hai trường hợp ấy, nếu các điểm tọa độ phân tán xa đường thẳng có độ dốc bằng 45° thì các sai lệch phân bố không chuẩn. Điều đó có nghĩa là mô hình giả định không phù hợp. Ngày nay những loại đồ thị sai lệch trên đây được xác định rất nhanh chóng bằng các phần mềm thống kê như Excel, SPSS, Statgraphics, Minitab...

Những quan sát ngoại lai

Những quan sát ngoại lai là những số không phù hợp với khuynh hướng biến đổi của các tập hợp quan sát khác. Để loại bỏ các số ngoại lai, chúng ta có thể sử dụng đồ thị phân tán (Scatter diagram) hoặc các sai lệch chuẩn. Nếu một quan sát phân tán xa các quan sát khác thì các sai lệch chuẩn của nó sẽ lớn hơn về trị tuyệt đối. Nói chung, những giá trị quan sát có sai lệch chuẩn lớn hơn $+2$ và nhỏ hơn -2 là những số ngoại lai.

Sự có mặt của các số ngoại lai sẽ làm tăng sai số chuẩn ước lượng (s) và sai tiêu chuẩn của sai lệch i ($s_{y_i - y_i}$). Khi $s_{y_i - y_i}$ tăng lên thì sai lệch chuẩn của quan sát i sẽ giảm. Điều đó có thể dẫn đến sự vi phạm quy tắc sai lệch chuẩn để xác định một quan sát là ngoại lai. Chúng ta có thể loại bỏ khó khăn này bằng cách sử dụng các sai lệch chuẩn Student hóa (Studentized deleted Residuals).

Các sai lệch và số ngoại lai được loại bỏ bằng Student hóa

Giả sử quan sát thứ i bị loại bỏ từ dãy dữ liệu có n quan sát và một hàm hồi quy mới được ước lượng từ $n - 1$ quan sát còn lại. Chúng ta kí hiệu $s_{(i)}$ là sai số chuẩn ước lượng dựa trên $n - 1$ quan sát. Nếu thay $s_{(i)}$ vào công thức $s_{y_i - y_i} = s\sqrt{1 - h_i}$, sau đó tính lại sai lệch chuẩn của quan sát i , thì kết quả nhận được sai lệch chuẩn Student hóa. Nếu quan sát thứ i là số ngoại lai thì $s_{(i)} < s$ và $|s_{(i)}| > |s|$. Bằng cách này có thể xác định được các số ngoại lai mà các sai lệch chuẩn không thể nhận ra được. Các sai lệch chuẩn Student hóa có thể được xác định rất nhanh chóng bằng các phần mềm thống kê SPSS, Statgraphics, Minitab...

Phân bố t cũng có thể được sử dụng để xác định các sai lệch chuẩn Student hóa. Nếu giá trị sai lệch chuẩn Student hóa nhỏ hơn $-t_{\alpha/2}$ hoặc lớn hơn $t_{\alpha/2}$ với $(n - 1) - p - 1$ độ tự do thì quan sát i là số ngoại lai.

Những quan sát ảnh hưởng

Những quan sát có ảnh hưởng đến hàm hồi quy ước lượng khi $h_i > 3(p + 1)/n$. Đối với hồi quy bậc nhất, $h_i > 6/n$. Tuy vậy, trong một số trường hợp giá trị h_i của quan sát i có thể lớn hơn h_i tiêu chuẩn $(3(p + 1)/n)$, nhưng quan sát ấy có thể ảnh hưởng không đáng kể đến hàm hồi quy ước lượng. Trong trường hợp ấy, việc sử dụng khi h_i để kết luận có thể phạm phải sai lầm. Nói chung, để loại bỏ những trường hợp này, người ta sử dụng thủ tục chuẩn đoán (Diagnostic procedure). Khi một mẫu bị ảnh hưởng lớn, thủ tục chuẩn đoán cho phép đưa cả hai (các sai lệch lớn và ảnh hưởng lớn) vào phân tích. Thủ tục chuẩn đoán bao gồm một số số đo như số đo khoảng cách của Cook, số đo khoảng cách của Mahalanobis, Dfits.

Số đo khoảng cách của Cook (Cook's Distance Measure)

$$D_i = \frac{(y_i - \hat{y}_i)^2}{(p - 1)s^2} \left[\frac{h_i}{(1 - h_i)^2} \right] \quad (5.55)$$

D_i = số đo khoảng cách của quan sát i

$y_i - \hat{y}_i$ = sai lệch của quan sát i

h_i = ảnh hưởng của quan sát i

p = số lượng biến độc lập

s = sai số ước lượng chuẩn.

Từ công thức 5.55 cho thấy, một quan sát có ảnh hưởng lớn đến hàm hồi quy ước lượng nếu $y_i - \hat{y}_i$ và h_i đều lớn. Nói chung, khi $D_i > 1$ thì quan sát i là số ngoại lai. Dfits là số đo mức độ thay đổi các hệ số của hàm hồi quy ước lượng khi mỗi quan sát ảnh hưởng bị loại bỏ. Nói chung, các số đo những quan sát ảnh hưởng (Cook, Dfits và Mahalanobis) có thể được tính toán dễ dàng bằng các phần mềm thống kê SPSS, Statgraphics, Minitab...

5.2.9. Phân tích vai trò của mỗi biến trong mô hình

Một câu hỏi đặt ra: Việc thêm vào hoặc loại bỏ một biến khỏi mô hình hồi quy đa biến có làm giảm tổng bình phương sai số (SSE) một cách có ý nghĩa thống kê hay không? Nói chung, để trả lời câu hỏi này người ta sử dụng thống kê F. Một cách tổng quát, thống kê F được viết như sau:

$$F = \frac{\frac{SSE(x_1, x_2, \dots, x_q) - SSE(x_1, x_2, \dots, x_q, x_{q+1}, \dots, x_p)}{(p - q)}}{\frac{SSE(x_1, x_2, \dots, x_q, x_{q+1}, \dots, x_p)}{(n - p - 1)}} \quad (5.56)$$

$q < p$ = số biến độc lập

q = số biến độc lập đưa vào mô hình ban đầu

$p - q$ = số biến thêm vào mô hình

Khi $q = 1$, $p = 2$, thì thống kê F có dạng

$$F = \frac{\frac{SSE(x_1) - SSE(x_1, x_2)}{1}}{\frac{SSE(x_1, x_2)}{(n - p - 1)}} \quad (5.57)$$

Giả thuyết H_0 : Các biến x_i thêm vào mô hình là không có ý nghĩa

Quy tắc quyết định

Bác bỏ H_0 : Nếu $F > F_a$ ($\alpha = 0,05$ hoặc $0,01$)

Sử dụng P - value. Bác bỏ H_0 khi p - value $< \alpha$ ($\alpha = 0,05$ hoặc $0,01$)

Nói chung, khi có nhiều biến độc lập đưa vào mô hình ($p \gg 6$), thủ tục chọn biến được thực hiện như sau:

- + Tính ma trận hệ số tương quan đơn giữa y với từng x_i và giữa các x_i với nhau. Tại bước này, hãy quan sát ma trận tương quan đơn giữa từng cặp biến x_i . Nếu các cặp x_i nào có quan hệ chặt với nhau ($r > 0,7$) thì chúng có hiện tượng cộng tuyến tính.
- + Chọn những biến độc lập có quan hệ chặt chẽ với y để xây dựng mô hình.

Trên thực tế, việc xây dựng mô hình bao gồm nhiều biến là hết sức phức tạp. Nhưng công việc này trở nên dễ dàng hơn khi sử dụng phương pháp đưa dần từng biến vào (Forward Selection) và phương pháp chọn biến từng bước (Stepwise selection) trong các phần mềm phân tích hồi quy đa biến của Minitab, SPSS và Statgraphics.

5.2.10. Tự tương quan và kiểm định Durbin - Watson

Khi tập hợp dữ liệu là một dãy số sắp xếp theo thời gian (ví dụ, chuỗi các vòng năm) thì cần phải xem xét vấn đề tự tương quan. Trong trường hợp này giá trị y_t (số đo vào thời gian t) có thể có quan hệ với $y_{t-1}, y_{t-2}, \dots$. Nếu y_t có quan hệ với $y_{t-1}, y_{t-2}, \dots$ thì người ta gọi tương ứng là tương quan trật tự (thứ nhất, thứ hai, thứ ba...).

Trong phân tích tự tương quan, các sai số e là không độc lập, nghĩa là sai số đo ở thời gian t có quan hệ với sai số đo ở thời gian trước đó.

Tự tương quan có thể dương hoặc âm. Tự tương quan lấy giá trị < 1 . Tự tương quan dương xuất hiện khi sai lệch dương hoặc âm ở thời kỳ này kéo theo sai lệch dương hoặc âm ở thời kỳ khác. Tự tương quan âm xuất hiện khi sai lệch dương ở thời kỳ này kéo theo sai lệch âm ở thời kỳ khác.

Thông kê Durbin - Watson cho phép dò tìm hiện tượng tự tương quan trong các số đo theo chuỗi thời gian.

Giả thiết e là không độc lập, nhưng có quan hệ như sau:

$$e_t = \rho e_{t-1} + z_t \quad (5.58)$$

trong đó $|\rho| < 1$; z_t = biến độc lập có phân bố $N(0, s^2)$.

Từ công thức 5.58 ta thấy:

+ Nếu $\rho = 0$ thì các kiểu sai số không có quan hệ với nhau và mỗi sai số có trung bình bằng zero và phương sai s^2 . Trong trường hợp này không có tự tương quan và các giả thuyết hồi quy được chấp nhận.

+ Nếu $\rho > 0$ thì các biến số có tương quan dương.

+ Nếu $\rho < 0$ thì các biến số có tương quan âm.

+ Khi ρ tồn tại thì các giả thuyết hồi quy bị vi phạm.

Thông kê Durbin - Watson sử dụng các sai lệch để kiểm định tự tương quan. Gọi $e_i = y_i - \hat{y}_i$, ta có:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad (5.59)$$

Giá trị d thay đổi từ 0 đến 4. Giá trị $d = 2$ cho biết không có tự tương quan.

Các giả định

$H_0: \rho = 0$; $H_a: \rho < 0$ hoặc $\rho > 0$.

Quy tắc quyết định

$d < d_L \Rightarrow$ tương quan dương

$d_L < d < d_U \Rightarrow$ kiểm định không thuyết phục

$d > d_U \Rightarrow$ không tồn tại tự tương quan dương

$d > 4 - d_L \Rightarrow$ tương quan âm

$4 - d_U \leq d \leq 4 - d_L \Rightarrow$ kiểm định không thuyết phục

$d < 4 - d_U \Rightarrow$ không tồn tại tự tương quan âm

Tóm lại, khi $d < d_L$ hoặc $d > 4 - d_L$ thì tự tương quan xuất hiện. Ở đây d_L và d_U tương ứng là giới hạn dưới và giới hạn trên của kiểm định Durbin - Watson. Bảng kiểm định Durbin - Watson được xây dựng cho các mức ý nghĩa $\alpha = 0,05$, $\alpha = 0,025$ và $\alpha = 0,01$. Nói chung, việc phân tích tự tương quan và xác định thống kê Durbin - Watson được thực hiện dễ dàng hơn khi sử dụng các phần mềm Minitab, SPSS và Statgraphics.

5.2.11. Sử dụng hồi quy đa biến để phân tích biến động và thiết kế thí nghiệm

Chúng ta đã xem xét việc sử dụng biến giả trong phân tích hồi quy. Phần này chúng ta sử dụng các biến giả để phân tích biến động và các vấn đề thí nghiệm.

Giả sử chúng ta có 3 nghiệm thức, mỗi nghiệm thức bao gồm 6 quan sát; tổng cộng có 18 quan sát. Dấu hiệu quan sát là chiều cao trung bình (H, cm) của cây con Sao đen 6 tháng tuổi (**Bảng 5.6**). Bây giờ chúng ta cần kiểm định giả thuyết H_0 cho rằng chiều cao của cây con trong các nghiệm thức là bằng nhau.

Vì có 3 nghiệm thức, chúng ta cần có $3 - 1 = 2$ biến giả. Các biến giả kí hiệu tương ứng x_1 và x_2 . Các biến giả x_1 và x_2 nhận các giá trị sau đây:

Biến giả	A	B	C
x_1	0	1	0
x_2	0	0	1

Từ hai biến x_1 và x_2 , chúng ta có thể lập được mối quan hệ giữa H với x_1 và x_2 :

$$E(H) = b_0 + b_1x_1 + b_2x_2$$

Nếu chúng ta quan tâm đến chiều cao kỳ vọng của các nghiệm thức, chúng ta cần xây dựng các hàm hồi quy đa biến như sau:

Đối với nghiệm thức A

$$E(H/A) = b_0 + b_1(0) + b_2(0) = b_0$$

Đối với nghiệm thức B

$$E(H/B) = b_0 + b_1(1) + b_2(0) = b_0 + b_1$$

Đối với nghiệm thức C

$$E(H/C) = b_0 + b_1(0) + b_2(1) = b_0 + b_2$$

Tham số b_0 chính là chiều cao trung bình của nghiệm thức A; tham số $b_0 + b_1$ là chiều cao trung bình của nghiệm thức B; tham số $b_0 + b_2$ là chiều cao trung bình của nghiệm thức C.

Bảng 5.6. Chiều cao (H, cm) của Sao đen 6 tháng tuổi

Thí nghiệm A			Thí nghiệm B			Thí nghiệm C		
x ₁	x ₂	H (cm)	x ₁	x ₂	H (cm)	x ₁	x ₂	H (cm)
0	0	85	1	0	71	0	1	59
0	0	75	1	0	75	0	1	64
0	0	82	1	0	73	0	1	62
0	0	76	1	0	74	0	1	69
0	0	71	1	0	69	0	1	75
0	0	85	1	0	82	0	1	67

Bảng 5.7. Phân tích hồi quy

	Hệ số	Sai số chuẩn	t	P-value
b ₀	79	2,1858	36,1422	0,000
b ₁	-5	3,0912	-1,6175	0,127
b ₂	-13	3,0912	-4,2055	0,001

Bảng 5.8. Phân tích biến động

	df	SS	MS	F	P
Hồi quy	2	516	258,00	9,0	0,003
Sai lệch	15	430	28,67		
Tổng	17	946			

Từ số liệu của **Bảng 5.6**, bằng thuật toán thống kê chúng ta nhận được Bảng phân tích biến động và hồi quy (**Bảng 5.7 và 5.8**). Từ **Bảng 5.7**, chúng ta có thể xây dựng hàm hồi quy biểu diễn ảnh hưởng của ba yếu tố thí nghiệm đến chiều cao trung bình E(H) của Sao đen 6 tháng tuổi như sau:

$$E(H) = 79 - 5x_1 - 13x_2; \text{ với } R^2 = 54,5\%; R = 0,74, s = 5,35.$$

Đối với thí nghiệm A

$$E(H/A) = 79 - 5(0) - 13(0) = 79$$

Đối với thí nghiệm B

$$E(H/B) = 79 - 5(1) - 13(0) = 74$$

Đối với thí nghiệm C

$$E(H/C) = 79 - 5(0) - 13(1) = 66$$

Bây giờ chúng ta cần kiểm định sự khác nhau giữa ba số trung bình (79, 74 và 66) của ba thí nghiệm.

Giả thuyết

$$H_0: E(H/B) = E(H/A) \text{ hay } E(H/B) - E(H/A) = 0$$

$$E(H/C) = E(H/A) \text{ hay } E(H/C) - E(H/A) = 0$$

Vì $b_0 = E(H/A)$ và $b_0 + b_1 = E(H/B)$, nên sự khác biệt giữa thí nghiệm A và B là b_1 . Tương tự, vì $b_0 + b_2 = E(H/C)$, nên giữa C với A khác nhau một lượng bằng b_2 . Chúng ta chỉ có thể kết luận ba số trung bình bằng nhau khi $b_1 = b_2 = 0$. Vì thế, giả thuyết cũng có thể đặt ra như sau:

$$H_0: b_1 = b_2 = 0$$

$$H_a: b_1 \neq b_2$$

Thống kê F

$$F = \frac{MSR}{MSE}$$

Quy tắc quyết định

Ở mức ý nghĩa $\alpha = 0,05$, bác bỏ H_0 khi $F > F_\alpha$.

Sử dụng p - value. Bác bỏ H_0 khi p - value $< \alpha$.

Ở ví dụ trên, $F = 9 > F(2,15) = 3,68$ (p - value = 0,003 $< 0,05$), nên giả thuyết H_0 bị bác bỏ.

5.3. PHÂN TÍCH TƯƠNG QUAN VÀ HỒI QUY GIỮA HAI BIẾN

Quan hệ giữa hai biến số có thể tồn tại ở dạng tuyến tính hoặc dạng phi tuyến tính. Thủ tục ước lượng quan hệ phi tuyến tính của SPSS 10.0 cung cấp các thống kê hồi quy ước lượng và các đồ thị quan hệ đối với 11 mô hình hồi quy ước lượng khác nhau. Mỗi mô hình được xây dựng cho một biến độc lập. Bạn cũng có thể sử dụng thủ tục save để lưu các giá trị dự đoán (Predicted values), các sai lệch (residuals) và các khoảng ước lượng (intervals) như những biến mới.

Các thống kê (Statistics)

Đối với mỗi mô hình, chương trình sẽ cung cấp cho bạn các thống kê sau đây: hệ số tương quan (multiple R) và hệ số xác định hay bình phương của hệ số tương quan (R^2); bình phương của hệ số tương quan điều chỉnh (Adjusted R^2); hệ số hồi quy (Regression coefficients); sai số chuẩn của ước lượng (Standard error of the estimate); Bảng phân tích biến động (Analysis-of-variance table); các giá trị dự đoán (Predicted values); các sai lệch giữa giá trị thực tế và giá trị của hàm hồi quy ước lượng (Residuals); các khoảng tin cậy của các giá trị dự đoán (Prediction intervals)...

Các mô hình (Models)

Chương trình SPSS 10.0 cung cấp sẵn 11 mô hình sau đây: tuyến tính hay đường thẳng (Linear), Logarit (Logarithmic), nghịch đảo của biến độc lập (Inverse = reciprocal – X), bậc 2 (Quadratic), bậc 3 (Cubic), lũy thừa (Power hay Multiplicative), đa hợp (Compound), đường cong Hình chữ S (S-curve), Logistic, tăng trưởng (Growth) và mũ (Exponential). Dưới đây là các dạng hàm trong phần mềm SPSS 10.0 và các hàm tương ứng trong phần mềm Stagraphics Plus Version 3.0 & 5.1.

Các hàm trong SPSS 10.0

Linear (mặc định): $Y = a + bX$

Logarithmic: $Y = a + b \times \ln X$

Inverse: $Y = a + b/X$

Compound: $Y = \alpha a^X$

Growth: $Y = \exp(a + bX)$

Power: $Y = aX^b$

S –Curve: $Y = \exp(a + b/X)$

Exponential: $Y = Ae^{\alpha x}$

Logistic: $N = K/(1 + \exp(a + bt))^{(\times)}$

Các hàm trong Stagraphics Plus Version 3.0

Linear (mặc định): $Y = a + bX$

Logarithmic: $Y = a + b \times \ln X$

Riciprocal - X: $Y = a + b/X$

Exponential: $Y = \exp(a + bX)$

Multiplicative: $Y = aX^b$

S –Curve: $Y = \exp(a + b/X)$

Logistic: $Y = \exp(a + bX)/(1 + \exp(a + bX))^{(\times)}$

Ước lượng hàm phi tuyến tính

Nếu hai biến Y và X có quan hệ phi tuyến tính, bạn có thể thử làm phù hợp số liệu của mình với 1 hoặc nhiều mô hình phi tuyến tính. Để xác định một mô hình, trước hết bạn sử dụng thủ tục đồ thị phân tán (Scatterplot) để mô tả số liệu. Nếu các biến xuất hiện mối liên hệ tuyến tính, bạn hãy sử dụng mô hình hồi quy tuyến tính đơn. Nếu các biến số không có mối liên hệ tuyến tính, bạn hãy cố gắng biến đổi số liệu của mình. Khi việc biến đổi số liệu cũng không giúp ích gì, bạn phải cần đến một mô hình phức tạp hơn. Hãy quan sát đồ thị phân tán (Scatterplot) mô tả dữ liệu. Nếu nhận thấy đồ thị tương đồng với một hàm toán học nào đó, bạn hãy thử số liệu của mình với mô hình đó. Ví dụ: Nếu số liệu phù hợp với hàm số mũ (exponential function), thì bạn có thể sử dụng mô hình mũ (exponential model). Chương trình cung cấp những mô hình hồi quy phi tuyến tính sau đây: inverse, quadratic, cubic, power, compound, S-curve, logistic, growth và exponential. Nếu bạn chưa tin tưởng vào một mô hình nào đó, bạn có thể thử số liệu của mình với một số mô hình khác. Một mô hình là phù hợp nếu nó mô tả số liệu với sai số ước lượng nhỏ nhất, nghĩa là min $S(Y_i - \hat{Y}_i)^2$.

Điều kiện của số liệu dùng trong ước lượng mô hình phi tuyến tính

Các biến phụ thuộc và biến độc lập (dự đoán) phải là biến định lượng. Nếu bạn chọn chuỗi thời gian (Time) thay cho chuỗi biến độc lập, thì thủ tục ước lượng mô hình phi tuyến tính đưa ra một biến thời gian mà trong đó độ dài của thời gian giữa các trường hợp là bằng nhau. Nếu thời gian được chọn, thì biến phụ thuộc cũng phải là những số đo theo chuỗi thời gian. Phân tích theo chuỗi thời gian đòi hỏi cấu trúc dữ liệu mà trong đó mỗi trường hợp (dòng) chỉ ra một tập hợp quan sát (dung lượng mẫu) ở những thời gian khác nhau và độ dài của thời gian giữa các trường hợp là bằng nhau. Ví dụ: Khi nghiên cứu sự biến đổi dân số của một quần thể sinh vật theo thời gian, thì dân số quần thể phải được thu thập theo từng thời điểm có độ dài thời gian như nhau.

Những giả định

Khi làm phù hợp số liệu thực nghiệm với một mô hình hồi quy nào đó (đường thẳng, mũ), thì những sai lệch (residuals) giữa những giá trị dự đoán (mô hình phù hợp) và những giá trị thực nghiệm phải tuân theo phân bố ngẫu nhiên và phân bố chuẩn. Nếu sử dụng một mô hình tuyến tính, thì những điều kiện sau đây phải thỏa mãn: (1) Đối với mỗi giá trị của biến độc lập, phân bố của biến phụ thuộc phải tuân theo phân bố chuẩn; (2) Phương sai của biến phụ thuộc Y phải bằng nhau đối với tất cả các giá trị của biến độc lập (X_i); (3) Y và X (hoặc đại diện của Y và X) có quan hệ tuyến tính; (4) Các quan sát là độc lập.

Thủ tục ước lượng phi tuyến tính

Từ menu chính, bạn hãy chọn (mũi tên chỉ thứ tự các bước): Analyze $\Rightarrow$ Regression $\Rightarrow$ Curve Estimation... Đến đây bạn chọn một hoặc nhiều biến phụ thuộc. Mỗi mô hình phải được cung cấp một biến phụ thuộc. Tiếp theo, chọn một biến độc lập. Biến độc lập có thể là biến số trong file dữ liệu của bạn, hoặc là biến số thời gian (Time). Điều đó tùy thuộc vào mục tiêu nghiên cứu của bạn. Sau khi thủ tục tính toán đã thực hiện xong, bạn nhấn nút Save để lưu các số dự đoán (predicted values), các sai lệch (residuals) và các khoảng dự đoán (prediction intervals) như một biến mới.

Khi xây dựng một hàm hồi quy ước lượng, bạn có thể yêu cầu chương trình hiển thị những tham số sau đây: hằng số của hàm hồi quy ước lượng (constant); đồ thị mô tả quan hệ giữa Y với X ; Bảng ANOVA.

Chương trình cũng giúp bạn lưu những kết quả sau đây: (1) các hệ số hồi quy (regression coefficients); (2) hệ số tương quan (multiple R), hệ số xác định hay bình phương của hệ số tương quan (R^2), bình phương của hệ số tương quan điều chỉnh (adjusted R^2); (3) sai

số chuẩn của ước lượng (standard error of the estimate); (4) Bảng phân tích biến động (ANOVA); (5) các trị dự đoán (predicted values); (6) các sai lệch giữa trị dự đoán và trị thực tế (residuals); (7) các khoảng dự đoán cho giới hạn trên và dưới của biến phụ thuộc (prediction intervals hoặc upper and lower bounds).

Các trường hợp dự đoán (Predict Cases)

Nếu chọn Time thay cho biến độc lập thì bạn có thể chỉ định một khoảng thời gian dự đoán nào đó thuộc một chuỗi thời gian. Bạn có thể chọn một trong những chọn lựa sau đây:

+ Dự đoán các giá trị Y dựa trên việc chỉ định một khoảng thời gian nào đó. Để đạt được mục đích này, bạn cần chỉ định thời gian cụ thể cho việc dự đoán.

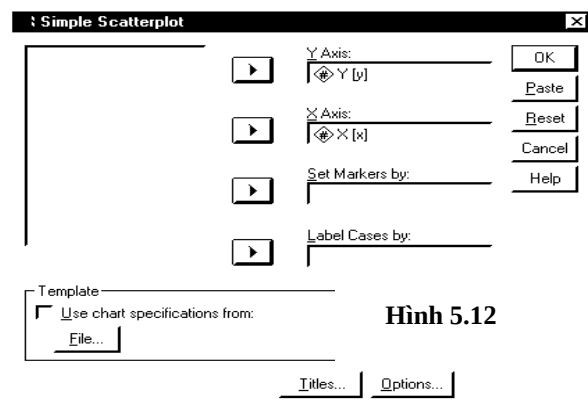
+ Dự đoán tất cả các trường hợp dựa trên toàn bộ chuỗi thời gian nghiên cứu. Nếu bạn không định nghĩa rõ khoảng thời gian dự đoán thì tất cả các trường hợp (dung lượng mẫu) sẽ được sử dụng để dự đoán.

5.3.1. Thực hành phân tích tương quan và hồi quy 2 biến

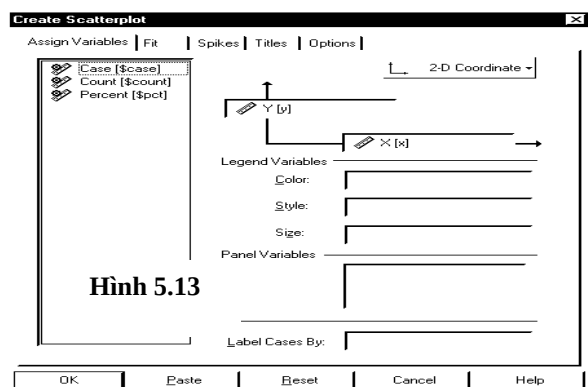
Bước 1. Trước hết, bạn hãy dùng biểu đồ phân tán (Scatter) để thăm dò mối quan hệ giữa biến phụ thuộc Y với biến độc lập X. Thủ tục này được thực hiện theo lệnh sau đây: Mở Graph → Scatter → Define → chọn biến Y và X để đưa vào tương ứng trục Y-Axis và X – Axis. Sau khi chọn OK, bạn hãy quan sát 3 nội dung chủ yếu sau đây:

+ **Phương hướng của mối liên hệ.** Quan sát phương hướng biến đổi của Y theo X và đưa ra phát biểu chính xác. Ví dụ: (1) Hàm lượng cao của phân super photphat có xu hướng làm giảm chiều cao và sinh khối của cây con Dầu song nòng 1 năm tuổi ở vườn ươm; (2) Nhiệt độ nắng cao có xu hướng làm giảm sinh trưởng của Thông ba lá...

+ **Dạng (form).** Để biết dạng của mối quan hệ, bạn có thể yêu cầu SPSS chỉ ra “đường hồi quy phù hợp nhất” bằng lệnh: Graphs → Scatter → Simple (trong cửa sổ Scatterplot) → Define → chọn Y và X để chuyển vào tương ứng trục Y Axis và X Axis (trong Simple Scatterplot) → OK (Hình 5.12). Nếu muốn thử các dạng hồi quy phi tuyến tính thì bạn nhấp vào Graphs → Interactive → Scatterplot. Đến đây chọn Y từ Assign Variable; sau đó nhấn con chuột vào Y, giữ và kéo Y thả vào ô chỉ trục tung (Y). Sau đó thực hiện tương tự cho biến X (Hình 5.13). Muốn tính toán các hệ số của phương trình thì chọn Fit. Tiếp theo trong cửa sổ Fit, bạn chọn Regression trong ô Method; sau đó đánh dấu các lệnh cần tính cho các thông số. Cuối cùng chọn OK. Kết quả cho ta dạng phương trình phù hợp với dãy số liệu.



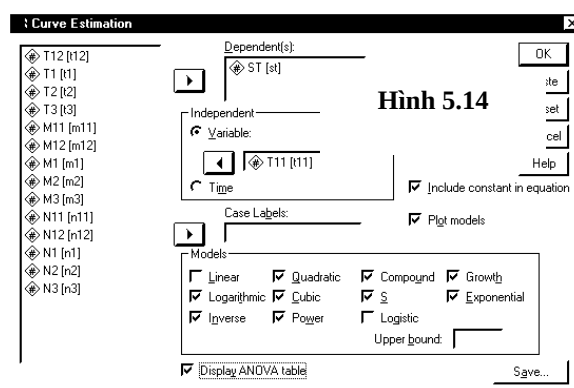
Hình 5.12



Hình 5.13

Lưu ý rằng, đồ thị và dạng phương trình cho ở bước này chỉ có ý nghĩa tham khảo. Trong mọi trường hợp, bạn có thể bắt đầu bằng cách làm phù hợp số liệu thực nghiệm với các dạng phương trình mà chương trình SPSS cung cấp sẵn, hoặc với các phương trình đã được nhiều người chấp nhận. Sau đó thực hiện thủ tục phân tích tương quan và hồi quy.

+ **Mức độ** (Strength). Khi phân tích tương quan, bạn phải trả lời câu hỏi: Quan hệ giữa Y với X mạnh hay yếu? Nhưng để trả lời được câu hỏi này thì bạn phải cần đến thủ tục Regression qua lệnh: Mở Analyze → Regression → Linear hay Curve Estimation → chọn Y đưa vào Dependent và chọn X đưa vào Variable. Tiếp theo chọn các mô hình cần thử mỗi liên hệ. Cuối cùng chọn OK để chương trình xử lý và hiển thị kết quả (**Hình 5.14**).



Kết quả hiển thị trên màn Hình bao gồm những tham số sau đây:

- + $Y_{bq} \pm S_y$, $X_{bq} \pm S_x$
- + r – Hệ số tương quan Pearson giữa Y với X và P – value 1 phía
- + R – Hệ số tương quan Multiple R
- + Bảng phân tích phương sai (ANOVA)
- + Các biến trong hàm hồi quy ước lượng.

Lưu ý rằng r và R giống nhau về trị số, nhưng R biểu thị quan hệ giữa $Y_{thực\ tể}$ với $Y_{lý\ thuyết}$ (dự đoán). R luôn luôn dương, còn r có thể dương hay âm. R^2 đo hiệu quả về mức độ chặt chẽ của mối quan hệ giữa Y_{tt} với Y_{lt} . $R^2 \times 100$ cho biết có bao nhiêu phần trăm giá trị Y biến đổi theo X hay bao nhiêu phần trăm giá trị Y được giải thích qua X .

Bước 2. Kiểm tra ý nghĩa của mối quan hệ và giải thích kết quả. Quan sát ma trận tương quan r và P -1 tailed Signification. Nếu $P < 0.05$ (mặc định) thì r tồn tại. Quan sát thống kê hồi quy. Nếu Sig – F hay P – value < 0.05 (mặc định) thì dạng hồi quy tồn tại.

Một số điều cần lưu ý

+ Khi mỗi quan hệ giữa hai biến số tồn tại ở dạng phi tuyến tính thì bạn nên tiến hành tuyến tính hóa bằng cách thay thế Y hoặc X bằng các “đại diện” của chúng như $\sqrt{X}$, $1/X$, $\ln X$, X^2 ...; $\sqrt{Y}$, $1/Y$, Y^2 ... Sau đó tiến hành phân tích tương quan tuyến tính giữa các đại diện này.

+ Hệ số tương quan r biểu thị mức độ chặt chẽ của mối quan hệ giữa đại diện của X với đại diện của Y .

+ Nếu sử dụng các hàm số mà chương trình SPSS cung cấp để xác định quan hệ phi tuyến tính giữa hai biến thì bạn có thể thực hiện theo các bước sau: Từ menu chính, mở Analyze ⇒ Regression ⇒ Curve Estimation. Khi cửa sổ Curve Estimation xuất hiện, bạn hãy chuyển biến Y vào ô Dependent, còn biến X vào ô Independent. Sau đó hãy chọn các hàm để thử mỗi liên hệ. Bạn có thể chọn cùng lúc nhiều hàm số để thử với số liệu của bạn.

5.3.2. Một số ví dụ về phân tích hồi quy đơn

1. Hàm hồi quy tuyến tính đơn: $Y = a + bX$

Giả thiết cần xác định mô hình hồi quy tuyến tính đơn để ước lượng chiều cao thân cây (H, m) trong quan hệ với tuổi cây (A, năm) theo số liệu của **Bảng 5.9**.

Bảng 5.9. Sự biến đổi của H theo A

A, năm	2	4	6	8	10
H, cm	4	6.5	7.8	9.5	11.8

Bảng 5.10a. Tóm tắt mô hình (Model Summary)^b

Model 1	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R ² Change	F Change	df1	df2	Sig. F Change
	.995 ^a	.990	.987	.343	.990	294.85	1	3	.000

a Predictors: (Constant), A (Năm); b Dependent Variable: H (m)

Sau khi thực hiện các bước gọi chương trình tính toán, kết quả hiển thị trên màn Hình những thông kê mô tả mô hình. Dưới đây chỉ trích dẫn những thông kê đáng quan tâm (**Bảng 5.10, 5.11 và 5.12**).

Bảng 5.10b. Bảng phân tích biến động (ANOVA)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	34.596	1	34.596	294.852	.000
	Residual	.352	3	.117		
	Total	34.948	4			

a Predictors: (Constant), X; b Dependent Variable: Y

Bảng 5.11. Các hệ số của mô hình (Coefficients)

Model 1		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
	(Constant)	2.340	.359		6.513	.007
	A (Năm)	.930	.054	.995	17.171	.000

Bảng 5.12. Các thống kê sai lệch (Residuals Statistics)

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	4.200	11.640	7.920	2.941	5
Std. Predicted Value	-1.265	1.265	.000	1.000	5
Standard Error of Predicted Value	.153	.265	.212	5.083E-02	5
Adjusted Predicted Value	4.500	11.400	7.924	2.825	5
Residual	-.280	.440	-3.553E-16	.297	5
Std. Residual	-.817	1.285	.000	.866	5

a Dependent Variable: H (m)

Giải thích và nhận xét kết quả

(1) Số liệu của **Bảng 5.10a** có ý nghĩa như sau:

+ P-value = 0.000 là giá trị làm tròn của 0.0005. Vì $P < 0,01$ nên giữa H với A tồn tại mối quan hệ tuyến tính rất chặt chẽ ở mức tin cậy 99%.

+ Thống kê R^2 chỉ ra 98,99% biến động của H được giải thích bởi biến tuổi (A, năm). Hệ số tương quan $R = 0,995$ và hệ số góc của phương trình đường (0,930); điều đó cho biết giữa H với A tồn tại mối liên hệ đường rất chặt chẽ.

+ Sai số chuẩn của ước lượng là Se (còn ký hiệu là S_{xy}) = 0,343. Bình phương của sai số chuẩn ước lượng (Se^2) được gọi là phương sai hồi quy nhỏ nhất. Đây là sai tiêu chuẩn của các sai lệch giữa $H_{thực\ tế}$ với $H_{ước\ lượng}$. Phương sai hồi quy nhỏ nhất được tính theo công thức:

$$S^2_e = \frac{1}{n-p-1} \times \sum_{i=1}^n (y_i - \hat{y}_i)^2; \quad (5.60)$$

trong đó : n - dung lượng quan sát; p - bậc hoặc số lượng biến độc lập; y_i và $\hat{y}_i$ - tương ứng là trị số quan sát và trị số hồi quy; n - p - 1 là độ tự do.

Giá trị Se ($Se = \sqrt{S^2_e}$) được dùng để tính khoảng tin cậy cho các trị dự đoán (y_p) và các giá trị trung bình của các mẫu mới (x_p) tương ứng với x_p nhất định.

Khoảng tin cậy $(1 - \alpha)100\%$ của các giá trị trung bình của biến y_p tương ứng với x_p được tính theo công thức:

$$y_p \pm t_{n-2, \alpha/2} Se \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x}_{bq})^2}{\sum (x_i - \bar{x}_{bq})^2}} \quad (5.61)$$

Khoảng tin cậy $(1 - \alpha)100\%$ của các giá trị dự đoán (Predicted Values) của biến y_p tương ứng với x_p được tính theo công thức:

$$y_p \pm t_{n-2, \alpha/2} Se \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x}_{bq})^2}{\sum (x_i - \bar{x}_{bq})^2}} \quad (5.62)$$

Từ công thức 5.61 và 5.62 cho thấy, khoảng tin cậy của các giá trị trung bình của biến y_p hẹp hơn so với khoảng tin cậy của các giá trị dự đoán y_p (Predicted Values).

(2) Bảng phân tích biến động (Bảng 5.10b) được đưa ra để kiểm định giả thuyết H_0 về sự tồn tại của mô hình hay tồn tại mối liên hệ tuyến tính giữa biến phụ thuộc Y với biến dự đoán X. Giả thuyết H_0 : Không tồn tại quan hệ tuyến tính giữa Y với X. Cột 3 là trị số tổng bình phương sai lệch của biến Y (SST). Trị số này được phân thành biến động do mô hình hay do hồi quy (SSR) và biến động không thể giải thích được (SSE); nghĩa là $SST = SSR + SSE$. Cột 4 là độ tự do. Cột 5 là trị số trung bình của các sai lệch bình phương do hồi quy (MSR) và do sai số (MSE). Giá trị của chúng tương ứng bằng cột 3 chia cho cột 4. Cột 6 và 7 báo cáo tương ứng giá trị của kiểm định F và mức ý nghĩa của kiểm định; trong đó $F = MSR/MSE$. Các trị số này được sử dụng để kiểm định giả thuyết về sự tồn tại của mô hình. Khi $P < 0,05$ (mặc định) thì giả thuyết H_0 bị bác bỏ.

(3) Bảng các hệ số của mô hình (Bảng 5.11). Cột Unstandardized Coefficients cho biết các hệ số chưa chuẩn hóa của mô hình $H = a + bA$. Các hệ số này dùng để ước lượng mô hình hồi quy giữa H với A. Ở ví dụ này, phương trình biểu thị mối liên hệ giữa H với A có dạng: $H = 2,340 + 0,930 \times A$. Cột Standardized Coefficients cho biết các hệ số hồi quy chuẩn hóa. Chúng được dùng để giải thích vai trò của mỗi biến độc lập trong mô hình. Ở đây chỉ có một biến A, nên trị số này có thể bỏ qua. Hai cột cuối ghi lại thống kê t và mức ý nghĩa P.

Chúng được dùng để kiểm định sự tồn tại của các hệ số hồi quy trong mô hình. Giả thuyết H_0 : Các hệ số a và $b = 0$. Khi $P < 0,05$ (mặc định) thì giả thuyết H_0 bị bác bỏ. Ở ví dụ này, vì các giá trị $P < 0,01$ nên các hệ số của phương trình tồn tại ở mức tin cậy rất cao (99%).

(4) Bảng các thống kê sai lệch (**Bảng 5.12**). Dòng đầu và dòng 4 cho biết các giá trị dự đoán ($H_{\text{lý thuyết}}$), trong đó H_{min} cho biết chiều cao ứng với tuổi 2, H_{max} cho biết chiều cao ứng với tuổi 10. Cột Mean cho biết chiều cao trung bình trong 10 năm. Cột Std. Deviation cho biết sai tiêu chuẩn về chiều cao dự đoán. Dòng hai có nhãn Std. Predicted Value cho biết giá trị dự đoán chuẩn. Dòng ba có nhãn Standard Error of Predicted Value cho biết sai số chuẩn của giá trị dự đoán. Hai hàng cuối có nhãn Residual và Std. Residual cho biết các sai lệch và sai lệch chuẩn.

2. Phân tích hồi quy phi tuyến tính đơn

Giả sử cần thăm dò mối quan hệ giữa tỉ lệ gia tăng giá trị gỗ hàng năm của rừng Keo lá tràm (Y) với tuổi rừng (A, năm). Số liệu ban đầu dẫn ra ở **Bảng 5.13**.

Bảng 5.13. Sự biến đổi của tỉ lệ gia tăng giá trị gỗ (Y) hàng năm theo tuổi rừng

A, năm	5	6	7	8	9	10	11	12
Y	0.3485	0.2485	0.1846	0.1419	0.1122	0.0908	0.0750	0.0629

Để chọn được một mô hình phù hợp với dữ liệu đã cho, trước hết bạn hãy thử số liệu của **Bảng 5.13** với tất cả các dạng phương trình mà chương trình cung cấp. Sau đó, bạn có thể chọn ra một số mô hình có hệ số tương quan cao nhất. Kết quả cho thấy những mô hình sau đây có hệ số tương quan cao :

Model	Correlation	Correlation R-Squared
Power	-0.9999	99.98%
Exponential	-0.9939	98.78%
S-curve	0.9902	98.04%
Logarithmic-X	-0.9749	95.04%
Linear	-0.9402	88.40%

Từ số liệu trên cho thấy mô hình mũ biến đổi (Power) có hệ số tương quan cao nhất. Do đó, bạn có thể làm phù hợp số liệu thực nghiệm với dạng mô hình Power. Mô hình Power có dạng: $Y = aX^b$; trong đó Y là tỉ lệ gia tăng giá trị gỗ hàng năm, còn X – tuổi rừng (A, năm). Phân tích tương quan và hồi quy cho kết quả như sau **Bảng 5.14, 5.15 và Hình 5.15**.

Bảng 5.14. Bảng phân tích biến động (Analysis of Variance)

Coefficient	Df	Sum of Squares	Mean Square	F	Signif F
Regression	1	2.49184	2.49184	27465.654	0.0000
Residuals	6	0.0005444	0.000091		

Multiple R = 0,9999; R Square = 0,9998; Adjusted R Square = 0,9998; Standard Error = 0,0095

Bảng 5.15. Các biến trong phương trình (Variables in the Equation)

Variable	B	SE B	Beta	t	Sig.
A	-1.962177	0.011840	-0.999891	-165.728	0.0000
(Constant)	8.321648	0.208906		39.834	0.0000

Giải thích và nhận xét kết quả

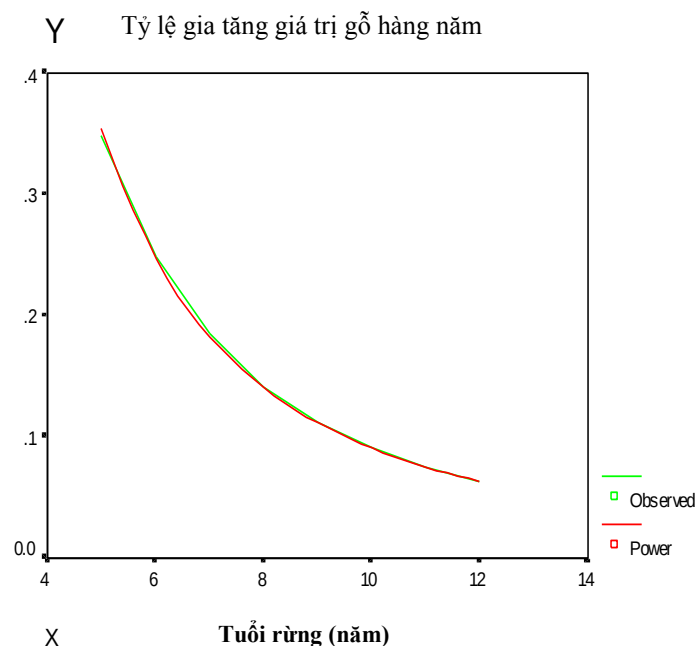
Hệ số R và Bảng ANOVA chỉ cho thấy sự phù hợp của mô hình. Ở ví dụ này, hệ số R biểu thị quan hệ giữa LnY với LnA theo phương trình: $\text{LnY} = \text{Lna} + b\text{LnA}$. Giá trị R âm (-0,999) cho biết khi tuổi rừng tăng lên thì tỉ lệ gia tăng giá trị gỗ hàng năm của rừng sẽ giảm.

Bình phương của hệ số tương quan bằng 99,9%; nó cho biết 99,9% biến động của tỉ lệ gia tăng giá trị gỗ hàng năm được giải thích bởi biến động của tuổi rừng sau khi đã tuyến tính hóa mô hình.

Phương trình của mô hình phù hợp có dạng:

$$\text{LnY} = 8.32165 - 1.96218 \times A$$

$$\text{hay } Y = 8.32165 \times A^{(-1.96218)}.$$



Bảng **Hình 5.15.** Quan hệ giữa tỷ lệ gia tăng giá trị gỗ hàng năm với tuổi rừng. Ở ví dụ này, hệ số R âm (-0,999) cho biết khi tuổi rừng tăng lên thì tỉ lệ gia tăng giá trị gỗ hàng năm của rừng sẽ giảm. Bình phương của hệ số tương quan bằng 99,9%; nó cho biết 99,9% biến động của tỉ lệ gia tăng giá trị gỗ hàng năm được giải thích bởi biến động của tuổi rừng sau khi đã tuyến tính hóa mô hình. Phương trình của mô hình phù hợp có dạng: $\text{LnY} = 8.32165 - 1.96218 \times A$ hay $Y = 8.32165 \times A^{(-1.96218)}$.

3. Phân tích hồi quy phi tuyến tính ở bậc cao

Để tìm được dạng phương trình phù hợp với số liệu đã cho ở các bậc cao hơn bậc 2 (mặc định), bạn có thể giải quyết vấn đề này bằng cách tuyến tính hóa các hàm bậc cao. Sau đó bạn có thể tiến hành phân tích các thống kê theo mô hình tuyến tính đa yếu tố (xem phần

5.4). Chẳng hạn, để làm phù hợp số liệu với hàm bậc 4 ($Y = a + bX + cX^2 + dX^3 + eX^4$), bạn thực hiện như sau :

Đặt $X = X_1$, $X^2 = X \times X = X_2$, $X^3 = X \times X \times X = X_3$, $X^4 = X \times X \times X \times X = X_4$. Từ đó ta có $Y = a + bX_1 + cX_2 + dX_3 + eX_4$.

Các bậc cao hơn bậc 4 cũng được thực hiện tương tự như trên. Sau khi phân tích thống kê, bạn có thể chuyển các hàm tuyến tính này về các hàm bậc k ban đầu.

Một số điều cần lưu ý

- Muốn biết bậc nào phù hợp nhất với số liệu đã cho, bạn có thể căn cứ vào những phân tích thống kê của Bảng ANOVA và Bảng phân tích hồi quy. Ở đây bạn cần lưu ý đến thống kê F, t và P - value. Khi bậc nào có $P < 0.01$, bậc đó được xem là phù hợp với dữ liệu đã cho, và bạn có thể dừng việc tìm kiếm các bậc cao hơn.
- Trong phần phân tích tương quan, các hệ số R và R^2 biểu thị cho hệ số tương quan đa phần giữa Y với các đại diện của X^k ...
- Khi cần dự đoán các giá trị Y lý thuyết, bạn hãy thực hiện các bước giống như phân tích tương quan và hồi quy đa yếu tố (xem phần 5.4).

5.4. PHÂN TÍCH TƯƠNG QUAN VÀ HỒI QUY ĐA TUYẾN TÍNH

5.4.1. Nội dung phân tích hồi quy tuyến tính đa biến tố

Hồi quy tuyến tính đa yếu tố có dạng tổng quát : $Y = A + B_1X_1 + B_2X_2 + B_3X_3 + \dots + B_KX_K + e$. Hồi quy tuyến tính đa yếu tố xuất hiện trong những trường hợp sau đây:

+ Một hoặc tất cả các biến dự đoán (X_i) có ảnh hưởng tuyến tính đến biến phụ thuộc Y. Điều này xảy ra khi giá trị biến đổi của Y trên một đơn vị biến đổi của X_i là hằng số trong khoảng biến thiên của X_i .

+ Ảnh hưởng của từng biến X_i lên Y là độc lập với biến khác. Điều này có nghĩa là độ lớn của sự thay đổi ở Y trên một đơn vị biến đổi của mỗi X_i là như nhau mà không phụ thuộc vào biến X_i khác.

Thủ tục phân tích tương quan và hồi quy tuyến tính đa yếu tố bao gồm những nội dung cơ bản sau đây:

Nội dung 1. Thăm dò mối quan hệ giữa Y với các X_i và xác định sơ bộ những biến X_i đưa vào mô hình. Mục đích của thủ tục này là nhằm xác định tương quan giữa Y với các X_i và tương quan giữa các X_i với nhau thông qua một ma trận tương quan. Ở bước này, ta cần chọn những hệ số tương quan cao giữa các biến Y với X_i , vì chúng có ảnh hưởng lớn đến kết quả phân tích hồi quy và mô hình dự đoán.

Nội dung 2. Phân tích hồi quy đa tuyến tính. Mục đích của bước này là nhằm tính các hệ số hồi quy riêng phần (Partial regression coefficients), kiểm định quan hệ tuyến tính giữa biến Y với các biến X_i , xác định các hệ số hồi quy và kiểm định cộng tuyến tính.

a. Xác định tầm quan trọng của từng biến X_i trong mô hình. Nội dung này được giải quyết bằng cách phân tích các hệ số hồi quy chuẩn hóa, các hệ số tương quan trật tự không (Zero-order Correlations), hệ số tương quan riêng phần hay cục bộ (Partial Correlation) và hệ số tương quan từng phần (Part Correlation).

Hệ số hồi quy chuẩn hóa Beta (Beta Standardized Coefficients) được đưa vào nhằm xác định mức ý nghĩa của từng biến X_i trong mô hình. Các hệ số hồi quy chuẩn hóa mô tả hồi quy giữa biến chuẩn hóa Y với các biến chuẩn hóa X_i . Bởi vì các hệ số hồi quy từng phần B_k (mỗi hệ số được điều chỉnh theo các biến độc lập khác) là các hệ số không được chuẩn hóa (nghĩa là chúng được đo đạc bằng các thang đo khác nhau), nên chúng không thể giải thích rõ tầm quan trọng của từng biến trong mô hình. Hệ số chuẩn hóa Beta tính theo công thức:

$$\text{Beta}_k = B_k(S_k/S_y) \quad (5.63)$$

Trong đó B_k là hệ số hồi quy từng phần của biến k , S_k – độ lệch chuẩn của biến độc lập thứ k , S_y – độ lệch chuẩn của biến Y . Cần nhận thấy rằng, hệ số Beta_k và hệ số B_k đều phụ thuộc không chỉ vào số lượng biến số đưa vào mô hình, mà còn bởi quan hệ giữa các biến độc lập. Vì thế, chúng phản ánh vẫn chưa chính xác tầm quan trọng của từng biến số độc lập.

Hệ số tương quan trật tự không là hệ số tương quan giữa Y với từng X_i nhưng chưa loại bỏ ảnh hưởng tuyến tính của các biến độc lập khác. Các hệ số tương quan này cũng phản ánh không chính xác tầm quan trọng của từng biến số độc lập.

Hệ số tương quan từng phần (Part Correlation) và hệ số tương quan riêng phần (Partial Correlation). Các hệ số này đo ảnh hưởng của từng biến độc lập trong mô hình. Hệ số tương quan từng phần đo mức độ quan hệ giữa Y với biến số độc lập thứ i sau khi đã loại bỏ ảnh hưởng tuyến tính của các biến độc lập khác đối với biến số độc lập thứ i . Hệ số tương quan riêng phần đo mức độ quan hệ giữa biến Y với biến độc lập thứ i sau khi đã loại bỏ ảnh hưởng tuyến tính của các biến độc lập khác đối với cả biến Y lẫn biến độc lập thứ i .

Bình phương của hệ số tương quan từng phần tính theo công thức:

$$R^2_{\text{change}} = R^2 - R^2_{(i)} \quad (5.64)$$

Bình phương của hệ số tương quan riêng phần tính theo công thức:

$$P_{Ri}^2 = [R^2 - R^2_{(i)}]/[1 - R^2_{(i)}] \quad (5.65)$$

Trong **công thức 5.64 và 5.65**, R^2 là bình phương của hệ số tương quan bội khi mô hình bao gồm tất cả các biến X_i , $R^2_{(i)}$ là bình phương của hệ số tương quan bội khi tất cả các biến đều được đưa vào mô hình trừ biến i . Bởi vì mẫu số của $P_{Ri}^2 \leq 1$, nên $P_{Ri}^2 > R^2_{\text{change}}$.

b. Thống kê cộng tuyến tính (Collinearity Statistics). Thống kê cộng tuyến tính đo mức độ quan hệ tuyến tính giữa các biến độc lập. Để đo mức độ cộng tuyến tính, bạn hãy chú ý đến thống kê độ chấp nhận (Tolerance) và yếu tố khuếch đại phương sai (Variance Inflation Factor - VIF).

$$\text{Độ chấp nhận của biến } X_i = 1 - R^2_i \quad (5.66)$$

R^2_i = hệ số tương quan giữa biến độc lập thứ i với các biến độc lập khác.

Khi biến độc lập thứ i có quan hệ yếu với các biến độc lập khác thì độ chấp nhận sẽ lớn. Ngược lại, khi biến độc lập thứ i có quan hệ chặt chẽ với các biến độc lập khác thì độ chấp nhận sẽ nhỏ. Cần lưu ý rằng, độ chấp nhận của biến thứ i được tính nhằm xác định biến i có ý nghĩa hay không, hay nó có thể được đưa vào mô hình hay không. Tiêu chuẩn độ chấp nhận mặc định là 0.0001. Vì thế, nếu một biến được đưa vào mô hình thì độ chấp nhận của nó phải lớn hơn 0.0001.

Yếu tố khuếch đại phương sai của biến X_i (VIF_i) tính theo công thức:

$$\text{VIF}_i = 1/[1 - R^2_i] \quad (5.67)$$

Như vậy, VIF là trị số nghịch đảo của độ chấp nhận. VIF_i được dùng để tính phương sai cho hệ số hồi quy thứ i . Khi độ chấp nhận nhỏ thì VIF sẽ lớn. Nói khác đi, khi VIF_i lớn thì hệ số hồi quy thứ i biến động mạnh. Sự biến động mạnh của hệ số hồi quy thứ i cho biết biến này có ý nghĩa kém trong mô hình.

Kiểm định cộng tuyến tính (Collinearity Diagnostics) được thực hiện theo các thống kê như các số đặc trưng (Eigenvalue), các chỉ số điều kiện (Condition Index) và phần đóng góp trong phương sai (Variance Proportions).

Giá trị đặc trưng (Eigenvalue) cho biết số lượng chiều (Dimensions) khác nhau giữa các biến độc lập (hay số lượng biến đưa vào mô hình). Khi giá trị đặc trưng $\cong 0$ thì các biến X_i có quan hệ chặt chẽ với nhau. Các chỉ số điều kiện là căn bậc 2 của các tỷ số giữa giá trị đặc trưng lớn nhất ($\text{Eigenvalue}_{\max}$) với từng giá trị đặc trưng kế tiếp với từng biến. Nếu chỉ số điều kiện của một biến nào đó lớn hơn 15 thì biến số đó có thể có hiện tượng cộng tuyến tính với những biến khác. Các phần đóng góp trong phương sai (Variance Proportions) là các phần phương sai của những ước lượng được giải thích bởi từng bộ phận có liên quan đến từng trị số Eigenvalue. Nếu một biến số nào đó có hiện tượng cộng tuyến tính với các biến độc lập khác thì phần đóng góp của nó trong phương sai của các biến khác sẽ lớn.

Nội dung 3. Xây dựng mô hình hồi quy đa tuyến tính. Sau khi chọn được những biến dự đoán (độc lập) có quan hệ chặt chẽ với biến phụ thuộc, bước tiếp theo là xây dựng mô hình và chọn mô hình phù hợp. Trước khi nghiên cứu các thủ tục xây dựng mô hình, bạn cần xem xét kết quả đưa vào và loại ra các biến trong phương trình hồi quy và những thông số thống kê của các biến không được đưa vào mô hình.

Những biến được đưa vào và loại ra khỏi mô hình

Khi có nhiều biến độc lập (dự đoán) được chọn để xây dựng mô hình, ta cần biết những biến nào sẽ tạo ra một mô hình tốt. Để chọn được những biến đưa vào và loại ra khỏi mô hình, chương trình SPSS thực hiện theo phương pháp chọn biến từng bước (Stepwise Method). Mục đích của chọn biến từng bước là nhằm xác định những biến dự đoán hữu ích để đưa vào mô hình. Trước khi xem xét các thủ tục xây dựng mô hình, chúng ta cần phân tích kết quả khi đưa vào hoặc loại bỏ các biến trong mô hình cũng như các thống kê đối với các biến không ở trong mô hình.

Phân tích một biến được đưa vào hay loại ra khỏi mô hình

Một biến được đưa vào mô hình khi nó làm thay đổi đáng kể hệ số R^2 . Để biết một biến đưa vào mô hình có làm thay đổi đáng kể hệ số R^2 hay không, người ta tính hệ số tương quan từng phần R^2_{change} .

Giả thuyết H_0 : Giá trị thực của hệ số tương quan từng phần trong tổng thể bằng không, nghĩa là $R^2_{\text{change}} = 0$. Giả thuyết này được kiểm định bằng thống kê F riêng phần (Partial F Test, kí hiệu là F_{change}). Thống kê F riêng phần được tính theo công thức:

$$F_{\text{change}} = \frac{R^2_{\text{change}}(N - p - 1)}{q(1 - R^2)} \quad (5.68)$$

ở đây N = dung lượng mẫu;

p = số lượng biến dự đoán (độc lập);

q = số lượng biến được đưa vào mô hình ở từng bước;

R^2 = bình phương của hệ số tương quan giữa biến phụ thuộc với những biến độc lập ở bước trước.

Quy tắc quyết định: Bác bỏ H_0 nếu ($\text{Sig. } F_{\text{change}} < \alpha$ ($\alpha = 0.05$ hoặc 0.01)).

Giả thuyết mức độ thay đổi của $R^2 = 0$ cũng tương đồng với giả thuyết các $B_i = 0$.

Tóm lại, một biến được đưa vào mô hình khi nó làm thay đổi đáng kể hệ số R^2 .

Những thông số thống kê của những biến không ở trong mô hình

Mục đích của việc xem xét những thông số thống kê của những biến không ở trong mô hình là xác định những biến nào sẽ được đưa vào mô hình ở bước tiếp theo. Khi xem xét những biến không được đưa vào mô hình ở từng bước, ta cần chú ý đến các thống kê sau: Beta In, t , mức ý nghĩa của t (Sig hay P) và hệ số tương quan riêng phần.

Thống kê Beta In tương ứng với mỗi biến là hệ số hồi quy chuẩn hóa của mỗi biến khi nó được đưa vào ở bước tiếp theo. Giả thuyết H_0 : Beta In = 0. Giả thuyết này được kiểm định bằng thống kê t. Khi mức ý nghĩa của thống kê t nhỏ (< 0.05 hay 0.01) thì giả thuyết H_0 bị bác bỏ. Khi giả thuyết H_0 bị bác bỏ thì biến ấy sẽ được đưa vào mô hình ở bước sau.

Xác định số lượng biến dự đoán “tối ưu”

Hệ số xác định R^2 được tính theo công thức:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (5.69)$$

trong đó:

SSR = tổng các chênh lệch bình phương do hồi quy;

SSE = tổng các chênh lệch bình phương do sai số;

SST = SSR + SSE - tổng các chênh lệch toàn bộ.

Về lý thuyết, khi đưa vào mô hình càng nhiều biến thì hệ số xác định (R^2) càng cao. Mặt khác, hệ số R^2 sẽ không bao giờ giảm khi các biến dự đoán tiếp tục được đưa vào mô hình. Nhưng một mô hình với nhiều biến không có nghĩa là nó sẽ phù hợp hơn với tổng thể. Một mô hình nhiều biến có thể dẫn đến làm tăng sai số chuẩn của ước lượng (dự đoán). Vì thế, việc xác định những biến dự đoán “tối ưu” là cần thiết. Để đạt được mục đích này, người ta sử dụng hệ số R^2 điều chỉnh ($\text{Adjusted } R^2 = R_a^2$).

$$R_a^2 = 1 - (1 - R^2) \frac{n - 1}{n - p - 1} \quad (5.70)$$

n = kích thước mẫu hay số cặp biến Y và X;

p = số biến dự đoán đưa vào mô hình.

Từ công thức 5.70 cho thấy, khi p tăng lên thì R_a^2 giảm xuống. Như vậy, hệ số R_a^2 không nhất thiết tăng lên khi đưa thêm một biến vào mô hình. Vì thế, R_a^2 là thước đo mức độ phù hợp của các biến độc lập trong mô hình.

Nội dung 4. Xác định phương pháp chọn biến. Trong phân tích tương quan và hồi quy tuyến tính bội có ba phương pháp chọn biến sau đây :

a. Phương pháp loại bỏ dần từng biến (Backward selection). Phương pháp này khởi đầu với tất cả các biến mà bạn dự định đưa vào xây dựng mô hình. Sau đó, chương trình sẽ tuần tự loại bỏ dần từng biến thông qua tiêu chuẩn loại trừ (Removal Criteria). Đây là phương pháp mặc định của chương trình. Để loại trừ dần từng biến, chương trình sử dụng hai tiêu chuẩn sau đây:

Tiêu chuẩn Fmin hay Fra (F to remove). Một biến được để lại trong mô hình khi F to remove của nó phải lớn hơn F to remove tiêu chuẩn (hay Fmin). Ngược lại, một biến bị loại ra khỏi mô hình khi F to remove của nó phải nhỏ hơn F to remove tiêu chuẩn (hay Fmin). Giá trị tiêu chuẩn mặc định của F ra là 2.71. Vì thế, khi sử dụng F to remove bằng 2.71 thì một biến được để lại trong mô hình nếu F to remove của nó lớn hơn 2.71. Ngược lại, khi một biến có F to remove nhỏ hơn 2.71 thì biến đó bị loại ra. Lưu ý rằng F to remove tiêu chuẩn có thể được chỉ định theo yêu cầu của nhà nghiên cứu.

Tiêu chuẩn xác suất tối đa của F ra (Probability of F to remove). Một biến được để lại trong mô hình khi xác suất của F tính phải nhỏ hơn xác suất tiêu chuẩn và ngược lại. Giá trị xác suất tiêu chuẩn mặc định là 0,10.

b. Phương pháp đưa dần từng biến vào (Forward Selection). Phương pháp này khởi đầu với mô hình không có biến nào. Sau đó, việc chọn những biến đưa vào mô hình sẽ được tiến hành qua nhiều bước. Ở bước thứ nhất, biến dự đoán đầu tiên có tương quan thuận

hay nghịch cao nhất với biến phụ thuộc sẽ được đưa vào phân tích. Ở những bước tiếp theo, biến nào có hệ số tương quan cục bộ cao nhất sẽ được đưa dần vào. Lưu ý rằng một biến Xi được đưa vào mô hình khi nó làm cho hệ số R^2 thay đổi đáng kể hoặc hệ số $B_i \neq 0$. Giả thuyết $B_i \neq 0$ được kiểm định theo thống kê F. Để biết một biến có được đưa vào mô hình hay không, ta so sánh $F_{tính}$ với $F_{tiêu chuẩn}$. Chúng ta có thể chỉ định thống kê F theo hai tiêu chuẩn :

Tiêu chuẩn F vào (F to enter). Giá trị mặc định của F to enter là 3,84. Vì thế, một biến được đưa vào khi F tương ứng với nó lớn hơn F tiêu chuẩn, nghĩa là $F_{tính} > F_{to enter} = 3,84$. Khi một biến không thỏa mãn điều kiện này thì biến ấy sẽ bị loại ra ở bước sau. Lưu ý rằng F to enter tiêu chuẩn có thể được chỉ định theo yêu cầu của nhà nghiên cứu.

Tiêu chuẩn xác suất của F to enter (Probability of F to enter). Xác suất mặc định của F to enter = 0.05. Khi một biến được đưa vào mô hình thì xác suất (kí hiệu Sig) tương ứng với F phải nhỏ hơn 0.05 (mặc định) hay một xác suất mà ta chỉ định.

Nếu những điều kiện trên đây được thỏa mãn thì những bước tiếp theo sẽ được thực hiện tiếp. Ngược lại, khi những điều kiện ấy không được thỏa mãn thì việc chọn biến sẽ dừng.

Một số lưu ý

- ✓ Trong Bảng kết quả, chương trình không đưa ra thống kê F mà đưa ra thống kê t. Hai thống kê này là tương đồng, vì $t^2 = F$.
- ✓ Khi chọn tiêu chuẩn F to enter thì kết quả có thể khác với việc chọn xác suất của F to enter. Điều này xảy ra là vì, khi đưa các biến vào mô hình, số bậc tự do của tổng bình phương phần dư (SS dư) sẽ giảm xuống, nhưng bậc tự do của hồi quy lại tăng lên. Vì thế, khi cho một giá trị F cố định thì mức ý nghĩa sẽ thay đổi theo số lượng biến đưa vào. Chỉ khi dung lượng quan sát lớn thì sự khác biệt này sẽ giảm.

c. Phương pháp chọn từng bước (Stepwise selection). Phương pháp này là sự kết hợp giữa hai phương pháp đưa dần vào từng biến độc lập và loại bỏ dần từng biến độc lập. Đây là phương pháp thông dụng nhất. Trước hết, chương trình bắt đầu chọn biến thứ 1 theo phương pháp chọn dần từng bước. Khi biến thứ 1 thỏa điều kiện vào ($F_{vào}$ – kí hiệu F_{IN} , và xác suất vào - kí hiệu P_{IN}) thì chương trình sẽ tiếp tục chọn biến thứ 2. Biến thứ 2 được chọn là biến có hệ số tương quan riêng phân cao nhất. Khác với thủ tục đưa dần vào từng biến, thủ tục chọn biến từng bước xem xét có nên loại bỏ biến thứ nhất hay không dựa theo tiêu chuẩn F ra ($F_{to remove}$) hoặc xác suất của F ra. Cách thức này giống như thủ tục loại trừ dần từng biến. Trong những bước kế tiếp, những biến nào không ở trong mô hình sẽ được xem xét để đưa dần vào. Sau đó các biến ở trong mô hình lại được xem xét để loại ra. Khi không còn biến nào thỏa điều kiện loại ra thì chương trình sẽ dừng. Cần lưu ý rằng, một biến có thể được đưa vào và loại ra nhiều lần. Để tránh hiện tượng này, xác suất vào phải nhỏ hơn xác suất ra hay $F_{IN} > F_{OUT}$. Thủ tục sẽ dừng khi không còn biến nào thỏa điều kiện vào và ra.

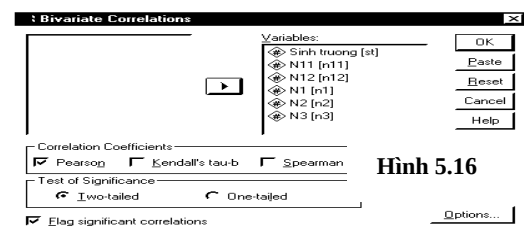
5.4.2. Thực hành phân tích hồi quy đa tuyến tính

(Áp dụng phương pháp chọn biến từng bước)

Giả thiết cần phân tích tương quan và hồi quy giữa sản lượng Vải (kí hiệu - ST) với số giờ năng từ tháng 11 năm trước (N_{11}) đến tháng 3 năm sau (N_3) tại Trại Phú Hộ (Vĩnh Phú cũ). Ví dụ này sử dụng từ nguồn dẫn liệu của Vũ Công Hậu (1980).

Bước 1. Tính ma trận tương quan giữa các biến số. Nếu số liệu đã được ghi lại

thành 1 file, thì từ menu chính, bạn mở File $\Rightarrow$ Open $\Rightarrow$ Data (tại đây chọn file name) $\Rightarrow$ OK.



Hình 5.16

Kế đến, mở Analyze $\Rightarrow$ Correlate $\Rightarrow$ Bivariate. Tiếp theo bạn chuyển tất cả các biến phụ thuộc và biến độc lập vào ô Variables; đồng thời đánh dấu chọn Pearson trong ô Correlation Coefficients như **Hình 5.16**. Cuối cùng chọn OK để chương trình tính toán và hiển thị ma trận tương quan (**Bảng 5.16**).

Bước 2. Phân tích mô hình (Model) với từng biến độc lập và chọn mô hình phù hợp. Để giải quyết vấn đề này, bạn hãy mở Analyze $\Rightarrow$ Regression $\Rightarrow$ Linear. Khi cửa sổ Linear Regression xuất hiện, bạn hãy chọn Y (ở đây là ST) chuyển vào ô Dependent Variable(s). Tiếp theo, bạn chọn từng biến độc lập để chuyển nó vào ô Independent(s) theo thủ tục sau :

+ Trước hết, chọn biến N₁₁ chuyển vào ô Independent(s). Kế đến nhấp Next, rồi chọn tiếp N₁₂ và chuyển nó vào ô Independent(s). Thao tác này được lặp lại cho đến khi đã chuyển xong tất cả 5 biến N_i vào ô Independent(s). Sau đó bạn chọn tiếp Method và đánh dấu chọn Stepwise (**Hình 5.17**). Thao tác này chỉ định cho chương trình chọn biến và phân tích hồi quy từng bước. Khi sử dụng phương pháp chọn biến từng bước, ở đây đã sử dụng tiêu chuẩn F vào ($F_{to\ enter} > 3.84$ (mặc định) và xác suất của F vào ≤ 0.05 (mặc định)).

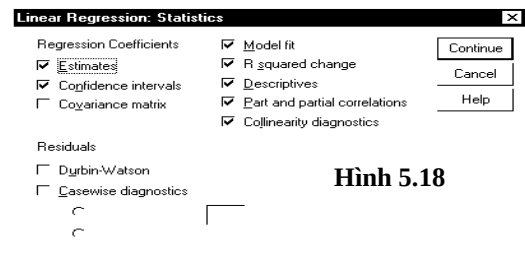
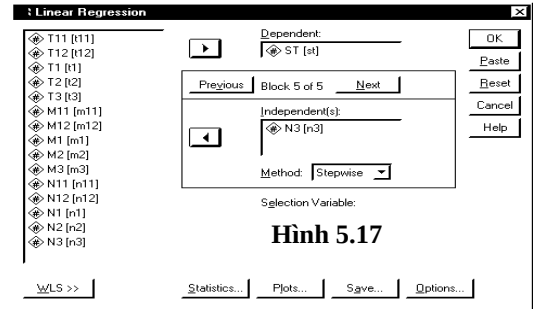
+ Từ cửa sổ Linear Regression, bạn chọn Statistics và đánh dấu các ô Estimates, Confidence intervals, Model fit, R – squared change, Descriptives, Part and Partial Correlations, Collinearity diagnostics (**Hình 5.18**). Sau đó bạn chọn Continue rồi chọn OK.

Đến đây kết quả hiển thị trên màn Hình bao gồm những Bảng sau đây : (1) Ma trận tương quan (**Bảng 5.16**), (2) tóm tắt mô hình (Model Summary) (**Bảng 5.17**), (3) phân tích biến động (ANOVA) (**Bảng 5.18**), (4) các hệ số hồi quy (**5.19**), (5) các biến loại bỏ khỏi mô hình (Excluded Variables) (**Bảng 5.20**), (6) dự đoán cộng tuyến tính (Collinearity Diagnostics) (**Bảng 5.21**), (7) các hệ số của mô hình phù hợp (**Bảng 5.22**).

Bảng 5.16. Ma trận tương quan

		Sinh truong	N11	N12	N1	N2	N3
Sinh truong	Pearson Correlation	1.000	-.609	-.602	-.611	.129	-.010
	Sig. (2-tailed)		.082	.086	.080	.741	.979
	N		9	9	9	9	9
N11	Pearson Correlation		1.000	.493	.147	-.215	.116
	Sig. (2-tailed)			.177	.705	.578	.767
	N			9	9	9	9
N12	Pearson Correlation			1.000	-.120	.314	-.540
	Sig. (2-tailed)				.758	.410	.134
	N				9	9	9
N1	Pearson Correlation				1.000	-.705	.126
	Sig. (2-tailed)					.034	.747
	N					9	9
N2	Pearson Correlation					1.000	-.022
	Sig. (2-tailed)						.954
	N						9
N3	Pearson Correlation						1.000
	Sig. (2-tailed)						
	N						

×× Correlation is significant at the 0.05 level (2-tailed).



Giải thích và nhận xét kết quả . Bảng 5.16 chỉ ra quan hệ giữa biến phụ thuộc ST với từng biến độc lập Ni và giữa các biến độc lập Ni với nhau. Từ Bảng 5.16, bạn có thể thấy những biến Ni nào có quan hệ chặt chẽ với biến ST; biến nào làm tăng ST, còn biến nào làm giảm ST. Ngoài ra, từ hệ số tương quan giữa các biến Ni, bạn cũng có thể dự đoán các hiện tượng cộng tuyến tính giữa các biến Ni. Trong Bảng 5.16, những biến Ni có quan hệ khá chặt chẽ với nhau được gạch ngang ở dưới. Biến N_3 bị loại ra khỏi mô hình, bởi vì trị số đặc trưng (Eigenvalue) < 0 , chỉ số điều kiện (Condition Index) $> 15,0$ và Beta In khác Zero (Bảng 5.20 và 5.21).

Bảng 5.17. Tóm tắt mô hình (Model Summary)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	2	3	4	5	6	7	8	9	10
1	.609a	.371	.281	1.0436	.371	4.126	1	7	.082
2	.701b	.491	.322	1.0137	.120	1.420	1	6	.278
3	.936c	.877	.803	.5467	.385	15.627	1	5	.011
4	.991d	.982	.964	.2340	.105	23.301	1	4	.008

a Predictors: (Constant), N11; b Predictors: (Constant), N11, N12; c Predictors: (Constant), N11, N12, N1. d Predictors: (Constant), N11, N12, N1, N2.

Bảng 5.18. Phân tích biến động (Tóm tắt)

Model		Sum of Squares	df	Mean Square	F	Sig.
1...	...	...	...	...	...	...
4	Regression	11.899	4	2.975	54.342	.001
	Residual	.219	4	5.474E-02		
	Total	12.118	8			

a Predictors: (Constant), N11; b Predictors: (Constant), N11, N12

c Predictors: (Constant), N11, N12, N1; d Predictors: (Constant), N11, N12, N1, N2

e Dependent Variable: ST

Bảng 5.19. Các hệ số hồi quy

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
4	(Constant)	5.815	.470		12.363	.000					
	N11	-.747	.160	-.407	-4.670	.010	-.609	-.919	-.314	.595	1.680
	N12	-.470	.121	-.352	-3.878	.018	-.602	-.889	-.261	.550	1.819
	N1	-1.869	.187	-.967	-10.009	.001	-.611	-.981	-.673	.484	2.065
	N2	-1.423	.295	-.530	-4.827	.008	.129	-.924	-.324	.375	2.664

Bảng 5.20. Các biến loại bỏ khỏi mô hình (Excluded Variables)

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
...	...							
4	N3	-.166	-1.406	.254	-.630	.262	3.818	.150

Bảng 5.17 cho biết tóm tắt mô hình khi các biến được đưa dần vào mô hình ở từng bước. Mô hình 1 mô tả quan hệ giữa ST với biến nắng tháng 11 (N_{11}); hệ số $R = 0.609$, $R^2 = 0.371$. Mô hình 2 mô tả quan hệ giữa ST với biến nắng tháng 11 và 12 (N_{11} , N_{12}); hệ số $R = 0.701$, $R^2 = 0.491$. So sánh hệ số R^2 của mô hình 1 và 2 cho thấy, khi đưa thêm biến N_{12} vào mô hình 2 thì hệ số R^2 tăng lên một lượng là 0.120. Tương tự như vậy, ở mô hình 3 và 4 đã đưa thêm tương ứng biến N_1 và N_2 vào mô hình, và từng biến này đã làm cho hệ số R^2 của mô hình mới tăng lên tương ứng là 0.385 và 0.105. Để biết biến N_{12} , N_1 và N_2 đưa vào mô hình

có làm cho R^2 thay đổi một cách có ý nghĩa hay không, bạn hãy quan sát mức ý nghĩa P (Sig. Fchange) của cột 10. Khi $P < 0.05$ (hoặc 0.01) thì biến độc lập đưa thêm vào mô hình đã làm thay đổi R^2 một cách có ý nghĩa lớn. Ở ví dụ này, các biến N_1 và N_2 đưa vào mô hình đã làm cho R^2 thay đổi có ý nghĩa lớn.

Bảng 5.21. Dự đoán cộng tuyến tính (Collinearity Diagnostics)

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions				
				(Constant)	N11	N12	N1	N2
...	...	...	...	...	...	...	...	...
4	1	4.186	1.000	.00	.01	.01	.00	.00
	2	.384	3.302	.00	.01	.08	.16	.04
	3	.291	3.794	.01	.22	.13	.01	.06
	4	.122	5.853	.01	.55	.63	.12	.02
	5	1.721E-02	15.597	.98	.22	.15	.71	.88

Bảng 5.22. Các hệ số của mô hình phù hợp

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
4	(Constant)	5.815	.470		12.363	.000	4.509	7.121					
	N11	-.747	.160	-.407	-4.670	.010	-1.191	-.303	-.609	-.919	-.314	.595	1.680
	N12	-.470	.121	-.352	-3.878	.018	-.807	-.134	-.602	-.889	-.261	.550	1.819
	N1	-1.869	.187	-.967	-10.01	.001	-2.388	-1.351	-.611	-.981	-.673	.484	2.065
	N2	-1.423	.295	-.530	-4.827	.008	-2.242	-.605	.129	-.924	-.324	.375	2.664

Phân tích mô hình 4 của **Bảng 5.18 và 5.19** cho thấy, mô hình với 4 biến (N_{11} , N_{12} , N_1 và N_2) là phù hợp ($P < 0.01$). Vì thế, quan hệ giữa ST với số giờ nắng (N_i) của các tháng có dạng: $ST = 5.815 - 0.747 \times N_{11} - 0.470 \times N_{12} - 1.869 \times N_1 - 1.423 \times N_2$; với $R\text{-Squared} = 0,982$. Điều đó cho thấy 98,2% biến động của sản lượng Vải được giải thích bởi biến động của số giờ nắng tháng 11 và tháng 12 năm trước đến tháng 1 và tháng 2 năm sau; trong đó biến động ST phụ thuộc lớn nhất vào biến động số giờ nắng tháng 1 và tháng 2. Phân tích chi tiết **Bảng 5.16 và 5.19** còn nhận thấy, khi số giờ nắng từ tháng 11 đến tháng 2 tăng lên thì sản lượng Vải giảm xuống.

5.4.3. Tương quan và hồi quy đa phi tuyến tính

Khi mối liên hệ giữa biến phụ thuộc Y với các biến độc lập X_i ($i = 1, 2, 3 \dots k$) không tuân theo quan hệ đa tuyến tính thì bạn có thể làm phù hợp số liệu với mối liên hệ đa phi tuyến tính. Mối liên hệ đa phi tuyến tính có thể xảy ra trong những trường hợp sau đây:

+ Có ít nhất một biến độc lập biểu hiện mối liên hệ phi tuyến tính với Y. Trong trường hợp này, mối liên hệ phi tuyến tính giữa Y với X_i có dạng:

$$Y = a + b_1X_1 + b_2X_1^2 + b_3X_2 + b_4X_2^2 \quad (5.71)$$

+ Có ít nhất hai biến độc lập biểu hiện mối liên hệ phi tuyến tính với Y. Trong trường hợp này, mối liên hệ phi tuyến tính giữa Y với X_i có dạng:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_1X_2 \quad (5.72)$$

+ Cả hai trường hợp trên cùng xảy ra. Trong trường hợp này, mối liên hệ phi tuyến tính giữa Y với X_i có dạng:

$$Y = a + b_1X_1 + b_2X_1^2 + b_3X_2 + b_4X_2^2 + b_5X_1X_2 \quad (5.73)$$

Một cách tổng quát: $Y = a + b_1X_1 + b_2X_1^2 + b_3X_2 + b_4X_2^2 + b_5X_1X_2 + b_6X_1^2X_2 + b_7X_1X_2^2 + b_8X_1^2X_2^2$. Để dễ dàng tính toán hồi quy đa phi tuyến, trước hết bạn có thể áp dụng kỹ thuật tuyến tính hóa. Ví dụ:

Từ phương trình 5.71, ta có:

$$\text{Đặt } Z_1 = X_1^2, Z_2 = X_2^2 \Rightarrow Y = a + b_1X_1 + b_2Z_1 + b_3X_2 + b_4Z_2$$

Từ phương trình 5.72 ta có:

$$\text{Đặt } Z_1 = X_1X_2 \Rightarrow Y = a + b_1X_1 + b_2X_2 + b_3Z_1 \dots$$

Sau đó, bạn thực hiện các bước phân tích thống kê giống như phân tích hồi quy đa tuyến tuyến.

Nói chung, trong trường hợp cần liên kết nhiều mối liên hệ riêng rẽ giữa y với x_i , bạn cần thực hiện theo các bước sau đây:

Bước 1. Thăm dò mối quan hệ riêng biệt giữa y với từng x_i nhằm xác lập Hình thái quan hệ giữa chúng. Kết quả của bước này cho ta các quan hệ sau đây:

$$y = f_1(x_1); y = f_2(x_2); \dots; y = f_m(x_m)$$

Bước 2. Xác lập quy tắc liên kết các hàm $y = f_i(x_i)$ sao cho quy tắc ấy có thể biểu diễn tốt quan hệ giữa y với các x_i .

Bước 3. Xây dựng hàm hồi quy biểu diễn quan hệ giữa y với các x_i .

Ví dụ. Giả sử cần xác lập hàm hồi quy $y = f(x_1, x_2)$. Cách xây dựng hàm hồi quy như sau:

+ Thăm dò mối quan hệ riêng biệt giữa y với từng x_i . Giả thiết quan hệ giữa y với x_1 có dạng bậc 2 ($y = a + bx_1 + cx_1^2$), còn giữa y với x_2 có dạng hàm số mũ ($y = ae^{-x_2}$).

+ Chọn quy tắc liên kết là tổ hợp tuyến tính:

$$y = Af_1(x_1) + Bf_2(x_2)$$

Thay các quan hệ tương ứng vào ta có:

$$y = Aa + Abx_1 + Acx_1^2 + Bae^{-x_2}$$

$$\text{hay } y = b_0 + b_1x_1 + b_2x_1^2 + b_3e^{-x_2} \quad (5.74)$$

+ Xây dựng hàm hồi quy. Đặt $Z_1 = x_1$, $Z_2 = x_1^2$, $e^{-x_2} = Z_3$.

Từ đó ta có $y = b_0 + b_1Z_1 + b_2Z_2 + b_3Z_3$. Sau đó bạn thực hiện các bước phân tích thống kê giống như phân tích hồi quy đa tuyến tuyến.

5.5. MỘT SỐ MÔ HÌNH MÔ TẢ QUAN HỆ GIỮA CÁC NHÂN TỐ ĐIỀU TRA CÂY CÁ THỂ VÀ QUẦN THỂ

5.5.1. Quan hệ giữa các nhân tố điều tra thân cây với tuổi cây

Mối quan hệ giữa đường kính thân cây ($D_{1.3}$, cm), chiều cao thân cây (H , m), tiết diện ngang thân cây (G , m^2) và thể tích thân cây (V , m^3) với tuổi cây (kí hiệu A hoặc T , năm) có thể được mô tả bằng nhiều dạng mô hình hồi quy phi tuyến khác nhau. Dưới đây chỉ giới thiệu một số mô hình hồi quy phi tuyến được sử dụng nhiều trong lâm học.

a. Mô hình Gomperts

$$Y = Y_{\max} \exp[-b \times \exp(-kT)]; \quad (5.75)$$

trong đó: T là tuổi cây; $Y_{\max}$ – chiều cao (hoặc đường kính) thân cây lớn nhất ở tuổi thành thực.

Để xác định các tham số của mô hình Gomperts, chúng ta thực hiện những biến đổi như sau:

$$\text{Đặt } Y/Y_{\max} = \text{Exp}[-b \times \exp(-kT)] \Rightarrow \text{Ln}(Y/Y_{\max}) = -b \times \exp(-kT)$$

$$\Rightarrow \text{Ln}(-\text{Ln}(Y/Y_{\max})) = \text{Lnb} - kT$$

$$\text{Đặt } Y' = \text{Ln}(-\text{Ln}(Y/Y_{\max})); A = \text{Lnb}; k = B.$$

$$\Rightarrow Y' = A + BT. \quad (5.76)$$

Giải phương trình (5.76) bằng phương pháp bình phương nhỏ nhất sẽ tìm được A và B. Khi biết $A = \text{Ln}b \Rightarrow b = e^A$. Từ đây có thể trở lại hàm phi tuyến.

b. Mô hình Schumacher

$$Y = Y_{\max} \exp(-b/T^m) \text{ hay } Y = Y_{\max} \exp(-bT^{-m}) \quad (5.77)$$

trong đó: T – tuổi rừng; $Y_{\max}$ – chiều cao (hoặc đường kính...) thân cây lớn nhất ở tuổi thành thực. Lưu ý: $m = 0,2 \div 2$; b luôn lấy giá trị âm.

Để xác định các tham số của mô hình Schumacher, chúng ta thực hiện những biến đổi như sau:

+ Cách thứ nhất. Đặt $Y/Y_{\max} = \exp(-bT^{-m})$

$$\Rightarrow \text{Ln}(Y/Y_{\max}) = -bT^{-m} \Rightarrow \text{Ln}(-\text{Ln}(Y/Y_{\max})) = \text{Ln}b - m\text{Ln}T.$$

$$\text{Đặt } Y' = \text{Ln}(-\text{Ln}(Y/Y_{\max})); \text{Ln}b = A; m = B; \text{Ln}T = T'$$

$$\Rightarrow Y' = A + BT'. \quad (5.78)$$

Giải phương trình (5.78) bằng phương pháp bình phương nhỏ nhất sẽ tìm được A và B. Khi biết $A = \text{Ln}b \Rightarrow b = e^A$; $\text{Ln}T = T' \Rightarrow T = e^{T'}$. Từ đây có thể trở lại hàm phi tuyến.

+ Cách thứ hai. Từ hàm 5.77, lấy logarit tự nhiên hai vế:

$$\text{Ln}Y = \text{Ln}Y_{\max} + b/T^m$$

$$\text{Đặt } \text{Ln}Y = Y'; \text{Ln}Y_{\max} = A; b = B; 1/T^m = T' \text{ với } m = 0,2 - 2,0.$$

$$\Rightarrow Y' = A + BT'.$$

Lưu ý rằng, giá trị m được chọn sao tổng bình phương các sai lệch giữa giá trị của biến y_i và giá trị ước lượng của biến y_i là nhỏ nhất, nghĩa là $\sum (y_i - \hat{y}_i)^2 = \min$.

c. Mô hình Korf

$$Y = Y_{\max} \exp(-C_1 T^{-C_2}) \quad (5.79)$$

trong đó: T – tuổi rừng; $Y_{\max}$ – chiều cao (hoặc đường kính) thân cây lớn nhất ở tuổi thành thực. Các tham số của phương trình Korf có quan hệ với nhau như sau: $Y_{\max} = e^{C_1}$, $C_1 = b(n-1)$, $C_2 = n-1$.

Để xác định các tham số của mô hình Korf, chúng ta thực hiện những biến đổi như sau:

$$\text{Đặt } Y/Y_{\max} = \exp(-C_1 T^{-C_2}) \Rightarrow -\text{Ln}(Y/Y_{\max}) = C_1 T^{-C_2}$$

$$\Rightarrow \text{Ln}(-\text{Ln}(Y/Y_{\max})) = \text{Ln}C_1 - C_2 \text{Ln}T$$

$$\Rightarrow \text{Đặt } \text{Ln}(-\text{Ln}(Y/Y_{\max})) = Y'; \text{Ln}T = X, \text{Ln}C_1 = A, C_2 = B$$

$$\Rightarrow Y = A + BX. \quad (5.80)$$

Giải phương trình (5.80) bằng phương pháp bình phương nhỏ nhất sẽ tìm được A và B. Khi biết $A \Rightarrow C_1 = 2.7182^A$; $C_2 = B$. Từ đây có thể trở lại hàm phi tuyến.

d. Những mô hình khác

Mô hình Jonhson-Schumacher

$$Y = Y_{\max} \times \exp\{-b_0/(X + b_1)\} \quad (1)$$

trong đó:

Y = biến số phụ thuộc (chẳng hạn D, H, G, V);
 $Y_{\max}$ = giá trị lớn nhất của biến phụ thuộc;
 Exp = cơ số Neper = 2,7182;
 b_0, b_1 = những tham số của mô hình;
 X = biến số độc lập (chẳng hạn tuổi cây A năm).

Hàm số mũ (Exponential Model)

$Y = \text{Exp}(a + b \times X)$
 Y = biến số phụ thuộc;
 Exp = cơ số Neper = 2,7182;
 a, b = những tham số của mô hình;
 X = biến số độc lập.

Hàm số mũ

$Y = \alpha \exp(\beta/X^k)$
 Y = biến số phụ thuộc;
 Exp = cơ số Neper = 2,7182;
 α, β và k (lấy giá trị từ 0.2 – 2) = những tham số của mô hình;
 X = biến số độc lập.

Mô hình Sigmaoid (S – Curve Model)

$Y = \text{Exp}(a + b/X)$
 Y = biến số phụ thuộc;
 Exp = cơ số Neper = 2,7182;
 a, b = những tham số của mô hình;
 X = biến số độc lập.

Hàm lũy thừa (Multiplicative, Power)

$Y = aX^b$
 Y = biến số phụ thuộc;
 a, b = những tham số của mô hình;
 X = biến số độc lập.

Hàm số nghịch đảo của Y (Reciprocal – Y)

$Y = 1/(a + b \times X)$
 Y = biến số phụ thuộc;
 a, b = những tham số của mô hình;
 X = biến số độc lập.

5.5.2. Quan hệ giữa tổng giá trị gỗ (S_A) với tuổi rừng (A)

+ Hàm số căn bậc hai của A (Square Root – A): $SA = a + b \times \text{SQRT}(A)$

- + Hàm số mũ (Exponential/Growth Model): $SA = \text{Exp}(a + b \times A)$
- + Đường cong Hình chữ S (S – Curve Model): $SA = \text{Exp}(a + b/A)$
- + Hàm số logarit (logarithmic): $SA = a + b \times \text{Ln}A$
- + Hàm số mũ biến đổi (Multiplicative/Power): $SA = aA^b$

5.5.3. Quan hệ giữa tỉ lệ gia tăng giá trị gỗ hàng năm ($\Delta Sa/Sa$) với tuổi rừng (A)

Quan hệ này có thể mô tả theo các dạng phương trình:

- + Hàm số căn bậc hai của $\Delta Sa/Sa$ (Square Root – Y): $\Delta Sa/Sa = (a + b \times A)^2$
- + Hàm số nghịch đảo của $\Delta Sa/Sa$ (Reciprocal – Y): $\Delta Sa/Sa = 1/(a + b \times A)$
- + Hàm số mũ (Exponential/Growth Model): $\Delta Sa/Sa = \text{Exp}(a + b \times A)$
- + Đường cong Hình chữ S (S – Curve Model): $\Delta Sa/Sa = \text{Exp}(a + b/A)$
- + Hàm số logarit (Logarithmic): $\Delta Sa/Sa = a + b \times \text{Ln}A$
- + Hàm số mũ biến đổi (Multiplicative/Power): $\Delta Sa/Sa = aA^b$
- + Hàm số logistic. Hàm logistic có nhiều dạng khác nhau.

Ví dụ: $\Delta Sa/Sa = Y = \text{Exp}(a+bx)/(1+\text{Exp}(a+bx))$. R^2 biểu thị % biến đổi của $\Delta Sa/Sa$ theo A sau khi biến đổi tỉ lệ $Y/(1-Y)$ thành mô hình tuyến tính.

5.5.4. Quan hệ giữa suất tăng trưởng đường kính (Pd) và chiều cao (Ph) với tuổi rừng (A, năm)

$Pd = K \times A^{-b}$. Giải: $\text{Ln}(Pd) = \text{Ln}K - b \text{Ln}A$

$Ph = K \times A^{-b}$. Giải: $\text{Ln}(Ph) = \text{Ln}K - b \text{Ln}A$

Ở đây Pd và Ph tính theo công thức:

a. Đối với những loài cây sinh trưởng nhanh

$$Pd = \frac{D_a - D_{a-n}}{D_a} \times 100; Ph = \frac{H_a - H_{a-n}}{H_a} \times 100$$

b. Đối với những loài cây sinh trưởng chậm

Nhân tố Pt (t là nhân tố điều tra trên thân cây) được tính theo công thức:

$$Pt = \frac{T_a - T_{a-n}}{T_a + T_{a-n}} \times \frac{200}{n}; Pv = Pg + Ph + Pf; Zt = T \times Pt\%; Ta \pm n = Ta(1 \pm nPt\%).$$

5.5.5. Quan hệ giữa thể tích thân cây với các nhân tố điều tra trên thân cây

Một số mô hình sau đây biểu thị tốt quan hệ giữa V với D và H:

- + $V = a + bH + cD^2 \times H/10^4$
- + $V = a + bD^2H$
- + $V = KD^aH^b$
- + $V = K(D^2H)^b$
- + $\text{Ln}V = a + b \text{Ln}D + c \text{Ln}D^2 + d \text{Ln}H + e \text{Ln}H^2$

5.5.6. Quan hệ giữa chiều cao thân cây (H) với đường kính thân cây (D)

$$H - 1.3 = a \text{Exp}(-b/D)$$

$$H = a^{(b \text{Ln}D - c \text{Ln}D^2)}$$

5.5.7. Dự đoán một số nhân tố điều tra quần thụ

- + $Dg = a + b \times H_0^{0.2}$
- + $M = a + bGH_0$
- + $M/G = a + bH_0$
- + $M/G = a + bH_0 + cH_0^2$
- + $\text{Ln}G = a + b \text{Ln}H_0 + c \text{Ln}N$
- + $\text{Ln}St = a + b \text{Ln}H_0 + c \text{Ln}N$
- + $\text{Ln}M = a + b \text{Ln}H_0 + c \text{Ln}N$
- + $\text{Ln}M = a + b \text{Ln}H + c \text{Ln}G$ (Trịnh Đức Huy, 1988)

5.5.8. Quan hệ giữa trữ lượng gỗ sản phẩm (M_{sf}) với tuổi rừng (A, năm)

- + $M_{sf} = a + b \ln A$
- + $M_{sf} = \exp(a + b/A)$
- + $\ln M_{sf} = a + b \ln D$, với D (cm) là đường kính thân cây.

5.5.9. Sử dụng thủ tục hồi quy phi tuyến để xác định các hệ số của mô hình

Để làm phù hợp những hàm số phi tuyến (Nonlinear functions) với dãy số liệu quan sát, yêu cầu phải xác định được những tham số của mô hình sao cho tổng sai lệch bình phương giữa những trị số quan sát và những trị số ước lượng (dự đoán, lý thuyết) là nhỏ nhất, nghĩa là min $S(y_i - \hat{y}_i)^2$.

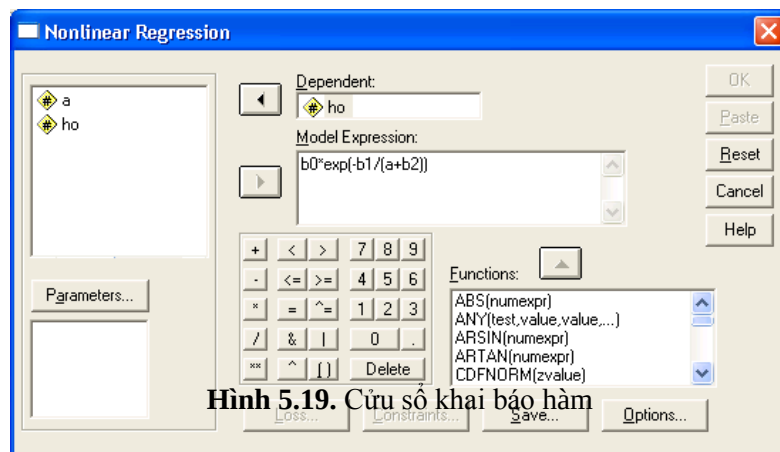
Các tham số của những mô hình phi tuyến có thể xác định bằng hai thủ tục: (1) hồi quy tuyến tính (Linear Regression), (2) hồi quy phi tuyến tính (Nonlinear Regression). Thủ tục hồi quy tuyến tính đã được giới thiệu ở **mục 5.1 của chương 5**. Phần dưới đây chúng ta sử dụng thủ tục hồi quy phi tuyến (Nonlinear Regression) để xác định các tham số của những mô hình phi tuyến mà không cần phải thực hiện bước tuyến tính hoá các mô hình phi tuyến.

Thủ tục hồi quy phi tuyến

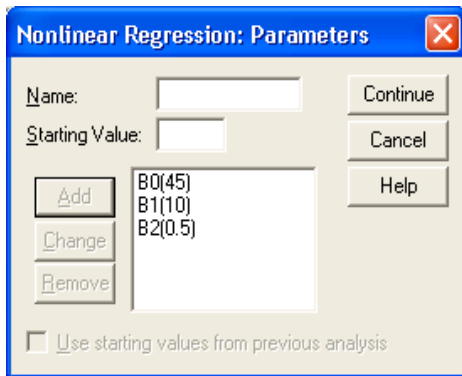
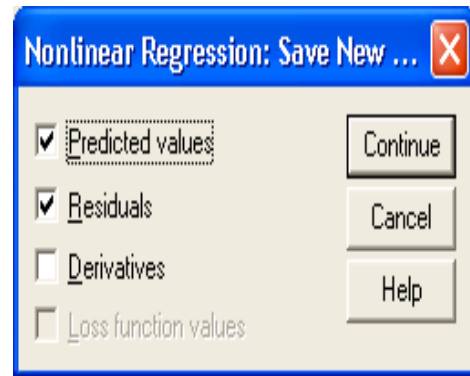
Trước hết, bạn vào đường lệnh hồi quy phi tuyến: *Analyze > Regression > Nonlinear*. Kế đến khai báo những điều kiện của mô hình phi tuyến như sau:

- Biến phụ thuộc Y. Biến này được khai báo trong cửa sổ Dependent bằng cách chọn Y và chuyển vào cửa sổ Dependent (**Hình 5.19**).
- Mô hình được chọn để làm phù hợp với dãy số liệu. Mô hình này được ghi trong cửa sổ Model expression (**Hình 5.19**).
- Những tham số giả định ban đầu của mô hình. Những tham số này được khai báo trong cửa sổ Parameters (**Hình 5.20**). Lưu ý rằng, sau mỗi lần khai báo giá trị ban đầu của một hệ số, bạn chọn Add để chương trình ghi nhận. Nếu khai báo sai hoặc cần thay đổi tham số, bạn chọn Remove và khai báo lại các tham số.
- Chọn những giá trị dự đoán (Predicted value) và những sai lệch (Residuals) trong thanh Save (**Hình 5.21**).
- Cuối cùng chọn OK.

Kết quả báo cáo những thống kê tóm tắt của hồi quy phi tuyến như phân tích ANOVA, hệ số tương quan và các tham số của mô hình.



Hình 5.19. Cửa sổ khai báo hàm

**Hình 5.20.** Khai báo tham số ban đầu**Hình 5.21.** Chọn giá trị dự đoán và sai

Lưu ý rằng, trong phân tích hồi quy phi tuyến việc khai báo những giá trị ban đầu của các tham số của mô hình là một thủ tục rất quan trọng, nhưng việc này cũng rất khó khăn. Bởi vì, nếu khai báo gần đúng những giá trị ban đầu của các tham số, thì thủ tục hồi quy phi tuyến cho phép dự đoán những giá trị của biến phụ thuộc chính xác hơn (min $S(y_i - y_i)^2$ nhỏ hơn) so với thủ tục hồi quy tuyến tính. Ngược lại, nếu những giá trị ban đầu của các tham số được khai báo không đúng hoặc chênh lệch quá lớn so với giá trị gần đúng nhất, thì thủ tục hồi quy phi tuyến hoặc báo lỗi hoặc báo cáo những giá trị dự đoán với sai lệch rất lớn.

Phân tích hồi quy trong lâm học là nhằm những mục đích sau đây: (1) tìm được một hàm số phản ánh đúng quy luật biến đổi của các nhân tố điều tra, (2) tìm được một hàm số để mô tả đúng mối quan hệ giữa các nhân tố điều tra với các biến môi trường; (3) tìm được một hàm số để làm phù hợp số liệu và sử dụng nó để dự đoán sao cho tổng sai lệch bình phương nhỏ nhất, nghĩa là min $S(y_i - y_i)^2$.

Nếu mục đích đặt ra là tìm được một hàm phi tuyến để làm phù hợp số liệu với sai lệch dự đoán nhỏ nhất, bạn có thể sử dụng thủ tục hồi quy phi tuyến. Tuy vậy, như đã nói ở trên, mục đích này chỉ có thể đạt được khi chúng ta khai báo gần đúng nhất những giá trị ban đầu của các tham số. Vậy, bằng cách nào có thể khai báo gần đúng nhất những giá trị ban đầu của các tham số của hàm phi tuyến? Đây là một công việc rất khó khăn, nhất là đối với những ai thiếu kinh nghiệm. Dưới đây giới thiệu một “mẹo vặt” để giúp bạn khai báo gần đúng nhất những giá trị ban đầu của các tham số của mô hình phi tuyến.

Trước hết, bạn hãy chọn những hàm phi tuyến để làm phù hợp với số liệu đã cho. Công việc này có thể thực hiện bằng nhiều cách khác nhau, chẳng hạn như dựa vào lý thuyết và kinh nghiệm, sử dụng thủ tục so sánh các hàm hồi quy trong Statgraphics và SPSS, vẽ đồ thị... Kế đến, biến đổi hàm phi tuyến sang dạng tuyến tính và tìm những hệ số của hàm phi tuyến bằng thủ tục hồi quy tuyến tính. Đến đây, bạn sử dụng những giá trị của những tham số vừa tìm được để khai báo trong cửa sổ các tham số (parameters) của thủ tục hồi quy phi tuyến. Cần lưu ý rằng, khi những tham số được giả định bằng những giá trị ban đầu khác nhau, thì các hàm phi tuyến sẽ nhận những hệ số khác nhau và tổng sai lệch bình phương cũng khác nhau. Do đó, để nhận được những tham số của hàm phi tuyến hợp lý nhất, bạn cần chú ý đến điều kiện “tổng sai lệch bình phương nhỏ nhất, nghĩa là min $S(y_i - y_i)^2$ ”. Để thỏa mãn điều kiện này, bạn cần phải khai báo lại một vài lần những giá trị ban đầu của các tham số của hàm phi tuyến. Mỗi lần khai báo lại, bạn chỉ cần thay đổi một vài đơn vị phần mười của mỗi tham số. Thủ tục sẽ dừng khi bạn đã nhận được tổng sai lệch bình phương nhỏ nhất.

Thủ tục xác định các hệ số của mô hình hồi quy phi tuyến

Bảng 5.23 ghi lại những số liệu về sinh trưởng đường kính thân cây (D, cm) của loài Keo lá tràm (*Acasia auriculiformis*) từ 2 – 21 tuổi, mọc trên đất xám phù sa cổ ở miền Đông Nam Bộ. Từ số liệu này, yêu cầu đặt ra là làm phù hợp hàm Gomperts và hàm Schumacher với số liệu đã cho để mô tả mối quan hệ giữa đường kính thân cây (D, cm) với tuổi cây (A, năm). Để đạt được yêu cầu trên đây, dưới đây chúng ta thử sử dụng hai thủ tục hồi quy tuyến tính và hồi quy phi tuyến để xác định các tham số của hàm Gomperts và hàm Schumacher.

Bảng 5.23. Sinh trưởng đường kính thân cây (D, cm) của loài Keo lá tràm (*Acasia auriculiformis*) từ 2 – 21 tuổi

A, năm	2	3	4	5	6	7	8	9	10
D, cm	4.73	6.19	7.51	8.84	10.12	11.31	12.26	13.29	14.21
A, năm	11	12	13	14	15	16	17	18	19
D, cm	15.16	16.01	16.76	17.5	18.28	19.17	20	20.61	21.47
A, năm	20	21							
D, cm	22.3	23.1							

Hàm Gomperts có dạng:

$$Y = Y_{\max} \times \exp(-b_0 \times \exp(-b_1 \times A))$$

Trong đó:

$Y_{\max}$ = D lớn nhất ở tuổi thành thực, cm;

Exp = cơ số Neper = 2,7182;

b_0 và b_1 = những tham số của mô hình cần tìm;

A = biến số tuổi cây, năm.

+ *Xác định các tham số của hàm Gomperts bằng thủ tục hồi quy tuyến tính*

Trước hết, giả định $Y_{\max} = 70$ cm và bằng thủ tục tuyến tính hoá, có thể xác định được hai tham số của hàm Gomperts như sau: $b_0 = -2.57685$; $b_1 = -0.04284$. Mô hình tuyến tính mô tả quan hệ giữa đường kính thân cây với tuổi có dạng:

$$Y' = 0.946596 - 0.04284 \times A, \text{ với } R = 0.981; S_e = 0.0507$$

$$\text{Hay } D_{1.3} = 70 \times \exp(-2.57685 \times \exp(-0.04284 \times A)) \quad (\times)$$

Tổng sai lệch bình phương giữa D thực tế và D tính theo mô hình (×) là 17,542.

+ *Xác định các tham số của hàm Gomperts bằng thủ tục hồi quy phi tính*

Trước hết, chuyển dữ liệu D và A vào Bảng tính SPSS 10.0. Kế đến mở đường dẫn vào thủ tục hồi quy phi tuyến: Analyze > Regression > Nonlinear. Tiếp theo chọn Y và chuyển vào khoảng Dependent. Sau đó khai báo mô hình Gomperts trong khoảng Model expression theo dạng: $Y_{\max} \times \exp(-b_0 \times \exp(-b_1 \times A))$. Cuối cùng khai báo những giá trị ban đầu của các tham số của hàm Gomperts trong thanh Parameters. Để tìm các tham số của hàm Gomperts và so sánh với cách xác định các tham số bằng thủ tục tuyến tính hoá, ở đây chúng

ta thử ba giả định: (1) $Y_{\max} = 70$, $b_0 = 3.0$ và $b_1 = 0.3$; (2) $Y_{\max} = 70$, $b_0 = 2.0$ và $b_1 = 0.2$; (3) $Y_{\max} = 70$, $b_0 = 2.5$ và $b_1 = 0.04$ (những tham số gần giống với thủ tục tuyến tính).

Sau khi thực hiện thủ tục hồi quy phi tuyến, kết quả nhận được các tham số $Y_{\max}$, b_0 , b_1 và tổng bình phương sai lệch như sau:

+ Đối với trường hợp 1, $Y_{\max} = 27.3938$, $b_0 = -2.0172$, $b_1 = -0.1111$, $R^2 = 0.9963$. Tổng sai lệch bình phương nhỏ nhất là 2.143, trung bình 0.1261.

+ Đối với trường hợp 2, $Y_{\max} = 27.39376$, $b_0 = -2.01713$, $b_1 = -0.11111$, $R^2 = 0.9963$. Tổng sai lệch bình phương nhỏ nhất là 2.143, trung bình 0.1261.

+ Đối với trường hợp 3, $Y_{\max} = 27.39377$, $b_0 = -2.01713$, $b_1 = -0.11111$, $r = 0.996$. Tổng sai lệch bình phương nhỏ nhất là 2.143, trung bình 0.1261.

So sánh những kết quả tính toán trên đây cho thấy, khi sử dụng thủ tục hồi quy phi tuyến để làm phù hợp hàm Gomperts với số liệu, thì tổng sai lệch bình phương trong cả ba trường hợp trên đều bằng nhau (2.143), và nhỏ hơn so với tổng sai lệch bình phương được tính theo thủ tục hồi quy tuyến tính (17.542) tương ứng là 8 lần.

Hàm Schumacher có dạng:

$Y = Y_{\max} \times \exp(-b_0 \times A^{-b_1})$
 Y = biến số phụ thuộc;
 $Y_{\max}$ = giá trị lớn nhất của biến phụ thuộc;
 $\exp$ = cơ số Neper = 2,7182;
 b_0 và b_1 = những tham số của mô hình;
 A = tuổi cây.

+ *Xác định các tham số của hàm Schumacher bằng thủ tục hồi quy tuyến tính*

Trước hết, giả định $Y_{\max} = 70$ cm và bằng thủ tục tuyến tính hoá, đã xác định được hai tham số của hàm Schumacher như sau: $b_0 = -3.76679$; $b_1 = 0.38444$. Mô hình tuyến tính mô tả quan hệ giữa đường kính thân cây với tuổi có dạng:

$$Y' = 1.32626 - 0.38444 \times \ln A, \text{ với } R = 0.993; S = 0.0292$$

$$\text{Hay } D_{1.3} = 70 \times \exp(-3.76679 \times A^{-0.38444}) \quad (xx)$$

Tổng sai lệch bình phương giữa D thực tế và D tính theo mô hình (xx) là 6,67.

+ *Xác định các tham số của hàm Schumacher bằng thủ tục hồi quy phi tính*

Để tìm các tham số của hàm Schumacher và so sánh với các tham số được xác định bằng thủ tục tuyến tính hoá, ở đây thử ba giả định: (1) $Y_{\max} = 70$, $b_0 = 4.0$ và $b_1 = 0.4$ (những tham số gần giống với thủ tục tuyến tính); (2) $Y_{\max} = 70$, $b_0 = -3.0$ và $b_1 = -0.3$; (3) $Y_{\max} = 70$, $b_0 = 2.0$ và $b_1 = 0.2$.

Sau khi thực hiện thủ tục hồi quy phi tuyến, kết quả nhận được các tham số $Y_{\max}$, b_0 , b_1 và tổng bình phương sai lệch như sau:

+ Đối với trường hợp 1, $Y_{\max} = 11171.2338$, $b_0 = -8.3702$, $b_1 = -0.09889$. Tổng sai lệch bình phương nhỏ nhất là 0.306, trung bình 0.01798; $R^2 = 0.9995$.

+ Đối với trường hợp 2, $Y_{\max} = 9477.26$, $b_0 = -8.21037$, $b_1 = -0.10138$. Tổng sai lệch bình phương nhỏ nhất là 0.316, trung bình 0.0186; $R^2 = 0.9995$.

+ Đối với trường hợp 3, $Y_{\max} = 8274.903$, $b_0 = -8.0787$, $b_1 = -0.10353$. Tổng sai lệch bình phương nhỏ nhất là 0.327, trung bình 0.01916; $R^2 = 0.9994$.

So sánh những kết quả tính toán trên đây cho thấy, khi sử dụng thủ tục hồi quy phi tuyến tính để làm phù hợp hàm Schumacher với số liệu, thì tổng sai lệch bình phương ở trường hợp 1 và 2 (tương ứng 0.306 và 0.316) nhỏ hơn so với tổng sai lệch bình phương được tính theo thủ tục hồi quy tuyến tính (6.67) tương ứng là 22 và 21 lần.

Lưu ý rằng, khi những tham số được giả định bằng những giá trị ban đầu khác nhau, thì các hàm phi tuyến sẽ nhận những hệ số khác nhau. Ngoài ra, đối với hai hàm Gomperts và Schumacher, nếu cố định hệ số b_0 và b_1 và thay đổi giá trị $Y_{\max}$ lớn dần, thì tổng sai lệch bình phương sẽ nhận giá trị nhỏ dần. Vì thế, để nhận được hàm phi tuyến mô tả tốt mối quan hệ giữa hai biến Y và X với tổng sai lệch bình phương nhỏ nhất, bạn có thể giả định những giá trị ban đầu của các tham số bằng những giá trị khác nhau. Thủ tục sẽ dừng khi mô hình ước lượng nhận được những giá trị dự đoán với sai số thoả mãn những điều kiện đặt ra.

Chương VI

QUAN HỆ GIỮA CÁC BIẾN SỐ CÓ THANG ĐO KHÁC NHAU

Để xác định mối quan hệ giữa các biến số được đo bằng các thang đo khác nhau, bạn có thể sử dụng thủ tục Crosstabs (lập Bảng chéo). Thủ tục Crosstabs là một công cụ thống kê rất hữu ích trong những trường hợp sau đây :

1. Xác định tần số lý thuyết và phần trăm các trường hợp (sự kiện) bất gặp trong một cuộc nghiên cứu về một vấn đề nào đó mà các đại lượng của chúng là những số đo không có thứ bậc. Ví dụ: Xác định tần số ô bất gặp và không bất gặp một hay nhiều loài cây nào đó qua một chuỗi các ô mẫu trong một quần xã thực vật.

2. Kiểm định về tính độc lập của hai biến định tính không có thứ bậc hơn kém. Ví dụ: Kiểm định sự kết nhóm sinh thái (Ecological Groups) giữa hai hay nhiều loài cây trong một quần xã thực vật.

3. Phân tích và đo đặc mối quan hệ giữa hai biến số (đại lượng) được đo bằng các thang đo khác nhau. Ví dụ: Một biến số định lượng, còn biến số kia được đo bằng thang đo định danh (có thứ bậc hoặc không có thứ bậc).

4. Kiểm tra dữ liệu để phát hiện những sai sót...

Dưới đây chúng ta xem một số cách thức và phương pháp xử lý thông tin trong lâm học bằng cách lập Bảng chéo.

6.1. KIỂM ĐỊNH χ^2 (Chi – square)

Thống kê Pearson Chi – square (χ^2) được sử dụng cho hai kiểm định: (1) Kiểm định về tính độc lập giữa biến cột với biến dòng của Bảng chéo 2 chiều (còn được gọi là Bảng 2x2), (2) Kiểm định về ngang bằng tỉ lệ dọc theo hàng hoặc theo cột của Bảng chéo 2 chiều.

6.1.1. Kiểm định về tính độc lập của các đại lượng

Một câu hỏi đặt ra là hai biến A và B được đo bằng hai dấu hiệu định tính không có thứ bậc hơn kém có quan hệ với nhau hay không ? Ví dụ: Khi nghiên cứu về các nhóm sinh học (Biological Groups), nhà lâm học thường thực hiện điều tra N ô tiêu chuẩn dọc theo một quần xã thực vật; trong đó dấu hiệu về sự có mặt của hai loài cây A và B trên các ô tiêu chuẩn chỉ được ghi nhận bằng 2 dấu hiệu định tính đối lập nhau là “có” và “không có”. Vấn đề đặt ra là hai loài cây A và B có quan hệ với nhau hay không ? Để trả lời câu hỏi này, nhà lâm học có thể xây dựng một Bảng chéo 2x2 về tần suất (%) bất gặp hai loài trên N ô tiêu chuẩn. Sau đó thực hiện phép kiểm định về tính độc lập của chúng bằng thống kê χ^2 .

Khi sử dụng thống kê χ^2 để kiểm định về tính độc lập của hai biến định tính không có thứ bậc, bạn có thể thực hiện theo trình tự như sau (trên ví dụ hai loài cây):

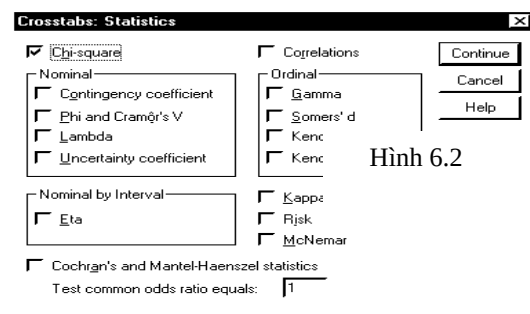
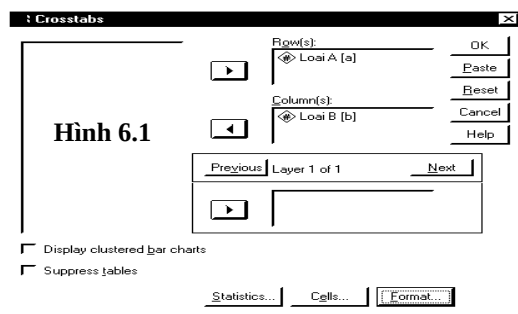
+ **Thu thập dữ liệu.** Trước hết, bạn phải bố trí một chuỗi N ô tiêu chuẩn dọc theo quần xã thực vật. Trong mỗi ô tiêu chuẩn, bạn ghi nhận sự có mặt của hai loài cây bằng hai dấu hiệu “có” và “không có”. Để đảm bảo cho phép phân tích thống kê đạt độ chính xác cao, số lượng ô tiêu chuẩn cần điều tra phải đủ lớn để tần số của các ô thuộc Bảng 2x2 ô (hoặc RxC ô) phải lớn hơn 5 hoặc trên 75% số ô thuộc Bảng 2x2 (hoặc RxC) có tần số lớn hơn 5. Nhiều nghiên cứu cho thấy, khi xác định sự kết nhóm giữa hai loài cây gỗ thì dung lượng quan sát không nhỏ hơn 200 ô tiêu chuẩn.

+ **Cách nhập dữ liệu vào Bảng tính SPSS.** Trước hết, bạn hãy mã hóa sự kiện “loài cây quan tâm có mặt trong ô tiêu chuẩn” bằng số 1, còn sự kiện “loài cây quan tâm không có

mặt trong ô tiêu chuẩn” bằng số 0. Sau đó, bạn có thể nhập số liệu này vào Bảng tính SPSS theo hai cách: (1) nhập trực tiếp vào Bảng tính SPSS, (2) nhập vào Bảng tính Excel rồi Copy sang Bảng tính SPSS (Bảng 6.1). Trong Bảng 6.1, cột 1 ghi thứ tự ô điều tra (có hay không cũng được), cột 2 nhập các dấu hiệu của loài A (theo mã hóa), còn cột 3 nhập các dấu hiệu của loài B (theo mã hóa).

Bảng 6.1. Cách thức nhập dữ liệu vào Bảng tính SPSS

TT	Loài cây A	Loài cây B
1	1	1
2	1	0
3	0	1
4	0	0
...	...	...



+ **Gọi chương trình tính toán.** Sau khi nhập xong dữ liệu vào Bảng tính SPSS, bạn thực hiện gọi chương trình tính toán. Để tạo được kết quả, từ thanh Menu chính bạn hãy thực hiện như sau (Dấu $\Rightarrow$ chỉ thứ tự chọn các thủ tục tính toán):

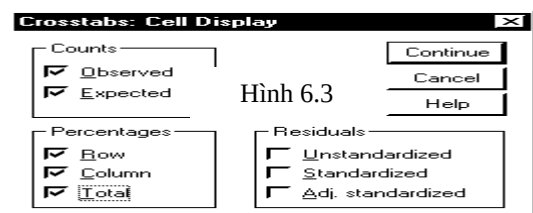
Bước 1. Mở Analyze $\Rightarrow$ Descriptive Statistics $\Rightarrow$ Crosstabs. Khi Bảng Crosstabs xuất hiện (Hình 6.1), bạn hãy chuyển biến “loài A” (cột 2) vào ô Row, còn “loài B” (cột 3) vào ô Column. Bạn cũng có thể chuyển biến “loài B” (cột 3) vào ô Row, còn “loài A” (cột 2) vào ô Column.

Bước 2. Từ cửa sổ Crosstabs, bạn chọn Statistics và đánh dấu chọn thủ tục Chi – square (Hình 6.2). Thủ tục này được thực hiện là nhằm mục đích kiểm định tính độc lập giữa hai loài cây bằng thống kê Chi – square. Tiếp theo chọn Cells và đánh dấu các mục Counts và Percentages (Hình 6.3). Thủ tục này được thực hiện là nhằm mục đích xác định các tần số (thực nghiệm và lý thuyết) của Bảng 2x2. Sau cùng bạn chọn OK để chương trình xử lý và hiển thị kết quả.

Kết quả kiểm định về tính độc lập của hai dấu hiệu báo cáo những Bảng sau đây :

+ Bảng tóm tắt xử lý các trường hợp (Case Processing Summary, Bảng 6.2). Bảng này ghi lại số lượng và phần trăm các trường hợp được đưa vào tính toán.

+ Bảng chéo hai chiều (Bảng 6.3). Nội dung của Bảng này ghi lại tần số (thực tế và lý thuyết hay kì vọng) của các sự kiện bất gặp trong một Bảng chéo hai chiều. Tần số kì vọng của mỗi ô trong Bảng 2x2 (hay RxC) được



tính như sau: Nhân tổng cột với tổng hàng tương ứng rồi chia cho tổng số trường hợp lấy mẫu.

Bảng 6.2. Tóm tắt xử lý các trường hợp (Case Processing Summary)

LoạiA × loạiB	Cases					
	Valid N	Percent	MissingN	Percent	TotalN	Percent
	120	100.0%	0	.0%	120	100.0%

Bảng 6.3. Bảng chéo hai chiều (LoạiA × LoạiB Crosstabulation)

			LoaiB		Total
			0	1	
LoaiA	0	Count	10	30	40
		Expected Count	20.0	20.0	40.0
	1	Count	50	30	80
		Expected Count	40.0	40.0	80.0
Total		Count	60	60	120
		Expected Count	60.0	60.0	120.0

+ Kiểm định Pearson Chi – Square (**Bảng 6.4**). Bảng này dẫn ra các kết quả kiểm định về tính độc lập của hai sự kiện bằng các thống kê Pearson Chi – Square, Likelihood Ratio (tỷ số tương đồng), Linear – by – Linear Association. Những thống kê này kiểm định giả thuyết H_0 : Các biến hàng và biến cột là độc lập với nhau. Đối thuyết H_0 : Các biến hàng và biến cột có quan hệ với nhau. Hai biến là độc lập với nhau khi xác suất mà một trường hợp rơi vào một ô của Bảng R×C ô là tích số của các xác suất biên (tổng cột×tổng hàng/tổng số mẫu N).

Bảng 6.4. Kiểm định Pearson Chi – Square (Chi-Square Tests)

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	15.000	1	0.000	0.000	0.000
Continuity Correction	13.538	1	0.000		
Likelihood Ratio	15.518	1	0.000		
Fisher's Exact Test					
Linear-by-Linear Association	14.875	1	0.000		
N of Valid Cases	120				

Giải thích và nhận xét kết quả. Căn cứ vào kết quả của **Bảng 6.4**, bạn có thể đi đến kết luận về tính độc lập của hai biến. Nếu $\chi^2_{tt} > \chi^2_{0.05}$ hoặc $P^1 < 0.05$ thì giả thuyết H_0 bị bác bỏ. Ngược lại, nếu $\chi^2_{tt} < \chi^2_{0.05}$ hoặc $P > 0.05$ thì giả thuyết H_0 được chấp nhận. Bậc tự do của phân bố $\chi^2 = (\text{số hàng} - 1) \times (\text{số cột} - 1)$.

Lưu ý rằng, bên cạnh kiểm định tính độc lập của hai biến bằng kiểm định Chi – Square, **Bảng 6.4** còn đưa ra 3 kiểm định khác là Likelihood Ratio (tỉ lệ tương đồng), Yates' Continuity Correction (hệ số hiệu chỉnh Yates) và Fisher's Exact Test (kiểm định chính xác Fisher). Khi dung lượng quan sát rất lớn thì kiểm định Likelihood Ratio Chi – Square = χ^2 .

¹ P = Asymp. Sig. (2-sided) – Mức ý nghĩa phù hợp (hai chiều)

Kiểm định Yates' Continuity Correction được dùng thay thế cho kiểm định χ^2 trong trường hợp dung lượng quan sát nhỏ; nó chỉ thích hợp với Bảng 2x2. Đối với Bảng 2x2, kiểm định Yates' Continuity Correction cho kết quả chính xác hơn. Kiểm định chính xác Fisher (Fisher's Exact Test) được đưa vào Bảng 2x2 nhằm xác định chính xác tính độc lập của hai sự kiện khi dung lượng quan sát nhỏ. Bởi vì, nếu dung lượng quan sát nhỏ (tần số trong mỗi ô < 5) thì kiểm định χ^2 cho kết quả kém chính xác. Thống kê Linear – by – Linear Association Chi – Square là một hàm của hệ số tương quan Pearson (Pearson Correlation Coefficient). Nó chỉ thích hợp với những biến định lượng. Vì thế, ở ví dụ này bạn có thể bỏ qua thống kê này.

6.1.2. Kiểm định ngang bằng về tỉ lệ

Giả sử bạn muốn biết tỉ lệ bị bệnh Y của nhóm cây thuộc cấp sinh trưởng trung bình đến tốt (cấp III – I theo phân cấp Kraft) có giống với những cây thuộc nhóm sinh trưởng kém (cấp IV và V) hay không? Từ những dữ liệu ghi thành Bảng 2x2, bạn có thể dùng tiêu chuẩn χ^2 để kiểm định ngang bằng về tỉ lệ (Equality of Proportions). Giả thuyết H_0 cho rằng nhóm cây thuộc cấp sinh trưởng trung bình đến tốt có tỉ lệ bệnh Y giống với những cây thuộc nhóm sinh trưởng kém. Đối thuyết H_0^- : Tỉ lệ bị bệnh Y của nhóm cây thuộc cấp sinh trưởng trung bình đến tốt khác với những cây thuộc nhóm sinh trưởng kém. Khi sử dụng kiểm định χ^2 về ngang bằng tỉ lệ, bạn hãy thực hiện như sau:

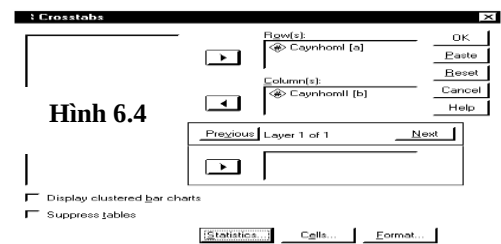
+ **Thu thập dữ liệu.** Trước hết, bạn cần phân chia cây rừng thành hai nhóm: (1) những cây thuộc cấp sinh trưởng tốt (cây cấp III - I), (2) những cây thuộc cấp sinh trưởng kém (cây cấp IV - V). Tiếp theo, bạn xác định bệnh Y trên mỗi cá thể thuộc hai cấp trên. Cây có bệnh Y được mã hóa là 1, còn cây không có bệnh Y được mã hóa là 0. Lưu ý rằng số lượng cây (N) cần điều tra phải đủ lớn để tần số của các ô thuộc Bảng 2x2 phải lớn hơn 5.

+ **Cách nhập dữ liệu.** Số liệu có thể được nhập trực tiếp vào Bảng tính SPSS hoặc nhập vào Bảng tính Excel rồi Copy sang SPSS (Bảng 6.5). Cột 1 ghi thứ tự cây, cột 2 nhập thông tin về cây nhóm 1 (theo mã hóa); cột 3 nhập thông tin về cây nhóm 2 (theo mã hóa).

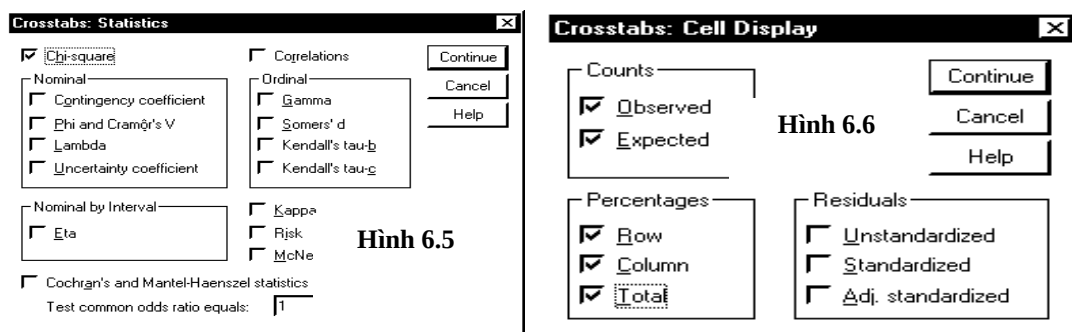
Bảng 6.5. Nhập dữ liệu để kiểm định ngang bằng tỉ lệ (tóm tắt)

TT	Cây nhóm I	Cây nhóm II
1	1	1
2	1	0
3	0	1
4	0	0
5...	1	1

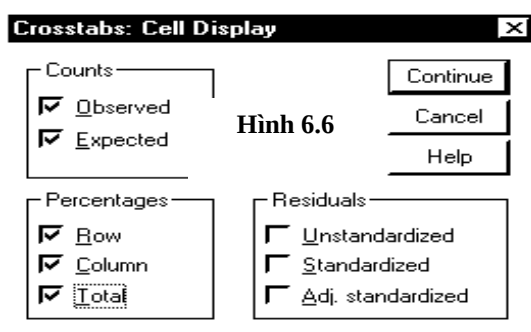
+ **Gọi chương trình tính toán.** Sau khi nhập xong dữ liệu, để tạo được kết quả, từ thanh Menu chính bạn hãy vào đường lệnh sau: Mở Analyze $\Rightarrow$ Descriptive Statistics $\Rightarrow$ Crosstabs. Khi Bảng Crosstabs xuất hiện (Hình 6.4), bạn hãy chọn biến “cây nhóm 1” và chuyển nó vào ô Row, còn biến “cây nhóm 2” được chuyển vào ô Column. Kế đến chọn Statistics và đánh dấu chọn ô Chi – Square (Hình 6.5). Tiếp theo chọn Cells (Hình 6.6) và đánh dấu chọn các mục Counts (tần số) và Percentages (phần trăm). Sau cùng chọn OK để chương trình tính toán và hiển thị kết quả.



Giải thích và nhận xét kết quả. Kết quả kiểm định ngang bằng về tỉ lệ đưa ra những Bảng có nội dung cũng giống với những Bảng của kiểm định về tính độc lập (Bảng 6.6, 6.7, 6.8). Từ kết quả của các Bảng này, bạn hãy căn cứ vào giá trị của thống kê Pearson Chi-Square (trong Bảng là χ^2 - value), độ tự do (df) và mức ý nghĩa thống kê P (trong Bảng là Asymp. Sig. (2-sided)) để kiểm định giả thuyết về ngang bằng tỉ lệ. Nếu mức ý nghĩa $P < 0.05$ (mặc định) thì giả thuyết H_0 bị bác bỏ. Ngược lại, khi mức ý nghĩa $P > 0.05$ (mặc định) thì giả thuyết H_0 được chấp nhận. Bậc tự do của phân bố $\chi^2 = (\text{số hàng} - 1) \times (\text{số cột} - 1)$. Để khẳng định độ chính xác của kiểm định χ^2 , bạn có thể quan sát thêm các kiểm định Yates' Continuity Correction (hệ số hiệu chỉnh Yates), Likelihood Ratio (tỉ lệ tương đồng) và Fisher's Exact Test (kiểm định chính xác Fisher). Nếu các kiểm định này cùng đưa ra kết quả với mức ý nghĩa P giống như kiểm định χ^2 thì bạn yên tâm hơn về việc chấp nhận hay phủ nhận giả thuyết H_0 .



Hình 6.5



Hình 6.6

Bảng 6.6. Tóm tắt xử lý các trường hợp (Case Processing Summary)

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Caynhom 1 × Caynhom 2	120	100.0%	0	.0%	120	100.0%

Bảng 6.7. Tần số xuất hiện bệnh của hai nhóm cây
(Cay nhóm I × Cay nhóm II Crosstabulation)

			Cây nhóm 2		Total
			0	1	
Cây nhóm 1	0	Count	10	30	40
		Expected Count	20.0	20.0	40.0
		% within CaynhomI	25.0%	75.0%	100.0%
		% within CaynhomII	16.7%	50.0%	33.3%
		% of Total	8.3%	25.0%	33.3%
	1	Count	50	30	80
		Expected Count	40.0	40.0	80.0
		% within CaynhomI	62.5%	37.5%	100.0%
		% within CaynhomII	83.3%	50.0%	66.7%
		% of Total	41.7%	25.0%	66.7%
Total		Count	60	60	120
		Expected Count	60.0	60.0	120.0
		% within CaynhomI	50.0%	50.0%	100.0%
		% within CaynhomII	100.0%	100.0%	100.0%
		% of Total	50.0%	50.0%	100.0%

Bảng 6.8. Kiểm định Chi-Square (Chi-Square Tests)

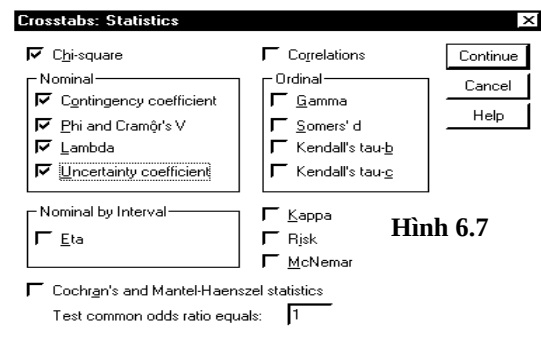
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	15.000	1	.000		
Continuity Correction	13.538	1	.000		
Likelihood Ratio	15.518	1	.000		
Fisher's Exact Test				.000	.000
N of Valid Cases	120				

6.2. QUAN HỆ GIỮA HAI BIẾN ĐỊNH TÍNH KHÔNG CÓ THỨ BẬC HƠN KÉM

1. Điều kiện áp dụng. (1) Kiểm định quan hệ giữa 2 biến định tính không có thứ bậc hơn kém bằng Bảng 2×2 hoặc Bảng $R \times C$ ô. (2) Vai trò của biến phụ thuộc và biến độc lập là không rõ ràng. Điều đó có nghĩa là có thể xem biến nào là biến phụ thuộc hay biến độc lập đều được. Lưu ý rằng trong sinh thái học, khi xác định mối quan hệ giữa các loài sinh vật với nhau, người ta gọi phép kiểm định này là kiểm định sự kết nhóm (nhóm sinh thái) giữa các loài.

2. Nội dung kiểm định. Khi kiểm định mối quan hệ giữa hai biến định tính không có thứ bậc hơn kém, bạn cần phải giải quyết hai nội dung. Một là, kiểm định mối quan hệ giữa hai đại lượng (hai loài cây) bằng thống kê Pearson Chi – Square. Kết quả của bước 1 chỉ cho biết hai đại lượng có quan hệ với nhau hay không. Hai là, nếu hai biến có quan hệ với nhau thì bước tiếp theo cần xác định cường độ của mối quan hệ.

3. Thực hành. Để thực hành thủ tục này, chúng ta sử dụng lại ví dụ ở mục 6.1.1. Nội dung thu thập dữ liệu, cách vào dữ liệu và gọi chương trình tính toán cũng tương tự như kiểm định tính độc lập giữa hai đại lượng (hai loài cây). Nhưng để kiểm định mối quan hệ giữa hai đại lượng, sau khi chọn Statistics, bạn thực hiện như sau :



Hình 6.7

+ Chọn Chi – Square. Thủ tục này nhằm kiểm định tính độc lập giữa hai biến.

+ Nếu hai biến số có quan hệ với nhau thì bước tiếp theo là tìm mức độ chặt chẽ trong quan hệ giữa hai biến số. Để đạt mục đích này, bạn chọn các mục Contingency coefficient, Phi and Cramer's V, Lambda, Uncertainty coefficient trong ô Nominal (Hình 6.7). Sau khi chọn các thống kê cần tính, bạn chọn OK để chương trình tính toán và hiển thị kết quả.

Kết quả kiểm định mối quan hệ giữa hai biến đưa ra 5 Bảng sau đây: (1) Thống kê tóm tắt xử lý các trường hợp, (2) Bảng chéo, (3) kiểm định Chi – square, (4) các đo đặc về xu hướng mỗi quan hệ (Directional Measures), (5) các đo đặc về tính đối xứng (Symmetric Measures). Nội dung của Bảng thống kê tóm tắt xử lý các trường hợp, Bảng chéo và kiểm định Chi – square cũng giống với kiểm định mối quan hệ giữa hai đại lượng. Dưới đây chúng ta chỉ xem xét kỹ nội dung của các đo đặc về xu hướng mỗi quan hệ (Directional Measures) (Bảng 6.9) và các đo đặc về tính đối xứng (Symmetric Measures) giữa hai biến số (Bảng 6.10).

4. Giải thích và nhận xét kết quả. Khi xét quan hệ giữa các biến số định danh không có thứ bậc, SPSS sử dụng hai nhóm đo đặc sau đây :

+ **Nhóm 1.** Những đo đặc về tính đối xứng trong cách sắp xếp của hai đại lượng của Bảng 2×2 (hay $R \times C$ ô). Những đo đặc này cũng dựa trên thống kê Pearson Chi – square, bao gồm thống kê Phi, Contingency coefficient và thống kê Cramer's V.

+ **Nhóm 2.** Những hệ số đo xu hướng của mức giảm tỉ lệ sai số khi chọn biến phụ thuộc và biến dự đoán khác nhau. Nhóm này bao gồm các hệ số Lambda, Goodman & Kruskal's tau, Uncertainty coefficient (hệ số đo mức không chắc chắn của mối quan hệ).

Những thống kê đo đặc mức giảm tỉ lệ trong sai số

(Áp dụng cho các biến phân nhóm không có thứ bậc)

+ **Những đo đặc về xu hướng** (Directional Measures). Những đo đặc này xác định mức độ của mối quan hệ giữa biến phụ thuộc với biến độc lập (hoặc giữa biến dòng với biến cột của Bảng R×C). Để đo xu hướng quan hệ giữa hai biến, người ta sử dụng các hệ số Lambda, Goodman & Kruskal's tau và Uncertainty coefficient. Chúng lấy các giá trị từ 0 – 1. Giá trị 0 cho biết biến độc lập không có quan hệ với biến phụ thuộc (hay biến phụ thuộc không thể được dự đoán bằng biến độc lập). Ngược lại, khi chúng lấy giá trị gần 1 hoặc bằng 1 thì biến độc lập có quan hệ chặt chẽ với biến phụ thuộc (hay biến phụ thuộc có thể được dự đoán chính xác bằng biến độc lập). **Bảng 6.9** ghi lại nội dung của những đo đặc về xu hướng.

Bảng 6.9. Những đo đặc về xu hướng quan hệ (Directional Measures)

Nominal by Nominal		Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Lambda	Symmetric	.200	.116	1.598	.110
	LoaiA Dependent	.000	.194	.000	1.000
	LoaiB Dependent	.333	.086	3.303	.001
Uncertainty Coefficient	Symmetric	.097	.047	2.059	.000 ^d
	LoaiA Dependent	.102	.049	2.059	.000 ^d

Lưu ý rằng, trong trường hợp hai biến định tính không có thứ bậc hơn kém, bạn có thể chọn bất kỳ biến nào trong hai biến (A và B) là biến độc lập hay biến phụ thuộc đều được. Điều đó là tùy theo mục tiêu nghiên cứu của bạn. Khi bạn chọn biến A là biến độc lập để dự đoán (xác định) biến B thì tỉ lệ (phần trăm) giảm sai số có thể khác với trường hợp bạn dùng biến B là biến độc lập để dự đoán biến A. Mức giảm sai số được trình bày ở cột Value của Bảng Directional Measures (**Bảng 6.9**). Mức giảm sai số được tính theo % bằng cách lấy cột Value × 100.

Thống kê Lambda đo mức độ quan hệ giữa hai biến. Nó cho biết yếu tố dòng (hoặc cột) hữu ích như thế nào trong việc dự đoán biến cột (hoặc dòng). Nói khác đi, Lambda đo mức giảm sai số khi lấy biến dòng là biến độc lập để dự đoán biến phụ thuộc theo cột và ngược lại. Từ **Bảng 6.9** cho thấy, nếu sử dụng loài A (biến độc lập) để dự đoán loài B (biến phụ thuộc) thì tỉ lệ giảm sai số là 33,3% (Lambda = 0,333). Ngược lại, khi lấy loài B để dự đoán loài A thì tỉ lệ giảm sai số là 0% (Lambda = 0.000). Lưu ý rằng khi hai biến (ví dụ hai loài cây) có quan hệ càng chặt chẽ với nhau thì tỉ lệ giảm sai số trong dự đoán càng lớn. Bạn hãy quan sát P – Value (Approx. Sig) để biết mức quan hệ giữa hai biến hay mức ý nghĩa của sự giảm trong sai số khi dự đoán. Nếu $P < 0.05$ (mặc định) thì biến cột (hay biến hàng) có quan hệ với biến hàng (hay cột). Ở ví dụ trên đây, thống kê P - value = 0,001; điều đó cho biết mối quan hệ giữa hai loài có ý nghĩa thống kê với mức tin cậy rất cao (99,9%). Tùy theo mục tiêu nghiên cứu, bạn cần chọn lựa biến dự đoán sao cho mức giảm tỉ lệ trong sai số dự đoán là lớn nhất.

Khi áp dụng thống kê Lambda trong nghiên cứu sự kết nhóm giữa các loài cây, bạn có thể chọn loài A để dự đoán sự có mặt của loài B hoặc chọn loài B để dự đoán sự có mặt của loài A. Khi chọn biến dự đoán, bạn phải căn cứ vào mức giảm tỉ lệ sai số dự đoán. Một biến được chọn là biến dự đoán nếu nó cho tỉ lệ giảm sai số trong dự đoán là cao hơn.

+ **Những đo đặc đối xứng hay cân đối** (Symmetric Measures). Những đo đặc này cũng đo đặc mức độ quan hệ giữa các biến hoặc đo đặc mức giảm tỉ lệ trong sai số (PRE =

Proportional Reduction in Error). Chúng bao gồm các hệ số Phi, hệ số ngẫu nhiên (Contingency Coefficient) và thống kê Cramer's V (Bảng 6.10).

Bảng 6.10. Những số đo về tính đối xứng (Symmetric Measures)

		Value	Approx. Sig.
Nominal by Nominal (hai biến không có thứ bậc)	Phi	-.354	.000
	Cramer's V	.354	.000
	Contingency Coefficient	.333	.000
N of Valid Cases		120	

Hệ số Phi, hệ số ngẫu nhiên và thống kê Cramer's V được đo bằng các thang đo khác nhau; trong đó :

- (a) Thống kê Cramer's V lấy giá trị từ 0 – 1. Giá trị zero cho biết hai biến không có quan hệ với nhau. Ngược lại, giá trị 1 cho biết hai biến có quan hệ với nhau rất chặt chẽ. Thống kê Cramer's V được dùng phổ biến nhất.
- (b) Thống kê Phi có thể lấy giá trị lớn hơn 1. Khi giá trị của thống kê Cramer's V bằng giá trị của thống kê Phi thì chúng cũng bằng hệ số tương quan Pearson (Pearson correlation coefficient).
- (c) Hệ số Contingency lấy giá trị từ 0 đến 1, nhưng không bao giờ đạt đến 1.
- (d) Hệ số Phi, hệ số ngẫu nhiên (Contingency) và thống kê Cramer's V lấy giá trị 0 khi và chỉ khi kiểm định Pearson $\chi^2 = 0$.

Cột Approx. Sig. (mức ý nghĩa gần đúng) được đưa ra nhằm kiểm định giả thuyết H_0 cho rằng từng đo đặc = 0. Đối thuyết H_0 : từng đo đặc $\neq 0$. Nếu $P > 0.05$ thì giả thuyết H_0 là đúng, nghĩa là các biến không có quan hệ với nhau. Ngược lại, $P < 0.05$ thì giả thuyết H_0 là sai, nghĩa là các biến có quan hệ với nhau ở mức tin cậy 95%.

Lưu ý rằng, thống kê χ^2 chỉ trả lời được câu hỏi là hai biến số có mối quan hệ với nhau hay không; nó không cho biết cường độ (mức độ) của mối quan hệ. Khi hai biến số không có thứ bậc có quan hệ với nhau thì bạn cần phải đo mức độ (cường độ) của mối quan hệ bằng các kiểm định Cramer's Test, hệ số Phi và hệ số Contingency.

5. Kiểm định sự kết nhóm riêng phần giữa các loài cây

Các hệ số kết nhóm (hoặc hệ số tương quan) tự chúng chưa cho biết đầy đủ nguyên nhân ảnh hưởng đến sự phân bố của các loài theo diện tích. Để tìm hiểu rõ nguyên nhân ảnh hưởng đến sự phân bố của hai loài, chúng ta cần phải thu thập thêm nhiều thông tin khác nhau hoặc phải so sánh nhiều hệ số kết nhóm khác nhau bằng cách phân chia giới hạn cấp khác nhau. Tuy nhiên, điều này có thể được giải quyết bằng phương pháp tính hệ số tương quan riêng phần. Bản chất của phương pháp này là tính hệ số kết nhóm khi các giá trị của một hoặc một số nhân tố được cố định. Điều đó cho phép loại trừ ảnh hưởng của những nhân tố này đến quan hệ của các đại lượng nghiên cứu khác.

Theo Cole (1957), hệ số kết nhóm riêng phần được xác định như sau:

Bước 1. Giả thiết trong khu vực nghiên cứu, bạn cần tính sự kết nhóm giữa 3 loài cây. Để đạt mục đích đặt ra, trước hết bạn cần thống kê sự có mặt hay vắng mặt của 3 loài cây quan tâm trên các ô nghiên cứu. Để tính hệ số kết nhóm riêng phần giữa 3 loài, số lượng ô nghiên cứu cần phải có ít nhất là 250 ô. Điều kiện này được đặt ra là vì các ô trong Bảng chéo đòi hỏi phải có tần số lớn hơn 5 hoặc số ô của Bảng C×R có tần số nhỏ hơn 5 không lớn hơn 25%.

Bước 2. Để tính hệ số kết nhóm riêng phần, bạn hãy chia tập hợp ô mẫu (N) mà trên cơ sở đã tính hệ số kết nhóm đầy đủ cho ba loài (hệ số kết nhóm có sự tham gia của tất cả ba loài cây trong các ô mẫu) thành hai phần: (1) những ô bắt gặp loài thứ 3 (N_1); (2) những ô vắng mặt loài thứ 3 (N_2). Kế đến, bạn tính hệ số kết nhóm giữa hai loài 1 và 2 trên những ô bắt gặp loài 3; kết quả nhận được hệ số kết nhóm đầy đủ hay đa phần C_1 . Tương tự như thế, bạn tính hệ số kết nhóm giữa hai loài 1 và 2 trên những ô không bắt gặp loài 3; kết quả nhận được hệ số kết nhóm riêng phần C_2 .

Bước 3. So sánh hệ số kết nhóm C_1 với C_2 sẽ cho phép đánh giá loài mà theo đó tiến hành phân chia các nhóm ô mẫu đóng vai trò như thế nào trong việc xác định hệ số kết nhóm đa phần. Tiêu chuẩn so sánh được tính theo công thức:

$$T = \frac{C_2 - C_1}{\sqrt{\delta_{C1} + \delta_{C2}}} \quad (6.1)$$

trong đó :

+ δ_{C1} là sai số của hệ số kết nhóm trên những ô có loài thứ 3 xuất hiện

$$\delta_{C1} = \sqrt{\frac{(a_1+b_1)*(b_1+d_1)}{(N_1*(a_1+c_1)*(c_1+d_1))}} \quad (6.2)$$

+ δ_{C2} là sai số của hệ số kết nhóm trên những ô không có loài thứ 3 xuất hiện

$$\delta_{C2} = \sqrt{\frac{(a_2+b_2)*(b_2+d_2)}{(N_2*(a_2+c_2)*(c_2+d_2))}} \quad (6.3)$$

Trong công thức 6.2 và 6.3, N_1 là tổng số ô bắt gặp loài thứ 3; N_2 là tổng số ô vắng mặt loài thứ 3; a_1, b_1, c_1 và d_1 - tần số ô trong Bảng chéo dùng để tính quan hệ giữa loài 1 và 2 trên những ô bắt gặp loài 3; a_2, b_2, c_2 và d_2 - tần số ô trong Bảng chéo dùng để tính quan hệ giữa loài 1 và 2 trên những ô không bắt gặp loài 3. So sánh giá trị T với $T_{(0.05 \text{ hoặc } 0.01)}$. Nếu $T < T_{0.05} = 2,0$ thì sự khác biệt giữa hệ số kết nhóm riêng phần với hệ số kết nhóm đa phần là không có ý nghĩa. Điều này có nghĩa là sự xuất hiện của loài thứ 3 trong các ô nghiên cứu không ảnh hưởng đáng kể đến sự phân bố của loài 1 và 2. Ngược lại, nếu $T > T_{0.05}$ thì sự xuất hiện của loài thứ 3 trong các ô nghiên cứu có ảnh hưởng đáng kể đến sự phân bố của loài 1 và 2.

6. Kiểm định sự kết nhóm đa biến

Một quần xã thực vật có rất nhiều loài cây, giữa chúng tồn tại những mối quan hệ rất phức tạp. Vì thế, khi phân tích quan hệ giữa các loài, chúng ta cần phải phân tích không chỉ quan hệ giữa loài cây này với loài cây khác mà còn giữa nó với hai, ba...loài cây khác nhau. Quan hệ của một loài cây với nhiều loài cây khác được gọi là tương quan (kết nhóm) đa biến. Cần nhận thấy rằng hệ số tương quan đa biến của loài này với nhiều loài khác bao giờ cũng cao hơn hệ số tương quan giữa hai loài. Bởi vì độ phong phú của mỗi loài luôn bị kiểm soát bởi rất nhiều loài khác trong quần xã. Chúng ta có thể vận dụng hệ số tương quan đa biến để xác định mối liên hệ giữa các loài cây khi chúng chỉ được đo bằng hai dấu hiệu định tính là có và không có.

Ví dụ: Giả thiết chúng ta cần nghiên cứu quan hệ của 3 loài cây 1, 2 và 3. Trước hết, chúng ta tính các quan hệ giữa loài 1 và 2, 1 và 3, 2 và 3. Nếu tính quan hệ giữa loài 1 với hai loài 2 và 3, thì chúng ta cũng lập Bảng 2×2 ; trong đó dấu hiệu 1 là loài 1, còn dấu hiệu 2 là loài 2 và 3. Những ô mẫu cùng xuất hiện cả hai loài 2 và 3 được gọi là ô có (kí hiệu 1), còn những ô vắng mặt cả hai loài hoặc vắng mặt một trong hai loài 2 và 3 được gọi là ô không có

(kí hiệu 0). Sau đó, bằng cách thông thường tính hệ số kết nhóm giữa loài 1 và 2 loài kia. Như vậy, hệ số kết nhóm đa biến chỉ ra mức độ liên hệ của các nhóm sinh thái¹.

Cần nhận thấy rằng, rừng mưa nhiệt đới có rất nhiều loài cây cùng chung sống trên một không gian nhất định. Vì thế, để phân tích rõ sự kết nhóm giữa nhiều loài cây (4 - 5 loài trở lên) đòi hỏi phải thu thập hàng ngàn ô mẫu. Chính khối lượng điều tra lớn như thế đã gây ra những khó khăn lớn cho việc phân tích cấu trúc quần xã.

Thực hành. Ở phụ biểu 17 dẫn ví dụ về kết quả điều tra sự kết nhóm của bốn loài cây (Trắc, Hà nu, Chò xót và Lành ngạnh) trên 250 ô mẫu với kích thước 200 m²/ô. Vấn đề đặt ra: (1) tìm hệ số kết nhóm đầy đủ giữa loài Hà nu (loài 1) và Chò xót (loài 2) khi có mặt cả loài Lành ngạnh (loài thứ 3); (2) hãy cho biết Lành ngạnh (loài thứ 3) đóng vai trò như thế nào trong quan hệ giữa Hà nu (loài 1) với Chò xót (loài 2). Để giải quyết hai nội dung đặt ra trên đây, bạn có thể thực hiện các thủ tục như dưới đây.

Bước 1. Nhập số liệu. Số liệu bắt gặp hay không bắt gặp bốn loài cây được ghi thành mỗi cột hay mỗi hàng một loài. Cột thứ 5, 6...ghi lại số ô cùng bắt gặp hay không bắt gặp từng cặp hai hoặc ba loài nào đó.

Bước 2. Gọi thủ tục xử lý sự kết nhóm giữa từng cặp loài. Trước hết, bạn cần tính sự kết nhóm giữa từng cặp loài. Để đạt được mục đích đặt ra, từ Menu chính bạn mở Analyze $\Rightarrow$ Descriptive Statistics $\Rightarrow$ Crosstabs. Khi cửa sổ Crosstabs xuất hiện, bạn đưa số liệu của loài 1 vào ô hàng (Row variable), còn loài 2 đưa vào ô cột (Column variable). Kế đến chọn Statistics và đánh dấu chọn Chi - square.

Để biết hai loài Hà nu và Chò xót có kết nhóm (quan hệ) với nhau hay không, bạn hãy quan sát thống kê Chi - square test. Giả thuyết H_0 : Hà nu và Chò xót không có sự kết nhóm với nhau. Đối thuyết H_0 ': Hà nu và Chò xót có sự kết nhóm với nhau. Khi $P < 0,05$ hoặc 0,01 thì giả thuyết H_0 bị bác bỏ. Ở ví dụ này, vì $P = 0,001 < 0,05$ nên hai loài có sự kết nhóm với nhau.

Để kiểm định sự kết nhóm giữa các cặp loài khác, bạn cũng thực hiện các thủ tục tương tự như đã làm với Hà nu và Chò xót. Lưu ý rằng, kết quả ở bước 2 chỉ cho biết hệ số kết nhóm giữa từng cặp loài.

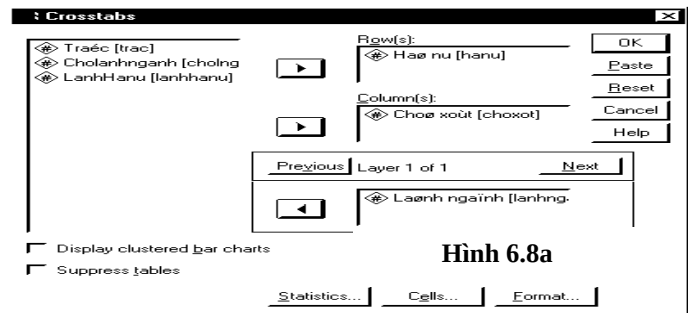
Nếu hai biến số có quan hệ với nhau thì bước tiếp theo là tìm mức độ chặt chẽ trong quan hệ giữa hai biến số. Để đạt mục đích này, bạn chọn các mục Contingency coefficient, Phi and Cramer's V, Lambda, Uncertainty coefficient trong ô Nominal. Sau khi chọn các thống kê cần tính, bạn chọn OK để chương trình tính toán và hiển thị kết quả. Kết quả ở bước này cho biết hệ số kết nhóm giữa từng cặp loài.

Bước 3. Gọi thủ tục xử lý hệ số kết nhóm riêng phần. Để đạt mục đích đặt ra, trước hết bạn đưa loài 1 vào ô Row variable, loài 2 vào ô Column, còn loài 3 (Lành ngạnh) vào ô Layer 1 of 1 (**Hình 6.8a**). Kế đến chọn Statistics và đánh dấu chọn Chi - square. Nếu hai biến số có quan hệ với nhau thì bước tiếp theo là tìm mức độ chặt chẽ trong quan hệ giữa hai biến số. Để đạt mục đích này, bạn chọn các mục Contingency coefficient, Phi and Cramer's V, Lambda, Uncertainty coefficient trong ô Nominal.

Ở **Bảng 6.11 - 6.13 dẫn** kết quả phân tích sự kết nhóm riêng phần giữa Hà nu với Chò xót trên những ô mẫu bắt gặp và không bắt gặp Lành ngạnh. Phân tích kết quả dẫn ra ở **Bảng 6.12 và 6.13** cho thấy, Hà nu và Chò xót chỉ kết nhóm với nhau khi có mặt Lành ngạnh ($\chi^2_{(1)} = 13,83$; $P = 0,000$). Khi không có mặt Lành ngạnh thì Hà nu và Chò xót không kết nhóm với nhau ($\chi^2_{(1)} = 1,03$; $P = 0,309$). Điều đó chứng tỏ rằng hai loài Hà nu và Chò xót không kết nhóm với nhau. Để khẳng định chính xác nhận định này, chúng ta cần thu thập thêm những thông tin trên những ô mẫu khác. Lưu ý rằng, khi phân tích sự kết nhóm riêng phần của những loài khác, bạn cũng thực hiện lặp lại các thủ tục như trên đây.

¹ Ecological groups

Bước 4. Kiểm định sự khác biệt giữa các hệ số kết nhóm riêng phần. Khi hai hệ số kết nhóm riêng phần đều tồn tại, chúng ta cần phải kiểm định sự khác biệt giữa hệ số kết nhóm của loài 1 và 2 trên những ô không có mặt loài 3 với hệ số kết nhóm của loài 1 và 2 trên những ô có mặt loài 3 theo **công thức 6.1**. Nếu $T_{tt} < T_{05} = 2,0$ thì sự khác biệt giữa các hệ số kết nhóm riêng phần là không có ý nghĩa. Điều này có nghĩa là sự xuất hiện của loài thứ 3 trong các ô nghiên cứu không ảnh hưởng đáng kể đến sự phân bố của loài 1 và 2. Ngược lại nếu $T_{tt} > T_{0,05}$ thì sự xuất hiện của loài thứ 3 trong các ô nghiên cứu có ảnh hưởng đáng kể đến sự phân bố của loài 1 và 2.



Hình 6.8a

Bảng 6.11. Sự kết nhóm riêng phần giữa Hà nu với Chò xốt khi có mặt (1) hoặc vắng mặt (0) Lành ngạnh

Lành ngạnh			Chò xốt		Tổng
			0	1	
0	Hà nu	0	10 12.4 ^(x)	23 20.6	33 33
		1	32 29.6	47 49.4	79 79
	Tổng		42 42	70 70	112 112
1	Hà nu	0	17 27.8	48 37.2	65 65
		1	42 31.2	31 41.8	73 73
	Tổng		59 59	79 79	138 138

(x) Mẫu số là tần số lý thuyết

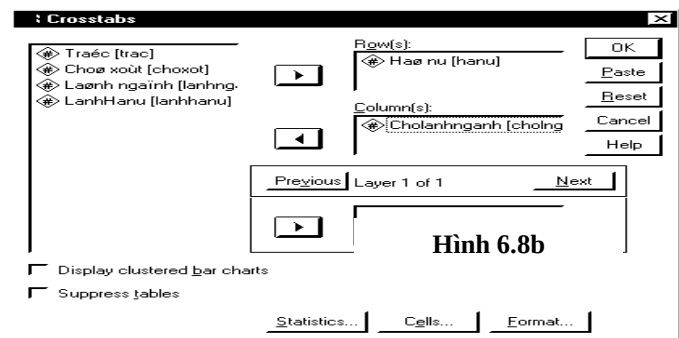
Bảng 6.12. Kiểm định tính độc lập giữa Hà nu với Chò xốt khi có mặt (1) hoặc vắng mặt (0) Lành ngạnh

Lành ngạnh	Thống kê	Giá trị	df	P	P _(Chính xác)
0	χ^2	1.034	1	.309	
	Hiệu chỉnh tính liên tục	.644	1	.422	
	Tỉ lệ Likelihood	1.053	1	.305	
	Kiểm định Fisher			.393	
	N	112			
1	χ^2	13.834	1	.000	0,000
	Hiệu chỉnh tính liên tục	12.582	1	.000	
	Tỉ lệ Likelihood	14.158	1	.000	
	Kiểm định Fisher				
	N	138			

Bảng 6.13. Mối quan hệ giữa Hà nu với Chò xốt khi có mặt (1) hoặc vắng mặt (0) Lành ngành

Lành ngành	Thống kê	Giá trị	P
0	Hệ số Phi	-.096	.309
	Hệ số Cramer's V	.096	.309
	Hệ số ngẫu nhiên	.096	.309
	Số ô mẫu N	112	
1	Hệ số Phi	-.317	.000
	Hệ số Cramer's V	.317	.000
	Hệ số ngẫu nhiên	.302	.000
	Số ô mẫu N	138	

Bước 5. Phân tích sự kết nhóm đa phần. Để phân tích sự kết nhóm đa phần giữa Hà nu với tổ hợp hai loài Chò xốt - Lành ngành, bạn đưa loài Hà nu vào ô Row variable, tổ hợp hai loài Chò xốt - Lành ngành vào ô Column (Hình 6.8b). Kế đến chọn Statistics và đánh dấu chọn Chi-square. Nếu hai biến số có quan hệ với nhau thì bước tiếp theo là tìm mức độ chặt chẽ trong quan hệ giữa hai biến số. Để đạt mục đích này, bạn chọn các mục Contingency coefficient, Phi and Cramer's V, Lambda, Uncertainty coefficient trong ô Nominal. Kết quả phân tích sự kết nhóm đa phần giữa Hà nu với tổ hợp Chò xốt - Lành ngành dẫn ra ở Bảng 6.14 và 6.15. Từ đó nhận thấy phân bố của Hà nu có mối liên hệ thực sự với tổ hợp hai loài Chò xốt và Lành ngành ($\chi^2_{(1)} = 11,05$; $P = 0,000$). Từ số liệu của Bảng 6.15 cho thấy, nếu căn cứ vào sự có mặt của Hà nu trong ô mẫu để dự đoán sự có mặt của cả hai loài Chò xốt và Lành ngành thì mức giảm sai số là zero phần trăm. Ngược lại, nếu dựa trên sự cùng có mặt của Chò xốt và Lành ngành trong ô mẫu để dự đoán sự có mặt của Hà nu thì mức giảm sai số là 17,3%. Mức độ quan hệ giữa Hà nu với tổ hợp hai loài Chò xốt - Lành ngành là chặt chẽ (hệ số Cramer's V = 0,300; $P = 0,000$).



Hình 6.8b

Bảng 6.14. Kiểm định tính độc lập giữa Hà nu với Chò xốt - Lành ngành

	Value	df	P	P chính xác
Pearson Chi-Square	11.052	1	.001	.001
Continuity Correction	10.192	1	.001	
Likelihood Ratio	11.311	1	.001	
Fisher's Exact Test				
N of Valid Cases	250			

Bảng 6.15. Quan hệ giữa Hà nu với tổ hợp loài Chò xốt - Lành ngành

6.15a. Xu hướng quan hệ

Thông kê	Biến phụ thuộc	Giá trị	Sai số chuẩn	T	P
Lambda	Hệ số đối xứng	.096	.047	1.927	.054
	Hà nu	.173	.082	1.927	.054
	Chò xốt + Lành ngành	.000	.000	.	.

6.15b. Mức độ quan hệ

Thông kê	Giá trị	P
Cramer's V	.300	.000
Hệ số ngẫu nhiên	.287	.000
Số ô mẫu (N)	250	

Hai vấn đề cần lưu ý

1. Bằng cách phân tích tương quan riêng phần hoặc sự kết nhóm riêng phần, chúng ta xác định được tương quan cặp giữa các loài cây và tìm ra loài cây này có ảnh hưởng đến sự phân bố của loài cây kia như thế nào. Nếu sau khi tính toán nhận được các hệ số tương quan riêng phần không khác nhau thực sự, thì ta có cơ sở kết luận sự kết nhóm này không phải do quan hệ qua lại giữa các loài. Ngược lại, sự cùng có mặt của các loài cây trên những ô nghiên cứu là do sự tương đồng về biên độ sinh thái của chúng.

2. Khi nghiên cứu tương quan đa biến giữa các loài, hoặc giữa các loài cây và các yếu tố môi trường, chúng ta cần phải định nghĩa rõ ràng về tổng thể nghiên cứu hay đối tượng thu thập dữ liệu. Việc mở rộng phạm vi nghiên cứu có thể bao trùm toàn bộ biên độ sinh thái của một số loài cây, hoặc tổng thể sẽ chứa những vùng nằm ngoài biên độ sinh thái của các loài. Điều đó có thể dẫn đến sự thay đổi của hệ số kết nhóm giữa các loài cây hoặc hệ số tương quan giữa các loài và các yếu tố môi trường.

6.3. MỐI QUAN HỆ GIỮA HAI BIẾN ĐỊNH TÍNH CÓ THỨ BẬC

Thủ tục lập Bảng chéo 2 chiều còn cho phép nghiên cứu quan hệ giữa hai biến định tính có thứ bậc hơn kém bằng cách xây dựng một Bảng ngẫu nhiên gồm $R \times C$ ô và thực hiện các kiểm định về mối quan hệ giữa biến dòng với biến cột.

1. Điều kiện áp dụng. (1) Cả hai biến lập Bảng là những biến định tính có thứ bậc hơn kém; (2) Các biến có thể lập thành nhóm (cấp) trong Bảng chéo gồm $R \times C$ ô.

2. Nội dung tính toán. Đo đạc về mối quan hệ giữa hai biến lập nhóm có thứ bậc hơn kém bao gồm hai nội dung.

Nội dung 1. Xác định tương quan giữa các biến. Để giải quyết vấn đề đặt ra, bạn có thể sử dụng hệ số tương quan hạng của Spearman, hệ số tương quan Pearson và kiểm định tuyến tính – tuyến tính χ^2 (Linear -by - Linear association Chi - Square). Nhưng cần lưu ý rằng, hệ số tương quan Pearson đòi hỏi hai đại lượng quan sát phải được rút ra từ một phân phối chuẩn nhị biến (Bivariate normal distribution), còn hệ số tương quan Spearman không đòi hỏi hai đại lượng quan sát có phân phối chuẩn. Vì thế, khi đo đạc mối quan hệ giữa hai biến lập nhóm có thứ bậc thì hệ số tương quan Spearman được sử dụng nhiều hơn.

Nội dung 2. Đo đạc tính phù hợp giữa các biến hay xu hướng quan hệ giữa các biến. Đo đạc này kiểm định tính phù hợp giữa hai biến lập Bảng. Nếu các trị số của cả hai biến cùng cấp này lớn hơn hoặc nhỏ hơn các trị số của hai biến cùng cấp kia thì cặp biến ấy được

gọi là phù hợp hay tương đồng. Khi điều kiện đó không thỏa mãn, người ta gọi các cặp biến đó là không phù hợp hay không tương đồng. Khi các biến có cùng trị số ở một hoặc cả hai cặp biến thì cặp đó được gọi là cặp biến bị ràng buộc. Nói khác đi, khi các hạng của biến Y có xu hướng tăng cùng với các hạng của biến X (Y và X đồng biến) thì các cặp biến ấy được gọi là phù hợp hay tương đồng. Ngược lại, khi các hạng của biến Y tăng lên còn hạng của biến X giảm xuống (Y và X nghịch biến) thì các cặp biến ấy được gọi là các cặp biến không phù hợp hay không tương đồng. Khi số cặp phù hợp bằng với số cặp không phù hợp thì hai biến số ấy không có quan hệ với nhau hay không tương đồng với nhau.

Để đo đặc sự tương đồng giữa các cặp biến số định tính có thứ bậc hơn kém, người sử dụng những thống kê sau đây :

a. Các đo đặc dựa trên các cặp biến tương đồng. Đó là các thống kê Kendall tau b, Kendall tau c, Goodman và Kruskal's Gamma.

Điều kiện áp dụng. Ba thống kê này chỉ áp dụng cho hai biến định tính được phân nhóm có thứ bậc hơn kém và vai trò của biến độc lập và biến phụ thuộc là không rõ ràng (nghĩa là có thể xem biến nào là biến phụ thuộc hay biến độc lập đều được).

+ Thống kê Kendall tau b tính theo công thức :
$$\tau_B = \frac{P - Q}{\sqrt{(P+Q+T_X)(P+Q+T_Y)}}$$

+ Thống kê Kendall tau c tính theo công thức :
$$\tau_C = \frac{2m*(P - Q)}{N^2*(m-1)}$$

Trong hai công thức trên, P là số cặp biến tương đồng với nhau (nghĩa là cùng tăng hoặc cùng giảm); Q là số cặp biến không tương đồng với nhau (nghĩa là những cặp tăng – giảm khác nhau); Tx – những cặp bị ràng buộc bởi yếu tố X; Ty – những cặp bị ràng buộc bởi yếu tố Y; m – số nhỏ nhất trong hai số cột và số hàng.

b. Thống kê Somers' D và Gamma

Điều kiện áp dụng. Hai thống kê này áp dụng cho hai biến lập nhóm có thứ bậc và vai trò của biến độc lập và biến phụ thuộc là rõ ràng.

+ Thống kê Gamma tính theo công thức :
$$G = \frac{(P - Q)}{(P + Q)}$$

+ Thống kê Somers'D tính theo công thức :
$$D_Y = \frac{(P - Q)}{(P + Q + T_Y)}$$

Trong hai công thức trên, P là số cặp biến tương đồng với nhau (nghĩa là cùng tăng hoặc cùng giảm); Q là số cặp biến không tương đồng với nhau (nghĩa là những cặp tăng – giảm khác nhau); Ty – những cặp bị ràng buộc bởi yếu tố Y.

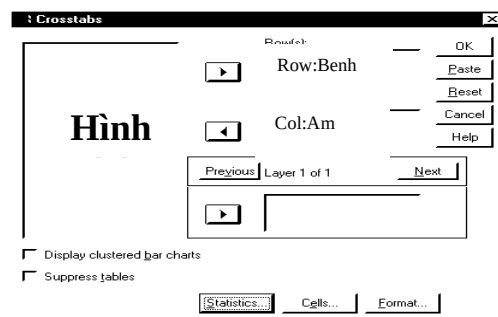
Để kiểm định sự tồn tại của các hệ số Kendall tau b, Kendall tau c, Somers'D, Goodman và Kruskal's Gamma, người ta sử dụng thống kê T.

Giả thuyết H_0 : Từng hệ số (hay đo đặc) = 0.

Đối thuyết H_0 : Từng hệ số (hay đo đặc) $\neq 0$.

Quy tắc quyết định. Khi $P < 0.05$ thì giả thuyết H_0 bị bác bỏ. Ngược lại, khi $P > 0.05$ thì giả thuyết H_0 được chấp nhận.

Thực hành. Giả sử bạn cần xác định quan hệ giữa tình trạng bị bệnh của cây với tình trạng độ ẩm lập địa. Bệnh cây được chia thành 5 cấp và mã hóa tương ứng như sau: không có bệnh (mã hóa = 1), bệnh nhẹ (mã hóa = 2), bệnh trung bình (mã hóa = 3), bệnh nặng (mã hóa = 4) và bệnh rất nặng (mã hóa = 5). Ẩm lập địa cũng được chia thành 5 cấp và mã hóa tương ứng như sau: rất ẩm (mã hóa = 1), ẩm (mã hóa = 2), mát (mã hóa = 3), khô (mã hóa = 4) và rất khô (mã hóa = 5).



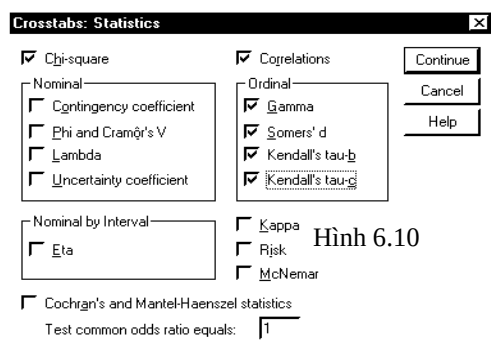
Nhập số liệu vào Bảng tính. Cột 1 (hoặc hàng 1) ghi thứ tự cây được nghiên cứu, cột 2 (hoặc hàng 2) ghi tình trạng bệnh theo mã hóa, cột 3 (hoặc hàng 3) ghi tình trạng âm lập địa theo mã hóa (**Bảng 6.16**).

Gọi chương trình tính toán. Sau khi nhập số liệu vào Bảng tính SPSS, để tạo được kết quả, từ thanh Menu chính bạn hãy thực hiện các thủ tục như sau :

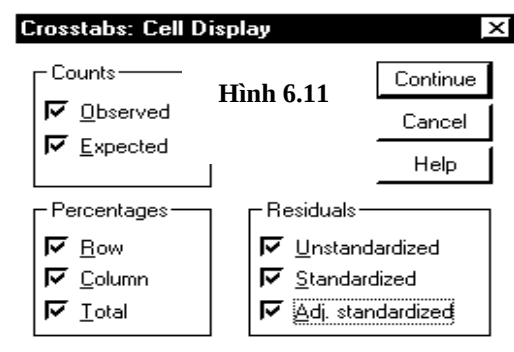
+ **Bước 1.** Mở Analyze $\Rightarrow$ Descriptive Statistics $\Rightarrow$ Crosstabs. Khi cửa sổ Crosstabs xuất hiện (**Hình 6.9**), bạn chọn tình trạng bệnh và chuyển nó vào ô Row, còn tình trạng âm lập địa được đưa vào ô Column.

Bảng 6.16. Cách nhập số liệu để xác định quan hệ giữa hai biến số

Thứ tự	1	2	3	4	5	6	...
Bệnh cây	1	2	1	3	5	4	...
Am lập địa	2	3	4	5	1	2	...



Hình 6.10



Hình 6.11

Bảng 6.17. Tóm tắt xử lý các trường hợp (Case Processing Summary)

	Cases					
	Valid N	Percent	Missing N	Percent	Total N	Percent
Bệnh - Âm	177	100.0%	0	.0%	177	100.0%

Bảng 6.18. Bảng chéo (Bệnh $\times$ Âm)

			Am lập địa					otal
			1	2	3	4	5	
Bệnh cây	1	Count	10	7	6	4	4	31
		Expected Count	5.1	7.9	7.9	5.1	5.1	31.0
	2	Count	3	16	8	3	4	34
		Expected Count	5.6	8.6	8.6	5.6	5.6	34.0
	3	Count	4	7	15	3	5	34
		Expected Count	5.6	8.6	8.6	5.6	5.6	34.0
	4	Count	7	9	7	14	4	41
		Expected Count	6.7	10.4	10.4	6.7	6.7	41.0
	5	Count	5	6	9	5	12	37
		Expected Count	6.1	9.4	9.4	6.1	6.1	37.0
T	otal	Count	29	45	45	29	29	177
		Expected Count	29.0	45.0	45.0	29.0	29.0	177.0

+ **Bước 2.** Từ cửa sổ Crosstabs, bạn vào Statistics và đánh dấu chọn các mục Chi – Square, Correlations, Gamma, Somer's D; kế đến chọn Continue (**Hình 6.10**). Tiếp theo, chọn

Cells và đánh dấu các mục Counts (tần số), Percentages (%) rồi chọn Continue (Hình 6.11). Sau cùng thì chọn OK để chương trình tính toán và hiển thị kết quả. Kết quả hiển thị trên màn Hình bao gồm những Bảng sau đây (Bảng 6.17 - 6.20):

Bảng 6.19. Xu hướng quan hệ giữa tình trạng bệnh cây với tình trạng ẩm lập địa (Directional Measures)

			Value	Asymp. Std. Error ^b	Approx. T ^a	Approx. Sig.
Ordinal by Ordinal	Somers' d	Symmetric	.177	.065	2.731	.006
		Bệnh Dependent	.178	.065	2.731	.006
		Ẩm Dependent	.177	.065	2.731	.006
Nominal by Interval	Eta	Bệnh Dependent	.217			
		Ẩm Dependent	.215			

Bảng 6.20. Đo quan hệ giữa hai biến (Symmetric Measures)

		Value	Asymp. Std. Error ^b	Approx. T ^a	Approx. Sig. ^c
Ordinal by Ordinal	Gamma	.220	.080	2.731	.006
	Spearman Correlation	.210	.077	2.838	.005
Interval by Interval	Pearson's R	.209	.078	2.826	.005
N of Valid Cases		177			

Giải thích và nhận xét kết quả. Căn cứ vào Bảng kết quả, bạn có thể đưa ra nhận xét của mình về những vấn đề mà bạn quan tâm. Chẳng hạn, bạn có thể đưa ra những nhận xét theo gợi ý sau đây:

+ Bảng Crosstabs (Bảng 6.18) trình bày tần số ô bất gặp cả tình trạng bệnh lẫn tình trạng ẩm lập địa. Từ Bảng này, bạn thấy từng cấp tình trạng bệnh rơi vào những cấp tình trạng ẩm lập địa nào ? Bạn có thấy những cây bị bệnh nặng và rất nặng nhưng lại xuất hiện trên những lập địa khô và rất khô hay không ? Câu trả lời sau có thể được làm rõ từ những con số to - nhỏ khác nhau nằm ở ngoài đường chéo của Bảng Crosstabs.

+ Bảng Directional Measures (Bảng 6.19) trình bày các số đo Somers'D (cột Value) về xu hướng của mối quan hệ giữa tình trạng bệnh với tình trạng ẩm lập địa. Thống kê T (cột Approx. T) và mức ý nghĩa phù hợp P (cột Approx. Sig) được đưa ra nhằm kiểm định giả thuyết H_0 : Các hệ số Somers'D = 0 (hay không tồn tại). Khi $P < 0.05$ thì hệ số Somers'D tồn tại thực sự. Từ Bảng 6.19, bạn có thể đi đến nhận xét :

- Giữa các cặp biến độc lập không bị ràng buộc thì tỉ lệ trội hơn thuộc về các cặp tương đồng hay các cặp không tương đồng ? Để trả lời câu hỏi này, bạn quan sát số đo Symmetric Somers'D và hai số đo quan hệ phía dưới. Sự chênh lệch lớn của những số đo này cho biết sự khác nhau giữa các cặp tương đồng (phù hợp) với các cặp không tương đồng. Nếu trị số của chúng gần bằng nhau thì các cặp tương đồng và không tương đồng là gần bằng nhau. Điều đó có nghĩa là hai biến không có quan hệ với nhau hoặc quan hệ với nhau rất yếu. Ở ví dụ này, vì thống kê Symmetric = Bệnh Dependent = Ẩm Dependent = 0.177, nên hai biến có quan hệ với nhau rất yếu.

- Khi bạn sử dụng tình trạng bệnh là biến phụ thuộc, còn tình trạng ẩm lập địa là biến độc lập thì thống kê Somers'D có thay đổi hay không ? Các hệ số Somers'D có tồn tại hay không ?

Lưu ý rằng, trong ví dụ dẫn ra trên đây bạn đã thấy rằng, tình trạng bệnh có quan hệ với tình trạng ẩm lập địa, còn tình trạng ẩm lập địa không phụ thuộc vào tình trạng bệnh. Vì thế, trong quan hệ này biến phụ thuộc là tình trạng bệnh, còn biến độc lập là tình trạng ẩm lập địa.

- Khi xem xét quan hệ giữa hai biến định tính có thứ bậc hơn kém, bạn cần chú ý đến cả thống kê tương quan hạng của Spearman và thống kê Eta. Thống kê Eta (tính theo phần trăm) đo sự thành công của phép dự đoán. Giá trị Eta^2 nhân với 100 cho biết có bao nhiêu phần trăm sự biến thiên của biến phụ thuộc được giải thích bởi sự biến thiên của biến độc lập. Ở ví dụ trên, nếu coi tình trạng ẩm lập địa là biến độc lập để dự đoán tình trạng bệnh cây thì sự thành công của việc dự đoán là 4,7% (trong Bảng Eta = 0,217). Lưu ý rằng, khi các trị số của biến phụ thuộc là biến định lượng, còn biến độc lập là biến phân nhóm có thứ bậc hay không có thứ bậc thì bạn cần chú ý đến hệ số Eta.

+ Bảng Symmetric Measures (Bảng 6.20) trình bày các số đo về tính đối xứng (Symmetric Measures) trong quan hệ giữa hai biến thông qua các hệ số Gamma, Spearman's correlation coefficient và Pearson's correlation coefficient (R). Hai hệ số đầu áp dụng cho thang đo định danh có thứ bậc (Ordinal by Ordinal), còn hệ số Pearson' R áp dụng cho thang đo khoảng (Interval by Interval). Bảng này còn đưa ra thống kê T để kiểm định giả thuyết H_0 : từng hệ số tương quan = 0. Khi $P < 0.05$ thì giả thuyết bị bác bỏ (nghĩa là từng hệ số khác 0). Ngược lại, khi $P > 0.05$ thì từng hệ số = 0. Ở ví dụ này, cả hai biến là thang đo định danh có thứ bậc. Vì thế, bạn hãy chú ý đến thống kê Conditional Gamma. Thống kê Conditional Gamma = 0.2204; điều đó cho biết quan hệ giữa hai biến rất yếu.

6.4. ĐO ĐẠC TÍNH THỐNG NHẤT ĐỐI VỚI MỘT BẢNG R×R Ô

Trong các cuộc điều tra rừng, khi hai điều tra viên cùng đo đạc một đối tượng, bạn cần phải biết họ có đưa ra kết quả giống nhau hay không? Để làm rõ vấn đề đặt ra, trước hết bạn có thể xây dựng một Bảng chéo 2 chiều với số hàng (R) bằng số cột (C) để mô tả kết quả đo đạc của hai điều tra viên. Sau đó, bạn có thể sử dụng đo đạc Kappa để kiểm định tính thống nhất đối với một Bảng R×C ô.

Điều kiện áp dụng. Thống kê Kappa đòi hỏi một Bảng 2 chiều đối xứng (nghĩa là số hàng bằng số cột) và số đo biến thứ nhất đối lập với biến thứ hai (hay biến ở dạng có và không có). Ví dụ: Hai người cùng chẩn đoán bệnh cho một loài cây; trong đó bệnh cây được phân thành 3 nhóm: có bệnh, không có bệnh và nghi ngờ (Bảng 6.21).

Để đo tính thống nhất, bạn có thể sử dụng thống kê Kappa. Thống kê Kappa tính theo công thức: $K = \frac{P_o - P_e}{1 - P_e}$; trong đó P_o và P_e tương ứng là tổng tỉ lệ quan sát và tổng tỉ lệ lí thuyết nằm trên đường chéo từ góc cao bên trái đến góc thấp bên phải.

Khi các ô nằm ngoài đường chéo của Bảng bằng zero thì $K = 1$. Khi $K > 0,75$ thì hai điều tra viên đo đạc cùng một đối tượng có sự thống nhất cao. Khi $K = 0,4 - 0,75$ thì hai điều tra viên đo đạc cùng một đối tượng có sự thống nhất khá đến cao. Khi $K < 0,4$ thì hai điều tra viên đo đạc cùng một đối tượng có sự nhất trí kém.

Bảng 6.21. Kết quả chẩn đoán bệnh K trên 1 loài cây
(Ví dụ về sơ đồ Bảng R×R ô)

		Điều tra viên B		
		Có bệnh	Không có bệnh	Nghi ngờ
Điều tra viên A	Có bệnh	A	B	C
	Không có bệnh	D	E	F
	Nghi ngờ	G	H	I

6.5. KIỂM ĐỊNH TRONG NỘI BỘ CÁC LỚP CỦA MỘT BẢNG NHIỀU CHIỀU

1. Khái niệm về một Bảng chéo nhiều chiều. Bảng chéo nhiều chiều là tập hợp của những Bảng chéo hai chiều; trong đó mỗi biến số là một chiều. Số liệu dùng trong phân tích bao gồm một biến hàng, một biến cột và một biến lớp (Layer variable). Biến lớp còn được gọi là biến điều khiển hay biến lập nhóm (Control/grouping variable). Từ số liệu đó, chương trình SPSS sẽ xây dựng các Bảng phụ 2 chiều cho từng trị số của biến lớp hoặc kết hợp giữa hai hoặc hơn hai biến lớp.

2. Mục đích. Kiểm định trong nội bộ các lớp của một Bảng nhiều chiều (Tests Within Layers of a Multiway Table) là nhằm tìm mối quan hệ giữa các biến trong nội bộ các lớp.

3. Nội dung của phương pháp. Kiểm định trong nội bộ các lớp bao gồm hai nội dung. Một là, kiểm định mối quan hệ giữa các biến số. Hai là, xác định cường độ của mối quan hệ giữa các biến số.

4. Thực hành. Giả sử bạn cần tìm hiểu quan hệ giữa mức độ bị bệnh của Sao đen và Dầu rái với chiều cao của cây (hoặc với tuổi cây). Ngoài ra, bạn cũng muốn biết mối quan hệ này có giống nhau giữa hai loài Sao đen và Dầu rái hay không? Ở ví dụ này, loài cây là yếu tố lớp (Layer) cho một Bảng quan hệ giữa mức độ bị bệnh của cây với chiều cao của cây (tuổi). Biến lớp được biểu thị bằng các giá trị sau: Sao đen, Dầu rái. Biến chiều cao của cây được đo bằng 4 giá trị sau: trên 6 m, 4 - 6 m, 2 - 4 m và nhỏ hơn 2 m. Biến mức độ bị bệnh được đo bằng 3 giá trị: nặng, trung bình, nhẹ.

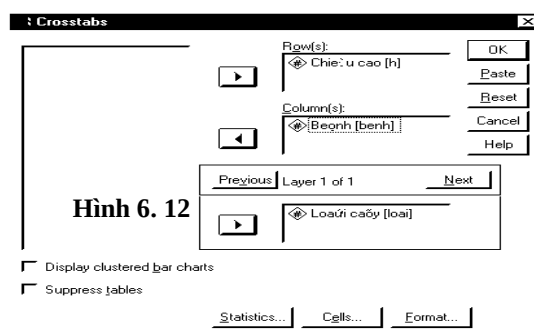
Cách thu thập và ghi chép dữ liệu

+ Trước hết, để đơn giản cho việc ghi chép số liệu tại rừng, bạn cần phải mã hóa số liệu. Biến loài cây được mã hóa là Sao đen = 0, Dầu rái = 1. Biến chiều cao của cây được mã hóa như sau: > 6 m = 1, 4 - 6 m = 2, 2 - 4 m = 3, < 2 m = 4. Bệnh cây được mã hóa như sau: nặng = 1, trung bình = 2, nhẹ = 3.

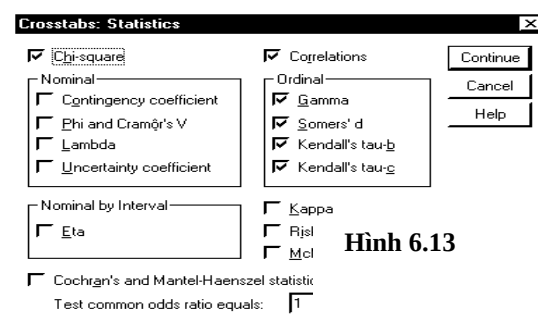
+ Khi tiến hành điều tra một đối tượng nào đó, bạn cần phải thu thập các dữ liệu về loài cây, tình trạng bị bệnh và chiều cao cây (theo cấp). Những biến số này được ghi vào biểu theo mã số quy ước. Để kết quả đạt được chính xác, bạn cần phải thu hàng trăm đến hàng ngàn quan sát khác nhau. Điều đó được đặt ra là vì các phép kiểm định đòi hỏi những tần số của Bảng chéo phải có trên 75% số ô với tần số lớn hơn 5. Kết cấu của phiếu quan sát phải được xây dựng rõ ràng, dễ ghi chép.

Nhập dữ liệu vào Bảng tính SPSS

+ Trước khi gọi chương trình tính toán, bạn hãy nhập dữ liệu vào Bảng tính SPSS như sau (**Bảng 6.22**): Cột 1 nhập thứ tự cây quan sát (có hay không cũng được); cột 2 nhập mã số của các loài cây; cột 3 nhập chiều cao của từng cây (theo mã hóa); cột 4 nhập mức độ bị bệnh của từng cây (theo mã hóa).



Hình 6.12



Hình 6.13

Gọi chương trình tính toán. Để tạo được kết quả, từ thanh Menu chính bạn hãy thực hiện như sau:

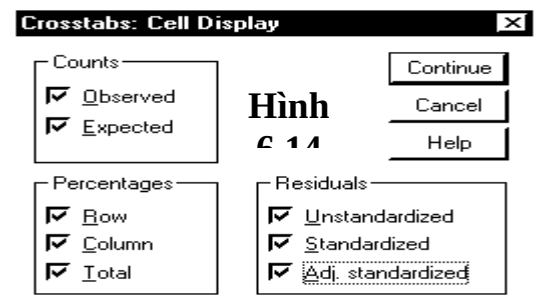
Bước 1. Mở Analyze $\Rightarrow$ Descriptive Statistics $\Rightarrow$ Crosstabs. Khi cửa sổ Crosstabs xuất hiện (Hình 6.12), bạn hãy chuyển biến “chiều cao” vào ô Row, biến “mức bị bệnh” vào ô Column, còn biến “loài cây” vào ô Layer.

Bước 2. Từ cửa sổ Crosstabs, bạn hãy chọn Statistics và đánh dấu chọn ô Chi – square và các mục của ô correlations (Hình 6.13); kế đến chọn continue. Tiếp theo, chọn Cells và đánh dấu chọn các mục Counts, Percentages và Residuals (Hình 6.14).

Bước 3. Sau khi đã chọn các mục cần tính, hãy nhấp nút OK để chương trình xử lý số liệu và hiển thị kết quả. Kết quả hiển thị bao gồm những kiểm định, các đo đặc và tần số trong Bảng 2 chiều. Nội dung của các Bảng như sau (Bảng 6.23 – 6.27).

Giải thích và nhận xét kết quả

1. **Bảng 6.24 báo cáo** kiểm định Chi – Square về tính độc lập của các biến số. Giả thuyết H_0 cho rằng mức độ bị bệnh của các loài cây không phụ thuộc vào chiều cao và loài cây. Để xác nhận tính đúng đắn của giả thuyết H_0 , bạn hãy quan sát giá trị Value và P – value (Asymp. Sig. (2-sided)) của kiểm định Pearson Chi – Square và kiểm định Likelihood Ratio Chi – Square (dùng thay thế cho kiểm định Pearson Chi – Square trong mô hình tuyến tính logarit). Khi $P > 0.05$ (mặc định) thì chấp nhận giả thuyết H_0 , nghĩa là mức độ bị bệnh của các loài cây không phụ thuộc vào chiều cao và loài cây. Ngược lại, $P < 0.05$ (mặc định) thì bác bỏ giả thuyết H_0 , nghĩa là mức độ bị bệnh của các loài cây phụ thuộc vào chiều cao và loài cây. Ở ví dụ này, $P < 0.05$ cho biết mức độ bị bệnh ở Dầu rái và Sao đen thay đổi tùy theo chiều cao (tuổi) và loài cây.



Bảng 6.22. Cách nhập dữ liệu của một Bảng nhiều chiều

TT	Loài cây	Cấp chiều cao	Mức bị bệnh
1	1	1	1
2	0	2	2
3	0	4	3
4	1	3	2
5	0	2	4
...	...	...	...

Bảng 6.23. Tóm tắt về các trường hợp (Case Processing Summary)

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Chiều cao $\times$ Bệnh $\times$ Loài cây	288	100.0%	0	.0%	288	100.0%

Bảng 6.24. Kiểm định Chi – Square về tính độc lập của bệnh với H cây

Loài cây	Thống kê	Value	df	Asymp. Sig. (2-sided)
Dầu rái	Pearson Chi-Square	96.084a	6	.000
	Likelihood Ratio	94.713	6	.000

	Linear-by-Linear Association N of Valid Cases	62.227 146	1	.000
Sao đen	Pearson Chi-Square	94.244b	6	.000
	Likelihood Ratio	99.837	6	.000
	Linear-by-Linear Association	61.528	1	.000
	N of Valid Cases	142		

Bảng 6.25. Tần số về tình trạng bệnh và chiều cao cây
(Chiều cao × Bệnh × Loài cây Crosstabulation)

Loài cây			Bệnh			Tổng
			Nặng	Trung bình	Nhẹ	
Dầu rái	Chiều cao	> 6 m	1	10	34	45
		4-6 m	1	28	8	37
		2-4m	10	14	10	34
		< 2 m	22	7	1	30
	Tổng		34	59	53	146
Sao đen	Chiều cao	>6m	2	14	29	45
		4-6m		25	10	35
		2-4m	8	8	6	22
		<2m	32	5	3	40
	Tổng		42	52	48	142

Bảng 6.26. Các đo đặc xu hướng của mối quan hệ (Directional Measures)

Loài cây	Thống kê		Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Dầu rái	Somers' d	Symmetric	-.584	.054	-10.459	.000
		Chiều cao Dependent	-.626	.056	-10.459	.000
		Bệnh Dependent	-.548	.052	-10.459	.000
Sao đen	Somers' d	Symmetric	-.586	.057	-10.164	.000
		Chiều cao Dependent	-.618	.059	-10.164	.000
		Bệnh Dependent	-.558	.056	-10.164	.000

Thống kê Linear-by-Linear Association chỉ dùng cho các biến định lượng và biến định tính có thứ bậc. Ở ví dụ này, nó đo quan hệ tuyến tính giữa chiều cao (tuổi) với mức độ bị bệnh của cây. Giả thuyết H_0^+ : Giữa chiều cao với mức độ bị bệnh của cây không tồn tại quan hệ tuyến tính. Đối thuyết H_0^- : Giữa chiều cao với mức độ bị bệnh của cây tồn tại quan hệ tuyến tính. Khi $P > 0.05$ (mặc định) thì chấp nhận giả thuyết H_0 . Ngược lại, khi $P < 0.05$ (mặc định) thì bác bỏ giả thuyết H_0 .

2. Khi theo dõi kết quả của **Bảng 6.25**, bạn có thể đưa ra nhận xét: (1) Tần số Sao đen và Dầu rái phân bố ở từng lớp chiều cao có khác nhau hay không? (2) Bao nhiêu Sao đen và Dầu rái bị bệnh (nặng, trung bình và nhẹ)? Sao đen bị bệnh nhiều hơn hay Dầu rái bị bệnh nhiều hơn? Tương tự như vậy, bạn có thể mở rộng các câu hỏi cho những vấn đề mà bạn quan tâm.

3. Biến chiều cao cây là một biến định lượng, nhưng ở đây nó đã được đo bằng thang đo thứ bậc. Biến mức độ bị bệnh của cây là biến định tính có thứ bậc hơn kém. Hai biến này phân biệt rõ ràng với nhau. Vì thế, nếu bạn muốn dự đoán biến chiều cao trên cơ sở biến mức độ bị bệnh, hoặc ngược lại dự đoán biến mức độ bị bệnh trên cơ sở biến chiều cao, thì bạn hãy quan sát thống kê Somer's trong Bảng Directional Measures (đo đặc xu hướng mối quan hệ, **Bảng 6.26**) và các thống kê Spearman Correlation trong Bảng Symmetric Measures (đo cường

độ của mối quan hệ). Từ Bảng 6.26 và 6.27 cho thấy, giữa chiều cao với mức độ bị bệnh của cây có quan hệ tuyến tính âm chặt chẽ (hệ số Somer = - 0.584 và - 0.586). Điều đó có nghĩa là mức độ bị bệnh của cả Sao đen lẫn Dầu rái đều giảm khi chiều cao hay tuổi cây tăng lên và ngược lại.

Bảng 6.27. Các đo đặc cường độ mối quan hệ (Symmetric Measures)

Loài	Thống kê	Value	Std. Error ^a	Approx. T ^b	Approx. Sig.
Dầu rái	Kendall's tau-b	-.586	.054	-10.459	.000
	Kendall's tau-c	-.611	.058	-10.459	.000
	Gamma	-.758	.057	-10.459	.000
	Spearman Correlation	-.646	.056	-10.152	.000c
	Pearson's R	-.655	.053	-10.405	.000c
	N	146			
Sao đen	Kendall's tau-b	-.587	.057	-10.164	.000
	Kendall's tau-c	-.616	.061	-10.164	.000
	Gamma	-.749	.061	-10.164	.000
	Spearman Correlation	-.650	.060	-10.119	.000c
	Pearson's R	-.661	.057	-10.411	.000c
	N	142			

Chương VII

SỬ DỤNG THỦ TỤC SO SÁNH CÁC SỐ TRUNG BÌNH

Thủ tục so sánh các số trung bình (Compare Means ➤ Means) cho phép bạn không chỉ phân tích các dữ liệu trong một Bảng chéo theo nhiều lớp (yếu tố lập nhóm hay yếu tố phân loại) khác nhau, mà còn thực hiện được phân tích phương sai kèm theo phép kiểm định về tuyến tính. Kiểm định về tuyến tính (Test of Linearity) thực hiện so sánh tương phản tuyến tính. Mục đích của so sánh tương phản tuyến tính là nhằm chỉ ra độ dốc của đường thẳng đi ngang qua các trung bình mẫu với thứ bậc hơn kém có khác Zero hay không. Nếu độ dốc của đường thẳng khác Zero thực sự (hay độ dốc của đường thẳng tồn tại thực sự) thì biến phụ thuộc có quan hệ tuyến tính với biến lập nhóm. Nếu điều đó không thỏa mãn thì mối quan hệ có thể tồn tại ở các dạng khác như quan hệ bậc 2, bậc 3... Dưới đây chúng ta lần lượt xem xét những vấn đề mà thủ tục Compare Means ➤ Means cho phép.

7.1. NHỮNG THÔNG KÊ MÔ TẢ TRONG MỘT BẢNG CHÉO NHIỀU YẾU TỐ

7.1.1. Đặt vấn đề

Trong thực hành, nhà lâm nghiệp thường có những yêu cầu như sau: (1) thống kê các đặc trưng của các quần thụ (thuần loài đồng tuổi, hỗn giao khác tuổi...) đưa vào chặt nuôi dưỡng hoặc khai thác chính...; (2) so sánh các nghiệm thức dựa trên nhiều biến số... Chẳng hạn, khi thống kê những đặc trưng của các quần thụ đưa vào chặt nuôi dưỡng, nhà lâm học thường phải ghi nhận đường kính thân cây ($D_{1.3}$, cm), chiều cao thân cây (H , m), trữ lượng thân cây (M , m^3)... tương ứng với những cây thuộc 5 cấp sinh trưởng của Kraft. Vấn đề đặt ra là hãy cho biết các đặc trưng thống kê mô tả đường kính thân cây ($D_{1.3}$, cm), chiều cao thân cây (H , m), trữ lượng thân cây (M , m^3)... tương ứng với 5 cấp sinh trưởng của Kraft. Ngoài ra, hãy cho biết các đặc trưng thống kê cơ bản (tần số xuất hiện các trường hợp của mỗi biến (D , H , M) trong một Bảng chéo, trị trung bình và sai số chuẩn của các biến tùy thuộc vào cấp sinh trưởng của chúng...) của nhóm cây để lại tiếp tục nuôi dưỡng và nhóm cây bị loại bỏ thông qua tía thưa.

Để đạt được mục đích đặt ra trên đây, bạn có thể sử dụng thủ tục Reports, Descriptives, Frequencies, Explore và Compare Means - Means. Các thủ tục Reports, Descriptives, Frequencies và Explore đã được trình bày ở chương 2. Ở đây chúng ta thử giải quyết vấn đề đặt ra theo thủ tục Compare Means - Means. Bằng thủ tục Compare Means - Means, bạn có thể phân tích không chỉ các đặc trưng thống kê theo từng biến lập nhóm (hay từng lớp - Layer) mà còn kiểm định sự khác biệt giữa các nhóm. Vì thế, bạn cần phải chọn biến số để phân nhóm (hay lớp). Khi sử dụng thủ tục Means, với ví dụ trên đây bạn có thể xem những cấp sinh trưởng cây rừng của Kraft là các biến lập nhóm (Grouping Variables) thuộc lớp 1 (Layer 1); cây để lại nuôi dưỡng hay chặt bỏ là biến lập nhóm thuộc lớp 2 (Layer 2). Tương tự như thế, bạn có thể xây dựng Bảng phân tích cho các trường hợp với hơn 2 lớp.

7.1.2. Thực hành thủ tục Compare Means - Means

Giả sử cần xác định những đặc trưng của các quần thụ đưa vào chặt nuôi dưỡng, đồng thời so sánh sự khác biệt giữa nhóm cây bị loại bỏ và nhóm cây để lại nuôi dưỡng. Để đạt được mục đích đặt ra, bạn có thể thực hiện như sau :

1. Thu thập và ghi chép số liệu. Trước hết, bạn hãy phân chia những cây rừng thành 5 cấp sinh trưởng theo cách phân cấp của Kraft; trong đó cây cấp I, II, III, IV và V được mã

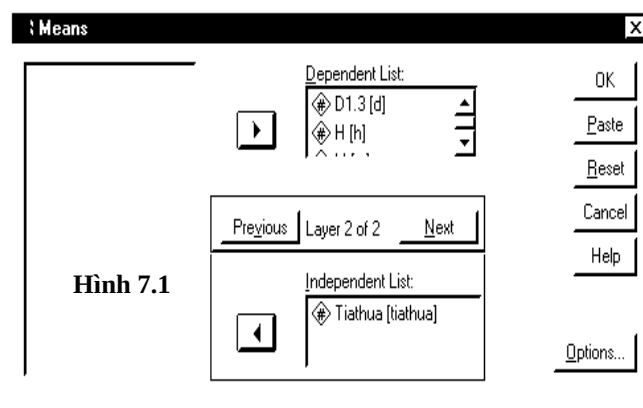
hóa tương ứng là 1, 2, 3, 4 và 5. Tiếp theo, bạn phân chia toàn bộ cây rừng trong quần thụ thành hai nhóm: nhóm 1 là những cây để lại tiếp tục nuôi dưỡng (mã hóa là 1), còn nhóm 2 là những cây đưa vào thu hoạch trung gian hay tía thừa (mã hóa là 0). Sau khi hoàn thành hai nội dung trên đây, bạn thống kê D, H và thể tích thân cây (V) của từng cá thể cây rừng trong quần thụ và ghi lại thành Bảng sau đây (**Bảng 7.1**):

Bảng 7.1. Phiếu đo đếm cây rừng trong quần thụ...

TT cây	D _{1.3} , cm	H, m	Thể tích (V, m ³)	Cấp sinh trưởng Kraft	Cây nuôi dưỡng và cây chặt bỏ
1	6.0	5.5	0.0140	5	0
2	7.5	6.5	0.0251	3	1
3	10.5	9.5	0.0680	1	1
...	...	...	...	...	...

2. Gọi chương trình tính toán. Sau khi đã nhập dữ liệu vào Bảng tính, để nhận được kết quả, bạn hãy thực hiện các bước sau đây :

Bước 1. Từ thanh Menu chính, chọn Analyze ⇒ Compare Means ⇒ Means. Khi cửa sổ Means xuất hiện (**Hình 7.1**), trước hết bạn hãy nhập 3 biến phụ thuộc (D, H và V) vào ô Dependent List. Kế đến nhập lần lượt từng biến độc lập (biến phân loại, phân nhóm, phân lớp) vào ô Independent List theo thứ tự như sau:



Hình 7.1

1. Chọn biến “*Cấp sinh trưởng Kraft*” và nhấp mũi tên bên cạnh để chuyển nó vào ô Independent List.

2. Sau khi nhập biến “*Cấp sinh trưởng*” vào ô Independent List, bạn thấy xuất hiện thêm hai nút Next và Previous (biến hay lớp trước); đồng thời dòng chữ Layer 1 of 1 đổi thành Layer 2 of 2. Dấu hiệu này cho biết rằng biến độc lập 1 hay lớp 1 (Layer 1 of 1) với tên là “*Cấp sinh trưởng*” đã được chương trình ghi nhận. Chữ Next xuất hiện có nghĩa là chương trình SPSS hỏi bạn có muốn đưa tiếp biến độc lập thứ 2 (lớp 2) vào hay không ? Nếu đồng ý, bạn hãy chọn Next. Khi đó dòng chữ Layer 1 of 1 đổi thành Layer 2 of 2, đồng thời xuất hiện thêm chữ Previous. Khi bạn đồng ý đưa biến độc lập thứ 2 (lớp 2) vào phân tích thì chọn Next, tiếp đến chọn biến độc lập thứ 2 (ở đây là nhóm cây nuôi dưỡng và cây chặt bỏ) và nhấp nút mũi tên để đưa nó vào ô Independent List.

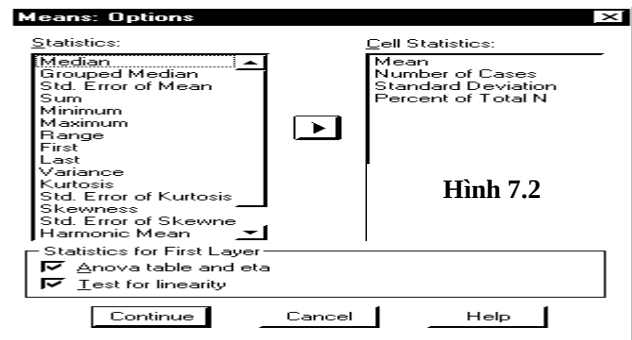
3. Nếu muốn đưa thêm các biến độc lập khác (lớp 3, 4...) vào phân tích thì bạn cũng thực hiện tương tự như việc đưa vào biến 1 và 2. Nếu bạn có ý định thay đổi hay xem xét lại biến độc lập đã được đưa vào ở những bước trước đó thì nhấp nút Previous và chọn Remove (bỏ biến). Đến đây việc nhập dữ liệu đã xong.

Bước 2. Để yêu cầu chương trình tính các đặc trưng thống kê, bạn hãy chọn Options. Khi cửa sổ Means: Option xuất hiện (**Hình 7.2**), bạn hãy chọn các thống kê mô tả cần quan tâm từ ô Statistics và nhấp nút mũi tên bên cạnh để đưa chúng sang ô Cell Statistics. Nếu cần phân tích sự khác biệt giữa các nhóm, bạn hãy chọn Anova table and eta và Test for linearity trong ô Statistics for First Layer. Sau đó nhấp nút Continue và nhấp tiếp nút OK để chương trình tính toán và hiển thị kết quả.

3. Giải thích và nhận xét kết quả. Khi bạn thực hiện những thủ tục trên đây thì kết quả hiển thị trên màn Hình những nội dung sau đây: (1) tóm tắt xử lý các trường hợp (**Bảng 7.2**), (2) báo cáo tóm tắt các đặc trưng thống kê mẫu (**Bảng 7.3**), (3) phân tích biến động (**Bảng 7.4**), (4) các số đo mối quan hệ (**Bảng 7.5**). Nội dung của các Bảng này như sau :

+ **Bảng 7.2** ghi lại tóm tắt các trường hợp xử lý (Case Processing Summary). Nội dung của Bảng này bao gồm những báo cáo về các trường hợp đưa vào tính toán, về những trường hợp bị loại bỏ và tổng số các trường hợp của các biến phụ thuộc được phân loại theo 2 lớp (2 biến phân loại) khác nhau.

+ **Bảng 7.3** báo cáo kết quả xử lý thống kê theo 2 lớp chính. Lớp 1 là những thống kê cho những nhóm cây rừng được phân theo cấp sinh trưởng của Kraft. Lớp 2 là những thống kê cho những nhóm cây đưa vào nuôi dưỡng. Lưu ý rằng lớp 2 bao gồm hai phân lớp: (1) những cây để lại nuôi dưỡng, (2) những cây sẽ chặt bỏ (tỉa thưa).



Hình 7.2

Bảng 7.2. Tóm tắt xử lý các trường hợp (Case Processing Summary)

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
D1.3 × Kraft × Tiathua	285	100.0%	0	.0%	285	100.0%
H × Kraft × Tiathua	285	100.0%	0	.0%	285	100.0%
M × Kraft × Tiathua	285	100.0%	0	.0%	285	100.0%

Bảng 7.3. Báo cáo tóm tắt các đặc trưng thống kê mẫu (Report)

Kraft	Tiathua		D1.3	H	M
Cây cấp I	Nuoiduong	Mean	8.600	8.340	4.82900E-02
		N	30	30	30
		Std. Deviation	.203	1.492	1.46584E-02
		% of Total N	10.5%	10.5%	10.5%
... (tương tự từ II – V)					
Total	Tiathua	Mean	5.045	4.145	8.20909E-03
		N	66	66	66
		Std. Deviation	1.224	1.642	7.31213E-03
		% of Total N	23.2%	23.2%	23.2%
	Nuoiduong	Mean	7.418	6.895	2.93260E-02
		N	219	219	219
		Std. Deviation	.821	1.305	1.30564E-02
		% of Total N	76.8%	76.8%	76.8%
	Total	Mean	6.868	6.258	2.44358E-02
		N	285	285	285
		Std. Deviation	1.366	1.810	1.49239E-02
		% of Total N	100.0%	100.0%	100.0%

Bảng 7.4. Phân tích biến động (ANOVA)

			Sum of Squares	df	Mean Square	F	Sig.
D1.3 × Kraft	Between Groups	(Combined)	452.262	4	113.066	408.214	.000

		Linearity	426.026	1	426.026	1538.129	.000
		Deviation from Linearity	26.237	3	8.746	31.575	.000
	Within Groups		77.553	280	.277		
	Total		529.816	284			
H × Kraft	Between Groups	(Combined)	528.441	4	132.110	92.138	.000
		Linearity	483.438	1	483.438	337.165	.000
		Deviation from Linearity	45.003	3	15.001	10.462	.000
	Within Groups		401.474	280	1.434		
	Total		929.915	284			
M×Kraft	...						

Khi phân tích **Bảng 7.3**, bạn có thể nhận được hầu hết các đặc trưng thống kê mô tả cho các biến D, H, M của từng nhóm cây thuộc các đối tượng khác nhau (cấp sinh trưởng và nhóm cây đưa vào chặt nuôi dưỡng).

Bảng 7.5. Các số đo mối quan hệ (Measures of Association)

	R	R Squared	Eta	Eta Squared
D _{1,3} × Kraft	-.897	.804	.924	.854
H × Kraft	-.721	.520	.754	.568
M × Kraft	-.780	.608	.788	.621

+ **Bảng 7.4** báo cáo kết quả phân tích biến động (ANOVA) nhằm so sánh sự khác nhau về D, H và M giữa các cấp sinh trưởng và giữa nhóm cây để lại nuôi dưỡng với nhóm cây bị loại bỏ. Khi phân tích **Bảng 7.4**, bạn có thể thấy:

- Các đại lượng D, H và M của những cây thuộc 5 cấp sinh trưởng có khác nhau hay không? Để trả lời câu hỏi này, bạn hãy quan sát giá trị F và mức ý nghĩa thống kê (P hay Sig.). Khi $P < 0.05$ thì giả thuyết cho rằng các đại lượng D, H và M của 5 cấp sinh trưởng là bằng nhau sẽ bị bác bỏ với mức tin cậy 95% và ngược lại.

- Các trị trung bình của từng biến (D, H và M) thuộc 5 nhóm cây có nằm trên cùng một đường thẳng hay không? Để trả lời câu hỏi này, bạn hãy quan sát giá trị F và mức ý nghĩa thống kê (P hay Sig.) tương ứng với dòng Linearity. Khi $P < 0.05$ (mặc định) thì giả thuyết Ho sẽ bị bác bỏ với mức tin cậy 95% và ngược lại.

+ **Bảng 7.5** báo cáo kết quả về phân tích mối quan hệ tuyến tính giữa từng nhân tố D, H và M với 5 cấp sinh trưởng cây rừng. Phương hướng và mức độ chặt chẽ của mối quan hệ được trình bày ở cột 2 đến 5 của **Bảng 7.5**. Ở ví dụ này, các biến D, H và M có mối quan hệ âm chặt chẽ với cấp sinh trưởng của cây rừng.

7.2. PHÂN TÍCH PHƯƠNG SAI MỘT YẾU TỐ CÙNG VỚI KIỂM ĐỊNH TUYẾN TÍNH

Thủ tục Compare Means ➤ Means còn cho phép bạn kiểm định giả thuyết về mối quan hệ tuyến tính giữa biến phụ thuộc với biến độc lập. Lưu ý rằng các biến phụ thuộc được đo bằng giá trị định lượng, còn các biến độc lập có thể được đo bằng giá trị định lượng hoặc định tính. Nếu biến độc lập được đo bằng giá trị định tính thì nó cũng có thể là biến định danh có thứ bậc hơn kém hay không có thứ bậc hơn kém. Nếu biến độc lập được đo bằng giá trị định lượng thì chúng phải được sắp xếp thành cấp; mỗi cấp được mã hóa bằng một con số nguyên liên tục có thứ bậc hơn kém hay không có thứ bậc hơn kém. Việc chọn cách mã hóa nào là tùy thuộc vào mục tiêu nghiên cứu hoặc ý tưởng của nhà nghiên cứu.

Để thực hành thủ tục Compare Means ➤ Means, bạn hãy xem xét ví dụ sau đây. Vấn đề đặt ra là hãy xác định quan hệ giữa năng suất rừng (biến phụ thuộc) với 5 cấp độ dày tầng đất khác nhau (biến độc lập dùng để phân loại cấp năng suất). Năng suất rừng được đo bằng trữ lượng thân cây (M , m^3/ha). Độ dày tầng đất được đo bằng đơn vị xentimét (X_i , cm) và được phân chia thành 5 cấp : < 50, 50 - 100, 101 - 150, 150 - 200 và trên 200 cm.

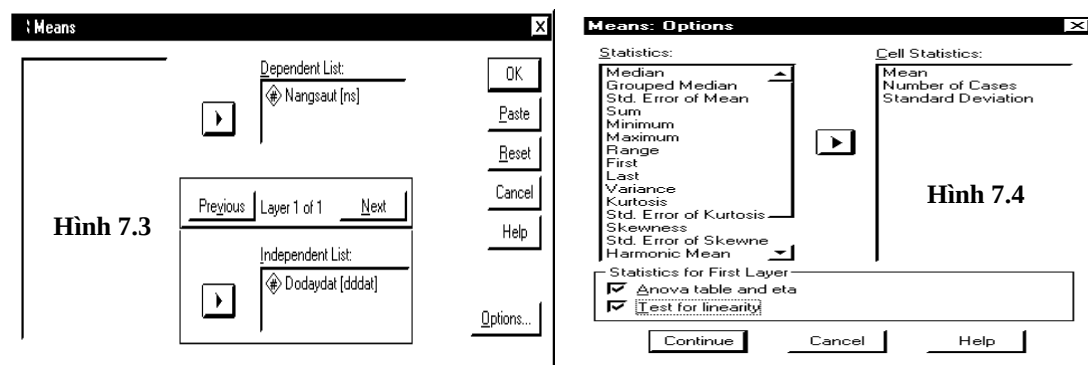
Để giải quyết bài toán đặt ra trên đây bằng thủ tục Compare Means ➤ Means, bạn hãy mã hóa 5 cấp độ dày tầng đất bằng các số theo thứ tự từ 1 đến 5. Ở đây cần lưu ý rằng, việc mã hóa cấp độ dày tầng đất từ 1 đến 5 chỉ là một quy ước. Nếu không mã hóa cấp độ dày tầng đất như trên, bạn cũng có thể mã hóa theo thứ tự ngược lại từ 5 đến 1. Nhưng ở đây việc quy ước cách mã hóa cấp độ dày tầng đất từ 1 đến 5 là nhằm mục đích tạo sự dễ dàng cho việc giải thích mối quan hệ giữa năng suất rừng với độ dày tầng đất. Bởi vì, nhà lâm học đã biết về cơ bản cấp độ dày tầng đất tăng lên thì năng suất rừng cũng tăng lên. Khi tiến hành đo đạc tại rừng, bạn phải lập ô đo đếm trữ lượng rừng (theo loài cây), đồng thời xác định cấp độ dày tầng đất tương ứng với mỗi cấp năng suất rừng. Để phản ánh chính xác quan hệ giữa năng suất rừng với độ dày tầng đất, bạn cần thu thập dữ liệu trên N ô tiêu chuẩn. Số lượng ô tiêu chuẩn cần đo đếm phụ thuộc vào biến động của năng suất rừng theo độ dày tầng đất. Khi đã thu thập đủ dữ liệu, bạn có thể nhập chúng vào Bảng tính SPSS như sau (Bảng 7.6): Cột 1 (hàng 1) nhập số thứ tự ô tiêu chuẩn (có hay không cũng được), cột 2 (hàng 3) nhập năng suất rừng, cột 3 nhập cấp độ dày tầng đất theo số mã hóa.

Bảng 7.6. Năng suất rừng tương ứng với các cấp độ dày tầng đất (tóm tắt)

M (m^3/ha)	26.0	18.0	19.0	25.0	24.0	22.0	22.5	23.5	23.5	17.0
X_i (số mã hóa)	5	3	3	4	4	3	3	4	4	2

Sau khi đã nhập xong số liệu vào Bảng tính SPSS, bạn hãy ghi chúng thành một File (để chống mất dữ liệu và có thể sử dụng cho những mục đích khác) và thực hiện các bước gọi chương trình tính toán như sau :

Bước 1. Từ thanh Menu chính, bạn hãy mở Analyze ➤ Compare Means ➤ Means. Khi cửa sổ Means xuất hiện, bạn hãy nhập biến phụ thuộc (năng suất rừng) vào Dependent List, nhập biến độc lập (cấp độ dày tầng đất) vào ô Independent List (Hình 7.3).



Bước 2. Từ cửa sổ Means, bạn hãy mở Options. Khi cửa sổ Means: Options xuất hiện, bạn hãy chuyển những thống kê cần tính từ ô Statistics sang ô Cell Statistics bằng cách chọn từng thống kê và nhấp mũi tên bên cạnh. Sau đó bạn đánh dấu chọn Bảng ANOVA, hệ số Eta và Test for Linearity (kiểm định tuyến tính) trong ô Statistics for First Layer (Hình 7.4). Thao tác này nhằm yêu cầu chương trình tính các thống kê mô tả và kiểm định mối quan hệ tuyến

tính giữa hai biến. Tiếp theo bạn hãy chọn Continue và nhấp OK để chương trình xử lý và hiển thị kết quả.

Giải thích và nhận xét kết quả. Kết quả xử lý thống kê đưa ra 4 nội dung: (1) tóm tắt các trường hợp xử lý (Case processing summary), (2) báo cáo kết quả về năng suất rừng bình quân theo độ dày tầng đất (Report), (3) Bảng phân tích ANOVA (ANOVA table), (4) các số đo mối quan hệ (Measures of Association). Dưới đây chỉ trích dẫn Bảng báo cáo kết quả (Report) về năng suất rừng bình quân theo độ dày tầng đất (**Bảng 7.6**), Bảng phân tích ANOVA (**Bảng 7.7**) và Bảng số đo mối quan hệ (Measures of Association) giữa hai biến (**Bảng 7.8**). Từ số liệu của **Bảng 7.6**, bạn có thể sơ bộ nhận xét về xu hướng biến đổi của năng suất rừng bình quân và mức biến động của chúng theo 5 cấp độ dày tầng đất.

Bảng 7.6. Báo cáo tóm tắt về năng suất rừng theo độ dày tầng đất (Report)

Độ dày tầng đất, cm	Mean	N	Std. Deviation	CV%
< 50	11.812	72	2.603	22.0
50 - 100	16.250	18	.772	4.8
100 - 150	20.375	36	1.943	9.5
150 - 200	24.000	36	.621	2.6
> 200	28.000	36	2.121	7.6
Total	18.932	198	6.582	34.8

Bảng 7.7. Phân tích phương sai (ANOVA Table)

			Sum of Squares	df	Mean Square	F	Sig.
Nangsaut × Dodaydat	Between Groups	(Combined)	7738.798	4	1934.700	469.959	.000
		Linearity	7731.376	1	7731.376	1878.033	.000
		Deviation from Linearity	7.422	3	2.474	.601	.615
	Within Groups		794.531	193	4.117		
	Total		8533.330	197			

Bảng 7.8. Các số đo mối quan hệ (Measures of Association)

	R	R Squared	Eta	Eta Squared
Nangsaut × Dodaydat	.952	.906	.952	.907

Bảng 7.7 ghi lại kết quả phân tích ANOVA nhằm kiểm định :

+ *Sự sai khác giữa các trung bình nhóm* (ở đây là năng suất rừng trên 5 cấp độ dày tầng đất). Giả thuyết H_0 : Trung bình năng suất rừng theo 5 cấp độ dày tầng đất khác nhau không có ý nghĩa thống kê. Để kiểm định giả thuyết này, bạn hãy quan sát thống kê F và mức ý nghĩa thống kê (Sig = P) tương ứng với dòng Between Groups. Nếu $P < 0.05$ (mặc định) thì các trung bình nhóm khác nhau thực sự với mức tin cậy 95%. Ngược lại, khi $P > 0.05$ thì các trung bình nhóm khác nhau không có ý nghĩa thống kê. Ở ví dụ này, $P < 0.000$ (số làm tròn của < 0.0005) nên giả thuyết H_0 bị bác bỏ.

+ *Quan hệ tuyến tính giữa hai biến số M với X* (ở đây M là năng suất rừng, còn X là cấp độ dày tầng đất). Dòng Linearity và Deviation from linearity báo cáo kết quả về kiểm định mối quan hệ tuyến tính giữa M với X. Giả thuyết H_0 : giữa M với X không tồn tại quan hệ tuyến tính (hoặc các trung bình về năng suất rừng (M_{itb}) không cùng nằm trên một đường

thẳng với độ dốc khác zero). Để xác định giả thuyết trên là đúng hay sai, bạn hãy quan sát thống kê F và mức ý nghĩa thống kê (Sig = P) của dòng Linearity. Khi $P < 0.05$ (mặc định) thì giả thuyết H_0 bị bác bỏ. Ngược lại, khi $P > 0.05$ (mặc định) thì giả thuyết H_0 được chấp nhận. Ở ví dụ này, vì $P = 0.000 < 0.01$, nên giả thuyết H_0 bị bác bỏ. Lưu ý rằng, để kiểm định về tuyến tính, bên cạnh thống kê Linearity, chương trình còn sử dụng thống kê độ lệch bình phương bình quân (phương sai) so với tuyến tính (Deviation from linearity mean square). Giả thuyết H_0 : Trị trung bình của các nhóm (năng suất rừng) không nằm trên cùng một đường thẳng. Để xác định giả thuyết trên là đúng hay sai, bạn hãy quan sát thống kê F và Sig (P) tương ứng với dòng Deviation from linearity. Khi $P > 0.05$ thì giả thuyết H_0 được chấp nhận. Ngược lại, nếu $P < 0.05$ thì giả thuyết H_0 bị bác bỏ. Ở ví dụ này, vì $P = 0.615$ nên giả thuyết H_0 được chấp nhận. Lưu ý rằng, nếu giả thuyết về quan hệ tuyến tính giữa M với X bị bác bỏ thì bạn có thể tiếp tục phân tích mối quan hệ này ở dạng bậc 2, bậc 3... Để phân tích mối quan hệ giữa hai biến số ở dạng bậc 2 và bậc 3..., bạn cần thực hiện thủ tục phân tích phương sai một yếu tố bằng cách mở Analyze - Compare means - Means – One Way Anova.

Bảng 7.8 trình bày 2 hệ số quan hệ :

- Hệ số R đo mức độ chặt chẽ của mối quan hệ tuyến tính giữa biến phụ thuộc với biến độc lập. R^2 nhân với 100 giải thích bao nhiêu phần trăm sự biến thiên của biến phụ thuộc do biến độc lập gây ra. Ở ví dụ này, 90.6% sự biến thiên của biến phụ thuộc gây ra bởi biến độc lập. Nói khác đi, biến độ dày tầng đất có ảnh hưởng thực sự đến năng suất rừng.

- Hệ số Eta^2 (tỷ tương quan) **bằng tổng các độ lệch bình phương giữa các tổ** (Between groups sum of square) **chia cho tổng các độ lệch bình phương toàn bộ** (Total sum of square). Từ **Bảng 7.8** ta có $Eta^2 = (7738,798/8533,330) = 0,907$. Hệ số Eta^2 cũng giải thích bao nhiêu phần trăm sự biến thiên của biến phụ thuộc là do biến độc lập gây ra. Lưu ý rằng, hệ số Eta thường được sử dụng trong phân tích phương sai, vì nó không đòi hỏi các nhóm của biến độc lập phải có thứ bậc.

Chương VIII

PHÂN TÍCH NHÓM

8.1. Khái quát về phân tích nhóm

8.1.1. Định nghĩa phân tích nhóm

Phân tích nhóm (Grouping Analysis) là phân chia một tập dữ liệu lớn thành nhiều nhóm nhỏ dựa trên các biến phân loại định lượng. Biến phân loại còn được gọi là biến lập nhóm, biến dự đoán, biến độc lập. Những biến có ý nghĩa phân loại được sử dụng để xây dựng các hàm lập nhóm (Discriminant, Grouping, Classification Functions). Hàm lập nhóm tuyến tính tổng quát có dạng như Hàm 8.1. Ở Hàm 8.1, $Z^{(k)}$ là hàm của nhóm thứ k ($k = 1, 2, \dots$); $X_1^{(k)}, X_2^{(k)}, \dots, X_i^{(k)}$ là biến phân loại thứ i của hàm lập nhóm thứ k ; $a, b, c, \dots, f$ là các hệ số của hàm lập nhóm. Các hệ số của hàm lập nhóm được xác định theo phương pháp lập nhóm. Hàm lập nhóm có thể bao gồm nhiều biến phân loại. Hàm lập nhóm tối ưu chỉ bao gồm một số biến phân loại hữu ích nhất hay khác biệt rõ rệt nhất giữa các nhóm.

$$Z^{(k)} = a \times X_1^{(k)} + b \times X_2^{(k)} + c \times X_3^{(k)} + \dots + f \times X_i^{(k)} \quad (8.1)$$

Để mô tả sự khác nhau giữa các nhóm, phương pháp phân tích nhóm xây dựng các hàm lập nhóm tuyến tính. Hàm này do Fisher đề xuất vào năm 1936, nên nó còn được gọi là hàm lập nhóm tuyến tính Fisher. Khi chỉ có hai biến phân loại, thì hàm lập nhóm tuyến tính Fisher có dạng phương trình của một mặt phẳng. Khi số nhóm lớn hơn 2, các biến hợp quy (Canonical Variables) là trung tâm của phép phân tích nhóm. Biến hợp quy là sự kết hợp tuyến tính của các biến phân loại. Biến hợp quy được sử dụng để làm khác nhau cực đại giữa các số trung bình của k nhóm theo một phương nào đó. Khi có k nhóm, thì số hàm hợp quy là $k - 1$. Biến hợp quy thứ nhất là sự kết hợp tuyến tính của các biến phân loại để mô tả sự khác nhau cực đại giữa các số trung bình của k nhóm theo phương thứ nhất. Biến hợp quy thứ 2 là sự kết hợp tuyến tính của các biến phân loại để mô tả sự khác nhau cực đại giữa các số trung bình của k nhóm theo phương thứ 2. Biến hợp quy thứ 3 là sự kết hợp tuyến tính của các biến phân loại để mô tả sự khác nhau cực đại giữa các số trung bình của k nhóm theo phương thứ ba vuông góc với hai phương 1 và 2. Biến hợp quy thứ $k - 1$ là sự kết hợp tuyến tính của các biến phân loại để mô tả sự khác nhau cực đại giữa các số trung bình của k nhóm theo một phương vuông góc với các phương trên. Mỗi biến hợp quy là một yếu tố để phân biệt các trung bình giữa các nhóm. Biến hợp quy còn được gọi là hàm phân loại hợp quy (Canonical Discriminant Function). Các hàm phân loại hợp quy có ba loại: (1) Các hàm phân loại hợp quy chuẩn hóa (Standardized Canonical Discriminant Functions); (2) Các hàm phân loại hợp quy chưa chuẩn hóa (Unstandardized Canonical Discriminant Functions); (3) Các hàm phân loại hợp quy ở trung tâm nhóm (Canonical Discriminant Functions at Group Centroids).

Hàm phân loại hợp quy chuẩn hóa được sử dụng để so sánh vai trò hay ý nghĩa của từng biến phân loại trong hàm hợp quy. Vai trò cao thấp của các biến phân loại được xác định theo giá trị tuyệt đối của các hệ số trong hàm lập nhóm. Hàm phân loại

hợp quy chưa chuẩn hóa được sử dụng để tính điểm số phân loại cho từng trường hợp trong một nhóm. Điểm số phân loại trung bình của các trường hợp trong mỗi nhóm được báo cáo qua hàm ở trung tâm nhóm. Các hàm ở trung tâm nhóm dùng để so sánh sự khác nhau giữa các nhóm. Các hàm hợp quy 1 và 2 được sử dụng để xây dựng đồ thị và bản đồ mô tả phân bố của các trường hợp trong các nhóm. Kết quả phân tích nhóm cung cấp các hàm để phân loại các đối tượng mới vào các nhóm khác nhau. Các đối tượng mới là các đối tượng nằm ngoài tập dữ liệu để xây dựng các hàm lập nhóm.

Sự thành công của việc phân loại các nhóm phụ thuộc vào tổ hợp các biến được chọn để xây dựng hàm phân loại. Các hàm phân loại với nhiều biến dự đoán cho kết quả phân loại cao hơn so với các hàm với ít biến dự đoán. Nguyên nhân là vì các đặc tính của một nhóm được phản ánh thông qua nhiều biến khác nhau. Tuy vậy, hàm phân loại với nhiều biến dự đoán không chỉ làm tăng tính phức tạp trong đo đạc các biến, mà còn gây ra khó khăn cho việc ứng dụng vào thực tế. Vì thế, xác định những biến dự đoán hữu ích để xây dựng hàm phân loại là một việc làm cần thiết. Phương pháp chọn biến từng bước (Stepwise Method) cho phép chọn những biến dự đoán hữu ích nhất từ một tập hợp biến dự tuyển. Các biến dự tuyển phản ánh các đặc trưng của tổng thể. Các biến dự tuyển là tập hợp các biến được chọn để xây dựng hàm lập nhóm. Theo phương pháp chọn biến từng bước, trước hết xác định một biến tốt nhất từ một tập hợp biến dự tuyển, nghĩa là biến có trung bình khác nhau nhiều nhất giữa các nhóm. Tiếp theo chọn những biến tốt khác bằng cách lần lượt đưa từng biến dự tuyển vào phân tích. Các biến dự tuyển được đưa vào hay loại ra khỏi mô hình ở từng bước được kiểm định theo tiêu chuẩn Wilk's Lambda, tiêu chuẩn khoảng cách bình phương của Mahalanobis (D^2), tiêu chuẩn tỷ lệ F nhỏ nhất (Smallest F Ratio) và tiêu chuẩn Rao's V; trong đó hai tiêu chuẩn Wilk's Lambda và Mahalanobis được sử dụng nhiều nhất.

Theo tiêu chuẩn Wilk's Lambda, những biến được sử dụng để xây dựng hàm lập nhóm là những biến có giá trị F_{Ra} (F to Remove) $> 2,71$ (Mặc định). Trái lại, khi những biến có giá trị $F_{Ra} < 2,71$, thì chúng bị loại ra khỏi mô hình phân loại. Một biến có thể được đưa vào mô hình phân loại khi nó làm thay đổi thống kê Wilk's Lambda một cách có ý nghĩa. Sự thay đổi của thống kê Wilk's Lambda được kiểm định theo tiêu chuẩn F thay đổi (F_{Change}) (Công thức 8.2). Ở công thức 8.2, N = tổng số các trường hợp đưa vào phân loại, g = số nhóm phân loại, p = số biến trong mô hình phân loại, λ_p = Wilk's Lambda ứng với p biến trước khi đưa thêm một biến vào hàm phân loại, λ_{p+1} = Wilk's Lambda ứng với $p + 1$ biến sau khi đưa thêm một biến vào hàm phân loại.

Phương pháp khoảng cách bình phương của Mahalanobis dựa trên cơ sở của hàm lập nhóm tuyến tính (Công thức 8.3); trong đó m là số biến để xây dựng hàm lập nhóm; k là số nhóm phân loại ($k = 1, 2, 3, \dots$); b_i là các tham số; $X_m^{(k)}$ là biến dự đoán thứ m thuộc nhóm phân loại k .

$$F_{Change} = \frac{N-g-p}{g-1} \frac{1 - \lambda_{p+1}/\lambda_p}{\lambda_{p+1}/\lambda_p} \quad (8.2)$$

$$Z^{(k)} = b_1 X_1^{(k)} + b_2 X_2^{(k)} + \dots + b_m X_m^{(k)} \quad (8.3)$$

Hàm lập nhóm tuyến tính chỉ được xác định khi các nhóm khác nhau rõ rệt ở mức ý nghĩa alpha cho trước, thông thường $\alpha = 0,10; 0,05; 0,01$. Vì thế, giả thuyết H_0 được đặt ra là các nhóm khác nhau không rõ rệt. Để kiểm định giả thuyết này, Mahalanobis sử dụng số đo khoảng cách D^2 . Sự khác nhau giữa các nhóm được kiểm định theo tiêu chuẩn F (Công thức 8.4); trong đó n_1 và n_2 là dung lượng quan sát của nhóm 1 và nhóm 2, m là số biến để xây dựng hàm lập nhóm, D là khoảng cách giữa các nhóm. Khi $F > F_{05}$ hay $P < 0,05$ (Mặc định), thì hai nhóm khác nhau rõ rệt. Khi hai nhóm khác nhau rõ rệt, thì bước tiếp theo là xây dựng hàm lập nhóm tuyến tính Fisher và biểu đồ để mô tả phân bố của các trường hợp trong các nhóm. Sau đó hàm lập nhóm tuyến tính Fisher được sử dụng để phân loại các trường hợp mới vào các nhóm khác nhau.

$$F = \frac{n_1 n_2}{(n_1 + n_2)} \frac{(n_1 + n_2 - m - 1)}{(n_1 + n_2 - 2)} D^2 \quad (8.4)$$

8.1.2. Mục đích và mục tiêu của phân tích nhóm

Mục đích của phân tích nhóm là cung cấp các hàm để dự đoán các đối tượng hay trường hợp mới thuộc về các nhóm khác nhau. Các đối tượng mới là các đối tượng nằm ngoài tập dữ liệu để xây dựng các hàm lập nhóm.

Phân tích nhóm có ba mục tiêu cơ bản: (1) Xác định những biến độc lập (biến dự đoán, biến phân loại) khác nhau nhiều nhất giữa các nhóm; (2) Xác định sự khác biệt có ý nghĩa thống kê giữa hai hay nhiều nhóm; (3) Xây dựng các hàm lập nhóm để phân loại những trường hợp mới vào các nhóm khác nhau.

8.1.3. Những ứng dụng của phân tích nhóm trong lâm học

Phân tích nhóm có thể được sử dụng để giải quyết những vấn đề sau đây: (1) Phân loại số liệu thành hai hay nhiều nhóm; (2) Tìm kiếm một hàm số từ các biến số dự đoán (định lượng hoặc định tính) để phân chia các nhóm. Trong lâm học, phân tích nhóm được sử dụng để giải quyết một số vấn đề dưới đây.

(1) Phân loại (rừng, cấp lập địa, cấp nguy cơ cháy rừng, cấp năng suất rừng...). Để giải quyết những vấn đề đặt ra, trước hết chúng ta cần phân loại sơ bộ các đối tượng vào các nhóm dựa theo một số đặc trưng nào đó. Tiếp theo xây dựng các hàm để phân loại các đối tượng vào các nhóm dựa trên những biến phân loại của các đối tượng đã biết. Sau đó sử dụng các hàm phân loại này để phân chia các đối tượng mới thuộc về các nhóm khác nhau.

(2) Kiểm định sự khác nhau giữa các nghiệm thức dựa trên nhiều biến số khác nhau. Chẳng hạn: (a) Phân tích ảnh hưởng của 4 phương thức xử lý lập địa đến sinh trưởng của rừng trồng dựa trên các biến sản lượng (D, H, G, V); (b) Xác định những biện pháp chặt nuôi rừng mang lại hiệu quả lâm sinh cao nhất dựa trên các biến sản lượng (D, H, G, V).

(3) Xác định những tập hợp biến hữu ích nhất để phân loại hoặc so sánh giữa các đối tượng. Những đối tượng cần phân loại thường có rất nhiều đặc trưng khác nhau. Vì thế, tùy theo mục tiêu đặt ra, nhà lâm học cần phải chọn ra những biến có ý nghĩa nhất để phân chia các nhóm.

(4) So sánh vai trò của những tập hợp biến dự đoán khác nhau được sử dụng để xây dựng hàm lập nhóm. Để giải quyết vấn đề đặt ra, nhà lâm học chọn những tập hợp biến phân loại khác nhau để xây dựng các hàm lập nhóm. Một tập hợp biến phân loại tốt hơn khi hàm phân loại được xây dựng từ những biến này cho kết quả phân loại cao hơn. Nói chung, không có phương pháp nào cho phép chọn được một tập hợp biến tốt nhất, mà chỉ có phương pháp cho phép chọn được một tập hợp biến tốt hơn. Vì thế, khi phân loại các đối tượng, nhà lâm học cần phải tìm kiếm những tập hợp biến cho kết quả phân loại cao hơn. Một hàm phân loại bao gồm nhiều biến dự đoán cho kết quả phân loại cao hơn so với hàm phân loại chỉ bao gồm ít biến. Tuy vậy, việc xây dựng những hàm nhiều biến không chỉ phức tạp, mà còn cần đến nhiều nhân lực, thời gian, kinh phí và khả năng ứng dụng. Chính vì thế, tùy theo yêu cầu đặt ra, nhà lâm học cần xác định một tập hợp biến hữu ích cho việc phân loại các nhóm.

(5) Tìm kiếm những giá trị bất thường hay ngoại lai. Những giá trị bất thường làm tăng biến động trong một nhóm. Những giá trị này sẽ nhận các điểm số bất thường.

8.1.4. Những điều kiện trong phân tích nhóm

Phân tích nhóm đòi hỏi phải thỏa mãn 6 điều kiện cơ bản. Một là số nhóm phân loại từ 2 trở lên. Hai là mỗi nhóm phải có hai biến phân loại trở lên. Các biến phân loại này phải nhận các giá trị liên tục. Ba là các biến phân loại là các biến định lượng và khác nhau thực sự giữa các nhóm. Nếu các biến phân loại là các biến định danh, thì chúng phải được mã hóa thành các biến giả (Dummy variables) hoặc biến nhị phân (0 và 1). Bốn là các biến phân loại của mỗi nhóm phải có phân bố chuẩn. Điều kiện này được kiểm định theo thống kê χ^2 . Năm là phương sai của các biến phân loại phải bằng nhau giữa các nhóm. Điều kiện này được kiểm định theo thống kê Levene. Sáu là ma trận hiệp phương sai trong nội bộ nhóm phải bằng nhau giữa các nhóm. Điều kiện này được kiểm định bằng tiêu chuẩn M của Box (Box's M Test).

8.1.5. Các bước cơ bản trong phân tích nhóm

Bước 1. Phân tích đối tượng nghiên cứu. Trong lâm học, đối đối tượng nghiên cứu có thể là rừng phân bố trên những môi trường khác nhau; rừng được xử lý theo những kỹ thuật lâm sinh khác nhau; các nghiệm thức trong gieo ươm và trồng rừng... Mục đích của bước này là chọn các biến dự đoán thích hợp (biến phân loại, biến độc lập); kiểm định sự phù hợp của các hàm lập nhóm; xác định phương pháp xây dựng các hàm lập nhóm và đánh giá ý nghĩa hay vai trò của các biến dự đoán trong các hàm lập nhóm.

Kết quả phân loại các đối tượng phụ thuộc vào số lượng biến dự đoán được sử dụng để xây dựng hàm phân loại nhóm. Khi các biến dự đoán khác nhau rõ rệt giữa các nhóm, thì hàm lập nhóm cho kết quả phân loại với độ chính xác cao. Trái lại, nếu các biến dự đoán kém hữu ích cho việc phân loại, thì hàm lập nhóm cho kết quả phân loại với độ chính xác thấp. Kết quả phân loại được tính theo phần trăm. Vì thế, xác định những biến hữu ích cho việc phân loại là việc làm cần thiết. Sự khác nhau của các biến số dự đoán giữa các nhóm được kiểm định bằng tiêu chuẩn Wilk's Lambda (Công thức 8.5). Tiêu chuẩn Wilk's Lambda là tỷ lệ giữa tổng các độ lệch bình phương trong các nhóm (SSE_{Group}) với tổng các độ lệch bình phương của tất cả các nhóm gộp lại

($SSE_{\text{Tổng}}$). Tiêu chuẩn Wilks's Lambda = $0 \div 1$. Giá trị Wilks's Lambda = 0 cho biết các biến dự đoán có ý nghĩa phân loại tốt nhất. Khi Wilks's Lambda $\ll 1$ hoặc $P < 0,01$, thì trung bình của các biến dự đoán khác nhau rõ rệt giữa các nhóm hay các nhóm khác nhau rõ rệt. Giá trị Wilks's Lambda = 1 cho biết các biến dự đoán không có ý nghĩa phân loại.

$$\text{Wilk's Lambda} = \frac{SSE_{\text{Trong nhóm}}}{SSE_{\text{Tổng}}} \quad (8.5)$$

Hàm lập nhóm có thể được xây dựng bằng nhiều phương pháp khác nhau. Một là hàm lập nhóm bao gồm tất cả các biến dự đoán. Hai là hàm lập nhóm chỉ bao gồm những biến dự đoán hữu ích cho việc phân loại. Những biến hữu ích được xác định theo phương pháp hàm lập nhóm từng bước (Stepwise Discriminant Function Method). Hàm này có thể được xây dựng theo 4 phương pháp cơ bản: Wilk's Lambda, khoảng cách bình phương của Mahalanobis (Mahalanobis Squared Distance), tỷ lệ F nhỏ nhất (Smallest F ratio) và Rao's V.

Bước 2. Xác định mục tiêu phân loại nhóm và độ chính xác của các hàm phân loại nhóm. Khi phân tích nhóm, trước hết cần định nghĩa rõ ràng mục tiêu hay lý do của việc xây dựng các hàm lập nhóm. Phương pháp xây dựng các hàm lập nhóm thay đổi tùy theo mục tiêu nghiên cứu.

Nếu mục tiêu nghiên cứu là sắp xếp các trường hợp vào những nhóm khác nhau, thì hàm lập nhóm có thể được xây dựng từ một tập hợp biến nào đó. Tập hợp biến này được chọn ra từ những biến dự tuyển. Lưu ý rằng không có phương pháp nào cho phép chọn được một tập hợp biến tốt nhất, mà chỉ có những phương pháp cho phép chọn được một tập hợp biến tốt hơn. Vì thế, để xây dựng hàm phân loại các đối tượng, nhà lâm học cần xác định những tập hợp biến cho kết quả phân loại cao hơn. Để đạt mục tiêu này, các biến trong hàm phân loại được xác định từ những tập hợp biến dự tuyển khác nhau.

Nếu mục tiêu nghiên cứu là so sánh sự khác nhau giữa các nhóm đã biết trước, thì hàm lập nhóm được xây dựng từ những nhóm đã biết này. Những nhóm đã biết trước là những nhóm đã được định nghĩa rõ ràng. Chẳng hạn: (a) Các nghiệm thức thí nghiệm trong gieo ươm; (b) Rừng trồng từ một loài cây gỗ nào đó theo ba phương thức khác nhau; (c) Rừng trồng trên các đất khác nhau... Trong những trường hợp này, chúng ta chỉ chọn những tập hợp biến hữu ích cho việc so sánh các nhóm. Nếu các nhóm khác nhau không rõ rệt hoặc có nhiều nhóm gần giống nhau, thì kết quả phân loại sẽ thấp. Ngược lại, khi các nhóm khác nhau rõ rệt, thì kết quả phân loại sẽ cao. Khi hai nhóm giao nhau một phần, thì tùy theo phần giao nhau mà chúng khác nhau nhiều hay ít. Nếu phần giao nhau càng lớn, thì hai nhóm càng giống nhau.

Nếu mục tiêu nghiên cứu là xây dựng các hàm lập nhóm để xác định các trường hợp mới thuộc về những nhóm đã biết trước, thì các bước xây dựng hàm lập nhóm được bắt đầu từ những biến của các nhóm đã biết này. Chẳng hạn: (a) Phân cấp sinh trưởng cho những cây hình thành rừng trồng thuần loài đồng tuổi; (b) Phân chia các ô mẫu theo những kiểu QXTV đã biết. Trong trường hợp này, trước hết xây dựng các hàm lập nhóm tuyến tính Fisher. Sau đó sử dụng các hàm lập nhóm này để phân loại những trường hợp mới thuộc về những nhóm khác nhau. Một trường hợp mới thuộc về

thành viên của một nhóm nào đó nếu hàm lập nhóm tại nhóm này nhận được giá trị lớn nhất. Sự thành công của việc phân loại ban đầu được báo cáo thông qua các thống kê cho từng trường hợp (Casewise Statistics), bản đồ phân loại (Territorial Map), đồ thị phân tán giữa các nhóm (All-groups Scatter Plot) và kết quả phân loại (Classification Results). Kết quả phân loại được báo cáo theo số lượng trường hợp và phần trăm trường hợp được phân loại đúng.

Bước 3. Phân tích các kết quả phân loại nhóm. Ở bước này xây dựng hai kiểu hàm là hàm lập nhóm hợp quy chuẩn hóa và hàm lập nhóm hợp quy chưa chuẩn hóa. Nếu có hai nhóm, thì chỉ có một hàm hợp quy. Nếu có k nhóm, thì số hàm hợp quy là $k - 1$. Hàm lập nhóm hợp quy chuẩn hóa được sử dụng để đánh giá tầm quan trọng hay ý nghĩa của từng biến phân loại trong hàm lập nhóm. Mức ý nghĩa của các biến phân loại được đánh giá theo độ lớn tuyệt đối của các hệ số trong hàm lập nhóm. Hàm lập nhóm hợp quy chưa chuẩn hóa được sử dụng để xác định điểm số hay khoảng cách của mỗi trường hợp đến trung tâm nhóm. Điểm trung bình của mỗi nhóm được báo cáo thông qua các hàm ở trung tâm nhóm (Functions at Group Centroids). Khi có nhiều hàm hợp quy ($k > 2$), thì hai hàm hợp quy 1 và 2 được sử dụng để xây dựng các bản đồ và đồ thị phân chia các nhóm.

Sự khác nhau giữa các nhóm được đánh giá bằng số đặc trưng (Eigenvalues) và hệ số tương quan. Thống kê Eigenvalues là tỷ số của tổng bình phương giữa các nhóm với tổng bình phương trong các nhóm. Tỷ lệ đóng góp của một hàm hợp quy theo một phương nào đó được đo bằng phần trăm phương sai. Phần trăm phương sai cộng dồn của các hàm hợp quy cho biết tỷ lệ đóng góp của chúng theo một biến dự đoán nào đó. Mỗi biến dự đoán là một phương.

Mối quan hệ giữa các điểm số lập nhóm được đo bằng hệ số tương quan hợp quy. Nếu chỉ có hai nhóm phân loại, thì hệ số tương quan hợp quy là hệ số tương quan Pearson. Khi có nhiều hơn hai nhóm phân loại, thì hệ số tương quan hợp quy là hệ số Eta hay tỷ tương quan. Đó là tỷ số của tổng bình phương giữa các nhóm với tổng bình phương toàn bộ. Nếu sử dụng phương pháp đưa tất cả các biến dự đoán vào hàm phân loại, thì sự khác nhau giữa các nhóm được xác định bằng tiêu chuẩn Wilk's Lambda và tiêu chuẩn χ^2 . Khi $P < 0,05$ hoặc $0,01$, thì trọng tâm giữa các nhóm là khác nhau. Nếu sử dụng phương pháp chọn biến từng bước (Stepwise Method), thì những biến được chọn hay loại bỏ khỏi mô hình cũng được xác định bằng tiêu chuẩn Wilk's Lambda.

Khi sử dụng phương pháp chọn biến từng bước, thì sự khác nhau giữa các cặp nhóm được kiểm định bằng tiêu chuẩn F. Tiêu chuẩn F cho từng cặp nhóm tỷ lệ với khoảng cách Mahalanobis bình phương (D^2). Vì thế, độ lớn và mức ý nghĩa (P – value) của tiêu chuẩn F cho biết sự khác nhau giữa các nhóm. Thống kê F chỉ thích hợp với một cặp nhóm so sánh. Khi có nhiều cặp nhóm cần so sánh, thì chúng ta phải sử dụng phương pháp so sánh đa kiểm định (Multiple Comparisons). Trong trường hợp này, sau khi mức ý nghĩa của các cặp so sánh được báo cáo, thì chúng phải được điều chỉnh bằng cách nhân mức ý nghĩa của mỗi cặp so sánh với số cặp so sánh. Chẳng hạn: Khi so sánh 5 nhóm, thì số cặp so sánh là 10. Giả sử chọn $P = 0,01$ đối với mỗi cặp so sánh, thì xác suất của một cặp so sánh được điều chỉnh là $10 \times 0,01 = 0,10$. Vì $P = 0,10 > 0,05$, nên cặp nhóm này khác nhau không rõ rệt.

8.2. Phân tích mô hình hai nhóm

Giả sử mục tiêu nghiên cứu là xây dựng hàm lập nhóm để so sánh sự khác nhau giữa hai quần thụ Đước 16 tuổi. Hàm lập nhóm bao gồm 3 biến phân loại (D, H và D_T) của những cá thể đại diện cho hai quần thụ này. Trong trường hợp này, hàm lập nhóm có dạng như Hàm 8.6; trong đó k = 2 nhóm.

$$Z^{(k)} = aD^{(k)} + bH^{(k)} + cD_T^{(k)} \quad (8.6)$$

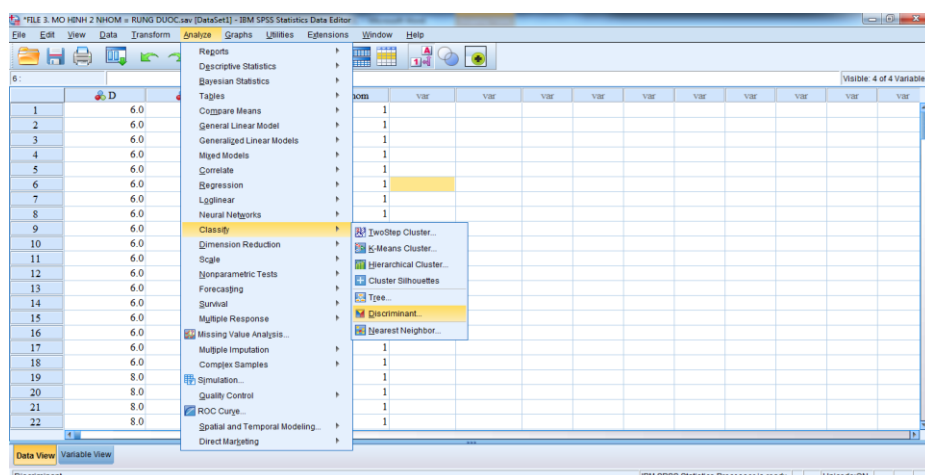
Để đạt được mục tiêu này, hàm lập nhóm được xây dựng thông qua những bước dưới đây. Ở đây sử dụng công cụ tính toán là phần mềm thống kê SPSS.

Bước 1. Nhập các giá trị của 3 biến dự đoán (D, H, D_T) vào 3 cột liên tiếp của bảng tính SPSS. Cột thứ 4 nhập số nhóm phân loại. Trong trường hợp này, số nhóm phân loại là 2 quần thụ Đước 16 tuổi (Hình 8.1).

	D	H	DT	Nhóm	V1	V2	V3	V4	V5	V6	V7	V8
1	6.0	6.0	1.5	1								
2	6.0	6.0	1.7	1								
3	6.0	6.0	2.1	1								
4	6.0	6.0	2.4	1								
5	6.0	6.0	2.7	1								
6	6.0	6.0	1.9	1								
7	6.0	6.0	2.2	1								
8	6.0	6.0	2.5	1								
9	6.0	6.0	2.7	1								
10	6.0	6.0	3.0	1								
11	6.0	6.0	3.2	1								
12	6.0	6.0	3.4	1								
13	6.0	6.0	3.6	1								
14	6.0	6.0	3.8	1								
15	6.0	6.0	4.0	1								
16	6.0	6.0	4.1	1								
17	6.0	6.0	4.3	1								
18	6.0	6.0	4.4	1								
19	8.0	6.0	4.5	1								
20	8.0	6.0	4.7	1								
21	8.0	6.0	4.8	1								
22	8.0	6.0	4.9	1								

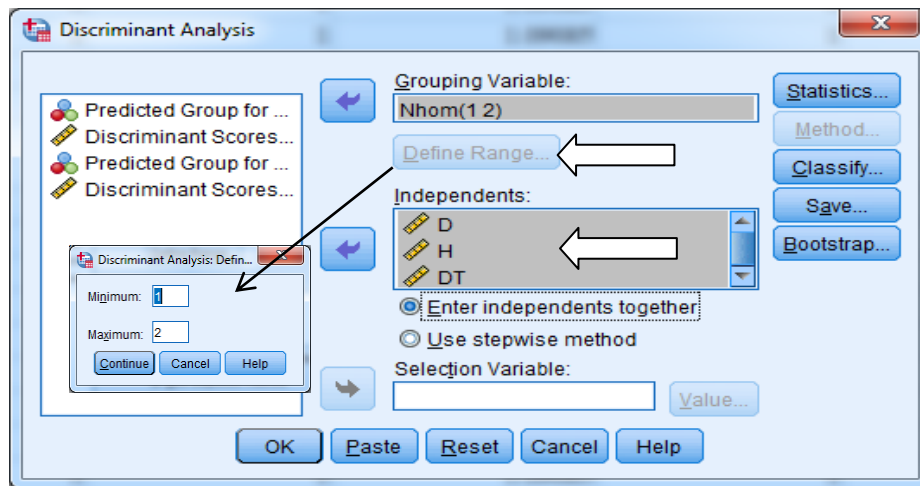
Hình 8.1. Đặc trưng của hai quần thụ 1 và 2.

Bước 2. Khai báo số nhóm và số lượng biến dự đoán của hàm lập nhóm tuyến tính Fisher theo Hàm 8.6. Trước hết mở tập số liệu của hai quần thụ Đước 16 tuổi. Kế đến, từ Menu chính, chọn Analyze ⇒ Classify ⇒ Discriminant (Hình 8.2).



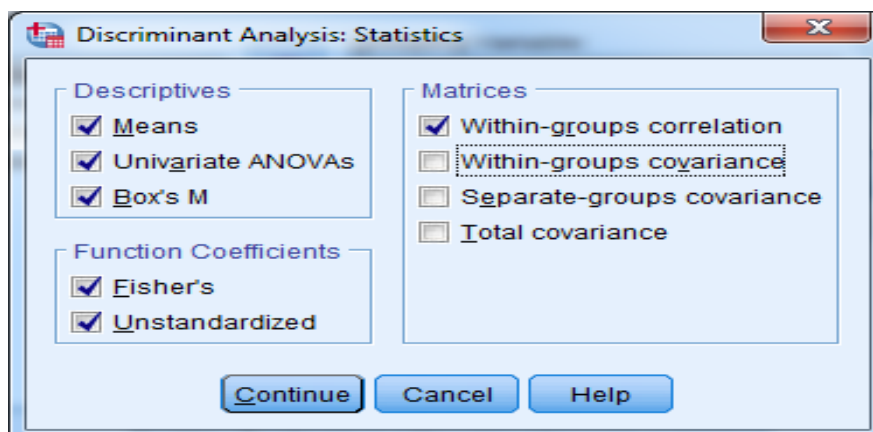
Hình 8.2. Đường dẫn vào Menu phân tích hàm lập nhóm.

Tại cửa sổ Discriminant Analysis, chọn và chuyển số nhóm phân loại vào ô Grouping Variable. Kế đến chọn Define Range để định nghĩa phạm vi của các nhóm phân loại. Lưu ý số nhóm phân loại phải là số nguyên dương liên tục. Ở trường hợp này chỉ có hai quần thể Đước 16 tuổi đưa vào phân tích. Để định nghĩa phạm vi của các nhóm, sau khi mở Define Range, nhập số 1 (Quần thể 1) vào ô Minimum và số 2 (Quần thể 2) vào ô Maximum. Tiếp theo chọn 3 biến dự đoán (D, H, D_T) và chuyển chúng vào ô Independents (Hình 8.3).



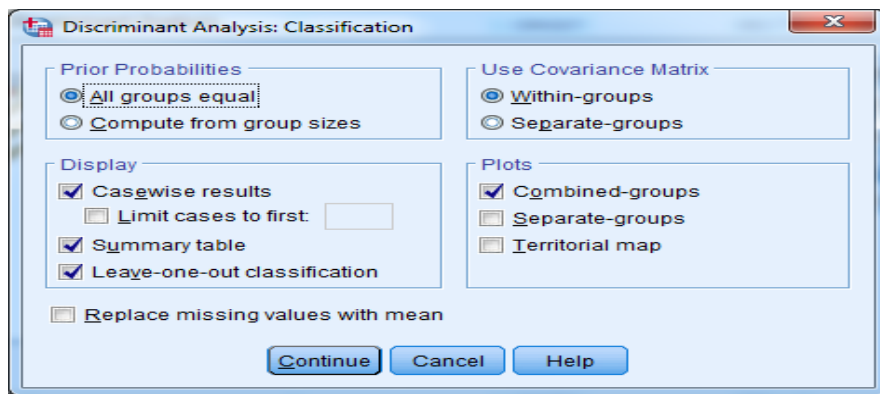
Hình 8.3. Phân tích hàm lập nhóm tuyến tính Fisher.

Bước 3. Xây dựng hàm lập nhóm tuyến tính Fisher theo Hàm 8.6. Từ cửa sổ Discriminant Analysis ở Hình 8.3, chọn Statistics và đánh dấu (Hình 8.4): (a) Các thống kê mô tả (Means; Univariate ANOVAs; Box's M = Kiểm định hiệp phương sai trong nội bộ nhóm); (b) Các hệ số của hàm phân loại (Hợp quy chuẩn hóa = Standardized Canonical Discriminant Function Coefficients; Hợp quy chưa chuẩn hóa = Unstandards Canonical Discriminant Function Coefficients và ma trận tương quan trong nhóm = Within-Group Correlation Matrices). Sau đó chọn Continue.



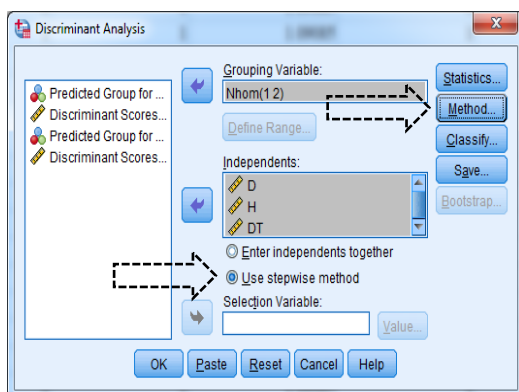
Hình 8.4. Những thống kê trong phân tích hàm lập nhóm.

Bước 4. Xác định kết quả phân tích nhóm. Từ cửa sổ Discriminant Analysis ở Hình 8.3, chọn Classify và đánh dấu All Group Equal trong ô Prior Probabilities, Casewise Results và Summary Table trong ô Display, Within Groups trong ô Use Covariance Matrix (Hình 8.5), chọn Combined Groups trong ô Plots. Mục All Group Equal trong ô Prior Probabilities là xác suất ưu tiên cho mỗi nhóm. Nếu có k nhóm, thì xác suất ưu tiên (%) của mỗi nhóm là $(100/k)$. Ở ví dụ này, bởi vì chỉ có hai nhóm, nên xác suất ưu tiên của mỗi nhóm là $(100/2) = 50\%$. Mục Casewise Results là kết quả phân tích nhóm; trong đó bao gồm số trường hợp phân tích, nhóm phân loại ban đầu (Actual Group), nhóm dự đoán (Predicted Group) và điểm số phân loại cho từng trường hợp (Discriminant Scores). Sau cùng chọn OK để nhận kết quả phân tích nhóm.

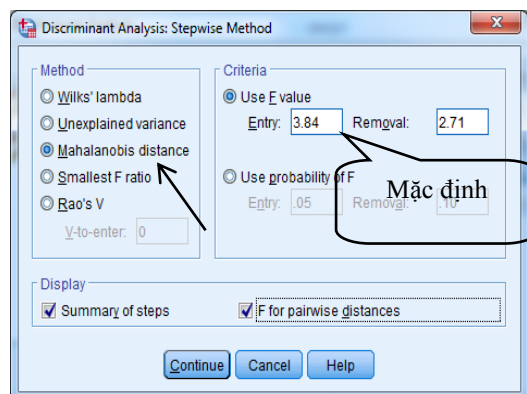


Hình 8.5. Xác định những thông tin của hàm lập nhóm.

Bốn bước trên đây được thực hiện khi hàm lập nhóm bao gồm tất cả 3 biến dự đoán (D, H và D_T). Nếu chỉ xây dựng hàm lập nhóm với những biến dự đoán hữu ích nhất, thì từ Hình 8.3 chọn phương pháp phân tích nhóm từng bước (Use Stepwise Method) theo phương pháp khoảng cách bình phương của Mahalanobis (Hình 8.6 và 8.7).



Hình 8.6. Phương pháp chọn biến từng bước.



Hình 8.7. Chọn phương pháp phân tích hàm lập nhóm.

Bước 5. Giải thích kết quả. Kết quả phân tích nhóm cung cấp nhiều nội dung khác nhau. Dưới đây chỉ giới thiệu tóm tắt những nội dung cơ bản.

Nội dung 1. Thống kê mô tả các nhóm. Phần này báo cáo: (a) Các đặc trưng thống kê mô tả (trung bình, sai tiêu chuẩn, số mẫu của các biến đưa vào phân tích) đối với từng biến dự đoán trong mỗi nhóm (Bảng 8.1).

Bảng 8.1. Các thống kê mô tả trong hai nhóm (Group Statistics)

Biến dự đoán của 2 nhóm		Mean	Std. Deviation	Valid N (listwise)	
				Unweighted	Weighted
1	D	8,3	2,1	110	110
	H	8,0	2,0	110	110
	DT	5,0	1,7	110	110
2	D	14,2	2,2	102	102
	H	14,2	2,1	102	102
	DT	3,8	1,3	102	102
Total	D	11,2	3,6	212	212
	H	11,0	3,7	212	212
	DT	4,5	1,6	212	212

Nội dung 2. Kiểm định ngang bằng trung bình của các biến dự đoán giữa các nhóm (Bảng 8.2). Giả thuyết H_0 : Trung bình của các biến dự đoán trong các nhóm là không khác nhau. Giả thuyết này được kiểm định bằng tiêu chuẩn Wilks' Lambda và tiêu chuẩn F. Tiêu chuẩn Wilks' Lambda nhận giá trị từ 0÷1. Đối với mỗi biến phân loại, nếu giá trị Wilks' Lambda rất nhỏ hoặc $P < 0,01$, thì giả thuyết bị bác bỏ. Kết quả ở Bảng 8.2 cho thấy giá trị Sig hay $P < 0,01$ chứng tỏ ba biến (D, H và D_T) khác nhau rất rõ rệt giữa hai quần thể. Lưu ý rằng một biến số độc lập chỉ có ý nghĩa phân loại khi nó khác nhau thực sự giữa các nhóm.

Bảng 8.2. Kiểm định ngang bằng trung bình của các biến dự đoán giữa hai nhóm

Biến dự đoán	Wilks' Lambda	F	df1	df2	Sig.
D	0,340	408,109	1	210	0,000
H	0,303	482,636	1	210	0,000
DT	0,860	34,221	1	210	0,000

Nội dung 3. Phân tích ý nghĩa của các biến dự tuyển hay biến ban đầu được đưa vào hàm phân loại theo phương pháp lập nhóm từng bước (Stepwise Method). Nội dung này phân tích và xác định những biến phân loại có ý nghĩa thực sự hay hữu ích để xây dựng hàm lập nhóm. Những biến phân loại có ý nghĩa được báo cáo ở mục phân tích những biến ở trong hàm lập nhóm (Variables Entered) và những biến bị loại ra khỏi hàm lập nhóm (Variables Removed). Ở ví dụ này, những biến được đưa vào hay loại ra khỏi hàm lập nhóm được kiểm định bằng phương pháp khoảng cách bình phương của Mahalanobis. Nếu sử dụng J biến dự tuyển, thì số bước kiểm định là J. Ở ví dụ này, hàm lập nhóm được xây dựng từ 3 biến dự tuyển (D, H, D_T), nên ý nghĩa của 3 biến này được kiểm định theo 3 bước. Giả thuyết H_0 : Biến độc lập thứ nhất (Biến đưa vào phân tích đầu tiên) không có ý nghĩa phân loại. Nếu $P < 0,01$, thì giả thuyết bị bác bỏ. Những bước tiếp theo kiểm định ý nghĩa của các biến khác. Sau đó chương trình báo cáo kết quả kiểm định những biến dự tuyển ở trong hàm lập nhóm và những biến dự tuyển bị loại ra khỏi hàm lập nhóm. Ý nghĩa của các biến dự tuyển

cũng được kiểm định bằng tiêu chuẩn Wilks' Lambda. Kết quả cho thấy ba biến (D, H và D_T) đều có ý nghĩa phân loại ($P < 0,01$) (Bảng 8.3).

Bảng 8.3. Kiểm định ý nghĩa của các biến số dự tuyển

Step	Number of Variables	Lambda	df1	df2	df3	Exact F			
						Statistic	df1	df2	Sig.
1	1	0,303	1	1	210	482,6	1	210	0,000
2	2	0,086	2	1	210	1103,8	2	209	0,000
3	3	0,084	3	1	210	753,0	3	208	0,000

Nội dung 4. Phân tích hàm lập nhóm hợp quy (Canominal Discriminant Function Analysis). Phần này cung cấp các thông tin về kiểm định sự tồn tại của các hàm hợp quy; xác định các hệ số của hàm lập nhóm hợp quy chuẩn hóa và chưa chuẩn hóa; đo đặc mức ảnh hưởng hay vai trò của từng biến trong hàm lập nhóm. Những kiểm định này được xác định theo số đặc trưng (Eigenvalue) và tiêu chuẩn Wilks' Lambda (Bảng 8.4).

Thống kê Eigenvalue đo biến động về điểm số giữa các nhóm và quan hệ giữa các điểm số lập nhóm (Discriminant Scores). Nếu có hai nhóm, thì tương quan hợp quy (Canonical Correlation) là hệ số tương quan Pearson HAD giữa các điểm số lập nhóm của hai nhóm được mã hóa là 0 và 1. Nếu có ba nhóm trở lên, thì tương quan hợp quy là thống kê Eta (Tỷ tương quan η) trong phân tích phương sai 1 yếu tố. Ở ví dụ này, tương quan hợp quy giữa các điểm số lập nhóm của hai nhóm là rất chặt chẽ ($R = 0,957$).

Bảng 8.4. Tóm tắt hàm lập nhóm hợp quy

(a) Thống kê Eigenvalue				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	10,862 ^a	100,0	100,0	0,957

(b) Tiêu chuẩn Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	0,084	515,690	3	0,000

Tiêu chuẩn Wilks' Lambda là tỷ lệ của tổng biến thiên trong các điểm số lập nhóm không thể giải thích được bởi sự khác nhau giữa các nhóm. Ở ví dụ này, Wilks' Lambda = 0,084 cho biết khoảng 8,4% sự biến thiên của điểm số lập nhóm không thể giải thích được bằng sự khác nhau giữa các nhóm. Hai tiêu chuẩn Wilks' Lambda và Chi square đều kiểm định giả thuyết H₀: Điểm số trung bình hay trọng tâm của các nhóm là bằng nhau. Nếu P value < 0,01, thì điểm số trung bình của các nhóm là khác nhau. Ở ví dụ này, giá trị P < 0,01 cho biết điểm số trung bình của hai nhóm khác nhau rõ rệt. Khi áp dụng vào lâm học, chúng ta kết luận hai quần thể 1 và 2 khác nhau rất rõ rệt.

Các hệ số của hàm lập nhóm hợp quy chuẩn hóa (Cột 1 của Bảng 8.5) được sử dụng để đánh giá vai trò của từng biến phân loại trong hàm lập nhóm. Vai trò của từng biến phân loại được đánh giá theo độ lớn tuyệt đối của các hệ số hồi quy chuẩn hóa. Ở ví dụ này, thứ bậc vai trò từ cao đến thấp là D_T, H và D. Vai trò của từng biến phân loại trong hàm lập nhóm hợp quy chuẩn hóa còn được đánh giá theo hệ số tương quan

(Bảng 8.6). Theo hệ số tương quan, biến H có quan hệ với các biến của hàm lập nhóm hợp quy chuẩn hóa lớn hơn so với biến D và D_T.

Bảng 8.5. Các hệ số của hàm lập nhóm hợp quy chuẩn hóa và chưa chuẩn hóa

Biến dự đoán	Standardized Function	Unstandardized Function
	(1)	(2)
D	0,526	0,247
H	1,267	0,617
D _T	-1,591	-1,053
(Constant)		-4,847

Hàm lập nhóm hợp quy chưa chuẩn hóa (Cột 2 của Bảng 8.5) được sử dụng để xác định điểm số của biến hợp quy (Canonical Variable Scores) cho từng trường hợp. Điểm số này không phụ thuộc vào dấu. Khi có hai nhóm, thì những điểm số âm thuộc về nhóm 1 hay hàm 1, còn điểm số dương thuộc về nhóm 2 hay hàm 2. Điểm số bình quân của mỗi nhóm được xác định bằng những hàm lập nhóm hợp quy chưa chuẩn hóa ở trung tâm nhóm (Functions at Group Centroids) (Bảng 8.7). Giá trị tuyệt đối của hàm lập nhóm hợp quy chưa chuẩn hóa ở trung tâm nhóm cho biết mức độ khác nhau giữa các nhóm. Điểm số bình quân của mỗi nhóm được xác định theo hàm ở trung tâm nhóm (Bảng 8.8).

Bảng 8.6. Các hệ số tương quan giữa các biến dự đoán với các biến của hàm lập nhóm hợp quy chuẩn hóa.

Biến dự đoán	Function
	1
H	0,460
D	0,423
D _T	-0,122

Bảng 8.7. Hàm lập nhóm hợp quy chưa chuẩn hóa ở trung tâm nhóm.

Cấp sinh trưởng	Function	
	1	2
1	1,155	0,622
2	0,891	0,114
3	0,025	-0,173
4	-0,602	0,022
5	-1,771	0,505

Bảng 8.8. Hàm ở trung tâm nhóm (Functions at Group Centroids)

Nhóm	Functions at Group Centroids of 1
1	-3,159
2	3,406

Nội dung 5. Phân tích các hàm phân loại. Phần này phân tích xác suất ưu tiên của các nhóm và các hệ số của hàm lập nhóm. Ở ví dụ này chỉ có 2 quần thụ, nên xác suất ưu tiên của mỗi nhóm là 50%. Mỗi nhóm có một hàm phân loại tuyến tính Fisher. Ở ví dụ này có hai nhóm nên chỉ có hai hàm phân loại nhóm (Bảng 8.9). Hàm phân loại đối với quần thụ Được (1) và (2) tương ứng có dạng hàm 8.7 và 8.8.

Bảng 8.9. Các hệ số của hàm phân loại (Classification Function Coefficients)

Biến dự đoán	Các hệ số của nhóm	
	1	2
D	0,978	2,602
H	0,876	4,929
DT	0,106	-6,810
(Constant)	-8,540	-41,172

$$Z^{(1)} = 0,978 \times D + 0,876 \times H + 0,106 \times D_T = 8,540 \quad (8.7)$$

$$Z^{(2)} = 2,602 \times D + 4,929 \times H - 6,810 \times D_T - 41,172 \quad (8.8)$$

Hai hàm phân loại này (Hàm 8.7 và 8.8) được sử dụng để tính khoảng cách khác nhau cực đại giữa hai quần thụ Đước. Hai hàm này cũng được sử dụng để phân loại những cây mới thuộc về quần thụ Đước khác nhau. Để biết một cây mới thuộc về quần thụ nào, trước hết đo đặc ba biến D, H và D_T của những cây cần phân loại trong mỗi quần thụ được nghiên cứu. Kế đến nhân các hệ số của mỗi hàm với các giá trị của các biến dự đoán (D, H và D_T). Kết quả nhận được khoảng cách khác nhau cực đại của mỗi quần thụ. Cây phân loại thuộc về quần thụ nhận giá trị khoảng cách lớn nhất. Bảng 8.10 dẫn ví dụ phân loại 4 cây Đước theo hàm 8.7 và 8.8. Cây số 1 và 2 nhận giá trị khoảng cách lớn nhất ở hàm 1 (13), nên nó thuộc quần thụ 1. Cây số 3 và 4 nhận giá trị khoảng cách lớn nhất ở hàm 2 (tương ứng 27,4 và 48), nên nó thuộc quần thụ 2.

Bảng 8.10. Xác định nhóm đối với 4 cây mới theo hai hàm lập nhóm ở Bảng 8.9

TT cây	Kích thước thân cây			Điểm số theo hàm:		Nhóm
	D (cm)	H (m)	D_T (m)	1	2	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	12	12	6,9	<u>13,0</u>	2,2	1
2	12	12	7	<u>13,0</u>	1,5	1
3	12	12	3,2	13,4	<u>27,4</u>	2
4	16	16	4,6	20,6	<u>48,0</u>	2

Nội dung 6. Phân tích những thống kê đối với các trường hợp (Casewise Statistics). Phần này so sánh các trường hợp được phân loại thực tế và dự đoán. Bảng báo cáo này bao gồm 2 phần (Hình 8.8). Phần 1 báo cáo các thông tin về kết quả phân loại lần đầu từ hàm lập nhóm (Highest Group). Phần 2 báo cáo các thống kê kiểm định lại các trường hợp phân loại lần đầu ở phần 1 có thể thuộc về nhóm hai hay không.

Phần 1 bao gồm 7 cột. Cột 1 là thứ tự các trường hợp được sử dụng để xây dựng hàm lập nhóm. Cột 2 và 3 tương ứng là nhóm thực tế và nhóm dự đoán của từng trường hợp. Cột 4 là xác suất có điều kiện ($P(D>d | G=g)$). Nó cho biết xác suất của điểm số quan sát thuộc nhóm đúng nhất. Cột 5 là bậc tự do. Cột 6 là xác suất ($P(G=g | D=d)$) chấp nhận từng trường hợp thuộc về nhóm đã được phân loại. Cột 7 là khoảng cách Mahalanobis bình phương của mỗi trường hợp đến trung tâm nhóm. Khoảng cách Mahalanobis càng nhỏ cho biết trường hợp ấy càng gần trung tâm nhóm.

File Edit View Data Transform Insert Format Analyze Graphs Utilities Extensions Window Help

Statistics
of Equality of Group Means
Within-Groups Matrices
sis 1
Title
X's Test of Equality of Covarian
Title
Log Determinants
Test Results
Stepwise Statistics
Title
Variables Entered/Removed
Variables in the Analysis
Variables Not in the Analysis
Wilks' Lambda
Summary of Canonical Discrimin
Title
Eigenvalues
Wilks' Lambda
Standardized Canonical Discrimi
Structure Matrix
Canonical Discriminant Function
Unions at Group Centroids
Classification Statistics
Title
Classification Processing Su
Prior Probabilities for Groups
Classification Function Coeff
Casewise Statistics
Separate-Groups Graphs
Title
Nhom = 1
Nhom = 2
Classification Results

Casewise Statistics

	Case Number	Actual Group	Predicted Group	Highest Group		df	P(G=g D=d)	Squared Mahalanobis Distance to Centroid	Second Highest Group		Discriminant Scores
				p	P(D=d G=g)				Group	P(G=g D=d)	
Original	1	1	1	.055	1	1.000	3.683	2	.000	21.585	-1.240
	2	1	1	.088	1	1.000	2.919	2	.000	23.587	-1.450
	3	1	1	.198	1	1.000	1.656	2	.000	27.857	-1.872
	4	1	1	.332	1	1.000	.943	2	.000	31.293	-2.188
	5	1	1	.512	1	1.000	.429	2	.000	34.929	-2.504
	6	1	1	.134	1	1.000	2.243	2	.000	25.678	-1.661
	7	1	1	.237	1	1.000	1.396	2	.000	28.980	-1.977
	8	1	1	.387	1	1.000	.749	2	.000	32.483	-2.293
	9	1	1	.512	1	1.000	.429	2	.000	34.929	-2.504
	10	1	1	.735	1	1.000	.115	2	.000	38.764	-2.820
	11	1	1	.898	1	1.000	.016	2	.000	41.432	-3.030
	12	1	1	.934	1	1.000	.007	2	.000	44.188	-3.241
	13	1	1	.769	1	1.000	.086	2	.000	47.034	-3.452
	14	1	1	.614	1	1.000	.254	2	.000	49.968	-3.662
	15	1	1	.475	1	1.000	.510	2	.000	52.991	-3.873
	16	1	1	.412	1	1.000	.672	2	.000	54.535	-3.978
	17	1	1	.303	1	1.000	1.062	2	.000	57.692	-4.189
	18	1	1	.256	1	1.000	1.290	2	.000	59.303	-4.294
	19	1	1	.455	1	1.000	.558	2	.000	53.461	-3.905
	20	1	1	.338	1	1.000	.917	2	.000	56.586	-4.116
	21	1	1	.288	1	1.000	1.129	2	.000	58.182	-4.221
	22	1	1	.243	1	1.000	1.364	2	.000	59.800	-4.327
	23	1	1	.969	1	1.000	.002	2	.000	43.610	-3.197

IBM SPSS Statistics Processor is ready

Unicond ON

Hình 8.8. Xác định những thống kê của các trường hợp.

Phần 2 bao gồm 3 cột. Cột 8÷10 là các thống kê kiểm định lại các trường hợp phân loại lần đầu ở phần 1 có thể thuộc về nhóm hai hay không (Second Highest Group). Cột 8 ghi lại nhóm của các trường hợp phân loại lần 2. Cột 9 là xác suất ($P(G=g | D=d)$) mà một trường hợp cụ thể thuộc về nhóm hai. Nếu $P < 0,01$, thì trường hợp đó không thuộc về nhóm hai. Cột 10 là khoảng cách Mahalanobis bình phương (D^2) của mỗi trường hợp đến trung tâm nhóm. Giá trị D^2 càng nhỏ cho biết trường hợp ấy càng gần trung tâm nhóm. Cột 11 là các điểm số phân loại (Discriminant Scores). Điểm số phân loại các trường hợp được tính từ hàm lập nhóm hợp quy chưa chuẩn hóa (Unstandardized canonical discriminant functions; Hàm 2 của Bảng 8.5). Những trường hợp nhận điểm số phân loại âm và dương tương ứng cho biết chúng thuộc về nhóm 1 và 2. Những trường hợp phân nhóm nhầm được ghi chú bằng dấu (××) ở nhóm dự đoán (Predicted Group).

Nội dung 7. Kết quả phân loại (Classification Results). Phần này cung cấp các thông tin về kết quả phân loại các trường hợp vào từng nhóm (Bảng 8.11).

Bảng 8.11. Kết quả phân loại (Classification Results)

Kết quả phân loại		Nhóm	Predicted Group Membership		Total
			1	2	
Original (Ban đầu)	Count	1	110		110
		2		102	102
	%	1	100,0		100
		2		100,0	100
Cross-validated ^b (Kiểm định chéo)	Count	1	110		110
		2		102	102
	%	1	100,0		100
		2		100,0	100

a. 100.0% of original grouped cases correctly classified.

b. 100.0% of cross-validated grouped cases correctly classified.

Mức độ thành công của việc phân loại các trường hợp vào từng nhóm được thống kê bằng tần số và tỷ lệ phần trăm. Bảng 8.11 bao gồm kết quả phân loại ban đầu (Original Classification Results) và kết quả kiểm định độ chính xác của phân loại ban đầu (Cross-validated Grouped Cases). Ở ví dụ này có 212 cây ban đầu được đưa vào phân loại; trong đó nhóm 1 là 110 cây, còn nhóm 2 là 102 cây. Kết quả dự đoán cho biết tất cả 110 cây được phân loại ban đầu vào nhóm 1 là chính xác 100%. Tương tự, tất cả 102 cây được phân loại ban đầu vào nhóm 2 là chính xác 100%. Bảng này cũng báo cáo số cây và tỷ lệ phần trăm số cây bị phân loại nhầm vào mỗi nhóm. Kết quả còn báo cáo tỷ lệ phần trăm trung bình đối với những cây được phân loại đúng. Ở ví dụ này tất cả 212 cây (100%) được phân loại đúng vào hai nhóm.

8.3. Phân tích mô hình nhiều nhóm

8.3.1. Khái quát chung về mô hình nhiều nhóm

Các mô hình lớn hơn hai nhóm là các mô hình nhiều nhóm. Các bước xây dựng mô hình nhiều nhóm cũng tương tự như mô hình hai nhóm. Mô hình nhiều nhóm có thể được xây dựng theo một số phương pháp khác nhau. Một mô hình có thể bao gồm nhiều biến dự đoán hoặc chỉ bao gồm một vài biến dự đoán. Tổ hợp các biến dự đoán khác nhau cho kết quả phân loại khác nhau.

8.3.2. Ứng dụng mô hình nhiều nhóm trong phân cấp sinh trưởng cây rừng

Giả sử mục tiêu nghiên cứu là xác định hàm để phân chia 5 cấp sinh trưởng từ I÷V đối với những cây hình thành rừng Đước (*Rhizophora apiculata*) trồng 16 tuổi. Các biến dự tuyến được sử dụng để phân chia 5 cấp sinh trưởng là D, H và D_T. Đặc trưng thống kê của rừng Đước trồng 16 tuổi được ghi lại ở Bảng 8.12. Trong trường hợp này, hàm lập nhóm có dạng như Hàm 8.9; trong đó k = 5 cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi.

$$Z^{(k)} = aD^{(k)} + bH^{(k)} + cD_T^{(k)} \quad (8.9)$$

Bảng 8.12. Đặc trưng thống kê D, H và D_T của rừng Đước trồng 16 tuổi

Chỉ tiêu	N (cây/ha)	Trung bình	Min	Max	±SEE	CV%
(1)	(2)	(3)	(4)	(5)	(6)	(7)
D (cm)	970	16,0	4,5	32,0	4,2	26,4
H (m)	970	14,5	3,0	30,5	4,2	28,8
D _T (m)	970	7,9	3,0	12,7	1,8	22,2
g (m ² /cây)	970	0,0215	0,0016	0,0804	0,0111	51,7
V (m ³ /cây)	970	0,1782	0,0024	1,2265	0,1460	81,9

Trước khi xây dựng hàm phân loại 5 cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi, chúng ta cần phải phân chia sơ bộ các cây vào 5 cấp sinh trưởng khác nhau. Đây là điều kiện để xây dựng hàm phân loại nhóm. Để thỏa mãn điều kiện này, chọn 1 biến phân loại nào đó và phân chia biến này thành 5 nhóm. Giả sử chọn biến D. Trước hết xác định các thống kê mô tả D. Kế đến xác định các hệ số D (KD) bằng cách chia D của từng cây (D_i) cho D bình quân (D_{Bq}) của rừng Đước 16 tuổi, nghĩa là $KD_i = D_i/D_{Bq}$. Tiếp theo phân chia chỉ số KD thành 5 cấp khác nhau. Ở ví dụ này, những cây thuộc cấp sinh trưởng I÷V được phân chia sơ bộ theo 5 nhóm hệ

số KD tương ứng như sau: $KD_I > 1,3$, $KD_{II} = 1,1 \div 1,3$; $KD_{III} = 0,9 \div 1,1$; $KD_{IV} = 0,7 \div 0,9$; $KD_V < 0,7$. Sau đó mã hóa các hệ số này bằng các số nguyên liên tục từ 1 ÷ 5.

Sau đó thực hiện các bước xây dựng và kiểm định 5 hàm phân cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi. Các bước xử lý ở đây cũng được thực hiện tương tự như mô hình 2 nhóm ở Mục 8.2. Kết quả xây dựng hàm phân loại 5 cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi được tóm tắt ở Bảng 8.13÷8.21. Bảng 8.13 cho biết kết quả kiểm định ngang bằng D, H và D_T giữa 5 cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi bằng tiêu chuẩn Wilks' Lambda. Khi $P < 0,01$, thì D, H và D_T khác nhau rõ rệt giữa 5 cấp sinh trưởng. Ở ví dụ này, giá trị $P < 0,01$, nên D, H và D_T khác nhau rõ rệt giữa 5 cấp sinh trưởng.

Bảng 8.13. Kiểm định ngang bằng D, H và D_T giữa 5 cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi

Biến dự đoán	Wilks' Lambda	F	df1	df2	Sig.
(1)	(2)	(3)	(4)	(5)	(6)
D	0,083	2666,6	4	965	0,000
H	0,082	2700,8	4	965	0,000
D_T	0,922	20,281	4	965	0,000

Bảng 8.14. Kiểm định các biến ở trong mô hình

Step		Tolerance	F to Remove	Min. D Squared	Between Groups
(1)		(2)	(3)	(4)	(5)
1	H	1,000	2700,834		
2	H	0,975	2624,548	0,010	1 and 2
	D_T	0,975	13,728	3,815	2 and 3
3	H	0,127	15,423	3,882	2 and 3
	D_T	0,975	13,318	3,857	2 and 3
	D	0,128	12,090	4,133	2 and 3

Bảng 8.15. Kiểm định các biến không ở trong mô hình

Step	Tolerance	Min. Tolerance	F to Enter	Min. D Squared	Between Groups
(1)	(2)	(3)	(4)	(5)	(6)
0	D	1,000	2666,588	3,601	2 and 3
	H	1,000	2700,834	3,815	2 and 3
	D_T	1,000	20,281	,010	1 and 2
1	D	0,128	12,496	3,857	2 and 3
	D_T	0,975	13,728	4,133	2 and 3
2	D	0,128	12,090	4,173	2 and 3

Bảng 8.16. Kiểm định ý nghĩa của các hàm phân loại hợp quy

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
(1)	(2)	(3)	(4)	(5)
1	11,618 ^a	99,4	99,4	0,960
2	0,058 ^a	0,5	99,9	0,235
3	0,014 ^a	0,1	100,0	0,116

Bảng 8.14 và 8.15 báo cáo kết quả kiểm định các biến ở trong hàm phân loại và những biến bị loại ra khỏi hàm phân loại bằng tiêu chuẩn độ chất nhận (Tolerance) và

tiêu chuẩn F_{Ra} (F to Remoe). Khi một biến được sử dụng để xây dựng hàm phân loại, thì độ chấp nhận phải cao, cao nhất là 1,0. Ở ví dụ này, ba biến D, H và D_T đều nhận Tolerance rất cao, nên chúng được chấp nhận để xây dựng hàm phân loại cấp sinh trưởng. Theo tiêu chuẩn F_{Ra} (F to Remoe), một biến được sử dụng để xây dựng hàm phân loại khi $F_{Ra} > 2,71$ (Mặc định). Ở ví dụ này, vì F_{Ra} (Bảng 8.15) của ba biến D, H và D_T đều lớn hơn 2,71, nên chúng được chấp nhận để xây dựng hàm phân loại cấp sinh trưởng.

Bảng 8.16 là kết quả kiểm định ý nghĩa của các hàm phân loại hợp quy. Các hàm này đo sự khác nhau giữa các nhóm. Giả định H_0 : Các nhóm là khác nhau. Sự khác nhau giữa các nhóm được đo bằng tỷ lệ (%) phương sai cộng dồn. Tỷ lệ này càng gần 100% thì các nhóm càng khác nhau. Ở ví dụ này, phần trăm phương sai cộng dồn (Cumulative %) ở các hàm đều gần bằng 100%, nên các nhóm khác nhau rất rõ rệt.

Bảng 8.17. Các hệ số của các hàm phân loại hợp quy chuẩn hóa

Biến dự đoán	Function		
	1	2	3
(1)	(2)	(3)	(4)
D	0,457	-1,379	2,394
H	0,570	1,321	-2,408
D_T	-0,099	0,862	0,522

Bảng 8.18. Các hệ số của các hàm phân loại hợp quy chưa chuẩn hóa

Biến dự đoán	Function		
	1	2	3
(1)	(2)	(3)	(4)
D	0,376	-1,133	1,966
H	0,476	1,103	-2,011
D_T	-0,059	,510	0,309
(Constant)	-12,447	-1,906	-4,742

Bảng 8.19. Điểm số trung bình của các hàm hợp quy ở trung tâm 5 cấp sinh trưởng đối với những cây hình thành rừng Được 16 tuổi

Cấp sinh trưởng	Function		
	1	2	3
(1)	(2)	(3)	(4)
1	6,207	-0,360	0,119
2	1,994	-0,004	-0,238
3	-0,012	0,248	0,054
4	-2,615	-0,072	0,044
5	-6,113	-0,342	-0,009

Bảng 8.17 là các hệ số của các hàm phân loại hợp quy chuẩn hóa. Chúng được sử dụng để xác định vai trò của các biến dự đoán trong hàm phân loại. Vai trò cao, thấp của các biến dự đoán được đánh giá theo độ lớn tuyệt đối của các hệ số hồi quy chuẩn hóa. Biến phân loại nào có hệ số hồi quy chuẩn hóa lớn hơn đóng vai trò lớn hơn so với biến khác. Ở ví dụ này, biến D và H có ý nghĩa lớn trong hàm phân loại thứ 3, còn biến D_T có ý nghĩa lớn trong hàm phân loại thứ 2. Bảng 8.18 là các hệ số của

các hàm phân loại hợp quy chưa chuẩn hóa. Chúng được sử dụng để tính điểm số cho các trường hợp đưa vào phân loại. Điểm số trung bình của các hàm hợp quy ở trung tâm 5 nhóm được báo cáo ở Bảng 8.19. Bảng 8.20 là các hàm phân loại 5 cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi. Các hàm này được sử dụng để phân loại các cây hình thành rừng Đước 16 tuổi vào 5 cấp sinh trưởng khác nhau.

Bảng 8.20. Các hàm phân loại 5 cấp sinh trưởng đối với rừng Đước 16 tuổi

Biến dự đoán	Các hàm phân cấp sinh trưởng				
	1	2	3	4	5
(1)	(2)	(3)	(4)	(5)	(6)
D	14,816	12,128	11,663	11,027	9,916
H	0,808	-0,086	-1,351	-2,922	-4,781
D _T	1,193	1,512	1,849	1,836	1,888
(Constant)	-188,590	-117,833	-92,745	-63,078	-34,091

Kết quả phân loại các cây vào 5 cấp sinh trưởng đối với rừng Đước 16 tuổi được tóm tắt ở Bảng 8.21. Từ đó cho thấy, tổng số 970 cây (100%) được sử dụng để xây dựng 5 hàm phân loại 5 cấp sinh trưởng; trong đó những cây thuộc cấp sinh trưởng 1, 2, 4 và 5 được phân loại chính xác 100%. Trái lại, tổng số cây ở cấp sinh trưởng 3 là 399 cây; trong đó số cây được phân loại chính xác là 369 (92,5%), còn lại 30 cây (7,5%) bị phân loại nhầm sang cấp sinh trưởng 2. Trung bình 96,9% (205 cây) số cây hình thành rừng Đước 16 tuổi được phân loại chính xác vào 5 cấp sinh trưởng. Điều đó chứng tỏ các hàm ở Bảng 5.20 được sử dụng để phân loại 5 cấp sinh trưởng đối với những cây hình thành rừng Đước 16 tuổi. Bảng 8.22 dẫn ví dụ ứng dụng 5 hàm phân loại để xác định 5 cây mới thuộc về những cấp sinh trưởng khác nhau.

Bảng 8.21. Kết quả phân loại 5 cấp sinh trưởng đối với rừng Đước 16 tuổi

	Cấp sinh trưởng	Predicted Group Membership					Tổng
		1	2	3	4	5	
Count	1	127					127
	2		176				176
	3		30	369			399
	4				144		144
	5					124	124
%	1	100.0					100.0
	2		100.0				100.0
	3		7.5	92.5			100.0
	4				100.0		100.0
	5					100.0	100.0

Bảng 8.22. Phân cấp sinh trưởng đối với 5 cây mới theo 5 hàm lập nhóm

Kích thước thân cây			Điểm số theo hàm:					Cấp sinh trưởng
D (cm)	H (m)	D _T (m)	1	2	3	4	5	
21,0	19,5	7,2	<u>147</u>	146	139	125	95	1
20,5	19,0	6,3	138	<u>139</u>	132	119	90	2
15,0	13,5	6,1	52	72	<u>75</u>	74	62	3
12,0	10,5	4,6	3	34	42	<u>47</u>	43	4
4,5	3,0	5,3	-113	-55	-35	-12	<u>6</u>	5

Bảng 8.23. Đặc trưng lâm học của rừng Đước 16 tuổi theo 5 cấp sinh trưởng thực tế.
Đơn vị tính: 1ha

Cấp sinh trưởng	N (cây)	D (cm)	H (m)	G (m ²)	M (m ³)	N (%)	G (%)	M (%)	Bqun (%)
I	127	23,4	21,7	5,5	61,1	13,1	26,5	35,4	25,0
II	176	18,3	16,9	4,6	39,2	18,1	22,2	22,7	21,0
III	399	16,0	14,5	8,0	58,7	41,1	38,5	33,9	37,9
IV	144	12,9	11,4	1,9	10,9	14,8	9,1	6,3	10,1
V	124	8,7	7,2	0,8	3,0	12,8	3,7	1,7	6,1
Tổng số	970	15,9	14,4	20,9	172,8	100,0	100,0	100,0	100,0

Bảng 8.24. Đặc trưng của rừng Đước 16 tuổi theo 5 cấp sinh trưởng được phân loại theo 5 hàm lập nhóm. Đơn vị tính: 1ha

Cấp sinh trưởng	N (cây)	D (cm)	H (m)	G (m ²)	M (m ³)	N (%)	G (%)	M (%)	Bqun (%)
I	127	23,4	21,7	5,5	61,1	13,1	26,5	35,4	25,0
II	206	18,2	16,7	5,3	44,8	21,2	25,6	25,9	24,3
III	369	15,9	14,4	7,3	53,0	38,0	35,1	30,7	34,6
IV	144	12,9	11,4	1,9	10,9	14,8	9,1	6,3	10,1
V	124	8,7	7,2	0,8	3,0	12,8	3,7	1,7	6,1
Tổng số	970	15,8	14,3	20,9	172,8	100,0	100,0	100,0	100,0

Đặc điểm của 5 cấp sinh trưởng thực tế được ghi lại tương ứng ở Bảng 8.23. Bảng 8.24 dẫn đặc trưng của rừng Đước 16 tuổi theo 5 cấp sinh trưởng được phân loại theo những hàm ở Bảng 8.20. So sánh số liệu ở Bảng 8.23 và 8.24 cho thấy kết quả phân cấp sinh trưởng thực tế và dự đoán đối với những cây hình thành rừng Đước 16 tuổi là tương tự như nhau. Sự khác biệt chỉ xảy ra ở cấp sinh trưởng III; trong đó tỷ lệ sai khác bình quân về số cây là 7,5%.

TÀI LIỆU THAM KHẢO

- Burr, I. W. (1942). Cumulative frequency functions. *Annals of Mathematical Statistics*. 13 (2): 215-232.
- David R. Anderson, Dennis J. Sweeney, Thomas A. Williams (2002). *Statistics for business and economics*, 8 th ed., BookMasters, Inc.
- Elliot, A (2006). *Statistical Analysis Quick Reference Guidebook: With SPSS Examples* 1st Edition. SAGE.
- Greig – Smith, P (1964). *Quantitative plant ecology*, Ed. 2. London.
- Gompertz, B (1825). On the nature of function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London*. DOI: 1-1098/rstl.0026. S2CID145157003.
- Nath, R. and Pavur, R (1985) A new statistic in the one way multivariate analysis of variance, *Computational Statistics and Data Analysis*, 2, 297–315.
- Nguyễn Hải Tuất (1982). *Thống kê toán học trong lâm nghiệp*, Nxb. Nông Nghiệp, Hà Nội.
- Nguyễn Ngọc Kiêng (1996). *Thống kê trong nghiên cứu khoa học*, Nxb. Giáo Dục.
- Nguyễn Ngọc Kiêng (2000). *Thống kê học ứng dụng: Các kiểu mẫu thí nghiệm*, Tủ sách Đại Học Nông Lâm Tp. Hồ Chí Minh.
- Nguyễn Văn Thêm (1996). *Thực hành sinh thái rừng*, Tủ sách Đại Học Nông Lâm Tp. Hồ Chí Minh.
- Nguyễn Văn Thêm (1998). *Phương pháp điều tra quần xã thực vật*, Tủ sách Đại Học Nông Lâm Tp. Hồ Chí Minh.
- Nguyễn Văn Liệu và các tác giả khác (2000). *SPSS- Ứng dụng phân tích dữ liệu trong quản trị kinh doanh và khoa học tự nhiên - xã hội*, Nxb. Giao Thông Vận Tải.
- Phan Hiếu Hiền (1996). *Phương pháp bố trí thí nghiệm và xử lý số liệu thực nghiệm*, Tủ sách Đại Học Nông Lâm Tp. Hồ Chí Minh.
- Rameshwar, D., Gupta và Debasis Kandu (1999). Generalized exponential distributions. *Austral. & New Zealand J. Statist.* 41(2), 1999, 173-188.
- Trần Bá Nhân và Đinh Thái Hoàng (1998). *Lý thuyết thống kê ứng dụng trong quản trị, kinh doanh và nghiên cứu kinh tế*, Nxb. Thống Kê.
- Todorov, V. and Filzmoser, P (2007) Robust statistic for the one-way MANOVA, submitted to the *Journal of Environmetrics*.
- Võ Văn Huy và các tác giả khác (1997). *Ứng dụng Spss for windows để xử lý và phân tích dữ kiện nghiên cứu marketing, quản trị, kinh tế, y học, tâm lý, xã hội*, Nxb. Khoa Học và Kỹ Thuật.
- Vũ Tiến Hình và các tác giả khác (1992). *Điều tra, quy hoạch và điều chế rừng*, Trường Đại Học Lâm Nghiệp.
- Vũ Tiến Hình (2003). *Sản lượng rừng*, Nxb. Nông Nghiệp, Hà Nội.
- Vasilevich, V.I (1969). *Những phương pháp thống kê trong địa thực vật*, Nxb. “Khoa học”, Leningrad (tiếng Nga).
- Ventsen, E.C (1969). *Lý thuyết xác suất*, Nxb. “Khoa học”, Moxkva (tiếng Nga).
- Zeide B (1993), *Analysis of growth equations*, *Forest Science*, 39, (3), tr. 594-61.

PHỤ LỤC

Phụ lục 1. ANOVA một yếu tố được bố trí ngẫu nhiên

TT	D, cm	Địa Hình	TT	D, cm	Địa Hình	TT	D, cm	Địa Hình
1	27.5	1	16	48	2	31	49.5	3
2	25.5	1	17	49	2	32	45.5	3
3	26.5	1	18	48.5	2	33	46.5	3
4	28.5	1	19	50	2	34	42.5	3
5	27.5	1	20	46.5	2	35	43.5	3
6	32.5	1	21	43.5	2	36	42.5	3
7	35.5	1	22	42.5	2	37	39.5	3
8	39.5	1	23	45.5	2	38	40.5	3
9	37.5	1	24	38.5	2	39	37.5	3
10	35.5	1	25	39.5	2	40	38.5	3
11	45.5	1	26	47.5	2	41	36.5	3
12	43.5	1	27	37.5	2	42	37.5	3
13	43.5	1	28	36.5	2	43	35.5	3
14	44.5	1	29	35.5	2	44	34.5	3
15	40.5	1	30	34.5	2	45	37.5	3

Phụ lục 2. ANOVA một yếu tố được bố trí theo khối ngẫu nhiên

TT	H, cm	Nghiem	Khối	TT	H, cm	Nghiem	Khối
1	73.2	1	1	11	81.2	3	3
2	81.0	1	2	12	81.0	3	4
3	90.4	1	3	13	81.6	4	1
4	76.6	1	4	14	98.6	4	2
5	82.8	2	1	15	83.0	4	3
6	84.8	2	2	16	94.4	4	4
7	86.8	2	3	17	85.2	5	1
8	84.8	2	4	18	97.2	5	2
9	76.4	3	1	19	84.2	5	3
10	83.0	3	2	20	99.2	5	4

Phụ lục 3. ANOVA một yếu tố được bố trí theo kiểu bình phương la tinh

TT	H, cm	Nghiem	TT. Hàng	TT. Cột	TT	H, cm	Nghiem	TT. Hàng	TT. Cột
1	67	1	1	1	19	24	4	1	4
2	46	2	2	1	20	2	6	2	4
3	47	3	3	1	21	13	5	3	4
4	28	4	4	1	22	90	1	4	4
5	15	5	5	1	23	42	3	5	4
6	7	6	6	1	24	86	2	6	4
7	62	2	1	2	25	9	5	1	5
8	62	1	2	2	26	54	3	2	5
9	5	6	3	2	27	35	4	3	5
10	43	3	4	2	28	3	6	4	5
11	42	4	5	2	29	81	2	5	5
12	23	5	6	2	30	105	1	6	5
13	30	3	1	3	31	5	6	1	6
14	15	5	2	3	32	19	4	2	6
15	65	1	3	3	33	57	2	3	6
16	61	2	4	3	34	17	5	4	6
17	8	6	5	3	35	87	1	5	6
18	58	4	6	3	36	45	3	6	6

Phụ lục 4. ANOVA hai yếu tố được bố trí theo khối ngẫu nhiên

TT	H, cm	Khối	Đạm	Lân	TT	H, cm	Khối	Đạm	Lân
1	60	1	1	1	24	82	4	2	2
2	56	2	1	1	25	84	5	2	2
3	58	3	1	1	26	100	1	2	3
4	62	4	1	1	27	80	2	2	3
5	60	5	1	1	28	78	3	2	3
6	70	1	1	2	29	76	4	2	3
7	76	2	1	2	30	74	5	2	3
8	84	3	1	2	31	104	1	3	1
9	80	4	1	2	32	96	2	3	1
10	76	5	1	2	33	98	3	3	1
11	96	1	1	3	34	90	4	3	1
12	92	2	1	3	35	92	5	3	1
13	92	3	1	3	36	90	1	3	2

(xem tiếp trang sau)

Phụ lục 4. ANOVA hai yếu tố được bố trí theo khối ngẫu nhiên (Tiếp)

TT	H, cm	Khối	Đạm	Lân	TT	H, cm	Khối	Đạm	Lân
14	94	4	1	3	37	84	2	3	2
15	86	5	1	3	38	86	3	3	2
16	116	1	2	1	39	80	4	3	2
17	120	2	2	1	40	82	5	3	2
18	124	3	2	1	41	78	1	3	3
19	112	4	2	1	42	80	2	3	3
20	120	5	2	1	43	76	3	3	3
21	92	1	2	2	44	74	4	3	3
22	90	2	2	2	45	80	5	3	3
23	96	3	2	2	46				

Phụ lục 5. ANOVA hai yếu tố được bố trí ngẫu nhiên

TT	H, cm	TTcây	Đạm	Lân	T	H, cm	TTcây	Đạm	Lân	T	H, cm	TTcây	Đạm	Lân
1	55	1	1	1	23	61	8	1	2	45	55	15	1	3
2	51	2	1	1	24	57	9	1	2	46	55	1	2	1
3	53	3	1	1	25	59	10	1	2	47	51	2	2	1
4	57	4	1	1	26	61	11	1	2	48	53	3	2	1
5	55	5	1	1	27	56	12	1	2	49	57	4	2	1
6	65	6	1	1	28	54	13	1	2	50	55	5	2	1
7	52	7	1	1	29	60	14	1	2	51	65	6	2	1
8	54	8	1	1	30	59	15	1	2	52	52	7	2	1
9	55	9	1	1	31	55	1	1	3	53	54	8	2	1
10	51	10	1	1	32	61	2	1	3	54	55	9	2	1
11	58	11	1	1	33	60	3	1	3	55	51	10	2	1
12	57	12	1	1	34	56	4	1	3	56	58	11	2	1
13	56	13	1	1	35	57	5	1	3	57	57	12	2	1
14	59	14	1	1	36	55	6	1	3	58	56	13	2	1
15	61	15	1	1	37	59	7	1	3	59	59	14	2	1
16	57	1	1	2	38	61	8	1	3	60	61	15	2	1
17	54	2	1	2	39	55	9	1	3	61	57	1	2	2
18	57	3	1	2	40	57	10	1	3	62	54	2	2	2
19	53	4	1	2	41	54	11	1	3	63	57	3	2	2
20	61	5	1	2	42	56	12	1	3	64	53	4	2	2
21	58	6	1	2	43	61	13	1	3	65	61	5	2	2
22	56	7	1	2	44	59	14	1	3	66	58	6	2	2

Phụ lục 5. ANOVA hai yếu tố được bố trí ngẫu nhiên (Tiếp)

TT	H, cm	TTcây	Đạm	Lân	TT	H, cm	TTcây	Đạm	Lân	TT	H, cm	TTcây	Đạm	Lân
67	56	7	2	2	92	61	2	3	1	113	71	8	3	2
68	61	8	2	2	93	63	3	3	1	114	67	9	3	2
69	57	9	2	2	94	67	4	3	1	115	69	10	3	2
70	59	10	2	2	95	65	5	3	1	116	71	11	3	2
71	61	11	2	2	96	75	6	3	1	117	65	12	3	2
72	56	12	2	2	97	62	7	3	1	118	63	13	3	2
73	54	13	2	2	98	64	8	3	1	119	70	14	3	2
74	60	14	2	2	99	65	9	3	1	120	69	15	3	2
75	59	15	2	2	89	59	14	2	3	121	65	1	3	3
76	55	1	2	3	90	55	15	2	3	122	71	2	3	3
77	61	2	2	3	91	65	1	3	1	123	70	3	3	3
78	60	3	2	3	92	61	2	3	1	124	65	4	3	3
79	56	4	2	3	93	63	3	3	1	125	67	5	3	3
80	57	5	2	3	94	67	4	3	1	126	65	6	3	3
81	55	6	2	3	95	65	5	3	1	127	69	7	3	3
82	59	7	2	3	96	75	6	3	1	128	71	8	3	3
83	61	8	2	3	97	62	7	3	1	129	65	9	3	3
84	55	9	2	3	98	64	8	3	1	130	67	10	3	3
85	57	10	2	3	99	65	9	3	1	131	64	11	3	3
86	54	11	2	3	107	64	2	3	2	132	65	12	3	3
87	56	12	2	3	108	67	3	3	2	133	71	13	3	3
88	61	13	2	3	109	63	4	3	2	134	69	14	3	3
89	59	14	2	3	110	71	5	3	2	135	65	15	3	3
90	55	15	2	3	111	67	6	3	2					
91	65	1	3	1	112	65	7	3	2					

Phụ lục 6. ANOVA hai yếu tố được bố trí theo kiểu bình phương la tinh

H, cm	Nghiệm	Hàng	Cột	KT. bầu	NPK	H, cm	Nghiệm	Hàng	Cột	KT. bầu	NPK
67.5	3	1	1	1	3	77.5	5	4	1	2	2
68.8	4	1	2	2	1	76.3	6	4	2	2	3
71.3	6	1	3	2	3	66.3	2	4	3	1	2
59	1	1	4	1	1	72.5	3	4	4	1	3
80	5	1	5	2	2	58.8	1	4	5	1	1
61.3	2	1	6	1	2	63.8	4	4	6	2	1
62.5	2	2	1	1	2	65	4	5	1	2	1
71.3	3	2	2	1	3	78.8	5	5	2	2	2
81.3	5	2	3	2	2	56	1	5	3	1	1
72.5	6	2	4	2	3	68.8	2	5	4	1	2

71.3	4	2	5	2	1	75	6	5	5	2	3
57.5	1	2	6	1	1	66.3	3	5	6	1	3
70	6	3	1	2	3	53.8	1	6	1	1	1
60	1	3	2	1	1	67.5	2	6	2	1	2
68.8	3	3	3	1	3	66.3	4	6	3	2	1
70	4	3	4	2	1	82.5	5	6	4	2	2
63.8	2	3	5	1	2	73.8	3	6	5	1	3
80	5	3	6	2	2	70	6	6	6	2	3

Phụ lục 7. ANOVA hai yếu tố được bố trí theo kiểu lô chính - lô phụ

TT	H, cm	Phương thức	Loài cây	Khối	TT	H, cm	Phương thức	Loài cây	Khối
1	54	1	1	1	19	67	2	2	3
2	58	1	1	2	20	63	2	2	4
3	63	1	1	3	21	81	2	3	1
4	55	1	1	4	22	75	2	3	2
5	60	1	2	1	23	74	2	3	3
6	64	1	2	2	24	78	2	3	4
7	66	1	2	3	25	57	3	1	1
8	63	1	2	4	26	68	3	1	2
9	82	1	3	1	27	66	3	1	3
10	83	1	3	2	28	62	3	1	4
11	87	1	3	3	29	56	3	2	1
12	79	1	3	4	30	65	3	2	2
13	63	2	1	1	31	62	3	2	3
14	53	2	1	2	32	58	3	2	4
15	55	2	1	3	33	53	3	3	1
16	52	2	1	4	34	57	3	3	2
17	69	2	2	1	35	55	3	3	3
18	64	2	2	2	36	54	3	3	4

Phụ lục 8. ANOVA ba yếu tố

TT	D, cm	Ánh sáng	Phân bón	Nước	Khối	TT	D, cm	Ánh sáng	Phân bón	Nước	Khối
1	2.49	1	1	1	1	25	3.23	2	3	1	1
2	2.49	1	1	1	2	26	3.26	2	3	1	2
3	2.55	1	1	2	1	27	3.34	2	3	2	1
4	2.55	1	1	2	2	28	3.38	2	3	2	2
5	2.54	1	2	1	1	29	3.38	2	4	1	1
6	2.59	1	2	1	2	30	3.38	2	4	1	2
7	2.63	1	2	2	1	31	3.52	2	4	2	1
8	2.66	1	2	2	2	32	3.48	2	4	2	2
9	2.86	1	3	1	1	33	3.19	3	1	1	1

10	2.8	1	3	1	2	34	3.13	3	1	1	2
11	2.77	1	3	2	1	35	3.31	3	1	2	1
12	2.66	1	3	2	2	36	3.31	3	1	2	2
13	3.58	1	4	1	1	37	3.88	3	2	1	1
14	3.64	1	4	1	2	38	3.82	3	2	1	2
15	3.74	1	4	2	1	39	3.78	3	2	2	1
16	3.44	1	4	2	2	40	3.87	3	2	2	2
17	2.81	2	1	1	1	41	4	3	3	1	1
18	2.76	2	1	1	2	42	4.08	3	3	1	2
19	2.88	2	1	2	1	43	4.07	3	3	2	1
20	2.8	2	1	2	2	44	4.12	3	3	2	2
21	3.03	2	2	1	1	45	3.99	3	4	1	1
22	3.03	2	2	1	2	46	4.05	3	4	1	2
23	3.28	2	2	2	1	47	4.31	3	4	2	1
24	3.09	2	2	2	2	48	4.4	3	4	2	2

Phụ lục 9. ANOCOVA một yếu tố được bố trí ngẫu nhiên

TT	D, cm	Zd, cm	Nghiệm	TT	D, cm	Zd, cm	Nghiệm	TT	D, cm	Zd, cm	Nghiệm
1	4.6	1.8	1	11	5.8	1.9	1	21	1.9	1.9	2
2	5.2	2.3	1	12	1.8	1.4	2	22	4.2	2.1	2
3	3.9	1.8	1	13	4.4	1.7	2	23	3	1.4	2
4	6.5	2.5	1	14	3.3	1.8	2	24	4.7	2	2
5	5.4	2.6	1	15	7.8	2.4	2	25	5.8	2.6	2
6	3.3	2	1	16	5.1	2.3	2	26	7	2.6	2
7	4.9	1.2	1	17	4.3	2.2	2	27	6.7	3.6	2
8	7.6	3.4	1	18	4.4	1.9	2	28	3.1	1.6	2
9	7.1	2.2	1	19	5.8	2.6	2	29	2.1	1.9	2
10	4.1	1.1	1	20	6.3	3.4	2	30	5.6	2	2

Phụ lục 10. ANOCOVA một yếu tố được bố trí ngẫu nhiên

TT	H, m	Zh, m	Nghiệm	TT	H, m	Zh, m	Nghiệm
1	79	1.96	1	21	62	1.22	3
2	65	1.77	1	22	73	1.39	3
3	57	1.62	1	23	58	1.28	3
4	51	1.76	1	24	43	1.28	3
5	57	1.88	1	25	50	1.45	3
6	66	1.5	1	26	44	1.22	3
7	44	1.6	1	27	48	1.31	3
8	41	1.49	1	28	51	1.57	3
9	44	1.77	1	29	40	1.21	3

10	36	1.27	1	30	38	1.06	3
11	61	1.4	2	31	71	1.15	4
12	59	1.79	2	32	60	1.28	4
13	59	1.61	2	33	54	1.4	4
14	53	1.47	2	34	50	1.37	4
15	56	1.69	2	35	60	1.19	4
16	50	1.48	2	36	61	1.18	4
17	45	1.4	2	37	44	1.2	4
18	39	1.42	2	38	53	0.96	4
19	38	1.29	2	39	41	1.13	4
20	45	1.26	2	40	38	1.12	4

Phụ lục 11. ANOCOVA một yếu tố được bố trí theo khối ngẫu nhiên

TT	N, cây	H, cm	Khối	Dòng	TT	N, cây	H, cm	Khối	Dòng
1	28	202	1	1	13	27	191	3	1
2	23	145	1	2	14	28	203	3	2
3	27	188	1	3	15	27	185	3	3
4	24	201	1	4	16	30	238	3	4
5	30	202	1	5	17	26	198	3	5
6	30	228	1	6	18	27	207	3	6
7	22	165	2	1	19	19	134	4	1
8	26	201	2	2	20	24	180	4	2
9	24	185	2	3	21	28	220	4	3
10	28	231	2	4	22	30	261	4	4
11	26	178	2	5	23	29	226	4	5
12	25	221	2	6	24	24	204	4	6

Phụ lục 12. ANOCOVA một yếu tố được bố trí theo kiểu bình phương la tinh

TT	Tỉ lệ bệnh	Zh, cm	Dòng	Khối	Nghiệm
1	29.2	19.7	1	1	1
2	34.7	20.7	2	2	1
3	30.2	20.1	3	3	1
4	48.2	14.7	4	4	1
5	4.3	26.7	1	1	2
6	19.3	21.3	2	2	2
7	10.1	28.3	3	3	2
8	14	25.1	4	4	2
9	14	26	1	1	3
10	7.2	24.9	2	2	3
11	6.3	29	3	3	3

12	1	28.7	4	4	3
13	8.9	29.8	1	1	4
14	6.7	29	2	2	4
15	6.4	27.3	3	3	4
16	5.6	34.1	4	4	4

Phụ lục 13. ANOMUCOVA được bố trí theo khối ngẫu nhiên

H, cm	Zh, cm	Nghiệm	Khối	H, cm	Zh, cm	Nghiệm	Khối
183	2.45	1	1	224	3.22	3	1
176	2.25	1	2	258	4.14	3	2
291	4.38	1	3	244	2.32	3	3
254	4.35	1	4	217	4.42	3	4
225	3.42	1	5	192	3.28	3	5
249	3.27	1	6	236	4	3	6
356	6.71	2	1	329	6.34	4	1
300	5.44	2	2	283	5.44	4	2
301	4.92	2	3	308	5.22	4	3
271	5.23	2	4	326	8	4	4
288	6.74	2	5	318	6.96	4	5
258	4.74	2	6	318	6.96	4	6

Phụ lục 14. ANOMUCOVA được bố trí ngẫu nhiên

TT	H, cm	A, tháng	Zh, cm	Nghiệm	TT	H, cm	A, tháng	Zh, cm	Nghiệm
1	78	6	1.4	1	21	78	8	1.67	3
2	90	6	1.79	1	22	83	6	1.41	3
3	94	8	1.72	1	23	79	6	1.73	3
4	71	5	1.47	1	24	70	5	1.23	3
5	99	6	1.26	1	25	85	6	1.29	3
6	80	5	1.28	1	26	83	4	1.22	3
7	83	6	1.34	1	27	71	5	1.39	3
8	75	5	1.55	1	28	66	4	1.59	3
9	62	4	1.57	1	29	67	4	1.56	3
10	67	4	1.26	1	30	67	4	1.36	3
11	78	7	1.61	2	31	77	6	1.4	4
12	99	8	1.31	2	32	71	6	1.47	4
13	80	6	1.12	2	33	78	6	1.37	4
14	75	5	1.35	2	34	70	4	1.15	4
15	94	6	1.29	2	35	95	6	1.22	4
16	91	4	1.24	2	36	96	5	1.48	4
17	75	5	1.29	2	37	71	4	1.31	4
18	63	4	1.43	2	38	63	4	1.27	4

19	62	5	1.29	2	39	62	5	1.22	4
20	67	4	1.26	2	40	67	4	1.36	4

Phụ lục 15. ANOMUCOVA được bố trí theo khối ngẫu nhiên

TT	H, cm	Lr, cm	Zh, cm	Khối	Nghiệm
1	25.6	14.9	19	1	1
2	28.5	12.8	25.3	1	2
3	25.4	13.3	22.2	1	3
4	33	14.7	32.8	1	4
5	30.8	4.6	35.3	1	5
6	28	7.5	35.8	1	6
7	25.4	7.2	32.4	2	1
8	25.9	9.2	28.3	2	2
9	28.3	9.5	32.2	2	3
10	32.4	9.7	35.7	2	4
11	35.2	6.8	43.7	2	5
12	24.2	7.5	35.2	2	6
13	27.9	18.6	26.2	3	1
14	23.7	14.4	20.6	3	2
15	34.4	22.2	34.7	3	3
16	27.5	17.6	29.6	3	4
17	32.5	10	40	3	5
18	32.9	7.9	47.2	3	6

Phụ lục 16. Nhập số liệu vào Bảng tính đối với kiểu thí nghiệm phân cấp

TT	Dược chất	Cây	Lá	TT	Dược chất	Cây	Lá
1	3.28	1	1	13	2.77	3	1
2	3.09	1	1	14	2.66	3	1
3	3.52	1	2	15	3.74	3	2
4	3.48	1	2	16	3.44	3	2
5	2.88	1	3	17	2.55	3	3
6	2.8	1	3	18	2.55	3	3
7	2.46	2	1	19	3.78	4	1
8	2.44	2	1	20	3.87	4	1
9	1.87	2	2	21	4.07	4	2
10	1.92	2	2	22	4.12	4	2
11	2.19	2	3	23	3.31	4	3
12	2.19	2	3	24	3.31	4	3

Phụ biểu 17. Sự kết nhóm giữa các loài cây gỗ

Thứ tự	Trắc	Hà nu	Chò xốt	Lành ngành	Chò xốt-Lành ngành	Hà nu - Lành ngành
1	1	1	0	0	0	0
2	0	1	0	0	0	0
3	1	1	1	0	0	0
4	1	1	1	1	1	1
5	1	1	0	1	0	1
6	1	1	0	1	0	1
7	1	1	0	0	0	0
8	0	1	0	1	0	1
9	1	0	1	1	1	0
10	1	0	1	1	1	0
11	1	1	1	0	0	0
12	1	1	1	1	1	1
13	0	0	1	1	1	0
14	0	1	1	1	1	1
15	1	1	1	0	0	0
16	0	1	0	1	0	1
17	1	1	1	0	0	0
18	0	1	1	1	1	1
19	1	1	0	0	0	0
20	1	0	1	1	1	0
21	1	0	1	1	1	0
22	1	1	1	1	1	1
23	0	1	1	1	1	1
24	0	1	1	0	0	0
25	0	1	0	1	0	1
26	0	1	0	1	0	1
27	1	1	0	0	0	0
28	1	0	1	1	1	0
29	1	1	0	1	0	1
30	0	0	1	1	1	0
31	1	0	1	1	1	0
32	1	1	0	1	0	1
33	1	1	1	0	0	0
34	1	0	0	1	0	0

Tiếp phụ biểu 17

Thứ tự	Trắc	Hà nu	Chò xốt	Lãnh ngành	Chò xốt-Lãnh ngành	Hà nu - Lãnh ngành
35	1	0	1	0	0	0
36	0	1	1	1	1	1
37	0	1	1	0	0	0
38	1	0	1	1	1	0
39	0	1	1	0	0	0
40	1	1	0	1	0	1
41	0	1	1	0	0	0
42	1	0	1	1	1	0
43	0	1	1	1	1	1
44	1	1	1	1	1	1
45	0	1	1	0	0	0
46	1	0	1	1	1	0
47	0	1	1	0	0	0
48	1	1	1	0	0	0
49	0	1	1	1	1	1
50	1	1	0	1	0	1
51	0	1	0	1	0	1
52	1	0	1	1	1	0
53	1	1	1	0	0	0
54	0	1	1	1	1	1
55	1	1	1	1	1	1
56	1	1	0	1	0	1
57	0	1	0	1	0	1
58	1	1	1	0	0	0
59	1	0	0	1	0	0
60	1	0	1	0	0	0
61	1	1	1	0	0	0
62	0	1	0	1	0	1
63	0	1	1	0	0	0
64	1	1	0	0	0	0
65	1	1	1	1	1	1
66	1	0	0	0	0	0
67	1	1	0	1	0	1

Tiếp phụ biểu 17

Thứ tự	Trắc	Hà nu	Chò xốt	Lãnh ngành	Chò xốt-Lãnh ngành	Hà nu - Lãnh ngành
68	0	1	1	1	1	1
69	1	1	0	0	0	0
70	1	0	1	0	0	0
71	0	1	1	1	1	1
72	1	1	1	1	1	1
73	0	1	1	1	1	1
74	0	1	1	0	0	0
75	1	0	1	1	1	0
76	0	0	1	0	0	0
77	0	1	1	0	0	0
78	0	0	0	0	0	0
79	1	1	0	0	0	0
80	1	1	1	1	1	1
81	0	1	1	0	0	0
82	0	0	1	1	1	0
83	1	0	1	0	0	0
84	0	0	1	1	1	0
85	1	1	0	0	0	0
86	1	0	1	0	0	0
87	1	0	0	1	0	0
88	0	1	0	1	0	1
89	0	1	1	0	0	0
90	1	1	0	0	0	0
91	1	0	0	1	0	0
92	1	0	1	0	0	0
93	1	0	1	1	1	0
94	1	1	1	1	1	1
95	1	1	1	0	0	0
96	1	0	0	1	0	0
97	1	1	0	1	0	1
98	0	0	0	0	0	0
99	0	1	1	0	0	0

Tiếp phụ biểu 17

Thứ tự	Trắc	Hà nu	Chò xốt	Lãnh ngành	Chò xốt-Lãnh ngành	Hà nu - Lãnh ngành
100	1	1	0	0	0	0
101	0	1	1	0	0	0
102	0	0	1	1	1	0
103	1	0	0	0	0	0
104	0	1	0	1	0	1
105	0	0	1	1	1	0
106	1	1	0	0	0	0
107	0	0	1	1	1	0
108	0	1	0	1	0	1
109	1	1	0	0	0	0
110	0	0	0	1	0	0
111	1	1	0	0	0	0
112	0	1	1	0	0	0
113	1	0	0	0	0	0
114	1	1	1	0	0	0
115	0	0	0	0	0	0
116	0	1	1	1	1	1
117	1	0	0	1	0	0
118	0	1	0	1	0	1
119	0	1	1	1	1	1
120	1	1	1	0	0	0
121	1	0	1	0	0	0
122	0	1	0	1	0	1
123	0	0	0	1	0	0
124	1	1	0	0	0	0
125	0	0	0	0	0	0
126	1	1	0	0	0	0
127	0	1	1	1	1	1
128	1	1	0	0	0	0
129	1	0	0	0	0	0
130	0	0	1	1	1	0
131	1	0	0	1	0	0

Tiếp phụ biểu 17

Thứ tự	Trắc	Hà nu	Chò xốt	Lãnh ngành	Chò xốt-Lãnh ngành	Hà nu - Lãnh ngành
132	1	0	0	0	0	0
133	0	0	1	0	0	0
134	1	1	0	0	0	0
135	1	1	1	0	0	0
136	0	1	1	0	0	0
137	0	1	1	0	0	0
138	1	1	1	0	0	0
139	1	1	0	1	0	1
140	0	0	1	0	0	0
141	0	1	1	0	0	0
142	1	1	0	0	0	0
143	0	0	1	1	1	0
144	1	1	0	1	0	1
145	1	1	1	0	0	0
146	1	0	0	1	0	0
147	1	0	1	1	1	0
148	1	0	1	0	0	0
149	0	1	1	1	1	1
150	1	0	1	1	1	0
151	1	0	0	1	0	0
152	1	1	1	0	0	0
153	1	1	0	0	0	0
154	0	0	1	1	1	0
155	1	1	0	1	0	1
156	0	0	1	1	1	0
157	1	1	0	0	0	0
158	0	0	1	1	1	0
159	0	0	1	1	1	0
160	1	0	1	0	0	0
161	1	1	0	1	0	1
162	1	0	0	1	0	0
163	1	0	1	1	1	0
164	1	0	1	1	1	0

Tiếp phụ biểu 17

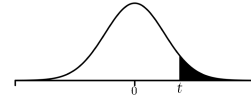
Thứ tự	Trắc	Hà nu	Chò xốt	Lãnh ngành	Chò xốt-Lãnh ngành	Hà nu - Lãnh ngành
165	1	0	1	1	1	0
166	1	0	1	0	0	0
167	0	1	1	0	0	0
168	0	0	1	1	1	0
169	1	1	0	1	0	1
170	0	1	1	0	0	0
171	1	0	1	0	0	0
172	1	0	0	1	0	0
173	1	0	1	1	1	0
174	0	0	1	0	0	0
175	1	1	1	1	1	1
176	1	0	1	1	1	0
177	1	0	1	0	0	0
178	0	0	1	0	0	0
179	0	1	1	0	0	0
180	1	0	1	1	1	0
181	1	1	0	0	0	0
182	1	0	1	1	1	0
183	0	0	1	1	1	0
184	1	0	1	0	0	0
185	1	0	1	1	1	0
186	1	0	1	1	1	0
187	0	1	1	1	1	1
188	0	1	0	1	0	1
189	0	1	1	1	1	1
190	1	1	0	0	0	0
191	1	1	0	1	0	1
192	1	1	0	1	0	1
193	1	1	0	1	0	1
194	0	1	1	0	0	0
195	0	1	0	1	0	1
196	1	0	0	1	0	0

Tiếp phụ biểu 17

Thứ tự	Trắc	Hà nu	Chò xốt	Lãnh ngành	Chò xốt-Lãnh ngành	Hà nu - Lãnh ngành
197	1	0	1	1	1	0
198	1	1	1	1	1	1
199	1	1	0	0	0	0
200	0	1	0	0	0	0
201	1	0	1	1	1	0
202	0	1	0	0	0	0
203	1	1	0	1	0	1
204	1	1	1	1	1	1
205	0	1	0	1	0	1
206	0	0	0	1	0	0
207	1	0	1	0	0	0
208	1	1	0	1	0	1
209	0	1	1	0	0	0
210	1	1	1	0	0	0
211	1	1	0	1	0	1
212	0	1	0	0	0	0
213	0	1	1	0	0	0
214	1	1	1	1	1	1
215	1	0	1	1	1	0
216	1	1	1	0	0	0
217	1	1	0	1	0	1
218	0	1	1	0	0	0
219	1	0	1	1	1	0
220	1	1	0	0	0	0
221	0	1	0	1	0	1
222	0	0	1	0	0	0
223	1	1	1	0	0	0
224	1	1	0	0	0	0
225	0	0	1	1	1	0
226	0	1	1	0	0	0
227	1	1	0	1	0	1
228	1	1	1	1	1	1

Tiếp phụ biểu 17

Thứ tự	Trắc	Hà nu	Chò xốt	Lãnh ngành	Chò xốt-Lãnh ngành	Hà nu - Lãnh ngành
229	0	0	1	0	0	0
230	0	1	0	1	0	1
231	1	0	1	1	1	0
232	1	1	0	1	0	1
233	1	0	1	1	1	0
234	0	0	0	1	0	0
235	0	1	1	0	0	0
236	1	1	0	0	0	0
237	1	0	1	1	1	0
238	0	1	0	0	0	0
239	1	1	0	1	0	1
240	1	0	1	0	0	0
241	0	1	1	0	0	0
242	0	0	1	0	0	0
243	1	0	1	1	1	0
244	1	0	0	1	0	0
245	0	1	1	1	1	1
246	1	1	0	0	0	0
247	1	1	1	0	0	0
248	1	0	0	0	0	0
249	1	1	0	1	0	1
250	1	0	1	1	1	0

Critical Values for Student's t -Distribution.

df	Upper Tail Probability: $\Pr(T > t)$									
	0.2	0.1	0.05	0.04	0.03	0.025	0.02	0.01	0.005	0.0005
1	1.376	3.078	6.314	7.916	10.579	12.706	15.895	31.821	63.657	636.619
2	1.061	1.886	2.920	3.320	3.896	4.303	4.849	6.965	9.925	31.599
3	0.978	1.638	2.353	2.605	2.951	3.182	3.482	4.541	5.841	12.924
4	0.941	1.533	2.132	2.333	2.601	2.776	2.999	3.747	4.604	8.610
5	0.920	1.476	2.015	2.191	2.422	2.571	2.757	3.365	4.032	6.869
6	0.906	1.440	1.943	2.104	2.313	2.447	2.612	3.143	3.707	5.959
7	0.896	1.415	1.895	2.046	2.241	2.365	2.517	2.998	3.499	5.408
8	0.889	1.397	1.860	2.004	2.189	2.306	2.449	2.896	3.355	5.041
9	0.883	1.383	1.833	1.973	2.150	2.262	2.398	2.821	3.250	4.781
10	0.879	1.372	1.812	1.948	2.120	2.228	2.359	2.764	3.169	4.587
11	0.876	1.363	1.796	1.928	2.096	2.201	2.328	2.718	3.106	4.437
12	0.873	1.356	1.782	1.912	2.076	2.179	2.303	2.681	3.055	4.318
13	0.870	1.350	1.771	1.899	2.060	2.160	2.282	2.650	3.012	4.221
14	0.868	1.345	1.761	1.887	2.046	2.145	2.264	2.624	2.977	4.140
15	0.866	1.341	1.753	1.878	2.034	2.131	2.249	2.602	2.947	4.073
16	0.865	1.337	1.746	1.869	2.024	2.120	2.235	2.583	2.921	4.015
17	0.863	1.333	1.740	1.862	2.015	2.110	2.224	2.567	2.898	3.965
18	0.862	1.330	1.734	1.855	2.007	2.101	2.214	2.552	2.878	3.922
19	0.861	1.328	1.729	1.850	2.000	2.093	2.205	2.539	2.861	3.883
20	0.860	1.325	1.725	1.844	1.994	2.086	2.197	2.528	2.845	3.850
21	0.859	1.323	1.721	1.840	1.988	2.080	2.189	2.518	2.831	3.819
22	0.858	1.321	1.717	1.835	1.983	2.074	2.183	2.508	2.819	3.792
23	0.858	1.319	1.714	1.832	1.978	2.069	2.177	2.500	2.807	3.768
24	0.857	1.318	1.711	1.828	1.974	2.064	2.172	2.492	2.797	3.745
25	0.856	1.316	1.708	1.825	1.970	2.060	2.167	2.485	2.787	3.725
26	0.856	1.315	1.706	1.822	1.967	2.056	2.162	2.479	2.779	3.707
27	0.855	1.314	1.703	1.819	1.963	2.052	2.158	2.473	2.771	3.690
28	0.855	1.313	1.701	1.817	1.960	2.048	2.154	2.467	2.763	3.674
29	0.854	1.311	1.699	1.814	1.957	2.045	2.150	2.462	2.756	3.659
30	0.854	1.310	1.697	1.812	1.955	2.042	2.147	2.457	2.750	3.646
31	0.853	1.309	1.696	1.810	1.952	2.040	2.144	2.453	2.744	3.633
32	0.853	1.309	1.694	1.808	1.950	2.037	2.141	2.449	2.738	3.622
33	0.853	1.308	1.692	1.806	1.948	2.035	2.138	2.445	2.733	3.611
34	0.852	1.307	1.691	1.805	1.946	2.032	2.136	2.441	2.728	3.601
35	0.852	1.306	1.690	1.803	1.944	2.030	2.133	2.438	2.724	3.591
36	0.852	1.306	1.688	1.802	1.942	2.028	2.131	2.434	2.719	3.582
37	0.851	1.305	1.687	1.800	1.940	2.026	2.129	2.431	2.715	3.574
38	0.851	1.304	1.686	1.799	1.939	2.024	2.127	2.429	2.712	3.566
39	0.851	1.304	1.685	1.798	1.937	2.023	2.125	2.426	2.708	3.558
40	0.851	1.303	1.684	1.796	1.936	2.021	2.123	2.423	2.704	3.551
41	0.850	1.303	1.683	1.795	1.934	2.020	2.121	2.421	2.701	3.544
42	0.850	1.302	1.682	1.794	1.933	2.018	2.120	2.418	2.698	3.538
43	0.850	1.302	1.681	1.793	1.932	2.017	2.118	2.416	2.695	3.532
44	0.850	1.301	1.680	1.792	1.931	2.015	2.116	2.414	2.692	3.526
45	0.850	1.301	1.679	1.791	1.929	2.014	2.115	2.412	2.690	3.520
46	0.850	1.300	1.679	1.790	1.928	2.013	2.114	2.410	2.687	3.515
47	0.849	1.300	1.678	1.789	1.927	2.012	2.112	2.408	2.685	3.510
48	0.849	1.299	1.677	1.789	1.926	2.011	2.111	2.407	2.682	3.505
49	0.849	1.299	1.677	1.788	1.925	2.010	2.110	2.405	2.680	3.500
50	0.849	1.299	1.676	1.787	1.924	2.009	2.109	2.403	2.678	3.496
60	0.848	1.296	1.671	1.781	1.917	2.000	2.099	2.390	2.660	3.460
70	0.847	1.294	1.667	1.776	1.912	1.994	2.093	2.381	2.648	3.435
80	0.846	1.292	1.664	1.773	1.908	1.990	2.088	2.374	2.639	3.416
90	0.846	1.291	1.662	1.771	1.905	1.987	2.084	2.368	2.632	3.402
100	0.845	1.290	1.660	1.769	1.902	1.984	2.081	2.364	2.626	3.390
120	0.845	1.289	1.658	1.766	1.899	1.980	2.076	2.358	2.617	3.373
140	0.844	1.288	1.656	1.763	1.896	1.977	2.073	2.353	2.611	3.361
180	0.844	1.286	1.653	1.761	1.893	1.973	2.069	2.347	2.603	3.345
200	0.843	1.286	1.653	1.760	1.892	1.972	2.067	2.345	2.601	3.340
500	0.842	1.283	1.648	1.754	1.885	1.965	2.059	2.334	2.586	3.310
1000	0.842	1.282	1.646	1.752	1.883	1.962	2.056	2.330	2.581	3.300
∞	0.842	1.282	1.645	1.751	1.881	1.960	2.054	2.326	2.576	3.291
	60%	80%	90%	92%	94%	95%	96%	98%	99%	99.9%
	Confidence Level									

Note: $t(\infty)_{\alpha/2} = Z_{\alpha/2}$ in our notation.

STATISTICAL TABLES

Cumulative normal distribution

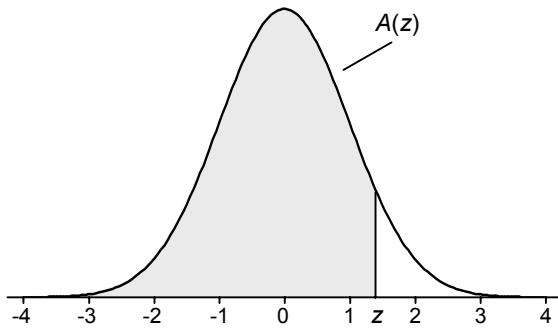
Critical values of the t distribution

Critical values of the F distribution

Critical values of the chi-squared distribution

TABLE A.1

Cumulative Standardized Normal Distribution



$A(z)$ is the integral of the standardized normal distribution from $-\infty$ to z (in other words, the area under the curve to the left of z). It gives the probability of a normal random variable not being more than z standard deviations above its mean. Values of z of particular importance:

z	$A(z)$	
1.645	0.9500	Lower limit of right 5% tail
1.960	0.9750	Lower limit of right 2.5% tail
2.326	0.9900	Lower limit of right 1% tail
2.576	0.9950	Lower limit of right 0.5% tail
3.090	0.9990	Lower limit of right 0.1% tail
3.291	0.9995	Lower limit of right 0.05% tail

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999							

TABLE A.2

t Distribution: Critical Values of t

<i>Degrees of freedom</i>	<i>Two-tailed test: One-tailed test:</i>	<i>Significance level</i>					
		10% 5%	5% 2.5%	2% 1%	1% 0.5%	0.2% 0.1%	0.1% 0.05%
1		6.314	12.706	31.821	63.657	318.309	636.619
2		2.920	4.303	6.965	9.925	22.327	31.599
3		2.353	3.182	4.541	5.841	10.215	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
6		1.943	2.447	3.143	3.707	5.208	5.959
7		1.894	2.365	2.998	3.499	4.785	5.408
8		1.860	2.306	2.896	3.355	4.501	5.041
9		1.833	2.262	2.821	3.250	4.297	4.781
10		1.812	2.228	2.764	3.169	4.144	4.587
11		1.796	2.201	2.718	3.106	4.025	4.437
12		1.782	2.179	2.681	3.055	3.930	4.318
13		1.771	2.160	2.650	3.012	3.852	4.221
14		1.761	2.145	2.624	2.977	3.787	4.140
15		1.753	2.131	2.602	2.947	3.733	4.073
16		1.746	2.120	2.583	2.921	3.686	4.015
17		1.740	2.110	2.567	2.898	3.646	3.965
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850
21		1.721	2.080	2.518	2.831	3.527	3.819
22		1.717	2.074	2.508	2.819	3.505	3.792
23		1.714	2.069	2.500	2.807	3.485	3.768
24		1.711	2.064	2.492	2.797	3.467	3.745
25		1.708	2.060	2.485	2.787	3.450	3.725
26		1.706	2.056	2.479	2.779	3.435	3.707
27		1.703	2.052	2.473	2.771	3.421	3.690
28		1.701	2.048	2.467	2.763	3.408	3.674
29		1.699	2.045	2.462	2.756	3.396	3.659
30		1.697	2.042	2.457	2.750	3.385	3.646
32		1.694	2.037	2.449	2.738	3.365	3.622
34		1.691	2.032	2.441	2.728	3.348	3.601
36		1.688	2.028	2.434	2.719	3.333	3.582
38		1.686	2.024	2.429	2.712	3.319	3.566
40		1.684	2.021	2.423	2.704	3.307	3.551
42		1.682	2.018	2.418	2.698	3.296	3.538
44		1.680	2.015	2.414	2.692	3.286	3.526
46		1.679	2.013	2.410	2.687	3.277	3.515
48		1.677	2.011	2.407	2.682	3.269	3.505
50		1.676	2.009	2.403	2.678	3.261	3.496
60		1.671	2.000	2.390	2.660	3.232	3.460
70		1.667	1.994	2.381	2.648	3.211	3.435
80		1.664	1.990	2.374	2.639	3.195	3.416
90		1.662	1.987	2.368	2.632	3.183	3.402
100		1.660	1.984	2.364	2.626	3.174	3.390
120		1.658	1.980	2.358	2.617	3.160	3.373
150		1.655	1.976	2.351	2.609	3.145	3.357
200		1.653	1.972	2.345	2.601	3.131	3.340
300		1.650	1.968	2.339	2.592	3.118	3.323
400		1.649	1.966	2.336	2.588	3.111	3.315
500		1.648	1.965	2.334	2.586	3.107	3.310
600		1.647	1.964	2.333	2.584	3.104	3.307
∞		1.645	1.960	2.326	2.576	3.090	3.291

TABLE A.3

F Distribution: Critical Values of F (5% significance level)

ν_1	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20
ν_2															
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.36	246.46	247.32	248.01
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.42	19.43	19.44	19.45
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.71	8.69	8.67	8.66
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.87	5.84	5.82	5.80
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.64	4.60	4.58	4.56
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.96	3.92	3.90	3.87
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.53	3.49	3.47	3.44
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.24	3.20	3.17	3.15
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.03	2.99	2.96	2.94
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.86	2.83	2.80	2.77
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.74	2.70	2.67	2.65
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.64	2.60	2.57	2.54
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.55	2.51	2.48	2.46
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.48	2.44	2.41	2.39
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.42	2.38	2.35	2.33
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.37	2.33	2.30	2.28
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.33	2.29	2.26	2.23
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.29	2.25	2.22	2.19
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.26	2.21	2.18	2.16
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.22	2.18	2.15	2.12
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.20	2.16	2.12	2.10
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.17	2.13	2.10	2.07
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.15	2.11	2.08	2.05
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.13	2.09	2.05	2.03
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.11	2.07	2.04	2.01
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.09	2.05	2.02	1.99
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.08	2.04	2.00	1.97
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.06	2.02	1.99	1.96
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.05	2.01	1.97	1.94
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.04	1.99	1.96	1.93
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	2.04	1.99	1.94	1.91	1.88
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.95	1.90	1.87	1.84
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.95	1.89	1.85	1.81	1.78
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.86	1.82	1.78	1.75
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97	1.89	1.84	1.79	1.75	1.72
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.88	1.82	1.77	1.73	1.70
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94	1.86	1.80	1.76	1.72	1.69
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.85	1.79	1.75	1.71	1.68
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.78	1.73	1.69	1.66
150	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.82	1.76	1.71	1.67	1.64
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.80	1.74	1.69	1.66	1.62
250	3.88	3.03	2.64	2.41	2.25	2.13	2.05	1.98	1.92	1.87	1.79	1.73	1.68	1.65	1.61
300	3.87	3.03	2.63	2.40	2.24	2.13	2.04	1.97	1.91	1.86	1.78	1.72	1.68	1.64	1.61
400	3.86	3.02	2.63	2.39	2.24	2.12	2.03	1.96	1.90	1.85	1.78	1.72	1.67	1.63	1.60
500	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.77	1.71	1.66	1.62	1.59
600	3.86	3.01	2.62	2.39	2.23	2.11	2.02	1.95	1.90	1.85	1.77	1.71	1.66	1.62	1.59
750	3.85	3.01	2.62	2.38	2.23	2.11	2.02	1.95	1.89	1.84	1.77	1.70	1.66	1.62	1.58
1000	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.76	1.70	1.65	1.61	1.58

TABLE A.3 (continued)

F Distribution: Critical Values of F (5% significance level)

ν_1	25	30	35	40	50	60	75	100	150	200
ν_2										
1	249.26	250.10	250.69	251.14	251.77	252.20	252.62	253.04	253.46	253.68
2	19.46	19.46	19.47	19.47	19.48	19.48	19.48	19.49	19.49	19.49
3	8.63	8.62	8.60	8.59	8.58	8.57	8.56	8.55	8.54	8.54
4	5.77	5.75	5.73	5.72	5.70	5.69	5.68	5.66	5.65	5.65
5	4.52	4.50	4.48	4.46	4.44	4.43	4.42	4.41	4.39	4.39
6	3.83	3.81	3.79	3.77	3.75	3.74	3.73	3.71	3.70	3.69
7	3.40	3.38	3.36	3.34	3.32	3.30	3.29	3.27	3.26	3.25
8	3.11	3.08	3.06	3.04	3.02	3.01	2.99	2.97	2.96	2.95
9	2.89	2.86	2.84	2.83	2.80	2.79	2.77	2.76	2.74	2.73
10	2.73	2.70	2.68	2.66	2.64	2.62	2.60	2.59	2.57	2.56
11	2.60	2.57	2.55	2.53	2.51	2.49	2.47	2.46	2.44	2.43
12	2.50	2.47	2.44	2.43	2.40	2.38	2.37	2.35	2.33	2.32
13	2.41	2.38	2.36	2.34	2.31	2.30	2.28	2.26	2.24	2.23
14	2.34	2.31	2.28	2.27	2.24	2.22	2.21	2.19	2.17	2.16
15	2.28	2.25	2.22	2.20	2.18	2.16	2.14	2.12	2.10	2.10
16	2.23	2.19	2.17	2.15	2.12	2.11	2.09	2.07	2.05	2.04
17	2.18	2.15	2.12	2.10	2.08	2.06	2.04	2.02	2.00	1.99
18	2.14	2.11	2.08	2.06	2.04	2.02	2.00	1.98	1.96	1.95
19	2.11	2.07	2.05	2.03	2.00	1.98	1.96	1.94	1.92	1.91
20	2.07	2.04	2.01	1.99	1.97	1.95	1.93	1.91	1.89	1.88
21	2.05	2.01	1.98	1.96	1.94	1.92	1.90	1.88	1.86	1.84
22	2.02	1.98	1.96	1.94	1.91	1.89	1.87	1.85	1.83	1.82
23	2.00	1.96	1.93	1.91	1.88	1.86	1.84	1.82	1.80	1.79
24	1.97	1.94	1.91	1.89	1.86	1.84	1.82	1.80	1.78	1.77
25	1.96	1.92	1.89	1.87	1.84	1.82	1.80	1.78	1.76	1.75
26	1.94	1.90	1.87	1.85	1.82	1.80	1.78	1.76	1.74	1.73
27	1.92	1.88	1.86	1.84	1.81	1.79	1.76	1.74	1.72	1.71
28	1.91	1.87	1.84	1.82	1.79	1.77	1.75	1.73	1.70	1.69
29	1.89	1.85	1.83	1.81	1.77	1.75	1.73	1.71	1.69	1.67
30	1.88	1.84	1.81	1.79	1.76	1.74	1.72	1.70	1.67	1.66
35	1.82	1.79	1.76	1.74	1.70	1.68	1.66	1.63	1.61	1.60
40	1.78	1.74	1.72	1.69	1.66	1.64	1.61	1.59	1.56	1.55
50	1.73	1.69	1.66	1.63	1.60	1.58	1.55	1.52	1.50	1.48
60	1.69	1.65	1.62	1.59	1.56	1.53	1.51	1.48	1.45	1.44
70	1.66	1.62	1.59	1.57	1.53	1.50	1.48	1.45	1.42	1.40
80	1.64	1.60	1.57	1.54	1.51	1.48	1.45	1.43	1.39	1.38
90	1.63	1.59	1.55	1.53	1.49	1.46	1.44	1.41	1.38	1.36
100	1.62	1.57	1.54	1.52	1.48	1.45	1.42	1.39	1.36	1.34
120	1.60	1.55	1.52	1.50	1.46	1.43	1.40	1.37	1.33	1.32
150	1.58	1.54	1.50	1.48	1.44	1.41	1.38	1.34	1.31	1.29
200	1.56	1.52	1.48	1.46	1.41	1.39	1.35	1.32	1.28	1.26
250	1.55	1.50	1.47	1.44	1.40	1.37	1.34	1.31	1.27	1.25
300	1.54	1.50	1.46	1.43	1.39	1.36	1.33	1.30	1.26	1.23
400	1.53	1.49	1.45	1.42	1.38	1.35	1.32	1.28	1.24	1.22
500	1.53	1.48	1.45	1.42	1.38	1.35	1.31	1.28	1.23	1.21
600	1.52	1.48	1.44	1.41	1.37	1.34	1.31	1.27	1.23	1.20
750	1.52	1.47	1.44	1.41	1.37	1.34	1.30	1.26	1.22	1.20
1000	1.52	1.47	1.43	1.41	1.36	1.33	1.30	1.26	1.22	1.19

TABLE A.3 (continued)

F Distribution: Critical Values of F (1% significance level)

ν_1	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20
ν_2															
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47	6055.85	6106.32	6142.67	6170.10	6191.53	6208.73
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.44	99.44	99.45
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.92	26.83	26.75	26.69
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.25	14.15	14.08	14.02
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.77	9.68	9.61	9.55
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.60	7.52	7.45	7.40
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.36	6.28	6.21	6.16
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.56	5.48	5.41	5.36
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	5.01	4.92	4.86	4.81
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.60	4.52	4.46	4.41
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.29	4.21	4.15	4.10
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.05	3.97	3.91	3.86
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.86	3.78	3.72	3.66
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.70	3.62	3.56	3.51
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.56	3.49	3.42	3.37
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.45	3.37	3.31	3.26
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.35	3.27	3.21	3.16
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.27	3.19	3.13	3.08
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.19	3.12	3.05	3.00
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.13	3.05	2.99	2.94
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.07	2.99	2.93	2.88
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	3.02	2.94	2.88	2.83
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.97	2.89	2.83	2.78
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.93	2.85	2.79	2.74
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.89	2.81	2.75	2.70
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.86	2.78	2.72	2.66
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.82	2.75	2.68	2.63
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.79	2.72	2.65	2.60
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.77	2.69	2.63	2.57
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.74	2.66	2.60	2.55
35	7.42	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96	2.88	2.74	2.64	2.56	2.50	2.44
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.56	2.48	2.42	2.37
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.56	2.46	2.38	2.32	2.27
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.39	2.31	2.25	2.20
70	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67	2.59	2.45	2.35	2.27	2.20	2.15
80	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55	2.42	2.31	2.23	2.17	2.12
90	6.93	4.85	4.01	3.53	3.23	3.01	2.84	2.72	2.61	2.52	2.39	2.29	2.21	2.14	2.09
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.37	2.27	2.19	2.12	2.07
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.23	2.15	2.09	2.03
150	6.81	4.75	3.91	3.45	3.14	2.92	2.76	2.63	2.53	2.44	2.31	2.20	2.12	2.06	2.00
200	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.27	2.17	2.09	2.03	1.97
250	6.74	4.69	3.86	3.40	3.09	2.87	2.71	2.58	2.48	2.39	2.26	2.15	2.07	2.01	1.95
300	6.72	4.68	3.85	3.38	3.08	2.86	2.70	2.57	2.47	2.38	2.24	2.14	2.06	1.99	1.94
400	6.70	4.66	3.83	3.37	3.06	2.85	2.68	2.56	2.45	2.37	2.23	2.13	2.05	1.98	1.92
500	6.69	4.65	3.82	3.36	3.05	2.84	2.68	2.55	2.44	2.36	2.22	2.12	2.04	1.97	1.92
600	6.68	4.64	3.81	3.35	3.05	2.83	2.67	2.54	2.44	2.35	2.21	2.11	2.03	1.96	1.91
750	6.67	4.63	3.81	3.34	3.04	2.83	2.66	2.53	2.43	2.34	2.21	2.11	2.02	1.96	1.90
1000	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.34	2.20	2.10	2.02	1.95	1.90

TABLE A.3 (continued)

F Distribution: Critical Values of F (1% significance level)

ν_1	25	30	35	40	50	60	75	100	150	200
ν_2										
1	6239.83	6260.65	6275.57	6286.78	6302.52	6313.03	6323.56	6334.11	6344.68	6349.97
2	99.46	99.47	99.47	99.47	99.48	99.48	99.49	99.49	99.49	99.49
3	26.58	26.50	26.45	26.41	26.35	26.32	26.28	26.24	26.20	26.18
4	13.91	13.84	13.79	13.75	13.69	13.65	13.61	13.58	13.54	13.52
5	9.45	9.38	9.33	9.29	9.24	9.20	9.17	9.13	9.09	9.08
6	7.30	7.23	7.18	7.14	7.09	7.06	7.02	6.99	6.95	6.93
7	6.06	5.99	5.94	5.91	5.86	5.82	5.79	5.75	5.72	5.70
8	5.26	5.20	5.15	5.12	5.07	5.03	5.00	4.96	4.93	4.91
9	4.71	4.65	4.60	4.57	4.52	4.48	4.45	4.41	4.38	4.36
10	4.31	4.25	4.20	4.17	4.12	4.08	4.05	4.01	3.98	3.96
11	4.01	3.94	3.89	3.86	3.81	3.78	3.74	3.71	3.67	3.66
12	3.76	3.70	3.65	3.62	3.57	3.54	3.50	3.47	3.43	3.41
13	3.57	3.51	3.46	3.43	3.38	3.34	3.31	3.27	3.24	3.22
14	3.41	3.35	3.30	3.27	3.22	3.18	3.15	3.11	3.08	3.06
15	3.28	3.21	3.17	3.13	3.08	3.05	3.01	2.98	2.94	2.92
16	3.16	3.10	3.05	3.02	2.97	2.93	2.90	2.86	2.83	2.81
17	3.07	3.00	2.96	2.92	2.87	2.83	2.80	2.76	2.73	2.71
18	2.98	2.92	2.87	2.84	2.78	2.75	2.71	2.68	2.64	2.62
19	2.91	2.84	2.80	2.76	2.71	2.67	2.64	2.60	2.57	2.55
20	2.84	2.78	2.73	2.69	2.64	2.61	2.57	2.54	2.50	2.48
21	2.79	2.72	2.67	2.64	2.58	2.55	2.51	2.48	2.44	2.42
22	2.73	2.67	2.62	2.58	2.53	2.50	2.46	2.42	2.38	2.36
23	2.69	2.62	2.57	2.54	2.48	2.45	2.41	2.37	2.34	2.32
24	2.64	2.58	2.53	2.49	2.44	2.40	2.37	2.33	2.29	2.27
25	2.60	2.54	2.49	2.45	2.40	2.36	2.33	2.29	2.25	2.23
26	2.57	2.50	2.45	2.42	2.36	2.33	2.29	2.25	2.21	2.19
27	2.54	2.47	2.42	2.38	2.33	2.29	2.26	2.22	2.18	2.16
28	2.51	2.44	2.39	2.35	2.30	2.26	2.23	2.19	2.15	2.13
29	2.48	2.41	2.36	2.33	2.27	2.23	2.20	2.16	2.12	2.10
30	2.45	2.39	2.34	2.30	2.25	2.21	2.17	2.13	2.09	2.07
35	2.35	2.28	2.23	2.19	2.14	2.10	2.06	2.02	1.98	1.96
40	2.27	2.20	2.15	2.11	2.06	2.02	1.98	1.94	1.90	1.87
50	2.17	2.10	2.05	2.01	1.95	1.91	1.87	1.82	1.78	1.76
60	2.10	2.03	1.98	1.94	1.88	1.84	1.79	1.75	1.70	1.68
70	2.05	1.98	1.93	1.89	1.83	1.78	1.74	1.70	1.65	1.62
80	2.01	1.94	1.89	1.85	1.79	1.75	1.70	1.65	1.61	1.58
90	1.99	1.92	1.86	1.82	1.76	1.72	1.67	1.62	1.57	1.55
100	1.97	1.89	1.84	1.80	1.74	1.69	1.65	1.60	1.55	1.52
120	1.93	1.86	1.81	1.76	1.70	1.66	1.61	1.56	1.51	1.48
150	1.90	1.83	1.77	1.73	1.66	1.62	1.57	1.52	1.46	1.43
200	1.87	1.79	1.74	1.69	1.63	1.58	1.53	1.48	1.42	1.39
250	1.85	1.77	1.72	1.67	1.61	1.56	1.51	1.46	1.40	1.36
300	1.84	1.76	1.70	1.66	1.59	1.55	1.50	1.44	1.38	1.35
400	1.82	1.75	1.69	1.64	1.58	1.53	1.48	1.42	1.36	1.32
500	1.81	1.74	1.68	1.63	1.57	1.52	1.47	1.41	1.34	1.31
600	1.80	1.73	1.67	1.63	1.56	1.51	1.46	1.40	1.34	1.30
750	1.80	1.72	1.66	1.62	1.55	1.50	1.45	1.39	1.33	1.29
1000	1.79	1.72	1.66	1.61	1.54	1.50	1.44	1.38	1.32	1.28

TABLE A.3 (continued)

F Distribution: Critical Values of F (0.1% significance level)

ν_1	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20
ν_2															
1	4.05e05	5.00e05	5.40e05	5.62e05	5.76e05	5.86e05	5.93e05	5.98e05	6.02e05	6.06e05	6.11e05	6.14e05	6.17e05	6.19e05	6.21e05
2	998.50	999.00	999.17	999.25	999.30	999.33	999.36	999.37	999.39	999.40	999.42	999.43	999.44	999.44	999.45
3	167.03	148.50	141.11	137.10	134.58	132.85	131.58	130.62	129.86	129.25	128.32	127.64	127.14	126.74	126.42
4	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	48.47	48.05	47.41	46.95	46.60	46.32	46.10
5	47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65	27.24	26.92	26.42	26.06	25.78	25.57	25.39
6	35.51	27.00	23.70	21.92	20.80	20.03	19.46	19.03	18.69	18.41	17.99	17.68	17.45	17.27	17.12
7	29.25	21.69	18.77	17.20	16.21	15.52	15.02	14.63	14.33	14.08	13.71	13.43	13.23	13.06	12.93
8	25.41	18.49	15.83	14.39	13.48	12.86	12.40	12.05	11.77	11.54	11.19	10.94	10.75	10.60	10.48
9	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	10.11	9.89	9.57	9.33	9.15	9.01	8.90
10	21.04	14.91	12.55	11.28	10.48	9.93	9.52	9.20	8.96	8.75	8.45	8.22	8.05	7.91	7.80
11	19.69	13.81	11.56	10.35	9.58	9.05	8.66	8.35	8.12	7.92	7.63	7.41	7.24	7.11	7.01
12	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.48	7.29	7.00	6.79	6.63	6.51	6.40
13	17.82	12.31	10.21	9.07	8.35	7.86	7.49	7.21	6.98	6.80	6.52	6.31	6.16	6.03	5.93
14	17.14	11.78	9.73	8.62	7.92	7.44	7.08	6.80	6.58	6.40	6.13	5.93	5.78	5.66	5.56
15	16.59	11.34	9.34	8.25	7.57	7.09	6.74	6.47	6.26	6.08	5.81	5.62	5.46	5.35	5.25
16	16.12	10.97	9.01	7.94	7.27	6.80	6.46	6.19	5.98	5.81	5.55	5.35	5.20	5.09	4.99
17	15.72	10.66	8.73	7.68	7.02	6.56	6.22	5.96	5.75	5.58	5.32	5.13	4.99	4.87	4.78
18	15.38	10.39	8.49	7.46	6.81	6.35	6.02	5.76	5.56	5.39	5.13	4.94	4.80	4.68	4.59
19	15.08	10.16	8.28	7.27	6.62	6.18	5.85	5.59	5.39	5.22	4.97	4.78	4.64	4.52	4.43
20	14.82	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24	5.08	4.82	4.64	4.49	4.38	4.29
21	14.59	9.77	7.94	6.95	6.32	5.88	5.56	5.31	5.11	4.95	4.70	4.51	4.37	4.26	4.17
22	14.38	9.61	7.80	6.81	6.19	5.76	5.44	5.19	4.99	4.83	4.58	4.40	4.26	4.15	4.06
23	14.20	9.47	7.67	6.70	6.08	5.65	5.33	5.09	4.89	4.73	4.48	4.30	4.16	4.05	3.96
24	14.03	9.34	7.55	6.59	5.98	5.55	5.23	4.99	4.80	4.64	4.39	4.21	4.07	3.96	3.87
25	13.88	9.22	7.45	6.49	5.89	5.46	5.15	4.91	4.71	4.56	4.31	4.13	3.99	3.88	3.79
26	13.74	9.12	7.36	6.41	5.80	5.38	5.07	4.83	4.64	4.48	4.24	4.06	3.92	3.81	3.72
27	13.61	9.02	7.27	6.33	5.73	5.31	5.00	4.76	4.57	4.41	4.17	3.99	3.86	3.75	3.66
28	13.50	8.93	7.19	6.25	5.66	5.24	4.93	4.69	4.50	4.35	4.11	3.93	3.80	3.69	3.60
29	13.39	8.85	7.12	6.19	5.59	5.18	4.87	4.64	4.45	4.29	4.05	3.88	3.74	3.63	3.54
30	13.29	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39	4.24	4.00	3.82	3.69	3.58	3.49
35	12.90	8.47	6.79	5.88	5.30	4.89	4.59	4.36	4.18	4.03	3.79	3.62	3.48	3.38	3.29
40	12.61	8.25	6.59	5.70	5.13	4.73	4.44	4.21	4.02	3.87	3.64	3.47	3.34	3.23	3.14
50	12.22	7.96	6.34	5.46	4.90	4.51	4.22	4.00	3.82	3.67	3.44	3.27	3.14	3.04	2.95
60	11.97	7.77	6.17	5.31	4.76	4.37	4.09	3.86	3.69	3.54	3.32	3.15	3.02	2.91	2.83
70	11.80	7.64	6.06	5.20	4.66	4.28	3.99	3.77	3.60	3.45	3.23	3.06	2.93	2.83	2.74
80	11.67	7.54	5.97	5.12	4.58	4.20	3.92	3.70	3.53	3.39	3.16	3.00	2.87	2.76	2.68
90	11.57	7.47	5.91	5.06	4.53	4.15	3.87	3.65	3.48	3.34	3.11	2.95	2.82	2.71	2.63
100	11.50	7.41	5.86	5.02	4.48	4.11	3.83	3.61	3.44	3.30	3.07	2.91	2.78	2.68	2.59
120	11.38	7.32	5.78	4.95	4.42	4.04	3.77	3.55	3.38	3.24	3.02	2.85	2.72	2.62	2.53
150	11.27	7.24	5.71	4.88	4.35	3.98	3.71	3.49	3.32	3.18	2.96	2.80	2.67	2.56	2.48
200	11.15	7.15	5.63	4.81	4.29	3.92	3.65	3.43	3.26	3.12	2.90	2.74	2.61	2.51	2.42
250	11.09	7.10	5.59	4.77	4.25	3.88	3.61	3.40	3.23	3.09	2.87	2.71	2.58	2.48	2.39
300	11.04	7.07	5.56	4.75	4.22	3.86	3.59	3.38	3.21	3.07	2.85	2.69	2.56	2.46	2.37
400	10.99	7.03	5.53	4.71	4.19	3.83	3.56	3.35	3.18	3.04	2.82	2.66	2.53	2.43	2.34
500	10.96	7.00	5.51	4.69	4.18	3.81	3.54	3.33	3.16	3.02	2.81	2.64	2.52	2.41	2.33
600	10.94	6.99	5.49	4.68	4.16	3.80	3.53	3.32	3.15	3.01	2.80	2.63	2.51	2.40	2.32
750	10.91	6.97	5.48	4.67	4.15	3.79	3.52	3.31	3.14	3.00	2.78	2.62	2.49	2.39	2.31
1000	10.89	6.96	5.46	4.65	4.14	3.78	3.51	3.30	3.13	2.99	2.77	2.61	2.48	2.38	2.30

TABLE A.3 (continued)

F Distribution: Critical Values of F (0.1% significance level)

ν_1	25	30	35	40	50	60	75	100	150	200
ν_2										
1	6.24e05	6.26e05	6.28e05	6.29e05	6.30e05	6.31e05	6.32e05	6.33e05	6.35e05	6.35e05
2	999.46	999.47	999.47	999.47	999.48	999.48	999.49	999.49	999.49	999.49
3	125.84	125.45	125.17	124.96	124.66	124.47	124.27	124.07	123.87	123.77
4	45.70	45.43	45.23	45.09	44.88	44.75	44.61	44.47	44.33	44.26
5	25.08	24.87	24.72	24.60	24.44	24.33	24.22	24.12	24.01	23.95
6	16.85	16.67	16.54	16.44	16.31	16.21	16.12	16.03	15.93	15.89
7	12.69	12.53	12.41	12.33	12.20	12.12	12.04	11.95	11.87	11.82
8	10.26	10.11	10.00	9.92	9.80	9.73	9.65	9.57	9.49	9.45
9	8.69	8.55	8.46	8.37	8.26	8.19	8.11	8.04	7.96	7.93
10	7.60	7.47	7.37	7.30	7.19	7.12	7.05	6.98	6.91	6.87
11	6.81	6.68	6.59	6.52	6.42	6.35	6.28	6.21	6.14	6.10
12	6.22	6.09	6.00	5.93	5.83	5.76	5.70	5.63	5.56	5.52
13	5.75	5.63	5.54	5.47	5.37	5.30	5.24	5.17	5.10	5.07
14	5.38	5.25	5.17	5.10	5.00	4.94	4.87	4.81	4.74	4.71
15	5.07	4.95	4.86	4.80	4.70	4.64	4.57	4.51	4.44	4.41
16	4.82	4.70	4.61	4.54	4.45	4.39	4.32	4.26	4.19	4.16
17	4.60	4.48	4.40	4.33	4.24	4.18	4.11	4.05	3.98	3.95
18	4.42	4.30	4.22	4.15	4.06	4.00	3.93	3.87	3.80	3.77
19	4.26	4.14	4.06	3.99	3.90	3.84	3.78	3.71	3.65	3.61
20	4.12	4.00	3.92	3.86	3.77	3.70	3.64	3.58	3.51	3.48
21	4.00	3.88	3.80	3.74	3.64	3.58	3.52	3.46	3.39	3.36
22	3.89	3.78	3.70	3.63	3.54	3.48	3.41	3.35	3.28	3.25
23	3.79	3.68	3.60	3.53	3.44	3.38	3.32	3.25	3.19	3.16
24	3.71	3.59	3.51	3.45	3.36	3.29	3.23	3.17	3.10	3.07
25	3.63	3.52	3.43	3.37	3.28	3.22	3.15	3.09	3.03	2.99
26	3.56	3.44	3.36	3.30	3.21	3.15	3.08	3.02	2.95	2.92
27	3.49	3.38	3.30	3.23	3.14	3.08	3.02	2.96	2.89	2.86
28	3.43	3.32	3.24	3.18	3.09	3.02	2.96	2.90	2.83	2.80
29	3.38	3.27	3.18	3.12	3.03	2.97	2.91	2.84	2.78	2.74
30	3.33	3.22	3.13	3.07	2.98	2.92	2.86	2.79	2.73	2.69
35	3.13	3.02	2.93	2.87	2.78	2.72	2.66	2.59	2.52	2.49
40	2.98	2.87	2.79	2.73	2.64	2.57	2.51	2.44	2.38	2.34
50	2.79	2.68	2.60	2.53	2.44	2.38	2.31	2.25	2.18	2.14
60	2.67	2.55	2.47	2.41	2.32	2.25	2.19	2.12	2.05	2.01
70	2.58	2.47	2.39	2.32	2.23	2.16	2.10	2.03	1.95	1.92
80	2.52	2.41	2.32	2.26	2.16	2.10	2.03	1.96	1.89	1.85
90	2.47	2.36	2.27	2.21	2.11	2.05	1.98	1.91	1.83	1.79
100	2.43	2.32	2.24	2.17	2.08	2.01	1.94	1.87	1.79	1.75
120	2.37	2.26	2.18	2.11	2.02	1.95	1.88	1.81	1.73	1.68
150	2.32	2.21	2.12	2.06	1.96	1.89	1.82	1.74	1.66	1.62
200	2.26	2.15	2.07	2.00	1.90	1.83	1.76	1.68	1.60	1.55
250	2.23	2.12	2.03	1.97	1.87	1.80	1.72	1.65	1.56	1.51
300	2.21	2.10	2.01	1.94	1.85	1.78	1.70	1.62	1.53	1.48
400	2.18	2.07	1.98	1.92	1.82	1.75	1.67	1.59	1.50	1.45
500	2.17	2.05	1.97	1.90	1.80	1.73	1.65	1.57	1.48	1.43
600	2.16	2.04	1.96	1.89	1.79	1.72	1.64	1.56	1.46	1.41
750	2.15	2.03	1.95	1.88	1.78	1.71	1.63	1.55	1.45	1.40
1000	2.14	2.02	1.94	1.87	1.77	1.69	1.62	1.53	1.44	1.38

TABLE A.4

 χ^2 (Chi-Squared) Distribution: Critical Values of χ^2

<i>Degrees of freedom</i>	<i>Significance level</i>		
	5%	1%	0.1%
1	3.841	6.635	10.828
2	5.991	9.210	13.816
3	7.815	11.345	16.266
4	9.488	13.277	18.467
5	11.070	15.086	20.515
6	12.592	16.812	22.458
7	14.067	18.475	24.322
8	15.507	20.090	26.124
9	16.919	21.666	27.877
10	18.307	23.209	29.588

BẢNG TRA PHÂN BỐ Q CỦA TUKEY

Studentized Range q Table

The following tables provide the critical values for $q(k, df, \alpha)$ when $\alpha = 0.10, 0.05, 0.025, 0.01, 0.005$ and $\alpha = 0.001$. See [Unplanned Comparisons for ANOVA](#) for more details.

Studentized T Table by Excel

Alpha 0.10

	k-->																			
df	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	8.929	13.437	16.358	18.488	20.150	21.504	22.642	23.621	24.477	25.237	25.918	26.536	27.100	27.618	28.097	28.542	28.958	29.347	29.713	
2	4.129	5.733	6.772	7.538	8.139	8.633	9.049	9.409	9.725	10.006	10.259	10.488	10.698	10.891	11.070	11.237	11.392	11.538	11.676	
3	3.328	4.467	5.199	5.738	6.162	6.511	6.806	7.062	7.287	7.487	7.667	7.831	7.982	8.120	8.248	8.368	8.479	8.584	8.683	
4	3.015	3.976	4.586	5.035	5.388	5.679	5.926	6.139	6.327	6.494	6.645	6.783	6.909	7.025	7.132	7.233	7.326	7.414	7.497	
5	2.850	3.717	4.264	4.664	4.979	5.238	5.458	5.648	5.816	5.965	6.100	6.223	6.336	6.439	6.536	6.626	6.710	6.788	6.863	
6	2.748	3.558	4.065	4.435	4.726	4.966	5.168	5.344	5.499	5.637	5.762	5.875	5.979	6.075	6.164	6.247	6.325	6.398	6.466	
7	2.679	3.451	3.931	4.280	4.555	4.780	4.971	5.137	5.283	5.413	5.530	5.637	5.735	5.826	5.910	5.988	6.061	6.130	6.195	
8	2.630	3.374	3.834	4.169	4.431	4.646	4.829	4.987	5.126	5.250	5.362	5.464	5.558	5.644	5.724	5.799	5.869	5.935	5.997	
9	2.592	3.316	3.761	4.084	4.337	4.545	4.721	4.873	5.007	5.126	5.234	5.333	5.423	5.506	5.583	5.655	5.722	5.786	5.845	
10	2.563	3.270	3.704	4.018	4.264	4.465	4.636	4.783	4.913	5.029	5.134	5.229	5.316	5.397	5.472	5.542	5.607	5.668	5.726	
11	2.540	3.234	3.658	3.965	4.205	4.401	4.567	4.711	4.838	4.951	5.053	5.145	5.231	5.309	5.382	5.450	5.514	5.573	5.630	
12	2.521	3.204	3.621	3.921	4.156	4.349	4.511	4.652	4.776	4.886	4.986	5.076	5.160	5.236	5.308	5.374	5.436	5.495	5.550	
13	2.504	3.179	3.589	3.885	4.116	4.304	4.464	4.602	4.724	4.832	4.930	5.019	5.100	5.175	5.245	5.310	5.371	5.429	5.483	
14	2.491	3.158	3.563	3.854	4.081	4.267	4.424	4.560	4.679	4.786	4.882	4.969	5.050	5.124	5.192	5.256	5.316	5.372	5.426	
15	2.479	3.140	3.540	3.828	4.052	4.235	4.390	4.524	4.641	4.746	4.841	4.927	5.006	5.079	5.146	5.209	5.268	5.324	5.376	
16	2.469	3.124	3.520	3.804	4.026	4.207	4.360	4.492	4.608	4.712	4.805	4.890	4.968	5.040	5.106	5.169	5.227	5.282	5.333	
17	2.460	3.110	3.503	3.784	4.003	4.182	4.334	4.464	4.579	4.681	4.774	4.857	4.934	5.005	5.071	5.133	5.190	5.244	5.295	
18	2.452	3.098	3.487	3.766	3.984	4.161	4.310	4.440	4.553	4.654	4.746	4.829	4.905	4.975	5.040	5.101	5.158	5.211	5.262	
19	2.445	3.087	3.474	3.751	3.966	4.142	4.290	4.418	4.530	4.630	4.721	4.803	4.878	4.948	5.012	5.072	5.129	5.182	5.232	
20	2.439	3.077	3.462	3.736	3.950	4.124	4.271	4.398	4.510	4.609	4.699	4.780	4.855	4.923	4.987	5.047	5.103	5.155	5.205	
21	2.433	3.069	3.451	3.724	3.936	4.109	4.255	4.380	4.491	4.590	4.678	4.759	4.833	4.901	4.965	5.024	5.079	5.131	5.180	
22	2.428	3.061	3.441	3.712	3.923	4.095	4.239	4.364	4.474	4.572	4.660	4.740	4.814	4.882	4.944	5.003	5.058	5.109	5.158	
23	2.424	3.054	3.432	3.701	3.911	4.082	4.226	4.350	4.459	4.556	4.644	4.723	4.796	4.863	4.926	4.984	5.038	5.089	5.138	
24	2.420	3.047	3.423	3.692	3.900	4.070	4.213	4.336	4.445	4.541	4.628	4.707	4.780	4.847	4.909	4.966	5.020	5.071	5.119	
25	2.416	3.041	3.416	3.683	3.890	4.059	4.201	4.324	4.432	4.528	4.614	4.693	4.765	4.831	4.893	4.950	5.004	5.055	5.102	
df	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
26	2.412	3.036	3.409	3.675	3.881	4.049	4.191	4.313	4.420	4.515	4.601	4.680	4.751	4.817	4.878	4.936	4.989	5.039	5.086	
27	2.409	3.030	3.402	3.667	3.873	4.040	4.181	4.302	4.409	4.504	4.590	4.667	4.739	4.804	4.865	4.922	4.975	5.025	5.072	
28	2.406	3.026	3.396	3.660	3.865	4.032	4.172	4.293	4.399	4.493	4.579	4.656	4.727	4.792	4.853	4.909	4.962	5.012	5.058	
29	2.403	3.021	3.391	3.654	3.858	4.024	4.163	4.284	4.389	4.484	4.568	4.645	4.716	4.781	4.841	4.897	4.950	4.999	5.046	
30	2.400	3.017	3.386	3.648	3.851	4.016	4.155	4.275	4.381	4.474	4.559	4.635	4.706	4.770	4.830	4.886	4.939	4.988	5.034	
31	2.398	3.013	3.381	3.642	3.845	4.009	4.148	4.268	4.372	4.466	4.550	4.626	4.696	4.760	4.820	4.876	4.928	4.977	5.023	
32	2.396	3.010	3.376	3.637	3.839	4.003	4.141	4.260	4.365	4.458	4.541	4.617	4.687	4.751	4.811	4.866	4.918	4.967	5.013	
33	2.393	3.006	3.372	3.632	3.833	3.997	4.135	4.253	4.357	4.450	4.533	4.609	4.679	4.743	4.802	4.857	4.909	4.957	5.003	
34	2.391	3.003	3.368	3.627	3.828	3.991	4.129	4.247	4.351	4.443	4.526	4.602	4.671	4.734	4.794	4.849	4.900	4.949	4.994	
35	2.389	3.000	3.364	3.623	3.823	3.986	4.123	4.241	4.344	4.436	4.519	4.594	4.663	4.727	4.786	4.841	4.892	4.940	4.986	
36	2.388	2.998	3.361	3.619	3.819	3.981	4.117	4.235	4.338	4.430	4.512	4.588	4.656	4.720	4.778	4.833	4.884	4.932	4.978	
37	2.386	2.995	3.357	3.615	3.814	3.976	4.112	4.230	4.332	4.424	4.506	4.581	4.650	4.713	4.771	4.826	4.877	4.925	4.970	
38	2.384	2.992	3.354	3.611	3.810	3.972	4.107	4.224	4.327	4.418	4.500	4.575	4.643	4.706	4.765	4.819	4.870	4.918	4.963	
39	2.383	2.990	3.351	3.608	3.806	3.967	4.103	4.220	4.322	4.413	4.495	4.569	4.637	4.700	4.758	4.812	4.863	4.911	4.956	
40	2.381	2.988	3.348	3.605	3.802	3.963	4.099	4.215	4.317	4.408	4.490	4.564	4.632	4.694	4.752	4.806	4.857	4.904	4.949	
48	2.372	2.973	3.330	3.583	3.778	3.937	4.070	4.185	4.285	4.375	4.455	4.528	4.595	4.656	4.713	4.766	4.816	4.863	4.907	
60	2.363	2.959	3.312	3.562	3.755	3.911	4.042	4.155	4.254	4.342	4.421	4.493	4.558	4.619	4.675	4.727	4.775	4.821	4.864	
80	2.353	2.945	3.294	3.541	3.731	3.885	4.014	4.125	4.223	4.309	4.387	4.457	4.521	4.581	4.636	4.687	4.735	4.780	4.822	
120	2.344	2.930	3.276	3.520	3.707	3.859	3.986	4.096	4.191	4.276	4.353	4.422	4.485	4.543	4.597	4.647	4.694	4.738	4.779	
240	2.335	2.916	3.258	3.499	3.684	3.834	3.959	4.066	4.160	4.244	4.319	4.386	4.448	4.505	4.558	4.607	4.653	4.696	4.737	
inf	2.326	2.902	3.240	3.478	3.661	3.808	3.931	4.037	4.129	4.211	4.285	4.351	4.412	4.468	4.519	4.568	4.612	4.654	4.694	

df	20	22	24	26	28	30	32	34	36	38	40	50	60	70	80	90	100
1	29.71	30.39	30.99	31.54	32.04	32.5	32.93	33.33	33.71	34.06	34.38	35.79	36.91	37.83	38.62	39.3	39.91
2	11.68	11.93	12.16	12.36	12.55	12.73	12.89	13.04	13.18	13.31	13.44	13.97	14.40	14.75	15.05	15.31	15.54
3	8.683	8.864	9.029	9.177	9.314	9.440	9.557	9.666	9.768	9.864	9.954	10.34	10.65	10.91	11.12	11.31	11.48
4	7.497	7.650	7.789	7.914	8.029	8.135	8.234	8.326	8.412	8.493	8.569	8.806	9.156	9.373	9.557	9.718	9.860
5	6.863	7.000	7.123	7.236	7.340	7.435	7.523	7.606	7.683	7.756	7.825	8.118	8.353	8.548	8.715	8.859	8.988
6	6.466	6.593	6.708	6.812	6.908	6.996	7.078	7.155	7.227	7.294	7.358	7.630	7.848	8.029	8.184	8.319	8.438
7	6.195	6.315	6.422	6.521	6.611	6.695	6.773	6.845	6.913	6.976	7.036	7.294	7.500	7.672	7.818	7.946	8.059
8	5.997	6.111	6.214	6.308	6.395	6.475	6.549	6.618	6.683	6.744	6.801	7.048	7.245	7.409	7.550	7.672	7.780
9	5.845	5.956	6.055	6.146	6.229	6.306	6.378	6.444	6.507	6.566	6.621	6.859	7.050	7.208	7.343	7.461	7.566
10	5.726	5.833	5.930	6.017	6.098	6.173	6.242	6.307	6.368	6.425	6.479	6.709	6.895	7.048	7.180	7.295	7.396
11	5.630	5.734	5.828	5.914	5.992	6.065	6.132	6.196	6.255	6.310	6.363	6.588	6.768	6.918	7.047	7.158	7.258
12	5.550	5.652	5.744	5.827	5.904	5.976	6.042	6.103	6.161	6.215	6.267	6.487	6.663	6.810	6.936	7.045	7.142
13	5.483	5.583	5.673	5.755	5.830	5.900	5.965	6.025	6.082	6.135	6.186	6.402	6.575	6.719	6.842	6.949	7.045
14	5.426	5.524	5.612	5.693	5.767	5.836	5.899	5.959	6.014	6.067	6.116	6.329	6.499	6.641	6.762	6.868	6.961
15	5.376	5.473	5.560	5.639	5.713	5.780	5.843	5.901	5.956	6.008	6.057	6.266	6.433	6.573	6.692	6.796	6.888
16	5.333	5.428	5.515	5.593	5.665	5.732	5.793	5.851	5.905	5.956	6.004	6.210	6.376	6.513	6.631	6.734	6.825
17	5.295	5.389	5.474	5.552	5.623	5.689	5.750	5.806	5.860	5.910	5.958	6.162	6.325	6.461	6.577	6.679	6.769
18	5.262	5.355	5.439	5.515	5.585	5.650	5.711	5.767	5.820	5.870	5.917	6.113	6.280	6.414	6.529	6.630	6.719
19	5.232	5.324	5.407	5.483	5.552	5.616	5.676	5.732	5.784	5.833	5.880	6.079	6.239	6.372	6.486	6.585	6.674
20	5.205	5.296	5.378	5.453	5.522	5.586	5.645	5.700	5.752	5.801	5.847	6.044	6.203	6.335	6.447	6.546	6.633
24	5.119	5.208	5.287	5.360	5.427	5.489	5.546	5.600	5.650	5.697	5.741	5.933	6.086	6.214	6.324	6.419	6.503
30	5.034	5.120	5.197	5.267	5.332	5.392	5.447	5.499	5.547	5.593	5.636	5.821	5.969	6.093	6.198	6.291	6.372
40	4.949	5.032	5.107	5.174	5.236	5.294	5.347	5.397	5.444	5.488	5.529	5.708	5.850	5.969	6.071	6.160	6.238
60	4.864	4.944	5.015	5.081	5.141	5.196	5.247	5.295	5.340	5.382	5.422	5.593	5.730	5.844	5.941	6.026	6.102
120	4.779	4.656	4.924	4.987	5.044	5.097	5.146	5.192	5.235	5.275	5.313	5.476	5.606	5.715	5.808	5.888	5.960
inf	4.694	4.767	4.832	4.892	4.947	4.997	5.044	5.087	5.128	5.166	5.202	5.357	5.480	5.582	5.669	5.745	5.812

Alpha = 0.05

	k -->																			
df	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	17.969	26.976	32.819	37.082	40.408	43.119	45.397	47.357	49.071	50.592	51.957	53.194	54.323	55.361	56.320	57.212	58.044	58.824	59.558	
2	6.085	8.331	9.798	10.881	11.734	12.435	13.027	13.539	13.988	14.389	14.749	15.076	15.375	15.650	15.905	16.143	16.365	16.573	16.769	
3	4.501	5.910	6.825	7.502	8.037	8.478	8.852	9.177	9.462	9.717	9.946	10.155	10.346	10.522	10.686	10.838	10.980	11.114	11.240	
4	3.926	5.040	5.757	6.287	6.706	7.053	7.347	7.602	7.826	8.027	8.208	8.373	8.524	8.664	8.793	8.914	9.027	9.133	9.233	
5	3.635	4.602	5.218	5.673	6.033	6.330	6.582	6.801	6.995	7.167	7.323	7.466	7.596	7.716	7.828	7.932	8.030	8.122	8.208	
6	3.460	4.339	4.896	5.305	5.628	5.895	6.122	6.319	6.493	6.649	6.789	6.917	7.034	7.143	7.244	7.338	7.426	7.508	7.586	
7	3.344	4.165	4.681	5.060	5.359	5.606	5.815	5.997	6.158	6.302	6.431	6.550	6.658	6.759	6.852	6.939	7.020	7.097	7.169	
8	3.261	4.041	4.529	4.886	5.167	5.399	5.596	5.767	5.918	6.053	6.175	6.287	6.389	6.483	6.571	6.653	6.729	6.801	6.869	
9	3.199	3.948	4.415	4.755	5.024	5.244	5.432	5.595	5.738	5.867	5.983	6.089	6.186	6.276	6.359	6.437	6.510	6.579	6.643	
10	3.151	3.877	4.327	4.654	4.912	5.124	5.304	5.460	5.598	5.722	5.833	5.935	6.028	6.114	6.194	6.269	6.339	6.405	6.467	
11	3.113	3.820	4.256	4.574	4.823	5.028	5.202	5.353	5.486	5.605	5.713	5.811	5.901	5.984	6.062	6.134	6.202	6.265	6.325	
12	3.081	3.773	4.199	4.508	4.750	4.950	5.119	5.265	5.395	5.510	5.615	5.710	5.797	5.878	5.953	6.023	6.089	6.151	6.209	
13	3.055	3.734	4.151	4.453	4.690	4.884	5.049	5.192	5.318	5.431	5.533	5.625	5.711	5.789	5.862	5.931	5.995	6.055	6.112	
14	3.033	3.701	4.111	4.407	4.639	4.829	4.990	5.130	5.253	5.364	5.463	5.554	5.637	5.714	5.785	5.852	5.915	5.973	6.029	
15	3.014	3.673	4.076	4.367	4.595	4.782	4.940	5.077	5.198	5.306	5.403	5.492	5.574	5.649	5.719	5.785	5.846	5.904	5.958	
16	2.998	3.649	4.046	4.333	4.557	4.741	4.896	5.031	5.150	5.256	5.352	5.439	5.519	5.593	5.662	5.726	5.786	5.843	5.896	
17	2.984	3.628	4.020	4.303	4.524	4.705	4.858	4.991	5.108	5.212	5.306	5.392	5.471	5.544	5.612	5.675	5.734	5.790	5.842	
18	2.971	3.609	3.997	4.276	4.494	4.673	4.824	4.955	5.071	5.173	5.266	5.351	5.429	5.501	5.567	5.629	5.688	5.743	5.794	
19	2.960	3.593	3.977	4.253	4.468	4.645	4.794	4.924	5.037	5.139	5.231	5.314	5.391	5.462	5.528	5.589	5.647	5.701	5.752	
20	2.950	3.578	3.958	4.232	4.445	4.620	4.768	4.895	5.008	5.108	5.199	5.282	5.357	5.427	5.492	5.553	5.610	5.663	5.714	
21	2.941	3.565	3.942	4.213	4.424	4.597	4.743	4.870	4.981	5.081	5.170	5.252	5.327	5.396	5.460	5.520	5.576	5.629	5.679	
22	2.933	3.553	3.927	4.196	4.405	4.577	4.722	4.847	4.957	5.056	5.144	5.225	5.299	5.368	5.431	5.491	5.546	5.599	5.648	
23	2.926	3.542	3.914	4.180	4.388	4.558	4.702	4.826	4.935	5.033	5.121	5.201	5.274	5.342	5.405	5.464	5.519	5.571	5.620	
24	2.919	3.532	3.901	4.166	4.373	4.541	4.684	4.807	4.915	5.012	5.099	5.179	5.251	5.319	5.381	5.439	5.494	5.545	5.594	
25	2.913	3.523	3.890	4.153	4.358	4.526	4.667	4.789	4.897	4.993	5.079	5.158	5.230	5.297	5.359	5.417	5.471	5.522	5.570	

df	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
26	2.907	3.514	3.880	4.141	4.345	4.511	4.652	4.773	4.880	4.975	5.061	5.139	5.211	5.277	5.339	5.396	5.450	5.500	5.548
27	2.902	3.506	3.870	4.130	4.333	4.498	4.638	4.758	4.864	4.959	5.044	5.122	5.193	5.259	5.320	5.377	5.430	5.480	5.528
28	2.897	3.499	3.861	4.120	4.322	4.486	4.625	4.745	4.850	4.944	5.029	5.106	5.177	5.242	5.302	5.359	5.412	5.462	5.509
29	2.892	3.493	3.853	4.111	4.311	4.475	4.613	4.732	4.837	4.930	5.014	5.091	5.161	5.226	5.286	5.342	5.395	5.445	5.491
30	2.888	3.486	3.845	4.102	4.301	4.464	4.601	4.720	4.824	4.917	5.001	5.077	5.147	5.211	5.271	5.327	5.379	5.429	5.475
31	2.884	3.481	3.838	4.094	4.292	4.454	4.591	4.709	4.812	4.905	4.988	5.064	5.134	5.198	5.257	5.313	5.365	5.414	5.460
32	2.881	3.475	3.832	4.086	4.284	4.445	4.581	4.698	4.802	4.894	4.976	5.052	5.121	5.185	5.244	5.299	5.351	5.400	5.445
33	2.877	3.470	3.825	4.079	4.276	4.436	4.572	4.689	4.791	4.883	4.965	5.040	5.109	5.173	5.232	5.287	5.338	5.386	5.432
34	2.874	3.465	3.820	4.072	4.268	4.428	4.563	4.680	4.782	4.873	4.955	5.030	5.098	5.161	5.220	5.275	5.326	5.374	5.420
35	2.871	3.461	3.814	4.066	4.261	4.421	4.555	4.671	4.773	4.863	4.945	5.020	5.088	5.151	5.209	5.264	5.315	5.362	5.408
36	2.868	3.457	3.809	4.060	4.255	4.414	4.547	4.663	4.764	4.855	4.936	5.010	5.078	5.141	5.199	5.253	5.304	5.352	5.397
37	2.865	3.453	3.804	4.054	4.249	4.407	4.540	4.655	4.756	4.846	4.927	5.001	5.069	5.131	5.189	5.243	5.294	5.341	5.386
38	2.863	3.449	3.799	4.049	4.243	4.400	4.533	4.648	4.749	4.838	4.919	4.993	5.060	5.122	5.180	5.234	5.284	5.331	5.376
39	2.861	3.445	3.795	4.044	4.237	4.394	4.527	4.641	4.741	4.831	4.911	4.985	5.052	5.114	5.171	5.225	5.275	5.322	5.367
40	2.858	3.442	3.791	4.039	4.232	4.388	4.521	4.634	4.735	4.824	4.904	4.977	5.044	5.106	5.163	5.216	5.266	5.313	5.358
48	2.843	3.420	3.764	4.008	4.197	4.351	4.481	4.592	4.690	4.777	4.856	4.927	4.993	5.053	5.109	5.161	5.210	5.256	5.299
60	2.829	3.399	3.737	3.977	4.163	4.314	4.441	4.550	4.646	4.732	4.808	4.878	4.942	5.001	5.056	5.107	5.154	5.199	5.241
80	2.814	3.377	3.711	3.947	4.129	4.277	4.402	4.509	4.603	4.686	4.761	4.829	4.892	4.949	5.003	5.052	5.099	5.142	5.183
120	2.800	3.356	3.685	3.917	4.096	4.241	4.363	4.468	4.560	4.641	4.714	4.781	4.842	4.898	4.950	4.998	5.043	5.086	5.126
240	2.786	3.335	3.659	3.887	4.063	4.205	4.324	4.427	4.517	4.596	4.668	4.733	4.792	4.847	4.897	4.944	4.988	5.030	5.069
inf	2.772	3.314	3.633	3.858	4.030	4.170	4.286	4.387	4.474	4.552	4.622	4.685	4.743	4.796	4.845	4.891	4.934	4.974	5.012

df	20	22	24	26	28	30	32	34	36	38	40	50	60	70	80	90	100
1	59.56	60.91	62.12	63.22	64.23	65.15	66.01	66.81	67.56	68.26	68.92	71.73	73.97	75.82	77.4	78.77	79.98
2	16.77	17.13	17.45	17.75	18.02	18.27	18.5	18.72	18.92	19.11	19.28	20.05	20.66	21.16	21.59	21.96	22.29
3	11.24	11.47	11.68	11.87	12.05	12.21	12.36	12.50	12.63	12.75	12.87	13.36	13.76	14.08	14.36	14.61	14.82
4	9.233	9.418	9.584	9.736	9.875	10.00	10.12	10.23	10.34	10.44	10.53	10.93	11.24	11.51	11.73	11.92	12.09
5	8.208	8.368	8.512	8.643	8.764	8.875	8.979	9.075	9.165	9.250	9.330	9.674	9.949	10.18	10.38	10.54	10.69
6	7.587	7.730	7.861	7.979	8.088	8.189	8.283	8.370	8.452	8.529	8.601	8.913	9.163	9.370	9.548	9.702	9.839
7	7.170	7.303	7.423	7.533	7.634	7.728	7.814	7.895	7.972	8.043	8.110	8.400	8.632	8.824	8.989	9.133	9.261
8	6.870	6.995	7.109	7.212	7.307	7.395	7.477	7.554	7.625	7.693	7.756	8.029	8.248	8.430	8.586	8.722	8.843
9	6.644	6.763	6.871	6.970	7.061	7.145	7.222	7.295	7.363	7.428	7.488	7.749	7.958	8.132	8.281	8.410	8.526
10	6.467	6.582	6.686	6.781	6.868	6.948	7.023	7.093	7.159	7.220	7.279	7.529	7.730	7.897	8.041	8.166	8.276
11	6.326	6.436	6.536	6.628	6.712	6.790	6.863	6.930	6.994	7.053	7.110	7.352	7.546	7.708	7.847	7.968	8.075
12	6.209	6.317	6.414	6.503	6.585	6.660	6.731	6.796	6.858	6.916	6.970	7.205	7.394	7.552	7.687	7.804	7.909
13	6.112	6.217	6.312	6.398	6.478	6.551	6.620	6.684	6.744	6.800	6.854	7.083	7.267	7.421	7.552	7.667	7.769
14	6.029	6.132	6.224	6.309	6.387	6.459	6.526	6.588	6.647	6.702	6.754	6.979	7.159	7.309	7.438	7.550	7.650
15	5.958	6.059	6.149	6.233	6.309	6.379	6.445	6.506	6.564	6.618	6.669	6.888	7.065	7.212	7.339	7.449	7.546
16	5.897	5.995	6.084	6.166	6.241	6.310	6.374	6.434	6.491	6.544	6.594	6.810	6.984	7.128	7.252	7.360	7.457
17	5.842	5.940	6.027	6.107	6.181	6.249	6.313	6.372	6.427	6.479	6.529	6.741	6.912	7.054	7.176	7.283	7.377
18	5.794	5.890	5.977	6.055	6.128	6.195	6.258	6.316	6.371	6.422	6.471	6.680	6.848	6.989	7.109	7.213	7.307
19	5.752	5.846	5.932	6.009	6.081	6.147	6.209	6.267	6.321	6.371	6.419	6.626	6.792	6.930	7.048	7.152	7.244
20	5.714	5.807	5.891	5.968	6.039	6.104	6.165	6.222	6.275	6.325	6.373	6.576	6.740	6.877	6.994	7.097	7.187
24	5.594	5.683	5.764	5.838	5.906	5.968	6.027	6.081	6.132	6.181	6.226	6.421	6.579	6.710	6.822	6.920	7.008
30	5.475	5.561	5.638	5.709	5.774	5.833	5.889	5.941	5.990	6.037	6.080	6.267	6.417	6.543	6.650	6.744	6.827
40	5.358	5.439	5.513	5.581	5.642	5.700	5.753	5.803	5.849	5.893	5.934	6.112	6.255	6.375	6.477	6.566	6.645
60	5.241	5.319	5.389	5.453	5.512	5.566	5.617	5.664	5.708	5.750	5.789	5.958	6.093	6.206	6.303	6.387	6.462
120	5.126	5.200	5.266	5.327	5.382	5.434	5.481	5.526	5.568	5.607	5.644	5.802	5.929	6.035	6.126	6.205	6.275
inf	5.012	5.081	5.144	5.201	5.253	5.301	5.346	5.388	5.427	5.463	5.498	5.646	5.764	5.863	5.947	6.020	6.085

Alpha = 0.025

df	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	18	20
1	35.99	54.00	65.69	74.22	80.87	86.29	90.85	94.77	98.20	101.3	104.0	106.5	108.8	110.8	112.7	116.2	119.2
2	8.776	11.94	14.01	15.54	16.75	17.74	18.58	19.31	19.95	20.52	21.03	21.49	21.91	22.30	22.67	23.32	23.89
3	5.907	7.661	8.808	9.660	10.34	10.89	11.37	11.78	12.14	12.46	12.75	13.01	13.26	13.48	13.69	14.06	14.39
4	4.943	6.244	7.088	7.716	8.213	8.625	8.976	9.279	9.548	9.788	10.01	10.20	10.39	10.55	10.71	10.99	11.23
5	4.474	5.558	6.257	6.775	7.186	7.527	7.816	8.068	8.291	8.490	8.670	8.834	8.984	9.124	9.253	9.486	9.693
6	4.199	5.158	5.772	6.226	6.586	6.884	7.138	7.359	7.554	7.729	7.887	8.031	8.163	8.286	8.399	8.605	8.787
7	4.018	4.897	5.455	5.868	6.194	6.464	6.695	6.895	7.072	7.230	7.373	7.504	7.624	7.735	7.839	8.025	8.191
8	3.892	4.714	5.233	5.616	5.919	6.169	6.382	6.568	6.732	6.879	7.011	7.132	7.244	7.347	7.443	7.616	7.769
9	3.797	4.578	5.069	5.430	5.715	5.950	6.151	6.325	6.479	6.617	6.742	6.856	6.961	7.058	7.148	7.311	7.455
10	3.725	4.474	4.943	5.287	5.558	5.782	5.972	6.138	6.285	6.416	6.534	6.643	6.742	6.834	6.920	7.075	7.212
11	3.667	4.391	4.843	5.173	5.433	5.648	5.831	5.989	6.130	6.256	6.369	6.473	6.568	6.657	6.739	6.887	7.019
12	3.620	4.325	4.762	5.081	5.332	5.540	5.716	5.869	6.004	6.125	6.235	6.335	6.427	6.512	6.591	6.734	6.861
13	3.582	4.269	4.694	5.004	5.248	5.449	5.620	5.769	5.900	6.017	6.123	6.220	6.309	6.392	6.468	6.607	6.730
14	3.550	4.222	4.638	4.940	5.178	5.374	5.540	5.684	5.811	5.926	6.029	6.123	6.210	6.290	6.364	6.499	6.619
15	3.522	4.182	4.589	4.885	5.118	5.309	5.471	5.612	5.737	5.848	5.949	6.041	6.125	6.203	6.276	6.407	6.523
16	3.498	4.148	4.548	4.838	5.066	5.253	5.412	5.550	5.672	5.781	5.879	5.969	6.052	6.128	6.199	6.328	6.441
17	3.477	4.118	4.512	4.797	5.020	5.204	5.361	5.496	5.615	5.722	5.818	5.907	5.987	6.062	6.132	6.258	6.370
18	3.458	4.092	4.480	4.761	4.981	5.162	5.315	5.448	5.565	5.670	5.765	5.852	5.931	6.004	6.073	6.197	6.306
19	3.442	4.068	4.451	4.728	4.945	5.123	5.275	5.405	5.521	5.624	5.718	5.803	5.881	5.954	6.020	6.142	6.250
20	3.427	4.047	4.426	4.700	4.914	5.089	5.238	5.368	5.481	5.583	5.675	5.759	5.836	5.907	5.974	6.093	6.200
24	3.381	3.983	4.347	4.610	4.816	4.984	5.216	5.250	5.358	5.455	5.543	5.623	5.697	5.764	5.827	5.941	6.043
30	3.337	3.919	4.271	4.523	4.720	4.881	5.017	5.134	5.238	5.330	5.414	5.490	5.560	5.624	5.684	5.792	5.888
40	3.294	3.858	4.197	4.439	4.627	4.780	4.910	5.022	5.120	5.208	5.288	5.360	5.426	5.487	5.544	5.646	5.737
60	3.251	3.798	4.124	4.356	4.536	4.682	4.806	4.912	5.006	5.089	5.164	5.232	5.295	5.352	5.406	5.503	5.588
120	3.210	3.739	4.053	4.276	4.447	4.587	4.704	4.805	4.894	4.972	5.043	5.107	5.166	5.221	5.271	5.362	5.442
inf	3.170	3.682	3.984	4.197	4.361	4.494	4.605	4.700	4.784	4.858	4.925	4.985	5.041	5.092	5.139	5.224	5.299
df	20	22	24	26	28	30	32	34	36	38	40	50	60	70	80	90	100
1	119.2	121.9	124.3	126.5	128.6	130.4	132.1	133.7	135.2	136.6	137.9	143.6	148.1	151.8	154.9	157.7	160.0
2	23.89	24.41	24.87	25.29	25.67	26.03	26.35	26.66	26.95	27.22	27.47	28.55	29.42	30.13	30.74	31.27	31.74
3	14.39	14.69	14.95	15.19	15.41	15.62	15.81	15.99	16.15	16.31	16.46	17.08	17.59	18.00	18.36	18.67	18.95
4	11.23	11.46	11.66	11.84	12.00	12.16	12.30	12.44	12.56	12.68	12.79	13.27	13.65	13.96	14.23	14.47	14.68
5	9.693	9.878	10.04	10.20	10.34	10.47	10.59	10.70	10.80	10.91	11.00	11.40	11.72	11.99	12.21	12.41	12.59
6	8.787	8.949	9.097	9.231	9.355	9.469	9.575	9.674	9.767	9.855	9.938	10.30	10.58	10.81	11.02	11.19	11.35
7	8.191	8.339	8.473	8.595	8.708	8.812	8.909	8.999	9.084	9.164	9.239	9.563	9.822	10.04	10.23	10.38	10.53
8	7.769	7.907	8.031	8.145	8.250	8.346	8.436	8.520	8.599	8.673	8.743	9.044	9.286	9.487	9.660	9.810	9.944
9	7.455	7.585	7.702	7.809	7.908	7.999	8.084	8.163	8.237	8.307	8.373	8.657	8.885	9.076	9.238	9.381	9.507
10	7.212	7.335	7.447	7.549	7.643	7.729	7.810	7.885	7.956	8.023	8.086	8.356	8.574	8.755	8.911	9.046	9.167
11	7.019	7.137	7.244	7.341	7.431	7.514	7.592	7.664	7.732	7.796	7.856	8.116	8.325	8.499	8.648	8.779	8.894
12	6.861	6.974	7.078	7.172	7.258	7.338	7.413	7.483	7.548	7.610	7.668	7.919	8.120	8.289	8.433	8.559	8.671
13	6.730	6.840	6.939	7.031	7.115	7.192	7.265	7.332	7.396	7.455	7.512	7.755	7.950	8.113	8.253	8.375	8.484
14	6.619	6.726	6.823	6.911	6.993	7.069	7.139	7.204	7.266	7.324	7.379	7.615	7.806	7.965	8.101	8.220	8.325
15	6.523	6.628	6.723	6.809	6.889	6.962	7.031	7.095	7.155	7.212	7.265	7.496	7.682	7.837	7.970	8.086	8.189
16	6.441	6.543	6.636	6.721	6.799	6.870	6.938	7.000	7.059	7.115	7.167	7.393	7.574	7.726	7.856	7.969	8.070
17	6.370	6.469	6.560	6.644	6.720	6.790	6.856	6.917	6.975	7.030	7.081	7.302	7.480	7.628	7.756	7.868	7.966
18	6.306	6.404	6.493	6.575	6.650	6.720	6.784	6.844	6.900	6.954	7.005	7.221	7.396	7.543	7.667	7.777	7.874
19	6.250	6.347	6.434	6.514	6.588	6.656	6.719	6.779	6.835	6.887	6.936	7.150	7.322	7.465	7.589	7.696	7.792
20	6.200	6.295	6.381	6.460	6.532	6.600	6.662	6.720	6.775	6.827	6.876	7.086	7.255	7.397	7.518	7.624	7.718
24	6.043	6.133	6.215	6.290	6.359	6.423	6.482	6.538	6.589	6.639	6.685	6.885	7.046	7.180	7.296	7.397	7.486
30	5.888	5.974	6.052	6.123	6.188	6.248	6.305	6.357	6.406	6.453	6.497	6.686	6.839	6.965	7.075	7.171	7.255
40	5.737	5.818	5.891	5.958	6.020	6.077	6.130	6.179	6.226	6.270	6.311	6.489	6.633	6.753	6.855	6.945	7.025
60	5.588	5.664	5.733	5.797	5.854	5.908	5.958	6.004	6.048	6.089	6.127	6.295	6.429	6.540	6.636	6.720	6.795
120	5.442	5.513	5.578	5.637	5.691	5.741	5.788	5.831	5.872	5.910	5.946	6.101	6.225	6.329	6.418	6.495	6.564
inf	5.299	5.365	5.425	5.480	5.530	5.577	5.620	5.660	5.698	5.733	5.766	5.909	6.023	6.118	6.199	6.270	6.333

Alpha = 0.01

	k -->																		
df	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	90.024	135.04	164.26	185.58	202.21	215.77	227.17	236.97	245.54	253.15	259.98	266.17	271.81	277.00	281.80	286.26	290.43	294.33	298.00
2	14.036	19.019	22.294	24.717	26.629	28.201	29.530	30.679	31.689	32.589	33.398	34.134	34.806	35.426	36.000	36.534	37.034	37.502	37.943
3	8.260	10.619	12.170	13.324	14.241	14.998	15.641	16.199	16.691	17.130	17.526	17.887	18.217	18.522	18.805	19.068	19.315	19.546	19.765
4	6.511	8.120	9.173	9.958	10.583	11.101	11.542	11.925	12.264	12.567	12.840	13.090	13.318	13.530	13.726	13.909	14.081	14.242	14.394
5	5.702	6.976	7.804	8.421	8.913	9.321	9.669	9.971	10.239	10.479	10.696	10.894	11.076	11.244	11.400	11.545	11.682	11.811	11.932
6	5.243	6.331	7.033	7.556	7.972	8.318	8.612	8.869	9.097	9.300	9.485	9.653	9.808	9.951	10.084	10.208	10.325	10.434	10.538
7	4.949	5.919	6.542	7.005	7.373	7.678	7.939	8.166	8.367	8.548	8.711	8.860	8.997	9.124	9.242	9.353	9.456	9.553	9.645
8	4.745	5.635	6.204	6.625	6.959	7.237	7.474	7.680	7.863	8.027	8.176	8.311	8.436	8.552	8.659	8.760	8.854	8.943	9.027
9	4.596	5.428	5.957	6.347	6.657	6.915	7.134	7.325	7.494	7.646	7.784	7.910	8.025	8.132	8.232	8.325	8.412	8.495	8.573
10	4.482	5.270	5.769	6.136	6.428	6.669	6.875	7.054	7.213	7.356	7.485	7.603	7.712	7.812	7.906	7.993	8.075	8.153	8.226
11	4.392	5.146	5.621	5.970	6.247	6.476	6.671	6.841	6.992	7.127	7.250	7.362	7.464	7.560	7.648	7.731	7.809	7.883	7.952
12	4.320	5.046	5.502	5.836	6.101	6.320	6.507	6.670	6.814	6.943	7.060	7.166	7.265	7.356	7.441	7.520	7.594	7.664	7.730
13	4.260	4.964	5.404	5.726	5.981	6.192	6.372	6.528	6.666	6.791	6.903	7.006	7.100	7.188	7.269	7.345	7.417	7.484	7.548
14	4.210	4.895	5.322	5.634	5.881	6.085	6.258	6.409	6.543	6.663	6.772	6.871	6.962	7.047	7.125	7.199	7.268	7.333	7.394
15	4.167	4.836	5.252	5.556	5.796	5.994	6.162	6.309	6.438	6.555	6.660	6.756	6.845	6.927	7.003	7.074	7.141	7.204	7.264
16	4.131	4.786	5.192	5.489	5.722	5.915	6.079	6.222	6.348	6.461	6.564	6.658	6.744	6.823	6.897	6.967	7.032	7.093	7.151
17	4.099	4.742	5.140	5.430	5.659	5.847	6.007	6.147	6.270	6.380	6.480	6.572	6.656	6.733	6.806	6.873	6.937	6.997	7.053
18	4.071	4.703	5.094	5.379	5.603	5.787	5.944	6.081	6.201	6.309	6.407	6.496	6.579	6.655	6.725	6.791	6.854	6.912	6.967
19	4.046	4.669	5.054	5.334	5.553	5.735	5.889	6.022	6.141	6.246	6.342	6.430	6.510	6.585	6.654	6.719	6.780	6.837	6.891
20	4.024	4.639	5.018	5.293	5.510	5.688	5.839	5.970	6.086	6.190	6.285	6.370	6.449	6.523	6.591	6.654	6.714	6.770	6.823
21	4.004	4.612	4.986	5.257	5.470	5.646	5.794	5.924	6.038	6.140	6.233	6.317	6.395	6.467	6.534	6.596	6.655	6.710	6.762
22	3.986	4.588	4.957	5.225	5.435	5.608	5.754	5.882	5.994	6.095	6.186	6.269	6.346	6.417	6.482	6.544	6.602	6.656	6.707
23	3.970	4.566	4.931	5.195	5.403	5.573	5.718	5.844	5.955	6.054	6.144	6.226	6.301	6.371	6.436	6.497	6.553	6.607	6.658
24	3.955	4.546	4.907	5.168	5.373	5.542	5.685	5.809	5.919	6.017	6.105	6.186	6.261	6.330	6.394	6.453	6.510	6.562	6.612
25	3.942	4.527	4.885	5.144	5.347	5.513	5.655	5.778	5.886	5.983	6.070	6.150	6.224	6.292	6.355	6.414	6.469	6.522	6.571

df	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
26	3.930	4.510	4.865	5.121	5.322	5.487	5.627	5.749	5.856	5.951	6.038	6.117	6.190	6.257	6.319	6.378	6.432	6.484	6.533
27	3.918	4.495	4.847	5.101	5.300	5.463	5.602	5.722	5.828	5.923	6.008	6.087	6.158	6.225	6.287	6.344	6.399	6.450	6.498
28	3.908	4.481	4.830	5.082	5.279	5.441	5.578	5.697	5.802	5.896	5.981	6.058	6.129	6.195	6.256	6.314	6.367	6.418	6.465
29	3.898	4.467	4.814	5.064	5.260	5.420	5.556	5.674	5.778	5.871	5.955	6.032	6.103	6.168	6.228	6.285	6.338	6.388	6.435
30	3.889	4.455	4.799	5.048	5.242	5.401	5.536	5.653	5.756	5.848	5.932	6.008	6.078	6.142	6.202	6.258	6.311	6.361	6.407
31	3.881	4.443	4.786	5.032	5.225	5.383	5.517	5.633	5.736	5.827	5.910	5.985	6.055	6.119	6.178	6.234	6.286	6.335	6.381
32	3.873	4.433	4.773	5.018	5.210	5.367	5.500	5.615	5.716	5.807	5.889	5.964	6.033	6.096	6.155	6.211	6.262	6.311	6.357
33	3.865	4.423	4.761	5.005	5.195	5.351	5.483	5.598	5.698	5.789	5.870	5.944	6.013	6.076	6.134	6.189	6.240	6.289	6.334
34	3.859	4.413	4.750	4.992	5.181	5.336	5.468	5.581	5.682	5.771	5.852	5.926	5.994	6.056	6.114	6.169	6.220	6.268	6.313
35	3.852	4.404	4.739	4.980	5.169	5.323	5.453	5.566	5.666	5.755	5.835	5.908	5.976	6.038	6.096	6.150	6.200	6.248	6.293
36	3.846	4.396	4.729	4.969	5.156	5.310	5.439	5.552	5.651	5.739	5.819	5.892	5.959	6.021	6.078	6.132	6.182	6.229	6.274
37	3.840	4.388	4.720	4.959	5.145	5.298	5.427	5.538	5.637	5.725	5.804	5.876	5.943	6.004	6.061	6.115	6.165	6.212	6.256
38	3.835	4.381	4.711	4.949	5.134	5.286	5.414	5.526	5.623	5.711	5.790	5.862	5.928	5.989	6.046	6.099	6.148	6.195	6.239
39	3.830	4.374	4.703	4.940	5.124	5.275	5.403	5.513	5.611	5.698	5.776	5.848	5.914	5.974	6.031	6.084	6.133	6.179	6.223
40	3.825	4.367	4.695	4.931	5.114	5.265	5.392	5.502	5.599	5.685	5.764	5.835	5.900	5.961	6.017	6.069	6.118	6.165	6.208
48	3.793	4.324	4.644	4.874	5.052	5.198	5.322	5.428	5.522	5.606	5.681	5.750	5.814	5.872	5.926	5.977	6.024	6.069	6.111
60	3.762	4.282	4.594	4.818	4.991	5.133	5.253	5.356	5.447	5.528	5.601	5.667	5.728	5.784	5.837	5.886	5.931	5.974	6.015
80	3.732	4.241	4.545	4.763	4.931	5.069	5.185	5.284	5.372	5.451	5.521	5.585	5.644	5.698	5.749	5.796	5.840	5.881	5.920
120	3.702	4.200	4.497	4.709	4.872	5.005	5.118	5.214	5.299	5.375	5.443	5.505	5.561	5.614	5.662	5.708	5.750	5.790	5.827
240	3.672	4.160	4.450	4.655	4.814	4.943	5.052	5.145	5.227	5.300	5.366	5.426	5.480	5.530	5.577	5.621	5.661	5.699	5.735
inf	3.643	4.120	4.403	4.603	4.757	4.882	4.987	5.078	5.157	5.227	5.290	5.348	5.400	5.448	5.493	5.535	5.574	5.611	5.645

df	20	22	24	26	28	30	32	34	36	38	40	50	60	70	80	90	100
1	298.0	304.7	310.8	316.3	321.3	326.0	330.3	334.3	338.0	341.5	344.8	358.9	370.1	379.4	387.3	394.1	400.1
2	37.95	38.76	39.49	40.15	40.76	41.32	41.84	42.33	42.78	43.21	43.61	45.33	46.70	47.83	48.80	49.64	50.38
3	19.77	20.17	20.53	20.86	21.16	21.44	21.70	21.95	22.17	22.39	22.59	23.45	24.13	24.71	25.19	25.67	25.99
4	14.40	14.68	14.93	15.16	15.37	15.57	15.75	15.92	16.08	16.23	16.37	16.98	17.46	17.86	18.20	18.50	18.77
5	11.93	12.16	12.36	12.54	12.71	12.87	13.02	13.15	13.28	13.40	13.52	14.00	14.39	14.72	14.99	15.23	15.45
6	10.54	10.73	10.91	11.06	11.21	11.34	11.47	11.58	11.69	11.80	11.90	12.31	12.65	12.92	13.16	13.37	13.55
7	9.646	9.815	9.970	10.11	10.24	10.36	10.47	10.58	10.67	10.77	10.85	11.23	11.52	11.77	11.99	12.17	12.34
8	9.027	9.182	9.322	9.450	9.569	9.678	9.779	9.874	9.964	10.05	10.13	10.47	10.75	10.97	11.17	11.34	11.49
9	8.573	8.717	8.847	8.966	9.075	9.177	9.271	9.360	9.443	9.521	9.594	9.912	10.17	10.38	10.57	10.73	10.87
10	8.226	8.361	8.483	8.595	8.698	8.794	8.883	8.966	9.044	9.117	9.187	9.486	9.726	9.927	10.10	10.25	10.39
11	7.952	8.080	8.196	8.303	8.400	8.491	8.575	8.654	8.728	8.798	8.864	9.148	9.377	9.568	9.732	9.875	10.00
12	7.731	7.853	7.964	8.066	8.159	8.246	8.327	8.402	8.473	8.539	8.603	8.875	9.094	9.277	9.434	9.571	9.693
13	7.548	7.665	7.772	7.870	7.960	8.043	8.121	8.193	8.262	8.326	8.387	8.648	8.859	9.035	9.187	9.318	9.436
14	7.395	7.508	7.611	7.705	7.792	7.873	7.948	8.018	8.084	8.146	8.204	8.457	8.661	8.832	8.978	9.106	9.219
15	7.264	7.374	7.474	7.566	7.650	7.728	7.800	7.869	7.932	7.992	8.049	8.295	8.492	8.658	8.800	8.924	9.035
16	7.152	7.258	7.356	7.445	7.527	7.602	7.673	7.739	7.802	7.860	7.916	8.154	8.347	8.507	8.646	8.767	8.874
17	7.053	7.158	7.253	7.340	7.420	7.493	7.563	7.627	7.687	7.745	7.799	8.031	8.229	8.377	8.511	8.630	8.735
18	6.968	7.070	7.163	7.247	7.325	7.398	7.465	7.528	7.587	7.643	7.696	7.924	8.107	8.261	8.393	8.508	8.611
19	6.891	6.992	7.082	7.166	7.242	7.313	7.379	7.440	7.498	7.553	7.605	7.828	8.008	8.159	8.288	8.401	8.502
20	6.823	6.922	7.011	7.092	7.168	7.237	7.302	7.362	7.419	7.473	7.523	7.742	7.919	8.067	8.194	8.305	8.404
24	6.612	6.705	6.789	6.865	6.936	7.001	7.062	7.119	7.173	7.223	7.270	7.476	7.642	7.780	7.900	8.004	8.097
30	6.407	6.494	6.572	6.644	6.710	6.772	6.828	6.881	6.932	6.978	7.023	7.215	7.370	7.500	7.611	7.709	7.796
40	6.209	6.289	6.362	6.429	6.490	6.547	6.600	6.650	6.697	6.740	6.782	6.960	7.104	7.225	7.328	7.419	7.500
60	6.015	6.090	6.158	6.220	6.277	6.330	6.378	6.424	6.467	6.507	6.546	6.710	6.843	6.954	7.050	7.133	7.207
120	5.827	5.897	5.959	6.016	6.069	6.117	6.162	6.204	6.244	6.281	6.316	6.467	6.588	6.689	6.776	6.852	6.919
inf	5.645	5.709	5.766	5.818	5.866	5.911	5.952	5.990	6.026	6.060	6.092	6.228	6.338	6.429	6.507	6.575	6.636

Alpha = 0.005

df	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	18	20
1	180.1	270.1	328.5	371.2	404.4	431.6	454.4	474.0	491.1	506.3	520.0	532.4	543.6	554.0	563.6	580.9	596.0
2	19.93	26.97	31.60	35.02	37.73	39.95	41.83	43.46	44.89	46.16	47.31	48.35	49.30	50.17	50.99	52.45	53.74
3	10.55	13.50	15.45	16.91	18.06	19.01	19.83	20.53	21.15	21.70	22.20	22.66	23.08	23.46	23.82	24.46	25.03
4	7.916	9.814	11.06	11.99	12.74	13.35	13.88	14.33	14.74	15.10	15.42	15.72	15.99	16.24	16.48	16.90	17.28
5	6.751	8.196	9.141	9.847	10.41	10.88	11.28	11.63	11.93	12.21	12.46	12.69	12.90	13.09	13.27	13.60	13.89
6	6.105	7.306	8.088	8.670	9.135	9.522	9.852	10.14	10.40	10.63	10.83	11.02	11.20	11.36	11.51	11.78	12.02
7	5.699	6.750	7.429	7.935	8.339	8.674	8.961	9.211	9.433	9.632	9.812	9.977	10.13	10.27	10.40	10.64	10.85
8	5.420	6.370	6.981	7.435	7.797	8.097	8.354	8.578	8.777	8.955	9.117	9.265	9.401	9.527	9.644	9.857	10.04
9	5.218	6.096	6.657	7.074	7.405	7.680	7.915	8.120	8.303	8.466	8.614	8.749	8.874	8.990	9.097	9.292	9.465
10	5.065	5.888	6.412	6.800	7.109	7.365	7.584	7.775	7.944	8.096	8.234	8.360	8.476	8.583	8.683	8.865	9.026
11	4.945	5.727	6.222	6.588	6.878	7.119	7.325	7.505	7.664	7.807	7.937	8.055	8.164	8.265	8.359	8.530	8.682
12	4.849	5.597	6.068	6.416	6.693	6.922	7.118	7.288	7.439	7.575	7.697	7.810	7.914	8.009	8.099	8.261	8.405
13	4.770	5.490	5.943	6.277	6.541	6.760	6.947	7.111	7.255	7.384	7.502	7.609	7.708	7.800	7.886	8.040	8.178
14	4.704	5.401	5.838	6.160	6.414	6.626	6.805	6.962	7.101	7.225	7.338	7.442	7.537	7.625	7.707	7.856	7.988
15	4.647	5.325	5.750	6.061	6.308	6.511	6.685	6.837	6.971	7.091	7.200	7.300	7.392	7.477	7.556	7.699	7.827
16	4.599	5.261	5.674	5.977	6.216	6.413	6.582	6.729	6.859	6.976	7.081	7.178	7.267	7.349	7.426	7.566	7.689
17	4.557	5.205	5.608	5.903	6.136	6.329	6.493	6.636	6.763	6.876	6.979	7.072	7.159	7.239	7.314	7.449	7.569
18	4.521	5.156	5.550	5.839	6.067	6.255	6.415	6.554	6.678	6.788	6.888	6.980	7.064	7.142	7.215	7.347	7.464
19	4.488	5.113	5.500	5.783	6.005	6.189	6.346	6.482	6.603	6.711	6.809	6.898	6.981	7.057	7.128	7.257	7.372
20	4.460	5.074	5.455	5.732	5.951	6.131	6.285	6.418	6.537	6.642	6.738	6.826	6.907	6.981	7.051	7.177	7.289
24	4.371	4.955	5.315	5.577	5.783	5.952	6.096	6.221	6.332	6.431	6.520	6.602	6.677	6.747	6.812	6.930	7.034
30	4.285	4.841	5.181	5.428	5.621	5.780	5.914	6.031	6.135	6.227	6.310	6.387	6.456	6.521	6.581	6.691	6.788
40	4.202	4.731	5.053	5.284	5.465	5.614	5.739	5.848	5.944	6.030	6.108	6.179	6.244	6.304	6.360	6.461	6.550
60	4.122	4.625	4.928	5.146	5.316	5.454	5.571	5.673	5.762	5.841	5.913	5.979	6.039	6.094	6.146	6.239	6.321
120	4.045	4.523	4.809	5.013	5.172	5.301	5.410	5.504	5.586	5.660	5.726	5.786	5.842	5.893	5.940	6.025	6.101
inf	3.970	4.424	4.694	4.886	5.033	5.154	5.255	5.341	5.418	5.485	5.546	5.602	5.652	5.699	5.742	5.820	5.889
df	20	22	24	26	28	30	32	34	36	38	40	50	60	70	80	90	100
1	596.0	609.5	621.7	632.6	642.7	652.0	660.6	668.5	676.0	683.0	689.6	717.8	740.2	758.8	774.5	788.2	800.3
2	53.74	54.89	55.92	56.86	57.73	58.52	59.26	59.95	60.59	61.19	61.76	64.19	66.13	67.74	69.10	70.29	71.35
3	25.03	25.54	26.00	26.42	26.80	27.15	27.48	27.79	28.07	28.34	28.60	29.68	30.55	31.27	31.88	32.42	32.90
4	17.28	17.61	17.91	18.19	18.44	18.68	18.89	19.09	19.28	19.46	19.63	20.36	20.93	21.42	21.83	22.18	22.50
5	13.89	14.14	14.38	14.59	14.79	14.96	15.13	15.29	15.44	15.58	15.71	16.27	16.72	17.09	17.41	17.69	17.94
6	12.02	12.23	12.43	12.61	12.77	12.92	13.06	13.19	13.32	13.43	13.54	14.02	14.40	14.71	14.98	15.21	15.43
7	10.85	11.03	11.21	11.36	11.50	11.64	11.76	11.88	11.99	12.09	12.18	12.60	12.93	13.21	13.44	13.65	13.84
8	10.04	10.22	10.37	10.51	10.64	10.76	10.87	10.97	11.07	11.16	11.25	11.63	11.93	12.18	12.39	12.58	12.75
9	9.465	9.620	9.761	9.890	10.01	10.12	10.22	10.32	10.41	10.49	10.58	10.92	11.20	11.43	11.63	11.80	11.96
10	9.026	9.170	9.302	9.422	9.532	9.635	9.730	9.820	9.904	9.983	10.06	10.38	10.64	10.86	11.04	11.20	11.35
11	8.682	8.818	8.941	9.055	9.159	9.256	9.345	9.430	9.509	9.583	9.654	9.957	10.20	10.41	10.59	10.74	10.88
12	8.405	8.534	8.652	8.759	8.858	8.950	9.036	9.116	9.191	9.262	9.328	9.617	9.850	10.04	10.21	10.36	10.49
13	8.178	8.302	8.414	8.516	8.611	8.699	8.781	8.857	8.929	8.997	9.061	9.337	9.560	9.747	9.907	10.05	10.17
14	7.988	8.107	8.215	8.314	8.404	8.489	8.568	8.641	8.710	8.775	8.837	9.103	9.317	9.497	9.652	9.787	9.907
15	7.827	7.942	8.046	8.141	8.229	8.311	8.387	8.458	8.524	8.587	8.647	8.904	9.111	9.285	9.434	9.565	9.680
16	7.689	7.800	7.901	7.994	8.078	8.158	8.231	8.300	8.365	8.425	8.483	8.733	8.933	9.102	9.247	9.373	9.486
17	7.569	7.677	7.775	7.865	7.948	8.024	8.096	8.163	8.226	8.285	8.341	8.583	8.779	8.943	9.084	9.206	9.316
18	7.464	7.570	7.665	7.753	7.833	7.908	7.978	8.043	8.104	8.162	8.217	8.452	8.643	8.803	8.940	9.061	9.167
19	7.372	7.474	7.568	7.653	7.732	7.805	7.873	7.937	7.996	8.053	8.106	8.337	8.523	8.679	8.813	8.931	9.036
20	7.289	7.390	7.481	7.565	7.642	7.713	7.780	7.842	7.901	7.956	8.008	8.234	8.416	8.569	8.700	8.815	8.917
24	7.034	7.128	7.213	7.291	7.362	7.429	7.491	7.549	7.603	7.655	7.704	7.914	8.083	8.226	8.348	8.455	8.551
30	6.788	6.875	6.954	7.026	7.093	7.154	7.212	7.265	7.316	7.364	7.409	7.603	7.760	7.893	8.006	8.105	8.193
40	6.550	6.631	6.704	6.770	6.832	6.889	6.942	6.991	7.038	7.082	7.123	7.302	7.447	7.568	7.672	7.763	7.845
60	6.321	6.396	6.462	6.523	6.580	6.632	6.681	6.726	6.769	6.808	6.846	7.010	7.143	7.252	7.347	7.431	7.504
120	6.101	6.169	6.230	6.286	6.337	6.385	6.428	6.470	6.508	6.545	6.580	6.728	6.846	6.946	7.032	7.107	7.173
inf	5.889	5.951	6.006	6.057	6.103	6.146	6.186	6.223	6.258	6.291	6.322	6.454	6.561	6.649	6.725	6.792	6.850

Alpha = 0.001

df	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	18	20
1	900.3	1351	1643	1856	2022	2158	2272	2370	2455	2532	2600	2662	2718	2770	2818	2904	2980
2	44.69	60.42	70.77	78.43	84.49	89.46	93.67	97.30	100.5	103.3	105.9	108.2	110.4	112.3	114.2	117.4	120.3
3	18.28	23.32	26.65	29.13	31.11	32.74	34.12	35.33	36.39	37.34	38.20	38.98	39.69	40.35	40.97	42.07	43.05
4	12.18	14.99	16.84	18.23	19.34	20.26	21.04	21.73	22.33	22.87	23.36	23.81	24.21	24.59	24.94	25.58	26.14
5	9.714	11.67	12.96	13.93	14.71	15.35	15.90	16.38	16.81	17.18	17.53	17.85	18.13	18.41	18.66	19.10	19.51
6	8.427	9.96	10.97	11.72	12.32	12.83	13.26	13.63	13.97	14.27	14.54	14.79	15.01	15.22	15.42	15.78	16.09
7	7.648	8.930	9.768	10.40	10.90	11.32	11.68	11.99	12.27	12.52	12.74	12.95	13.14	13.32	13.48	13.78	14.04
8	7.130	8.250	8.978	9.522	9.958	10.32	10.64	10.91	11.15	11.36	11.56	11.74	11.91	12.06	12.21	12.47	12.70
9	6.762	7.768	8.419	8.906	9.295	9.619	9.897	10.14	10.36	10.55	10.73	10.89	11.03	11.18	11.30	11.54	11.75
10	6.487	7.411	8.006	8.450	8.804	9.099	9.352	9.573	9.769	9.946	10.11	10.25	10.39	10.52	10.64	10.85	11.03
11	6.275	7.136	7.687	8.098	8.426	8.699	8.933	9.138	9.319	9.482	9.630	9.766	9.892	10.01	10.12	10.31	10.49
12	6.106	6.917	7.436	7.821	8.127	8.383	8.601	8.793	8.962	9.115	9.254	9.381	9.498	9.606	9.707	9.891	10.06
13	5.970	6.740	7.231	7.595	7.885	8.126	8.333	8.513	8.673	8.817	8.948	9.068	9.178	9.281	9.376	9.550	9.704
14	5.856	6.594	7.062	7.409	7.685	7.915	8.110	8.282	8.434	8.571	8.696	8.809	8.914	9.012	9.103	9.267	9.414
15	5.760	6.470	6.920	7.252	7.517	7.736	7.925	8.088	8.234	8.365	8.483	8.592	8.693	8.786	8.872	9.030	9.170
16	5.678	6.365	6.799	7.119	7.374	7.585	7.766	7.923	8.063	8.189	8.303	8.407	8.504	8.593	8.676	8.828	8.963
17	5.608	6.275	6.695	7.005	7.250	7.454	7.629	7.781	7.916	8.037	8.148	8.248	8.342	8.427	8.508	8.654	8.784
18	5.546	6.196	6.604	6.905	7.143	7.341	7.510	7.657	7.788	7.906	8.012	8.110	8.199	8.283	8.361	8.502	8.628
19	5.492	6.127	6.525	6.817	7.049	7.242	7.405	7.549	7.676	7.790	7.893	7.988	8.075	8.156	8.232	8.369	8.491
20	5.444	6.065	6.454	6.740	6.966	7.154	7.313	7.453	7.577	7.688	7.788	7.880	7.966	8.044	8.118	8.251	8.370
24	5.297	5.877	6.238	6.503	6.712	6.884	7.031	7.159	7.272	7.374	7.467	7.551	7.629	7.701	7.768	7.890	7.999
30	5.156	5.698	6.033	6.278	6.470	6.628	6.763	6.880	6.984	7.077	7.162	7.239	7.310	7.375	7.437	7.548	7.647
40	5.022	5.528	5.838	6.063	6.240	6.386	6.509	6.616	6.711	6.796	6.872	6.942	7.007	7.067	7.122	7.223	7.312
60	4.894	5.365	5.653	5.860	6.022	6.155	6.268	6.366	6.451	6.528	6.598	6.661	6.720	6.774	6.824	6.914	6.995
120	4.771	5.211	5.476	5.667	5.815	5.937	6.039	6.128	6.206	6.276	6.339	6.396	6.448	6.496	6.542	6.623	6.695
inf	4.654	5.063	5.309	5.484	5.619	5.730	5.823	5.903	5.973	6.036	6.092	6.144	6.191	6.234	6.274	6.347	6.411
df	20	22	24	26	28	30	32	34	36	38	40	50	60	70	80	90	100
1	2980	3047	3108	3163	3213	3260	3303	3343	3380	3415	3448	3589	3701	3794	3873	3941	4002
2	120.3	122.9	125.2	127.3	129.3	131.0	132.7	134.2	135.7	137.0	138.3	143.7	148.0	151.6	154.7	157.4	159.7
3	43.05	43.92	44.70	45.42	46.07	46.68	47.24	47.77	48.26	48.72	49.16	51.02	52.51	53.75	54.81	55.72	56.53
4	26.14	26.65	27.10	27.51	27.89	28.24	28.57	28.88	29.16	29.43	29.68	30.78	31.65	32.37	32.98	33.52	34.00
5	19.51	19.86	20.19	20.48	20.75	21.01	21.24	21.46	21.66	21.86	22.03	22.82	23.45	23.97	24.41	24.80	25.15
6	16.09	16.38	16.64	16.87	17.08	17.28	17.47	17.64	17.81	17.96	18.10	18.73	19.22	19.64	20.00	20.31	20.58
7	14.04	14.29	14.50	14.70	14.88	15.05	15.20	15.35	15.49	15.62	15.74	16.27	16.69	17.04	17.35	17.61	17.85
8	12.70	12.91	13.09	13.26	13.42	13.57	13.71	13.84	13.96	14.07	14.18	14.64	15.01	15.32	15.59	15.82	16.02
9	11.75	11.93	12.10	12.25	12.39	12.53	12.65	12.77	12.87	12.97	13.07	13.49	13.82	14.10	14.34	14.55	14.74
10	11.03	11.20	11.36	11.50	11.63	11.75	11.87	11.97	12.07	12.16	12.25	12.63	12.94	13.20	13.42	13.61	13.78
11	10.49	10.65	10.79	10.92	11.04	11.16	11.26	11.35	11.45	11.53	11.62	11.97	12.25	12.49	12.70	12.88	13.04
12	10.06	10.20	10.34	10.46	10.57	10.68	10.78	10.87	10.96	11.03	11.11	11.44	11.71	11.94	12.13	12.29	12.45
13	9.704	9.843	9.969	10.09	10.19	10.29	10.39	10.47	10.55	10.63	10.70	11.01	11.27	11.48	11.66	11.82	11.97
14	9.414	9.546	9.666	9.776	9.878	9.972	10.06	10.14	10.22	10.30	10.37	10.66	10.91	11.11	11.28	11.43	11.57
15	9.170	9.296	9.411	9.517	9.613	9.703	9.788	9.867	9.940	10.01	10.08	10.37	10.59	10.79	10.96	11.10	11.23
16	8.963	9.084	9.194	9.295	9.388	9.475	9.556	9.631	9.702	9.769	9.833	10.11	10.34	10.52	10.68	10.82	10.95
17	8.784	8.900	9.007	9.104	9.194	9.277	9.355	9.429	9.497	9.562	9.623	9.888	10.10	10.29	10.44	10.58	10.70
18	8.628	8.741	8.844	8.938	9.025	9.106	9.181	9.251	9.318	9.381	9.440	9.696	9.904	10.08	10.23	10.36	10.48
19	8.491	8.601	8.701	8.792	8.876	8.955	9.028	9.096	9.161	9.221	9.279	9.528	9.730	9.899	10.04	10.17	10.29
20	8.370	8.477	8.574	8.663	8.745	8.821	8.892	8.959	9.021	9.081	9.137	9.379	9.575	9.740	9.881	10.01	10.12
24	7.999	8.097	8.185	8.267	8.342	8.411	8.476	8.537	8.594	8.648	8.700	8.921	9.100	9.250	9.380	9.494	9.596
30	7.647	7.735	7.816	7.890	7.958	8.021	8.080	8.135	8.188	8.237	8.283	8.484	8.647	8.783	8.901	9.004	9.096
40	7.312	7.393	7.466	7.533	7.594	7.651	7.704	7.754	7.801	7.845	7.887	8.067	8.214	8.337	8.442	8.535	8.618
60	6.995	7.067	7.133	7.193	7.248	7.299	7.347	7.392	7.433	7.473	7.510	7.671	7.802	7.924	8.025	8.108	8.181
120	6.695	6.760	6.818	6.872	6.921	6.966	7.008	7.048	7.085	7.121	7.153	7.296	7.411	7.507	7.590	7.662	7.726
inf	6.411	6.469	6.520	6.568	6.611	6.651	6.689	6.723	6.756	6.787	6.816	6.941	7.041	7.124	7.196	7.259	7.314

53 thoughts on “Studentized Range q Table”

