PHAN THANH KIẾM

CƠ SỞ TOÁN HỌC

CỦA CÁC PHÉP XỬ LÝ THỐNG KÊ TRONG NGHIÊN CỨU KHOA HỌC NÔNG NGHIỆP

NHÀ XUẤT BẢN NÔNG NGHIỆP

PGS. TS. PHAN THANH KIẾM

CƠ SỞ TOÁN HỌC

CỦA CÁC PHÉP XỬ LÝ THỐNG KÊ TRONG NGHIÊN CỨU KHOA HỌC NÔNG NGHIỆP

NHÀ XUẤT BẢN NÔNG NGHIỆP

Tp. Hồ Chí Minh – 2010

MỤC LỤC

MỘT SỐ THUẬT NGỮ VÀ KÝ HIỆU

Thuật ngữ	**Tiếng Anh**
Dấu hiệu (đặc trưng) định lượng	Quantitative characteristics
Dấu hiệu (đặc trưng) định tính	Qualitative characteristics
Dung lượng (kích thước) mẫu	Size of sample
Đại lượng (biến) ngẫu nhiên	Random variable
Độ lệch chuẩn	Standard deviation
Độ tin cậy	Degree of confidence
Độ tự do	Degree of freedom
Giả thiết thống kê	Statisticcal hypothesis
Đối thiết	Alternative hypothesis
Hàm phân phối	Distribution function
Hàm mật độ xác suất	Probability density function
Hệ số góc	Slope
Hệ số đường	Path coefficient
Hệ số tương quan	Correlation coefficient
Hiệp phương sai (hiệp sai)	Covariance
Hồi quy tuyến tính	Linear regression
Hồi quy phi tuyến tính	Non - linear regression
Kỳ vọng (kỳ vọng toán)	Mathematical expectation
Mẫu	Sample
Phân tích đường	Path analysis
Phương pháp (nguyên tắc) bình phương tối thiểu	Method (principle) of least squares
Phương sai	Variance (dispersion)
Sai lầm	Risk

Sai số tiêu chuẩn (sai số chuẩn)	Standard error
Tham số (thông số) thống kê	Statistical parameter
Thống kê mô tả	Descriptive statistics
Tổng thể	Population
Tương quan	Correlation
Trung bình (trung bình cộng)	Mean, sample mean, average
Ước lượng điểm	Point estimate
Ước lượng khoảng	Interval estimate

Ký hiệu	**Nghĩa**
AB	Nghiệm thức phối hợp giữa hai yếu tố A với B
X_{ij}	Giá trị nghiệm thức A_iB_j
A×B	Tương tác giữa hai yếu tố A với B
ab_{ij}	Giá trị hiệu quả tương tác $A_i \times B_j$
ABC	Nghiệm thức phối hợp giữa ba yếu tố A, B với C
X_{ijl}	Giá trị nghiệm thức $A_iB_jC_l$
A×B×C	Tương tác giữa ba yếu tố A với B với C
abc_{ijl}	Giá trị hiệu quả tương tác $A_i \times B_j \times C_l$

LỜI NÓI ĐẦU

Thống kê toán học ra đời rất sớm và có mặt ở hầu hết các lĩnh vực hoạt động của con người, từ khoa học tự nhiên, kinh tế học đến khoa học xã hội và nhân văn. A. Ketle (1796 – 1874), F. Galton (1822 – 1911), K. Pearson (1857 – 1936), W. S. Gosset (Student, 1876 – 1937), R. A. Fisher (1890 – 1962), M. Mitrel (1874 – 1948) là những người đặt nền móng cho thống kê sinh học hiện đại.

Trong quá trình phát triển, thống kê sinh học không dừng lại ở việc mô tả, suy đoán mà đã trở thành môn "khoa học về các tiêu chuẩn của việc tính toán". Trong sự lớn mạnh của thống kê sinh học có sự đóng góp đáng kể của các nhà khoa học thực nghiệm.

Năm 1973, khi đề cập đến công tác cải cách giáo dục, UNESCO đã khẳng định rằng Xác suất – Thống kê là một trong 9 vấn đề chủ chốt để xây dựng nền học vấn hiện đại.

Để giúp cho các sinh viên, học viên cao học và những nghiên cứu viên am hiểu cơ sở toán học của các phép xử lý số liệu trong nghiên cứu khoa học nông nghiệp, cuốn sách này được biên soạn. Nội dung của sách gồm hai phần:

- Phần đầu là các phương pháp lấy mẫu, điều tra thu thập và xử lý số liệu, từ thống kê mô tả, ước lượng các tham số thống kê đến việc so sánh và phân tích mối quan hệ giữa các tham số.

- Phần hai là các kiểu bố trí thí nghiệm, các phương pháp xử lý số liệu và cách trình bày báo cáo khoa học.

Để giúp bạn đọc không chuyên ngành thống kê có thể dễ nắm bắt được các nội dung, trong phần đầu tác giả đã trình bày dưới dạng ứng dụng, hạn chế việc lạm dụng các thuật ngữ thống kê. Tuy nhiên các nội dung vẫn đảm bảo tính khoa học, tính logic và tính thực tiễn. Ở phần hai tác giả đã cố gắng để làm rõ *cơ sở lý luận của các kiểu bố trí thí nghiệm, phương pháp phân tích số liệu giúp cho người đọc có thể nắm bắt được và ứng dụng để bố trí và xử lý số liệu các thí nghiệm trong chậu, trong phòng và thí nghiệm đồng ruộng.*

Mặc dù ngày càng có nhiều phần mềm tính toán ra đời làm cho việc xử lý các số liệu tiến hành nhanh chóng, nhưng những hiểu biết về cơ sở của các phép tính toán là rất quan trọng, nó giúp cho việc kiểm tra các kết quả tính toán, phân tích và đánh giá đúng các hiện tượng trong nghiên cứu, tránh những sai sót trong sử dụng các phần mềm thống kê.

Tác giả xin chân thành cảm ơn Thầy Nguyễn Đình Hiền Đại học Nông nghiệp Hà Nội, người đã đóng góp nhiều ý kiến quý báu cho nội dung của cuốn sách.

Không thể tránh khỏi những thiếu sót về nội dung và hình thức, rất mong được sự góp ý của bạn đọc. Mọi góp ý xin gửi về:

Bộ môn Di truyền – Chọn giống
Khoa Nông học, Đại học Nông Lâm Tp. HCM.

hoặc E-mail: ptkiem@hotmail.com
ptkiem1@gmail.com

Xin giới thiệu cùng bạn đọc.

Tác giả

Phần 1

XỬ LÝ SỐ LIỆU ĐIỀU TRA KHẢO SÁT

Chương 1

THỐNG KÊ MÔ TẢ - CÁC THAM SỐ THỐNG KÊ

Để nghiên cứu các đối tượng, công việc đầu tiên là điều tra, thu thập số liệu và dùng các tham số thống kê để mô tả đối tượng nghiên cứu. Chương này sẽ đề cập đến các vấn đề:

- Tổng thể và mẫu;

- Các tham số đặc trưng của mẫu và tổng thể.

1.1. TỔNG THỂ VÀ MẪU

1.1.1. Tổng thể

1.1.1.1. Khái niệm

Theo quan điểm thống kê, tổng thể nghiên cứu hay tổng thể là toàn bộ các phần tử hay cá thể có cùng một hay một số đặc trưng (dấu hiệu) định tính hay định lượng nào đó của đối tượng nghiên cứu.

Trong nông học, một tổng thể có thể là một quần thể cây trồng gồm nhiều cá thể. Một tổng thể cũng có thể là một nhân tố cụ thể liên quan đến cây trồng cần được nghiên cứu như một khu đất canh tác khi giả thiết rằng nó bao gồm vô số mẫu đất cần được khảo sát, đánh giá.

Số lượng các phần tử hay cá thể (dưới đây được gọi chung là cá thể) trong tổng thể được gọi là kích thước, cỡ hay dung lượng (dưới đây được gọi là dung lượng) tổng thể, ký hiệu là N. Thường thì dung lượng tổng thể là một số hữu hạn, nhưng nếu tổng thể quá lớn hoặc không thể nắm được toàn bộ các cá thể, ta có thể coi dung lượng của tổng thể là vô hạn. Điều này dựa trên cơ sở, rằng khi dung lượng của tổng thể tăng lên khá lớn thì ảnh hưởng không đáng kể đến kết quả tính toán cho tổng thể từ số liệu thu được trên từng bộ phận rút ra từ tổng thể đó.

1.1.1.2. Các loại dấu hiệu của tổng thể

Có thể chia các dấu hiệu tổng thể thành hai loại: các dấu hiệu định tính và các dấu hiệu định lượng.

- Các dấu hiệu định tính, còn được gọi là các dấu hiệu về chất (hay dấu hiệu chất lượng) là các dấu hiệu có thể phân biệt sự khác nhau giữa các cá thể hay nhóm cá thể bằng mắt, nếm hay thử. Ví dụ như có lông, râu hoặc không có, màu vàng hay màu xanh, hạt trần hay có màng, tròn hay dài, trơn hay nhăn, nhiễm hay kháng bệnh v.v. Đối với loại dấu hiệu này người ta có phương pháp nghiên cứu riêng biệt.

- Các dấu hiệu định lượng, còn được gọi là các dấu hiệu về lượng (hay dấu hiệu số lượng) là các dấu hiệu không thể phân biệt sự khác nhau giữa các cá thể hay nhóm cá thể bằng mắt, mà phải tiến hành cân, đo, đếm và phân biệt được nhờ sử dụng các phép toán thống kê. Ví dụ như khối lượng hạt, củ, quả, thân, rễ, độ lớn, độ dài của các bộ phận, số lượng hạt, củ, quả, v.v.

Sự phân chia này có tính tương đối vì bất kỳ một dấu hiệu chất lượng nào cũng có thể lượng hóa bằng các mức độ khác nhau, và có nhiều dấu hiệu số lượng cũng có thể phân biệt bằng mắt được như to, trung bình hay nhỏ, cao, trung bình hay thấp, dài hay ngắn, nhiều hay ít.

1.1.1.3. Các phương pháp mô tả tổng thể

- ***Bằng bảng phân bố tần số***

Nếu gọi các trị số x_i nhận được từ phép xác định nào đó và n_i $(i=\overline{1,n})$ là các tần số (n_i là số cá thể của tổng thể có cùng trị số x_i) thì tổng thể có thể mô tả:

Trị số	x_1	x_2	x_3	...	x_i	...	x_n
Tần số	n_1	n_2	n_3	...	n_i	...	n_n

Hiển nhiên

$$\begin{cases} 0 \le n_i \le N, \text{ với } \forall i \\ \sum_{i=1}^{k} n_i = N \end{cases}$$

- ***Bằng liệt kê bảng phân bố tần suất***

Nếu ký hiệu p_i $(i=\overline{1,k})$ là tần suất của x_i, $p_i = \dfrac{n_i}{N}$ $(i=\overline{1,k})$ thì tổng thể có thể mô tả:

Trị số	x_1	x_2	x_3	...	x_i	...	x_n
Tần suất	p_1	p_2	p_3	...	p_i	...	p_n

với:

$$\begin{cases} 0 \le p_i \le 1, \text{ với } \forall i \\ \sum_{i=1}^{k} p_i = 1 \end{cases}$$

- ***Bằng bảng ghép***

Trị số	x_1	x_2	x_3	...	x_i	...	x_n
Tần số	n_1	n_2	n_3	...	n_i	...	n_n
Tần suất	p_1	p_2	p_3	...	p_i	...	p_n

Đây là những phương pháp mô tả các dấu hiệu lấy các trị số rời rạc.

- ***Bằng tần suất tích lũy***

Nếu w_i $(i = \overline{1,k})$ là tần số tích lũy của các $x_j < x_i$ thì:

$$w_i = \sum_{x_j < x_i} N_j$$

và $f(x_i)$ là tần suất tích lũy của các $x_j < x_i$ thì:

$$f(x_i) = \frac{w_i}{N} = \sum_{x_j < x_i} \frac{N_j}{N}$$

Tần suất tích lũy là một hàm của x_i có tính chất giống như hàm phân phối xác suất của đại lượng ngẫu nhiên rời rạc.

- ***Bằng đồ họa***

Để mô tả tổng thể, từ kết quả điều tra mẫu người ta xây dựng các loại đồ thị, các loại biểu đồ thực nghiệm và tổng thể.

Như vậy, việc mô tả tổng thể bằng bảng phân bố tần số, bảng phân bố tần suất, tần suất tích lũy hay đồ họa cho thấy những dấu hiệu định lượng hoàn toàn có thể mô hình hóa bằng một đại lượng ngẫu nhiên rời rạc. Điều đó cũng đúng cho các tổng thể có dấu hiệu phân phối liên tục.

1.1.2. Mẫu

1.1.2.1. Khái niệm

Mẫu là một bộ phận hữu hạn của tổng thể gồm n cá thể (n < N) được gọi là dung lượng mẫu, trên đó người ta tiến hành điều tra, khảo sát, đo đếm và thu thập các số liệu.

Từ các số liệu thu thập được, người ta sử dụng các thuật toán theo lý thuyết xác suất để suy đoán những hiện tượng, quy luật của tổng thể. Nội dung chính của sự suy đoán này là:

- Ước lượng các tham số của tổng thể thông qua các tham số của mẫu và kiểm định độ tin cậy của các tham số.

- Tìm hiểu mối quan hệ giữa các dấu hiệu nghiên cứu trong tổng thể thông qua mối quan hệ giữa các dấu hiệu trong mẫu và kiểm định độ tin cậy về mối quan hệ.

1.1.2.2. Các phương pháp chọn mẫu

Để việc suy đoán có độ chính xác cao, các mẫu được rút ra để nghiên cứu phải đại diện được cho toàn bộ các cá thể trong tổng thể.

- ***Với tổng thể thuần nhất***

Với loại tổng thể này, áp dụng các phương pháp rút mẫu sau đây.

Rút ngẫu nhiên trực tiếp từ tổng thể

Đây là cách chọn mẫu một cách ngẫu nhiên có hoàn lại và không hoàn lại. Thông thường, có 4 phương pháp chọn ngẫu nhiên:

- Rút mẫu ngẫu nhiên đơn giản: Mỗi cá thể trong tổng thể đều có cơ hội như nhau trong lựa chọn. Các cá thể được quy định trước theo một thứ tự nào đó (có thể đánh số trực tiếp hay quy ước), sau đó tiến hành bốc thăm.

- Rút ngẫu nhiên hệ thống: Quy định lấy mẫu ở các vị trí nào đó được định trước. Đây cũng coi như là phép lấy mẫu ngẫu nhiên, bởi vì cá thể được chọn đứng ở vị trí đó là ngẫu nhiên, trước khi lấy mẫu điều tra, ta cũng không hề biết tình trạng của cá thể này. Người ta có thể định vị trí lấy mẫu trên đường chéo góc, trên đường dích dắc hay các kiểu quy định nào đó. Ví dụ: trong quy phạm khảo nghiệm giống ngô, người ta quy định theo dõi 10 cây/1 giống ở mỗi lần nhắc lại, lấy 5 cây liên tiếp nhau từ cây thứ 5 đến cây thứ 9 tính từ đầu hàng thứ 2 và từ cây thứ 5 đến cây thứ 9 tính từ cuối hàng thứ 3 của ô.

- Dùng bảng số ngẫu nhiên: Có thể sử dụng các bảng số ngẫu nhiên sau để chọn mẫu: Bảng Tippett (các số có 4 chữ số), bảng Fisher và Yates, các bảng của Kendall và Babington Smith (các số có 4 chữ số), bảng của Burke Haton.

- Dùng phần mềm Excel (theo cú pháp ghi ở chương 8).

Chọn cá thể điển hình trực tiếp từ tổng thể

Đây là phương pháp chọn mẫu không ngẫu nhiên. Từ quan sát tổng thể, chọn các cá thể điển hình, đại biểu cho tổng thể theo mục tiêu nghiên cứu.

Rút từ các phần của tổng thể (chia nhóm rồi chọn mẫu)

Người ta chia tổng thể thành các nhóm một cách cơ giới theo một quy tắc nào đó, từ mỗi nhóm lấy ra một số cá thể theo một cách thống nhất để nghiên cứu.

- ***Với tổng thể không thuần nhất***

Có những tổng thể không có từng cá thể điển hình mà chỉ có tập hợp mẫu điển hình. Ví dụ, tổng thể là quần thể phân ly được tạo ra từ phép lai hay tác nhân đột biến hoặc là quần thể tạo được từ kỹ thuật di truyền. Để nghiên cứu chúng ta không thể áp dụng phương pháp chọn từng cá thể điển hình. Tốt nhất là theo dõi toàn thể quần thể hoặc lấy một bộ phận liên tục có dung lượng mẫu lớn (nếu quần thể quá lớn), hoặc sử dung một trong 4 phương pháp chọn ngẫu nhiên đã trình bày trong mục 1.2.1 trên đây.

1.2. CÁC THAM SỐ ĐẶC TRƯNG CỦA MẪU VÀ TỔNG THỂ

1.2.1. Các tham số đặc trưng cho sự tập trung

1.2.1.1. Số cực trị:

Số cực trị là số bé nhất và lớn nhất trong mẫu, ký hiệu là X_{min} và X_{max}.

1.2.1.2. Mốt

Mốt là trị số có tần số cao nhất trong một mẫu. Nếu mẫu đã phân tổ thì tổ mốt là tổ có tần số cao nhất và trị số giữa của tổ mốt là trị số mốt của mẫu.

Trong một tổng thể quan sát nhiều mẫu, mỗi mẫu gồm một số cá thể xác định, khi theo dõi một chỉ tiêu nào đấy ta nhận được trị số mốt của các mẫu xấp xỉ bằng nhau thì tổng thể đó đồng nhất theo chỉ tiêu này, ngược lại nếu trị các trị số mốt của các mẫu khác nhau thì tổng thể đó không đồng nhất. Nếu các chỉ tiêu khác cũng cho kết quả tương tự, ta có thể đánh giá được tính đồng nhất hay không đồng nhất của tổng thể. Người ta thường áp dụng tính chất này để đánh giá độ thuần của giống và mức độ đồng đều của đất.

1.2.1.3. Trung bình và kỳ vọng

Trung bình (trung bình mẫu hay trung bình thực nghiệm), thường ký hiệu là $\overline{X}$, là tham số đặc trưng cho sự tập trung của mẫu và kỳ vọng (trung bình tổng thể hay trung bình lý luận), thường ký hiệu là E(X), MX, μ hay m, là tham số đặc trưng cho sự tập trung của tổng thể.

Bản chất của trị trung bình các giá trị quan sát là gần bằng kỳ vọng, nó phản ánh giá trị trung tâm của phân phối xác suất của đại lượng ngẫu nhiên. Vì vậy, người ta thường sử dụng trị trung bình của mẫu để ước lượng kỳ vọng của tổng thể.

$$E(X) = \mu$$

Khi dung lượng càng lớn, trị trung bình càng gần với kỳ vọng, vì vậy để ước lượng đúng kỳ vọng, dung lượng mẫu phải đủ lớn.

Trong thực nghiệm, khi x_i lấy các trị số rời rạc, $\overline{X}$ được tính theo các công thức sau:

$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Ví dụ, nếu có các số đo x_i là:

20	24	24	23	25	14	21	20	31	16
18	21	19	20	19	13	20	24	18	20

thì $\overline{X}$ = (20 + 24 + 24 + 23 + ... + 18 + 20) : 20 = 20,5

Nếu x_i lấy n_i lần với $n = \sum n_i$ thì $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i n_i$

Nếu xác suất bắt gặp của x_i là p_i ($p_i = n_i / n$) và k là số nhóm x_i thì $\overline{X} = \sum_{i=1}^{k} x_i p_i$. Ví dụ, nếu các số đo x_i có n_i lần bắt gặp với xác suất p_i như sau:

x_i:	17	18	19	20	21	22	23	24	25	26
n_i:	2	5	8	10	20	16	15	14	7	3
p_i :	0,02	0,05	0,08	0,10	0,20	0,16	0,15	0,14	0,07	0,03

thì $\overline{X}$ = $\frac{1}{100}$ [(17 × 2) + (18 × 5) + ... + (26 × 0,3)] = 21,82

hoặc $\overline{X}$ = (17 × 0,02) + (18 × 0,05) + ... + (26 × 0,03) =

21,82 Khi biết các x_i và n_i thì tính theo công thức $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i n_i$, còn khi chỉ biết x_i và p_i thì tính theo công thức $\overline{X} = \sum_{i=1}^{k} x_i p_i$.

Với X là đại lượng ngẫu nhiên liên tục:

$$E(X) = \mu = \int_{-\infty}^{\infty} xf(x)dx$$

Các tính chất của kỳ vọng:

1. Kỳ vọng của một hằng số C bằng chính hằng số đó:

$$E(C) = C$$

2. Kỳ vọng của tích giữa một hằng số và một đại lượng ngẫu nhiên bằng tích của hằng số với kỳ vọng của đại lượng ngẫu nhiên đó:

$$E(CX) = CE(X)$$

3. Kỳ vọng của tổng một hằng số C với một đại lượng ngẫu nhiên bằng tổng của hằng số với kỳ vọng của đại lượng ngẫu nhiên đó:

$$E(X + C) = E(X) + C$$

4. Kỳ vọng của tổng các đại lượng ngẫu nhiên bằng tổng các kỳ vọng thành phần:

$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

5. Kỳ vọng của tích hai đại lượng ngẫu nhiên độc lập bằng tích của hai kỳ vọng của hai đại lượng ngẫu nhiên đó:

$$E(X_1.X_2) = E(X_1).\ E(X_2)$$

Tất cả các tính chất này đều đúng cho số trung bình thực nghiệm.

1.2.2. Các tham số đặc trưng cho độ phân tán của các dấu hiệu định lượng

1.2.2.1. Khoảng biến thiên

Khoảng biến thiên là khoảng cách giữa hai cực trị:

$$R = X_{max} - X_{min}$$

1.2.2.2. Phương sai mẫu, phương sai tổng thể và độ lệch chuẩn

- ***Phương sai mẫu và phương sai tổng thể***

Trung bình và kỳ vọng chỉ là một số bình quân của đại lượng ngẫu nhiên của mẫu và tổng thể. Do khoảng biến thiên R chỉ đo khoảng cách từ hai trị số lớn nhất và nhỏ nhất, chưa xét đến các giá trị khác, vì vậy khoảng biến thiên không đặc trưng cho độ phân tán của mẫu hay tổng thể xung quanh trị bình quân. Hãy xét hai mẫu sau đây:

Mẫu 1:

20	24	24	23	25	14	21	20	31	16
18	21	19	20	19	13	20	24	18	20

Mẫu 2:

26	25	29	14	23	13	14	22	28	24
15	31	14	13	29	16	28	14	17	15

Hai mẫu này cùng có trị trung bình và khoảng biến thiên bằng nhau ($\overline{X}$ = 20,5; R = 18) nhưng không thể nói hai mẫu giống nhau do độ đồng đều của hai mẫu khác nhau rõ ràng, tức là độ phân tán của các số đo so với trị trung bình của từng mẫu khác nhau. Vậy tham số nào đặc trưng cho độ phân tán của các số trong mẫu xung quanh trị trung bình của chúng.

Nếu $(X - \overline{X})$ là độ lệch của mỗi số X với số trung bình $\overline{X}$, theo tính chất 3 và 1 của kỳ vọng, ta có:

$$E[X - E(X)] = E(X) - E[E(X)]$$

$$= E(X) - E(X) = 0$$

tức là: trung bình độ lệch từ mỗi giá trị X với trung bình mẫu luôn bằng không. Nói cách khác: do tổng đại số các độ lệch từ mỗi giá trị của mẫu với trung bình mẫu luôn bằng 0 nên trung bình độ lệch cũng luôn bằng 0. Vì vậy trung bình độ lệch không phản ánh độ phân tán.

Người ta sử dụng tổng bình phương độ lệch và trung bình bình phương để nghiên cứu độ phân tán.

Tổng bình phương độ lệch $\sum_{i=1}^{n}[X - M(X)]^2 = 0$ khi mọi X đều bằng nhau và $\sum_{i=1}^{n}[X - M(X)]^2$ càng tăng khi các giá trị X càng khác nhau.

Trung bình bình phương thực nghiệm, còn gọi là phương sai mẫu hay phương sai, ký hiệu là MS (Mean Square), S^2, s^2 hay V(X) hoặc Var(X), là tham số đặc trưng cho độ phân tán của các cá thể trong mẫu theo dấu hiệu nghiên cứu và trung bình bình phương lý luận, còn gọi phương sai tổng thể, thường ký hiệu là V(X) hoặc Var(X), DX, σ_X^2 hay σ^2 (nói chung), là tham số đặc trưng cho độ phân tán của các cá thể trong tổng thể.

Bản chất của phương sai mẫu là trung bình số học của bình phương các độ lệch giữa các giá trị của đại lượng ngẫu nhiên so với trị trung bình, phản ánh mức độ phân tán của các giá trị quan sát của đại lượng ngẫu nhiên xung quanh giá trị trung bình của chúng. Nếu trị trung bình mẫu dùng để ước lượng kỳ vọng của tổng thể thì phương sai mẫu dùng để ước lượng phương sai tổng thể. Khi dung lượng mẫu càng lớn, phương sai mẫu càng gần với phương sai tổng thể, vì vậy để ước lượng đúng phương sai tổng thể, dung lượng mẫu phải đủ lớn.

$$V(X) = E[X - E(X)]^2 = \sigma_X^2$$

Phương sai có đơn vị đo là bình phương đơn vị đo của đại lượng ngẫu nhiên.

Trong thực nghiệm, khi x_i lấy các giá trị rời rạc, V(X) được tính theo các công thức sau:

$$V(X) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right).$$

Ở mẫu 1:

$$V(X) = \frac{1}{20-1}[(20 - 20{,}5)^2 + (24 - 20{,}5)^2 +$$

$$... + (20 - 20{,}5)^2] = 16{,}37$$

Tương tự, phương sai ở mẫu 2 là: V(x) = 42,789

Khi x_i lấy n_i lần (như ví dụ sau trong mục 1.2.1.3), công thức tính phương sai có dạng:

$$V(x) = \frac{1}{n-1}\left[n\sum_{i=1}^{k} n_i x_i^2 - (\sum_{k=1}^{k} n_i x_i)^2 \right]$$

$$= \frac{1}{n(n-1)}\left[n\sum_{i=1}^{k} x_i^2 p_i - (\sum_{k=1}^{k} x_i p_i)^2 \right] \text{với } n = \sum n_i .$$

Kết quả tính được: V(x) = 4,452.

Với X là đại lượng ngẫu nhiên liên tục:

$$V(X) = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$$

Các tính chất của phương sai:

1. Phương sai của một hằng số C thì bằng 0:

$$V(C) = 0$$

Thật vậy: $V(C) = E[C - E(C)]^2 = E[C - C]^2 = E(0) = 0$

2. Phương sai của tích một hằng số và một đại lượng ngẫu nhiên bằng tích giữa bình phương hằng số và phương sai của đại lượng ngẫu nhiên đó:

$$V(CX) = C^2V(X)$$

Thật vậy:

$$V(CX) = E[CX - E(CX)]^2 = E[CX - CE(X)]^2$$

$$= E\{C^2[X - E(X)]^2\} = C^2E[X - E(X)]^2$$

$$= C^2V(X)$$

3. Phương sai của tổng một hằng số C với một đại lượng ngẫu nhiên thì bằng chính phương sai của đại lượng ngẫu nhiên đó. Nói cách khác nếu cộng một hằng số C với một đại lượng ngẫu nhiên thì phương sai không đổi:

$$V(X + C) = V(X)$$

Thật vậy:

$$V(X + C) = E[(X + C) - E(X + C)]^2$$
$$= E[(X + C) - E(X) - C]^2$$
$$= E[X - E(X)]^2 = V(X)$$

4. Phương sai của một tổng hai đại lượng ngẫu nhiên độc lập bằng tổng các phương sai thành phần:

$$V(X_1 + X_2) = V(X_1) + V(X_2)$$

Thật vậy:

$$V(X_1 + X_2) = E[(X_1 + X_2) - E(X_1 + X_2)]^2$$
$$= E\{[X_1 - E(X_1)] - [X_2 - E(X_2)]\}^2$$
$$= E\{[X_1 - E(X_1)]^2 + 2[X_1 - E(X_1)].[X_2 - E(X_2)]$$
$$+ [X_2 - E(X_2)]^2\}$$
$$= E[X_1 - E(X_1)]^2 + 2E\{[X_1 - E(X_1)].[X_2 - E(X_2)]\}$$
$$+ E[X_2 - E(X_2)]^2$$
$$= V(X_1) + V(X_2) + 2E[X_1X_2 - X_1E(X_1)$$
$$- X_2E(X_1) + E(X_1)E(X_2)]$$
$$= V(X_1) + V(X_2) + 2[E(X_1X_2) - E(X_1)E(X_2)$$
$$- E(X_2)E(X_1)] + E(X_1)E(X_2)]$$
$$= V(X_1) + V(X_2) + 2[E(X_1)E(X_2)$$
$$- E(X_1)E(X_2)] = V(X_1) + V(X_2)$$

Hệ quả:

1. $$V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i)$$

2. $$V(X_1 - X_2) = V(X_1) + V(X_2)$$

Độ lớn nhỏ của phương sai không phụ thuộc vào độ lớn của số trung bình. Các tập hợp mẫu có cùng trị trung bình nhưng phương sai có thể khác nhau (trong so sánh mẫu 1 và mẫu 2 trên đây) và các số trung bình có thể khác nhau nhưng phương sai có thể bằng nhau (tính chất 3 của phương sai).

Như vậy, phương sai đặc trưng cho độ phân tán (hay là độ khác biệt giữa các số). Khi các số càng gần bằng nhau thì phương sai càng nhỏ, mẫu càng đồng đều, ngược lại, khi các số càng khác xa nhau thì phương sai càng lớn, mẫu càng kém đồng đều.

Phương sai cũng như các tính chất của nó được sử dụng như là phương pháp hữu hiệu trong nhiều phép phân tích, đánh giá các số liệu thu thập (sẽ được đề cập ở các phần sau).

Như trên đã nói, μ và σ^2 là kỳ vọng và phương sai tổng thể, người ta đã chứng minh được (không dẫn):

- Kỳ vọng của trung bình mẫu bằng trung bình tổng thể:

$$E(\overline{X}) = \mu$$

- Kỳ vọng phương sai mẫu bằng phương sai tổng thể:

$$E[V(X)] = E(S^2) = \sigma^2$$

Do đó người ta lấy $\overline{X}$ để ước lượng μ và lấy S^2 để ước lượng σ^2 ở mức tin cậy nào đó.

- ***Độ lệch chuẩn***

Độ lệch chuẩn mẫu, ký hiệu là S hay sd (standard deviation) là căn bậc hai của phương sai ($S = \sqrt{S^2}$), còn độ lệch chuẩn tổng thể, ký hiệu là σ_x hay σ (nói chung) là căn bậc hai của phương sai tổng thể ($\sigma = \sqrt{\sigma^2}$).

Đơn vị tính của độ lệch chuẩn là đơn vị đo của đại lượng ngẫu nhiên.

Độ lệch chuẩn phản ánh mức sai lệch trung bình của các cá thể xung quanh trị trung bình. Mẫu và tổng thể càng đồng đều, S và σ càng bé và ngược lại.

1.2.2.3. Hệ số biến động

Do độ lệch chuẩn là một số tuyệt đối không phụ thuộc vào số trung bình nên không phản ánh mức độ biến động xung quanh trị trung bình. Hai mẫu có cùng độ lệch chuẩn không thể coi chúng biến động như nhau khi chúng có hai trị trung bình khác nhau. Người ta dùng hệ số động (ký hiệu là CV – Coefficient of Variation) để đánh giá mức sai lệch lớn hay nhỏ so với trung bình của nó và được tính bằng %:

$$CV(\%) = \frac{S}{\bar{\bar{X}}}.100$$

Hệ số biến động được sử dụng trong các trường hợp sau:

- Đánh giá độ biến động của các cá thể trong mẫu và tổng thể theo một chỉ tiêu nào đó, ví dụ chiều cao cây, chiều dài của các bộ phận, khối lượng hạt, củ, quả, số lượng hạt, củ, quả. Để đánh giá độ đồng đều của hạt của một giống, sau khi lấy mẫu phân tích, người ta đếm và cân ít nhất 8 mẫu, mỗi mẫu 100 hạt hay 1.000 hạt (tùy hạt lớn hay nhỏ), rồi tính CV(%). Nếu hạt đồng đều, khối lượng các mẫu ít khác biệt nhau và độ biến động thấp. Nếu CV(%) $\leq$ 5 là biến động ít – hạt đồng đều, 6 – 10 là biến động vừa phải – hạt tương đối đều và > 10 là biến động nhiều và rất nhiều – hạt không đều.

- Đánh giá sự khác nhau giữa các nhóm cá thể (quần thể) như: giữa các giống, giữa các nghiệm thức theo đặc trưng nào đó. Giá trị hệ số biến động càng cao chứng tỏ chúng càng khác biệt nhau.

- Chọn ruộng (đất) thí nghiệm. Khi chưa biết được lịch sử canh tác của khu đất, có thể chọn đất thí nghiệm bằng cách lấy mẫu đất, phân tích và đánh giá nhanh sự đồng nhất bằng phương pháp phi tham số hoặc phương pháp so sánh phương sai một số chỉ tiêu chính giữa các lô lấy mẫu trong khu đất. Tuy nhiên có thể chọn đất bằng cách thực hiện một “thí nghiệm trắng”. Gọi là “thí nghiệm trắng” vì thí nghiệm không nghiên cứu điều gì ngoài việc chọn đất. Trong thí nghiệm này, người ta sử dụng chỉ dùng một giống để gieo lên các ô đã được thiết kế theo kiểu CRD hoặc RCBD cho các “nghiệm thức” giả định, thu năng suất từng ô như một thí nghiệm thông thường, hoặc là lấy mẫu năng suất từ ruộng đã được gieo trồng sẵn một giống nào đó. Nếu kết quả phân tích số liệu cho thấy giữa các “nghiệm thức” không khác biệt nhau và hệ số biến động của sai số $\leq$ 10% (càng nhỏ càng tốt) thì đất này sẽ được chọn làm thí nghiệm.

- Đánh giá độ chính xác (ít sai số) của một thí nghiệm. Trong bảng phân tích phương sai (ANOVA), CV phản ánh độ biến động do sai số gây ra:

$$CV(\%) = \frac{\sqrt{MS_e}}{\bar{X}}.100$$

Do $\sqrt{MS_e}$ là độ lệch chuẩn của sai số nên CV(%) càng nhỏ thí nghiệm càng chính xác.

1.2.2.4. Hệ số nhọn của phân phối xác suất

Hệ số nhọn của phân phối xác suất, ký hiệu là α_4, cho thấy các giá trị x_i của đại lượng biến thiên tập trung nhiều hay ít xung quanh kỳ vọng, tương ứng với phương sai nhỏ hay lớn.

$$\alpha_4 = \frac{\mu^4}{\sigma^4}$$

trong đó: μ^4 là mô men trung tâm bậc 4: $\mu^4 = E[X - E(X)]^4$
σ^4 là bình phương của phương sai

Nếu $\mu^4 = 3$ thì đồ thị phân phối xác suất là bình thường, nếu $\mu^4 > 3$ thì đồ thị nhọn (các x_i tập trung nhiều xung quanh kỳ vọng μ), còn nếu $\mu^4 < 3$ thì đồ thị tù (không nhọn).

Với ví dụ ở mục 2.2.2 trên đây:

mẫu 1: $\mu^4 = 973{,}80$, $\sigma^4 = 267{,}91$ và $\mu^4 = 3{,}6$

mẫu 2: $\mu^4 = 2.436{,}91$, $\sigma^4 = 1.830{,}90$ và $\mu^4 = 1{,}3$

Rõ ràng mẫu 1 số liệu rất tập trung còn mẫu 2 số khá rải rác mặc dù trung bình mẫu đều bằng nhau.

1.2.3. Các tham số đặc trưng cho độ phân tán của các dấu hiệu định tính

Các dấu hiệu định tính, còn gọi là các dấu hiệu chất lượng, thường được biểu thị dưới dạng tần suất. Với loại dấu hiệu này, để biết được độ phân tán người ta đánh giá độ lệch chuẩn của các tần suất phân phối ở các mức chất lượng khác nhau và hệ số biến động biểu thị mức độ sai khác (%) giữa độ lệch chuẩn và độ lệch chuẩn cao nhất.

Hãy xét ví dụ ở Bảng 1.1 sau đây.

Theo mức độ lông của lá, Bảng 1.1 cho thấy có 7,1% số giống không hay rất ít lông, 14,3% - ít lông, 21,4% -

lông vừa, 28,6% - lông nhiều, 28,6% - lông rất nhiều. trong 20 giống kháng rầy có 5% thuộc loại ít lông, 20% thuộc loại lông vừa, 35% thuộc loại nhiều lông, còn 40% thuộc loại rất nhiều lông.

Để đánh giá độ lệch chuẩn biểu thị mức độ khác nhau về các tần suất theo độ lông của các giống ta dùng công thức:

Bảng 1.1: Kết quả điều tra mức độ rầy xanh hại trên 28 giống bông tại Đại học Nông Lâm Tp. HCM, 2009

Mức độ lông của lá	Giống có lông		Giống kháng	
	Số giống (n_i)	Tần suất (p_i)	Số giống kháng (n_i)	Tần suất (p_i)
Không hay rất ít	2	0,071	0	0,000
Ít	4	0,143	1	0,050
Vừa	6	0,214	3	0,150
Nhiều	8	0,286	7	0,350
Rất nhiêu	8	0,286	8	0,400
Tổng	28	1,000	20	1,000

$$S_p = \sqrt[k]{p_1.p_2....p_k} \qquad (1-1)$$

trong đó, S_p là độ lệch chuẩn, p_1, p_2, ..., p_k là tần suất các nhóm chất lượng ($\Sigma p_i = 1$) và k là số nhóm.

Theo Bảng 1.1, về cơ cấu giống có lông:

$$S_p = \sqrt[5]{0,071\times 0,143\times 0,214\times 0,286\times 0,286}$$

= 0.178 (hay 17,8%)

và hệ số biến động được tính theo công thức:

$$CV(\%) = \frac{S_p}{S_{p\max}}.100 \qquad (1 - 2)$$

Trong ví dụ này là: $CV(\%) = \frac{0,178}{0,200}.100$ = 89,0, tức là biến động rất nhiều.

Cách xác định giá trị $S_{p\ max}$ ở công thức (1 - 2) khá dễ dàng vì: theo công thức (1 - 1), S_p lấy giá trị cao nhất khi các các tần suất p_i của các nhóm bằng nhau. Do $\Sigma p_i = 1$ nên $S_{p\ max} = 1/k$, vì vậy ở ví dụ này $1/5 = S_{p\ max} = 0,20$.

Hãy tính S_p cho cơ cấu tính kháng.

Với số liệu Bảng 1.1, ta chỉ có 4 nhóm chất lượng: ít, vừa, nhiều và rất nhiều, tần suất của 4 nhóm này theo thứ tự là 0,050; 0,150; 0,350 và 0,400.

Một cách tương tự ta có: S_p = 0,180 và CV(%) = 72,0.

So sánh với chỉ tiêu cơ cấu giống có lông, độ lệch chuẩn của cơ cấu tính kháng lớn hơn nhưng hệ số biến động thấp hơn vì nó có ít nhóm hơn nên độ lệch chuẩn tối đa lớn hơn (0,25 so 0,20), thành thử CV(%) nhỏ hơn.

Về việc xác định nhóm, ở cột đầu có 5 nhóm theo độ lông khác nhau, chỉ tiêu cơ cấu giống có lông có tần số và tần suất theo 5 nhóm này, nhưng với chỉ tiêu cơ cấu giống kháng tất cả các giống kháng (20 giống) chỉ nằm trong 4 nhóm với $\Sigma p_i = 1$. Do nhóm không và rất ít lông có $n_i = 0$ nên với chỉ tiêu này chỉ xét cho 4 nhóm.

Trong trường hợp chỉ có 2 nhóm (k = 2), khi đó $p_1 + p_2 = 1$, tức là $p_1 = 1 - p_2$, $0 < S_p \leq 0,5$ và $S_{p\ max} = 0,5$.

Việc tính toán trên đây khá dễ dàng nhờ sự trợ giúp của phần mềm Excel trên máy điện toán.

Một cách tính toán khác có thể được áp dụng là biến công thức (1– 1) thành: $\log S_p = \frac{1}{k}(\log p_1 + \log p_2 + ... + \log p_k)$, sau khi tính được $\log S_p$ ta sẽ có được S_p.

Cách tính như sau, ví dụ cho chỉ tiêu giống có lông:

$$\log S_p = \frac{1}{5}(\log 0,071 + \log 0,143 + \log 0,214 + \log 0,286 + \log 0,286)$$

$$= \frac{1}{5}(-1,14874 - 0,84466 - 0,66958 - 0,54363 - 0,54363)$$

$= - 0,75005$. Từ đó: $S_p = 0,178$ (hay 17,8%)

Như vậy, với các đặc trưng định tính có thể đánh giá được độ phân tán của các tần suất các nhóm chất lượng qua độ lệch chuẩn và và hệ số biến động khi so sánh với độ lệch chuẩn cao nhất.

1.2.4. Các tham số đặc trưng cho mối quan hệ giữa các đại lượng ngẫu nhiên

1.2.4.1. Hiệp phương sai

Hiệp phương sai, thường ký hiệu là Cov(X,Y), Covar(X,Y) hoặc W(X,Y), là kỳ vọng của tích các độ lệch của các đại lượng ngẫu nhiên với kỳ vọng (hay trung bình thực nghiệm) của chúng, biểu thị mức độ quan hệ giữa hai đại lượng ngẫu nhiên và được tính theo công thức:

$$W(X,Y) = E\{[(X - E(X)][(Y - E(Y)]\}$$

Hiệp phương sai có đơn vị đo là tích đơn vị đo của các đại lượng ngẫu nhiên X và Y.

Trong thực nghiệm, công thức tính hiệp phương sai giữa biến X và Y được viết:

$$W_{x,y} = \frac{1}{n-1}\sum\left(X - \bar{X}\right)\left(Y - \bar{Y}\right)$$

hay: $$\operatorname{Cov}\left(X, Y\right) = \frac{1}{n-1}\left[\sum XY - \left(\sum X\right)\left(\sum Y\right)/n\right]$$

Các tính chất của hiệp phương sai:

1. Hiệp phương sai của hai đại lượng ngẫu nhiên lấy các giá trị là hằng số thì bằng 0:

$$W(C_1, C_2) = 0$$

Ví dụ:

c_1:	12	12	12	12	12	12	12	12	12	12
c_2:	15	15	15	15	15	15	15	15	15	15

2. Nếu nhân mỗi đại lượng ngẫu nhiên với mỗi hằng số khác nhau, như C_1 và C_2 thì:

$$W(C_1X, C_2Y) = C_1C_2W(X,Y)$$

3. Nếu cộng mỗi đại lượng ngẫu nhiên với mỗi hằng số khác nhau, như C_1 và C_2 thì hiệp phương sai không đổi:

$$W(X + C_1, Y + C_2) = W(X,Y)$$

4. Hiệp phương sai của hai đại lượng ngẫu nhiên độc lập, như X độc lập với Y, thì bằng 0:

$$W(X,Y) = 0 \qquad (X \neq Y)$$

5. Nếu hai đại lượng ngẫu nhiên X, Y có quan hệ, thì:

$$V(X + Y) = V(X) + V(Y) - 2W(X,Y)$$

1.2.4.2. Hệ số tương quan tuyến tính và hệ số hồi quy

Mục này sẽ được trình bày trong chương 4.

Chương 2

ƯỚC LƯỢNG CÁC THAM SỐ

2.1. KHÁI NIỆM

Các tham số thống kê là những thông tin phản ánh bản chất của tổng thể theo một dấu hiệu (chỉ tiêu) nào đó. Thường thì không thể nghiên cứu toàn bộ số cá thể trong tổng thể. Vậy, để tìm hiểu tổng thể ta phải tìm các phương pháp để suy đoán các tham số thống kê của tổng thể.

Phương pháp tiếp cận thường dùng, như trên đã nói, là phương pháp rút mẫu và từ kết quả nghiên cứu mẫu để suy đoán cho tổng thể bằng phép quy nạp thống kê gọi là ước lượng. Kết quả ước lượng là xác định một cách gần đúng giá trị của các tham số thống kê tổng thể ở độ tin cậy nào đó. Có hai phương pháp sử dụng tham số mẫu để ước lượng cho tham số tổng thể là phương pháp ước lượng điểm và phương pháp ước lượng khoảng.

Ước lượng điểm: là phương pháp dùng trị số của hàm ước lượng được tính toán ở mẫu để thay một cách gần đúng cho tham số tổng thể.

Công thức tổng quát của phương pháp ước lượng điểm như sau:

$$\theta = T_n$$

trong đó: - θ là tham số tổng thể cần ước lượng;

- T_n là hàm ước lượng của tham số θ.

Để ước lượng đúng nhất, phải chọn được hàm ước lượng tốt nhất. Muốn vậy, hàm ước lượng này phải thỏa mãn: không chệch, hội tụ và hiệu nghiệm.

- Ước lượng T_n gọi là ước lượng không chệch cho θ nếu $E(T_n) = \theta$.

Ước lượng không chệch cho biết hàm ước lượng T_n không có sai số hệ thống.

- Ước lượng T_n gọi là ước lượng vững cho θ nếu với mọi $\varepsilon > 0$

$$\lim_{n\to\infty} P\{|T_n - \theta| < \varepsilon\} = 1$$

hay
$$\lim_{n\to\infty} P\{\theta - \varepsilon < T_n < \theta + \varepsilon\} = 1$$

Ước lượng vững có xác xuất cao khi dung lượng mẫu đủ lớn.

- Ước lượng T_n gọi là ước lượng hiệu quả cho θ nếu T_n là ước lượng không chệch và có phương sai nhỏ nhất so với mọi ước lượng không chệch khác cho θ.

Ước lượng khoảng: là phương pháp mà tham số ước lượng của tổng thể nằm trong một khoảng với một xác suất (hay độ tin cậy) cho trước. Khoảng này xác định được nhờ những kết quả khi nghiên cứu ở mẫu.

Công thức tổng quát của phương pháp ước lượng khoảng như sau:

$$P(G_1 \le \theta \le G_2) = 1 - \alpha$$

trong đó:

- P là xác suất của sự ước lượng cho tham số θ của tổng thể;
- G_1 và G_2 là giới hạn dưới và giới hạn trên của

khoảng ước lượng được xác định từ kết quả quan sát ở mẫu;

- 1 – α là mức tin cậy của ước lượng, α thường chọn là 0,05; 0,01 hay 0,001 (mức sai lầm).

Hiệu số $G_2 - G_1$ được gọi là độ dài khoảng ước lượng. Độ dài khoảng ước lượng càng nhỏ thì độ chính xác của ước lượng càng cao và ngược lại.

Nếu ký hiệu $\varepsilon = \dfrac{G_2 - G_1}{2}$ thì khoảng tin cậy sẽ được viết ($\theta - \varepsilon$; ($\theta + \varepsilon$). ε gọi là sai số tới hạn của ước lượng còn được gọi là độ chính xác của ước lượng và $\varepsilon\% = \dfrac{\varepsilon}{\overline{X}}.100$ gọi là sai số tương đối hay độ chính xác của ước lượng.

Người ta chia phương pháp ước lượng khoảng ra hai trường hợp:

- Ước lượng khoảng một phía (một chiều - one-tail):

Tham số θ của phân phối lý thuyết được nằm trong một khoảng:

$$P(-\infty < \theta < G_2) = 1 - \alpha \text{ (nằm phải)}$$

hay

$$P(G_1 < \theta < +\infty) = 1 - \alpha \text{ (nằm trái)}$$

- Ước lượng khoảng hai phía (hai chiều - two-tail):

$$P(G_1 \le \theta \le G_2) = 1 - \alpha$$

Đó là khoảng tin cậy cần tìm.

Trong thực tế, người ta thường yêu cầu độ tin cậy 1 – α, chẳng hạn 1 – α = 0,95 nên theo nguyên lý xác suất số lớn, biến cố ($G_1 < \theta < G_2$) hầu như chắc chắn xẩy ra. Khi tiến hành rút mẫu quan sát, giá trị của G_1 và G_2 ứng với mẫu sẽ được viết g_1 và g_2 và $P(g_1 < \theta < g_2) = 1 - \alpha$ hay $P(\overline{X} - \varepsilon \leq \theta \leq \overline{X} + \varepsilon) = 1 - \alpha$.

2.2. ƯỚC LƯỢNG TRUNG BÌNH TỔNG THỂ

2.1. Ước lượng điểm trung bình tổng thể

Giả sử có một tổng thể, để ước lượng trị trung bình tổng thể theo biến X, người ta rút ngẫu nhiên một mẫu độc lập với dung lượng mẫu *n* đủ lớn và quan sát được các số đo $x_1, x_2, x_3, \ldots, x_n$.

Người ta chứng minh được rằng trị trung bình mẫu $\overline{X} = \sum_{i=1}^{n} x_i$ là trị ước lượng hiệu nghiệm nhất đối với trị trung bình tổng thể (kỳ vọng μ):

$$\lim_{n \to \infty} P\left\{\left|\overline{X} - \mu\right| > \varepsilon\right\} = 0$$

2.2. Ước lượng khoảng trung bình tổng thể

Giả sử đại lượng ngẫu nhiên X tuân theo quy luật phân phối chuẩn $N(\mu, \sigma^2)$ nhưng chưa biết tham số trung bình μ. Để ước lượng μ ta xét các trường hợp sau.

2.2.2.1. Khi đã biết phương sai σ^2 của tổng thể

Khi đó việc ước lượng khoảng μ được tiến hành theo luật phân phối chuẩn tắc N(0,1):

$$U = \frac{\left(\overline{X} - \mu\right)\sqrt{n}}{\sigma} \qquad (2 - 1)$$

trong đó $\overline{X}$ là trung bình mẫu. Do U có phân phối chuẩn tắc nên với độ tin cậy $1 - \alpha$ cho trước có thể tìm được cặp giá trị α_1 và α_2 sao cho $\alpha_1 + \alpha_2 = \alpha$. Từ đó tìm được hai giá trị tới hạn tương ứng $u_{1-\alpha_1}$ và u_{α_2} thỏa mãn điều kiện:

$$P(U < u_{1-\alpha_1}) = \alpha_1$$

và $$P(U > u_{\alpha_2}) = \alpha_2$$

Từ đó $P(u_{1-\alpha_1} < U < u_{\alpha_2}) = 1 - (\alpha_1 + \alpha_2) = 1 - \alpha$

Vì $u_{\alpha_1} = -\ u_{1-\alpha_1}$ nên có thể viết

$$P(-u_{\alpha_1} < U < u_{\alpha_2}) = 1 - \alpha \qquad (2 - 2)$$

Thay (2 - 1) vào (2 - 2) và giải ra μ ta có:

$$P(\bar{X} - u_{\alpha_2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + u_{\alpha_1}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

$(\bar{X} - u_{\alpha_2}\frac{\sigma}{\sqrt{n}}; \bar{X} + u_{\alpha_1}\frac{\sigma}{\sqrt{n}})$ là khoảng ước lượng của μ với độ tin cậy $1 - \alpha$. Do có vô số cặp α_1 và α_2 thỏa mãn $\alpha_1 + \alpha_2 = \alpha$ và vì vậy có vô số khoảng ước lượng. Trong thực tế người ta thường sử dụng một số trường hợp sau để ước lượng:

- Ước lượng khoảng đối xứng:

Nếu lấy $\alpha_1 = \alpha_2 = \alpha/2$ thì:

$$P(\bar{X} - u_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + u_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha \quad (2 - 3)$$

$\varepsilon = u_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ được gọi là sai số tới hạn hay độ chính xác của ước lượng và khoảng tin cậy $(\bar{X} - u_{\alpha/2}\frac{\sigma}{\sqrt{n}}; \bar{X} + u_{\alpha/2}\frac{\sigma}{\sqrt{n}})$ có thể viết:

$$I = 2\varepsilon = 2u_{\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad (2 - 4)$$

Công thức (2 - 4) cho thấy:

- Khi tăng dung lượng mẫu lên và giữ nguyên độ tin cậy 1 – α cho trước thì ε sẽ giảm xuống, độ chính xác của ước lượng tăng lên.

- Khi tăng độ tin cậy 1 – α lên và giữ nguyên dung lượng mẫu thì giá trị tới hạn $u_{\alpha/2}$ tăng lên, do đó sai số tới hạn ε cũng tăng lên làm cho độ chính xác của ước lượng giảm đi.

Trong thực, tùy yêu cầu về độ chính xác của cuộc điều tra để xác định dung lượng mẫu phù hợp.

$$\varepsilon(\%) = \frac{\varepsilon}{\overline{x}}.100 = u_{\alpha/2}\frac{\sigma}{\overline{x}\sqrt{n}}.100$$

$= u_{\alpha/2}\dfrac{CV(\%)}{\sqrt{n}}$ là sai số tương đối biểu thị mức độ chính xác của ước lượng nên còn gọi là độ chính xác tương đối.

Dung lượng mẫu cần thiết để đạt được độ chính xác tương đối cho trước $\varepsilon_0(\%)$ là:

$$n_{min} = \left(\frac{u_{\alpha/2}CV(\%)}{\varepsilon_0(\%)}\right)^2$$

Nếu α = 0,10 thì $u_{\alpha/2}$ = 1,645; α = 0,05 thì $u_{\alpha/2}$ = 1,960

α = 0,02 thì $u_{\alpha/2}$ = 2,326; α = 0,01 thì $u_{\alpha/2}$ = 2,576

2.2.2.2. Khi chưa biết phương sai σ^2 của tổng thể nhưng có dung lượng mẫu lớn (n > 30)

Khi dung lượng mẫu đủ lớn thì S ≈ σ nên có thể thay thế S cho σ, khi đó việc ước lượng được tiến hành theo luật phân phối chuẩn theo công thức (2 - 3).

Với độ tin cậy 95%:

$$P(\overline{X} - 1,96\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + 1,96\frac{S}{\sqrt{n}}) = 0,95$$

Có thể phát biểu: ở độ tin cậy 95%, trung bình tổng thể μ nằm trong khoảng $(\overline{X} - 1,96\frac{S}{\sqrt{n}}; \overline{X} + 1,96\frac{S}{\sqrt{n}})$.

$$\varepsilon = 1,96\frac{S}{\sqrt{n}}$$

$$\varepsilon(\%) = \frac{\varepsilon}{\overline{x}}.100 = 1,96\frac{S}{\overline{x}\sqrt{n}}.100 = 1,96\frac{CV(\%)}{\sqrt{n}}$$

$$n_{\min} = \left(\frac{1,96CV(\%)}{\varepsilon_0(\%)}\right)^2$$

Với độ tin cậy 99%:

$$P(\overline{x} - 2,58\frac{S}{\sqrt{n}} \leq \mu \leq \overline{x} + 2,58\frac{S}{\sqrt{n}}) = 0,99$$

và trung bình tổng thể μ nằm trong khoảng:

$$(\overline{x} - 2,58\frac{S}{\sqrt{n}}; \overline{x} + 2,58\frac{S}{\sqrt{n}})$$

$$\varepsilon = 2,58\frac{S}{\sqrt{n}}$$

$$\varepsilon(\%) = \frac{\varepsilon}{\overline{x}}.100 = 2,58\frac{S}{\overline{x}\sqrt{n}}.100 = 2,58\frac{CV(\%)}{\sqrt{n}}$$

và: $$n_{\min} = \left(\frac{2,58CV(\%)}{\varepsilon_0(\%)}\right)^2$$

Nếu mẫu được chọn từ tổng thể hữu hạn theo cách rút mẫu không lặp (không hoàn lại), khi n > 0,1N, để bảo đảm độ chính xác của ước lượng thì sai số tới hạn sẽ nhân thêm hệ số điều chỉnh $\sqrt{\frac{N-n}{N-1}}$, khi đó công thức (2 - 3) trở thành:

$$P(\bar{X} - u_{\alpha/2}\frac{\sigma}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}} \le \mu \le \bar{X} + u_{\alpha/2}\frac{\sigma}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}}) = 1-\alpha \quad (2 - 5)$$

và dựa vào đó để tính toán ε, ε(%), từ đó sẽ xác định được dung lượng mẫu cần thiết để đạt được độ chính xác $\varepsilon_0(\%)$ theo yêu cầu.

Ví dụ: Ước lượng trung bình năng suất cá thể và khoảng tin cậy của tổ hợp bông lai F_1 S02-13/TM1 trồng tại Đại học Nông Lâm Tp. HCM, 2008 theo số liệu bảng sau:

Bảng 2.1: Năng suất cá thể 50 cây (g/cây)

84,4	73,5	84,5	93,5	75,2	74,3	93,5	82,1	68,6	85,4
66,8	57,7	79,4	91,0	129,0	74,6	61,4	91,4	99,5	82,7
95,2	88,3	28,4	77,0	81,2	39,7	86,4	51,5	51,2	80,5
101,0	77,7	90,0	92,9	80,8	67,2	57,1	57,3	34,2	79,5
80,3	88,0	61,1	63,8	101,0	70,2	95,1	97,0	50,3	73,7
Trung bình: 76,92, Phương sai: 351,68, Độ lệch chuẩn: 18,75									

Ở trường hợp này n = N.

Với độ tin cậy 95%:

$$P(76,92 - 1,96\frac{18,75}{\sqrt{50}} \le \mu \le 76,92 + 1,96\frac{18,75}{\sqrt{50}}) = 0,95$$

$$= \text{P}(76,92 - 5,20 \le \mu \le 76,92 + 5,20) = 0,95$$

$$= \text{P}(71,72 \le \mu \le 82,12) = 0,95$$

Tức là: ở độ tin cậy 95% trung bình năng suất cá thể của tổ hợp lai S02-13/TM1 nằm trong khoảng 71,72 g/cây đến 82,12 g/cây (76,92 ± 5,20 g/cây) và với mật độ 41.667 cây/ha thì năng suất dao động trong khoảng 29,9 – 34,2 tạ/ha, trung bình là 32,1 tạ /ha.

Với kết quả điều tra này, sai số tới hạn sai ε = 5,20 g/cây, sai số tương đối ε(%) là 6,76.

Để sai số tương đối $\varepsilon_0(\%)$ cho trước không vượt quá 5% thì số mẫu điều tra tối thiểu phải đạt:

$$n_{\min} = \left(\frac{2\text{CV}(\%)}{\varepsilon_0(\%)}\right)^2 = \left(\frac{2 \times 24,38}{5}\right)^2 = 95 \text{ cây}$$

Bây giờ, ta hãy giả sử rằng N = 95, n = 50 thì số hiểu chỉnh sai số $\sqrt{\frac{\text{N}-n}{\text{N}-1}}$ = 0,692, khi đó sai số tới hạn sẽ là: 5,20 × 0,692 = 3,60. Từ đây, ε(%) = (3,60/76,92) × 100 = 4,7. Như vậy nếu tổ hợp lai được gieo 95 cây thì sai số tương đối không vượt quá 5%.

2.2.2.3. Khi chưa biết phương sai σ^2 của tổng thể và có dung lượng mẫu nhỏ (n < 30)

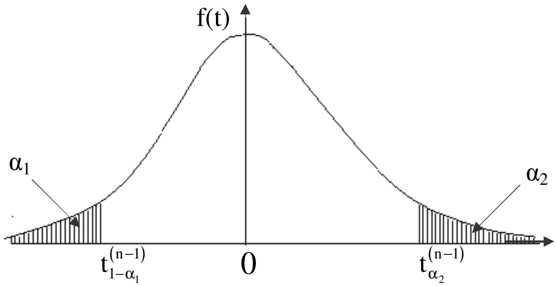
Trong trường hợp mẫu nhỏ việc ước lượng được tiến hành theo luật phân phối Student.

Đại lượng ngẫu nhiên liên tục T phân phối theo luật Student với k bậc tự do, nếu hàm mật độ xác suất của nó được xác định bằng công thức :

$$f(t) = \frac{\Gamma\left(\frac{k}{2}\right)}{\sqrt{\pi(k-1)\Gamma\frac{(k-1)}{2}}}\left(1+\frac{t^2}{n-1}\right)^{-\frac{k}{2}} \text{ với } \forall\, t$$

trong đó Γ(x) là hàm Gamma.

Và, đồ thị hàm mật độ xác suất có dạng như hình 2.1.



Hình 2.1: Đồ thị hàm mật độ xác suất theo luật phân phối Student

Theo luật phân phối Student, với dung lượng mẫu *n*, số bậc tự do là *n* – 1 và độ tin cậy 1 – α cho trước, có thể tìm được cặp α_1 và α_2 sao cho $\alpha_1 + \alpha_2 = \alpha$ và hai giá trị tới hạn Student tương ứng là $t_{1-\alpha_1}^{(n-1)}$ và $t_{\alpha_2}^{(n-1)}$ thỏa mãn điều kiện:

$P(T \leq t_{1-\alpha_1}^{(n-1)}) = \alpha_1$ và

$P(T \geq t_{\alpha_2}^{(n-1)}) = \alpha_2$

Từ đó:

$$P(-t_{1-\alpha_1}^{(n-1)} \leq T \leq t_{\alpha_2}^{(n-1)}) = 1-(\alpha_1+\alpha_2) = 1-\alpha$$

Vì giá trị tới hạn có tính chất $t_{\alpha_1}^{(n-1)}$ = - $t_{1-\alpha_1}^{(n-1)}$ nên có thể viết:

$$P(-t_{\alpha_1}^{(n-1)} \le T \le t_{\alpha_2}^{(n-1)}) = 1 - \alpha \qquad (2 - 6)$$

Thay $T = \dfrac{(\overline{x} - \mu)\sqrt{n}}{S}$ vào công thức (2 - 6) thì công thức ước lượng khoảng số trung bình tổng thể theo luật Student là:

$$P(\overline{x} - t_{\alpha_2}^{(n-1)} \frac{S}{\sqrt{n}} \le \mu \le \overline{x} + t_{\alpha_1}^{(n-1)} \frac{S}{\sqrt{n}}) = 1 - \alpha \quad (2 - 7)$$

và khoảng tin cậy của μ sẽ là:

$$\left(\overline{x} - t_{\alpha_2}^{(n-1)} \frac{S}{\sqrt{n}} ; \overline{x} + t_{\alpha_1}^{(n-1)} \frac{S}{\sqrt{n}}\right)$$

Khi ước lượng khoảng tin cậy đối xứng thì $\alpha_1 = \alpha_2 = \alpha/2$, công thức (2 - 7) trở thành:

$$P(\overline{x} - t_{\alpha/2}^{(n-1)} \frac{S}{\sqrt{n}} \le \mu \le \overline{x} + t_{\alpha/2}^{(n-1)} \frac{S}{\sqrt{n}}) = 1 - \alpha \quad (2 - 8)$$

Đây là công thức thường áp dụng nhất để ước lượng khoảng tin cậy của trung bình tổng thể khi $n < 30$.

Từ công thức (2 - 8) sai số tới hạn (hay độ chính xác) là $\varepsilon = t_{\alpha/2}^{(n-1)} \dfrac{S}{\sqrt{n}}$, khoảng tin cậy sẽ là:

$$I = 2\varepsilon = 2t_{\alpha/2}^{(n-1)} \frac{S}{\sqrt{n}},$$

$$\varepsilon(\%) = \frac{\varepsilon}{\overline{x}}.100 = t_{\alpha/2}^{(n-1)} \frac{S}{\overline{x}\sqrt{n}}.100 = t_{\alpha/2}^{(n-1)} \frac{CV(\%)}{\sqrt{n}}$$

Dung lượng mẫu cần thiết để đạt được độ chính xác

$\varepsilon_0(\%)$ là:

$$n_{\min} = \left(t_{\alpha/2}^{(n-1)} \frac{CV(\%)}{\varepsilon_0(\%)} \right)^2$$

Nếu mẫu được chọn từ tổng thể hữu hạn theo cách rút mẫu không lặp (không hoàn lại), khi n > 0,1N, để bảo đảm độ chính xác của ước lượng thì sai số tới hạn sẽ nhân thêm hệ số điều chỉnh $\sqrt{\frac{N-n}{N-1}}$, khi đó công thức (2 - 8) trở thành:

$$P(\overline{x} - t_{\alpha/2}^{(n-1)} \frac{S}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}} \le \mu \le \overline{x} + t_{\alpha/2}^{(n-1)} \frac{S}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}}) = 1-\alpha \quad (2 - 9)$$

Dựa vào đây để tính toán ε, ε(%), từ đó sẽ xác định được dung lượng mẫu cần thiết để đạt được độ chính xác xác tương đối cho trước $\varepsilon_0(\%)$.

Ví dụ: Từ ví dụ ở Bảng 2.1, mục 2.2.2.2 trên đây, nếu chỉ điều tra ngẫu nhiên 10 cây trong tổng số 50 cây. Hãy ước lượng khoảng tin cậy của tổng thể (N = 50).

Bảng 2.2 sau đây là kết quả tính toán các tham số từ 7 cách chọn mẫu có hoàn lại, mỗi mẫu 10 cây.

Bảng 2.2 cho các nhận xét:

- Dù rút mẫu có hoàn lại nhưng với n < 30, giá trị trung bình của các mẫu chênh lệch nhau đáng kể, $\overline{x}_{max} - \overline{x}_{min}$ = 16,82 g/cây, dao động trong khoảng 66,53 g/cây đến 83,35 g/cây.

- Giá trị độ lệch chuẩn S khác nhau không nhiều. Nếu so với trung bình tổng thể thì hệ số biến động tương đối ổn định và khác biệt với hệ số biến động tổng thể không nhiều.

Bảng 2.2: Các tham số của tổng thể và 7 mẫu chọn ngẫu nhiên có hoàn lại

Tham số	Tổng thể	Mẫu						
		1	2	3	4	5	6	7
$\overline{x}$	76,92	83,35	73,76	78,05	82,07	66,53	82,78	69,22
CV(%)	24,38	25,21	27,22	21,87	13,69	33,47	23,81	28,24
S	18,76	21,01	20,08	17,07	11,24	22,27	19,71	19,55
ε	5,20	15,03	14,36	12,21	8,04	15,93	14,10	13,98
I=2ε	10,40	30,06	28,73	24,42	16,08	31,86	28,19	27,96
ε(%)	6,76	18,03	19,47	15,65	9,79	23,94	17,03	20,20
$\varepsilon_{\text{điểm}}$	2,65	6,65	6,35	5,40	3,55	7,04	6,23	6,18
$\varepsilon_{đ}$(%)	3,45	7,97	8,61	6,92	4,33	10,59	7,53	8,93

Điều đó chứng tỏ rằng có thể sử dụng CV(%) của mẫu và độ chính xác ε_0(%) cho trước theo yêu cầu để dự tính số lượng cá thể cần thiết cho cuộc điều tra.

- Khoảng tin cậy của ước lượng (I = 2ε) theo luật Student trong trường hợp mẫu nhỏ lớn và dao động từ 16,08 - 31,86, trong khi đó khoảng tin cậy của ước lượng theo luật chuẩn là 10,40. Tất cả các mẫu đều có sai số tương đối khá lớn, hầu hết đều lớn, từ 16 - 20%, chỉ có mẫu 4 là 9,8%, cao hơn đáng kể so với sai số tổng thể (6,76%).

- Với dung lượng mẫu n = 10, ước lượng điểm mẫu tuy có chênh lệch với ước lượng điểm tổng thể, nhưng có thể áp dụng phương pháp ước lượng điểm để thu thập giá trị trung bình các ô trong thí nghiệm trên đồng ruộng.

Với trung bình tổng thể μ = 76,92, độ lệch chuẩn tổng thể S = 18,76 (Bảng 2.1), Bảng 2.3 sau đây sẽ làm rõ sự khác biệt giữa luật Student và luật chuẩn trong ước lượng trung bình tổng thể.

Từ Bảng 2.3 có thể thấy rằng:

- Khi dung lượng mẫu n tăng lên, phân phối Student sẽ hội tụ rất nhanh về phân phối chuẩn. Do đó nếu $n > 30$ có thể dùng phân phối chuẩn thay cho phân phối Student.

Bảng 2.3: Ước lượng một số tham số thống kê theo luật Student khi dung lượng mẫu thay đổi với $\alpha/2 = 0,025$.

Độ tự do (n-1)	$t_{0,025}^{n-1}$ (*)	Tham số thống kê				
		ε	I = 2ε	ε (%)	ε điểm	$\varepsilon_đ$(%)
Theo luật Student với các độ tự do khác nhau						
1	25,452	337,63	675,3	438,93	13,27	17,25
3	4.177	39,18	78,35	50,93	9,38	12,19
5	3,163	24,23	48,46	31,50	7,66	9,96
10	2,634	14,90	29,80	19,37	5,66	7,35
20	2,423	9,92	19,84	12,90	4,09	5,32
30	2,360	7,95	15,90	10,34	3,37	4,38
40	2,329	6,82	13,65	8,87	2,93	3,81
45	2,319	6,41	12,83	8,34	2,77	3,60
49	2,312	6,13	12,27	7,98	2,65	3,45
Theo luật chuẩn						
	$u_{0,05}^{n-1} = 1,96$	5,20	10,40	6,76	2,65	3,44

(*) Tra trong phần mềm Excel: = tinv(0.025,df$_{(n-1)}$)

- Khi dung lượng mẫu nhỏ ($n < 30$) việc thay thế luật Student bằng luật chuẩn có thể dẫn đến sai lầm lớn. Chẳng hạn, với hàm phân phối chuẩn với $\alpha = 0,05$, $u_{0,05} = 1,96$, trong khi đó với $n = 4$ (df = 3), giá trị tới hạn Student $t_{0,025}^{(3)} = 4,18$, tức là sai lệch đến 2,22.

- Về ước lượng điểm: Một quần thể của một giống (có kiểu gen đồng nhất), hệ số biến động giữa các cá thể của các chỉ tiêu sinh trưởng như chiều cao cây, số hạt, quả/cây ở trên ruộng phụ thuộc chủ yếu vào sự đồng đều của đất trồng, thông thường là 8 – 15%, có khi lên tới 20 – 25% (như Bảng 2.3) hoặc cao hơn. Với các chỉ tiêu này, để có độ chính xác của ước lượng điểm đạt 6 – 8%, dung lượng mẫu khoảng 10 cây. Tuy nhiên có nhiều chỉ tiêu mà giữa các cá thể trong cùng một giống rất ít khác biệt nhau kể cả khi trồng trên ruộng độ đồng đều về đất không cao. Với các chỉ tiêu này hệ số biến động lại rất thấp, không vượt quá 8%, thậm chí chỉ 1 – 2%. Như vậy chỉ cần theo dõi khoảng 3 cá thể thì có thể đạt được độ chính xác rất cao. Chẳng hạn, hệ số biến động khối lượng 100 hạt của một giống đậu nành là 5%, chỉ cần cân 3 mẫu (n = 3) đã đạt độ chính xác $\varepsilon(\%) = \dfrac{CV(\%)}{\sqrt{n}} = \dfrac{5}{\sqrt{3}}$ = 2,9.

Những điều cần lưu ý khi ước lượng trung bình tổng thể:

- Trước hết, cần xem đối tượng nghiên cứu để quyết định phương pháp lấy mẫu. Nếu quần thể đồng nhất (ví dụ như một giống thuần hay giống lai F_1) có thể lấy một mẫu nhỏ hoặc mẫu lớn. Nếu quần thể không đồng nhất (ví dụ như quần thể phân ly các thế hệ lai, quần thể đột biến hay sản phẩm của kỹ thuật di tuyền), khi ước lượng trung bình quần thể cần phải lấy nhiều mẫu, hay lấy mẫu thử theo phép rút mẫu ngẫu nhiên rồi tính toán dung lượng mẫu cần thiết để bảo đảm độ chính xác của ước lượng.

- Căn cứ vào độ lớn tổng thể để quyết định dung lượng mẫu cần thiết. Nếu quần thể không lớn có thể quan sát toàn bộ tổng thể.

- Nên áp dung phương pháp lấy mẫu ngẫu nhiên.

- Chỉ nên áp dụng phương pháp ước lượng điểm khi dung lượng mẫu nhỏ.

2.3. ƯỚC LƯỢNG PHƯƠNG SAI TỔNG THỂ (σ^2)

2.3.1. Ước lượng điểm phương sai tổng thể

Giả sử là đại lượng ngẫu nhiên với phương sai tổng thể σ^2 chưa biết. Kết quả rút mẫu với dung lượng *n* ta được các giá trị quan sát: $x_1, x_2, ..., x_n$. Giá trị phương sai mẫu chưa hiệu chỉnh sẽ là

$$\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$$

Nếu sử dụng phương sai mẫu chưa hiệu chỉnh để ước lượng σ^2 ta có:

$$E\left\{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2\right\} = E\left\{\frac{1}{n}\sum_{i=1}^{n}\left(x_i^2 - 2x_i\overline{x} + \overline{x}^2\right)\right\}$$

$$= E\left\{\frac{1}{n}\left[\sum_{i=1}^{n}x_i^2 - 2\overline{x}\sum_{i=1}^{n}x_i + \sum_{i=1}^{n}\overline{x}^2\right]\right\}$$

$$= \frac{1}{n}E\left\{\sum_{i=1}^{n}\left(x_i^2 - 2n\overline{x}^2 + \overline{x}^2\right)\right\}$$

$$= \frac{1}{n}E\left\{\sum_{i=1}^{n}\left(x_i^2 - n\overline{x}\right)^2\right\} = \frac{1}{n}\left\{\sum_{i=1}^{n}E\left(x_i^2\right) - nE\left(\overline{x}^2\right)\right\}$$

Do $E\left(x_i^2\right) = V\left(x_i\right) + \left[E\left(x_i\right)\right]^2$ và $E\left(\overline{x}^2\right) = V\left(\overline{x}\right) + \left[E\left(\overline{x}\right)\right]^2$.

Cần chứng minh $\hat{S}^2 = V\left(\overline{x}\right) = V\left(X\right) = \sigma^2$

Theo tính chất 2 và 4 của phương sai ta có:

$$V(\overline{x}) = V\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(x_i) = \frac{1}{n^2}\sum_{i=1}^{n} V(X)$$

$$= \frac{1}{n}\sigma^2$$

Như vậy phương sai chưa hiệu chỉnh là một ước lượng chệch.

Phương sai hiệu chỉnh $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$ là ước lượng không chệch cho phương sai tổng thể. Thật vậy:

$$E\left\{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2\right\} = \frac{n}{n-1}E\left\{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2\right\}$$

$$= \frac{n}{n-1}.\frac{n-1}{n}V(X) = \sigma^2$$

2.3.2. Ước lượng khoảng phương sai tổng thể

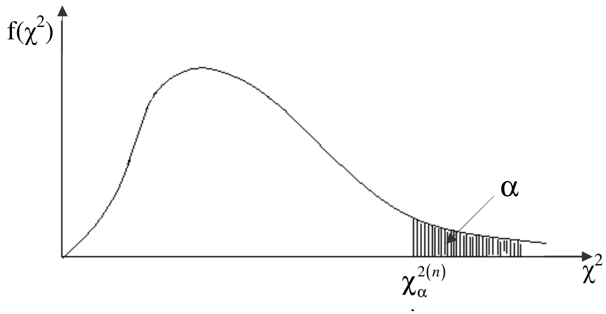
Việc ước lượng khoảng phương sai tổng thể được tiến hành theo luật phân phối "khi bình phương".

Đại lượng ngẫu nhiên liên tục χ^2 phân phối theo luật "khi bình phương" với n bậc tự do, nếu hàm mật độ xác suất của nó được xác định bằng công thức:

$$f(x) = \begin{cases} 0 & \text{với } x \leq 0 \\ \dfrac{1}{2^{\frac{n}{2}}.\Gamma\left(\dfrac{n}{2}\right)} e^{\frac{x}{2}}.x^{\frac{n}{2}-1} & \text{với } x > 0 \end{cases}$$

trong đó $\Gamma(x)$ là hàm Gamma.

Và, đồ thị hàm mật độ xác suất có dạng như hình 2.2.



Hình 2.2: Đồ thị hàm mật độ xác suất theo luật khi bình phương

Người ta đã chứng minh được rằng nếu đại lượng ngẫu nhiên χ^2 phân phối theo luật khi bình phương với n bậc tự do thì kỳ vọng $E(\chi^2) = n$ và phương sai $V(\chi^2) = 2n$.

Theo luật phân phối khi bình phương, với dung lượng mẫu n, số bậc tự do là $n - 1$ và độ tin cậy $1 - \alpha$ cho trước, có thể tìm được cặp α_1 và α_2 sao cho $\alpha_1 + \alpha_2 = \alpha$ và hai giá trị tới hạn khi bình phương tương ứng là $\chi_{1-\alpha_1}^{2(n-1)}$ và $\chi_{\alpha_2}^{2(n-1)}$ thỏa mãn điều kiện:

$$P(\chi^2 \leq \chi_{1-\alpha_1}^{2(n-1)}) = \alpha_1 \text{ và}$$

$$P(\chi^2 \geq \chi_{\alpha_2}^{2(n-1)}) = \alpha_2$$

Từ đó:

$$P(\chi_{1-\alpha_1}^{2(n-1)} \leq \chi^2 \leq \chi_{\alpha_2}^{2(n-1)}) = 1 - (\alpha_1 + \alpha_2) = 1 - \alpha \quad (2 - 10)$$

Thay $\chi^2 = \dfrac{(n-1)S^2}{\sigma^2}$ vào công thức (2 - 10) thì công thức ước lượng phương sai tổng thể theo luật khi bình phương là:

$$P\left(\frac{(n-1)S^2}{\chi^{2(n-1)}_{\alpha_2}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^{2(n-1)}_{1-\alpha_2}}\right) = 1-\alpha \qquad (2 - 11)$$

và khoảng tin cậy của μ là:

$$\left(\frac{(n-1)S^2}{\chi^{2(n-1)}_{\alpha_2}};\frac{(n-1)S^2}{\chi^{2(n-1)}_{1-\alpha_2}}\right)$$

Khi ước lượng khoảng tin cậy đối xứng thì $\alpha_1 = \alpha_2 = \alpha/2$, công thức (2 - 11) trở thành:

$$P\left(\frac{(n-1)S^2}{\chi^{2(n-1)}_{\alpha/2}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^{2(n-1)}_{1-\alpha/2}}\right) = 1-\alpha \qquad (2 - 12)$$

Ví dụ: Hãy ước lượng phương sai tổng thể về năng suất (cá thể) của tổ hợp bông lai F_1 S02-13/TM1 trồng tại Đại học Nông Lâm Tp. HCM, 2008 theo số liệu Bảng 2.2 với S = 18,76 với độ tin cậy 0,95.

Ở đây $\alpha = 1 - 0{,}95 = 0{,}05$; $\alpha/2 = 0{,}025$; $1 - \alpha/2 = 0{,}975$.

Tra giá trị χ^2 trong phần mềm Excel, ta có $\chi^{2(49)}_{0,025}$ = 70,222 và $\chi^{2(49)}_{0,975}$ = 31,555. Thay các giá trị này vào công thức (2 - 12) ta có:

$$\left(\frac{49\times 18{,}76^2}{70{,}222};\frac{49\times 18{,}76^2}{31{,}555}\right) = (546{,}24\ ;\ 245{,}46)$$

và: $$P(245{,}46 \le \sigma^2 \le 546{,}24) = 0{,}95$$

Như vậy ở mức tin cậy 0,95 phương sai tổng thể năng suất cá thể của tổ hợp bông lai F1 S02-13/TM1 nằm trong khoảng từ 245,46 đến 546,24.

2.4. ƯỚC LƯỢNG KHOẢNG XÁC SUẤT CÁC DẤU HIỆU ĐỊNH TÍNH CỦA MỘT TỔNG THỂ

Việc ước lượng khoảng xác suất các dấu hiệu định tính của một tổng thể được tiến hành theo luật "không – một".

- ***Luật "không - một"A(p)***

Giả sử để thử nẩy mầm một lô hạt giống, một hạt giống chỉ có một trong hai khả năng xẩy ra, hoặc là nẩy mầm (biến cố A) hoặc là không nảy mầm (biến cố $\overline{A}$). Vậy xác suất để có m hạt nảy mầm trong n hạt thử là $p = \frac{m}{n}$ còn xác suất hạt không nảy mầm là $q = \frac{n-m}{n} = 1 - \frac{m}{n} = 1 - p$.

Một cách tổng quát, giả sử ta làm một phép thử, trong đó biến cố A có thể xảy ra với xác suất bằng p. Gọi X là số lần xuất hiện A. Như vậy X là đại lượng ngẫu nhiên rời rạc với hai giá trị hoặc là bằng 0 (nếu không xuất hiện A) hoặc là bằng 1 (nếu xuất hiện A) với các xác suất tương ứng được biểu thị bằng công thức:

$$P_x = p^x q^{1-x} \text{ với } x = 0; 1, \text{ trong đó } q = 1 - p.$$

Phân phối thỏa mãn với công thức trên đây được gọi là phân phối theo quy luật "không – một" cho đại lượng ngẫu nhiên X với tham số p.

Bảng phân phối xác suất của đại lượng ngẫu nhiên X theo quy luật "không – một" có dạng:

X	0	1
P	q	p

$(q = 1 - p)$

Từ bảng phân phối xác suất ta có:

Kỳ vọng: $E(X) = p$

Phương sai: $V(X) = pq$

Độ lệch chuẩn: $\sigma_x = \sqrt{pq}$

- ***Khoảng ước lượng và độ chính xác của ước lượng***

Với dung lượng mẫu đủ lớn, hoặc nếu $n > 5$ và $\dfrac{\left|\sqrt{\dfrac{p_m}{1-p_m}} - \sqrt{\dfrac{1-p_m}{p_m}}\right|}{\sqrt{n}} < 0{,}3$ (p_m là xác suất mẫu) thì luật phân phối "không – một" sẽ có phân phối gần phân phối chuẩn tắc N(0, 1), do đó vẫn có thể ước lượng khoảng tin cậy tần suất p của tổng thể. Với độ tin cây $1 - \alpha$ cho trước, có thể tìm được cặp α_1 và α_2 sao cho $\alpha_1 + \alpha_2 = \alpha$ và hai giá trị tới hạn chuẩn $u_{1-\alpha_1}$ và u_{α_2} thỏa mãn điều kiện:

$$P(U \leq u_{1-\alpha_1}) = \alpha_1 \text{ và } P(U \geq u_{\alpha_2}) = \alpha_2$$

Từ đó:

$$P(u_{1-\alpha_1} \leq U \leq u_{\alpha_2}) = 1 - (\alpha_1 + \alpha_2) = 1 - \alpha \qquad (2 - 13)$$

Thay $U = \dfrac{(p_m - p)\sqrt{n}}{\sqrt{p_m(1-p_m)}}$ vào công thức (2 - 13) và áp dụng tính chất $u_{\alpha_1} = -u_{1-\alpha_1}$ thì công thức ước lượng xác suất p

tổng thể là:

$$P\left(p_m - u_{\alpha_2}\sqrt{\frac{p_m(1-p_m)}{n}} \le p \le p_m + u_{\alpha_1}\sqrt{\frac{p_m(1-p_m)}{n}}\right) = 1-\alpha \quad (2 - 14)$$

và khoảng tin cậy của *p* là:

$$\left(p_m - u_{\alpha_2}\sqrt{\frac{p_m(1-p_m)}{n}}; p_m + u_{\alpha_1}\sqrt{\frac{p_m(1-p_m)}{n}}\right)$$

Khi ước lượng khoảng tin cậy đối xứng thì $\alpha_1 = \alpha_2 = \alpha/2$, công thức (2 - 14) trở thành:

$$P\left(p_m - u_{\alpha/2}\sqrt{\frac{p_m(1-p_m)}{n}} \le p \le p_m + u_{\alpha/2}\sqrt{\frac{p_m(1-p_m)}{n}}\right) = 1-\alpha \quad (2\text{-}15)$$

Đây là công thức thường áp dụng nhất để ước lượng khoảng tin cậy xác suất *p*. Với $\alpha = 0{,}05$, $u_{\alpha/2} = 1{,}96$; $\alpha = 0{,}01$, $u_{\alpha/2} = 2{,}58$ và $\alpha = 0{,}001$, $u_{\alpha/2} = 3{,}29$.

Từ công thức (2 - 15), sai số tới hạn của ước lượng (độ chính xác) là $\varepsilon = u_{\alpha/2}\sqrt{\frac{p_m(1-p_m)}{n}}$ và khoảng tin cậy là $I = 2\varepsilon$. Dung lượng mẫu cần thiết để đạt được độ chính xác cho trước ε_0 là:

$$n_{\min} = (u_{\alpha/2})^2 \frac{p_m(1-p_m)}{\varepsilon_0^{\,2}}$$

Ví dụ: Để kiểm nghiệm tỷ lệ nảy mầm một giống bắp lai, người ta đã tiến hành thử 4 mẫu, mỗi mẫu 100 hạt. Kết quả như sau:

Mẫu thử	1	2	3	4	Tổng
Số hạt nảy mầm	93	89	87	96	365

Hãy ước lượng khoảng xác suất nảy mầm của lô hạt giống và số lượng hạt cần thử để đạt sai số không vượt quá 3% với độ tin cậy 95%.

Giải: n = 400, p_m = 365/400 = 0,913, α = 1 – 0,95 = 0,05, $u_{0,025}$ = 1,96 và ε_0 = 0,03.

Với dung lượng mẫu đủ lớn, theo công thức (2 - 15) ta có:

$$P\left(0,885 \le p \le 0,940\right) = 0,95$$

Như vậy với mức tin cậy 95% tỷ lệ nảy mầm của hạt giống nằm trong khoảng 88,5% đến 94,0%.

Số hạt cần thử: $n_{\min} = (u_{\alpha/2})^2 \dfrac{p_m\left(1 - p_m\right)}{{\varepsilon_0}^2}$

$$= 1,96^2 \frac{0,913\left(1 - 0,913\right)}{0,03^2} = 340,8 \approx 341$$

Để đạt sai số không vượt quá 3% cần thử tối thiểu 341 hạt.

Chương 3
SO SÁNH CÁC THAM SỐ

Giá trị trung bình, phương sai và các tham số khác theo các dấu hiệu định lượng hoặc xác suất các dấu hiệu định tính của các tổng thể được ước lượng từ những số liệu qua các cuộc điều tra khảo sát mẫu là những kết quả mô tả tổng thể. Để cung cấp những dữ liệu cần thiết phục vụ cho mục tiêu nghiên cứu hoặc ứng dụng trong sản xuất, cần phải đánh giá, so sánh và phân tích mối quan hệ giữa các tham số của các tổng thể. Chương này sẽ đề cập đến việc so sánh các tham số tổng thể từ kết quả các cuộc điều tra khảo sát các mẫu, gồm:

- So sánh hai trung bình và mở rộng (phương pháp tham số và phi tham số);

- So sánh hai phương sai và mở rộng;

- Đánh giá tính độc lập của các dấu hiệu định tính.

Việc so sánh trong các thí nghiệm sẽ được đề cập ở phần 2.

3.1. SO SÁNH HAI TRUNG BÌNH VÀ MỞ RỘNG

3.1.1. Phương pháp tham số

3.1.1.1. Cơ sở lý luận

Để so sánh hai trung bình mẫu $\overline{X}_i$ và $\overline{X}_2$ không đơn giản là so sánh hiệu số $\overline{X}_1$ - $\overline{X}_2$, bởi vì mỗi số trung bình đều có sai số ε_i: $\varepsilon_i = u_{\alpha/2}\dfrac{S_i}{\sqrt{n_i}}$ (trường hợp mẫu lớn) hay $\varepsilon_i = t_\alpha \dfrac{S_i}{\sqrt{n_i}}$ (cho mọi trường hợp, theo chương 2).

Như vậy để so sánh hai hay nhiều trung bình ở độ tin cậy 1 - α nào đó cần phải xác định khoảng khác biệt tối thiểu có ý nghĩa phân biệt (Least Significant Difference - LSD) giữa chúng: $LSD_\alpha = t_\alpha Sd$, trong đó t_α là giá trị tới hạn phân phối Student ở mức α (thường gọi là $t_{bảng}$ hay $t_{lý\ thuyết}$), Sd là sai số thực nghiệm giữa hai trung bình (giá trị Sd sẽ được nêu cụ thể trong từng trường hợp). Khi có được Sd dễ dàng tính được LSD_α.

Giả thuyết H_0: $\bar{X}_1 = \bar{X}_2$ được chấp nhận với độ tin cậy 1 - α (mức sai lầm α) khi $|\bar{X}_1 - \bar{X}_2| < LSD_\alpha$, còn giả thuyết H_1: $\bar{X}_1 \neq \bar{X}_2$ được chấp nhận với độ tin cậy 1 - α khi $|\bar{X}_1 - \bar{X}_2| \geq LSD_\alpha$. Do $T_{TN} = \dfrac{|\bar{X}_1 - \bar{X}_2|}{Sd}$ và $t_\alpha = \dfrac{LSD_\alpha}{Sd}$ nên thay vì kiểm định sự chênh lệch giữa hai trung bình $|\bar{X}_1 - \bar{X}_2|$ so với LSD_α, người ta chuyển sang kiểm định T_{TN} so với $t_{bảng}$. Khi $T_{TN} < t_{bảng}$ giả thuyết H_0 được chấp nhận, còn khi $T_{TN} \geq t_{bảng}$ giả thuyết H_1 được chấp nhận.

Trong trường hợp dung lượng mẫu lớn hoặc đã biết phương sai của hai tổng thể, phép nghiệm U sẽ được sử dụng để so sánh hai trung bình: $U_{TN} = \dfrac{|\bar{X}_1 - \bar{X}_2|}{Sd}$ được so với $u_{\alpha/2}$.

Để phép nghiệm đúng cần lưu ý yếu tố so sánh. Ví dụ: Khi so sánh giống dừa Xiêm được trồng ở Bình Định và giống dừa Xiêm đó được trồng ở Bến Tre tức là ta muốn đánh giá ảnh hưởng của hai điều kiện trồng đến giống dừa. Yếu tố so sánh ở đây là sinh thái trồng dừa. Còn khi so sánh các giống dừa thì chúng phải được trong một điều kiện như nhau (cùng vùng và cùng một loại đất). Và, để đánh giá giống dừa nào tốt cho vùng nào thì thí nghiệm đó phải được trồng trên nhiều vùng (thí nghiệm hai yếu tố: giống và vùng). Tuy nhiên để đánh giá giống dừa đặc sản của Bình Định được trồng ở Bình Định và giống dừa đặc sản của Bến Tre được trồng ở Bến Tre (hai giống này khác nhau) thì yếu tố so sánh ở đây là dừa của Bình Định và dừa của Bến Tre.

Ở ngay trong một nơi, nếu trồng hai giống ở hai lô khác nhau, mỗi lô lấy một số mẫu (một số cây). Việc so sánh sẽ không chính xác vì đất của hai lô không thể đồng nhất (khác nhau không nhiều thì ít) nên giá trị trung bình của hai giống không chỉ do giống tạo nên, chưa nói đến ngay trong một lô các cá thể trong một giống có sự khác nhau do yếu tố đất trồng. Tuy nhiên có thể chấp nhận khi đã xác định được độ đồng đều của đất ở mức cho phép.

Trong so sánh hai trung bình, nếu so sánh T_{TN} với t_{α} (t tới hạn hai phía – t Critical two-tail), khi đó độ tin cậy sẽ là 1- α. Nếu so sánh T_{TN} với $t_{\alpha/2}$ (t tới hạn một phía – t Critical one-tail), khi đó độ tin cậy sẽ là 1- α/2.

Trong thực nghiệm, người ta thường lấy mức α = 0,05 (độ tin cậy là 95%) khi so sánh T_{TN} với $t_{0,05}$ hoặc α = 0,01 (độ tin cậy sẽ là 99%) khi so sánh T_{TN} với $t_{0,01}$.

3.1.1.2. So sánh hai trung bình khi đã biết phương sai của hai tổng thể σ_1^2 và σ_2^2

Khi đã biết phương sai của hai tổng thể, việc so sánh giữa hai trung bình được thực hiện theo công thức:

$$U_{TN} = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad (3 - 1)$$

Trong đó:

$\bar{X}_1$ và $\bar{X}_2$ là trung bình của hai mẫu mẫu quan sát

σ_1^2 và σ_2^2 là phương sai của hai mẫu quan sát

n_1 và n_2 là dung lượng của hai mẫu quan sát

$$\text{ở đây } Sd = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Giá trị U_{TN} sẽ được so sánh với giá trị tới hạn $u_{\alpha/2}$.

Nếu $U_{TN} < u_{\alpha/2}$ thì $\bar{X}_1$ không khác $\bar{X}_2$ ở độ tin cậy $1 - \alpha$.

Nếu $U_{TN} \geq u_{\alpha/2}$ thì $\bar{X}_1$ khác $\bar{X}_2$ ở độ tin cậy $1 - \alpha$, hoặc là $\bar{X}_1 > \bar{X}_2$ hoặc là $\bar{X}_1 < \bar{X}_2$.

Giá trị $u_{\alpha/2}$ được ghi ở mục 2.2.1, chương 2.

Thường thì trong sinh học rất hiếm trường hợp biết được phương sai tổng thể.

3.1.1.3. So sánh hai trung bình khi chưa biết phương sai của hai tổng thể nhưng biết rằng chúng bằng nhau ($\sigma_1^2 = \sigma_2^2$)

Việc kiểm tra sự bằng nhau (khác nhau không có ý nghĩa) của σ_1^2 và σ_2^2 thông qua việc kiểm tra hai phương sai mẫu S_1^2 và S_2^2 nhờ phép nghiệm F (sẽ được đề cập ở mục 3.3). Khi phương sai của hai tổng thể không khác nhau có ý nghĩa, thì việc so sánh giữa hai trung bình được thực hiện theo công thức:

$$T_{TN} = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \quad (3 - 2)$$

$$\text{ở đây } Sd = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

t_α được tra với $(n_1 + n_2 - 2)$ độ tự do.

Độ tin cậy của phép so sánh khác nhau khi lấy các giá trị tới hạn của t khác nhau (giá trị tới hạn một phía và hai phía).

Ví dụ: So sánh năng suất cá thể của tổ hợp bông lai F_1 C92-52/C118A theo số liệu sau đây và F_1 S02-13/TM1 (ở ví dụ Bảng 2.1, chương 2) với các thông tin sau:

Tổ hợp C92-52/C118A: $n_1 = 45$; $\bar{x} = 63{,}56$, $S^2 = 387{,}90$

Tổ hợp S02-13/TM1: $n_2 = 50$; $\bar{x} = 76{,}92$, $S^2{}_1 = 351{,}68$

Giải:

Trước hết phải kiểm tra hai phương sai. Để kiểm tra ta dùng tiêu chuẩn F:

$$F_{TN} = \frac{387{,}90}{351{,}68} = 1{,}10$$

tra bảng F với hai độ tự do 49 và 44 ta có $F_{0,05} = 1{,}63$. Như vậy có thể coi hai phương sai bằng nhau.

Áp dụng công thức (3 - 2) ta có:

$$T_{TN} = \frac{\left|\overline{X}_1 - \overline{X}_2\right|}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$= \frac{76,92 - 63,56}{\sqrt{\frac{(50-1)(351,68)+(45-1)(387,90)}{50+45-2}\left(\frac{1}{50}+\frac{1}{45}\right)}}$$

$= 3,38$

t_α được tra với 50 + 45 – 2 = 93 độ tự do ta được:

$$t_{0,05}^{93} = 1,99\ ,\ t_{0,01}^{93} = 2,63\,.$$

Như vậy, T_{TN} = 3,38 > $t_{0,01}^{93} = 2,63$, năng suất F_1 tổ hợp S02-13/TM1 cao hơn tổ hợp C92-52/C118A với độ tin cậy 99%.

Kết quả so sánh trung bình F_1 S02-13/TM1 và F_1 C92-52/C118A trên phần mềm Excel:

t-Test: Two-Sample Assuming Equal Variances

	C92-52/C118A	S02-13/TM1
Mean	63.56	76.922
Variance	387.90	351.68
Observations	45	50
Pooled Variance	368.8169	
Hypothesized Mean Difference	0	
df	93	
t Stat	-3.3861	
P(T<=t) one-tail	0.0005	
t Critical one-tail	1.6614	
P(T<=t) two-tail	0.0010	
t Critical two-tail	1.9858	

3.1.1.4. So sánh hai trung bình khi chưa biết phương sai của hai tổng thể nhưng biết rằng chúng khác nhau ($\sigma_1^2 \neq \sigma_2^2$)

Với dung lượng mẫu đủ lớn (n > 30), khi phương sai của hai tổng thể khác nhau, việc so sánh giữa hai trung bình được thực hiện theo công thức:

$$T_{TN} = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \qquad (3 - 3)$$

trong đó: $\bar{X}_1$ và $\bar{X}_2$ là trung bình của hai mẫu mẫu quan sát; S_1^2 và S_2^2 là phương sai của hai mẫu quan sát; n_1 và n_2 là dung lượng của hai mẫu quan sát.

Giá trị tới hạn phân phối Student t_α được tra với k độ tự do lấy số nguyên từ công thức sau:

$$k = \frac{(n_1 - 1)(n_2 - 1)\left(\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right)^2}{(n_1 - 1)\left(\dfrac{S_1^2}{n_1}\right)^2 + (n_2 - 1)\left(\dfrac{S_2^2}{n_2}\right)^2} \qquad (3 - 4)$$

Nếu $n_1 = n_2 = n$ thì công thức (3 - 3) trở thành:

$$T_{TN} = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{\sqrt{\dfrac{S_1^2 + S_2^2}{n}}}.$$

Người ta đã chứng minh được rằng nếu $\bar{X}_1$ khác $\bar{X}_2$ một cách ngẫu nhiên thì cứ 100 lần rút mẫu không quá 5 lần T_{TN} < $t_{0,05}$ (giá trị tới hạn t hai chiều) với độ tin cậy giả thuyết H_1

là 95% và không quá 2,5 lần $T_{TN} < t_{0,025}$ (giá trị tới hạn t một chiều) với độ tin cậy 97,5%. Ngược lại, khi $\bar{X}_1$ không khác $\bar{X}_2$ một cách ngẫu nhiên thì cứ 100 lần rút mẫu không quá 5 lần $T_{TN} > t_{0,05}$ với độ tin cậy giả thuyết H_0 là 95% và không quá 2,5 lần $T_{TN} > t_{0,025}$ với độ tin cậy 97,5%.

Như vậy, nếu $T_{TN} \geq t_{bảng}$ ở mức α thì kết luận rằng $\bar{X}_1$ khác $\bar{X}_2$ ở độ tin cậy 1 - α. Và, nếu $T_{TN} < t_{bảng}$ ở mức α thì kết luận rằng $\bar{X}_1$ không khác với $\bar{X}_2$ ở độ tin cậy 1 - α.

Lưu ý rằng, khi *n* càng lớn t_α càng gần đến $u_{\alpha/2}$, phân phối Student càng gần phân phối chuẩn. Khi đó việc so sánh hai trung bình theo công thức:

$$U_{TN} = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Khi đó ta sử dụng phép nghiệm U.

Ví dụ: So sánh năng suất cá thể thế hệ F_1 và F_2 của tổ hợp bông lai C92-52/C118A trồng tại Đại học Nông Lâm Tp. HCM 2008 theo kết quả theo dõi sau đây.

Năng suất cá thể (g/cây) của 45 cây F_1

50,7	30,0	32,9	78,1	41,3	72,9	57,1	52,0	94,5	87,7
69,7	64,6	72,9	79,6	91,2	46,6	42,9	42,9	29,4	76,4
72,0	65,8	58,1	50,1	53,1	71,0	54,5	52,1	62,3	94,0
59,2	38,5	57,9	66,0	39,6	78,6	37,4	54,8	78,4	48,6
98,0	68,0	96,8	97,8	94,2					

Trung bình: 63,56; Phương sai: 387,90; Độ lệch chuẩn:19,70

Năng suất cá thể (g/cây) của 110 cây F_2

20,9	69,4	42,5	21,3	45,6	21,5	14,9	10,7	11,3	20,3
103,0	10,4	97,0	53,0	57,5	41,8	79,5	91,0	44,5	37,0
42,0	11,7	54,9	41,8	49,2	52,4	55,1	91,0	47,5	43,0
49,6	64,3	132,0	60,7	94,0	4,5	99,0	96,3	89,4	96,0
49,5	59,1	44,9	42,9	62,8	49,7	73,8	46,9	75,8	62,0
40,2	57,9	87,7	53,3	98,5	3,2	98,2	41,9	58,8	79,1
49,5	52,3	63,8	17,4	77,6	69,9	65,5	59,6	79,5	48,5
17,7	38,0	20,5	35,9	47,4	37,0	85,8	45,5	29,0	62,8
28,9	31,6	16,6	34,6	48,5	37,4	64,2	50,4	26,5	94,0
78,0	16,6	37,8	38,1	83,3	86,4	29,6	25,5	33,4	11,3
74,2	19,9	75,8	59,1	33,1	66,4	139	52,4	31,8	98,0

Trung bình: 53,45; Phương sai: 768,61; Độ lệch chuẩn: 27,72

Rõ ràng hai tập hợp số liệu có phương sai khác nhau. Áp dụng công thức (3 - 3) ta có:

$$T_{TN} = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$= \frac{63,56 - 53,45}{\sqrt{\frac{387,9}{45} + \frac{768,61}{110}}} = 2,56$$

Thay các giá trị vào công thức (3 - 4) ta được k = 136 độ tự do và với độ tự do này tiêu chuẩn T ≈ tiêu chuẩn U, $t_{0,05}^{136} \approx u_{0,025} = 1,98$, còn $t_{0,01}^{136} \approx u_{0,015} = 2,61$.

Như vậy, năng suất F_1 cao hơn năng suất F_2 với độ tin cậy trên 95% gần 99%.

Kết quả so sánh trung bình F_1 và F_2 trên phần mềm Excel:

t-Test: Two-Sample Assuming Unequal Variances

	F1	F2
Mean	63.56	53.45
Variance	387.90	768.61
Observations	45	110
Hypothesized Mean Difference	0	
df	114	
t Stat	2.5596	
P(T<=t) one-tail	0.0059	
t Critical one-tail	1.6583	
P(T<=t) two-tail	0.0118	
t Critical two-tail	1.9810	

Để so sánh chính xác hai trung bình khi hai phương sai mẫu S_1^2 và S_2^2 khác nhau ở mức tin cậy 95%, đòi hỏi phải có dung lượng mẫu lớn (n > 30).

Với dung lượng mẫu nhỏ (n < 30), khi hai phương sai mẫu S_1^2 và S_2^2 khác nhau việc so sánh sẽ kém chính xác. Trong trường hợp này có thể áp dụng phương pháp rút mẫu ngẫu nhiên có hoàn lại từ mẫu đã có rất nhiều lần (hàng trăm lần) để ước lượng trung bình mới của hai mẫu và tiến hành so sánh như trường hợp dung lượng mẫu lớn.

3.1.1.5. So sánh hai trung bình lấy mẫu theo cặp (Paired two samples)

Cơ sở của phép so sánh này là: Nếu hai đại lượng ngẫu nhiên X_1, X_2 có quan hệ phân phối theo luật Student thì đại lượng ngẫu nhiên tổng hay hiệu của chúng cũng phân phốitheo luật Student.

Trong trường hợp này, việc so sánh hai trung bình $\overline{X}_1$ và $\overline{X}_2$ được kiểm định theo tiêu chuẩn T sau đây:

$$T_{TN} = \frac{\overline{d}}{S_d}\sqrt{n} \qquad (3 - 5)$$

trong đó:

$\overline{d} = \frac{1}{n}\sum_{i=1}^{n}(x_{1i} - x_{2i})$ là trung bình độ lệch;

$S_d = \sqrt{V(d_i)}$ là độ lệch chuẩn của các độ lệch $d_i = x_{1i} - x_{2i}$

n là dung lượng mẫu ($n = n_{x_1} = n_{x_2}$)

Nếu $T_{TN} \geq t_{0,05}$ với n – 1 bậc tự do thì $\overline{X}_1 \neq \overline{X}_2$, ngược lại $T_{TN} < t_{0,05}$ thì $\overline{X}_1 \approx \overline{X}_2$.

Độ tin cậy của phép so sánh khác nhau khi lấy các giá trị tới hạn của t khác nhau (giá trị tới hạn một phía và hai phía).

Ví dụ: Kết quả học tập của 26 sinh viên năm thứ nhất và năm thứ 2 được ghi ở Bảng 3.1.

Loại trừ những trường hợp may rủi trong thi cử, kết quả học tập ở Bảng 3.1 là do sự cố gắng của các em.

Số liệu Bảng 3.1 cho thấy trong 26 sinh viên, có 14 em có điểm năm 2 cao hơn năm 1 (+), 11 em có điểm năm 2 thấp hơn năm 1 (-) và 1 em điểm 2 năm bằng nhau. Khó có thể so sánh kết quả học tập 2 năm học.

Để áp dụng công thức (3 - 5) ta tính:

$$\overline{d} = \frac{1}{n}\sum_{i=1}^{n}(x_{1i} - x_{2i}) = (-\,0,4 + 0,2 - 0,1 + \ldots$$

$$-\,0,4) = 0,073$$

$$S_d = \sqrt{V(d_i)} = 0,346$$

Bảng 3.1: Điểm trung bình các môn của 26 sinh viên

TT	Điểm các năm		$X_2 - X_1$	TT	Điểm các năm		$X_2 - X_1$
	1 (X_1)	2(X_2)	(d_i)		1 (X_1)	2(X_2)	(d_i)
1	8,3	7,9	– 0,4	14	6,7	7,2	+ 0,5
2	8,4	8,6	+ 0,2	15	6,9	7,1	+ 0,2
3	8,2	8,1	– 0,1	16	9,3	9,4	+ 0,1
4	6,5	7,2	+ 0,7	17	5,8	5,6	– 0,2
5	7,8	7,3	– 0,5	18	8,6	8,8	+ 0,2
6	6,9	7,2	+ 0,3	19	6,2	5,8	– 0,4
7	7,1	7,1	0,0	20	7,9	8,2	+ 0,3
8	8,4	8,7	+ 0,3	21	7,8	7,6	– 0,2
9	7,6	7,9	+ 0,3	22	9,0	8,7	– 0,3
10	7,8	7,7	– 0,1	23	8,2	8,6	+ 0,4
11	7,5	7,7	+ 0,2	24	8,4	8,2	– 0,2
12	6,4	7,2	+ 0,8	25	7,3	7,2	– 0,1
13	6,8	7,1	+ 0,3	26	5,8	5,4	– 0,4

Cuối cùng: $T_{TN} = \frac{\bar{d}}{S_d}\sqrt{n} = \frac{0,073}{0,346}\sqrt{26} = 1,08$

$$T_{TN} = 1,08 < t_{0,05}^{(26-1)} = 2,06.$$

Như vậy kết quả học tập hai năm như nhau ($\bar{X}_1 \approx \bar{X}_2$).

3.1.2. Phương pháp phi tham số

Các phép so sánh hai trung bình bằng phương pháp tham số trên đây được thực hiện với điều kiện là các tổng thể có phân phối theo luật chuẩn hoặc là có dung lượng mẫu lớn để có thể áp dụng định lý giới hạn trung tâm. Nếu điều kiện này bị vi phạm, việc kiểm định theo các tiêu chuẩn trên kém hiệu nghiệm. Trong trường hợp này cần phải sử dụng tiêu chuẩn phi tham số.

Với phương pháp phi tham số, các tiêu chuẩn kiểm định dựa vào thứ hạng xếp theo độ lớn nhỏ của các giá trị quan sát, không sử dụng tham số trung bình và phương sai.

Do các tiêu chuẩn phi tham số không chính xác bằng các tiêu chuẩn tham số, nên nếu điều kiện kiểm định tham số được thỏa mãn thì không nên sử dụng tiêu chuẩn phi tham số.

Phương pháp phi tham số dùng để so sánh hai hay nhiều trị trung bình của hai hay nhiều mẫu rút từ các tổng thể có nguồn gốc khác nhau (mẫu độc lập) hoặc có cùng nguồn gốc (mẫu phụ thuộc).

3.1.2.1. So sánh các trung bình các mẫu độc lập

- ***So sánh trung bình hai mẫu độc lập***

Tiêu chuẩn U của Mann và Whitney

Để áp dụng tiêu chuẩn U, hãy xét ví dụ sau đây.

Ví dụ: Để đánh giá tính đồng nhất của khu đất thí nghiệm tại Trại thực nghiệm Đại học Nông Lâm Tp. HCM, chúng tôi đã tiến hành đo chiều cao cây của giống bông S02-13 được trồng ở ba vị trí (ba lô), mỗi lô theo dõi chiều cao 10 cây. Kết quả như sau:

Chiều cao cây (cm) của các cây trong ba lô

Lô 1:	72	87	71	70	80	67	80	80	82	66
Lô 2:	97	95	90	81	92	91	95	96	84	72
Lô 3:	62	68	73	85	69	79	77	76	83	84

Để đánh giá lô 1 và 2:

Bước 1: Xếp hạng số liệu

Trước hết hãy sắp xếp hạng từ nhỏ đến lớn các số đo của cả hai lô theo thứ tự 1, 2, 3, ..., 20. Có thể sắp xếp thủ công hay nhờ phần mềm Excel trên máy vi tính. Trường hợp các số có cùng độ lớn thì thứ hạng được chia đều cho mỗi số. Ở ví dụ này kết quả xếp hạng nhờ Excel như sau:

Lô 1	72	87	71	70	80	67	80	80	82	66
Hạng	*(5)*	*13*	*4*	*3*	*(7)*	*2*	*(7)*	*(7)*	*11*	*1*
Lô 2	97	95	90	81	92	91	95	96	84	72
Hạng	*20*	*17*	*14*	*10*	*16*	*15*	*17*	*19*	*12*	*(5)*

Ở đây có 2 hạng 5 cho số 72 theo thứ tự 5, 6; 3 hạng 7 cho số 80 theo thứ tự 7, 8, 9, vì thế mỗi số 72 có thứ hạng mới là 5,5, tức là (5 + 6)/2 và mỗi số 80 có thứ hạng mới là 8, tức là (7 + 8 + 9)/3. Việc xếp hạng đúng khi: $\mathrm{R}_1 + \mathrm{R}_2 = \dfrac{n(n+1)}{2}$; R_1 là tổng thứ hạng của lô 1 và R_2 là tổng thứ hạng của lô 2.

Kết quả xếp hạng lại như sau:

Lô 1	72	87	71	70	80	67	80	80	82	66	
Hạng	*5,5*	*13*	*4*	*3*	*8*	*2*	*8*	*8*	*11*	*1*	R_1=63,5
Lô 2	97	95	90	81	92	91	95	96	84	72	
Hạng	*20*	*17*	*14*	*10*	*16*	*15*	*17*	*19*	*12*	*5,5*	R_2=146,5

Kiểm tra: $\mathrm{R}_1 + \mathrm{R}_2 = \dfrac{n(n+1)}{2} = 63{,}5 + 146{,}5$

$$= \frac{20(20+1)}{2} = 210$$

Bước 2: Kiểm tra và đánh giá kết quả

Người ta đã chứng minh được rằng khi n_1 và n_2 ($n_1 + n_2 = n$) các phân phối U_1 (cho tổng thể của mẫu 1) và U_2 (cho tổng thể của mẫu 2) tiệm cận phân phối chuẩn với:

$$E(U) = \frac{n_1 n_2}{2} \text{ và } V(U) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

Khi đó sử dụng phép thử sau để đánh giá:

$$U_{TN} = \frac{\left| U_1 - \frac{n_1 n_2}{2} \right|}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \quad (3 - 6)$$

với $$U_1 = n_1 n_2 + \frac{n_1 (n_1 + 1)}{2} - R_1 \quad (3 - 7)$$

và $$U_2 = n_1 n_2 + \frac{n_2 (n_2 + 1)}{2} - R_2$$

hoặc U_2 có thể tính:

$$U_2 = n_1 n_2 - U_1 \quad (3 - 7')$$

và $$U_{TN}(U_1) = - U_{TN}(U_2).$$

Nếu $U_{TN} \geq 1,96$ thì $U_1 \neq U_2$, ngược lại $U_{TN} < 1,96$ thì $U_1 \approx U_2$.

Ở ví dụ này: $R_1 = 63,5$; $R_2 = 146,5$. Thay vào công thức (3 - 7) và (3 - 7') ta có: $U_1 = 91,5$ và $U_2 = 8,5$.

Theo công thức (3 - 6) ta có:

$$U_{TN} = \frac{\left| 91,5 - \frac{10 \times 10}{2} \right|}{\sqrt{\frac{10 \times 10 (10 + 10 + 1)}{12}}} = 3,14 > 1,96$$

$U_1 \neq U_2$ cho thấy hai lô này có đất khác nhau.

Một cách tương tự, kết quả kiểm tra U_1 (lô 1) và U_3 (lô 3):

$R_1 = 104$; $R_3 = 106$; $U_1 = 51{,}0$; $U_3 = 49{,}0$; $U_{TN} = 0{,}08$

và giữa U_2 (lô 2) và U_3 (lô 3):

$R_2 = 143{,}5$; $R_3 = 66{,}5$; $U_1 = 11{,}5$; $U_3 = 88{,}5$; $U_{TN} = 2{,}91$

Với các kết quả này thì đất lô 1 và lô 3 đồng nhất và khác với lô 2 về độ tốt xấu.

Bây giờ hãy đánh giá độ đồng đều của các lô qua phép so sánh phương sai (phương pháp tham số) theo số liệu sau:

Lô	$\overline{x}$	S^2	S	CV(%)
1	75,5	51,17	7,15	9,5
2	89,3	64,01	8,00	9,0
3	75,6	57,82	7,60	10,1

So sánh phương sai lô 2 với lô 1:

$$F_{2/1} = \frac{64{,}01}{51{,}17} = 1{,}25 < F_{0,05} = 3{,}18$$

Hiển nhiên là $F_{3/1}$ và $F_{2/3}$ đều nhỏ hơn 3,18.

Điều đó cho thấy, tuy lô 2 đất có tốt hơn hai lô còn lại nhưng độ đồng đều trong từng lô của cả ba lô như nhau.

Có thể so sánh trung bình lô 1 và 2 theo công thức (3 - 2):

$$T_{TN} = \frac{|75{,}5 - 89{,}3|}{\sqrt{\frac{(10-1)51{,}17 + (10-1)64{,}01}{10+10-2}\left(\frac{1}{10} + \frac{1}{10}\right)}}$$

$= 4{,}07 > t_{0,05}$ với 18 độ tự do là 2,10.

Rõ ràng trung bình lô 2 khác với lô 1 và 3.

Mặc dù độ tốt xấu có khác nhau nhưng cả ba lô đều có hệ số biến động có thể chấp nhận được (CV $\leq$ 10%).

Như vậy, ở mỗi lô có thể bố trí mỗi lần nhắc lại cho thí nghiệm đồng ruộng.

Tiêu chuẩn U của Siegel và Tukey

Để kiểm tra tính đồng nhất của hai mẫu (hai lô) từ hai tổng thể có nguồn gốc khác nhau, Siegel và Tukey cũng xếp hạng chung cho cả hai lô hoàn toàn giống Mann và Whitney nhưng ký hiệu R_1 cho lô có dung lượng mẫu nhỏ, còn R_2 cho lô có dung lượng mẫu lớn. Nếu n_1 và n_2 đều > 9 hoặc $n_1 > 2$ và $n_2 > 20$ thì việc kiểm tra tính đồng nhất của hai lô được thực hiện theo tiêu chuẩn U sau đây:

$$U_{TN} = \frac{\left|2R_1 - n_1(n_1 + n_2 + 1) + 1\right|}{\sqrt{n_1(n_1 + n_2 + 1).\frac{n_2}{3}}} \qquad (3 - 8)$$

Nếu $U_{TN} \geq 1{,}96$ thì $U_1 \neq U_2$, ngược lại $U_{TN} < 1{,}96$ thì $U_1 \approx U_2$.

Theo ví dụ trên:

$$U_{TN} = \frac{\left|2\times 63{,}6 - 10(10+10+1)+1\right|}{\sqrt{10(10+10+1).\frac{10}{3}}} = 3{,}10$$

Như vậy: $U_1 \neq U_2$.

Trong trường hợp $n_1 = n_2$ thì có thể thay R_1 hoặc R_2 vào công thức (3 - 8) và kết quả tương đương nhau. Ở đây, nếu thay $R_1 = 63{,}5$ bằng $R_2 = 146.5$, $U_{TN} = 3{,}17$.

- ***So sánh các trung bình nhiều mẫu độc lập***

Tiêu chuẩn H của Kruskal và Wallis

Để kiểm tra tính đồng nhất của nhiều mẫu từ nhiều tổng thể có nguồn gốc khác nhau, việc xây dựng tiêu H cũng thực hiện sau khi xếp hạng chung cho tập hợp tất cả các số liệu các mẫu. Phương pháp xếp hạng hoàn toàn giống Mann và Whitney.

Người ta đã chứng minh được rằng một đại lượng ngẫu nhiên gồm n biến số được xếp hạng từ 1 đến n, tập hợp từ k mẫu, phân phối theo quy luật "khi bình phương" với $k - 1$ bậc tự do:

$$\mathrm{H} = \frac{12}{n(n+1))}\sum_{k=1}^{k}\frac{\mathrm{R}_\mathrm{i}^2}{n_\mathrm{i}} - 3(n+1) \qquad (3 - 9)$$

Trong đó: $n = \Sigma n_\mathrm{i}$ là tổng dung lượng mẫu

$\mathrm{R_i}$ là tổng các hạng trong mẫu i $(i = \overline{1,k})$

k là số mẫu quan sát

Nếu $\mathrm{H} \geq \chi^2_{0,05}$ thì các mẫu không thuần nhất.

Nếu $\mathrm{H} < \chi^2_{0,05}$ thì các mẫu thuần nhất, tức là các mẫu xuất phát từ một tổng thể.

Ví dụ: Từ kết quả phân tích hàm lượng mùn (%) trong 3 lô thí nghiệm ở bảng ở trang sau, hãy so sánh tính đồng nhất của khu đất.

Lô 1:	12,3	12,5	13,1	13,6	13,8	14,2	14,7	14,9	15,3	n_1=9
Lô 2:	12,8	13,9	14,2	14,7	15,3	15,3	15,8	16,8	17,4	n_2=9
Lô 3:	13,9	14,9	15,7	15,7	15,8	16,5	16,8	17,3	18,5	n_3=9

Áp dụng phương pháp xếp hạng như trên ta có:

Lô 1:	12,3	12,5	13,1	13,6	13,8	14,2	14,7	14,9	15,3	n_1=9
Hạng	*1*	*2*	*4*	*5*	*6*	*9,5*	*11,5*	*13,5*	*16*	R_1=68,5
Lô 2:	12,8	13,9	14,2	14,7	15,3	15,3	15,8	16,8	17,4	n_2=9
Hạng	*3*	*7,5*	*9,5*	*11,5*	*16*	*16*	*20,5*	*23*	*26*	R_2=133
Lô 3:	13,9	14,9	15,7	15,7	15,8	16,5	16,8	17,3	18,5	n_3=9
Hạng	*7,5*	*13,5*	*18,5*	*18,5*	*20,5*	*22*	*24*	*25*	*27*	R_3=176,5

$$\sum_{i=1}^{k} R_i = 68,5 + 133 + 186,5 = 378$$

Kiểm tra lại kết quả xếp hạng:

$$\sum_{i=1}^{k} R_i = \frac{27(27+1)}{2} = 378$$

Như vậy việc xếp hạng là đúng.

Theo công thức (3 - 9) ta có: H = 10,34, $\chi^2_{0,05}$ = 5,99.

Như vậy, không có sự đồng nhất của 3 lô đất thí nghiệm.

3.1.2.2. So sánh trung bình hai mẫu phụ thuộc - Tiêu chuẩn tổng hạng theo dấu của Wilcoxon

Nếu dung lượng mẫu không đủ lớn, các tổng thể lại không theo luật phân phối chuẩn thì việc so sánh được thực hiện bằng phép nghiệm phi tham số bằng tổng hạng của Wilcoxon sau đây trên cơ sở các giả thiết:

- Các đại lượng ngẫu nhiên X_1 và X_2 của hai tổng thể nghiên cứu có thể phân phối theo một quy luật bất kỳ.

- Hai mẫu ngẫu nhiên rút ra từ hai tổng thể phải độc lập và có dung lượng tùy ý.

Khi đó giả thuyết kiểm định hai trung bình sẽ là:

H_0: Hai tổng thể có kỳ vọng bằng nhau với độ tin cậy $1 - \alpha$;

H_1: Hai tổng thể có kỳ vọng khác nhau với độ tin cậy $1 - \alpha$ (hoặc là $X_1 > X_2$ hoặc là $X_2 > X_1$).

Để đánh giá hai trung bình mẫu, trước hết phải xếp hạng từ nhỏ đến lớn các số đo của cả hai mẫu theo thứ tự 1, 2, ..., *n* như phương pháp xếp hạng đã nêu trong mục

2.2.1 trên đây. n là tổng dung lượng của hai mẫu: n_1 của X_1 và n_2 của X_2 ($n = n_1 + n_2$).

Nếu H_0 đúng thì tổng hạng (ký hiệu là T) của mẫu sẽ tỷ lệ thuận với dung lượng mẫu (cũng tức là nếu $n_1 = n_2$ thì T của mẫu 1 bằng T của mẫu 2) và T sẽ có kỳ vọng và phương sai như sau:

$$\mu_T = \frac{n_1(n_1 + n_2 + 1)}{2} \qquad (3 - 10)$$

và
$$\sigma_T^2 = \frac{n_1 n_2}{12}(n_1 + n_2 + 1) \qquad (3 - 11)$$

Trong kiểm định tổng hạng của Wilcoxon, hai đại lượng ngẫu nhiên X_1 và X_2 phân phối liên tục và việc xếp hạng sẽ không có thứ hạng trùng nhau. Thực tế không thể có hai cá thể/phần tử hoàn toàn giống nhau, nhưng khi quan trắc do lấy độ chính xác của đơn vị đo nên có sự trùng nhau. Lúc đó thứ hạng sẽ được chia đều (như đã thực hiện ở 3.1.2.1 trên đây). Giá trị phương sai thực nghiệm sẽ được tính:

$$\sigma_T^2 = \frac{n_1 n_2}{12}[n_1 + n_2 + 1] - \frac{\sum_j t_j(t_j^2 - 1)}{(n_1 + n_2)(n_1 + n_2 - 1)} \qquad (3 - 12)$$

Trong đó t_j là tần số của hạng ghép nhóm thứ j. Hiển nhiên, khi không có hạng ghép thì công thức (3 - 12) lại trở thành công thức (3 - 11).

Nếu $n_1 \leq 10$ và $n_2 \leq 10$

Sau khi tính được tổng hạng của mỗi mẫu, tra bảng giá trị tổng hạng Wilcoxon để tìm các giá trị tới hạn T_L và T_u và xác định:

- Nếu kỳ vọng của hai tổng thể giống nhau thì $T < T_u$ hoặc $T > T_L$.

- Nếu kỳ vọng của hai tổng thể khác nhau thì $T > T_u$ hoặc $T < T_L$.

Nếu $n_1 > 10$ và $n_2 > 10$

Trong kiểm định phi tham số tổng hạng, Wilcoxon đã chứng minh được rằng khi cả n_1 và n_2 đều lớn hơn 10 thì phân phối T sẽ tiệm cận với phân phối chuẩn. Khi đó việc so sánh trung bình của hai mẫu theo tiêu chuẩn U:

$$U_{TN} = \frac{T - \mu_T}{\sigma_T} \qquad (3 - 13)$$

Nếu $U_{TN} \geq u_{\alpha/2}$ thì kết luận rằng $\overline{X}_1$ khác $\overline{X}_2$ ở độ tin cậy 1 - α. Và, nếu $U_{TN} < u_{\alpha/2}$ thì kết luận rằng $\overline{X}_1$ không khác với $\overline{X}_2$ ở độ tin cậy 1 - α.

Ví dụ: Để đánh giá một giống bắp mới trong sản xuất một công ty đã hợp đồng với 15 gia đình trong vùng, yêu cầu mỗi gia đình chọn một đám đất đồng đều và gieo 2 giống: giống đang sản xuất phổ biến trong vùng làm đối chứng (ĐC) và giống mới do công ty mang xuống (GM). Họ còn yêu cầu việc thực hiện các công việc gieo trồng, chăm sóc phải thự hiện như nhau cho cả hai giống. Kết quả năng suất và phân hạng (trong ngoặc) được ghi trong Bảng 3.2 sau đây. Hãy so sánh năng suất 2 giống này ?

Giải:

Ở đây, $n_1 = n_2 = 15$; $n = n_1 + n_2 = 30$; T = 465

Thay số liệu vào (3-10) và (3-11) ta được:

$$\mu_T = (15 \times 31)/2 = 232,5$$

Bảng 3.2: Năng suất và phân hạng hai giống bắp

Gia đình	Năng suất (tạ/ha)		Gia đình	Năng suất (tạ/ha)	
	ĐC	GM		ĐC	GM
1	68,5 (*5,5*)	77,5 (*18*)	9	69,0 (*7*)	71,8 (*12*)
2	73,3 (*15*)	83,0 (*22*)	10	72,0 (*13*)	74,8 (*16*)
3	57,5 (*1*)	62,5 (*2*)	11	79,0 (*20*)	88,0 (*29*)
4	81,4 (*21*)	87,6 (*28*)	12	69,5 (*9*)	72,2 (*14*)
5	77,8 (*19*)	69,3 (*8*)	13	87,3 (*27*)	83,8 (*24*)
6	85,5 (*25*)	93,5 (*30*)	14	70,2 (*10*)	63,8 (*3*)
7	68,5 (*5,5*)	71,5 (*11*)	15	68,0 (*4*)	76,5 (*17*)
8	87,2 (*26*)	83,5 (*23*)		T = 465	

$$\sigma_T^2 = \frac{225}{12}\times 31 - \frac{6}{30\times 29} = 581{,}24 \text{ và } \sigma_T = 24{,}11$$

Từ đó: $U_{TN} = \dfrac{T - \mu_T}{\sigma_T} = \dfrac{465 - 232{,}5}{24{,}11} = 9{,}64$

Với $\alpha = 0{,}05$, $u_{0,025} = 1{,}96$ và $\sigma = 0{,}01$, $u_{0,005} = 2{,}58$

Như vậy: giống GM > ĐC với độ tin cậy 99%.

3.1.2.3. So sánh các trung bình nhiều mẫu phụ thuộc - Tiêu chuẩn Friedman

Việc kiểm tra tính đồng nhất của các mẫu bằng phép nghiệm χ^2.

- Trước hết, xếp hạng thứ tự 1, 2, 3, ... giữa các phương án trong từng nơi (hoặc từng thời điểm), mỗi nơi một hàng.

- Sau đó tính tổng số hạng cho từng phương án theo từng cột.

- Cuối cùng là kiểm tra sự giống hay khác nhau giữa các phương án theo tiêu chuẩn χ^2:

$$\chi^2_{TN} = \frac{12}{ab(a+1)}\sum R_i^2 - 3b(a+1) \qquad (3 - 14)$$

Trong đó: *a* là số phương án

b là số nơi (số thời điểm)

R_i là tổng hạng của từng phương án $(i = \overline{1,a})$.

Nếu $\chi^2_{TN} \geq \chi^2_{0,05}$ với *a* – 1 độ tự do thì các phương án cho kết quả khác nhau, còn $\chi^2_{TN} < \chi^2_{0,05}$ thì các phương án khác nhau không đủ tin cậy.

Ví dụ 1: Trong một thử nghiệm 5 giống đậu xanh tại 3 xã Phước Tiến, Phước Thắng và Phước Đại, huyện Bác Ái, tỉnh Ninh Thuận vụ Hè Thu năm 2009, năng suất các giống tại các điểm thí nghiệm được ghi nhận trong Bảng 3.3. Câu hỏi đặt ra: năng suất của các xã này khác nhau không ?

Bảng 3.3: Năng suất (tạ/ha) của 5 giống đậu xanh và kết quả xếp hạng từng xã cho mỗi giống

Giống\Xã	Phước Tiến /hạng		Phước Thắng /hạng		Phước Đại /hạng	
NP 305	13,9 b	*3*	12,3 c	*1*	13,8 bc	*2*
ĐX208	17,8 a	*3*	16,4 a	*2*	15,6 ab	*1*
HL 89-E3	16,2 a	*3*	15,1 b	*1*	16,1 a	*2*
V 99-1	12,3 b	*2*	12,2 c	*1*	12,5 c	*3*
Agredec-01	13,6 b	*2*	12,9 c	*1*	14,6 ab	*3*
P	< 0,05		< 0,05		< 0,05	
ΣR_i		13		6		11

Giải:

Ở đây: a = 3, b = 5, ΣR_1 = 13, ΣR_2 = 6, ΣR_3 = 11

Thay giá trị vào công thức (3-14) ta được:

$$\chi^2_{TN} = \frac{12}{3\times5(3+1)}\left(13^2+6^2+11^2\right)-3\times5\times4$$

$= 5{,}20 < \chi^{2(2)}_{0,05} = 6{,}0$.Như vậy năng suất đậu xanh của 3 xã này không có sự khác nhau.

Ví dụ 2: Kết quả cân khối lượng 100 cây mầm một giống cải bẹ xanh *Brassica juncea* L. được gieo trên 4 loại giá thể khác nhau (TN1, NT2, NT3 và TN4) tại Bảo Lộc, Lâm Đồng được ghi ở bảng 3.4. Theo kết quả này, có sự khác nhau hay không về khối lượng cây mầm trên các loại giá thể?

Bảng 3.4: Khối lượng trung bình 100 cây mầm (g) của 4 nghiệm thức (NT) trên các lần lặp lại.

Lặp lại	Khối lượng trung bình 100 cây mầm (g)			
	NT1	NT2	NT3	NT4
I	4,5 (*1*)	5,5 (*4*)	5,2 (*2,5*)	5,2 (*2,5*)
II	4,7 (*1*)	5,3 (*4*)	4,8 (*2*)	5,0 (*3*)
III	4,7 (*1*)	5,1 (*4*)	5,0 (*3*)	4,8 (*2*)
ΣR_i	3	12	7,5	7,5

Ghi chú: Số liệu trong dấu ( ) là hạng từ nhỏ đến lớn của các NT cho mỗi lần lặp lại.

Giải:

Ở đây: a = 4, b = 3, ΣR_1 = 3, ΣR_2 = 12, ΣR_3 = ΣR_5 = 7,5

Thay giá trị vào công thức (3-14) ta được:

$$\chi^2_{TN} = \frac{12}{4\times3(4+1)}\left(3^2+12^2+2\times7{,}5^2\right)-3\times3\times5$$

$$= 8{,}1 > \chi^{2(3)}_{0,05} = 7{,}8.$$

Như vậy, khối lượng cây mầm trên các loại giá thể có khác nhau.

Để biết được sự tốt xấu của 4 loại giá thể ta chỉ cần kiểm tra sự khác nhau giữa NT1 và NT3, NT2 và NT3 theo tiêu chuẩn U của Mann và Whitney hay của Siegel và Tukey đã nêu trên đây.

Kết quả kiểm tra theo tiêu chuẩn U của Mann và Whitney cho biết:

U_1 và U_3: $R_1 = 6$; $R_3 = 15$; $U_1 = 9$; $U_3 = 0$; $U_{TN} = 1{,}96$

U_2 và U_3: $R_2 = 14$; $R_3 = 7$; $U_1 = 1$; $U_3 = 8$; $U_{TN} = 1{,}53 < 1{,}96$

Như vậy $U_1 \neq U_3$ và $U_2 \approx U_3$

Từ đây, có thể nói NT2 là tốt nhất, NT1 là xấu nhất, NT2 hơn NT3 và NT4 không đủ mức tin cậy 95%.

3.2. SO SÁNH HAI PHƯƠNG SAI VÀ MỞ RỘNG

3.2.1. Cơ sở lý luận

Như đã đề cập ở chương 1, phương sai là tham số đặc trưng cho độ phân tán của đại lượng ngẫu nhiên, nói cách khác là sự khác biệt giữa các giá trị x_i của trong một tập hợp số liệu quan sát so với số trung bình. Nếu các x_i là số đo của một tổng thể thuần nhất thì phương sai phản ánh độ đồng đều của tổng thể. Việc so sánh hai phương sai của hai tổng thể loại này là so sánh sự đồng nhất của tổng thể này với sự đồng nhất của tổng thể khác, ví dụ như giữa giống này với giống khác. Nếu các x_i là các giá trị của một tổng thể không thuần nhất này (ví dụ như giữa các giống), còn các y_i là giá trị của một tổng thể không thuần nhất khác (ví dụ như giữa các mức bón), thì việc so sánh phương sai của hai tổng thể này cho biết sự khác nhau giữa các mức của yếu tố này (giữa các giống) giống hay là khác với sự khác nhau giữa các mức của yếu tố khác (giữa các mức phân), tức là yếu tố nào khác nhau nhiều hơn. Nếu so sánh phương sai gây ra do sự khác nhau giữa các nghiệm thức với một loại phương sai khác do sai số (ngẫu nhiên) gây ra, thì việc so sánh đó cho biết giữa các nghiệm thức có khác nhau hay không. Trong trường hợp này, nếu như phương sai do các nghiệm thức khác với phương sai ngẫu nhiên không đủ tin cậy ở mức nào đó, thì ta nói rằng sự khác nhau giữa các nghiệm thức là do sai số (hay, cũng như sai số), thực sự là chúng không khác biệt nhau. Đây là phép phân tích phương sai (ANOVA) các nghiệm thức trong các loại thí nghiệm, sẽ được nói tới ở phần 2.

Việc so sánh phương sai σ_1^2 và σ_2^2 của hai tổng thể có hai trung bình tổng thể μ_1 và μ_2 không thể đo trực tiếp bằng khoảng hiệu số $\sigma_1^2 - \sigma_2^2$ như so sánh hai trung bình, bởi vì: phương sai đặc trưng cho độ phân tán, không là đặc trưng về vị trí, có đơn vị tính là bình phương của số đo theo dấu hiệu nào đó.

Người ta đã dùng phép tỷ số để so sánh hai phương sai mẫu, từ đó suy đoán cho tổng thể. Nếu hai phương sai bằng nhau, thương số sẽ bằng 1. Tuy nhiên do trong ước lượng, các phương sai tổng thể nằm trong những khoảng khác nhau, tỷ lệ giữa hai phương sai tuân theo luật phân phối Fisher – Snedecor $F(n_1, n_2)$ nên được trắc nghiệm bằng tiêu chuẩn F.

3.2.2. So sánh hai phương sai

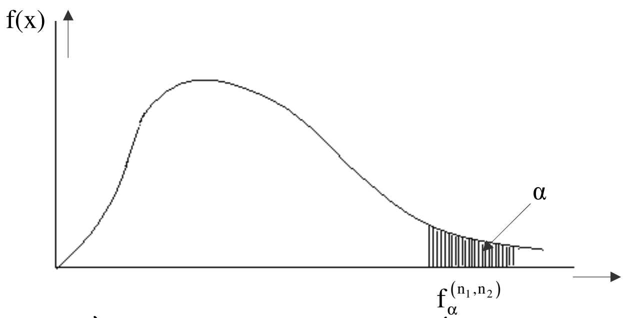
3.2.2.1. Luật phân phối Fisher – Snedecor F(n_1, n_2)

Đại lượng ngẫu nhiên liên tục F phân phối theo luật Fisher – Snedecor với n_1 và n_2 bậc tự do, hàm mật độ xác suất của nó được xác định bằng công thức :

$$f(x)=\begin{cases} 0 & \text{với } x \le 0 \\ \dfrac{\Gamma\left(\dfrac{n_1+n_2}{2}\right) n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}}}{\Gamma\left(\dfrac{n_1}{2}\right)\Gamma\left(\dfrac{n_2}{2}\right)} \cdot \dfrac{x^{\frac{n_1-n_2}{2}}}{(n_2+n_1x)^{\frac{n_1+n_2}{2}}} & \text{với } x > 0 \end{cases}$$

trong đó $\Gamma(n)$ là hàm Gamma.

Và, đồ thị hàm mật độ xác suất có dạng như hình 3.1.



Hình 3.1: Đồ thị hàm mật độ xác suất theo luật phân phối Fisher – Snedecor

Người ta đã chứng minh được rằng nếu đại lượng ngẫu nhiên F phân phối theo luật phân phối Fisher – Snedecor với n_1 và n_2 bậc tự do thì kỳ vọng:

$$E(F)=\frac{n_2}{n_2-2}$$

và phương sai $V(F) = \dfrac{2n_2^2\left(n_1 + n_2^2 - 2\right)}{n_1\left(n_2 - 2\right)^2\left(n_2 - 4\right)}$.

3.2.2.2. Phép nghiệm Fisher – Snedecor F(n_1 -1, n_2 - 1)

Theo luật phân phối Fisher – Snedecor, với dung lượng mẫu của hai đại lượng ngẫu nhiên n_1 và n_2, số bậc tự do là (n_1 – 1) và (n_2 – 1), thỏa mãn điều kiện:

$$P(F \geq f_{\alpha}^{[(n_1-1)(n_2-1)]}) = \alpha$$

và có tính chất: $f_{\alpha}^{[(n_1-1)(n_2-1)]} = \dfrac{1}{f_{1-\alpha}^{[(n_1-1)(n_2-1)]}}$.

Với độ tin cậy 1 – α cho trước, có thể tìm được giá trị tới hạn Fisher – Snedecor f_α với (n_1 – 1) và (n_2 – 1) bậc tự do để kiểm định sự khác nhau của hai phương sai hai đại lượng ngẫu nhiên:

$$F_{TN} = \frac{S_1^2}{S_2^2} \qquad (S_1^2 > S_2^2) \qquad (3 - 15)$$

Nếu $F_{TN} \geq f_\alpha$ thì $S_1^2 > S_2^2$ ở độ tin cậy 1 - α, còn $F_{TN} < f_\alpha$ thì $S_1^2 \approx S_2^2$.

Ví dụ: Hãy so sánh phương sai F$_1$ (S_1^2 = 387,9, n_1 = 45) và phương sai F$_2$ (S_2^2 =768,61, n_2 = 110) của tổ hợp bông lai C92-52/C118A với số liệu trong ví dụ ở mục 3.2.1.4.

Giải: Ở đây giá trị $S_2^2 > S_1^2$, công thức (3 - 15) trở thành: $F_{TN} = \dfrac{S_2^2}{S_1^2}$. Thay S_1^2 và S_2^2 vào ta có: $F_{TN} = \dfrac{768,61}{387,90} =$ 1,98. Giá trị $f_{0,05}$ = 1,55; $f_{0,01}$ = 1,88 và $f_{0,001}$ = 2,34 (giá trị $f_{bảng}$ được tra trên phần mềm Excel với độ tự do tử số 109 và mẫu số 44).

Như vậy, phương sai F_2 lớn hơn phương sai F_1 với độ tin cậy 99%.

Đó là điều đương nhiên vì phương sai F_1 do ngẫu nhiên (môi trường đất) gây ra. Nếu đất hoàn toàn đồng nhất thì $S_1^2 = 0$ vì các cá thể F1 có cùng kiểu gen, còn phương sai S_2^2 của F_2 là vừa do sự phân ly về kiểu gen vừa do môi trường đất gây ra. Chênh lệch phương sai do sự khác nhau kiểu gen gây ra là: $S_2^2 - S_1^2$ = 768,61 – 387,90 = 380,71 và hệ số di truyền năng suất trong F_2 là: H^2 = 380,71/768,61 = 0,495 (49,5%). Đây là hệ số di truyền nghĩa rộng.

3.2.3. Đánh giá sự đồng nhất các phương sai của nhiều tổng thể

3.2.3.1. Khi dung lượng mẫu rút ra từ các tổng thể khác nhau

Nếu dung lượng mẫu của k phương sai mẫu S_1^2, S_2^2, $...S_k^2$ là n_1, n_2, ... n_k $(i = \overline{1,k})$, $n_1 \neq n_2 \neq ... \neq n_k$, $h_i = (n_i - 1)$, $h = \Sigma h_i$ và $\overline{S}^2$ là trung bình số học của k phương sai:

$$\overline{S}^2 = \frac{\sum h_i S_i^2}{h}$$

Tiêu chuẩn Bartlett dùng để kiểm định sự đồng nhất của các phương sai là:

$$B = \frac{V}{C}, \text{ trong đó:}$$

$$V = 2,303\left[h \lg \overline{S}^2 - \sum h_i \lg S_i^2\right]$$

$$C = 1 + \frac{1}{3(k-1)}\left[\sum \frac{1}{h_i} - \frac{1}{h}\right]$$

Các phương sai mẫu đồng nhất khi $B < \chi^{2(k-1)}_{0,05}$, ngược lại khi $B \geq \chi^{2(k-1)}_{0,05}$ thì các phương sai không đồng nhất.

***Ví dụ*:** Kết quả điều tra biến động năng suất cá thể (g/cây) của 5 giống bông thuần với n_1 = 26, n_2 = 32, n_3 = 29, n_4 = 30, n_5 = 19 và các phương sai tương ứng S_1^2 = 623,8, S_2^2 = 420,4, S_3^2 = 630,6, S_4^2 = 461,0 và S_5^2 = 586,6. Hãy kiểm định tính đồng nhất của các phương sai với độ tin cậy 95%.

Để tính B ta lập bảng sau

Mẫu	h_i	S_i^2	$h_i S_i^2$	$\lg S_i^2$	$h_i \lg S_i^2$	$1/h_i$
1	25	623,8	15.595,0	2,795	69,9	0,040
2	31	420,4	13.032,4	2,624	81,3	0,032
3	28	630,6	17.656,8	2,800	78,4	0,036
4	29	461,0	13.369,0	2,664	77,2	0,034
5	18	586,6	10.558,8	2,768	49,8	0,056
Σ	131		70.212,0		356,7	0,198

Từ đó:

$$\overline{S}^2 = \frac{\sum h_i S_i^2}{h} = \frac{70.212,0}{131} = 536,0$$

$$\lg \overline{S}^2 = 2,729$$

$$V = 2,303\left[h \lg \overline{S}^2 - \sum h_i \lg S_i^2\right]$$

$$= 2,303[(131)(2,729) - 356,7] = 1,982$$

$$C = 1 + \frac{1}{3(k-1)}\left[\sum \frac{1}{h_i} - \frac{1}{h}\right]$$

$$= 1 + \frac{1}{3(5-1)}\left[0,198 - \frac{1}{131}\right] = 1,016$$

$$B = \frac{V}{C} = \frac{1,982}{1,016} = 1,898 < \chi^{2(4)}_{0,05} = 9,488$$

Như vậy, các phương sai được xem là đồng nhất, tức là các giống đều thuần chủng.

3.2.3.2. Khi dung lượng mẫu rút ra từ các tổng thể bằng nhau

Trong trường hợp này, có thể dùng tiêu chuẩn Bartlett để kiểm tra. Tuy nhiên do phân phốixác suất theo tiêu chuẩn Bartlett chỉ là xấp xỉ nên kém chính xác. Khi kích thước mẫu bằng nhau, ta dùng tiêu chuẩn Cochran (G).

$$G = \frac{S^2_{max}}{S^2_1 + S^2_2 + \ldots + S^2_k}$$

Các phương sai mẫu đồng nhất khi $G < g_\alpha^{(n-1,k)}$, ngược lại khi $G \geq g_\alpha^{(n-1,k)}$ thì các phương sai không đồng nhất.

$g_\alpha^{(n-1,k)}$ là giá trị tới hạn của phân phối Cochran được tra trong bảng phụ lục 7.

Ví dụ: Cũng với các giống trên, nếu lấy dung lượng mẫu bằng nhau và bằng 19, các phương sai mẫu sẽ là 653,2; 466,1; 671,3; 581,4 và 586,6. Hãy kiểm định tính đồng nhất của các phương sai với độ tin cậy 95% và ước lượng phương sai tổng thể nếu các phương sai đồng nhất.

Giá trị Cochran từ mẫu quan sát:

$$G = \frac{671,3}{653,2 + 466,1 + 671,3 + 581,4 + 586,6} = 0,227$$

Với α = 0,05, số bậc tự do là 19 – 1 = 18, số lượng mẫu là 5, giá trị tới hạn tra được là $g_{\alpha}^{(n-1,k)} = g_{0,05}^{(18,5)} = 0,3645$

$G < g_{0,05}^{(18,5)}$ cho thấy các phương sai là đồng nhất, độ đồng đều các giống là như nhau.

Phương sai tổng thể được ước lượng:

$$\sigma^2 = \overline{S}^2 = \frac{653,2 + 466,1 + 671,3 + 581,4 + 586,6}{5}$$

$$= 591,7$$

3.3. ĐÁNH GIÁ TÍNH ĐỘC LẬP CỦA CÁC DẤU HIỆU ĐỊNH TÍNH

Người ta sử dụng tiêu chuẩn khi bình phương (χ^2) để xác định mối quan hệ giữa hai dấu hiệu định tính.

Giả sử có một tổng thể có hai dấu hiệu định tính A và B. Giả thiết H_0 : A và B độc lập và H_1: A và B phụ thuộc.

Để kiểm tra các giả thiết này, từ tổng thể có dung lượng mẫu n, lập bảng trình bày các đặc trưng A, B và tần số tương ứng:

A \ B	B_1	B_2	...	B_j	Tổng A_i
A_1	n_{11}	n_{12}	...	n_{1j}	$n_{1.}$
A_2	n_{21}	n_{22}	...	n_{2j}	$n_{2.}$
...	...	...	...	...	...
A_i	n_{i1}	n_{i2}	...	n_{ij}	$n_{i.}$
Tổng B_j	$n_{.1}$	$n_{.2}$	...	$n_{.j}$	$\Sigma n_{ij} = n$

Trong bảng:

n là dung lượng mẫu.

n_{ij} là là tần số ứng với các mức độ của A_i $(i = \overline{1, i})$ và B_j $(j = \overline{1, j})$.

$n_{i.}$ là là tần số ứng với các mức độ của dấu hiệu A.

$n_{.j}$ là là tần số ứng với các mức độ của dấu hiệu B.

Tính độc lập của hai dấu hiệu A và B được kiểm tra theo tiêu chuẩn khi bình phương (χ^2):

$$\chi^2_{TN} = n\left(\sum_{i=1}^{i}\sum_{j=1}^{j}\frac{n_{ij}^2}{n_{i.}n_{.j}} - 1\right) \qquad (3 - 16)$$

Nếu $\chi^2_{TN} \geq \chi^2_{\alpha}$ với $(i - 1)(j - 1)$ độ tự do thì bác bỏ H_0 và chấp nhận H_1, còn $\chi^2_{TN} < \chi^2_{\alpha}$ thì chấp nhận H_0 và bác bỏ H_1 với độ tin cậy $1 - \alpha$.

Ví dụ: Kết quả điều tra mức độ lông của lá bông và mức độ kháng rầy xanh được ghi ở Bảng 3.5. Vậy, tính có lông có quan hệ với mức độ kháng rầy không ?

Bảng 3.5: Kết quả điều tra tính kháng rầy 50 giống bông

Mức độ lông của lá	KTB	K	RB	Tổng, A_i
Ít	3	0	0	3
Vừa	7	3	0	10
Nhiều	5	8	5	18
Rất nhiêu	3	5	11	19
Tổng, B_j	18	16	16	$\Sigma n_{ij} = 50$

Ghi chú : KTB – kháng trung bình; K – kháng ; RK – rất kháng.

Giải:

Ở đây: i = 4; j = 3; n = 50; $n_{i.}$ = 3, 10, 18 và 19; $n_{.j}$ = 18, 16 và 16.

Thay giá trị vào công thức (3 - 16) ta được:

$$\chi^2_{TN} = 50 \times \{[3^2/(18 \times 3) + 7^2/(18 \times 10) + 5^2/(18 \times 18) + ...$$

$$+ 5^2/(16 \times 18) + 11^2/(16 \times 19)] - 1\} = 19{,}40$$

$$\chi^2_{TN} = 19{,}40 > \chi^{2(6)}_{0,01} = 16{,}81.$$

Như vậy, tính có lông có quan hệ chặt chẽ với mức độ kháng rầy với độ tin cậy 99%.

Để giải bài này có thể sử dụng phương pháp sau:

Trước hết tính tần số lý thuyết n'_{ij} cho các n_{ij}: $n'_{ij} = \frac{n_{i.}n_{.j}}{n}$. Ví dụ cho n'_{32} là: (16 × 18)/50 = 5,76 trong khi đó n_{32} = 8. Sau khi tính xong n'_{ij}, kiểm tra sự khác nhau giữa tẩn số lý thuyết với tần số thực nghiệm theo công thức:

$$\chi^2_{TN} = \sum \frac{\left(n_{ij} - n'_{ij}\right)^2}{n'_{ij}} \qquad (3 - 17)$$

Kết quả tính các tần số n'_{ij} được ghi ở bảng sau (số trong ngoặc):

Độ lông	KTB	K	RK	Tổng, A_i
Ít	3 (1,08)	0 (0,96)	0 (0,96)	3
Vừa	7 (3,60)	3 (3,20)	0 (3,20)	10
Nhiều	5 (6,48)	8 (5,76)	5 (5,76)	18
Rất nhiêu	3 (6,84)	5 (6,08)	11 (6,08)	19
Tổng, B_j	18	16	16	T = 50

Thay giá trị vào công thức (3 - 17) ta được:

χ^2_{TN} = [(3 – 1,08)^2/1,08 + (7 – 3,6)^2/3,6 + (5 – 6,48)^2/6,48 + ... + (5 – 5,76)^2/5,76 + (11 – 6,08)^2/6,08)] = 19,40 > $\chi^{2(6)}_{0,01}$ = 16,81.

Tiêu chuẩn χ^2 có thể sử dụng rộng rãi để kiểm tra các tần số thực nghiệm khi đã biết các tần số lý thuyết.

Chương 4

PHÂN TÍCH MỐI QUAN HỆ

4.1. CÁC LOẠI QUAN HỆ

Phép ước lượng các tham số thống kê là phép mô tả tổng thể từ các mẫu điều tra khảo sát theo một chỉ tiêu đó, còn phép so sánh là phép phân định sự khác nhau hay giống nhau giữa các tổng thể theo một hay một số tham số nhất định. Chương này sẽ đề cập đến mối quan hệ giữa các đặc trưng trong một tổng thể và giữa các tổng thể.

Nếu coi các số đo nhận được từ điều tra khảo sát về một chỉ tiêu nào đó của mẫu rút ra từ tổng thể là các biến số của một đại lượng ngẫu nhiên thì mối quan hệ giữa hai hay nhiều tổng thể, giữa hai hay nhiều đặc trưng trong một tổng thể là mối quan hệ giữa hai hay nhiều đại lượng ngẫu nhiên.

Các mối quan hệ được biểu thị bởi các hàm số phụ thuộc gọi là phương trình hồi quy. Tùy các mối quan hệ khác nhau người ta chia ra:

1. Tương quan và hồi quy tuyến tính (đường thẳng), gồm:

- Tương quan và hồi quy và tuyến tính một biến
- Tương quan và hồi quy và tuyến tính đa biến

2. Tương quan và hồi quy phi tuyến tính (đường cong), gồm:

- Tương quan và hồi quy và phi tuyến tính một biến

- Tương quan và hồi quy và phi tuyến tính đa biến.

4.2. QUAN HỆ TUYẾN TÍNH

4.2.1. Các dạng quan hệ tuyến tính

Phương trình biểu thị mối quan hệ tuyến tính giữa X và Y có dạng:

$$y = f(x) = a + bx \qquad (4 - 1)$$

hoặc $$y = f(x_i) = b_0 + b_1x_1 + b_2x_2 + ... + b_nx_n \qquad (4 - 2)$$

Phương trình (4 - 1) gọi là phương trình hồi quy tuyến tính một biến, trong đó *y* là hàm số (số phụ thuộc), *x* là đối số (số độc lập). Ứng với mỗi giá trị của *x* ta có một giá trị xác định tương ứng của *y*, b là hệ số góc, còn gọi là hệ số hồi quy và a là hằng số. Trên đồ thị hai chiều (trục tung *y* và trục hoành *x*), đồ thị hàm số cắt trục tung tại điểm a. Nếu a > 0 đồ thị đi qua phía trên góc tọa độ, ngược lại, nếu a < 0 đồ thị đi qua phía dưới góc tọa độ và đồ thị đi qua góc tọa độ khi a = 0. Về chiều hướng và mức độ của mối quan hệ phụ thuộc vào hệ số tương quan *r* (sẽ xét cụ thể ở mục hệ số tương quan).

Người ta gọi là tương quan tuyến tính vì đường biểu diễn của phương trình hồi quy $\overline{y} = a + b\overline{x}$ từ các mẫu số quan sát là một đường thẳng và $\hat{y} = \hat{a} + b\hat{x}$ là phương trình của đường hồi quy lý luận.

- Phương trình (4 - 2) gọi là phương trình hồi quy tuyến tính bội (đa biến), trong đó y là hàm số - số phụ thuộc, còn x_i $\left(i = \overline{1, n}\right)$ là các đối số - các số độc lập. Phương trình của đường hồi quy lý luận được ước lượng từ các mẫu số quan sát là: $\hat{y} = \hat{b}_0 + \hat{b}_1\hat{x}_1 + \hat{b}_1\hat{x}_2 + ... + \hat{b}_n\hat{x}_n$

4.2.2. Mô hình tuyến tính đơn các đặc trưng định lượng

4.2.2.1. Hệ số tương quan đơn và đánh giá sự tồn tại của hệ số tương quan

- ***Hệ số tương quan đơn***

Để đo mức độ quan hệ tuyến tính giữa hai đại lượng ngẫu nhiên X và Y, người ta đưa ra khái niệm *hệ số tương quan*. *Hệ số tương quan lý luận* giữa X và Y, ký hiệu là ρ, được định nghĩa bởi công thức:

$$\rho = \frac{E(X-\mu_X)(Y-\mu_Y)}{\sigma_X \sigma_Y}$$

trong đó: μ_X, σ_X là kỳ vọng và độ lệch chuẩn lý luận của X và μ_Y, σ_Y là là kỳ vọng và độ lệch chuẩn lý luận của Y.

Người ta đã chứng minh được $-1 \le \rho \le 1$.

Trong thực nghiệm, hệ số tương quan giữa X và Y, ký hiệu là *r* hoặc r_{XY}, được tính bằng công thức:

$$r_{XY} = \frac{\mathrm{Cov}(X,Y)}{S_x S_y}$$

$$= \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]\left[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2\right]}}$$

$$= \frac{\overline{xy} - \overline{x}.\overline{y}}{\sqrt{\left[\overline{x}^2 - (\overline{x})^2\right]\left[\overline{y}^2 - (\overline{y})^2\right]}} = \frac{\overline{xy} - \overline{x}.\overline{y}}{\sqrt{\mathrm{Var}(x)\mathrm{Var}(y)}}$$

Tính chất của hệ số tương quan

1. $-1 \le r_{XY} \le 1$

2. Nếu S_x và S_y đều khác 0 và r_{XY} càng gần 1 hay càng gần -1 thì X và Y có quan hệ hàm số chặt chẽ.

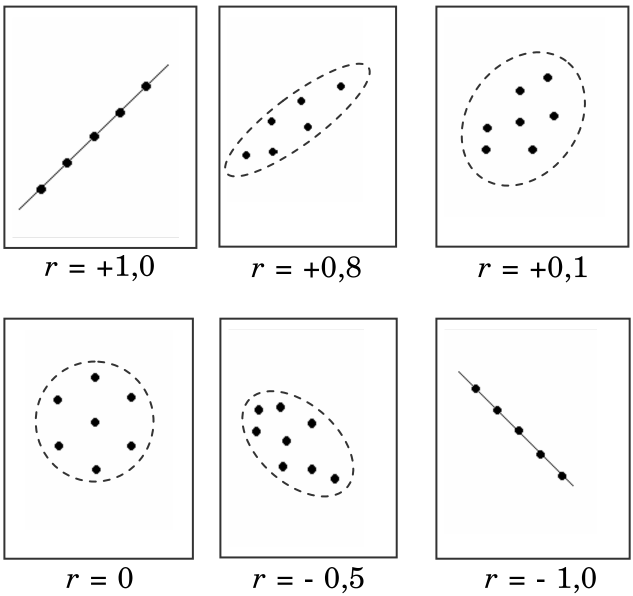
3. Nếu S_x và S_y đều khác 0 và r_{XY} càng gần 0 thì X và Y không có quan hệ tuyến tính hoặc độc lập với nhau.

4. Nếu r_{XY} càng gần 0 và:

- $S_x \approx 0$ thì X độc lập với Y
- $S_y \approx 0$ thì Y độc lập với X

5. Nếu $r_{XY} > 0$ thì X và Y đồng biến

$r_{XY} < 0$ thì X và Y nghịch biến



Hình 4.1: Sơ đồ biểu thị hệ số tương quan r đơn với sự biến động các giá trị

Có thể dựa vào độ lớn của hệ số tương quan để xác định giá mức độ quan hệ giữa hai đại lượng.

$|r| \simeq 0$: X và Y độc lập hoặc có quan hệ phi tuyến tính

$0 < |r| \leq 0,3$: X và Y có quan hệ yếu

$0,3 < |r| \leq 0,5$: X và Y có quan hệ vừa

$0,5 < |r| \leq 0,7$: X và Y có quan hệ tương đối chặt

$0,7 < |r| \leq 0,9$: X và Y có quan hệ chặt

$0,9 < |r| \leq 1$: X và Y có quan hệ rất chặt

- ***Đánh giá sự tồn tại của hệ số tương quan***

Hệ số tương quan biểu thị mức độ quan hệ giữa hai đại lượng. Tuy nhiên, mối quan hệ có thực (tồn tại) hay không phụ thuộc vào độ tin cậy (nói cách khác là mức ý nghĩa) của r. Có những mối quan hệ "chặt" nhưng chưa đủ tin cậy (do dung lượng mẫu quá ít), nhưng có những mối quan hệ yếu, thậm chí hai đại lượng không quan hệ với nhau ($|r| \simeq 0$) nhưng đáng tin cậy (dung lượng mẫu lớn).

Trong thực nghiệm, có hai cách để đánh giá độ tin cậy của hệ số tương quan r:

1. So sánh hệ số tương quan thực nghiệm r_{xy} với giá trị r lý luận:

Hệ số tương quan lý luận được tính sẵn với độ tự do $n - 2$ ở các mức tin cậy khác nhau (phụ lục 4). Nếu $r_{xy} \geq r_{\alpha}$ với $n - 2$ bậc tự do, ta nói r_{xy} tồn tại với độ tin cậy $1 - \alpha$.

Ví dụ: với n = 20, độ tự do là 18, nếu $r_{xy} \geq 0,444$ thì r tồn tại với độ tin cậy 95% (sai lầm < 5%); nếu $r_{xy} \geq 0,561$ thì r tồn tại với độ tin cậy 99% (sai lầm < 1%) và nếu $r_{xy} \geq 0,679$ thì r tồn tại với độ tin cậy 99,9% (sai lầm < 0,1%).

2. Tra độ tin cậy *r* qua giá trị tới hạn *t* trong bảng Student:

Đặt: $$T = r_{xy}\sqrt{\frac{n-2}{1-r_{xy}^2}}$$

Tra bảng T với *n* - 2 bậc tự do ta được các trị tới hạn *t* với các mức á khác nhau. Nếu $T \geq t_{\alpha}^{(n-2)}$, ta nói r_{xy} tồn tại với độ tin cậy 1 – α.

Ví dụ: với n = 20, độ tự do là 18, $t_{0,05}^{(18)} = 2,101$; $t_{0,01}^{(18)} = 2,878$; $t_{0,001}^{(18)} = 3,922$. Nếu $r_{xy} \geq 0,486$, tính theo công thức trên ta có: T = 2,359 > 2,101, ta nói *r* tồn tại với độ tin cậy 95%. Nếu $r_{xy} \geq 0,582$ ta có: T = 3,036 > 2,878, ta nói r tồn tại với độ tin cậy 99% và nếu $r_{xy} \geq 0,701$ ta có T = 4,170 > 3,922, ta nói *r* tồn tại với độ tin cậy 99,9%.

Thông thường, người ta ghi sau hệ số tương quan thực nghiệm các ký hiệu : (*), (**) và (***) tương ứng với các độ tin cậy 95%, 99% và 99,9% của *r*.

4.2.2.2. Xác định các hệ số a, b và ý nghĩa của a, b

- ***Xác định các hệ số*** $\hat{a}$, **b**

Để dự đoán mô hình tuyến tính đơn $\hat{y} = \hat{a} + b\hat{x}$, phải xác định giá trị tối thích $\hat{a}$ và b. Để tìm các giá trị này ta sử dụng phương pháp bình phương tối thiểu do Karl Pearson (1857 – 1936) đề xuất.

Theo phương pháp bình phương tối thiểu, $\hat{a}$ và b tối thích khi tổng bình phương độ lệch từ các giá trị thực nghiệm y_i và giá trị lý luận $\hat{y}$ là nhỏ nhất, tức là:

$$Q = \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \text{ là nhỏ nhất.}$$

Do $Q \geq 0$ nên nhất định sẽ tìm được $\hat{a}$ và b khi Q nhỏ nhất.

Tại điểm $\hat{a}$ và điểm b: $\frac{\partial Q}{\partial \hat{a}} = 0$ và $\frac{\partial Q}{\partial b} = 0$, tức là:

$$\frac{\partial \sum_{i=1}^{n} \left(y_i - \hat{a}x_i - b \right)^2}{\partial \hat{a}} = 0$$

$$\frac{\partial \sum_{i=1}^{n} \left(y_i - \hat{a}x_i - b \right)^2}{\partial b} = 0$$

hay

$$\begin{cases} nb + \hat{a}\sum x_i = \sum y_i \\ b\sum x_i + \hat{a}\sum x_i^2 = \sum x_i y_i \end{cases}$$

Giải hệ phương trình ta được:

$$\hat{a} = \frac{\sum x_i y_i - \sum x_i \sum y_i}{\sum x_i^2 - \left(\sum x_i\right)^2} = \frac{\overline{xy} - \overline{x}.\overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2}$$

$$b = \frac{\sum y_i}{n} - \hat{a}\frac{\sum x_i}{n} = \overline{y} - \hat{a}\overline{x}$$

- ***Kiểm định ý nghĩa của*** $\hat{a}$**, b**

Để kiểm định $\hat{a}$ ta sử dụng:

$$T = \frac{\hat{a} - a_0}{s_e\left(\hat{a}\right)} \text{ so với } t_{\alpha/2}^{(n-2)}$$

trong đó: $$s_e(\hat{a}) = \sqrt{\frac{\sum(y_i - \hat{y})^2}{(n-2)\sum(x_i - \bar{x})^6}}$$

Để kiểm định b ta sử dụng:

$$T = \frac{b - b_0}{s_e(b)} \text{ so với } t_{\alpha/2}^{(n-2)}$$

trong đó: $$s_e(b) = \sqrt{\frac{\sum(y_i - \hat{y})^2 \sum x_i^2}{n(n-2)\sum(x_i - \bar{x})^2}}$$

4.2.2.3. Kiểm định độ tin cậy của hàm hồi quy tuyến tính

Để kiểm định sự tồn tại của hàm hồi quy ta sử dụng phép nghiệm F:

$$F = \frac{R^2}{\frac{(1-R^2)}{(n-2)}} \text{ so với } f_\alpha(1, n\text{-}2)$$

trong đó: $R^2 = r_{xy}^2 = \frac{SS_{Regression}}{SS_T}$ hay $R^2 = 1 - \frac{SS_{Residual}}{SS_T}$

$$SS_T = \sum(y_i - \bar{y})^2 = \sum y_i^2 - n(\bar{y})^2$$

$$SS_{Regr.} = SS_T - SS_{Residual}$$

$$SS_{Residual} = (\hat{a})^2\left[\sum x_i^2 - n(\bar{x})^2\right]$$

R^2 được gọi là *hệ số xác định* (determinant coefficient). R^2 cho biết tỷ lệ phần biến động do X gây nên trong quan hệ tuyến tính ($SS_{Regression}$) so với tổng số biến động (SS_T) của Y. Tỷ lệ này được tính bằng phần trăm (%)

khi nhân R^2 với 100. Phần trăm còn lại là do các yếu tố khác gây nên trong quan hệ đa biến (nếu có) và do sai số ngẫu nhiên.

R^2 càng lớn thì Y càng phụ thuộc vào X.

Trong phép nghiệm F, nếu $F > f_\alpha(1, n\text{-}2)$ thì hàm hồi quy tuyến tính có độ tin cậy $1 - \alpha$.

Ví dụ:

Số quả /cây (x)	9,0	10,7	10,7	13,1	9,8	14,4	10,6	11,4	8,6	12,2
NS, tạ/ha (y)	23,6	31,2	28,8	37,3	28,5	37,1	30,8	28,1	18,3	26,2

Kết quả xử lý trên phần mềm Excel

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.831
R Square	0.691
Adjusted R Square	0.652
Standard Error	3.377
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Sig. F*
Regression	1	203.93	203.93	17.88	0.0029
Residual	8	91.24	11.41		
Total	9	295.17			

	Coefficients	*SE*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-0.27	7.001	-0.039	0.9702	-16.41	15.88
x	2.65	0.626	4.229	0.0029	1.20	4.09

Ở ví dụ này, hệ số tương quan $r = 0,831^{**}$

Hàm hồi quy tuyến tính y = 2,65x – 0,27 tồn tại với độ tin cậy 99% (mức ý nghĩa của F = 0,0029).

4.2.3. Mô hình tuyến tính đa biến

4.2.3.1. Hệ số tương quan riêng

Hệ số tương quan riêng được tính toán trong nghiên cứu quan hệ đa biến.

Có *m* đại lượng ngẫu nhiên X_1, X_2. ... X_m có quan hệ với nhau. Hệ số tương quan riêng của X_1 với X_2 khi cố định X_3, ký hiệu $r_{12,3}$ sẽ là:

$$r_{12,3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{\left(1 - r_{13}^2\right)\left(1 - r_{23}^2\right)}} \qquad (4 - 3)$$

Tương tự, ta có hệ số tương quan riêng của X_i với X_j khi cố định X_k ($r_{ij,k}$) sẽ là:

$$r_{ij,k} = \frac{r_{ij} - r_{ik}r_{jk}}{\sqrt{\left(1 - r_{ik}^2\right)\left(1 - r_{jk}^2\right)}}$$

Hệ số tương quan riêng của X_1 với X_2 khi cố định X_3 và X_4 ($r_{12,34}$) là:

$$r_{12,34} = \frac{r_{12,3} - r_{14,3}r_{24,3}}{\sqrt{\left(1 - r_{14,3}^2\right)\left(1 - r_{24,3}^2\right)}} \qquad (4 - 4)$$

Một cách tổng quát, hệ số tương quan riêng của X_1 với X_2 khi cố định X_3, X_4, ... X_m ($r_{12,34...m}$) là:

$$r_{12,34...m} = \frac{\mathrm{r}_{12,34...(m-1)} - \mathrm{r}_{1m,34...(m-1)}\mathrm{r}_{2n,34...(m-1)}}{\sqrt{\left(1 - \mathrm{r}_{1m,34...(m-1)}^2\right)\left(1 - \mathrm{r}_{2m,34...(m-1)}^2\right)}}$$

Ví dụ: Tính $r_{12,3}$ và $r_{13,2}$ theo Bảng 4.1 sau.

Thay số liệu vào công thức (4-3 và (4-4) ta có:

$$r_{12,3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{\left(1 - r_{13}^2\right)\left(1 - r_{23}^2\right)}}$$

Bảng 4.1: Hệ số tương quan đơn giữa 7 đại lượng

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	1						
x_2	0,244	1					
x_3	0,627	-0,228	1				
x_4	0,609	-0,363	0,794	1			
x_5	0,461	-0,249	0,894	0,689	1		
x_6	0,636	-0,069	0,701	0,628	0,687	1	
x_7	0,638	0,705	0,524	0,240	0,438	0,438	1

$$= \frac{0,244 - 0,627 \times (-0,228)}{\sqrt{\left(1 - 0,627^2\right)\left[1 - (-0,228)^2\right]}} = 0,510$$

và:

$$r_{13,2} = \frac{r_{13} - r_{12}r_{32}}{\sqrt{\left(1 - r_{12}^2\right)\left(1 - r_{32}^2\right)}}$$

$$= \frac{0,627 - 0,244 \times (-0,228)}{\sqrt{\left(1 - 0,244^2\right)\left[1 - (-0,228)^2\right]}} = 0,723$$

4.2.3.2. Xác định phương trình hồi quy tuyến tính

- ***Cách tính hệ số hồi quy***

Như đã nêu ở trên, phương trình hồi quy tuyến tính lý luận có dạng:

$$\hat{y} = \hat{b}_0 + \hat{b}_1\hat{x}_1 + \hat{b}_1\hat{x}_2 + \ldots + \hat{b}_n\hat{x}_n$$

Theo nguyên lý của phương pháp bình phương tối thiểu, có thể tìm được các hệ số $\hat{b}_0$, $\hat{b}_1$, $\hat{b}_2$, ..., $\hat{b}_n$ sao cho:

$$Q = \sum_i \left(y_i - \hat{y}_i\right)^2 = \sum \left(y_i - \hat{b}_0 - \hat{b}_1 x_{i1} - \ldots - \hat{b}_n x_{in}\right)^2$$

là tối thiểu.

Ví dụ, với phương trình $\hat{y} = \hat{b}_0 + \hat{b}_1 \hat{x}_1 + \hat{b}_2 \hat{x}_2$

Ta tìm $\hat{b}_0$, $\hat{b}_1$, $\hat{b}_2$ sao cho $Q = \sum \left(y_i - \hat{b}_0 - \hat{b}_1 x_{i1} - \hat{b}_2 x_{i2}\right)^2$ nhỏ nhất.

Tại ba điểm $\hat{b}_0$, $\hat{b}_1$ và $\hat{b}_2$:

$$\begin{cases} \dfrac{\partial Q}{\partial b_0} = \sum \left(y_i - \hat{b}_0 - \hat{b}_1 x_{i1} - \hat{b}_2 x_{i2}\right) = 0 \\ \dfrac{\partial Q}{\partial b_1} = \sum \left(y_i - \hat{b}_0 - \hat{b}_1 x_{i1} - \hat{b}_2 x_{i2}\right) x_{i1} = 0 \\ \dfrac{\partial Q}{\partial b_2} = \sum \left(y_i - \hat{b}_0 - \hat{b}_1 x_{i1} - \hat{b}_2 x_{i2}\right) x_{i2} = 0 \end{cases}$$

hay

$$\begin{cases} n b_0 + \left(\sum_{i=1}^{n} x_{i1}\right) b_1 + \left(\sum_{i=1}^{n} x_{i2}\right) b_2 = \sum_{i=1}^{n} y \\ \left(\sum_{i=1}^{n} x_{i1}\right) b_0 + \left(\sum_{i=1}^{n} x_{i1}^2\right) b_1 + \left(\sum_{i=1}^{n} x_{i1} x_{i2}\right) b_2 = \sum_{i=1}^{n} x_{i1} y \\ \left(\sum_{i=1}^{n} x_{i2}\right) b_0 + \left(\sum_{i=1}^{n} x_{i2} x_{i1}\right) b_1 + \left(\sum_{i=1}^{n} x_{i2}^2\right) b_2 \\ \quad = \sum_{i=1}^{n} x_{i2} y \end{cases}$$

Giải hệ phương trình này ta được $\hat{b}_0$, $\hat{b}_1$ và $\hat{b}_2$

- ***Kiểm định độ tin cậy của hàm hồi quy tuyến tính***

Để kiểm định sự tồn tại của hàm hồi quy ta sử dụng phép nghiệm F:

$$F = \frac{R^2(n-p)}{(1-R^2)(p-1)} \text{ so với } f_\alpha(p-1, n-p)$$

Trong đó: $R^2 = 1 - \frac{SS_{Residual}}{SS_T}$

p là số đại lượng ngẫu nhiên được nghiên cứu trong quan hệ với y.

$$SS_T = \|y_i - \bar{y}\|^2$$

$$SS_{Regr.} = SS_T - SS_{Residual}$$

$$SS_{Residual} = \|\hat{y} - \bar{y}\|^2$$

- ***Kiểm định ý nghĩa của các hệ số*** $\hat{b}_i \left(i = \overline{0,1,n}\right)$

Để kiểm định sự tồn tại $\hat{b}_i \neq 0$ ta sử dụng:

$$T = \frac{\hat{b}_i}{s_e(\hat{b}_i)} \text{ so với } t_{\alpha/2}^{(n-p)}$$

Trong đó: $s_e(\hat{b}_i) = \sqrt{MS_{(\hat{b}_i)}}$

Ví dụ: Kết quả theo dõi 7 chỉ tiêu trên 36 giống lúa được ghi ở Bảng 4.2. Hãy xác định phương trình quan hệ với năng suất (g/bụi).

Bảng 4.2: Kết quả theo dõi 7 chỉ tiêu trên 36 giống lúa

Giống	Cao cây (cm)	TS nhánh	Nhánh HH	Dài bông (cm)	Số hạt /bông	M100 hạt (g)	NSTT (g/bụi)	NS hồi quy (g/bụi)
1	109,5	46,1	41,9	20,3	52,5	3,9	85,7	86,2
2	119,1	36,2	33,4	20,0	63,2	4,0	83,9	85,4
3	114,4	47,6	44,6	21,7	48,6	4,4	96,4	94,3
4	119,7	57,2	55,2	20,0	52,0	4,1	117,0	114,1
5	117,0	44,0	42,1	21,2	64,1	4,3	115,8	108,4
6	121,5	35,9	34,0	22,1	47,6	4,3	68,7	69,8
7	120,9	53,1	50,4	22,1	57,1	4,8	108,5	122,4
8	119,5	33,3	30,8	20,0	40,9	4,0	50,7	50,7
9	126,8	46,7	43,8	19,8	61,1	3,7	98,3	98,2
10	111,7	43,7	42,3	20,2	77,4	3,3	108,1	111,8
11	120,1	41,5	40,5	21,7	51,1	4,3	88,7	87,8
12	121,2	52,5	50,7	22,0	51,2	3,9	101,4	101,5
13	127,8	30,6	28,3	22,5	66,8	3,8	72,3	78,2
14	127,7	42,9	40,1	21,9	52,6	4,4	92,7	90,7
15	131,8	49,7	47,5	20,1	41,5	4,3	84,4	88,4
16	111,8	42,1	37,3	18,7	43,2	4,6	74,1	75,6
17	117,5	40,2	38,1	19,4	56,7	4,7	101,7	97,1
18	109,8	47,6	45,9	20,1	45,5	4,4	91,3	92,1
19	122,4	57,0	54,8	19,0	43,5	4,2	100,2	103,7
20	110,2	50,0	47,4	23,1	52,8	4,4	110,0	105,2
21	129,3	55,7	52,8	20,3	52,8	4,2	116,6	112,4
22	126,1	45,4	41,2	20,3	52,2	4,3	91,6	90,5
23	120,7	46,9	44,2	20,6	48,2	4,0	84,3	85,7
24	111,9	49,8	45,9	19,4	55,5	4,2	107,2	103,0
25	100,0	36,9	34,2	21,3	39,8	4,5	61,1	63,1
26	116,4	29,9	26,7	20,8	41,7	4,2	56,8	46,8
27	130,7	34,9	32,5	20,0	40,5	4,1	54,1	55,2
28	117,1	45,1	42,7	20,8	44,0	4,6	86,1	86,8
29	110,4	34,6	33,0	20,1	53,2	4,6	80,3	80,6
30	123,4	39,9	38,2	21,9	53,6	4,6	94,1	91,5
31	133,5	57,2	52,8	19,7	43,6	4,4	100,4	102,8
32	120,4	58,5	56,1	21,0	50,3	4,2	118,6	115,5
33	116,2	30,2	27,8	20,9	41,4	5,0	57,0	60,1
34	126,1	49,0	43,9	20,2	48,3	4,3	91,0	90,5
35	130,1	53,5	51,2	21,9	52,8	4,5	121,7	114,1
36	126,3	49,8	46,7	20,0	42,3	4,2	82,0	85,6

Ghi chú: TS – tổng số, HH – hữu hiệu, M – khối lượng, NSTT- năng suất thực thu

Kết quả tính tốn nhờ phần mềm Excel 5.0 cho thấy:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.974
R Square	0.949
Adjusted R Square	0.938
Standard Error	4.794
Observations	36

ANOVA

	df	*SS*	*MS*	*F*	*Sig.F*
Regression	6	12412.126	2068.688	90.02	2.1E-17
Residual	29	666.403	22.979		
Total	35	13078.529			

	Coefficients	*St-dard Error*	*t Stat*	*P-value*
Intercept	-125.236	25.808	-4.85	3.82E-05
X Variable 1	-0.046	0.112	-0.41	6.82E-01
X Variable 2	0.130	0.936	0.14	8.91E-01
X Variable 3	1.822	0.943	1.93	6.32E-02
X Variable 4	0.244	0.865	0.28	7.80E-01
X Variable 5	1.354	0.116	11.67	1.78E-12
X Variable 6	15.129	2.943	5.14	1.72E-05

Phương trình hồi quy theo tính toán trên đây :

$$y = -\ 125{,}24 - 0{,}05\ x_1 + 0{,}13\ x_2 + 1{,}82x_3 + 0{,}24\ x_4 + 1{,}35x_5 + 15{,}13x_6$$

Kết quả trắc nghiệm sự tồn tại của các hệ số hồi quy ở độ tin cậy 95% cho thấy, chiều cao cây (x_1), tổng số nhánh (x_2), chiều dài bông (x_4) không quan hệ với năng suất (P > 0,68 và P > 0,89). Các tính trạng số hạt/bông (x_5) và khối lượng 100 hạt (x_6) có quan hệ chắc chắn với năng suất với độ tin cậy của b_5 và b_6 đều lớn hơn 99%. Riêng số nhánh hữu hiệu (x_3) có thể có quan hệ với năng suất. Do sự có mặt của chiều cao cây, tổng số nhánh và chiều dài bông trong mô hình nghiên cứu có thể đã ảnh hưởng đến độ tin cậy của hệ số hồi quy b_3 (gần 94%).

Sau khi loại trừ ba tính trạng không quan hệ với năng suất, kết quả xử lý lại với các tính trạng số nhánh hữu hiệu (x_1), số hạt/bông (x_2) và khối lượng 100 hạt (x_3) cho thấy:

	Coefficients	*Std Error*	*t Stat*	*P-value*
Intercept	-126.90	15.495	-8.189	2.4E-09
X Variable 1	1.94	0.096	20.293	7.7E-20
X Variable 2	1.36	0.105	13.028	2.4E-14
X Variable 3	15.47	2.701	5.728	2.4E-06

Rõ ràng, cả ba tính trạng này đều quan hệ với năng suất. Phương trình hồi quy được xác định là:

$$y = -126{,}90 + 1{,}94x_1 + 1{,}36x_2 + 15{,}47x_3 \quad (4 - 5)$$

Thay các giá trị x_1, x_2, x_3 của ba tính trạng số nhánh hữu hiệu, số hạt/bông và khối lượng 100 hạt vào phương trình (4 - 5) ta được năng suất hồi quy ở bảng 4.1.

4.2.4. Vai trò của từng biến trong quan hệ đa biến

Phương trình hồi quy đa biến: $y = b_0 + b_1x_1 + ... + b_nx_n$ cho biết các biến X_1, X_2, ... X_n có quan hệ với Y. Do đơn vị tính của các biến khác nhau nên độ lớn nhỏ của hệ số hồi quy b_i không biểu thị vai trò của từng biến X_i đối với Y. Ví dụ, xét vai trò của biến x_1 (số nhánh hữu hiệu) và x_3 (khối lượng 100 hạt) với năng suất (y), ta thấy mặc dù hệ số hồi quy của x_1 nhỏ (1,94) nhưng giá trị của x_1 lớn (trên dưới 40), còn hệ số hồi quy của x_3 lớn (15,47) nhưng giá trị của x_3 nhỏ (khoảng 4g). Vì thế, không thể nói khối lượng 100 hạt có vai trò quan trọng hơn số nhánh hữu hiệu trong quan hệ với năng suất.

Để xác định vai trò của từng biến trong quan hệ đa biến ta sử dụng phép phân tích đường (Path Analysis).

Phép phân tích đường do S.Wright đề xuất năm 1921 là phương pháp phân tích mối quan hệ nhân quả giữa các đại lượng ngẫu nhiên. Theo đó, trong mối quan hệ của một cặp đại lượng có thể là kết quả ảnh hưởng của một số mối quan hệ thành phần với chiều hướng và mức độ khác nhau.

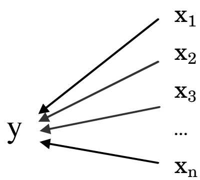
4.2.4.1. Hệ số đường

- ***Các mối quan hệ***

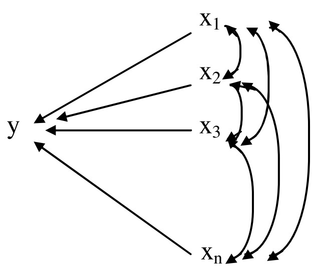
Quan hệ giữa hai hay nhiều đại lượng, nói chung có hai loại:

- Quan hệ nhân quả: một bên là nguyên nhân còn một bên là kết quả và không thể đảo ngược. Ví dụ, quan hệ giữa số quả, số hạt với năng suất.

- Quan hệ ngang bằng: có ảnh hưởng qua lại, không phân biệt nguyên nhân và kết quả. Ví dụ, chiều dài quả và to quả, số quả và khối lượng quả.



Hình 4.1: Quan hệ đồng thời x_i lên y



Hình 4.2: Quan hệ giữa nhiều đại lượng

- ***Khái niệm hệ số đường***

Dùng y, x_1, x_2 biểu thị giá trị thực của đại lượng kết quả và các đại lượng nguyên nhân. Giả sử y có quan hệ với x_1, x_2 theo phương trình:

$$y = b_0 + b_1x_1 + b_2x_2$$

b_1 và b_2 là hệ số hồi quy riêng của x_1, x_2. Hệ số b_1 phản ánh ảnh hưởng của x_1 tới y khi cố định x_2, tương tự, hệ số b_2 phản ánh ảnh hưởng của x_2 tới y khi cố định x_1. Như vậy khi cố định x_1, x_2 là đơn vị của y, còn khi cố định x_2, x_1 là đơn vị của y. Do x_1 và x_2 có đơn vị tính khác nhau và khác với y nên không thể dùng b_1 và b_2 để đánh giá vai trò của x_1 và x_2 đối với y.

Để đánh giá vai trò của x_1 và x_2 đối với y phải chuyển đổi đơn vị của y, x_1, x_2 thành số tương đối.

Nếu gọi $\overline{y}$, $\overline{x}_1$ và $\overline{x}_2$ là trị số trung bình và s_y, s_1 và s_2 là độ lệch chuẩn của y, x_1 và x_2, ta có:

$$(y - \overline{y}) = b_1(x_1 - \overline{x}_1) + b_2(x_2 - \overline{x}_2)$$

Chia hai vế cho s_y và sự dung s_1, s_2 trong thuật toán:

$$\frac{(y-\overline{y})}{s_y} = b_1.\frac{s_1}{s_y}.\frac{(x_1-\overline{x}_1)}{s_1} + b_2.\frac{s_2}{s_y}.\frac{(x_2-\overline{x}_2)}{s_2}$$

Đặt $y' = \dfrac{y-\overline{y}}{s_y}$, $x_1' = \dfrac{x_1-\overline{x}_1}{s_1}$ và $x_2' = \dfrac{x_2-\overline{x}_2}{s_2}$ ta có:

$$y' = P_{y.1}.x_1' + P_{y.2}.x_2'$$

trong đó: $P_{y.1} = b_1.\dfrac{s_1}{s_y}$ và $P_{y.2} = b_2.\dfrac{s_2}{s_y}$. $P_{y.1}$ và $P_{y.2}$ được gọi là *hệ số đường* biểu thị ảnh hưởng của nguyên nhân x_1 và

x_2 đến kết quả y.

Một cách tổng quát, khi đại lượng y có *n* đại lượng nguyên nhân x_i (i = 1, 2, ...n) thì hệ số đường ảnh hưởng của x_i đến y là: $P_{y.i} = b_i.\frac{S_i}{S_y}$

Từ công thức $y' = P_{y.1}.x'_1 + P_{y.2}.x'_2$ không khó để nhận ra hệ số đường $P_{y.i}$ là hệ số hồi quy riêng của x'_i đến y'. Đó cũng là số hồi quy riêng của x_i đến y sau khi đã chuyển hóa đơn vị.

4.2.4.2. Xác định hệ số đường

Mối quan hệ giữa các x_i với y và giữa các x_i với nhau tác động đến y là hết sức phức tạp. Trong tài liệu này ta chỉ xét vai trò của từng đại lượng x_i khi chúng tác động đồng thời đến y theo sơ đồ hình 4.1.

Theo tính chất 1 của hệ số đường, nếu quan hệ giữa các đại lượng x_1, x_2, ... x_n với y là quan hệ nhân quả theo hình 4.1 thì hệ số tương quan giữa đại lượng nguyên nhân x_i với đại lượng kết quả y sẽ là:

$$r_{yi} = \sum_{j=1}^{n} P_{y.j} r_{ij} \qquad (4 - 6)$$

Công thức này là phương trình thứ *i* trong hệ phương trình tuyến tính. Theo tính chất này, hệ số tương quan giữa x_i và y có thể chia thành *n* số hạng, mỗi số hạng là $P_{y.j} r_{ji}$.

Nếu đại lượng x_i và x_j độc lập, tức là $r_{ij} = 0$ thì công thức (4 - 6) trở thành:

$$r_{yi} = P_{y.j}$$

Để xác định các hệ số đường ($P_{y.x}$) biểu thị mức độ ảnh hưởng của các đại lượng nguyên nhân x_i đến đại lượng y phải giải hệ phương trình $r_{yx} = [A]P_{y.x}$, trong đó $[A]$ là ma trận của các hệ số tương quan giữa các đại lượng:

$$[A] = \begin{pmatrix} r_{11} & r_{12} & r_{13} & \dots & r_{1n} \\ r_{12} & r_{22} & r_{23} & \dots & r_{2n} \\ r_{13} & r_{32} & r_{33} & \dots & r_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ r_{1n} & r_{2n} & r_{3n} & \dots & r_{nn} \end{pmatrix}$$

và r_{yx} là véctơ hệ số tương quan trực tiếp giữa các đại lượng x_1, x_2, ... với y:

$$r_{y.x} = \begin{pmatrix} r_{y1} \\ r_{y2} \\ r_{y3} \\ \dots \\ r_{yn} \end{pmatrix}$$

Nghiệm sẽ là: $P_{y.x} = [A]^{-1} r_{y.x}$, $[A]^{-1}$ là ma trận đảo của $[A]$. Đại lượng nào có $P_{y.x}$ càng cao, vai trò đối với y càng lớn.

4.2.4.3. Kiểm tra mức độ quan hệ của các đại lượng nguyên nhân với đại lượng kết quả

Để kiểm tra mức độ ảnh hưởng của các đại lượng nguyên nhân X_i lên đại lượng kết quả Y, ta xét giá trị Bx: $Bx = \sum r_{y.x}P_{y.x}$ hoặc giá trị R: $R = (1 - Bx)^{0,5}$.

Bx gọi là *hệ số xác định,* cho biết tỷ lệ (mức độ) biến động của các đại lượng X_i gây nên so với tổng số biến động của Y. Tỷ lệ này được tính bằng phần trăm (%) khi nhân

Bx với 100. Giá trị R cho biết mức độ của phần còn lại do các đại lượng khác chưa được nghiên cứu (nếu có) hoặc là sai số ngẫu nhiên.

Nếu $0,8 \le Bx \le 1$ thì các đại lượng ảnh hưởng đến Y, về cơ bản, đã được nghiên cứu đầy đủ, khi đó $R \le 0,4$;

Nếu $Bx < 0,7$ thì còn có một hoặc một số đại lượng khác ảnh hưởng đến Y mà chưa được nghiên cứu, khi đó $R \ge 0,6$.

Ví dụ ứng dụng

Hãy tìm hiểu vai trò của sáu chỉ tiêu $x_1, x_2, ..., x_6$ đến năng suất lúa từ số liệu Bảng 4.1.

Xử lý số liệu trên Excel ta xác định được ma trận tương quan giữa các x_i $([A])$ và vector r_{yx} như sau :

Hệ số tương quan giữa các x_i $([A])$ và r_{yx}:

$$
\begin{array}{c}
 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6
\end{array}
\begin{array}{c}
\begin{array}{rrrrrr} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \end{array} \\
\begin{pmatrix}
1 & 0,252 & 0,244 & 0,001 & -0,037 & -0,108 \\
0,252 & 1 & 0,993 & -0,113 & 0,014 & -0,060 \\
0,244 & 0,993 & 1 & -0,072 & 0,042 & -0,068 \\
0,001 & -0,113 & -0,072 & 1 & 0,185 & 0,123 \\
-0,037 & 0,014 & 0,042 & 0,185 & 1 & -0,468 \\
-0,108 & -0,060 & -0,068 & 0,123 & -0,468 & 1
\end{pmatrix}
\end{array}
\qquad
r_{yx} = \begin{pmatrix}
0,134 \\ 0,802 \\ 0,822 \\ 0,091 \\ 0,503 \\ -0,072
\end{pmatrix}
$$

Xác định $[A]^{-1}$ bằng hàm MINVERSE trong Excel:

$$
[A]^{-1} = \begin{pmatrix}
1,098 & -0,875 & 0,600 & -0,103 & 0,132 & 0,182 \\
-0,875 & 90,219 & -89,246 & 3,509 & 1,575 & -0,446 \\
0,600 & -89,246 & 89,359 & -3,380 & -1,631 & 0,440 \\
-0,103 & 3,509 & -3,380 & 1,246 & -0,293 & -0,321 \\
0,132 & 1,575 & -1,631 & -0,293 & 1,436 & 0,707 \\
0,182 & -0,446 & 0,440 & -0,321 & 0,707 & 1,394
\end{pmatrix}
$$

Nhân $[A]^{-1}$ với r_{yx} bằng hàm MMULT ta được vector các hệ số đường $P_{y.x}$:

$$\vec{P}_{y.x} = \begin{pmatrix} -0{,}018 \\ 0{,}055 \\ 0{,}765 \\ 0{,}013 \\ 0{,}586 \\ 0{,}254 \end{pmatrix}$$

Ta thấy, các hệ số đường của chiều cao cây (x_1), tổng số nhánh (x_2), chiều dài bông (x_4) là bằng 0 cho thấy các tính trạng này không quan hệ với năng suất. Trong các tính trạng còn lại, số nhánh hữu hiệu (x_3) có hệ số đường cao nhất, cho thấy số bông có vai trò quan trong nhất đến năng suất, kế đến là số hạt/bông (x_5) sau cùng là khối lượng 100 hạt (x_6).

Xét hệ số tương quan đơn giữa các tính trạng với năng suất (vector r_{yx}) thì chỉ có tổng số nhánh (x_2) và số nhánh hữu hiệu (x_3) có quan hệ chặt chẽ với năng suất, số hạt/ bông (x_5) quan hệ trung bình còn khối lượng hạt (x_6) không có quan hệ với năng suất. Điều đó cho thấy hệ số tương quan đơn phản ánh không đầy đủ và đúng đắn các mối quan hệ.

Để đánh giá mức độ quan hệ giữa ba tính trạng số nhánh hữu hiệu, số hạt/ bông và khối lượng hạt đến năng suất, ta tính Bx. Ở ví dụ này:

$$Bx = P_{y.x}\, r_{yx} = (-\ 0{,}018 \times 0{,}134) + (0{,}055 \times 0{,}802) + \ldots$$

$$+ (0{,}586 \times 0{,}503) + (0{,}254 \times -\ 0{,}072 = 0{,}95$$

Trong Excel, dùng hàm SUMPRODUCT để nhân vector $P_{y.x}$ với vector r_{yx}. Bx = 0,95 cho thấy trong nghiên cứu này ba tính trạng số nhánh hữu hiệu, số hạt/ bông và khối lượng hạt kiểm soát toàn bộ năng suất lúa.

4.3. QUAN HỆ PHI TUYẾN TÍNH

4.3.1. Tỷ số tương quan

Hệ số tương quan *r* biểu thị mức độ quan hệ tuyến tính giữa hai đại lượng ngẫu nhiên. Tuy nhiên, trong nhiều trường hợp, hệ số tương quan *r* rất nhỏ, thậm chí bằng 0, tức là không có quan hệ chặt chẽ hay độc lập với nhau theo quan hệ tuyến tính nhưng lại có quan hệ chặt chẽ theo một mối quan hệ khác – quan hệ phi tuyến tính.

Để biểu thị mức độ quan hệ khác giữa hai đại lượng Y theo X, người ta sử dụng tỷ số tương quan, ký hiệu là η_{yx}:

$$\eta_{yx} = \sqrt{1 - \frac{E\left[Y - E\left(Y_X\right)\right]^2}{DY}}$$

hay

$$= \sqrt{\frac{\sum\left(Y - \bar{Y}\right)^2 - \sum\left(Y - \bar{Y}_X\right)^2}{\sum\left(Y - \bar{Y}\right)^2}}$$

trong đó: $E\left(Y_X\right)$ hay $\bar{Y}_X$ là kỳ vọng của Y khi cố định một giá trị X, còn gọi là kỳ vọng của Y với điều kiện X.

Người ta đã chứng minh được:

$0 \le \eta_{yx} \le 1$ và $\rho \le \eta_{yx}$, ρ là hệ số tương quan lý luận và $0 \le r^2 \le \eta^2$.

Nếu r^2 là hệ số xác định biểu thị mức độ biến động của X trong quan hệ tuyến tính so với biến động của Y, thì

η^2 biểu thị mức độ biến động của X trong quan hệ phi tuyến tính so với biến động của Y.

$(\eta_{yx} - \rho)$ biểu thị mức độ quan hệ phi tuyến giữa Y và X. $(\eta_{yx} - \rho)$ càng lớn thì Y và X có tương quan phi tuyến càng chặt.

Để xác định tỷ tương quan ta thực hiện các bước sau đây:

1. Lập bảng số liệu:

Y \ X	x_1	x_2	...	x_k	
y_1	y_{11}	y_{12}	...	y_{1k}	
y_2	y_{21}	y_{22}	...	y_{2k}	
...	...	...	...	...	
y_n	$y_{n_1 1}$	$y_{n_1 2}$	...	$y_{n_1 k}$	
	n_1	n_2		n_k	$n = \Sigma n_i$
	T_1	T_2		T_k	$T = \Sigma T_i$

Trong bảng: $T_i = \sum_{i=1}^{n_i} y_{ij}$; n_i là số các số liệu các cột x_j

2. Tính tổng bình phương tổng số (SS_T):

$$SS_T = \sum_1^n y_{ij}^2 - \frac{T^2}{n}$$

3 . Tính tổng bình phương Y do sai khác giữa các x_j (SS_X):

$$SS_{Y_X} = \sum_1^n \frac{T_i^2}{n_i} - \frac{T^2}{n}$$

4. Tỷ số tương quan:

$$\eta_{yx} = \eta = \sqrt{\frac{SS_{Y_X}}{SS_T}}$$

4.3.2. Đánh giá sự tồn tại của tỷ số tương quan

Để kiểm định sự tồn tại của tỷ số tương quan ta sử dụng phép nghiệm F:

$$F_{TN} = \frac{(\eta^2 - r^2)(n-k)}{(1-\eta^2)(k-2)} \qquad (4-7)$$

Trong đó: n là tổng số số liệu

k là số nhóm của đại lượng X

F_{TN} được so sánh với so với F_α ($k - 2$, $n - k$).

Nếu $F_{TN} \geq F_\alpha$ thì tồn tại quan hệ phi tuyến tính, ngược lại, $F_{TN} < F_\alpha$ thì tồn tại quan hệ tuyến tính.

Ví dụ ứng dụng

Kết quả theo dõi giữa thời gian sinh trưởng (TGST, ngày) và năng suất bông (g/cây) được ghi ở Bảng 4.3. Hãy đánh giá quan hệ giữa TGST với năng suất.

Bảng 4.3: TGST và năng suất (NS) cá thể

	Thời gian sinh trưởng, X (ngày)					
	98	99	100	101	102	103
NS cá thể, Y (g)	64,4	59,2	50,0	57,4	60,9	38,2
	39,7	51,7	42,4	50,3	50,4	37,4
	59,0	54,1	51,2		41,3	61,7
		60,5			67,1	64,0
					51,9	
					54,1	
n_i	3	4	3	2	6	4
T_i	163,1	225,5	143,6	107,7	325,7	201,3

(tiếp)	Thời gian sinh trưởng, X (ngày)					
	104	105	106	107	108	109
NS cá thể, Y (g)	66,0 67,2 72,2 73,8	59,8 48,1	66,7 53,5 56,7	63,3 63,0 75,1	64,9 68,2	45,8 63,5 51,3
n_i	4	2	3	3	2	3
T_i	279,2	107,9	176,9	201,4	133,1	160,6
$\Sigma n_i = 39$; $\Sigma T_i = 2.226,0$						

Theo số liệu bảng 4.3:

1. Tổng bình phương tổng số:

$$SS_T = \sum_1^n y_{ij}^2 - \frac{T^2}{n} = [(64,4)^2 + (29,7)^2 + \ldots + (51,3)^2] - (2.226,0)^2/39 = 3.685,769$$

2 . Tính tổng bình phương Y do sai khác giữa các x_j (SS_X):

$$SS_{Y_X} = \sum_1^n \frac{T_i^2}{n_i} - \frac{T^2}{n} = [(163,1)^2/3 + (225,5)^2/4 + \ldots + (160,6)^2/3] - (2.226,0)^2/39 = 1.726,818$$

3. Tỷ số tương quan:

$$\eta_{yx} = \eta = \sqrt{\frac{SS_{Y_X}}{SS_T}} = \sqrt{\frac{1.726,818}{3.685,769}} = 0,684$$

4. Kiểm định η

Tính hệ số tương quan theo mục 4.2.2.1 ta được:

$r = 0,281$

Thay η và r vào công thức (4 – 7), ta có:

$$F_{TN} = \frac{\left(\eta^2 - r^2\right)\left(n-k\right)}{\left(1-\eta^2\right)\left(k-2\right)}$$

$$= \frac{\left[\left(0,684\right)^2 - \left(0,281\right)^2\right]\left(39-12\right)}{\left[1-\left(0,684\right)^2\right]\left(12-2\right)} = 1,98$$

$F_{TN} = 1,98 < F_{0,05}(10, 27) = 2,20$

Như vậy, với dung lượng mẫu n = 39, giữa thời gian sinh trưởng và năng suất bông chưa cho thấy có quan hệ phi tuyến tính.

Kiểm tra mức ý nghĩa của hệ số tương quan tuyến tính:

$$T = r_{xy}\sqrt{\frac{n-2}{1-r_{xy}^2}} = 1,78 < t_{0,05} = 2,03$$

Kết quả này cũng cho thấy chưa đủ cơ sở để kết luận về sự tồn tại mối quan hệ tuyến tính giữa thời gian sinh trưởng với năng suất.

Như vậy, để xác định đúng mối quan hệ này cần phải có dung lượng mẫu lớn hơn.

4.3.4. Chuyển hàm hồi quy phi tuyến tính về dạng tuyến tính

Để ước lượng các tham số của mô hình phi tuyến tính, một số hàm có thể chuyển thành dạng tuyến tính và thực hiện như mô hình tuyến tính, cuối cùng lại chuyển về dạng ban đầu.

1. Hàm lũy thừa

$$y = ax_1^{b_1}x_2^{b_2}...x_n^{b_n}$$

Logarit hai vế:

$$\ln y = \ln a + b_1 \ln x_1 + b_2 \ln x_2 + \ldots + b_n \ln x_n$$

Đặt $\ln y = z$, $\ln a = b_0$, $\ln x_i = k_i$ và chuyển thành hàm tuyến tính:

$$z = b_0 + b_1 k_1 + b_2 k_2 + \ldots + b_n k_n$$

2. Hàm mũ

$$y = e^{b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n}$$

Logarit hai vế:

$$\ln y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n$$

Đặt $\ln y = z$ ta có:

$$z = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n$$

3. Hàm parabol

$$y = ax^2 + bx + c$$

Đặt $x^2 = k_1$, $x = k_2$ ta có:

$$y = ak_1 + bk_2 + c$$

4. Hàm lorarit

$$\ln y = b_0 + b_1 \ln x_1 + b_2 \ln x_2 + \ldots + b_n \ln x_n$$

Đặt $\ln y = z$, $\ln x_i = k_i$ ta có :

$$z = z = b_0 + b_1 k_1 + b_2 k_2 + \ldots + b_n k_n$$

4.4. QUAN HỆ GIỮA CÁC DẤU HIỆU ĐỊNH TÍNH

4.4.1. Hai dấu hiệu phân phối số liệu hai chiều

Dạng phân phối số liệu như sau:

B / A	b_1	b_2	Tổng
a_1	n_1	n_2	$n_1 + n_2$
a_2	n_3	n_4	$n_3 + n_4$
Tổng	$n_1 + n_3$	$n_2 + n_4$	$n_1 + n_2 + n_3 + n_4$

Mối quan hệ giữa A và B được xác định bởi hệ số F:

$$F = \frac{n_1 n_4 - n_2 n_3}{\sqrt{(n_1 + n_2)(n_1 + n_3)(n_2 + n_4)(n_3 + n_4)}}$$

Ví dụ: Nghiên cứu mối quan hệ giữa độ lông của lá bông với mức độ kháng rầy xanh theo kết quả điều tra ghi ở Bảng 4.4 dưới đây.

Bảng 4.4: Kết quả điều tra mức độ kháng rầy trên các giống bông có mức độ lông của lá khác nhau

Đơn vị: số giống

Mức độ lông của lá (A)	Mức độ kháng rầy (B)		
	Kháng vừa	Kháng	Rất kháng
Ít	3	0	0
Vừa	7	3	0
Nhiều	5	8	5
Rất nhiều	11	5	3

Với số liệu trong bảng có thể nghiên cứu mối quan hệ giữa mức độ lông của lá bông với mức độ kháng rầy qua rất nhiều mối quan hệ. Chẳng hạn:

- Giữa lông ít và lông vừa với kháng vừa và kháng
- Giữa lông ít và lông nhiều với kháng vừa và kháng
- Giữa lông ít và lông rất nhiều với kháng vừa và kháng

- Giữa lông vừa và lông nhiều với kháng vừa và kháng

...

Giữa lông ít và lông nhiều với rất kháng và kháng vừa ta có bảng sau:

Mức độ lông của lá (A)	Mức độ kháng rầy (B)		
	Kháng vừa	Rất kháng	Tổng
Ít	3	0	3
Nhiều	5	5	10
Tổng	8	5	13

Ta có: $$F = \frac{(3)(5)-(0)(5)}{\sqrt{(3)(8)(5)(10)}} = 0,43$$

Giữa lông ít và lông rất nhiều với rất kháng và kháng vừa ta có bảng sau:

Mức độ lông của lá (A)	Mức độ kháng rầy (B)		
	Kháng vừa	Rất kháng	Tổng
Ít	3	0	3
Rất nhiều	3	11	14
Tổng	6	11	17

Ta có: $$F = \frac{(3)(11)-(0)(3)}{\sqrt{(3)(6)(11)(14)}} = 0,63$$

Kết quả cho thấy, lá càng nhiều lông thì càng kháng rầy.

4.4.2. Tương quan theo thứ hạng

Kết quả theo dõi lượng bón phân tổng hợp (tạ/ha) và năng suất lúa (tấn/ha) như sau:

Lượng bón	0,8	0,9	1,0	1,1	1,2	1,3	1,4	1,5	1,6
Năng suất	5,0	5,1	5,3	5,4	5,4	5,5	5,6	5,6	5,7

Với cách tính hệ số tương quan theo dấu hiệu định lượng, từ số liệu ở bảng ta được: r = 0,97.

Để tính hệ số tương quan theo thứ hạng, năng suất được xếp hạng theo thứ tự hạng lượng phân bón từ nhỏ đến lớn như sau:

Lượng bón	1	2	3	4	5	6	7	8	9
Năng suất	1	2	3	4	4	6	7	7	9

Hệ số tương quan thứ hạng được tính theo công thức của Spearman (1904):

$$r_s = 1 - \frac{6\sum d_i^2}{n\left(n^2 - 1\right)}$$

trong đó: d_i là hiệu số của hai thứ hạng theo cặp, n là số cặp hạng. Ví dụ: $d_1 = 1 - 1 = 0$, $d_5 = 5 - 4 = 1$.

Từ số liệu phân hạng ta có:

$$r_s = 1 - \frac{6\left[(0)^2 + (0)^2 + \ldots(1)^2 + (0)^2\right]}{9\left[(9)^2 - 1\right]} = 0,98$$

Hai kết quả khác nhau không nhiều và đều được chấp nhận.

Phần 2

BỐ TRÍ THÍ NGHIỆM VÀ XỬ LÝ SỐ LIỆU

Chương 5
NHỮNG VẤN ĐỀ CHUNG

5.1. CÁC LOẠI THÍ NGHIỆM

Thí nghiệm là một hình thức nghiên cứu khoa học mà con người tạo ra những hiện tượng, tìm hiểu, phát hiện bản chất và nguồn gốc của hiện tượng, xác minh những quy luật trong tự nhiên để giải đáp các mục tiêu đặt ra.

Trong sinh học nói chung, nông nghiệp nói riêng, theo điều kiện và tính chất thí nghiệm, người ta chia thí nghiệm thành hai nhóm: Thí nghiệm trong chậu, trong phòng (dưới đây gọi là thí nghiệm trong phòng) và thí nghiệm ngoài đồng.

- *Thí nghiệm trong phòng* là thí nghiệm mà cây được gieo trồng ở trong chậu trên những giá thể, dung dịch hoặc cấy trong môi trường dinh dưỡng ở các lọ (bình). Với thí nghiệm trong phòng, con người có thể kiểm soát được các loại giá thể, dung dịch, môi trường nền, làm cho các nhân tố ngoài yếu tố thí nghiệm không ảnh hưởng đến kết quả nghiên cứu. Vì vậy có thể nghiên cứu và kết luận một cách chính xác và nhanh chóng tác động của từng yếu tố và tương tác giữa các yếu tố thí nghiệm.

Kết quả của các thí nghiệm trong phòng là luận cứ khoa học giải thích các hiện tượng tự nhiên. Tuy nhiên, do mang tính nhân tạo nên không thể áp dụng trực tiếp kết quả nghiên cứu vào sản xuất mà phải qua thử nghiệm lại trên đồng ruộng.

- *Thí nghiệm ngoài đồng* là thí nghiệm được tiến hành trong điều kiện tự nhiên trên đồng ruộng. Cây trồng chịu tác động không chỉ do yếu tố thí nghiệm mà còn bị ảnh hưởng của môi trường sống như đất đai, khí hậu thời tiết. Trong điều kiện đồng ruộng, những kết quả của các thí nghiệm trong phòng được kiểm chứng. Tuy bị tác động của nhân tố môi trường nhưng thí nghiệm đồng ruộng có thể nghiên cứu ảnh hưởng của từng yếu tố, tương tác giữa các yếu tố thí nghiệm, ảnh hưởng của môi trường cũng như tương tác giữa các yếu tố thí nghiệm với môi trường. Kết quả của các thí nghiệm đồng ruộng mang tính thực tiễn cao nên được chuyển giao trực tiếp ra sản xuất.

Theo mục đích nghiên cứu, người ta chia thí nghiệm thành 2 loại: *Thí nghiệm một yếu tố* và *thí nghiệm nhiều yếu tố*.

Với thí nghiệm trong phòng, trên cùng một điều kiện ngoại cảnh, vị trí đặt chậu, bình (nghiệm thức) không ảnh hưởng đến sự sinh trưởng phát triển của cây trồng. Sự khác biệt giữa các chậu trong cùng một nghiệm thức là do sai số gây ra (lẽ ra các giá trị các chậu phải bằng nhau). Trong trường hợp này, nếu sắp xếp các chậu theo kiểu khối như thí nghiệm đồng ruộng thì sự khác biệt giữa các "khối" cũng do sai số gây ra (lẽ ra giá trị của các "khối" này phải bằng nhau). Vì vậy, *với thí nghiệm trong phòng, khi điều kiện ngoại cảnh như nhau, việc sắp xếp khối, hàng, cột một cách "tùy tiện" để loại trừ sự sai khác này ra khỏi tổng bình phương độ lệch thực chất là loại trừ "một phần sai số" ra khỏi "sai số" và đều dẫn đến kết luận sai lầm*. Tuy nhiên, với thí nghiệm trong phòng có thể bố trí theo kiểu khối, ở đó các nghiệm thức trong từng khối có cùng một chế độ và giữa các khối có chế độ khác nhau. Và như vậy, sự khác nhau giữa các khối được xem như sự khác nhau giữa các mức của một yếu tố.

Với thí nghiệm đồng ruộng, các loại thí nghiệm khác nhau có các kiểu bố trí khác nhau.

Có thể biểu thị hai cách phân loại và các kiểu bố trí thí nghiệm thông thường trong nghiên cứu cây trồng ở Hình 5 - 1.

Hình 5-1: Các loại thí nghiệm và các kiểu bố trí trong nghiên cứu cây trồng

5.2. CÁC YÊU CẦU CỦA THÍ NGHIỆM ĐỒNG RUỘNG

Để thực hiện tốt thí nghiệm đồng ruộng, cần thực hiện các yêu cầu được xem như là các nguyên tắc sau đây.

5.2.1. Nguyên tắc điển hình

Thí nghiệm phải được bố trí trong điều kiện điển hình:

- Điển hình về khí hậu thời tiết.

- Điển hình về đất đai.

- Điển hình về kỹ thuật canh tác.

Nguyên tắc này bảo đảm ý nghĩa thực tiễn các kết quả nghiên cứu. Chỉ có thực hiện thí nghiệm trong điều kiện điển hình thì các kết luận rút ra từ thí nghiệm mới được áp dụng rộng rãi trong sản xuất.

5.2.2. Nguyên tắc khác nhau duy nhất

Người ta còn gọi nguyên tắc này là nguyên tắc "một sai khác" hay nguyên tắc "đồng nhất". Nội dung cơ bản của nó là ngoài yếu tố thí nghiệm, tất cả các yếu tố đất đai, kỹ thuật canh tác, ... tác động lên cây trồng trong thí nghiệm phải đồng nhất. Nếu thực hiện một cách triệt để nguyên tắc này, sự khác biệt trong thí nghiệm chỉ do yếu tố thí nghiệm gây ra. Trong thực tế đồng ruộng không thể khống chế một cách triệt để tất các tác động khác nhau của môi trường sống lên cây trồng mà chỉ có thể hạn chế tác động của các yếu tố chính như chọn đất đồng đều, tưới nước chủ động, ánh sáng đều (không bị che bóng), áp dụng các biện pháp canh tác cùng thời gian, cùng chế độ (ngoài yếu tố thí nghiệm) vv.

Cần nói thêm rằng, có những thí nghiệm mà mỗi nghiệm thức là một tổ hợp các yếu tố thì ngoài tổ hợp các yếu tố đó các tác động khác phải như nhau. Ở khía cạnh khác, nguyên tắc này không hoàn toàn cứng nhắc. Ví dụ, có một thí nghiệm so sánh các giống mà một vài giống trong đó yêu cầu mức mật độ, phân bón khác các giống còn lại thì không thể bố trí cùng một loại khoảng cách, một loại mật độ và một mức phân bón như nhau cho mọi giống. Tính đồng nhất ở đây là áp dụng các biện pháp kỹ thuật tối thích theo yêu cầu của mỗi giống. Do mức đầu tư khác nhau nên ngoài việc so sánh năng suất phải so sánh hiệu quả kinh tế của các giống.

5.2.3. Nguyên tắc về độ chính xác

Nguyên tắc này đòi hỏi các thí nghiệm phải đạt được độ chính xác nhất định. Vì vậy các thí nghiệm phải được xử lý thống kê, tính độ chính xác và kiểm định sự khác biệt giữa các nghiệm thức ở mức tin cậy nhất định. Thường thì các thí nghiệm đồng ruộng được bố trí một số lần nhắc lại và nhờ phép phân tích ANOVA để loại bỏ các yếu tố chi phối, xác định sai số thí nghiệm, làm rõ và đánh giá chính xác tác động của các yếu tố thí nghiệm.

5.2.4. Nguyên tắc khẳng định kết quả

Theo nguyên tắc này, những kết quả của thí nghiệm là chính xác, tức là nếu thực hiện lại thí nghiệm trong điều kiện tương tự thì kết quả sẽ không thay đổi. Với các thí nghiệm trong phòng việc kiểm tra lại kết quả có thể thực hiện được, nhưng với các thí nghiệm đồng ruộng, để khẳng định kết quả thí nghiệm phải được tiến hành một số vụ tương tự.

Ngoài ra, để làm tốt thí nghiệm đồng ruộng điều cần thiết là phải điều tra, nghiên cứu, đánh giá khu đất trước khi tiến hành thí nghiệm, đảm bảo lô đất được chon có độ đồng đều cao, đủ điều kiện để làm thí nghiệm.

Các nguyên tắc trên đây cũng đúng cho các thí nghiệm trong phòng.

5.3. CÁC THÀNH PHẦN CỦA MỘT THÍ NGHIỆM ĐỒNG RUỘNG

5.3.1. Nghiệm thức

5.3.1.1. Nghiệm thức thí nghiệm

Tùy mục tiêu nghiên cứu để đặt các nghiệm thức thích hợp sao cho phân tích và đánh giá được các yếu tố thí nghiệm. Với thí nghiệm một yếu tố nghiệm thức có thể là các giống, có thể là các mật độ gieo trồng, các mức phân bón khác nhau. Mức phân có thể là lượng bón khác nhau của một loại phân, cũng có thể là các tổ hợp phân này độc lập với tổ hợp phân khác. Với thí nghiệm nhiều yếu tố, nghiệm thức là sự phối hợp của các yếu tố theo các cách khác nhau sao cho có thể phân tích được ảnh hưởng của từng yếu tố và tương tác giữa chúng. Các nghiệm thức phải rõ ràng, không để các yếu tố ngoài thí nghiệm chi phối, đan xen không làm rõ yếu tố thí nghiệm.

5.3.1.2. Nghiệm thức đối chứng

Nghiệm thức đối chứng thông thường là biện pháp đang phổ biến nhất trong sản xuất của vùng. Trong thí nghiệm giống thì lấy giống sản xuất đại trà trong vùng làm giống đối chứng, trong thí nghiệm kỹ thuật thì lấy biện pháp kỹ thuật tiến bộ đang được áp dụng rộng rãi trong sản xuất làm đối chứng. Trong chọn giống, tùy loại thí nghiệm, có thể có một, hai hoặc ba giống đối chứng. Ví dụ, trong quá trình chọn giống, ngoài giống sản xuất đại trà có thể sử dụng giống khởi đầu, giống bố mẹ làm đối chứng. Với mục đích chọn giống kháng sâu bệnh, ngoài giống sản xuất đại trà người ta còn lấy giống chuẩn kháng, giống chuẩn nhiễm làm đối chứng và tương tự cho chọn giống chịu mặn, chịu phèn, chịu hạn.

5.3.2. Ô thí nghiệm

Ô thí nghiệm là nơi đặt một nghiệm thức trong một lần lặp lại.

5.3.2.1. Diện tích ô

Diện tích ô lớn nhỏ tùy loại cây trồng. Theo quan điểm thống kê, ô thí nghiệm đại diện cho tổng thể của một nghiệm thức ở một lần lặp lại. Để ước lượng tổng thể, số lượng cá thể phải đủ lớn. Trong bố trí thí nghiệm, không phải ô càng lớn càng tốt, mà là ô vừa đủ số lượng cá thể để ước lượng tổng thể trong ô. Một khi ô quá lớn thì vị trí giữa các ô càng xa nhau, nếu đất kém đồng đều thì giữa các ô trong một lần lặp lại có nguy cơ khác nhau về đất làm cho thí nghiệm kém chính xác. Ngược lại nếu ô quá nhỏ thì không đủ số cây cần thiết làm cho việc ước lượng tổng thể thiếu chính xác. Những cây nhỏ như cây lạc, chỉ cần diện tích ô 7 – 8 m^2 (gieo 4 hàng dài 5 m), ở hai hàng giữa đã có đủ cá thể để ước lượng và đánh giá tổng thể. Với cây ngô diện tích ô chỉ cần 15 m^2 (gieo 4 hàng dài 5 m). Đây cũng là cơ sở để xác định diện tích ô trong Quy phạm khảo nghiệm giống cây trồng, Bộ Nông nghiệp và PTNT (với cây lạc 6,5 – 7,5 m^2, cây lúa - 10 m^2, cây ngô – 15 m^2, cây bông – 25 m^2).

Với cây lâu năm, mỗi cây chiếm một diện tích lớn nên nếu ô bố trí nhiều cây, không thể tránh khỏi nguy cơ không đồng đều về đất giữa các ô. Nếu trong khu thí nghiệm đất tương đối đồng đều, mỗi ô có thể 10 – 20 cây cho loại cây to và 20 – 40 cây cho các loại cây nhỏ, bố trí tối thiểu 3 lần lặp lại. Tuy nhiên nếu đất không đều thì mỗi ô có số cây ít hơn, có thể 1 – 3 cây cho loại cây lớn, 3 – 5 cây cho loại cây nhỏ nhưng phải bố trí 5 – 8 lần lặp lại ở những vị trí đất khác nhau mới đảm bảo độ chính xác của thí nghiệm.

Diện tích ô còn phụ thuộc loại thí nghiệm và mục đích nghiên cứu.

Cần phải nhấn mạnh rằng, khi ô thí nghiệm đã có đủ số lượng cây cần thiết để ước lượng và đánh giá tổng thể thì việc bố trí diện tích ô tăng lên không làm tăng độ chính xác của thí nghiệm, trong nhiều trường lại làm giảm độ chính xác. Trong trường hợp này, độ chính xác của thí nghiệm phụ thuộc vào độ đồng đều của đất thí nghiệm, số lần lặp lại, việc kiểm soát các yếu tố thí nghiệm và các yếu tố ngoài thí nghiệm.

5.3.2.2. Hình dạng ô

Để các ô bố trí trong một lần lặp lại ít khác nhau về đất, các ô phải có hình chữ nhật dài và các ô nằm cạnh nhau theo chiều dài ô. Hình dạng của một lần lặp lại (khối) càng vuông, ô của các nghiệm thức càng gần nhau. Trong trường hợp đất có độ biến thiên đều về hai phía vuông góc, số nghiệm thức không nhiều, có thể bố trí theo phương pháp ô vuông Latin (Latin Square Design). Để kích thước lần lặp lại theo hàng bằng kích thước lần lặp lại theo cột thì ô thí nghiệm có hình vuông.

5.3.3. Lần lặp lại, khối

5.3.3.1. Nguyên tắc bố trí lần lặp lại và các ô trong một khối

Lần lặp lại là cách bố trí để hạn chế sai số do đất không đều. Với thí nghiệm trong phòng, lần lặp lại của một nghiệm thức có thể đặt ngẫu nhiên trong khu thí nghiệm. Với thí nghiệm đồng ruộng kiểu hoàn toàn ngẫu nhiên, do đất không thể hoàn toàn đồng nhất nên các lần lặp lại của một nghiệm thức cần bố trí rãi đều trên khu thí nghiệm.

Với kiểu bố trí lần lặp lại theo khối đầy đủ, mỗi khối đất phải có độ đồng đều nhất định. Trong một khối các nghiệm thức (các ô) đều có cơ hội như nhau về vị trí đặt ô, sự biểu hiện khác nhau là do yếu tố thí nghiệm gây ra. Tùy chiều biến thiên của đất để bố trí khối lặp lại. Nếu khu đất đồng đều, hướng ô bố trí theo hướng vuông góc với chiều dài hay chiều rộng của khu đất. Nếu đất không đều thì chọn những lô đất đồng đều trong khu làm các khối lặp lại. Do tổng bình phương sai khác giữa các khối lặp lại sẽ được loại trừ trong phân tích ANOVA nên điều kiện đất đai của các khối không nhất thiết phải giống nhau. Nếu đất trong khu thí nghiệm biến thiên đều về một phía (có thể theo độ phì hoặc theo độ dốc) thì các khối lặp lại bố trí vuông góc với chiều biến thiên.

Nguyên tắc chung về sắp xếp ô:

- Các nghiệm thức trong một lần lặp lại (khối đầy đủ) cần bố trí gần nhau. Với số nghiệm thức không nhiều, một khối có thể bố trí trong một băng (dãy), nhưng nếu có quá nhiều nghiệm thức thì nên bố trí khối có nhiều băng, sao cho các ô trong khối được gần nhau.

- Các ô của mỗi nghiệm thức phải bố trí sao cho được nằm đều trong khu thí nghiệm, ở mọi độ phì khác nhau.

Tùy kiểu bố trí thí nghiệm mà sắp xếp vị trí các ô. Theo kiểu hoàn toàn ngẫu nhiên (Completely Randomized Design – CRD) hoặc khối đầy đủ ngẫu nhiên (Randomized Complete Block Design – RCBD), vị trí các ô có thể bố trí ngẫu nhiên bằng cách gắp thăm, hay theo chủ ý của nhà nghiên cứu sao cho các ô của một nghiệm thức được rải đều trong lô thí nghiệm. Các thí nghiệm bố trí theo kiểu đặc biệt thì thì các được đặt theo sơ đồ bố trí sẵn. Với các thí nghiệm nghiên cứu di truyền, chọn giống, do các nghiệm thức có cơ hội như nhau ở mọi vị trí trong một khối nên có thể bố trí tuần tự. Khi số lượng nghiệm thức quá nhiều, tốt nhất là bố trí theo kiểu mạng lưới (Lattice Design), đặc biệt là mạng cân bằng từng phần (Parially Balanced Lattices) vừa giảm bớt lần lặp lại vừa đảm bảo độ chính xác của thí nghiệm.

- Tạo điều kiện để các nghiệm thức cần ở gần nhau thì gần nhau.

5.3.3.2. Số lần lặp lại

Nếu một nghiệm thức chỉ đặt ở một chỗ, biểu hiện của nghiệm thức bị chi phối không chỉ do yếu tố thí nghiệm mà còn do đất tại nơi đặt ô. Khi đó giá trị của ô bằng giá trị thực do tác động của yếu tố thí nghiệm cộng với giá trị do môi trường đất tạo nên. Để loại bỏ tác động của đất không đều cần phải có lần lặp lại. Khi đó giá trị thực của chỉ tiêu theo dõi do tác động của yếu tố thí nghiệm là số trung bình của các lần lặp lại, còn ở nơi đất thuận lợi giá trị cao hơn trung bình (do đất) và ở nơi không thuận lợi giá trị sẽ thấp hơn trung bình (cũng do đất). Trong di truyền, số liệu ở các lần lặp lại gọi là giá trị kiểu hình còn số trung bình từ các lần lặp lại là giá trị kiểu gen do tác động của yếu tố thí nghiệm. Vậy cần bao nhiêu lần lặp lại để đánh giá chính xác số trung bình.

Nếu coi giá trị trung bình của một ô như là một mẫu, theo lý thuyết ước lượng số trung bình, để có độ chính xác mong đợi $s_{\bar{x}}(\%)$ và hệ số biến động của sai số là CV(%), số lần lặp lại sẽ là:

$$r = \left(\frac{CV(\%)}{s_{\bar{x}}(\%)} \right)^2$$

trong đó: $CV(\%) = \frac{\sqrt{MS_e}}{\overline{x}} \times 100$ và $s_{\overline{x}}(\%) = \frac{s_{\overline{x}}}{\overline{x}} \times 100$, $s_{\overline{x}}$ là sai số trung bình: $s_{\overline{x}} = \sqrt{\frac{MS_e}{r}}$

Nếu hệ số biến động sai số CV(%) của thí nghiệm đồng ruộng 10%, muốn có độ chính xác không vượt quá 5% thì tối phải có: $r_{min} = \left(\frac{10}{5}\right)^2 = 4$ lần lặp lại. Nếu bố trí 5 lần lặp lại thì $s_{\overline{x}}(\%)$ sẽ là 4,5%, còn chỉ có 3 lần lặp lại thì $s_{\overline{x}}(\%)$ sẽ là 5,8 %.

Như vậy nếu số lần lặp lại tăng lên thì sai số giảm xuống, ngược lại nếu giảm số lần lặp lại xuống thì sai số tăng lên. Từ thực tế đồng ruộng và yêu cầu về độ chính xác của thí nghiệm có thể tính được số lần lặp lại cần thiết.

5.3.4. Số cây theo dõi

Số cây theo dõi phụ thuộc vào yêu cầu của thí nghiệm cho các chỉ tiêu cụ thể.

5.3.4.1. Để ước lượng số trung bình

Thực tế nghiên cứu trên đồng ruộng cho thấy, dù đất khá đồng đều nhưng ngay trên một giống thuần chủng các giá trị theo dõi của các cá thể có sự biến động khá lớn, thông thường CV khoảng 10 – 15%, thậm chí lên tới 30 - 35% hoặc cao hơn. Theo lý thuyết ước lượng điểm, nếu CV khoảng 15%, để có sai số 5%, số cây cần theo dõi trong một ô khoảng 10 cây. Nếu CV tăng lên khoảng 20% và yêu cầu sai số không đổi thì số cây theo dõi khoảng 15 cây.

Như vậy, để xác định số cây theo dõi phải dựa vào độ biến động của chỉ tiêu quan sát và độ chính xác của phép ước lượng cần đạt được. Có những loại chỉ tiêu chỉ cần theo dõi 3 cây, thậm chí chỉ 1 cây, nhưng cũng có những chỉ tiêu cần phải theo dõi nhiều hơn mới đạt được độ chính xác mong muốn. Thông thường để ước lượng giá trị trung bình ô thí nghiệm cho các chỉ tiêu sinh trưởng, người ta theo dõi 10 cây, còn với năng suất thì tối thiểu là 30 cây.

5.3.4.2. Để ước lượng số trung bình và phương sai

Trong công tác nghiên cứu, đặc biệt về lĩnh vực di truyền, chọn giống, có những ô thí nghiệm là quần thể phân ly (do lai, đột biến hoặc sản phẩm của kỹ thuật di truyền). Theo luật Student, để ước lượng đúng số trung bình và phương sai tổng thể số cây theo dõi tối thiểu là 30 cây. Do yêu cầu về độ chính xác, trong nhiều trường hợp ở đời bố mẹ, F_1 số cây cần theo dõi $\geq$ 50 cây, ở các thế hệ phân ly $\geq$ 100 cây.

5.4.4.3. Xác định vị trí cây theo dõi

Nguyên tắc chung trong việc chọn cây theo dõi là chọn các cây điển hình cho ô. Có thể áp dụng phương pháp chọn mẫu ngẫu nhiên các cây bất kỳ trong ô hoặc định các cây trên đường chéo góc. Thông thường, ở hàng theo dõi đầu tiên người ta bỏ một số cây gần bờ, sau đó theo dõi liên tục số cây cần thiết, đến hàng kế tiếp thì làm tương tự theo chiều ngược lại. Đó cũng được coi là một cách ngẫu nhiên. Tuy nhiên khi gặp những cây không điển hình do gieo dặm hoặc do tác động nào đó từ bên ngoài làm mất tính điển hình thì phải loại bỏ và có thể thay bằng cây khác nếu số cây theo dõi quá ít.

Với quần thể không đồng nhất, không chọn cây điển hình mà chọn quần thể mẫu điển hình. Để đánh giá được chính xác số trung bình và phương sai quần thể mẫu đại diện cho tổng thể, ngay từ đầu phải đảm bảo khoảng cách gieo trồng, không để mất cây hoặc gieo dặm làm sai lệch kết quả thí nghiệm. Việc định cây theo dõi có thể áp dụng các phương pháp đã nêu.

5.3.5. Mẫu phân tích

Với các thí nghiệm đồng ruộng, trước khi thu hoạch ô thường thu mẫu phân tích. Cách lấy mẫu, số lượng và độ lớn mẫu tùy thuộc vào yêu cầu cho từng loại thí nghiệm và từng loại cây trồng.

Nguyên tắc chung đối với việc thu mẫu là đảm bảo tính điển hình. Kết quả phân tích các chỉ tiêu từ mẫu phải điển hình cho các phép thử theo yêu cầu của từng thí nghiệm.

Chương 6

PHÂN TÍCH PHƯƠNG SAI THÍ NGHIỆM MỘT YẾU TỐ

Tính chất: Với loại thí nghiệm này, chỉ duy nhất một yếu tố thay đổi được bố trí vào các ô riêng biệt với các mức khác nhau. Ví dụ như các giống khác nhau, các liều phân bón khác nhau, các loại thuốc trừ sâu khác nhau và các mật độ gieo khác nhau. Tất cả các biện pháp kỹ thuật khác đều áp dụng như nhau cho mọi ô.

Các kiểu bố trí: Tùy điều kiện và mục đích, thí nghiệm một yếu tố có thể bố trí các kiểu: Hoàn toàn ngẫu nhiên, khối đầy đủ ngẫu nhiên, ô vuông Latin, chữ nhật Latin hay kiểu mạng lưới. Mỗi kiểu có mô hình toán và cách tính toán khác nhau.

6.1. THÍ NGHIỆM MỘT YẾU TỐ KIỂU HOÀN TOÀN NGẪU NHIÊN (Completely Randomized Design – CRD)

6.1.1. Bố trí thí nghiệm

Kiểu thí nghiệm này đòi hỏi điều kiện ngoài yếu tố thí nghiệm hoàn toàn đồng nhất. Điều này chỉ có thể thực hiện được trong chậu, trong dung dịch và trong phòng thí nghiệm. Ở đó mỗi lần nhắc lại (mỗi chậu, mỗi bình) của mỗi một nghiệm thức đều có cơ hội như nhau ở mọi chỗ đặt, nghĩa là ta có thể đổi chỗ tùy ý cho nhau giữa các chậu bất kỳ trong một nghiệm thức và giữa các nghiệm thức mà không hề thay đổi tình hình sinh trưởng, phát triển các chậu, bình và do đó không ảnh hưởng đến kết quả thí nghiệm. Do vị trí đặt chậu, đặt bình (như nuôi cấy mô) không ảnh hưởng đến kết quả thí nghiệm nên để tiện theo dõi ta có thể đặt các chậu của một nghiệm thức nằm cạnh nhau.

Với thí nghiệm đồng ruộng, kiểu thí nghiệm này đòi hỏi phải chọn khu đất đồng đều. Tuy nhiên, dù đất đồng đều đến đâu vẫn không thể đáp ứng đòi hỏi "lý tưởng" như thí nghiệm trong chậu, trong phòng. Vì vậy số ô cho mỗi nghiệm thức phải đủ lớn và bố trí sao cho được nằm đều trong khu thí nghiệm.

1 C	2 A	3 D	4 B
5 D	6 B	7 A	8 C
9 A	10 C	11 B	12 D
13 B	14 D	15 C	16 A
17 D	18 A	19 B	20 C

Hình 6.1: Sơ đồ thí nghiệm đồng ruộng một yếu tố, kiểu CRD, 4 nghiệm thức (A, B, C, D) mỗi nghiệm thức 5 lần lặp lại

Trong thí nghiệm kiểu CRD, số lần lặp lại cho các nghiệm thức có thể bằng nhau hay khác nhau.

6.1.2. Phân tích phương sai (ANOVA)

6.1.2.1. Mô hình toán học và nguyên lý tính toán

Với thí nghiệm một yếu tố kiểu CRD, giá trị của ô nghiệm thức mức *i* ở lần lặp thứ *j* (X_{ij}) được tính theo mô hình:

$$X_{ij} = m + t_i + e_{ij}$$

trong đó: m là giá trị trung bình toàn thí nghiệm, t_i là giá trị do nghiệm thức i tạo ra (đóng góp vào giá trị ô), e_{ij} là sai số ngẫu nhiên phân phối chuẩn tắc $N(0, \sigma^2)$.

Theo mô hình này tổng bình phương độ lệch giữa giá trị mỗi ô với giá trị trung bình thí nghiệm $SS_{TO} = \sum_1^n \left(x_{ij} - m\right)^2$ tạo bởi hai nguyên nhân:

- Sai khác giữa các nghiệm thức (giữa các mức) gây ra (Treatment);

- Sai khác giữa các ô trong mỗi nghiệm thức - Sai số (Error).

Trong trường hợp số ô lặp lại của các nghiệm thức bằng nhau, bảng phân tích phương sai có dạng

Nguồn biến động	Bậc tự do (DF)	Tổng bình phương (SS)	Phương sai (MS)	F_{TN}	$F_{bảng}$
Nghiệm thức (t)	t - 1	SS_t	SS_t/df_t	MS_t/MS_e	$F_\alpha^{(df_t, df_e)}$
Sai số (e)	$t(r$ -1)	SS_e	SS_e/df_e		
Tổng số (TO)	rt - 1	SS_T			

Trong đó, t là số nghiệm thức, r là số ô lặp lại của mỗi nghiệm thức.

Tổng bình phương tổng số: $SS_{TO} = \sum_1^n x_{ij}^2 - CF$;

$CF = \dfrac{\left(\sum_1^n x_{ij}\right)^2}{n}$ gọi là số điều chỉnh (hay phần trừ)

Tổng bình phương nghiệm thức: $SS_t = \frac{\sum_{i=1}^{t} T_i^2}{r} - CF$

(T_i là tổng của các ô nghiệm thức i $\left(i = \overline{1,t}\right)$

Tổng bình phương sai số: $SS_e = SS_{TO} - SS_t$

Bậc tự do tổng số: $df_{TO} = rt - 1$

Bậc tự do nghiệm thức: $df_t = t - 1$

Bậc tự do sai số: $df_e = (rt - 1)\text{-}(t - 1) = t(r - 1)$

Các công thức tính phương sai của nghiệm thức và sai số được ghi trong bảng.

Phương sai nghiệm thức (MS_t) phát sinh do sự khác biệt giữa các nghiệm thức. Các nghiệm thức càng khác nhau MS_t càng lớn.

Để xác minh MS_t thực sự là do sai khác giữa các nghiệm thức gây ra hay do sai số ta sử dụng phép nghiệm F (đã được đề cập ở mục 3.3.1, chương 3).

$$F_{TN} = \frac{MS_t}{MS_e}$$

$f_{bảng}$ ($f_\alpha^{(df_t, df_e)}$) được tra trong bảng F với độ tự do tử số là df_t và độ tự do mẫu số là df_e. Có thể sử dụng phần mềm Excel để biết giá trị $f_{bảng}$ theo cú pháp =finv(α, df_t,df_e) với α = 0,05; 0,01 và 0,001.

Nếu $F_{TN} < f_\alpha$: các nghiệm thức khác nhau không có ý nghĩa thống kê ở đô tin cậy 1 - α. Kết luận này đã thỏa mãn yêu cầu của thí nghiệm, không cần tiếp tục tính toán.

Nếu $F_{TN} \geq f_{0,05}$: có sự khác nhau giữa các nghiệm thức với độ tin cậy 95%. Người ta thường đánh dấu * vào giá trị F_{TN} để biểu thị có sự sai khác giữa phương sai nghiệm thức khi so sánh với phương sai sai số ở mức $\alpha = 0,05$.

Nếu $F_{TN} \geq f_{0,01}$: có sự khác nhau giữa các nghiệm thức với độ tin cậy 99%. Khi đó giá trị F_{TN} được đánh ** để biểu thị có sự sai khác giữa phương sai nghiệm thức khi so sánh với phương sai sai số ở mức $\alpha = 0,01$.

Nếu $F_{TN} \geq f_{0,001}$: có sự khác nhau giữa các nghiệm thức với độ tin cậy 99,9%, tương tự giá trị F_{TN} được đánh *** để biểu thị có sự sai khác giữa phương sai nghiệm thức khi so sánh với phương sai sai số ở mức $\alpha = 0,001$.

Thông thường, người ta thường trắc nghiệm F ở mức $\alpha = 0,05$ hoặc $\alpha = 0,01$.

Có thể kiểm định sự sự khác biệt giữa các nghiệm thức qua xác suất sai lầm của F, ký hiệu là P hay Prob. Khi P > 0,05, tức là $F_{TN} < F_{0,05}$, các nghiệm thức khác nhau không có ý nghĩa thống kê, độ tin cậy < 95%, khi $0,01 < P \leq 0,05$, tức là $F_{TN} \geq F_{0,05}$, các nghiệm thức khác nhau có ý nghĩa thống kê với độ tin cậy 95% và khi $P < 0,01$ ($F_{TN} > F_{0,01}$), các nghiệm thức khác nhau rất có ý nghĩa với độ tin cậy 99%. Vấn đề này sẽ được nói lại trong mục 9.2.2, chương 9.

Khi $F_{TN} \geq F_{0,05}$ hay $F_{0,01}$ ($P < 0,05$ hay $< 0,01$), để phân biệt sự khác nhau giữa các nghiệm thức ta sử dụng "thước" LSD_α($\alpha = 0,05$ hay 0,01).

$$LSD_\alpha = t_\alpha Sd;\ Sd = \sqrt{\frac{2MS_e}{r}}$$

Khi hiệu số giá trị hai nghiệm thức $\geq LSD_\alpha$ thì chúng khác nhau và nếu $< LSD_\alpha$ thì sự khác nhau giữa chúng là không đủ tin cậy.

Sự khác biệt giữa các nghiệm thức có thể biểu thị bằng bảng phối hợp hai chiều hiệu số giữa các nghiệm thức và đánh mức ý nghĩa.

Sai số trung bình: $s_{\bar{x}} = \sqrt{\frac{MS_e}{r}}$;

Sai số tương đối: $s_{\bar{x}}(\%) = \frac{s_{\bar{x}}}{m} \times 100$

Hệ số biến động: $CV(\%) = \frac{\sqrt{MS_e}}{m} \times 100$

Có thể phân tích phương sai trong trường hợp số ô lặp lại của các nghiệm thức bằng nhau hoặc không bằng nhau trong hai ví dụ ứng dụng sau đây.

6.1.2.2. Ví dụ ứng dụng

- ***Số ô lặp lại bằng nhau***

Ví dụ 1: So sánh ảnh hưởng của các mức phối hợp phân bón N : P_2O_5 : K_2O khác nhau đến năng suất cà chua trồng trong chậu theo số liệu ở Bảng 6.1 dưới đây.

Bảng 6.1: Năng suất quả cà chua (g/chậu)

Nghiệm thức@	Năng suất (x_{ij})				Tổng NT (T)	Trung bình
1(ĐC)	454	470	430	500	1.854	463,5
2	502	550	490	507	2.049	512,2
3	601	670	550	607	2.428	607,0
4	407	412	475	402	1.626	424,0
5	418	470	460	412	1.760	440,0
			Tổng cộng		9.787	*m*=489,35

@ Tỷ lệ N : P_2O_5 : K_2O của các nghiệm thức 1 – 1:1:1 (ĐC); 2 – 1:2:1; 3 – 1:2:2; 4 – 2:1:1; 5 – 2:2:1

Giải:

Ở đây, số nghiệm thức t = 5, số chậu cho mỗi nghiệm thức bằng nhau r = 4, tổng số chậu n = 20.

1. Tính số điều chỉnh CF và tổng các bình phương

$$CF = \frac{\left(\sum_{1}^{n} x_{ij}\right)^2}{n} = (9.787)^2 : 20 = 4.789.268{,}45$$

$$SS_{TO} = \sum_{1}^{n} x_{ij}^2 - CF$$

$$= (454)^2 + (470)^2 + \ldots + (460)^2 + (412)^2$$

$$- 4.789.268{,}45 = 104.940{,}550$$

$$SS_t = \frac{\sum_{i=1}^{t} T_i^2}{r} - CF$$

$$= \frac{(1.854)^2 + (2.049)^2 + \ldots + (1.760)^2}{4}$$

$$- 4.789.268{,}45 = 86.960{,}800$$

$$SS_e = SS_{TO} - SS_t = 104.940{,}550 - 86.960{,}800$$

$$= 17.979{,}750$$

2. Tính phương sai các nguồn biến động

Bậc tự do: df_{TO} = 20 – 1 = 19; df_t = 5 – 1 = 4; df_e = 15

$$MS_t = \frac{SS_t}{df_t} = \frac{86.960{,}80}{4} = 21.740{,}20$$

$$MS_e = \frac{SS_e}{df_e} = \frac{17.979{,}75}{15} = 1.198{,}65$$

Bảng 6.2: Kết quả phân tích phương sai

Nguồn biến động	DF	SS	MS	F_{TN}	$F_{bảng}$	
					0,05	0,01
Ngh. thức	4	86.960,80	21.740,20	18,14	3,06	4,89
Sai số	15	17.979,75	1.198,65			
Tổng số	19	104.940,55				

$F_{TN} >> F_{0,01}^{(4,15)} = 4,89$ khẳng định rằng giữa các mức phối hợp phân bón có sự khác nhau rất rõ ràng. Ở thí nghiệm này P = $1,29 \times 10^{-5} << 0,01$. Ở các bảng sau sẽ sử dụng P thay cho sử dụng $F_{bảng}$.

3. Sai số thí nghiệm và so sánh các nghiệm thức

Sai số trung bình:

$$s_{\bar{x}} = \sqrt{\frac{MS_e}{r}} = \sqrt{\frac{1.198,65}{4}} = 17,3$$

Sai số tương đối:

$$s_{\bar{x}}(\%) = \frac{s_{\bar{x}}}{m} \times 100 = \frac{17,3}{489,35} \times 100 = 3,3$$

Hệ số biến động sai số:

$$CV(\%) = \frac{\sqrt{MS_e}}{m} \times 100 = 7,1$$

Sai khác giữa hai nghiệm thức:

$$Sd = \sqrt{\frac{2MS_e}{r}} = \sqrt{\frac{2 \times 1.198,65}{4}} = 24,5$$

Với $t_{0,05} = 2,13$, $t_{0,01} = 2,95$ và $t_{0,001} = 4,07$

$$LSD_{0,05} = 2,13 \times 24,5 = 52,2$$

$$LSD_{0,01} = 2,95 \times 24,5 = 72,1$$

$$LSD_{0,001} = 4,07 \times 24,5 = 99,7$$

Kết quả phân hạng trung bình các nghiệm thức theo $LSD_{0,05}$, $LSD_{0,01}$ như sau:

Mức ý nghĩa α = 0,05		Mức ý nghĩa α = 0,01	
NT3	607,0 A	NT3	607,0 A
NT2	512,3 B	NT2	512,3 B
NT1(ĐC)	463,5 BC	NT1(ĐC)	463,5 BC
NT5	440,0 C	NT5	440,0 C
NT4	424,0 C	NT4	424,0 C

Do độ dài các "thước" khác nhau: $LSD_{0,05} < LSD_{0,01}$ nên mức 0,05 thường nhiều hạng hơn mức 0,01.

Căn cứ vào các ký tự phân hạng để kết luận cao thấp của các nghiệm thức và so sánh với nghiệm thức đối chứng. Các giá trị trung bình có ký tự giống nhau không khác nhau về ý nghĩa thống kê.

Trong thực tế người ta thường phân hạng theo mức ý nghĩa 0,05 và 0,01 và sử dụng dấu * (α = 0,05), ** (α = 0,01) và ns để so sánh các nghiệm thức với nhau.

NT	Trung bình	Chênh lệch và mức ý nghĩa			
		2	3	4	5
1	463,5	-48.8^{ns}	-143.5^{**}	39.5^{ns}	23.5^{ns}
2	512,3		-94.7^{**}	88.3^{**}	72.3^{**}
3	607,0			183^{**}	167^{**}
4	424,0				-16^{ns}
5	440,0				

- ***Số ô lặp lại không bằng nhau***

Ví dụ 2: So sánh ảnh hưởng 4 mức bón đạm khác nhau đến năng suất kiều mạch trồng trong chậu với số liệu ở Bảng 6.3 dưới đây.

Bảng 6.3: Năng suất kiều mạch (g/chậu)

NT	Năng suất (x_{ij})						Σ (T)	TB
1	16,0	17,2	14,4	15,8	-	-	63,4	15,8
2	29,4	30,4	30,3	28,1	-	-	118,2	29,5
3	26,0	29,2	26,7	27,1	26,0	28,1	163,1	27,2
4	25,3	24,8	26,1	23,2	25,7	24,0	149,1	24,8
					Tổng cộng		943,8	
						m =	24,7	

Giải:

Số nghiệm thức t = 4, nếu gọi r_i là số chậu của các nghiệm thức ($i = \overline{1,4}$), thì $r_1 = r_2 = 4$, $r_3 = r_4 = 6$, tổng số chậu n = 20.

1. Tính số điều chỉnh CF và tổng các bình phương

$$CF = \frac{\left(\sum_{1}^{n} x_{ij}\right)^2}{n} = (943{,}8)^2 : 20 = 12.191{,}922$$

$$SS_{TO} = \sum_{1}^{n} x_{ij}^2 - CF$$

$$= (16{,}0)^2 + (17{,}2)^2 + ... + (25{,}7)^2 +$$

$$(24{,}0)^2 - 12.191{,}922 = 465{,}758$$

$$SS_t = \sum_{i=1}^{t} \frac{T_i^2}{r_i} - CF$$

$$= \left[\frac{(64,3)^2}{4} + \frac{(118,2)^2}{4} + \frac{(163,1)^2}{6} + \frac{(149,1)^2}{6} \right]$$

$$- 12.191,922 = 444,515$$

$SS_e = SS_{TO} - SS_t = 465,758 - 444,515 = 21,243$

2. Tính phương sai các nguồn biến động

Bậc tự do: $df_{TO} = 20 - 1 = 19$; $df_t = 4 - 1 = 3$; $df_e = df_T - df_t = 19 - 3 = 16$

$$MS_t = \frac{SS_t}{df_t} = \frac{444,51}{3}$$

$$= 148,17$$

$$MS_e = \frac{SS_e}{df_e} = \frac{21,24}{16} = 1,33$$

Bảng 6.4: Kết quả phân tích phương sai

Nguồn biến động	DF	SS	MS	F_{TN}	Prob.
Nghiệm thức	3	444,51	148,17	111,60	0,0000
Sai số	16	21,24	1,33		
Tổng số	19	465,76			

P < 0,01 khẳng định rằng giữa các mức bón đạm có sự khác nhau rất rõ ràng.

3. Cách tính sai số trung bình, sai số tương đối và hệ số biến động sai số tương tự như ví dụ 1.

Do số chậu lặp lại của NT1, NT2 khác với NT3, NT4 nên sai khác giữa NT1 và NT2, NT3 và NT4 được tính như sau:

$$Sd_{(1,2)} = \sqrt{\frac{2MS_e}{r_1}} = \sqrt{\frac{2 \times 1,33}{4}} = 0,81$$

$$Sd_{(3,4)} = \sqrt{\frac{2MS_e}{r_3}} = \sqrt{\frac{2 \times 1,33}{6}} = 0,66$$

và giữa NT1, NT2 với NT3, NT4 sẽ là:

$$Sd_{(12,34)} = \sqrt{MS_e \frac{r_1 + r_3}{r_1 r_3}} = \sqrt{(1,33)\frac{4+6}{4 \times 6}} = 0,74$$

Với $t_{0,05} = 2,12$, $t_{0,01} = 2,92$

Giữa NT1 với NT2:

$LSD_{0,05} = 2,12 \times 0,81 = 1,7$

$LSD_{0,01} = 2,92 \times 0,81 = 2,4$

Giữa NT3 với NT4:

$LSD_{0,05} = 2,12 \times 0,66 = 1,4$

$LSD_{0,01} = 2,92 \times 0,66 = 1,9$

Giữa NT1, NT2 với NT4:

$LSD_{0,05} = 2,12 \times 0,74 = 1,6$

$LSD_{0,01} = 2,92 \times 0,74 = 2,2$

Việc so sánh theo phân hạng A, B, ... hay sử dụng dấu * (α = 0,05), ** (α = 0,01), *** (α = 0,001) và ns để xác nhận các mức ý nghĩa tiến hành tương tự như ví dụ 1.

Một số lưu ý

- Trong kiểu bố trí CRD chỉ có 2 nguồn biến động tạo ra tổng bình phương độ lệch tổng số SS_{TO}. Đó là: tổng bình phương do sự sai khác giữa các nghiệm thức và tổng bình phương do sai số. Nếu các lần lặp lại của mỗi nghiệm thức bằng nhau, tức là không có sai số thì tổng bình phương nghiệm thức bằng tổng bình phương tổng số.

- Trong thí nghiệm đồng ruộng, nếu các ô của một nghiệm thức không nằm rải đều trong khu thí nghiệm mà tập trung ở một lô nào đó, giá trị trung bình các nghiệm thức bị chi phối bởi yếu tố đất làm cho kết quả thí nghiệm bị sai lệch.

6.2. THÍ NGHIỆM MỘT YẾU TỐ KIỂU KHỐI ĐẦY ĐỦ NGẪU NHIÊN (Randomized Complete Block Design – RCBD)

6.2.1. Bố trí thí nghiệm

Theo kiểu thí nghiệm khối đầy đủ ngẫu nhiên, mỗi lần lặp lại được bố trí trong các khối hoàn chỉnh (đầy đủ - Complete Block), tức là trong một lần lặp lại có đầy đủ ô của tất cả các nghiệm thức và tạo thành một khối. Trong một khối, mỗi ô của một nghiệm thức đều có cơ hội như nhau ở mọi chỗ (ngẫu nhiên), nghĩa là ta có thể đổi chỗ tùy ý cho nhau giữa các ô trong một khối. Điều đó nói rằng điều kiện ngoài yếu tố thí nghiệm, đặc biệt là đất đai phải ảnh hưởng như nhau cho mọi ô trong khối. Trong thực tế dù đất có "lý tưởng" đến đâu cũng không thể nào đồng đều một cách tuyệt đối, vì vậy các ô của các nghiệm thức trong một khối cần phải bố trí gần nhau và các ô của mỗi nghiệm thức phải được nằm đều trong khu thí nghiệm để hạn chế sự sai khác do đất gây ra.

Kiểu bố trí RCBD cho phép sự khác nhau giữa các khối. Cùng với tổng bình phương độ lệch của yếu tố thí nghiệm, tổng bình phương độ lệch giữa các khối là một trong các nguồn biến động sẽ được loại trừ ra khỏi tổng bình phương tổng số để xác định tổng bình phương sai số thí nghiệm.

Sau đây là một số sơ đồ bố trí thí nghiệm.

I	2	3	1	5	4
II	5	1	4	2	3
III	4	3	2	1	5
IV	1	3	2	5	4

Hình 6.2: Sơ đồ RCBD
5 nghiệm thức, 4 lần lặp lại

I	4	1	6	7	5	3	2
II	2	5	3	1	6	4	7
III	3	6	7	4	2	1	5
IV	7	2	1	5	3	6	4

Hình 6.3:
Sơ đồ RCBD 7 nghiệm thức 4 khối (lần lặp lại)

I			II		
6	1	3	2	3	4
2	5	4	5	6	1
III			**IV**		
4	1	6	3	2	4
5	3	2	1	5	6

Hình 6.4:
Sơ đồ RCBD 6 nghiệm thức
4 lần lặp lại

Lần lặp I							Lần lặp II							Lần lặp III						
1	16	6	9	4	21	11	12	8	14	5	7	17	19	20	2	10	18	15	13	3
5	12	8	13	2	20	18	4	15	10	1	6	3	9	17	19	11	7	21	16	14
17	10	3	15	7	14	19	13	2	20	16	11	18	21	6	4	8	12	1	5	9

Hình 6.5: Sơ đồ RCBD, 21 nghiệm thức, 3 lần lặp lại

6.2.2. Phân tích phương sai

6.2.2.1. Mô hình toán học và nguyên lý tính toán

Với thí nghiệm một yếu tố kiểu RCBD, giá trị của ô nghiệm thức *i* ở lần lặp thứ *j* (X_{ij}) được tính theo mô hình:

$$X_{ij} = m + t_i + r_i + e_{ij}$$

trong đó: m là giá trị trung bình toàn thí nghiệm, a_i là giá trị do nghiêm thức *i* tạo ra (đóng góp vào giá trị ô), r_i giá trị của lần lặp lại (khối) *i* đóng góp, e_{ij} là sai số phân phối chuẩn tắc $N(0, \sigma^2)$.

Theo mô hình toán học kiểu RCBD, tổng bình phương độ lệch giữa giá trị mỗi ô với giá trị trung bình thí nghiệm $\sum_1^n \left(x_{ij} - m \right)^2$ do ba nguyên nhân:

- Sự khác nhau giữa các nghiệm thức (giữa các mức) gây ra (Treatment);

- Sự sai khác giữa các khối (Block) do khác nhau về đất gây ra;

- Sai khác ngẫu nhiên gây ra - sai số (Error).

Bảng phân tích phương sai có dạng

Nguồn	DF	SS	MS	F_{TN}	Prob.
Khối (r)	r - 1	SS_r	SS_r/df_r	MS_r/MS_e	
Nghiệm thức	t - 1	SS_t	SS_t/df_t	MS_t/MS_e	
Sai số (e)	$df_t \times df_r$	SS_e	SS_e/df_e		
Tổng số (TO)	rt - 1	SS_T			

Trong bảng: t là số nghiệm thức, r là số khối;

$$CF = \left(\sum_1^n x_{ij} \right)^2 / n$$

Tổng bình phương tổng số:

$$SS_{TO} = \sum_{1}^{n} x_{ij}^2 - \left(\sum_{1}^{n} x_{ij}\right)^2 / n;$$

Tổng bình phương khối:

$$SS_r = \frac{\sum_{i=1}^{r} R_i^2}{t} - CF$$

(R_i là tổng của các ô trong khối i $\left(i = \overline{1,r}\right)$

Tổng bình phương nghiệm thức:

$$SS_t = \frac{\sum_{i=1}^{t} T_i^2}{r} - CF$$

(T_i là tổng của các ô nghiệm thức i $\left(i = \overline{1,t}\right)$)

Tổng bình phương sai số:

$$SS_e = SS_{TO} - SS_t - SS_r$$

Bậc tự do tổng số: $df_{TO} = rt - 1$

Bậc tự do nghiệm thức: $df_t = t - 1$

Bậc tự do khối: $df_r = r - 1$

Bậc tự do sai số: $df_e = (r - 1)(t - 1)$

Các công thức tính phương sai của nghiệm thức và sai số được ghi trong bảng.

Việc kiểm tra sai khác giữa các nghiệm thức, sai số trung bình của thí nghiệm, sai số tương đối, hệ số biến động và so sánh giữa các nghiệm thức được tính toán như thí nghiệm kiểu hoàn toàn ngẫu nhiên.

6.2.2.2. Vai trò của lần lặp lại

Mặc dù sự sai khác do các lần lặp lại gây ra đã bị loại trừ ra khỏi sai số thí nghiệm, nhưng có thể đánh giá độ đồng đều về đất thí nghiệm qua phép nghiệm F cho phương sai khối lặp lại.

$$\text{F (lần lặp lại)} = \frac{\text{MS}_\text{r}}{\text{MS}_\text{e}}$$

Nếu F (lần lặp lại) < $F_{bảng}$ (tra với độ tự do tử số là df_r và độ tự do mẫu số là df_e) thì sự sai khác giữa các lần lặp lại là không đáng kể, đất tương đối đồng đều. Nếu F(lần lặp lại) ≥ $F_{bảng}$ thì đất không được đồng đều. Có thể sử dụng phần mềm Excel để biết giá trị $F_{bảng}$ theo cú pháp =finv(α, df_r,df_e) với α = 0,05 và α = 0,01.

Để đánh giá tác dụng làm giảm sai số của thí nghiệm kiểu RCBD so với kiểu CRD ta tính từ mô hình của kỳ vọng MS trong bảng sau:

Nguồn	DF	MS	Kỳ vọng MS
Khối	$r - 1$	MS_r	$\sigma_e^2 + t\sigma_r^2$
Nghiệm thức	$t - 1$	MS_t	$\sigma_e^2 + r\sigma_t^2$
Sai số	$(t-1)(r-1)$	MS_e	σ_e^2

Từ kỳ vọng MS ta có:

$$\sigma_r^2 = \frac{\text{MS}_\text{r} - \text{MS}_\text{e}}{t} \qquad (\text{MS}_\text{e} = \sigma_e^2)$$

Nếu coi thí nghiệm được bố trí kiểu CRD thì kỳ vọng MS của sai số CRD ($\sigma^2_{e(CRD)}$) sẽ là:

$$\sigma^2_{e(CRD)} = \sigma_r^2 + \sigma_e^2 = \frac{\text{MS}_\text{r} - \text{MS}_\text{e}}{t} + \text{MS}_\text{e}$$

$$= \frac{\text{MS}_\text{r} + (t-1)\text{MS}_\text{e}}{t}$$

Nhờ bố trí theo khối đầy đủ ngẫu nhiên, phương sai sai số đã giảm (r_e,%):

$$r_e(\%) = \frac{\sigma^2_{e(CRD)} - \sigma^2_e}{\sigma^2_{e(CRD)}} \times 100$$

$$= \frac{MS_r - MS_e}{MS_r + (t-1)MS_e} \times 100$$

Đó chính là ưu điểm của kiểu bố trí RCBD so với kiểu bố trí CRD.

6.2.2.3. Ví dụ ứng dụng

Ví dụ: So sánh năng suất bố mẹ và tổ hợp bông lai F_1 (2008) được bố trí thí nghiệm khối đầy đủ ngẫu nhiên 3 lần lặp lại với số liệu ở Bảng 6.5.

Trong thí nghiệm, số nghiệm thức t = 21, số lần lặp lại r = 3, tổng số ô (n) = 63.

1. Tính số điều chỉnh CF và các tổng các bình phương

$$CF = \frac{\left(\sum_1^n x_{ij}\right)^2}{n} = (1.578,4)^2 : 63 = 39.545,183$$

$$SS_{TO} = \sum_1^n x_{ij}^2 - CF = (30,3)^2 + (19,3)^2 + ...$$

$$+ (19,7)^2 + (15,8)^2 - 39.545,183 = 1.827,477$$

$$SS_r = \frac{\sum_{i=1}^r R_i^2}{t} - CF = \frac{(510,3)^2 + (567,6)^2 + (500,5)^2}{21}$$

$$- 39.545,183 = 125,107$$

Bảng 6.5: Năng suất các giống bông bố mẹ và con lai F_1

TT	Nghiệm thức	Năng suất (x_{ij}, tạ/ha)			Tổng NT (T)	Trung bình
		I	II	III		
1	1. C92-52	30,3	19,3	22,2	71,8	23,9
2	2. C118A	12,8	15,9	12,5	41,2	13,7
3	3. S02-13	25,7	28,0	27,1	80,8	26,9
4	4. TM1	24,5	22,1	18,5	65,1	21,7
5	5. NH04-2	16,7	17,0	15,2	48,9	16,3
6	6. 1354	25,1	19,7	15,8	60,6	20,2
7	1 × 2	25,1	26,6	27,0	78,7	26,2
8	1 × 3	27,5	32,1	32,5	92,1	30,7
9	1 × 4	24,7	30,1	29,5	84,3	28,1
10	1 × 5	23,6	31,5	33,7	88,8	29,6
11	1 × 6	24,1	30,7	26,7	81,5	27,2
12	2 × 3	26,1	28,4	24,0	78,5	26,2
13	2 × 4	20,5	25,9	27,6	74,0	24,7
14	2 × 5	23,3	22,5	18,7	64,5	21,5
15	2 × 6	25,7	30,5	26,8	83,0	27,7
16	3 × 4	35,3	30,3	28,8	94,4	31,5
17	3 × 5	22,5	36,8	21,9	81,2	27,1
18	3 × 6	26,3	30,4	25,5	82,2	27,4
19	4 × 5	25,7	29,6	18,5	73,8	24,6
20	4 × 6	24,5	27,4	27,8	79,7	26,6
21	5 × 6	20,3	32,8	20,2	73,3	24,4
Tổng		510,3	567,6	500,5	1.578,4	
Trung bình chung, *m*						25,05

$$SS_t = \frac{\sum_{i=1}^{t} T_i^2}{r} - CF$$

$$= \frac{(71,8)^2 + (78,7)^2 + \ldots + (73,3)^2 + (60,6)^2}{3}$$

$$- 39.545,183 = 1.159,397$$

$$SS_e = SS_{TO} - SS_t - SS_r = 1.827,477 -$$

$$1.159,397 - 125,107 = 542,973$$

2. Tính phương sai các nguồn biến động

Bậc tự do: $df_{TO} = 63 - 1 = 62$; $df_r = 3 - 1 = 2$;

$$df_t = 21 - 1 = 20; \quad df_e = 40$$

Áp dụng cách tính như các ví dụ trên ta có các giá trị phương sai nghiệm thức (MS_t), phương sai khối (MS_r) và phương sai sai số (MS_e) ở Bảng 6.6 sau.

Bảng 6.6: Kết quả phân tích phương sai

Nguồn	DF	SS	MS	F_{TN}	Prob.
Khối	2	125,107	62,553	4,61	
Nghiệm thức	20	1159,397	57,970	4,27	0,0000
Sai số	40	542,973	13,574		
Tổng số	62	1827,477			

P < 0,01 cho thấy giữa các nghiệm thức có sự khác nhau rất rõ ràng.

3. Tính sai số thí nghiệm và so sánh các nghiệm thức

Sai số trung bình: $s_{\overline{x}} = \sqrt{\dfrac{MS_e}{r}} = \sqrt{\dfrac{13,574}{3}} = 2,13$

Sai số tương đối: $s_{\overline{x}}(\%) = \dfrac{s_{\overline{x}}}{m} \times 100 = \dfrac{2,13}{25,05} \times 100 = 8,5$

Hệ số biến động sai số:

$$CV(\%) = \frac{\sqrt{13,574}}{25,05} \times 100 = 14,7$$

Sai khác giữa hai nghiệm thức:

$$Sd = \sqrt{\frac{2MS_e}{r}} = \sqrt{\frac{2 \times 13,574}{3}} = 3,01$$

với $t_{0,05} = 2,02$, $t_{0,01} = 2,70$

$$LSD_{0,05} = 2,02 \times 3,01 = 6,1$$
$$LSD_{0,01} = 2,70 \times 3,01 = 8,2$$

Kết quả phân hạng trung bình các nghiệm thức theo $LSD_{0,05}$ và phân hạng Duncan với $\alpha = 0,05$ như sau:

Nghiệm thức	Giá trị	Phân hạng	
		$LSD_{0,05}$	Duncan (0,05)[@]
16	31,5	A	A
8	30,7	AB	AB
10	29,6	ABC	AB
9	28,1	ABC	ABC
15	27,7	ABCD	ABC
18	27,4	ABCDE	ABCD
11	27,2	ABCDE	ABCD
17	27,1	ABCDE	ABCD
3	26,9	ABCDE	ABCD
20	26,6	ABCDE	ABCD
7	26,2	ABCDEF	ABCD
12	26,2	ABCDEF	ABCD
13	24,7	BCDEF	ABCD
19	24,6	CDEF	ABCD
21	24,4	CDEF	ABCD
1	23,9	CDEF	BCD
4	21,7	DEFG	CDE
14	21,5	EFG	CDE
6	20,2	FG	DE
5	16,3	GH	EF
2	13,7	H	F

@: Phân hạng Duncan được trình bày ở mục 9.2.3 (chương 9)

4. Vai trò của lần lặp lại

$$F(\text{lần lặp lại}) = \frac{MS_r}{MS_e} = \frac{62,553}{13,574} = 4,61$$

với $F_{0,05}^{(2,40)} = 3,23$ và $F_{0,01}^{(2,40)} = 5,18$ thì có thể kết luận rằng sự khác nhau giữa các khối là có ý nghĩa thống kê ở độ tin cậy 95%.

Việc bố trí kiểu RCBD phương sai sai số đã giảm xuống (%):

$$r_e(\%) = \frac{62,553 - 13,574}{62,553 + 20 \times 13,574} \times 100 = 14,7\%$$

Ở một số thí nghiệm, phương sai sai số giảm xuống một cách đáng kể. Ví dụ, kết quả phân tích phương sai năng suất (tạ/ha) thí nghiệm 5 giống lúa mì bố trí theo kiểu RCBD, 4 lần lặp lại như sau (Dospekhov, 1985):

Nguồn biến động	DF	MS	F_{TN}	$F_{bảng}$	
				0,05	0,01
Khối	3	11,04			
Nghiệm thức	4	48,56	30,35	3,25	5,41
Sai số	12	1,60			
Tổng số	19				

Ở thí nghiệm này, phương sai sai số kiểu RCBD đã giảm đi (%) so với phương sai sai số kiểu CRD là:

$$r_e(\%) = \frac{MS_r - MS_e}{MS_r + (t-1)MS_e} \times 100$$

$$= \frac{11,04 - 1,60}{11,04 + 4 \times 1,60} \times 100 = 54,1\%$$

Như vậy, nhờ bố trí kiểu RCBD nên phương sai sai số đã giảm hơn một nửa so với phương sai sai số theo kiểu CRD.

6.3. THÍ NGHIỆM MỘT YẾU TỐ KIỂU Ô VUÔNG LA TINH (Latin Square Design)

6.3.1. Bố trí thí nghiệm

Bố trí thí nghiệm kiểu ô vuông La tinh không đơn thuần là kiểu bố trí mà số khối lặp lại nhiều bằng số nghiệm thức mà chính là cách đặt ô, sao cho ngoài lần lặp lại theo hàng như kiểu RCBD còn có cả lần lặp lại theo cột. Đặc trưng chính của bố trí thí nghiệm kiểu ô vuông La tinh là khả năng khai thác nguồn biến động theo hai hướng: hướng hàng và hướng cột.

Như vậy, để bố trí kiểu ô vuông La tinh đất thí nghiệm phải đều theo hai chiều ngang dọc.

Về vị trí đặt ô, có thể sử dụng phương pháp chọn ngẫu nhiên để chọn vị trí ô theo hàng và cột nhưng các ô của cùng một nghiệm thức không được nằm trên cùng một hàng hay trên cùng một cột.

Tuy đất chọn để bố trí kiểu ô vuông La tinh là đồng đều nhưng không thể đồng đều một cách lý tưởng. Vì vậy, để hạn chế sai số của đất do hàng dài hơn cột hoặc cột dài hơn hàng thì kích thước các lần lặp theo hàng bằng kích thước các lần lặp theo cột, tức là khoảng cách từ ô đầu đến ô cuối của hàng cũng bằng khoảng cách từ ô đầu đến ô cuối của cột, do vậy ô có hình vuông.

Sau đây là một số sơ đồ bố trí thí nghiệm.

3 × 3		
A	B	C
B	C	A
C	A	B

4 × 4

Cách 1			
B	A	D	C
A	B	C	D
C	D	B	A
D	C	A	B

Cách 2			
A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

Cách 3			
A	B	C	D
C	D	A	B
B	A	D	C
D	C	B	A

Hình 6.5: Sơ đồ kiểu ô vuông La tinh 3 × 3 và 4 × 4

Số cột 1	2	3	4	5	Số hàng
E	C	B	A	D	1
A	D	C	B	E	2
C	B	D	E	A	3
B	E	A	D	C	4
D	A	E	C	B	5

Hình 6.6: Sơ đồ kiểu ô vuông La tinh 5 × 5

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

Hình 6.7: Sơ đồ kiểu ô vuông La tinh 6 × 6

Số hàng	Số cột 1	2	3	4	5	6	7
1	A	B	C	D	E	F	G
2	D	E	F	G	A	B	C
3	F	G	A	B	C	D	E
4	B	C	D	E	F	G	A
5	E	F	G	A	B	C	D
6	G	A	B	C	D	E	F
7	C	D	E	F	G	A	B

Hình 6.8: Sơ đồ kiểu 7×7 (các ô cần nằm cạnh nhau theo hàng)

6.3.2. Phân tích phương sai (ANOVA)

6.3.2.1. Mô hình toán học và nguyên lý tính toán

Với thí nghiệm một yếu tố kiểu ô vuông La tinh, giá trị của ô nghiệm thức hàng thứ i , cột thứ j (X_{ij}) được tính theo mô hình:

$$X_{ij} = m + t_k + r_i + c_j + e_{ij}$$

trong đó: m là giá trị trung bình toàn thí nghiệm, t_i là giá trị do nghiệm thức k đóng góp, r_i là giá trị do hàng i đóng góp, c_j giá trị do cột j đóng góp, e_{ij} là sai số ngẫu nhiên phân phối chuẩn tắc $N(0, \sigma^2)$.

Theo mô hình này tổng bình phương độ lệch giữa giá trị mỗi ô với giá trị trung bình thí nghiệm SS_{TO} do bốn nguyên nhân:

- Sự khác nhau giữa các nghiệm thức (giữa các mức) gây ra (Treatment);

- Sự sai khác giữa các lặp lại theo hàng;

- Sự sai khác giữa các lặp lại theo cột;

- Sai khác ngẫu nhiên gây ra - sai số (Error).

Bảng phân tích phương sai có dạng:

Nguồn	DF	SS	MS	F_{TN}	Prob.
Hàng (r)	$t - 1^{(*)}$	SS_r	SS_r/df_r	MS_r/MS_e	
Cột (c)	$t - 1^{(*)}$	SS_c	SS_c/df_c	MS_c/MS_e	
Nghiệm thức	$t - 1^{(*)}$	SS_t	SS_t/df_t	MS_t/MS_e	
Sai số (e)	$t^2 - 3t + 2^{(*)}$	SS_e	SS_e/df_e		
Tổng số	$t^2 - 1$	SS_T			

(*) : Số nghiệm thức (t) = số hàng (r) = số cột (c)

Tổng bình phương tổng số:

$$SS_{TO} = \sum_{1}^{n} x_{ij}^2 - \left(\sum_{1}^{n} x_{ij} \right)^2 / n ;$$

Tổng bình phương của hàng:

$$SS_r = \frac{\sum_{i=1}^{r} R_i^2}{t} - CF$$

(R_i là tổng số của các ô trong hàng i $\left(i = \overline{1, t} \right)$

Tổng bình phương của cột:

$$SS_c = \frac{\sum_{i=1}^{c} C_i^2}{t} - CF$$

(C_i là tổng số của các ô trong cột i $\left(i = \overline{1, t} \right)$

Tổng bình phương nghiệm thức:

$$SS_t = \frac{\sum_{i=1}^{t} T_i^2}{r} - CF$$

Tổng bình phương sai số:

$$SS_e = SS_{TO} - SS_t - SS_r - SS_c$$

Bậc tự do tổng số: $df_{TO} = t^2 - 1$

Bậc tự do hàng = Bậc tự do cột: $df_r = df_c = t - 1$

Bậc tự do nghiệm thức: $df_t = t - 1$

Bậc tự do sai số: $df_e = t^2 - 3t + 2$

Các công thức tính phương sai của nghiệm thức và sai số được ghi trong bảng.

Việc kiểm tra sai khác giữa các nghiệm thức, sai số trung bình của thí nghiệm, sai số tương đối, hệ số biến động và so sánh giữa các nghiệm thức được tính toán như thí nghiệm kiểu khối đầy đủ ngẫu nhiên.

6.3.2.2. Vai trò của hàng và cột

Để xác nhận sự khác nhau giữa các hàng với nhau và sự khác nhau giữa các cột với nhau ta tính:

$$\text{F(hàng)} = \frac{MS_r}{MS_e}\text{ ; F(cột)} = \frac{MS_c}{MS_e}$$

và kiểm tra bằng cách so sánh với $F_{\alpha}^{(df_f, df_e)}$ (α = 0,05 và α = 0,01)

Nếu F(hàng) hoặc F(cột) hay cả hai đều lớn hơn $F_{0,05}^{(df_f, df_e)}$ thì chúng có vai trò thực sự chi phối tổng bình phương SS_{TO} và nhờ việc tách chúng ra khỏi SS_{TO} mà làm giảm sai số thí nghiệm.

Để đánh giá vai trò làm giảm sai số so với phương pháp CRD và RCBD hãy phân tích các vấn đề sau đây.

- Hiệu quả so sánh với phương pháp CRD:

Từ bảng phân tích phương sai, nếu coi kiểu bố trí ô vuông La tinh là một kiểu CRD thì tổng bình phương sai sai số là:

$$SS_{e(CRD)} = (t-1)(MS_r + MS_c) + (t^2 - 3t + 2)MS_e$$

và
$$MS_{e(CRD)} = \frac{SS_{e(CRD)}}{t(t-1)}$$

Phương sai sai số kiểu ô vuông La tinh đã giảm đi (%) so với phương sai sai số kiểu CRD:

$$r_e(\%) = \frac{MS_{e(CRD)} - MS_e}{MS_{e(CRD)}} \times 100$$

$$= \frac{(t-1)(\mathrm{MS_r} + \mathrm{MS_c}) - 2(t-1)\mathrm{MS_e}}{(t-1)(\mathrm{MS_r} + \mathrm{MS_c}) + (t^2 - 3t + 2)\mathrm{MS_e}} \times 100$$

- Hiệu quả so sánh với phương pháp RCBD:

Từ bảng phân tích phương sai, nếu coi kiểu bố trí ô vuông La tinh là một kiểu RCBD với lần lặp lại theo hàng thì tổng bình phương sai sai số ($\mathrm{SS_{e(RCBD_r)}}$) là:

$$\mathrm{SS_{e(RCBD_r)}} = (t-1)\mathrm{MS_c} + (t^2 - 3t + 2)\mathrm{MS_e}$$

và $$\mathrm{MS_{e(RCBD_r)}} = \frac{\mathrm{SS_{e(RCBD_r)}}}{(t-1)^2}$$

Phương sai sai số kiểu ô vuông La tinh đã giảm đi (%) so với phương sai sai số kiểu RCBD khi coi hàng là lần lặp lại:

$$\mathrm{r_e}(\%) = \frac{\mathrm{MS_{e(RCBD_r)}} - \mathrm{MS_e}}{\mathrm{MS_{e(RCBD_r)}}} \times 100$$

$$= \frac{(t-1)(\mathrm{MS_c} - \mathrm{MS_e})}{(t-1)\mathrm{MS_c} + (t^2 - 3t + 2)\mathrm{MS_e}} \times 100$$

Một cách tương tự, nếu coi cột là lần lặp lại thì phương sai sai số kiểu ô vuông La tinh sẽ giảm đi (%) so với phương sai sai số kiểu RCBD khi coi hàng là lần lặp lại:

$$\mathrm{r_e}(\%) = \frac{\mathrm{MS_{e(RCBD_c)}} - \mathrm{MS_e}}{\mathrm{MS_{e(RCBD_c)}}} \times 100$$

$$= \frac{(t-1)(\mathrm{MS_r} - \mathrm{MS_e})}{(t-1)\mathrm{MS_r} + (t^2 - 3t + 2)\mathrm{MS_e}} \times 100$$

Đó chính là ưu điểm của kiểu bố trí ô vuông La tinh so với kiểu CRD và RCBD.

6.1.2.2. Ví dụ ứng dụng

Ví dụ: So sánh năng suất 4 giống bông lai được bố trí thí nghiệm kiểu ô vuông La tinh với số liệu ở Bảng 6.7 dưới đây.

Bảng 6.7: Năng suất 4 giống bông lai A, B, C và D (tạ/ha)

Hàng	Cột				Σ (R_i)	Σ (T_i)	TB_t
	1	2	3	4			
1	24,6D	31,5B	27,3C	25,2A	108,6	103,0A	25,8
2	30,5B	26,0D	23,1A	28,7C	108,3	122,7B	30,7
3	25,4C	28,0A	24,9D	31,4B	109,7	105,0C	26,3
4	26,7A	23,6C	29,3B	23,5D	103,1	99,0D	24,8
Σ(C_j)	107,2	109,1	104,6	108,8			
$\Sigma\Sigma$					429,7	*m* =	26,9

Ở đây, số nghiệm thức (t) = số hàng (r) = số cột (c) = 4, tổng số ô (n) = t^2 = 16.

1. Tính số điều chỉnh CF và tổng các bình phương

$$CF = \left(\sum_{1}^{n} x_{ij} \right)^2 / n = (429,7)^2 : 16 = 11.540,131$$

Tổng bình phương tổng số:

$$SS_{TO} = \sum_{1}^{n} x_{ij}^2 - CF = (24,6)^2 + (31,5)^2 + ...$$

$$+ (29,3)^2 + (23,5)^2 - 11.540,131 = 116,879$$

Tổng bình phương hàng:

$$SS_r = \frac{\sum_{i=1}^{r} R_i^2}{t} - CF$$

$$= \frac{\left[(108,6)^2 + (108,3)^2 + \ldots + (103,1)^2\right]}{4}$$

$$- 11.540,131 = 6,507$$

Tổng bình phương cột:

$$SS_c = \frac{\sum_{i=1}^{c} C_i^2}{t} - CF$$

$$= \frac{\left[(107,2)^2 + (109,1)^2 + \ldots + (108,8)^2\right]}{4}$$

$$- 11.540,131 = 3,182$$

Tổng bình phương nghiệm thức:

$$SS_t = \frac{\sum_{i=1}^{t} T_i^2}{t} - CF$$

$$= \frac{\left[(103,0)^2 + (122,7)^2 + \ldots + (99,0)^2\right]}{4}$$

$$- 11.540,131 = 82,442$$

Tổng bình phương sai số:

$$SS_e = SS_{TO} - SS_r - SS_c - SS_t$$

$$= 116,879 - 6,507 - 3,182 - 82,442$$

$$= 24,749$$

2. Tính phương sai các nguồn biến động

Bậc tự do tổng số: $df_{TO} = t^2 - 1 = 15$

Bậc tự do hàng = bậc tự do cột: $df_r = df_c = t - 1 = 3$

Bậc tự do nghiệm thức: $df_t = t - 1 = 3$

Bậc tự do sai số: $df_e = t^2 - 3t + 2 = 16 - 12 + 2 = 6$

Áp dụng cách tính như các ví dụ trên ta có các giá trị phương sai nghiệm thức (MS_t), phương sai lần lặp lại (MS_r) và phương sai sai số (MS_e) ở Bảng 6.8 sau.

Bảng 6.8: Kết quả phân tích phương sai

Nguồn	DF	SS	MS	F_{TN}	Prob.
Hàng	3	6,507	2,169		
Cột	3	3,182	1,061		
Nghiệm thức	3	82,442	27,481	6,66	0,0245
Sai số	6	24,749	4,125		
Tổng số	15	116,879			

P < 0,05 cho thấy có sự khác nhau giữa các nghiệm thức ở độ tin cậy 95%.

3. Tính sai số thí nghiệm và so sánh các nghiệm thức

Sai số trung bình: $s_{\bar{x}} = \sqrt{\dfrac{MS_e}{r}} = \sqrt{\dfrac{4,125}{4}} = 1,02$

Sai số tương đối: $s_{\bar{x}}(\%) = \dfrac{s_{\bar{x}}}{m} \times 100 = \dfrac{1,02}{26,9} \times 100 = 3,8\%$

Hệ số biến động sai số: $CV(\%) = \dfrac{\sqrt{4,125}}{26,9} \times 100 = 7,6\%$

Sai khác giữa hai nghiệm thức:

$$Sd = \sqrt{\frac{2MS_e}{r}} = \sqrt{\frac{2 \times 4,125}{4}} = 1,44$$

với $t_{0,05}$ = 2,45, $LSD_{0,05}$ = 2,45 × 1,44 = 3,5

Kết quả so sánh sự chênh lệch giữa các nghiệm thức:

NT	TB (tạ/ha)	Chênh lệch và mức ý nghĩa		
		B	C	D
A	25,8	-4,92*	-0,50	1,00
B	30,7		4.43*	5.93*
C	26,3			1.50
D	24,8			

Kiểu phân hạng theo A, B:

NTB	30,7	A
NTC	26,3	B
NTA	25,5	B
NTD	24,8	B

Như vậy, giống B có năng suất cao nhất và cao hơn 3 giống còn lại (A, C, D) có ý nghĩa thống kê.

4. Vai trò của hàng và cột

F(hàng) và F(cột) đều nhỏ hơn 1 cho thấy đất không ảnh hưởng đến kết quả thí nghiệm nên không cần thiết so sánh giữa các phương pháp bố trí.

6.4. THÍ NGHIỆM MỘT YẾU TỐ KIỂU CHỮ NHẬT LA TINH (Latin Rectangular Design)

6.4.1. Bố trí thí nghiệm

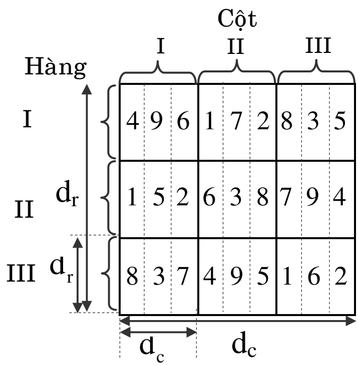
Kiểu bố trí chữ nhật La tinh là biến dạng của kiểu ô vuông La tinh, nghĩa là kiểu bố trí vẫn đáp ứng những đòi hỏi như phương pháp ô vuông La tinh nhưng ô không là hình vuông mà là hình chữ nhật.

Kiểu chữ nhật La tinh cho phép bố trí nhiều nghiệm thức nhưng số ô ít hơn nhiều so với kiểu ô vuông La tinh.

Để bố trí một thí nghiệm kiểu chữ nhật La tinh, số nghiệm thức (t), số hàng (r) và số cột (c) có sự ràng buộc sau đây:

- Số nghiệm thức (t) chia hết cho số hàng (r);

- Số cột chính bằng số hàng ($c = r$), mỗi cột chính có $\frac{t}{r}$ cột phụ sao cho kích thước của cột chính bằng kích thước của hàng. Sau đây là một số sơ đồ bố trí thí nghiệm.



Hình 6.9:
Kiểu chữ nhật La tinh
$3 \times 3 \times 3$
($d_r = d_c$; $d'_r = d'_c$)

I				II				III				
4	9	11	1	7	2	6	10	12	8	3	5	I
12	6	8	3	4	9	1	5	11	2	7	10	II
2	7	10	5	12	11	8	3	6	4	1	9	III

Hình 6.10:
Kiểu chữ nhật La tinh $3 \times 3 \times 4$

	I				II				III				IV			
I	7	9	12	3	15	16	1	6	2	4	14	13	11	8	5	10
II	8	6	5	14	11	2	4	7	12	3	1	10	9	16	13	15
III	2	11	10	4	5	13	9	8	6	15	16	7	3	14	12	1
IV	15	1	16	13	3	10	12	14	11	8	5	9	2	7	4	6

Hình 6.11: Kiểu chữ nhật La tinh 4 × 4 × 4

6.1.2. Phân tích phương sai

Mô hình toán học và phương pháp phân tích phương sai của kiểu bố trí chữ nhật La tinh hoàn toàn giống kiểu ô vuông La tinh.

Ví dụ: So sánh năng suất 9 giống ngô lai được bố trí thí nghiệm kiểu chữ nhật La tinh 3 × 3 × 3 (Hình 6.9) với số liệu ở Bảng 6.9 dưới đây.

Bảng 6.9: Năng suất 9 giống ngô lai (tấn/ha)

Hàng	Cột 1			Cột 2			Cột 3			ΣR_i
(1)	(4)	(9)	(6)	(1)	(7)	(2)	(8)	(3)	(5)	
	7,6	6,9	8,4	6,9	5,7	5,6	8,1	6,3	6,1	61,6
(2)	(1)	(5)	(2)	(6)	(3)	(8)	(7)	(9)	(4)	
	7,0	6,4	5,4	7,0	5,8	7,6	5,8	6,7	8,1	59,7
(3)	(8)	(3)	(7)	(4)	(9)	(5)	(1)	(6)	(2)	
	7,4	6,5	6,5	8,2	5,9	6,6	5,2	12,0	5,3	63,6
ΣC_j		62,1			59,3			63,6		

$\Sigma\Sigma = 184{,}9$

$m = 6{,}85$

NT	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ΣNS	19,1	16,3	18,6	23,9	19,1	27,4	18,0	23,1	19,6
TB	6,4	5,4	6,2	8,0	6,4	9,1	6,0	7,7	6,5

Ghi chú: Các số trong ( ) là các nghiệm thức từ 1 đến 9;
ΣNS là tổng năng suất của mỗi nghiệm thức (tấn/ha).

Ở đây, số nghiệm thức (t) = 9; số hàng (r) = 3; số cột chính (c) = 3, tổng số ô (n) = 27.

1. Tính CF và các tổng bình phương

$$CF = \left(\sum_{1}^{n} x_{ij}\right)^2 / n = (184{,}9)^2 : 27 = 1.266{,}829$$

$$SS_{TO} = \sum_{1}^{n} x_{ij}^2 - CF = (7{,}6)^2 + (6{,}9)^2 + \ldots$$

$$+ (12{,}0)^2 + (5{,}3)^2 - 1.266{,}829 = 49{,}936$$

$$SS_r = \frac{\sum_{i=1}^{r} R_i^2}{t} - CF$$

$$= \frac{\left[(61{,}6)^2 + (59{,}7)^2 + (63{,}6)^2\right]}{9} - 1.266{,}829$$

$$= 0{,}827$$

$$SS_c = \frac{\sum_{i=1}^{c} C_i^2}{t} - CF$$

$$= \frac{\left[(62{,}1)^2 + (59{,}3)^2 + (63{,}6)^2\right]}{9} - 1.266{,}829$$

$$= 1{,}066$$

$$SS_t = \frac{\sum_{i=1}^{t} T_i^2}{r} - CF$$

$$= \frac{\left[(19{,}1)^2 + (16{,}3)^2 + \ldots + (19{,}6)^2\right]}{3}$$

$$- 1.266{,}829 = 32{,}589$$

$$SS_e = SS_{TO} - SS_r - SS_c - SS_t = 49{,}936 - 0{,}827 - 1{,}066 - 32{,}589 = 15{,}454$$

2. Tính phương sai các nguồn biến động

Bậc tự do tổng số: $df_{TO} = rt - 1 = 26$

Bậc tự do hàng = bậc tự do cột = $r - 1 = 2$

Bậc tự do nghiệm thức: $df_t = t - 1 = 9 - 1 = 8$

Bậc tự do sai số: $df_e = df_T - df_r - df_c = 26 - 2 - 2 - 8 = 14$

Áp dụng cách tính phương sai hàng (MS_r), phương sai cột (MS_c), phương sai nghiệm thức (MS_t) và phương sai sai số (MS_e) như đã biết. Kết quả được ghi ở Bảng 6.10.

Bảng 6.10: Kết quả phân tích phương sai

Nguồn	DF	SS	MS	F_{TN}	Prob.
Hàng	2	0,827	0,414		
Cột	2	1,066	0,533		
Nghiệm thức	8	32,589	4,074	3,69	0,0160
Sai số	14	15,454	1,104		
Tổng số	26	49,936			

$P < 0,05$ cho thấy có sự khác nhau giữa các nghiệm thức ở độ tin cậy 95%.

3. Sai số thí nghiệm và so sánh các nghiệm thức

Sai số trung bình:

$$s_{\bar{x}} = \sqrt{\frac{MS_e}{r}} = \sqrt{\frac{1,104}{3}} = 0,61$$

Sai số tương đối:

$$s_{\bar{x}}(\%) = \frac{s_{\bar{x}}}{m} \times 100 = \frac{0,61}{6,85} \times 100 = 8,9$$

Hệ số biến động sai số: $CV(\%) = \frac{\sqrt{1,104}}{6,85} \times 100 = 15,3$

Sai khác giữa hai nghiệm thức:

$$Sd = \sqrt{\frac{2MS_e}{r}} = \sqrt{\frac{2 \times 1,104}{3}} = 0,86\,;$$

với $t_{0,05}$ = 2,145, $LSD_{0,05}$ = 2,145 × 0,86 = 1,8

Kết quả so sánh sự chênh lệch giữa các nghiệm thức:

NT	TB	Chênh lệch và mức ý nghĩa							
		2	3	4	5	6	7	8	9
1	6,3	1,0	0,2	-1,6	0,0	-2,7*	0,4	-1,3	-0,1
2	5,3		-0,8	-2,5*	-0,9	-3,7*	-0,6	-2,3*	-1,1
3	6,1			-1,8	-0,2	-2,9*	0,2	-1,5	-0,3
4	7,9				1,6	-1,2	2,0*	0,3	1,4
5	6,3					-2,8*	0,4	-1,3	-0,2
6	9,0						3,1*	1,4	2,6*
7	5,9							-1,7	-0,5
8	7,6								1,2
9	6,4								

Kiểu phân hạng:

NT	Giá trị	Kiểu chữ	Kiểu vạch
6	9,0	A	
4	7,9	AB	
8	7,6	ABC	
9	6,4	BCD	
1	6,3	BCD	
5	6,3	BCD	
3	6,1	CD	
7	5,9	CD	
2	5,3	D	

Như vậy, giống số có năng suất cao nhất và cao hơn giống số 9, 1, 5, 3, 7 và 2 có ý nghĩa thống kê nhưng không hơn hẳn giống số 4 và 8.

6.5. THÍ NGHIỆM MỘT YẾU TỐ THEO KIỂU MẠNG LƯỚI (Lattice Design)

So với kiểu CRD, kiểu bố trí RCBD hạn chế được sai số do đất không đều bằng cách loại trừ sai khác do khối gây ra, kiểu ô vuông La tinh và chữ nhật La tinh do loại trừ được sai khác cả hàng và cột nên sai số thí nghiệm thấp hơn kiểu CRD và RCBD. Tuy nhiên với số lượng lớn đến hàng mấy chục nghiệm thức, nếu kiểu bố trí CRD và RCBD thí nghiệm sẽ có sai số lớn, còn kiểu ô vuông La tinh hay chữ nhật La tinh sẽ có quá nhiều ô nên thực tế không thể bố trí được. Kiểu bố trí mạng lưới (sau đây được gọi là kiểu mạng) sẽ khắc phục được hạn chế này.

Có nhiều cách bố trí mạng khác nhau:

- Mạng cân bằng (Balanced Lattices);

- Mạng cân bằng từng phần (cân bằng cục bộ - Partially Balanced Lattices);

- Mạng chữ nhật (Rectangular Lattices), và

- Mạng khối lập phương (Cubic Lattices).

Trong kiểu mạng, các khối được bố trí không đầy đủ (incomplete block) và trên một lần lặp lại bao gồm một số khối. Do các nghiệm thức trong khối không đầy đủ nên trong tính toán tổng bình phương sai khác giữa các khối sẽ được điều chỉnh từ tất cả các khối trong các lần lặp lại, đụng đến tất cả mọi ô trong thí nghiệm như “mạng lưới”. Vì vậy, thuật ngữ “Lattice” giải thích phương thức tính toán này. Mỗi kiểu mạng khác nhau sẽ có cách tính toán khác nhau.

Sau đây sẽ đề cập đến một vài kiểu của các loại này.

6.5.1. Kiểu mạng cân bằng (Balanced Lattices)

6.5.1.1. Bố trí thí nghiệm

Để bố trí kiểu mạng cân bằng, số nghiệm thức phải là một số chính phương, số ô trong một khối (kích thước khối) bằng căn bậc hai của số nghiệm thức. Mọi khối đều không đầy đủ được cấu thành từ một số ô nằm trên một hàng trong các lần lặp lại. Đặc tính của mạng cân bằng (cũng như các loại mạng khác) là hai nghiệm thức chỉ ở cạnh nhau một lần trong một khối. Vì vậy tất cả các cặp đứng cạnh nhau được so sánh cùng một độ chính xác.

Số lần lặp lại là một số cố định phụ thuộc vào số nghiệm thức và được xác định như sau:

Số nghiệm thức	9	16	25	49	64	81
Số ô trong khối	3	4	5	7	8	9
Số lần lặp lại	4	5	6	8	9	10

Với kiểu này không thể bố trí cho 36 nghiệm thức và số nghiệm thức trên 100.

Sau đây là một số sơ đồ chính:

Khối	Lần lặp I			Khối	Lần lặp II			Khối	Lần lặp III			Khối	Lần lặp IV		
(1)	1	2	3	(4)	1	4	7	(7)	1	5	9	(10)	1	8	6
(2)	4	5	6	(5)	2	5	8	(8)	7	2	6	(11)	4	2	9
(3)	7	8	9	(6)	3	6	9	(9)	4	8	3	(12)	7	5	3

Hình 6.12: Sơ đồ bố trí mạng cân bằng 3 × 3
$t = 9, k = 3, r = 4, b = 12$

Khối	Lần lặp I				Khối	Lần lặp II				Khối	Lần lặp III			
(1)	1	2	3	4	(5)	1	5	9	13	(9)	1	6	11	16
(2)	5	5	7	8	(6)	2	6	10	14	(10)	5	2	15	12
(3)	9	10	11	12	(7)	3	7	11	15	(11)	9	14	3	8
(4)	13	14	15	16	(8)	4	8	12	16	(12)	13	10	7	4

Khối	Lần lặp IV				Khối	Lần lặp V			
(13)	1	14	7	12	(17)	1	10	15	8
(14)	13	2	11	8	(18)	9	2	7	16
(15)	5	10	3	16	(19)	13	6	3	12
(16)	9	6	15	4	(20)	4	8	12	16

Hình 6.13: Sơ đồ bố trí mạng cân bằng
3×3, $t = 16$, $k = 4$, $r = 5$, $b = 20$

Khối	Lần lặp I						Lần lặp II					Khối	Lần lặp III				
(1)	1	2	3	4	5	(6)	1	6	11	16	21	(11)	1	7	13	9	25
(2)	6	7	8	9	10	(7)	2	7	12	17	22	(12)	21	2	8	14	20
(3)	11	12	13	14	15	(8)	3	8	13	18	23	(13)	16	22	3	9	15
(4)	16	17	18	19	20	(9)	4	9	14	19	24	(14)	11	17	23	4	10
(5)	21	22	23	24	25	(10)	5	10	15	20	25	(15)	6	12	18	24	5

	Lần lặp IV						Lần lặp V						Lần lặp VI				
(16)	1	12	23	9	20	(21)	1	17	8	24	15	(26)	1	22	18	14	10
(17)	16	2	13	24	10	(22)	11	2	18	9	25	(27)	6	2	23	19	15
(18)	6	17	3	14	25	(23)	21	12	3	19	10	(28)	11	7	3	24	20
(19)	23	7	18	4	15	(24)	6	22	13	4	20	(29)	16	12	8	4	25
(20)	11	22	8	19	5	(25)	16	7	23	14	5	(30)	21	17	13	9	5

Hình 6.14: Sơ đồ bố trí mạng cân bằng 5×5
$t = 25$, $k = 5$, $r = 6$, $b = 30$

6.5.1.2. Phân tích phương sai

Với thí nghiệm một yếu tố kiểu mạng cân bằng, giá trị của ô nghiệm thức *i* ở khối thứ *j*, lần lặp lại *k* (X_{jk}) được tính theo mô hình:

$$X_{jk} = m + t_k + b_i' + r_j + e_{ij}$$

trong đó: m là giá trị trung bình toàn thí nghiệm, t_k là giá trị do nghiệm thức k đóng góp, b'_i là giá trị do khối i đã được điều chỉnh đóng góp, r_j giá trị do lần lặp lại j đóng góp, e_{ij} là sai số ngẫu nhiên phân phối chuẩn tắc $N(0, \sigma^2)$.

Theo mô hình này tổng bình phương độ lệch giữa giá trị mỗi ô với giá trị trung bình thí nghiệm SS_{TO} do bốn nguyên nhân:

- Sự khác nhau giữa các nghiệm thức;
- Sự sai khác giữa các lặp lại;
- Sự sai khác giữa các khối sau khi điều chỉnh;
- Sai khác ngẫu nhiên gây ra - sai số (Error).

Để minh họa nguyên lý tính toán, xét ví dụ ở Bảng 6.11 sau đây được bố trí theo sơ đồ Hình 6.12.

Số ô trong khối = k = 3; số khối = 12

Số nghiệm thức = $k \times k$ = 9

Số lần lặp lại = k + 1 = 4

Bảng 6.11: Năng suất 9 giống lúa mì (tạ/ha)

Khối	Lần lặp I			ΣB_t	Khối	Lần lặp II			ΣB_t
(1)	(1)	(2)	(3)		(4)	(1)	(4)	(7)	
	47,7	42,8	54,6	145,1		46,9	42,2	38,8	127,9
(2)	(4)	(5)	(6)		(5)	(2)	(5)	(8)	
	47,7	59,2	48,0	154,9		40,0	47,0	51,7	138,7
(3)	(7)	(8)	(9)		(6)	(3)	(6)	(9)	
	41,5	43,8	37,9	123,2		50,0	39,5	45,5	135,0
ΣR				**423,2**					**401,6**

Khối	Lần lặp III			ΣB_t	Khối	Lần lặp IV			ΣB_t
(7)	(1)	(5)	(9)		(10)	(1)	(6)	(8)	
	45,4	45,9	45,2	136,5		44,3	44,3	47,0	135,6
(8)	(2)	(6)	(7)		(11)	(2)	(4)	(9)	
	43,1	40,5	45,5	129,1		41,0	40,7	45,0	126,7
(9)	(3)	(4)	(8)		(12)	(3)	(5)	(7)	
	50,5	43,4	38,1	132,0		44,4	45,7	43,8	133,9
ΣR				**397,6**					**396,2**

Tổng toàn bộ thí nghiệm: G = 1.618,6

Bảng 6.12: Tổng năng suất các giống và kết quả điều chỉnh

NT	Tổng năng suất (T)	Khối (B_t)	Quyền số (W_i)	Tổng NS điều chỉnh (T'_i)	TB năng suất điều chỉnh
1	184,3	545,1	-8,9	183,7	45,9
2	166,9	539,6	-39,1	164,4	41,1
3	199,5	546,0	33,1	201,6	50,4
4	174,0	541,5	-25,4	172,4	43,1
5	197,8	564,0	-44,0	195,0	48,8
6	172,3	554,6	-82,9	167,1	41,6
7	169,6	514,1	71,0	174,1	43,5
8	180,6	529,5	42,4	183,3	45,8
9	173,6	521,4	53,8	177,0	44,3
Tổng	1618,6	4855,8	0,0	1618,6	

Các bước tính toán như sau:

1. Tính tổng của tất cả các khối ΣB_t, tổng từng nghiệm thức (T) và tổng toàn bộ thí nghiệm (G).

2. Cho mỗi nghiệm thức, tính tổng B_t từ tất cả các khối, ở đó có nghiệm thức tham gia.

Cho giống 1: B_1 = 145,1 + 127,9 + 136,5 + 135,6 = 545,1

Cho giống 2: B_2 = 145,1 + 138,7 + 129,1 + 126,7 = 539,6

Một cách tương tự ta được số liệu Bảng 6.12.

3. Tính quyền số W_t cho mỗi nghiệm thức.

$$W_t = kT - (k+1)B_t + G$$

Cho giống 1: $W_1 = 3 \times 184{,}3 - 4 \times 545{,}1 + 1.618{,}6 = -8{,}9$

Cho giống 2: $W_2 = 3 \times 166{,}9 - 4 \times 539{,}6 + 1.618{,}6 = -39{,}1$

Một cách tương tự ta được số liệu Bảng 6.12.

4. Tính toán các nguồn biến động và phân tích phương sai

$$CF = \left(\sum_1^n x_{ij}\right)^2 / n = (1.592{,}4)^2 : 36 = 72.774{,}054$$

$$SS_{TO} = \sum_1^n x_{ij}^2 - CF = 712{,}846$$

$$SS_r = \frac{\sum_{i=1}^r R_i^2}{k^2} - CF = 52{,}723$$

$$SS_t = \frac{\sum_{i=1}^t T_i^2}{r} - CF = 282{,}936$$

Tổng bình phương khối tính theo quyền số W_i:

$$SS_b = \frac{\sum_{i=1}^r W_i^2}{k^3(k+1)}$$

$$= \frac{(-8,9)^2 + (-39,1)^2 + ... + (53,8)^2}{108}$$

$= 202,689$

$$SS_e = SS_T - SS_t - SS_r - SS_b = 174,498$$

Bảng 6.13: Kết quả phân tích phương sai sai số

Nguồn biến động	Bậc tự do		SS	MS
	Công thức	k =3		
Lần lặp lại (r)	k	3	52,723	
Nghiệm thức (giống)	$(k^2 - 1)$	8	282,936	
Khối (b)	$(k^2 - 1)$	8	202,689	25,336
Sai số ngoài khối (e)	$(k-1)(k^2 - 1)$	16	174,498	10,906
Tổng số (TO)	$k^3 + k^2 - 1$	35	712,846	

5. Tính hệ số điều chỉnh μ, tổng giá trị nghiệm thức và trung bình nghiệm thức sau khi điều chỉnh.

$$\mu = \frac{(MS_b - MS_e)}{k^2 MS_b} = \frac{(25,336 - 10,906)}{9 \times 25,336} = 0,0633$$

Tổng giá trị nghiệm thức điều chỉnh của mỗi nghiệm thức T_i' được tính theo công thức: $T_i' = T_i + \mu W_i$:

Cho giống 1: T_1' = 184,3 + 0,0633 × (-8,9) = 183,7

Trung bình giống 1: 183,7 : 4 = 45,9

Cho giống 2: T_2' = 166,9 + 0,0633 × (- 39,1) = 164,4

Trung bình giống 2: 164,4 : 4 = 41,1

Một cách tương tự ta được số liệu Bảng 6.12.

6. Kiểm tra mức ý nghĩa của phương sai nghiệm thức

Tổng phương sai nghiệm thức điều chỉnh:

$$SS'_t = \frac{\sum_{i=1}^{t} T'^2_i}{r} - CF = 307{,}055$$

Phương sai nghiệm thức điều chỉnh:

$$MS'_t = \frac{307{,}055}{8} = 38{,}382$$

Tổng bình phương sai số điều chỉnh:

$$MS'_e = MS_e\ (k\mu + 1)$$

$$= 10{,}906\ (3 \times 0{,}0633 + 1) = 12{,}977$$

$$F_{TN} = \frac{38{,}382}{12{,}977} = 2{,}96$$

F_{TN} = 2,96 > $F_{0,05}^{(8,16)}$ = 2,59 nhưng nhỏ hơn $F_{0,01}^{(8,16)}$ = 3,89 cho thấy có sự khác nhau giữa các nghiệm thức ở độ tin cậy 95%.

Cách tính sai số trung bình, sai số tương đối, hệ số biến động sai số và so sánh giữa các nghiệm thức theo $LSD_{0,05}$ và cách phân hạng như đã biết.

7. Hiệu quả của kiểu bố trí mạng cân bằng cũng được tính tương tự như các phương pháp trước.

Nếu coi kiểu bố trí ô vuông La tinh là một kiểu RCBD với 4 lần lặp lại thì tổng bình phương sai sai số ($SS_{e(RCBD)}$) là:

$$SS_{e(RCBD)} = \left(k^2 - 1\right)MS_b + \left(k - 1\right)\left(k^2 - 1\right)MS_e$$
$$= \left(k^2 - 1\right)\left[MS_b + \left(k - 1\right)MS_e\right]$$

và $$MS_{e(RCBD)} = \frac{\left(k^2-1\right)\left[MS_b + \left(k-1\right)MS_e\right]}{k\left(k^2-1\right)}$$

$$= \frac{MS_b + \left(k-1\right)MS_e}{k}$$

Phương sai sai số kiểu mạng cân bằng 3 ×3 đã giảm đi (%) so với phương sai sai số kiểu RCBD:

$$r_e\left(\%\right) = \frac{MS_{e(RCBD)} - MS'_e}{MS_{e(RCBD)}} \times 100$$

$$= \frac{MS_b - (k^2\mu + 1)MS_e}{MS_b + \left(k-1\right)MS_e} \times 100$$

Cho thí nghiệm này:

$$r_e\left(\%\right) = \frac{25,336 - (9\times 0,0633 + 1)\times 10,906}{25,336 + 2\times 10,906} \times 100$$

$$= 17,4$$

6.5.2. Kiểu mạng cân bằng từng phần (Parially Balanced Lattices)

6.5.2.1. Bố trí thí nghiệm

Mặc dù mạng cân bằng có ưu điểm nhưng trong trường hợp có quá nhiều nghiệm thức, số lần lặp lại vẫn lớn. Mạng cân bằng từng phần cho phép giảm tối đa số lần lặp lại. Nếu trong mạng cân bằng có 25 nghiệm thức phải bố trí 6 lần lặp lại thì trong mạng đơn cân bằng từng phần chỉ cần bố trí 2 lần lặp lại. Phương pháp này cho phép bố trí các thí nghiệm có 36 và trên 100 nghiệm thức mà phương pháp mạng cân bằng không bố trí được.

Mạng cân bằng từng phần gồm các kiểu: Mạng đơn (hai lần lặp lại), mạng ba (ba lần lặp lại), mạng bốn (bốn lần lặp lại), mạng năm (năm lần lặp lại) và mạng rất nhiều lần lặp lại.

Sau đây là một số sơ đồ chính:

Khối	Lần lặp I					Khối	Lần lặp II				
(1)	1	2	3	4	5	(6)	1	6	11	16	21
(2)	6	7	8	9	10	(7)	2	7	12	17	22
(3)	11	12	13	14	15	(8)	3	8	13	18	23
(4)	16	17	18	19	20	(9)	4	9	14	19	24
(5)	21	22	23	24	25	(10	5	10	15	20	25

Hình 6.15: Sơ đồ bố trí mạng đơn cân bằng từng phần 5 × 5

	Lần lặp I							Lần lặp II							Lần lặp III					
(1)	1	2	3	4	5	6	(7)	1	7	13	19	25	31	(13)	1	8	15	22	29	36
(2)	7	8	9	10	11	12	(8)	2	8	14	20	26	32	(14)	31	2	9	16	23	30
(3)	13	14	15	16	17	18	(9)	3	9	15	21	27	33	(15)	25	32	3	10	17	24
(4)	19	20	21	22	23	24	(10)	4	10	16	22	28	34	(16)	19	26	33	4	11	18
(5)	25	26	27	28	29	30	(11)	5	11	17	23	29	35	(17)	13	20	27	34	5	12
(6)	31	32	33	34	35	36	(12)	6	12	18	24	30	36	(18)	7	14	21	28	35	6

Hình 6.16: Sơ đồ bố trí mạng ba cân bằng từng phần 6 × 6

6.5.1.2. Phân tích phương sai

Theo mô hình này tổng bình phương độ lệch giữa giá trị mỗi ô với giá trị trung bình thí nghiệm SS_{TO} do bốn nguyên nhân:

- Sự khác nhau giữa các nghiệm thức;

- Sự sai khác giữa các lặp lại;

- Sự sai khác giữa các khối sau khi điều chỉnh;

- Sai khác ngẫu nhiên gây ra – sai số (Error).

Để minh họa nguyên lý tính toán, xét ví dụ số liệu ở

Bảng 6.14, 6.15 và 6.16 sau đây được thu thập từ một thí nghiệm bố trí theo sơ đồ Hình 6.15.

Trong thí nghiệm này:

Số ô trong khối = k = 5; số khối = 10

Số nghiệm thức = $k \times k$ = 25; Số lần lặp lại = 2

Bảng 6.14: Năng suất 25 giống bông lai F_1 (tạ/ha)

Khối	Lần lặp I					ΣB	C	μC
(1)	(1)	(2)	(3)	(4)	(5)			
	23,2	26,7	21,3	24,1	24,2	119,5	13,9	1,720
(2)	(6)	(7)	(8)	(9)	(10)			
	24,5	23,4	24,6	20,2	28,0	120,7	-0,1	-0,012
(3)	(11)	(12)	(13)	(14)	(15)			
	24,3	25.2	23,9	27,7	25,4	126,5	27,6	3,415
(4)	(16)	(17)	(18)	(19)	(20)			
	24,9	21,0	27,8	23,2	25,1	122,0	3,6	0,445
(5)	(21)	(22)	(23)	(24)	(26)			
	35,2	19,3	18,4	28,2	22,1	123,2	-12,7	-1,571
ΣR						**611,9**	32.3	3,996
Khối	Lần lặp II					ΣB	C	μC
(6)	(1)	(6)	(11)	(16)	(21)			
	24,4	24,1	25,2	27,1	29,1	129,9	2,2	0,272
(7)	(2)	(7)	(12)	(17)	(22)			
	28,4	22,5	28.3	22,1	19,3	122,6	-5,0	-0,619
(8)	(3)	(8)	(13)	(18)	(23)			
	23,3	21,1	35,8	24,9	18,4	123,5	-7,5	-0,928
(9)	(4)	(9)	(14)	(19)	(24)			
	27,6	24,8	32,9	20,4	23,3	129,0	-5,6	-0,693
(10)	(5)	(10)	(15)	(20)	(25)			
	29,7	28,1	31,9	31,1	20,4	141,2	-16,4	-2,029
ΣR						**644,2**	-32,3	-3,996
			Tổng toàn thí nghiệm G =			**1.256,1**		

Bảng 6.15: Tổng năng suất thực thu các giống

						μC
	(1) 47,6	(2) 55,1	(3) 44,6	(4) 51,7	(5) 53,9	1,720
	(6) 48,6	(7) 45,9	(8) 45,7	(9) 45,0	(10) 56,1	-0,012
	(11) 49,5	(12) 53,5	(13) 59,7	(14) 60,6	(15) 57,3	3,415
	(16) 52,0	(17) 43,1	(18) 52,7	(19) 43,6	(20) 56,2	0,445
	(21) 64,3	(22) 38,6	(23) 36,8	(24) 51,5	(25) 42,5	-1,571
μC	0,272	-0,619	-0,928	-0,693	-2,029	0,0

Các bước tính toán như sau:

1. Tính tổng của tất cả các khối (B), tổng từng nghiệm thức (T) và tổng toàn bộ thí nghiệm (G).

2. Cho mỗi nghiệm thức, tính tổng C của tất cả các khối.

C = Tổng (cho tất cả các lần lặp) của tất cả các nghiệm thức trong khối trừ đi rB. Số liệu để tính C lấy từ Bảng 6.15 và tính cho từng khối theo sơ đồ thí nghiệm.

Cho khối 1 ở lần lặp 1:

$C = 47{,}6 + 55{,}1 + 44{,}6 + 51{,}7 + 53{,}9 - 2 \times 119{,}5 = 13{,}9$

Cho khối 2 ở lần lặp 1:

$C = 48{,}6 + 45{,}9 + 45{,}7 + 45{,}0 + 56{,}1 - 2 \times 120{,}7 = -0{,}1$

Một cách tương tự ta được số liệu Bảng 6.14

3. Tính tổng bình phương các nguồn biến động và phân tích phương sai sai số

Tổng bình phương biến động tổng số, tổng bình phương biến động do lần lặp lại và nghiệm thức (chưa điều chỉnh) được tính toán theo cách thông thường. Riêng tổng bình phương biến động khối được tính theo công thức:

$$SS'_b = \frac{\sum C^2}{kr(r-1)} - \frac{\sum R_c^2}{k^2 r(r-1)}$$

$$= \frac{(13,9)^2 + (-0,1)^2 + ... + (-16,4)^2}{10}$$

$$- \frac{(32,3)^2 + (-32,3)^2}{50} = 20,866$$

Trong công thức này, R_c là tổng của các lần lặp lại theo các giá trị C.

Kết quả tính toán tổng bình phương tổng số, tổng bình phương các nguồn biến động, phương sai khối và phương sai sai số được ghi ở Bảng 6.16.

Bảng 6.16: Tổng bình phương các nguồn biến động, phương sai khối và phương sai sai số

Nguồn biến động	Bậc tự do		SS	MS
	Công thức			
Lần lặp lại	*(r – 1)*	1	20,866	
Nghiệm thức (giống)	*(k^2 – 1)*	24	577,371	
Khối	*r(k – 1)*	8	109,832	13,729
Sai số ngoài khối	*(k–1)(rk–k–1)*	16	83,777	5,236
Tổng số	*r(k^2 – 1)*	49	791,846	

4. Tính hệ số điều chỉnh μ, tổng giá trị nghiệm thức và trung bình nghiệm thức sau khi điều chỉnh.

$$\mu = \frac{(MS_b - MS_e)}{k(r-1)MS_b}$$

$$= \frac{(13,729 - 5,236)}{5 \times 5,236} = 0,1237$$

Tổng giá trị nghiệm thức điều chỉnh của mỗi nghiệm thức T_i' được tính theo công thức: $T_i' = T_i + \mu W_i$ như kiểu mạng cân đối đã biết. Kết quả tính toán được ghi ở Bảng 6.17.

Bảng 6.17: Tổng năng suất điều chỉnh (tử số) và năng suất trung bình (mẫu số)

(1)	$\frac{49,6}{24,8}$	(2)	$\frac{56,2}{28,1}$	(3)	$\frac{45,4}{22,7}$	(4)	$\frac{52,7}{26,4}$	(5)	$\frac{53,6}{26,8}$
(6)	$\frac{48,9}{24,4}$	(7)	$\frac{45,3}{22,6}$	(8)	$\frac{44,8}{22,4}$	(9)	$\frac{44,3}{22,2}$	(10)	$\frac{54,1}{27,0}$
(11)	$\frac{53,2}{26,6}$	(12)	$\frac{56,3}{28,1}$	(13)	$\frac{62,2}{31,1}$	(14)	$\frac{63,3}{31,7}$	(15)	$\frac{58,7}{29,3}$
(16)	$\frac{52,7}{26,4}$	(17)	$\frac{42,9}{21,5}$	(18)	$\frac{52,2}{26,1}$	(19)	$\frac{43,4}{21,7}$	(20)	$\frac{54,6}{27,3}$
(21)	$\frac{63,0}{31,5}$	(22)	$\frac{36,4}{18,2}$	(23)	$\frac{34,3}{17,2}$	(24)	$\frac{49,2}{24,6}$	(25)	$\frac{38,9}{19,4}$

5. Kiểm tra mức ý nghĩa của phương sai nghiệm thức

- Tổng phương sai nghiệm thức được điều chỉnh theo công thức:

SS_t (điều chỉnh) = SS_t (chưa điều chỉnh) – A

A được tính theo công thức:

$$A = k\mu(r-1)\left[\frac{r}{(r-1)(k\mu+1)}B_u - B_a\right]$$

trong đó: B_u là tổng bình phương khối chưa điều chỉnh còn B_a là tổng bình phương khối sau khi đã điều chỉnh.

$$A = 5\times 0,1237\times\left[\frac{2}{(2-1)(5\times 0,1237+1)}\times 55,848 - 109,832\right]$$

$$= -25,255$$

SS_t (điều chỉnh) = 577,371 + 25,255 = 602,626

MS_t (điều chỉnh) = 602,626 : 24 = 25,109

- Tổng bình phương sai sai số điều chỉnh:

$$MS'_e = \left(1 + \frac{rk\mu}{k+1}\right)MS_e$$

$$= \left(1 + \frac{2\times5\times0,1237}{5+1}\right)\times5,236 = 6,316$$

$$F_{TN} = \frac{25,109}{6,316} = 3,98$$

F_{TN} = 3,98 > $F_{0,05}^{(24,16)}$ = 2,24 và lớn hơn cả $F_{0,01}^{(24,16)}$ = 3,18 cho thấy có sự khác nhau rất đáng tin cậy giữa các nghiệm thức.

Cách tính sai số trung bình, sai số tương đối, hệ số biến động sai số được tính như các kiểu thí nghiệm trước đây. Tuy nhiên việc so sánh giữa các nghiệm thức như sau:

- Sai khác giữa hai nghiệm thức cùng khối:

$$Sd_1 = \sqrt{\left[1 + (r-1)\mu\right]\frac{2MS_e}{r}}$$

$$= \sqrt{(1 + 0,1237)\times\frac{2\times5,236}{2}} = 2,4$$

- Sai khác giữa hai nghiệm thức khác khối:

$$Sd_2 = \sqrt{(1 + r\mu)\frac{2MS_e}{r}}$$

$$= \sqrt{(1 + 2\times0,1237)\times\frac{2\times5,236}{2}} = 2,6$$

6. Hiệu quả của kiểu bố trí mạng đơn cân bằng từng phần

Từ bảng phân tích phương sai, dễ dàng tính được phương sai sai số kiểu mạng đơn cân bằng từng phần đã giảm đi so với kiểu RCBD là:

$$\mathrm{r_e}(\%) = \frac{r(k-1)(\mathrm{MS_b} - \mathrm{MS_e})}{r(k-1)\mathrm{MS_b} + \left[k^2(r-1) - rk + 1\right]\mathrm{MS_e}} \times 100$$

Với $k = 5$, $r = 2$ thì:

$$\mathrm{r_e}(\%) = \frac{\mathrm{MS_b} - \mathrm{MS_e}}{\mathrm{MS_b} + 2\mathrm{MS_e}} \times 100$$

Trong thí nghiệm này:

$$\mathrm{r_e}(\%) = \frac{13,729 - 5,236}{13,729 + 2 \times 5,236} \times 100 = 35,1$$

6.6. THÍ NGHIỆM MỘT YẾU TỐ KIỂU MẠNG LƯỚI VUÔNG (Lattice Squares)

6.6.1. Bố trí thí nghiệm

Trong kiểu lưới vuông, số nghiệm thức cũng là một số chính phương. Trong một lần lặp lại, k^2 nghiệm thức được sắp xếp trên một ô vuông $k \times k$. Cách đặt các hàng và cột trong lần lặp sau được điều chỉnh sao cho các ô của cùng nghiệm thức không cùng nằm trong một hàng hoặc một cột. Điều đó cho phép tiếp tục loại trừ ra khỏi sai số nguồn biến động “kiểm soát kép” trong mỗi lần lặp lại, tương tự như kiểu ô vuông La tinh.

Theo phương này có thể bố trí mạng lưới vuông cho 9, 25, 49, 81, 121 và 169 nghiệm thức. Số lần lặp lại cho k^2 nghiệm thức là $(k + 1)/2$.

Sau đây là một số sơ đồ.

Lần lặp I			Lần lặp II		
1	2	3	1	6	8
4	5	6	9	2	4
7	8	9	5	7	3

Hình 6.17: Lưới vuông cân bằng 3 × 3

t = 9, k = 3, r = 2, hàng = 6, cột = 6, λ = 1[@]

Lần lặp I					Lần lặp II					Lần lặp III				
1	2	3	4	5	1	10	14	18	22	1	8	15	17	24
6	7	8	9	10	23	2	6	15	19	25	2	9	11	18
11	12	13	14	15	20	24	3	7	11	19	21	3	10	12
16	17	18	19	20	12	16	25	4	8	13	20	22	4	6
21	22	23	24	25	9	13	17	21	5	7	14	16	23	5

Hình 6.18a: Lưới vuông cân bằng 5 × 5

t = 25, k = 5, r = 3, hàng = 15, cột = 15, λ = 1

Lần lặp I					Lần lặp II					Lần lặp III				
18	9	11	2	25	20	17	19	16	18	19	15	23	6	2
24	15	17	8	1	15	12	14	11	13	11	7	20	3	24
12	3	10	21	19	25	22	24	21	23	22	18	1	14	10
6	22	4	20	13	5	2	4	1	3	5	21	9	17	13
5	16	23	14	7	10	7	9	6	8	8	4	12	25	16

Hình 6.18b: Lưới vuông cân bằng 5 × 5

t = 25, k = 5, r = 3, hàng = 15, cột = 15, λ = 1

Lần lặp I							Lần lặp II						
1	2	3	4	5	6	7	1	38	26	14	44	32	20
8	9	10	11	12	13	14	21	2	39	27	8	45	33
15	16	17	18	19	20	21	34	15	3	40	28	9	46
22	23	24	25	26	27	28	47	35	16	4	41	22	10
29	30	31	32	33	34	35	11	48	29	17	5	42	23
36	37	38	39	40	41	42	24	12	49	30	18	6	36
42	44	45	46	47	48	49	37	25	13	43	31	19	7

@: Số lần để hai nghiệm thức cùng có mặt trên một hàng hoặc một

Lần lặp III

1	19	30	48	10	28	39
40	2	20	31	49	11	22
23	41	3	21	32	43	12
13	24	42	4	15	33	44
45	14	25	36	5	16	34
35	46	8	26	37	6	17
18	29	47	9	27	38	7

Lần lặp IV

1	42	27	12	46	31	16
17	2	36	28	13	47	32
33	18	3	37	22	14	48
49	34	19	4	38	23	8
9	43	35	20	5	39	24
25	10	44	29	21	6	40
41	26	11	45	30	15	7

Hình 6.19: Lưới vuông cân bằng 7 × 7
$t = 49$, $k = 7$, $r = 4$, hàng = 28, cột = 28, $\lambda = 1$

6.6.2. Phân tích phương sai

Theo kiểu thí nghiệm này tổng bình phương độ lệch giữa giá trị mỗi ô với giá trị trung bình thí nghiệm SS_{TO} do các nguyên nhân:

- Sự sai khác giữa các lặp lại;

- Sự khác nhau giữa các nghiệm thức;

- Sự khác nhau giữa các hàng từ số liệu điều chỉnh nghiệm thức và cột;

- Sự khác nhau giữa các cột từ số liệu điều chỉnh nghiệm thức và hàng;

- Sai khác ngẫu nhiên gây ra - Sai số.

Để minh họa nguyên lý tính toán, xét ví dụ ở Bảng 6.18, 6.19 sau đây được thu thập từ thí nghiệm bố trí theo sơ đồ Hình 6.18b.

Trong thí nghiệm này:

Số nghiệm thức = $k \times k$ = 25

Số lần lặp lại r = 3

Số hàng = số cột = 15

Bảng 6.18: Số quả trên/cây (quả) của các giống bông trong thí nghiệm bố trí mạng lưới vuông 5 × 5, 3 lần lặp lại (trong ngoặc đơn là số của nghiệm thức)

	Lần lặp I					Σ	L	δ
	(18)	(9)	(11)	(2)	(25)			
	12,9	14,6	12,5	12,2	17,0	69,2	-5,83	-0,371
	(24)	(15)	(17)	(8)	(1)			
	17,1	9,8	9,1	15,5	12,3	63,8	-1,88	-0,120
	(12)	(3)	(10)	(21)	(19)			
	9,5	13,3	13,0	13,1	10,6	59,4	18,96	1,208
	(6)	(22)	(4)	(20)	(13)			
	10,8	10,9	9,1	12,9	8,7	52,4	21,15	1,348
	(5)	(16)	(23)	(14)	(7)			
	16,2	11,5	11,9	10,3	9,2	59,1	16,23	1,034
Σ	**66,40**	**60,13**	**55,58**	**64,01**	**57,81**	**303,93**		
M	-1,63	3,34	41,08	0,38	5,46		48,64	3,099
ε	-0,11	0,22	2,75	0,03	0,36			

	Lần lặp II					Σ	L	δ
	(20)	(17)	(19)	(16)	(18)			
	12,0	14,0	12,9	13,6	14,1	66,6	- 13,97	-0,890
	(15)	(12)	(14)	(11)	(13)			
	10,0	11,5	13,4	15,1	9,4	59,4	-6,00	-0,382
	(25)	(22)	(24)	(21)	(23)			
	14,2	14,3	15,6	13,2	16,2	73,4	0,44	0,028
	(5)	(2)	(4)	(1)	(3)			
	10,6	8,3	15,0	8,4	10,0	52,3	29,97	1,910
	(10)	(7)	(9)	(6)	(8)			
	14,2	12,9	14,3	10,9	13,6	65,8	-2,81	-0,179
Σ	**60,89**	**60,99**	**71,13**	**61,34**	**63,25**	**317,60**		
M	20,13	-5,95	-10,7	-3,16	7,30		7,62	0,486
ε	1,35	-0,40	-0,72	-0,21	0,49			

	Lần lặp III					Σ	L	δ
	(10)	(15)	(23)	(6)	(2)			
	13,6	10,4	18,1	10,1	11,6	63,8	-13,91	-0,886
	(11)	(7)	(20)	(3)	(24)			
	13,9	12,1	14,1	16,0	14,9	71,0	-11,36	-0,723
	(22)	(18)	(1)	(14)	(10)			
	13,6	12,7	11,5	12,7	16,5	67,0	-10,27	-0,655
	(5)	(21)	(9)	(17)	(13)			
	15,4	14,4	11,8	12,3	9,6	63,4	-3,63	-0,231
	(8)	(4)	(12)	(25)	(16)			
	15,2	17,0	15,5	16,5	9,6	73,7	-17,09	-1,089
Σ	**71,55**	**66,66**	**70,90**	**67,72**	**62,07**	**338,90**		
M	-10,89	-14,06	-18,23	-12,45	-0,62		-56,26	-3,584
ε	-0,73	-0,94	-1,22	-0,83	-0,04			
			Tổng cộng (G)			**960,43**	0,0	0,0

Bảng 6.19: Tổng số quả thực tế của các giống

(1) 32,2	(2) 32,1	(3) 39,4	(4) 41,1	(5) 42,1
(6) 31,9	(7) 34,3	(8) 44,2	(9) 40,7	(10) 43,7
(11) 41,5	(12) 36,5	(13) 27,7	(14) 36,4	(15) 30,3
(16) 34,6	(17) 35,4	(18) 39,6	(19) 37,1	(20) 39,0
(21) 40,7	(22) 38,8	(23) 46,1	(24) 47,5	(25) 47,7

Các bước tính toán như sau:

1. Tính tổng của tất cả các hàng trên các lần lặp lại, tổng của tất cả các cột trên các lần lặp lại, tổng từng nghiệm thức, tổng của các lần lặp lại và tổng toàn bộ thí nghiệm (bảng 6.18, 6.19).

2. Tính các giá trị L và M cho mỗi hàng:

L = Tổng (từ tất cả các lần lặp) của tất cả các nghiệm thức có trong hàng đó trừ đi *r* lần tổng hàng đó. Số liệu để tính L và M lấy từ Bảng 6.19 và sắp xếp theo sơ đồ thí nghiệm.

Cho hàng 1:

$L_1 = 39{,}6 + 40{,}7 + 41{,}5 + 32{,}1 + 47{,}7 - 3 \times 69{,}2 = -\ 5{,}83$

Cho hàng 2:

$L_2 = 47{,}5 + 30{,}3 + 35{,}7 + 44{,}2 + 32{,}2 - 3 \times 63{,}8 = -\ 1{,}88$

Một cách tương tự ta được số liệu Bảng 6.18

Kiểm tra lại tổng số của L bằng cách:

Tổng L = Tổng G – *r* × tổng lần lặp lại

Như, $48{,}64 = 960{,}43 - 3 \times 303{,}93$

M = Tổng (từ tất cả các lần lặp) của tất cả các nghiệm thức có trong cột đó trừ đi *r* lần tổng cột đó.

Cho cột 1, lần lặp 1:

$M_{1.1} = 39{,}6 + 47{,}5 + 36{,}5 + 31{,}9 + 42{,}1 - 3 \times 66{,}4 = -\ 0{,}11$

Cho cột 3 lần lặp 2:

$M_{3.2} = 37{,}1 + 36{,}4 + 47{,}5 + 41{,}1 + 40{,}7 - 3 \times 71{,}13 = -\ 0{,}72$

Một cách tương tự ta được số liệu Bảng 6.18

3. Tính tổng bình phương các nguồn biến động và phân tích phương sai sai số

Tổng bình phương biến động tổng số, tổng bình phương biến động do lần lặp lại và nghiệm thức (chưa điều chỉnh) được tính toán như các kiểu thí nghiệm đã biết.

Tổng bình phương biến động hàng:

$$SS'_r = \frac{\sum L^2}{kr(r-1)} - \frac{\sum L_r^2}{k^2 r(r-1)}$$

$$= \frac{(-5,83)^2 + (-1,88)^2 + ... + (-17,09)^2}{30}$$

$$- \frac{(48,64)^2 + (7,62)^2 + (-56,26)^2}{150} = 62,028$$

Trong công thức này, L_r là tổng của các lần lặp lại theo các giá trị L.

Tổng bình phương biến động cột:

$$SS'_c = \frac{\sum M^2}{kr(r-1)} - \frac{\sum M_r^2}{k^2 r(r-1)}$$

$$= \frac{(-1,63)^2 + (3,44)^2 + ... + (-0,62)^2}{30}$$

$$- \frac{(48,64)^2 + (7,62)^2 + (-56,26)^2}{150} = 67,867$$

Tổng M_r = Tổng L_r

Kết quả tính toán tổng bình phương các nguồn biến động được ghi trong Bảng 6.20.

Bảng 6.20: Kết quả phân tích phương sai

Nguồn biến động	Bậc tự do		SS	MS
	Công thức			
Lần lặp lại	*(r – 1)*	2	24,837	2,419
Nghiệm thức (giống)	*(k^2 – 1)*	24	228,833	9,535
Hàng (điều chỉnh)	*r(k – 1)*	12	62,028	5,169
Cột (điều chỉnh)	*r(k – 1)*	12	67,867	5,656
Sai số ngoài khối	*(k–1)(rk–r-k–1)*	24	45,016	1,876
Tổng số	*r(k^2 – 1)*	74	428,581	

4. Tính các hệ số điều chỉnh hàng λ', hệ số điều chỉnh cột μ' để điều chỉnh giá trị nghiệm thức theo hàng

và cột.

$$\lambda' = \frac{MS_r - MS_e}{k(r-1)MS_r} = \frac{5{,}169 - 1{,}876}{5 \times 2 \times 5{,}169} = 0{,}0637$$

$$\mu' = \frac{MS_c - MS_e}{k(r-1)MS_c} = \frac{5{,}656 - 1{,}876}{5 \times 2 \times 5{,}656} = 0{,}0668$$

Nếu MS_r hoặc MS_c nhỏ hơn MS_e thì λ' hoặc μ' coi như bằng 0.

5. Nhân các L với λ' sẽ được δ và các M với μ' sẽ được ε. Kết quả tính được ở Bảng 6.18.

6. Tổng giá trị nghiệm thức điều chỉnh của mỗi nghiệm thức T_i' được tính theo công thức:

T_i' = Tổng T_i + tổng δ + tổng ε tất cả các hàng và cột của nghiệm thức ở tất cả các lần lặp lại.

Cho nghiệm thức 1: T_1' = 32,2 + (- 0,12) + (0,36) + (1,91) + (- 0,21) + (- 0,655) + (- 1,22) = 33,0

Kết quả tính toán được ghi ở Bảng 6.21.

Bảng 6.21: Tổng số quả/cây các giống đã điều chỉnh (ở tử số) và trung bình (ở mẫu số)

(1)	$\frac{33,0}{11,0}$	(2)	$\frac{32,4}{10,8}$	(3)	$\frac{41,6}{13,9}$	(4)	$\frac{44,4}{14,8}$	(5)	$\frac{45,4}{15,1}$
(6)	$\frac{31,0}{10,3}$	(7)	$\frac{33,4}{11,1}$	(8)	$\frac{42,6}{14,2}$	(9)	$\frac{38,2}{12,7}$	(10)	$\frac{48,1}{16,0}$
(11)	$\frac{41,8}{13,9}$	(12)	$\frac{34,5}{11,5}$	(13)	$\frac{29,2}{9,7}$	(14)	$\frac{34,9}{11,6}$	(15)	$\frac{29,5}{9,8}$
(16)	$\frac{33,7}{11,2}$	(17)	$\frac{35,7}{11,9}$	(18)	$\frac{37,2}{12,4}$	(19)	$\frac{35,4}{11,8}$	(20)	$\frac{38,9}{13,0}$
(21)	$\frac{40,5}{13,5}$	(22)	$\frac{38,6}{12,9}$	(23)	$\frac{48,3}{16,1}$	(24)	$\frac{45,8}{15,3}$	(25)	$\frac{47,1}{15,7}$

7. Kiểm tra mức ý nghĩa của phương sai nghiệm thức

- Tổng phương sai nghiệm thức điều chỉnh:

$$\mathrm{SS}'_{\mathrm{t}} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{t}} \mathrm{T}_{\mathrm{i}}'^{2}}{r} - \mathrm{CF} = 294{,}671$$

và phương sai nghiệm thức điều chỉnh:

$$\mathrm{MS}'_{\mathrm{t}} = \frac{294{,}671}{24} = 12{,}278$$

- Tổng bình phương sai số điều chỉnh:

$$\mathrm{MS}'_{\mathrm{e}} = \mathrm{MS}_{\mathrm{e}}\left[1 + \frac{rk}{k+1}(\lambda' + \mu')\right]$$

$$= 1{,}876\left[1 + \frac{15}{6} \times 0{,}1305\right] = 2{,}488$$

$$\mathrm{F}_{\mathrm{TN}} = \frac{12{,}278}{2{,}488} = 4{,}93$$

F_{TN} = 4,93 lớn hơn cả $\mathrm{F}_{0,05}^{(24,24)}$ = 1,98 và $\mathrm{F}_{0,01}^{(24,24)}$ = 2,66 cho thấy có sự khác nhau rất đáng tin cậy giữa các nghiệm thức.

8. So sánh các nghiệm thức

- Sai khác giữa hai nghiệm thức điều chỉnh:

$$\mathrm{Sd}_1 = \left\{\frac{2\mathrm{MS}_{\mathrm{e}}}{r}\left[1 + \frac{rk}{k+1}(\lambda' + \mu')\right]\right\}^{0,5}$$

$$= \left\{\frac{(2)(1{,}876)}{3}\left[1 + \frac{15}{6}(0{,}1305)\right]\right\}^{0,5} = 1{,}3$$

- Sai khác giữa hai nghiệm thức cùng hàng:

$$\mathrm{Sd}_2\left\{\frac{2\mathrm{MS_e}}{r}\left[1+\left(r-1\right)\lambda'+r\mu'\right]\right\}^{0,5}$$

$$=\left\{\frac{2\times1,876}{3}\left[1+\left(2\right)\left(0,0637\right)+\left(3\right)\left(0,0668\right)\right]\right\}^{0,5}$$

= 1,3

- Sai khác giữa hai nghiệm thức cùng cột:

$$\mathrm{Sd}_3=\left\{\frac{2\mathrm{MS_e}}{r}\left[1+r\lambda'+\left(r-1\right)\mu'\right]\right\}^{0,5}$$

$$=\left\{\frac{2\times1,876}{3}\left[1+\left(3\right)\left(0,0637\right)+\left(2\right)\left(0,0668\right)\right]\right\}^{0,5} = 1,3$$

9. Hiệu quả của kiểu bố trí mạng vuông (Lattice Squares - LS) so với mạng thường (Lattice Design - LD)

Nếu coi thí kiểu mạng vuông như là một trường hợp của mạng thường thì phương sai MS_r được xem là phương sai khối (MS_b = MS_r = 5,169). Khi đó, phương sai sai số của mạng thường $MS_{e(LD)}$ sẽ được tính từ tổng bình phương hai nguồn cột và sai số với độ tự do là $r\left(k-1\right)+\left(k-1\right)\left(rk-r-k-1\right)$.

$$\mathrm{MS_{e(LD)}}=\frac{r\left(k-1\right)\mathrm{MS_c}+\left(k-1\right)\left(rk-r-k-1\right)\mathrm{MS_e}}{r\left(k-1\right)+\left(k-1\right)\left(rk-r-k-1\right)}$$

Với r = 3, k = 5 thì:

$$\mathrm{MS_{e(LD)}}=\frac{\mathrm{MS_c}+2\mathrm{MS_e}}{3}=\frac{5,656+\left(2\right)\left(1,876\right)}{3}=3,136$$

Phương sai sai số mạng thường điều chỉnh ($\mathrm{MS'_{e(LD)}}$):

$$MS'_{e(LD)} = MS_e\left[1+\frac{rk\mu}{k+1}\right]$$

$$= MS_{e(LD)}\left[1+\frac{r\left(MS_b - MS_{e(LD)}\right)}{(k+1)(r-1)MS_b}\right]$$

Trong ví dụ này:

$$MS'_{e(LD)} = (3{,}136)\left[1+\frac{(3)(5{,}169-3{,}136)}{(6)(2)(5{,}169)}\right]$$

$$= 3{,}444$$

So với phương sai sai số mạng thường, phương sai sai số của mạng vuông đã giảm đi (%):

$$r_e(\%) = \frac{MS'_{e(LD)} - MS'_e}{MS'_{e(LD)}}.100$$

Trong thí nghiệm này:

$$r_e(\%) = \frac{3{,}444-2{,}488}{3{,}444}\times 100 = 27{,}8$$

Rõ ràng: $MS_{e(CRD)} > MS_{e(RCBD)} > MS_{e(\text{vuông LT})}$

$> MS_{e(LD)} > MS_{e(LS)}$

6.7. THÍ NGHIỆM MỘT YẾU TỐ Ở NHIỀU NƠI HOẶC NHIỀU NĂM

6.7.1. Bố trí thí nghiệm

Đây là thí nghiệm một yếu tố được bố trí ở nhiều nơi khác nhau trong một thời điểm hoặc được lặp lại nhiều năm. Các thí nghiệm được bố trí giống nhau cho mọi điểm hoặc giống nhau hàng năm.

6.7.2. Phân tích phương sai

Đây là loại thí nghiệm mà ở mỗi nơi (địa điểm) là thí nghiệm một yếu tố bố trí theo kiểu RCBD, được thực hiện ở nhiều điểm, vì vậy nếu nhìn tổng quát thì đây là thí nghiệm hai yếu tố: các nghiệm thức và các địa điểm.

Tổng bình phương tổng số mỗi điểm do:

- Khác nhau giữa các lần lặp lại;;
- Khác nhau giữa các nghiệm thức;
- Sai số (ngẫu nhiên).

Tổng bình phương tổng số ở tất cả các điểm do:

- Khác nhau giữa các địa điểm thí nghiệm (hoặc các năm);
- Tương tác giữa lần lặp lại với địa điểm (hoặc với năm);
- Khác nhau giữa các nghiệm thức;
- Tương tác giữa nghiệm thức với điểm (hoặc với năm);
- Sai số (ngẫu nhiên).

Bảng phân tích phương sai có dạng

Nguồn	DF	SS	MS	F_{TN}
Địa điểm (L)	L - 1			
Lần lặp × Điểm	$L(r-1)$			
Nghiệm thức (t)	t - 1			
N.thức × Điểm	$(t-1)(L-1)$			
Sai số (e)	$L(t-1)(r-1)$			
Tổng số	$Ltr-1$			

Để làm rõ cách tính toán, xét ví dụ sau đây.

Một thí nghiệm khảo nghiệm 12 giống lạc, bố trí theo kiểu RCBD trên sáu điểm, mỗi điểm có bốn lần lặp lại. Kết quả năng suất lạc được ghi ở Bảng 6.22 dưới đây.

Bảng 6.22: Năng suất lạc (tấn/ha) ở 6 điểm

Giống	I	II	III	IV	Giống	I	II	III	IV
		<u>**Điểm 1**</u>					<u>**Điểm 2**</u>		
1	4,10	4,00	4,20	4,70	1	3,34	3,73	3,83	4,19
2	4,20	4,10	3,80	3,90	2	3,96	4,34	4,13	3,92
3	4,20	4,20	4,00	3,60	3	4,28	4,32	3,74	4,03
4	3,80	3,40	3,00	3,10	4	3,55	3,16	3,20	3,45
5	3,50	3,50	3,40	3,20	5	3,10	3,27	3,15	3,55
6	4,00	4,20	4,00	4,00	6	3,34	4,02	3,52	3,38
7	4,10	4,30	3,80	4,20	7	3,25	3,20	3,30	3,27
8	3,60	3,20	3,80	3,50	8	3,14	3,17	2,85	2,92
9	3,80	3,90	3,60	3,90	9	3,63	4,14	4,01	3,43
10	4,00	4,20	4,40	4,00	10	3,38	3,37	3,93	3,36
11	4,00	4,00	4,20	4,60	11	3,72	3,86	3,93	3,22
12	3,70	3,50	3,80	3,10	12	2,93	3,53	3,65	2,95
Cộng	47,00	46,50	46,00	45,80	Cộng	78,00	38,60	38,00	40,70
		<u>**Điểm 3**</u>					<u>**Điểm 4**</u>		
1	3,35	3,60	3,00	3,35	1	3,10	3,40	3,10	3,20
2	3,55	3,20	3,30	3,30	2	3,40	3,50	3,30	3,30
3	3,20	3,20	3,60	3,50	3	3,60	3,60	3,60	3,20
4	3,40	3,80	3,30	3,50	4	2,80	3,00	3,40	2,90
5	3,50	2,70	3,00	3,07	5	3,30	3,00	3,40	3,00
6	2,40	2,50	2,45	2,65	6	3,20	3,60	3,80	2,80
7	3,30	3,30	3,30	3,15	7	3,40	2,80	3,60	2,60
8	3,20	3,50	3,60	3,50	8	3,50	3,00	3,50	2,90
9	3,30	3,70	2,95	3,20	9	3,00	3,20	3,80	3,40
10	3,30	3,45	3,40	3,35	10	3,20	3,30	2,90	3,20
11	3,20	3,50	3,50	3,25	11	3,00	2,80	3,20	3,20
12	3,25	3,30	3,00	2,80	12	3,10	2,80	3,10	3,00
Cộng	41,62	44,11	43,24	41,67	Cộng	78,00	35,06	34,81	32,68

	Điểm 5					Điểm 6			
1	3,26	3,40	3,05	2,95	1	3,50	3,10	3,20	3,50
2	3,63	3,20	3,35	3,51	2	3,20	3,60	3,20	3,60
3	3,00	2,80	3,05	3,19	3	3,50	3,00	3,50	3,00
4	2,85	3,22	2,68	2,40	4	3,20	2,80	2,70	3,20
5	2,46	1,90	1,95	2,00	5	2,20	2,20	2,20	2,20
6	2,45	1,95	2,20	2,15	6	2,00	2,50	2,20	2,70
7	3,00	3,20	2,85	2,68	7	2,90	3,20	3,00	3,50
8	2,80	3,25	3,10	2,75	8	3,30	3,20	2,80	3,00
9	2,56	2,88	2,50	3,10	9	3,20	3,00	3,30	3,80
10	3,35	3,68	3,15	3,20	10	3,20	3,70	3,60	3,10
11	2,80	2,77	2,50	3,25	11	2,80	3,10	2,80	2,50
12	2,90	2,56	2,30	2,45	12	3,20	3,10	2,80	3,10
Cộng	38,95	39,75	38,40	38,62	Cộng	78,00	36,20	36,50	35,30
Tổng cộng (ΣΣ)									**947,03**

Tổng năng suất (ΣNS) và năng suất trung bình (m) của các giống trên toàn thí nghiệm

Giống	*1*	*2*	*3*	*4*	*5*	*6*
ΣNS	84,15	86,49	84,91	75,81	68,75	72,01
m	3,51	3,60	3,54	3,16	2,86	3,00
Giống	*7*	*8*	*9*	*10*	*11*	*12*
ΣNS	79,2	77,08	81,3	83,72	79,7	73,92
m	3,30	3,21	3,39	3,49	3,32	3,08

Việc phân tích phương sai được thực hiện hai bước chính: Phân tích, đánh giá từng điểm và phân tích tổng hợp tất cả các điểm (gọi tắt là toàn thí nghiệm).

Bước 1: Phân tích phương sai và đánh giá từng điểm

Việc phân tích phương sai từng điểm được thực hiện như kiểu thí nghiệm RCBD. Kết quả phân tích được trình bày ở Bảng 6.23.

Bảng 6.23: Kết quả phân tích ANOVA ở sáu điểm

BANG PHAN TICH PHUONG SAI DIA DIEM 1

Nguon	SS	DF	MS	Ftn	Prob.
Giong	4.842	11	0.440	7.051	0.0000
Lap lai	0.072	3	0.024	0.386	
Ngau nhien	2.060	33	0.062		
Toan bo	6.975	47			

So dieu chinh 715.335
Tong các binh phuong 722.310

BANG PHAN TICH PHUONG SAI DIA DIEM 2

Nguon	SS	DF	MS	Ftn	Prob.
Giong	5.154	11	0.469	6.965	0.0000
Lap lai	0.375	3	0.125	1.858	
Ngau nhien	2.220	33	0.067		
Toan bo	7.750	47			

So dieu chinh 606.625
Tong các binh phuong 614.375

BANG PHAN TICH PHUONG SAI DIA DIEM 3

Nguon	SS	DF	MS	Ftn	Prob.
Giong	3.133	11	0.285	6.524	0.0000
Lap lai	0.088	3	0.029	0.669	
Ngau nhien	1.441	33	0.044		
Toan bo	4.662	47			

So dieu chinh 505.162
Tong các binh phuong 509.824

BANG PHAN TICH PHUONG SAI DIA DIEM 4

Nguon	SS	DF	MS	Ftn	Prob.
Giong	1.087	11	0.099	1.602	0.1441
Lap lai	0.695	3	0.232	3.757	
Ngau nhien	2.035	33	0.062		
Toan bo	3.817	47			

So dieu chinh 494.083
Tong các binh phuong 497.900

BANG PHAN TICH PHUONG SAI DIA DIEM 5

Nguon	SS	DF	MS	Ftn	Prob.
Giong	7.413	11	0.674	11.762	0.0000
Lap lai	0.303	3	0.101	1.761	
Ngau nhien	1.891	33	0.057		
Toan bo	9.607	47			

So dieu chinh 386.297

Tong các binh phuong 395.904

BANG PHAN TICH PHUONG SAI DIA DIEM 6

Nguon	SS	DF	MS	Ftn	Prob.
Giong	6.880	11	0.625	9.759	0.0000
Lap lai	0.155	3	0.052	0.806	
Ngau nhien	2.115	33	0.064		
Toan bo	9.150	47			

So dieu chinh 439.230

Tong các binh phuong 448.380

Bước 2: Phân tích phương sai toàn thí nghiệm

1. Tính số điều chỉnh toàn thí nghiệm (CF_{TO})

$$CF_{TO} = \left(\sum_{1}^{Ltr} x_{ij} \right)^2 / Ltr = (947,03)^2 : 288 = 3.114,098$$

2. Tổng bình phương địa điểm (SS_L)

Tổng bình phương SS_L = Tổng số điều chỉnh ở các điểm (CF_L) – Số điều chỉnh toàn thí nghiệm (CF_{TO}):

$$SS_L = \sum (CF_L) - CF_{TO}$$

$$= (715,335 + 606,625 + ... + 439,230) - 3.114,098 = 32.635$$

2. Tổng bình phương lần lặp toàn thí nghiệm ($SS_{R/L}$)

Tổng bình phương $SS_{R/L}$ = Tổng các tổng bình phương lần lặp lại ở các điểm:

$$SS_{R/L} = \Sigma SS_r = 0{,}072 + 0{,}375 + 0{,}088 + 0{,}695$$
$$+ 0{,}303 + 0{,}155 = 1{,}688$$

4. Tổng bình phương nghiệm thức (giống) toàn thí nghiệm (SS_v)

$$SS_v = \frac{\sum\left(\sum_{i=1}^{L} V_i\right)^2}{rL} - CF_{TO}$$

$$= \frac{(84{,}15)^2 + (86{,}49)^2 + \ldots + (73{,}92)^2}{24}$$

$$- 3.114{,}098 = 14{,}133$$

5. Tổng bình phương tương tác nghiệm thức (giống) × điểm

Tổng bình phương tương tác giữa nghiệm thức (giống) × điểm ($SS_{v\times L}$) = Tổng các tổng bình phương nghiệm thức (giống) từ mỗi điểm ($SS_{v/L}$) – Tổng bình phương nghiệm thức (giống) toàn thí nghiệm (SS_v):

Tổng các tổng bình phương nghiệm thức (giống) từ mỗi điểm:

$$SS_{v/L} = \sum_{i=1}^{L} SS_{v(L)} = 4{,}842 + 5{,}154 + 3{,}133$$
$$+ 1{,}087 + 7{,}413 + 6{,}880 = 28{,}510$$

và: $$SS_{v\times L} = SS_{v/L} - SS_v = 28{,}510 - 14{,}133 = 14{,}377$$

6. Tổng bình phương sai số toàn thí nghiệm

Tổng bình phương sai số toàn thí nghiệm (SS_e) = Tổng bình phương sai số các điểm:

$$SS_e = \sum_{i=1}^{L} SS_{e(L)} = 2{,}060 + 2{,}220 + 1{,}441 +$$

$2{,}035 + 1{,}891 + 2{,}115 = 11{,}762$

7. Tổng bình phương tổng số toàn thí nghiệm

Tổng bình phương tổng số toàn thí nghiệm (SS_{TO}) = Tổng của tổng các bình phương các điểm – CF_{TO}:

$$SS_{TO} = (722{,}310 + 614{,}375 + 509{,}824 + 497{,}900 + 395{,}904 + 448{,}380) - 3.114{,}098 = 74{,}595$$

8. Tính phương sai các nguồn biến động

$$MS_{nguồn} = SS_{nguồn} / DF_{nguồn}$$

Cho giống: $MS_v = SS_v / DF_v = (14{,}133)/(11) = 1{,}285$

Tương tự ta có kết quả ở Bảng 6.24

9. Tính giá trị F_{TN} địa điểm, nghiệm thức (giống), nghiệm thức × điểm và kết quả phân tích phương sai.

$$F_{TN(L)} = \frac{MS_L}{MS_{r/L}} = \frac{6{,}527}{0{,}094} = 69{,}61$$

$$F_{TN(v)} = \frac{MS_v}{MS_{v\times L}} = \frac{1{,}285}{0{,}261} = 4{,}92$$

$$F_{TN(v\times L)} = \frac{MS_{v\times L}}{MS_e} = \frac{0{,}026}{0{,}059} = 4{,}40$$

Bảng 6.24: Kết quả phân tích phương sai toàn thí nghiệm

Nguon	SS	DF	MS	Ftn	Prob.
Dia diem	32.635	5	6.527	69.612	0.0000
Lap lai/diem	1.688	18	0.094	1.578	0.0685
Giong	14.133	11	1.285	4.915	0.0000
Giong*Diem	14.377	55	0.261	4.400	0.0000
Ngau nhien	11.762	198	0.059		
Toan bo	74.595	287			

10. So sánh giữa các nghiệm thức và tương tác

- So sánh các nghiệm thức mỗi điểm được tính toán như kiểu RCBD.

- So sánh giữa các nghiệm thức (giống) và tương tác toàn thí nghiệm:

Sai khác giữa hai nghiệm thức (giống):

$$Sd_v = \sqrt{\frac{2MS_{v\times L}}{rL}} = \sqrt{\frac{(2)(0,261)}{24}} = 0,15$$

$$LSD_{0,01} = t_{0,01}Sd_v = (2,668)(0,15) = 0,39$$

$t_{0.01}$ tra với df = 55.

Sai khác giữa hai mức tương tác nghiệm thức × điểm:

$$Sd_{v\times L} = \sqrt{\frac{2MS_e}{r}} = \sqrt{\frac{(2)(0,059)}{4}} = 0,17$$

$$LSD_{0,01} = t_{0,01}Sd_{v\times L} = (2,601)(0,17) = 0,45$$

$t_{0.01}$ tra với df = 198.

Phân hạng

Theo giống			Theo giống × điểm (16 vị trí đầu)			
Giống	TB (tấn/ha)	Hạng	Giống	Điểm	NS (tấn/ha)	Hạng
2	3,60	A	1	1	4,25	A
3	3,54	AB	11	1	4,20	AB
1	3,51	AB	10	1	4,15	AB
10	3,49	AB	7	1	4,10	AB
9	3,39	ABC	3	3	4,09	AB
11	3,32	ABCD	2	3	4,09	AB
7	3,30	ABCD	6	1	4,05	AB
8	3,21	BCDE	2	1	4,00	ABC
4	3,16	BCDE	3	1	4,00	ABC
12	3,08	CDE	9	3	3,80	BCD
6	3,00	CDE	9	1	3,80	BCD
5	2,86	E	1	3	3,77	BCD
			6	3	3,57	CD
			8	1	3,53	D
			12	1	3,53	D
			10	3	3,51	D

Chương 7

PHÂN TÍCH PHƯƠNG SAI THÍ NGHIỆM NHIỀU YẾU TỐ

Tính chất: Thí nghiệm nhiều yếu tố cho phép đánh giá tác động của từng yếu tố đơn lẻ và tương tác của các yếu tố đến đối tượng nghiên cứu. Ví dụ như nghiên cứu mật độ gieo trồng và chế độ bón phân cho các giống; nghiên cứu các liều lượng đạm, lân và kali và phối hợp giữa chúng.

Các kiểu bố trí: Tùy điều kiện và mục đích, có thể bố trí thí nghiệm hai, ba, bốn yếu tố với các kiểu: Hoàn toàn ngẫu nhiên (CRD), khối đầy đủ ngẫu nhiên (RCBD), lô phụ (Split Plot Design), lô sọc (Strip Plot Design), lô phụ kép (Split-Split Plot Design), phối hợp lô sọc - lô phụ (Strip-Split Plot Design) hay lũy thừa (Fractional Factorial Design).

Dưới đây chỉ đề cập đến các kiểu thông dụng trong nghiên cứu hai yếu tố và ba yếu tố.

7.1. THÍ NGHIỆM HAI YẾU TỐ KIỂU HOÀN TOÀN NGẪU NHIÊN (Two-Factorial CRD)

7.1.1. Hiệu quả phối hợp và bố trí thí nghiệm

Thí nghiệm hai yếu tố là thí nghiệm cùng một lúc người ta phối hợp hai yếu tố, mỗi yếu tố tối thiểu là hai mức. Nếu một trong hai yếu tố chỉ có một mức thì sẽ trở thành thí nghiệm một yếu tố. Hai yếu tố được phối hợp ở bảng sau.

Phối hợp hai yếu tố		Yếu tố B			
		B_1	B_2	...	B_j
Yếu tố A	A_1	A_1B_1	A_1B_2	...	A_1B_j
	A_2	A_2B_1	A_2B_2	...	A_2B_j
	...	...	...	...	...
	A_i	A_iB_1	A_iB_2	...	A_iB_j

Hiệu quả của các mức yếu tố A (nếu có) khi cố định B:

$$A_1B_n \neq A_2B_n \neq \ldots \neq A_nB_n$$

Hiệu quả của các mức yếu tố B (nếu có) khi cố định A:

$$A_nB_1 \neq A_nB_2 \neq \ldots \neq A_nB_j$$

Hiệu quả của các mức tương tác $A_i \times B_j$ (nếu có) khi:

$$A_iB_j - m - a_i - b_j > 0$$

trong đó: A_iB_j là giá trị nghiệm thức phối hợp mức i của yếu tố A và mức j của yếu tố B; m là trung bình của tất cả các nghiệm thức A_iB_j; a_i là hiệu quả mức i của yếu tố A; b_j là hiệu quả mức j của yếu tố B.

Ở ví dụ sau đây, hiệu quả A_i (a_i), B_j (b_i) và tương tác $A_i \times B_j$ (ab_{ij}) được tính như sau:

B / A	B_1	B_2	B_3	TB A	Hiệu quả A_i
A_1	84,33	79,43	77,33	80,36	-1,44
A_2	88,70	93,60	85,17	89,16	7,35
A_3	84,43	85,67	74,43	81,51	-0,30
A_4	79,43	85,00	64,17	76,20	-5,61
TB B	84,22	85,93	75,28	81,81 = m	
Hiệu quả B_j	2,42	4,12	-6,53		

Ở bảng trên:

Hiệu quả A_1: a_1 = TB $A_1 - m = 80{,}36 - m = -1{,}44$

Hiệu quả A_2: a_2 = TB $A_2 - m = 89{,}16 - m = 7{,}35$ ...

và:

Hiệu quả B_1: b_1 = TB $B_1 - m = 84{,}22 - m = 2{,}42$

Hiệu quả B_2: b_2 = TB $B_2 - m = 85{,}93 - m = 4{,}12$ ...

Hiệu quả tương tác $A_i \times B_j$ được tính:

Hiệu quả $A_1 \times B_1$: $ab_{11} = 84{,}33 - 81{,}81 - (-1{,}44) - 2{,}42 = 1{,}54$

Hiệu quả $A_2 \times B_1$: $ab_{21} = 88{,}70 - 81{,}81 - 7{,}35 - 2{,}42 = -2{,}87$

Để đánh giá được sai số thí nghiệm và mức ý nghĩa phân biệt LSD_α giữa các mức trong từng yếu tố và giữa các mức tương tác cần phải có giá trị của các lần lặp lại của các nghiệm thức.

Cũng như thí nghiệm kiểu hoàn toàn ngẫu nhiên một yếu tố, thí nghiệm kiểu CRD hai yếu tố cũng yêu cầu điều kiện ngoài yếu tố thí nghiệm hoàn toàn đồng nhất. Điều kiện tốt nhất để tiến hành kiểu thí nghiệm này là những thí nghiệm trong chậu, hay trong phong thí nghiệm. Các chậu của mỗi một nghiệm thức đều có cơ hội như nhau ở mọi chỗ đặt. Cũng do vị trí đặt chậu không ảnh hưởng đến kết quả thí nghiệm nên để tiện theo dõi ta có thể đặt các chậu trong một nghiệm thức nằm cạnh nhau như kiểu CRD một yếu tố.

Để bố trí kiểu thí nghiệm này trên đồng ruộng phải chọn khu đất đồng đều và bố trí sao cho mọi nghiệm thức có các ô nằm rãi đều trong khu đất thí nghiệm để hạn chế sai số do đất gây ra.

Số lần lặp lại (số chậu, hay số ô) CRD hai yếu tố có thể bằng nhau hay khác nhau. Tuy nhiên để thuận lợi cho việc tính toán nên bố trí số lần lặp lại bằng nhau.

7.1.2. Mô hình toán và phân tích phương sai

7.1.2.1. Mô hình toán học và nguyên lý tính toán

Như đã nêu ở mục trên, giá trị của ô nghiệm thức mức *i* của yếu tố A phối hợp với mức *j* của yếu tố B ở ô lặp *k* (X_{ijk}) được tính theo mô hình:

$$X_{ijk} = m + a_i + b_j + ab_{ij} + e_{ijk}$$

trong đó: m là giá trị trung bình toàn thí nghiệm, a_i là giá trị do mức *i* của yếu tố A tạo ra (đóng góp vào giá trị ô), b_j là giá trị do mức *j* của yếu tố B tạo ra (đóng góp), ab_{ij} là giá trị do tương tác *ij* của A×B tạo ra, e_{ij} là sai số do ô *k* tạo nên.

Theo mô hình này tổng bình phương tổng số (SS_{TO}) được tạo ra do bốn nguyên nhân:

- Sự khác nhau giữa các mức yếu tố A;

- Sự khác nhau giữa các mức yếu tố B;

- Sự khác nhau giữa các tương tác A×B với các mức phối hợp khác nhau;

- Sự sai khác giữa các ô lặp lại trong mỗi nghiệm thức do sai số (Error).

Trong trường hợp số ô lặp lại của các nghiệm thức bằng nhau, bảng phân tích phương sai có dạng

Nguồn	DF	SS	MS	F_{TN}
Nghiệm thức AB	$ab - 1$	SS_t	SS_t/df_t	
- Yếu tố A	$a - 1$	SS_a	SS_t/df_a	
- Yếu tố B	$b - 1$	SS_b	SS_t/df_b	
- Tương tác A×B	$(a - 1)(b - 1)$	SS_{ab}	SS_t/df_{ab}	
Sai số (e)	$ab(r - 1)$	SS_e	SS_e/df_e	
Tổng số (TO)	$abr - 1$	SS_T		

Trong bảng:

Yếu tố A có *a* mức, độ tự do $df_a = a - 1$;

Yếu tố B có *b* mức, độ tự do $df_b = b - 1$;

Có *ab* mức tương tác A×B, độ tự do $df_{ab} = (a - 1)(b - 1)$;

Số nghiệm thức t = *ab*, độ tự do $df_t = ab - 1$;

Tổng số ô thí nghiệm (TO) = *abr*, độ tự do $df_T = abr - 1$; *r* là số chậu (ô) của mỗi nghiệm thức;

Việc phân tích phương sai tiến hành 2 bước:

1. Tính tổng bình phương tổng số, tổng bình phương và phương sai nghiệm thức, tổng bình phương và phương sai sai số, kiểm tra sự khác biệt giữa các nghiệm thức bằng phép nghiệm F.

$$F_{TN(t)} = \frac{MS_t}{MS_e}$$

2. Khi $F_{TN(t)} < f_{0,05}$, tức là sự khác nhau giữa các nghiệm thức nằm trong phạm vi sai số thì không tiếp tục trắc nghiệm F cho bất cứ yếu tố nào. Khi xử lý số liệu số liệu bằng các phần mềm tính toán (hay tính trên Excel), trong nhiều trường hợp, mặc dù $F_{TN(t)} < f_{0,05}$ nhưng có thể thấy giá trị F_{TN} của yếu tố nào đó lớn hơn $f_{0,05}$. Điều đó không được thừa nhận vì khi sai khác giữa các mức của yếu tố đó đã không có ý nghĩa thì việc tính toán sau đó chỉ là kết quả tính toán từ các nguồn biến động do sai số gây ra.

Khi $F_{TN(t)} \geq F_{0,05}$, tức là có sự khác nhau giữa các nghiệm thức, tiếp tục tính tổng bình phương và phương sai của từng yếu tố A và B và tương tác A×B, đánh giá sự sai giữa các mức trong mỗi yếu tố A và B và giữa các mức tương tác A×B bằng phép nghiệm F.

$$F_{TN(A)} = \frac{MS_a}{MS_e};\ F_{TN(B)} = \frac{MS_b}{MS_e} \text{ và } F_{TN(A\times B)} = \frac{MS_{a\times b}}{MS_{e(b)}}$$

Sai khác giữa hai mức yếu tố A:

$$Sd_A = \sqrt{\frac{2MS_e}{br}}$$

Sai khác giữa hai mức yếu tố B:

$$Sd_B = \sqrt{\frac{2MS_e}{ar}}$$

Sai khác giữa hai mức tương tác A×B:

$$Sd_{A\times B} = \sqrt{\frac{2MS_e}{r}}$$

LSD_α (cho A) = $t_\alpha Sd_A$
LSD_α (cho B) = $t_\alpha Sd_B$
LSD_α (cho A×B) = $t_\alpha Sd_{A\times B}$

Sai khác giữa các nghiệm thức (t):

$$Sd_t = \sqrt{\frac{2MS_e}{r}}$$

LSD_α (cho t) = $t_\alpha Sd_t$

Theo mô hình này $Sd_t = Sd_{A\times B}$

Việc xác định sai số trung bình, sai số tương đối, hệ số biến động theo cách đã biết.

Để làm rõ phương pháp phân tích phương sai cho thí nghiệm hai yếu tố kiểu hoàn toàn ngẫu nhiên, hãy tiến hành với ví dụ sau.

7.1.2.2. Ví dụ ứng dụng

Ví du: Đánh giá ảnh hưởng của 3 kiểu bấm ngọn khác nhau cho 4 giống bông trồng trong chậu với số liệu ở Bảng 7.1.

Bảng 7.1: Năng suất bông ở các nghiệm thức (g/chậu)

Bấm ngọn (Yếu tố B)(*)		Giống (Yếu tố A)				Tổng A_iB_j	Trung bình
			A_1 (giống 1)				
B_1	8 cành	66	69	69	70	274	68,5
B_2	12 cành	67	69	71	70	277	69,3
B_3	16 cành	68	73	72	74	287	71,8
			A_2 (giống 2)				
B_1	8 cành	64	70	71	74	279	69,8
B_2	12 cành	65	70	72	70	277	69,3
B_3	16 cành	68	73	72	73	286	71,5
			A_3(giống 3)				
B_1	8 cành	71	73	74	71	289	72,3
B_2	12 cành	69	76	75	72	292	73,0
B_3	16 cành	69	70	72	73	284	71,0
			A_4 (giống 4)				
B_1	8 cành	72	73	70	67	282	70,5
B_2	12 cành	70	72	73	70	285	71,3
B_3	16 cành	74	75	75	76	300	75,0
Tổng toàn bộ						**3.412**	

(*) *Bấm lúc có 8, 12 và 16 cành.*

Bảng 7.2: Tổng năng suất 2 yếu tố và phối hợp

Phối hợp		$A_1(V_1)$	$A_2(V_2)$	$A_3(V_3)$	$A_4(V_4)$	Tổng B_j
B_1	8 cành	274	279	289	282	1.124
B_2	12 cành	277	277	292	285	1.131
B_3	16 cành	287	286	284	300	1.157
Tổng A_i		838	842	865	867	3.412

Giải:

Ở đây, yếu tố A có 4 giống, yếu tố B có 3 kiểu bấm, tổng số nghiệm thức t = 12, mỗi nghiệm thức có r = 4 chậu, tổng số chậu (T) rt = 48.

Bước 1: Kiểm tra sự khác biệt giữa các nghiệm thức

$$CF = \frac{\left(\sum_{i=1}^{ab} x_{ij}\right)^2}{abr} = \frac{(3.412)^2}{(4)(3)(4)} = 242.536,333$$

Tổng bình phương tổng số:

$$SS_{TO} = \sum_{i=1}^{ab} x_{ij}^2 - CF = (66)^2 + (69)^2 + \dots$$
$$+ (75)^2 + (76)^2 - 242.536,333 = 357,667$$

Tổng bình phương nghiệm thức:

$$SS_t = \frac{\sum_{i=1}^{ab} T_i^2}{r} - CF$$
$$= \frac{(274)^2 + (277)^2 + \dots + (285)^2 + (300)^2}{4}$$
$$- 242.536,333 = 146,167$$

Tổng bình phương sai số:

$$SS_e = SS_T - SS_t = 357,667 - 146,167 = 211,500$$

Bảng 7.3: Kết quả phân tích phương sai bước đầu

Nguồn	DF	SS	MS	F_{TN}	Prob.
Nghiệm thức	11	146,167	13,288	2,262	0,0322
Sai số	36	211,500	5,875		
Tổng số	47	357,167	13,288	2,262	

Kết quả cho thấy $F_{TN(t)} > F_{0,05}$ ($P < 0,05$) cho phép tiếp tục tiến hành bước 2.

***Bước 2*: Đánh giá sự sai khác của các nguồn biến động và so sánh giữa các mức của từng yếu tố, các mức tương tác và giữa các nghiệm thức.**

1. Đánh giá sự sai khác của các nguồn biến động

Tổng bình phương yếu tố A (giống):

$$SS_A = \frac{\sum_{i=1}^{a} A_i^2}{br} - CF$$

$$= \frac{(838)^2 + (842)^2 + (865)^2 + (867)^2}{12}$$

$$- 242.536,333 = 57,167$$

Tổng bình phương yếu tố B (bấm ngọn):

$$SS_B = \frac{\sum_{j=1}^{b} B_j^2}{ar} - CF = \frac{(1.124)^2 + (1.131)^2 + (1.157)^2}{16}$$

$$- 242.536,333 = 37,792$$

Tổng bình phương tương tác A×B:

$$SS_{A \times B} = SS_t - SS_A - SS_B$$

$$= 146,167 - 57,167 - 37,792 = 51,208$$

Kết quả phân tích phương sai được ghi ở bảng 7.4.

Bảng 7.4 cho thấy:

$P_{(F/A)} < 0,05$ và $P_{(F/B)} = 0,05$ cho thấy có sự khác nhau giữa các mức A và giữa các mức B. $P_{(F/A\times B)} > 0,05$ cho thấy không có sự khác nhau giữa các mức tương tác.

Bảng 7.4: Kết quả phân tích phương sai

Nguồn	DF	SS	MS	F_{TN}	Prob.
Nghiệm thức	11	146,167	13,288	2,26	0,0322
A	3	57,167	19,056	3,24	0,0331
B	2	37,794	18,896	3,22	0,0519
A×B	6	51,208	8,535	1,45	0,2222
Sai số	36	211,500	5,875		
Tổng số	47	357,167			

2. So sánh từng yếu tố và tương tác

- Sai khác giữa hai mức yếu tố A:

$$Sd_A = \sqrt{\frac{2MS_e}{br}} = \sqrt{\frac{(2)(5,875)}{12}} = 0,99$$

$$LSD_{0,05}\ (A) = t_{0,05}Sd_A$$

$$= 2,028 \times 0,99 = 2,0$$

- Sai khác giữa hai mức yếu tố B:

$$Sd_B = \sqrt{\frac{2MS_e}{ar}} = \sqrt{\frac{(2)(5,875)}{16}} = 0,86$$

$$LSD_{0,05}\ (B) = t_{0,05}Sd_B$$

$$= 2,028 \times 0,86 = 1,7$$

(do $t_{0,05}$ với độ tự do 36 = 2,029 khác nhau không nhiều với $t_{0,0519}$ = 2,010 nên khi lấy một số lẻ thì giá trị LSD bằng nhau và bằng 1,7).

Bảng 7.5: Kết quả xếp hạng các mức A và các mức B với độ tin cậy 95% và hiệu quả của các mức A và các mức B

Yếu tố A	Yếu tố B			Tr. bình A	Hiệu quả A_i (a_i)
	B_1	B_2	B_3		
A_1	68,5	69,3	71,8	$69,87^{C}$	- 1,24
A_2	69,8	69,3	71,5	$70,20^{BC}$	- 0,91
A_3	72,3	73,0	71,0	$72,10^{AB}$	0,99
A_4	70,5	71,3	75,0	$72,27^{A}$	1,16
Tr. bình B	$70,28^{B}$	$70,73^{AB}$	$72,33^{A}$	m = 71,11	
Hiệu quả B_j (b_j)	- 0,83	- 0,38	1,22		

Trong bảng: Hiệu quả A_i (a_i) = Trung bình A_i - m

Hiệu quả B_j (b_j) = Trung bình B_j – m

3. Hiệu quả tương tác $A_i \times B_j$

Giả thiết rằng tất cả các tương tác ab_{ij} đều bằng 0, giá trị ($X_{ij} - m$) lý thuyết của các nghiệm thức (NT) được tính như sau:

$$
\begin{array}{c}
\text{NT} \\
\\
\begin{pmatrix} A_1B_1 \\ A_1B_2 \\ A_1B_3 \\ A_2B_1 \\ A_2B_2 \\ A_2B_3 \\ A_3B_1 \\ A_3B_2 \\ A_3B_3 \\ A_4B_1 \\ A_4B_2 \\ A_4B_3 \end{pmatrix}
\end{array}
=
\begin{array}{c}
\begin{array}{ccccccc} a_1 & a_2 & a_3 & a_4 & b_1 & b_2 & b_3 \\ -1{,}24 & -0{,}91 & 0{,}99 & 1{,}16 & -0{,}83 & -0{,}38 & 1{,}22 \end{array} \\
\begin{pmatrix}
1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 1
\end{pmatrix}
\end{array}
=
\begin{array}{c}
X_{ij} - m \\
\\
\begin{pmatrix} -2{,}07 \\ -1{,}62 \\ -0{,}02 \\ -1{,}74 \\ -1{,}29 \\ 0{,}31 \\ 0{,}16 \\ 0{,}61 \\ 2{,}21 \\ 0{,}33 \\ 0{,}78 \\ 2{,}38 \end{pmatrix}
\end{array}
$$

Gía trị chênh lệch giữa ($X_{ij} - m$) thực tế và ($X_{ij} - m$) lý thuyết chính là hiệu quả tương tác:

Nghiệm thức (A_iB_j)	Giá trị thực tế (X_{ij})	$X_{ij} - m$ Thực tế	$X_{ij} - m$ Lý thuyết	Hiệu quả tương tác (ab_{ij})
A_1B_1	68,5	-2,61	-2,07	-0,54
A_1B_2	69,3	-1,81	-1,62	-0,19
A_1B_3	71,8	0,69	-0,02	0,71
A_2B_1	69,8	-1,31	-1,74	0,43
A_2B_2	69,3	-1,81	-1,29	-0,52
A_2B_3	71,5	0,39	0,31	0,08
A_3B_1	72,3	1,19	0,16	1,03
A_3B_2	73,0	1,89	0,61	1,28
A_3B_3	71,0	-0,11	2,21	-2,32
A_4B_1	70,5	-0,61	0,33	-0,94
A_4B_2	71,3	0,19	0,78	-0,59
A_4B_3	75,0	3,89	2,38	1,51

Bảng 7.6: Hiệu quả tương tác ab_{ij}

Yếu tố A	Yếu tố B		
	B_1	B_2	B_3
A_1	- 0,54	- 0,19	0,71
A_2	0,43	- 0,52	0,08
A_3	1,03	1,28	- 2,32
A_4	- 0,94	- 0,59	1,51

Thực chất, các số liệu bảng 7.6 được tính từ bảng 7.5 bằng cách lấy mỗi giá trị $X_{ij} - m - a_i - b_j$.

Độ tin cậy của tương tác được xác định:

$$LSD_{0,05} = t_{0,05}\sqrt{\frac{MS_e}{r}} = (2,028)\sqrt{\frac{5,875}{4}} = 2,5$$

Không có một sự tương tác nào tin cậy do phương sai SS_{AB} không có ý nghĩa thống kê (P > 0,05).

Những tính toán trên đây là phương pháp xác định hiệu quả tương tác để áp dụng trong tính toán các thí nghiệm tương tự.

3. Phân hạng các nghiệm thức

Sai khác giữa các nghiệm thức:

$$Sd_t = \sqrt{\frac{2MS_e}{r}} = \sqrt{\frac{2\times5,875}{4}} = 1,7$$

$$LSD_{0,05}\ (\text{cho t}) = t_{0,05}\, Sd_t = 3,5$$

Kết quả phân hạng các nghiệm thức như sau:

Nghiệm thức	Giá trị	Phân hạng	
		$LSD_{0,05}$	Duncan (α = 0,05)
A_4B_3	75,0	A	A
A_3B_2	73,0	AB	AB
A_3B_1	72,3	ABC	ABC
A_1B_3	71,8	ABCD	ABC
A_2B_3	71,5	BCD	ABC
A_4B_2	71,3	BCD	ABC
A_3B_3	71,0	BCD	BC
A_4B_1	70,5	BCD	BC
A_2B_1	69,8	BCD	BC
A_2B_2	69,3	CD	BC
A_1B_2	69,3	CD	BC
A_1B_1	68,5	D	C

Kết quả phân tích phương sai trên phần mềm SAS 9.1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	146.1667	13.2878788	2.26	0.0322
Error	36	211.5000	5.8750000		
Corr.Total	47	357.6667			

R-Square	Coeff Var	Root MSE	Y Mean
0.4087	3.4098	2.4238	71.0833

Source	DF	Type I SS	Mean Square	F Value	Pr > F
A	3	57.16667	19.0555	3.24	0.0331
B	2	37.79167	18.8958	3.22	0.0519
A*B	6	51.20833	8.5347	1.45	0.2222

Alpha	0.05
Critical Value of t	2.0281
Least Significant Difference	2.0069

Means with the same letter are not significantly different.

t Grouping			Mean	N	A
	A		72.25	12	4
B	A		72.08	12	3
B		C	70.17	12	2
		C	69.83	12	1

NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Critical Value of t	2.0281
Least Significant Difference	1.738

Means with the same letter are not significantly different.

t Grouping		Mean	N	B
	A	72.31	16	3
B	A	70.68	16	2
B		70.25	16	1

NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Critical Value of t	2.0281
Least Significant Difference	3.476

Means with the same letter are not significantly different.

t Grouping			Mean	N	AB
		A	75.000	4	A4B3
B		A	73.000	4	A3B2
B		A C	72.250	4	A3B1
B	D	A C	71.750	4	A1B3
B	D	C	71.500	4	A2B3
B	D	C	71.250	4	A4B2
B	D	C	71.000	4	A3B3
B	D	C	70.500	4	A4B1
B	D	C	69.750	4	A2B1
	D	C	69.250	4	A2B2
	D	C	69.250	4	A1B2
	D		68.500	4	A1B1

Means with the same letter are not significantly different.

Duncan Grouping			Mean	N	AB
	A		75.000	4	A4B3
B	A		73.000	4	A3B2
B	A	C	72.250	4	A3B1
B	A	C	71.750	4	A1B3
B	A	C	71.500	4	A2B3
B	A	C	71.250	4	A4B2
B		C	71.000	4	A3B3
B		C	70.500	4	A4B1
B		C	69.750	4	A2B1
B		C	69.250	4	A2B2
B		C	69.250	4	A1B2
		C	68.500	4	A1B1

Thảo luận về sai sót thường gặp trong bố trí thí nghiệm kiểu CRD

Mục đích của việc phân tích phương sai là xác định được các nguồn biến động gây ra trong tổng số biến động SS_{TO}, từ đó loại trừ ra khỏi SS_{TO} các tổng bình phương các nguồn biến động để xác định tổng bình phương do sai số. Sau đó sử dụng phép nghiệm F để so sánh phương sai do sự sai khác giữa các nghiệm thức với phương sai sai số. Khi giữa các nghiệm thức có sự khác nhau ở mức á thì tính toán $LSD_á$ và phân hạng các nghiệm thức.

Các thí nghiệm trong chậu, trong phòng là thí nghiệm kiểu hoàn toàn ngẫu nhiên (CRD), vị trí đặt chậu (thí nghiệm trong chậu) hay đặt bình, đĩa petri (thí nghiệm trong phòng) không ảnh hưởng đến các chỉ tiêu sinh trưởng phát triển của các nghiệm thức. Tuy nhiên có người đã "vô ý" đặt các chậu (các bình, đĩa petri) theo khối (lần lặp lại) như là một thí nghiệm đồng ruộng và khi sử dụng các phần mềm tính toán đã loại trừ tổng bình phương giữa các "khối giả" này ra khỏi tổng bình phương tổng số. Các khối này y hệt nhau về số nghiệm thức và giá thể vì vậy sự sai khác giữa các "khối" là do sai số. Việc trừ đi nguồn "biến động giả" này đã làm thấp sai số, làm sai lệch kết quả thí nghiệm.

Với ví dụ ở Bảng 7.1 trên đây, nếu coi 4 cột là 4 khối với tổng số các cột là "823", "863", "866" và "860". Kết quả phân tích phương sai sẽ có bảng sau:

Nguồn	DF	SS	MS	F_{TN}	Prob.
Nghiệm thức	11	146,167	13,288	2,26	0,001
"Khối"	3	101,500	33,833		
A	3	57,167	19,056	5,72	0,003
B	2	37,794	18,896	5,67	0,008
A×B	6	51,208	8,535	2,56	0,038
Sai số	33	110,000	3,333		
Tổng số	47	357,167			

Với kết quả này, các nghiệm thức, các mức của A, các mức của B đều khác nhau rất đáng tin cậy (P <0,01), còn giữa các mức tương tác cũng khác nhau có ý nghĩa.

Rõ ràng, chỉ bằng "thủ thuật" đặt chậu tạo khối đã đi đến kết luận sai lầm. Tính một cách đơn giản, để tạo thành "khối" đầu tiên, với mỗi nghiệm thức có 4 chậu sẽ có 4 khả năng, với 12 nghiệm thức sẽ có 48 kiểu sắp xếp và chỉ thế đã có 48 kết luận sai lầm, chưa kể tới các khả năng để sắp xếp các "khối" còn lại.

7.2. THÍ NGHIỆM HAI YẾU TỐ KIỂU KHỐI ĐẦY ĐỦ NGẪU NHIÊN (Two-Factorial RCBD)

7.2.1. Bố trí thí nghiệm

Thí nghiệm hai yếu tố kiểu RCBD là thí nghiệm đồng ruộng được bố trí như kiểu thí nghiệm một yếu tố RCBD, tức là có đầy đủ các ô của các nghiệm thức nằm trong khối và được lặp lai *r* lần. Trong một khối mỗi ô của một nghiệm thức nằm ngẫu nhiên.

Khác với thí nghiệm một yếu tố là mỗi nghiệm thức (mỗi ô) được tạo thành từ hai yếu tố với sự phối hợp giữa các mức khác nhau.

Sau đây là một số sơ đồ bố trí thí nghiệm.

I	(9) A_2B_4	(5) A_1B_5	(10) A_2B_5	(2) A_1B_2	(4) A_1B_4	(8) A_2B_3	(1) A_1B_1	(6) A_2B_1	(7) A_2B_2	(3) A_1B_3
II	(1) A_1B_1	(6) A_2B_1	(7) A_2B_2	(9) A_2B_4	(2) A_1B_2	(3) A_1B_3	(5) A_1B_5	(10) A_2B_5	(4) A_1B_4	(8) A_2B_3
III	(4) A_1B_4	(3) A_1B_3	(1) A_1B_1	(8) A_2B_3	(10) A_2B_5	(6) A_2B_1	(7) A_2B_2	(2) A_1B_2	(9) A_2B_4	(5) A_1B_5
IV	(2) A_1B_2	(8) A_2B_3	(5) A_1B_5	(6) A_2B_1	(7) A_2B_2	(9) A_2B_4	(3) A_1B_3	(1) A_1B_1	(10) A_2B_5	(4) A_1B_4

Hình 7.1: Sơ đồ thí nghiệm 2 yếu tố kiểu RCBD
(10 nghiệm thứ , 2 mức yếu tố A, 5 mức yếu tố B,
4 lần lặp lại, mỗi lần lặp lại trên một băng)

I					II				
(9)	(5)	(10)	(2)	(4)	(1)	(6)	(7)	(9)	(2)
(8)	(1)	(6)	(7)	(3)	(3)	(5)	(10)	(4)	(8)
III					**IV**				
(4)	(3)	(1)	(8)	(10)	(2)	(8)	(5)	(6)	(7)
(6)	(7)	(2)	(9)	(5)	(9)	(3)	(1)	(10)	(4)

Hình 7.2: Sơ đồ thí nghiệm 2 yếu tố kiểu RCBD
(10 nghiệm thứ , 2 mức yếu tố A, 5 mức yếu tố B,
4 lần lặp lại, mỗi lần lặp lại trên 2 băng)

I				
A_3B_2	A_2B_1	A_1B_4	A_1B_1	A_2B_3
A_1B_3	A_3B_5	A_3B_4	A_1B_2	A_3B_3
A_2B_4	A_3B_1	A_2B_5	A_1B_5	A_2B_2

II				
A_1B_1	A_3B_5	A_1B_5	A_3B_1	A_1B_4
A_2B_2	A_1B_2	A_1B_3	A_2B_4	A_3B_4
A_2B_5	A_3B_2	A_2B_1	A_2B_3	A_3B_3

III				
A_3B_3	A_3B_3	A_1B_1	A_2B_5	A_2B_1
A_1B_3	A_3B_2	A_1B_2	A_1B_4	A_2B_4
A_1B_5	A_3B_4	A_2B_2	A_3B_1	A_3B_5

IV				
A_1B_2	A_2B_2	A_2B_4	A_1B_5	A_2B_5
A_1B_3	A_3B_1	A_1B_4	A_1B_1	A_2B_3
A_3B_5	A_2B_1	A_3B_2	A_3B_3	A_3B_4

Hình 7.3: Sơ đồ thí nghiệm 2 yếu tố kiểu RCBD
(15 nghiệm thứ , 3 mức yếu tố A, 5 mức yếu tố B,
4 lần lặp lại, mỗi lần lặp lại trên 3 băng)

7.2.2. Mô hình toán và phân tích phương sai

7.2.2.1. Mô hình toán học và nguyên lý tính toán

Trong thí nghiệm 2 yếu tố kiểu RCBD, giá trị của ô nghiệm thức mức *i* của yếu tố A phối hợp với mức *j* của yếu tố B ở lần lặp lại *k* (X_{ijk}) được tính theo mô hình:

$$X_{ijk} = m + a_i + b_j + ab_{ij} + r_k + e_{ijk}$$

trong đó: *m* là giá trị trung bình toàn thí nghiệm, a_i là giá trị do mức *i* của yếu tố A tạo ra (đóng góp vào giá trị ô), b_j là giá trị do mức *j* của yếu tố B tạo ra, ab_{ij} là giá trị do tương tác giữa A_i với B_j ($A_i \times B_j$) tạo ra, r_k là giá trị tạo ra do ảnh hưởng của lần lặp lại *k*, e_{ijk} là sai số tại ô *ijk*.

A và B là cách đặt tên cho các yếu tố. Tùy cách gọi, người ta có thể đặt tên bằng các chữ cái khác nhau.

Theo mô hình này tổng bình phương tổng số (SS_{TO}) được tạo ra do bốn nguyên nhân:

- Sự khác nhau giữa các mức yếu tố A;
- Sự khác nhau giữa các mức yếu tố B;
- Sự khác nhau giữa các tương tác A×B với các mức

phối hợp khác nhau;

- Sự sai khác giữa các lần lặp lại;
- Sai số (Error).

Bảng phân tích phương sai có dạng

Nguồn	DF	SS	MS	F_{TN}
Lần lặp lại	$r - 1$	SS_r		
Nghiệm thức	$ab - 1$	SS_t	SS_t/df_t	
- Yếu tố A	$(a - 1)$	SS_a	SS_t/df_a	
- Yếu tố B	$(b - 1)$	SS_b	SS_t/df_b	
- Tương tác A×B	$(a - 1)(b - 1)$	SS_{ab}	SS_t/df_{ab}	
Sai số (e)	$(r - 1)(ab - 1)$	SS_e	SS_e/df_e	
Tổng số (TO)	$abr - 1$	SS_T		

Trong bảng:

Yếu tố A có a mức, độ tự do $df_a = a - 1$;

Yếu tố B có b mức, độ tự do $df_b = b - 1$;

Có ab mức tương tác A×B, độ tự do $df_{ab} = (a - 1)(b - 1)$;

Số nghiệm thức t = ab, độ tự do $df_t = ab - 1$;

Số lần lặp lại = r , độ tự do $df_r = r - 1$

Tổng số ô thí nghiêm (TO) = abr, độ tự do $df_{TO} = abr - 1$;

- Sai số, độ tự do $df_e = (r - 1)(ab - 1)$.

Việc phân tích kết quả nghiên cứu được tiến hành như kiểu thí nghiệm 2 yếu tố CRD nhưng trong tổng bình phương biến động có nguồn tổng bình phương lần lặp lại gây ra.

7.2.2.2. Ví dụ ứng dụng

Ví dụ: Một thí nghiệm nghiên cứu ba giống lúa, mỗi giống bón năm liều đạm: 0, 40, 70, 100, và 130 kg N/ha phối hợp thành 15 nghiệm thức (Bảng 7.7), được bố trí theo kiểu RCBD, 4 lần lặp lại theo Hình 7.3. Hãy phân tích vai trò của giống và lân đối với năng suất lúa theo số liệu ở Bảng 7.8.

Bảng 7.7: 15 nghiệm thức được phối hợp từ ba giống lúa và năm liều đạm N

Liều đạm (Yếu tố B)	Giống (Yếu tố A)		
	A_1	A_2	A_3
B_0 (0 kg N/ha)	A_1B_0	A_2B_0	A_3B_0
B_1 (40 kg N/ha)	A_1B_1	A_2B_1	A_3B_1
B_2 (70 kg N/ha)	A_1B_2	A_2B_2	A_3B_2
B_3 (100 kg N/ha)	A_1B_3	A_2B_3	A_3B_3
B_4 (130 kg N/ha)	A_1B_3	A_2B_3	A_3B_3

Bảng 7.8: Năng suất các nghiệm thức (tấn/ha)

Liều đạm	Năng suất, tấn/ha				Tổng NT (T)
	I	II	III	IV	
			A_1		
B_0	3,85	2,61	3,14	2,89	12,49
B_1	4,79	4,94	4,56	4,61	18,90
B_2	4,58	4,45	4,88	3,92	17,83
B_3	6,03	5,28	5,91	5,65	22,87
B_4	5,87	5,92	5,98	5,52	23,29
			A_2		
B_0	2,85	3,79	4,11	3,44	14,19
B_1	4,96	5,13	4,15	4,99	19,23
B_2	5,93	5,70	5,81	4,31	21,75
B_3	5,66	5,36	6,46	5,47	22,95
B_4	5,46	5,55	5,79	5,93	22,73

		A$_3$			
B$_0$	4,19	3,75	3,74	3,43	15,11
B$_1$	5,25	4,58	4,90	4,29	19,02
B$_2$	5,82	4,85	5,68	4,93	21,28
B$_3$	5,89	5,52	6,04	4,76	22,21
B$_4$	5,86	6,26	6,06	5,36	23,54
Tổng LLL (R)	76,99	73,69	77,21	69,50	
Tổng cộng					**297.39**

Bảng 7.9: Tổng năng suất 2 yếu tố và phối hợp

Yếu tố B	Tổng năng suất (AB)			Tổng N (B)
	A$_1$	A$_2$	A$_3$	
B$_0$	12,49	14,19	15,11	41,79
B$_1$	18,90	19,23	19,02	57,15
B$_2$	17,83	21,75	21,28	60,86
B$_3$	22,87	22,95	22,21	68,03
B$_4$	23,29	22,73	23,54	69,56
Tổnggiống (A)	95,38	100,85	101,16	**297,39**

Giải:

Bước 1: Kiểm tra sự khác biệt giữa các nghiệm thức

$$CF = \frac{\left(\sum_{i=1}^{ab} x_{ij}\right)^2}{abr} = \frac{(297,39)^2}{(3)(5)(4)} = 1.474,014$$

Tính tổng nghiệm thức (T), tổng lần lặp lại (R), tổng toàn thí nghiệm, tính tổng bình phương tổng số, tổng bình phương và phương sai nghiệm thức và lần lặp lại và trắc nghiệm F cho phương sai nghiệm thức.

Tổng bình phương tổng số:

$$SS_{TO} = \sum_{i=1}^{ab} x_{ij}^2 - CF = (3,85)^2 + (2,61)^2 + ...$$

$$+ (5,36)^2 - 1.474,014 = 53,540$$

Tổng bình phương lần lặp lại:

$$SS_r = \frac{\sum_{i=1}^{r} R_i^2}{ab} - CF$$

$$= \frac{(76,99)^2 + (73,69)^2 ... + (69,51)^2}{15}$$

$$- 1.474,014 = 2,607$$

Tổng bình phương nghiệm thức:

$$SS_t = \frac{\sum_{i=1}^{ab} T_i^2}{r} - CF$$

$$= \frac{(12,49)^2 + (18,90)^2 + ... + (23,54)^2}{4}$$

$$- 1.474,014 = 44,605$$

Tổng bình phương sai số:

$$SS_e = SS_{TO} - SS_t$$

$$= 53,540 - 2,607 - 44,605 = 6,328$$

Bảng 7.10: Kết quả phân tích phương sai bước đầu

Nguồn	DF	SS	MS	F_{TN}	Prob.
Lần lặp lại	3	2,607	0,869	5,77	0,0022
Nghiệm thức	14	44,605	3,186	21,15	0,0000
Sai số	42	6,328	0,151		
Tổng số	59	53,540			

Kết quả $F_{TN(t)} > F_{0,01}$ ($P < 0,01$) cho phép tiếp tục tiến hành bước 2.

Bước 2:

1. Kiểm tra sự sai khác của các nguồn biến động

Tổng bình phương yếu tố A (giống):

$$SS_A = \frac{\sum_{i=1}^{a} A_i^2}{rb} - CF$$

$$= \frac{(95,38)^2 + (100,85)^2 + (101,16)^2}{20}$$

$$- 1.474,014 = 1,057$$

Tổng bình phương yếu tố B (liều đạm):

$$SS_B = \frac{\sum_{j=1}^{b} B_j^2}{ra} - CF$$

$$= \frac{(41,79)^2 + 57,15 + ... + (69,56)^2}{12}$$

$$- 1.474,014 = 41,248$$

Tổng bình phương tương tác A×B:

$$SS_{A\times B} = SS_t - SS_A - SS_B$$

$$= 44,605 - 1,057 - 41,248 = 2,300$$

Phương sai yếu tố A:

$$MS_A = \frac{SS_A}{a-1} = \frac{1,057}{2} = 0,529$$

Phương sai yếu tố B:

$$MS_b = \frac{SS_b}{b-1} = \frac{41,248}{4} = 10,312$$

Phương sai tương tác A×B:

$$MS_{A\times B} = \frac{SS_{A\times B}}{(a-1)(b-1)} = \frac{2,300}{8} = 0,287$$

Phương sai sai số:

$$MS_e = \frac{SS_e}{(r-1)(ab-1)} = \frac{6,328}{42} = 0,151$$

Bảng 7.11: Kết quả phân tích phương sai

Nguồn	DF	SS	MS	F_{TN}	Prob.
Lần lặp lại	3	2,607	0,869	5,73	0,0022
Nghiệm thức	14	44,605	3,186	21,09	0,0000
A	2	1,057	0,529	3,51	0,0400
B	4	41,248	10,312	68,45	0,0000
A×B	8	2,300	0,287	1,91	0,0872
Sai số	42	6,328	0,151		
Tổng số	59	53,540			

Như vậy:

- Năng suất các giống khác nhau, độ tin cậy 95% ($P < 0,05$)

- Năng suất các mức đạm khác nhau cho năng suất khác nhau, độ tin cậy 99,9% ($P < 0,001$)

- Các mức tương tác giữa giống lúa và mức bón đạm khác nhau không đủ tin cậy 95% ($P > 0,05$).

2. So sánh các mức của từng yếu tố, giữa các nghiệm thức và hệ số biến động

- Sai khác giữa hai mức của yếu tố A:

$$Sd_A = \sqrt{\frac{2MS_e}{rb}} = \sqrt{\frac{(2)(0,151)}{(4)(5)}} = 0,123$$

$$LSD_{0,05} = t_{0,05}Sd_A = (2,018)(0,123) = 0,25$$

- Sai khác giữa hai mức của yếu tố B:

$$Sd_B = \sqrt{\frac{2MS_e}{ra}} = \sqrt{\frac{(2)(0,151)}{(4)(3)}} = 0,159$$

$$LSD_{0,05} = t_{0,05}Sd_B = (2,018)(0,159) = 0,32$$
$$LSD_{0,01} = t_{0,01}Sd_B = (2,698)(0,159) = 0,43$$

- Sai khác giữa hai nghiệm thức:

$$Sd_t = \sqrt{\frac{2MS_e}{r}} = \sqrt{\frac{(2)(0,151)}{4}} = 0,27$$

$$LSD_{0,05} = t_{0,05}Sd_t = (2,018)(0,27) = 0,55$$
$$LSD_{0,01} = t_{0,01}Sd_t = (2,698)(0,27) = 0,73$$

- Hệ số biến động:

$$CV(\%) = \frac{\sqrt{MS_e}}{m} \times 100 = \frac{\sqrt{0,151}}{4,96} \times 100 = 7,8$$

Bảng 7.12: Kết quả phân hạng LSD theo năng suất

Yếu tố B	Năng suất (tấn/ha)			Tổng N (B)
	A_1	A_2	A_3	
B_0	$3,12^{f}$	$3,55^{ef}$	$3,78^{e}$	$3,48^{C}$
B_1	$4,73^{d}$	$4,81^{cd}$	$4,76^{d}$	$4,76^{B}$
B_2	$4,46^{d}$	$5,44^{ab}$	$5,32^{bc}$	$5,07^{B}$
B_3	$5,72^{ab}$	$5,74^{ab}$	$5,55^{ab}$	$5,67^{A}$
B_4	$5,82^{ab}$	$5,68^{ab}$	$5,89^{a}$	$5,80^{A}$
Tổng giống (A)	$4,77^{B}$	$5,05^{A}$	$5,06^{A}$	

Ghi chú: Phân hạng yếu tố A theo $LSD_{0,05}$, yếu tố B theo $LSD_{0,01}$ (chữ in hoa) và phân hạng các nghiệm thức theo $LSD_{0,05}$ (chữ thường)

3. Hiệu quả của các yếu tố và tương tác

Phương pháp tính hiệu quả của các mức A_i, các mức B_j và tương tác $A_i \times B_j$ như mục 1.

7.3. THÍ NGHIỆM HAI YẾU TỐ KIỂU CHIA LÔ (kiểu lô phụ, Split-Plot Design)

7.3.1. Bố trí thí nghiệm

Trong thí nghiệm đồng ruộng hai yếu tố kiểu RCBD với i mức của yếu tố A ($i=\overline{1,a}$) và j mức của yếu tố B ($j=\overline{1,b}$) được phối hợp thành ab nghiệm thức A_iB_j khác nhau. Mỗi ô nghiệm thức nằm ngẫu nhiên ở các vị trí khác trong một lần lặp lại (khối). Mặc dù đây là kiểu khá phổ biến trong thí nghiệm nghiên cứu hai yếu tố nhưng trong nhiều trường hợp các ô của nghiệm thức này có thể bị ảnh hưởng bởi các yếu tố nghiên cứu từ ô bên cạnh làm cho việc đánh giá thiếu chính xác. Ví dụ, khi muốn nghiên cứu chế độ tưới khác nhau cho các giống, nếu bố trí kiểu RCBD chế độ tưới ô này sẽ ảnh hưởng sang ô khác. Điều đó cũng có thể xẩy ra cho các thí nghiệm phân bón, thí nghiệm phun thuốc trừ sâu, vv.

Khi nghi ngờ về ảnh hưởng qua lại như trường hợp nêu trên (không phải tất cả), người ta bố trí kiểu lô phụ. Trong thí nghiệm hai yếu tố kiểu lô phụ, một trong hai yếu tố được cho là yếu tố phụ được bố trí thành các lô chính, mỗi lô chính là một mức của yếu tố phụ. Trên lô chính được chia ra một số ô khác nhau tương ứng với số mức của yếu tố thứ hai, được coi là yếu tố chính. các ô nghiệm thức của yếu tố chính được đặt ngẫu nhiên trong lô chính. Người ta thường nói ngắn gọn: yếu tố phụ bố trí thành lô chính còn yếu tố chính thì bố trí lô phụ.

Trong nghiên cứu chế độ tưới, phun thuốc bảo vệ thực vật, và một số yếu tố khác, nếu bố trí kiểu RCBD việc khống chế các yếu tố này rất khó khăn nên tốt nhất lấy các yếu tố này làm yếu tố phụ và vai trò này không thể đảo ngược, không phải vì các yếu tố này không quan trọng mà để kiểm soát tốt các yếu tố.

Để đánh giá chính xác mục tiêu nghiên cứu cần lưu ý:

- Việc coi yếu tố chính và phụ không hoàn toàn phụ thuộc vào ý muốn của nhà nghiên cứu mà còn phụ thuộc vào mức độ ảnh hưởng của các yếu tố khi bố trí cạnh nhau.

- Xác định kích thước của lô chính và lô phụ sao cho hạn chế tối đa ảnh hưởng qua lại giữa các yếu tố ở các ô.

- Trong quá trính thí nghiệm, khống chế tốt các yếu tố nghiên cứu.

Sau đây là một số sơ đồ bố trí thí nghiệm.

Lần lặp lại I

a_3	a_1	a_2	a_0	a_4
b_0	b_1	b_0	b_1	b_0
b_1	b_0	b_1	b_0	b_1

Lần lặp lại II

a_1	a_4	a_0	a_2	a_3
b_1	b_0	b_1	b_0	b_0
b_0	b_1	b_0	b_1	b_1

Lần lặp lại III

a_1	a_3	a_0	a_2	a_4
b_1	b_0	b_0	b_0	b_1
b_0	b_1	b_1	b_1	b_0

Hình 7.4: Sơ đồ thí nghiệm kiểu lô phụ 5 × 2
(5 mức A (lô chính), 2 mức B (lô phụ), 3 lần lặp lại)

I

a_1	a_2
b_2	b_1
b_1	b_5
b_4	b_3
b_3	b_2
b_5	b_4

II

a_1	a_2
b_4	b_1
b_3	b_2
b_5	b_1
b_1	b_5
b_2	b_4

III

a_1	a_2
b_3	b_4
b_5	b_2
b_4	b_5
b_1	b_1
b_2	b_3

IV

a_1	a_2
b_1	b_4
b_5	b_2
b_3	b_5
b_2	b_1
b_4	b_3

Hình 7.5: Sơ đồ thí nghiệm kiểu lô phụ 2 × 5
(2 mức A (lô chính), 5 mức B (lô phụ), 3 lần lặp lại)

I

N_4	N_3	N_1	N_0	N_5	N_2
V_2	V_1	V_1	V_2	V_4	V_3
V_1	V_4	V_2	V_3	V_3	V_2
V_3	V_2	V_4	V_1	V_2	V_1
V_4	V_3	V_3	V_4	V_1	V_4

II

N_1	N_0	N_5	N_2	N_4	N_3
V_1	V_4	V_3	V_1	V_1	V_3
V_3	V_1	V_4	V_2	V_4	V_2
V_2	V_2	V_1	V_4	V_2	V_4
V_4	V_3	V_2	V_3	V_3	V_1

III

N_0	N_1	N_4	N_5	N_3	N_2
V_4	V_3	V_3	V_1	V_2	V_1
V_2	V_4	V_2	V_3	V_3	V_4
V_1	V_1	V_4	V_2	V_4	V_2
V_3	V_2	V_1	V_4	V_1	V_3

Hình 7.6: Sơ đồ thí nghiệm kiểu lô phụ 6 × 4
(6 mức đạm, N (lô chính), 4 giống, V (lô phụ), 3 lần lặp lại)

7.3.2. Mô hình toán và phân tích phương sai

7.3.2.1. Mô hình toán học và nguyên lý tính toán

Nếu A là yếu tố phụ và B là yếu tố chính thì giá trị của ô nghiệm thức mức *i* của yếu tố A phối hợp với mức *j* của yếu tố B ở lần lặp lại *k* (X_{ijk}) được tính theo mô hình:

$$X_{ijk} = m + a_i + b_j + ab_{ij} + r_k + e_{ijk}$$
$$e_{ijk} = e(a)_{ik} + e(b)_{ijk}$$

trong đó: m là giá trị trung bình toàn thí nghiệm, a_i là giá trị do mức *i* của yếu tố phụ A tạo ra (đóng góp vào giá trị ô), b_j là giá trị do mức *j* của yếu tố chính B tạo ra, ab_{ij} là giá trị do tương tác giữa A_i với B_j ($A_i \times B_j$) tạo ra, r_k là giá trị tạo nên do ảnh hưởng của lần lặp lại *k*, e_{ijk} là sai số tại ô *ijk* được tạo bởi $e(a)_{ik}$ (sai số của A) và $e(b)_{ijk}$ (sai số của B và A×B).

Điều khác nhau giữa kiểu lô phụ và kiểu RCBD là trong phân tích phương sai, kiểu lô phụ cho phép tách tổng bình phương sai số kiểu RCBD thành hai phần khác nhau, cho phép đánh chính xác hơn tác động của từng yếu tố và tương tác.

Theo mô hình này tổng bình phương tổng số (SS_{TO}) được tạo ra do bốn nguyên nhân theo bảng sau:

Nguồn biến động	DF	SS	MS	F_{TN}
1. Lần lặp lại	$r - 1$	SS_r		
2. Lô chính (A)	$a - 1$	SS_a	SS_a/df_a	
Sai số (*a*)	$(r - 1)(a - 1)$	$SS_{e(a)}$	$SS_{e(a)}/df_{e(a)}$	
3. Lô phụ (B)	$(b - 1)$	SS_b	SS_b/df_b	
4. Tương tác A×B	$(a - 1)(b - 1)$	SS_{ab}	SS_{ab}/df_{ab}	
Sai số (*b*)	$a(r - 1)(b - 1)$	$SS_{e(b)}$	$SS_{e(b)}/df_{e(b)}$	
Tổng số (TO)	$abr - 1$	SS_{TO}		

Việc phân tích kết quả thí nghiệm được minh họa trong ví dụ sau.

7.3.2.2. Ví dụ ứng dụng

Ví dụ: Để nghiên cứu ảnh hưởng của vôi và phân lân đến năng suất một loại cỏ lâu năm làm thức ăn gia súc, người ta đã bố trí một thí nghiệm kiểu lô phụ với 2 yếu tố: Yếu tố phụ là bón vôi, 2 mức: bón vôi và không bón vôi, yếu tố chính là bón lân, 5 liều: 0, 30, 60, 90 và 120 kg P_2O_5/ha. Hãy phân tích hiệu quả của vôi và lân theo kết quả Bảng 7.13, 7.14 sau.

Bảng 7.13: Năng suất cỏ khô (tạ/ha)

Mức lân, B	Năng suất ở các lần lặp lại, x_{ij} I	II	III	IV	Tổng NT (T)	Trung bình
		L_0 (0 bón)				
P_0	22	20	24	26	92	23,0
P_1	26	23	26	29	104	26,0
P_2	29	28	31	31	119	29,8
P_3	31	35	30	31	127	31,8
P_4	31	30	32	30	123	30,8

	L$_1$ (30 kg/ha)					
P_0	25	22	28	24	99	24,8
P_1	28	29	32	28	117	29,3
P_2	29	31	34	36	130	32,5
P_3	34	36	37	32	139	34,8
P_4	36	40	42	36	154	38,5
Trung bình chung					m =	30,1

Bảng 7.14: Tổng năng suất của các mức vôi trên các lần lặp lại (RA)

Vô (L)	Tổng năng suất (RA, kg/ha)				Tổng (A)
	I	II	III	IV	
L_0	139	136	143	147	565
L_1	152	158	173	156	639
Tổng (R)	291	294	316	303	
Tổng cộng				G =	1.204

Tổng năng suất của các mức lân (B)

Mức lân	P_0	P_1	P_2	P_3	P_4
Tổng (B)	191	221	249	266	277

Giải:

1. Kiểm tra sự khác biệt giữa các nghiệm thức

$$CF = \frac{\left(\sum_{i=1}^{ab} x_{ij}\right)^2}{abr} = \frac{(1.204)^2}{(2)(5)(4)} = 36.240,40$$

Tổng bình phương tổng số:

$$SS_{TO} = \sum_{i=1}^{ab} x_{ij}^2 - CF = (22^2 + (20)^2 + \dots + (42)^2 +$$

$$(36)^2 - 36.240,4 = 953,60$$

Tổng bình phương lần lặp lại:

$$SS_r = \frac{\sum_{i=1}^{r} R_i^2}{ab} - CF$$

$$= \frac{(291)^2 + (294)^2 + (316)^2 + (303)^2}{10}$$

$$- 36.240,400 = 37,80$$

Tổng bình phương nghiệm thức:

$$SS_t = \frac{\sum_{i=1}^{ab} T_i^2}{r} - CF$$

$$= \frac{(92)^2 + (104)^2 + ... + (139)^2 + (154)^2}{4}$$

$$- 36.240,40 = 791,10$$

Tổng bình phương sai số:

$$SS_e = SS_{TO} - SS_t - SS_r = 953,60 - 791,10 - 37,80$$

$$= 124,70$$

Bảng 7.15: Kết quả phân tích phương sai bước đầu

Nguồn biến động	DF	SS	MS	F_{TN}	Prob.
Lần lặp lại	3	37,80	12,60	2,73	
Nghiệm thức	9	791,10	87,90	19,03	< 0,001
Sai số	27	124,70	4,62		
Tổng số	39	953,60			

Kết quả $F_{TN(t)} > F_{0,001}$ ($P < 0,001$) cho phép tiếp tục tiến hành bước 2.

2. Tính tổng bình phương các nguồn biến động A, B, AB và tổng bình phương các loại sai số.

Tổng bình phương yếu tố A (bón vôi):

$$SS_A = \frac{\sum_{i=1}^{a} A_i^2}{rb} - CF = \frac{(565)^2 + (639)^2}{20} -$$

$$36.240,40 = 136,90$$

Tổng bình phương yếu tố B (bón lân):

$$SS_B = \frac{\sum_{j=1}^{b} B_j^2}{ra} - CF$$

$$= \frac{(191)^2 + (221)^2 + ... + (277)^2}{8}$$

$$- 36.240,40 = 610,60$$

Tổng bình phương tương tác A×B:

$$SS_{A\times B} = SS_t - SS_A - SS_B$$

$$= 791,10 - 136,90 - 610,60 = 43,60$$

Tổng bình phương sai số (a):

$$SS_{e(a)} = \frac{\sum_{i=1}^{a} (RA)^2}{b} - CF - SS_r - SS_A$$

$$= \frac{(139)^2 + (136)^2 + ... + (173)^2 + (156)^2}{5}$$

$- 36.240,40 - 37,80 - 136,90 = 26,50$

Tổng bình phương sai số (*b*):

$$SS_{e(b)} = SS_e - SS_{e(a)} = 124,70 - 26,50 = 98,20$$

3. Tính phương sai các nguồn biến động A, B, A×B và phương sai các loại sai số.

Phương sai yếu tố A:

$$MS_A = \frac{SS_A}{a-1} = \frac{136,90}{1} = 136,90$$

Phương sai yếu tố B:

$$MS_b = \frac{SS_b}{b-1} = \frac{610,60}{4} = 152,65$$

Phương sai tương tác A×B:

$$MS_{A\times B} = \frac{SS_{A\times B}}{(a-1)(b-1)} = \frac{43,60}{8} = 10,90$$

Phương sai sai số *(a)*:

$$MS_{e(a)} = \frac{SS_{e(a)}}{(r-1)(a-1)} = \frac{26,50}{3} = 8,83$$

Phương sai sai số *(b)*:

$$MS_{e(b)} = \frac{SS_{e(b)}}{a(r-1)(b-1)} = \frac{98,20}{(2)(3)(4)} = 4,09$$

4. Đánh giá sự sai khác của các nguồn biến động

$$F_{TN(A)} = \frac{MS_A}{MS_{e(a)}} = \frac{136,90}{8,83} = 15,50$$

$$F_{TN(B)} = \frac{MS_B}{MS_{e(b)}} = \frac{152,65}{4,09} = 37,31$$

$$F_{TN(A\times B)} = \frac{MS_{A\times B}}{MS_{e(b)}} = \frac{10,90}{4,09} = 2,66$$

Bảng 7.16: Kết quả phân tích phương sai

Nguồn	DF	SS	MS	F_{TN}	Prob.
Bón vôi (A)	1	136,90	136,90	15,50	0,0292
Sai số (*a*)	3	26,50	8,83		
Bón lân (B)	4	610,60	152,65	37,31	0,0000
Tương tác A×B	4	43,60	10,90	2,66	0,0571
Sai số (*b*)	24	98,20	4,09		
Tổng số	39	953,60			

Như vậy:

- Có sự khác nhau giữa bón vôi và không bón, độ tin cậy 95% ($P < 0,05$)

- Có sự khác nhau rất chắc chắn giữa các mức lân với năng suất, độ tin cậy 99,9% ($P < 0,001$)

- Sự phối hợp giữa mức bón vôi và mức bón lân khác nhau không đủ tin cậy 95% ($P > 0,05$).

5. So sánh từng yếu tố và tương tác

- Sai khác giữa hai mức của yếu tố A:

$$Sd_A = \sqrt{\frac{2MS_{e(a)}}{rb}} = \sqrt{\frac{(2)(8,83)}{(5)(4)}} = 0,94$$

$$LSD_{0,05} = t_{0,05}Sd_A = (3,182)(0,94) = 2,99$$

- Sai khác giữa hai mức của yếu tố B:

$$Sd_B = \sqrt{\frac{2MS_{e(b)}}{ra}} = \sqrt{\frac{(2)(4,09)}{(4)(2)}} = 1,01$$

$$LSD_{0,05} = t_{0,05}Sd_B = (2,064)(1,01) = 2,08$$

$$LSD_{0,01} = t_{0,01}Sd_B = (2,797)(1,01) = 2,83$$

- Sai khác giữa hai mức B trên cùng một mức A:

$$Sd_{B/A} = \sqrt{\frac{2MS_{e(b)}}{r}} = \sqrt{\frac{(2)(4,09)}{4}} = 1,43$$

$$LSD_{0,05} = t_{0,05}Sd_{B/A} = (3,182)(1,43) = 4,55$$

$$LSD_{0,01} = t_{0,01}Sd_{B/A} = (5,841)(1,43) = 8,35$$

- Sai khác giữa hai mức A trên cùng một mức B hay khác mức B:

$$Sd_{A/B} = \sqrt{\frac{2\left[(b-1)MS_{e(b)} + MS_{e(a)}\right]}{rb}}$$

$$= \sqrt{\frac{2\left[(5-1)(4,09)+8,83\right]}{(4)(5)}} = 1,59$$

$$LSD_{0,05} = t_{0,05}Sd_{A/B} = (2,093)(1,59) = 3,32$$

- Hệ số biến động:

$$CV(a) = \frac{\sqrt{MS_{e(a)}}}{m} \times 100 = \frac{\sqrt{8,83}}{30,1} \times 100 = 9,9\%$$

$$CV(b) = \frac{\sqrt{MS_{e(b)}}}{m} \times 100 = \frac{\sqrt{4,09}}{30,1} \times 100 = 6,7\%$$

Bảng 7.17: Kết quả phân hạng

Yếu tố B	Tổng năng suất (AB)		Tổng P
	$A_1^{(2)}$	$A_2^{(2)}$	$(B)^{(1)}$
P_0	$23{,}00^{c}$	$24{,}75^{d}$	$23{,}88^{D}$
P_1	$26{,}00^{bc}$	$29{,}25^{cd}$	$27{,}63^{C}$
P_2	$29{,}75^{ab}$	$32{,}50^{bc}$	$31{,}13^{B}$
P_3	$31{,}75^{a}$	$34{,}75^{ab}$	$33{,}25^{A}$
P_4	$30{,}75^{a}$	$38{,}5^{a}$	$34{,}63^{A}$
Tổng L $(A)^{(1)}$	$28{,}25^{B}$	$31{,}95^{A}$	

(1) *Phân hạng các mức yếu tố A, B với $\alpha = 0{,}05$*

(2) *Phân hạng các mức B trên cùng mức A (cùng A_1 và cùng A_2) với $\alpha = 0{,}05$*

Cách tính LSD_α giữa các nghiệm thức và phân hạng được thực hiện như đã biết. Việc tính hiệu quả của các mức của yếu tố A, B và tương tác $A_i \times B_j$ thực hiện như ở mục 7.1.2.2.

7.4. THÍ NGHIỆM HAI YẾU TỐ KIỂU LÔ NGANG DỌC (kiểu lô sọc, Strip-Plot Design)

7.4.1. Bố trí thí nghiệm

Kiểu lô sọc là kiểu bố trí đặc biệt của thí nghiệm hai yếu tố, có độ chính xác hơn kiểu lô phụ. Trong kiểu lô phụ, thứ tự các ô phụ (yếu tố chính) trong các lô chính (yếu tố phụ) thay đổi nên chỉ loại trừ được sự khác biệt giữa các lô chính (lô dọc). Trong kiểu lô sọc do thứ tự các ô trong lô dọc và lô ngang không thay đổi nên đã loại trừ được sự sai khác bởi:

- Sai khác giữa các lô nằm dọc của yếu tố dọc (vertical factor);

- Sai khác giữa các lô nằm ngang của yếu tố ngang (horizontal factor);

- Sai khác giữa các ô giao nhau giữa lô dọc và lô ngang biểu thị sự tương tác giữa hai yếu tố.

Sau đây là một số sơ đồ bố trí thí nghiệm.

I

	A_3	A_1	A_2	A_0	A_4
B_2					
B_0					
B_1					

II

	A_3	A_1	A_2	A_0	A_4
B_2					
B_0					
B_1					

III

	A_3	A_1	A_2	A_0	A_4
B_2					
B_0					
B_1					

Hình 7.7: Sơ đồ kiểu lô sọc 5×3 (5 mức A - lô dọc; 2 mức B - lô ngang; 3 lần lặp lại)

I

	A_1	A_3	A_2
B_6			
B_5			
B_3			
B_2			
B_4			
B_1			

II

	A_3	A_2	A_1
B_4			
B_2			
B_6			
B_3			
B_1			
B_5			

III

	A_3	A_1	A_2
B_5			
B_2			
B_3			
B_4			
B_6			
B_1			

Hình 7.8: Sơ đồ kiểu lô sọc 3×6

(3 mức A - lô dọc; 6 mức B - lô ngang; 3 lần lặp lại)

7.4.2. Mô hình toán và phân tích phương sai

7.4.2.1. Mô hình toán học và nguyên lý tính toán

Nếu A là yếu tố dọc và B là yếu tố ngang thì giá trị của ô nghiệm thức mức *i* của yếu tố A phối hợp với mức *j* của yếu tố B ở lần lặp lại *k* (X_{ijk}) được tính theo mô hình:

$$X_{ijk} = m + a_i + b_j + ab_{ij} + r_k + e_{ijk}$$

$$e_{ijk} = e(a)_{ik} + e(b)_{jk} + e(c)_{ijk}$$

trong đó: m là giá trị trung bình toàn thí nghiệm, a_i là giá trị do mức *i* của yếu tố phụ A tạo ra, b_j là giá trị do mức *j* của yếu tố chính B tạo ra, ab_{ij} là giá trị do tương tác A_i với B_j ($A_i \times B_j$) tạo ra, r_k là giá trị tạo nên do ảnh hưởng của lần lặp lại k, e_{ijk} là sai số tại ô *ijk* được tạo bởi $e(a)_{ik}$ (sai số của A), $e(b)_{jk}$ (sai số của B) và $e(c)_{ijk}$ (sai số của A×B).

Trong phân tích phương sai, kiểu lô sọc cho phép tách tổng bình phương sai số thành ba phần khác nhau, cho phép đánh chính xác hơn tác động của từng yếu tố và tương tác.

Theo mô hình này tổng bình phương tổng số (SS_{TO}) được tạo ra do bốn nguyên nhân:

- Sự sai khác giữa các lần lặp lại;
- Sự khác nhau giữa các lô dọc (yếu tố A);
- Sai số (a);
- Sự khác nhau giữa các lô ngang (yếu tố B);
- Sai số (b);
- Sự khác nhau về tương tác giữa mức yếu tố lô dọc và mức yếu tố lô ngang (tương tác A×B);

- Sai số (c).

Bảng phân tích phương sai có dạng

Nguồn	DF	SS	MS	F_{TN}
Lần lặp lại	$r - 1$	SS_r		
Lô dọc (A)	$a - 1$	SS_a	SS_a/df_a	
Sai số (a)	$(r - 1)(a - 1)$	$SS_{e(a)}$	$SS_{e(a)}/df_{e(a)}$	
Lô ngang (B)	$(b - 1)$	SS_b	SS_b/df_b	
Sai số (b)	$(r - 1)(b - 1)$	$SS_{e(b)}$	$SS_{e(b)}/df_{e(b)}$	
Tương tác A×B	$(a - 1)(b - 1)$	SS_{ab}	SS_{ab}/df_{ab}	
Sai số (c)	$(r-1)(a-1)(b-1)$	$SS_{e(c)}$	$SS_{e(c)}/df_{e(c)}$	
Tổng số (TO)	$rab - 1$	SS_T		

Việc phân tích kết quả thí nghiệm được minh họa trong ví dụ sau.

7.4.2.2. Ví dụ ứng dụng

Ví dụ: Để nghiên cứu ảnh hưởng của ba mức bón đạm đến năng suất sáu giống lúa, người ta đã bố trí một thí nghiệm kiểu lô sọc: yếu tố lô dọc là mức bón đạm và yếu tố lô ngang là giống lúa, ba lần lặp lại. Kết quả được ghi nhận trong Bảng 7.18 và các bảng tổng hợp 7.19, 7.20 sau đây. Hãy phân tích hiệu quả của hai yếu tố và tương tác giữa chúng.

Bảng 7.18: Năng suất sáu giống lúa với ba mức đạm

Liều đạm (B)	Năng suất (tấn/ha) I	II	III	Tổng NT (T)
		Giống 1 (A_1)		
B_1 (0 kg N/ha)	2,37	3,96	4,38	10,72
B_2 (60 kg N/ha)	4,08	6,43	4,89	15,40
B_3 (120 kg N/ha)	7,25	6,81	8,58	22,64

		Giống 2 (A_2)		
B_1 (0 kg N/ha)	4,01	5,80	5,00	14,80
B_2 (60 kg N/ha)	5,63	7,33	7,18	20,14
B_3 (120 kg N/ha)	7,05	8,28	6,30	21,63
		Giống 3 (A_3)		
B_1 (0 kg N/ha)	2,62	4,51	5,62	12,75
B_2 (60 kg N/ha)	4,68	6,67	7,02	18,37
B_3 (120 kg N/ha)	7,67	7,33	8,61	23,61
		Giống 4 (A_4)		
B_1 (0 kg N/ha)	2,73	5,63	3,82	12,18
B_2 (60 kg N/ha)	4,84	7,01	4,82	16,66
B_3 (120 kg N/ha)	6,88	7,74	6,67	21,28
		Giống 5 (A_5)		
B_1 (0 kg N/ha)	4,45	3,73	4,58	12,76
B_2 (60 kg N/ha)	5,55	5,34	6,01	16,90
B_3 (120 kg N/ha)	6,88	5,08	6,08	18,04
		Giống 6 (A_6)		
B_1 (0 kg N/ha)	2,57	3,72	3,33	9,62
B_2 (60 kg N/ha)	3,90	2,82	4,43	11,14
B_3 (120 kg N/ha)	1,56	2,71	3,12	7,39
Tổng cộng	84,70	100,89	100,43	**286,02**
			m =	5,30

Bảng 7.19: Tổng năng suất các giống ở các lần lặp lại (RA)

Giống	Năng suất (tấn/ha)			Tổng A
	I	II	III	
A_1	13,70	17,20	17,86	48,76
A_2	16,69	21,41	18,48	56,58
A_3	14,96	18,51	21,25	54,72
A_4	14,45	20,37	15,30	50,12
A_5	16,88	14,15	16,67	47,69
A_6	8,02	9,25	10,88	28,15

Bảng 7.20: Tổng năng suất các mức đạm ở các lần lặp lại (RB)

Giống	Năng suất (tấn/ha)			Tổng B
	I	II	III	
B_1	18,75	27,34	26,74	72,82
B_2	28,67	35,61	34,34	98,61
B_3	37,29	37,94	39,36	114,59

Giải:

1. Kiểm tra sự sai khác giữa các nghiệm thức

$$CF = \left(\sum_{i=1}^{ab} X_{ij}\right)^2 / abr = \frac{(286,02)^2}{(3)(6)(3)} = 1.514,921$$

Tổng bình phương tổng số:

$$SS_{TO} = \sum_{i=1}^{ab} X_{ij}^2 - CF = (2,37)^2 + (3,96)^2 + \ldots$$

$$+ (3,12)^2 - 1.514,921 = 165,775$$

Tổng bình phương lần lặp lại:

$$SS_r = \frac{\sum_{i=1}^{r} R_i^2}{ab} - CF$$

$$= \frac{(84,70)^2 + (100,89)^2 + (100,43)^2}{18}$$

$$- 1.514,921 = 9,438$$

Tổng bình phương nghiệm thức:

$$SS_t = \frac{\sum_{i=1}^{ab} T_i^2}{r} - CF$$

$$= \frac{(10,72)^2 + (15,40)^2 + ... + (7,39)^2}{3}$$

$$- 1.514,921 = 131,156$$

Tổng bình phương sai số:

$$SS_e = SS_T - SS_t - SS_r = 165,775 - 131,156$$

$$- 9,438 = 25,181$$

Bảng 7.21: Kết quả phân tích phương sai bước đầu

Nguồn	DF	SS	MS	F_{TN}	Prob.
Lần lặp lại	2	9,438	4,719	6,37	0,0045
Nghiệm thức	17	131,156	7,715	10,42	0,0000
Sai số	34	25,181	0,741		
Tổng số	53	165,775			

Kết quả $F_{TN(t)} > F_{0,01}$ ($P < 0,01$) cho phép tiếp tục tiến hành bước 2.

2. Tính tổng bình phương các nguồn biến động A, B, AB và tổng bình phương các loại sai số.

Tổng bình phương yếu tố A (giống):

$$SS_A = \frac{\sum_{i=1}^{a} A_i^2}{rb} - CF$$

$$= \frac{(48,76)^2 + (56,58)^2 + ... + (28,15)^2}{9}$$

$$- 1.514,921 = 57,472$$

Tổng bình phương yếu tố B (đạm):

$$SS_B = \frac{\sum_{j=1}^{b} B_j^2}{ra} - CF$$

$$= \frac{(72,82)^2 + (98,61)^2 + (114,59)^2}{18}$$

$$- 1.514,921 = 49,348$$

Tổng bình phương tương tác A×B:

$$SS_{A\times B} = SS_t - SS_A - SS_B$$

$$= 131,156 - 57,472 - 49,348 = 24,336$$

Tổng bình phương sai số (a):

$$SS_{e(a)} = \frac{\sum_{i=1}^{a} (RA)^2}{b} - CF - SS_r - SS_A$$

$$= \frac{(13,70)^2 + (17,20)^2 + ... + (10,88)^2}{3}$$

$$- 1.514,921 - 9,438 - 57,472 = 14,044$$

Tổng bình phương sai số (b):

$$SS_{e(b)} = \frac{\sum_{i=1}^{a} (RB)^2}{a} - CF - SS_r - SS_B$$

$$= \frac{(18,75)^2 + (27,34)^2 + ... + (39,36)^2}{6}$$

$$- 1.514,921 - 9,438 - 49,348 = 3,160$$

Tổng bình phương sai số (c):

$$SS_{e(c)} = SS_e - SS_{e(a)} - SS_{e(b)}$$

$$= 25,181 - 14,044 - 3,160 = 7,977$$

3. Tính phương sai các nguồn biến động A, B, AB và phương sai các loại sai số.

Phương sai yếu tố A:

$$MS_A = \frac{SS_A}{a-1} = \frac{57,472}{5} = 11,494$$

Phương sai yếu tố B:

$$MS_b = \frac{SS_b}{b-1} = \frac{49,348}{3} = 24,674$$

Phương sai tương tác A×B:

$$MS_{A\times B} = \frac{SS_{A\times B}}{(a-1)(b-1)} = \frac{24,336}{10} = 2,434$$

Phương sai sai số a:

$$MS_{e(a)} = \frac{SS_{e(a)}}{(r-1)(a-1)} = \frac{14,044}{10} = 1,404$$

Phương sai sai số b:

$$MS_{e(b)} = \frac{SS_{e(b)}}{(r-1)(b-1)} = \frac{3,160}{(2)(2)} = 0,790$$

4. Đánh giá sự sai khác của các nguồn biến động

$$F_{TN(A)} = \frac{MS_A}{MS_{e(a)}} = \frac{11,494}{1,404} = 8,18$$

$$F_{TN(B)} = \frac{MS_B}{MS_{e(b)}} = \frac{24,674}{0,790} = 31,24$$

$$F_{TN(A\times B)} = \frac{MS_{A\times B}}{MS_{e(b)}} = \frac{2,434}{0,399} = 6,10$$

Bảng 7.22: Kết quả phân tích phương sai

Nguồn	DF	SS	MS	F_{TN}	Prob.
Giống (A)	5	57,472	11,494	8,18	0,0026
Sai số (*a*)	10	14,044	1,404		
Đạm (B)	2	49,348	24,674	31,24	0,0036
Sai số (*b*)	4	3,160	0,790		
Tương tác A×B	10	24,336	2,434	6,10	0,0003
Sai số (*c*)	20	7,977	0,399		
Tổng số	53	165,775			

Như vậy:

- Có sự khác nhau rất đáng tin cậy giữa các giống, giữa các mức bón đạm ($P < 0,01$)

- Hiệu quả sử dụng đạm giữa các giống rất khác nhau, vì vậy sự khác nhau giữa các mức tương tác A×B rất đáng tin cậy ($P < 0,001$).

5. So sánh từng yếu tố và tương tác

- Sai khác giữa hai mức của yếu tố A:

$$Sd_A = \sqrt{\frac{2MS_{e(a)}}{rb}} = \sqrt{\frac{(2)(1,404)}{(3)(3)}} = 0,56$$

$$LSD_{0,05} = t_{0,05}Sd_A = (2,228)(0,56) = 1,25$$

$$LSD_{0,01} = t_{0,01}Sd_A = (3,169)(0,56) = 1,77$$

- Sai khác giữa hai mức của yếu tố B:

$$Sd_B = \sqrt{\frac{2MS_{e(b)}}{ra}} = \sqrt{\frac{(2)(0,790)}{(3)(6)}} = 0,30$$

$$LSD_{0,05} = t_{0,05}Sd_B = (2,776)(0,30) = 0,83$$

$$LSD_{0,01} = t_{0,01}Sd_B = (4,604)(0,30) = 1,36$$

- Sai khác giữa hai mức B trên cùng một mức A:

$$Sd_{B/A} = \sqrt{\frac{2\left[(b-1)MS_{e(c)} + MS_{e(a)}\right]}{rb}}$$

$$= \sqrt{\frac{2\left[(3-1)(0,399)+(1,404)\right]}{(3)(3)}} = 0,70$$

$$LSD_{á} = t'_{\alpha}Sd_{B/A}$$

$$t'_{\alpha} = \frac{(b-1)t_{c(\alpha)}MS_{e(c)} + t_{a(\alpha)}MS_{e(a)}}{(b-1)MS_{e(c)} + MS_{e(a)}}$$

$$t'_{0,05} = \frac{(3-1)(2,086)(0,399)+(2,228)(1,404)}{(3-1)(0,399)+(1,404)}$$

$$= 2,176$$

$$t'_{0,01} = \frac{(3-1)(2,845)(0,399)+(3,169)(1,404)}{(3-1)(0,399)+(1,404)}$$

$$= 3,051$$

Như vậy: $LSD_{0,05} = (2,176)(0,70) = 1,52$

$$LSD_{0,01} = (3,051)(0,70) = 2,13$$

- Sai khác giữa hai mức A trên cùng một mức B:

$$\mathrm{Sd}_{\mathrm{A/B}} = \sqrt{\frac{2\left[(\mathrm{a}-1)\mathrm{MS}_{\mathrm{e}(c)} + \mathrm{MS}_{\mathrm{e}(b)}\right]}{ra}}$$

$$= \sqrt{\frac{2\left[(3-1)(0,399)+(0,790)\right]}{(3)(6)}} = 0,42$$

$$\mathrm{LSD}_{á} = t'_{\alpha}\,\mathrm{Sd}_{\mathrm{B/A}}$$

$$t'_{\alpha} = \frac{(\mathrm{b}-1)\,t_{c(\alpha)}\mathrm{MS}_{\mathrm{e(c)}} + t_{a(\alpha)}\mathrm{MS}_{\mathrm{e(b)}}}{(\mathrm{b}-1)\mathrm{MS}_{\mathrm{e(c)}} + \mathrm{MS}_{\mathrm{e(b)}}}$$

Cách tính $\mathrm{LSD}_{0,05}$ và $\mathrm{LSD}_{0,01}$ tương tự như trên.

- Sai khác giữa hai mức tương tác A×B bất kỳ:

$$\mathrm{Sd}_{\mathrm{A\times B}} = \sqrt{\frac{2\mathrm{MS}_{\mathrm{e}(c)}}{r}} = \sqrt{\frac{(2)(0,399)}{3}} = 0,52$$

Cách tính $\mathrm{LSD}_{0,05}$ và $\mathrm{LSD}_{0,01}$ như đã biết.

Cách tính LSD_{α} giữa các nghiệm thức và phân hạng như đã biết.

- Hệ số biến động:

$$\mathrm{CV}(a) = \frac{\sqrt{\mathrm{MS}_{\mathrm{e}(a)}}}{m}\times 100 = \frac{\sqrt{1,404}}{5,297}\times 100 = 22,3\%$$

$$\mathrm{CV}(b) = \frac{\sqrt{\mathrm{MS}_{\mathrm{e}(b)}}}{m}\times 100 = \frac{\sqrt{0,790}}{5,297}\times 100 = 16,8\%$$

$$\mathrm{CV}(c) = \frac{\sqrt{\mathrm{MS}_{\mathrm{e}(c)}}}{m}\times 100 = \frac{\sqrt{0,339}}{5,297}\times 100 = 11,0\%$$

6. Hiệu quả các mức của từng yếu tố và tương tác $A_i \times B_j$

Trung bình các nghiệm thức và hiệu quả các mức A (a_i) và các mức B (b_j)

	B_1	B_2	B_3	TB A	a_i
A_1	3,57	5,13	7,55	5,42	0,12
A_2	4,93	6,71	7,21	6,29	0,99
A_3	4,25	6,12	7,87	6,08	0,78
A_4	4,06	5,55	7,09	5,57	0,27
A_5	4,25	5,63	6,01	5,30	0,00
A_6	3,21	3,71	2,46	3,13	-2,17
TB B	4,05	5,48	6,37	5,30	
b_j	-1,25	0,18	1,07		

Giả thiết rằng tất cả các tương tác a_ib_j đều bằng 0, giá trị ($X_{ij} - m$) lý thuyết của các nghiệm thức được tính như sau:

$$
\begin{array}{c}
\text{NT} \\
\begin{pmatrix}
A_1B_1 \\ A_1B_2 \\ A_1B_3 \\ A_2B_1 \\ A_2B_2 \\ A_2B_3 \\ A_3B_1 \\ A_3B_2 \\ A_3B_3 \\ A_4B_1 \\ A_4B_2 \\ A_4B_3 \\ A_5B_1 \\ A_5B_2 \\ A_5B_3 \\ A_6B_1 \\ A_6B_2 \\ A_6B_3
\end{pmatrix}
\end{array}
=
\begin{array}{c}
\begin{array}{ccccccccc} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & b_1 & b_2 & b_3 \end{array} \\
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1
\end{pmatrix}
\end{array}
=
\begin{array}{c}
X_{ij} - m \\
\begin{pmatrix}
-1{,}14 \\ 0{,}30 \\ 1{,}18 \\ -0{,}27 \\ 1{,}16 \\ 2{,}05 \\ -0{,}47 \\ 0{,}96 \\ 1{,}85 \\ -0{,}99 \\ 0{,}45 \\ 1{,}34 \\ -1{,}26 \\ 0{,}18 \\ 1{,}07 \\ -3{,}43 \\ -1{,}99 \\ -1{,}11
\end{pmatrix}
\end{array}
$$

Hiệu quả tương tác a_ib_j = $(X_{ij} - m)$ thực – $(X_{ij} - m)$ lý thuyết

Nghiệm thức (A_iB_j)	Giá trị thực tế (X_{ij})	$X_{ij} - m$ Thực tế	$X_{ij} - m$ Lý thuyết	Hiệu quả tương tác (ab_{ij})
A_1B_1	3,57	-1,72	-1,14	-0,59
A_1B_2	5,13	-0,16	0,30	-0,46
A_1B_3	7,55	2,25	1,18	1,07*
A_2B_1	4,93	-0,36	-0,27	-0,09
A_2B_2	6,71	1,42	1,16	0,25
A_2B_3	7,21	1,91	2,05	-0,14
A_3B_1	4,25	-1,05	-0,47	-0,57
A_3B_2	6,12	0,83	0,96	-0,13
A_3B_3	7,87	2,57	1,85	0,73
A_4B_1	4,06	-1,24	-0,99	-0,25
A_4B_2	5,55	0,26	0,45	-0,19
A_4B_3	7,09	1,80	1,34	0,46
A_5B_1	4,25	-1,04	-1,26	0,21
A_5B_2	5,63	0,34	0,18	0,16
A_5B_3	6,01	0,72	1,07	-0,35
A_6B_1	3,21	-2,09	-3,43	1,34*
A_6B_2	3,71	-1,58	-1,99	0,41
A_6B_3	2,46	-2,83	-1,11	-1,73*

$$LSD_{0,05} = t_{0,05}\sqrt{\frac{MS_e}{r}} = (2,086)\sqrt{\frac{0,399}{3}} = 0,76$$

Như vậy, trong số 18 mối tương tác, chỉ có ba mối tương tác $A_1 \times B_3$, $A_6 \times B_1$ và $A_6 \times B_3$ là có ý nghĩa ở mức α = 0,05.

7.5. THÍ NGHIỆM HAI YẾU TỐ BỐ TRÍ Ở NHIỀU NƠI HOẶC NHIỀU NĂM

Các thí nghiệm hai yếu tố được bố trí giống nhau cho mọi điểm hoặc được tiến hành lặp lại trong một số năm.

Về nguyên lý tính, thí nghiệm một yếu tố, hai hay nhiều yếu tố bố trí ở nhiều điểm hoặc nhiều năm có cách tính toán tương tự nhau:

- Phân tích phương sai từng điểm cho mỗi kiểu thí nghiệm;

- Phân tích phương sai toàn thí nghiệm.

Để minh họa cách tính toán cho thí nghiệm hai yếu tố ở nhiều điểm hoặc nhiều năm, hãy xét ví dụ sau.

Một thí nghiệm kiểu lô phụ có sự tham gia của hai giống lúa (V_1 và V_2) – yếu tố B, mỗi giống được bón sáu mức đạm (N_1, N_2, ... N_6) – yếu tố A được trồng ở ba nơi (L_1, L_2 và L_3), mỗi nơi bố trí ba lần lặp lại. Kết quả năng suất được ghi ở bảng 7.23 dưới đây.

Bảng 7.23: Năng suất lúa thí nghiệm (tấn/ha)

Điểm	Đạm	Năng suất (tấn/ha)						Tổng	
		I		II		III			
		V_1	V_2	V_1	V_2	V_1	V_2	V_1	V_2
L_1	N_1	1,98	5,30	1,51	1,88	3,66	3,57	7,15	10,75
	N_2	4,57	5,66	4,34	5,10	4,13	5,39	13,04	16,15
	N_3	5,63	6,34	6,78	6,62	4,93	6,33	17,34	19,29
	N_4	7,15	8,11	6,50	8,58	6,33	7,64	19,98	24,33
	N_5	7,22	7,53	7,11	7,10	6,05	6,67	20,38	21,30
	N_6	7,24	7,85	6,83	7,11	5,87	7,44	19,94	22,40
L_2	N_1	3,62	3,45	3,58	3,56	3,94	3,52	11,14	10,53
	N_2	6,07	5,91	5,46	5,97	5,44	6,03	16,97	17,91
	N_3	6,09	5,32	6,57	5,88	6,08	6,48	18,74	17,68
	N_4	5,92	6,51	6,98	6,56	7,15	7,85	20,05	20,92
	N_5	7,19	8,15	6,11	7,21	7,97	6,69	21,27	22,05
	N_6	5,81	7,29	6,89	6,56	7,11	7,4	19,81	21,25

L_3	N_1	4,32	4,89	4,07	2,58	3,86	4,54	12,25	12,01
	N_2	5,86	6,01	4,63	6,63	4,91	5,67	15,40	18,31
	N_3	5,14	6,71	5,84	6,69	4,90	6,8	15,88	20,20
	N_4	6,34	6,46	5,46	6,68	5,66	6,64	17,46	19,78
	N_5	5,57	5,68	5,85	6,87	5,53	5,69	16,95	18,24
	N_6	6,77	6,34	5,26	6,06	3,91	5,95	15,94	18,35

Bước 1: Phân tích phương sai và đánh giá từng điểm

Việc phân tích phương sai từng điểm được thực hiện như kiểu thí nghiệm lô phụ. Kết quả được trình bày ở Bảng 7.24.

Bước 2: Phân tích phương sai toàn thí nghiệm

Với kiểu bố trí này các nguồn biến dị được phân tích theo Bảng 7.25 sau.

Bảng 7.25: Bảng phân tích phương sai

Nguồn biến động	DF@	SS	MS	F_{TN}
Điểm (L)	$m - 1 = 2$			
Lần lặp các điểm	$m(r - 1) = 6$			
Yếu tố lô chính (A)	$a - 1 = 5$			
L × A	$(m - 1)(a - 1) = 10$			
Sai số chung (*a*)	$m(r - 1)(a - 1) = 30$			
Yếu tố lô phụ (B)	$b - 1 = 1$			
L × B	$(m - 1)(b - 1) = 2$			
A × B	$(a - 1)(b - 1) = 5$			
L × A × B	$(m-1)(a-1)(b-1) = 10$			
Sai số chung (*b*)	$ma(r - 1)(b - 1) = 36$			
Tổng số	$mrab - 1 = 107$			

@: a là số nghiệm thức lô chính, b là nghiệm thức lô phụ, r là số lần lặp lại, và m là số điểm (nơi) thí nghiệm.

Bảng 7.24: Kết quả phân tích phương sai thí nghiệm tại ba điểm

Nguồn biến động	DF	Điểm 1 (L_1)			Điểm 2 (L_2)			Điểm 3 (L_3)		
		SS	MS	F_{TN}	SS	MS	F_{TN}	SS	MS	F_{TN}
Lần lặp lại	2	1,986	0,993		1,042	0,521		1,527	0,763	
Nghiệm thức	11	94,259	8,569		52,598	4,782		24,930	2,266	
Liều đạm (N)	5	85,549	17,110	$25,24^{**}$	51,628	10,326	$37,48^{**}$	18,256	3,651	$8,49^{**}$
Sai số *a*	10	6,780	0,678		2,755	0,276		4,301	0,430	
Giống (V)	1	7,462	7,462	$17,99^{**}$	0,155	0,155	0,51	4,702	4,702	$11,77^{**}$
N × V	5	1,247	0,249	0,602	0,815	0,163	0,54	1,972	0,394	0,998
Sai số *b*	12	4,976	0,415		3,648	0,304		4,793	0,399	
CV (*a*), %			14,0			8,7			11,8	
CV (*b*), %			10,9			9,1			11,3	

***: Ý nghĩa ở mức α = 0,01*

1. Tính tổng bình phương lần lặp lại các điểm, tổng bình phương sai số chung (a) và tổng bình phương sai số chung (b):

$$SS_{r/L} = \Sigma SS_r = 1{,}986 + 1{,}042 + 1{,}527 = 4{,}555$$

$$SS_{e(a)} = \sum_{i=1}^{L} SS_{e(a)} = 6{,}780 + 2{,}755 + 4{,}301$$
$$= 13{,}836$$

$$SS_{e(b)} = \sum_{i=1}^{L} SS_{e(b)} = 4{,}976 + 3{,}648 + 4{,}793$$
$$= 13{,}417$$

2. Tính các tổng bình SS trong bảng 7.25

Trước khi tính toán, lập bảng 7.26 , 7.27 và 7.28 sau đây.

Bảng 7.26: Tổng năng suất L × N toàn thí nghiệm

Mức N	Tổng năng suất			Tổng N (A)
	L_1	L_2	L_3	
N_1	17,90	21,67	24,26	63,83
N_2	29,19	34,88	33,71	97,78
N_3	36,63	36,42	36,08	109,13
N_4	44,31	40,97	37,24	122,52
N_5	41,68	43,32	35,19	120,19
N_6	42,34	41,06	34,29	117,69
Tổng G	212,05	218,32	200,77	631,14

Bảng 7.27: Tổng năng suất L × V toàn thí nghiệm

Giống (V)	Tổng năng suất			Tổng V (B)
	L1	L2	L3	
V1	97,83	107,98	93,88	299,69
V2	114,22	110,34	106,89	331,45

Bảng 7.28: Tổng năng suất N × V toàn thí nghiệm

Mức N (A)	Tổng năng suất	
	V_1	V_2
N_1	30,54	33,29
N_2	45,41	52,37
N_3	51,96	57,17
N_4	57,49	65,03
N_5	58,60	61,59
N_6	55,69	62,00

$$CF = \frac{\left(\sum_{i=1}^{m} G_i\right)^2}{mabr} = \frac{(631{,}14)^2}{(3)(6)(2)(3)} = 3.688{,}312$$

$$SS_L = \frac{\sum_{i=1}^{m} G_i^2}{abr} - CF$$

$$= \frac{(212.05)^2 + (218.32)^2 + (200.77)^2}{(6)(2)(3)}$$

$$- 3.688{,}312 = 4{,}349$$

$$SS_A = \frac{\sum_{i=1}^{m} A_j^2}{mbr} - CF$$

$$= \frac{(63{,}83)^2 + (97{,}78)^2 + \ldots + (117{,}69)^2}{(3)(2)(3)}$$

$$- 3{,}688{,}312 = 136{,}814$$

$$SS_{(L\times A)} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{a} (LA)_{ij}^2}{br} - CF - SS_L - SS_A$$

$$= \frac{(17,90)^2 + (21.67)^2 + ... + (34.29)^2}{(2)(3)}$$

$$- 3.688,312 - 4,349 - 136,814 = 18,619$$

$$SS_B = \frac{\sum_{k=1}^{b} B_k^2}{mar} - CF$$

$$= \frac{(299.69)^2 + (331.45)^2}{(3)(6)(3)} - 3.688,312 = 9,340$$

$$SS_{(L\times B)} = \frac{\sum_{i=1}^{m}\sum_{k=1}^{k}(LB)_{ik}^2}{ar} - CF - SS_L - SS_B$$

$$= \frac{(97.83)^2 + (107,98)^2 + ... + (106,89)^2}{(6)(3)}$$

$$- 3.688,312 - 4,349 - 9,340 = 2,979$$

$$SS_{(A\times B)} = \frac{\sum_{j=1}^{a}\sum_{k=1}^{b}(AB)_{jk}^2}{mr} - CF - SS_A - SS_B$$

$$= \frac{(30,54)^2 + (33,29)^2 + ... + (62,00)^2}{(3)(3)}$$

$$- 3.688,312 - 136,814 - 9,340 = 1,147$$

$$SS_{(L\times A\times B)} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{a}\sum_{k=1}^{b}(LAB)_{ijk}^2}{r} - CF - SS_L$$

$$- SS_A - SS_B - SS_{(L\times A)} - SS_{(L\times B)} - SS_{(A\times B)}$$

$$= \frac{(7,15)^2 + (10,75)^2 + ... + (18,35)^2}{3}$$

$$- 3.688,312 - 4,349 - 136,814 - 9,340$$

$$- 18,619 - 2,979 - 1,147 = 2,888$$

3. Tính phương sai các nguồn biến động

$$MS_{nguồn} = SS_{nguồn} / DF_{nguồn}$$

Cho điểm: $MS_L = SS_L / DF_L = (4,349/(2) = 2,197$

Cho lần lặp các điểm: $MS_{r/L} = SS_{r/L} / DF_{r/L}$

$$= (4,555)/(6) = 0,759$$

Tương tự ta có kết quả ở bảng 7.29.

4. Tính các giá trị F_{TN} và mức ý nghĩa

$$F_{TN(L)} = \frac{MS_L}{MS_{r/L}} = \frac{2.200.532}{757.486} = 2,91$$

$$F_{TN(A)} = \frac{MS_A}{MS_{e(a)}} = \frac{27.369.819}{461.480} = 59,31$$

$$F_{TN(L\times A)} = \frac{MS_{L\times A}}{MS_{e(a)}} = \frac{1.859.166}{461.480} = 4,03$$

$$F_{TN(B)} = \frac{MS_B}{MS_{e(b)}} = \frac{9.335.676}{373.204} = 25,01$$

$$F_{TN(L\times B)} = \frac{MS_{L\times B}}{MS_{e(b)}} = \frac{1.481.662}{373.204} = 3,97$$

Tra bảng hoặc sử dụng phần mềm Excel để xác định mức ý nghĩa của F_{TN}.

Bảng 7.29: Kết quả phân tích phương sai

Nguồn	DF	SS	MS	F_{TN}	Prob.
Điểm (L)	2	4,394	2,197		0,131
Lần lặp × điểm	6	4,555	0,759		
Yếu tố lô chính (A)	5	136,814	27,363	59,33	0,000
L × A	10	18,619	1,862	4,04	0,001
Sai số chung (*a*)	30	13,836	0,461		
Yếu tố lô phụ (B)	1	9,340	9,340	25,06	0,000
L × B	2	2,979	1,489	4,00	0,028
A × B	5	1,147	0,229	0,62	
L × A × B	10	2,888	0,289	0,77	
Sai số chung (*b*)	36	13,417	0,373		
Tổng số	107				

5. So sánh các mức của từng yếu tố và tương tác

So sánh và phân hạng các mức của mỗi yếu tố và tương tác ở từng điểm và theo trung bình toàn thí nghiệm được thực hiện như kiểu thí nghiệm lô phụ.

6. Xác định hiệu quả các mức của từng yếu tố và các tương tác A_iB_j được thực hiện như thí nghiệm hai yếu tố đã đề cập ở mục 7.1 và 7.4.

7.6. THÍ NGHIỆM BA YẾU TỐ 2^3 (2 × 2× 2) KIỂU KHỐI ĐẦY ĐỦ NGẪU NHIÊN

7.6.1. Mô tả phương pháp bố trí và phân tích nguồn biến động

Thí nghiệm ba yếu tố A, B, C 2^3 = 8 nghiệm thức kiểu khối đầy đủ ngẫu nhiên, *r* lần lặp lại là thí nghiệm mà trong một khối có đầy đủ tất cả các nghiệm thức, mỗi nghiệm thức được phối hợp các mức khác nhau từ ba yếu tố, vị trí ô thí nghiệm được bố trí ngẫu nhiên. Với kiểu bố trí này, tổng bình phương tổng số được tạo bởi thành phần:

- Sai khác giữa các lần lặp lại;
- Sai khác giữa các mức A;
- Sai khác giữa các mức B;
- Sai khác giữa các mức C;
- Sai khác giữa các mức A×B;
- Sai khác giữa các mức A×C;
- Sai khác giữa các mức B×C;
- Sai khác giữa các mức A×B×C;
- Sai khác ngẫu nhiên (sai số).

Với mô hình:

$$X_{ijlk} = m + a_i + b_j + c_l + r_k + ab_{ij} + ac_{il} + bc_{jl} + abc_{ijl} + e_{ijlk}$$

Một khối (một lần lặp lại) có thể bố trí trên một băng hay trên các băng khác nhau sao cho trong một lần lặp lại tất cả các nghiệm thức càng gần nhau càng tốt.

Sau đây là sơ đồ bố trí hai băng cho một lần lặp lại.

I

BC	ABC	C	AC
A	B	0	AB

II

0	AB	A	BC
B	AC	ABC	C

III

AC	0	ABC	A
AB	B	C	BC

IV

AB	A	BC	AC
C	ABC	B	C

Hình 7.9: Sơ đồ thí nghiệm 3 yếu tố 2 × 2 × 2 kiểu RCBD, 4 lần lặp lại

(Trong đó: 0: $A_0B_0C_0$; A: $A_1B_0C_0$; B: $A_0B_1C_0$; C: $A_0B_0C_1$ AB: $A_1B_1C_0$; AC: $A_1B_0C_1$; BC: $A_0B_1C_1$; ABC: $A_1B_1C_1$)

Bảng phân tích phương sai có dạng

Nguồn	DF	SS	MS	F_{TN}
Lần lặp lại	$r - 1 = 7$			
Nghiệm thức	$abc - 1 = 7$			
A	$a - 1 = 1$			
B	$b - 1 = 1$			
C	$c - 1 = 1$			
AB	$(a - 1)(b - 1) = 1$			
AC	$(a - 1)(c - 1) = 1$			
BC	$(b - 1)(c - 1) = 1$			
ABC	$(a-1)(b-1)(c-1) = 1$			
Sai số	$3abc - r + 1 = 17$			
Tổng số	$4abc - 1 = 31$			

7.6.2. Phương pháp tính toán

Ví dụ: Để nghiên cứu ảnh hưởng của hai mức bón phân (0 và 1 - yếu tố A), hai loại thuốc diệt cỏ (0 và 1 - yếu tố B) và hai mức bón vôi (0 và 1 - yếu tố C) đến năng suất kiều mạch (Buckwheat), người ta đã bố trí thí nghiệm $2 \times 2 \times 2$, kiểu RCBD, 3 lần lặp lại. Kết quả được ghi ở Bảng 7.30 sau đây (Dospekhov).

Bảng 7.30: Năng suất kiều mạch (tạ/ha)

Yếu tố			Lần lặp lại (x_{ij})			ΣT	TB $(\bar{x})$
A	B	C	I	II	III		
0	0	0	16,2	19,5	18,3	54,0	18,0 (0)
		1	16,1	18,1	17,3	51,5	17,2 (*c*)
	1	0	17,5	23,6	20,5	61,6	20,5 (*b*)
		1	17,7	24,3	21,3	63,3	21,1 (*bc*)
1	0	0	22,3	31,4	27,6	81,3	27,1 (*a*)
		1	29,1	35,2	32,7	97,0	32,3 (*ac*)
	1	0	29,5	35,6	32,8	97,9	32,6 (*ab*)
		1	31,4	39,4	35,8	106,6	35,5 (*abc*)
ΣR			179,8	227,1	206,3	G = 613,2; m = 25,5	

Giải:

1. Tính tổng các nghiệm thức (ΣT), tổng các khối (ΣR), tổng toàn thí nghiệm (G) và trung bình chung (m).

2. Kiểm tra sự sai khác giữa các nghiệm thức

$$\mathrm{CF} = \frac{\left(\sum_{i=1}^{ab} \mathrm{x_{ij}}\right)^2}{abcr} = \frac{(613,2)^2}{24} = 15.667,260$$

Tổng bình phương tổng số:

$$\mathrm{SS_{TO}} = \sum_{i=1}^{abc} \mathrm{x_{ij}^2} - \mathrm{CF}$$

$$= (16,2)^2 + (16,1)^2 + \ldots + (32,8)^2 + (35,8)^2$$

$$- 15.667,260 = 1.270,920$$

Tổng bình phương lần lặp lại:

$$\mathrm{SS}_r = \frac{\sum_{i=1}^{r} \mathrm{R_i^2}}{abc} - \mathrm{CF} = \frac{(179,8)^2 + (227,1)^2 + (206,3)^2}{8}$$

$$- 15.667,260 = 140,507$$

Tổng bình phương nghiệm thức:

$$\mathrm{SS_t} = \frac{\sum_{i=1}^{abc} \mathrm{T_i^2}}{r} - \mathrm{CF} = \frac{(54,0)^2 + (51,5)^2 + \ldots + (106,6)^2}{3}$$

$$- 15.667,260 = 1.111,527$$

Tổng bình phương sai số:

$$\mathrm{SS_e} = \mathrm{SS_{TO}} - \mathrm{SS_t} - \mathrm{SS}_r = 1.270,920 -$$

$$1.111,527 - 140,507 = 18,886$$

Bảng 7.31: Kết quả phân tích phương sai bước đầu

Nguồn	DF	SS	MS	F_{TN}	Prob.
Lần lặp lại	2	140,508	70,254	52,079	-
Ngh. thức	7	1.111,527	158,790	117,710	0,0000
Sai số	14	18,886	1,349		
Tổng số	23	1270,920			

Kết quả $F_{TN(t)} > F_{0,01}$ ($P < 0,01$) cho phép tiếp tục tiến hành việc phân tích.

2. Tính tổng bình phương các nguồn biến động A, B, C, A×B, A×C, B×C và A×B×C.

Để tính các tổng bình phương này phải lập Bảng 7.32 sau:

Bảng 7.32: Tổng năng suất các yếu tố và tương tác

Tổng năng suất nghiệm thức				Tổng cộng các yếu tố và tương tác				
A	B	C: 0	C: 1	A	B	AB	AC	BC
0	0	54,0	51,5	A_0 230,4	B_0 283,8	A_0B_0 105,5	A_0C_0 115,6	B_0C_0 135,3
	1	61,6	63,3			A_0B_1 124,9	A_0C_1 114,8	B_0C_1 148,5
1	0	81,3	97,0	A_1 382,8	B_1 329,4	A_1B_0 178,3	A_1C_0 179,2	B_1C_0 159,5
	1	97,9	106,6			A_1B_1 204,5	A_1C_1 203,6	B_1C_1 169,9
Tổng C		C_0 294,8	C_1 318,4					
Tổng cộng G		613,2		613,2	613,2	613,2	613,2	613,2

Tổng bình phương yếu tố A (phân bón):

$$SS_A = \frac{\sum_{i=1}^{a} A_i^2}{rbc} - CF = \frac{(230,4)^2 + (382,8)^2}{(3)(2)(2)}$$

$$- 15.667,260 = 967,740$$

Tổng bình phương yếu tố B (thuốc diệt cỏ):

$$SS_B = \frac{\sum_{j=1}^{b} B_j^2}{rac} - CF$$

$$= \frac{(283,8)^2 + (329,4)^2}{(3)(2)(2)} - 15.667,260$$

$$= 86,640$$

Tổng bình phương yếu tố C (kali):

$$SS_C = \frac{\sum_{j=1}^{c} C_j^2}{rab} - CF$$

$$= \frac{(294,8)^2 + (318,4)^2}{(3)(2)(2)} - 15.667,260$$

$$= 23,207$$

Tổng bình phương $SS_{A\times B}$:

$$SS_{A\times B} = \frac{\sum (AB)^2}{rc} - CF - SS_A - SS_B$$

$$= \frac{(105,5)^2 + (124,9)^2 + (178,3)^2 + (204,5)^2}{6}$$

$$- 15.667,260 - 967,740 - 86,640$$

$$= 1,927$$

Tổng bình phương $SS_{A\times C}$:

$$SS_{A\times C} = \frac{\sum (AC)^2}{rb} - CF - SS_A - SS_C$$

$$= \frac{(115,6)^2 + (114,8)^2 + (179,2)^2 + (203,6)^2}{6}$$

$$- 15.667,260 - 967,740 - 23,207$$

$$= 26,460$$

Tổng bình phương $SS_{B\times C}$:

$$SS_{B\times C} = \frac{\sum (BC)^2}{ra} - CF - SS_B - SS_C$$

$$= \frac{(135.3)^2 + (148.5)^2 + (159.5)^2 + (169.9)^2}{6}$$

$$- 15.667,260 - 86,640 - 23,207$$

$$= 0,326$$

Tổng bình phương tương tác A×B×C ($SS_{A\times B\times C}$):

$SS_{A\times B\times C} = SS_t - SS_A - SS_B - SS_C - SS_{A\times B} - SS_{A\times C} - SS_{B\times C} = 1.111,527 - 967,740 - 86,640 - 23,207 - 1,927 - 26,460 - 0,326 = 5,227$

3. Tính phương sai các nguồn biến động A, B, A×B, A×B×C và phương sai sai số được theo cách tính đã biết. Kết quả tính toán được ghi ở bảng 7.33.

4. Đánh giá sự sai khác giữa các mức của các yếu tố và giửa các mức tương tác

Tính giá trị F_{TN} của các yếu tố và các loại tương tác bằng cách so sánh phương sai của chúng với phương sai sai số. Kết quả tính toán được ghi ở Bảng 7.33.

Bảng 7.33: Kết quả phân tích phương sai

Nguồn	DF	SS	MS	F_{TN}	Prob.
Lần lặp lại	2	140,508	70,254	52,079	-
Nghiệm thức	7	1.111,527	158,790	117,710	0,0000
A	1	967,740	967,74	717,376	0,0000
B	1	86,640	86,64	64,225	0,0000
C	1	23,207	23,207	17,203	0,0010
A×B	1	1,927	1,927	1,428	0,2519
A×C	1	26,460	26,46	19,615	0,0006
B×C	1	0,326	0,326	0,242	0,6304
A×B×C	1	5,227	5,227	3,875	0,0691
Sai số	14	18.886	1,349		
Tổng số	23	1270,920			

Như vậy:

- Có sự khác nhau rất đáng tin cậy giữa hai mức bón phân, giữa hai loại thuốc trừ sâu, giữa hai mức vôi và giữa các mức tương tác giữa bón phân và bón vôi ($P < 0,01$)

- Không có sự khác nhau giữa các mức trong các loại tương tác khác ($P > 0,05$).

5. So sánh từng yếu tố và tương tác

- Sai khác giữa hai nghiệm thức:

$$Sd = \sqrt{\frac{2MS_e}{r}} = \sqrt{\frac{(2)(1.349)}{3}} = 0,948$$

$$LSD_{0,05} = t_{0,05}Sd = (2,145)(0,948) = 2,0$$

$$LSD_{0,01} = t_{0,01}Sd = (2,977)(0,948) = 2,8$$

- Sai khác giữa các yếu tố:

$$Sd' = \sqrt{\frac{2MS_e}{abr}} = \sqrt{\frac{(2)(1,349)}{12}} = 0,474$$

$$LSD_{0,05} = t_{0,05}Sd = (2,145)(0,474) = 1,0$$
$$LSD_{0,01} = t_{0,01}Sd = (2,977)(0,474) = 1,4$$

- Sai khác giữa các mức tương tác:

$$Sd'' = \sqrt{\frac{2MS_e}{ar}} = \sqrt{\frac{2MS_e}{br}} = \sqrt{\frac{2MS_e}{cr}}$$

$$= \sqrt{\frac{(2)(1,349)}{6}} = 0,671$$

$$LSD_{0,05} = t_{0,05}Sd = (2,145)(0,671) = 1,4$$
$$LSD_{0,01} = t_{0,01}Sd = (2,977)(0,474) = 2,0$$

6. Tính hiệu quả của các yếu tố và tương tác giữa chúng

Để tính hiệu quả của từng yếu tố và tương tác được xác định nhờ giải hệ phương trình sau:

$$\begin{array}{cccccccc} A_0B_0C_0 & A_1B_0C_0 & A_0B_1C_0 & A_0B_0C_1 & A_1B_1C_0 & A_1B_0C_1 & A_0B_1C_1 & A_1B_1C_1 \\ 18,0 & 27,1 & 20,5 & 17,2 & 32,6 & 32,3 & 21,1 & 35,5 \end{array}$$

$$\begin{pmatrix} A \\ B \\ C \\ AB \\ AC \\ BC \\ ABC \end{pmatrix} = \begin{pmatrix} -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \end{pmatrix} = \begin{pmatrix} 12,7^{*} \\ 3,8^{*} \\ 2,0^{*} \\ 0,6 \\ 2,1^{*} \\ -0,2 \\ -0,9 \end{pmatrix}$$

$$\overline{X} = 25,5$$

(*) Độ tin cậy 95%

Trong hệ phương trình trên, ma trận 7 × 8 là các hệ số của hệ phương trình tính toán hiệu quả của các yếu tố

A, B, C và các tương tác được tính từ thành phần các nghiệm thức.

Như vậy, trong sự hình thành năng suất, phân bón đóng vai trò quan trong nhất, kế đến là thuốc trừ sâu, cuối cùng là vôi và tương tác giữa bón phân và bón vôi.

7. Việc tính LDS_{α} giữa các nghiệm thức và phân hạng được thực hiện theo cách đã biết.

7.7. THÍ NGHIỆM BA YẾU TỐ KIỂU 2^3 THEO KHỐI CÂN BẰNG CÁC YẾU TỐ

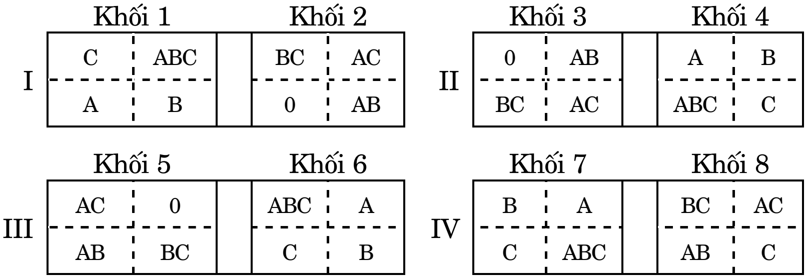
7.7.1. Mô tả phương pháp bố trí và phân tích ANOVA

Thí nghiệm ba yếu tố A, B, C kiểu 2^3 nếu bố trí theo sơ đồ RCBD với $2 \times 2 \times 2 = 8$ nghiệm thức, 4 lần lặp lại như mục 5 trên đây thì tổng bình phương tổng số được tạo bởi thành phần:

- Sai khác giữa các lần lặp lại;
- Sai khác giữa các mức A;
- Sai khác giữa các mức B;
- Sai khác giữa các mức C;
- Sai khác giữa các mức A×B;
- Sai khác giữa các mức A×C;
- Sai khác giữa các mức B×C;
- Sai khác giữa các mức A×B×C;
- Sai khác ngẫu nhiên (sai số).

Tuy nhiên việc bố trí theo khối cân bằng giữa các yếu tố (Hình 7.10) để tăng số lần lặp lại nên sai số sẽ giảm xuống và độ chính xác tăng lên.

Sơ đồ bố trí



Hình 7.10: Sơ đồ thí nghiệm 3 yếu tố
$2 \times 2 \times 2$, 8 khối, 4 lần lặp lại
(Trong đó: 0: $A_0B_0C_0$; A: $A_1B_0C_0$; B: $A_0B_1C_0$; C: $A_0B_0C_1$
AB: $A_1B_1C_0$; AC: $A_1B_0C_1$; BC: $A_0B_1C_1$; ABC: $A_1B_1C_1$)

Bảng phân tích phương sai có dạng

Nguồn	DF	SS	MS	F_{TN}
Khối	$k - 1 = 7$			
Nghiệm thức	$abc - 1 = 7$			
A	$a - 1 = 1$			
B	$b - 1 = 1$			
C	$c - 1 = 1$			
A×B	$(a - 1)(b - 1) = 1$			
A×C	$(a - 1)(c - 1) = 1$			
B×C	$(b - 1)(c - 1) = 1$			
A×B×C	$(a - 1)(b - 1)(c - 1) = 1$			
Sai số	$3abc - k + 1 = 17$			
Tổng số	$4abc - 1 = 31$			

7.7.2. Phương pháp tính toán

Ví dụ: Phân tích phương sai năng suất lúa mỳ mùa xuân trong thí nghiệm ba yếu tố $2 \times 2 \times 2$: hai mức đạm (0, 1), hai mức lân (0, 1) và hai mức kali (0, 1) được bố trí trong 8 khối, 4 lần lặp lại theo Hình 7.9 với số liệu ở Bảng 7.34 sau đây (Dospekhov, 1985).

Bảng 7.34: Năng suất lúa mỳ mùa xuân (tạ/ha)

Ngh. thức	Lần lặp lại (*r*) I	II	III	IV	Tổng (ΣT)	Tr. bình
			Khối			
	1	*4*	*6*	*7*		
a	27,2	27,8	26,4	25,2	106,6	26,6
b	26,0	25,8	24,8	25,2	101,8	25,4
c	26,1	26,6	24,2	23,0	99,9	25,0
abc	30,2	32,6	30,4	28,7	121,9	30,5
ΣK	109,5	112,8	105,8	102,1		
			Khối			
	2	*3*	*5*	*8*		
0	23,4	25,0	23,4	22,0	93,8	23,4
ab	29,4	29,8	28,0	27,9	115,1	28,8
ac	26,9	28,0	27,0	26,8	108,7	27,2
bc	24,9	28,0	27,0	25,8	105,7	26,4
ΣK	104,6	110,8	105,4	102,5	m =	**26,7**
ΣR	214,1	223,6	211,2	204,6	G =	**853,5**

Giải:

1. Tính tổng các nghiệm thức (ΣT), tổng các khối (ΣK), tổng các lần lặp lại (ΣR), tổng toàn thí nghiệm (G) và trung bình chung (*m*).

2. Kiểm tra sự sai khác giữa các nghiệm thức

$$CF = \frac{\left(\sum_{i=1}^{ab} x_{ij}\right)^2}{abcr} = \frac{(853,5)^2}{32} = 22.764,445$$

Tổng bình phương tổng số:

$$SS_{TO} = \sum_{i=1}^{abc} x_{ij}^2 - CF = (27,2)^2 + (27,8)^2 + \ldots$$

$$+ (27,0)^2 + (25,8)^2 - \ 22.764,445 = 170,444$$

Tổng bình phương khối:

$$SS_k = \frac{\sum_{i=1}^{k} K_i^2}{4} - CF = \frac{(109,5)^2 + (112,8)^2 + ... + (102,5)^2}{4}$$

$$- \ 22.764,445 = 26,892$$

Tổng bình phương nghiệm thức:

$$SS_t = \frac{\sum_{i=1}^{ab} T_i^2}{r} - CF = \frac{(106,6)^2 + (101,8)^2 + ... + (105,7)^2}{4}$$

$$- \ 22.764,445 = 135,817$$

Tổng bình phương sai số:

$$SS_e = SS_{TO} - SS_t - SS_k$$

$$= 170,444 - 135,817 - 26,892 = 7,735$$

Bảng 7.35: Kết quả phân tích phương sai bước đầu

Nguồn	DF	SS	MS	F_{TN}	Prob.
Khối	7	26,892	3,842	8,44	-
Nghiệm thức	7	135,817	19,402	42,64	0,0000
Sai số	17	7,735	0,455		
Tổng số	31	170,444			

Kết quả $F_{TN(t)} > F_{0,01}$ ($P < 0,01$) cho phép tiếp tục tiến hành việc phân tích.

2. Tính tổng bình phương các nguồn biến động A, B, C, A×B, A×C, B×C và A×B×C.

Để tính các tổng bình phương này phải lập bảng sau:

Bảng 7.36: Tổng năng suất các yếu tố và tương tác

Tổng năng suất các nghiệm thức				Tổng cộng các yếu tố và tương tác				
A	B	C: 0	C: 1	A	B	AB	AC	BC
0	0	93,8	99,9	A_0 401,2	B_0 409,0	A_0B_0 193,7	A_0C_0 195,6	B_0C_0 200,4
	1	101,8	105,7			A_0B_1 207,5	A_0C_1 205,6	B_0C_1 208,6
1	0	106,6	108,7	A_1 452,3	B_1 444,5	A_1B_0 215,3	A_1C_0 221,7	B_1C_0 216,9
	1	115,1	121,9			A_1B_1 237,0	A_1C_1 230,6	B_1C_1 227,6
Tổng C		C_0 417,3	C_1 436,2					
Tổng cộng G		853,5		853,5	853,5	853,5	853,5	853,5

Tổng bình phương yếu tố A (đạm):

$$SS_A = \frac{\sum_{i=1}^{a} A_i^2}{rbc} - CF = \frac{(401,2)^2 + (452,3)^2}{(4)(2)(2)}$$

$$- 22.764,445 = 81,600$$

Tổng bình phương yếu tố B (lân):

$$SS_B = \frac{\sum_{j=1}^{b} B_j^2}{rac} - CF = \frac{(409,0)^2 + (444,5)^2}{(4)(2)(2)}$$

$$- 22.764,445 = 39,400$$

Tổng bình phương yếu tố C (kali):

$$SS_C = \frac{\sum_{j=1}^{c} C_j^2}{rab} - CF$$

$$= \frac{(417,3)^2 + (436,2)^2}{(4)(2)(2)} - 22.764,445 = 11,100$$

Tổng bình phương SS_{A+B+AB}:

$$SS_{A+B+AB} = \frac{\sum_{j=1}^{b} AB^2}{rc} - CF$$

$$= \frac{(193,7)^2 + (207,5)^2 + (215,3)^2 + (237,0)^2}{8}$$

$$- 22.764,445 = 122,933$$

Tổng bình phương tương tác A×B ($SS_{A\times B}$):

$$SS_{A\times B} = SS_{A+B+AB} - SS_A - SS_B$$

$$= 122,933 - 81,600 - 39,400 = 1,933$$

Tổng bình phương SS_{A+C+AC}:

$$SS_{A+C+AC} = \frac{\sum_{j=1}^{b} AC^2}{rb} - CF$$

$$= \frac{(195,6)^2 + (205,6)^2 + (221,7)^2 + (230,6)^2}{8}$$

$$- 22.764,445 = 92,801$$

Tổng bình phương tương tác A×C ($SS_{A×C}$):

$$SS_{A\times C} = SS_{A+C+AC} - SS_A - SS_C$$

$$= 92,801 - 81,600 - 11,100 = 0,101$$

Tổng bình phương SS_{B+C+BC}:

$$SS_{B+C+BC} = \frac{\sum_{j=1}^{b} BC^2}{ra} - CF$$

$$= \frac{(200,4)^2 + (208,6)^2 + (216,9)^2 + (227,6)^2}{8}$$

$$- 22.764,445 = 50,741$$

Tổng bình phương tương tác B×C ($SS_{B×C}$):

$$SS_{B\times C} = SS_{B+C+BC} - SS_B - SS_C$$

$$= 50,741 - 39,400 - 11,100 = 0,241$$

Tổng bình phương tương tác ABC (SS_{ABC}):

$SS_{ABC} = SS_t - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC} = 135,817 - 81,600 - 39,400 - 11,100 - 1,993 - 0,101 - 0,241 = 1,442$

3. Tính phương sai các nguồn biến động A, B, A×B, A×B×C và phương sai sai số được theo cách tính đã biết. Kết quả tính toán được ghi ở bảng 7.37.

4. Đánh giá sự sai khác giữa các mức của các yếu tố và giữa các mức tương tác

Tính giá trị F_{TN} của các yếu tố và các loại tương tác bằng cách so sánh phương sai của chúng với phương sai sai số. Kết quả tính toán được ghi ở Bảng 7.37.

Bảng 7.37: Kết quả phân tích phương sai

Nguồn	DF	SS	MS	F_{TN}	Prob.
Khối	7	26,892	3,842	8,44	-
Ngh. thức	7	135,817	19,402	42,64	0,0000
A	1	81,600	81,600	179,33	0,0000
B	1	39,400	39,400	86,59	0,0000
C	1	11,100	11,100	24,39	0,0001
A×B	1	1,933	1,33	4,25	0,0549
A×C	1	0,101	0,101	0,22	0,6450
B×C	1	0,241	0,241	0,53	0,4765
A×B×C	1	1,442	1,442	3,17	0,0929
Sai số	17	7,735	0,455	1,00	
Tổng số	31	170,444			

Như vậy:

- Có sự khác nhau rất đáng tin cậy giữa các mức đạm, giữa các mức lân và giữa các mức kali ($P < 0,01$)

- Không có sự khác nhau giữa các loại tương tác ($P > 0,05$).

5. So sánh từng yếu tố và tương tác

- Sai khác giữa hai nghiệm thức bất kỳ:

$$Sd = \sqrt{\frac{2MS_e}{r}} = \sqrt{\frac{(2)(0,455)}{4}} = 0,48$$

$$LSD_{0,05} = t_{0,05}Sd = (2,110)(0,48) = 1,0$$
$$LSD_{0,01} = t_{0,01}Sd = (2,898)(0,48) = 1,4$$

- Sai khác giữa hai mức của yếu tố A, B, C và tương tác:

$$Sd' = \sqrt{\frac{2MS_e}{abr}} = \sqrt{\frac{(2)(0,455)}{16}} = 0,24$$

$$(abr = acr = bcr)$$

$$LSD_{0,05} = t_{0,05}Sd' = (2,110)(0,24) = 0,5$$

$$LSD_{0,01} = t_{0,01}Sd = (2,898)(0,24) = 0,7$$

6. Việc tính LDS_{α} giữa các nghiệm thức và phân hạng được thực hiện theo cách đã biết.

7. So sánh hiệu quả tính theo kiểu 8 khối và tính theo 4 lần lặp lại.

Nếu xử lý theo 4 lần lặp lại (xem Hình 7.10), tức là lần lặp lại I gồm khối 1 và 2, lần lặp lại II gồm khối 3 và 4, lần lặp lại III gồm khối 5 và 6 và lần lặp lại IV gồm khối 7 và 8 ta được:

Tổng bình phương lần lặp lại:

$$SS_r = \frac{\sum_{i=1}^{r} R_i^2}{abc} - CF$$

$$\frac{(214,1)^2 + (223,6)^2 + (211,2)^2 + (204,6)^2}{8}$$

$$- 22.764,445 = 23,351$$

Tổng bình phương sai sai số

$$SS_e = SS_T - SS_t - SS_r$$

$$= 170,444 - 135,817 - 23,351 = 11,277$$

với $(abc - 1)(r - 1) = 21$ độ tự do.

Phương sai sai số sẽ là:

$$MS_e = SS_e / DF_e = 11,277 : 21 = 0,537$$

Như vậy, tính theo khối phương sai sai số sẽ giảm:

$$r_e(\%) = \frac{0,537 - 0,455}{0,537} \times 100 = 15,3\%$$

7.8. THÍ NGHIỆM BA YẾU TỐ KIỂU PHỐI HỢP LÔ PHỤ - LÔ SỌC (Strip-Split- Plot Design)

7.8.1. Mô tả phương pháp bố trí

Kiểu phối hợp lô phụ - lô sọc là kiểu mở rộng của kiểu lô sọc, trong đó ở các lô giao nhau giữa hai yếu tố kiểu lô sọc được bố trí các lô chính của yếu tố thứ ba. Kiểu phối hợp lô phụ - lô sọc có hai điểm đặc biệt sau:

- Có bốn loại lô: lô sọc ngang, lô sọc dọc, lô giao nhau và lô phụ.

- Có bốn cách so sánh và đánh giá các nguồn biến động (cho ba yếu tố nghiên cứu và tương tác giữa chúng).

Sau đây là sơ đồ bố trí kiểu phối hợp lô phụ - lô sọc trong thí nghiệm ba yếu tố: 6 giống , 3 liều bón đạm và 2 liều bón lân với 3 lần lặp lại.

I

	N_1	N_3	N_2
V_6	P_2	P_1	P_2
	P_1	P_2	P_1
V_5	P_2	P_1	P_1
	P_1	P_2	P_2
V_3	P_1	P_2	P_2
	P_2	P_1	P_1
V_2	P_1	P_2	P_1
	P_2	P_1	P_2
V_4	P_2	P_1	P_2
	P_1	P_2	P_1
V_1	P_1	P_2	P_2
	P_2	P_1	P_1

II

	N_3	N_2	N_1
V_4	P_2	P_1	P_2
	P_1	P_2	P_1
V_2	P_1	P_2	P_2
	P_2	P_1	P_1
V_6	P_2	P_2	P_2
	P_1	P_1	P_1
V_3	P_1	P_1	P_2
	P_2	P_2	P_1
V_1	P_1	P_2	P_2
	P_2	P_1	P_1
V_5	P_1	P_2	P_2
	P_2	P_1	P_1

III

	N_3	N_1	N_2
V_5	P_2	P_2	P_2
	P_1	P_1	P_1
V_2	P_1	P_2	P_2
	P_2	P_1	P_1
V_3	P_1	P_1	P_2
	P_2	P_2	P_1
V_4	P_1	P_2	P_2
	P_2	P_1	P_1
V_6	P_1	P_2	P_2
	P_2	P_1	P_1
V_1	P_1	P_1	P_2
	P_2	P_2	P_1

Hình 7.11: Sơ đồ phối hợp lô phụ - lô sọc 3 × 6× 2 với 6 giống (V_1 đến V_6 - lô ngang), 3 mức đạm (N_1 đến N_3 - lô dọc) và 2 mức lân (lô phụ), 3 lần lặp lại

Bảng phân tích phương sai có dạng

Nguồn	DF	SS	MS	F_{TN}
Lần lặp lại	$r - 1$			
Yếu tố dọc (A)	$a - 1$			
Sai số (a)	$(r - 1)(a - 1)$			
Yếu tố ngang (B)	$b - 1$			
Sai số (b)	$(r - 1)(b - 1)$			
Tương tác A×B	$(a - 1)(b - 1)$			
Sai số (c)	$(r - 1)\ )(a - 1)\ (b - 1)$			
Yếu tố lô phụ (C)	$c - 1$			
Tương tác A×C	$(a - 1)(c - 1)$			
Tương tác B×C	$(b - 1)(c - 1)$			
Tương tác A×B×C	$(a - 1)(b - 1)(c - 1)$			
Sai số	$ab(r - 1)(c - 1)$			
Tổng số	$rabc - 1$			

7.6.2. Phương pháp phân tích phương sai

7.6.2.1. Mô hình toán học

Trong thí nghiệm 3 yếu tố kiểu phối hợp lô phụ - lô sọc, giá trị của ô nghiệm thức mức *i* của yếu tố A, mức *j* của yếu tố B và mức *l* của yếu tố C ở lần lặp lại *k* (X_{ijlk}) được tính theo mô hình:

$$X_{ijlk} = m + a_i + b_j + c_l + r_k + ab_{ij} + ac_{il}$$
$$+ ar_{ik} + bc_{jl} + br_{jk} + abc_{ijl} + e_{ijlk}$$

trong đó: m là giá trị trung bình toàn thí nghiệm, a_i, b_j, c_l là giá trị do mức *i* của yếu tố A, mức *j* của yếu tố B và mức l của yếu tố C tạo ra; r_k là giá trị tạo nên bởi ảnh hưởng của lần lặp lại *k*; ab_{ij}, ac_{il}, bc_{jl} là giá trị do tương tác giữa hai yếu tố $A_i \times B_j$, $A_i \times C_l$, $B_j \times C_l$; abc_{ijl} là giá trị do tương tác giữa ba yếu tố $A_i \times B_j \times C_l$; ar_{ik}, br_{jk} là giá trị tương tác giữa A_i, B_j với lần lặp lại k tạo ra, e_{ijlk} là sai số tại ô $ijlk$.

7.6.2.2. Phân tích phương sai

Để minh họa cách phân tích phương sai, xét ví dụ ở bảng 7.38 và các bảng tổng hợp sau đây.

Bảng 7.38: Năng suất lúa trong thí nghiệm kiểu phối hợp lô phụ - lô sọc 6 giống, 3 liều bón đạm và 2 liều bón lân với 3 lần lặp lại.

Phương pháp trồng	Giống	Năng suất (tấn/ha)			Tổng NT (T)
		I	II	III	
P_1 (sạ hạt)		N_1 (0 kg N/ha)			
	V_1	2,37	3,96	4,38	10,71
	V_2	4,01	5,80	5,00	14,81
	V_3	2,62	4,51	5,62	12,75
	V_4	2,73	5,63	3,82	12,18
	V_5	4,45	3,28	4,58	12,31
	V_6	2,57	3,72	3,33	9,62
		N_2 (60 kg N/ha)			
	V_1	4,08	6,43	4,89	15,40
	V_2	5,63	7,33	7,18	20,14
	V_3	4,68	6,67	7,02	18,37
	V_4	4,84	7,01	4,82	16,67
	V_5	5,55	5,34	6,01	16,90
	V_6	3,90	2,82	4,43	11,15
		N_3 (120 kg N/ha)			
	V_1	7,25	6,81	8,58	22,64
	V_2	7,05	8,28	6,30	21,63
	V_3	7,67	7,33	8,61	23,61
	V_4	6,88	7,74	6,67	21,29
	V_5	6,88	5,08	6,08	18,04
	V_6	1,56	2,71	3,21	7,48

P_2 (cấy)			N_1 (0 kg N/ha)		
	V_1	2,29	3,53	2,54	8,36
	V_2	4,04	4,89	4,58	13,51
	V_3	4,53	4,87	3,63	13,03
	V_4	5,27	6,20	4,04	15,51
	V_5	4,66	2,80	3,74	11,20
	V_6	4,54	5,46	3,54	13,54
			N_2 (60 kg N/ha)		
	V_1	3,09	7,50	4,36	14,95
	V_2	3,73	7,43	5,38	16,54
	V_3	4,95	7,61	6,14	18,70
	V_4	4,88	6,93	4,83	16,64
	V_5	4,65	5,01	4,67	14,33
	V_6	4,63	4,46	4,77	13,86
			N_3 (120 kg N/ha)		
	V_1	6,66	6,35	7,76	20,77
	V_2	6,44	7,65	5,74	19,83
	V_3	8,63	7,10	7,42	23,15
	V_4	6,55	9,84	7,25	23,64
	V_5	7,00	4,49	6,56	18,05
	V_6	5,37	7,22	6,37	18,96

Bảng 7.39: Tổng năng suất các liều đạm ở các lần lặp lại (RA)

Đạm	Tổng năng suất (RA,tấn/ha)			Tổng N (A)
	I	II	III	
N_1	44,08	54,65	48,80	147,53
N_2	54,61	74,54	64,50	193,65
N_3	77,94	80,60	80,55	239,09
ΣR	176,63	209,79	193,85	
ΣΣ (G)				**580,27**
m				5,37

Bảng 7.40: Tổng năng suất các giống ở các lần lặp lại (RB)

Giống	Tổng năng suất (RB, tấn/ha)			Tổng giống (B)
	I	II	III	
V_1	25,74	34,58	32,51	92,83
V_2	30,90	41,38	34,18	106,46
V_3	33,08	38,09	38,44	109,61
V_4	31,15	43,35	31,43	105,93
V_5	33,19	26,00	31,64	90,83
V_6	22,57	26,39	25,65	74,61

Bảng 7.41: Tổng năng suất các giống × các liều đạm (AB)

Giống	Tổng năng suất (AB, tấn/ha)		
	N_1	N_2	N_3
V_1	19,07	30,35	43,41
V_2	28,32	36,68	41,46
V_3	25,78	37,07	46,76
V_4	27,69	33,31	44,93
V_5	23,51	31,23	36,09
V_6	23,16	25,01	26,44

Bảng 7.42: Tổng năng suất giống × đạm ở ở các lần lặp lại (RAB)

Giống	Năng suất (tấn/ha)		
	I	II	III
		N_1 (0 kg N/ha)	
V_1	4,66	7,49	6,92
V_2	8,05	10,69	9,58
V_3	7,15	9,38	9,25
V_4	8,00	11,83	7,86
V_5	9,11	6,08	8,32
V_6	7,11	9,18	6,87

		N_2 (60 kg N/ha)	
V_1	7,17	13,93	9,25
V_2	9,36	14,76	12,56
V_3	9,63	14,28	13,16
V_4	9,72	13,94	9,65
V_5	10,20	10,35	10,68
V_6	8,53	7,28	9,20
		N_3 (120 kg N/ha)	
V_1	13,91	13,16	16,34
V_2	13,49	15,93	12,04
V_3	16,30	14,43	16,03
V_4	13,43	17,58	13,92
V_5	13,88	9,57	12,64
V_6	6,93	9,93	9,58

Bảng 7.43: Tổng năng suất đạm × phương pháp trồng

Mức đạm	Phương pháp trồng (AC)	
	Sạ (P_1)	Cấy (P_2)
N_1	72,38	75,15
N_2	98,63	95,02
N_3	114,69	124,40
ΣC	285,70	294,57

Bảng 7.44: Tổng năng suất giống × phương pháp trồng

Mức đạm	Phương pháp trồng (AC)	
	Sạ (P_1)	Cấy (P_2)
V_1	48,75	44,08
V_2	56,58	49,88
V_3	54,73	54,88
V_4	50,14	55,79
V_5	47,25	43,58
V_6	28,25	46,36

Giải:

1. Tính tổng bình phương các nguồn biến động

$$CF = \frac{\left(\sum_{i=1}^{abc} x_{ij}\right)^2}{rabc} = \frac{(580,27)^2}{(3)(3)(6)(2)} = 3.117,715$$

Tổng bình phương tổng số:

$$SS_{TO} = \sum_{i=1}^{abc} x_{ij}^2 - CF$$

$$= (2,37)^2 + (3.9)^2 + \ldots + (6,37)^2$$

$$- 3.117.715 = 307,169$$

Tổng bình phương lần lặp lại:

$$SS_r = \frac{\sum_{i=1}^{r} R_i^2}{abc} - CF$$

$$= \frac{(176,63)^2 + (209,79)^2 + (193,85)^2}{36}$$

$$- 3.117,715 = 15,280$$

Tổng bình phương nghiệm thức:

$$SS_t = \frac{\sum_{i=1}^{ab} T_i^2}{r} - CF$$

$$= \frac{(10,71)^2 + (14,81)^2 + \ldots + (18,96)^2}{r}$$

$$- 3.117,715 = 224,588$$

Tổng bình phương yếu tố A (đạm):

$$SS_A = \frac{\sum_{i=1}^{a} A_i^2}{rbc} - CF$$

$$= \frac{(147,53)^2 + (193,65)^2 + (239,09)^2}{36}$$

$- \ 3.117,715 = 116,436$

Tổng bình phương sai số (a):

$$SS_{e(a)} = \frac{\sum_{i=1}^{ra} (RA)^2}{bc} - CF - SS_r - SS_A$$

$$= \frac{(44,08)^2 + (54,65)^2 + ... + (80,55)^2}{12}$$

$- \ 3.117,715 - 15,280 - 116,436 = 6,330$

Tổng bình phương yếu tố B (giống):

$$SS_B = \frac{\sum_{j=1}^{rb} B_j^2}{rac} - CF$$

$$= \frac{(92,83)^2 + (106,46)^2 + ... + (74,61)^2}{18}$$

$- \ 3.117,715 = 49,140$

Tổng bình phương sai số (b):

$$SS_{e(b)} = \frac{\sum_{i=1}^{ra} (RB)^2}{ac} - CF - SS_r - SS_B$$

$$= \frac{(25,74)^2 + (34,58)^2 + ... + (25,65)^2}{6}$$

– 3.117,715 – 15,280 – 49,140 = 26,731

Tổng bình phương tương tác A×B:

$$SS_{A\times B} = \frac{\sum (AB)^2}{rc} - CF - SS_A - SS_B$$

$$= \frac{(19,07)^2 + (30,35)^2 + ... + (26,44)^2}{6}$$

– 3.117,715 – 116,436 – 49,140 = 24,602

Tổng bình phương sai số (c):

$$SS_{e(c)} = \frac{\sum_{i=1}^{rab} (RAB)^2}{c} - CF - SS_r - SS_a - SS_{e(a)}$$
$$- SS_B - SS_{e(b)} - SS_{A\times B}$$

$$= \frac{(4,66)^2 + (7,49)^2 + ... + (9,58)^2}{2}$$

– 3.117,715 – 15,280 – 116,436 – 6,330

– 49,140 – 26,731 – 24,602 = 19,086

Tổng bình phương yếu tố C (phương pháp trồng):

$$SS_C = \frac{\sum_{j=1}^{rab} C_j^2}{rab} - CF = \frac{(258,70)^2 + (294,57)^2}{54}$$

– 3.117,715 = 0,728

Tổng bình phương tương tác A×C:

$$SS_{A\times C} = \frac{\sum(AC)^2}{rb} - CF - SS_A - SS_C$$

$$= \frac{(72,38)^2 + (75,15)^2 + ... + (124,40)^2}{18}$$

$- \; 3.117,715 - 116,436 - 0,728 = 2,466$

Tổng bình phương tương tác B×C:

$$SS_{B\times C} = \frac{\sum(BC)^2}{ra} - CF - SS_B - SS_C$$

$$= \frac{(48,75)^2 + (44,08)^2 + ... + (46,36)^2}{9}$$

$- \; 3.117,715 - 49,140 - 0,728 = 23,721$

Tổng bình phương tương tác A×B×C:

$SS_{A\times B\times C} = SS_t - SS_A - SS_B - SS_C$

$- \; SS_{A\times B} - SS_{A\times C} - SS_{B\times C}$

$= 224,588 - 116,436 - 49,140 - 0,728$

$- 24,602 - 2,466 - 23,721 = 7,495$

Tổng bình phương sai số (d):

$SS_{e(d)} = SS_{TO} - SS_t - SS_r - SS_{e(a)} - SS_{e(b)} - SS_{e(c} = 307,169 - 224,588 - 15,280 - 6,330 - 26,731 - 19,086 = 15,155$

3. Tính phương sai các nguồn biến động A, B, A×B, A×B×C và phương sai sai số theo cách tính đã biết. Kết quả tính toán được ghi ở Bảng 7.45.

Bảng 7.45: Kết quả phân tích phương sai

Nguồn	DF	SS	MS	F_{TN}	Prob.
Lần lặp lại	2	15,280	7,640		
Đạm (A)	2	116,436	58,218	36,79	0,0027
Sai số (*a*)	4	6,330	1,582		
Giống (B)	5	49,140	9,828	3,68	0,0378
Sai số (*b*)	10	26,731	2,673		
Tương tác A×B	10	24,602	2,460	2,58	0,0347
Sai số (*c*)	20	19,086	0,954		
PP gieo (C)	1	0,728	0,728	1,73	0,1993
Tương tác A×C	2	2,466	1,233	2,93	0,0662
Tương tác B×C	5	23,721	4,744	11,27	0,0000
Tương tác A×B×C	10	7,495	0,750	1,78	0,1003
Sai số (*d*)	36	15,155	0,421		
Tổng số	107	307,169			

4. Tính giá trị F_{TN}

$$F_{TN(A)} = \frac{MS_A}{MS_{e(a)}}\ ;\ F_{TN(B)} = \frac{MS_B}{MS_{e(b)}}$$

$$F_{TN(A\times B)} = \frac{MS_{A\times B}}{MS_{e(c)}}\ F_{TN(C)} = \frac{MS_C}{MS_{e(d)}}$$

Các F_{TN} của tương tác A×C, B×C và A×B×C được tính bằng cách chia MS của chúng cho $MS_{e(d)}$. Kết quả tính toán được ghi ở bảng 7.44.

Như vậy:

- Có sự khác nhau rất đáng tin cậy giữa các mức bón đạm (P < 0,001), có sự khác nhau giữa các giống, tương tác giữa giống và mức bón đạm (P < 0,05)

- Các giống khác nhau phương pháp gieo trồng (sạ và cấy) khác nhau, vì vậy sự khác nhau giữa các mức tương tác giữa giống và phương pháp gieo trồng (AC) rất đáng tin cậy ($P < 0{,}001$).

5. So sánh các mức của từng yếu tố và tương tác

- Sai khác giữa hai mức của yếu tố A:

$$Sd_A = \sqrt{\frac{2MS_{e(a)}}{rbc}} = \sqrt{\frac{(2)(1{,}582)}{(3)(6)(2)}} = 0{,}297$$

$$LSD_{0,05}\text{ (cho A)} = t_{0,05}Sd_A = (2{,}776)(0{,}297)$$

$$= 0{,}82$$

$$LSD_{0,01}\text{ (cho A)} = t_{0,01}Sd_A = (4{,}604)(0{,}297)$$

$$= 1{,}37$$

- Sai khác giữa hai mức của yếu tố B:

$$Sd_B = \sqrt{\frac{2MS_{e(b)}}{rac}} = \sqrt{\frac{(2)(2{,}673)}{(3)(3)(2)}} = 0{,}545$$

$$LSD_{0,05} = t_{0,05}Sd_B = (2{,}228)(0{,}545) = 1{,}21$$

- Sai khác giữa hai mức tương tác A×B:

$$Sd_{A\times B} = \sqrt{\frac{2MS_{e(c)}}{rc}} = \sqrt{\frac{(2)(0{,}954)}{(3)(2)}} = 0{,}564$$

$$LSD_{0,05}\text{ (cho A×B)} = t_{0,05}Sd_{A\times B}$$

$$= (2{,}086)(0{,}564) = 1{,}18$$

- Sai khác giữa hai mức tương tác B×C:

$$Sd_{B\times C} = \sqrt{\frac{2MS_{e(d)}}{ra}} = \sqrt{\frac{(2)(0{,}421)}{(3)(3)}} = 0{,}306$$

$$LSD_{0,05} = t_{0,05}Sd_{B\times C} = (2,028)(0,306) = 0,62$$

$$LSD_{0,01} = t_{0,01}Sd_{B\times C} = (2,719)(0,306) = 0,83$$

- Cách tính LSD_α giữa các nghiệm thức và phân hạng được thực hiện như đã biết.

Bảng 7.46 : Phân hạng năng suất theo A (mức đạm), B (giống) và nghiệm thức ABC

(B)\(C)	Đạm (A)						TB (B)
	A_1		A_2		A_3		
	C_1	C_2	C_1	C_2	C_1	C_2	
B1	$3,57^{lm-o}$	$2,79^{on}$	$5,13^{fg-l}$	$4,98^{gh-l}$	$7,55^{abc}$	$6,92^{ab-e}$	$5,16^{AB}$
B2	$4,94^{gh-l}$	$4,50^{ij-m}$	$6,71^{ab-f}$	$5,51^{ef-j}$	$7,21^{ab-d}$	$6,61^{ab-f}$	$5,91^{A}$
B3	$4,25^{jk-n}$	$4,34^{jk-m}$	$6,12^{bc-h}$	$6,23^{bc-g}$	$7,87^{a}$	$7,72^{ab}$	$6,09^{A}$
B4	$4,06^{jk-o}$	$5,17^{fg-k}$	$5,56^{ef-j}$	$5,55^{ef-j}$	$7,10^{ab-d}$	$7,88^{a}$	$5,89^{A}$
B5	$4,10^{jk-n}$	$3,73^{kl-o}$	$5,63^{de-j}$	$4,78^{gh-m}$	$6,01^{cd-i}$	$6,02^{cd-i}$	$5,05^{AB}$
B6	$3,21^{omn}$	$4,51^{ij-m}$	$3,72^{kl-o}$	$4,62^{hi-m}$	$2,49^{o}$	$6,32^{ab-g}$	$4,15^{B}$
TB A	$4,10^{C}$		$5,38^{B}$		$6,64^{A}$		
TB C	5,29			5,45			

Ghi chú: *Phân hạng giữa các mức A, giữa các mức B (chữ hoa) và giữa các nghiệm thức (chữ thương) theo $LSD_{0,05}$*

Bảng 7.47 : Kết quả phân hạng năng suất AB và BC theo LSD_α

B	A			C	
	A_1	A_2	A_3	C_1	C_2
B_1	$3,18^{j}$	$5,06^{d-h}$	$7,24^{ab}$	$5,42^{B}$	$4,90^{B}$
B_2	$4,72^{e-j}$	$6,11^{b-d}$	$6,91^{abc}$	$6,29^{A}$	$5,54^{AB}$
B_3	$4,29^{g-j}$	$6,18^{b-d}$	$7,79^{a}$	$6,08^{A}$	$6,10^{A}$
B_4	$4,62^{e-j}$	$5,55^{de}$	$7,49^{a}$	$5,57^{AB}$	$6,20^{A}$
B_5	$3,92^{ij}$	$5,20^{d-g}$	$6,01^{cd}$	$5,25^{B}$	$4,84^{B}$
B_6	$3,86^{ij}$	$4,17^{g-j}$	$4,41^{e-i}$	$3,14^{C}$	$5,15^{B}$

Ghi chú: *Phân hạng AB theo $LSD_{0,05}$ và BC theo $LSD_{0,01}$*

6. Hệ số biến động:

$$\mathrm{CV}(a)=\frac{\sqrt{\mathrm{MS}_{\mathrm{e}(a)}}}{m}\times 100=\frac{\sqrt{1,582}}{5,372}\times 100=23,5\%$$

$$\mathrm{CV}(b)=\frac{\sqrt{\mathrm{MS}_{\mathrm{e}(b)}}}{m}\times 100=\frac{\sqrt{2,673}}{5,372}\times 100=30,4\%$$

$$\mathrm{CV}(c)=\frac{\sqrt{\mathrm{MS}_{\mathrm{e}(c)}}}{m}\times 100=\frac{\sqrt{0,954}}{5,372}\times 100=18,2\%$$

$$\mathrm{CV}(d)=\frac{\sqrt{\mathrm{MS}_{\mathrm{e}(d)}}}{m}\times 100=\frac{\sqrt{0,421}}{5,372}\times 100=12,1\%$$

Chương 8

XỬ LÝ SỐ LIỆU NGHI NGỜ, CHUYỂN ĐỔI SỐ LIỆU VÀ LÀM VIỆC VỚI EXCEL

8.1. XỬ LÝ SỐ LIỆU NGHI NGỜ

Trong quá trình thực hiện thí nghiệm có thể có một số số liệu thu thập không hoàn toàn do yếu tố thí nghiệm chi phối mà bị sai lệch bởi sự can thiệp của các yếu tố bên ngoài, làm cho kết quả phân tích phương sai sai lệch, trong nhiều trường hợp kết luận rút ra từ thí nghiệm khác hẳn với thực tế.

Yếu tố ảnh hưởng có thể là do đất đai (tốt, xấu cục bộ), do động vật, con người (làm thiệt hại, mất mát) và do những sai sót trong thu hoạch thí nghiệm gây nên. Khi biết chắc chắn số liệu của ô thí nghiệm bị ảnh hưởng (gọi chung là số liệu nghi ngờ) thì phải loại bỏ số liệu ô đó ra khỏi bảng số liệu và thực hiện một trong hai cách sau đây:

1. Phân tích phương sai cho trường hợp thiếu ô (ô trống).

2. Tính toán và thay số liệu ước tính vào ô bị mất, sau đó phân tích phương sai cho trường hợp có ô ước tính số liệu.

Với các kiểu thí nghiệm khác nhau việc tính toán và phân tích phương sai khác nhau.

Sau đây là cách xử lý cho hai kiểu thí nghiệm thường gặp.

8.1.1. Kiểu khối đầy đủ ngẫu nhiên (RCBD)

Giả sử trong bảng 6.5, chương 6 số liệu 30,4 của nghiệm thức 18 lần lặp lại II là số liệu nghi ngờ.

Bảng 8.1: Năng suất các giống bông bố mẹ và con lai F_1

TT	Nghiệm thức	Năng suất (tạ/ha, x_{ij})			Tổng NT (T)	Trung bình
		I	II	III		
1	1. C92-52	30,3	19,3	22,2	71,8	23,9
2	2. C118A	12,8	15,9	12,5	41,2	13,7
3	3. S02-13	25,7	28,0	27,1	80,8	26,9
4	4. TM1	24,5	22,1	18,5	65,1	21,7
5	5. NH04-2	16,7	17,0	15,2	48,9	16,3
6	6. 1354	25,1	19,7	15,8	60,5	20,2
7	1 × 2	25,1	26,6	27,0	78,7	26,2
8	1 × 3	27,5	32,1	32,5	92,1	30,7
9	1 × 4	24,7	30,1	29,5	84,3	28,1
10	1 × 5	23,6	31,5	33,7	88,8	29,6
11	1 × 6	24,1	30,7	26,7	81,5	27,2
12	2 × 3	26,1	28,4	24,0	78,4	26,1
13	2 × 4	20,5	25,9	27,6	74,0	24,7
14	2 × 5	23,3	22,5	18,7	64,5	21,5
15	2 × 6	25,7	30,5	26,8	83,0	27,7
16	3 × 4	35,3	30,3	28,8	94,5	31,5
17	3 × 5	22,5	36,8	21,9	81,2	27,1
18	3 × 6	26,3	(-)	25,5	(51,8)	(25,9)
19	4 × 5	25,7	29,6	18,5	73,8	24,6
20	4 × 6	24,5	27,4	27,8	79,6	26,5
21	5 × 6	20,3	32,8	20,2	73,3	24,4
Tổng		510,3	(537,2)	500,5	1.548,0	
Trung bình chung, *m*						24,97

8.1.1.1. Cách 1: Phân tích phương sai thiếu số liệu nghiệm thức 18 lần lặp lại II.

1. Tính tổng của các nghiệm thức, các lần lặp lại, tổng toàn thí nghiệm và trung bình của các nghiệm thức khi bỏ trống số liệu nghiệm thức 18 lần lặp lại II. Kết quả tính được ghi trong bảng 8.1.

2. Tính CF, tổng bình phương tổng số, tổng bình phương các nguồn, phương sai nghiệm thức và lập bảng phân tích phương sai.

$$CF = \left(\sum_{1}^{n} x_{ij}\right)^2 / n$$

$$= (1.548,0)^2 : 62 = 38.650,065$$

$$SS_T = \sum_{1}^{n} x_{ij}^2 - CF = 1.798,435$$

$$SS_r = \left[\frac{(510,4)^2}{21} + \frac{(537,2)^2}{20} + \frac{(500,5)^2}{21}\right]$$

$$- 38.650,065 = 108,001$$

$$SS_t = \frac{\sum_{i=1}^{t} T_i^2}{r} - CF$$

$$= \left[\frac{(71,8)^2 + ... + (81,2)^2}{3} + \frac{(51,8)^2}{2}\right.$$

$$\left.+ \frac{(65,1)^2 + ... + (60,5)^2}{3}\right] - 38.650,065$$

$$= 1.143,855$$

$$SS_e = SS_T - SS_t - SS_r = 546,579$$

Bảng 8.2: Kết quả phân tích phương sai

Nguồn	DF	SS	MS	F_{TN}	Prob.
Lần lặp lại	2	108,001	54,000	3,85	
Nghiệm thức	20	1.143,855	57,193	4,08	0,0001
Sai số	39	546,579	14,015		
Tổng số	61	1.798,435			

P < 0,01 khẳng định rằng giữa các nghiệm thức có sự khác nhau rất rõ ràng.

3. Tính sai số thí nghiệm và so sánh các nghiệm thức

Sai số trung bình: $s_{\bar{x}} = \sqrt{\dfrac{MS_e}{r}} = \sqrt{\dfrac{14,015}{3}}$

$= 2,16$

Sai số tương đối: $s_{\bar{x}}(\%) = \dfrac{s_{\bar{x}}}{m} \times 100 = \dfrac{2,16}{24,97} \times 100$

$= 8,7\%$

Hệ số biến động sai số: $CV(\%) = \dfrac{\sqrt{14,015}}{24,97} \times 100$

$= 15,0\%$

Sai khác giữa hai nghiệm thức đủ số liệu:

$$Sd = \sqrt{\frac{2MS_e}{r}} = \sqrt{\frac{2 \times 14,015}{3}} = 3,06$$

Sai khác giữa hai nghiệm thức 15 với các nghiệm thức khác:

$$Sd' = \sqrt{MS_e \frac{r_{(15)} + r}{r_{(15)} r}} = \sqrt{\frac{(14,015)(5)}{6}} = 3,42$$

So sánh giữa các nghiệm thức bình thường:

$$LSD_{0,05} = t_{0,05}Sd = 2,023 \times 3,06 = 6,2$$

$$LSD_{0,01} = t_{0,01}Sd' = 2,808 \times 3,06 = 8,3$$

So sánh giữa các nghiệm thức bình thường với nghiệm thức mất số liệu:

$$LSD_{0,05} = t_{0,05}Sd' = 2,023 \times 3,42 = 6,9$$

$$LSD_{0,01} = t_{0,01}Sd' = 2,808 \times 3,42 = 9,2$$

8.1.1.2. Cách 2: Tính năng suất ước tính ô bị mất và phân tích phương sai.

Trường hợp mất một số liệu

1. Tính năng suất cho ô bị mất và thay vào ô trống, tính tổng của các nghiệm thức, các lần lặp lại, tổng toàn thí nghiệm và trung bình của các nghiệm thức kể cả số liệu mới của nghiệm thức 18 lần lặp lại II.

Ước tính giá trị ô bị mất theo công thức:

$$X = \frac{rB' + tT' - G'}{(r-1)(t-1)}$$

Trong đó: X là số liệu ước tính của ô bị mất
t là số nghiệm thức
r là số lần lặp lại
B′ là tổng lần lặp lại có ô bị mất
T′ là tổng nghiệm thức có ô bị mất
G′ là tổng toàn thí nghiệm.

Thay các giá trị vào công thức ta có :

$$X = \frac{(3)(537,2)+(21)(51,8)-(1.548,0)}{(3-1)(21-1)} = 28,8$$

và bảng số liệu mới sẽ là:

Bảng 8.3: Năng suất các giống bông bố mẹ và con lai

TT	Nghiệm thức	Năng suất (tạ/ha, x_{ij})			Tổng NT (T)	Trung bình
		I	II	III		
1	1. C92-52	30,3	19,3	22,2	71,8	23,9
2	2. C118A	12,8	15,9	12,5	41,2	13,7
3	3. S02-13	25,7	28,0	27,1	80,8	26,9
4	4. TM1	24,5	22,1	18,5	65,1	21,7
5	5. NH04-2	16,7	17,0	15,2	48,9	16,3
6	6. 1354	25,1	19,7	15,8	60,5	20,2
7	1 × 2	25,1	26,6	27,0	78,7	26,2
8	1 × 3	27,5	32,1	32,5	92,1	30,7
9	1 × 4	24,7	30,1	29,5	84,3	28,1
10	1 × 5	23,6	31,5	33,7	88,8	29,6
11	1 × 6	24,1	30,7	26,7	81,5	27,2
12	2 × 3	26,1	28,4	24,0	78,4	26,1
13	2 × 4	20,5	25,9	27,6	74,0	24,7
14	2 × 5	23,3	22,5	18,7	64,5	21,5
15	2 × 6	25,7	30,5	26,8	83,0	27,7
16	3 × 4	35,3	30,3	28,8	94,5	31,5
17	3 × 5	22,5	36,8	21,9	81,2	27,1
18	3 × 6	26,3	(28,8)	25,5	(80,6)	(27,1)
19	4 × 5	25,7	29,6	18,5	73,8	24,6
20	4 × 6	24,5	27,4	27,8	79,6	26,5
21	5 × 6	20,3	32,8	20,2	73,3	24,4
Tổng		510,3	(566,0)	500,5	1.548,0	
Trung bình chung, m						24,97

2. Tính CF, tổng bình phương tổng số, tổng bình phương các nguồn, phương sai nghiệm thức và lập bảng phân tích phương sai.

Giá trị CF trong trường hợp này bằng tổng của CF tính theo cách bình thường (toàn bảng) cộng CF bổ sung cho ô có số liệu ước tính (CF'):

Giá trị CF toàn bảng = $\mathrm{CF} = \left(\sum_{1}^{n} \mathrm{X}_{ij}\right)^2 / n$ = 39.465,051

$$\mathrm{CF}' = \frac{\left[\mathrm{B}' - (t-1)\mathrm{X}\right]^2}{t(t-1)} = \frac{\left[537,2 - (20)(28,8)\right]^2}{(21)(20)} = 3,584$$

Tổng CF = 39.468,636

Giá trị CF này sử dụng để tính tổng bình phương tổng số và tổng bình phương nghiệm thức, không sử dụng trong tính cho tổng bình phương lần lặp lại.

Bảng 8.4: Kết quả phân tích phương sai

Nguồn	DF	SS	MS	F_{TN}	Prob.
Lần lặp lại	2	118,870	59,435	4,28	
Ng. thức	20	1.149,118	57,456	4,14	0,0001
Sai số	39	541,317	13,880		
Tổng số	61	1.809,304			

3. Tính sai số thí nghiệm và so sánh các nghiệm thức

Sai số thí nghiệm, hệ số biến động được tính theo cách bình thường.

Sai khác giữa hai nghiệm thức được tính theo công thức:

$$\mathrm{Sd}' = \sqrt{\mathrm{MS_e}\left[\frac{2}{r} + \frac{t}{r(r-1)(t-1)}\right]}$$

$$= \sqrt{13,880\left[\frac{2}{3} + \frac{21}{(3)(2)(20)}\right]} = 3,42$$

$$LSD_{0,05} = t_{0,05}\,Sd' = 2,023 \times 3,42 = 6,9$$
$$LSD_{0,01} = t_{0,01}\,Sd' = 2,808 \times 3,42 = 9,2$$

Trường hợp mất hơn một số liệu

Cũng bảng 6.5, chương 6 nhưng mất ba số liệu: nghiệm thức 3 ở lần lặp III, nghiệm thức 15 ở lần lặp II và nghiệm thức 18 ở lần lặp II. Số liệu còn lại như sau:

Bảng 8.5: Năng suất các giống bông bố mẹ và con lai

TT	Nghiệm thức	Năng suất (tạ/ha, x_{ij})			Tổng NT (T)	Trung bình
		I	II	III		
1	1. C92-52	30,3	19,3	22,2	71,8	23,9
2	2. C118A	12,8	15,9	12,5	41,2	13,7
3	3. S02-13	25,7	28,0	(-)	53,7	26,9
4	4. TM1	24,5	22,1	18,5	65,1	21,7
5	5. NH04-2	16,7	17,0	15,2	48,9	16,3
6	6. 1354	25,1	19,7	15,8	60,6	20,2
7	1 × 2	25,1	26,6	27,0	78,7	26,2
...	...	...	...	...	...	...
14	2 × 5	23,3	22,5	18,7	64,5	21,5
15	2 × 6	25,7	(-)	26,8	52,5	26,3
16	3 × 4	35,3	30,3	28,8	94,5	31,5
17	3 × 5	22,5	36,8	21,9	81,2	27,1
18	3 × 6	26,3	(-)	25,5	51,8	25,9
19	4 × 5	25,7	29,6	18,5	73,8	24,6
20	4 × 6	24,5	27,4	27,8	79,6	26,5
21	5 × 6	20,3	32,8	20,2	73,3	24,4

Các bước tính toán như sau:

1. Thành lập bảng 8.6 để ước tính các ô bị mất

Bảng 8.6: Ước tính các ô bị mất

Nghiệm thức (NT)	Năng suất (tạ/ha) I	II	III	ΣNT (T)	TB của các NT 3	15	18
3	25,7	28,0	-	53,7	26,9		
15	25,7	-	26,8	52,5		26,3	
18	26,3	-	25,5	51,8			25,9
Trung bình NT không mất	24,0	26,6	23,4		25,3	23,7	23,7
Hiệu quả NT					1,6	2,6	2,2
			Năng suất ước tính ô mất				
3			25,0				
15		29,2					
18		28,8					

Ở bảng 8.6, các số 24,0; 26,6 và 23,4 là trung bình các nghiệm thức không mất ô tính theo từng lần lặp lại, còn các số 25,3 là trung bình các nghiệm thức không mất ô của các lần lặp lại mà nghiệm thức 3 không mất ô, 23,7 là trung bình các nghiệm thức không mất ô của các lần lặp lại mà nghiệm thức 15 không mất ô và 23,7 là trung bình các nghiệm thức không mất ô của các lần lặp lại mà nghiệm thức 18 không mất ô. Cụ thể:

25,3 = (24,0 + 26,6)/2 (cho nghiệm thức 3)

23,7 = (24,0 + 23,4)/2 (cho nghiệm thức 15 và 18)

Hiệu quả nghiệm thức 3 = 26,9 – 25,3 = 1,6

Hiệu quả nghiệm thức 15 = 26,3 – 23,7 = 2,6

Hiệu quả nghiệm thức 18 = 25,9 – 23,7 = 2,2

Và: Ước tính năng suất ô mất của nghiệm thức 3 là 23,4

+ 1,6 = 25,0, của nghiệm thức 15 là 26,6 + 2,6 = 29,2 và của nghiệm thức 18 là 26,6 + 2,2 = 28,8.

2. Thay các giá trị ước tính vào bảng số liệu và phân tích phương sai theo cách thông thường. Trong bảng phân tích độ tự do của sai số (df_e) bằng $(r - 1)(t - 1)$ – số ô mất. Ở ví dụ này $df_e = 2 \times 20 - 3 = 37$. Kết quả phân tích phương sai như sau:

Bảng 8.7: Kết quả phân tích phương sai

Nguồn	DF	SS	MS	F_{TN}	Prob.
Lần lặp lại	2	118,612	59,306	4,08	
Ng. thức	20	1.139,244	56,962	3,92	0,0002
Sai số	37	537,188	14,519		
Tổng số	59	1.795,044			

3. Sai số thí nghiệm và so sánh các nghiệm thức

- Sai số chung của thí nghiệm:

$$s_{\bar{x}} = \sqrt{\frac{MS_e}{\left(\sum r_i\right)/t}}$$

$$= \sqrt{\frac{14,519}{(18 \times 3 + 3 \times 2)/21}}$$

$$= 2,25$$

(18 giống có 3 lần lặp lại và 3 giống có 2 lần lặp lại)

- Sai khác giữa hai nghiệm thức đủ lần lặp lại:

$$Sd' = \sqrt{\frac{2MS_e}{r}} = \sqrt{\frac{(2)(14,519)}{3}} = 3,11$$

$$LSD_{0,05} = t_{0,05}\,Sd' = 2{,}026 \times 3{,}11 = 6{,}3$$

$$LSD_{0,01} = t_{0,01}\,Sd' = 2{,}715 \times 3{,}11 = 8{,}4$$

- Sai khác giữa hai nghiệm thức đủ lần lặp lại và nghiệm thức mất số liệu:

$$Sd'' = \sqrt{MS_e \frac{r + r_1}{r r_1}}$$

$$= \sqrt{(14{,}519)\frac{(3+2)}{(3)(2)}} = 3{,}48$$

(r_1 là số lần lặp lại cùa nghiệm thức mất số liệu)

$$LSD_{0,05} = t_{0,05}\,Sd' = 2{,}026 \times 3{,}48 = 7{,}0$$

$$LSD_{0,01} = t_{0,01}\,Sd' = 2{,}715 \times 3{,}48 = 9{,}4$$

Sai khác giữa hai nghiệm thức mất số liệu:

$$Sd''' = \sqrt{MS_e \frac{r_1 + r_2}{r_1 r_2}} = \sqrt{(14{,}519)\frac{(2+2)}{(2)(2)}} = 3{,}81$$

(ở ví dụ này các nghiệm thức đều mất một ô nên $r_1 = r_2$)

$$LSD_{0,05} = t_{0,05}\,Sd' = 2{,}026 \times 3{,}81 = 7{,}7$$

$$LSD_{0,01} = t_{0,01}\,Sd' = 2{,}715 \times 3{,}81 = 10{,}3$$

8.1.2. Kiểu ô vuông La tinh (Latin Square)

Với kiểu ô vuông La tinh, có thể phân tích phương sai khi có ô trống như cách 1 đã nêu với kiểu RCBD. Ở đây, xin đề cập đến việc ước tính giá trị ô bị mất và phương pháp phân tích phương sai.

Giả sử rằng bảng 6.7, chương 6, ô ở hàng 2, cột 2 của nghiệm thức D bị loại bỏ, các số liệu thí nghiệm được ghi ở bảng 8.8. Để phân tích phương sai cần thực hiện các bước sau:

Bảng 8.8: Năng suất 4 giống bông lai (tạ/ha)

Hàng	Cột 1	Cột 2	Cột 3	Cột 4	Σ (R_i)	Σ (T_i)	TB_t
1	24,6D	31,5B	27,3C	25,2A	108,6	103,0A	25,8
2	30,5B	-	23,1	28,7C	82,3	122,7B	30,7
3	25,4C	28,0A	24,9D	31,4B	109,7	105,0C	26,3
4	26,7A	23,6C	29,3B	23,5D	103,1	73,0D	24,3
Σ(C_j)	107,2	83,1	104,6	108,8			
ΣΣ					403,7		

1. Tính năng suất cho ô bị mất và thay vào ô trống

Năng suất ô bị mất được tính theo công thức:

$$X = \frac{t\left(R' + C' + T'\right) - 2G'}{(t-1)(t-2)}$$

Trong đó: X là số liệu ước tính của ô bị mất
t là số nghiệm thức
r là số lần lặp lại
R' là tổng hàng có ô bị mất
C' là tổng cột có ô bị mất
T' là tổng nghiệm thức có ô bị mất
G' là tổng toàn thí nghiệm.

Thay số liệu bảng 8.8 vào công thức ta có:

$$X = \frac{(4)(82,3 + 83,1 + 73,0) - (2)(403,7)}{(3)(2)} = 24,4$$

2. Thay số liệu ước tính vào ô trống, tính tổng hàng, tổng cột, tổng các nghiệm thức và tổng toàn thí nghiệm kể cả số liệu mới của nghiệm thức D, sau đó phân tích phương sai. Giá trị CF trong trường hợp này bằng tổng của CF tính theo cách bình thường cộng CF bổ sung cho ô có số liệu ước tính (CF'):

$$CF = \left(\sum_{1}^{n} x_{ij}\right)^2 / n = 11.454,351$$

$$CF' = \left\{\frac{\left[G' - R' - C' - (t-1)T'\right]}{\left[(t-1)(t-2)\right]}\right\}^2$$

$$= \left\{\frac{\left[403,7 - 82,3 - 83,1 - (3)(73,0)\right]}{\left[(3)(2)\right]}\right\}^2$$

$$= 10,347$$

$$\text{CF tổng số} = 11.454,351 + 10,347$$

$$= 11.464,698$$

Giá trị CF tổng số được sử dụng để tính tổng bình phương nghiệm thức và tổng bình phương tổng só. Kết quả phân tích phương được ghi ở bảng 8.9 sau:

Bảng 8.9: Kết quả phân tích phương sai

Nguồn	DF	SS	MS	F_{TN}	Prob.
Hàng	3	6,287	2,096		
Cột	3	2,322	0,774		
Nghiệm thức	3	79,315	26,438	5,57	0,0474
Sai số	6	23,749	4,750		
Tổng số	15	111,672			

3. Sai khác giữa các nghiệm thức

- Sai khác giữa hai nghiệm thức bình thường:

$$Sd = \sqrt{\frac{2MS_e}{r}}$$

- Sai khác giữa nghiệm thức mất ô với nghiệm thức bình thường:

$$Sd' = \sqrt{MS_e\left[\frac{2}{t} + \frac{1}{(t-1)(t-2)}\right]}$$

Cho ví dụ này: $Sd' = \sqrt{(4,750)\left[\frac{2}{4} + \frac{1}{6}\right]} = 1,78$

8.2. CHUYỂN ĐỔI SỐ LIỆU

8.2.1. Nguyên tắc chuyển đổi

Trong nhiều trường hợp, các số liệu quan sát khác xa nhau đến mức làm cho sai số $s_{\bar{x}}$ và độ lệch chuẩn sai số $\sqrt{MS_e}$ quá lớn dẫn đến hệ số biến động sai số CV,% quá cao, trong nhiều trường hợp cao hơn 100%. Hiện tượng này thường thấy ở các bảng số liệu là các số đếm, chỉ số, hệ số, tỷ lệ biểu thị bằng số thập phân hay số %, không phải là đơn vị đo lường thông dụng. Khi đó số liệu phải được chuyển đổi cho phù hợp. Với kỹ thuật này, số liệu gốc được chuyển về một cấp độ khác để tạo thành một bộ số liệu mới. Việc chuyển đổi số liệu không áp dụng phổ biến, nó chỉ áp dụng cho các trường hợp đặc biệt phụ thuộc vào quan hệ giữa hệ số biến động và số trung bình. Thông thường người ta áp dụng ba cách chuyển đổi số liệu: logarit, căn bậc hai và arc sin các số liệu. Tuy nhiên việc chuyển đổi số liệu đơn giản chỉ là thuật xử lý số liệu, không làm thay đổi kết quả nghiên cứu. Sau khi kết thúc việc xử lý, giá trị trung bình của các nghiệm thức theo đơn vị mới phải được chuyển về giá trị theo đơn vị gốc và các ký hiệu phân hạng sẽ được ghi trên giá trị số liệu gốc.

8.2.1.1. Chuyển đổi bằng logarit các số liệu gốc (Logarit hóa)

Thực hiện việc logarit hóa cho các số liệu mà kết quả quan sát được với các giá trị quá lớn hay quá nhỏ, tức là các số liệu có độ phân tán rộng làm cho độ lệch chuẩn quá lớn so với số trung bình. Trường hợp này thường thấy do các tác động vào đặc trưng (chỉ tiêu) theo dõi một cách bất thường. Mật số sâu hại trên lô, số ổ trứng trên cây (không có, có quá ít và có rất nhiều) là những ví dụ điển hình.

Để chuyển đổi số liệu, đơn giản là logarit hóa toàn bộ số liệu của mỗi ô và các nghiệm thức. Khi đó mức độ phân tán của các số liệu giảm xuống rất nhiều so với số liệu gốc và kết quả phân tích phương sai sẽ hợp lý. Điều đó cho thấy hiệu quả của phép logarit hóa. Tính không hợp lý có thể thấy khi xử lý với số liệu gốc là độ lệch chuẩn bằng hoặc cao hơn số trung bình – nói lại.

Nếu các số liệu gốc x_i bao gồm các số có giá trị nhỏ hơn 10, khi logarit các số liệu này phải logarit cho $(x_i + 1)$. Khi phân tích phương sai được thực hiện với các số liệu chuyển đổi và so sánh, xếp hạng cho các giá trị trung bình chuyển đổi, tuy nhiên các ký hiệu phân hạng phải được ghi trên giá trị số liệu gốc (nói lại). Thông thường, việc xếp hạng này hầu như không thay đổi so với việc xếp hạng với số liệu gốc. Trong trường hợp hiếm hoi, ở cặp so sánh nào đó có thay đổi thì kết quả đó vẫn được thừa nhận. Có thể chuyển từ giá trị trung bình, LSD_α về đơn vị số liệu gốc bằng cách tính đối logarit (antilog).

8.2.1.2. Chuyển đổi bằng căn bậc hai số liệu gốc (Căn bậc hai hóa)

Việc chuyển đổi số liệu về căn bậc hai được thực hiện cho các số liệu là các số nguyên nhỏ, các số liệu thu thập từ những hiện tượng hiếm thấy, ví dụ, số cây bị hại (do sâu), số côn trùng bị sa bẩy hoặc số cây cỏ dại trong lô. Với các số liệu này, sự biến động có khuynh hướng tỷ lệ thuận với số trung bình.

Chuyển đổi căn bậc hai cũng được áp dụng cho các số liệu % mà các số liệu nằm trong khoảng 0 đến 30% hoặc là từ 70 đến 100%.

Nếu như hầu hết các số liệu gốc x_i nhỏ hơn 10, đặc biệt là có số 0, việc sử dụng $(x_i + 0,5)^{(1/2)}$ là phù hơn $(x_i)^{(1/2)}$. Việc phân tích phương sai, phân hạng được tiến hành với các số đã được chuyển đổi, sau đó chuyển ngược về đơn vị của số liệu gốc.

8.2.1.3. Chuyển đổi bằng arc sin các số liệu (Arc sin hóa)

Arc sin hóa được áp dụng cho các số liệu tỷ lệ, các số đếm, các số biểu thị bằng số thập phân hoặc số %. Số % ở đây là sự so sánh từ các số đếm, ví dụ, tỷ lệ sâu hại (%), tỷ lệ cây bị bệnh (%). Chuyển đổi arc sin không thực hiện cho các số % không có nguồn gốc từ số đếm như : tỷ lệ xơ bông (%), hàm lượng protein (%), hàm lượng hydrat-cacbon (%).

Giá trị arc sin được tra trong bảng chuyển đổi arc sin (phụ lục 8). Nếu có số 0 thì thay thế bằng 1/4*n* và 100 được thay bằng (100 – 1/4*n*) trước khi arc sin hóa, trong đó *n* là mẫu số của số tỷ lệ được tính ra % (ví dụ, 8% được tính tình 56/70 thì *n* = 70).

Không phải tất cả các bảng số liệu % cần phải được chuyển đổi. Các quy tắc sau đây có thể vận dụng để lựa chọn trong việc chuyển đổi số liệu % được tính từ số đếm:

1. Với bảng có các số nằm trong khoảng từ 30 đến 70% không cần chuyển đổi.

2. Với bảng có số liệu từ 0 đến 30% hoặc là từ 70 đến 100% có thể chuyển đổi bằng căn bậc hai.

3. Các số liệu không thuộc hai trường hợp trên thì nên chuyển đổi bằng arc sin.

8.2.2. Một vài ví dụ

Ví dụ 1: Trong một thí nghiệm hiệu lực trừ cỏ của bốn loại thuốc bố trí theo kiểu khối đầy đủ ngẫu nhiên, bốn lần lặp lại, kết quả tính toán số cỏ còn sống/m^2 được ghi ở bảng 8.10 (phần trên). Hãy phân tích kết quả.

Bảng 8.10: Số cỏ còn sống

Nghiệm thức	Lần lặp lại				Tổng	Số cỏ trung bình/m^2	
	I	II	III	IV		Thực tế	Tính toán[@]
			Số liệu gốc, X (số cỏ/m^2)				
1 (ĐC)	169	132	280	105	686	172	
2	210	172	358	125	865	216	
3	160	83	103	65	411	103	
4	42	40	84	28	194	48	
Cách 1:			Số liệu chuyển đổi, X_1 = lgX				
1 (ĐC)	2,23	2,12	2,45	2,02	8,82	2,21	162^A
2	2,32	2,24	2,55	2,10	9,21	2,30	200^A
3	2,20	1,92	2,01	1,81	7,95	1,99	98^B
4	1,62	1,60	1,92	1,45	6,60	1,65	45^C

Cách 2:	Số liệu chuyển đổi, $X_2 = X^{(1/2)}$						
1 (ĐC)	13,0	11,5	16,7	10,2	51,4	12,9	166^{A}
2	14,5	13,1	18,9	11,2	57,7	14,4	207^{A}
3	12,6	9,1	10,1	8,1	39,9	10,0	100^{B}
4	6,5	6,3	9,2	5,3	27,3	6,8	46^{C}
Tổng R	46,6	40,0	54,9	34,8	176,3		

@: Phân hạng theo $LSD_{0,05}$

Giải:

Khoảng biến thiên về số cỏ rất lớn (R = 280 – 28 = 252) cho thấy cần phải chuyển đổi số liệu trước khi phân tích phương sai. Ở đây có thể chuyển các số liệu X bằng một trong hai cách logaritX hay $\sqrt{X}$. Bảng 8.10 trình bày các số liệu gốc, các số liệu sau khi chuyển đổi theo hai cách và kết quả so sánh giữa các nghiệm thức. Kết quả phân tích phương sai theo số liệu chuyển đổi được ghi ở bảng 8.11.

Bảng 8.11: Kết quả phân tích phương sai

Nguồn	DF	SS	MS	F_{TN}	Prob.
Cách 1:	Theo số liệu X_1 = lgX				
Lần lặp lại	3	0,336	0,112		
Nghiệm thức	3	1,005	0,335	39,32	0,0001
Sai số	9	0,077	0,009		
Tổng số	15	1,418			
	CV(%) = 4,5				
Cách 2:	Theo số liệu $X_2 = X^{(1/2)}$				
Lần lặp lại	3	56,547	18,849		
Nghiệm thức	3	134,532	44,844	21,39	0,0002
Sai số	9	18,865	2,096		
Tổng số	15	209,944			
	CV(%) = 13,1				

Với cách logarit hóa:

$$LSD_{0,05} = t_{0,05}Sd = 2,262 \times 0,065 = 0,15$$

$$LSD_{0,01} = t_{0,01}\ Sd = 3,250 \times 0,065 = 0,21$$

Với cách căn bậc hai X:

$$LSD_{0,05} = t_{0,05}Sd = 2,262 \times 1,02 = 2,3$$

$$LSD_{0,01} = t_{0,01}\ Sd = 3,250 \times 1,02 = 3,3$$

Kết quả so sánh với đối chứng được ghi ở bảng 8.10. Số liệu trung bình tính toán được chuyển về đơn vị số liệu gốc:

Với cách logarit hóa:

159 = đối lg2,2; 200 = đối lg2,3,

98 = đối lg1,99; 45 = đối lg1,65

Với cách căn bậc hai X:

$165 = (12,9)^2$; $208 = (14,4)^2$,

$100 = (10,0)^2$; $47 = (6,8)^2$.

Ví dụ 2: Để đánh giá năm giống của một loại cây ăn quả trong một thí nghiệm theo kiểu khối đầy đủ ngẫu nhiên, bốn lần lặp lại, người ta đã tiến hành cho điểm theo thang điểm quy định. Kết quả tính toán điểm trung bình cho các giống được ghi ở bảng 8.12 (phần trên). Hãy phân tích kết quả.

Giải:

Do có số liệu có giá trị 0 nên cần phải chuyển đổi số liệu trước khi phân tích phương sai. Ở đây có thể chuyển các số liệu X thành $X_1 = (1 + X)^{(1/2)}$. Sau khi chuyển đổi, kết quả được ghi trong bảng 8.12 (phần dưới).

Bảng 8.12: Điểm trung bình của năm giống

Giống	Lần lặp lại				Tổng	Điểm trung bình	
	I	II	III	IV		Thực tế	Tính toán@
	Số liệu gốc, X (điểm)						
1	0,2	1,0	0,8	0,0	2,0	0,5	
2	2,7	3,2	2,7	1,2	9,8	2,5	
3(ĐC)	3,4	3,1	4,0	3,2	13,7	3,4	
4	4,1	4,0	3,7	3,2	15,0	3,8	
5	4,2	5,0	5,0	4,3	18,5	4,6	
	Số liệu chuyển đổi, $X_1 = (1 + X)^{(1/2)}$						
1	1,10	1,41	1,34	1,00	4,85	1,21	$0,5^D$
2	1,92	2,05	1,92	1,48	7,37	1,84	$2,4^C$
3(ĐC)	2,10	2,02	2,24	2,05	8,41	2,10	$3,4^B$
4	2,26	2,24	2,17	2,05	8,72	2,18	$3,8^B$
5	2,28	2,45	2,45	2,30	9,48	2,37	$4,6^A$
Tổng R	9,66	10,17	10,12	8,88	38,83	$\overline{x} = 1,94$	

@: *Phân hạng theo $LSD_{0,05}$*

Kết quả phân tích phương sai theo số liệu chuyển đổi được ghi ở bảng 8.13.

Bảng 8.13: Kết quả phân tích phương sai

Nguồn	DF	SS	MS	F_{TN}	Prob.
Lần lặp lại	3	0,214	0,071		
Nghiệm thức	4	3,231	0,808	58,17	0,0000
Sai số	12	0,167	0,014		
Tổng số	19	3,611			

$$CV(\%) = \frac{\sqrt{MS_e}}{\overline{x}} \times 100 = \frac{\sqrt{0,014}}{1,94} \times 100 = 6,1$$

$$s_{\overline{x}} = \sqrt{\frac{MS_e}{r}} = \sqrt{\frac{0,014}{4}} = 0,06$$

$$Sd = \sqrt{\frac{2MS_e}{r}} = \sqrt{\frac{(2)(0,014)}{4}} = 0,08$$

$$LSD_{0,05} = t_{0,05} Sd = 2,179 \times 0,08 = 0,18$$

$$LSD_{0,01} = t_{0,01}\ Sd = 3,054 \times 0,08 = 0,25$$

Kết quả so sánh với đối chứng được ghi ở bảng 8.12. Số liệu trung bình tính toán được chuyển về đơn vị số liệu gốc:

$0,5 = (1,21)^2 - 1$; $2,4 = (1,84)^2 - 1$; $3,4 = (2,10)^2 - 1$; $3,8 = (2,18)^2 - 1$, $4,6 = (2,37)^2 - 1$.

Ví dụ 3: Kết quả theo dõi tỷ lệ (%) số bông bị hại do bệnh than (*Ustilago panicimiliacei*) trên tám giống kê (*Setaria italica* – Foxtail millet) trong một thí nghiệm được bố trí theo kiểu khối đầy đủ ngẫu nhiên, ba lần lặp lại được ghi ở bảng 8.14 (phần trên) (theo Dospekhov, 1985). Hãy phân tích kết quả.

Giải:

Theo quy tắc 3 nêu trên cần phải chuyển đổi số liệu thành X_1 = Arc sin$\sqrt{\%}$ trước khi phân tích phương sai. Sau khi chuyển đổi, kết quả được ghi trong bảng 8.14 (phần dưới).

Kết quả phân tích phương sai theo số liệu chuyển đổi được ghi ở bảng 8.15.

Bảng 8.14: Tỷ lệ % bông kê bị hại do bệnh than

Giống	Lần lặp lại			Tổng	Trung bình (%)	
	I	II	III		Thực tế	Tính toán@
	Số liệu gốc, X (%)					
1(ĐC)	64,6	66,7	69,4	200,7	66,9	
2	67,0	64,2	68,0	199,2	66,4	
3	1,0	0,4	0,5	1,9	0,6	
4	12,2	10,6	13,3	36,1	12,0	
5	2,7	1,4	2,4	6,5	2,2	
6	63,8	62,2	59,6	185,6	61,9	
7	1,2	1,0	0,8	3,0	1,0	
8	55,8	55,4	52,8	164,0	54,7	
	Số liệu chuyển đổi, X_1 = Arc sin $\sqrt{\%}$					
1(ĐC)	53,5	54,8	56,4	164,7	54,9	$67,0^A$
2	54,9	53,2	55,6	163,7	54,6	$66,5^A$
3	5,7	3,6	4,0	13,3	4,4	$0,6^F$
4	20,4	19,0	21,4	60,8	20,3	$12,0^D$
5	9,5	6,8	8,9	25,2	8,4	$2,1^E$
6	53,0	52,1	50,5	155,6	51,9	$62,0^B$
7	6,3	5,7	5,1	17,1	5,7	$1,0^F$
8	48,3	48,1	46,6	143,0	47,7	$54,7^C$
Tổng R	251,6	243,3	248,5	743,4	$\bar{x} = 30,98$	

@: Phân hạng theo $LSD_{0,05}$

Bảng 8.15: Kết quả phân tích phương sai

Nguồn	DF	SS	MS	F_{TN}	Prob.
Lần lặp lại	2	4,398	2,199		
Nghiệm thức	7	11434,892	1633,556	1275,09	0,0000
Sai số	14	17,936	1,281		
Tổng số	23	11457,225			

$$CV(\%) = \frac{\sqrt{MS_e}}{\overline{x}} \times 100 = \frac{\sqrt{1,281}}{30,98} \times 100 = 3,7$$

$$s_{\overline{x}} = \sqrt{\frac{MS_e}{r}} = \sqrt{\frac{1,281}{3}} = 0,65$$

$$Sd = \sqrt{\frac{2MS_e}{r}} = \sqrt{\frac{(2)(1,281)}{3}} = 0,92$$

$$LSD_{0,05} = t_{0,05}Sd = 2,145 \times 0,92 = 2,0$$

$$LSD_{0,01} = t_{0,01}\, Sd = 2,977 \times 0,92 = 2,8$$

Kết quả so sánh với đối chứng được ghi ở bảng 8.14. Số liệu trung bình tính toán được chuyển về đơn vị số liệu gốc (tra bảng arc sin$\sqrt{\%}$).

8.3. LÀM VIỆC VỚI EXCEL

Excel là phần mềm đa chức năng, có thể sử dụng để:

- Tính toán và phân tích sô liệu nhờ các hàm toán học.

- Giải các phương trình phức tạp về đại số tuyến tính, tích phân, vi phân và thực hiện việc phân tích tối ưu hóa một cách tự động.

- Lập trình các đặc tính (Property)

- Vẽ đồ thị.

và nhiều tiện ích khác.

Trong xử lý số liệu, có thể sử dụng Excel để tính toán cho tất cả các kiểu thí nghiệm. Cùng một kiểu thí nghiệm, nếu có cùng số nghiệm thức, cách bố trí giống hệt nhau ta chỉ cần thiết lập cách tính toán một lần, sau đó "dán số liệu" của các chỉ tiêu khác nhau cùng thí nghiệm hay thí nghiệm tương tự sẽ nhận được ngay kết quả.

Không thể phủ nhận tốc độ xử lý mau lẹ, có khi chưa đến một giây sau khi nhập số liệu của một số phần mềm chuyên dụng hiện nay, nhưng nếu thiếu hiểu biết về mô hình toán học, nguyên lý tính toán nguồn biến động thì có thể nhận được kết quả sai.

Có khá nhiều tài liệu giới thiệu và hướng dẫn sử dụng Excel. Tài liệu này chỉ giới thiệu một số tiện ích, một số hàm thường dùng trong tính toán số liệu điều tra và số liệu thí nghiệm, cách tra một số giá trị.

8.3.1. Một số tiện ích có thể sử dụng để xử lý số liệu

Chọn số ngẫu nhiên

Phương pháp: Tạo dãy số liệu tổng thể \ Tools\ Data Analysis\ Sampling\ OK\ Input Range (khai vùng tổng thể)\ Number of Samples (ghi dung lượng mẫu chọn)\ Output (khai vùng xuất hiện mẫu chọn)\ OK.

Lập bảng và biểu đồ phân phối thực nghiệm

Quy trình:

- Nạp số liệu vào bảng.
- Chọn Tools\ Data Analysis
- Chọn Histogram:
  - Input Range: Khai khối dữ liệu của mẫu
  - Bin Range: Khai khối dữ liệu cự ly các tổ

• Output Range: Khai vùng xuất hiện kết quả

• Cumulative Percentage: Tính % tần số tích lũy

• Chart Output: Vẽ biểu đồ phân phốitần suất và tích lũy

• OK

Ước lượng điểm số trung bình tổng thể

Phương pháp:

- Lập bảng số liệu

- Dùng hàm = AVERAGE(number1, number2, ...) để tính trung bình $\overline{x}$; hàm = STDEV(number1, number2, ...) để tính độ lệch chuẩn S

Ước lượng lượng khoảng của số trung bình tổng thể

Phương pháp:

- Lập bảng số liệu

- Dùng hàm =AVERAGE(number1, number2, ...) để tính trung bình $\overline{x}$; hàm = STDEV(number1, number2, ...) để tính độ lệch chuẩn S; = 1.96*S/SQRT(n) để giới hạn một phía G

- Kết quả: $G_1 \leq \mu \leq G_2$

Ước lượng phương sai tổng thể

Phương pháp:

- Lập bảng số liệu

- Dùng hàm = STDEV(number1, number2, ...) để tính độ lệch chuẩn S; tra giá trị $\chi^2_{\frac{\alpha}{2}}$ và $\chi^2_{1-\frac{\alpha}{2}}$ theo cách tra ở mục 8.3.2 dưới đây.

- Tính giới hạn trên dưới.

So sánh hai trung bình bằng phép nghiệm T

Trước hết, so sánh hai phương sai bằng phép nghiệm F theo quy trình:

- Nạp số liệu hai mẫu vào bảng
- Chọn ô cho kết quả
- Vào *fx* và chọn hàm F-test
- Chọn Next và khai từng dãy số (array)\ OK

Nếu kết quả > 0,05 thì hai phương sai bằng nhau và $\leq$ 0,05 thì hai phương sai khác nhau nhau.

Dựa vào kết quả này để chọn hàm t-test trong so sánh hai trung bình.

Quy trình: Sử dụng bảng số liệu đã lập

- Chọn Tools\ Data Analysis
- Chọn t-Test: Two-Sampe Assuming Equal Variances hoặc Two-Sampe Assuming Unequal Variances (theo kết quả so sánh phương sai)
- Khai 1 dãy số liệu vào Variable 1 range và dãy thú hai vào Variable 2 range, ghi 0 vào Hypothesized Mean Difference
- Chọn vùng xuất hiện số liệu vào Output range\ OK.

So sánh hai trung bình bằng phép nghiệm U

Nếu dung lượng hai mẫu đều > 30, quy trình kiểm nghiệm giống như trên nhưng vào hàm z-Test.

Kiểm tra sự đồng nhất của nhiều mẫu độc lập bằng tiêu H của Kruskal và Wallis

Quy trình:

- Lập bảng số liệu các mẫu theo các cột và dùng lệnh Sort từ Menu Data để xếp hạng từ nhỏ đến lớn cho từng mẫu vào cột bên phải mỗi mẫu. Nếu trùng hạng thì chia đều hạng cho các số theo lý thuyết đã trình bày ở mục 2.2, chương 3.

- Dùng hàm sum để cộng tổng hạng R_1, R_2, ...R_l cho các mẫu.

- Tính H và so sánh với $\chi^2_{0,05}$

So sánh phương sai bằng phép nghiệm F

Như đã trình bày trong so sánh hai trung bình

Phân tích phương sai thí nghiệm một yếu tố kiểu CRD

Quy trình:

- Lập bảng số liệu:

Nghiệm thức	Kết quả				
1	x_{11}	x_{12}	x_{13}	...	x_{1n}
2	x_{21}	x_{22}	x_{23}	...	x_{2n}
...	...	...	...	...	...
t	x_{t1}	x_{t2}	x_{t3}	...	x_{tn}

- Chọn Tools\ Data Analysis

- Chọn Single Factor

- Khai vùng dữ liệu vào Input Range. Nếu nghiệm thức là hàng thí chọn Row, là cột thì chọn Columns

- Khai vùng xuất hiện kết quả vào Output\ OK

Từ đó tính $s_{\bar{x}}$, CV(%)

LSD_α = tinv(α, df_e)$\sqrt{\dfrac{2MS_e}{r}}$ (nếu lần *r* bằng nhau) hoặc

LSD_α = tinv(α, df_e)$\sqrt{MS_e \dfrac{r_1 + r_2}{r_1 r_2}}$ (nếu lần *r* khác nhau nhau)

Phân tích phương sai thí nghiệm một yếu tố kiểu RCBD

Quy trình:

- Lập bảng số liệu:

Lần lặp lại	Nghiệm thức				
	1	2	3	...	t
1	x_{11}	x_{12}	x_{13}	...	x_{1t}
2	x_{21}	x_{22}	x_{23}	...	x_{2t}
...	...	...	...	...	...
r	x_{r1}	x_{r2}	x_{r3}	...	x_{rt}

- Chọn Tools\ Data Analysis

- Chọn Two-Facror Without Replication

- Khai vùng dữ liệu vào Input Range và đánh dấu vào Label (có thể quét cả số thứ tự nghiệm thức và lần lặp lại)

Từ đó tính $s_{\bar{x}}$, CV(%)

LSD_α = tinv(α, df_e)$\sqrt{\dfrac{2MS_e}{r}}$

Có thể phân tích ANOVA cho các thí nghiệm kiểu ô vuông La tinh, thí nghiệm hai nhân tố, phân tích tương quan hồi quy hai hay nhiều biến số, phân tích quan hệ phi tuyến tính, mô hình hóa và phân tích mô hình tối ưu, phân loại theo khoảng cách D của Mahalanobis, phân tích đường (Path Analysis) và vẽ đồ thị.

8.3.2. Một số hàm trên Excel

Excel có khoảng 300 hàm được chia thành nhiều nhóm. Một số hàm thường dùng:

Các hàm thống kê

- Trung bình mẫu: = AVERAGE(number1, number2, ...)

- Phương sai mẫu: = VAR(number1, number2, ...)

- Độ lệch chuẩn mẫu: STDEV(number1, number2, ...)

- Hiệp phương sai: = COVAR(array1, array 2)

- Trị cực đại: = MAX(number1, number2, ...)

- Trị cự tiểu: = MIN(number1, number2, ...)

- Trung vị: = MEDIAN(number1, number2, ...)

- Trị phổ biến nhất: = MOD(number1, number2, ...)

- Tương quan tuyến tính: = LINEST(Known-y s, Known- x s, const. Stats)

- Hệ số tương quan: = CORREL(array1, array 2, ...)

- Hệ số tương quan bình phương:
 = RSQ(Known-y s, Known-x s)

- Hệ số hồi quy: = SLOPE(Known-y s, Known-x s)

- Hồi quy đa biến: = GRGRESSION(input y Range, input x Range)

- Tiêu chuẩn khi bình phương: = CHITEST(actual_ range, expected_ range)

- Tiêu chuẩn Student, t: = TTEST(variabe 1 Range, variabe2 Range)

- Độ lệch phân bố: = SKEW(number1, number2, ...)

- Phân phốiWeibull : = WEIBULL(x, alpha, beta, cumulative)

Các hàm toán học

1. Hàm tính tổng

- Tổng: = SUM(number1, number2, ...)
- Tổng các tích: = SUMPRODUCT(array1, array 2, ...)
- Tổng các bình phương: = SUMSQ (number1, number2, ...)
- Tổng của bình phương các độ lệch:

  = DEVSQ(number1, number2, ...)
- Tổng các hiệu các bình phương:

  = SUMX2MY2((array-x, array-y)

  Công thức: $\sum\left(x^2 - y^2\right)$
- Tổng các tổng các bình phương:

  = SUMX2PY2((array-x, array-y)

  Công thức: $\sum\left(x^2 + y^2\right)$
- Tổng các bình phương một hiệu:

  = SUMXMY2((array-x, array-y)

  Công thức: $\sum\left(x - y\right)^2$

2. Hàm nhân

- Tích các số = PRODUCT(number1, number2, ...)
- Bội số chung nhỏ nhất: = LCM(number1, number2, ...)

3. Hàm logarit và hàm mũ

- Logarit: = LOG(number, base)

- Logarit cơ số 10: = LOG10(number)

- Logarit tự nhiên (nepe): = LN(number)

- Đưa về e^n: = EXP(number)

4. Một số hàm khác

- Tìm ma trận nghịch đảo: = MINVERSE(array)

- Tích hai ma trận: = MMULT(array1, array 2)

- Chuyển vị ma trận: = TRANSPOSE(cell range)

- Căn bậc hai: = SQRT(number)

- Sin của một góc: = SIN(number)

- Cosin của một góc: = COS(number)

- Tang của một góc: = TAN(number)

8.3.3. Cách tra một số giá trị

- Xác suất sai lầm của F (P, Prob.):

  = FDIST(x, deg_freedom 1, deg_freedom 2)

- Giá trị tới hạn t:

  = TINV(probability, deg_freedom)

- Giá trị tới hạn χ^2:

  = CHIINV(probability, deg_freedom)

- Giá trị tới hạn F:

  = FINV((probability, deg_freedom 1, deg_freedom 2)

Chương 9

TRÌNH BÀY BÁO CÁO KHOA HỌC

9.1. BỐ CỤC CỦA MỘT BÁO CÁO KHOA HỌC

Một báo cáo khoa học gồm ba phần chính:

- Phần khai tập (*Front Matter*).

- Phần chính (*Main Text*).

- Phần phụ lục (*Back Matter*)

9.1.1. Phần khai tập

Phần khai tập gồm:

1) Bìa ngoài, bìa trong của báo cáo: Trình bày theo quy định của cơ quan chủ quản, nhưng về cơ bản có những nội dung sau:

+ Tên cơ quan chủ trì đề tài, tên chương trình dự án hay mã số.

+ Tên đề tài bằng chữ in hoa lớn.

+ Tên chủ nhiệm đề tài (trên bìa ngoài), tên chủ nhiệm đề tài và các thành viên tham gia (trên bìa trong).

+ Địa danh và năm bảo vệ công trình (phía dưới).

2) Trang ghi ơn (nếu có)

3) Lời nói đầu (nếu cần): Trình bày một cách vắn tắt lý do, bối cảnh, ý nghĩa lý thuyết và thực tiễn của đề tài. Có thể ghi lời ghi ơn cuối lời nói đầu.

4) Mục lục: Nếu không có trang ghi ơn, lời nói đầu, mục lục thường được đặt sau bìa trong của báo cáo.

5) Tóm tắt kết quả đề tài (cả tiếng Anh nếu có yêu cầu). Tóm tắt viết ngắn gọn bao gồm tên đề tài, thời gian và địa điểm nghiên cứu, mục tiêu đề tài, tóm lược cách bố trí thí nghiệm/nghiên cứu/điều tra và trình bày kết quả chủ yếu đã đạt được theo mục tiêu đặt ra, không nêu các thảo luận và khuyến nghị.

6) Danh mục các ký hiệu và viết tắt theo thứ tự a, b, c ... tiếng Việt và tiếng Anh (nếu có). Do ngôn ngữ của báo cáo viết bằng tiếng Việt nên phải ghi tên đầy đủ bằng tiếng Việt cho chữ viết tắt tiếng nước ngoài.

7) Danh sách các bảng số liệu, các biểu đồ, đồ thị và hình vẽ, hình chụp.

Tùy quy định của cơ quan chủ quản một số mục trên không nhất thiết phải trình bày.

Các trang trong phần khai tập được đánh số La mã kiểu chữ nhỏ (i, ii, iii, iv, v...).

9.1.2. Phần chính

Phần chính gồm các nội dung sau:

9.1.2.1. Mở đầu

- Lý do nghiên cứu (tại sao nghiên cứu đề tài).

- Mục tiêu nghiên cứu (những vấn đề cần đạt được sau khi thực hiện đề tài): Nêu những mục tiêu cụ thể. Tùy theo yêu cầu có thể nêu mục tiêu tổng quát trước khi nêu các mục tiêu cụ thể.

- Các yêu cầu (tùy mục tiêu cụ thể của đề tài).

- Phạm vi nghiên cứu: Nêu tóm tắt giới hạn về thời gian nghiên cứu, địa điểm nghiên cứu, nội dung và quy mô từng nội dung.

9.1.2.2. Tổng quan tài liệu

- Nêu các kết quả nghiên cứu, những thành tựu đạt được của các tác giả trong và ngoài nước về các vấn đề liên quan đến đề tài nghiên cứu.

- Những hạn chế, tồn tại của các nghiên cứu trước đây (trước khi đề tài được thực hiện).

- Những vấn đề cần phải tiếp tục nghiên cứu (trong đó có vấn đề mà đề tài nghiên cứu).

Cách trích dẫn tài liệu tham khảo:

Khi sử dụng kết quả nghiên cứu của đồng nghiệp phải ghi rõ xuất xứ của tài liệu trích dẫn như là một nguyên tắc bắt buộc. Việc ghi tên tác giả tài liệu tham khảo, tùy cách diễn đạt, có thể nằm sau hay trước lời trích dẫn.

Nếu sau lời trích dẫn, có hai cách ghi: i) ghi số thứ tự trong danh mục tài liệu tham khảo và đặt trong dấu ngoặc vuông, như [25] khi danh mục tài liệu tham khảo có đánh số thứ tự; ii) ghi họ tác giả và năm công bố đặt trong dấu ngoặc đơn như (Mandenhall, 1990) nếu tài liệu bằng tiếng nước ngoài và ghi đầy đủ họ tên tác giả, năm công bố đặt trong dấu ngoặc đơn nếu là tài liệu tiếng Việt như (Vũ Cao Đàm, 2005).

Nếu trước lời trích dẫn, là tài liệu tiếng Việt, tác giả là người Việt phải ghi họ tên tác giả, đặt năm công bố trong dấu ngoặc đơn, ví dụ: Theo Vũ Cao Đàm (2005), ... hoặc Vũ Cao Đàm (2005) cho rằng ... . Tương tự với tác giả nước ngoài thì: Theo Mandenhall (1990),.... hoặc: Mandenhall (1990) đã chứng minh được ....

Công dụng của trích dẫn:

- Trích dẫn để phân tích những vấn đề chưa được nghiên cứu hoặc đã nghiên cứu nhưng chưa có kết luận thống nhất, hoặc chỉ mới nghiên cứu ở điều kiện nước ngoài, để đề xuất vấn đề nghiên cứu mới.

- Trích dẫn để dùng làm luận cứ cho việc chứng minh một luận điểm.

- Trích dẫn để bác bỏ khi phát hiện chỗ sai trong nghiên cứu của đồng nghiệp.

9.1.2.3. Nội dung và phương pháp nghiên cứu

Nội dung nghiên cứu là những vấn đề cần phải nghiên cứu để trả lời những mục tiêu đề tài đặt ra. Để giải quyết một mục tiêu có thể cần một hoặc một số nội dung. Một nội dung có thể cần một hoặc một số thí nghiệm, một hoặc một số cuộc điều tra. Ngược lại, một thí nghiệm, một cuộc điều ta có thể giải quyết một hoặc một số mục tiêu.

Phương pháp nghiên cứu là cách làm rõ các vấn đề trong từng nội dung để trả lời các mục tiêu đề tài đặt ra. Nếu sử dụng phương pháp mới (hoặc ít phổ biến) thì nên trình bày chi tiết đầy đủ và nếu nội dung phương pháp quá dài thì nội dung chi tiết có thể trình bày đầy đủ trong phần phụ lục.

Từ nội dung và phương pháp, có thể biết được một cách đầy đủ rằng những nội dung được chọn và phương pháp nghiên cứu có khả năng để giải quyết tất cả các mục tiêu đề tài không.

9.1.2.4. Kết quả và thảo luận

- Kết quả: Nêu đầy đủ các kết quả đạt được trong từng nội dung, hướng tới giải quyết các mục tiêu của đề tài.

- Thảo luận: Luận giải điều kiện để đạt được các kết quả mong đợi, những yếu tố hạn chế, so sánh phân tích kết quả của đề tài với kết quả của các nghiên cứu trước, làm rõ những điểm mới về lý thuyết và thực tiễn của đề tài.

9.1.2.5. Kết luận và khuyến nghị

- Kết luận: Đúc kết và khẳng định được những kết quả đạt được theo mục tiêu đề tài đặt ra. Kết luận cần viết ngắn gọn, không có lời bàn và bình luận. Chỉ kết luận những vấn đề mà đề tài thực hiện.

- Khuyến nghị: Trong khoa học nên dùng khái niệm “khuyến nghị” thay cho “kiến nghị” và “đề nghị”. Khuyến nghị mang ý nghĩa một lời khuyên trên cơ sở các kết quả của đề tài. Người nhận lời khuyến nghị có thể tiếp nhận hoặc không tùy điều kiện thực tế. Kiến nghị hoặc đề nghị mang ý nghĩa sức ép đối với người nhận. Các khuyến nghị phải cụ thể, rõ ràng. Có thể có các loại khuyến nghị sau:

+ Khuyến nghị bổ sung về lý thuyết.

+ Khuyến nghị về sử dụng kết quả.

+ Khuyến nghị về hướng tiếp tục nghiên cứu.

9.1.2.6. Ghi danh mục tài liệu tham khảo

Có nhiều cách sắp xếp danh mục các tài liệu tham khảo, tốt nhất nên theo quy định của cơ quan quản lý đề tài. Thông thường người ta ghi tài liệu tiếng Việt trước theo thứ tự a, b, c tên tác giả chính, sau đó ghi tài liệu tiếng nước ngoài theo thứ tự a, b, c họ tác giả chính. Nếu tài liệu tham khảo có nhiều thứ tiếng thì cũng ghi theo thứ tự họ của tác giả tiếng Anh, tiếng Pháp. Tiếng Nga, tiếng Trung Quốc nên phiên âm theo mẫu tự latin.

9.1.3. Phần phụ lục

Phụ lục là phần đính kèm nằm cuối cùng của báo cáo khoa học. Ở đây có thể là các số liệu gốc, các hình vẽ, hình ảnh hay kết quả xử lý số liệu theo các phần mềm thống kê và được đánh số thứ tự theo số La Mã hay sồ Ả Rập. Ví dụ: Phụ lục I, Phụ lục II hay Phụ lục 1, Phụ lục 2. Kết quả xử lý thông kê theo các phần mềm nên để nguyên kiểu chữ gốc, nếu quá nhiều số lẻ có thể lược bớt nhưng phải trung thành với bản gốc.

Các trang của phần chính gồm cả tài liệu tham khảo và phụ lục phải được đánh số liên tục bằng số Ả rập (1,2,3...).

9.2. TRÌNH BÀY KẾT QUẢ

9.2.1. Các dạng bảng số liệu

Bảng số liệu hai chiều

Thông thường, bảng số liệu được trình bày hai chiều: chiều ngang (hàng, ví dụ, các chỉ tiêu) và chiều dọc (cột, ví dụ, các hạng mục), mỗi chiều ít nhất có hai hàng hay hai cột số liệu. Bảng số liệu hai chiều thông dụng như Bảng ... dưới đây.

Trường hợp bảng chỉ có một hàng hay một cột thì không nên lập bảng mà nên trình bày bằng lời và dẫn số liệu. Cũng có thể liệt kê số liệu kiểu bảng nhưng không nên ghi số bảng.

Bảng ...: Năng suất các giống bông bố mẹ và con lai F_1

TT	Nghiệm thức(*)	Năng suất (x_{ij}, tạ/ha) I	II	III	Tổng NT (T)	Trung bình
1	1. C92-52	30,3	19,3	22,2	71,8	23,9
2	2. C118A	12,8	15,9	12,5	41,2	13,7
3	3. S02-13	25,7	28,0	27,1	80,8	26,9
4	4. TM1	24,5	22,1	18,5	65,1	21,7
5	5. NH04-2	16,7	17,0	15,2	48,9	16,3
6	6. 1354	25,1	19,7	15,8	60,6	20,2
7	1 × 2	25,1	26,6	27,0	78,7	26,2
8	1 × 3	27,5	32,1	32,5	92,1	30,7
9	1 × 4	24,7	30,1	29,5	84,3	28,1
10	1 × 5	23,6	31,5	33,7	88,8	29,6
11	1 × 6	24,1	30,7	26,7	81,5	27,2
12	2 × 3	26,1	28,4	24,0	78,5	26,2
13	2 × 4	20,5	25,9	27,6	74,0	24,7
14	2 × 5	23,3	22,5	18,7	64,5	21,5
15	2 × 6	25,7	30,5	26,8	83,0	27,7
16	3 × 4	35,3	30,3	28,8	94,4	31,5
17	3 × 5	22,5	36,8	21,9	81,2	27,1
18	3 × 6	26,3	30,4	25,5	82,2	27,4
19	4 × 5	25,7	29,6	18,5	73,8	24,6
20	4 × 6	24,5	27,4	27,8	79,7	26,6
21	5 × 6	20,3	32,8	20,2	73,3	24,4
Tổng		510,3	567,6	500,5	1.578,4	
Trung bình chung, *m*						25,05

Bảng liệt kê một tập hợp số

Để phân tích số liệu có thể trình bày dạng bảng dưới hình thức một tập hợp số. Trong trường hợp này, tùy theo cách dẫn số liệu, bảng có thể không ghi số. Ví dụ:

Năng suất cá thể 50 cây (g/cây)

84,4	73,5	84,5	93,5	75,2	74,3	93,5	82,1	68,6	85,4
66,8	57,7	79,4	91,0	129,0	74,6	61,4	91,4	99,5	82,7
95,2	88,3	28,4	77,0	81,2	39,7	86,4	51,5	51,2	80,5
101,0	77,7	90,0	92,9	80,8	67,2	57,1	57,3	34,2	79,5
80,3	88,0	61,1	63,8	101,0	70,2	95,1	97,0	50,3	73,7

Trung bình: 76,92, phương sai: 351,68, độ lệch chuẩn: 18,75

Nếu trong quá trình phân tích muốn chỉ ra hay phân tích số liệu nào đó trong bảng thì cần phải ghi “địa chỉ”. Trong trường hợp này nên ghi số bảng.

Bảng ghép hoặc ô số ghép

Sử dụng bảng ghép hoặc ô có số ghép khi muốn làm rõ quan hệ giữa các chỉ tiêu. Người ta thường ghép hai hoặc ba bảng vào trong một bảng.

Ví dụ 1 (bảng có ô số ghép):

Bảng ...: Năng suất cá thể F_1 (g/cây), ưu thế lai thực (%) và sự suy giảm trong F_2 (%) của một số tổ hợp bông lai

♂ / ♀	S4880	175F	S8282	Karshi 2	Ash 25
S4880	**66,9**	$\frac{84{,}3}{26{,}0^{*}}$	$\frac{61{,}2}{-}$	$\frac{90{,}3}{35{,}0^{*}}$	$\frac{70{,}9}{6{,}0^{ns}}$
175F	- 12,6	**57,6**	$\frac{67{,}6}{4{,}6^{ns}}$	$\frac{83{,}7}{45{,}3^{*}}$	$\frac{72{,}9}{26{,}6^{*}}$
S8282	- 1,0	- 2,1	**64,6**	$\frac{87{,}5}{35{,}4^{*}}$	$\frac{70{,}3}{8{,}8^{ns}}$
Karshi 2	- 56,4	- 44,9	- 56,1	**48,7**	$\frac{64{,}8}{33{,}1^{*}}$
Ash 25	- 50,6	- 51,7	- 59,2	- 29,5	**39,9**

Ghi chú: - Đường chéo (số đậm) là năng suất bố mẹ;

- Phía trên đường chéo: Tử số là năng suất F_1 và mẫu số là ưu thế lai và mức ý nghĩa.

- Phía dưới đường chéo: mức suy giảm F_2 so F_1.

Ví dụ 2 (bảng ghép từ nhiều bảng):

Bảng ...: Số hạt lan Hồ Điệp nảy mầm ở các kỳ theo dõi

Đơn vị tính: Số mầm

Môi trường	Giống			Trung bình B
	V1	V2	V3	
<u>Lúc 30 ngày sau gieo</u>				
M1	$22,7^{e}$	$14,0^{f}$	$28,7^{d}$	$21,8^{C}$
M2	$31,0^{d}$	$12,7^{f}$	$28,7^{d}$	$24,1^{C}$
M3	$38,3^{bc}$	$10,7^{f}$	$37,7^{c}$	$28,9^{B}$
M4	$43,7^{b}$	$32,7^{cd}$	$63,0^{a}$	$46,4^{A}$
Trung bình V	$33,9^{B}$	$17,5^{C}$	$39,5^{A}$	
F(V) = $265,24^{**}$; F(M) = $189,41^{**}$; F(V×M) = $19,57^{**}$; F(VM) = $110,56^{**}$; CV (%) = 8,0				
<u>Lúc 60 ngày sau gieo</u>				
M1	$93,3^{d}$	$131,3^{ab}$	$120,7^{abcd}$	115,1
M2	$93,7^{d}$	$112,0^{bcd}$	$124,0^{abcd}$	109,9
M3	$95,7^{d}$	$107,3^{bcd}$	$129,7^{ab}$	110,9
M4	$97,0^{cd}$	$127,3^{abc}$	$147,3^{a}$	123,9
Trung bình V	$94,9^{B}$	$119,5^{A}$	$130,4^{A}$	
F(V) = $26,09^{**}$; F(M) = $3,76^{ns}$; F(V×M) = $2,42^{ns}$; F(VM) = $6,25^{**}$ CV (%) = 8,6				
<u>Lúc 80 ngày sau gieo</u>				
M1	$144,0^{g}$	$154,0^{e}$	$159,7^{c}$	$152,6^{C}$
M2	$145,3^{g}$	$157,0^{d}$	$160,7^{c}$	$154,3^{B}$
M3	$146,3^{g}$	$151,0^{f}$	$165,3^{b}$	$154,2^{B}$
M4	$150,7^{f}$	$160,3^{c}$	$175,0^{a}$	$162,0^{A}$
Trung bình V	$146,7^{C}$	$155,6^{B}$	$165,2^{A}$	
F(V) = $1008,35^{**}$; F(M) = $156,47^{**}$; F(V×M) = $28,93^{**}$; F(VM) = $241,79^{**}$ CV (%) = 0,7				

*Ghi chú: - ns: P > 0,05, *: 0,01 < P < 0,05 và **: P < 0,01*

- Trong cùng một cột, sau các giá trị trung bình có cùng mẫu tự không có sự khác biệt có ý nghĩa ở mức 0,01 theo phân hạng Ducan.

- Phân hạng giữa các mức của từng yếu tố (chữ in hoa) và phân hạng giữa các nghiệm thức (chữ thường)

9.2.2. Ghi các tham số thông kê

Thông tin về sự khác nhau giữa các nghiệm thức

Để biết giữa các nghiệm thức có sự khác nhau theo một chỉ tiêu nào đó hay không phải căn cứ vào giá trị F_{TN} hay xác suất sai lầm của F (ký hiệu là P hay Prob.). Có thể xác định P bằng hàm Fdist trong phần mềm Excel với cú pháp: = Fdist(trị F_{TN}, $df_{\text{nghiệm thức}}$, df_e).

Quan hệ giữa P và F_{TN} như sau:

Khi $P > 0{,}05$ thì $F_{TN} < F_{0,05}$.

Khi $0{,}01 < P \leq 0{,}05$ thì $F_{0,01} > F_{TN} \geq F_{0,05}$.

Khi $P \leq 0{,}01$ thì $F_{TN} \geq F_{0,01}$.

Từ giá trị P không cần tra bảng cũng biết được quan hệ giữa F_{TN} và $F_{\text{bảng}}$ và ngược lại.

Như vậy, để xác nhận có sự khác nhau hay không giữa các nghiệm thức, trên bảng số liệu có thể ghi một trong hai tham số, hoặc là giá trị P, hoặc là giá trị F_{TN} kèm mức ý nghĩa thống kê. Người ta đánh giá mức ý nghĩa của F_{TN} bằng các ký hiệu: *ns* (không có ý nghĩa ở độ tin cậy 95%), * (có ý nghĩa ở độ tin cậy 95%) và ** (có ý nghĩa ở độ tin cậy 99%) như: F_{TN} (A) = $36{,}63^{**}$; F_{TN}(B) = $3{,}68^{*}$; F_{TN} (C) = $1{,}71^{ns}$. Việc thông báo mức ý nghĩa của F_{TN}, để đơn giản, cũng có thể chỉ cần ghi bằng các ký hiệu: ns, * , ** mà không cần ghi giá trị.

Độ chính xác của thí nghiệm

Độ chính xác của thí nghiệm có thể đánh giá qua sai số tương đối hoặc hệ số biến động (CV,%):

Sai số tương đối: $s_{\overline{x}}(\%) = \frac{s_{\overline{x}}}{\overline{x}} \times 100 = \sqrt{\frac{MS_e}{r\overline{x}^2}} \times 100$

Người ta còn gọi sai số tương đối là độ chính xác.

Hệ số biến động: $CV(\%) = \frac{\sqrt{MS_e}}{\overline{x}} \times 100$

9.2.3. Phân hạng giữa các nghiệm thức

Phân hạng theo LSD_α

Về nguyên lý có thể phân hạng ở mức α thấp nhất (có thể) qua "thước đo" LSD_α, nhưng thông thường người ta dùng α = 0,05 hoặc 0,01. Khi $P \leq 0{,}01$, tức là $F_{TN} \geq F_{0,01}$ cho phép phân hạng theo $LSD_{0,01}$, nhưng ở mức α = 0,01, trong nhiều trường hợp không cho thấy sự khác nhau giữa hai nghiệm thức mặc dù ở mức α = 0,05 chúng lại khác nhau. Vì vậy, trong trường hợp này có thể phân hạng theo $LSD_{0,05}$.

Kết quả so sánh giữa các nghiệm thức có thể biểu thị:

- Bằng các hiệu số từng cặp trên bảng phối hợp hai chiều và đánh mức ý nghĩa.

- Đánh số a, b, c từ lớn đến nhỏ hoặc ngược lại. Những trị trung bình có cùng ký tự không khác nhau ở mức ý nghĩa so sánh.

- Đánh ký mức ý nghĩa ns, *, ** sau số trung bình khi so sánh với nghiệm thức đối chứng.

Phân hạng Duncan

Ngoài phân hạng theo LSD_α người ta còn phân hạng theo tiêu chuẩn xếp hạng Duncan (DMRT - Duncan's

Multiple Range Test).

Theo phương pháp này, các giá trị trung bình của các nghiệm thức được sắp xếp theo thứ tự từ lớn đến nhỏ hoặc ngược lại. Việc so sánh, phân hạng giữa các trung bình dựa vào các khoảng cách R_p.

$$R_p = \frac{r_{p(\alpha)}Sd}{\sqrt{2}} \text{ với } p = 2, 3, ..., t$$

trong đó: $r_{p(\alpha)}$ là giá trị được tra trong bảng phân vị thứ hạng kiểu phân phối Student (Significant Studentized Ranges) ở mức ý nghĩa α ở hàng df_e cụ thể cho từng thí nghiệm; *p* là khoảng cách theo vị trí của các trung bình nghiệm thức, *p = 2* là ở vị trí thứ 2 và *p = t* tức là vị trí cuối cùng theo thứ tự và $Sd = (2MS_e/r)^{(1/2)}$.

Ví dụ: Phân hạng Duncan cho 6 nghiệm thức từ B_1 đến B_6 (số liệu ở Bảng 7.46) với Sd = 0,545, df_e = 10.

NT	NS (tấn/ha)	*p*	r_p (0,05)	R_p	Hạng
B3	6,09				A
B2	5,91	2	3,15	1,21	A
B4	5,89	3	3,30	1,27	A
B1	5,16	4	3,37	1,30	AB
B5	5,05	5	3,43	1,32	AB
B6	4,15	6	3,46	1,33	B

So sánh, phân hạng được thực hiện như sau:

6,09 (B3) – 1,33 ($R_{p(6)}$) = 4,76. Như vậy những nghiệm thức có năng suất cao hơn 4,76 tấn/ha thì không khác biệt với B3.

Giữa B2 và B6: 5,91 – 4,15 = 1,76 > 1,33

Giữa B4 và B6: 5,89 – 4,15 = 1,74 > 1,33

Giữa B1 và B6: 5,16 – 4,15 = 1,01 < 1,33

Như vậy B2 và B4 khác hạng với B6 nhưng B1 và B5 không khác hạng với B6.

Kết quả phân hạng được ghi ở bảng trên.

Với 6 nghiệm thức này, kết quả phân hạng theo $LSD_{0,05}$ (ở Bảng 7.46) và phân hạng Duncan giống nhau.

Một số cách phân hạng khác

Có một số cách phân hạng khác cho *n* số trung bình của *n* nghiệm thức (NT).

Kiểu trắc nghiệm	Khoảng khác biệt tin cậy (D)
Tukey (còn gọi là Tukey – Cramer)	$D = w_\alpha Sd$; $w_\alpha = \dfrac{q(\alpha, n, df_e)}{\sqrt{2}}$; $q(\alpha, n, df_e)$ tra trong bảng: *Selected percentiles of Studentized Ranges Distributions*
Scheffé	$D = s_\alpha Sd$; $s_\alpha = \sqrt{df_{NT} \times f(\alpha, df_{NT}, df_e)}$
Bonferroni	$D = t'Sd$; $t' = t(\alpha', df_e)$; $\alpha' = \dfrac{\alpha}{2} \times n$

9.2.4. Trình bày kết quả của các loại thí nghiệm

Bảng số liệu so sánh một yếu tố

Ví dụ:

Bảng ...: Chiều cao cây, các yếu tố cấu thành năng suất và năng suất 15 giống bông

Giống	Chiều cao cây (cm)	Số quả/ cây (quả)	Khối lượng quả (g)	NS thực thu (tạ/ha)
1	$67,4^{bc}$	$13,4^{cde}$	$5,0^{ab}$	$20,0^{bc}$
2	$76,1^{ab}$	$15,7^{abcd}$	$5,3^{a}$	$24,4^{ab}$
3	$76,1^{ab}$	$13,1^{de}$	$5,3^{a}$	$22,9^{abc}$
4	$75,8^{ab}$	$17,4^{abc}$	$4,6^{bc}$	$18,1^{cd}$
5	$73,2^{ab}$	$10,8^{e}$	$4,8^{b}$	$20,8^{bc}$
6	$76,0^{ab}$	$17,9^{ab}$	4,9ab	$27,6^{a}$
7	$69,8^{ab}$	$16,8^{abcd}$	4,6bc	$12,9^{d}$
8	$71,3^{ab}$	$14,1^{bcde}$	4,3cde	$24,2^{ab}$
9	$74,8^{ab}$	$15,3^{abcd}$	4,2def	$17,7^{cd}$
10	$78,5^{a}$	$15,0^{abcd}$	$4,6^{bcd}$	$24,1^{ab}$
11	$70,3^{ab}$	$13,3^{de}$	$4,0^{ef}$	$18,0^{cd}$
12	$74,1^{ab}$	$17,5^{abc}$	$4,9^{ab}$	$24,2^{ab}$
13	$72,1^{ab}$	$18,8^{a}$	$4,2^{def}$	$19,9^{bc}$
14	$70,2^{ab}$	$17,6^{ab}$	$4,5^{bcd}$	$19,9^{bc}$
15	$59,5^{c}$	$17,0^{abcd}$	$3,8^{f}$	$21,8^{bc}$
F_{TN}	$2,14^{*}$	$2,64^{*}$	$4,9^{**}$	$4,01^{**}$
CV(%)	7,7	15,5	11,0	14,9

Ghi chú: Phân hạng các chỉ tiêu theo $LSD_{0,05}$

Bảng số liệu so sánh hai yếu tố

Ví dụ:

Bảng ...: Năng suất ba giống lúa (A) với năm mức bón đạm (B) và phân hạng

Yếu tố B	Năng suất (tấn/ha)			Tổng N (B)
	A_1	A_2	A_3	
B_0	$3,12^{f}$	$3,55^{ef}$	$3,78^{e}$	$3,48^{C}$
B_1	$4,73^{d}$	$4,81^{cd}$	$4,76^{d}$	$4,76^{B}$
B_2	$4,46^{d}$	$5,44^{ab}$	$5,32^{bc}$	$5,07^{B}$
B_3	$5,72^{ab}$	$5,74^{ab}$	$5,55^{ab}$	$5,67^{A}$
B_4	$5,82^{ab}$	$5,68^{ab}$	$5,89^{a}$	$5,80^{A}$
Tổng giống (A)	$4,77^{B}$	$5,05^{A}$	$5,06^{A}$	
$F(A) = 3,48^{*}$; $F(B) = 68,26^{**}$; $F(A\times B) = 1,89^{ns}$; $F(AB)= 21,15^{**}$				

Ghi chú: Phân hạng yếu tố A theo $LSD_{0,05}$, yếu tố B theo $LSD_{0,01}$ (chữ hoa) và phân hạng các nghiệm thức AB theo $LSD_{0,05}$ (chữ thường)

Bảng số liệu so sánh ba yếu tố

Ví dụ:

Bảng ...: Năng suất sáu giống lúa (B), ba mức bón đạm (A) với hai phương pháp gieo trồng (C) và phân hạng

(B)\(C)	Đạm (A)						TB (B)
	A_1		A_2		A_3		
	C_1	C_2	C_1	C_2	C_1	C_2	
B1	$3,57^{lm-o}$	$2,79^{on}$	$5,13^{fg-l}$	$4,98^{gh-l}$	$7,55^{abc}$	$6,92^{ab-e}$	$5,16^{AB}$
B2	$4,94^{gh-l}$	$4,50^{ij-m}$	$6,71^{ab-f}$	$5,51^{ef-j}$	$7,21^{ab-d}$	$6,61^{ab-f}$	$5,91^{A}$
B3	$4,25^{jk-n}$	$4,34^{jk-m}$	$6,12^{bc-h}$	$6,23^{bc-g}$	$7,87^{a}$	$7,72^{ab}$	$6,09^{A}$
B4	$4,06^{jk-o}$	$5,17^{fg-k}$	$5,56^{ef-j}$	$5,55^{ef-j}$	$7,10^{ab-d}$	$7,88^{a}$	$5,89^{A}$
B5	$4,10^{jk-n}$	$3,73^{kl-o}$	$5,63^{de-j}$	$4,78^{gh-m}$	$6,01^{cd-i}$	$6,02^{cd-i}$	$5,05^{AB}$
B6	$3,21^{omn}$	$4,51^{ij-m}$	$3,72^{kl-o}$	$4,62^{hi-m}$	$2,49^{o}$	$6,32^{ab-g}$	$4,15^{B}$
TB A	$4,10^{C}$		$5,38^{B}$		$6,64^{A}$		
TB C	5,29			5,45			

- $F(A) = 36,63^{**}$; $F(B) = 3,68^{*}$; $F(C) = 1,71^{ns}$; $F(A\times B) = 2,57^{*}$; $F(A\times C) = 2,93^{ns}$; $F(B\times C) = 11,27^{**}$; $F(A\times B\times C) = 1,78^{ns}$; $F(ABC) = 6,67^{**}$

- Phân hạng yếu tố A, yếu tố B (chữ hoa) và các nghiệm thức ABC (chữ thường) theo $LSD_{0,05}$

9.2.5. Lấy độ chính xác của các số liệu

- Độ chính xác của các số liệu phụ thuộc vào mức cho phép của công cụ cân đo. Với mỗi số đo ta có thể lấy độ chính xác đến mức cho phép. Thông thường, nếu đơn vị đo chiều cao cây là cm, ta nên lấy một số lẻ, ví dụ như 18,3 (tức là 183 mm hay khoảng trên 18 cm), còn 18,7 (tức là 187 mm hay gần 19 cm). Nếu lấy tròn số, ví dụ: 18 cm, ta không biết được trước khi làm tròn nó là 17,5 hay 18,4. Có thể lấy tối đa hai số lẻ khi không muốn làm tròn số đến mm. Tương tự, nếu đơn vị đo là g, ta cũng nên lấy một số lẻ, ví dụ, 5,6 (tức là gần 6) còn 5,4 (tức là khoảng trên 5g). Với hai đơn vị đo này, nếu lấy quá nhiều số lẻ lại làm cho số liệu kém chính xác vì không thể đo tới 1/100, 1/1000 cm hay 1/100, 1/1000 g. Nếu số liệu là thời gian như ngày, giờ, phút hoặc số đếm như số cây, số lá, số cành, số quả, số hạt chỉ được lấy một số lẻ. Số lẻ này không có ý nghĩa về đơn vị đo lường mà cho biết nếu số liệu đó là bao nhiêu nếu được làm tròn số. Nếu đơn vị đo là tấn, có thể lấy tối đa 3 số lẻ, tức là đến kg hoặc ít số lẻ hơn tùy độ chính xác. Có những phép cân đo cần độ chính xác cao nào đó, công cụ đo phải đạt được độ chính xác cần thiết và số liệu nên lấy một số lẻ sau đơn vị đo.

Tuy nhiên, trong quá trình tính toán không hạn chế số lẻ của các số liệu trung gian. Việc giữ lại số lẻ được thực hiện sau khi tính toán xong.

Số lẻ của phương sai phụ thuộc vào độ chính các của đơn vị đo. Khi xử lý số liệu trên phần mềm Excel và các phần mềm khác, sau khi có kết quả giá trị phương sai nên giữ lại 3 số lẻ.

- Trên một cột số liệu phải lấy cùng độ chính xác (số số lẻ của mỗi số đo phải bằng nhau).

- Với số phần trăm (%) chỉ nên lấy một số lẻ vì số lẻ thứ hai không có ý nghĩa đo lường. Ví dụ: 70,6%, tức là gần 71%, còn 69,7%, tức là gần 70%.

9.2.6. Trình bày đồ thị, biểu đồ

Để làm rõ chiều hướng hay quy mô, từ các bảng số liệu có thể biểu thị bằng đồ thị, biểu đồ. Tùy theo từng loại chỉ tiêu mà chọn hình thức trình bày phù hợp.

Một số đồ thị, biểu đồ thường gặp:

- Đồ thị biểu thị động thái.
- Đồ thị biểu thị quan hệ.
- Biểu đồ biểu thị mức độ
- Biểu đồ biểu thị tỷ lệ.

TÀI LIỆU THAM KHẢO

Tiếng Việt

1. Trần Văn Diễn, Tô Cẩm Tú, 1995. Di truyền số lượng (Giáo trình Cao học Nông nghiệp), NXB Nông nghiệp.

2. Lê Tiến Dũng, 2009. Điều tra tình hình sản xuất cây ngắn ngày và đánh giá khả năng cho năng suất của một số giống ngô, lạc và đậu xanh vụ Hè Thu năm 2008 tại huyện Bác Ái tỉnh Ninh Thuận, Luận văn Thạc sĩ Khoa học Nông nghiệp, Đại học Nông Lâm Tp. HCM.

3. Vũ Cao Đàm, 2006. Phương pháp luận nghiên cứu khoa học, NXB Khoa học và Kỹ thuật.

4. Vũ Thị Kim Điệp, 2009. Nghiên cứu một số biện pháp kỹ thuật trong sản suất hạt và rau mầm cây cải bẹ xanh (*Brassica juncea* L.) tại Bảo Lộc, Lâm Đồng, Luận văn Thạc sĩ Khoa học Sinh học, Đại học Đà Lạt.

5. Trần Văn Hãn, 1978. Đại số tuyến tính trong kỹ thuật, NXB Đại học và Trung học chuyên nghiệp.

6. Nguyễn Đình Hiền, 2007. Bài giảng về xử lý số liệu trong sinh học, NXB Nông nghiệp.

7. Nguyễn Quang Minh, 2006. Các bài thực hành Excel dùng cho kỹ sư và người làm khoa học, NXB Thanh niên.

8. Phạm Chí Thành, 1976. Giáo trình Phương pháp thí nghiệm đồng ruộng, NXB Nông nghiệp.

9. Đặng Hùng Thắng, 2008. Thống kê và ứng dụng, NXB Giáo dục.

10. Tô Cẩm Tú, Nguyễn Huy Hoàng, 2003. Phân tích số liệu nhiều chiều, NXB Khoa học và Kỹ thuật.

11. Nguyễn Hải Tuất, Ngô Kim Khôi, 1996. Xử lý thống kê kết quả nghiên cứu thực nghiệm trong nông lâm nghiệp trên máy tính, NXB Nông nghiệp.

12. Nguyễn Cao Văn, Trần Thái Ninh, 2002. Lý thuyết xác suất và thống kê toán, NXB Giáo dục.

13. Phạm Viết Vượng, 2004. Phương pháp luận nghiên cứu khoa học, NXB Đại học Quốc gia Hà Nội.

Tiếng nước ngoài

14. Cochran W., Cox G. M., 1964. Experimental Designs, John Wiley & Sons, New York.

15. Gomez K. A., Gomez A. A., 1984. Statistical Procedures for Agricultural Research, 2nd Edition, John Wiley & Sons, New York.

16. Dospekhov B. A., 1985. Metodica Polevogo Opita, "Agropromizdat", Moskva (Tiếng Nga).

17. Gnedenco B. V., 1969. Teoria Veroiatnostei, Moskva (Tiếng Nga).

18. Gruman V. E., 1972. Teoria Veroiatnostei i Matemati-cheskaia Statistica, Moskva (Tiếng Nga).

19. Liu L. F., Mao S. X., Huang Y. Z., 1984. Zuowu shuliang yichuan. Nongye Chubanshe, Beijing, 342 ye (Tiếng Trung).

20. Mandenhall W., 1990. Mathematical Statistics with Applications, Boston.

21. Soong T.T., 2004. Fundamentals of Probability and Statistics for Engineers, John Wiley & Sons, New York.

PHỤ LỤC

Phụ lục 1: Giá trị tới hạn T

Df (n)	Mức ý nghĩa α				Df (n)	Mức ý nghĩa α			
	.100	.050	.010	.001		.100	.050	.010	.001
1	6.314	12.706	63.657	636.62	21	1.721	2.080	2.831	3.819
2	2.920	4.303	9.925	31.598	22	1.717	2.074	2.819	3.792
3	2.353	3.182	5.841	12.941	23	1.714	2.069	2.807	3.767
4	2.132	2.776	4.604	8.610	24	1.711	2.064	2.797	3.745
5	2.015	2.571	4.032	6.859	25	1.708	2.060	2.787	3.725
6	1.943	2.447	3.707	5.959	26	1.706	2.056	2.779	3.707
7	1.895	2.365	3.499	5.405	27	1.703	2.052	2.771	3.690
8	1.860	2.306	3.355	5.041	28	1.701	2.048	2.763	3.674
9	1.833	2.262	3.250	4.781	29	1.699	2.045	2.756	3.659
10	1.812	2.228	3.169	4.587	30	1.697	2.042	2.750	3.646
11	1.796	2.201	3.106	4.437	35	1.690	2.030	2.724	3.591
12	1.782	2.179	3.055	4.318	40	1.684	2.021	2.704	3.551
13	1.771	2.160	3.012	4.221	45	1.680	2.014	2.690	3.520
14	1.761	2.145	2.977	4.140	50	1.676	2.008	2.678	3.496
15	1.753	2.131	2.947	4.073	55	1.673	2.004	2.669	3.476
16	1.746	2.120	2.921	4.015	60	1.671	2.000	2.660	3.460
17	1.740	2.140	2.898	3.965	70	1.667	1.994	2.648	3.435
18	1.734	2.101	2.878	3.922	80	1.665	1.989	2.638	3.416
19	1.729	2.093	2.861	3.883	90	1.662	1.986	2.631	3.402
20	1.725	2.086	2.845	3.850	100	1.661	1.982	2.625	3.390
					120	1.658	1.980	2.617	3.373
					∞	1.645	1.960	2.576	3.291

Phụ lục 2: Giá trị tới hạn Fisher (F)

df	α	1	2	3	4	5	6	7	8	9	10
1	.05	161	200	216	225	230	234	237	239	241	242
	.01	4.052	4.999	5.403	5.625	5.764	5.859	5.928	5.981	6.022	6.056
2	.05	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.39
	.01	98.5	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40
3	.05	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.78
	.01	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23
4	.05	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
	.01	21.2	18.00	16.69	15.98	15.52	15.21	14.98	14.8	14.66	14.55
5	.05	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
	.01	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05
6	.05	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
	.01	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87
7	.05	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
	.01	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62
8	.05	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
	.01	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81
9	.05	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
	.01	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26
10	.05	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
	.01	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	.05	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
	.01	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	.05	4.75	3.49	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
	.01	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	.05	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
	.01	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	.05	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
	.01	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
15	.05	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
	.01	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	.05	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
	.01	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	.05	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
	.01	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	.05	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
	.01	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51

11	12	15	20	24	30	40	75	100	∞	α	df
243	244	246	248	249	250	251	253	253	254	.05	1
6.038	6.016	6.517	6.209	6.234	6.261	6.286	6.323	6.334	6.366	.01	
19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50	.05	2
99.41	99.42	99.43	99.45	99.46	99.47	99.48	99.49	99.49	99.50	.01	
8.76	8.74	8.70	8.66	8.64	8.62	8.60	8.57	8.56	8.53	.05	3
27.13	27.05	26.87	26.69	26.60	26.50	26.41	26.27	26.23	26.12	.01	
5.40	5.91	5.86	5.80	5.77	5.74	5.71	5.68	5.66	5.63	.05	4
14.45	14.37	14.20	14.02	13.93	13.83	13.74	13.61	13.57	13.46	.01	
4.70	4.68	4.62	4.56	4.53	4.50	4.46	4.42	4.40	4.36	.05	5
9.96	9.89	9.72	9.55	9.47	9.38	9.29	9.17	9.13	9.02	.01	
4.03	4.00	3.94	3.87	3.84	3.81	3.77	3.72	3.71	3.67	.05	6
7.79	7.72	7.56	7.40	7.31	7.23	7.14	7.02	6.99	6.88	.01	
3.60	3.57	3.51	3.44	3.41	3.38	3.34	3.29	3.28	3.23	.05	7
6.54	6.47	6.31	6.16	6.07	5.98	5.90	5.78	5.75	5.65	.01	
3.31	3.28	3.22	3.15	3.12	3.08	3.05	3.00	2.98	2.93	.05	8
5.73	5.67	5.52	5.36	5.28	5.20	5.11	5.00	4.96	4.86	.01	
3.10	3.07	3.01	2.94	2.90	2.86	2.82	2.77	2.76	2.71	.05	9
5.18	5.11	4.96	4.71	4.73	4.64	4.56	4.46	4.41	4.31	.01	
2.94	2.91	2.85	2.77	2.74	2.70	2.67	2.61	2.59	2.54	.05	10
4.77	4.71	4.56	4.41	4.33	4.25	4.17	4.06	4.01	3.91	.01	
2.82	2.79	2.72	2.65	2.61	2.57	2.53	2.47	2.45	2.40	.05	11
4.46	4.40	4.25	4.10	4.02	3.94	3.86	3.74	3.70	3.60	.01	
2.72	2.69	2.62	2.54	2.50	2.46	2.42	2.36	2.35	2.30	.05	12
4.22	4.16	4.01	3.86	3.78	3.70	3.61	3.49	3.46	3.36	.01	
2.63	2.60	2.53	2.46	2.42	2.38	2.34	2.28	2.26	2.21	.05	13
4.02	3.96	3.82	3.66	3.59	3.51	3.42	3.30	3.27	3.16	.01	
2.57	2.53	2.46	2.39	2.35	2.31	2.27	2.21	2.19	2.13	.05	14
3.86	3.80	3.66	3.51	3.43	3.34	3.26	3.14	3.11	3.00	.01	
2.51	2.48	2.40	2.33	2.29	2.25	2.21	2.15	2.12	2.07	.05	15
3.73	3.67	3.57	3.37	3.29	3.20	3.12	3.00	2.97	2.87	.01	
2.46	2.42	2.35	2.28	2.24	2.20	2.16	2.09	2.07	2.01	.05	16
3.62	3.55	3.41	3.26	3.18	3.10	3.01	2.98	2.86	2.75	.01	
2.41	2.38	2.31	2.23	2.19	2.15	2.11	2.04	2.02	1.96	.05	17
3.52	3.46	3.31	3.16	3.08	3.00	2.92	2.79	2.76	2.65	.01	
2.37	2.34	2.27	2.19	2.15	2.11	2.07	2.00	1.98	1.92	.05	18
3.43	3.37	3.23	3.08	3.00	2.91	2.83	2.71	2.68	2.57	.01	

df	α	1	2	3	4	5	6	7	8	9	10
19	.05	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
	.01	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	.05	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
	.01	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
21	.05	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
	.01	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	.05	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
	.01	7.95	5.72	4.82	3.31	3.99	3.76	3.59	3.45	3.35	3.26
23	.05	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
	.01	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	.05	4.26	3.40	3.01	2.87	2.62	2.51	2.42	2.36	2.30	2.25
	.01	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	2.26	3.17
25	.05	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
	.01	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13
26	.05	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
	.01	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
27	.05	4.21	3.35	2.96	2.73	2.57	5.46	2.37	2.31	2.25	2.20
	.01	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06
28	.05	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
	.01	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
29	.05	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
	.01	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00
30	.05	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
	.01	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
32	.05	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14
	.01	7.50	5.34	4.46	3.97	3.65	3.43	3.26	3.13	3.02	2.93
34	.05	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17	2.12
	.01	7.44	5.29	4.42	3.93	3.61	3.39	3.22	3.09	2.98	2.89
36	.05	4.11	3.26	2.87	2.63	2.48	2.36	2.28	2.21	2.15	2.11
	.01	7.40	5.25	4.38	3.89	3.57	3.35	3.18	3.05	2.95	2.86
38	.05	4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09
	.01	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.92	2.83
40	.05	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
	.01	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
42	.05	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06
	.01	7.28	5.15	4.29	3.80	3.49	3.27	3.10	2.97	2.86	2.78

11	12	15	20	24	30	40	75	100	∞	α	df
2.34	2.31	2.23	2.16	2.11	2.07	2.02	1.96	1.94	1.88	.05	19
3.36	3.30	3.15	3.00	2.92	2.84	2.76	2.63	2.60	2.49	.01	
2.31	2.28	2.20	2.12	2.08	2.04	1.99	1.92	1.90	1.84	.05	20
3.29	3.23	3.09	2.94	2.86	2.77	2.69	2.56	2.53	2.42	.01	
2.28	2.25	2.18	2.10	2.05	2.00	1.96	1.89	1.87	1.81	.05	21
3.24	3.17	3.03	2.88	2.80	2.72	2.63	2.51	2.47	2.36	.01	
2.26	2.23	2.15	2.07	2.03	1.98	1.93	1.87	1.84	1.78	.05	22
3.18	3.12	2.98	2.83	2.75	2.67	2.58	2.46	2.42	2.31	.01	
2.24	2.20	2.13	2.05	2.00	1.96	1.91	1.84	1.82	1.76	.05	23
3.14	3.07	2.93	2.78	2.70	2.62	2.53	2.41	2.37	2.26	.01	
2.22	2.18	2.11	2.03	1.98	1.94	1.89	1.82	1.80	1.73	.05	24
3.09	3.03	2.89	2.74	2.66	2.58	2.49	2.36	2.33	2.21	.01	
2.20	2.14	2.09	2.01	1.96	1.92	1.87	1.80	1.77	1.71	.05	25
3.06	2.94	0.85	2.70	2.62	2.54	2.45	2.32	2.29	2.17	.01	
2.18	2.12	2.07	1.99	1.95	1.90	1.85	1.78	1.76	1.69	.05	26
3.02	2.90	2.81	2.66	2.58	2.50	2.41	2.28	2.25	2.13	.01	
2.17	2.10	2.06	1.97	1.93	1.88	1.84	1.76	1.74	1.67	.05	27
2.99	2.87	2.78	2.63	2.56	2.47	2.38	2.25	2.21	2.10	.01	
2.15	2.09	2.04	1.96	1.91	1.87	1.81	1.75	1.72	1.65	.05	28
2.96	2.84	2.75	2.60	2.52	2.44	2.35	2.22	2.18	2.06	.01	
2.14	2.08	2.03	1.94	1.90	1.85	1.80	1.73	1.71	1.64	.05	29
2.93	2.81	2.73	2.57	2.49	2.41	2.32	2.19	2.15	2.03	.01	
2.13	2.09	2.01	1.93	1.89	1.84	1.7	1.72	1.69	1.62	.05	30
2.91	2.84	2.70	2.55	2.47	2.38	2.29	2.16	2.13	2.01	.01	
2.10	2.07	1.99	1.91	1.86	1.82	1.76	1.69	1.67	1.59	.05	32
2.86	2.80	2.65	2.50	2.42	2.34	2.25	2.12	2.08	1.96	.01	
2.08	2.05	1.97	1.89	1.84	1.80	1.74	1.67	1.64	1.57	.05	34
2.82	2.76	2.61	2.46	2.38	2.30	2.21	2.08	2.04	1.91	.01	
2.07	2.03	1.95	1.87	1.82	1.78	1.72	1.65	1.62	1.55	.05	36
2.79	2.72	2.58	2.43	2.36	2.26	2.17	2.04	2.00	1.87	.01	
2.05	2.02	1.94	1.85	1.80	1.76	1.71	1.63	1.60	1.53	.05	38
2.75	2.69	2.55	2.40	2.32	2.22	2.14	2.00	1.97	1.84	.01	
2.04	2.00	1.92	1.84	1.79	1.74	1.69	1.61	1.59	1.51	.05	40
2.73	2.66	2.52	2.37	2.29	2.20	2.11	1.97	1.94	1.81	.01	
2.03	1.99	1.91	1.83	1.78	1.73	1.68	1.60	1.57	1.49	.05	42
2.70	2.64	2.50	2.34	2.26	2.17	2.08	1.94	1.91	1.78	.01	

Df	α	1	2	3	4	5	6	7	8	9	10
44	.05	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05
	.01	7.25	5.12	4.26	3.78	3.47	3.24	3.08	2.95	2.84	2.75
46	.05	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.15	2.09	2.04
	.01	7.22	5.10	4.24	3.76	3.44	3.22	3.06	2.93	2.82	2.73
48	.05	4.04	3.19	2.80	2.57	2.14	2.29	2.21	2.14	2.08	2.03
	.01	7.19	5.08	4.22	3.74	3.43	3.20	3.04	2.91	2.80	2.71
50	.05	4.03	3.18	2.79	2.56	2.40	2.29	2.02	2.13	2.07	2.03
	.01	7.17	5.06	4.20	3.72	3.41	3.09	3.02	2.89	2.78	2.70
55	.05	4.02	3.16	2.77	2.54	2.38	2.27	2.18	2.11	2.06	2.01
	.01	7.12	5.01	4.16	3.68	3.37	3.15	2.98	2.85	2.75	2.66
60	.05	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
	.01	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
65	.05	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.03	1.98
	.01	7.04	4.95	4.10	3.62	3.31	3.09	2.93	2.80	2.69	2.61
70	.05	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97
	.01	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67	2.59
80	.05	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95
	.01	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55
100	.05	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93
	.01	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50
120	.05	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91
	.01	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47
150	.05	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89
	.01	6.81	4.75	3.91	3.45	3.14	2.92	2.76	2.63	2.53	2.44
200	.05	3.89	3.01	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88
	.01	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41
400	.05	3.86	3.02	2.63	2.39	2.24	2.12	2.03	1.96	1.90	1.85
	.01	6.70	4.66	3.83	2.27	3.06	2.85	2.68	2.56	2.45	2.37
1000	.05	3.85	3.00	2.61	2.28	2.22	2.11	2.02	1.95	1.89	1.84
	.01	6.66	4.63	3.80	3.34	3.04	3.82	2.66	2.53	2.43	2.34
∞	.05	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83
	.01	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32

11	12	15	20	24	30	40	75	100	∞	α	df
2.01	1.98	1.90	1.81	1.76	1.72	1.66	1.58	1.56	1.48	.05	44
2.68	2.62	2.47	2.32	2.24	2.15	2.06	1.92	1.88	1.75	.01	
2.00	1.97	1.89	1.80	1.75	1.71	1.65	1.57	1.54	1.46	.05	46
2.66	2.60	2.45	2.30	2.22	2.13	2.04	1.90	1.86	1.72	.01	
1.99	1.96	1.88	1.79	1.74	1.70	1.64	1.56	1.53	1.45	.05	48
2.64	2.58	2.44	2.28	2.20	2.11	2.02	1.88	1.84	1.70	.01	
1.99	1.95	1.87	1.78	1.74	1.69	1.63	1.55	1.52	1.44	.05	50
2.63	2.56	2.42	2.27	2.18	2.10	2.00	1.86	1.82	1.68	.01	
1.97	1.93	1.85	1.76	1.72	1.67	1.61	1.52	1.50	1.41	.05	55
2.59	2.53	2.38	2.23	2.15	2.06	1.96	1.82	1.78	1.64	.01	
1.95	1.92	1.84	1.75	1.70	1.65	1.59	1.50	1.48	1.39	.05	60
2.56	2.50	2.35	2.20	2.12	2.03	1.93	1.79	1.74	1.60	.01	
1.94	1.90	1.82	1.73	1.68	1.63	1.57	1.49	1.46	1.37	.05	65
2.53	2.47	2.33	2.17	2.09	2.00	1.90	1.76	1.71	1.56	.01	
1.93	1.98	1.81	1.72	1.67	1.62	1.56	1.47	1.46	1.36	.05	70
2.51	5.46	2.31	2.15	2.07	1.98	1.88	1.74	1.69	1.53	.01	
1.91	1.88	1.79	1.70	1.65	1.60	1.54	1.45	1.42	1.32	.05	80
2.48	2.42	2.27	2.12	2.03	1.94	1.84	1.70	1.65	1.49	.01	
1.89	1.85	1.77	1.68	1.63	1.57	1.51	1.42	1.39	1.28	.05	100
2.43	2.37	2.22	2.07	1.98	1.89	1.79	1.64	1.59	1.43	.01	
1.87	1.83	1.75	1.66	1.61	1.55	1.50	1.40	1.37	1.25	.05	120
2.40	2.34	2.19	2.03	1.95	1.86	1.76	1.61	1.56	1.37	.01	
1.85	1.82	1.73	1.64	1.59	1.54	1.47	1.37	1.34	1.22	.05	150
2.37	2.31	2.16	2.00	1.91	1.83	1.72	1.56	1.51	1.33	.01	
1.84	1.80	1.72	1.62	1.57	1.52	1.45	1.35	1.32	1.19	.05	200
2.34	2.27	2.13	1.97	1.88	1.79	1.69	1.53	1.48	1.28	.01	
1.81	1.78	1.69	1.60	1.54	1.49	1.42	1.32	1.28	1.13	.05	400
2.29	2.23	2.08	1.92	1.84	1.74	1.64	1.47	1.42	1.19	.01	
1.80	1.76	1.68	1.58	1.53	1.47	1.41	1.30	1.26	1.08	.05	1000
2.27	2.20	2.06	1.90	1.81	1.71	1.61	1.44	1.38	1.11	.01	
1.79	1.75	1.67	1.57	1.52	1.46	1.40	1.28	1.24	1.00	.05	∞
2.25	2.18	2.04	1.88	1.79	1.69	1.59	1.41	1.36	1.00	.01	

Phụ lục 3: Giá trị tới hạn khi bình phương χ^2

Bậc tự do	Mức ý nghĩa							
	0.99	0.95	0.90	0.50	0.10	0.05	0.01	0.001
1	0.0002	0.00393	0.0158	0.455	2.706	3.841	6.635	10.827
2	0.0201	0.103	0.211	1.386	4.605	5.991	9.21	13.815
3	0.115	0.352	0.584	2.366	6.251	7.815	11.345	16.268
4	0.297	0.711	1.064	3.357	7.779	9.488	13.277	18.465
5	0.554	1.145	1.61	4.351	9.236	11.70	15.086	20.517
6	0.872	1.635	2.204	5.348	10.645	12.592	16.812	22.457
7	1.239	2.167	2.833	6.346	12.017	14.067	18.475	24.322
8	1.646	2.733	3.49	7.344	13.362	15.507	20.90	26.125
9	2.088	3.325	4.168	8.343	14.684	16.919	21.666	27.877
10	2.558	3.940	4.865	9.342	15.987	18.307	23.209	29.588
11	3.053	4.575	5.578	10.341	17.275	19.675	24.725	31.264
12	3.571	5.226	6.304	11.304	18.549	21.026	26.217	32.909
13	4.107	5.892	7.042	12.340	19.812	22.362	27.688	34.528
14	4.60	6.571	7.79	13.339	21.064	23.685	29.141	36.123
15	5.229	7.261	8.547	14.339	22.307	24.996	30.578	37.697
16	5.812	7.962	9.312	15.338	23.542	26.296	32.000	29.252
17	6.408	8.672	10.085	16.338	24.769	27.587	33.409	40.790
18	7.015	9.390	10.865	17.338	25.989	28.869	34.805	42.312
19	7.633	10.117	11.651	18.338	27.204	30.144	36.191	43.820
20	8.260	10.851	12.443	19.337	28.412	31.410	37.566	45.315
21	8.897	11.591	13.24	20.337	29.615	32.671	38.932	46.797
22	9.452	12.338	14.041	21.337	30.813	33.924	40.289	48.268
23	10.196	13.091	14.848	22.337	32.007	35.172	41.638	49.268
24	10.856	13.848	15.659	23.337	33.196	36.415	42.980	51.179
25	11.524	14.611	16.473	24.337	34.382	37.652	44.314	52.620
26	12.198	15.379	17.292	25.336	35.563	38.885	45.642	54.052
27	12.879	16.151	18.114	26.336	36.741	40.113	46.963	55.476
28	13.565	16.928	18.939	27.336	37.916	41.337	48.278	56.893
29	14.256	17.708	19.768	28.336	39.087	42.557	49.588	58.302
30	14.953	18.493	20.599	29.336	40.256	43.773	50.892	59.703

Phụ lục 4: Giá trị tới hạn *r*

Bậc tự do	Mức ý nghĩa			
	.1	.05	.01	.001
1	.9879	.9969	.9999	1.0000
2	.9000	.9500	.9900	.9990
3	.8054	.8783	.9587	.9912
4	.7293	.8114	.9172	.9741
5	.6694	.7545	.8745	.9507
6	.6215	.7067	.8343	.9349
7	.5822	.6664	.7977	.8982
8	.5494	.6319	.7646	.8721
9	.5214	.6021	.7348	.8471
10	.4973	.5760	.7079	.8233
11	.4762	.5529	.6835	.8010
12	.4575	.5324	.6614	.7800
13	.4409	.5139	.6411	.7603
14	.4259	.4973	.6226	.7420
15	.4124	.4821	.6055	.7246
16	.4000	.4683	.5897	.7084
17	.3887	.4555	.5751	.6932
18	.3783	.4438	.5614	.6787
19	.3687	.4329	.5487	.6652
20	.3598	.4227	.5368	.6524
25	.3233	.3809	.4869	.5974
30	.2960	.3494	.4487	.5541
35	.2746	.3246	.4182	.5189
40	.2573	.3044	.3932	.4896
45	.2428	.2875	.3721	.4648
50	.2306	.2732	.3541	.4433
60	.2108	.2500	.3248	.4078
70	.1954	.2319	.3017	.3799
80	.1829	.2172	.2830	.3568
90	.1726	.2050	.2673	.3375
100	.1638	.1946	.2540	.3211

Phụ lục 5: Giá trị kiểm định tổng hạng Wilcoxon

α = .025 một phía và α = .05 hai phía

n_2 \ n_1	3		4		5		6		7		8		9		10	
	T_l	T_u	T_l	T_u	T_l	T_u	T_l	T_u	T_l	T_u	T_l	T_u	T_l	T_u	T_l	T_u
3	5	16	6	18	6	21	7	23	7	26	8	28	8	31	9	33
4	6	18	11	25	12	28	12	32	13	35	14	38	15	41	16	44
5	6	21	12	28	18	37	19	41	20	45	21	49	22	53	24	56
6	7	23	12	32	19	41	26	52	28	56	29	61	31	65	32	70
7	7	26	13	35	20	45	28	56	37	68	39	73	41	78	43	83
8	8	28	14	38	21	49	29	61	39	73	49	87	51	93	54	98
9	8	31	15	41	22	53	31	65	41	78	51	93	63	108	66	114
10	9	33	16	44	24	56	32	70	43	83	54	98	66	114	79	131

α = .005 một phía và α = .01 hai phía

n_2 \ n_1	3		4		5		6		7		8		9		10	
	T_l	T_u	T_l	T_u	T_l	T_u	T_l	T_u	T_l	T_u	T_l	T_u	T_l	T_u	T_l	T_u
3	6	15	7	17	7	20	8	22	9	24	9	27	10	29	11	31
4	7	17	12	24	13	27	14	30	15	33	16	36	17	39	18	42
5	7	20	13	27	19	36	20	40	22	43	24	46	25	50	26	54
6	8	22	14	30	20	40	28	50	30	54	32	58	33	63	35	67
7	9	24	15	33	22	43	30	54	39	66	41	71	43	76	46	80
8	9	27	16	36	24	46	32	58	41	71	52	84	54	90	57	95
9	10	29	17	39	25	50	33	63	43	76	54	90	66	106	69	111
10	11	31	18	42	26	54	35	67	46	80	57	95	69	111	83	127

Phụ lục 6: Giá trị kiểm định tổng hạng theo dấu Wilcoxon

Một phía	Hai phía	Dung lượng mẫu n									
		5	6	7	8	9	10	11	12	13	14
p = .1	p = .2	2	3	5	8	10	14	17	21	26	31
p = .05	p = .1	0	2	3	5	8	10	13	17	21	25
p = .025	p = .05		0	2	3	5	8	10	13	17	21
p = .01	p = .02			0	1	3	5	7	9	12	15
p = .005	p = .01				0	1	3	5	7	9	12
p = .0025	p = .005					0	1	3	5	7	9
p = .001	p = .002						0	1	2	4	6
		15	16	17	18	19	20	21	22	23	24
p = .1	p = .2	36	42	48	55	62	69	77	86	94	104
p = .05	p = .1	30	35	41	47	53	60	67	75	83	91
p = .025	p = .05	25	29	34	40	46	52	58	65	73	81
p = .01	p = .02	19	23	27	32	37	43	49	55	62	69
p = .005	p = .01	15	19	23	27	32	37	42	48	54	61
p = .0025	p = .005	12	15	19	23	27	32	37	42	48	54
p = .001	p = .002	8	11	14	18	21	26	30	35	40	45
		25	26	27	28	29	30	31	32	33	34
p = .1	p = .2	113	124	134	145	157	169	181	194	207	221
p = .05	p = .1	100	110	119	130	140	151	163	175	187	200
p = .025	p = .05	89	98	107	116	126	137	147	159	170	182
p = .01	p = .02	76	84	92	101	110	120	130	140	151	162
p = .005	p = .01	68	75	83	91	100	109	118	128	138	148
p = .0025	p = .005	60	67	74	82	90	98	107	116	126	136
p = .001	p = .002	51	58	64	71	79	86	94	103	112	121
		35	36	37	38	39	40	41	42	43	44
p = .1	p = .2	235	250	265	281	297	313	330	348	365	384
p = .05	p = .1	213	227	241	256	271	286	302	319	336	353
p = .025	p = .05	195	208	221	235	249	264	279	294	310	327
p = .01	p = .02	173	185	198	211	224	238	252	266	281	296
p = .005	p = .01	159	171	182	194	207	220	233	247	261	276
p = .0025	p = .005	146	157	168	180	192	204	217	230	244	258
p = .001	p = .002	131	141	151	162	173	185	197	209	222	235
		35	36	37	38	39	40	41	42	43	44
p = .1	p = .2	402	422	441	462	482	503	525	547	569	592
p = .05	p = .1	371	389	407	426	446	466	486	507	529	514
p = .025	p = .05	343	361	378	396	415	434	453	473	494	514
p = .01	p = .02	312	328	345	362	379	397	416	434	454	473
p = .005	p = .01	291	307	322	339	355	373	390	408	427	445
p = .0025	p = .005	272	287	302	318	334	350	367	384	402	420
p = .001	p = .002	249	263	277	292	307	323	339	355	372	389

Phụ lục 7a: Giá trị tới hạn $g_{\alpha}^{(n-1,k)}$ phân phối Cocharan với k bậc tự do, l số lượng mẫu, α = .05

l \ k	1	2	3	4	5	6	7
2	.9850	.9750	.9392	.9057	.8772	.8534	.8332
3	.9669	.8709	.7977	.7457	.7071	.6771	.6530
4	.9065	.7679	.6841	.6287	.5895	.5598	.5365
5	.8412	.6338	.5981	.5440	.5063	.4783	.4564
6	.7808	.6161	.5321	.4803	.4447	.4184	.3980
7	.7271	.5612	.4800	.4307	.3974	.3726	.3535
8	.6798	.5157	.4377	.3910	.3595	.3362	.3185
9	.6385	.4775	.4027	.3584	.3286	.3067	.2901
10	.6020	.4450	.3733	.3311	.3029	.2823	.2666
12	.5410	.3924	.3624	.2880	.2624	.2439	.2299
15	.4709	.3346	.2758	.2419	.2195	.2034	.1911
20	.3894	.2705	.2205	.1921	.1735	.1602	..1501
24	.3434	.2354	.1907	.1656	.1493	.1374	.1286
30	.2929	.1980	.1593	.1377	.1237	.1137	.1061
40	.2370	.1576	.1259	.1082	.0968	.0887	.0827
60	.1737	.1131	.0895	.0765	.0682	.0623	.0583
120	.0998	.0632	.0495	.0419	.0371	.0337	.0312
+ ∞	.0000	.0000	.0000	.0000	.0000	.0000	.0000

l \ k	8	9	10	16	36	144	+ ∞
2	.8159	.8010	.7880	.7341	.6602	.5813	.5000
3	.6333	.6167	.6025	.5466	.4748	.4031	.3333
4	.5175	.5017	.4884	.4366	.3720	.3093	.2500
5	.4387	.4241	.4118	.3645	.3066	.2013	.2000
6	.3817	.3682	.3568	.3145	.2612	.2119	.1667
7	.3384	.3259	.3154	.2756	.2278	.1833	.1429
8	.3043	.2926	.2829	.2462	.2022	.1616	.1250
9	.2768	.2659	.2568	.2226	.1820	.1446	.1111
10	.2541	.2434	.2353	.2032	.1655	.1308	.1000
12	.2187	.2098	.2020	.1737	.1403	.1100	.0833
15	.1815	.1736	.1671	.1429	.1144	.0889	.0667
20	.1422	.1357	.1303	.1108	.0879	.0675	.0500
24	.1216	.1160	.1113	.0942	.0743	.0567	.0417
30	.1002	.0958	.0921	.0771	.0604	.0457	.0333
40	.0780	.0745	.0713	.0595	.0462	.0347	.0250
60	.0552	.0520	.0497	.0411	.0316	.0234	.0167
120	.0292	.0279	.0266	.0218	.0165	.0120	.0083
+ ∞	.0000	.0000	.0000	.0000	.0000	.0000	.0000

Phụ lục 7b: Giá trị tới hạn $g_{\alpha}^{(n-1,k)}$ phân phối Cocharan với k bậc tự do, l số lượng mẫu, α = .01

l \ k	1	2	3	4	5	6	7
2	.9999	.9950	.9794	.9586	.9373	.9172	.8988
3	.9933	.9423	.8831	.8335	.7938	.7606	.7335
4	.9676	.8643	.7841	.7212	.6761	.6410	.6129
5	.9279	.7885	.6957	.6329	.5875	.5531	.5259
6	.8828	.7218	.6258	.5635	.5195	.4866	.4608
7	.8376	.6644	.5685	.5080	.4659	.4347	.4105
8	.7945	.6152	.5209	.4627	.4226	.3932	.3704
9	.7544	.5727	.4810	.4251	.3870	.3592	.3378
10	.7175	.5358	.4469	.3934	.3572	.3308	.3106
12	.6528	.4751	.3919	.3428	.3099	.2861	.2680
15	.5747	.4069	.3317	.2882	.2593	.2386	.2228
20	.4799	.3297	.2654	.2288	.2048	.1877	.1748
24	.4247	.2871	.2295	.1970	.1759	.1608	.1495
30	.3632	.2412	.1913	.1635	.1454	.1327	.1232
40	.2940	.1915	.1508	.1281	.1135	.1033	.0957
60	.2151	.1371	.1069	.0902	.0796	.0722	.0668
120	.1225	.0759	.0585	.0489	.0429	.0387	.0357
+ ∞	.0000	.0000	.0000	.0000	.0000	.0000	.0000

l \ k	8	9	10	16	36	144	+ ∞
2	.8823	.8674	.8539	.7949	.7067	.6062	.5000
3	.7107	.6912	.6743	.6059	.5153	.4230	.3333
4	.5897	.5702	.5536	.4884	.4057	.3251	.2500
5	.5037	.4854	.4697	.4094	.3351	.2644	.2000
6	.4401	.4229	.4084	.3529	.2858	.2229	.1667
7	.3911	.3751	.3616	.3105	.2494	.1929	.1429
8	.3522	.3373	.3248	.2779	.2214	.1700	.1250
9	.3207	.3007	.2950	.2514	.1992	.1521	.1111
10	.2945	.2810	.2704	.2297	.1811	.1376	.1000
12	.2535	.2413	.2320	.1961	.1535	.1157	.0833
15	.2104	.2002	.1918	.1612	.1251	.0934	.0667
20	.1646	.1567	.1501	.1248	.0960	.0709	.0500
24	.1406	.1338	.1283	.1060	.0810	.0595	.0417
30	.1157	.1100	.1054	.0867	.0658	.0480	.0333
40	.0898	.9853	.0816	.0668	.0503	.0363	.0250
60	.0625	.0594	.0567	.0461	.0344	.0245	.0167
120	.0334	.0316	.0302	.0242	.0178	.0125	.0083
+ ∞	.0000	.0000	.0000	.0000	.0000	.0000	0000

Phụ lục 8: Giá trị chuyển đổi Arc sin$\sqrt{\%}$

%	0	1	2	3	4	5	6	7	8	9
0.0	0	0.57	0.81	0.99	1.15	1.28	1.40	1.52	1.62	1.72
0.1	1.81	1.90	1.99	2.07	2.14	2.22	2.29	2.36	2.43	2.50
0.2	2.56	2.63	2.69	2.75	2.81	2.87	2.92	2.98	3.03	3.09
0.3	3.14	3.19	3.24	3.29	3.34	3.39	3.44	3.49	3.53	3.58
0.4	3.63	3.67	3.72	3.76	3.80	3.85	3.89	3.93	3.97	4.01
0.5	4.05	4.09	4.13	4.17	4.21	4.25	4.29	4.33	4.37	4.40
0.6	4.44	4.48	4.52	4.55	4.59	4.62	4.66	4.69	4.73	4.76
0.7	4.80	4.83	4.87	4.90	4.93	4.97	5.00	5.03	5.07	5.10
0.8	5.13	5.16	5.20	5.23	5.26	5.29	5.32	5.35	5.38	5.41
0.9	5.44	5.47	5.50	5.53	5.56	5.59	5.62	5.65	5.68	5.71
1	5.74	6.02	6.29	6.55	6.80	7.04	7.27	7.49	7.71	7.92
2	8.13	8.33	8.53	8.72	8.91	9.10	9.28	9.46	9.63	9.81
3	9.98	10.14	10.31	10.47	10.63	10.78	10.94	11.09	11.24	11.39
4	11.54	11.68	11.83	11.97	12.11	12.25	12.39	12.52	12.66	12.79
5	12.92	13.05	13.18	13.31	13.44	13.56	13.69	13.81	13.94	14.06
6	14.18	14.30	14.42	14.54	14.65	14.77	14.89	15.00	15.12	15.23
7	15.34	15.45	15.56	15.68	15.79	15.89	16.00	16.11	16.22	16.32
8	16.43	16.54	16.64	16.74	16.85	16.95	17.05	17.16	17.26	17.36
9	17.46	17.56	17.66	17.76	17.85	17.95	18.05	18.15	18.24	18.34
10	18.44	18.53	18.63	18.72	18.81	18.91	19.00	19.09	19.19	19.28
11	19.37	19.46	19.55	19.64	19.73	19.82	19.91	20.00	20.09	20.18
12	20.27	20.36	20.44	20.53	20.62	20.70	20.79	20.88	20.96	21.05
13	21.13	21.22	21.30	21.39	21.47	21.56	21.64	21.72	21.81	21.89
14	21.97	22.06	22.14	22.22	22.30	22.38	22.46	22.55	22.63	22.71
15	22.79	22.87	22.95	23.03	23.11	23.19	23.26	23.34	23.42	23.50
16	23.58	23.66	23.73	23.81	23.89	23.97	24.04	24.12	24.20	24.27
17	24.35	24.43	24.50	24.58	24.65	24.73	24.80	24.88	24.95	25.03
18	25.10	25.18	25.25	25.33	25.40	25.48	25.56	25.62	25.70	25.77
19	25.84	25.92	25.99	26.06	26.13	26.21	26.28	26.35	26.42	26.49
20	26.56	26.64	26.71	26.78	26.85	26.92	26.99	27.06	27.13	27.20
21	27.28	27.35	27.42	27.49	27.56	27.63	27.69	27.76	27.83	27.90
22	27.97	28.04	28.11	28.18	28.25	28.32	28.38	28.45	28.52	28.59
23	28.66	28.73	28.79	28.86	28.93	29.00	29.06	29.13	29.20	29.27
24	29.33	29.40	29.47	29.53	29.60	29.67	29.73	29.80	29.87	29.93
25	30.00	30.07	30.13	30.20	30.26	30.33	30.40	30.46	30.53	30.59
26	30.66	30.72	30.79	30.85	30.92	30.98	31.05	31.11	31.18	31.24
27	31.31	31.37	31.44	31.50	31.56	31.63	31.69	31.76	31.82	31.88
28	31.95	32.01	32.08	32.14	32.20	32.27	32.33	32.39	32.46	32.52
29	32.58	32.65	32.71	32.77	32.83	32.90	32.96	33.02	33.09	33.15
30	33.21	33.27	33.34	33.40	33.46	33.52	33.58	33.65	33.71	33.77
31	33.83	33.89	33.96	34.02	34.08	34.14	34.20	34.27	34.33	34.39
32	34.45	34.51	34.57	34.63	34.70	34.76	34.82	34.88	34.94	35.00
33	35.06	35.12	35.18	35.24	35.30	35.37	35.43	35.49	35.55	35.61
34	35.67	35.73	35.79	35.85	36.91	35.97	36.03	36.09	36.15	36.21
35	36.27	36.33	36.39	36.45	36.51	36.57	36.63	36.69	36.75	36.81

%	0	1	2	3	4	5	6	7	8	9
36	36.87	36.93	36.99	37.05	37.11	37.17	37.23	37.29	37.35	37.41
37	37.47	37.52	37.58	37.64	37.70	37.76	37.82	37.88	37.94	38.00
38	38.06	38.12	38.17	38.23	38.29	38.35	38.41	38.47	38.53	38.59
39	38.65	38,70	38.76	38.82	38.88	38.94	39.00	39.06	39.11	39.17
40	39.23	39.29	39.35	39.41	39.47	39.52	39.58	39.64	39.70	39.76
41	39.82	39.87	39.93	39.99	40.05	40.11	40.16	40.22	40.28	40.34
42	40.40	40.56	40.51	40.57	40.63	40.69	40.74	40.80	40.86	40.92
43	40.98	41.03	41.09	41.15	41.21	41.27	41.32	41.38	41.44	41.50
44	41.55	41.61	41.67	41.73	41.78	41.84	41.90	41.96	42.02	42.07
45	42.13	42.19	42.25	42.30	42.36	42.42	42.48	42.53	42.59	42.65
46	42.71	42.76	42.82	42.88	42.94	42.99	43.05	43.11	43.17	43.22
47	43.28	43.34	43.39	43.45	43.51	43.57	43.62	43.68	43.74	43.80
48	43.85	43.91	43.97	44.03	44.08	44.14	44.20	44.25	44.51	44.37
49	44.43	44.46	44.54	44.60	44.66	44.71	44.77	44.83	44.89	44.94
50	45.00	45.06	45.11	45.17	45.23	45.29	45.34	45.40	45.46	45.52
51	45.57	45.63	45.69	45.75	45.80	45.86	45.92	45.97	46.03	46.09
52	46.15	46.20	46.26	46.32	46.38	46.43	46.49	46.55	46.61	46.66
53	46.72	46.78	46.83	46.89	46.95	47.01	47.06	47.12	47.18	47.24
54	47.29	47.35	47.41	47.47	47.52	47.58	47.64	47.70	47.75	47.81
55	47.87	47.93	47.98	48.04	48.10	48.16	48.22	48.27	48.33	48.39
56	48.45	48.50	48.56	48.62	48.68	48.73	48.79	48.85	48.91	48.97
57	49.02	49.08	49.14	49.20	49.26	49.31	49.37	49.43	49.49	49.54
58	49.60	49.66	49.72	49.78	49.84	49.89	49.95	50.01	50.07	50.13
59	50.18	50.24	50.30	50.36	50.42	50.48	50.53	50.59	50.65	50.71
60	50.77	50.83	50.89	50.94	51.00	51.06	51.12	51.18	51.24	51.30
61	51.35	51.41	51.47	51.53	51.59	51.65	51.71	51.77	51.83	51.88
62	51.94	52.00	52.06	52.12	52.18	52.24	52.30	52.36	52.42	52.48
63	52.53	52.59	52.65	52.71	52.77	52.83	52.89	52.95	53.01	53.07
64	53.13	53.19	53.25	53.31	53.37	53.43	53.49	53.55	53.61	53.67
65	53.73	53.79	53.85	53.91	53.97	54.03	54.09	54.15	54.21	54.27
66	54.33	54.39	54.45	54.51	54.57	54.63	54.70	54.76	54.82	54.88
67	54.94	55.00	55.06	55.12	55.18	55.24	55.30	55.37	55.43	55.49
68	55.55	55.61	55.67	55.73	55.80	55.86	55.92	55.98	56.04	56.11
69	56.17	56.23	56.29	56.35	56.42	56.48	56.54	56.60	56.66	56.73
70	56.79	56.85	56.91	56.98	57.04	57.10	57.17	57.23	57.29	57.35
71	57.42	57.48	57.54	57.61	57.67	57.73	57.80	57.86	57.92	57.99
72	58.05	58.12	58.18	58.24	58.31	58.37	58.44	58.50	58.56	58.63
73	58.69	58.76	58.82	58.89	58.95	59.02	59.08	59.15	59.21	59.28
74	59.34	59.41	59.47	59.54	59.60	59.67	59.74	59.80	59.87	59.93
75	60.00	60.07	60.13	60.20	60.27	60.33	60.40	60.47	60.53	60.60
76	60.67	60.73	60.80	60.87	60.94	61.00	61.07	61.14	61.21	61.27
77	61.34	61.41	61.48	61.55	61.62	61.68	61.75	61.82	61.89	61.96
78	62.03	62.10	62.17	62.24	62.31	62.37	62.44	62.51	62.58	62.65
79	62.72	62.80	62.87	62.94	63.01	63.08	63.15	63.22	63.29	63.36
80	63.44	63.51	63.58	63.65	63.72	63.79	63.87	63.94	64.01	64.08

%	0	1	2	3	4	5	6	7	8	9
81	64.16	64.23	64.30	64.38	64.45	64.52	64.60	64.67	64.75	64.82
82	64.90	64.97	65.05	65.12	65.20	65.27	65.35	65.42	65.50	65.57
83	65.65	65.73	65.80	65.88	65.96	66.03	66.11	66.19	66.27	66.34
84	66.42	66.50	66.58	66.66	66.74	66.81	66.89	66.97	67.05	67.13
85	67..21	67.29	67.37	67.45	67.54	67.62	67.70	67.78	67.86	67.94
86	68.03	68.11	68.19	68.28	68.36	68.44	68.53	68.61	68.70	68.78
87	68.87	68.95	69.04	69.12	69.21	69.30	69.38	69.47	69.56	69.64
88	69.73	69.82	69.91	70.00	70.09	70.18	70.27	70.36	70.45	70.54
89	70.63	70.72	70.81	70.91	71.00	71.09	71.19	71.23	71.37	71.47
90	71.56	71.66	71.76	71.85	71.95	72.05	72.15	72.24	72.34	72.44
91	72.54	72.64	72.74	72.84	72.95	73.05	73.15	73.26	73.36	73.46
92	73.57	73.68	73.78	73.89	74.00	74.11	74.21	74.32	74.44	74.55
93	74.66	74.77	74.88	75.00	75.11	75.23	75.35	75.46	75.58	75.70
94	75.82	75.94	76.06	76.19	76.31	76.44	76.55	76.69	76.82	76.95
95	77.08	77.21	77.34	77.48	77.61	77.75	77.89	78.03	78.17	78.32
96	78.46	78.61	78.76	78.91	79.06	79.22	79.37	79.53	79.69	79.86
97	80.02	80.19	80.37	80.54	80.72	80.90	81.09	81.28	81.47	81.67
98	81.87	82.08	82.29	82.51	82.73	82.96	83.20	83.45	83.71	83.98
99	84.26	84.29	84.32	84.35	84.38	84.41	84.44	84.47	84.50	84.53
99.1	84.56	84.59	84.62	84.65	84.68	84.71	84.74	84.77	84.80	84.84
99.2	84.87	84.90	84.93	84.97	85.00	85.03	85.07	85.10	85.13	85.17
99.3	85.20	85.24	85.27	85.31	85.34	85.38	85.41	85.45	85.48	85.52
99.4	85.56	85.60	85.63	85.67	85.71	85.75	85.79	85.83	85.87	85.91
99.5	85.95	85.99	86.03	86.07	86.11	86.15	86.20	86.24	86.28	86.33
99.6	86.37	86.42	86.47	86.51	86.56	86.61	86.66	86.71	86.76	86.81
99.7	86.86	86.91	86.97	87.02	87.08	87.13	87.19	87.25	87.31	87.37
99.8	87.44	87.50	87.57	87.64	87.71	87.78	87.86	87.93	88.01	88.10
99.9	88.19	88.28	88.38	88.48	88.60	88.72	88.85	89.01	89.19	89.43
100	90.00									

Phụ lục 9: Bảng phân vị thứ hạng kiểu phân phối Student (Significant Studentized Ranges)

df_e	α	p = số nghiệm thức cần so sánh									
		2	3	4	5	6	7	8	9	10	12
1	.05	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0
	.01	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0
2	.05	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09
	.01	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0
3	.05	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50
	.01	8.26	8.50	8.60	8.70	8.80	8.90	8.90	9.00	9.00	9.00
4	.05	3.93	4.01	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02
	.01	6.51	6.80	6.90	7.00	7.10	7.10	7.20	7.20	7.30	7.30
5	.05	3.64	3.74	3.79	3.83	3.83	3.83	3.83	3.83	3.83	3.83
	.01	5.70	5.96	6.11	6.18	6.26	6.33	6.40	6.44	6.50	6.60
6	.05	3.46	3.58	3.64	3.68	3.68	3.68	3.68	3.68	3.68	3.68
	.01	5.24	5.51	5.65	5.73	5.81	5.88	5.95	6.00	6.00	6.10
7	.05	3.35	3.47	3.54	3.58	3.60	3.61	3.61	3.61	3.61	3.61
	.01	4.95	5.22	5.37	5.45	5.53	5.61	5.69	5.73	5.80	5.80
8	.05	3.26	3.39	3.47	3.52	3.55	3.56	3.56	3.56	3.56	3.56
	.01	4.74	5.00	5.14	5.23	5.32	5.40	5.47	5.51	5.50	5.60
9	.05	3.20	3.34	3.41	3.47	3.50	3.52	3.52	3.52	3.52	3.52
	.01	4.60	4.86	4.99	5.08	5.17	5.25	5.32	5.36	5.40	5.50
10	.05	3.15	3.30	3.37	3.43	3.46	3.47	3.47	3.47	3.47	3.47
	.01	4.48	4.73	4.88	4.96	5.06	5.13	5.20	5.24	5.28	5.36
11	.05	3.11	3.27	3.35	3.39	3.43	3.44	3.45	3.46	4.46	3.46
	.01	4.39	4.63	4.77	4.86	4.94	5.01	5.06	5.12	5.15	5.24
12	.05	3.08	3.23	3.33	3.36	3.40	3.42	3.44	3.44	3.46	3.46
	.01	4.32	4.55	4.68	4.76	4.81	4.92	4.96	5.02	5.07	5.13
13	.05	3.06	3.21	3.30	3.35	3.38	3.41	3.42	3.44	3.45	3.45
	.01	4.26	4.48	4.62	4.69	4.74	4.84	4.88	4.94	4.98	5.04
14	.05	3.03	3.18	3.27	3.33	3.37	3.39	3.41	3.42	3.44	3.45
	.01	4.21	4.42	4.55	4.63	4.70	4.78	4.83	3.87	4.91	4.96
15	.05	3.01	3.16	3.25	3.31	3.36	3.38	3.40	3.42	3.43	3.44
	.01	4.17	4.37	4.50	4.58	4.64	4.72	4.77	4.81	4.84	4.90

df_e	α	p = số nghiệm thức cần so sánh									
		2	3	4	5	6	7	8	9	10	12
16	.05	3.00	3.15	3.23	3.30	3.34	3.37	3.39	3.41	3.43	3.44
	.01	4.13	4.34	4.45	4.54	4.60	4.67	4.72	4.76	4.79	4.84
17	.05	2.98	3.13	3.22	3.28	3.33	3.36	3.38	3.40	3.42	3.44
	.01	4.10	4.30	4.41	4.50	4.56	4.63	4.68	4.72	4.75	4.80
18	.05	2.97	3.12	3.21	3.27	3.32	3.35	3.37	3.39	3.41	3.43
	.01	4.07	4.27	4.38	4.46	4.53	4.59	4.64	4.68	4.71	4.76
19	.05	2.96	3.11	3.19	3.26	3.31	3.35	3.37	3.39	3.41	3.43
	.01	4.05	4.24	4.35	4.43	4.50	4.56	4.61	4.64	5.67	4.72
20	.05	2.95	3.10	3.18	3.25	3.30	3.34	3.36	3.38	3.40	3.43
	.01	4.02	4.22	4.33	4.40	4.47	4.53	4.58	4.61	4.65	4.69
22	.05	2.93	3.08	3.17	3.24	3.29	3.32	3.35	3.37	3.39	3.42
	.01	3.99	4.17	4.28	4.36	4.42	4.48	4.53	4.57	4.60	4.65
24	.05	2.92	3.07	3.15	3.22	3.28	3.31	3.34	3.37	3.38	3.41
	.01	3.96	4.14	4.24	4.33	4.39	4.44	4.49	4.53	4.57	4.62
26	.05	2.91	3.06	3.14	3.21	3.27	3.30	3.34	3.36	3.38	3.41
	.01	3.93	4.11	4.21	4.30	4.36	4.41	4.46	4.50	4.53	4.58
28	.05	2.90	3.04	3.13	3.20	3.26	3.30	3.33	3.35	3.37	3.40
	.01	3.91	4.08	4.18	4.28	4.34	4.39	4.43	4.47	4.51	4.56
30	.05	2.89	3.04	3.12	3.20	3.25	3.29	3.32	3.35	3.37	3.40
	.01	3.89	4.06	4.16	4.22	4.32	4.36	4.41	4.45	4.48	4.54
40	.05	2.86	3.01	3.10	3.17	3.22	3.27	3.30	3.33	3.35	3.39
	.01	3.82	3.99	4.10	4.17	4.21	4.30	4.34	4.37	4.41	4.46
60	.05	2.83	2.98	3.08	3.14	3.20	3.24	3.28	3.31	3.33	3.37
	.01	3.76	3.92	4.03	4.12	4.17	4.23	4.27	4.31	4.34	4.39
100	.05	2.80	2.95	3.05	3.12	3.18	3.22	3.26	3.29	3.32	3.36
	.01	3.71	3.86	3.98	4.06	4.11	4.17	4.21	4.25	4.29	4.35
∞	.05	2.77	2.92	3.02	3.09	3.15	3.19	3.23	3.26	3.29	3.34
	.01	3.64	3.80	3.90	3.98	4.04	4.09	4.14	4.17	4.20	4.26

CƠ SỞ TOÁN HỌC
CỦA CÁC PHÉP XỬ LÝ THỐNG KÊ TRONG NGHIÊN CỨU KHOA HỌC NÔNG NGHIỆP

PGS. TS. PHAN THANH KIẾM

Chịu trách nhiệm xuất bản:
TS. LÊ QUANG KHÔI

Phụ trách bản thảo : **Thành Vinh**
Biên tập : **Thành Vinh**
Trình bày – bìa : **Khánh Hà**

NHÀ XUẤT BẢN NÔNG NGHIỆP
167/6 – Phương Mai – Đống Đa – Hà Nội
ĐT: (04) 38523887 – 35760656 – 38521940
Fax: (04) 35760748. E-mail: nxbnn@hn.vnn.vn

CHI NHÁNH NHÀ XUẤT BẢN NÔNG NGHIỆP
58 Nguyễn Bỉnh Khiêm Q.1, TP. Hồ Chí Minh
ĐT: (08) 39111603 – 38297157 – 38299521
Fax: (08) 39101036. E-mail: cnnxbnn@yahoo.com.vn

In 1.030 bản khổ 14,5 × 20,5 cm tại Xưởng in NXB Nông nghiệp.
Đăng ký KHXB số do Cục xuất bản cấp
ngày . In xong và nộp lưu chiểu quý /2010