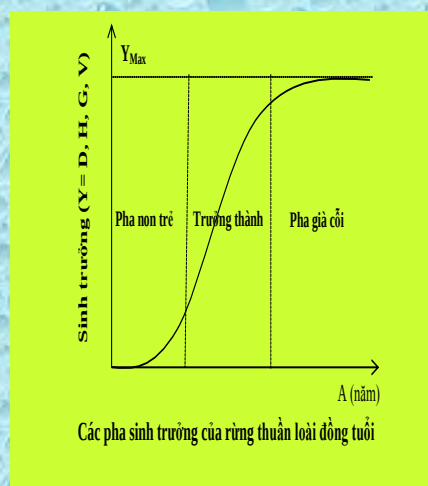


TRƯỜNG ĐẠI HỌC NÔNG LÂM Tp. HỒ CHÍ MINH

PGS. TS. Nguyễn Văn Thêm

oOo

Hướng dẫn sử dụng Statgraphics 3.0 & 5.1
để xử lý và phân tích các thông tin trong lâm học
(Sách tham khảo)



NHÀ XUẤT BẢN NÔNG NGHIỆP

Chi nhánh Tp. Hồ Chí Minh, 2010

LỜI NÓI ĐẦU

Hiện nay có rất nhiều tài liệu về lý thuyết xác suất và thống kê toán học ứng dụng trong các nghiên cứu lâm nghiệp và nhiều lĩnh vực khoa học khác. Những tài liệu này đã trình bày khá đầy đủ về lý thuyết xác suất và thống kê dựa trên nền toán học. Chúng thực sự là những tài liệu bổ ích cho những ai quan tâm đến lý thuyết xác suất và thống kê toán học. Nhưng nói chung những tài liệu này vẫn chỉ đi sâu về một số khía cạnh và coi nhẹ những phần khác, đặc biệt những khía cạnh mang tính đặc thù của lâm nghiệp. Điều này đã đem đến không ít khó khăn cho những cán bộ và sinh viên lâm nghiệp có mong muốn ứng dụng lý thuyết xác suất và thống kê để xử lý và phân tích các thông tin trong lâm nghiệp. Trước tình hình đó, cuốn sách “*Hướng dẫn sử dụng Statgraphics Plus Version 3.0 & 5.1 để xử lý và phân tích các thông tin trong lâm học*” được biên soạn như là một tài liệu thực hành nhằm trợ giúp cho những ai có mong muốn ứng dụng lý thuyết thống kê để xử lý và phân tích các thông tin trong lâm nghiệp. Sách được bố cục thành 8 chương. Nội dung của sách trình bày những cách thức và phương pháp xử lý và phân tích các số liệu trong lâm nghiệp với sự trợ giúp của thống kê toán học và phần mềm Statgraphics Plus Version 3.0 & 5.1. Cuốn sách này không trình bày sâu về lý thuyết thống kê toán học và những lĩnh vực khoa học khác trong lâm nghiệp, mà chỉ giới hạn ở phần thực hành những vấn đề có liên quan đến lĩnh vực lâm học và một số khoa học khác trong lâm nghiệp. Mặc dù vậy, nhiều nội dung của cuốn sách này vẫn có thể ứng dụng cho các lĩnh vực khoa học khác nhau (nông học, chăn nuôi, thú y, kinh tế, y học...).

Để giúp bạn đọc dễ dàng ứng dụng lý thuyết thống kê để xử lý các thông tin trong lâm học, ở mỗi chương, trước hết tác giả đưa ra một số vấn đề mà nhà lâm học cần quan tâm nghiên cứu. Kế đến dẫn ra những ví dụ cụ thể. Sau đó hướng dẫn sử dụng Statgraphics Plus Version 3.0 & 5.1 để xử lý và phân tích các thông tin trong lâm học. Vì sách chỉ trình bày những vấn đề thực hành, do đó muốn hiểu sâu từng vấn đề, bạn đọc cần đọc thêm những tài liệu khác có liên quan, đặc biệt là thống kê toán học. Tác giả cũng lưu ý rằng, những dữ liệu của các ví dụ dẫn ra trong cuốn sách này chỉ là những dữ liệu giả định. Một phần dữ liệu được trích dẫn từ các tài liệu tham khảo số 5, 6, 10, 11... Để tiện cho việc thực hành và kiểm tra kết quả tính toán, một số dữ liệu dùng trong phân tích phương sai, hiệp phương sai và phân tích mối quan hệ giữa các đại lượng có thang đo khác nhau (Chẳng hạn: Sự kết nhóm giữa các loài sinh vật) được ghi lại ở phần phụ lục. Tất cả các ví dụ dẫn ra trong tài liệu này chỉ nhằm mục đích hướng dẫn cách thức ứng dụng lý thuyết thống kê với sự trợ giúp của phần mềm Statgraphics Plus Version 3.0 & 5.1 để xử lý và phân tích các thông tin trong lâm học. Hiện nay phần mềm Statgraphics đã được nâng cấp thành nhiều phiên bản cao hơn. Tuy vậy, bạn đọc vẫn có thể sử dụng những hướng dẫn trong cuốn sách này để xử lý và phân tích các thông tin trong lâm học.

Nhân đây tác giả cũng xin trân trọng cảm ơn TS. Nguyễn Ngọc Kiêng (Giảng viên chính - Khoa lâm nghiệp – Trường ĐHNH Tp. HCM) đã cho phép tác giả sử dụng những số liệu của mình để làm ví dụ minh họa. Sau cùng, tác giả cũng nhận thấy rằng việc biên soạn một tài liệu với những nội dung có liên quan đến nhiều lĩnh vực khoa học khác nhau sẽ không thể tránh khỏi những thiếu sót. Tác giả rất mong nhận được những trao đổi và ý kiến góp ý chân thành của bạn đọc. Những ý kiến đóng góp và trao đổi xin gửi về địa chỉ: PGS. TS. Nguyễn Văn Thêm, ĐT: 0918.204.950; Email: nvthem2009@gmail.com.

Tp. Hồ Chí Minh, tháng 09/2010
Tác giả

Nguyễn Văn Thêm

Chương I

CÁCH NHẬP SỐ LIỆU VÀO BẢNG TÍNH

1.1. CÁC THANG ĐO

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: 6 = Soc, 5 = Cop₃, 4 = Cop₂, 3 = Cop₁, 2 = Sol, 1 = Sp và 0 = 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, chúng 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, chúng 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 =

¹ Case

² Variables

³ Variable name

⁴ Variable label

⁵ Value label

⁶ Nominal

⁷ Nominal variable

⁸ Unordered categories

⁹ Ordinal

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 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... Các biến định danh có thứ bậc được gọi 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: (1) Đường kính thân cây được phân chia thành cấp, trong đó mỗi cấp là 1, 2...k (cm); (2) chiều cao thân cây được phân thành 2 m một cấp...

d. Thang đo tỷ lệ³. Đó là thang đo khoảng với một điểm Zero lấy làm gốc để so sánh. Thang đo này cũng sử dụng cho những biến 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 tỷ 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 thể đo đếm bằng một đơn vị vật lý nào đó được 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 bắt 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 tần suất được biểu thị bằng đơn vị phần trăm thì tổng số 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: (1) bảng phân phối một chiều (bảng 1.1), (2) 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.

¹ Ordered categorical Variable, Controlling Variable, Grouping Variable

² Interval

³ Ratio

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

	Tần số	Tần suất	Tỷ 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

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. Đặc tính này bao gồm 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...k của một đặc tính, còn cột bên cạnh (hoặc hàng dưới) ghi nhận các tần số, tần suất và tỷ lệ phần trăm tương ứng với chúng (bảng 1.3).

Bảng 1.3. Phân bố cây tái sinh Dầu song nành trên mặt đất

Số cây/ô	Tần số	Tần suất	Tỷ 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

Cần lưu ý rằng, trong một vài trường hợp, người ta còn tính các tần số dồn, tần suất dồn và tỷ lệ dồn bằng cách cộng dồn tần số (hay tần suất, tỷ 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 biết có bao nhiêu ô tương ứng với 0, 1...k cây cụ thể, còn cách thứ 2 cho biết có bao nhiêu ô ít hơn (hay nhiều hơn) k cây/ô.

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

Số cây/ô	Tần số đôn	Tần suất đôn	Tỷ 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	Tỷ 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

b. Đặc tính định lượng liên tục. Những đặc tính định lượng 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 chỉ 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 tỷ lệ của các nhóm. Theo Brooks và Carruther, số nhóm (m) và cự ly giữa các nhóm (k) được 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. Theo Hoppman, cự ly giữa các nhóm được tính theo công thức 1.3; trong đó N là dung lượng quan sát.

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

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. Thật vậy, 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 chúng ta dùng dụng cụ đo cao với độ chính xác là 0,5 - 1,0 m. Do đó, chúng 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 thuần loài đồng tuổi, 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 thuần loài đồng tuổi, 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, còn các cột bên cạnh ghi vào tương ứng tần số, tần suất và tỷ 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 lâm phần có thể được ghi như sau (bảng 1.4 và 1.5). 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%.

Bảng 1.4. Phân bố số cây theo cấp đường kính

Cấp D, cm	10	14	18	22	26	30	34	38	42	≥ 46	Σ
N, cây/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	13	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	

1.2. CÁCH NHẬP SỐ LIỆU VÀO BẢNG TÍNH

Số liệu đưa vào tính toán trong bảng tính Statgraphics plus version 3.0 & 5.1 được thực hiện theo một trong những cách dưới đây (*Dấu $\Rightarrow$ được sử dụng để chỉ thứ tự các bước thực hiện những thủ tục trong Statgraphics Plus Version 3.0 & 5.1*).

Cách thứ nhất. Nhập trực tiếp số liệu vào bảng tính Statgraphics Plus Version 3.0 & 5.1. Mỗi cột (hoặc dòng) là một biến số riêng biệt. Kế đến bạn đặt tên cho mỗi biến số bằng cách chọn tiêu đề cột; tiếp theo nhấn nút phải chuột và chọn Modify Column. Tại cửa sổ Modify Column, bạn hãy ghi tên biến vào ô Name. Tên biến là các kí tự (có thể dài hơn 8 kí tự), nhưng không nên bỏ dấu tiếng Việt. Những tên biến này sẽ được Statgraphics Plus Version 3.0 & 5.1 ghi nhận trong phần hiển thị kết quả.

Cách thứ hai. Nhập số liệu vào bảng tính Microsoft Excel, trong đó mỗi cột (hoặc dòng) là một biến số riêng biệt. Sau đó Copy số liệu từ bảng tính Microsoft Excel sang Statgraphics Plus Version 3.0 & 5.1. Những bước tiếp theo như đặt tên biến, chỉnh sửa số liệu... được thực hiện tương tự như cách thứ 1.

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 tính các thống kê mô tả, 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, bạn mở Describe $\Rightarrow$ Numeric data $\Rightarrow$ One - variable analysis. Đến đây, bạn chuyển tên biến cần lọc số liệu vào hộp Data và select ở bên phải bằng cách nhấp nút mũi tên bên cạnh (hình 1.1).

+ Kế đến, bạn đặt con trỏ chuột vào ô Select và mở Transform. Tiếp theo, bạn chọn các toán tử từ ô vuông bên dưới và ghi số cần lọc rồi nhấp OK (hình 1.2).

Tùy theo số lượng biến số cần lọc, bạn có thể thực hiện theo một trong những thủ tục dưới đây.

Trường hợp có một biến số.

Khi có một biến số cần lọc, bạn hãy thực hiện như sau :

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.

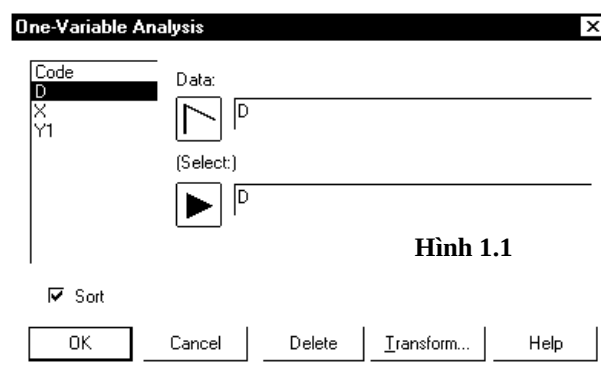
2. Nếu bạn ghi $D > 10 \& H > 8$ thì chương trình chỉ chọn những biến thỏa điều kiện $D > 10$ cm nhưng có $H > 8$ m.

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

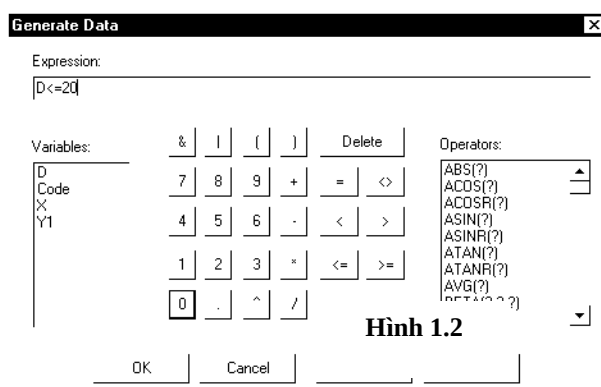
4. Nếu bạn ghi $D = 8 | D = 10$ thì chương trình sẽ chọn những biến $D = 8$ và $D = 10$.

Trường hợp có hai biến số trở lên. Giả sử có hai biến D và H. Yêu cầu đặt ra là chọn những cây có D bất kỳ nhưng $H \leq 20$ cm. Để đạt mục đích đặt ra, trước hết bạn chọn cả hai biến D và H để chuyển vào ô Data. Kế đến chọn biến H rồi chuyển vào ô Select. Tiếp theo bạn đặt con trỏ chuột vào ô Select rồi chọn Transform. Sau cùng bạn chọn $H \leq 20$ trong ô Expression rồi chọn OK. Cách làm cũng tương tự như trên cho trường hợp chọn những cây có D bất kỳ nhưng $H \geq 20$ cm.

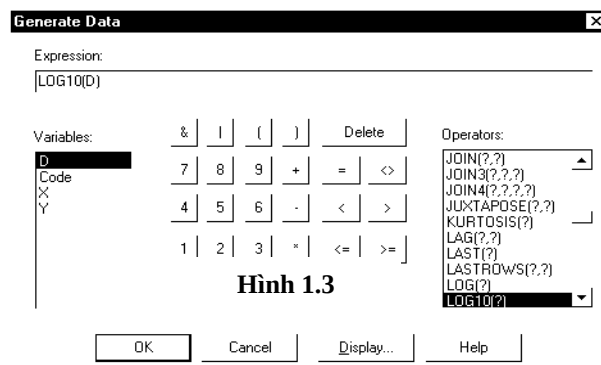
2. Biến đổi dữ liệu. Thủ tục này được đặt ra khi bạn cần phải biến đổi số



Hình 1.1



Hình 1.2



Hình 1.3

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 (*), chia (/), bình phương (**), lũy thừa bậc k (**k), căn bậc 2 (SQRT), logarit tự nhiên (Ln), logarit cơ số 10 (Log10)...Việc biến đổi số liệu được thực hiện theo trình tự sau đây :

+ Chuyển biến số cần biến đổi vào ô Data; kế đến chọn Transform. Tiếp theo bạn chọn hàm số từ ô Functions và nhập mũi tên bên trên để đưa nó vào ô Expression.

+ Khi hàm số xuất hiện trong ô Expression, bạn đưa biến số cần biến đổi vào hàm (nơi có dấu ?) bằng cách nhấp đúp vào biến số trong ô Variable.

+ Sau cùng thì bạn chọn OK (hình 1.3). Đến đây kết quả biến đổi số liệu sẽ được hiển thị.

+ Nếu muốn chọn những thống kê cần quan tâm (thống kê mô tả, bảng tần số...), bạn cần mở thủ tục Tabular Options.

1.4. COPY SỐ LIỆU TỪ BẢNG KẾT QUẢ CỦA STATGRAPHICS PLUS VERSION 3.0 SANG MICROSOFT WORD

Sau khi tính toán xong, bạn có thể Copy các kết quả từ bảng Statgraphics Plus Version 3.0 sang Microsoft Word để tiện cho việc chỉnh sửa. Việc Copy cũng được thực hiện theo cách thông thường. Nhưng lưu ý rằng, khi Paste số liệu từ Statgraphics Plus Version 3.0 sang Microsoft Word thì bạn có thể thực hiện theo ba dạng: (1) SGWIN StatFolios Object, (2) Picture, (3) Unformatted Text. Nếu Paste ở dạng SGWIN StatFolios Object và Picture thì việc chỉnh sửa số liệu trở nên rất khó khăn. Ngược lại, nếu muốn chỉnh sửa số liệu thì bạn phải Paste ở dạng Unformatted Text.

Chương II

PHÂN TÍCH CÁC ĐẶC TRƯNG THỐNG KÊ MẪU

2.1. THỦ TỤC PHÂN TÍCH CÁC ĐẶC TRƯNG THỐNG KÊ MẪU

Statgraphics Plus Version 3.0 phân tích tất cả các đặc trưng thống kê mẫu thông qua những thủ tục dưới đây.

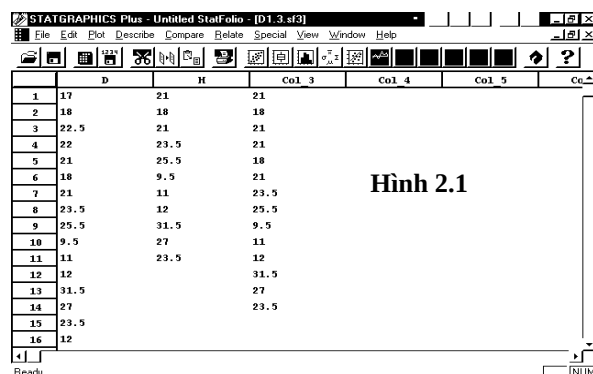
Bước 1. Nhập số liệu, trong đó mỗi cột là một nghiệm thức.

Bước 2. Mở Describe $\Rightarrow$ Numeric Data $\Rightarrow$ One – variable Analysis (hoặc Multiple variable Analysis) $\Rightarrow$ đưa các biến số vào ô Data $\Rightarrow$ OK. Đến đây bạn có thể chọn các đặc trưng thống kê tóm tắt bằng cách mở Tabular Options hoặc Graphics Options hoặc nhấp nút phải chuột và chọn Pane Option. Từ bảng Summary Statistics Options, bạn có thể chọn những thống kê mô tả cần quan tâm sau đây:

+ **Thống kê tóm tắt** (Summary Statistics): X_{bq} , S^2 , $V\%$, độ lệch (Skewness), độ nhọn (Kurtosis), phạm vi biến động (Range)...

+ **Lập bảng tần số.** Thủ tục này cho phép bạn lập bảng tần số (Tabulation) và tính các đặc trưng thống kê cho các đại lượng. Ví dụ: Lập bảng phân bố đường kính (N - D), chiều cao (N - H)... cho từng lâm phần. Để lập một bảng phân bố tần số, bạn cần thực hiện các bước sau đây:

Bước 1. Nhập số liệu. Trước hết, nhấp nút chuột vào đầu cột, chọn Modify Column. Khi cửa sổ Modify Column xuất hiện, bạn hãy ghi tên các biến số vào ô Name và chọn Type kiểu Numeric. Thao tác này giúp bạn ghi tên biến vào đầu cột và các số vào bảng tính một cách dễ dàng. Sau đó bạn vào số liệu như sau: Cột 1 ghi các giá trị của biến số 1; cột 2, 3...ghi tiếp tục giá trị của các biến số khác (hình 2.1). Bạn cũng có thể nhập dữ liệu vào bảng tính Excel, sau đó copy sang Statgraphics Plus Version 3.0 và 5.1.

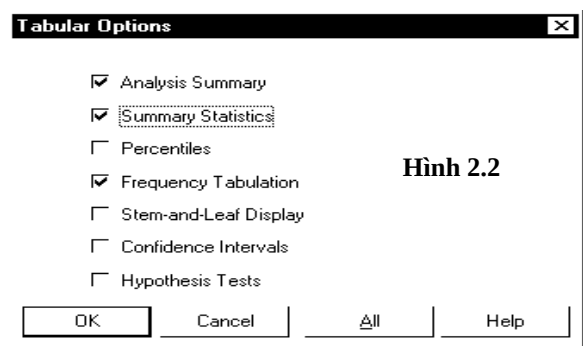


	D	H	Col. 3	Col. 4	Col. 5	Col. 6
1	17	21	21			
2	18	18	18			
3	22.5	21	21			
4	22	23.5	21			
5	21	25.5	18			
6	18	9.5	21			
7	21	11	23.5			
8	23.5	12	25.5			
9	25.5	31.5	9.5			
10	9.5	27	11			
11	11	23.5	12			
12	12		31.5			
13	31.5		27			
14	27		23.5			
15	23.5					
16	12					

Hình 2.1

Bước 2. Chọn Describe $\Rightarrow$ Numeric Data $\Rightarrow$ One Variable Analysis. Tại cửa sổ One Variable Analysis, bạn chọn từng biến số (D, H, G, M) và chuyển nó vào ô Data; sau đó chọn OK.

Bước 3. Mở Tabular Options, chọn Summary Statistics và Frequency Tabulation; cuối cùng chọn OK (hình 2.2). Kết quả sẽ hiển thị các thống kê mô tả, bảng tần số... Tại cửa sổ Frequency Tabulation, bạn có thể nhấp nút chuột phải và chọn Pane Options để thay đổi những thông số theo yêu cầu của bạn. Chẳng hạn, khi lập bảng phân bố N – D, bạn có thể chọn số tổ, cận trên - dưới của chuỗi số liệu sao cho những trị trung tâm (Midpoint) của mỗi tổ rơi vào các trị số chẵn (8, 10, 12...).



Tabular Options	
☒	Analysis Summary
☒	Summary Statistics
☐	Percentiles
☒	Frequency Tabulation
☐	Stem-and-Leaf Display
☐	Confidence Intervals
☐	Hypothesis Tests

OK Cancel All Help

Hình 2.2

Bước 4. Lập biểu đồ tần số (Frequency Histogram)... Để đạt được mục đích đặt ra, bạn hãy mở biểu tượng Graphical Options. Nếu cần vẽ biểu đồ đa giác tần số (Polygon) và biểu đồ tần số tích lũy (Cumulative), thì tại cửa sổ Frequency Histogram bạn nhấp nút chuột phải rồi chọn Polygon và Cumulative.

Dưới đây dẫn ví dụ về nội dung của một bảng phân phối đường kính thân cây được xử lý theo thủ tục lập bảng tần số (bảng 2.1).

Bảng 2.1. Phân bố N - D

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		7.0		2	0.0128	2	0.0128
1	7.0	9.0	8.0	4	0.0256	6	0.0385
2	9.0	11.0	10.0	4	0.0256	10	0.0641
3	11.0	13.0	12.0	8	0.0513	18	0.1154
4	13.0	15.0	14.0	11	0.0705	29	0.1859
5	15.0	17.0	16.0	10	0.0641	39	0.2500
6	17.0	19.0	18.0	26	0.1667	65	0.4167
7	19.0	21.0	20.0	17	0.1090	82	0.5256
8	21.0	23.0	22.0	21	0.1346	103	0.6603
9	23.0	25.0	24.0	18	0.1154	121	0.7756
10	25.0	27.0	26.0	16	0.1026	137	0.8782
11	27.0	29.0	28.0	13	0.0833	150	0.9615
12	29.0	31.0	30.0	5	0.0321	155	0.9936
above	31.0			1	0.0064	156	1.0000

Mean = 20.4679 Standard deviation = 5.64163

2.2. KIỂM ĐỊNH VỀ DẠNG PHÂN BỐ CỦA DÃY SỐ LIỆU

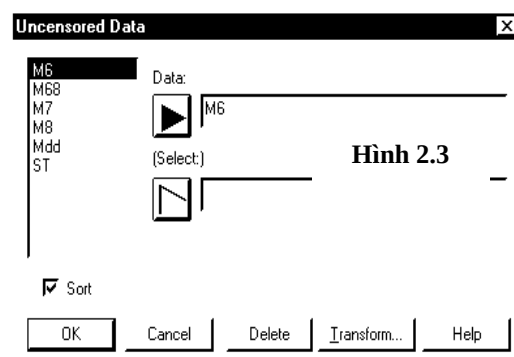
Thủ tục làm phù hợp dạng phân bố của dãy số liệu (Distribution Fitting) cho phép bạn thực hiện không chỉ các kiểm định về phân bố chuẩn (Tests For Normality) và nhiều dạng phân bố khác (trên 20 phân bố khác nhau), mà còn xác định xác suất và mật độ xác suất và các giá trị tiêu chuẩn (Critical value)...Lưu ý rằng, khi mô tả phân bố của một đại lượng nào đó, bạn cần phải báo cáo những thống kê sau đây:

+ Trị trung bình (X_{bq}), trung vị (Me), mode (Mo), trị lớn nhất (X_{max}) và nhỏ nhất (X_{min}), phương sai (S^2) và sai tiêu chuẩn (S), hệ số biến động (V%).

+ Độ lệch của dãy phân bố (S_k) và sai tiêu chuẩn của độ lệch (S_{sk}).

+ Độ nhọn của dãy phân bố (Ex) và sai tiêu chuẩn của độ nhọn (S_{Ex}).

Khi phân bố đối xứng thì $X_{bq} = Me = Mo$. Khi $X_{bq} > Me$ thì phân bố có dạng bất đối xứng. Điều đó chứng tỏ đại lượng nghiên cứu có biên độ biến động lớn. Khi $X_{bq} > Me > Mo$ thì đỉnh phân bố lệch trái hay đuôi phải của phân bố dài hơn. Ngược lại, khi $X_{bq} < Me < Mo$ thì đỉnh phân bố lệch phải hay đuôi trái của phân bố dài hơn. Hệ số S_k dương cho biết đỉnh đường cong lệch trái so với số trung bình. Ngược lại, hệ số S_k âm cho biết đỉnh đường cong lệch phải so với số trung bình. Khi Ex lấy giá trị âm thì đỉnh đường cong tù. Ngược lại, khi Ex lấy giá trị dương thì đỉnh đường cong nhọn.

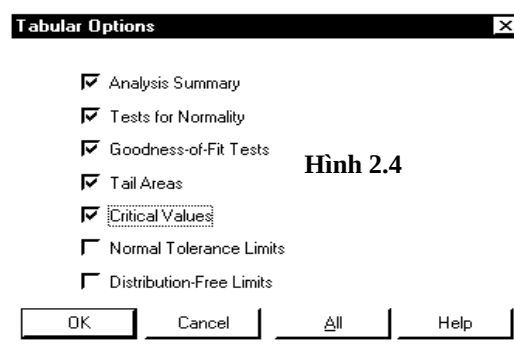


Để thực hiện kiểm định dạng phân bố của một đại lượng quan sát nào đó, bạn hãy thực hiện theo thủ tục sau đây:

Bước 1. Nhập số liệu cần kiểm định dạng phân bố vào bảng tính Statgraphics plus version 3.0. Mỗi biến số được ghi vào một cột hoặc một hàng. Sau đó bạn lưu lại thành một file để tránh mất dữ liệu và có thể sử dụng về sau.

Bước 2. Mở Describe $\Rightarrow$ Distribution Fitting $\Rightarrow$ Uncensored Data (hình 2.3). Đến đây bạn hãy chọn số liệu và nhấp nút mũi tên bên cạnh để chuyển nó vào ô Data. Sau khi đã vào xong số liệu, bạn chọn OK để chương trình xử lý.

Bước 3. Bạn hãy mở hộp thoại Tabular Options (hình 2.4) và chọn những thống kê cần tính toán.



Nghiên cứu cấu trúc rừng cho thấy, phân bố đường kính thân cây (N - D) thay đổi tùy theo kiểu rừng, cấu trúc tuổi lâm phần, biện pháp lâm sinh, quan hệ cạnh tranh giữa những cá thể cây rừng. Tuổi lâm phần tăng lên thì độ nhọn của phân bố sẽ giảm. Cấu trúc lâm phần càng phức tạp thì độ nhọn mang trị số âm càng lớn. Vì thế, xác định kiểu phân bố đường kính thân cây cho các loại rừng là việc làm cần thiết. Dưới đây chúng ta xem xét thủ tục kiểm định một số dạng phân bố N - D phổ biến nhất.

1. Kiểm định phân bố chuẩn (Tests For Normality)

Giả thuyết H_0 : Phân bố thực nghiệm phù hợp với phân bố chuẩn.

Đối thuyết H_0 : Phân bố thực nghiệm không phù hợp với phân bố chuẩn.

Quy tắc quyết định:

Khi P-value $< \alpha$ (0.01, 0.05) thì giả thuyết H_0 bị bác bỏ.

Khi P-value $> \alpha$ (0.01, 0.05) thì giả thuyết H_0 được chấp nhận.

Để nhận được kết quả về kiểm định phân bố chuẩn, bạn hãy thực hiện theo các thủ tục sau đây :

Bước 1. Mở hộp thoại Tabular Options. Khi hộp thoại Tabular Options mở ra, bạn đánh dấu chọn ô Tests For Normality; sau đó nhấp nút OK. Khi kết quả hiển thị, bạn hãy quan sát thống kê phù hợp Chi – Square (Chi-Square goodness-of-fit statistic), Shapiro-Wilks W Statistic và P-value. Khi P-value $< \alpha$ (0.01, 0.05) thì giả thuyết H_0 bị bác bỏ. Ngược lại, khi P-value $> \alpha$ (0.01, 0.05) thì giả thuyết H_0 được chấp nhận.

Việc kiểm định về phân bố chuẩn còn được trình bày thông qua bảng kiểm định tính phù hợp (Goodness – of – fit Tests). Bảng này phân chia dãy số liệu thành K nhóm (cấp) với mỗi nhóm gồm có giới hạn dưới và giới hạn trên. Ngoài ra, trong bảng còn bao gồm cả tần số thực nghiệm và tần số lí thuyết của mỗi nhóm, bên cạnh là kiểm định tính phù hợp với phân bố chuẩn bằng tiêu chuẩn phù hợp Chi – Square. Lưu ý rằng, số nhóm có ảnh hưởng đến kết quả phân tích thống kê và dạng phân bố. Nói chung, số nhóm đường kính không nên nhỏ hơn 7 và lớn hơn 12.

Bạn có thể thay đổi số nhóm được phân chia bằng cách: Tại cửa sổ kiểm định tính phù hợp (Goodness – of – fit Tests), nhấp nút phải chuột và chọn Pane Option. Sau đó bạn điều chỉnh lại số nhóm và xác định cận trên và cận dưới của dãy số liệu sao cho phù hợp với số liệu thực tế; cuối cùng chọn OK. Đến đây một bảng kết quả mới sẽ hiển thị,

trong đó mọi thông số sẽ được tính lại theo số liệu mà bạn vừa điều chỉnh. Bạn hãy quan sát tiêu chuẩn phù hợp Chi – Square và kiểm định Shapiro –Wilks W Statistic để đưa ra kết luận phù hợp.

Bước 2. Xác định xác suất và mật độ xác suất của phân bố chuẩn. Để tìm xác suất của phân bố chuẩn, sau khi mở hộp thoại Tabular Options, bạn hãy chọn mục Tail Areas (hình 2.4). Khi cửa sổ Tail Areas mở ra, bạn đặt con chuột vào cửa sổ Tail Areas for Col-? và nhấp nút phải chuột; tiếp theo chọn Pane Options. Từ cửa sổ Tail Areas Options, bạn hãy ghi những giá trị cần tìm xác suất vào ô Critical Values; sau đó chọn OK. Lưu ý rằng, những giá trị cần tìm xác suất phải là cận trên của mỗi tổ. Nếu cần xác định mật độ xác suất nằm trong một khoảng giữa hai giá trị hay cận trên và cận dưới của mỗi tổ ($X_2 - X_1$) thì bạn chỉ việc lấy hiệu số giữa xác suất tương ứng với giá trị cận trên và cận dưới của tổ đó.

Nếu cần xác định các giá trị tiêu chuẩn (Critical values) thì tại hộp thoại Tabular Options, bạn hãy đặt con chuột vào ô Critical Values for Col-? và nhấp nút phải chuột; tiếp theo chọn Pane Options. Từ cửa sổ Critical Values Options, bạn hãy ghi các giá trị tiêu chuẩn cần tìm vào ô Tail Areas; sau đó chọn OK.

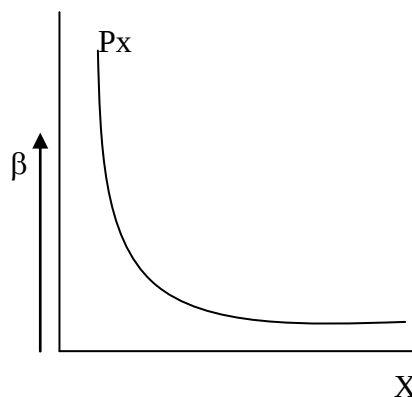
2. Kiểm định phân bố mũ

Mật độ xác suất của phân bố mũ được tính theo công thức:

$$P(x) = \beta e^{-\beta x} \quad (2.1)$$

trong đó β là tham số của phân bố mũ hay hệ số giảm. Dạng đường cong thay đổi tùy thuộc vào giá trị của tham số β . Khi β lấy giá trị càng lớn thì đường cong càng lõm. Khi $X = 0$ thì $P(x) = \beta$. Đường cong phân bố mũ có dạng như hình 2.5. Phân bố mũ có thể được sử dụng để mô tả phân bố N - D của rừng tự nhiên hỗn giao khác tuổi hoặc những lâm phần thuần loài khác tuổi.

Phân bố mũ có những đặc trưng sau đây: Kỳ vọng toán $E(x) = 1/\beta$; phương sai $D(x) = 1/\beta^2$; sai tiêu chuẩn $= 1/\beta$. Do đó khi một phân bố thực nghiệm có $E(x) = 1/\beta$ thì ta có thể mô phỏng phân bố ấy ở dạng phân bố mũ. Mật độ xác suất của phân bố mũ tính theo công thức:



Hình 2.5. Phân bố xác suất của phân bố mũ

$$\begin{aligned} P(X_1 < X < X_2) &= \beta \int_{X_1}^{X_2} e^{-\beta x} dx \\ &= -e^{-\beta X_1} + e^{-\beta X_2} \end{aligned} \quad (2.2)$$

Sau khi xác định được các đặc trưng thống kê của phân bố thực nghiệm, bạn thực hiện việc đồng hóa nó với phân bố mũ. Tính phù hợp của phân bố thực nghiệm với phân bố mũ được thực hiện bằng kiểm định Chi - square. Nếu P-value > 0.05 (mặc định) thì giả thuyết về phân bố mũ được chấp nhận ở mức tin cậy 95% (mặc định).

3. Kiểm định phân bố mũ

Phân bố mũ có dạng:

$$Y_i = \alpha e^{\beta X_i}; \quad (2.3)$$

trong đó Y_i là tần số quan sát, còn X_i là trị số quan sát. Để xác định các tham số của phân bố Mũ, bạn thực hiện phép tuyến tính hóa như sau:

$$\ln Y_i = \ln \alpha - \beta X_i. \text{ Đặt } \ln Y_i = Y, \ln \alpha = A, \beta = B.$$

$$\text{Từ đó ta có } Y = A + BX_i.$$

Từ đây bạn có thể áp dụng phương pháp bình phương nhỏ nhất để tìm các tham số A và B . Sau đó tra đối logarit bạn sẽ nhận được các tham số Y_i , α và β . Sau khi xác định được các đặc trưng thống kê của phân bố thực nghiệm, bạn thực hiện việc đồng hóa nó với phân bố Mũ. Tính phù hợp của phân bố thực nghiệm với phân bố mũ được thực hiện bằng kiểm định Chi - square. Nếu $P\text{-value} > 0.05$ (mức định) thì giả thuyết về phân bố Mũ được chấp nhận ở mức tin cậy 95% (mức định). Phân bố Mũ có thể được sử dụng để mô tả phân bố đường kính của các lâm phần hỗn giao khác tuổi tự nhiên.

4. Kiểm định phân bố Weibull

Bailey và Dell (1973) đã mô tả phân bố đường kính của các lâm phần bằng hàm Weibull. Hàm Weibull có dạng:

$$f(x) = (c/b)(x/b)^{c-1} \exp\{-(x/b)^c\} \quad (2.4)$$

với $c > 0$, $b > 0$, $x \geq 0$; c là tham số đặc trưng cho hình dạng đường cong; b là tham số tỷ lệ. Ở dạng chung nhất, phân bố Weibull bao gồm tham số dịch chuyển điểm đầu toạ độ a :

$$g(x) = (c/b)\{(y-a)/b\}^{c-1} \exp\{-(y-a)/b\}^c, \quad (2.5)$$

$$y \geq a, b > 0, c > 0,$$

ở đây y là đường kính thân cây, $g(x)$ là hàm Weibull, tham số $a \geq 0$ có thể hiểu như là đường kính thân cây nhỏ nhất. Hai mô hình 2.4 và 2.5 có mối liên hệ tương quan với nhau; trong đó $y = x + a$. Vì thế, để xác định phân bố đường kính ta chỉ cần sử dụng mô hình 2.4. Để dễ dàng tính các đặc trưng thống kê của phân bố Weibull, ta lấy tích phân phương trình 2.4 bằng cách thay đổi biến

$$F(x) = \int_0^x (c/b)(t/b)^{c-1} \exp\{-(t/b)^c\} dt = \int_0^x \exp\{-(t/b)^c\} d[(t/b)^c] =$$

$$= \int_0^{(x/b)^c} \exp(-\mu) d\mu = 1 - \exp\{-(X/b)^c\}; \quad (2.6)$$

trong đó $t = X$, X là đường kính thân cây, $\mu = (t/b)^c$. Ngoài ý nghĩa thông thường, hàm phân bố xác suất $F(x)$ còn cho phép xác định các tham số của phân bố. Hai tham số c và b có miền xác định từ $0 < c < +\infty$; $0 < b < +\infty$. Khi $c < 1$ thì đường cong có dạng hình chữ J ngược. Khi $c = 1$ thì phân bố Weibull có dạng phân bố mũ:

$$f(x) = (1/b)\exp(-X/b) \text{ với } X \geq 0, b > 0. \quad (2.7)$$

Giá trị $c < 1$ rất ít gặp. Khi $1 < c < 3,6$ thì hàm phân bố mật độ xác suất có đỉnh lệch trái. Khi $c \approx 3,6$ thì phân bố Weibull tiến đến phân bố chuẩn. Khi $c \gg 3,6$ thì hàm phân bố mật độ xác suất có đỉnh lệch phải. Cần lưu ý rằng tham số c không hoàn toàn xác định hình dạng đường cong. Hình dạng đường cong cũng phụ thuộc vào tham số b . Mặt khác, tham số b còn ảnh hưởng đến độ lồi lõm của đường cong. Khi b có giá trị càng nhỏ thì đỉnh đường cong càng cao. Khi b lấy giá trị càng lớn thì đỉnh đường cong càng bẹt hay đường cong càng trải dài theo trục hoành.

Bailey và Dell lưu ý rằng, chính nhờ khả năng biến đổi hình dạng đường cong theo sự biến đổi các đặc trưng của lâm phần mà phân bố Weibull có thể được ứng dụng để mô tả phân bố đường kính thân cây (N - D) của các lâm phần. Tùy theo đặc trưng lâm phần (thuần loài hay hỗn giao, đồng tuổi hay khác tuổi) mà hình dạng đường cong phân bố N - D có biến đổi khác nhau. Hình dạng đường cong phân bố N - D phụ thuộc vào hai tham số c và b . Trị số c và b phụ thuộc vào nhiều yếu tố như kiểu rừng, tuổi, mật độ và biện pháp chặt nuôi dưỡng rừng... Trường hợp $c < 1$ chỉ gặp ở những lâm phần đã qua chặt nuôi dưỡng. Ở những lâm phần non, phân bố đường kính có đỉnh lệch trái, nghĩa là tham số $1 < c < 3,6$. Ở những lâm phần gần thành thực, phân bố đường kính có đỉnh lệch phải ($c \gg 3,6$). Khi $c \approx 3,6$ thì hàm Weibull tiến đến phân bố chuẩn. Do đó, ở những lâm phần chưa qua tác động, tham số c đặc trưng cho kiểu rừng.

Để làm phù hợp phân bố N - D với hàm Weibull, trước hết bạn cần lưu ý đến việc phân cấp đường kính thân cây. Số cấp có ảnh hưởng đến kết quả phân tích thống kê và hình dạng phân bố. Nói chung, số cấp đường kính không nên nhỏ hơn 7 và lớn hơn 12; tốt nhất là 12 cấp. Số cây phân bố vào mỗi cấp đường kính xác định xác suất P xuất hiện cây trong mỗi cấp. Trong trường hợp này, phương trình 2.6 có dạng :

$$P = F(x) = 1 - \exp\{-(X_p/b)^c\} \quad (2.8)$$

trong đó $X_p = b\{-\ln(1 - P)\}^{1/c}$. Đồ thị của hàm $F(x)$ xác định xác suất của biến đường kính (D) bằng hoặc nhỏ hơn X . Vì thế, khi xác định $F(x)$ thì giá trị của biến D phải lấy ở cận trên của mỗi cấp. Mật độ xác suất $f(x)$ tương ứng với mỗi cấp đường kính được xác định bằng hiệu số xác suất của cận trên và cận dưới trong mỗi cấp đường kính.

Phương pháp ước lượng hai tham số b và c của hàm Weibull phụ thuộc vào hiệu quả thống kê mong đợi và kỹ thuật tính toán. Theo Dubey (1967), khi tham số c chưa biết thì ước lượng tốt nhất của tham số b có thể được xác định từ giá trị đứng ở vị trí thứ 40% và 82% của tài liệu mẫu. Lưu ý là các giá trị mẫu phải được sắp xếp theo thứ tự tăng dần. Trong trường hợp này hiệu quả ước lượng tham số b đạt 82%. Đối với hai giá trị phần trăm mẫu X_{p1} và X_{p2} ($0 < P_1 < 1$; $0 < P_2 < 1$), tham số b được tính theo công thức

$$b = \exp\{[\ln(X_{p1}) - \ln(X_{p2})\ln(-\ln(1 - P_1))/\ln(-\ln(1 - P_2))]/[1 - \ln(-\ln(1 - P_1))/\ln(-\ln(1 - P_2))]\} \quad (2.9)$$

Theo Dubey (1967), nếu sử dụng giá trị đường kính đứng ở vị trí thứ 17% và 97% của tài liệu mẫu để ước lượng tham số c trong khi chưa biết tham số b thì hiệu quả ước lượng c là 66%. Trong trường hợp này tham số c được tính theo công thức:

$$c = \ln\{\ln(1 - P_1)/\ln(1 - P_2)\}/\ln\{X_{p1}/X_{p2}\}. \quad (2.10)$$

Khi đã biết tham số c thì sử dụng giá trị đứng ở vị trí thứ 80% của tài liệu mẫu để ước lượng tham số b là tốt nhất. Hiệu quả ước lượng b đạt 65%. Trong trường hợp này b được tính theo công thức:

$$b = \exp\{\ln(X_{0,80}) - \ln[-\ln(0,20)]/c\}. \quad (2.11)$$

Từ công thức 2.9, 2.10 và 2.11 cho thấy hai tham số b và c có thể được tính từ một bộ phận mẫu hoặc từ hầu hết giá trị mẫu. Khi biết b và c ta có thể tính được hàm phân bố mật độ xác suất $f(x)$ và hàm phân bố xác suất $F(x)$ của phân bố Weibull. Như đã thấy, sau khi dò tìm được các trị số X_{p1} và X_{p2} ($0 < P_1 < 1$; $0 < P_2 < 1$), thì hai tham số b và c có thể được tính dễ dàng chỉ bằng bảng logarit và bảng tính Excel. Hiệu quả ước lượng hai tham số b và c phụ thuộc vào bộ phận mẫu được chọn.

Ngày nay hai tham số b và c đã được lập thành chương trình tính toán trên nhiều phần mềm thống kê khác nhau. Để xác định c và b bằng phần mềm thống kê Statgraphics Plus Version 3.0 và 5.1, bạn có thể thực hiện như sau :

Giả thuyết H_0 : Phân bố thực nghiệm phù hợp với phân bố Weibull.

Đôi thuyết H_0 : Phân bố thực nghiệm không phù hợp với phân bố Weibull.

Quy tắc quyết định:

Khi $P\text{-value} < \alpha$ (0.01, 0.05) thì giả thuyết H_0 bị bác bỏ.

Khi $P\text{-value} > \alpha$ (0.01, 0.05) thì giả thuyết H_0 được chấp nhận.

+ Trước hết, bạn nhập số liệu D vào cột 1 của bảng tính Statgraphics Plus Version 3.0 hoặc 5.1. Kế đến từ menu chính, bạn chọn Describe $\Rightarrow$ Distribution Fitting $\Rightarrow$ Uncensored Data. Đến đây một cửa sổ mở ra như hình 2.6. Tại cửa sổ này bạn chọn biến D và chuyển nó vào ô Data rồi nhấn nút OK để chương trình xử lý.

+ Để kiểm định phân bố $N - D$, bạn chọn Tabular Options và đánh dấu các mục như hình 2.7. Sau khi bạn nhấn OK, các thống kê mô tả và kết quả kiểm định dạng phân bố sẽ hiển thị trên màn hình.

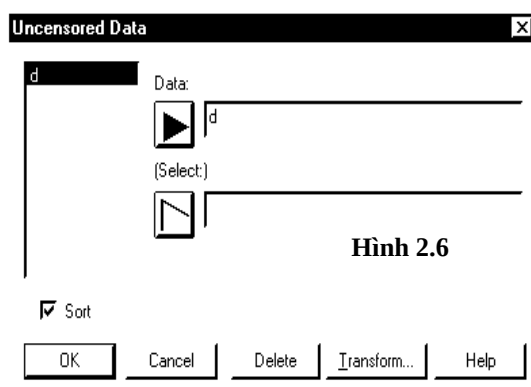
+ Đến đây bạn vào cửa sổ Goodness - of - Fit Tests và nhấn nút phải chuột rồi chọn Analysis Options. Tại cửa sổ Propability Distribution, bạn chọn phân bố Weibull rồi nhấn OK. Một bảng kết quả hiển thị, trong đó số liệu được phân chia thành 2 nhóm. Tại đây bạn cần phải khai báo lại số nhóm, giới hạn dưới và giới hạn trên của dãy phân bố. Thủ tục khai báo như sau:

- Tại cửa sổ kiểm định tính phù hợp (Goodness - of - fit Tests), nhấn nút phải chuột và chọn Pane Option.

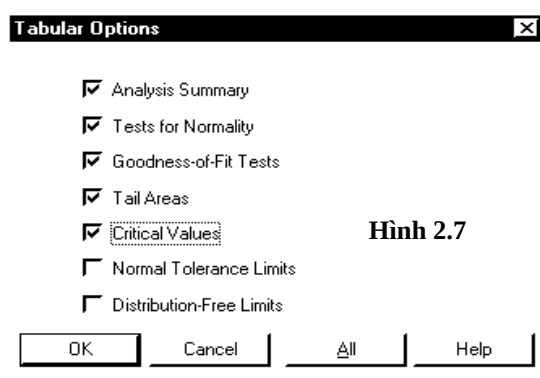
- Khi cửa sổ Pane Option hiện ra, bạn cần khai báo lại số nhóm, cận trên và cận dưới của toàn bộ dãy số liệu thực tế.

- Sau khi khai báo xong, bạn chọn OK. Đến đây một bảng kết quả mới sẽ hiển thị, trong đó báo cáo những nội dung sau đây: giới hạn dưới và giới hạn trên của mỗi nhóm, tần số thực nghiệm và tần số lý thuyết của mỗi nhóm, tiêu chuẩn phù hợp Chi - Square (bảng 2.2).

Từ bảng 2.2, bạn hãy quan sát tiêu chuẩn phù hợp Chi - Square và các thống kê Kolmogorov - Smirnov. Nếu $P\text{-value} < 0.05$ (mặc định) thì giả thuyết H_0 cho rằng dãy số



Hình 2.6



Hình 2.7

liệu phù hợp với phân bố Weibull sẽ bị bác bỏ. Ngược lại, nếu P-value > 0.05 (mặc định) thì giả thuyết H_0 được chấp nhận ở mức tin cậy 95% (mặc định).

Xác định xác suất của phân bố Weibull. Khi cần xác định xác suất và mật độ xác suất của phân bố Weibull (Tail Areas), bạn hãy đặt con chuột vào cửa sổ Tail Areas for D và nhấp nút phải chuột; tiếp theo chọn Pane Options. Từ cửa sổ Tail Areas Options, bạn hãy ghi những giá trị cần tìm xác suất vào ô Critical Values; sau đó chọn OK. Lưu ý rằng những giá trị cần tìm xác suất phải là cận trên của mỗi tổ. Nếu cần xác định mật độ xác suất nằm trong một khoảng giữa hai giá trị hay cận trên và cận dưới của mỗi tổ ($D_2 - D_1$) thì bạn chỉ việc lấy hiệu số giữa xác suất tương ứng với giá trị cận trên và cận dưới của tổ đó.

Bảng 2.2. Kiểm định tính phù hợp của phân bố N - D với phân bố Weibull

Goodness-of-Fit Tests for d

Chi-Square Test					
	Lower Limit	Upper Limit	Observed Frequency	Expected Frequency	Chi-Square
at or below		12.0	12	9.48	0.67
	12.0	16.0	19	21.83	0.37
	16.0	20.0	44	39.27	0.57
	20.0	24.0	43	45.22	0.11
	24.0	28.0	29	30.06	0.04
above	28.0		10	11.13	0.12

Chi-Square = 1.86761 with 3 d.f. P-Value = 0.600332

Estimated Kolmogorov statistic DPLUS = 0.0882776

Estimated Kolmogorov statistic DMINUS = 0.0625758

Estimated overall statistic DN = 0.0882776

Approximate P-Value = 0.173165

Xác định các giá trị tiêu chuẩn (Critical values). Để tìm các giá trị tiêu chuẩn, sau khi mở hộp thoại Tabular Options, bạn hãy đặt con chuột vào ô Critical Values for Col-? và nhấp nút phải chuột; tiếp theo chọn Pane Options. Từ cửa sổ Critical Values Options, bạn hãy ghi các giá trị tiêu chuẩn cần tìm vào ô Tail Areas; sau đó chọn OK. Lưu ý rằng, giá trị tiêu chuẩn phải nằm trong khoảng 0.01 - 0.99.

Đồ thị xác suất và mật độ xác suất của hàm Weibull. Thủ tục này được thực hiện bằng cách chọn các mục Weibull Plot và Frequency Histogram trong biểu tượng Graphical Options.

Thực hành làm phù hợp phân bố N - D với hàm Weibull. Giả sử chúng ta cần làm phù hợp phân bố đường kính của lâm phần Tách 22 tuổi mọc trên đất feralit đỏ vàng phiến sét ở miền Đông Nam Bộ với phân bố Weibull. Đường kính thân cây của lâm phần có các đặc trưng như bảng 2.3 và 2.4. Phân tích số liệu thực nghiệm và các đặc trưng thống kê của bảng 2.3 cho thấy, phân bố N - D của lâm phần Tách 22 tuổi có dạng một đỉnh bất đối xứng ($D_{bq} \neq Me \neq Mo$), đỉnh đường cong hơi lệch phải ($Sk < 0$) và tù ($Ex < 0$). Để làm phù hợp phân bố N - D của lâm phần này với phân bố Weibull, chúng ta có thể phân chia $D_{1.3}$ thành 7 cấp, mỗi cấp cách nhau 4 cm. Ở bảng 2.4 dẫn kết quả làm phù hợp phân bố đường kính của lâm phần Tách 22 tuổi với phân bố Weibull. Các ước lượng của tham số b và c tương ứng là 22,609 và 4,057.

Bảng 2.3. Các đặc trưng của phân bố N - D ở lâm phần Tách 22

Số cây (N)	$D_{1.3bp}$, cm	Me, cm	Mo, cm	s^2	S, cm	Sd, cm	Dmin, cm	Dmax, cm	$\frac{Dmax}{Dmin}$	Sk	S_{sk}	Ex	S_{sex}	V%
156	20,5	21,0	18,0	34,3	5,8	0,47	6,0	34,0	28,0	-0,333	-1,699	-0,3	-0,766	28,6

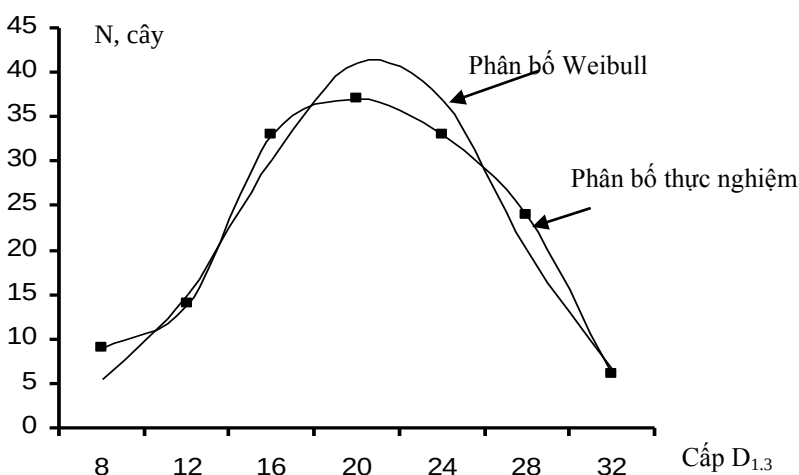
Bảng 2.4. Kiểm định tính phù hợp của phân bố N - D

Giới hạn :		Giữa cấp	Tần số :		χ^2
dưới	trên		thực nghiệm	lí thuyết	
-	10	8	9	5,6	2,07
10	14	12	14	15,2	0,09
14	18	16	33	30,3	0,25
18	22	20	37	41,2	0,43
22	26	24	33	37,0	0,43
26	30	28	24	20,1	0,76
> 30		32	6	6,7	0,07
			156	156	4,10

Từ bảng 2.4, bạn hãy quan sát tiêu chuẩn phù hợp Chi – Square và các thống kê Kolmogorov - Smirnov. Nếu P-value < 0.05 (mặc định) thì giả thuyết H_0 cho rằng phân bố N - D của lâm phần Tách 22 tuổi phù hợp với phân bố Weibull sẽ bị bác bỏ. Ngược lại, nếu P-value > 0.05 (mặc định) thì giả thuyết H_0 được chấp nhận ở mức

tin cậy 95% (mặc định). Ở ví dụ này, vì $\chi^2 = 4,1$ với 4 độ tự do và $P = 0,392$, nên phân bố N - D của lâm phần Tách 22 tuổi phù hợp với phân bố Weibull. Đồ thị mô tả phân bố thực nghiệm và phân bố lí thuyết dẫn ra ở hình 2.8.

Khi cần kiểm định dãy số liệu có phù hợp với một dạng phân bố nào đó hay không (ví dụ phân bố Gamma, Poisson...) thì bạn hãy thực hiện lại tất cả các bước như đã làm với phân bố chuẩn. Một cách khác là bạn đặt con chuột vào bất kì cửa sổ nào của bảng kết quả kiểm định phân bố chuẩn, sau đó nhấp nút phải chuột và chọn Analysis


Hình 2.8. Đồng hóa phân bố N - D của rừng Tách 22 tuổi với hàm Weibull: $b = 22,609$; $c = 4,057$.

Options. Khi cửa sổ Probability Distributions xuất hiện, bạn hãy chọn dạng phân bố cần thử. Những nội dung khác cần tính toán cũng được thực hiện giống như đã làm với phân bố chuẩn.

Chương III

MỐI QUAN HỆ GIỮA HAI BIẾN ĐỊNH TÍNH

Để xác định mối quan hệ giữa những biến số được đo bằng các thang đo định tính, 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: Tính tần số ô bất gặp và không bất gặp một hay nhiều loài cây nào đó trong một quần xã.
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.
3. Phân tích và đo đạc 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...

Để xác định mối quan hệ giữa những biến số, bạn hãy thực hiện nhập số liệu và gọi chương trình tính toán theo các thủ tục sau đây:

Cách thứ nhất. Các tần số của bảng 2*2 (hoặc R*C) đã được thống kê đầy đủ. Những tần số này được ghi vào bảng tính như sau: Hàng 1 ghi a và b, hàng 2 ghi c và d; trong đó cột 1 là a và c, còn cột 2 là b và d. Ví dụ:

a	b	(a + b)
c	d	(c + d)
(a + c)	(b + d)	N

Sau khi vào số liệu, bạn thực hiện gọi thủ tục tính toán như sau: Từ Menu chính, mở Describe $\Rightarrow$ Categorical data $\Rightarrow$ Contingency Tables (bảng số ngẫu nhiên). Tiếp theo đưa số liệu của cột và hàng vào bảng Contingency Tables; sau đó chọn OK. Đến đây kết quả của bảng 2*2 và các thống kê về mối quan hệ giữa các đại lượng sẽ hiển thị trên màn hình.

Cách thứ hai. Khi số liệu còn ở dạng các mã số (1 và 0) thì các tần số bất gặp sự kiện A và B được ghi tương ứng vào cột 1 và cột 2 hay hàng 1 và 2 của bảng tính như sau (ví dụ kiểm định sự kết nhóm giữa hai loài cây A và B):

Loài A	1	0	1	1	0	0	1	0	...
Loài B	1	1	0	0	1	0	0	0	...

Sau khi vào số liệu, bạn gọi thủ tục tính toán như sau: Từ Menu chính, mở Describe $\Rightarrow$ Categorical Data $\Rightarrow$ Crosstabulation. Sau đó vào số liệu theo cột; trong đó cột 1 vào số liệu theo hàng (Row), cột 2 vào số liệu theo cột (Column); cuối cùng chọn OK. Đến đây kết quả của bảng 2*2 và các thống kê về mối quan hệ giữa các đại lượng sẽ hiển thị trên màn hình.

Đối với bảng R*C ô, cách thực hiện cũng tương tự như bảng 2*2. Nhưng lưu ý rằng, các số liệu của R (hàng) và C (cột) được nhập thành R (hàng) và C (cột). Ví dụ: Trường hợp có 5 hàng và 3 cột thì cách vào số liệu như sau:

Hàng	Cột 1	Cột 2	Cột 3
1	12	28	36
2	24	35	18
3	32	10	24
4	12	24	6
5	10	20	12

Dưới đây giới thiệu một số nội dung thống kê được xử lý bằng cách lập bảng chéo hai chiều hoặc nhiều chiều.

3.1. 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 (được gọi là bảng 2*2); (2) Kiểm định về ngang bằng tỷ lệ dọc theo hàng hoặc theo cột của bảng chéo 2 chiều.

3.1.1. 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; 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 2*2 về tần số hoặc tần suất (%) bắt gặp hai loài trên N ô tiêu chuẩn. Sau đó thực hiện phép kiểm định về tính độc lập của chúng bằng kiểm định 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 như sau (trên ví dụ hai loài cây):

+ **Thu thập dữ liệu.** Trước hết, bạn phải bố trí 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”. Những ô tiêu chuẩn xuất hiện loài quan tâm, dù chỉ 1 cá thể, được gọi là ô có (kí hiệu 1), còn những ô không xuất hiện loài quan tâm được gọi là ô có (kí hiệu 0). 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 đảm bảo tần số của các ô thuộc bảng 2*2 ô (hoặc R*C ô) phải lớn hơn 5 hoặc không lớn hơn 25% số ô của bảng chéo có tần số nhỏ hơn 5. Nhiều nghiên cứu cho thấy điều kiện này được đảm bảo khi dung lượng quan sát không nhỏ hơn 200 ô tiêu chuẩn.

+ **Cách nhập dữ liệu vào bảng tính.** 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 Zero. Sau đó, bạn có thể nhập số liệu này vào bảng tính theo hai cách: (1) nhập trực tiếp vào bảng tính Statgraphics Plus Version 3.0 hoặc 5.1; (2) nhập vào bảng tính Excel rồi Copy sang bảng tính Statgraphics Plus

Version 3.0 hoặc 5.1 (bảng 3.1). Trong bảng 3.1, 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).

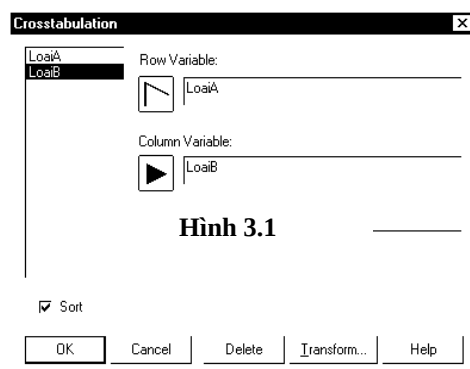
+ **Gọi chương trình tính toán.** Sau khi nhập xong dữ liệu vào bảng tính Statgraphics Plus Version 3.0 hoặc 5.1, bạn thực hiện gọi chương trình tính toán. Để tạo được kết quả, từ thanh Menu chính bạn hãy thực hiện như sau (kí hiệu $\Rightarrow$ chỉ thứ tự chọn các thủ tục tính toán):

Bước 1. Mở Describe $\Rightarrow$ Categorical Data $\Rightarrow$ Crosstabulation. Khi bảng Crosstabulation xuất hiện (hình 3.1), bạn hãy nhập loài A (cột 1) vào ô Row, nhập loài B (cột 2) vào ô Column.

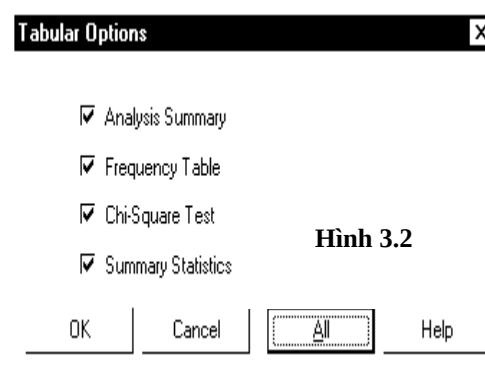
Bước 2. Từ cửa sổ Crosstabulation - LoaiA by LoaiB, bạn hãy chọn Tabular Option rồi chọn tiếp All và OK ở cửa sổ tiếp theo (hình 3.2). Các lệnh này yêu cầu chương trình lập bảng tần số, thực hiện kiểm định Chi – Square và tính các thống kê kiểm định về tính độc lập và quan hệ giữa hai loài. Đến đây kết quả hiển thị trên màn hình bao gồm: (1) tóm tắt xử lý các trường hợp (bảng 3.2), (2) bảng chéo hai chiều (bảng 3.3), (3) kiểm định Pearson Chi – Square (bảng 3.4).

Bảng 3.1. Cách thức nhập dữ liệu vào bảng tính Statgraphics Plus Version 3.0

Thứ tự	Loài cây A	Loài cây B
1	1	1
2	1	0
3	0	1
...	...	...



Hình 3.1



Hình 3.2

Bảng 3.2. Tóm tắt xử lý các trường hợp (Case Processing Summary)

LoaiA * loaiB	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
	120	100.0%	0	0%	120	100.0%

Giải thích và nhận xét kết quả. Bảng 3.2 ghi lại số lượng và phần trăm các trường hợp được đưa vào tính toán. Từ số và mẫu số của các ô thuộc bảng 3.3 cho biết tương ứng tần số và phần trăm số ô bắt gặp hay không bắt gặp hai loài cây trên cùng một ô đo đếm.

Bảng 3.4 dẫn kết quả kiểm định về tính độc lập của hai loài cây (sự kiện) bằng thống kê Pearson Chi – Square. Căn cứ vào kết quả của bảng 3.4, bạn có thể đi đến kết luận về tính độc lập của hai biến.

Giả thuyết H_0 : Hai loài sống độc lập với nhau.

Đối thuyết H_0 : Hai loài có quan hệ (kết nhóm) với nhau.

Quy tắc quyết định

Nếu $P^1 < \alpha$ (0.05 hay 0.01) thì giả thuyết H_0 bị bác bỏ. Ngược lại, nếu $P > \alpha$ (0.05 hay 0.01) thì giả thuyết H_0 được chấp nhận. Bậc tự do của phân bố $\chi^2 = (\text{số hàng} - 1) * (\text{số cột} - 1)$. Ở ví dụ này, vì $\chi^2_{tt} = 8.09$ với $P = 0.0044 < 0.01$, nên hai loài có quan hệ với nhau.

Bảng 3.3. Bảng chéo hai chiều (LoaiA * LoaiB Crosstabulation)

Frequency Table for LoaiA by LoaiB

	0	1	Row Total
0	25	21	46
	20.83	17.50	38.33
1	21	53	74
	17.50	44.17	61.67
Column	46	74	120
Total	38.33	61.67	100.00

Bảng 3.4. Kiểm định Pearson Chi – Square (Chi-Square Tests)

Chi-Square Test

Chi-Square	Df	P-Value
8.09	1	0.0044
7.03	1	0.0080 (with Yates' correction)

Lưu ý rằng, bên cạnh kiểm định tính độc lập của hai biến bằng kiểm định Chi – Square, bảng 3.4 còn đưa ra kiểm định Yates' Correction (hệ số hiệu chỉnh Yates). Kiểm định Yates' Correction được dùng thay thế cho kiểm định χ^2 trong trường hợp dung lượng quan sát nhỏ. Kiểm định Yates' Correction chỉ thích hợp với bảng 2*2.

3.1.2. Kiểm định ngang bằng về tỷ lệ

Giả sử bạn muốn biết tỷ 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 2*2, bạn có thể dùng tiêu chuẩn χ^2 để kiểm định ngang bằng về tỷ lệ (Equality of Proportions).

Giả thuyết H_0 : Hai nhóm cây có tỷ lệ bệnh bằng nhau.

¹ P = Asymp. Sig. (2-sided) – Mức ý nghĩa phù hợp (hai chiều)

Đối thuyết H_0 : Hai nhóm cây có tỷ lệ bệnh khác nhau.

Quy tắc quyết định. Nếu mức ý nghĩa $P < \alpha$ (0.05 hoặc 0.01) thì giả thuyết H_0 bị bác bỏ. Ngược lại, khi mức ý nghĩa $P > \alpha$ (0.05 hoặc 0.01) thì giả thuyết H_0 được chấp nhận. Bậc tự do của phân bố $\chi^2 = (\text{số hàng} - 1) * (\text{số cột} - 1)$.

Khi sử dụng kiểm định χ^2 về ngang bằng tỷ lệ, bạn có thể thực hiện như sau:

+ **Thu thập dữ liệu.** Trước hết, hãy phân chia cây rừng thành hai nhóm. Nhóm một là những cây thuộc cấp sinh trưởng tốt (cây cấp III - I). Nhóm hai là những cây thuộc cấp cây sinh trưởng kém (cây cấp IV - V). Tiếp theo, bạn hãy 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 2*2 phải lớn hơn 5 hoặc ít nhất có 75% số ô có tần số lớn hơn 5.

+ **Cách nhập dữ liệu, gọi chương trình tính toán và nhận xét kết quả.** Số liệu có thể được nhập trực tiếp vào bảng tính Statgraphics Plus Version 3.0 và 5.1 hoặc nhập vào bảng tính Excel rồi Copy sang Statgraphics. Cột 1 ghi thứ tự cây, cột 2 nhập cây nhóm 1 (theo mã hóa); cột 3 nhập cây nhóm 2 (theo mã hóa) (bảng 3.5).

Bảng 3.5. Cách thức nhập dữ liệu để kiểm định ngang bằng tỷ lệ (tóm tắt)

Thứ tự	Cây nhóm I	Cây nhóm II
1	1	1
2	1	0
3	0	1
4	0	0
5...	1	1

Sau khi nhập xong dữ liệu, để tạo được kết quả, từ thanh Menu chính bạn hãy vào đường lệnh sau: **Mở Describe $\Rightarrow$ Categorical Data $\Rightarrow$ Crosstabulation**. Khi bảng Crosstabulation xuất hiện, bạn hãy nhập nhóm I (cột 1) vào ô Row, nhập nhóm II (cột 2) vào ô Column. Tiếp theo, từ cửa sổ Crosstabulation, bạn chọn Tabular Option rồi chọn tiếp All và OK ở cửa sổ tiếp theo. Các lệnh này yêu cầu chương trình lập bảng tần số, thực hiện kiểm định Chi – Square và tính các thống kê kiểm định về tỷ lệ ngang bằng giữa hai đại lượng. Ở ví dụ này, bạn chỉ cần chú ý đến bảng tần số (bảng 3.6) và kiểm định Chi – Square (bảng 3.7).

Giải thích và nhận xét kết quả. Bảng chéo hai chiều (bảng 3.6) báo cáo tần số bất gặp hay không bất gặp hai nhóm cây. Mẫu số trong các ô của bảng này là trị số phần trăm của tần số ô tương ứng so với tổng số. Kết quả kiểm định ngang bằng về tỷ lệ báo cáo những bảng có nội dung cũng giống với những bảng của kiểm định về tính độc lập. Từ kết quả của các bảng 3.6 và 3.7, bạn hãy căn cứ vào giá trị của thống kê Pearson Chi - Square, 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 tỷ lệ. Ở ví dụ này, vì Chi-Square = 203.87 với P = 0.0000 nên hai nhóm cây có tỷ lệ bị bệnh khác nhau.

Bảng 3.6. Tần số xuất hiện bệnh của hai nhóm cây I và II

Frequency Table for NHOMBENHI by NHOMBENHII

	0	1	Row Total
0	126 57.27	2 0.91	128 58.18
1	2 0.91	90 40.91	92 41.82
Column Total	128 58.18	92 41.82	220 100.00

Bảng 3.7. Kiểm định Chi – Square về ngang bằng tỷ lệ

Chi-Square Test

Chi-Square	Df	P-Value
203.87	1	0.0000
199.93	1	0.0000 (with Yates' correction)

3.2. MỐI QUAN HỆ GIỮA HAI BIẾN ĐỊNH TÍNH KHÔNG CÓ THỨ BẬC

(1) Điều kiện áp dụng. Một là, 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. Hai là, việc kiểm định quan hệ giữa 2 biến định tính không có thứ bậc được thực hiện bằng bảng 2*2 hoặc bảng R*C ô.

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 mối quan hệ giữa hai biến định tính không có thứ bậc 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 làm rõ hai câu hỏi sau đây:

+ Một là, hai đại lượng (hai loài cây) có quan hệ với nhau hay không ? Câu hỏi này được làm rõ bằng kiểm định Pearson Chi – Square. Nếu $P > \alpha$ (0.05 hay 0.01) thì hai đại lượng có quan hệ với nhau. Ngược lại, $P < \alpha$ (0.05 hay 0.01) thì hai đại lượng độc lập với nhau.

+ Hai là, nếu hai đại lượng có quan hệ với nhau thì cường độ của mối quan hệ giữa hai đại lượng mạnh hay yếu ? Câu hỏi này được làm rõ thông qua các đo đặc về xu hướng mối quan hệ (Directional Measures) và các đo đặc về tính đối xứng (Symmetric Measures) trong cách sắp xếp của hai đại lượng của bảng 2*2 (hay R*C ô). Nếu các trị số của cả hai biến mang cùng dấu thì hai cặp biến ấy được gọi là phù hợp hay đối xứng (tương đồng). Khi điều kiện đó không thỏa mãn, người ta gọi các cặp biến đó là không phù hợp hay không đối xứng.

(3) Thực hành. Để thực hành thủ tục kiểm định mối quan hệ giữa hai biến định tính không có thứ bậc, chúng ta sử dụng lại ví dụ ở mục 3.1.1. 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). Kết quả kiểm định mối quan hệ giữa hai biến sẽ hiển thị 4 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 mối quan hệ (Directional Measures) và các đo đặc về tính đối xứng (Symmetric Measures) trong cách sắp xếp của hai đại lượng của bảng 2*2 (hay R*C ô). 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. Ở mục 3.1.1, kiểm định cho thấy hai loài có quan hệ với nhau. Vì thế, để làm rõ cường độ của mối quan hệ giữa hai biến bạn cần phải tiếp tục phân tích nội dung của các đo đặc về xu hướng mối quan hệ và các đo đặc về tính đối xứng giữa hai biến số (bảng 3.8).

Bảng 3.8a. Các hệ số đo xu hướng (Derectional Measures) giảm tỷ lệ sai số

Summary Statistics

Statistic	Symmetric	With Rows Dependent	With Columns Dependent
Lambda	0.0870	0.0870	0.0870
Uncertainty Coeff.	0.0504	0.0504	0.0504

Bảng 3.8b. Đo đặc về tính đối xứng trong cách sắp xếp của hai đại lượng

Statistic	Value	P-Value	Df
Contingency Coeff.	0.2514		
Cramer's V	0.2597		

Giải thích và nhận xét kết quả. Khi các biến số là những biến định danh không có thứ bậc hơn kém thì Statgraphics Plus Version 3.0 cung cấp hai nhóm đo đặc sau đây:

+ **Nhóm 1.** Các hệ số đo xu hướng (Derectional Measures) của mức giảm tỷ lệ sai số khi chọn biến phụ thuộc khác nhau (bảng 3.8a). Nhóm này bao gồm các hệ số Lambda và Uncertainty coefficient (hệ số đo mức không chắc chắn của mối quan hệ).

Thống kê Lambda cho biết những giá trị của một biến có tập trung quanh những giá trị của một số biến số nào đó hay không. Thống kê Lambda tính theo công thức:

$$\lambda = \frac{(\text{Sai số trước} - \text{Sai số sau})}{\text{Sai số trước}} \quad (3.1)$$

Trong công thức 3.1, sai số trước là số những sai lầm phạm phải khi dự đoán các trị số của biến phụ thuộc mà không xem xét đến biến độc lập. Sai số sau là số những sai lầm phạm phải khi dự đoán các trị số của biến phụ thuộc mà có xem xét đến biến độc lập.

+ **Nhóm 2.** Các đo đặc về tính đối xứng (Symmetric Measures) trong cách sắp xếp của hai đại lượng của bảng 2*2 (hay R*C ô). Các đo đặc này cũng dựa trên thống kê Pearson Chi – square, bao gồm thống kê Contingency coefficient (hệ số ngẫu nhiên) và thống kê Cramer's V (bảng 3.8b).

Hệ số ngẫu nhiên (Coefficient of Contingency) tính theo công thức:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}} \quad (3.2)$$

Giá trị C thay đổi trong khoảng $0 \leq C \leq 1$.
 Thống kê Cramer V tính theo công thức:

$$V = \sqrt{\frac{\chi^2}{N(k-1)}} \quad (3.3)$$

Các thành phần trong công thức 3.2 và 3.3 có nghĩa như sau:

$$\chi^2 = \frac{(F_{\text{quan sát}} - F_{\text{lý thuyết}})^2}{F_{\text{lý thuyết}}}$$

$$F_{\text{lý thuyết}} = \frac{R_i * C_j}{N} = (\text{Tổng số hàng} * \text{tổng số cột}) / \text{Tổng số quan sát};$$

N = số quan sát trong mẫu;

K là số nhỏ nhất trong số hàng hoặc số cột của bảng R*C. Khi số hàng bằng số cột thì k bằng số hàng hay số cột. Khi số hàng nhỏ hơn số cột thì k bằng số hàng, và ngược lại. Hệ số V nhận giá trị $0 \leq V < 1$.

Những thống kê đo đặc mức giảm tỷ lệ sai số

(Ap dụng cho các biến phân nhóm không có thứ bậc)

+ **Các đo đặc về xu hướng** (Derectional 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). Những đo đặc này bao gồm các hệ số Lambda và Uncertainty. Chúng lấy các giá trị từ $0 \div 1$. Giá trị 0 cho biết biến độc lập không có quan hệ với biến phụ thuộc (hay biến phụ thuộc không thể được dự báo 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 báo chính xác bằng biến độc lập).

Lưu ý rằng, khi xác định mối quan hệ giữa 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 lấy biến A là biến độc lập để dự đoán (xác định) biến B thì sai số dự đoán (tính theo %) có thể khác với trường hợp bạn dùng biến B là biến độc lập để dự đoán biến A. Mức sai số được trình bày ở hai cột cuối của bảng Derectional Measures (bảng 3.8a).

Thống kê Lambda đo mức độ quan hệ giữa hai biến. Nó cho biết biến 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. Ở ví dụ này, từ bảng 3.8a cho thấy, nếu sử dụng loài A (biến độc lập, ở đây là biến dòng) để dự đoán loài B (biến phụ thuộc, ở đây là biến cột) thì mức giảm tỷ lệ sai số là 0.087 hay 8.7% ($\text{Lambda} = 0,087$). Nếu lấy loài B để dự đoán loài A thì tỷ lệ giảm sai số cũng là 8.7%. Bạn hãy quan sát P – Value (trong bảng là Approx. Sig) để biết mức ý nghĩa của sự giảm sai số trong khi dự đoán. Nếu $P < 0.05$ (mặc định) thì biến cột với biến hàng có quan hệ với nhau. Ở ví dụ trên đây thống kê P - value $\ll 0,05$. Điều đó cho biết mối quan hệ giữa hai loài có ý nghĩa thống kê với mức tin cậy rất cao (95%). Tùy theo mục tiêu nghiên cứu, bạn cần chọn lựa biến dự báo sao cho mức giảm tỷ lệ trong sai số dự báo 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 sai số dự đoán. Một biến được chọn là biến dự đoán nếu nó cho tỷ lệ giảm sai số trong dự đoán là cao hơn.

+ **Các đo đặc tính đối xứng hay cân đối** (Symmetric Measures). Các đo đặc này cũng đo mức độ quan hệ giữa các biến hoặc đo mức giảm tỷ lệ sai số (PRE = Proportional Reduction in Error). Chúng bao gồm hệ số ngẫu nhiên (Contingency Coefficient) và thống kê Cramer's V (bảng 3.8b). Hệ số ngẫu nhiên và thống kê Cramer's V được đo bằng các thang đo khác nhau.

- (a) Thống kê Cramer's V lấy giá trị từ 0 ÷ 1. Giá trị Cramer's V = 0 cho biết hai biến không có quan hệ. Ngược lại, giá trị Cramer's V = 1 cho biết hai biến có quan hệ với nhau rất chặt chẽ. Thống kê Cramer's V được dùng phổ biến nhất.
- (b) Hệ số Contingency cũng lấy giá trị từ 0 ÷ 1, nhưng không bao giờ đạt đến 1.
- (c) Hệ số ngẫu nhiên và thống kê Cramer's V lấy giá trị 0 khi và chỉ khi kiểm định Pearson $\chi^2 = 0$.

Lưu ý rằng, thống kê χ^2 chỉ trả lời được câu hỏi là hai yếu tố có mối quan hệ với nhau hay không; nó không cho biết cường độ (mức độ) của mối quan hệ. Khi hai yếu tố 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 và hệ số Contingency. Trong sinh thái học, người ta còn kiểm định mức độ chặt chẽ của mối quan hệ giữa các loài sinh vật bằng nhiều kiểm định khác như kiểm định Iul, Cole, Sorensen...

(4) 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 kết nhóm (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 đại lượng này đến quan hệ của những 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, chúng ta cần tính sự kết nhóm giữa 3 loài cây. Để đạt mục đích đặt ra, trước hết chúng ta cần thống kê sự có mặt hay vắng mặt của 3 loài cây quan tâm trên các ô nghiên cứu. Theo Cole, để tính hệ số kết nhóm riêng phần giữa 3 loài cây, số lượng ô nghiên cứu cần phải có ít nhất là 250 ô. Điều kiện này được đặt ra là vì những 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 thì không được lớn hơn 25%.

Bước 2. Để tính hệ số kết nhóm riêng phần, chúng ta 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, chúng ta 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 C_1 hay còn gọi là hệ số kết nhóm đa phần. Tương tự như thế, chúng ta tiếp tục 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 (ở đây là loài 3) 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:

$$T = \frac{C_2 - C_1}{\sqrt{\delta_{C1} + \delta_{C2}}} ; \quad (3.4)$$

+ δ_{C1} là sai số của hệ số kết nhóm trên những ô có loài thứ 3 xuất hiện. Sai số δ_{C1} được tính theo công thức:

$$\delta_{C1} = \sqrt{\frac{(a_1+b_1)*(b_1+d_1)}{(N_1*(a_1+c_1)*(c_1+d_1))}} \quad (3.5)$$

+ δ_{C2} là sai số của hệ số kết nhóm trên những ô không có loài thứ 3 xuất hiện. Sai số δ_{C2} được tính theo công thức:

$$\delta_{C2} = \sqrt{\frac{(a_2+b_2)*(b_2+d_2)}{(N_2*(a_2+c_2)*(c_2+d_2))}} \quad (3.6)$$

Trong công thức 3.5 và 3.6, 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 là 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 là 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 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.

(5) Kiểm định sự kết nhóm đa biến

Như đã biết, một quần xã thực vật có rất nhiều loài cây, giữa chúng tồn tại những mối quan hệ rất phức tạp. Vì thế, nhiệm vụ đặt ra không chỉ phân tích quan hệ của loài cây này với loài cây khác mà còn phải nghiên cứu quan hệ của 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.

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 một trong hai loài 2 và 3 được gọi là ô không có (kí hiệu 0).

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.

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.

Thực hành. Ở phụ biểu 17 dẫn ví dụ về kết quả điều tra sự kết nhóm của bốn loài (Trắc, Hà nu, Chò xốt và Lành ngạnh) trên 279 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 xốt (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 xốt (loài 2). Để giải quyết hai nội dung đặt ra trên đây, bạn có thể thực hiện các thủ tục như dưới đây.

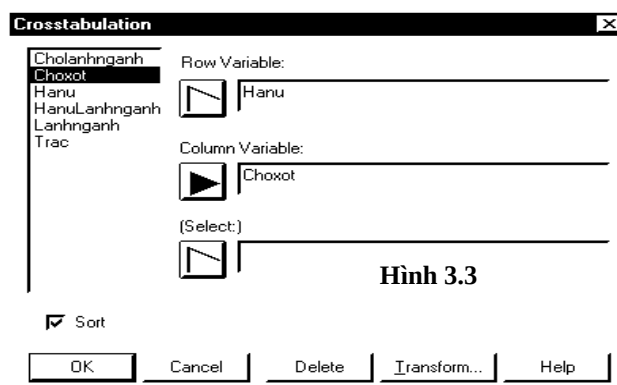
Bước 1. Nhập số liệu. Số liệu bắt gặp hay không bắt gặp bốn loài cây được ghi thành mỗi cột hay mỗi hàng một loài. Cột thứ 5, 6...ghi lại số ô cùng bắt gặp hay không bắt gặp từng cặp hai hoặc ba loài nào đó.

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 Describe ⇒ Categorical Data ⇒ Crosstabulation. Khi cửa sổ Crosstabulation xuất hiện, bạn đưa số liệu của loài 1 (ở đây là Hà nu) vào ô hàng (Row variable), còn loài 2 (ở đây là Chò xốt) vào ô cột (Column variable); rồi chọn OK. Kế đến, bạn chọn Tabular options và đánh dấu chọn All hoặc chọn từng thống kê Chi - square test, Frequency table, Summary statistics.

Để biết hai loài Hà nu và Chò xốt có kết nhóm (quan hệ) với nhau hay không, bạn hãy quan sát thống kê Chi - square test (bảng 3.10). Giả thuyết H₀: Hà nu và Chò xốt không có sự kết nhóm với nhau. Đối thuyết H₀⁻: Hà nu và Chò xốt có sự kết nhóm với nhau. Khi P < 0,05 hoặc 0,01 thì giả thuyết H₀ bị bác bỏ. Ở ví dụ này, vì P = 0,0009 < 0,05 nên hai loài có sự kết nhóm với nhau.

Nếu hai loài (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 hãy quan sát các thống kê Contingency coefficient, Phi and Cramer's V, Lambda, Uncertainty coefficient trong cửa sổ Summary Statistics (bảng 3.11).

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 xốt. Lưu ý rằng, kết quả ở bước 2 cho biết hệ số kết nhóm giữa từng cặp loài.



Hình 3.3

¹ Ecological groups

Bảng 3.10. Kiểm định sự kết nhóm giữa Hà nu với Chò xót

Chi-Square Test

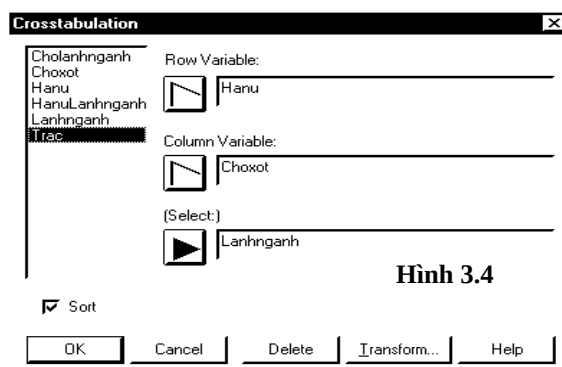
Chi-Square	Df	P-Value
11.04	1	0.0009
10.23	1	0.0014 (with Yates' correction)

Bảng 3.11. Cường độ kết nhóm giữa Hà nu với Chò xót

Summary Statistics		Hà nu	Chò xót
Statistic	Symmetric	With Rows Dependent	With Columns Dependent
Lambda	0.0000	0.0000	0.0000
Uncertainty Coeff.	0.0299	0.0300	0.0298
Somer's D	-0.1989	-0.1981	-0.1998
Eta		0.1989	0.1989
Statistic	Value	P-Value	Df
Contingency Coeff.	0.1951		
Cramer's V	0.1989		
Conditional Gamma	-0.4037		
Pearson's R	-0.1989	0.0004	277
Kendall's Tau b	-0.1989	0.0009	
Kendall's Tau c	-0.1915		

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 (ở đây là Hà nu) vào ô Row variable, loài 2 (ở đây là Chò xót) vào ô Column, còn loài 3 (ở đây là Lành ngạnh) vào ô select (hình 3.4); kế đến chọn OK. Tiếp theo bạn chọn Tabular options và đánh dấu chọn All hoặc chọn từng thống kê Chi - square test, Frequency table, Summary statistics. Ở bảng 3.12 - 3.14 dẫn kết quả phân tích sự kết nhóm riêng phần giữa Hà nu với Chò xót trên những ô mẫu bất gặp và không bất gặp Lành ngạnh.

Bước 4. Kiểm định sự khác biệt giữa hệ số kết nhóm của loài 1 và 2 trên những ô không có loài 3 với hệ số kết nhóm của loài 1 và 2 trên những ô có loài 3 theo công thức 3.1. Nếu $T_{tt} < T_{0.05} = 2,0$ 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 $T > T_{0.05}$ 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 3.4

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 chúng 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.

Bảng 3.12. Sự kết nhóm riêng phần giữa Hà nu với Chò xốt khi có mặt Lành ngành

Frequency Table for Hanu by Choxot

	0	1	Row Total
0	21	51	72
	13.91	33.77	47.68
1	46	33	79
	30.46	21.85	52.32
Column Total	67	84	151
	44.37	55.63	100.00

(*) Mẫu số là tần số lí thuyết

Bảng 3.13. Kiểm định tính độc lập giữa Hà nu với Chò xốt khi có mặt Lành ngành

Chi-Square Test

Chi-Square	Df	P-Value
12.89	1	0.0003
11.74	1	0.0006 (with Yates' correction)

Bảng 3.14. Mối quan hệ giữa Hà nu với Chò xốt khi có mặt Lành ngành

Summary Statistics		Hà nu	Chò xốt
Statistic	Symmetric	With Rows Dependent	With Columns Dependent
Lambda	0.2230	0.2500	0.1940
Uncertainty Coeff.	0.0630	0.0628	0.0633
Somer's D	-0.2922	-0.2937	-0.2906
Eta		0.2922	0.2922
Statistic	Value	P-Value	Df
Contingency Coeff.	0.2804		
Cramer's V	0.2922		
Conditional Gamma	-0.5439		
Pearson's R	-0.2922	0.0001	149
Kendall's Tau b	-0.2922	0.0003	
Kendall's Tau c	-0.2900		

Bảng 3.15. Kiểm định tính độc lập giữa Hà nu với Chò xót khi có mặt Lành ngạnh

Chi-Square Test

Chi-Square	Df	P-Value
21.98	1	0.0000
20.74	1	0.0000 (with Yates' correction)

Bảng 3.16. Xu hướng quan hệ giữa Hà nu với tổ hợp loài Chò xót - Lành ngạnh

Summary Statistics		Hà nu	Chò xót+Lành ngạnh
Statistic	Symmetric	With Rows Dependent	With Columns Dependent
Lambda	0.0923	0.1622	0.0000
Uncertainty Coeff.	0.0608	0.0580	0.0638
Somer's D	-0.2801	-0.2995	-0.2630
Eta		0.2807	0.2807
Statistic	Value	P-Value	Df
Contingency Coeff.	0.2702		
Cramer's V	0.2807		
Conditional Gamma	-0.5533		
Pearson's R	-0.2807	0.0000	277
Kendall's Tau b	-0.2807	0.0000	
Kendall's Tau c	-0.2521		

Bước 5. Phân tích sự kết nhóm đa phần. Để phân tích sự kết nhóm giữa Hà nu với tổ hợp hai loài Chò xót và Lành ngạnh, trước hết bạn chọn biến Hà nu và chuyển nó vào ô Row variable; kế đến chọn biến Chò xót + Lành ngạnh (những ô cùng xuất hiện Chò xót và Lành ngạnh) và chuyển nó vào ô Column variable; tiếp theo chọn OK. Đến đây bạn lại chọn Tabular Option, rồi chọn All. Từ bảng Frequency table, thống kê Chi-square test (bảng 3.15) và Summary statistics (bảng 3.16), bạn có thể rút ra những nhận xét về quan hệ giữa Hà nu với tổ hợp hai loài Chò xót + Lành ngạnh.

Từ kết quả của bảng 3.15, chúng ta có thể nhận thấy phân bố của Hà nu có mối liên hệ thực sự với tổ hợp hai loài Chò xót và Lành ngạnh ($\chi^2_{(1)} = 21,98$; $P = 0,000$). Từ số liệu của bảng 3.16 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 xót và Lành ngạnh thì mức giảm sai số là zero phần trăm. Ngược lại, nếu dựa trên sự cùng có mặt của Chò xót và Lành ngạnh trong ô mẫu để dự đoán sự có mặt của Hà nu thì mức giảm sai số là 16,2%. Mức độ quan hệ giữa Hà nu với tổ hợp hai loài Chò xót - Lành ngạnh là chặt chẽ (hệ số Cramer's V = 0,281; $P = 0,000$).

Lưu ý rằng, bạn có thể thực hiện các thủ tục như trên để xem xét sự kết nhóm đa phần giữa các cặp loài khác. 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 rất nhiều (hàng ngàn) ô mẫu. Chính khối lượng điều tra lớn như thế đã gây ra nhiều khó khăn cho việc phân tích cấu trúc quần xã.

3.3. MỐI QUAN HỆ GIỮA HAI BIẾN ĐỊNH TÍNH CÓ THỨ BẬC

Thủ tục lập bảng chéo 2*2 (hay R*C) 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*C ô và thực hiện các kiểm định về mối quan hệ giữa biến dòng với biến cột.

1. Điều kiện áp dụng. Một là, cả hai biến lập bảng là những biến định tính có thứ bậc hơn kém. Hai là, các biến có thể lập thành nhóm (cấp) trong bảng chéo gồm R*C ô.

2. Nội dung tính toán. Đo đạc về tương quan và mối quan hệ giữa hai biến lập nhóm có thứ bậc bao gồm hai nội dung.

Nội dung 1. Xác định mối quan hệ giữa các biến. Để tính tương quan giữa các biến định tính có thứ bậc hơn kém, 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). 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, hệ số tương quan Spearman thường được sử dụng để kiểm định quan hệ giữa hai biến định tính có thứ bậc hơn kém.

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 của 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 hai 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 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 đạc sự tương đồng giữa các cặp biến số định tính có thứ bậc hơn kém, người ta sử dụng các thống kê sau đây:

a. Các đo đạc dựa trên các cặp biến tương đồng. Đó là các thống kê Kendall tau b, Kendall tau c, Goodman và Kruskal's Gamma. 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).

$$+ \text{Thống kê Kendall tau b: } \tau_B = \frac{P - Q}{\sqrt{(P+Q+T_x)(P+Q+T_y)}} \quad (3.7)$$

$$+ \text{Thống kê Kendall tau c: } \tau_C = \frac{2m*(P - Q)}{N^2*(m-1)} \quad (3.8)$$

Trong công thức 3.7 và 3.8, P – số cặp biến tương đồng với nhau (nghĩa là cùng tăng hoặc cùng giảm); Q – số cặp không tương đồng với nhau (nghĩa là những cặp tăng – giảm khác nhau); T_x – những cặp bị ràng buộc bởi yếu tố X; T_y – 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 Conditional Gamma. Hai thống kê này chỉ được áp dụng cho hai biến lập nhóm có thứ bậc và vai trò của biến độc lập và biến phụ thuộc là rõ ràng.

$$+ \text{Thống kê Conditional Gamma: } G = \frac{(P - Q)}{(P + Q)} \quad (3.9)$$

$$+ \text{Thống kê Somers' D: } D_y = \frac{(P - Q)}{(P + Q + T_y)} \quad (3.10)$$

Trong hai công thức 3.9 và 3.10, P – số cặp biến tương đồng với nhau (nghĩa là cùng tăng hoặc cùng giảm); Q – số cặp không tương đồng với nhau (nghĩa là những cặp tăng – giảm khác nhau); T_y – 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 Conditional 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 gỗ 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). Am 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. Khi 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 số mã hóa, cột 3 (hoặc hàng 3) ghi tình trạng ẩm lập địa theo mã hóa (bảng 3.17).

Gọi chương trình tính toán. Sau khi nhập số liệu vào bảng tính, để tạo được kết quả, từ thanh Menu chính bạn hãy thực hiện các thủ tục như sau:

+ **Bước 1.** Mở Describe $\Rightarrow$ Categorical Data $\Rightarrow$ Crosstabulation. Khi cửa sổ Crosstabulation xuất hiện, bạn hãy nhập tình trạng bệnh vào ô Row, nhập tình trạng ẩm lập địa vào ô Column.

+ **Bước 2.** Từ cửa sổ Crosstabulation, chọn OK. Khi bảng Crosstabulation Bệnh by Am xuất hiện, bạn chọn Tabular Option rồi chọn All và OK trong cửa sổ này để chương trình tính toán và hiển thị kết quả. Khi kết quả hiển thị trên màn hình, bạn cần lưu ý đến bảng chéo hai chiều (bảng 3.18), bảng đo xu hướng quan hệ giữa tình trạng bệnh cây với tình trạng ẩm lập địa (bảng 3.19) và bảng đo quan hệ giữa hai biến (bảng 3.20).

Bảng 3.17. Cách nhập số liệu

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	...

Bảng 3.18. Bảng chéo (Bệnh * Âm)

Frequency Table for BENH by Amlapdia

		Âm lập địa					Row Total
		1	2	3	4	5	
Bệnh cây	1	10 5.65	7 3.95	6 3.39	4 2.26	4 2.26	31 17.51
	2	3 1.69	16 9.04	8 4.52	3 1.69	4 2.26	34 19.21
	3	4 2.26	7 3.95	15 8.47	3 1.69	5 2.82	34 19.21
	4	7 3.95	9 5.08	7 3.95	14 7.91	4 2.26	41 23.16
	5	5 2.82	6 3.39	9 5.08	5 2.82	12 6.78	37 20.90
Column Total		29 16.38	45 25.42	45 25.42	29 16.38	29 16.38	177 100.00

Note: cell counts are shown with percentages of the entire table.

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 3.18) trình bày tần số ô bắt gặp cả tình trạng bệnh lẫn tình trạng âm lập địa. Từ bảng này, bạn nhận thấy từng cặp tình trạng bệnh rơi vào những cặp tình trạng âm 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 cho câu hỏi sau có thể được làm rõ từ những con số to - nhỏ khác nhau nằm ở ngoài đường chéo của bảng Crosstabs.

+ Bảng 3.19 trình bày các số đo về xu hướng (Directional Measures) của mối quan hệ giữa tình trạng bệnh với tình trạng âm lập địa. Để đo mối quan hệ này, bạn cần chú ý đến thống kê Somers'D. Từ bảng bảng 3.19, bạn có thể đi đến nhận xét:

- Giữa các cặp không bị ràng buộc trong biến độc lập thì tỷ lệ trội hơn thuộc về các cặp tương đồng hay các cặp không tương đồng? Để trả lời câu hỏi này, bạn hãy quan sát số đo Symmetric Somers'D và hai số đo quan hệ ở hai cột tiếp theo (With Rows Dependent và With Columns Dependent). 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 có quan hệ với nhau rất yếu. Ở ví dụ này, thống kê Symmetric Somers'D = 0.1774. Khi lấy tình trạng bệnh (biến dòng) là biến phụ thuộc, còn tình trạng âm là biến độc lập (biến cột) thì thống kê Somers'D = 0.1783. Ở trường hợp ngược lại thì thống kê Somers'D = 0.1765. Ở ví dụ này, vì ba giá trị gần bằng nhau, nên quan hệ giữa hai biến là kém chặt chẽ hay hai biến biến đổi không tương đồng với nhau.

Bảng 3.19. Đo xu hướng quan hệ giữa tình trạng bệnh cây với tình trạng ẩm lập địa
(Directional Measures)

Summary Statistics			
Statistic	Symmetric	With Rows Dependent	With Columns Dependent
Lambda	0.1791	0.1912	0.1667
Uncertainty Coeff.	0.0610	0.0606	0.0614
Somer's D	0.1774	0.1783	0.1765
Eta		0.2168	0.2150

Bảng 3.20. Đo quan hệ giữa hai biến (Symmetric Measures)

Statistic	Value	P-Value	Df
Contingency Coeff.	0.4264		
Cramer's V	0.2357		
Conditional Gamma	0.2204		
Pearson's R	0.2089	0.0026	175
Kendall's Tau b	0.1774	0.0034	
Kendall's Tau c	0.1761		

Lưu ý rằng, trong ví dụ dẫn ra trên đây bạn đã thấy rằng, tình trạng bệnh phụ thuộc vào tình trạng ẩm lập địa, nhưng tình trạng ẩm lập địa không phụ thuộc vào tình trạng bệnh. Vì thế, trong quan hệ này bạn cần lấy 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 cũng 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 cho biết rằng 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 21,7% (trong bảng là 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 3.20 trình bày các số đo về tính đối xứng (Symmetric Measures) trong quan hệ giữa hai biến thông qua các hệ số Conditional Gamma và Pearson's correlation coefficient (R). Hệ số Conditional Gamma áp dụng cho thang đo định danh có thứ bậc (Ordinal by Ordinal), còn hệ số Pearson's R áp dụng cho thang đo khoảng (Interval by Interval). 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.

3.4. 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 Statgraphics Plus Version 3.0&5.1 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 là nhằm tìm hiểu 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 cây 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: > 6 m, $4 \div 6$ m, $2 \div 4$ m và < 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, $4 \div 6$ m = 2, $2 \div 4$ 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 có tần số lớn hơn 5. Kết cấu của phiếu quan sát phải được xây dựng rõ ràng, để ghi chép.

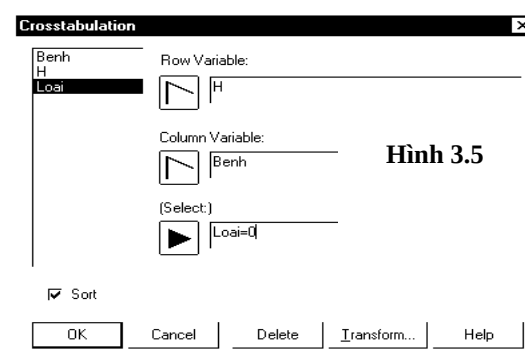
Bảng 3.21. Cách nhập dữ liệu của một bảng nhiều chiều

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
...	...	...	...

Cách nhập dữ liệu vào bảng tính.

Trước hết, bạn hãy nhập dữ liệu vào bảng tính như sau (bảng 3.21): 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). Để tạo được kết quả, từ thanh Menu chính bạn hãy thực hiện như sau:

Bước 1. Mở Describe $\Rightarrow$ Categorical



Hình 3.5

Data $\Rightarrow$ Crosstabulation. Khi cửa sổ Crosstabulation xuất hiện, bạn hãy nhập chiều cao vào ô Row, mức bị bệnh vào ô Column, còn loài cây vào ô Select. Đến đây, nếu chỉ muốn phân tích cho loài Sao đen thì bạn đặt con trỏ chuột bên cạnh chữ select và đánh dấu = 0. Tương tự như thế, nếu phân tích cho loài Dầu rái thì bạn chọn Select = 1 (hình 3.5).

Bước 2. Từ cửa sổ Crosstabulation, chọn OK. Khi bảng Crosstabulation – H by Bệnh xuất hiện, bạn chọn Tabular Option rồi chọn All và OK trong cửa sổ này để chương trình tính toán và hiển thị kết quả. Kết quả hiển thị trên màn hình bao gồm các bảng sau đây: Bảng chéo hai chiều (bảng 3.22), kiểm định Chi – Square (bảng 3.23), thống kê tóm tắt về kiểm định mối quan hệ giữa tình trạng bệnh với chiều cao cây (bảng 3.24).

Bảng 3.22. Tần số về tình trạng bệnh và chiều cao cây

3.22a. Đối với Sao đen

Frequency Table for H by Bệnh

	1	2	3	Row Total
1	1 0.68	10 6.85	34 23.29	45 30.82
2	1 0.68	28 19.18	8 5.48	37 25.34
3	10 6.85	14 9.59	10 6.85	34 23.29
4	22 15.07	7 4.79	1 0.68	30 20.55
Column Total	34 23.29	59 40.41	53 36.30	146 100.00

Note: cell counts are shown with percentages of the entire table.

3.22b. Đối với Dầu rái

Frequency Table for H by Bệnh

	1	2	3	Row Total
1	2 1.41	14 9.86	29 20.42	45 31.69
2	0 0.00	25 17.61	10 7.04	35 24.65
3	8 5.63	8 5.63	6 4.23	22 15.49
4	32 22.54	5 3.52	3 2.11	40 28.17
Column Total	42 29.58	52 36.62	48 33.80	142 100.00

Note: cell counts are shown with percentages of the entire table.

Bảng 3.23. Kiểm định Chi – Square về tính độc lập của bệnh với H cây

3.23a. Đối với Sao đen

Chi-Square Test		
Chi-Square	Df	P-Value
96.08	6	0.0000

3.23b. Đối với Dầu rái

Chi-Square Test		
Chi-Square	Df	P-Value
94.24	6	0.0000

Bảng 3.24. Các đo đặc về mối quan hệ giữa tình trạng bệnh với H cây

3.24a. Đối với Sao đen

Summary Statistics			
Statistic	Symmetric	With Rows Dependent	With Columns Dependent
Lambda	0.4149	0.3861	0.4483
Uncertainty Coeff.	0.2649	0.2359	0.3022
Somer's D	-0.5843	-0.6264	-0.5476
Eta		0.6653	0.6683
Statistic	Value	P-Value	Df
Contingency Coeff.	0.6300		
Cramer's V	0.5736		
Conditional Gamma	-0.7584		
Pearson's R	-0.6551	0.0000	144
Kendall's Tau b	-0.5857	0.0000	
Kendall's Tau c	-0.6114		

3.24b. Đối với Dầu rái

Summary Statistics			
Statistic	Symmetric	With Rows Dependent	With Columns Dependent
Lambda	0.4439	0.4227	0.4667
Uncertainty Coeff.	0.2870	0.2594	0.3211
Somer's D	-0.5864	-0.6178	-0.5579
Eta		0.7003	0.6677
Statistic	Value	P-Value	Df
Contingency Coeff.	0.6316		
Cramer's V	0.5761		
Conditional Gamma	-0.7490		
Pearson's R	-0.6606	0.0000	140
Kendall's Tau b	-0.5871	0.0000	
Kendall's Tau c	-0.6155		

Giải thích và nhận xét kết quả. Bảng 3.22 báo cáo tần số bị bệnh của Sao đen và Dầu rái ở từng lớp chiều cao khác nhau. Theo dõi kết quả của bảng 3.22(a,b), bạn cần phải trả lời một số câu hỏi sau đây: (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)? (3) 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.

Ở bảng 3.23(a,b) báo cáo kiểm định Chi – Square về tính độc lập của các biến. Giả thuyết H_0 cho rằng biến loài cây độc lập với hai biến chiều cao và bệnh. 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. Khi $P > 0.05$ (mặc định) thì giả thuyết H_0 được chấp nhận. Ngược lại, $P < 0.05$ (mặc định) thì giả thuyết H_0 bị bác bỏ.

Biến chiều cao cây là một biến định lượng, nhưng ở đây nó được đo bằng thang đo khoảng. 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 cây 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 và các thống kê Kendall's Tau b, Kendall's Tau c và Gamma. Từ bảng 3.24(a,b) cho thấy, giữa mức độ bị bệnh với chiều cao của cây có quan hệ tuyến tính âm chặt chẽ (hệ số Somer = -0.584 và -0.5864). Điều đó có nghĩa là khi chiều cao hay tuổi cây tăng lên thì mức độ bị bệnh của cả Sao đen lẫn Dầu rái đều giảm. Lưu ý rằng, ở ví dụ này bạn cũng có thể sử dụng tương quan hạng của Spearman để xác định quan hệ giữa biến mức độ bị bệnh với chiều cao cây.

TÓM TẮT PHÂN TÍCH QUAN HỆ GIỮA HAI BIẾN ĐỊNH TÍNH

(1) Đo đặc mối quan hệ giữa hai biến định tính không có thứ bậc

Các thống kê đo mối quan hệ giữa hai biến. Nếu chỉ cần biết hai biến số có quan hệ với nhau hay không và mức ý nghĩa của mối quan hệ ấy thì bạn chỉ cần sử dụng kiểm định χ^2 . Khi $P > 0.05$ (mặc định) thì hai biến số độc lập với nhau. Ngược lại, $P < 0.05$ thì hai biến số có quan hệ với nhau ở mức tin cậy 95%.

Các đo đặc về mức giảm tỷ lệ trong sai số hay cường độ của mối quan hệ. Các đo đặc này gồm có hệ số Phi, hệ số ngẫu nhiên (Contingency Coefficient) và thống kê Cramer's V. Thống kê Cramer's V lấy giá trị từ 0 – 1. Giá trị 0 cho biết không có quan hệ, còn 1 cho biết quan hệ rất chặt chẽ. Thống kê Cramer's V được dùng phổ biến nhất. Giả thuyết H_0 : Từng đo đặc = 0; H_0^- : từng đo đặc $\neq 0$. Nếu $P > 0.05$ thì các biến không có quan hệ với nhau. Ngược lại, khi $P < 0.05$ thì các biến có quan hệ với nhau ở mức tin cậy 95%.

Nội dung kiểm định mối quan hệ giữa hai biến định tính không có thứ bậc được tóm tắt như sau đây:

Nội dung	Thống kê kiểm định	Điều kiện	Giả thuyết	Chấp nhận giả thuyết	Bác bỏ giả thuyết
Kiểm định về tính độc lập của hai đại lượng A (cột) và B (hàng)	Pearson χ^2 Continuity Correction Likelihood Ratio (tỷ lệ đồng nhất) Fisher's Exact Test	N > N < N > N <	H_0 : A và B độc lập	$P > \alpha = 0.05$ hoặc 0.01	$P < \alpha = 0.05$ hoặc 0.01

Đo cường độ mối quan hệ	Phi Contingency coefficient Cramer's V Lambda	A và B là các biến định tính không có thứ bậc	* H_0 : A và B độc lập	$P > \alpha = 0.05$ hoặc 0.01	$P < \alpha = 0.05$ hoặc 0.01
-------------------------	---	---	--------------------------	-------------------------------	-------------------------------

(2) Đo mối quan hệ giữa hai biến định tính có thứ bậc

Nội dung kiểm định mối quan hệ giữa hai biến định tính có thứ bậc được ghi lại ở bảng sau đây:

Nội dung	Thống kê kiểm định	Điều kiện	Giả thuyết	Chấp nhận giả thuyết	Bác bỏ giả thuyết
Kiểm định mối quan hệ	Chi – square	A và B là các biến định tính có thứ bậc	H_0 : A và B độc lập; Hay các hệ số = 0	$P > \alpha = 0.05$ hoặc 0.01	$P < \alpha = 0.05$ hoặc 0.01
Đo cường độ mối quan hệ	Somers'd Spearman	Vai trò Y-X rõ ràng	H_0 : các hệ số = 0	$P > \alpha = 0.05$ hoặc 0.01	$P < \alpha = 0.05$ hoặc 0.01
Đo sự thành công của việc dự đoán	Kendall tau b, Kendall tau c, Goodman và Kruskal's Gamma Eta^2	Vai trò Y-X không rõ ràng	H_0 : các hệ số = 0	$P > \alpha = 0.05$ hoặc 0.01	$P < \alpha = 0.05$ hoặc 0.01

Các thống kê đo mối quan hệ giữa hai biến. Nếu chỉ cần biết hai biến có quan hệ với nhau hay không và mức ý nghĩa của mối quan hệ giữa hai biến số thì bạn chỉ cần sử dụng tiêu chuẩn χ^2 . Khi $P > 0.05$ thì hai biến độc lập với nhau. Ngược lại, khi $P < 0.05$ thì hai biến số có quan hệ với nhau ở mức tin cậy 95%.

Đo xu hướng của mối quan hệ giữa hai biến. Khi vai trò của biến độc lập và biến phụ thuộc là rõ ràng thì sử dụng thống kê Somers'D. Giả thuyết H_0 : Các hệ số Somers'D = 0. Khi $P < 0.05$ thì hệ số Somers'D tồn tại thực sự, nghĩa là biến A và biến B có quan hệ với nhau. Khi vai trò của biến độc lập và biến phụ thuộc là không rõ ràng thì sử dụng thống kê Kendall tau b, Kendall tau c, Goodman và Kruskal's Gamma.

Đo cường độ của mối quan hệ. Cường độ của mối quan hệ được đo bằng 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). 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.

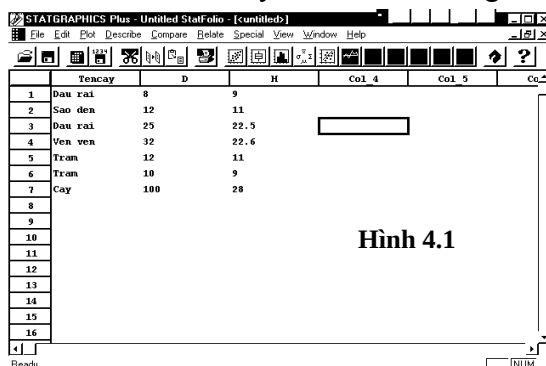
Đánh giá sự thành công của phép dự đoán. Sự thành công của việc dự đoán biến này từ biến kia được đánh giá thông qua thống kê Eta^2 . Nó cho biết có bao nhiêu phần trăm sự biến thiên của biến phụ thuộc được giải thích bởi sự biến thiên của biến độc lập.

Chương IV

PHÂN TÍCH THEO NHÓM

Khi bạn thu thập được một tập hợp các thông tin về thảm thực vật (loài cây cùng với các giá trị biểu thị cho vai trò của loài như D, H, G, M), bạn có yêu cầu xác định không chỉ các đặc trưng thống kê cho từng loài cây mà còn cho tập hợp nhiều loài cây. Để đạt được mục đích này, bạn có thể sử dụng thủ tục Subset Analysis. Cách sử dụng thủ tục Subset Analysis như sau :

Bước 1. Nhập những số liệu kiểu kí tự (Character). Trước hết, bạn hãy nhấp nút chuột trái vào đầu cột 1 hoặc hàng 1, rồi chọn Modify Column. Khi cửa sổ Modify Column xuất hiện, bạn hãy chọn Type theo kiểu Character. Sau đó trở lại bảng tính và ghi tên loài cây vào cột 1 hoặc hàng 1, nhưng không được bỏ dấu tiếng Việt (hình 4.1).



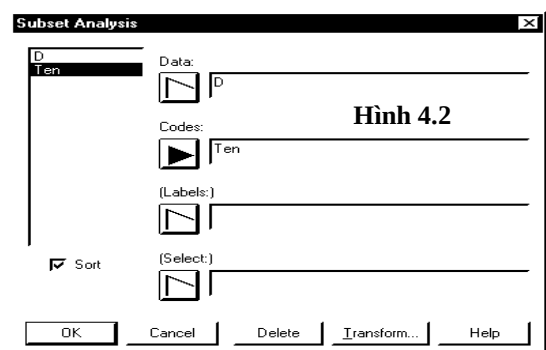
	Ten cay	D	H	Col 4	Col 5	Col 6
1	Dau rai	8	9			
2	Sao đen	12	11			
3	Dau rai	25	22.5			
4	Ven ven	32	22.6			
5	Tran	12	11			
6	Tran	10	9			
7	Cay	100	28			
8						
9						
10						
11						
12						
13						
14						
15						
16						

Hình 4.1

Bước 2. Nhập những số liệu kiểu số (Numeric). Trước hết, ở mỗi cột kế tiếp (cột 2, 3...) bạn hãy nhập các giá trị của D, H, G, M, N...(hình 4.1). Tiếp theo bạn bấm nút chuột trái vào đầu cột và chọn Modify Column. Kế đến ghi tên ngắn gọn cho các tham số (D – Đường kính, H – Chiều cao ...) vào ô Name và chọn Type theo kiểu Numeric.

Bước 3. Chọn Describe ⇒ Numeric Data ⇒ Subset Analysis (hình 4.2). Đến đây xuất hiện một cửa sổ Subset Analysis. Tại cửa sổ Subset Analysis, bạn hãy chọn từng tham số (D, H, G, M, N) và chuyển vào ô Data, còn tên loài cây chuyển vào ô Codes. Cuối cùng chọn OK.

Bước 4. Mở Tabular Options. Khi cửa sổ Tabular Options hiện ra, bạn chọn Summary Statistics và Mean Table (hoặc All); cuối cùng chọn OK. Đến đây trên màn hình sẽ hiển thị kết quả tính toán; trong đó bao gồm các thống kê mô tả, bảng số trung bình, khoảng trên dưới của các tham số, sai số chuẩn...Lưu ý rằng, tại mỗi cửa sổ hiển thị kết quả, bạn có thể điều chỉnh các thống kê bằng cách bấm nút chuột phải và chọn Pane Options. Ví dụ: Tại cửa sổ Summary Statistics, bạn bấm nút chuột phải và chọn Pane Options. Khi cửa sổ Summary Statistics Options xuất hiện, bạn có thể chọn các thống kê cần tính. Dưới đây là một số thống kê mô tả các đặc trưng mẫu (bảng 4.1 và 4.2).



Subset Analysis

Data: D

Codes: Ten

(Labels:)

(Select:)

☒ Sort

OK Cancel Delete Transform... Help

Hình 4.2

Bảng 4.1. Các thống kê tóm tắt

Summary Statistics

Code	Count	Average	Median	Standard Deviation
Dau	2	17.5	17.5	3.53553
Sao	3	8.0	9.0	2.64575
Sem	2	5.5	5.5	0.707107
Tram	1	12.0	12.0	0.0
Total	8	10.25	9.5	5.28475

Bảng 4.2. Các trị trung bình cùng với khoảng sai số chuẩn

Means Table with Standard Error Intervals

Code	Count	Mean	Standard Error	Lower Limit	Upper Limit
Dau	2	17.5	2.5	15.0	20.0
Sao	3	8.0	1.52753	6.47247	9.52753
Sem	2	5.5	0.5	5.0	6.0
Tram	1	12.0	0.0		
Total	8	10.25	1.86844	8.38156	12.1184

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

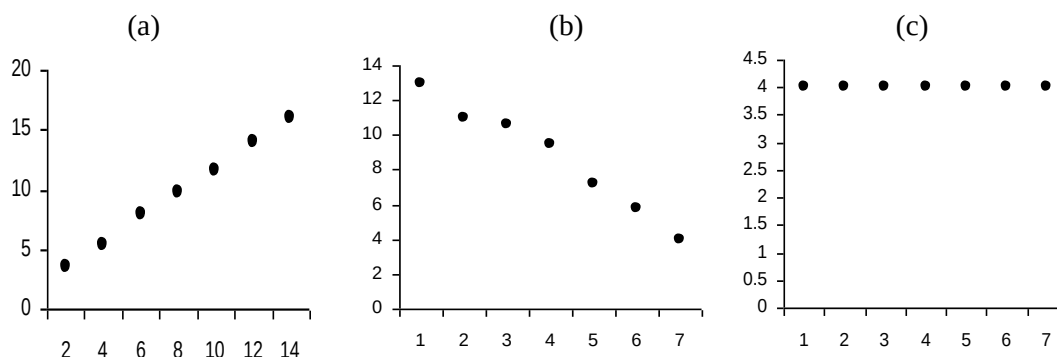
Hàm số mô tả quan hệ giữa y với x và sai số e được gọi là mô hình hồi quy. Mô hình hồi quy tuyến tính đơn có dạng:

$$Y = b_0 + b_1x + e \quad (5.1)$$

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:

$$\hat{y} = b_0 + b_1x \quad (5.3)$$

ở đây b_0 là điểm chặn, còn b_1 là độ dốc của đường hồi quy ước lượng, $\hat{y}$ 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 .

$$\text{Tiêu chuẩn bình phương nhỏ: } \min \sum (y_i - \hat{y}_i)^2 \quad (5.4)$$

Để $\sum (y_i - \hat{y}_i)^2$ 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 $y_i - \hat{y}_i$ 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)

$$SSE = \sum (y_i - \hat{y}_i)^2 \quad (5.7)$$

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

$$SSR = \sum (\hat{y}_i - y_{bq})^2 \quad (5.8)$$

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 $y_{\hat{a}}$ 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 ($SSR = \text{Regression Sum of Square}$).

Giữa SST, SSR và SSE có quan hệ với nhau như sau:

$$SST = SSR + SSE \quad (5.10)$$

Nếu tất cả các điểm quan sát y_i đều nằm trên đường $y_{\hat{a}}$ thì hàm hồi quy ước lượng phù hợp hoàn toàn. Điều đó có nghĩa là $y_i - y_{\hat{a}_i} = 0$ 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.

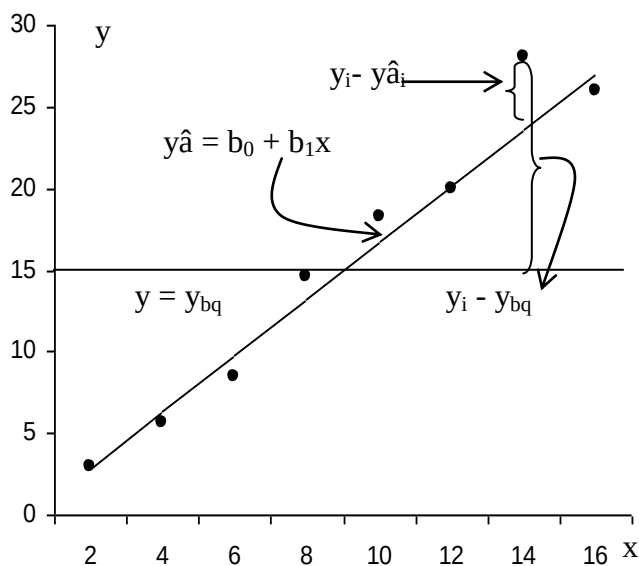
Hệ số tương quan mẫu (Sample Correlation Coefficient)

$$r_{xy} = (\text{dấu của } b_1) \sqrt{\frac{SSR}{SST}} = \sqrt{r^2} \quad (5.12)$$

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 $H_0: \rho_{xy} = 0$; $H_a: \rho_{xy} \neq 0$



Hình 5.2. Sai lệch giữa đường hồi quy ước lượng y và $y = y_{bq}$.

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)

$$H_0^-: \text{Nếu } t < -t_{\alpha/2(n-2)} \text{ hoặc } t > t_{\alpha/2(n-2)}$$

ở đây $t_{\alpha/2}$ dựa trên phân bố t với $n-2$ bậc tự do; n = dung lượng quan sát.

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:

$$\hat{y} = b_0 + b_1 x$$

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 (δ^2)

Phương sai tổng thể (δ^2) được ước lượng gần đúng thông qua sai số trung bình bình phương (MSE).

Phương sai tổng thể (δ^2) ước lượng

$$\delta^2 = s^2 = \text{MSE} = \frac{\text{SSE}}{n-2} \quad (5.14)$$

Sai số chuẩn ước lượng (Standard Error of the Estimate)

$$s = \sqrt{MSE} = \sqrt{\frac{SSE}{n-2}} = \sqrt{s^2} \quad (5.15)$$

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)

$$s_{b1} = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}} \quad (5.17)$$

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 $p - \text{value} < \alpha$.

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

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 lâm phần ở tuổi 10 năm. Từ $y = 60 + 5x$, khi thay $x = 10$, ta có $y = 60 + 5 \cdot 10 = 110$ m³/ha.

Giả sử chủ rừng lại có yêu cầu dự đoán trữ lượng rừng (y, m³/ha) đối với một lâm phần nhất định ở tuổi 10 năm. Trong trường hợp này, chủ rừng không quan tâm đến trữ lượng trung bình của tất cả các lâm phần, mà chỉ quan tâm đến việc dự đoán trữ lượng gỗ cho một lâm phần 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 $y = 60 + 5 \cdot 10 = 110$ m³/ha.

2. Ước lượng khoảng

Như trên đây đã thấy, 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 .

$y_{\hat{a}} = b_0 + b_1x_p$ = 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 lâm phần khi $x_p = 10$. Ước lượng của $E(y_p) = y_{\hat{a}} = 60 + 5 \cdot 10 = 110 \text{ m}^3/\text{ha}$.

Nói chung, chúng ta không thể mong đợi $y_{\hat{a}} = E(y_p)$. Nếu muốn biết $y_{\hat{a}}$ là một giá trị trung bình thực của $E(y_p)$, chúng ta cần ước lượng phương sai của $y_{\hat{a}}$ (kí hiệu là $s^2_{y_{\hat{a}}}$).

Phương sai ước lượng của $y_{\hat{a}}$

$$s^2_{y_{\hat{a}}} = s^2 \left\{ \frac{1}{n} + \frac{(x_p - x_{bq})^2}{\sum (x_i - x_{bq})^2} \right\} \quad (5.19)$$

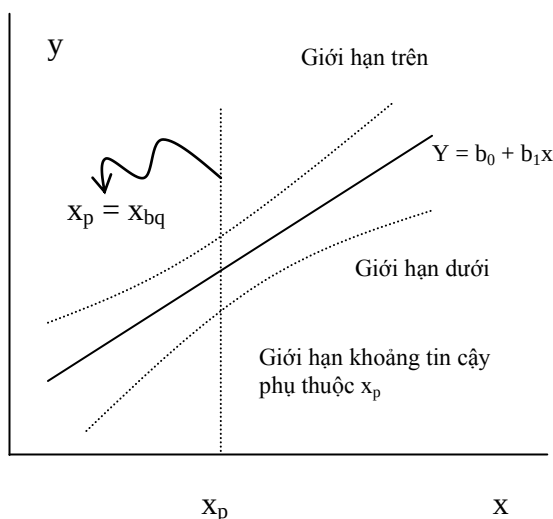
Sai tiêu chuẩn ước lượng của $y_{\hat{a}}$ (Estimated Standard Deviation of $y_{\hat{a}}$)

$$s_{y_{\hat{a}}} = s \sqrt{\frac{1}{n} + \frac{(x_p - x_{bq})^2}{\sum (x_i - x_{bq})^2}} \quad (5.20)$$

Ước lượng khoảng tin cậy của $E(y_p)$ (Confidence Interval Estiamte of $E(y_p)$)

$$y_{\hat{a}} \pm t_{\alpha/2(n-2)} s_{y_{\hat{a}}} \quad (5.21)$$

ở đây hệ số tin cậy là $1 - \alpha$, $t_{\alpha/2}$ dựa trên phân bố t với $n - 2$ độ tự do.



Hình 5.3. Khoảng tin cậy của số trung bình y

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)

$= 10 - 2 = 8$, $s_{y\hat{a}_p} = 4,95$. Từ bảng phân bố t, với $\alpha/2 = 0,025$, $df = 8$, $t_{\alpha/2} = 2,306$ và $t_{\alpha/2} s_{y\hat{a}_p} = 2,306(4,95) = 11,415$ thì ước lượng khoảng tin cậy 95% là $110 \pm 11,415$.

Lưu ý rằng, $s_{y\hat{a}_p}$ nhỏ nhất khi $x_p = x_{bq}$ hay $s_{y\hat{a}_p} = s\sqrt{\frac{1}{n}}$. 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 $y_{\hat{a}_p}$ để ước lượng $E(y_p)$ mà ước lượng của nó là $s_{y\hat{a}_p}^2$.

Phương sai của giá trị y_i riêng biệt (kí hiệu s^{2*}):

$$\begin{aligned} s^{2*} &= s^2 + s_{y\hat{a}_p}^2 = s^2 + s^2 \left\{ \frac{1}{n} + \frac{(x_p - x_{bq})^2}{\sum (x_i - x_{bq})^2} \right\} \\ &= s^2 \left\{ 1 + \frac{1}{n} + \frac{(x_p - x_{bq})^2}{\sum (x_i - x_{bq})^2} \right\} \end{aligned} \quad (5.22)$$

Sai tiêu chuẩn ước lượng của giá trị y_i :

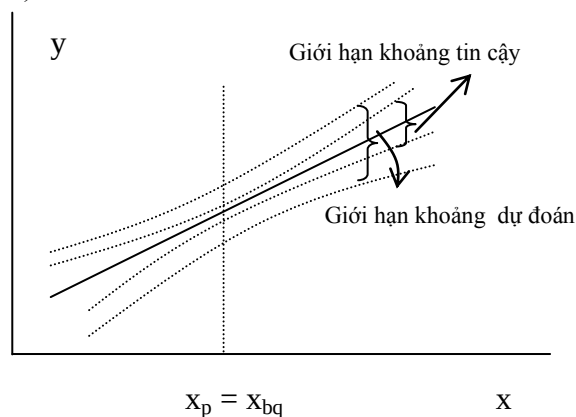
$$s^* = s\sqrt{1 + \frac{1}{n} + \frac{(x_p - x_{bq})^2}{\sum (x_i - x_{bq})^2}} \quad (5.23)$$

Ước lượng khoảng dự đoán của y_p (Prediction Interval Estiamte of y_p)

$$y_{\hat{a}_p} \pm t_{\alpha/2(n-2)} s^* \quad (5.24)$$

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).



Hình 5.4. Khoảng tin cậy và khoảng dự đoán của y tương với giá trị x nhất định

5.1.8. Phân tích các sai lệch

(1) 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 ($\hat{y}_i$). Nói khác đi, sai lệch thứ i là sai số thu được khi sử dụng hàm hồi quy ước lượng ($\hat{y}_i$) để dự đoán giá trị y_i .

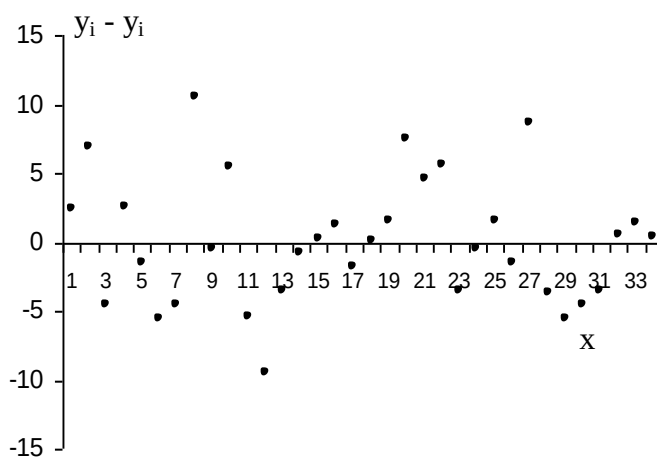
Sai lệch của quan sát i (Residual for Observation i)

$$y_i - \hat{y}_i \quad (5.25)$$

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 $\hat{y}$; (3) đồ thị các sai lệch chuẩn; (4) đồ thị xác suất chuẩn.



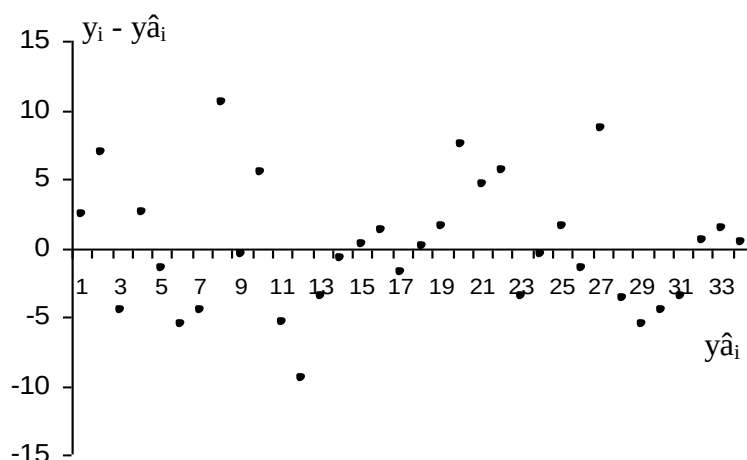
Hình 5.5. Đồ thị sai lệch đối với biến độc lập x

(2) 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 $y_i - \hat{y}_i$ (hình 5.5). Mô hình phù hợp nhất khi các sai lệch $y_i - \hat{y}_i$ nằm trên cùng một đường thẳng đi qua zero.

(3) 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 $\hat{y}_i$, còn trục tung là $y_i - \hat{y}_i$ (hình 5.6). Mô hình phù hợp nhất khi các sai lệch $y_i - \hat{y}_i$ nằm trên cùng một đường thẳng đi qua zero.



Hình 5.6. Đồ thị sai lệch đối với biến y

(4) 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 Diviation of ith 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 ith Observation)

$$h_i = \frac{1}{n} + \frac{(x_i - x_{bq})^2}{\sum (x_i - x_{bq})^2} \quad (5.28)$$

Sai lệch chuẩn của quan sát thứ i (Standardezed Residual of ith 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.

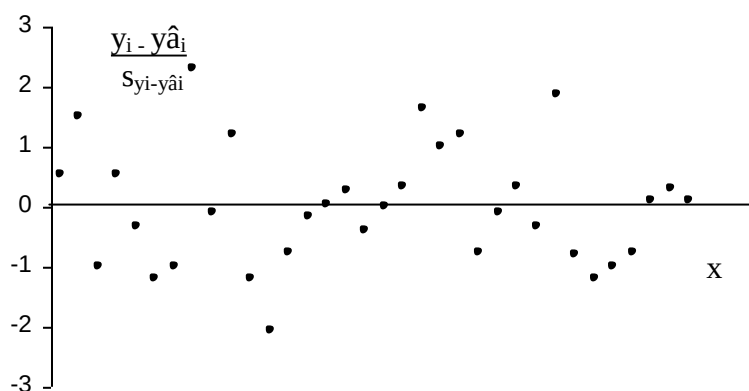
(5) Đồ thị xác suất chuẩn

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 này, 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...



Hình 5.7. Đồ thị sai lệch chuẩn

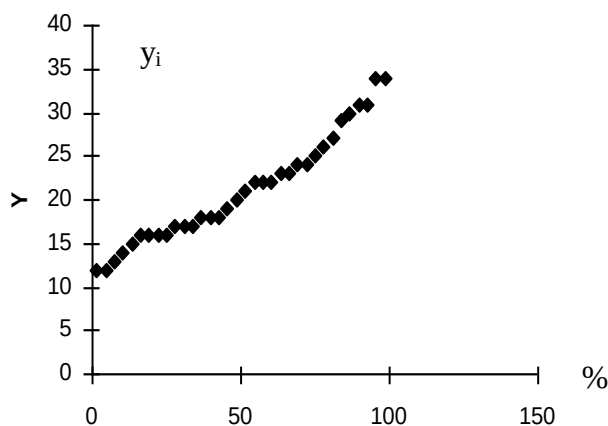
5.1.9. Phân tích các sai lệch: Các số ngoại lai và các quan sát ảnh hưởng

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.

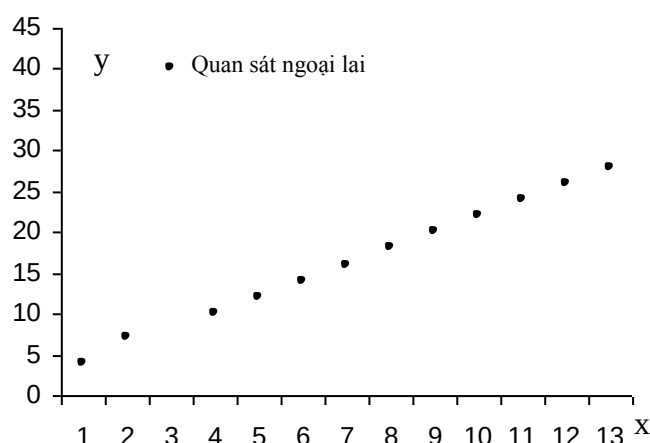
(1) Những quan sát ngoại lai

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.



Hình 5.8. Đồ thị xác suất chuẩn



(2) Các quan sát ảnh hưởng

Hình 5.9. Tập hợp số liệu với một quan sát ảnh hưởng

Đôi khi chúng ta gặp những quan sát có ảnh hưởng lớn đến hàm hồi quy ước lượng. Hình 5.9 là đồ thị chỉ ra quan sát có ảnh hưởng. Khi loại bỏ trị số này thì các thống 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.

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)

$$D_i = \frac{(y_i - \hat{y}_i)^2}{(p - 1)s^2} \left[\frac{h_i}{(1 - h_i)^2} \right] \quad (5.31)$$

D_i = số đo khoảng cách của quan sát i

$y_i - \hat{y}_i$ = 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 5.31 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 quan sát ảnh hưởng (Cook, Dfits và Mahalanibis) 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:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p \quad (5.34)$$

(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

$$SSE = \sum (y_i - \hat{y}_i)^2 \quad (5.36)$$

SSR là tổng bình phương do hồi quy

$$SSR = \sum(\hat{y}_i - y_{bq})^2 \quad (5.37)$$

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 tỷ lệ biến thiên của biến phụ thuộc y có thể được giải thích bởi hàm hồi quy đa biến ước lượng. Giá trị $R^2 \cdot 100$ 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.

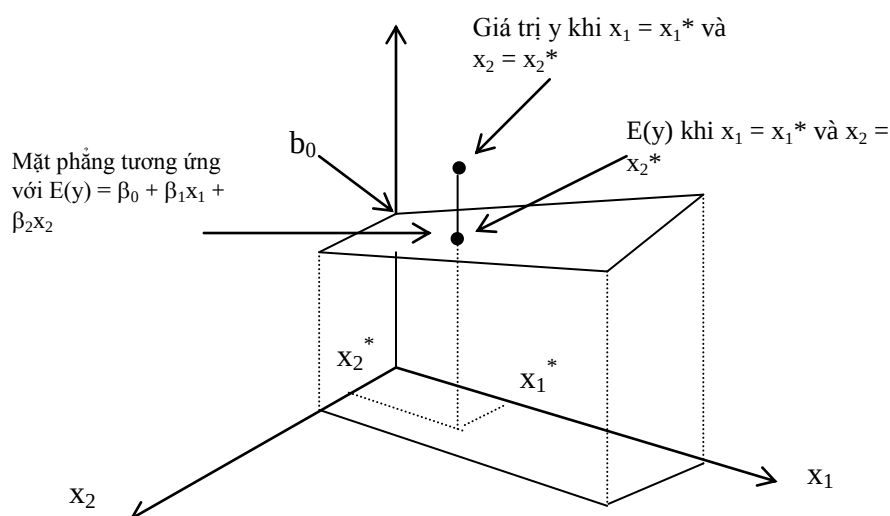
5.2.3. Các giả định của mô hình hồi quy đa biến

Các giả định trong mô hình hồi quy đa tuyến tính

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + e \quad (5.42)$$

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 tất cả các biến độc lập.
3. Các giá trị của e là độc lập.
4. 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 $\hat{y}_i$, 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 phụ thuộc. Thuật ngữ biến phản hồi ngụ ý rằng, những thay đổi của biến y phản ánh lại những ảnh hưởng của các 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).



Hình 5.10. Đồ thị hàm hồi quy để phân tích hồi quy đa biến với 2 biến phản hồi

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 chúng cùng đưa ra những kết luận giống nhau. Nhưng trong hồi quy đa tuyến tính, thống kê t và F được sử dụng với những ý nghĩa khác nhau.

- 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.
- 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

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p + e$$

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

$$s^2 = \text{MSE} = \frac{\text{SSE}}{n - p - 1} \quad (5.44)$$

Sai số chuẩn ước lượng

$$s = \sqrt{\text{MSE}} = \sqrt{\frac{\text{SSE}}{n - p - 1}} = \sqrt{s^2} \quad (5.45)$$

Thống kê kiểm định F

$$F = \frac{\text{MSR}}{\text{MSE}} \quad (5.46)$$

Quy tắc quyết định

Bác bỏ H_0 nếu $F > F_\alpha$ ($\alpha = 0,05$ hoặc $0,01$) hoặc $p\text{-value} < \alpha$; trong đó F_α 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	$\text{MSR} = \frac{\text{SSR}}{p}$	$\frac{\text{MSR}}{\text{MSE}}$
Sai số (Error)	SSE	$n - p - 1$	$\text{MSE} = \frac{\text{SSE}}{n - p - 1}$	
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

$$s_{b_i} = \frac{s}{\sqrt{\sum (x_{ij} - x_{bqj})^2}} \quad (5.48)$$

Quy tắc quyết định

Bác bỏ H_0 nếu $t > t_{\alpha/2}$ hoặc $t < -t_{\alpha/2}$ ($\alpha = 0,05$ hoặc $0,01$); trong đó $t_{\alpha/2}$ 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 những 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.

(1) Trường hợp có hai biến định tính

Giả sử cần xác định quan hệ giữa chiều cao của rừng Tách (y , m) với tuổi lâm phần (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 feralít 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.

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:

$$H = 4,260 + 0,941x_1 - 1,200x_2$$

$$\text{với } R^2 = 0,937; s = \pm 1,37; t_{b1} = 17,6 \text{ (} p = 0,000\text{); } t_{b2} = -1,96 \text{ (} p = 0,06\text{)}.$$

Ở mức ý nghĩa $\alpha = 0,05$, $p = 0,000$ với $F = 156,95$. Điều đó cho thấy giữa chiều cao với tuổi lâm phần 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 $t_{b1} = 17,6$ với $p = 0,000$; $t_{b2} = -1,96$ với $p = 0,06$. Điều đó chứng tỏ x_1 tồn tại trong mô hình với mức 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 $R^2 = 94,8\%$ và $R_a^2 = 94,2\%$ 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 lâm phần tương ứng với tuổi và loại đất khác nhau.

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

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 lâm phần Têch 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 lâm phần Têch 2 - 20 tuổi trên đất Fs như sau:

$$\begin{aligned} E(H/Fs) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 \\ &= b_0 + b_1 x_1 + b_2(1) \\ &= (b_0 + b_2) + b_1 x_1 \end{aligned} \quad (5.51)$$

Khi so sánh (5.50) và (5.51), chúng ta thấy chiều cao lâm phần 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 lâm phần trên đất Bazan thì điểm chặn là b_0 . Ngược lại, nếu xét trường hợp H của lâm phần 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 lâm phần Têch trên hai loại đất. Nếu $b_2 < 0$, thì $E(H/Bazan) > E(H/Fs)$. Khi $b_2 = 0$ thì sinh trưởng chiều cao của các lâm phần trên hai loại đất là không khác nhau. Từ ví dụ trên, chúng ta có:

$$\begin{aligned} &+ \text{ khi } x_2 = 0 \\ &H = 4,260 + 0,941x_1; \\ &+ \text{ khi } x_2 = 1 \end{aligned}$$

$$H = 3,060 + 0,941x_1.$$

So sánh hai hàm này dễ dàng nhận thấy chiều cao trung bình của các lâm phần Tách mọc trên đất bazan lớn hơn chiều cao trung bình của các lâm phần 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 lâm phần Tách phụ thuộc vào tuổi và loại đất.

(2) Trường hợp có 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 lâm phần 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 lâm phần trên ba loại đất. Giả sử chiều cao lâm phần 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 lâm phần 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:

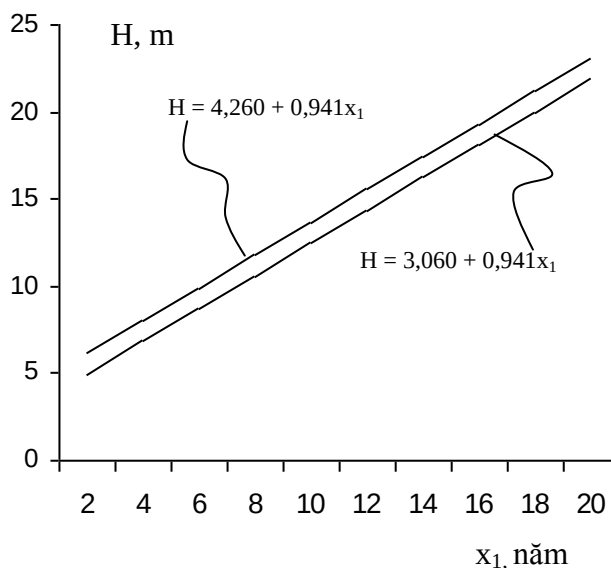
- $x_1 = 1$ nếu loại đất là B
- $x_1 = 0$ nếu loại đất khác B (đất A và C)
- $x_2 = 1$ nếu loại đất là C
- $x_2 = 0$ nếu loại đất khác C (đất A và B)

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 so sánh từng cặp chiều cao của lâm phần Tách trên loại đất B và C với chiều cao của lâm phần 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:



Hình 5.11. Chiều cao lâm phần Tách trên hai loại đất

$$E(H) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

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:

$$E(H/\text{loại đất A}) = \beta_0 + \beta_1(0) + \beta_2(0) = b_0$$

$$E(H/\text{loại đất B}) = \beta_0 + \beta_1(1) + \beta_2(0) = b_0 + b_1$$

$$E(H/\text{loại đất C}) = \beta_0 + \beta_1(0) + \beta_2(1) = b_0 + b_2$$

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:

$$E(H/\text{loại đất A}) = \beta_0 + \beta_1(1) + \beta_2(0) = b_0 + b_1$$

$$E(H/\text{loại đất B}) = \beta_0 + \beta_1(0) + \beta_2(0) = b_0$$

$$E(H/\text{loại đất C}) = \beta_0 + \beta_1(0) + \beta_2(1) = b_0 + b_2$$

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:

$$x_1 = 1 \text{ nếu loại đất là B}$$

$$x_1 = 0 \text{ nếu loại đất khác B (đất A, C, D)}$$

$$x_2 = 1 \text{ nếu loại đất là C}$$

$$x_2 = 0 \text{ nếu loại đất khác C (đất A, B, D)}$$

$$x_3 = 1 \text{ nếu loại đất là D}$$

$$x_3 = 0 \text{ nếu loại đất khác D (đất A, B, C)}$$

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:

$$E(H/\text{loại đất A}) = \beta_0 + \beta_1(0) + \beta_2(0) + \beta_3(0) = b_0$$

$$E(H/\text{loại đất B}) = \beta_0 + \beta_1(1) + \beta_2(0) + \beta_3(0) = b_0 + b_1$$

$$E(H/\text{loại đất C}) = \beta_0 + \beta_1(0) + \beta_2(1) + \beta_3(0) = b_0 + b_2$$

$$E(H/\text{loại đất D}) = \beta_0 + \beta_1(0) + \beta_2(0) + \beta_3(1) = b_0 + b_3$$

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:

$$s_{y_i - \hat{y}_i} = s \sqrt{1 - h_i} \quad (5.53)$$

Sai lệch chuẩn đối với quan sát thứ i được tính theo công thức:

$$\frac{y_i - \hat{y}_i}{s_{y_i - \hat{y}_i}} \quad (5.54)$$

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 , $s_{y_i - \hat{y}_i}$ 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 $(y_i - \hat{y}_i)/s_{y_i - \hat{y}_i}$. 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. 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. Nói chung những loại đồ thị sai lệch trên đây có thể được xác định rất nhanh chóng bằng các phần mềm thống kê như 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 ($s_{y_i - \hat{y}_i}$). Khi $s_{y_i - \hat{y}_i}$ 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 $s_{y_i - \hat{y}_i} = s\sqrt{1 - h_i}$, 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 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:

$$F = \frac{\frac{SSE(x_1, x_2, \dots, x_q) - SSE(x_1, x_2, \dots, x_q, x_{q+1}, \dots, x_p)}{(p - q)}}{\frac{SSE(x_1, x_2, \dots, x_q, x_{q+1}, \dots, x_p)}{(n - p - 1)}} \quad (5.56)$$

Trong công thức 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 $F > F_\alpha$ ($\alpha = 0,05$ 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 (đo đạc) theo thời gian (ví dụ, chuỗi các vòng năm), thì chúng ta 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.

Thống 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, chúng 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 δ^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

Nghiệm thức A			Nghiệm thức B			Nghiệm thức 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 nghiệm thức A

$$E(H/A) = 79 - 5(0) - 13(0) = 79$$

Đối với nghiệm thức B

$$E(H/B) = 79 - 5(1) - 13(0) = 74$$

Đối với nghiệm thức 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 nghiệm thức.

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 nghiệm thức 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, vì $F = 9 > F(2,15) = 3,68$ (p - value = 0,003 $< 0,05$), nên giả thuyết H_0 bị bác bỏ.

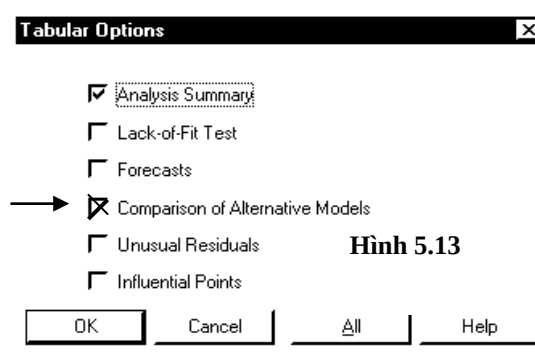
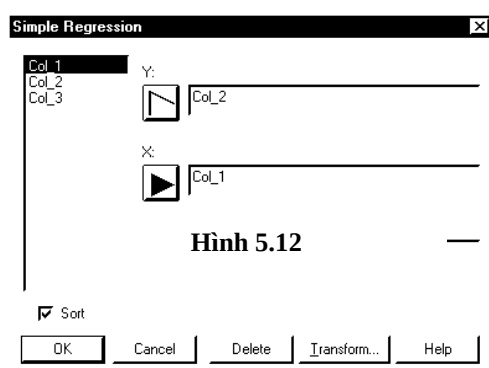
B. PHÂN TÍCH HỒI QUY BẰNG PHẦN MỀM STATGRAPHICS

5.3. PHÂN TÍCH TƯƠNG QUAN VÀ HỒI QUY GIỮA HAI BIẾN

5.3.1. Phân tích tương quan và hồi quy tuyến tính đơn

Khi thực hiện thủ tục phân tích tương quan và hồi quy tuyến tính đơn ($Y = f(X)$), bạn hãy thực hiện theo trình tự các bước sau đây (*mũi tên chỉ thứ tự chọn các lệnh*):

Bước 1. Trước hết, bấm nút chuột vào đầu cột, chọn Modify Column. Khi cửa sổ Modify Column xuất hiện, hãy ghi tên các biến số phụ thuộc (Y) và độc lập (X) vào ô Name và chọn Type kiểu Numeric. Thao tác này giúp bạn ghi tên biến vào đầu cột và các giá trị của biến số vào bảng tính một cách dễ dàng. Sau đó bạn vào số liệu như sau: Cột 1 ghi các giá trị của biến phụ thuộc (Y), cột 2 ghi các giá trị của biến độc lập (X).



Bước 2. Mở **Relate** $\Rightarrow$ **Simple Regression** (hình 5.12) $\Rightarrow$ **Dependent Variable** (đến đây vào biến Y) $\Rightarrow$ **Independent Variable** (đến đây vào biến X) $\Rightarrow$ **OK**. Sau đó chọn biểu tượng **Tabular Options** (hình 5.13). Khi cửa sổ **Tabular Options** xuất hiện, bạn hãy chọn những thống kê cần tính; cuối cùng chọn **OK**. Đến đây kết quả sẽ hiển thị trên màn hình phương trình hồi quy $Y = a + bX$.

Lưu ý rằng dạng phương trình tuyến tính $Y = a + bX$ là phương trình mặc định. Nó có thể mô tả tốt mối quan hệ của hai biến số đã chỉ định, nhưng cũng có thể là không phù hợp. Muốn biết chính xác quan hệ giữa biến phụ thuộc (Y) với biến độc lập (X), bạn có thể thực hiện các thủ tục như dưới đây:

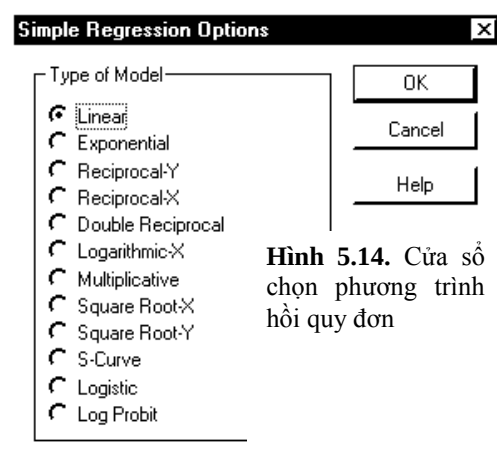
1. Biến đổi biến Y và X như lấy căn bậc (2, 3) hoặc lấy logarit...bằng cách dùng thủ tục **Transform** trong cửa sổ **Simple Regression** (hình 5.12).

2. Kết quả hiển thị trên màn hình bao gồm bảng phân tích hồi quy và bảng ANOVA kèm theo phần tư vấn thống kê (StatAdvisor). Tại đây, nếu muốn thử lại tính phù hợp của số liệu với một số mô hình hồi quy đơn thì bạn hãy mở biểu tượng của sổ **Tabular Options** và chọn mục so sánh với các mô hình khác (**Comparison of Alternative Models**, hình 5.14). Sau khi chọn **OK**, kết quả phân tích thống kê sẽ hiển thị trên màn hình những hệ số tương quan (R) của các dạng phương trình hồi quy phù hợp với số liệu đã cho. Bạn có thể căn cứ vào trị số R (hệ số tương quan đơn) hoặc bảng kiểm định các hệ số hồi quy để sơ bộ chọn dạng phương trình phù hợp.

3. Để chọn một dạng mô hình phù hợp nhất với số liệu đã cho, bạn cũng có thể thực hiện bằng cách: Đặt con chuột vào cửa sổ chính của màn hình, nhấp nút chuột phải và chọn **Analysis Options**. Tiếp theo bạn có thể chọn 1 trong 12 dạng mô hình có

trong cửa sổ Simple Regression Options để quan sát (hình 5.14). Căn cứ vào kết quả, bạn hãy chọn một mô hình phù hợp nhất.

4. Sau mỗi phương trình dự định chọn, bạn hãy mở biểu tượng Graphical Options để quan sát đồ thị quan hệ giữa biến phụ thuộc (Y) với biến độc lập (X) (hoặc đại diện của Y với đại diện của X). Trên đồ thị, bạn sẽ thấy xuất hiện 5 đường quan hệ. Đường trong cùng là đường Y bình quân ước lượng (lí thuyết). Hai đường nằm gần với Y ước lượng là khoảng giới hạn mức tin tưởng $(1 - \alpha)100\%$ (mặc định là 95%) cho số trung bình của các mẫu mới (95% Lower – Upper Confidence Limits) (xem công thức 5.59). Hai đường nằm ở phía ngoài đường khoảng giới hạn mức tin tưởng 95% cho số trung bình của các mẫu mới là khoảng giới hạn các giá trị dự đoán ở mức tin cậy $(1 - \alpha)100\%$ (mặc định là 95%) cho mẫu mới (95% Lower – Upper Prediction Limits). Khoảng giới hạn các giá trị dự đoán được tính theo công thức 5.60.



Hình 5.14. Cửa sổ chọn phương trình hồi quy đơn

Lưu ý rằng, trong kết quả của bảng ANOVA, trị số Se (Standard Error of Estimates) = $\sqrt{S_e^2}$. Đó là sai tiêu chuẩn của các sai lệch giữa $H_{\text{thực tế}}$ và $H_{\text{ước lượng}}$. Nó được dùng để tính khoảng tin cậy $(1 - \alpha)100\%$ cho các giá trị trung bình của biến y_i tương ứng với x_p nhất định (công thức 5.60) và cho các giá trị dự đoán y_i (Predicted Values) (công thức 5.61).

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:

$$y_p \pm t_{n-2, \alpha/2} Se \sqrt{\frac{1}{n} + \frac{(x_p - x_{bq})^2}{\sum (x_i - x_{bq})^2}} \quad (5.60)$$

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:

$$y_p \pm t_{n-2, \alpha/2} Se \sqrt{1 + \frac{1}{n} + \frac{(x_p - x_{bq})^2}{\sum (x_i - x_{bq})^2}} \quad (5.61)$$

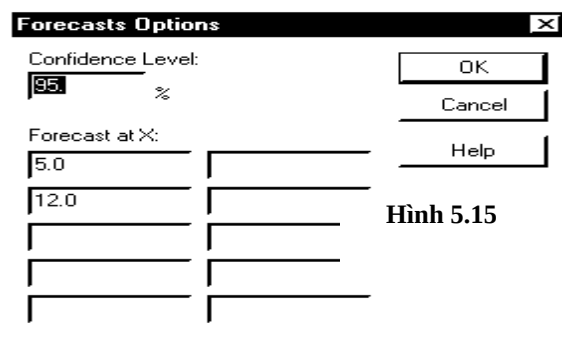
Từ công thức 5.60 và 5.61 cho thấy, khoảng tin cậy của các giá trị trung bình của biến y_p hẹp hơn so với khoảng tin cậy của các giá trị dự đoán y_p (Predicted Values).

5. Khi cần sử dụng mô hình để dự đoán biến phụ thuộc Y, bạn hãy mở Tabular Options và chọn Forecasts (dự đoán). Tại cửa sổ Forecasts, bạn nhấp nút phải chuột và chọn Pane Options. Khi cửa sổ Forecasts Options xuất hiện, bạn hãy ghi các trị số dự đoán (X_i) vào hộp forecast at X (hình 5.15). Trong hộp Forecasts Options, các giá trị mặc định là mức tin cậy 95% (95% confidence level) và cận trên – cận dưới của biến độc lập (X) trong dãy số liệu đưa vào tính toán ban đầu. Nếu cần thay đổi các trị số dự đoán này (X_i), thì bạn có thể thực hiện bằng cách xóa chúng đi và chỉ định những giá trị khác. Sau

khi chọn OK, trên màn hình sẽ hiển thị bảng kết quả của biến Y (dự đoán) theo biến dự đoán Xi đã chỉ định.

6. Trên màn hình có thể cùng lúc xuất hiện nhiều cửa sổ khác nhau. Khi muốn xem kết quả của cửa sổ nào, bạn hãy nhấp đúp nút chuột trái vào cửa sổ đó. Khi cần xem những nội dung khác thì bạn nhấp nút phải chuột và chọn Pane Options hoặc Analysis Options.

7. Bên dưới mỗi bảng kết quả, chương trình còn hiển thị những lời tư vấn về thống kê (The StatAdvisor). Ở đây bạn có thể đọc được những lời giải thích kết quả theo dạng phương trình mà bạn đã chọn ở từng bước. Lưu ý rằng, những lời tư vấn về thống kê chỉ gợi ý cho bạn về sự phù hợp của số liệu với dạng phương trình mà bạn đã chỉ định. Khi bạn chọn một dạng phương trình hồi quy khác thì chương trình cũng sẽ hiển thị một lời tư vấn khác. Vì thế, bạn phải luôn là người chủ động quyết định dạng phương trình hồi quy phù hợp nhất với số liệu đã chỉ định. Để đạt được điều đó, bạn có thể căn cứ vào: (1) phương trình có hệ số tương quan cao nhất, (2) khoảng xác định của các hệ số hồi quy, (3) các thống kê sai lệch, (4) những dạng phương trình đã được nhiều tác giả thừa nhận là hợp lý... Một phương trình là phù hợp khi tổng bình phương các sai lệch giữa giá trị thực tế và giá trị lý thuyết là nhỏ nhất.



Hình 5.15

Những dạng phương trình hồi quy đơn mặc định trong Statgraphics

Exponential	Hàm số mũ	$Y = \text{Exp}(a + bx)$
Reciprocal – y	Hàm số nghịch đảo của Y	$Y = 1/(a + bX)$
Reciprocal – x	Hàm số nghịch đảo của X	$Y = a + b/X$
Double Reciprocal	Hàm số 2 lần nghịch đảo của X	$Y = 1/(a + b/X)$
Logarithmic – x	Hàm số Logarit của X	$Y = a + b\text{Ln}X$
Multiplicative	Hàm số mũ biến đổi	$Y = aX^b$
Square root-x	Hàm số căn bậc 2 của X	$Y = a + b*\text{SQRT}(X)$
Square Root-y	Hàm số căn bậc 2 của Y	$Y = (a + b*X)^2$
S – curve	Đường cong hình chữ S	$Y = \text{Exp}(a + b/X)$
Logistic	Hàm số Logistic (có nhiều dạng)	$N = K/(1 + \text{Exp}(a + bt))$
Compound	Hàm đa hợp	$Y = \alpha a^X$

5.3.2. Kiểm định thuần nhất nhiều hệ số hồi quy tuyến tính đơn

Giả sử bạn cần kiểm định sự sai khác giữa các điểm chặn (a_i) và các hệ số hồi quy (b_i) của nhiều mô hình bậc 1: $Y_i = a_i + b_i x_i$ ($i = 1, 2, \dots, k$). Để làm rõ vấn đề đặt ra, bạn có thể sử dụng thủ tục so sánh các đường hồi quy (Comparison of Regression Lines).

1. Trường hợp có hai đường hồi quy tuyến tính đơn

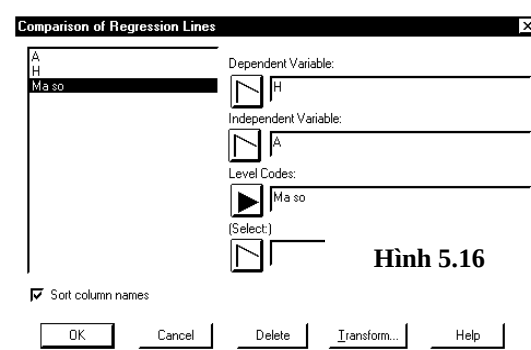
Giả sử bạn đã đo đếm được chiều cao (H, m) và tuổi (A, năm) của hai lâm phần Thông ba lá mọc trên hai lập địa khác nhau. Các số liệu được ghi lại thành bảng sau đây:

A	2	3	4	5	6	7	8	9	10	2	3	4	5	6	7	8	9	10
H	2	3.5	4.9	6.7	8.1	9.7	11.2	12.8	14.5	1.9	3.4	4.8	6.6	8.1	9.6	11.15	12.7	14.2
Mã số	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2

Từ số liệu trên, bạn đã tìm được phương trình của đường thẳng thứ nhất có dạng $H = -1.15 + 1.55 \cdot A$, còn đường thẳng thứ hai có dạng $H = -1.25 + 1.55 \cdot A$. Một vấn đề đặt ra là các điểm chặn (a_i) và các hệ số hồi quy (b_i) của hai phương trình trên có sai khác thực sự hay không? Để trả lời câu hỏi đặt ra, bạn có thể thực hiện các kiểm định theo thủ tục sau đây:

Bước 1. Nhập số liệu vào bảng tính. Trước hết, bạn hãy nhập biến phụ thuộc (H) vào cột 1, còn cột 2 và 3 nhập tương ứng biến độc lập (A) và mã số (bằng số nguyên liên tục) biểu thị cho mỗi đường hồi quy.

Bước 2. Gọi chương trình tính toán. Trước hết, mở Special $\Rightarrow$ Advanced Regression $\Rightarrow$ Comparison of Regression Lines. Đến đây bạn chọn biến H và chuyển chúng vào ô Dependent Variable. Kế đến bạn chọn biến A và chuyển chúng vào ô Independent Variable. Tiếp theo bạn chọn mã số (Maso) và chuyển chúng vào ô Level Codes. Sau cùng bạn chọn OK (hình 5.16).



Hình 5.16

Bước 3. Giải thích kết quả tính

toán. Kết quả xử lý thống kê báo cáo những nội dung sau đây :

+ **Những thống kê mô tả quan hệ giữa H với A**

Nội dung này được báo cáo qua hai bảng 5.9 và 5.10.

Bảng 5.9. Phân tích hồi quy (Multiple Regression Analysis)

Parameter	Estimate	Standard Error	T Statistic	P - value
Constant	-1.15	0.059628	- 19.286	0.000
A	1.55	0.009128	169.794	0.000
Maso=2	- 0.1	0.084327	- 1.185	0.255
A*Maso=2	0.0	0.012909	0.000	1.000

Bảng 5.10. Phân tích phương sai (Analysis of Variance)

Source	Sum of Squares	Df	Mean Square	F - ratio	P - value
Model	288.345	3	96.115	19223.0	0.000
Residual	0.07	14	0.005		
Total	288.415	17			

R-squared = 99.975 percent; R-squared (Adjusted for d.f) = 99.9705 percent

Standard error of estimate = 0.07071; Mean absolute error = 0.03333

+ Bảng 5.9 báo cáo kết quả phân tích hồi quy (Multiple regression Analysis). Nội dung của bảng này cho biết kết quả làm phù hợp mô hình hồi quy đường thẳng để mô tả mối quan hệ giữa H với A. Kí hiệu Constant = 1.15 (cột 2) cho biết điểm chặn của mô

hình ứng với biến Maso = 1 (phương trình 1). Điểm chặn của mô hình 2 bằng Constant của mô hình 1 cộng với giá trị (-0.1) ứng với hàng có Maso = 2. Kí hiệu $A = 1.55$ (cột 2) cho biết hệ số góc của mô hình ứng với biến Maso = 1 (mô hình 1). Hệ số góc của mô hình 2 bằng hệ số góc của mô hình 1 cộng với giá trị (0.0) ứng với hàng có kí hiệu $A * Maso = 2$. Như vậy, khi Maso = 1 thì phương trình của đường thẳng thứ nhất có dạng $H = -1.15 + 1.55 * A$. Ngược lại, khi Maso = 2 thì phương trình của đường thẳng thứ hai có dạng $H = -1.25 + 1.55 * A$.

Cột có nhãn T Statistic và P - value báo cáo kết quả kiểm định giả thuyết H_0 cho rằng các hệ số của phương trình = 0. Để kiểm định giả thuyết này, bạn hãy chú ý đến giá trị P - value trong bảng phân tích hồi quy. Khi P - value > 0.05 (hoặc 0.01) thì giả thuyết H_0 được chấp nhận. Ngược lại, khi P - value < 0.05 (hoặc 0.01) thì giả thuyết H_0 bị bác bỏ. Ở ví dụ này, vì P - value ứng với hai hệ số Constant và A đều bằng 0.000, nên các hệ số của phương trình tồn tại thực sự.

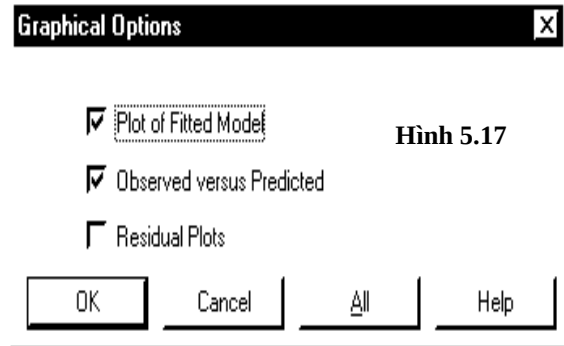
+ Bảng 5.10 trình bày kết quả phân tích phương sai. Bảng phân tích phương sai được đưa ra nhằm kiểm định sự tồn tại của mô hình đường thẳng giữa H với A. Giả thuyết H_0 : Không tồn tại mối quan hệ đường thẳng giữa H với A. Để kiểm định giả thuyết này, bạn hãy chú ý đến giá trị P - value trong bảng phân tích phương sai. Khi P - value > 0.05 (hoặc 0.01) thì giả thuyết H_0 được chấp nhận. Ngược lại, khi P - value < 0.05 (hoặc 0.01) thì giả thuyết H_0 bị bác bỏ. Ở ví dụ này, vì P - value = 0.000 < 0.01, nên hai biến H và A có quan hệ với nhau ở mức tin cậy 99%. Giá trị $R^2 = 99.975\%$ (R-squared = 99.975 percent) cho biết 99.9% biến động của H là do A gây ra.

+ Nếu cần quan sát đồ thị mô tả mối liên hệ giữa Y với A và kiểm định tính phù hợp của mô hình bằng đồ thị thì bạn có thể chọn các thủ tục tương ứng trong biểu tượng đồ thị Graphical Options (hình 5.17).

+ Kiểm định sự sai khác của hai hệ số góc và hai điểm chặn

Để kiểm định sự sai khác của hai hệ số góc (Slopes) và hai điểm chặn (Intercepts) của hai phương trình bậc 1, bạn hãy chọn thủ tục Conditional Sums of Squares trong biểu tượng Tabular Options. Kết quả kiểm định về sự sai khác của hai hệ số góc (Slopes) và hai điểm chặn (Intercepts) được ghi lại ở bảng 5.11.

+ Kiểm định sự sai khác của hai hệ số góc (Slopes). Giả thuyết H_0 : Hệ số góc của hai đường thẳng là bằng nhau. Khi P - value > 0.05 (hoặc 0.01) thì giả thuyết H_0 được chấp nhận. Ngược lại, khi P - value < 0.05 (hoặc 0.01) thì giả thuyết H_0 bị bác bỏ. Ở ví dụ này, vì P - value ứng với Slopes = 1.0 >> 0.10, nên hệ số góc của hai đường thẳng khác nhau không có ý nghĩa thống kê.



Bảng 5.11. Phân tích phương sai cho các biến trong mô hình

Source	Sum of Squares	Df	Mean Square	F - ratio	P - value
A	288.3	1	288.3	57660.0	0.000
Intercepts	0.045	1	0.045	9.0	0.0096
Slopes	0.00	1	0.00	0.0	1.000
Model	288.345	3			

+ Kiểm định sự sai khác của hai điểm chặn (Intercepts). Giả thuyết H_0 : Điểm chặn (Intercepts) của hai đường thẳng là bằng nhau. Khi $P - \text{value} > 0.05$ (hoặc 0.01) thì giả thuyết H_0 được chấp nhận. Ngược lại, khi $P - \text{value} < 0.05$ (hoặc 0.01) thì giả thuyết H_0 bị bác bỏ. Ở ví dụ này, vì $P - \text{value}$ ứng với Intercepts = 0.0096 $\ll$ 0.05, nên điểm chặn của hai đường thẳng khác nhau ở mức tin cậy 95%.

Vì điểm chặn của hai đường thẳng khác nhau, nên bạn không thể ghép chúng thành một phương trình có điểm chặn bằng nhau. Trong trường hợp của ví dụ này, hai phương trình có dạng :

$$\text{Phương trình 1 : } H = -1.15 + 1.55 \cdot A$$

$$\text{Phương trình 2 : } H = -1.25 + 1.55 \cdot A.$$

Lưu ý rằng, khi điểm chặn và hệ số góc của hai phương trình hồi quy giữa Y với X sai khác không rõ rệt thì bạn có thể san phẳng sự khác biệt giữa chúng bằng cách chọn tương ứng Assume equal intercepts và Assume equal slopes trong thủ tục Analysis Options. Thủ tục này được thực hiện bằng cách: Đặt con trỏ chuột lên một bảng bất kỳ hoặc đồ thị và nhấp nút chuột phải rồi chọn Assume equal intercepts và Assume equal slopes. Nếu chỉ có điểm chặn (hoặc độ dốc) của hai phương trình khác nhau không rõ rệt thì cách tiến hành san phẳng sự khác biệt của hai tham số này cũng được thực hiện tương tự như trên.

2. Trường hợp có nhiều đường hồi quy tuyến tính đơn

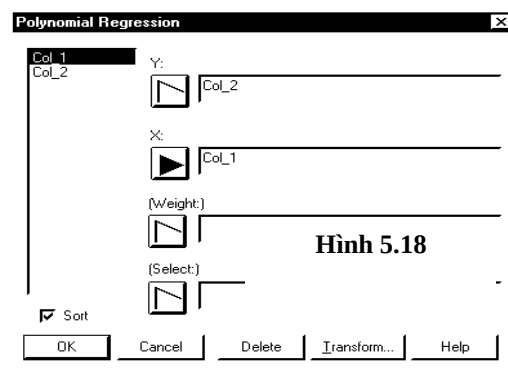
Khi kiểm định các hệ số góc và các điểm chặn của trên hai phương trình hồi quy, bạn cũng thực hiện các bước nhập số liệu, gọi chương trình tính toán và giải thích kết quả tương tự như trường hợp có hai đường hồi quy đường thẳng. Trong trường hợp có hai hoặc hơn hai hệ số góc của các đường thẳng sai khác không rõ rệt, bạn có thể gộp các hệ số này thành một hệ số chung bằng cách chọn Assume equal slopes. Nếu có hai hoặc hơn hai điểm chặn của các đường thẳng sai khác không rõ rệt, bạn có thể gộp các hệ số này thành một hệ số chung bằng cách chọn Assume equal intercepts.

5.3.3. Phân tích tương quan bậc cao

Để phân tích tương quan và hồi quy bậc cao (2, 3...), bạn hãy thực hiện theo các bước sau đây:

Bước 1. Nhập số liệu: Cột 1 nhập biến phụ thuộc Y, cột 2 nhập biến độc lập X.

Bước 2. Mở Relate $\Rightarrow$ Polynomial Regression (hình 5.18). Khi cửa sổ Polynomial Regression xuất hiện, bạn hãy nhập vào biến Y và biến X; sau đó chọn OK

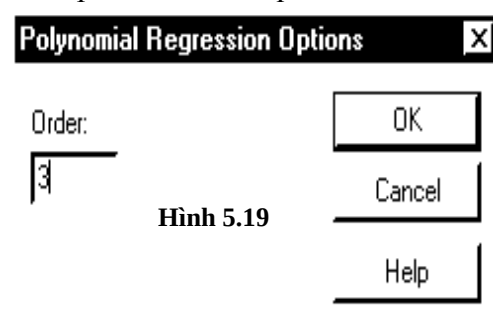


Hình 5.18

để hiển thị kết quả. Đến đây trên màn hình sẽ hiển thị bảng phân tích hồi quy và tương quan, bảng ANOVA và phần tư vấn thống kê (StatAdvisor) theo hàm số mặc định là hàm bậc 2. Từ bảng phân tích hồi quy và tương quan và bảng ANOVA, bạn hãy đưa ra những nhận xét về mối quan hệ giữa biến Y với biến X.

Bước 3. Muốn xem đồ thị mô tả quan hệ giữa biến phụ thuộc Y với biến độc lập X, bạn hãy mở biểu tượng đồ thị (Graphical Options). Khi cửa sổ Graphical Options xuất hiện, bạn hãy chọn Plot of fitted Model và những thành phần khác cần quan tâm.

Lưu ý rằng, với những số liệu đã cho, chương trình bao giờ cũng chỉ tính toán và đưa ra bảng phân tích tương quan và hồi quy cũng như bảng phân tích phương sai cho hàm bậc hai. Đây là hàm số mặc định. Vì thế, để tìm được dạng phương trình phù hợp với số liệu đã cho, bạn có thể thử sự phù hợp của số liệu với các bậc 1, 3, 4... Sau mỗi lần thử, bạn hãy quan sát bảng phân tích tương quan và hồi quy; đồng thời tính tổng chênh lệch bình phương giữa các



Hình 5.19

giá trị thực tế và các giá trị lý thuyết. Từ đó hãy chọn một dạng mô hình phù hợp nhất. Để đạt được mục đích này, bạn có thể đặt con chuột vào bất kỳ cửa sổ nào và nhấp nút phải chuột. Sau đó chọn Analysis Options. Tại cửa sổ Polynomial Regression Options (hình 5.19), bạn xóa số 2 trong ô Order và ghi lại bậc cần thử. Ví dụ: Bạn có thể thử số liệu với bậc 1, bậc 3... Sau khi ghi bậc mới cần thử, bạn nhấp OK. Trên màn hình sẽ hiển thị kết quả phân tích hồi quy và tương quan cùng với bảng phân tích ANOVA và phần tư vấn thống kê (StatAdvisor). Từ bảng phân tích hồi quy và tương quan và bảng ANOVA, bạn hãy phân tích và đưa ra những nhận xét về mối quan hệ giữa Y với X.

Định bậc của đa thức. Muốn biết bậc nào phù hợp nhất với số liệu đã cho, bạn có thể thực hiện theo các thủ tục sau đây:

1. Mở biểu tượng đồ thị và chọn Plot of fitted Model để quan sát đường biểu diễn Y lý thuyết theo sự biến đổi của X. Bên cạnh đường Y lý thuyết luôn có hai đường khác là khoảng giới hạn mức tin cậy 95% cho số trung bình của nhiều mẫu (95% Lower – Upper Confidence Limits) (đường nằm gần nhất với Y lý thuyết) và khoảng giới hạn các giá trị dự đoán cho mẫu mới ở mức tin cậy 95% (95% Lower – Upper Prediction Limits) (đường nằm phía ngoài Y lý thuyết). Một hàm số phù hợp nhất với số liệu đã cho sẽ là hàm số có khoảng giới hạn mức tin cậy 95% cho số trung bình của nhiều mẫu (95% Lower – Upper Confidence Limits) (đường nằm gần nhất với Y lý thuyết) bám sát với đường thực nghiệm.

2. Thử làm phù hợp mối liên hệ giữa hai biến số với một số bậc từ thấp lên cao. Ở mỗi bậc bạn cần ghi lại trị số phương sai thừa. Bậc phù hợp nhất là bậc có trị số phương sai hồi quy nhỏ nhất hoặc tổng các sai lệch bình phương nhỏ nhất.

3. Kiểm định sự tồn tại của các hệ số đa thức bằng thống kê T hoặc F. Trước hết, bạn hãy thử làm phù hợp mối liên hệ giữa hai biến số với một số bậc. Ở mỗi bậc bạn hãy quan sát các 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 $P < 0.01$ thì bậc đó được xem là phù hợp với dữ liệu đã cho và bạn có thể dừng việc tìm kiếm các bậc cao hơn. Bạn cũng có thể quan sát thống kê kiểm định sự tồn tại của các hệ số bi. Khi bậc thứ i có hệ số bi không tồn tại thì bậc thứ i - 1 là phù hợp. Đây là cách thức định bậc đa thức được dùng phổ biến nhất.

Bước 4. Khi cần dự đoán các giá trị Y lý thuyết, bạn hãy mở biểu tượng Tabular Options và chọn Forecasts, sau đó chọn OK. Khi cửa sổ Forecasts Options xuất hiện, bạn hãy ghi vào các ô những giá trị dự đoán (Xi) mới; sau đó nhấp OK. Trên màn hình sẽ xuất hiện các giá trị Y lý thuyết kèm theo khoảng giới hạn các giá trị dự đoán 95% cho mẫu mới (95% Lower – Upper Prediction Limits) và khoảng giới hạn mức tin cậy 95% cho số trung bình của nhiều mẫu (95% Lower – Upper Confidence Limits) trong cửa sổ Predicted Values. Khi thực hiện xong các phép thử, bạn có thể copy số liệu từ Statgraphics Plus Version 3.0 sang Microsoft Winword để xử lý văn bản.

5.4. PHÂN TÍCH TƯƠNG QUAN TUYẾN TÍNH ĐA BIẾN

Hồi quy đa tuyến tính có dạng tổng quát:

$$Y = A + B_1X_1 + B_2X_2 + B_3X_3 + \dots + B_KX_K + e$$

Hồi quy đa tuyến tính xuất hiện trong những trường hợp sau đây:

+ Một hoặc tất cả các biến dự đoán (Xi) 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 Xi là hằng số trong khoảng biến thiên của Xi.

+ Ảnh hưởng của từng biến Xi 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 Xi là như nhau mà không phụ thuộc vào biến Xi khác.

Thủ tục phân tích tương quan và hồi quy tuyến tính đa yếu tố (bội) bao gồm những nội dung cơ bản sau đây:

Nội dung 1. Thăm dò mối quan hệ giữa Y với các Xi và xác định sơ bộ những biến Xi đưa vào mô hình. Mục đích của thủ tục này là nhằm xác định tương quan giữa Y với các Xi và tương quan giữa các Xi với nhau thông qua một ma trận tương quan. Ở bước này, bạn cần chọn những hệ số tương quan cao giữa các biến, 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 Xi, xác định các hệ số hồi quy và kiểm định cộng tuyến tính. Những nội dung này được giải quyết bằng cách phân tích các hệ số hồi quy, 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 tương quan trật tự không.** Đó là hệ số tương quan giữa Y với từng Xi nhưng chưa loại bỏ ảnh hưởng tuyến tính của các biến độc lập khác. Các hệ số tương quan này phản ánh không chính xác tầm quan trọng của từng biến số độc lập.

- **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 khi 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 khi loại bỏ ảnh hưởng tuyến tính của các biến độc lập khác đối với cả biến Y lẫn biến độc lập thứ i.

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.

Xác định số lượng biến dự đoán “tối ưu”

Hệ số xác định R^2 được tính theo công thức:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (5.62)$$

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 lí 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 ($\text{adjusted } R^2 = R^2_a$).

Hệ số R^2 điều chỉnh (R^2_a)

$$R^2_a = 1 - (1 - R^2) \frac{n - 1}{n - p - 1} \quad (5.63)$$

trong đó, n là kích thước mẫu hay số cặp biến Y và X, còn p là số biến dự đoán đưa vào mô hình.

Từ công thức 5.63 cho thấy, khi p tăng lên thì R^2_a giảm xuống. Như vậy, hệ số R^2_a không nhất thiết tăng lên khi đưa thêm một biến vào mô hình. Vì thế, R^2_a 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 đa tuyến tính 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 xây dựng mô hình từ chúng. 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 tiêu chuẩn $F_{\min}$ hay F ra (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 $F_{\min}$) và ngược lại. Giá trị mặc định của F ra = 4. Vì thế, khi sử dụng F to remove bằng 4.0 thì một biến được để lại trong mô hình nếu F to remove của nó lớn hơn 4.0. Ngược lại, khi một biến có F to remove nhỏ hơn 4.0 thì biến đó bị loại ra.

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 tương quan 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 Bi $\neq 0$. Giả thuyết Bi $\neq 0$ được kiểm định theo thống kê F. Để biết một biến có được đưa vào mô hình hay không, ta so sánh $F_{\text{tính}}$ với $F_{\text{tiêu chuẩn}}$. Chúng ta có thể chỉ định thống kê F theo tiêu chuẩn F vào (F to enter). Giá trị mặc định của F to enter

là 4. Một biến được đưa vào khi F tương ứng với nó lớn hơn F tiêu chuẩn, nghĩa là $F_{tính} > F_{to\ enter} = 4$. Khi một biến không thỏa mãn điều kiện này thì biến ấy sẽ bị loại ra ở bước sau. Lưu ý rằng F tiêu chuẩn có thể được chỉ định theo yêu cầu của nhà nghiên cứu.

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.

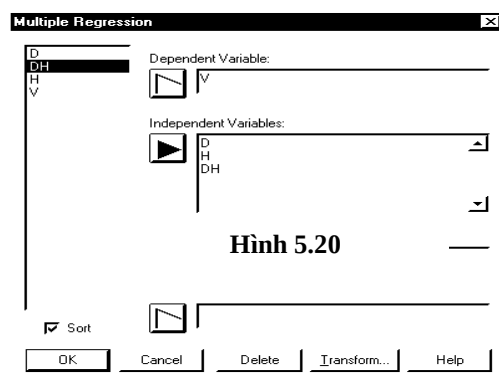
c. Phương pháp chọn tất cả các biến (All Variables). Chương trình sử dụng tất cả các biến mà bạn chỉ định để xây dựng hàm hồi quy ước lượng. Để biết những biến nào là phù hợp, bạn hãy quan sát thống kê T và giá trị P – Value. Khi P – value > 0.05 (mặc định) thì biến đó bị loại bỏ khỏi mô hình. Ngược lại, khi P – value < 0.05 (mặc định) thì biến đó được đưa vào mô hình.

Thủ tục phân tích tương quan và hồi quy tuyến tính đa nhân tố được thực hiện theo các bước sau đây:

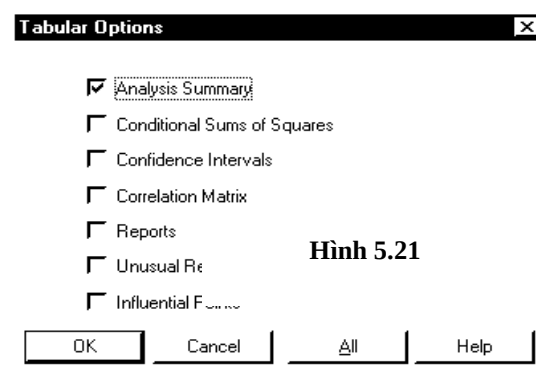
Bước 1. Nhập số liệu; trong đó cột 1 ghi trị số của biến phụ thuộc Y, còn các cột bên cạnh ghi các trị số của các biến độc lập Xi.

Bước 2. Mở Relate $\Rightarrow$ Multiple Regressions (hình 5.20) $\Rightarrow$ Dependent Variable (đến đây vào biến Y) $\Rightarrow$ Independent Variable (đến đây vào các biến Xi) $\Rightarrow$ OK. Đến đây kết quả sẽ hiển thị trên màn hình. Nội dung bao gồm phần phân tích hồi quy đa nhân tố (Multiple Regression Analysis) và phân tích biến động (ANOVA).

Bước 3. Muốn xem phần phân tích thống kê và ma trận tương quan giữa các nhân tố, bạn hãy mở biểu tượng Tabular Options và chọn những thống kê cần tính (hình 5.21). Bạn cũng có thể mở biểu tượng Graphical Options để quan sát những ảnh hưởng của các thành phần, đồ thị khoảng tin cậy của Y...



Hình 5.20



Hình 5.21

Bước 4. Nếu cần biết các thống kê tóm tắt về các nhân tố, tương quan giữa Y với từng Xi, khoảng tin cậy (Confidence Intervals), tương quan hạng (Rank Correlations), tương quan riêng phần (Partial Correlation)... thì mở biểu tượng Summary Statistics. Để biết thêm về những vấn đề khác, bạn có thể thực hiện các thủ tục giống như cách đã làm với hồi quy tuyến tính đơn.

5.5. MỘT SỐ MÔ HÌNH SẢN LƯỢNG RỪNG

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) (dưới đây ký hiệu là Y) với tuổi cây (kí hiệu A hoặc T , năm) có thể được mô tả bằng các dạng phương trình sau đây :

a. Mô hình Gomperts

$$Y = Y_{\max} \exp(-b \cdot e^{-kT}); \quad (5.64)$$

trong đó: T là tuổi cây; $Y_{\max}$ – chiều cao (hoặc đường kính) thân cây lớn nhất ở tuổi thành thực.

$$\begin{aligned} \text{Giải: Đặt } Y/Y_{\max} &= E^{(-b \cdot e^{-kT})} \Rightarrow \ln(Y/Y_{\max}) = -b \cdot e^{-kT} \\ &\Rightarrow \ln(-\ln(Y/Y_{\max})) = \ln b - kT \\ \text{Đặt } Y' &= \ln(-\ln(Y/Y_{\max})); A = \ln b; k = B. \\ &\Rightarrow Y' = A + BT. \end{aligned} \quad (5.65)$$

Giải phương trình (5.65) bằng phương pháp bình phương nhỏ nhất sẽ tìm được A và B . Khi biết $A = \ln b \Rightarrow b = e^A$. Từ đây có thể trở lại hàm phi tuyến.

b. Mô hình Schumacher

$$Y = Y_{\max} \exp(-b/T^m) \text{ hay } Y = Y_{\max} \exp(-bT^{-m}) \quad (5.66)$$

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.

$$\begin{aligned} \text{Giải:} \\ + \text{ Cách thứ nhất. Đặt } Y/Y_{\max} &= e^{(-bT^{-m})} \\ &\Rightarrow \ln(Y/Y_{\max}) = -bT^{-m} \Rightarrow \ln(-\ln(Y/Y_{\max})) = \ln b - m \ln T. \\ \text{Đặt } Y' &= \ln(-\ln(Y/Y_{\max})); \ln b = A; m = B; \ln T = T' \\ &\Rightarrow Y' = A + BT'. \end{aligned} \quad (5.67)$$

Giải phương trình (5.67) bằng phương pháp bình phương nhỏ nhất sẽ tìm được A và B . Khi biết $A = \ln b \Rightarrow b = e^A$; $\ln T = T' \Rightarrow T = e^{T'}$. Từ đây có thể trở lại hàm phi tuyến.

$$\begin{aligned} + \text{ Cách thứ hai. Từ hàm 5.66, lấy logarit tự nhiên hai vế:} \\ \ln Y &= \ln Y_{\max} + bT^{-m} \\ \text{Đặt } \ln Y &= Y'; \ln Y_{\max} = A; b = B; T^{-m} = T' \text{ với } m = 0,2 - 2,0. \\ &\Rightarrow Y' = A + BT'. \end{aligned}$$

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

$$Y = Y_{\max} \exp(-C_1 T^{-C_2}) \quad (5.68)$$

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: $Y_{\max} = e^{C_1}$, $C_1 = b(n-1)$, $C_2 = n-1$.

$$\begin{aligned}
 &\text{Giải: Đặt } Y/Y_{\max} = \exp(-C_1 T^{-C_2}) \Rightarrow -\ln(Y/Y_{\max}) = C_1 T^{-C_2} \\
 &\Rightarrow \ln(-\ln(Y/Y_{\max})) = \ln C_1 - C_2 \ln T \\
 &\Rightarrow \text{Đặt } \ln(-\ln(Y/Y_{\max})) = Y'; \ln T = X, \ln C_1 = A, C_2 = B \\
 &\Rightarrow Y = A + BX.
 \end{aligned}
 \tag{5.69}$$

Giải phương trình (5.69) 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 dạng phương trình mặc định trong Statgraphics Plus Version 3.0 và 5.1

Mối quan hệ H – A , D – A, M – A, LogN – LogD cũng có thể được mô tả bằng các hàm mặc định trong Statgraphics Plus Version 3.0 và 5.1 sau đây (hình 5.22, 5.23, 5.24 và 5.25):

Comparison of Alternative Models		
Model	Correlation	R-Squared
Multiplicative	0.9997	99.94%
Square root-X	0.9992	99.85%
Double reciprocal	0.9955	99.11%
Linear	0.9942	98.84%
Logarithmic-X	0.9808	96.20%
Square root-Y	0.9780	95.65%
Exponential	0.9494	90.14%
S-curve	-0.9487	90.00%
Reciprocal-X	-0.8642	74.69%
Reciprocal-Y	-0.8572	73.48%

Hình 5.22

Comparison of Alternative Models		
Model	Correlation	R-Squared
Linear	-0.9995	99.91%
Square root-Y	-0.9987	99.74%
Square root-X	-0.9984	99.69%
Exponential	-0.9974	99.49%
Logarithmic-X	-0.9967	99.33%
Reciprocal-Y	0.9937	98.74%
Multiplicative	-0.9925	98.50%
Reciprocal-X	0.9911	98.23%
S-curve	0.9849	96.99%
Double reciprocal	-0.9770	95.46%

Hình 5.23

Comparison of Alternative Models		
Model	Correlation	R-Squared
Multiplicative	-0.9999	99.98%
Reciprocal-X	0.9944	98.88%
Reciprocal-Y	0.9923	98.46%
Exponential	-0.9907	98.16%
S-curve	0.9863	97.28%
Logarithmic-X	-0.9670	93.51%
Square root-Y	-0.9630	92.73%
Square root-X	-0.9454	89.38%
Linear	-0.9198	84.61%
Double reciprocal	-0.9130	83.35%

Hình 5.24

Comparison of Alternative Models		
Model	Correlation	R-Squared
Reciprocal-X	-0.9843	96.88%
S-curve	-0.9802	96.08%
Double reciprocal	0.9753	95.12%
Logarithmic-X	0.9484	89.95%
Multiplicative	0.9405	88.45%
Square root-X	0.9224	85.09%
Linear	0.8925	79.65%
Square root-Y	0.8871	78.69%
Exponential	0.8814	77.69%
Reciprocal-Y	-0.8697	75.63%

Hình 5.25

Quan hệ giữa H - A và D - A còn tồn tại dưới dạng sau đây:

$$\begin{aligned}
 &+ Y = a + b \log A \\
 &+ Y = \alpha \exp(\beta/A^k)
 \end{aligned}$$

trong đó Y = H, D_{1.3}; A = tuổi cây; k = hệ số (lấy giá trị từ 0,2 – 2) tùy theo loài cây.

5.5.2. Quan hệ giữa tổng giá trị gỗ (S_A) với tuổi rừng (A)

$$+ \text{Hàm số căn bậc hai của A (Square Root – A): } S_A = a + b \cdot \text{SQRT}(A)$$

- + Hàm số mũ (Exponential Model): $SA = \text{Exp}(a + b \cdot A)$
- + Đường cong hình chữ S (S – Curve Model): $SA = \text{Exp}(a + b/A)$
- + Hàm số logarit (logarithmic – A Model): $SA = a + b \cdot \text{Ln}A$
- + Hàm số mũ biến đổi (Multiplicative): $SA = aA^b$

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): $\Delta Sa/Sa = (a + b \cdot A)^2$
- + Hàm số nghịch đảo của $\Delta Sa/Sa$ (Reciprocal – Y): $\Delta Sa/Sa = 1/(a + b \cdot A)$
- + Hàm số mũ (Exponential Model): $\Delta Sa/Sa = \text{Exp}(a + b \cdot A)$
- + Đường cong hình chữ S (S – Curve Model): $\Delta Sa/Sa = \text{Exp}(a + b/A)$
- + Hàm số logarit (Logarithmic – A Model): $\Delta Sa/Sa = a + b \cdot \text{Ln}A$
- + Hàm số mũ biến đổi (Multiplicative): $\Delta Sa/Sa = aA^b$
- + 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 tỷ 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)

$Pd = K \cdot A^{-b}$. Giải: $\text{Ln}(Pd) = \text{Ln}K - b \cdot \text{Ln}A$

$Ph = K \cdot A^{-b}$. Giải: $\text{Ln}(Ph) = \text{Ln}K - b \cdot \text{Ln}A$

Ở đây Pd và Ph tính theo công thức:

a. Đối với những loài cây sinh trưởng nhanh

$$Pd = \frac{D_a - D_{a-n}}{D_a} \cdot 100 ; Ph = \frac{H_a - H_{a-n}}{H_a} \cdot 100$$

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:

$$Pt = \frac{T_a - T_{a-n}}{T_a + T_{a-n}} \cdot \frac{200}{n} ; Pv = Pg + Ph + Pf ; Zt = T \cdot Pt\% ; Ta \pm n = Ta(1 \pm nPt\%).$$

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:

- + $V = a + bH + cD^2 \cdot H/10^4$
- + $V = a + bD^2H$
- + $V = KD^3H^b$
- + $V = K(D^2H)^b$
- + $\text{Ln}V = a + b \cdot \text{Ln}D + c \cdot \text{Ln}D^2 + d \cdot \text{Ln}H + e \cdot \text{Ln}H^2$

5.5.6. Quan hệ giữa chiều cao thân cây (H) với đường kính thân cây (D)

- + $H - 1.3 = a \cdot \text{Exp}(-b/D)$

$$+ H = a^{(b \ln D - c \ln D^2)}$$

5.5.7. Dự đoán một số nhân tố điều tra lâm phần

$$\begin{aligned} + Dg &= a + b \cdot H_0^{0.2} \\ + M &= a + bGH_0 \\ + M/G &= a + bH_0 \\ + M/G &= a + bH_0 + cH_0^2 \\ + \ln G &= a + b \ln H_0 + c \ln N \\ + \ln St &= a + b \ln H_0 + c \ln N \\ + \ln M &= a + b \ln H_0 + c \ln N \\ + \ln M &= a + b \ln H + c \ln G \text{ (Trịnh Đức Huy, 1988)} \end{aligned}$$

5.5.8. Quan hệ giữa trữ lượng gỗ sản phẩm (Msf) với tuổi rừng (A, năm)

$$\begin{aligned} + Msf &= a + b \ln A \\ + Msf &= e^{(a + b/A)} \\ + \ln Msf &= a + b \ln D, \text{ với Msf là tỷ lệ gỗ sản phẩm (\%).} \end{aligned}$$

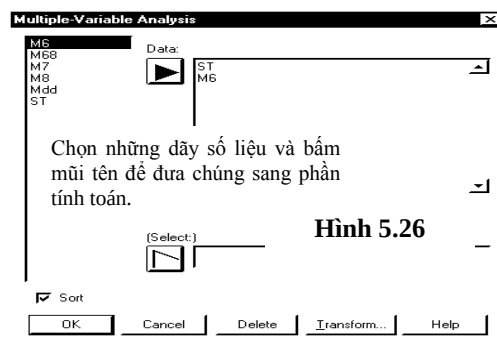
5.6. TƯƠNG QUAN HẠNG

Thủ tục tương quan hạng (Rank Correlation) cho phép xác định mối liên hệ giữa hai biến định tính được xếp hạng. Để đạt được mục đích này, bạn cần phải sử dụng tương quan hạng của Spearman (Spearman Rank Correlation). Trình tự thực hiện như sau:

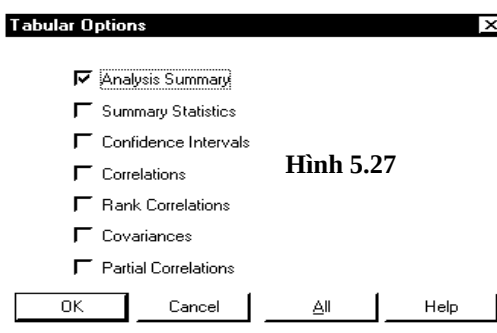
Bước 1. Nhập số liệu của hai biến trong hai cột. Nếu số liệu đo đếm là những dấu hiệu định tính (có hoặc không có thứ bậc) thì bạn phải tiến hành mã hóa chúng trước khi đưa vào tính tương quan hạng (xem hướng dẫn sử dụng SPSS 10.0).

Bước 2. Mở Describe $\Rightarrow$ Numeric Data $\Rightarrow$ Multiple Variable Analysis. Đến đây màn hình hiển thị bảng Multiple Variable Analysis. Tại đây, bạn hãy chọn những dãy số liệu từ bảng Multiple Variable Analysis và bấm mũi tên để đưa chúng sang ô Data (hình 5.26).

Bước 3. Chọn OK; tiếp theo chọn Tabular Options. Từ cửa sổ Tabular Options, bạn hãy chọn Rank Correlations và những thống kê cần tính (hình 5.27).



Hình 5.26



Hình 5.27

Chương VI

PHÂN TÍCH PHƯƠNG SAI

6.1. KHÁI QUÁT

Phân tích phương sai (ANOVA) là một kỹ thuật rất hữu ích cho những vấn đề thống kê mà bạn quan tâm 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 bạn vẫn có thể sử dụng kỹ thuật này 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 phân tích ảnh hưởng của độ dày tầng đất (được phân thành cấp có thứ bậc) đến năng suất rừng. Nhưng lưu ý rằng, muốn nhận được nhiều thông tin hơn trong mối quan hệ giữa năng suất rừng với độ dày tầng đất, bạn nên sử dụng phép phân tích tương quan và hồi quy.

Phân tích biến động (1 yếu tố, đa yếu tố và thiết kế hỗn hợp) giả định rằng số liệu tuân theo phân bố chuẩn với phương sai bằng nhau. Chương trình Statgraphics Plus Version 3.0 còn cung cấp hai thủ tục phân tích biến động phi tham số – đó là phân tích 1 yếu tố theo hạng của Kruskal – Wallis và phân tích 2 yếu tố theo hạng của Friedman. Thủ tục Kruskal – Wallis giúp bạn nghiên cứu những thí nghiệm một yếu tố cân bằng hoặc không cân bằng, còn thủ tục Friedman giúp bạn nghiên cứu những thí nghiệm 2 yếu tố cân bằng. Năm 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 theo kiểu phân cấp (Nested designs)	Nhiều	Số	Phân tích ảnh hưởng của 1 hoặc nhiều yếu tố định tính hoặc định lượng đến 1 yếu tố phụ thuộc khi số mẫu ở tất cả các tổ hợp của các mức yếu tố là bằng nhau.
Phân tích theo hạng 1 yếu tố của Kruskal – Wallis	2	Số	Phân tích ảnh hưởng của 1 yếu tố phân loại cho thí nghiệm 1 yếu tố cân bằng hoặc không cân bằng.
Phân tích theo hạng nhiều yếu tố của Friedman	2	Số	Phân tích ảnh hưởng của 2 yếu tố phân loại cho thí nghiệm 2 yếu tố cân bằng.

Chương 6 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.

6.2. 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:

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

Trong đó:

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

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

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

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

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

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 (I) 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: $T_p^{(a)}(k) = \sqrt{(l - 1)F_p(l - 1; k)}$. 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 $S_{Ci} \times T_{\alpha}(k)$ là sai khác nhỏ nhất có ý nghĩa so sánh (LSD). Quy tắc quyết định: (1) Nếu $C_i^{(*)} < S_{Ci} \times T_{\alpha}(k)$, thì giá trị trung bình của 2 nghiệm thức khác nhau không rõ rệt ở mức ý nghĩa sai lầm P . (2) Nếu $C_i^{(*)} > S_{Ci} \times T_{\alpha}(k)$, 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 (Mức 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 biến 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 mối 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.

6.2.1. 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 toàn**¹

Kiểu thí nghiệm ngẫu nhiên hoàn toàn (CRD) có thể được thiết kế theo sơ đồ hình 6.1. Trong kiểu bố trí thí nghiệm ngẫu nhiên hoàn toàn (CRD), số liệu quan sát có thể bằng nhau hoặc không bằng nhau và trật tự của các nghiệm thức không ảnh hưởng đến kết quả phân tích biến động. Ở thí nghiệm này có hai 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). Giả thuyết H_0 đặt ra là các nghiệm thức (yếu tố thí nghiệm) ảnh hưởng đồng đều đến kết quả thí nghiệm; nghĩa là $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_r = 0$ hay $\mu_1 = \mu_2 = \dots = \mu_r = \mu$. Đối thuyết $H_a: \mu_1 \neq \mu_2 \neq \dots \neq \mu_r$. Khi nguồn biến động của các nghiệm thức lớn hơn biến động ngẫu nhiên thì giả thuyết H_0 bị bác bỏ.

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

Hình 6.1. Sơ đồ kiểu bố trí thí nghiệm ngẫu nhiên hoàn toàn

¹ Complete Randomized Design = CRD

Mô hình phân tích thống kê có dạng: $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$; trong đó: Y_{ij} – biến phụ thuộc, μ là 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 ($\varepsilon_{ij} \rightarrow N(0, \sigma^2)$). Nếu phân tích phương sai 1 yếu tố theo thí nghiệm kiểu CRD bằng mô hình tuyến tính tổng quát thì mô hình có dạng như sau: $Y = \beta_0 + \beta_1 X_1 + \varepsilon$; trong đó X_1 – yếu tố thí nghiệm, ε – sai số ngẫu nhiên.

Ví dụ 1. Giả sử cần kiểm định ảnh hưởng của 3 cấp địa hình (chân, sườn, đỉnh) đến sinh trưởng $D_{1.3}$ (cm) của rừng Sao đen. 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 một cách chính xác, mỗi cấp địa hình có thể bố trí lặp lại r ô đo đếm ($r = 3, 4 \dots$). Sau khi thu thập xong kết quả thí nghiệm, nếu số cây đo đếm nhiều (vài trăm cây/cấp địa hình) thì bạn có thể gộp số liệu của r ô đo đếm trong một cấp địa hình và tính các đặc trưng thống kê cơ bản như trị số trung bình, phương sai... Số liệu này có thể được đưa vào phân tích phương sai. Nhưng nếu số cây đo đếm không nhiều (vài chục cây/cấp địa hình) thì bạn không cần phải tính toán trung gian. Trong trường hợp này, bạn có thể nhập trực tiếp số liệu này vào bảng Statgraphics để thực hiện phân tích ANOVA.

Bước 2. Nhập số liệu vào bảng tính. Số liệu nhập vào bảng tính Statgraphics bao gồm hai cột: Cột 1 nhập biến phụ thuộc ($D_{1.3} = Y_i$) hay kết quả thí nghiệm theo thứ tự từ thí nghiệm thứ 1 đến thí 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 ghi lại bằng các số nguyên giống nhau hoặc ký tự giống nhau. Số lượng mã số lặp lại giống nhau cho biết số lượng quan sát của mỗi thí nghiệm thứ (bảng 6.1). Sau đó bạn thực hiện mã hóa cho các tên biến.

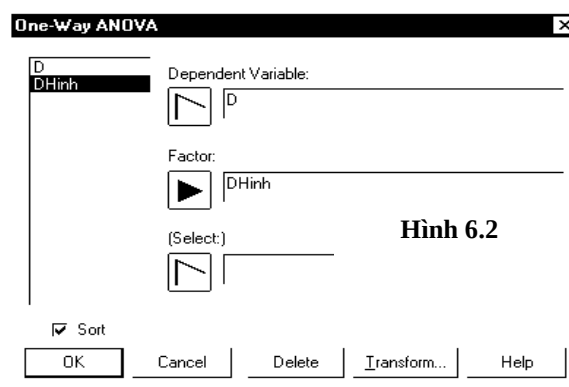
Bảng 6.1. Cách nhập dữ liệu trong phân tích phương sai một yếu tố
(Chi tiết số liệu xem ở phụ lục 1)

D	27.5	25.5	44.5	40.5		48	49		35.5	34.5	49.5	45.5	37.5	...
Địa hình	1	1	1	1	...	2	2	...	2	2	3	3	3	...

Bước 3. Thực hiện các kiểm định về phân bố chuẩn và ngang bằng phương sai mẫu. Kiểm định ngang bằng phương sai mẫu bằng cách chọn Variance Check trong biểu tượng Tabular Option. Thủ tục này còn tính toán các tham số thống kê mô tả, phân tích ANOVA, kiểm định phân hạng các thí nghiệm thứ (Multiple Range Tests)...

Giả định $H_0: X_{bq1} = X_{bq2} = \dots = X_{bqk}$; $H_0: X_{bq1} \neq X_{bq2} \neq \dots \neq X_{bqk}$. Để kiểm định giả định này, bạn tiến hành như sau:

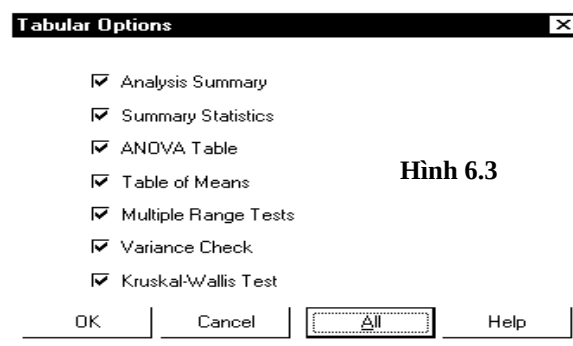
Từ menu chính, chọn Compare → Analysis of variance → One way ANOVA → Dependent list (đến đây hãy đưa biến phụ thuộc hay kết quả thí nghiệm vào bằng cách chọn nó và nhấn nút mũi tên bên cạnh) → factor (đưa biến lập nhóm vào bằng cách chọn



nó và nhấn nút mũi tên bên cạnh) (hình 6.2). Để thực hiện các phân tích thống kê, từ cửa sổ One way ANOVA, bạn hãy thực hiện như sau:

+ Chọn OK, kể đến chọn tiếp biểu tượng Tabular Options. Trong cửa sổ này chọn All (hình 6.3).

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 6.2), (2) kiểm định sự bằng nhau của phương sai (bảng 6.3), (3) bảng ANOVA (bảng 6.4), (4) so sánh giữa các cặp số trung bình nghiệm thức và phân hạng các nhóm thuần nhất (bảng 6.5).



Hình 6.3

Bảng 6.2. Thống kê mô tả (Descriptives)

DHinh	Count	Average	Variance
1	15	35.5667	51.4952
2	15	42.8667	29.481
3	15	40.5	19.1429
Total	45	39.6444	41.3139

Bảng 6.3. Kiểm định sự bằng nhau của phương sai mẫu (Variance Check for D_{1,3} cm)

Cochran's C test: 0.51434 P-Value = 0.132736
 Bartlett's test: 1.08548 P-Value = 0.188331
 Hartley's test: 2.69005

Bảng 6.4. Phân tích phương sai (ANOVA)

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Between groups	416.144	2	208.072	6.23	0.0043
Within groups	1401.67	42	33.373		
Total (Corr.)	1817.81	44			

Bảng 6.5. Kiểm định phân hạng các cặp số trung bình nghiệm thức

METHOD: 95.0 PERCENT LSD

DHinh	Count	Mean	Homogeneous Groups
1	15	35.5667	x
3	15	40.5	x
2	15	42.8667	x
Contrast		Difference	+/- Limits
1 - 2		*-7.3	4.25703

1 - 3	*-4.93333	4.25703
2 - 3	2.36667	4.25703

* denotes a statistically significant difference.

Giải thích và nhận xét kết quả. Bảng 6.3 ghi lại kết quả kiểm định sự bằng nhau của phương sai mẫu. Vì $P > 0.05$ nên phương sai các mẫu là bằng nhau. Bảng 6.4 báo cáo kết quả phân tích phương sai (ANOVA). Dòng có nhãn “Between groups” có $P = 0.0043 < 0.01$ nên sinh trưởng của rừng Sao đen trên những địa hình khác nhau có sự khác nhau rất rõ rệt. Bảng 6.5 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 ấy khác nhau ở mức ý nghĩa 0.05. Bảng 6.5 cũng dẫn kết quả kiểm định phân hạng các nghiệm thức. Những nghiệm thức thuần nhất 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 mọc tốt ở địa hình từ chân đến sườn đồi, còn ở đỉnh đồi thì sinh trưởng kém nhất.

Phân tích phương sai một yếu tố với đa quan sát. Đa quan sát được hiểu là biến phụ thuộc ở một đơn vị thí nghiệm được thu thập từ nhiều trường hợp hoặc mỗi trường hợp được đo ở những thời điểm khác nhau. Ví dụ: (1) Một nghiệm thức được thu thập từ K trường hợp; (2) Một trường hợp được đo đạc ở những thời điểm khác nhau trong suốt thời gian thí nghiệm.

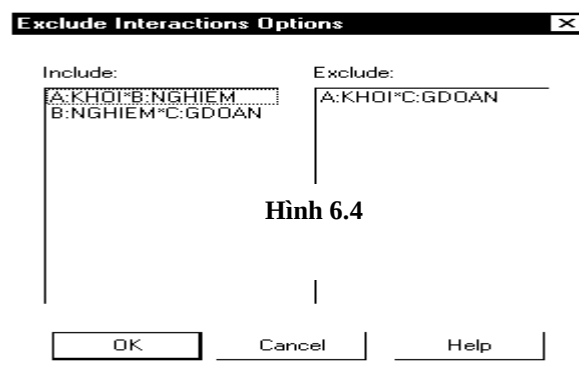
Ví dụ 2. Nghiên cứu ảnh hưởng của 5 chế độ che bóng khác nhau đến sinh trưởng về chiều cao của Dầu song nằng sau 6 tháng. Thí nghiệm 1 yếu tố với 3 lần lặp lại. Mỗi nghiệm thức đo 30 cây. Khi phân tích phương sai, bạn lưu ý rằng mỗi nghiệm thức có 90 quan sát (cây); tổng cộng 5 nghiệm thức có 450 quan sát. Hãy nhập tất cả 450 số liệu của 5 nghiệm thức vào cột 1. Cột 2 nhập thứ tự nghiệm thức từ 1 đến 5. Cột 3 nhập thứ tự khối (1 đến 3). Sau đó bạn thực hiện mở chương trình tính toán tương tự như ở ví dụ 1 trên đây.

Ví dụ 3. Nghiên cứu ảnh hưởng của 8 chế độ bón phân khác nhau đến hàm lượng đạm trong đất. Thí nghiệm 1 yếu tố với 4 lần lặp lại. Số liệu thu thập trên mỗi nghiệm thức theo 3 thời kỳ sinh trưởng của rừng: 3 tháng, 6 tháng và 12 tháng tuổi. Khi giải bài toán này, bạn cần thực hiện theo trình tự 2 bước như sau:

Bước 1. Phân tích ảnh hưởng của phân bón đến hàm lượng đạm trong đất theo từng giai đoạn tuổi.

Bước 2. Phân tích ảnh hưởng của phân bón đến hàm lượng đạm trong đất qua 3 giai đoạn tuổi. Ở thí nghiệm này, bạn cần phân tích theo mô hình thí nghiệm có lô chính - lô phụ. Bạn có thể xem nghiệm thức được bố trí ở lô chính, còn giai đoạn đo được bố trí ở lô phụ. Sai lệch của nghiệm thức ở lô chính là sự tương tác giữa “khối với nghiệm thức”. Do đó, trình tự phân tích phương sai ở bước 2 như sau:

+ Ở mỗi giai đoạn tuổi, bạn hãy tính chiều cao trung bình cho từng lô thí nghiệm. Sau đó tập hợp lại theo khối.



Hình 6.4

+ Nhập số liệu về chiều cao trung bình ở cột 1 theo thứ tự từ giai đoạn 3 tháng tuổi đến 12 tháng tuổi. Ở cột 2 bạn nhập thứ tự các giai đoạn tuổi. Ở cột 3 bạn nhập thứ tự các nghiệm thức. Ở cột 4 bạn nhập thứ tự các khối. Sau đó bạn thực hiện mã hóa cho các tên biến. Đến đây bạn gọi chương trình tính toán như sau:

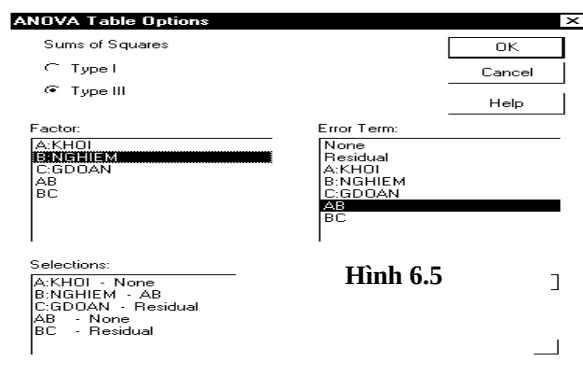
+ Mở Compare $\Rightarrow$ Analysis of Variance $\Rightarrow$ Multifactor ANOVA $\Rightarrow$ Dependent Variable (vào biến phụ thuộc – cột 1) $\Rightarrow$ Factor (chọn và chuyển vào cả 3 biến phân loại (cột 2, 3 và 4) $\Rightarrow$ OK. Đến đây trên màn hình sẽ hiển thị bảng thống kê tóm tắt (Analysis Summary).

+ Tại cửa sổ Analysis Summary, bạn hãy mở biểu tượng Tabular Options

để xem Summary Statistics, bảng ANOVA, bảng các trị số trung bình nghiệm thức (Table of means), bảng kiểm định phân hạng các nghiệm thức (Multiple Range Tests)...

+ Xác định sự tương tác giữa các yếu tố thí nghiệm. Ở thí nghiệm này có 2 nhóm tương tác sau đây: (1) tương tác giữa nghiệm thức với giai đoạn thu mẫu, (2) tương tác giữa khối với nghiệm thức. Nhóm tương tác giữa khối với nghiệm thức biểu thị cho những sai lệch ở lô chính. Để xem các yếu tố thí nghiệm có cùng phối hợp ảnh hưởng đến kết quả thí nghiệm hay không, từ cửa sổ phân tích phương sai (Analysis of Variance) bạn mở Analysis Options. Khi cửa sổ Multifactor ANOVA Options xuất hiện, bạn xoá số 1 và điền vào số 2 trong ô Maximum Order Interaction. Sau đó bạn chọn nút Exclude. Khi cửa sổ Exclude Interaction Options xuất hiện, bạn thấy trong ô Include đã có ba yếu tố xuất hiện ở dạng A: Khối, B: Nghiệm, C: Giai đoạn (hình 6.4). Điều này có nghĩa là chương trình hỏi bạn cần chọn các tương tác giữa hai yếu tố nào? Mặc định thì chương trình tính tương tác giữa từng cặp yếu tố. Nếu bạn đồng ý với đề nghị ấy thì hãy nhấp nút OK trong cửa sổ Maximum Order Interaction, rồi lại chọn OK trong cửa sổ Multifactor ANOVA Option. Ở ví dụ này, bạn chỉ cần chọn hai tương tác sau đây: Khối với nghiệm thức (hay A với B); nghiệm thức với giai đoạn (hay B với C). Để bỏ tương tác giữa khối với giai đoạn, bạn chọn nó và nhấp đúp nút trở chuột. Sau thao tác này, tương tác giữa khối với giai đoạn sẽ chuyển sang ô loại bỏ (Exclude). Cuối cùng bạn chọn OK để hiển thị kết quả. Khi kết quả hiển thị trên màn hình, bạn sẽ thấy những ảnh hưởng chính và ảnh hưởng tương tác giữa hai yếu tố thí nghiệm đến yếu tố phụ thuộc. Để tính kiểm định F cho nghiệm thức, bạn đặt con trỏ chuột vào bảng ANOVA, sau đó bấm nút phải chuột và chọn Pane options. Trong ô Factors, bạn chọn B: Nghiệm thức, rồi chọn AB trong ô Error term; cuối cùng chọn OK (hình 6.5).

Kiểm định phân hạng các nhóm thuần nhất (trung bình nhóm) theo yếu tố thí nghiệm. Trước hết, từ Multiple Range Tests, bấm nút chuột phải và chọn Pane Options. Tiếp theo chọn LSD (mặc định) hoặc Tukey's Test... Muốn phân loại kết quả theo yếu tố nào thì chọn yếu tố đó trong ô Factor. Cuối cùng bạn chọn OK để chương trình tính toán và hiển thị kết quả. Ở đây chỉ dẫn kết quả phân tích phương sai đa quan sát để bạn tham khảo (bảng 6.6).



Hình 6.5

Bảng 6.6. Phân tích phương sai (ANOVA)

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
--------	----------------	----	-------------	---------	---------

MAIN EFFECTS					
A:KHOI	0.845742	3	0.281914		
B:GDOAN	52.0429	2	26.0214	715.87(0)	0.0000
C:NGHIEM	1.26583	7	0.180833	4.94(1)	0.0020
INTERACTIONS					
AC	0.769492	21	0.0366425		
BC	3.56644	14	0.254746	7.01(0)	0.0000
RESIDUAL	1.74477	48	0.0363493		

TOTAL	60.2351	95			

F-ratios are based on the following mean squares: (0) Residual; (1) AC

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

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 như hình 6.6.

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

Hình 6.6. Sơ đồ bố trí thí nghiệm 1 yếu tố theo khối ngẫu nhiên

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. Giả thuyết H_0 đặt ra là các nghiệm thức và các khối ảnh hưởng đồng đều đến kết quả thí nghiệm; nghĩa là $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_r = 0$ hay $\mu_1 = \mu_2 = \dots = \mu_r = \mu$, đối thuật $H_a: \mu_1 \neq \mu_2 \neq \dots \neq \mu_r$. 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: $Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$; 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, $i = 1, 2, \dots, r$; β_j – tham số đặc trưng cho ảnh hưởng của khối, $j = 1, 2, \dots, k$; ε_{ij} – sai số ngẫu nhiên ($\varepsilon_{ij} \rightarrow N(0, \sigma^2)$).

Khi sử dụng thủ tục GLM trong Statgraphics 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: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$; trong đó 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: $Y = \beta_0 + \beta_1 X_1 + \varepsilon$; trong đó β_0 – điểm chặn (Intercept), X_1 – yếu tố thí nghiệm, ε – sai số ngẫu nhiên.

Ví dụ. 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 nòng 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). Để giải quyết vấn đề đặt ra, bạn thực hiện các bước như sau:

Bảng 6.7. Cách nhập số liệu trong phân tích phương sai 1 yếu tố được bố trí theo khối (Tóm tắt)
(Chi tiết số liệu xem ở phụ lục 2)

Thứ tự	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 1. Cách vào số liệu để phân tích phương sai 1 yếu tố được bố trí theo khối. Số liệu nhập vào bảng tính Statgraphics theo 3 cột: cột 1 nhập thứ tự đại lượng quan sát (có hay không cũng được), cột 2 nhập biến phụ thuộc H; 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 6.7). Sau đó bạn thực hiện mã hóa cho các tên biến.

Bước 2. Kiểm định giả thuyết về ngang bằng phương sai mẫu, phân tích ANOVA, so sánh sự khác nhau giữa các cặp số trung bình mẫu và kiểm định phân hạng nhóm thuần nhất... được thực hiện tương tự như phân tích phương sai 1 yếu tố được bố trí ngẫu nhiên. Ở đây chỉ dẫn kết quả của bảng ANOVA để các bạn tham khảo bảng 6.8.

Giải thích và nhận xét kết quả. Kiểm định ảnh hưởng chính (yếu tố nghiệm thức và khối). Giả thuyết H_0 : các yếu tố nghiên cứu ảnh hưởng đồng đều đến kết quả thí nghiệm. 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 H). 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, vì $P = 0.042 < 0.05$ nên giả thuyết H_0 bị bác bỏ. Hàng có nhãn Residual cho biết những sai lệch không thể giải thích được bởi mô hình. Ghi chú All F-ratios cho biết tất cả thống kê F được tính từ các sai số bình phương trung bình của các sai lệch.

Bảng 6.8. Kiểm định ảnh hưởng của các yếu tố thí nghiệm

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A:Khoi	232.774	3	77.5913	2.57	0.1027
B:Nghiem	415.552	4	103.888	3.45	0.0428
RESIDUAL	361.776	12	30.148		
TOTAL (CORRECTED)					
		1010.1	19		

All F-ratios are based on the residual mean square error.

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

		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 6.7. Sơ đồ bố trí thí nghiệm 1 yếu tố theo kiểu bình phương la tinh

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.

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 Dầu song nà 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 hình 6.7.

Trong kiểu bố trí thí nghiệm 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. 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.

Khi sử dụng mô hình GLM để phân tích biến động 1 yếu tố theo thí nghiệm kiểu bình phương la tinh, mô hình tuyến tính đầy đủ có dạng: $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \varepsilon$; 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: $Y = \beta_0 + \beta_1X_1 + \varepsilon$; trong đó β_0 – điểm chặn (Intercept), X_1 – yếu tố thí nghiệm, ε – sai số ngẫu nhiên.

Ví dụ. 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. 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. Kí hiệu A = 0, B = 1, C = 2, D = 3, E = 4, F = 5% Super photphat so với trọng lượng bầu. Hãy phân tích kết quả thí nghiệm. Để giải quyết vấn đề đặt ra trên đây, bạn thực hiện các bước sau đây:

Bước 1. Nhập số liệu vào bảng tính Statgraphics hoặc bảng tính Excel như sau (bảng 6.8): Cột 1 nhập kết quả thí nghiệm (H, cm), cột 2 nhập thứ tự nghiệm thức (khi

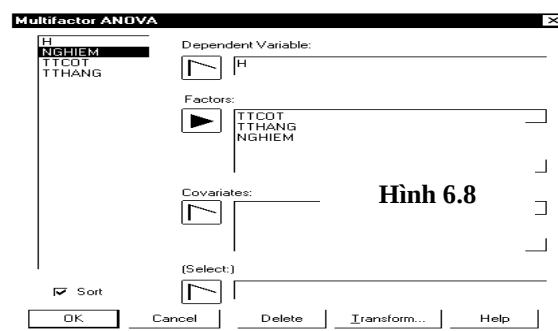
nhập phải chú ý đến sơ đồ bố trí các nghiệm thức trong mỗi cột và hàng), cột 3 nhập thứ tự hàng (khối hàng), cột 4 nhập thứ tự cột (khối cột). Sau đó bạn thực hiện mã hóa cho các tên biến và lưu số liệu này thành một file.

Bảng 6.8. Ảnh hưởng của hàm lượng Super photphat đến sinh trưởng H của Dầu song nòng
(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

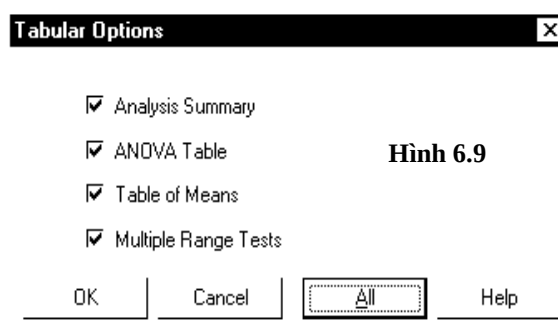
Bước 2. Gọi chương trình tính toán. Trình tự kiểm định giả thuyết về sự khác biệt giữa các trung bình mẫu (nghiệm thức), so sánh sự khác nhau giữa các cặp số trung bình mẫu và kiểm định phân hạng các trung bình mẫu được thực hiện theo các thủ tục sau đây:

+ Mở Compare → Analysis of Variance → Multifactor ANOVA → Dependent (đến đây đưa vào biến phụ thuộc H) → factor (đưa vào các biến nghiệm thức, cột và hàng) bằng cách chọn 1 hay nhiều biến và nhấn nút mũi tên bên cạnh (hình 6.8). Sau đó chọn OK.



Hình 6.8

Bước 3. Phân tích thống kê. Ở đây bạn cần yêu cầu chương trình xây dựng bảng ANOVA, lập bảng trung bình mẫu và kiểm định phân hạng các trung bình mẫu. Để đạt được các mục đích này, trước hết bạn hãy mở biểu tượng Tabular Options; kế đến tại cửa sổ Tabular Options bạn chọn All rồi chọn tiếp OK (hình 6.9). Đến đây kết quả hiển thị trên màn hình những nội dung mà bạn đã chỉ định. Ở đây chỉ dẫn bảng ANOVA (bảng 6.9) và bảng phân hạng các nghiệm thức (bảng 6.10) để các bạn tham khảo.



Hình 6.9

Giải thích và nhận xét kết quả.

Bảng 6.9 báo cáo kết quả kiểm định ảnh hưởng của yếu tố thí nghiệm đến yếu tố phụ thuộc. Vì $P - \text{value (Sig)} < 0.05$ nên giả thuyết cho rằng hàm lượng phân super photphat khác nhau ảnh hưởng đồng đều đến sinh trưởng chiều cao của Dầu song nòng đã bị bác bỏ. Theo kết quả kiểm định phân hạng, bạn có thể thấy những cặp mẫu nào khác nhau không có ý nghĩa thống kê.

Bảng 6.9. Kiểm định ảnh hưởng của yếu tố thí nghiệm đến yếu tố phụ thuộc

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A:TTCOT	577.0	5	115.4	1.57	0.2151
B:TTHANG	2028.0	5	405.6	5.50	0.0024
C:NGHIEM	24442.0	5	4888.4	66.33	0.0000
RESIDUAL	1474.0	20	73.7		
TOTAL	28521.0	35			

All F-ratios are based on the residual mean square error.

Bảng 6.10. Kiểm định phân hạng các nhóm thuần nhất

Method: 95.0 percent LSD

TTCOT Count LS Mean Homogeneous Groups

1	6	35.0	x
6	6	38.3333	xx
3	6	39.5	xx
2	6	39.5	xx
4	6	42.8333	xx
5	6	47.8333	x

6.2.2. PHÂN TÍCH PHƯƠNG SAI NHIỀU YẾU TỐ

Thủ tục phân tích phương sai hai hay nhiều yếu tố kiểm định ảnh hưởng của hai hay nhiều yếu tố định tính hoặc định lượng đến biến phụ thuộc. Dưới đây chúng ta xem xét một số ví dụ về phân tích phương sai nhiều yếu tố.

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

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; $(\alpha\beta)_{ij}$ - tham số 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 ($k = 1 \dots r$).

	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	Y ₃₃₁	...	Y _{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	Y ₃₃₂	...	Y _{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	Y ₃₃₃	...	Y _{3b3}

Hình 6.10. Sơ đồ bố trí thí nghiệm hai yếu tố theo khối ngẫu nhiên hoàn toàn

Khi sử dụng mô hình tuyến tính tổng quát để phân tích phương sai 2 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: $Y = \beta_0 + \beta_1A + \beta_2B + \beta_3AB + \beta_5\text{Khối} + \varepsilon$; 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: $Y = \beta_0 + \beta_1A + \beta_2B + \beta_3AB + \varepsilon$.

Các 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. Kiểu thí nghiệm hai yếu tố được bố trí theo khối ngẫu nhiên hoàn toàn (RCBD) có r lần lặp lại ở mỗi tổ hợp cấp của yếu tố A và yếu tố B được thiết kế theo sơ đồ như hình 6.10.

Ví dụ. 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. 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.

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) có ảnh hưởng đồng đều đến sinh trưởng H của Dầu song năng; (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 năng. Để kiểm định các giả thuyết trên, bạn hãy thực hiện các bước sau:

Bước 1. Nhập số liệu vào bảng tính Statgraphics như sau (bảng 6.11): cột 1 hoặc hàng 1 nhập biến kết quả thí nghiệm (H, cm); cột 2 nhập số lần lặp của mỗi nghiệm thức (số khối); cột 3 hoặc hàng 3 nhập số lần lặp lại của yếu tố đạm (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 thực hiện mã hóa cho các tên biến và ghi lại thành 1 file để có thể dùng về sau.

Bảng 6.11. Cách nhập dữ liệu trong phân tích phương sai 2 yếu tố bố trí theo khối ngẫu nhiên
(Chi tiết số liệu xem ở phụ lục 4)

H, cm	TT. khối	Đạm (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
...	...	...	...

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:

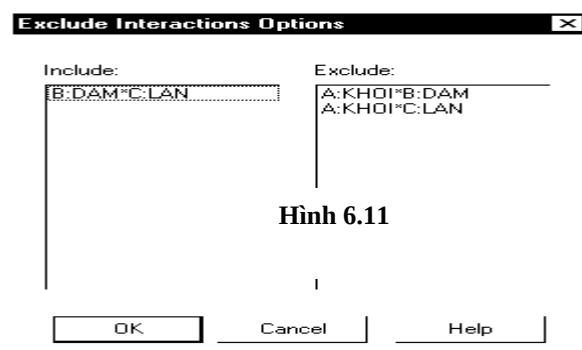
+ Mở Compare → Analysis of Variance → Multifactor ANOVA → Dependent List (đến đây đưa vào biến phụ thuộc H) → factor (đưa vào các biến phân loại cho

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 factors).

+ 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, hãy chọn biểu tượng Tabular options; kế đến tại cửa sổ Tabular options, bạn chọn All rồi chọn tiếp OK.

+ Từ màn hình hiển thị kết quả, bạn nhấp đúp nút chuột trái lên trang Analysis of Variance for H - Type III Sums of Squares. Tiếp theo đặt con trỏ chuột lên trang này và nhấp nút chuột phải để chọn Analysis

options. Khi cửa sổ Analysis options hiện ra, bạn xóa số 1 và ghi vào số 2 trong ô Maximum Order Interactions; kế đến chọn Exclude. Khi cửa sổ Exclude Interactions Options xuất hiện (hình 6.11), bạn nhấp đúp vào các dòng A: khối*B:Đậm và A: khối*C: Lân để loại bỏ kết quả tương tác giữa chúng (vì các yếu tố này không tương tác với nhau). Sau đó chọn OK để hiển thị kết quả. Kết



Hình 6.11

quả hiển thị trên màn hình bao gồm các bảng sau đây: các thống kê mô tả (Descriptive Statistics), kiểm định ảnh hưởng giữa các yếu tố (ANOVA), ước lượng trung bình nghiệm thức và trung bình khối, so sánh từng cặp đối với các nghiệm thức và khối (Multiple Comparisons), phân hạng các nghiệm thức.

Lưu ý rằng kiểm định LSD là mặc định để so sánh giữa các cặp số trung bình và kiểm định phân hạng các trung bình nghiệm thức. Nếu muốn sử dụng các kiểm định Bonferroni và Tukey's Test thì bạn đặt con trỏ chuột vào trang Multiple Range Tests; kế đến nhấn nút chuột phải và chọn Pane Options. Tại đây bạn hãy chọn các kiểm định và yếu tố cần so sánh. Cuối cùng chọn OK để chương trình tính toán và hiển thị kết quả. Ở đây chỉ trích dẫn bảng phân tích biến động (bảng 6.12) và bảng phân hạng các nghiệm thức (bảng 6.13) để các bạn tham khảo.

Giải thích và nhận xét kết quả. 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 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. Từ bảng 7.12 cho thấy, từng yếu tố và tổ hợp hai yếu tố có ảnh hưởng rõ rệt đến kết quả thí nghiệm. Từ bảng 7.13 cho thấy, so với nghiệm thức 1 – 2% đậm (so với trọng lượng bầu) sinh trưởng của Dầu song năng trên môi trường có hàm lượng đậm bằng 3% là trội hơn.

Bảng 6.12. Kiểm định ảnh hưởng của các yếu tố đậm và lân đến H của Dầu song năng
Analysis of Variance for H - Type III Sums of Squares

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A:KHỐI	257.422	4	64.3556	2.59	0.0550
B:DAM	3040.53	2	1520.27	61.29	0.0000
C:LAN	578.133	2	289.067	11.65	0.0002
INTERACTIONS					
BC	6789.33	4	1697.33	68.43	0.0000

RESIDUAL 793.778 32 24.8056

TOTAL (CORRECTED) 11459.2 44

Bảng 6.13. Phân nhóm thuần nhất về H theo yếu tố đậm

Method: 95.0 percent Bonferroni

DAM Count LS Mean Homogeneous Groups

DAM	Count	LS Mean	Homogeneous Groups
1	15	76.1333	x
3	15	86.0	x
2	15	96.2667	x

Contrast	Difference	+/- Limits
1 - 2	*-20.1333	4.59464
1 - 3	*-9.86667	4.59464
2 - 3	*10.2667	4.59464

B. 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 ($k = 1 \dots m$).

Khi sử dụng thủ tục GLM để phân tích phương sai 2 yếu tố theo kiểu thí nghiệm hoàn toàn ngẫu nhiên (RCD), bạn có thể xây dựng mô hình tuyến tính tổng quát có dạng: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$; trong đó: X_1 và X_2 - yếu tố thí nghiệm 1 và 2; ε - sai số thí nghiệm.

Các giả thuyết. 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. Kiểu thí nghiệm hai yếu tố ngẫu nhiên hoàn toàn 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 hình 6.12.

Yếu tố A	Yếu tố B				
	1	2	3	...	b
A1	X111	X121	X131		X1b1
	X112	X122	X132		X1b2
	X11m	X12m	X13m		X1bm
A2	X211	X221	X231		X2b1
	X212	X222	X232		X2b2
	X21m	X22m	X23m		X2bm
A3	X311	X321	X331		X3b1
	X312	X322	X332		X3b2
	X31m	X32m	X33m		X3bm

Hình 6.12. Sơ đồ bố trí thí nghiệm hai yếu tố ngẫu nhiên hoàn toàn

Ví dụ. Giả sử cần đánh giá sinh trưởng về chiều cao (H, cm) của cây con Dầu song nòng 6 tháng tuổi dưới ảnh hưởng của 3 mức đậm (kí hiệu yếu tố A: 1, 2 và 3% đậm); mỗi mức đậm được sủi sung thêm 3 mức lần (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.

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 Statgraphics như sau (bảng 6.14): Cột 1 hoặc hàng 1 nhập biến phụ thuộc (H, cm); cột 2 nhập số lần lặp của mỗi nghiệm thức; cột 3 hoặc hàng 3 nhập số lần lặp lại của yếu tố đạm (biến phân loại A); cột 4 hoặc hàng 4 nhập số lần lặp lại của yếu tố lân (biến phân loại B). Sau đó bạn thực hiện mã hóa cho các tên biến và ghi lại thành 1 file để có thể dùng về sau.

Bảng 6.14. Cách nhập dữ liệu trong phân tích phương sai 2 yếu tố 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
53	3	1	1
57	4	1	1
...	...	...	...
54	15	3	3

Bước 2. Kiểm định giả thuyết 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 tương tự như phân tích phương sai 2 yếu tố được bố trí theo khối. Ở đây chỉ trích dẫn bảng phân tích biến động (bảng 6.15), so sánh cặp đối với các nghiệm thức (đạm, lân) và phân hạng các nghiệm thức để các bạn tham khảo (bảng 6.16).

Giải thích và nhận xét kết quả. Bảng 6.15 cho biết kết quả phân tích biến động (ANOVA). 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 rõ rệt đến H của Dầu song nòng. Ngược lại, ở dòng Đạm*Lân có $P = 0,9998 > 0.05$, nên ảnh hưởng phối hợp của đạm với lân đến H của Dầu song nòng là không rõ rệt. Từ bảng 6.16 cho thấy, so với nghiệm thức 1 – 2% đạm (so với trọng lượng bầu) sinh trưởng H của Dầu song nòng trên môi trường có hàm lượng đạm bằng 3% là trội hơn.

Bảng 6.15. Kiểm định ảnh hưởng của các yếu tố đạm và lân đến H của Dầu song nòng

Analysis of Variance for H - Type III Sums of Squares

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A:DAM	2920.53	2	1460.27	154.18	0.0000
B:LAN	64.7111	2	32.3556	3.42	0.0359
INTERACTIONS					
AB	0.355556	4	0.0888889	0.01	0.9998
RESIDUAL	1193.33	126	9.4709		

TOTAL (CORRECTED) 4178.93 134

Bảng 6.16. Phân hạng các nghiệm thức (trên ví dụ yếu tố đạm)

Method: 95.0 percent LSD

DAM Count LS Mean Homogeneous Groups

2	45	56.9556	x
1	45	56.9556	x
3	45	66.8222	x

Contrast	Difference	+/- Limits
1 - 2	0.0	1.28394
1 - 3	*-9.86667	1.28394
2 - 3	*-9.86667	1.28394

* denotes a statistically significant difference.

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

Thí nghiệm 2 yếu tố được bố trí 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 A, (4) nguồn biến động gây ra bởi yếu tố B, (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. Mô hình phân tích thống kê có dạng tổng quát: $Y_{ijch} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_c + \delta_h + \varepsilon_{ijch}$; 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; $(\alpha\beta)_{ij}$ - tham số đặc trưng cho 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, bạn có thể thiết kế mô hình như sau: $Y = \beta_0 + \beta_1(\text{khối hàng}) + \beta_2(\text{khối cột}) + \beta_3A + \beta_4B + \beta_5AB + \varepsilon$; 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.

Các giả thuyết: (1) Các yếu tố độc lập ảnh hưởng đồng đều đến biến phụ thuộc (kết quả thí nghiệm); (2) Không có ảnh hưởng tương tác giữa các yếu tố độc lập đến biến phụ thuộc; (3) Các khối hàng và khối cột ảnh hưởng đồng đều đến biến phụ thuộc.

Ví dụ. Giả thiết cần đánh giá ảnh hưởng của hai yếu tố (kích thước bầu và phân tổng hợp NPK) đến sinh trưởng chiều cao (H, cm) của Đậu nành 8 tháng tuổi: (1) Yếu tố kích thước bầu (kí hiệu X) với 2 mức độ: 15*25 cm và 20*30 cm; (2) 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*3 = 6 nghiệm thức. Mỗi nghiệm thức được lặp lại 6 lần. Các nghiệm thức được kí hiệu như sau: A = X₁P₁; B = X₁P₂; C = X₁P₃; D = X₂P₁; E = X₂P₂; F = X₂P₃. Hãy phân tích và so sánh kết quả 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 6.17): cột 1 hoặc hàng 1 nhập biến kết quả thí nghiệm (H, cm); cột 2 nhập số thứ tự nghiệm thức; cột 3 và 4 nhập tương ứng các khối theo hàng và theo cột; cột 5 nhập mã số biểu thị cho kích thước bầu theo khối hàng; cột 6 nhập mã số biểu thị cho phân tổng hợp NPK theo khối cột. Sau khi nhập xong dữ liệu, bạn thực hiện mã hóa cho các tên biến và ghi lại thành 1 file để có thể dùng về sau.

Bước 2. Kiểm định giả thuyết 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 những thủ tục tương tự như phân tích phương sai 2 yếu tố được bố trí theo khối. Ở đây chỉ trích dẫn bảng phân tích biến động (bảng 6.18) và bảng phân hạng các nghiệm thức (bảng 6.19) theo yếu tố kích thước bầu và NPK để các bạn tham khảo.

Bảng 6.17. 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
67.5	3	1	1	1	3
68.8	4	1	2	2	1
71.3	6	1	3	2	3
...	...	...	...	...	...

Bảng 6.18. Kiểm định ảnh hưởng giữa các yếu tố (ANOVA)

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A:COT	137.573	5	27.5145	8.97	0.0001
B:HANG	8.70917	5	1.74183	0.57	0.7237
C:KTBAU	755.334	1	755.334	246.18	0.0000
D:NPK	712.5	2	356.25	116.11	0.0000
INTERACTIONS					
CD	237.669	2	118.834	38.73	0.0000
RESIDUAL	61.3633	20	3.06817		
TOTAL (CORRECTED)					
		1913.15	35		

All F-ratios are based on the residual mean square error.

Bảng 6.19. Kiểm định phân hạng các nghiệm thức (Multiple Range Tests for H by NPK)

Method: 95.0 percent LSD			
NPK	Count	LS Mean	Homogeneous Groups
1	12	62.525	x
3	12	71.275	x
2	12	72.525	x
Contrast			
		Difference	+/- Limits
1 - 2		*-10.0	1.49167
1 - 3		*-8.75	1.49167
2 - 3		1.25	1.49167

* denotes a statistically significant difference.

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

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, bạn có thể thiết kế mô hình như sau: $Y = \beta_0 + \beta_1(\text{khối hàng}) + \beta_2(\text{khối cột}) + \beta_3A + \beta_4B + \beta_5AB + \varepsilon$; 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.

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 A với B đến kết quả thí nghiệm; (3) Các khối ảnh hưởng đồng đều đến kết quả thí nghiệm.

Ví dụ. Giả thiết cần xác định sinh trưởng về 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. Để đạt được mục đích đặt ra, bạn hãy thực hiện các bước sau:

Bước 1. Nhập số liệu vào bảng tính Statgraphics như sau (bảng 6.20): cột 1 hoặc hàng 1 nhập biến kết quả thí nghiệm (H, cm); cột 2 nhập số thứ tự phương thức lâm sinh (lô chính); cột 3 nhập loài cây (lô phụ); cột 4 nhập số khối (lần lặp). Sau đó bạn thực hiện mã hóa cho các tên biến và ghi lại thành 1 file để có thể dùng về sau.

Bảng 6.20. Phân tích phương sai 2 yếu tố được bố trí theo kiểu 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
...			

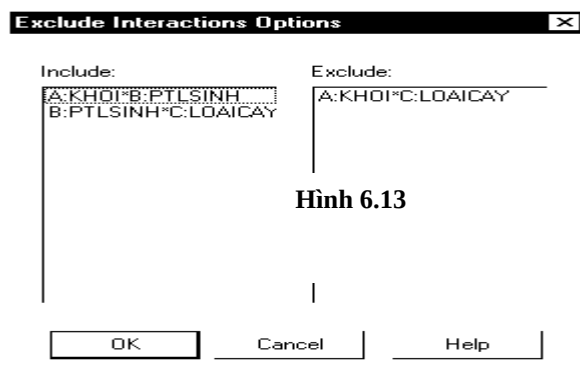
Bước 2. Kiểm định giả thuyết về sự khác biệt giữa các trung bình mẫu, so sánh sự khác nhau giữa các cặp số trung bình mẫu và kiểm định phân hạng các nghiệm thức được thực hiện theo những thủ tục sau đây:

+ Mở Compare → Analysis of Variance → Multifactor ANOVA → Dependent List (đến đây đưa vào biến phụ thuộc H) → factor (đưa vào các biến “Phương thức lâm sinh, Khối, Loài cây” bằng cách chọn 1 hay nhiều biến và nhấn nút mũi tên bên cạnh để chuyển chúng sang ô factors).

+ 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 H.

Trước hết, bạn hãy chọn biểu tượng Tabular options; kể đến tại cửa sổ Tabular options, bạn chọn All rồi chọn tiếp OK. Từ màn hình hiển thị kết quả, bạn nhấp đúp nút chuột trái lên trang Analysis of Variance for H - Type III Sums of Squares. Tiếp theo đặt con trỏ chuột lên trang này và nhấp nút chuột phải để chọn Analysis options. Khi cửa sổ Analysis options hiện ra, bạn xóa số 1 và ghi vào số 2 trong ô Maximum Order

Interactions; kể đến chọn Exclude. Khi cửa sổ Exclude Interactions Options xuất hiện (hình 6.13), bạn chỉ nhấp đúp nút chuột trái vào dòng (A: khối*C: Loài cây) để loại bỏ kết quả tương tác giữa chúng (vì các yếu tố này không tương tác với nhau). Bạn giữ lại cặp tương tác “A: Khối*B: Phương thức lâm sinh” để tính thống kê F cho kiểm định ảnh hưởng của phương thức lâm sinh trong lô chính. Để có thống kê F trong kiểm định ảnh hưởng của phương thức lâm sinh, bạn đặt con trỏ chuột vào bảng ANOVA và nhấp nút chuột phải, rồi chọn Pane options. Khi cửa sổ ANOVA table options xuất hiện, bạn chọn B: PTLSINH trong ô factor; cùng lúc đó một vệt sáng chuyển xuống dòng AB trong ô Error term. Lựa chọn này là mặc định; nó cho biết giá trị F bằng tỷ số giữa trung bình bình phương ứng với yếu tố phương thức lâm sinh và trung bình bình phương ứng với tương tác giữa phương thức lâm sinh với khối. Để nhận được cấu trúc của bảng ANOVA như bảng 6.21, bạn có thể chọn lần lượt A: Khối và AB trong ô Factor, rồi chọn None trong Error Term.



Hình 6.13

Bảng 6.21. Kiểm định ảnh hưởng của các yếu tố đến sinh trưởng của các loài cây

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A:KHỐI	61.6389	3	20.5463		
B:PTLSINH	476.722	2	238.361	6.22 (1)	0.0345
C:LOAICAY	997.389	2	498.694	138.81(0)	0.0000
INTERACTIONS					
AB	229.944	6	38.3241		
BC	1456.61	4	364.153	101.36 (0)	0.0000
RESIDUAL	4.6667	18	3.59259		
TOTAL (CORRECTED)					
		3286.97	35		

Đến đây bạn có thể chọn OK để hiển thị kết quả. Kết quả hiển thị trên màn hình bao gồm các bảng sau đây: ANOVA, 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 6.21), bảng kiểm định phân hạng các phương thức lâm sinh (bảng 6.22) và phân hạng sinh trưởng giữa ba loài cây (bảng 6.23) để các bạn tham khảo.

Bảng 6.22. Kiểm định phân hạng các phương thức lâm sinh

Method: 95.0 percent LSD			
PTLSINH	Count	LS Mean	Homogeneous Groups
3	12	59.4167	x
2	12	66.1667	x
1	12	67.8333	x
Contrast			
		Difference	+/- Limits
1 - 2		1.66667	6.18415
1 - 3		*8.41667	6.18415
2 - 3		*6.75	6.18415

* denotes a statistically significant difference.

Bảng 6.23. Kiểm định phân hạng sinh trưởng giữa ba loài cây

Method: 95.0 percent LSD			
LOAICAY	Count	LS Mean	Homogeneous Groups
1	12	58.8333	x
2	12	63.0833	x
3	12	71.5	x
Contrast			
		Difference	+/- Limits
1 - 2		*-4.25	1.6257
1 - 3		*-12.6667	1.6257
2 - 3		*-8.41667	1.6257

6.2.3. PHÂN TÍCH PHƯƠNG SAI BA YẾU TỐ

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”.

Mô hình phân tích thống kê có dạng:

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

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 A; β_j ($j = 1 \dots b$) - tham số đặc trưng cho ảnh hưởng của yếu tố B; γ_k - tham số đặc trưng cho ảnh hưởng của yếu tố C, $(\alpha\beta)_{ij}$ - tham số đặc trưng cho tương tác giữa yếu tố A với yếu tố B; $(\alpha\gamma)_{ik}$ - tham số đặc trưng cho tương tác giữa yếu tố A với yếu tố C, $(\beta\gamma)_{jk}$ - tham số đặc trưng cho tương tác giữa yếu tố B với yếu tố C, $(\beta\alpha\gamma)_{ijk}$ - tham số đặc trưng cho tương tác giữa 3 yếu tố (A, B và C), k_k - tham số đặc trưng cho ảnh hưởng của khối (lặp); ε_{ijk} - sai số ngẫu nhiên.

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, 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 của các yếu tố.

Thực hành. Giả sử cần thăm dò phản ứng sinh trưởng về D (mm) của Dầu song nòng 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 6.24.

Bảng 6.24. Sinh trưởng về D của Dầu song nòng 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	...	...

Phân tích phương sai ba yếu tố bằng chương trình Statgraphics Plus version 3.0 được thực hiện theo các bước sau đây:

Bước 1. Nhập số liệu vào bảng tính. 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 cho các tên biến và ghi lại 1 file để có thể dùng về sau.

Bước 2. Kiểm định giả thuyết 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 những thủ tục sau đây:

+ Mở Compare → Analysis of Variance → Multifactor ANOVA → Dependent (đến đây đưa vào biến phụ thuộc D) → factor (đưa vào các biến phân loại (khối, ánh sáng, phân bón và nước tưới) bằng cách chọn 1 hay nhiều biến và nhấn nút mũi tên bên cạnh để chuyển chúng sang ô factors. Sau đó chọn OK để chương trình xử lý và hiển thị kết quả ra màn hình.

+ 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 độc lập đến biến phụ thuộc. Trước hết, hãy chọn biểu tượng Tabular options; kể đến tại cửa sổ Tabular options, bạn chọn All rồi chọn tiếp OK. Từ màn hình hiển thị kết quả, bạn nhấp đúp nút chuột trái lên trang phân tích biến động (Analysis of Variance for D - Type III Sums of Squares). Khi trang phân tích biến động hiện ra, bạn đặt con trỏ chuột lên trang này và nhấp nút chuột phải để chọn Analysis options. Khi cửa sổ Analysis options hiện ra, bạn xóa số 1 và ghi vào số 3 trong ô Maximum Order Interactions. Thao tác này yêu cầu chương trình tính sự tương tác giữa từng cặp yếu tố và giữa 3 yếu tố thí nghiệm.

Bảng 6.25. Phân tích biến động (Analysis of Variance for D - Type III Sums of Squares)

Analysis of Variance for Do - Type III Sums of Squares

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A:KHỐI	0.00285208	1	0.00285208	0.68	0.4169
B:ANH SANG	7.70945	2	3.85472	923.55	0.0000
C:NUOC	0.0776021	1	0.0776021	18.59	0.0003
D:PHAN BON	4.95461	3	1.65154	395.69	0.0000
INTERACTIONS					
BC	0.0382167	2	0.0191083	4.58	0.0212
BD	1.0271	6	0.171183	41.01	0.0000
CD	0.0245896	3	0.00819653	1.96	0.1476
BCD	0.0762667	6	0.0127111	3.05	0.0242
RESIDUAL	0.0959979	23	0.00417382		
TOTAL (CORRECTED)	14.0067	47			

All F-ratios are based on the residual mean square error.

+ Cuối cùng bạn có thể chọn OK để hiển thị kết quả của các bảng sau đây: kiểm định ảnh hưởng giữa các yếu tố (bảng 6.25), ước lượng trung bình nghiệm thức, so sánh từng cặp nghiệm thức (Multiple Comparisons) và phân hạng các nghiệm thức (bảng 6.26)...

Bảng 6.26. Phân hạng các nghiệm thức theo yếu tố ánh sáng

Method: 95.0 percent LSD			
ANHSANG	Count	LS Mean	Homogeneous Groups
1	16	2.87437	x
2	16	3.16562	x
3	16	3.83188	x
Contrast	Difference		+/- Limits
1 - 2	*-0.29125		0.0468303
1 - 3	*-0.9575		0.0468303
2 - 3	*-0.66625		0.0468303

* denotes a statistically significant difference.

Giải thích và nhận xét kết quả. Bảng 6.25 báo cáo kết quả phân tích biến động. 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 6.26 báo cáo kiểm định phân hạng các nghiệm thức. Từ bảng 6.26 cho thấy, sinh trưởng của Dầu song nòng dưới nghiệm thức 3 là trội hơn nghiệm thức 2 và 1.

6.3. PHÂN TÍCH HIỆP PHƯƠNG SAI

Phân tích hiệp phương sai (Analysis of covariance = ANOCOVA) được áp dụng cho những đối tượng thí nghiệm mà ngay từ ban đầu đã không có sự đồng nhất về những đặc tính nào đó. Những đặc tính này có thể ảnh hưởng đến kết quả thí nghiệm. Ví dụ: Khi xác định ảnh hưởng của độ tàn che đến sinh trưởng của Dầu song nòng, nhưng cây con đưa vào thí nghiệm lại có chiều cao khác nhau hoặc đường kính khác nhau.

Lưu ý rằng, trong phân tích ANOCOVA, vì thí nghiệm được thực hiện trên những đối tượng không đồng nhất về một hay nhiều đặc tính nào đó (kí hiệu Y_i), nên trước khi so sánh các trung bình nghiệm thức 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 hay tu chỉnh (Y_{dc}). Để đạt được mục đích đặt ra, trước hết ta phải đồng nhất những giá trị của các đặc tính ban đầu đưa vào nghiên cứu. Sau đó tính những giá trị trung bình điều chỉnh (Y_{dc}) - đó là những giá trị trung bình được tính bằng cách san phẳng sự khác biệt giữa các giá trị của đặc tính ban đầu đưa vào nghiên cứu. Để tính được Y_{dc} của từng yếu tố, trước hết ta tính Y' trung bình của các đặc tính ban đầu. Nếu gọi Y'_i (với $i = 1 \div n$) là giá trị thứ i của một đặc tính ban đầu nào đó ở các nghiệm thức thì $Y' = \sum Y'_i / N$; trong đó N là tổng dung lượng của hai mẫu. Sau đó giá trị trung bình của biến phụ thuộc Y trong từng mẫu được điều chỉnh (Y_{dc}) theo công thức:

$$+ \text{Đối với nghiệm thức A: } Y_{dc}(A) = Y_{bq}(A) + b(Y' - Y_{Abq})$$

$$+ \text{Đối với nghiệm thức B: } Y_{dc}(B) = Y_{bq}(B) + b(Y' - Y_{Bbq})$$

Trong các công thức trên, b là hệ số hồi quy của mô hình tuyến tính bậc 1 giữa biến Y_i với biến Y . Dưới đây chúng ta xem xét một số mô hình phân tích hiệp phương sai thường được áp dụng.

6.3.1. 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 cần nghiên cứu phản ứng tăng trưởng về đường kính (ZD, mm) của cây con Dầu song năng dưới ảnh hưởng của hai biện pháp chăm sóc khác nhau. Biện pháp thứ nhất (A) được thực hiện trên 11 cây con, còn biện pháp thứ hai (B) được thực hiện trên 19 cây con. Những cây con đưa vào thí nghiệm có kích thước đường kính ban đầu (D, cm) khác nhau. Thí nghiệm được theo dõi trong 3 tháng. Số liệu về D ban đầu và lượng tăng trưởng đường kính (Zd, mm) của Dầu song năng sau 3 tháng thí nghiệm được ghi lại ở bảng 6.27. Để giải bài toán này, bạn hãy thực hiện theo các bước sau đây:

Bước 1. Nhập dữ liệu vào bảng tính. Trước hết, 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 chăm sóc cây con. Sau đó bạn thực hiện mã hóa cho các tên biến.

Bảng 6.27. Cách nhập dữ liệu trong phân tích Anocova 1 nhân tố được bố trí ngẫu nhiên
(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

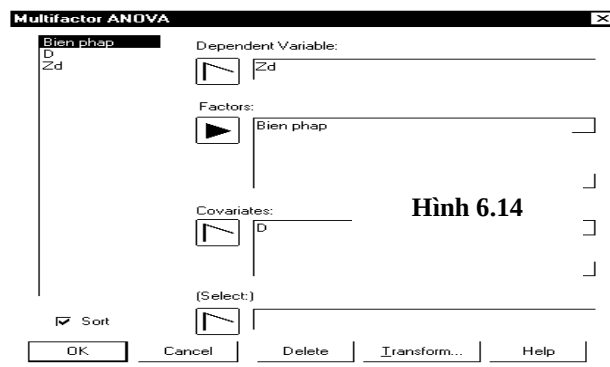
Bước 2. Gọi chương trình tính toán. Từ thanh Menu chính, bạn hãy chọn Compare $\Rightarrow$ Analysis of Variance $\Rightarrow$ Multifactor ANOVA. Khi cửa sổ Multifactor ANOVA xuất hiện (hình 6.14), bạn hãy thực hiện như sau:

- Chọn biến Zd và chuyển nó sang ô Dependent Variable;
- Chọn biến D và chuyển nó vào ô Covariate;
- Chọn biến biện pháp nuôi dưỡng và chuyển nó sang ô factor(s);
- Chọn $\Rightarrow$ OK $\Rightarrow$ Tabular Options $\Rightarrow$ All $\Rightarrow$ OK.

Đến đây kết quả phân tích thống kê sẽ hiển thị ra màn hình những nội dung sau đây: bảng phân tích biến động (bảng 6.28), bảng số trung bình đã được điều chỉnh (bảng 6.29), kiểm định phân hạng các nghiệm thức (bảng 6.30).

Giải thích và nhận xét kết quả. Bảng 6.28 báo cáo kết quả phân tích biến động của tăng trưởng đường kính thân cây (Zd) thành các thành phần gây ra biến động (ở đây là biện pháp lâm sinh và D ban đầu). Vì bạn chọn mặc định là tổng bình phương

kiểu III, nên sự đóng góp của mỗi yếu tố đã được điều chỉnh theo các yếu tố khác. Thống kê F và xác suất P kiểm định mức ý nghĩa thống kê của mỗi yếu tố (các biện pháp nuôi



đường cây và kích thước cây ban đầu). Giả thuyết H_0 cho rằng các biện pháp nuôi dưỡng cây có ảnh hưởng đồng đều đến tăng trưởng đường kính thân cây. Ở dòng có nhãn D, vì $P = 0.000 < 0.01$ nên đường kính thân cây ban đầu có ảnh hưởng rất rõ rệt đến tăng trưởng đường kính thân cây. Ngược lại, các biện pháp nuôi dưỡng cây sau 3 tháng có ảnh hưởng đồng đều đến tăng trưởng đường kính thân cây (vì $P = 0.086 > 0.05$).

Bảng 6.29 dẫn các trị trung bình Zđ đã được điều chỉnh cho từng mức của yếu tố (biện pháp lâm sinh), các sai số chuẩn và khoảng tin cậy 95% của từng số trung bình mẫu. Bảng 6.30 báo cáo kết quả kiểm định phân hạng các nghiệm thức theo thủ tục so sánh bội (Multiple comparisons) nhằm xác định sự khác biệt có ý nghĩa thống kê giữa nghiệm thức này với nghiệm thức khác. Ở phần dưới của bảng 6.30 báo cáo sự khác biệt giữa các cặp số trung bình đã được điều chỉnh. Ở phần trên của bảng 6.30 báo cáo kết quả kiểm định thuần nhất hai trung bình mẫu. Những cặp số trung bình không có sự khác biệt có ý nghĩa về mặt thống kê được sắp xếp trên cùng một cột. Phương pháp dùng để phân định các trung bình mẫu là thống kê LSD. Thủ tục LSD có 5% sai số (mạo hiểm) khi cho rằng mỗi cặp số trung bình khác nhau có ý nghĩa trong khi đó chúng lại không khác biệt có ý nghĩa.

Bảng 6.28. Kiểm định ảnh hưởng giữa các yếu tố (ANOCOVA)

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
COVARIATES					
D	5.4319	1	5.4319	28.86	0.0000
MAIN EFFECTS					
A:Biện pháp	0.597604	1	0.597604	3.18	0.0860
RESIDUAL	5.0815	27	0.188204		
TOTAL					
	10.592	29			

All F-ratios are based on the residual mean square error.

Bảng 6.29. Các ước lượng đã được điều chỉnh (Estimates)

Level	Count	Mean	Std. Error	Lower Limit	Upper Limit
GRAND MEAN					
	30	2.10002			
Biện pháp					
1	11	1.9501	0.13278	1.67766	2.22254
2	19	2.24994	0.1004	2.04394	2.45595

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

Bảng 6.30. So sánh các nghiệm thức (Pairwise Comparisons)

Multiple Range Tests for Zđ by Biện pháp

Method: 95.0 percent LSD			
Bien phap	Count	LS Mean	Homogeneous Groups
1	11	1.9501	x
2	19	2.24994	x
Contrast		Difference	+/- Limits
1 - 2		-0.299842	0.337244

Bảng 6.31. Chiều cao ban đầu (H, cm) và lượng tăng trưởng đường kính (Zh, cm) của Dầu song nà sau hai tháng

Hỗn hợp ruột bầu (thí nghiệm)	A	H	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

2. Trường hợp có nhiều nghiệm thức. Giả sử cần nghiên cứu phản ứng sinh trưởng về chiều cao (Zh, cm) của cây con Dầu con rái dưới ảnh hưởng của 4 hỗn hợp ruột bầu khác nhau. Mỗi nghiệm thức có 10 cây. Cây con đưa vào thí nghiệm có chiều cao ban đầu (H, cm) khác nhau. Các nghiệm thức được bố trí ngẫu nhiên trong cùng hoàn cảnh. Thí nghiệm được theo dõi trong hai tháng. Số liệu về H ban đầu và lượng tăng trưởng về chiều cao (Zh, cm) của Dầu con rái sau 2 tháng thí nghiệm được ghi lại ở bảng 6.31.

Để giải bài toán này, bạn hãy thực hiện như sau:

Bước 1. Nhập dữ liệu vào bảng tính (*xem phụ lục 10*). Trước hết, bạn nhập lần lượt biến H và Zh của 4 nghiệm thức (hỗn hợp ruột bầu) vào cột 1 và 2; cột 3 nhập mã số biểu thị cho 4 nghiệm thức. Sau đó bạn mã hóa cho các tên biến và ghi lại thành 1 file để có thể dùng về sau.

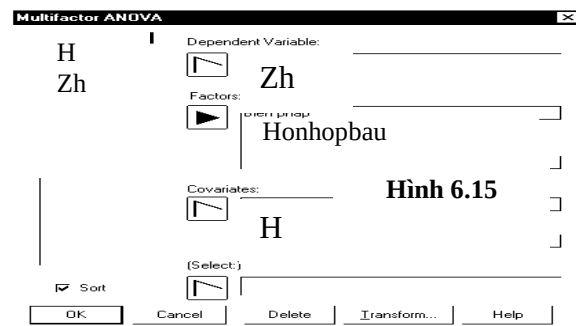
Bảng 6.32. Phân tích ANOCOVA về ảnh hưởng của hỗn hợp ruột bầu đến sinh trưởng chiều cao của Dầu con rái

Analysis of Variance for Zh - Type III Sums of Squares

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
COVARIATES					
H	0.229503	1	0.229503	10.81	0.0023
MAIN EFFECTS					
A:HONHOP	1.21985	3	0.406617	19.15	0.0000
RESIDUAL	0.742997	35	0.0212285		
TOTAL (CORRECTED)	2.2306	39			

All F-ratios are based on the residual mean square error.

Bước 2. Gọi chương trình tính toán. Từ thanh Menu chính, bạn hãy chọn Compare $\Rightarrow$ Analysis of Variance $\Rightarrow$ Multifactor ANOVA. Khi cửa sổ Multifactor ANOVA xuất hiện (hình 6.15), bạn hãy chọn biến Zh và chuyển nó sang ô Dependent Variable. Kế đến chọn biến H và chuyển nó vào ô Covariates. Tiếp theo chọn biến hỗn hợp ruột bầu và chuyển nó sang ô factor(s). Cuối cùng bạn chọn OK $\Rightarrow$ Tabular Options $\Rightarrow$ All $\Rightarrow$ OK.



Hình 6.15

Đến đây kết quả xử lý sẽ hiển thị ra màn hình bảng phân tích biến động (bảng 6.32), bảng số trung bình đã được điều chỉnh (bảng 6.33), so sánh cặp và kiểm định phân hạng các nghiệm thức (bảng 6.34). Nội dung và cách giải thích kết quả của các bảng này cũng tương tự như trường hợp có 2 mẫu.

Bảng 6.33. Bảng trung bình nghiệm thức đã điều chỉnh

Table of Least Squares Means for Zh
with 95.0 Percent Confidence Intervals

Level	Count	Mean	Std. Error	Lower Limit	Upper Limit
GRAND MEAN	40	1.41			
HONHOP					
1	10	1.64806	0.0462691	1.55413	1.74199
2	10	1.49274	0.0462125	1.39892	1.58656
3	10	1.30927	0.0461802	1.21552	1.40302
4	10	1.18993	0.0461397	1.09626	1.2836

Bảng 6.34. So sánh các cặp nghiệm thức và kiểm định phân hạng

Multiple Range Tests for Zh by HONHOP

Method: 95.0 percent LSD

HONHOP	Count	LS Mean	Homogeneous Groups
4	10	1.18993	X
3	10	1.30927	X
2	10	1.49274	X
1	10	1.64806	X
Contrast		Difference	+/- Limits
1 - 2		*0.155317	0.13228
1 - 3		*0.338785	0.13228
1 - 4		*0.45813	0.13228
2 - 3		*0.183468	0.13228
2 - 4		*0.302813	0.13228
3 - 4		0.119345	0.13228

* denotes a statistically significant difference.

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

Giả sử cần so sánh sinh trưởng về chiều cao (H, mm) của 6 dòng Gió bầu sau 2 tháng thí nghiệm dưới ảnh hưởng của mật độ cây (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 6.35. Trình tự phân tích ANOCOVA như sau:

Bước 1. Nhập dữ liệu vào bảng tính (xem phụ lục 11). Trước hết, 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 Gió bầu. Sau đó bạn mã hóa cho các tên biến.

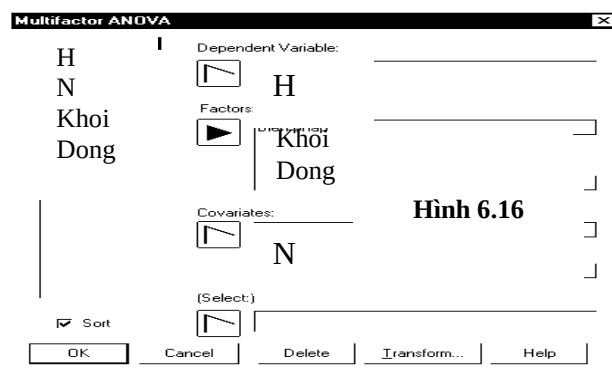
Bảng 6.35. Mật độ cây và chiều cao (H, cm) của Gió bầu sau một tháng

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

Bước 2. Gọi chương trình tính toán. Từ thanh Menu chính, hãy chọn Compare ⇒ Analysis of Variance ⇒ Multifactor ANOVA. Khi cửa sổ Multifactor ANOVA xuất hiện (hình 6.16), bạn hãy thực hiện như sau:

- Chọn biến H và chuyển nó sang ô Dependent Variable; chọn biến N và chuyển nó vào ô Covariates. Sau đó chọn hai biến “khối và dòng” rồi chuyển chúng sang ô factor(s).
- Chọn OK ⇒ Tabular Options ⇒ All ⇒ OK.

Đến đây kết quả xử lý sẽ hiển thị ra màn hình bảng phân tích biến động (bảng 6.36), bảng số trung bình đã được điều chỉnh (bảng 6.37), so sánh cặp và kiểm định phân hạng các nghiệm thức (bảng 6.38).



Hình 6.16

Bảng 6.36. Phân tích ANOCOVA về ảnh hưởng của hỗn hợp ruột bầu đến sinh trưởng chiều cao của Dầu con rái

Analysis of Variance for H - Type III Sums of Squares					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
COVARIATES					
N	7391.05	1	7391.05	76.01	0.0000
MAIN EFFECTS					
A:KHOI	1502.39	3	500.798	5.15	0.0132
B:DONG	3227.28	5	645.455	6.64	0.0023
RESIDUAL	1361.29	14	97.2347		
TOTAL (CORRECTED)	18678.5	23			

All F-ratios are based on the residual mean square error.

Bảng 6.37. Bảng trung bình nghiệm thức đã điều chỉnh

Table of Least Squares Means for H
with 95.0 Percent Confidence Intervals

Level	Count	Mean	Std. Error	Lower Limit	Upper Limit
GRAND MEAN	24	199.75			
KHOI					
1	6	188.961	4.07252	180.227	197.696
2	6	206.234	4.16754	197.296	215.173
3	6	194.266	4.16754	185.327	203.204
4	6	209.539	4.07252	200.804	218.273
DONG					
1	4	191.802	5.38138	180.26	203.344
2	4	190.979	5.03102	180.189	201.77
3	4	193.157	4.93279	182.577	203.737
4	4	219.32	5.16541	208.242	230.399
5	4	189.585	5.10127	178.644	200.526
6	4	213.657	4.93279	203.077	224.237

Bảng 6.38. Kiểm định phân hạng các dòng Gió bầu

Multiple Range Tests for H by DONG

Method: 95.0 percent LSD			
DONG	Count	LS Mean	Homogeneous Groups
5	4	189.585	X
2	4	190.979	X
1	4	191.802	X
3	4	193.157	X
6	4	213.657	X
4	4	219.32	X

Giải thích và nhận xét kết quả. Bảng 6.36 dẫn kết quả phân tích ANOCOVA. Bạn hãy chú ý đến dòng có nhãn “DONG”. Vì giá trị $P = 0.002 < 0.01$, nên sinh trưởng chiều cao của các dòng Gió bầu là khác nhau rất có ý nghĩa. Ở dòng có nhãn N, giá trị $P =$

0.000; điều đó cho biết mật độ cây có ảnh hưởng đến sinh trưởng chiều cao của các dòng Gió bầu. Số liệu của bảng 6.37 là chiều cao trung bình điều chỉnh và sai tiêu chuẩn về H của 6 dòng Gió bầu thuộc từng khối và của chung 4 khối. Bảng 6.38 là các so sánh phân hạng về H trung bình đã điều chỉnh của 6 dòng Gió bầu.

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 6.39. Để giải bài toán này, bạn hãy thực hiện như sau:

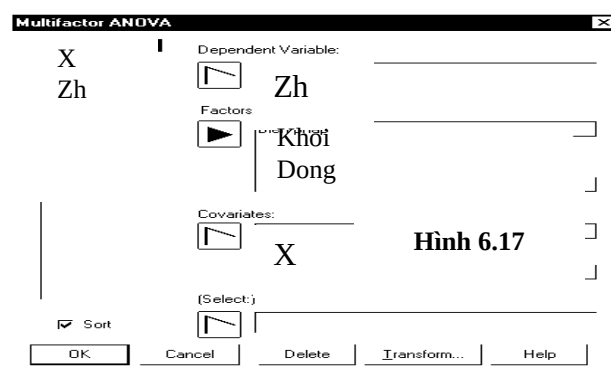
Bước 1. Nhập dữ liệu vào bảng tính (xem phụ lục 12). Trước hết, bạn nhập lần lượt biến X và Zh vào cột 1 và 2; cột 3 và 4 nhập tương ứng mã số biểu thị cho 4 khối cột và 4 khối dòng; cột 5 nhập mã số biểu thị cho dòng Keo lai. Sau đó thực hiện mã hóa cho các tên biến.

Bảng 6.39. Tỷ lệ bệnh và tăng trưởng chiều cao (Zh, cm) của Keo lai sau một tháng

			Khối cột (Dòng Keo lai)			
			I	II	III	IV
Khối dòng (Đợt thu hoạch)	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

Bước 2. Gọi chương trình tính toán. Từ thanh Menu chính, hãy chọn Compare $\Rightarrow$ Analysis of Variance $\Rightarrow$ Multifactor ANOVA. Khi cửa sổ Multifactor ANOVA xuất hiện (hình 6.17), bạn hãy thực hiện như sau:

+ Chọn biến Zh và chuyển nó sang ô Dependent Variable; chọn biến X và chuyển nó vào ô Covariates. Sau đó chọn hai biến khối và dòng rồi chuyển chúng sang ô factor(s).



Hình 6.17

+ Cuối cùng chọn OK $\Rightarrow$ Tabular Options $\Rightarrow$ All $\Rightarrow$ OK.

Đến đây kết quả xử lý sẽ hiển thị ra màn hình bảng phân tích biến động (bảng 6.40), bảng số trung bình đã được điều chỉnh (bảng 6.41), so sánh cặp và kiểm định phân hạng các nghiệm thức (bảng 6.42).

Bảng 6.40. Phân tích ANOCOVA

Analysis of Variance for Zh - Type III Sums of Squares

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
COVARIATES					
X	45.3242	1	45.3242	22.89	0.0050
MAIN EFFECTS					
A:KHOI	6.42239	3	2.1408	1.08	0.4367
B:DONG	29.4019	3	9.80063	4.95	0.0587
C:DOT	23.1397	3	7.71324	3.90	0.0887
RESIDUAL	9.90083	5	1.98017		
TOTAL (CORRECTED)	360.178	15			

All F-ratios are based on the residual mean square error.

Bảng 6.41. Giá trị chiều cao trung bình của 4 dòng Keo lai đã được điều chỉnh

Level	Count	Mean	Std. Error	Lower Limit	Upper Limit
GRAND MEAN	16	25.3375			
DONG					
1	4	23.7636	0.710899	21.9362	25.591
2	4	27.5236	0.705508	25.71	29.3372
3	4	24.8691	0.70394	23.0595	26.6786
4	4	25.1937	0.70725	23.3757	27.0118
DOT					
1	4	25.7845	1.62059	21.6186	29.9504
2	4	24.1546	0.746643	22.2353	26.0739
3	4	24.2944	0.922663	21.9226	26.6662
4	4	27.1166	0.933268	24.7175	29.5156
KHOI					
1	4	25.1068	0.709663	23.2826	26.9311
2	4	24.5262	0.712964	22.6935	26.359
3	4	25.4379	0.720265	23.5864	27.2894
4	4	26.2791	0.715772	24.4391	28.119

Giải thích và nhận xét kết quả. Bảng 6.39 dẫn kết quả phân tích ANOCOVA. Bạn hãy chú ý đến dòng có nhãn DONG. Giá trị $P = 0.059 \approx 0.05$; điều đó cho biết tăng trưởng chiều cao của các dòng Keo lai là khác nhau. Bảng 6.40 là các giá trị tăng trưởng về chiều cao của 4 dòng Keo lai đã được điều chỉnh. Bảng 6.41 là kiểm định phân hạng và các so sánh từng cặp nghiệm thức với nhau (Zh đã được điều chỉnh). Dấu (*) ở bên cạnh các con số của cột Difference cho biết các trung bình mẫu khác nhau có ý nghĩa thống kê ở mức 0.05.

Bảng 6.42. So sánh từng cặp và phân hạng trung bình nghiệm thức

Multiple Range Tests for Zh by DONG

Method: 95.0 percent LSD			
DONG	Count	LS Mean	Homogeneous Groups
1	4	23.7636	X
3	4	24.8691	X
4	4	25.1937	XX
2	4	27.5236	X
Contrast	Difference		+/- Limits
1 - 2	*-3.75998		2.55781
1 - 3	-1.10546		2.55781
1 - 4	-1.4301		2.55781
2 - 3	*2.65452		2.55781
2 - 4	2.32988		2.55781
3 - 4	-0.324636		2.55781

* denotes a statistically significant difference.

6.3.2. 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 khác nhau (X). Kết quả thí nghiệm được ghi lại ở bảng 6.43. Để giải bài toán này, bạn hãy thực hiện như sau:

Bảng 6.43. Chiều cao (H, cm) của Vên vên năng 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

Bước 1. Nhập dữ liệu vào bảng tính (xem phụ lục 13). Trước hết, 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 thực hiện mã hóa cho các tên biến.

Bước 2. Gọi chương trình tính toán. Từ thanh Menu chính, hãy chọn Compare $\Rightarrow$ Analysis of Variance $\Rightarrow$ Multifactor ANOVA. Khi cửa sổ Multifactor ANOVA xuất hiện, bạn hãy thực hiện 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 Factor(s). Cuối cùng chọn OK $\Rightarrow$ Tabular Options $\Rightarrow$ All $\Rightarrow$ OK.

Đến đây kết quả xử lý sẽ hiển thị ra màn hình bảng phân tích biến động (bảng 6.44), bảng số trung bình đã được điều chỉnh (bảng 6.45), so sánh cặp và kiểm định phân hạng các nghiệm thức (bảng 6.46).

Bảng 6.44. Kiểm định ảnh hưởng giữa các yếu tố (ANOCOVA)

Analysis of Variance for H - Type III Sums of Squares					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
COVARIATES					
X	8.61551	1	8.61551	38.22	0.0000
MAIN EFFECTS					
A:KHOI	7.22683	5	1.44537	6.41	0.0027
B:NGHIEM	2.03505	3	0.67835	3.01	0.0658
RESIDUAL	3.1555	14	0.225393		
TOTAL (CORRECTED)	59.3377	23			

All F-ratios are based on the residual mean square error.

Bảng 6.45. Các ước lượng Zh đã tu chỉnh (Estimates)

NGHIEM					
1	6	4.19411	0.236768	3.6863	4.70193
2	6	4.97941	0.220542	4.50639	5.45242
3	6	4.43048	0.239243	3.91735	4.9436
4	6	5.42933	0.258481	4.87495	5.98372

Giải thích và nhận xét kết quả. Bảng 6.44 dẫn kết quả phân tích ANOCOVA để kiểm định ảnh hưởng của các yếu tố phân bón đến sinh trưởng về chiều cao (H, cm) của Vên vên sau 2 tháng tuổi. 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 thực sự khác nhau với xác suất sai lầm $P = 0.07$. Bảng 6.45 là các giá trị chiều cao trung bình đã được điều chỉnh của 4 nghiệm thức. Bảng 6.46 là các so sánh từng cặp và kiểm định phân hạng trung bình nghiệm thức. Dấu (*) ở bên cạnh các con số của cột Difference cho biết các trung bình mẫu khác nhau có ý nghĩa thống kê ở mức 0.05. Từ số liệu của bảng 6.46 cho thấy, ở những nghiệm thức có bón lân, lượng tăng trưởng về chiều cao của Vên vên sau 2 tháng thí nghiệm đều vượt trội hơn so với các nghiệm thức khác.

Bảng 6.46. So sánh từng cặp và kiểm định phân hạng trung bình nghiệm thức

Multiple Range Tests for H by NGHIEM

Method: 95.0 percent LSD

NGHIEM	Count	LS Mean	Homogeneous Groups
1	6	4.19411	X
3	6	4.43048	XX
2	6	4.97941	XX
4	6	5.42933	X

Contrast	Difference	+/- Limits
1 - 2	*-0.785292	0.587888
1 - 3	-0.236363	0.587888
1 - 4	*-1.23522	0.587888
2 - 3	0.548929	0.587888
2 - 4	-0.449928	0.587888
3 - 4	*-0.998858	0.587888

* denotes a statistically significant difference.

6.3.3. 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 năng sau 3 tháng thí nghiệm. Những cây con Dầu song năng 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 6.47. Để giải bài toán này, bạn hãy thực hiện như sau:

Bước 1. Nhập dữ liệu vào bảng tính (xem phụ lục 14). Trước hết, 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 bạn mã hóa cho các tên biến.

Bảng 6.47. Lượng tăng trưởng về đường kính (Zd, mm) của Dầu song năng sau 3 tháng tuổi

Thí nghiệm thứ	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

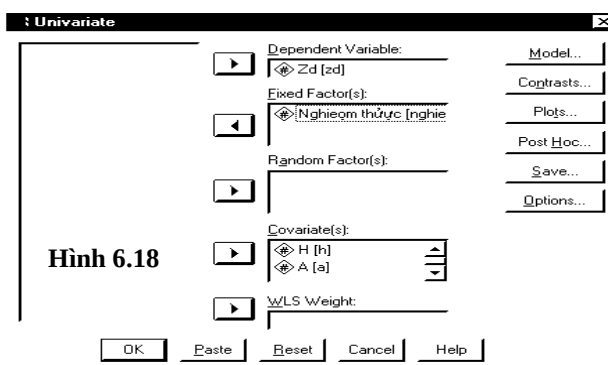
Bảng 6.48. Kiểm định ảnh hưởng giữa các yếu tố (ANOMUCOVA)

Analysis of Variance for zd - Type III Sums of Squares

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
COVARIATES					
A	0.148005	1	0.148005	6.83	0.0132
H	0.044082	1	0.044082	2.04	0.1628
MAIN EFFECTS					
A:NGHIEM	0.1599	3	0.0533	2.46	0.0794
RESIDUAL	0.736494	34	0.0216616		
TOTAL (CORRECTED)	1.06278	39			

All F-ratios are based on the residual mean square error.

Bước 2. Gọi chương trình tính toán. Từ thanh Menu chính, hãy chọn Compare ⇒ Analysis of Variance ⇒ Multifactor ANOVA. Khi cửa sổ Multifactor ANOVA xuất hiện (hình 6.18), 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; sau đó chọn biến thí nghiệm rồi chuyển chúng sang ô Factor(s). Đến đây chọn OK ⇒ Tabular Options ⇒ All ⇒ OK. Ở bảng 6.48 – 6.50 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 cũng tương tự như phân tích ANOCOVA.



Hình 6.18

Bảng 6.49. Các ước lượng trung bình mẫu đã điều chỉnh (Estimates)

Table of Least Squares Means for zd
with 95.0 Percent Confidence Intervals

Level	Count	Mean	Std. Error	Lower Limit	Upper Limit
GRAND MEAN	40	1.38825			
NGHIEM					
1	10	1.46015	0.0469672	1.3647	1.5556
2	10	1.31595	0.046645	1.22116	1.41075
3	10	1.4416	0.0468063	1.34648	1.53673
4	10	1.33529	0.0468967	1.23999	1.4306

Bảng 6.50. So sánh từng cặp và kiểm định phân hạng các nghiệm thức

Multiple Range Tests for zd by NGHIEM

Method: 95.0 percent LSD			
NGHIEM	Count	LS Mean	Homogeneous Groups
2	10	1.31595	X
4	10	1.33529	XX
3	10	1.4416	XX
1	10	1.46015	X
Contrast	Difference		+/- Limits
1 - 2	*0.144199		0.133763
1 - 3	0.018546		0.133763
1 - 4	0.124859		0.133763
2 - 3	-0.125653		0.133763
2 - 4	-0.0193399		0.133763
3 - 4	0.106313		0.133763

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 nà 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 ghi lại ở bảng 6.51.

Để giải bài toán này, bạn hãy thực hiện các bước giống như trường hợp phân tích ANOMUCOVA với thí nghiệm 1 yếu tố được bố trí ngẫu nhiên. Bạn có thể tham khảo cách nhập dữ liệu vào bảng tính ở phụ lục 15. Ở đây chỉ lưu ý rằng, bạn phải chọn cùng lúc cả biến H và L_R rồi chuyển chúng vào ô Covariance. Một số kết quả phân tích thống kê ở ví dụ này được ghi lại ở bảng 6.52 – 6.54. Việc giải thích các nội dung của các bảng này cũng được thực hiện tương tự như phân tích ANOVA với thí nghiệm 1 yếu tố được bố trí theo khối ngẫu nhiên.

Bảng 6.51. Lượng tăng trưởng về chiều cao của Dầu song năng 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 6.52. Kiểm định ảnh hưởng giữa các yếu tố (ANOMUCOVA)

Analysis of Variance for Zh - Type III Sums of Squares

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
COVARIATES					
H	194.475	1	194.475	55.61	0.0001
Lr	29.2239	1	29.2239	8.36	0.0202
MAIN EFFECTS					
A:Khoi	92.1583	2	46.0792	13.18	0.0029
B:nghiem	100.717	5	20.1435	5.76	0.0151
RESIDUAL	27.9772	8	3.49715		
TOTAL (CORRECTED)	982.298	17			

All F-ratios are based on the residual mean square error.

Bảng 6.53. Các ước lượng trung bình mẫu đã điều chỉnh (Estimates)

Level	Count	Mean	Std. Error	Lower Limit	Upper Limit
GRAND MEAN	18	32.0111			
Khoi					
1	6	28.7851	0.76837	27.0133	30.557
2	6	33.0302	1.04443	30.6217	35.4386
3	6	34.2181	1.08331	31.7199	36.7162
nghiem					
1	3	30.6719	1.29379	27.6885	33.6554
2	3	28.9692	1.2161	26.1649	31.7735
3	3	31.3859	1.31395	28.356	34.4159
4	3	31.6357	1.22405	28.8131	34.4584
5	3	31.7325	1.70046	27.8112	35.6537
6	3	37.6714	1.38032	34.4883	40.8544

Bảng 6.54. So sánh từng cặp và kiểm định phân hạng nghiệm thức

Multiple Range Tests for Zh by nghiem

Method: 95.0 percent LSD

nghiem	Count	LS Mean	Homogeneous Groups
2	3	28.9692	X
1	3	30.6719	X
3	3	31.3859	X
4	3	31.6357	X
5	3	31.7325	X
6	3	37.6714	X

Chương VII

SO SÁNH NHIỀU MẪU

Thủ tục so sánh nhiều mẫu (2 hay nhiều nghiệm thức) báo cáo các thống kê mô tả, phân nhóm, kiểm định sự bằng nhau của các phương sai mẫu, kiểm định sự bằng nhau của hai hay nhiều số bình quân bằng kiểm định T (T – Test), kiểm định kiểu phân bố của hai hay nhiều mẫu theo kiểm định Kolmogorov – Smirnov Test. Những chi tiết về các kiểm định này được trình bày trong “Hướng dẫn sử dụng chương trình SPSS 10.0”.

7.1. SO SÁNH HAI MẪU BẰNG KIỂM ĐỊNH KOLMOGOROV – SMIRNOV

Phương pháp so sánh hai mẫu bằng kiểm định Kolmogorov – Smirnov có nội dung như sau:

- + Đem hai dãy quan sát chia tổ với cùng cự ly như nhau.
- + Thống kê tần số quan sát trong mỗi tổ của từng dãy quan sát (f_i) và tổng số quan sát của mỗi tổ (N_i).
- + Tại mỗi chuỗi quan sát, tính tỷ lệ F_i/N_i cho mỗi tổ, trong đó F_i là tần số tích lũy từ tổ 1 đến tổ i .
- + Tính hiệu $D_i = F_i/N_i$ (tổ 1) – F_i'/N_i' (tổ 2) giữa hai tổ thuộc hai dãy quan sát tương ứng và ghi nhận giá trị $D_{\max}$.

+ So sánh $D_{\max}$ với $K_\alpha \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$ ứng với mức xác suất α (0.05, 0.01...).

+ Nếu $D_{\max} < K_\alpha \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$ thì hai dãy quan sát là thuần nhất, và ngược lại.

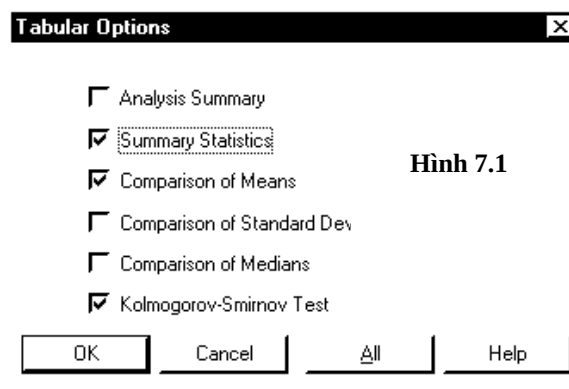
Khi sử dụng Statgraphics plus version 3.0 để phân tích hai mẫu bằng kiểm định Kolmogorov – Smirnov, bạn hãy thực hiện theo các bước sau đây:

Bước 1. Nhập số liệu; trong đó mỗi cột là một nghiệm thức.

Bước 2. Từ menu chính, mở Compare $\Rightarrow$ Two Sample $\Rightarrow$ Two Sample Comparisons.

Bước 3. Mở biểu tượng Tabular Options và chọn kiểm định Kolmogorov – Smirnov Test (hình 7.1). Sau cùng bạn chọn OK để chương trình tính toán và hiển thị kết quả.

Bước 4. Nhận xét kết quả. Hãy quan sát bảng kiểm định Kolmogorov – Smirnov Test. Khi $P > 0.05$ (mặc định) thì giả thuyết H_0 cho rằng hai trung bình mẫu bằng nhau được chấp nhận. Ngược lại, khi $P < 0.05$ (mặc định) thì hai trung bình mẫu khác nhau ở mức tin cậy 95%.



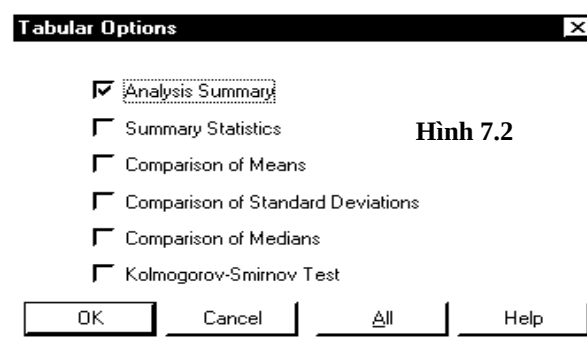
Hình 7.1

7.2. KIỂM ĐỊNH SỰ BẰNG NHAU CỦA HAI TRUNG BÌNH MẪU BẰNG T-TEST

Điều kiện. Hai mẫu có phân bố chuẩn và phương sai bằng nhau. Kiểm định sự bằng nhau của hai trung bình mẫu bằng t-test được thực hiện theo thủ tục sau đây:

Bước 1. Nhập hai dãy số liệu cần so sánh, trong đó mỗi cột là một nghiệm thức.

Bước 2. Mở Compare $\Rightarrow$ Two Samples $\Rightarrow$ Two Sample Comparison. Đến đây bạn hãy chọn mẫu 1 và đưa nó vào Sample 1 bằng cách nhấp mũi tên bên cạnh Sample 1; tiếp theo chọn mẫu 2 và đưa nó vào Sample 2. Sau khi vào số liệu, hãy chọn OK. Tiếp theo bạn chọn Tabular Options, rồi chọn các thống kê cần tính (hình 7.2).



Hình 7.2

Bước 3. Nhận xét kết quả. Hãy quan sát bảng kiểm định T. Khi $P > 0.05$ (mặc định) thì giả thuyết H_0 cho rằng hai trung bình mẫu bằng nhau sẽ được chấp nhận. Ngược lại, khi $P < 0.05$ (mặc định) thì hai trung bình mẫu khác nhau. Ở đây ta cũng có thể căn cứ vào khoảng xác định mức tin cậy 95% của sự khác biệt giữa hai trung bình mẫu để so sánh. Nếu khoảng xác định này không chứa số Zero thì hai trung bình mẫu khác nhau ở mức tin cậy 95%.

7.3. SO SÁNH NHIỀU MẪU (HƠN HAI MẪU)

Khi các mẫu có phân bố chuẩn với phương sai bằng nhau thì bạn hãy dùng phương pháp phân tích phương sai (chọn ANOVA Table) để thực hiện việc so sánh nhiều mẫu. Ngược lại, khi các mẫu không tuân theo luật phân phối chuẩn và phương sai của các mẫu khác nhau thì bạn cần sử dụng phương pháp kiểm định phi tham số của Kruskal – Wallis. Thống kê Kruskal – Wallis kiểm định giả thuyết H_0 cho rằng các số trung vị (Medians) của K mẫu là bằng nhau. Kiểm định này dựa trên thống kê kiểm định χ^2 . Nếu $P\text{-Value} < 0.05$ thì K mẫu là khác nhau ở mức tin cậy 95% (mặc định). Để thấy sự khác nhau giữa các số trung vị bằng trực quan, bạn có thể dùng biểu đồ Box – and Whisker Plot để quan sát. Ngược lại, nếu $P\text{-Value} > 0.05$ thì K mẫu khác nhau không có ý nghĩa thống kê ở mức tin cậy 95% (mặc định).

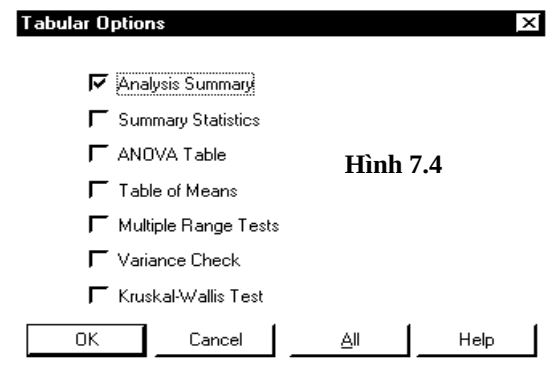
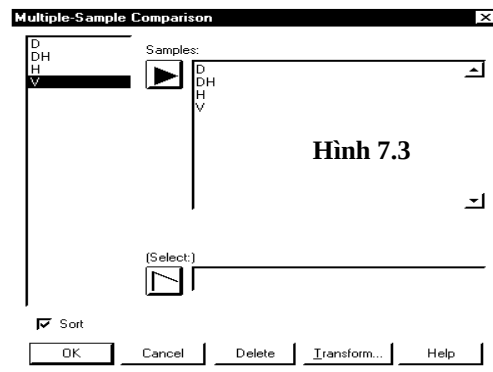
Thực hành. Giả sử cần kiểm định sự khác biệt về đường kính của 4 lâm phần Sao đen được xử lý theo 4 phương thức lâm sinh khác nhau. Để nhận được kết quả phân tích sự khác biệt giữa 4 nghiệm thức bằng thủ tục so sánh nhiều mẫu của Kruskal – Wallis, bạn hãy thực hiện theo các bước sau đây:

Bước 1. Nhập 4 nghiệm thức vào 4 cột khác nhau.

Bước 2. Mở Compare $\Rightarrow$ Multiple Samples $\Rightarrow$ Multiple Sample Comparison (hình 7.3). Tiếp theo hãy mở Tabular Option và chọn những thống kê cần tính (hình 7.4). Đến đây kết quả hiển thị trên màn hình như bảng 7.1.

Giải thích và nhận xét kết quả. 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 nhóm có phân phối giống nhau thì $H \approx \chi^2$. Giả thuyết H_0 cho rằng trung vị của K mẫu là bằng nhau. Để kiểm định giả thuyết này, các dãy số liệu được tập hợp lại và xếp

hạng từ nhỏ nhất đến lớn nhất. Sau đó hạng bình quân được tính cho mỗi nghiệm thức (cột 3). Nếu $P < 0.05$ (mặc định) thì giả thuyết bị bác bỏ. Ngược lại, nếu $P > 0.05$ (mặc định) thì giả thuyết H_0 được chấp nhận. Ở ví dụ này, giá trị kiểm định Kruskal Wallis = 30.2586 với $P = 0.0000 < 0.01$; do đó 4 mẫu (trung vị) khác nhau có ý nghĩa thống kê ở mức tin cậy 99%.



Bảng 7.1. Kiểm định Kruskal – Wallis

Kruskal-Wallis Test

	Sample Size	Average Rank
TN1	148	352.348
TN2	151	335.93
TN3	156	284.974
TN4	156	254.083
Test statistic = 30.2586 P-Value = 0.00000121759		

Bảng 7.2. Kiểm định ngang bằng phương sai

Cochran's C test: 0.270742 P-Value = 0.800286
 Bartlett's test: 1.0036 P-Value = 0.53717
 Hartley's test: 1.20913

Bảng 7.3. Phân tích phương sai (ANOVA)

ANOVA Table

Analysis of Variance					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Between groups	1148.83	3	382.942	13.01	0.0000
Within groups	17869.3	607	29.4387		
Total (Corr.)	19018.1	610			

Bảng 7.4. Kiểm định phân hạng các nghiệm thức

Multiple Range Tests

Method: 95.0 percent LSD

	Count	Mean	Homogeneous Groups
TN4	156	17.6679	X
TN3	156	18.6679	X
TN2	151	20.543	X
TN1	148	21.0473	X
Contrast	Difference		+/- Limits
TN1 - TN2	0.504251		1.23251
TN1 - TN3	*2.37935		1.2227
TN1 - TN4	*3.37935		1.2227
TN2 - TN3	*1.8751		1.21645
TN2 - TN4	*2.8751		1.21645
TN3 - TN4	1.0		1.2065

* denotes a statistically significant difference.

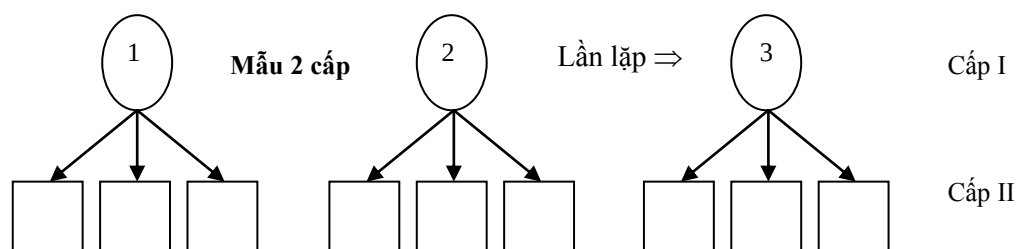
Lưu ý rằng, với số liệu của ví dụ trên đây bạn cũng có thể xử lý bằng phương pháp phân tích phương sai 1 nhân tố được bố trí ngẫu nhiên. Để đạt được mục đích này, từ cửa sổ Tabular Option ở bước 2, bạn chọn All và OK. Kết quả sẽ báo cáo các thống kê mô tả, kiểm định ngang bằng phương sai mẫu (bảng 7.2), bảng ANOVA (bảng 7.3) và kiểm định phân hạng các nghiệm thức (bảng 7.4). Từ số liệu của bảng phân tích ANOVA, bạn cũng nhận được kết quả giống như kiểm định Kruskal Wallis. Ngoài ra, từ số liệu của bảng 7.4, bạn còn biết được những nghiệm thức nào là khác nhau có ý nghĩa thống kê.

Chương VIII

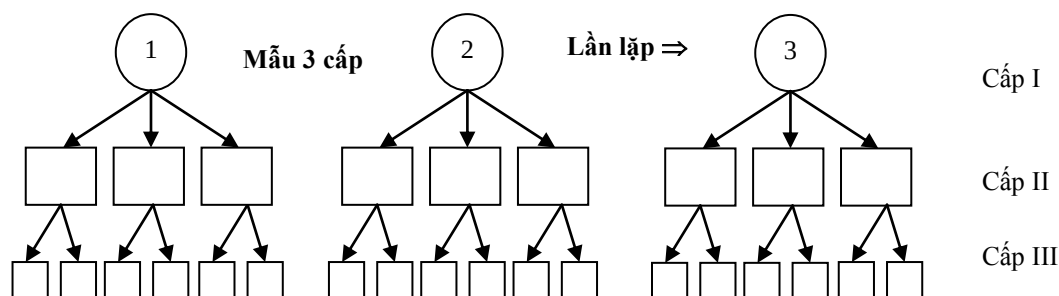
KIỂU THÍ NGHIỆM PHÂN CẤP

Kiểu thí nghiệm phân cấp (Nested designs) được dùng để phân tích ảnh hưởng của 1 hoặc nhiều yếu tố định tính đến một biến phản hồi (phụ thuộc) khi số liệu được thu thập theo một trật tự nhất định (theo thứ bậc) và số mẫu là bằng nhau ở tất cả các tổ hợp của các mức yếu tố. Mô hình phải được thiết kế cân bằng, nghĩa là có số mẫu bằng nhau đối với mỗi tổ hợp của các mức yếu tố. Dưới đây là một số ví dụ về mẫu phân cấp.

+ **Mẫu 2 cấp**. Mẫu cấp I là cây; mẫu cấp II là lá cây.



+ **Mẫu 3 cấp**. Mẫu cấp I: Cây; mẫu cấp II: lá; mẫu cấp III: hàm lượng khoáng.



Phương pháp thu thập mẫu phân cấp

Bảng số liệu của mẫu 2 cấp

Mẫu cấp I	Số lần đo lặp lại (Mẫu cấp II)			Tổng
	1	k	n	
1	Y ₁₁	Y _{1k}	Y _{1n}	T _{1*}
i	Y _{i1}	Y _{ik}	Y _{in}	T _{i*}
a	Y _{a1}	Y _{ak}	Y _{an}	T _{a*}
Tổng	T*1	T*k	T*n	ΣT

Bảng số liệu của mẫu 3 cấp

Mẫu cấp I (yếu tố A)	Mẫu cấp II (Yếu tố B)	Mẫu cấp III (Lần đo lặp)			Tổng		
					T _{ij} *	T _i **	T***
1	1	Y ₁₁₁	Y _{11k}	Y _{11n}			
	j	Y _{1j1}	Y _{1jk}	Y _{1jn}			
	b	Y _{1b1}	Y _{1bk}	Y _{1bn}			
i	1	...	...	...			
	j		Y _{ijk}				
	b						
a	1						
	j						
	b						

+ Khi lấy mẫu 2 cấp thì cấu trúc của bảng ANOVA có dạng như sau:

Nguồn biến động	df	SS	MS	F _{tt}
A	a - 1	SSA	MSA	MSA/MSB(A)
B(A)	a(n - 1)	SSB	MSB(A)	
Tổng	an - 1	SST		

+ Khi lấy mẫu 3 cấp thì cấu trúc của bảng ANOVA có dạng như sau:

Nguồn biến động	df	SS	MS	F _{tt}
A	a - 1	SSA	MSA	MSA/MSB(A)
B(A)	a(b - 1)	SSB(A)	MSB(A)	MSB(A)/MSK(B)
K(B)	ab(n - 1)	SSK(B)	MSK(B)	
Tổng	abn - 1	SST		

Lưu ý rằng, trong thí nghiệm phân cấp, việc xử lý số liệu như thế nào là tùy thuộc vào mục tiêu nghiên cứu của bạn. Nếu bạn quan tâm đến yếu tố hay mức nào có ảnh hưởng nhiều hơn đến biến phản hồi thì bạn cần phân tích theo mô hình phân cấp. Ngược lại, nếu bạn mong muốn phát hiện ảnh hưởng tương tác giữa các mức yếu tố đến biến phản hồi thì bạn cần phân tích theo mô hình đa yếu tố.

Thực hành. Giả sử bạn cần thăm dò sự khác biệt về hàm lượng tinh dầu ở Bạch đàn chanh mọc trên 4 lập địa khác nhau, đồng thời bạn cũng muốn biết hàm lượng tinh dầu có khác nhau giữa những lá trên cùng một cây hay không. Để giải quyết vấn đề đặt ra, người ta đã thu thập mẫu như sau: (1) Từ 4 cây Bạch đàn chanh mọc trên 4 lập địa khác nhau, mỗi cây lấy 3 lá ở 3 vị trí khác nhau trên thân cây, mỗi lá thực hiện 2 thí nghiệm phân tích hàm lượng tinh dầu. Kết quả hàm lượng tinh dầu trong lá cây dẫn ra ở bảng 8.1. Để phân tích phương sai, bạn thực hiện như sau:

Bảng 8.1. Hàm lượng tinh dầu trong lá Bạch đàn chanh

TT. Cây	TT. Lá	Hàm lượng tinh dầu	
1	1	3.28	3.09
	2	3.52	3.48
	3	2.88	2.8
2	1	2.46	2.44
	2	1.87	1.92
	3	2.19	2.19
3	1	2.77	2.66
	2	3.74	3.44
	3	2.55	2.55
4	1	3.78	3.87
	2	4.07	4.12
	3	3.31	3.31

Bước 1. Nhập số liệu vào bảng tính. Số liệu nhập vào bảng tính bao gồm hai cột: Cột 1 nhập biến phụ thuộc (tinh dầu = Y_i) hay kết quả thí nghiệm theo thứ tự từ nghiệm thức 1 đến nghiệm thức 4 (cây); cột 2 nhập mã biến độc lập (X_i) hay mã biến lập nhóm (biến phân loại - Grouping Variable) biểu thị cho cây; cột 2 nhập mã biến độc lập biểu thị cho lá cây (bảng 8.2). Sau đó bạn thực hiện mã hóa cho các tên biến.

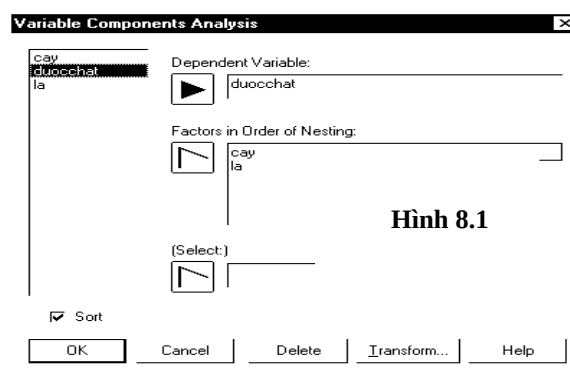
Bảng 8.2. Cách nhập dữ liệu trong phân tích phương sai một yếu tố
(Chi tiết số liệu xem ở phụ lục 16)

Tinh dầu	3.28	3.09	3.52	3.48	2.88	2.8	2.46	2.44	1.87	1.92	2.19	2.19	..
Cây	1	1	1	1	1	1	2	2	2	2	2	2	...
Lá	1	1	2	2	3	3	1	1	2	2	3	3	

Bước 2. Thực hiện phân tích ANOVA, so sánh các cặp giá trị trung bình mẫu, kiểm định phân hạng các nghiệm thức (Multiple Range Tests).

+ Trước hết, nếu bạn quan tâm đến thành phần (yếu tố) nào gây ra biến động hàm lượng tinh dầu lớn hơn thì bạn hãy phân tích các thành phần đóng góp trong biến động của hàm lượng tinh dầu. Ở trường hợp này, bạn cần kiểm định giả thuyết cho rằng, cây mọc ở những lập địa khác nhau gây ra biến động của hàm lượng tinh dầu lớn hơn so với lá trên các vị trí khác nhau của cây. Để kiểm định giả định này, bạn tiến hành như sau:

Từ menu chính, chọn Compare → Analysis of variance → Variance



Hình 8.1

Components → Dependent Variable (đến đây hãy đưa biến phụ thuộc vào bằng cách chọn nó và nhấn nút mũi tên bên cạnh) → Factor in Order of Nesting (đưa các biến lập nhóm (cây và lá) vào bằng cách chọn chúng và nhấn nút mũi tên bên cạnh)(hình 8.1).

Bảng 8.3. Phân tích biến động tinh dầu

Analysis of Variance for Tinh dau

Source	Sum of Squares	Df	Mean Square	Var. Comp.	Percent
TOTAL (CORRECTED)	10.2704	23			
cay	7.56035	3	2.52012	0.365223	68.53
la	2.6302	8	0.328775	0.16106	30.22
ERROR	0.07985	12	0.00665417	0.00665417	1.25

Bảng 8.4. Phân tích ANOVA

Analysis of Variance for Tinh dau - Type III Sums of Squares

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A:cay	7.56035	3	2.52012	378.73	0.0000
B:la	1.21106	2	0.605529	91.00	0.0000
INTERACTIONS					
AB	1.41914	6	0.236524	35.55	0.0000
RESIDUAL	0.07985	12	0.00665417		
TOTAL (CORRECTED)	10.2704	23			

All the error.

Bảng 8.5. Kiểm định phân hạng nghiệm thức

+ Kiểm định phân hạng theo cây

Multiple Range Tests for Tinh dau by cay

Method: 95.0 percent LSD

cay	Count	LS Mean	Homogeneous Groups
2	6	2.17833	X
3	6	2.95167	X
1	6	3.175	X
4	6	3.74333	X

Contrast	Difference	+/- Limits
1 - 2	*0.996667	0.102614
1 - 3	*0.223333	0.102614
1 - 4	*-0.568333	0.102614
2 - 3	*-0.773333	0.102614
2 - 4	*-1.565	0.102614
3 - 4	*-0.791667	0.102614

* denotes a statistically significant difference.

+ Kiểm định phân hạng theo lá cây

Multiple Range Tests for Tinh dầu by la

Method: 95.0 percent LSD

la	Count	LS Mean	Homogeneous Groups
3	8	2.7225	X
1	8	3.04375	X
2	8	3.27	X

Contrast	Difference	+/- Limits
1 - 2	*-0.22625	0.0888664
1 - 3	*0.32125	0.0888664
2 - 3	*0.5475	0.0888664

* denotes a statistically significant difference.

Để thực hiện các phân tích thống kê, từ cửa sổ Variable Component Analysis, bạn hãy chọn OK để chương trình xử lý và hiển thị kết quả phân tích các thành phần biến động. Tại bảng tính các thành phần biến động, bạn mở biểu tượng Tabular Options và chọn ALL; kết quả bạn sẽ nhận được các thống kê mô tả và bảng phân tích phương sai (bảng 8.3).

Giải thích và nhận xét kết quả. Bảng 8.3 báo cáo kết quả phân tích biến động của tinh dầu. Trong bảng 8.3, biến động của tinh dầu được chia ra 2 thành phần: một là biến động tinh dầu do cây gây ra và 2 là biến động tinh dầu do lá gây ra. Sự đóng góp của mỗi thành phần được đánh giá bằng phần trăm (Percent), và được ghi lại ở cột 6. Trong ví dụ này, cây đóng góp vào sự biến động của tinh dầu (68.53%) lớn hơn lá (30.22%).

+ Nếu bạn mong muốn phát hiện ảnh hưởng tương tác giữa các mức yếu tố (cây và lá) đến biến động của hàm lượng tinh dầu thì bạn cần phân tích dữ liệu trên theo mô hình đa yếu tố. Cách thực hiện gọi chương trình tính toán giống như phân tích phương sai hai yếu tố được bố trí ngẫu nhiên. Ở ví dụ này, sau khi gọi chương trình tính toán, bạn sẽ nhận được kết quả phân tích phương sai (bảng 8.4) và kiểm định phân hạng các nghiệm thức (bảng 8.5).

TÀI LIỆU THAM KHẢO CHÍNH

- 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.
- 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 lí, xã hội, Nxb. Khoa Học và Kỹ Thuật.
- Vũ Tiến Hinh** và các tác giả khác, 1992. Điều tra, quy hoạch và điều chế rừng, Trường Đại Học Lâm Nghiệp.
- Vũ Tiến Hinh**, 2003. Sản lượng rừng, Nxb. Nông Nghiệp, Hà Nội.
- Nguyễn Ngọc Kiểng**, 1996. Thống kê học trong nghiên cứu khoa học, Nxb. Giáo Dục.
- Nguyễn Ngọc Kiểng**, 2000. Thống kê học ứng dụng: Các kiểu mẫu thí nghiệm, 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.
- Trần Bá Nhẫn và Đinh Thái Hoàng**, 1998. Lí thuyết thống kê ứng dụng trong quản trị, kinh doanh và nghiên cứu kinh tế, Nxb. Thống Kê.
- Nguyễn Hải Tuất**, 1982. Thống kê toán học trong lâm nghiệp, Nxb. Nông Nghiệp, Hà Nội.
- 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.
- Phan Văn Tân**, 2003. Các phương pháp thống kê trong khí hậu, Nxb. Đại Học Quốc Gia Hà Nội.
- V.I. Vasilevich**, 1969. Những phương pháp thống kê trong địa thực vật, Nxb. “Khoa học”, Leningrad (tiếng Nga).
- E.C. Ventsen**, 1969. Lí thuyết xác suất, Nxb. “Khoa học”, Moxkva (tiếng Nga).
- Greig – Smith P**, 1964. Quantitative plant ecology, Ed. 2. London.
- David R. Anderson, Dennis J. Sweeney, Thomas A. Williams**, 2002. Statistics for business and economics, 8th ed., BookMasters, Inc.

PHỤ LỤC

Phụ lục 1. ANOVA một nhân 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 nhân tố được bố trí theo khối ngẫu nhiên

TT	H, cm	Nghiệm	Khối	TT	H, cm	Nghiệm	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 nhân tố được bố trí theo kiểu bình phương la tinh

TT	H, cm	Nghiệm	TT. Hàng	TT. Cột	TT	H, cm	Nghiệm	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 nhân 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 nhân 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 nhân tố được bố trí ngẫu nhiên

TT	H, cm	TTcây	Đạm	Lân	TT	H, cm	TTcây	Đạm	Lân	TT	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

(xem tiếp trang sau)

Phụ lục 5. ANOVA hai nhân 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 nhân 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

(xem tiếp trang sau)

Phụ lục 7. ANOVA hai nhân tố được bố trí theo kiểu lô chính - lô phụ (Tiếp)

TT	H, cm	Phương thức	Loài cây	Khối	TT	H, cm	Phương thức	Loài cây	Khối
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 nhân 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 nhân 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 nhân 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 nhân 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 nhân tố được bố trí theo kiểu bình phương la tinh

TT	Tỷ 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	14.0	25.1	4	4	2
9	14.0	26.0	1	1	3
10	7.2	24.9	2	2	3
11	6.3	29	3	3	3
12	1.0	28.7	4	4	3
13	8.9	29.8	1	1	4
14	6.7	29.0	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	4.00	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	8.00	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	Nghiem
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	2.80	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 ngạnh	Chò xốt-Lành ngạnh	Hà nu - Lành ngạnh
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 xót	Lành ngành	Chò xót-Lành ngành	Hà nu - Lành 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	Lành ngành	Chò xốt-Lành ngành	Hà nu - Lành 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
100	1	1	0	0	0	0
101	0	1	1	0	0	0

Tiếp phụ biểu 17

Thứ tự	Trắc	Hà nu	Chò xốt	Lành ngành	Chò xốt-Lành ngành	Hà nu - Lành ngành
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
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

Tiếp phụ biểu 17

Thứ tự	Trắc	Hà nu	Chò xót	Lành ngành	Chò xót-Lành ngành	Hà nu - Lành ngành
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
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

Tiếp phụ biểu 17

Thứ tự	Trắc	Hà nu	Chò xốt	Lành ngành	Chò xốt-Lành ngành	Hà nu - Lành ngành
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
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

Tiếp phụ biểu 17

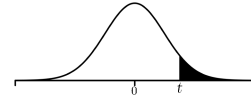
Thứ tự	Trắc	Hà nu	Chò xót	Lành ngành	Chò xót-Lành ngành	Hà nu - Lành ngành
204	1	1	1	1	1	1
205	0	1	0	1	0	1
206	0	0	0	1	0	0
207	1	0	1	0	0	0
208	1	1	0	1	0	1
209	0	1	1	0	0	0
210	1	1	1	0	0	0
211	1	1	0	1	0	1
212	0	1	0	0	0	0
213	0	1	1	0	0	0
214	1	1	1	1	1	1
215	1	0	1	1	1	0
216	1	1	1	0	0	0
217	1	1	0	1	0	1
218	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
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

Tiếp phụ biểu 17

Thứ tự	Trắc	Hà nu	Chò xót	Lành ngành	Chò xót-Lành ngành	Hà nu - Lành ngành
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
251	0	1	0	0	0	0
252	0	1	1	0	0	0
253	1	1	0	1	0	1
254	1	1	1	0	0	0
255	1	0	0	1	0	0
256	0	1	1	0	0	0
257	1	0	0	0	0	0
258	1	1	1	1	1	1
259	1	0	0	1	0	0
260	1	1	0	0	0	0
261	0	0	1	0	0	0
262	0	1	1	1	1	1
263	1	0	1	0	0	0
264	1	1	0	0	0	0
265	1	0	1	1	1	0
266	0	0	1	0	0	0
267	1	0	1	0	0	0
268	1	1	0	1	0	1
269	0	0	0	1	0	0
270	0	0	1	1	1	0
271	1	0	0	1	0	0

Tiếp phụ biểu 17

Thứ tự	Trắc	Hà nu	Chò xót	Lành ngành	Chò xót-Lành ngành	Hà nu - Lành ngành
272	1	1	1	0	0	0
273	1	1	0	0	0	0
274	0	0	1	0	0	0
275	1	1	0	1	0	1
276	1	1	1	0	0	0
277	1	1	0	1	0	1
278	0	0	1	1	1	0
279	1	1	1	0	0	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

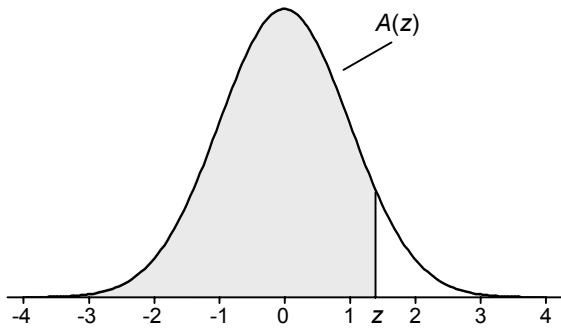
Critical values of the t distribution

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”

