

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC NÔNG NGHIỆP I HÀ NỘI**

NGUYỄN THỊ LAN (Chủ biên) & PHẠM TIẾN DŨNG



**GIÁO TRÌNH
PHƯƠNG PHÁP THÍ NGHIỆM**

HÀ NỘI – 2005

MỞ ĐẦU

Phương pháp thí nghiệm là một ngành khoa học được dạy trong một số trường đại học có liên quan đến lĩnh vực sinh học như: Nông nghiệp, lâm nghiệp và thủy sản. Tuy nhiên, nội dung dạy có khác nhau tùy thuộc vào ngành cụ thể.

Cuốn giáo trình phương pháp thí nghiệm của chúng tôi lần này được viết cho sinh viên Đại học Nông nghiệp mà chủ yếu là cho ngành nông học. Nội dung của cuốn giáo trình cung cấp cho sinh viên một số kiến thức cơ bản về phương pháp thiết kế và bố trí thí nghiệm cũng như các thuật toán thống kê mô tả các tham số, các tiêu chuẩn kiểm định thống kê trong xử lý kết quả thí nghiệm, nghiên cứu các mối quan hệ đơn giản trong phân tích kết quả nghiên cứu và cách trình bày một báo cáo khoa học.

Giáo trình viết cho người học nên khi dạy cán bộ giảng dạy cần tham khảo thêm các giáo trình và sách khác viết kỹ và sâu hơn về phương pháp thí nghiệm cũng như các kiến thức toán xác suất thống kê.

Để nắm được kiến thức của môn học viết trong giáo trình này sinh viên phải được học và nắm vững kiến thức xác suất thống kê, tin học và một số môn khoa học khác.

Nội dung của cuốn giáo trình bao gồm các chương:

Chương I - Trình bày các bước cần thiết trong quá trình nghiên cứu khoa học và các nhóm phương pháp nghiên cứu trong nông nghiệp.

Chương II - Trình bày các yêu cầu trong thiết kế thí nghiệm, các loại thí nghiệm đồng ruộng cùng các nội dung khác có liên quan đến thiết kế thí nghiệm (công thức thí nghiệm, diện tích ô, nhắc lại, bảo vệ, hàng biên...). Cách xây dựng và viết một đề cương nghiên cứu khoa học. Đây là chương quan trọng nhất của phần phương pháp nghiên cứu.

Chương III - Giới thiệu cách triển khai một thí nghiệm cụ thể từ thiết kế (đề cương) ra ngoài thực địa nhằm đảm bảo tính khách quan và tôn trọng nguyên tắc “sai khác duy nhất”, chăm sóc thí nghiệm và trình tự thu hoạch thí nghiệm.

Chương IV - Trình bày các loại số liệu trong nghiên cứu khoa học, các tham số thống kê cơ bản của mẫu và các công thức tính các tham số đó.

Chương V - Trình bày ngắn gọn các cách ước lượng một số tham số thống kê cơ bản thường được sử dụng trong nghiên cứu và thực tiễn sản xuất nông nghiệp (ước lượng điểm và ước lượng khoảng của hai tham số cơ bản nhất của tổng thể là kỳ vọng và xác suất của đặc tính sinh học nào đó).

Chương VI - Trình bày ngắn gọn bài toán kiểm định các giả thiết thống kê thông thường như: kiểm định hai trung bình (hai kỳ vọng) và kiểm định hai xác suất của hai tổng thể và kiểm định tính độc lập.

Chương VII - Trình bày các phương pháp bố trí thí nghiệm 1, 2 nhân tố và cách phân tích phương sai. Công bố kết quả và đánh giá các kết quả thí nghiệm. Phần thí nghiệm hai nhân tố mới chỉ đề cập sơ bộ trong giáo trình này. Đây được coi là chương quan trọng nhất trong phần thiết kế thí nghiệm và thống kê ứng dụng. Chương này giúp các nhà khoa học đánh giá một cách đầy đủ kết quả cụ thể của mỗi thí nghiệm.

Chương VIII - Giới thiệu tương quan và hồi quy, chủ yếu là tương quan và hồi quy tuyến tính đơn. Nội dung chương này sẽ giúp cho người học cách đánh giá mối quan hệ của các đặc trưng (chỉ tiêu) trong thí nghiệm qua hệ số tương quan. Xây dựng phương trình hồi quy mô tả mối quan hệ tương quan, tính hồi quy tuyến tính đơn. Bước đầu giới thiệu quan hệ phi tuyến.

Chương IX - Giới thiệu cho người học cách trình bày số liệu trong báo cáo và trình tự viết một báo cáo khoa học.

Ngoài ra, giáo trình còn trình bày một số bảng số thống kê thông dụng giúp các nhà thực nghiệm xử lý kết quả nghiên cứu của thí nghiệm: các bảng này rất cần thiết cho chương ước lượng, kiểm định cũng như phân tích phương sai và hồi quy. Khi dạy giáo viên phải hướng dẫn cho học sinh biết cách sử dụng các bảng số.

CHƯƠNG I - ĐẠI CƯƠNG VỀ CÔNG TÁC NGHIÊN CỨU KHOA HỌC NÔNG NGHIỆP

Chương này trang bị cho người học những nhận thức cơ bản về phương pháp nghiên cứu Nông nghiệp nói chung và Nông học nói riêng như các bước cần được tiến hành trong quá trình nghiên cứu khoa học để trả lời câu hỏi mà thực tế đặt ra.

1.1. Vai trò của công tác nghiên cứu khoa học nông nghiệp

Theo nghĩa rộng của quan điểm triết học duy vật: "Thí nghiệm là một phần của sự nghiệp sản xuất trong xã hội loài người, nhằm khám phá ra các quy luật khách quan của thế giới vật chất với mục đích nắm vững và bắt các điều bí mật của thiên nhiên phục vụ cho cuộc sống con người".

Con người đã biết làm thí nghiệm (Experiment) từ bao giờ?

Như chúng ta đã biết, từ cổ xưa loài người đã phải kiếm ăn để sinh sống, do đó, con người phải biết lựa chọn, so sánh để tìm kiếm thức ăn. Song cũng chính từ đó mà họ đã tạo ra một kho tàng các kinh nghiệm quý báu thúc đẩy xã hội phát triển. Khi xã hội tiến lên đòi hỏi con người cũng phải nắm bắt, vận dụng các quy luật khách quan của tự nhiên có hiệu quả hơn. Muốn làm được điều này cần phải có phương pháp và từ đó phương pháp thí nghiệm ra đời.

Nghiên cứu khoa học là quá trình nghiên cứu và giải thích đến cùng các hiện tượng khoa học xuất phát từ lý luận và thực tiễn. Từ đó sẽ ứng dụng các kết quả nghiên cứu được vào thực tiễn sản xuất phục vụ cho cuộc sống con người.

Nghiên cứu khoa học nói chung và khoa học nông nghiệp nói riêng hay cụ thể hơn là nghiên cứu khoa học trong lĩnh vực Nông học phụ thuộc rất mật thiết với điều kiện tự nhiên và các điều kiện kinh tế - xã hội nên việc vận dụng các phương pháp và kết quả nghiên cứu của các nước trên thế giới có tính kế thừa chọn lọc cho phù hợp với điều kiện cụ thể của Việt Nam là rất cần thiết. Ngay trong phạm vi của đất nước chúng ta cũng không thể có tính đồng nhất về các điều kiện cụ thể, cho các thực nghiệm nông nghiệp, vậy nhiệm vụ của các nhà khoa học nông nghiệp phải nghiên cứu và đề xuất được những biện pháp kỹ thuật cụ thể, thích hợp cho vùng nơi mình phụ trách nhằm khai thác bền vững và hiệu quả các điều kiện ấy. Để có kết quả nghiên cứu đúng và khách quan cần phải có kiến thức tổng hợp của nhiều lĩnh vực: toán học, hoá học, thổ nhưỡng, khí tượng, sinh học và kinh tế học và phương pháp nghiên cứu đúng, khách quan, phù hợp với các quy luật tự nhiên, quy luật xã hội và cả tính sáng tạo đúng đắn.

1.2. Các bước trong quá trình nghiên cứu khoa học nông nghiệp

Để có thể xây dựng được một đề tài nghiên cứu khoa học nông nghiệp nói chung và cụ thể hơn nữa là xây dựng được một thí nghiệm về một biện pháp kỹ thuật nào đó như: Giống, phân bón, tưới nước, thời vụ hay bảo vệ thực vật... cho một vùng đòi hỏi nhà khoa học (người làm công tác nghiên cứu) cần phải tiến hành theo các bước sau đây.

1.2.1. Thu thập thông tin (Bước 1)

Mục đích của thu thập thông tin là giúp cho nhà khoa học hiểu rõ được vấn đề sẽ được nghiên cứu đã có ai, nơi nào nghiên cứu chưa và nếu có thì vấn đề được nghiên cứu đến đâu. Xem xét tính khả thi để từ đó hình thành hướng nghiên cứu thích hợp.

Nội dung thông tin thu thập gồm:

* Các tài liệu có liên quan trực tiếp và gián tiếp tới vấn đề dự định nghiên cứu.

* Kinh nghiệm sản xuất của người dân.

Việc thu thập các thông tin bao gồm:

- Đọc các tài liệu tại thư viện cụ thể là các sách báo gồm các giáo trình, sách chuyên khảo, sách hướng dẫn phổ biến khoa học kỹ thuật, các tạp chí khoa học, các kết quả nghiên cứu khoa học của các nhà khoa học khác. Các nguồn số liệu này bao gồm cả trong nước và trên thế giới.

- Tham dự các hội nghị, hội thảo và các hoạt động khoa học khác.

- Tìm hiểu thực tiễn sản xuất của nông dân để thấy rõ kinh nghiệm cũng như biện pháp xử lý của nông dân với vấn đề sẽ được nghiên cứu.

- Thu thập thông tin qua các phương tiện thông tin đại chúng khác như: Vô tuyến truyền hình, đài phát thanh, báo khoa học, báo nông thôn cũng như các loại báo khác.

1.2.2. Xây dựng giả thiết khoa học (Bước 2)

Giả thiết khoa học là những giả định mà theo nhà khoa học là có nhiều khả năng đúng nhất về một sự vật hay một hiện tượng nào đó. Nó giúp cho ta có thể phát hiện và giải thích những cái mới mà những giả thiết khác trước đây chưa giải thích được.

Vì vậy, giả thiết khoa học không được phép chung chung mà phải cụ thể, phải thực sự xuất phát từ các nguồn thông tin thu thập được (mục 1.2.1). Giả thiết này cũng chính là xuất phát điểm để xây dựng kế hoạch nghiên cứu thực nghiệm.

Giả thiết khoa học phải tránh viển vông, song không nên sợ cái mới, phải xuất phát từ quy luật khách quan của tự nhiên, đầu tư công sức, trí tuệ để tìm hiểu cái mới, thậm chí có thể khó khăn gai góc. Có như vậy con người mới có thể tìm ra được cái mới, cái đổi thay trong khoa học và có thể cắt nghĩa nó hoàn toàn có cơ sở, theo đúng logic của các quá trình các mối quan hệ qua lại lẫn nhau trong tự nhiên đầy bí hiểm và đa dạng.

1.2.3. Chứng minh giả thiết khoa học (Bước 3)

Chứng minh giả thiết khoa học là quá trình quan sát, quá trình làm thí nghiệm. Trên cơ sở các số liệu (các chỉ tiêu nghiên cứu thể hiện qua kết quả theo dõi hay quan sát) có được và suy luận nhằm gạt bỏ cái không đúng, sàng lọc lấy cái đúng có tính quy luật và những cái có thể coi là chân lý.

Kiểm chứng giả thiết khoa học có hai cách, đó là: **Quan sát hay điều tra và làm thí nghiệm thực nghiệm.**

* Quan sát hay điều tra là việc tìm hiểu, theo dõi thực tế, đây là cả một quá trình bắt nguồn từ việc thu thập những cái đơn giản, những cái đã có trong thực tế sản xuất và trong tự nhiên, giúp ta phân biệt được cái đặc trưng của sự việc, so sánh giữa các sự việc và tiến đến suy luận xây dựng căn cứ khoa học cho các sự việc đó. Hay nói một cách khác: quan sát là tìm hiểu, mô tả diện mạo bên ngoài của sự việc hay hiện tượng để từ đó suy ra bản chất của chúng dựa trên cơ sở nhận thức của người nghiên cứu. Như vậy, quan sát là đi từ bên ngoài sự việc vào trong nhận thức. Do đó, yêu cầu của quan sát là "kiên trì", chỉ có kiên trì mới có thể hy vọng thu được những thông tin, những tài liệu và có như vậy tài liệu mới đầy đủ, khách quan và mang tính chính xác. Quan sát (điều tra) phải được thực hiện sao cho đại diện, khách quan để đảm bảo độ tin cậy của những thông tin thu được về đối tượng nghiên cứu.

* **Làm thí nghiệm**

Thí nghiệm là những công việc mà con người tự xây dựng để tạo ra những hiện tượng làm thay đổi một cách nhân tạo bản chất của sự việc nhằm phát hiện được đầy đủ bản chất và

nguyên nhân (nguồn gốc) của hiện tượng hay sự việc đó, cũng như nghiên cứu mối quan hệ tương hỗ giữa các hiện tượng (hay các sinh vật).

Như vậy, thí nghiệm là xuất phát từ những nhận thức của con người thông qua những giả thiết khoa học (đã nêu mục 1.2.2.), sau đó xác minh bằng hành động của mình (thực hiện thí nghiệm, đo đếm, quan sát các chỉ tiêu trên đối tượng thí nghiệm). Quá trình xác minh (làm thí nghiệm có thể được thực hiện ở trong phòng thí nghiệm, trong các nhà lưới, nhà kính, các chậu, vại, ô xi măng hay trên đồng ruộng) sẽ đưa tới nhận thức chặt chẽ hơn.

Như vậy, con người không phải chỉ chờ đợi vào những cái đã có sẵn mà ngược lại, có thể tự mình tạo ra những ý tưởng cụ thể, thực hiện ý tưởng đó để bắt đối tượng nghiên cứu phải tự bộc lộ và phát sinh tính quy luật của mình. Hay có thể nhận thức theo nghĩa rộng mà Papiôp đã nói "Quan sát là thu thập những gì mà thiên nhiên cho ta, còn thí nghiệm là lấy từ thiên nhiên những gì mà ta muốn".

1.2.4. Biện luận để rút ra kết luận và xây dựng lý thuyết khoa học (Bước 4)

Thông qua các kết quả của quan sát, điều tra cũng như thí nghiệm, người làm nghiên cứu thực hiện việc kiểm chứng giả thiết khoa học để rút ra những kết luận và đánh giá vấn đề mà mình quan tâm, không thể chỉ nghiên cứu mà bỏ qua công tác biện luận và rút ra kết luận. Tất nhiên đây là công việc không hề đơn giản. Đề xuất ra được những kết luận và biện luận cho các kết luận đó đòi hỏi nhà khoa học phải có trình độ kiến thức và hiểu sâu sắc đối tượng mình nghiên cứu. Có như vậy, các kết luận và biện luận mới khách quan có cơ sở khoa học phù hợp với môi trường và hệ sinh thái cụ thể của đối tượng đó.

Nếu như các nhà khoa học chỉ dừng lại ở việc rút ra những kết luận trực tiếp từ thí nghiệm thì những kết luận đó chỉ mang tính chất kinh nghiệm cụ thể của duy nhất một lần thí nghiệm nên chưa thể ứng dụng rộng rãi trong thực tế sản xuất được. Do đó, nhiệm vụ tiếp của các nhà khoa học là từ những kết quả của thí nghiệm được làm lại nhiều lần tập hợp thành các kết luận và biện luận nhằm tìm ra chân lý, tìm ra tính quy luật để nâng lên thành lý luận khoa học.

1.3. Các nhóm phương pháp thí nghiệm trong nông nghiệp

Hiện nay trong thực tiễn nghiên cứu của ngành nông học người ta đã sử dụng các nhóm phương pháp nghiên cứu và được chia cụ thể như sau:

1.3.1. Nhóm các thí nghiệm cứu trong phòng

Đây là loại nghiên cứu mà những thí nghiệm được thực hiện trong các phòng thí nghiệm. Điều kiện để thực hiện được các thí nghiệm đó gồm các loại dụng cụ như: hoá chất, các máy móc phân tích, các bình, hộp, khay... mang tính chất riêng biệt (chuyên sâu). Nhóm các thí nghiệm này hầu như độc lập với điều kiện tự nhiên của môi trường bên ngoài. Do điều kiện thực hiện trong phòng, cho nên các kết quả từ các thí nghiệm này được kiểm tra, điều khiển bằng các dụng cụ có độ chính xác cao. Tuy nhiên, những số liệu này chưa được áp dụng vào thực tế. Bởi vì, ở những môi trường nghiên cứu khác mà nhất là trên đồng ruộng thì có rất nhiều nhân tố sinh thái ảnh hưởng tới sinh vật (cây trồng) và có nhiều nhân tố khó có thể kiểm soát được cụ thể và chính xác.

Thí dụ: Muốn tìm hiểu quá trình trao đổi vật chất và năng lượng trong cây để xem dòng vật chất năng lượng ấy vận chuyển tới đâu, nhanh hay chậm, bao nhiêu được tích lũy trong sản phẩm của cây và bao nhiêu được loại thải ra môi trường bên ngoài? hoặc việc nghiên cứu tốc độ phát triển về mặt số lượng của rễ, chiều dài rễ. Nhưng nhóm nghiên cứu trong phòng có nhược điểm là số lượng cá thể ít (không mang tính đại diện) và điều kiện

nghiên cứu nhân tạo không phải là điều kiện thực tại mà đối tượng nghiên cứu sẽ được gieo trồng.

1.3.2. Nhóm các thí nghiệm trong chậu vại

Các thí nghiệm thuộc nhóm phương pháp nghiên cứu này có đối tượng nghiên cứu là cây trồng được gieo trồng trong các chậu, vại bằng sành, sứ trên nền đất hay dung dịch hoặc trồng trong các ô xi măng, trong nhà lưới, nhà polyetylen nền đất hoặc cát. Về điều kiện thì đối với nhóm này cây trồng đã được sống trong một phần là điều kiện tự nhiên, còn một phần là điều kiện nhân tạo.

Đây là loại hình thí nghiệm thường làm tại các cơ quan nghiên cứu nông nghiệp như các Viện, các Trường Đại học, Cao đẳng và các Trung tâm nghiên cứu. Nhóm thí nghiệm này cũng đã có lịch sử lâu đời, kết quả nghiên cứu của nhóm này phần lớn nhằm giải thích cơ chế, bản chất của cây.

Thí dụ: Để xác định lượng nước cần cho quá trình sinh trưởng và phát triển của cây ở các giai đoạn, lượng bốc thoát hơi nước qua các bộ phận của cây người ta thường làm thí nghiệm trong chậu vại.

Hoặc nghiên cứu mối quan hệ giữa lượng nước tiêu thụ, lượng dinh dưỡng khoáng với khả năng tích lũy vật chất của cây. Từ kết quả của những nghiên cứu này có thể giúp cho các nhà khoa học điều chỉnh các nhân tố ảnh hưởng tới quá trình sinh trưởng và phát triển của cây trồng trong điều kiện tự nhiên.

So với nhóm các nghiên cứu trong phòng thì nhóm này có số lượng cá thể nhiều hơn. Do vậy, tính đại diện quần thể sinh vật mang tính chính xác cao hơn, và lại cây trồng lại có thể sống trong một điều kiện cụ thể và chúng có thể phải chịu cả những rủi ro của điều kiện thời tiết. Song mức độ chính xác thì chưa chắc chắn được như nhóm thí nghiệm trong phòng. Các nhà khoa học cũng cần nhận thức được rằng: kết quả của nhóm này tuy đã gắn với điều kiện sản xuất, song không thể thay thế cho nhóm thí nghiệm đồng ruộng được.

Khi thực hiện thí nghiệm nhóm này cần chú ý:

- Ngoài các nhân tố nghiên cứu cũng cần phải chú ý tất cả các nhân tố khác có ảnh hưởng tới kết quả của thí nghiệm.
- Cần phải chọn lọc dòng, giống cây đưa vào thí nghiệm có độ đồng đều cao, giống tốt (trừ thí nghiệm chọn, tạo dòng trong ngành chọn giống) để tránh ảnh hưởng cá thể vì số giống và số cá thể của giống, dòng nghiên cứu còn hạn chế.
- Số lần lặp lại (nhắc lại) phải cao hơn có thể tới chục lần. Thiết kế các vị trí để sao cho các đối tượng nghiên cứu chịu ảnh hưởng đồng đều của môi trường bên ngoài.
- Tạo điều kiện để có thể hạn chế đến mức cao nhất ảnh hưởng của rủi ro do thời tiết như gió, bão, nắng và tác hại của động vật hại cây trồng cũng như dịch bệnh khác.

1.3.3. Nhóm phương pháp nghiên cứu trên đồng ruộng

Nhóm nghiên cứu này bao gồm những thí nghiệm mà cây trồng được sống trong điều kiện tự nhiên. Do vậy, nó chịu sự chi phối của nhiều nhân tố (các nhân tố sinh thái) từ môi trường bên ngoài. Những nhân tố đó là: các điều kiện thời tiết, đất đai, các biện pháp kỹ thuật canh tác... Loại thí nghiệm này có ưu điểm là:

- Số lượng cá thể lớn (tính đại diện của quần thể sinh vật hay cây trồng cao).
- Gắn với điều kiện sản xuất. Vì vậy, có thể nghiên cứu được mối quan hệ tương hỗ giữa cây với nhiều nhân tố khác.

Cũng qua những kết quả của thí nghiệm đồng ruộng có thể nhận định rõ thêm kết quả và các kết luận của thí nghiệm trong phòng cũng như trong chậu, vại hoặc nhà lưới. Những kết luận của thí nghiệm đồng ruộng sẽ được coi là cơ sở xây dựng biện pháp kỹ thuật cho quy trình sản xuất và thâm canh cây trồng.

Nhà khoa học người Nga Ivanov (1969) đã nhận định: "Thí nghiệm đồng ruộng là phương pháp cơ bản và trung tâm của thí nghiệm nông nghiệp. Còn các phương pháp thí nghiệm nghiên cứu khác có tính chất thăm dò hay kết hợp giải quyết vấn đề".

Nhóm thí nghiệm đồng ruộng được tiến hành trong điều kiện tự nhiên nên khối lượng nghiên cứu lớn và có khả năng giải quyết các vấn đề mà các nhà khoa học đặt ra tùy thuộc vào mục đích nghiên cứu sao cho phù hợp với điều kiện sinh thái cũng như kinh tế - xã hội của một vùng nào đó. Các trang thiết bị và phương pháp nghiên cứu có độ chính xác thấp hơn so với hai nhóm phương pháp nghiên cứu trong chậu vại và trong phòng. Do đó, yêu cầu độ chính xác cũng thấp hơn (chúng tôi sẽ đề cập ở chương sau về nội dung này).

Trong phạm vi của cuốn giáo trình này chúng tôi chỉ tập trung đề cập chủ yếu các nội dung có liên quan tới thí nghiệm đồng ruộng. Còn các nội dung khác trong công tác nghiên cứu khoa học nói chung và nông học nói riêng được giới thiệu cụ thể ở từng môn học của chương trình đào tạo tùy theo chuyên ngành mà sinh viên theo học như: Cây trồng, Bảo vệ thực vật, Giống cây trồng, Làm vườn, Dâu tằm...

Bài tập: (Do giáo viên sẽ nêu ra cho sinh viên tùy điều kiện cụ thể).

CHƯƠNG II - THIẾT KẾ THÍ NGHIỆM TRÊN ĐỒNG RUỘNG

Đây là chương quan trọng nhất trong phần phương pháp thí nghiệm. Sau khi học, người học phải biết xây dựng một đề cương nghiên cứu, cách chọn công thức đối chứng, chọn đất thí nghiệm và chọn cây thí nghiệm.

Như chúng tôi đã đề cập thì thí nghiệm đồng ruộng là rất quan trọng. Bởi vì, để đẩy mạnh sản xuất nông nghiệp nói chung và ngành trồng trọt hay nông học nói riêng thì cây phải được hoàn toàn sống trong điều kiện tự nhiên phù hợp với bản năng sinh vật của nó. Vì vậy, để cho kết quả của thí nghiệm đồng ruộng được sát với điều kiện sản xuất, khi tiến hành thiết kế thí nghiệm người chủ trì hay người làm thí nghiệm phải nắm vững các vấn đề sau

2.1. Các yêu cầu của thí nghiệm đồng ruộng

Naidin (1968) đã đánh giá: "Thí nghiệm đồng ruộng là thí nghiệm nghiên cứu trong điều kiện tự nhiên, trên những mảnh đất đặc biệt, có mục đích xác định về số lượng các điều kiện và các biện pháp canh tác đến năng suất cây trồng".

Mỗi quá trình sinh học diễn ra trong cây đều có quan hệ chặt chẽ và có tác động qua lại lẫn nhau mà điều kiện ngoại cảnh (điều kiện sinh thái như các nhân tố khí hậu và các nhân tố có mặt trong đất) là rất quan trọng. Nếu như một nhân tố nào đó thay đổi sẽ làm cho các nhân tố khác cũng như hoạt động sống của cây thay đổi theo.

Cây trồng được sống trong điều kiện tự nhiên của nó sẽ bộc lộ những đặc trưng, đặc tính một cách rõ nét, trong đó, có cả những lợi thế và những hạn chế của các biện pháp kỹ thuật canh tác hoặc bản chất giống cây trồng giúp cho các nhà khoa học khẳng định giá trị của các biện pháp hay giống trước khi chuyển giao cho sản xuất. Vì vậy, thí nghiệm đồng ruộng phải tôn trọng các yêu cầu sau đây:

2.1.1. Yêu cầu về tính đại diện

Thí nghiệm đồng ruộng phải mang tính đại diện. Cơ sở của vấn đề này là:

- Mỗi một thành tựu nghiên cứu đều gắn liền với điều kiện kinh tế - xã hội nhất định.
- Khi thay đổi điều kiện tự nhiên, điều kiện kinh tế - xã hội thì biện pháp kỹ thuật phải thay đổi theo.

Tính đại diện được thể hiện qua hai mặt là:

*** Đại diện về điều kiện sinh thái**

Có nghĩa là thí nghiệm phải được thiết kế và làm cụ thể tại một vùng đất đai, trong điều kiện khí hậu của vùng đó tương tự như điều kiện sau này sẽ áp dụng.

*** Đại diện về điều kiện kinh tế - xã hội**

Tuỳ theo thời gian và tuỳ thuộc vào các điều kiện cụ thể khác về mặt xã hội mà người nông dân có các nhận thức cũng như khả năng tiếp thu các tiến bộ khoa học kỹ thuật vào sản xuất là khác nhau. Vì vậy, các nhà nghiên cứu phải có những thông tin từ đó xây dựng biện pháp (nhân tố thí nghiệm) cho phù hợp để sau một thời gian nghiên cứu thành công thì biện pháp đó có thể được sản xuất chấp nhận. Vì lẽ đó mà biện pháp kỹ thuật phải cao hơn điều kiện sản xuất một mức, mức này tuỳ thuộc vào từng địa phương, từng cộng đồng dân tộc và từng thời gian cụ thể. Nó hoàn toàn không có một mức chung cho tất cả.

2.1.2. Yêu cầu về sai khác duy nhất

Trong lôgic học "suy luận" nếu giữa trường hợp có phát sinh hiện tượng và trường hợp không phát sinh hiện tượng mà chỉ khác nhau có tình hình thì tình hình là nguyên nhân của hiện tượng.

Hiểu một cách cụ thể là trong thí nghiệm sẽ phân biệt hai loại yếu tố: yếu tố thí nghiệm (dùng để nghiên cứu) và yếu tố không thí nghiệm (hay còn gọi là nền thí nghiệm). Trong hai loại yếu tố này thì duy nhất chỉ có yếu tố thí nghiệm được quyền sai khác (thay đổi). Còn yếu tố không thí nghiệm (không cần so sánh) thì phải càng đồng nhất càng tốt.

Có triết để tôn trọng nguyên tắc này mới tìm được sự khác nhau của kết quả thí nghiệm là do nhân tố nào của yếu tố thí nghiệm gây ra. Tuy nhiên, sự đồng nhất tuyệt đối trên đồng ruộng là điều không thể có được.

Thí dụ: Nghiên cứu ảnh hưởng của liều lượng phân lân bón khác nhau tới năng suất lúa trên đất trũng.

Như vậy lượng lân bón cho lúa ở các công thức phải khác nhau, còn các biện pháp kỹ thuật khác là đồng nhất. Cụ thể giống lúa gì, cây hay gieo vãi ở vụ nào, mật độ bao nhiêu, lượng phân bón khác ngoài lân là bao nhiêu và cách bón lượng phân này ra sao, tưới nước, chăm sóc cũng như phòng trừ sâu hại... cũng đều phải đồng nhất.

Song có một điều cần lưu ý: trong thí nghiệm đồng ruộng không thể loại trừ hoàn toàn được một nhân tố nào đấy mà chỉ có khả năng hạn chế nó mà thôi. Trong thí nghiệm nêu trên ta chỉ biết được lượng lân cho thêm vào nghiên cứu là bao nhiêu ở các công thức, còn trong phân chuồng hoặc các dạng phân tổng hợp khác và cả trong đất cũng đã tồn tại một lượng lân nhất định. Tuy nhiên, điều này không có ảnh hưởng nhiều vì các công thức đều có nền thí nghiệm như nhau.

Một đặc điểm khác nữa đó là trong tự nhiên hay trong thí nghiệm đồng ruộng còn tồn tại mối quan hệ "kéo theo" có nghĩa là khi thay đổi nhân tố A thì nhân tố B cũng thay đổi.

Thí dụ: Nghiên cứu ảnh hưởng của lượng nước tưới khác nhau tới năng suất mía.

Như vậy nước là nhân tố của yếu tố thí nghiệm và được thay đổi ở mức độ khác nhau. Tuy nhiên, nếu mức nước tưới khác nhau kéo theo những thay đổi khác như số lượng, chủng loại vi sinh vật cũng như sinh vật đất, nhiệt độ đất, ẩm độ đất cũng thay đổi không giống nhau. Từ đó có thể làm quá trình sinh học của cây sẽ không giống nhau. Vì vậy, khi nhận định đánh giá các ảnh hưởng của các nhân tố thí nghiệm nói riêng và các điều kiện thí nghiệm nói chung phải tìm ra được nguyên nhân chính ảnh hưởng đến kết quả thí nghiệm, có như vậy mới đưa ra được biện pháp kỹ thuật có tính chất then chốt để đạt hiệu quả mong muốn.

Song cần phải lưu ý là tránh không được hiểu nguyên tắc này một cách quá "máy móc".

2.1.3. Yêu cầu về độ chính xác

Khi xây dựng nội dung nghiên cứu, nhà khoa học luôn mong muốn và đòi hỏi độ chính xác của thí nghiệm phải cao. Vì độ chính xác này ảnh hưởng đến kết quả nghiên cứu và có thể cả hiệu quả kinh tế. Song không thể có một độ chính xác chung cho tất cả các nhóm phương pháp thí nghiệm. Độ chính xác của thí nghiệm phụ thuộc vào rất nhiều mặt (khía cạnh), có thể nêu ra một vài khía cạnh là:

a) Điều kiện tiến hành thí nghiệm (thí nghiệm trong phòng khác với thí nghiệm trong chậu vại hay nhà lưới; thí nghiệm ngoài đồng lại khác với thí nghiệm trong phòng và thí nghiệm chậu vại...)

b) Những sai khác về kỹ thuật khi thực hiện thí nghiệm.

c) Độ đồng đều của đất thí nghiệm.

d) Những vết thương cơ giới và tác hại của sâu bệnh.

Những sai khác là không thể tránh được, song sai khác càng nhỏ thì càng tốt. Vì vậy mỗi nhóm phương pháp thí nghiệm khác nhau cho phép có độ chính xác khác nhau thể hiện qua hệ số biến động CV% (Coefficient of variation).

- Nhóm thí nghiệm trong phòng cho phép sai số thí nghiệm $CV\% \leq 1\%$.

- Nhóm thí nghiệm trong chậu, vại, nhà lưới $CV\% \leq 5\%$

- Nhóm thí nghiệm ngoài đồng cho phép sai số thí nghiệm:

+ Các thí nghiệm giống $CV\%$ từ 6% - 8 %.

+ Các thí nghiệm phân bón từ 10 - 12%.

+ Thí nghiệm bảo vệ thực vật (BVTV) từ 13 - 15%.

+ Thí nghiệm cây ăn quả $CV\% \leq 20\%$.

+ Thí nghiệm về lúa $CV\%$ khoảng 10%.

+ Các thí nghiệm điều tra thì thay đổi trong khoảng 20 - 30%.

Ngoài ra người chủ trì thí nghiệm cần phải nắm vững và thường xuyên bao quát thí nghiệm. Các cán bộ kỹ thuật khi thực hiện thí nghiệm phải thành thạo công việc và hết sức có trách nhiệm với việc được phân công và khi thiết kế phải sát với điều kiện đặt ra của thí nghiệm. Theo các nguyên nhân đã nêu có thể coi các sai số thí nghiệm là sai số ngẫu nhiên. Tuy nhiên, trong thực tế không chỉ có duy nhất sai số ngẫu nhiên mà còn tồn tại hai loại sai số khác nữa là: sai số thô (hay còn gọi là sai lầm) và sai số hệ thống.

Khi gặp phải sai số thô thì phải loại bỏ các số liệu ra khỏi dãy kết quả nghiên cứu (thí dụ như đo sai, cân sai, ghi nhầm). Sai số thô không phải là phổ biến.

Sai số hệ thống là do dụng cụ thí nghiệm.

Thí dụ như cân nhẹ hơn hoặc nặng hơn tiêu chuẩn. Thước đo chưa chuẩn hoặc hoá chất pha không được chuẩn như hướng dẫn của hoá chất tiêu chuẩn đặt ra... Loại sai số này tuy không làm ảnh hưởng tới việc so sánh kết quả thí nghiệm giữa các công thức với nhau nhưng khi công bố các giá trị cụ thể của từng chỉ tiêu (ước lượng điểm, ước lượng khoảng của các tham số thống kê) thì không chính xác. Vì vậy, có thể dẫn tới việc nhận định đánh giá sai lệch và điều này cũng không có lợi, nhất là khi chuyển giao kết quả nghiên cứu cho sản xuất. Để tránh sai số này tốt nhất trước khi làm thí nghiệm phải chỉnh các dụng cụ và vật tư theo tiêu chuẩn đo lường cho phép. Hoặc nếu như đã mắc phải sai số hệ thống phải tìm cách hiệu chỉnh giá trị quan sát (các số liệu) về giá trị có được với thước đo tiêu chuẩn.

2.1.4. Yêu cầu diễn lại

Khả năng năng diễn lại của thí nghiệm có nghĩa là: khi thực hiện lại thí nghiệm đó với số lượng công thức, nội dung các công thức như cũ cùng trên khoảng không gian (mảnh đất cũ và thời vụ tương tự) sẽ cho kết quả tương tự.

Thí dụ: Theo dõi thời gian sinh trưởng và các thời kỳ phát dục của hai giống ngô lai LVN20 và LVN24 qua các năm 1997 và 1998, các số liệu được thể hiện ở bảng sau (bảng 2.1):

Bảng 2.1. Thời gian sinh trưởng qua các giai đoạn phát dục của 2 giống ngô từ 1997 & 1998 (ngày)

Tên giống	TG từ gieo - tung phần		TG từ gieo - phun râu		Thời gian từ gieo - chín sinh lý	
	T.97	T.98	T.97	T.98	T.97	T.98
LVN20(đ/c)	54	47	56	48	98	90
LVN 24	56	48	58	49	102	96

Ghi chú:

T.97: vụ thu 1997

T.98: vụ thu 1998

TG: thời gian

Qua các số liệu có thể cho thấy kết quả của hai giống đã có tính diễn lại (có chiều hướng thể hiện tính quy luật).

Tuy nhiên, không nên hiểu nguyên tắc này một cách cứng nhắc, bởi vì điều kiện ngoại cảnh không thể hoàn toàn như nhau khi làm thí nghiệm. Chính vì vậy phải làm lại thí nghiệm trong vài năm (hoặc vài vụ) liên tiếp, hy vọng từ đó sẽ tìm ra tính quy luật của vấn đề nghiên cứu.

Thí nghiệm có khả năng diễn lại càng cao thì việc rút ra kết luận càng chắc chắn. Có nghĩa là đã giải quyết được mối quan hệ giữa các nhân tố thí nghiệm (yếu tố thí nghiệm) với ngoại cảnh trong sự biểu hiện của cây trồng thí nghiệm. Thí nghiệm không có khả năng diễn lại thì không thể đưa ra được kết luận làm cơ sở xây dựng các biện pháp kỹ thuật canh tác và lại càng không thể xây dựng được lý thuyết khoa học. Theo chúng tôi phải lặp lại thí nghiệm ít nhất là 3 vụ (hay 3 năm).

2.1.5. Yêu cầu về lịch sử khu đất canh tác

Thí nghiệm phải được đặt trên các khu đất có lịch sử canh tác rõ ràng. Đây là yêu cầu hết sức cần thiết đối với mỗi thí nghiệm đồng ruộng. Hầu hết trong nội dung thí nghiệm thì đất đai nơi đặt thí nghiệm là yếu tố không thí nghiệm. Đất chỉ là điều kiện (giá đỡ) cho cây mà thôi. Một số biện pháp kỹ thuật có ảnh hưởng tới đất cũng có thể làm cho đất tốt hơn (khoẻ hơn) nếu như biết sử dụng và ngược lại có thể làm cho đất bị thoái hoá. Vì vậy, cần phải biết rõ quá trình canh tác của khu đất trước khi đặt thí nghiệm nghiên cứu.

Thí dụ: Có một giống lúa được cấy thử nghiệm tại 2 thửa ruộng khác nhau trên cùng 1 cánh đồng trong vụ xuân. Kết quả thí nghiệm đã chỉ ra rằng:

Tại ruộng thứ nhất: bón 1 kg N (nguyên chất) thì bội thu 5 kg thóc

Tại ruộng thứ hai: bón 1 kg N (nguyên chất) thì bội thu 10 kg thóc

Khi xem xét nguyên nhân của việc tăng hiệu quả bón N thấy mảnh ruộng thứ nhất trước đây trồng lạc còn ruộng thứ hai trồng ngô nên hiệu quả bón N trên mảnh ruộng thứ hai rõ ràng là cao hơn trên mảnh ruộng thứ nhất.

Do đó khi xem xét lịch sử canh tác của ruộng thí nghiệm cần lưu ý:

- Không đặt ruộng thí nghiệm nằm kề sát các trục đường giao thông lớn, mà nên cách từ 10 - 20m.

- Không đặt ruộng thí nghiệm nằm sát các hệ thống dẫn nước thải của các khu dân cư, bệnh viện, các khu công nghiệp.

- Không đặt thí nghiệm trên đất mới khai hoang, đất này phải làm thí nghiệm trắng vài vụ để san bằng độ đồng đều sau đó mới làm thí nghiệm.

Các thí dụ nêu trên cho thấy đất định đặt thí nghiệm cần phải có lịch sử canh tác rõ ràng.

2.2. Các loại thí nghiệm ngoài đồng ruộng

Hiện nay có nhiều cách phân loại thí nghiệm đồng ruộng. Thông thường có thể phân thành các loại sau:

2.2.1. Thí nghiệm thăm dò

Hay còn gọi là thí nghiệm sơ bộ, thí nghiệm khảo sát. Mục đích của loại thí nghiệm này là nhằm xây dựng những nhận thức ban đầu về đối tượng nghiên cứu trên đồng ruộng để có cơ sở xây dựng các nội dung nghiên cứu chính sau này được tốt hơn. Do đó, thí nghiệm này thường làm trên diện tích nhỏ nhắc lại ít lần và có thể không nhắc lại. Có nghĩa là một công thức có thể chỉ có 1 ô thí nghiệm, số công thức ít và không đi sâu phân tích về cây và đất đai, chỉ quan sát, đánh giá các biểu hiện của cây với các biện pháp thí nghiệm và theo dõi một số chỉ tiêu có tính chất cơ bản về năng suất.

2.2.2. Thí nghiệm chính thức

Nhóm này còn có tên gọi là thí nghiệm chủ yếu. Đây là thí nghiệm đặt ra nhằm giải quyết nội dung cơ bản của vấn đề nghiên cứu. Do đó, thí nghiệm này phải thực hiện đúng như thiết kế đã xây dựng, phải tuân thủ các yêu cầu đặt ra (ở phần 2.1). Tùy thuộc vào loại cây trồng (cây hàng năm hay cây lâu năm; cây hàng hẹp hay cây hàng rộng), loại hình thí nghiệm, mục đích nghiên cứu có thể chia thí nghiệm chính thành các loại khác nhau theo số lượng nhân tố, thời gian và khối lượng nghiên cứu.

2.2.2.1. Theo số lượng nhân tố thí nghiệm

*** Thí nghiệm một nhân tố**

Là thí nghiệm mà trong thành phần của yếu tố thí nghiệm chỉ có một nhân tố tham gia (nhân tố này có quyền thay đổi giữa các công thức) để nghiên cứu tác động của nó đến sự thay đổi của kết quả thí nghiệm.

Thí dụ: Nghiên cứu ảnh hưởng của phân lân đến quá trình sinh trưởng, phát triển và năng suất đậu tương.

Như vậy yếu tố thí nghiệm ở đây là lân nên được phép thay đổi ở các mức bón khác nhau. Còn các nhân tố khác đều phải được thực hiện đồng đều (yếu tố không thí nghiệm).

*** Thí nghiệm nhiều nhân tố**

Đây là thí nghiệm mà trong thành phần của yếu tố nghiên cứu có mặt từ hai nhân tố thí nghiệm trở lên.

Trong thí nghiệm này người ta nghiên cứu ảnh hưởng đồng thời của các nhân tố đối với cây trồng. Đây là những thí nghiệm phức tạp và thường là bước nghiên cứu tiếp của các thí nghiệm một nhân tố.

Để giúp cho thí nghiệm này có độ chính xác cao đôi khi phải chia cụ thể thí nghiệm hai nhân tố, 3 nhân tố... Bởi vì phải có cách sắp xếp ngoài đồng cho phù hợp với số lượng nhân tố thì mới xử lý kết quả bằng các mô hình thống kê tương ứng nhằm tăng tính chính xác..

2.2.2.2. Chia theo thời gian nghiên cứu

*** Thí nghiệm ngắn hạn**

Thường gọi là thí nghiệm ít năm. Đây là loại thí nghiệm nghiên cứu trong một thời gian ngắn đã có thể rút ra được kết luận.

Thông thường loại này được áp dụng để nghiên cứu tác dụng của một biện pháp kỹ thuật cụ thể với cây trồng (thường là các cây hàng năm).

Thí dụ: Nghiên cứu về mật độ cấy, về thời vụ của một giống cây trồng như lúa, ngô, đậu đỗ, rau...

Những giống này cũng chỉ có tính chất thời vụ và thường là không tồn tại dài hàng chục năm.

*** Thí nghiệm dài hạn (thí nghiệm lâu năm)**

Đây là loại hình thí nghiệm cần có thời gian hàng chục năm nghiên cứu liên tục mới có thể đưa ra kết luận. Cá biệt có thí nghiệm phải hàng trăm năm.

Thí dụ: Nghiên cứu tác động của thuốc bảo vệ thực vật (hoá học) đến đất. Hay nghiên cứu hiệu quả cải tạo đất của phương thức canh tác hữu cơ trong nông nghiệp.

Một thí dụ nữa là nghiên cứu hiệu lực của phân lân đến năng suất và chất lượng của nhãn vải, hay xoài...

Với khoảng thời gian từ 5 năm trở lên thì các tác động của nhân tố nghiên cứu mới thể hiện kết quả.

Các thí nghiệm và công trình chọn tạo giống cây trồng vật nuôi đều thuộc loại này.

2.2.2.3. Theo khối lượng nghiên cứu

Có thể chia ra

*** Thí nghiệm đơn độc (độc lập)**

Các thí nghiệm làm ở nhiều nơi và không có liên quan gì với nhau cả.

Thường thì kết quả có tính chính xác cao, đúng cho một điều kiện cụ thể. Song tính phổ biến lại hẹp thậm chí rất hẹp.

*** Thí nghiệm hệ thống**

Đây là những thí nghiệm làm ở nhiều nơi và có liên hệ với nhau theo những khía cạnh nhất định mà người chủ trì đặt ra.

Thí nghiệm này có nhược điểm là khối lượng lớn, tốn công sức và vật chất, tốn thời gian; có thể cách xa nhau về địa lý, khác nhau thời tiết và đất đai (điều kiện sinh thái), về tập quán và điều kiện kinh tế - xã hội.

Ưu điểm của nhóm này là thí nghiệm mang tính đa dạng và khi kết quả thành công có phổ áp dụng rộng rãi.

2.2.3. Thí nghiệm làm trong điều kiện sản xuất

Loại thí nghiệm này còn có tên gọi là thực nghiệm khoa học, thực nghiệm đồng ruộng. Với chuyên ngành chọn giống và nhân giống còn gọi là thí nghiệm khảo nghiệm hay khu vực hoá giống mới đó.

Đây là những thực nghiệm cần phải được thẩm định lại trong điều kiện tự nhiên trước khi chuyển giao kỹ thuật sản xuất cho nông dân.

Loại này khối lượng lớn có thể nhắc lại nhiều hay ít tùy thuộc vào điều kiện kinh tế và đất đai. Không cần theo dõi quá chi tiết các chỉ tiêu về sinh trưởng của cây mà chủ yếu là quan sát tình hình sinh trưởng, nhiễm sâu bệnh để đưa ra các nhận định chung về phản ứng của cây với điều kiện tự nhiên, nhưng cần quan tâm cụ thể đến năng suất và hiệu quả kinh tế.

2.3. Xây dựng chương trình thí nghiệm

2.3.1. Một số vấn đề liên quan đến xây dựng chương trình thí nghiệm

2.3.1.1. Các loại công thức trong thí nghiệm

Trong một thí nghiệm có hai loại công thức đó là:

* *Loại 1:* Công thức đối chứng hay còn gọi là công thức tiêu chuẩn. Công thức đối chứng được đặt ra làm tiêu chuẩn cho các công thức khác trong thí nghiệm so sánh để rút ra hiệu quả cụ thể của nhân tố (biện pháp) nghiên cứu.

Công thức đối chứng được xây dựng như sau

- Nếu là thí nghiệm giống thì giống trong công thức đối chứng thường là giống tốt đang được sản xuất của địa phương chấp nhận, cũng có thể là một giống tiêu chuẩn của Nhà nước hay của một vùng nhất định đã được chính thức công nhận (Bộ Nông nghiệp và PTNT công nhận).

- Nếu là biện pháp kỹ thuật khác như mật độ, thời gian gieo cấy thì công thức đối chứng là biện pháp phù hợp với điều kiện sản xuất và đang được sản xuất nơi đó chấp nhận.

- Nếu là các nhân tố như bón phân, phun thuốc bảo vệ thực vật, tưới nước... mà những nhân tố này là cho thêm vào nghiên cứu bằng việc cân, đong, đo, đếm cụ thể được thì công thức đối chứng thường được xây dựng ở mức 0 (không cho thêm vào).

Thí dụ:

- Thí nghiệm về thời vụ thì công thức đối chứng là chính vụ.

- Thí nghiệm về mật độ thì công thức đối chứng là mật độ phù hợp với cây trồng trong vụ chính đó.

- Thí nghiệm về thuốc bảo vệ thực vật thì công thức đối chứng là không phun (nhưng phải phun nước thông thường được dùng pha thuốc cho các công thức có nồng độ khác nhau). Nhưng nếu là thuốc dùng bón vào đất thì công thức đối chứng là không dùng thuốc (không cần phun nước)

Trong một thí nghiệm ít nhất phải xây dựng một công thức đối chứng. Còn tùy thuộc vào các điều kiện cụ thể khi làm thí nghiệm và nội dung nghiên cứu mà có thể tới hai hay ba công thức đối chứng, để khi so sánh được chính xác.

* *Loại 2:* Công thức nghiên cứu

Công thức nghiên cứu là công thức được tác động biện pháp kỹ thuật (nhân tố thí nghiệm) ở các mức độ khác nhau.

Kết quả của công thức này được so sánh với kết quả của công thức đối chứng.

Cả hai loại công thức đối chứng và công thức nghiên cứu đều gọi chung là các công thức thí nghiệm hay nghiệm thức.

Để xây dựng công thức thí nghiệm cần lưu ý những vấn đề sau:

Một là: Cần tạo điều kiện để các công thức thí nghiệm so sánh được với nhau và so sánh được với công thức đối chứng.

Thí dụ: Nghiên cứu hiệu lực của phân lân đến sinh trưởng, phát triển, khả năng nhiễm sâu bệnh và năng suất lúa.

Như vậy tùy thuộc vào điều kiện cụ thể của thí nghiệm lúa ở đất nào, vụ nào mà xây dựng số lượng công thức thí nghiệm là mấy công thức.

Song ít nhất cũng phải có 3 công thức, trong đó

Công thức I: là không bón lân vào đất ($0 P_2O_5$) (CT I là đối chứng)

Công thức II: là bốn lần ở mức độ 1 (CT II)

Công thức III: là bốn lần ở mức độ 2. (CT III)

Như vậy, sẽ so sánh được hiệu lực của việc có bốn lần so với không bốn lần tới các chỉ tiêu nghiên cứu ở cây lúa. Sau đó cho phép so sánh là trong hai mức bốn lần cụ thể thì mức nào đạt kết quả cao hơn và mức nào đạt hiệu quả đầu tư lên cao hơn.

Hai là: Cần tạo điều kiện cho các công thức thí nghiệm thể hiện được quy luật số lượng.

Để thí nghiệm thành công người chủ trì xây dựng nội dung nghiên cứu hiểu rất rõ đối tượng nghiên cứu và tác dụng (công năng) của biện pháp kỹ thuật đến đối tượng nghiên cứu muốn vậy phải có kiến thức tổng hợp các môn học từ cơ sở đến chuyên môn trong điều kiện cụ thể

Thí dụ: Nghiên cứu hiệu quả của việc trồng xen giữa cà phê với (Coffee Canephora) với sầu riêng trong vườn cà phê tại Đắk Lắk

Công thức I: Cà phê thuần mật độ $(3 \times 3)m = 1111$ cây/ha (đ/c)

Công thức II: Sầu riêng xen cà phê mật độ $(6 \times 9)m$

Công thức III: Sầu riêng xen cà phê mật độ $(9 \times 9)m$

Công thức IV: Sầu riêng xen cà phê với mật độ $(9 \times 12)m$.

Công thức V: Sầu riêng xen cà phê mật độ $(12 \times 12)m$

Công thức VI: Sầu riêng xen cà phê mật độ $(12 \times 15)m$

Như vậy giữa các công thức nghiên cứu (các mật độ trồng) có khoảng cách là 3 m, với khoảng cách này hy vọng giữa các công thức khác nhau thể hiện kết quả khác nhau và cũng từ kết quả nghiên cứu có thể tìm ra khoảng cách trồng xen hợp lý nhất.

2.3.1.2. Xác định số lượng công thức trong một thí nghiệm

Số lượng công thức thí nghiệm được xác định tùy thuộc vào nội dung và mục đích của người chủ trì thí nghiệm. Thông thường thí nghiệm một nhân tố có số lượng công thức ít hơn thí nghiệm hai nhân tố.

Trong thí nghiệm có một nguyên tắc cơ bản giúp các nhà khoa học xác định được có bao nhiêu công thức là vừa đủ cho nội dung nghiên cứu của mình. Nguyên tắc đó là: "Dựa vào giả thiết khoa học để lập ra công thức trung tâm, từ công thức trung tâm sẽ xê dịch lên phía trên một số mức và xuống phía dưới một số mức".

Còn khoảng cách giữa các mức tùy thuộc vào tác động của nhân tố nghiên cứu tới đối tượng được sử dụng trong thí nghiệm (vật liệu thí nghiệm).

Như vậy, với thí nghiệm một nhân tố có bao nhiêu mức (liều lượng) thì có bấy nhiêu công thức kể cả mức đối chứng có thể là 0 (không cho thêm vào).

Thí dụ: Nghiên cứu ảnh hưởng của phân N tới năng suất lúa

Đây là thí nghiệm một nhân tố. Nếu như lấy công thức trung tâm là 100 N hoặc 150N thì các công thức được xây dựng xung quanh công thức trung tâm như sau:

Công thức 1 (CTI): 0 N (đối chứng)

CTII: bốn 50 N

CTIII: bốn 100N

CTIV: bốn 150 N

CTV: bốn 200 N

CTVI: bốn 250 N

Như vậy thí nghiệm có thể có 5 công thức (dùng lại mức 200N) hoặc có thể 6 công thức như đã viết. Không nên đặt ít hơn 5 hoặc nhiều hơn 6. Độ rộng hay hẹp của N giữa các công thức phụ thuộc vào vai trò của đạm (N) và đối tượng nghiên cứu. Thí dụ đối với cây lúa và cây ngô độ rộng như thí nghiệm đã nêu là phù hợp, còn nếu là cây họ đậu (đậu tương, lạc..., thì độ rộng như thí nghiệm trên lại cao và mức bón 250N có thể là thừa. Còn trong thí nghiệm nhiều nhân tố thì một nguyên tắc chung là: Để nghiên cứu được đầy đủ tác động phối hợp các nhân tố, số lượng công thức thí nghiệm là tích của số mức của mỗi nhân tố thí nghiệm.

Thí dụ: Nghiên cứu hiệu lực của lân tới năng suất ở hai giống đậu tương vụ đông tại vùng Đồng Bằng sông Hồng.

Như vậy đây là thí nghiệm hai nhân tố (lân là nhân tố ký hiệu là A với số mức là L_a).

Giống đậu tương là B với số giống tham gia là L_b).

Vậy gọi số công thức nghiên cứu là K thì

$$K = L_a \times L_b \quad (2.1)$$

Bảng 2.2. Có thể mô tả tổ hợp các công thức như sau

Giống (A)	Lân (P_2O_5) (B)	Công thức	STT CT	Giống	Lân (P_2O_5)	Công thức	STT CT
Giống a_1	b_0	a_1b_0	I	Giống a_2	b_0	a_2b_0	V
	b_1	a_1b_1	II		b_1	a_2b_1	VI
	b_2	a_1b_2	III		b_2	a_2b_2	VII
	b_3	a_1b_3	IV		b_3	a_2b_3	VIII

Thí nghiệm có 8 công thức trong đó có 2 công thức đối chứng là a_1b_0 : đối chứng 1

a_2b_0 : đối chứng 2. Trong thí dụ này số công thức $K = 2 \times 4 = 8$ công thức

Với thí nghiệm hai nhân tố thông thường có nhiều cách thiết kế (sắp xếp ô thí nghiệm trên đồng ruộng sẽ đề cập kỹ hơn ở hệ đào tạo sau đại học).

Nếu thí nghiệm ba nhân tố có ký hiệu (nhân tố 1 là A có L_a mức, nhân tố 2 là B có L_b mức và nhân tố thứ 3 là C có L_c mức). Vậy số lượng công thức sẽ là:

$$K = L_a \times L_b \times L_c \quad (2.2)$$

Trong giáo trình này chúng tôi tập trung chủ yếu vào thí nghiệm một nhân tố. Còn thí nghiệm từ 2 nhân tố trở lên sẽ được đề cập ở hệ đào tạo sau đại học.

Thực tế các nhà khoa học đều biết rằng thí nghiệm nghiên cứu tác dụng đồng thời của nhiều nhân tố đến cây trồng là tốt nhất vì theo quan điểm sinh thái học tác dụng tổng hợp không phải là cấp số cộng đơn thuần của các tác động riêng lẻ độc lập. Song nhiều nhân tố sẽ có nhiều công thức thí nghiệm và như vậy, sẽ nảy sinh một loạt các vấn đề khác mà hầu hết các vấn đề đó lại nằm trong yếu tố không thí nghiệm, rất khó tạo được tính đồng nhất cao. Cho nên không nên quá nhiều nhân tố trong một thí nghiệm và các nhà khoa học chỉ nên nghiên cứu mối quan hệ đồng thời của 2 hay 3 nhân tố mà thôi.

2.3.1.3. Một số chỉ tiêu kỹ thuật có liên quan khi thiết kế thí nghiệm

* Diện tích ô thí nghiệm

Ô thí nghiệm là thành phần cơ bản của thí nghiệm trên đó các nhân tố thí nghiệm được thực hiện theo một nội dung đã xây dựng. Do đó, diện tích ô (cụ thể là kích thước ô) phải như

nhau. Nếu lấy độ chính xác của thí nghiệm là hệ số biến động CV% để khảo sát thì giá trị này phụ thuộc vào những khía cạnh sau:

- Diện tích ô thí nghiệm
- Hình dáng ô thí nghiệm (kích thước ô là vuông hay là chữ nhật...).
- Địa hình đặt thí nghiệm.
- Loại đất thí nghiệm.
- Loại cây trồng.
- Loại thí nghiệm.

Chúng tôi xin đưa ra những dẫn liệu cụ thể của một số tác giả xung quanh vấn đề này.

Sanhina (1965) đưa ra mối quan hệ giữa sai số thí nghiệm với việc tăng diện tích ô thí nghiệm ở lúa mì tại Nga, kết quả như số liệu (bảng 2.3).

Bảng 2.3. Quan hệ giữa sai số thí nghiệm với diện tích ô thí nghiệm

m ²	20	30	40	50	60	70	80	90	100
Sai số %	3.63	3.17	2.96	2.82	2.61	2.45	2.37	2.31	2.28

Mudra (1958) cho thấy trên hai loại đất được trồng khoai tây và lúa mì khi diện tích ô thí nghiệm thay đổi thì sai số cũng thay đổi theo (bảng 2.4).

Bảng 2.4. Quan hệ giữa diện tích ô thí nghiệm, loại cây trồng trên hai loại đất với sai số thí nghiệm

Ngô trồng trên đất bằng		Lúa mì trồng trên đất cát	
Diện tích ô (m ²)	Sai số (CV%)	Diện tích ô (m ²)	Sai số(CV%)
1	19,6	1	30,0
2	17,7	2	28,0
3	15,2	4	27,0
8	9,8	6	26,0
10	7,5	8	25,0
12	6,3	10	24,0
16	3,8	12	23,0
20	2,9	16	22,0
25	2,3	25	22,0
33	1,8	31	22,0
56	1,3	48	22,0

Phạm Chí Thành và Nguyễn Thị Lan (1983) nghiên cứu trên lúa cây tại đất phù sa trong đê sông Hồng cho mối quan hệ giữa sai số thí nghiệm, với diện tích ô (cụ thể là hình dáng ô thí nghiệm) qua các dẫn liệu cụ thể sau:

Bảng 2.5. Mối quan hệ giữa sai số thí nghiệm với diện tích ô khi ô có dạng hình vuông

Độ biến động đất TN CV%)	Diện tích ô (m ²)								
	4	9	16	25	36	49	64	81	100
7.76	7,58	8,18	7,27	8,42	5,75	5,64	5,29	5,99	1,65
9.07	10,64	10,08	9,32	9,14	10,28	10,21	10,92	8,19	5,83
16.18	22,23	20,20	17,90	18,00	18,30	18,90	19,40	18,90	19,10

Ghi chú: CV% là độ biến động của đất thí nghiệm

Bảng 2.6. Khi ô thí nghiệm có dạng hình chữ nhật 2 : 1

Độ biến động đất TN (CV%)	Diện tích ô (m ²)					
	2	8	18	32	50	72
7.76	14,08	8,27	6,52	6,19	4,82	5,54
9.07	13,15	10,28	10,87	9,18	7,16	6,77
16.18	18,29	17,90	18,19	20,75	17,57	18,75

Bảng 2.7. Khi ô thí nghiệm có dạng hình chữ nhật 3 : 1

Độ biến động đất TN (CV%)	Diện tích ô (m ²)			
	3	12	27	48
7.76	11,26	8,53	6,29	5,90
9.07	12,09	12,67	9,59	5,68
16.18	20,00	15,85	16,45	18,96

Như vậy, qua dẫn liệu trong các bảng 2.5; bảng 2.6 và bảng 2.7 với lúa cây cho thấy:

- Khi diện tích ô thí nghiệm tăng lên thì sai số thí nghiệm có giảm khá rõ. Nhưng nhận xét này chỉ đúng khi đất thí nghiệm tương đối đồng đều (CV% < 10%).

Còn khi nền đất thí nghiệm có biến động đất CV% > 15% thì nhận xét trên không còn đúng nữa.

Điều này cho thấy sẽ không thể cứ tăng diện tích ô thí nghiệm thì nâng cao độ chính xác của thí nghiệm đồng ruộng khi đất thí nghiệm có độ đồng đều kém.

Thực tế cho thấy là không có giải pháp nào mang tính tuyệt đối được. Mà những mối quan hệ đã nêu cũng chỉ hài hoà ở mức độ nhất định. Vì vậy, diện tích ô thí nghiệm lớn hay nhỏ tùy thuộc vào:

(1) Mục đích và yêu cầu của thí nghiệm

Thí dụ: thí nghiệm lưu giữ, phục hồi một số giống lúa đặc sản địa phương quý hiếm.

Như vậy tùy thuộc vào số lượng nguồn hạt giống thu được là nhiều hay ít mà diện tích ô thí nghiệm cho mỗi giống to nhỏ khác nhau. Tuy nhiên diện tích quá nhỏ sẽ khó giám định và nhận xét.

Những thí nghiệm so sánh giống hoặc một biện pháp kỹ thuật nào đó để chuẩn bị phổ biến ra sản xuất thì diện tích ô phải đủ lớn.

Nếu thí nghiệm trong điều kiện sản xuất (thí nghiệm khảo nghiệm) thì diện tích ô phải lớn hơn thí nghiệm thăm dò và thí nghiệm chính.

(2) Phụ thuộc vào loại cây trồng và mật độ gieo cấy sao cho mỗi một ô thí nghiệm có khả năng tạo ra một quần thể cây trồng mang tính đại diện, ổn định và đảm bảo số lượng cá thể trong quá trình nghiên cứu được đánh giá khách quan gần tương tự như ngoài sản xuất.

(3) Phụ thuộc vào phương tiện kỹ thuật khi thực hiện thí nghiệm

Nếu làm bằng thủ công thì diện tích ô nhỏ, còn nếu làm bằng máy (cơ giới) diện tích ô phải rộng hơn. Như vậy diện tích ô phải phù hợp để có thể hoàn thành công việc thí nghiệm trong khuôn khổ thời gian mà yêu cầu thí nghiệm đặt ra.

Cũng chính do những vấn đề nêu trên mà hiện nay việc xác định một cách chính xác diện tích ô thí nghiệm vẫn còn nhiều ý kiến tranh luận. Xin nêu ý kiến của các tác giả quan tâm tới vấn đề này như sau:

Trên thế giới:

+ Theo Zimmerman (1961) với thí nghiệm về nông học thì:

- Với cây họ hoà thảo diện tích ô thay đổi từ 10 - 20 m²
- Với khoai tây thì diện tích ô từ 20 - 25 m².
- Với củ cải đường thì diện tích ô từ 12 - 20 m²
- Với cây làm phân xanh thì ô có diện tích từ 5 - 25 m².

+ Theo Whyte (1964) và Ivanov (1969):

Tùy thuộc vào điều kiện cụ thể mà xác định diện tích ô thí nghiệm. Song tối thiểu phải đạt từ 15 - 20 m².

+ Theo Kwanchai, A. Gomez và Arturo.A. A. Gomez (1986) với thí nghiệm lúa diện tích ô có thể thay đổi từ 5 - 15 m².

Các nhà khoa học của Liên Xô cũ thì trong thí nghiệm so sánh giống

- Cây trồng dày: 100 m².
- Cây hàng rộng: 150 - 200 m².
- Với rau lại quy định theo đơn vị số cây/ô như sau

+ Cải bắp: 140 cây.

+ Cà chua: 115 cây.

+ Cà: 100 cây.

+ Ớt: 180 cây.

+ Dưa: 50 cây

Các nhà khoa học Việt Nam đề xuất

+ Theo Đào Thế Tuấn (1960) thì:

Bảng 2.8. Diện tích ô thí nghiệm (đơn vị m²/ô)

Loại thí nghiệm	Cây trồng	
	Lúa	Cây hàng rộng
- Thí nghiệm ngoài đồng kết hợp với thí nghiệm trong phòng	10 - 20	20 - 40
- Thí nghiệm nhỏ	20 - 50	40 - 100
- Thí nghiệm lớn	50 - 100	100 - 200
- Thí nghiệm cơ giới	100 - 200	200 - 400

+ Tác giả Đinh Văn Lữ đề nghị:

- Với các loại cây có kích thước tương đối to như: Bông, ngô thì diện tích ô là 50 – 100 m².

- Với các cây nhỏ như: Lúa, lúa mì... thì diện tích ô thường là từ 30 – 60 m² trên 1 ô.

Theo ý kiến của một số tác giả và chúng tôi thì diện tích ô thí nghiệm rộng hay hẹp tùy thuộc vào số lượng cây cần thiết để quan sát ở các chỉ tiêu nghiên cứu trong suốt quá trình thí nghiệm. Nên đề xuất như sau:

- Với cây hàng năm diện tích ô từ 50 – 100 m².

- Cây rau diện tích ô từ 30 - 40 m².

- Cây ăn quả lâu năm: 10 - 20 cây.

- Thí nghiệm với cây thuốc: 10m².

* *Hình dạng ô thí nghiệm*

Hình dạng ô thí nghiệm được xây dựng bởi tỷ lệ giữa 2 cạnh của ô thí nghiệm, nếu tỷ lệ là 1 (1:1) ô có dạng hình vuông. Còn các tỷ lệ khác 1 là hình chữ nhật.

Ở nội dung này cũng một phần có liên quan tới diện tích ô như đã nêu trên.

Theo Ivanov (1969), với một diện tích ô xác định thì ô càng dài (tỷ lệ rộng/dài ô càng < 1) thì càng chính xác. Do dễ dàng gieo trồng, chăm sóc và độ phì nhiêu của đất càng gần nhau giữa các ô và trong cùng ô.

Whyte (1964) lại đề xuất: Nếu diện tích ô thí nghiệm nhỏ thì hình dạng ô nên vuông để giảm diện tích phân bảo vệ.

Theo Phạm Chí Thành và Nguyễn Thị Lan (1983) thay đổi hình dạng ô thí nghiệm cũng là một biện pháp để nâng cao độ chính xác của thí nghiệm. Tuy nhiên, nó phụ thuộc vào độ đồng đều của đất thí nghiệm. Điều này thể hiện qua dẫn liệu trong bảng sau:

Bảng 2.9. Quan hệ giữa hình dạng ô với độ chính xác của thí nghiệm lúa cây khi có độ đồng đều đất khác nhau (diện tích ô = 50 m²)

Biến động đất thí nghiệm (CV%)	Ô hình vuông (7m x 7m)	Ô chữ nhật 1 : 2 (5m x 10m)	Ô chữ nhật 1 : 3 (4m x 12m)
7,76	5,64	4,82	5,90
9,07	10,21	7,16	5,60
16,18	18,90	17,57	18,96

Qua đây cho phép nhận xét:

- Ở điều kiện đất tương đối đồng đều việc thay đổi hình dạng ô thí nghiệm không có ảnh hưởng đáng kể đến độ chính xác của thí nghiệm.

- Ở điều kiện đất ít đồng đều khi ô hình chữ nhật thì độ chính xác sẽ càng cao (vì bao gồm được các vị trí không đồng đều của đất trong 1 ô).

- Khi đất không đồng đều thì việc thay đổi hình dạng ô thí nghiệm không làm cho độ chính xác của thí nghiệm nâng lên.

Tác giả Phạm Chí Thành (1971) qua kết quả nghiên cứu về thuốc bảo vệ thực vật đã rút ra nhận xét:

Trong thí nghiệm này nếu ô thí nghiệm càng chữ nhật thì độ chính xác của thí nghiệm càng kém (ảnh hưởng kế cận giữa các ô).

Phạm Chí Thành và Nguyễn Thị Lan (1983) đã rút ra kết luận: trong thí nghiệm về phân bón với lúa cây có sự di chuyển dinh dưỡng từ ô nọ sang ô kia (với thí nghiệm không đắp bờ

ngăn giữa các ô, mà chỉ để khoảng cách 40 cm). Sự di chuyển này là nhiều hay ít, phụ thuộc vào tốc độ di chuyển của nước ruộng thí nghiệm.

Như vậy, tùy thuộc vào loại hình thí nghiệm (đúng hơn là tính chất của thí nghiệm theo nội dung nghiên cứu) mà chọn hình dạng ô cho phù hợp để vừa đảm bảo độ chính xác vừa thoả mãn được điều kiện đất làm thí nghiệm và thuận tiện trong công tác nghiên cứu.

Nhìn chung từ đề xuất của các tác giả nêu trên. Chúng tôi thấy:

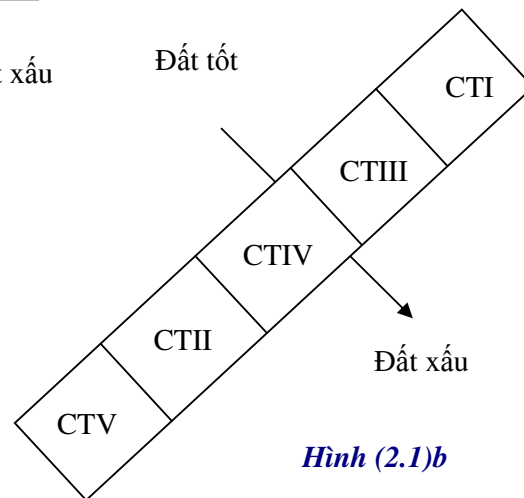
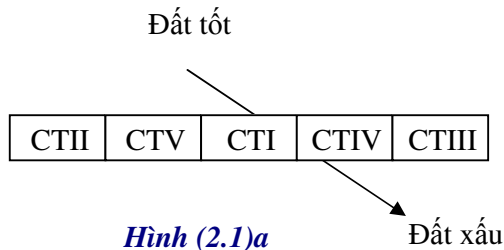
- Thí nghiệm về cách làm đất; tưới nước và so sánh giống nên làm dài (ô chữ nhật). Với cây hàng hẹp chiều rộng là 1,50 m, cây hàng rộng có 4 - 6 hàng.

- Thí nghiệm về thuốc (BVTV) và phân bón nên làm hình vuông hoặc gần vuông.

Song đây cũng chỉ là những đề xuất còn vấn đề cụ thể phải xem xét ở nhiều khía cạnh khác nhau cho hài hoà.

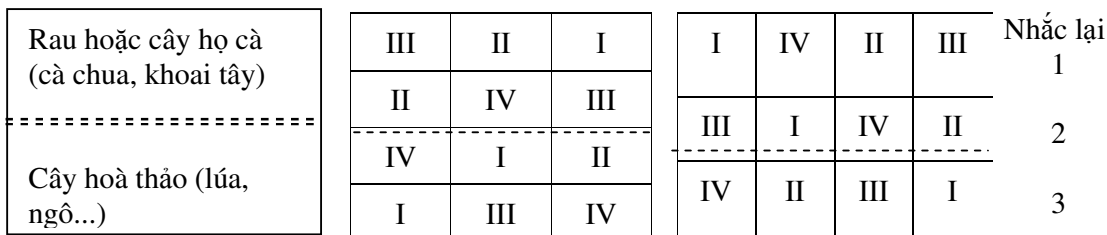
* *Hướng của ô thí nghiệm*

Nhìn chung vấn đề này ít có liên quan tới độ chính xác của thí nghiệm nếu như đất thí nghiệm đồng đều. Còn khi đất thí nghiệm có độ biến động lớn thì hướng ô thí nghiệm có ảnh hưởng đến độ chính xác của thí nghiệm. Và trong thí nghiệm này, ô thí nghiệm nên là hình chữ nhật. Nếu đất biến đổi về độ đồng đều theo một hướng xác định thì chiều dài ô thí nghiệm nên song song với hướng biến động của đất. Minh hoạ hình 2.1. và 2.2.



Trong sơ đồ (2.1): hình (2.1)a là không hợp lý có thể 1 công thức nào đó nằm vào đất tốt, hoặc lại rơi vào đất xấu hoàn toàn. Còn hình (2.1)b là hợp lý, do các công thức đều có các vị trí đất tốt, xấu là tương đương.

Còn ở đất đồi dốc độ phì của đất (độ đồng đều) thay đổi từ đỉnh đồi xuống chân đồi - nên chiều dài của ô nên vuông góc với đường đồng mức. Các lần nhắc lại càng phải nằm ở cùng một bình độ chạy vòng quanh đồi. Còn nếu như cần phải làm thí nghiệm trên mảnh đất trước đó trồng các loại cây khác nhau nên sắp đặt thí nghiệm như hình (2.2)c là phù hợp nhất, còn hình 2.b là không phù hợp.



Hình (2.2) a
Khu đất thí nghiệm

Nhắc lại
1 2 3
I; II; III; IV là ký hiệu
công thức

Hình (2.2) c

Hình (2.2) b

** Số lần nhắc lại*

Nhắc lại của thí nghiệm có nghĩa là: mỗi công thức thí nghiệm được thực hiện ở một số ô trong cùng mảnh đất thí nghiệm hoặc khu vực thí nghiệm (gọi là nhắc lại không gian), trong cùng một thời gian thí nghiệm.

Thí nghiệm phải được nhắc lại (nhất là thí nghiệm chính). Đây có thể coi là một tiêu chuẩn bắt buộc khi nghiên cứu. Song số lần nhắc lại bao nhiêu là hợp lý?

Theo tác giả Remera dẫn từ Dospekhov (1979) thì:

- Khi tăng số lần nhắc lại độ chính xác của thí nghiệm tăng nhiều hơn khi tăng diện tích ô thí nghiệm.

- Khi tăng số lần nhắc lại thì ban đầu độ chính xác tăng nhanh, nhưng nếu tiếp tục tăng nhắc lại nhiều lần (lớn hơn 4 lần) thì độ chính xác có giảm, song giảm chậm.

Theo Phạm Chí Thành và Nguyễn Thị Lan (1983) ở thí nghiệm lúa cấy, cho thấy mối quan hệ giữa diện tích ô, số lần nhắc lại với sai số thí nghiệm như sau:

Bảng 2.10. Quan hệ giữa tăng diện tích ô (m²), số lần nhắc lại với độ chính xác của thí nghiệm (ở nền đất thí nghiệm CV% = 9,07%).

Diện tích ô (m ²)	CV% (không nhắc lại)	Cố định ô 25 m ²	
		Số lần nhắc	CV%
25	10.0	1	10.0
50	8.3	2	7.1
75	7.6	3	5.8
100	7.1	4	5.0

125	6.7	5	4.5
150	6.4	6	4.1
175	6.1	7	3.8
200	5.9	8	3.5
225	5.7	9	3.3
250	5.6	10	3.2

Ghi chú: CV% là sai số của thí nghiệm

Theo Đào Thế Tuấn (1960) tùy các nhóm nghiên cứu thí nghiệm mà số lần nhắc lại có thể dao động từ 2 - 6 lần.

Từ các ý kiến trên chúng ta có thể nhận định rằng số lần nhắc lại nhiều hay ít của mỗi công thức thí nghiệm phụ thuộc vào:

- Độ đồng đều của đất thí nghiệm.

- Kích thước cây trồng (Có liên quan tới diện tích ô). Cây có kích thước nhỏ có thể nhắc lại ít hơn. Vì trên 1 ô thí nghiệm có nhiều cây. Còn cây có kích thước lớn nhắc lại nhiều, vì trên một ô có số lượng cây ít (thậm chí nhắc lại tới 10 lần).

- Diện tích ô thí nghiệm: Khi diện tích ô nhỏ thì nhắc lại nhiều hơn diện tích ô lớn.

Nhưng tựu chung lại là: Theo chúng tôi số lần nhắc lại với thí nghiệm nên ≥ 3 , tùy theo kiểu sắp xếp thí nghiệm. Nên thường thường từ 3 - 4 lần nhắc lại là vừa. Còn với thí nghiệm thăm dò và thí nghiệm khảo nghiệm số lần nhắc lại có thể < 3 .

* Dải bảo vệ và hàng biên

+ Dải bảo vệ là phần diện tích bao quanh mảnh đất (khu đất) làm thí nghiệm. Đây là nội dung đặt ra nhằm đảm bảo nguyên tắc "Sai khác duy nhất". Tùy thuộc vào điều kiện của khu đất thí nghiệm mà độ rộng của dải bảo vệ có thể rộng bao nhiêu. Song nhìn chung tối thiểu độ rộng này phải đạt từ 1,5 m - 2,0 m.

+ Hàng biên

Hàng biên chính là phần bảo vệ của ô thí nghiệm nhằm loại trừ ảnh hưởng giữa các ô với nhau (đặc biệt các ô khác công thức).

Theo K.A.Gomez và A.A. Gomez (1986) trong thí nghiệm lúa có bốn phân N và giữa các ô không có bờ ngăn chỉ để khoảng trống giữa các ô thay đổi từ 40 - 100 cm thì năng suất hàng biên thay đổi từ 50% đến 100%. Phạm vi ảnh hưởng chỉ ở hàng thứ nhất. Riêng bốn 120N/ha có ảnh hưởng tới hàng thứ hai.

Nếu thí nghiệm so sánh giống mà giữa các ô không để cách bờ (không để khoảng cách 40 cm mà chỉ khoảng 20 cm là khoảng cách giữa hàng với hàng) thì có ảnh hưởng tới hàng thứ ba. Vì vậy, với thí nghiệm lúa khoảng cách để loại bỏ hàng biên không lấy mẫu theo dõi là hai hàng xung quanh ô (tương đương 40 cm). Với cây trồng cạn cũng nên loại bỏ một hàng xung quanh (tương đương 40 - 60 cm) là phù hợp với điều kiện thí nghiệm của Việt Nam. Tuyệt đối không lấy mẫu theo dõi tại các vị trí xung quanh rìa ô thí nghiệm để (hàng biên) để thí nghiệm mang tính chính xác và khách quan.

* Cách sắp xếp các công thức thí nghiệm (phần này sẽ được mô tả ở chương VII)

2.3.2. Xây dựng nền thí nghiệm

Nền thí nghiệm là bao gồm tất cả các điều kiện canh tác được thực hiện đồng nhất giữa tất cả các công thức của thí nghiệm.

Nền thí nghiệm không phải là nhân tố so sánh giữa các công thức với nhau. Tuy nhiên, nó giúp cho các công thức nghiên cứu thể hiện được tác dụng. Vì vậy, chọn nền thí nghiệm căn cứ vào các điều kiện sau:

+ Phải đại diện về điều kiện kinh tế - xã hội của thời điểm các kết quả nghiên cứu dự kiến sẽ được áp dụng vào sản xuất.

+ Phải là điều kiện để cho nhân tố thí nghiệm phát huy được hiệu lực của nó.

Do đó, phải hết sức cẩn thận khi xây dựng nền thí nghiệm, để nền thí nghiệm không phải là một tác nhân làm giảm tác dụng của các nhân tố nghiên cứu.

2.3.3. Chọn đất thí nghiệm

Hầu hết các thí nghiệm trong nông học đất là nhân tố không nghiên cứu, mà đất chỉ đóng vai trò làm giá đỡ và là môi trường để cho các nhân tố nghiên cứu thể hiện tác dụng của nó với cây trồng. Nên yêu cầu của đất thí nghiệm phải mang tính đại diện về các mặt như: Loại đất nào? Trong loại đất ấy phải biết được địa hình ra sao? Thành phần hoá tính, lý tính đất có đại diện cho vùng mà sau này thí nghiệm phục vụ không? Có như vậy thí nghiệm mới có tính thực tiễn cao.

Để đạt được các yêu cầu nêu trên, trước khi làm thí nghiệm phải có được các thông tin về đất bao gồm các nội dung sau:

2.3.3.1. Điều tra về địa hình

Địa hình được coi là nhân tố chính quyết định (chi phối) độ đồng đều của đất. Vì vậy, đất thí nghiệm phải phẳng (nhìn bề mặt) song có thể thấy như sau:

Với thí nghiệm lúa nước làm trong điều kiện đất bằng, tùy vào thí nghiệm với lúa cấy hay gieo vào mà yêu cầu cụ thể về độ phẳng mặt ruộng cao hay thấp. Lúa cấy thì độ phẳng thấp hơn so với lúa gieo vãi. Nhìn chung ruộng thí nghiệm có bề mặt làm đất khi gieo cấy càng phẳng càng tốt. Song thực tế không phải là dễ dàng thực hiện được như vậy. Độ phẳng mặt ruộng khi gieo cấy chênh lệch ± 5 cm với thí nghiệm lúa. Nếu làm thí nghiệm lúa nước trên đất dốc phải làm trên các ruộng bậc thang. Còn với thí nghiệm của các cây trồng cạn khác thì chênh lệch độ phẳng mặt ruộng cho phép đạt mức ± 10 cm.

Với các thí nghiệm cây trồng trên đất dốc thì yêu cầu độ dốc cho phép để làm thí nghiệm là 25% và mặt ruộng hay đôi phải dốc đều và tốt nhất cũng nên làm đường đồng mức. Tùy thuộc vào đặc điểm sinh học của từng loại cây trồng mà chọn đất dốc cho phù hợp. Nên bố trí gôn thí nghiệm trên một khu vực (có thể 1 lần nhắc lại trên đất có độ dốc tương tự nhau). Nếu không có điều kiện thì một hoặc hai lần nhắc lại ở một khu vực còn bên cạnh là lần nhắc lại khác. Tuyệt đối tránh việc bố trí gôn một hoặc một số công thức nào đó tại một khu vực và khu vực khác sẽ là một hay một số công thức khác còn lại của thí nghiệm.

2.3.3.2. Điều tra lý tính đất và hoá tính đất

Điều tra đất được đặt ra với mục đích là sau khi thí nghiệm thành công, các kết quả sẽ được ứng dụng ở những vùng đất cùng loại tương tự.

Trước hết nếu có điều kiện nên vẽ bản đồ vùng thí nghiệm với các tỷ lệ phù hợp 1 : 5000 hoặc 1 : 1000. Thí nghiệm đòi hỏi độ chính xác càng cao thì càng phải làm kỹ vấn đề này.

Rất cần quan tâm đến lịch sử canh tác của đất trước lúc đặt thí nghiệm. Nhất là các biện pháp kỹ thuật canh tác trước có khả năng làm thay đổi tới kết cấu và các chỉ tiêu lý, hoá tính của đất.

Nhìn chung khi chọn đất thí nghiệm nên chọn xa các rừng cây (hàng cây to); xa trục đường giao thông, xa nơi chứa nước thải hay mương dẫn, nước thải của các khu dân cư, nhà máy, bệnh viện... với khoảng cách từ 40 – 50 m.

Tránh làm thí nghiệm trên đất mới được khai hoang. Tùy từng điều kiện cụ thể mà bố trí thí nghiệm tránh vài vụ để san bằng độ đồng đều của đất thí nghiệm.

Thí nghiệm trồng là một trong những kỹ thuật được thực hiện nhằm san bằng độ đồng đều của đất trước khi bố trí thí nghiệm chính. Đất thí nghiệm không đồng đều có thể do một trong những nguyên nhân sau:

- Quá trình sử dụng đất đã áp dụng các biện pháp kỹ thuật như: cày sâu, bón phân (gây ảnh hưởng kéo theo).

- Do chênh lệch về địa hình có thể dẫn tới sự khác nhau về thành phần cơ giới và hóa tính của đất.

Ta có thể san bằng độ đồng đều của đất bằng cách:

+ Giao cấy cùng một loại cây trồng trong vài vụ liên tục và thường dùng các loại cây hàng dày như lúa (đối với đất ngập nước) hoặc cây họ đậu, cây rau (đối với đất màu).

+ Kỹ thuật chăm sóc bón phân theo kiểu bón vá áo vào những chỗ cây trồng xấu hoặc chỗ xấu nhiều bón nhiều, chỗ xấu ít bón ít, chỗ tốt thì không bón.

Các biện pháp kỹ thuật thực hiện trong quá trình làm thí nghiệm trồng phải giống nhau. Khi thu hoạch năng suất chia mảnh thí nghiệm thành nhiều ô có kích thước $10m^2$ (2m X 5m), sau đó tính hệ số biến động qua năng suất của các ô nhỏ. Nếu hệ số biến động $\leq 10\%$ cho phép bố trí thí nghiệm chính trên khu đất đó.

2.4. Xây dựng đề tài nghiên cứu

2.4.1. Cơ sở để xây dựng đề tài

Đề cương nghiên cứu khoa học được xây dựng trên cơ sở đã xác định được một nội dung nghiên cứu của đề tài. Vì vậy, cơ sở để xây dựng đề cương dựa vào:

a) Mục đích và yêu cầu mà nội dung nghiên cứu đặt ra.

b) Xuất phát từ điều kiện tự nhiên (khí hậu, đất đai) và kinh tế - xã hội nơi nghiên cứu (vùng sau này ứng dụng các kết quả nghiên cứu vào sản xuất).

c) Kinh phí và thời gian nghiên cứu

2.4.2. Nội dung của đề cương nghiên cứu

Để có được nội dung của đề cương nghiên cứu khoa học, người nghiên cứu phải xác định cho được tên của đề tài nghiên cứu (có thể chi tiết tới tên của thí nghiệm).

Yêu cầu của tên đề tài là: phát biểu ngắn, gọn, song chính xác.

Trong đề cương nghiên cứu phải thể hiện rõ các vấn đề sau:

(1) Phải phản ánh được đòi hỏi của thực tiễn sản xuất. Trong thực tế rất đa dạng và phong phú, song người chủ trì đề tài phải biết chọn lọc vấn đề cơ bản và thiết thực để nghiên cứu.

(2) Phải biết kế thừa một cách chọn lọc và đặc biệt phải nêu rõ mục đích và yêu cầu đề tài đặt ra.

(3) Phải thể hiện rõ quy hoạch và quy trình thí nghiệm.

(4) Thể hiện được sự phối hợp nghiên cứu giữa các cơ quan, phải hạch toán kinh phí, phải xác định rõ người chủ trì và người thực hiện.

Vì vậy, nội dung của một đề cương bao gồm các phần sau:

Tên đề tài hay tên thí nghiệm phải ghi ngắn gọn, rõ ràng, tóm tắt được mục tiêu thí nghiệm

Ghi cụ thể tên và mã chuyên ngành hoặc mã ngành.

Tên người chủ trì

Tên người thực hiện

Cơ quan quản lý khoa học của đề tài.

Phần Mở đầu

1. Đặt vấn đề

Ghi ngắn gọn nhằm nêu lên được tầm quan trọng của vấn đề nghiên cứu.

2. Mục đích và yêu cầu của đề tài

a. Mục đích

Viết rõ ràng, không viết dài. Đây sẽ là cơ sở để bố trí thí nghiệm được chặt chẽ và đầy đủ.

b. Yêu cầu

Phần này đòi hỏi viết thật cụ thể là đề tài nhằm đạt được những gì.

Phần thứ nhất: Nội dung và phương pháp nghiên cứu

1.1. Tổng quan về vấn đề nghiên cứu

Phần này viết sơ lược những cơ sở lý luận và thực tiễn của đề tài và tình hình nghiên cứu trong nước và thế giới một cách ngắn gọn tóm tắt có liên quan tới vấn đề nghiên cứu.

1.2. Nội dung nghiên cứu

1.2.1. Địa điểm, thời gian và đối tượng nghiên cứu

(nếu có thể ghi thêm giới hạn của vấn đề nghiên cứu)

1.2.2. Nội dung thí nghiệm

Nêu số lượng và nội dung của các công thức thí nghiệm (viết rõ cho từng công thức, đây là yếu tố thí nghiệm).

- Các biện pháp thuộc yếu tố không thí nghiệm (nền thí nghiệm), nên viết tỷ mỉ, những biện pháp chính, còn những biện pháp thứ yếu nên viết tóm tắt.

- Diện tích ô thí nghiệm (ghi cả diện tích và kích thước ô).

- Số lần nhắc lại.

- Cách sắp xếp các công thức (nếu vẽ sơ đồ thí nghiệm thì càng tốt).

1.3. Phương pháp nghiên cứu

Ghi cụ thể các phương pháp sử dụng trong nghiên cứu đề tài.

1.4. Chỉ tiêu nghiên cứu

Ghi rõ tên, thời gian bắt đầu theo dõi, phương pháp theo dõi của từng chỉ tiêu, cách lấy mẫu và dung lượng mẫu lấy.

1.5. Dự trù kinh phí nghiên cứu

Gồm vật tư, công (hạch toán bằng tiền). Lưu ý các vật tư nên có cùng nguồn gốc và thời hạn sử dụng là tốt nhất.

1.6. Phân công thực hiện và kế hoạch hợp tác

(ghi rõ tên người thực hiện các công việc trong thời gian nào, cơ quan nào hợp tác nghiên cứu).

Phần thứ hai: Dự kiến kết quả thu được

Tóm lại: Đề cương phải đầy đủ, nghiêm túc, không được tùy tiện, nhưng phải linh hoạt khi có diễn biến do yếu tố khách quan như khí hậu, hay cụ thể là thời tiết gây ra...

2.5. Một số loại thí nghiệm riêng biệt (thí nghiệm chuyên sâu)

2.5.1. Thí nghiệm với cây ăn quả và cây công nghiệp lâu năm

Đặc điểm của loại thí nghiệm này là cây to, mỗi cây chiếm một diện tích đất lớn và được chăm sóc riêng biệt cho từng cây. Chu kỳ khai thác hiệu quả kinh tế dài (hay nói cách khác là chu kỳ sống kéo dài hàng chục năm và cá biệt có thể tới 100 năm). Vì vậy, rất khó có thể đạt được độ chính xác của thí nghiệm cao như các loại thí nghiệm với cây hàng năm.

Bên cạnh đó trong chu kỳ sống của cây có thể thực hiện nhiều thí nghiệm khác nhau trên cùng một cây. Do đó, dễ có những ảnh hưởng dư thừa lại của thí nghiệm trước đó để lại cho thí nghiệm sau. Nếu như thời gian làm thí nghiệm trắng (không làm thí nghiệm) là chưa đủ thời hạn cho phép.

Phần lớn các cây lâu năm lại được trồng trong điều kiện đất có địa hình dốc (trung du và miền núi), nên độ đồng đều của đất nói chung là thấp hơn đất bằng. Vì vậy, để khắc phục những vấn đề trên nhằm nâng cao độ chính xác của thí nghiệm cần phải quan tâm đến những nội dung sau:

2.5.2. Chuẩn bị đất thí nghiệm

Đặt thí nghiệm ở vườn ươm

Điều này chỉ đặt ra với thí nghiệm làm mới từ đầu. Tùy thuộc vào loại cây, loại thí nghiệm mà cần có các biện pháp chăm sóc, chọn cây cho đồng đều.

Sẽ có 2 trường hợp cụ thể khi đặt thí nghiệm gồm:

** Thiết kế thí nghiệm mới*

Bắt đầu chuyển cây con từ vườn ươm ra vườn sản xuất (hay điều kiện thực tại của cây sẽ sống). Vì vậy, chuẩn bị đất là khâu hết sức có ý nghĩa và quan trọng. Nên chọn đất phải phù hợp với cây định nghiên cứu, đất phải đồng đều về tính chất đất, chế độ nước, địa hình. Để có được đất đồng đều thì trước khi đặt vườn thí nghiệm nên làm thí nghiệm trắng với thời hạn từ 3 năm - 5 năm. Nếu trong thí nghiệm có nghiên cứu về tưới nước thì nên làm đường đồng mức cho mỗi lần nhắc lại, áp dụng các biện pháp chống xói mòn và sắp xếp ô thí nghiệm không được sử dụng kiểu hoàn toàn ngẫu nhiên - CRD (nếu thí nghiệm trên đất dốc). Còn thí nghiệm trên đất bằng nên đào rãnh (muong) vừa để thoát nước và dự trữ nước tưới cho cây khi cần.

** Với thí nghiệm đặt trên vườn đã có sẵn cây thì phải có đủ điều kiện sau:*

- Vườn phải đại diện cho vùng định phổ biến kết quả thí nghiệm.
- Trước khi tiến hành thí nghiệm phải làm thí nghiệm trắng để san bằng độ phì nhiêu của đất, loại trừ các ảnh hưởng dư thừa của các biện pháp kỹ thuật trước đó (thời gian có thể từ 3 - 5 năm).
- Giữa các ô thí nghiệm nên để khoảng cách (khoảng trống hoặc có cây) dùng làm ranh giới giữa các ô và đây là dải bảo vệ của ô (hay gọi là hàng biên).

2.5.1.2. Cây thí nghiệm

a) *Với thí nghiệm trồng mới:* thì hết sức lưu ý tới nguồn gốc của cây. Phải thống nhất cây con đem trồng nghiên cứu có nguồn gốc từ hạt, từ ghép mắt, ghép cành, giâm cành hay chiết. Khi nguồn gốc cây con là từ sinh sản vô tính nên chú ý thêm đến gốc ghép, mắt ghép, cành ghép, cành giâm.

Cây con đem trồng thí nghiệm phải đảm bảo đồng đều về chiều cao cây, đường kính gốc, số cành, số lá/cành.

b) Với thí nghiệm trên vườn đã có sẵn: cần lưu ý chọn các cây có độ đồng đều cao về giống, tuổi, chiều cao, tình trạng phát triển (được thể hiện qua đường kính gốc, chu vi tán cây). Nên có thêm các thông tin về các biện pháp kỹ thuật đã sử dụng trước khi thí nghiệm nhất là các biện pháp có liên quan nhiều tới cây như (đốn đầu, tạo tán, tỉa cành...)

Với thí nghiệm này diện tích ô phải lớn. Cũng có thể xác định diện tích ô qua số lượng cây, tùy theo kích thước của cây mà có thể cho phép mỗi ô có từ 10 - 50 cây. Nhưng số lần nhắc lại phải đạt ít nhất 3 - 4 lần. Nếu trên mỗi ô số cây ít thì số lần nhắc lại phải tăng lên 5 - 8 lần, thậm chí có thể 10 lần nhắc lại cho một công thức.

Việc lấy mẫu quan sát cũng hết sức phức tạp, nên tùy mục đích nghiên cứu mà phải lấy mẫu mang tính đại diện (ở các hướng của tán cây và các vị trí cành của tán cây).

2.5.2. Thí nghiệm về rau

Nguyên lý chung của thí nghiệm này cũng như các cây khác mà chúng tôi đã đề cập trong các nội dung trước. Song vẫn có những đặc điểm khác biệt khi làm thí nghiệm với cây rau các nhà nghiên cứu cần nắm được đặc điểm là:

- Rau rất đa dạng: Về chủng loại, thời gian sinh trưởng, nhu cầu dinh dưỡng, giá trị (bộ phận sử dụng...). Vì vậy thí nghiệm với cây rau cũng rất đa dạng. Hầu như trồng rau không thể thiếu việc tưới nước (nhân tạo) một cách thường xuyên.

- Đa số các loại rau đều phải trải qua thời kỳ vườn ươm, sau đó mới trồng chính thức.

- Vấn đề chăm sóc cũng rất khác nhau và phức tạp. Tùy thuộc vào thí nghiệm rau nhằm cung cấp thực phẩm cho tiêu dùng hay sản xuất hạt giống mà quyết định các biện pháp kỹ thuật cho phù hợp. Bên cạnh đó công tác BVTV phải được quan tâm thường xuyên để phát hiện kịp thời mức độ bị hại của côn trùng hay bệnh mà quyết định biện pháp phòng, trừ đảm bảo an toàn cho người tiêu dùng nhưng vẫn mang lại hiệu quả kinh tế cũng như hiệu quả môi trường.

2.5.3. Thí nghiệm bảo vệ thực vật (BVTV)

Mục đích của thí nghiệm này là đi sâu nghiên cứu các biện pháp đấu tranh chống lại các tác nhân gây sâu, bệnh và cách phòng chống tác nhân gây hại đó cho cây trồng.

Bên cạnh các thí nghiệm ngoài đồng ruộng thông thường thì các thí nghiệm trong phòng cũng thường xuyên sử dụng.

Thí nghiệm trong phòng với loại này là các điều kiện thực hiện có thể kiểm soát không toàn diện hoặc kiểm soát toàn diện như (điều chỉnh nhiệt độ, độ ẩm, cường độ ánh sáng, thành phần ánh sáng...)

Trong ngành bảo vệ thực vật các nghiên cứu trong phòng nhằm xây dựng sự phối hợp của các loại thuốc mới với hiệu quả bảo vệ môi trường. Xác định tính độc hại của thuốc tới côn trùng, nấm bệnh cũng như vi khuẩn gây bệnh, ngưỡng an toàn của chúng với con người và vật nuôi khi sử dụng sản phẩm được áp dụng các loại thuốc nghiên cứu. Tiếp theo là nghiên cứu cơ chế tác động của thuốc, phổ hoạt động của nó, liều lượng, thời kỳ và điều kiện sử dụng. Đồng thời cũng có thể tiến hành nghiên cứu sự xuất hiện hoặc không xuất hiện của sâu hại hoặc thời kỳ phát sinh bệnh, khả năng đấu tranh sinh học chống lại sâu bệnh của các kiểu gen ở cây trồng. Tất cả các kết quả thí nghiệm trong phòng đều sẽ được xác minh kiểm tra ở thí nghiệm ngoài đồng.

Thông thường các thí nghiệm đồng ruộng về BVTV thường sử dụng các loại thuốc (hoá học hay sinh học) ở liều lượng tối thiểu trong các thời kỳ phun tối ưu theo kết quả thu được từ thí nghiệm trong phòng.

Bên cạnh những nguyên tắc thông thường như các thí nghiệm nông học nói chung, thí nghiệm BVTV có khác biệt về diện tích ô hoặc nói cụ thể là hình dạng ô, mức độ di động hay phát tán của côn trùng, nấm bệnh... cũng sẽ quy định diện tích ô khác nhau. Điều khác biệt của thí nghiệm BVTV là diện tích ô càng lớn thì sự khác biệt trong phạm vi ô sẽ càng nhỏ. Cho nên nếu thí nghiệm với côn trùng có mức độ di chuyển hẹp (chậm) diện tích ô nên 25 m² còn côn trùng di động nhanh thì diện tích ô từ 25 – 100 m².

Thí nghiệm lây bệnh nhân tạo ô nhỏ hơn (10 m²/ô). Vấn đề hình dáng ô thí nghiệm ở thí nghiệm BVTV có ảnh hưởng khá rõ tới độ chính xác của thí nghiệm. Cho nên chọn hình dáng ô không thích hợp sẽ làm giảm độ chính xác của thí nghiệm như mục (2.4.1.3.) đã nêu với loại thí nghiệm về BVTV ô thí nghiệm hình vuông là tốt nhất và số lần nhắc lại cũng nên từ 3 lần trở lên. Số cây theo dõi trong mỗi ô thí nghiệm (mỗi lần nhắc lại) phải đạt 30 cây. Trong thí nghiệm lây bệnh nhân tạo trên đồng ruộng khi quan sát sự nhiễm bệnh thì trong mỗi ô của mỗi lần nhắc lại phải theo dõi ở 5 điểm, mỗi điểm (vị trí) có diện tích đạt 0,25 m² - 0,50 m². Để thí nghiệm chính xác giữa hai ô kề nhau nên để hàng biên rộng và tốt nhất có thể đặt một công thức đối chứng.

Thí nghiệm BVTV được chia thành 3 nhóm:

Nhóm 1: thí nghiệm về thuốc BVTV. Trong nhóm này có hai dạng cụ thể, bao gồm:

+ Thuốc phòng trừ sâu bệnh hại cây trồng.

+ Thuốc trừ cỏ dại

Nhóm 2: thí nghiệm nghiên cứu các giống có khả năng kháng sâu và kháng bệnh.

Nhóm 3: thí nghiệm nghiên cứu về mặt sinh học của các tác nhân gây bệnh và tạo các điều kiện để xuất hiện côn trùng gây hại. Tuy nhiên, kết quả của nhóm này phụ thuộc vào điều kiện môi trường.

Xin lưu ý với thí nghiệm về thuốc trừ cỏ dại cần có diện tích cho các điểm theo dõi cỏ xuất hiện sau khi sử dụng thuốc diện tích cho mỗi điểm là từ 0,5 - 1 m²/điểm. Và việc theo dõi này phải liên tục cho tới khi thu hoạch cây trồng và tiếp sau các cây trồng khác ít nhất là từ 1 - 2 năm. Bên cạnh việc theo dõi cỏ dại cần thiết nên có các kết quả liên quan đến các sinh vật đất trong các công thức nghiên cứu. Nhìn chung thí nghiệm về thuốc BVTV thường gây hiệu quả kéo theo tạo nên ảnh hưởng gián tiếp đến trước hết là môi trường đất, các hệ sinh thái và môi trường sống nói chung.

2.5.4. Thí nghiệm nuôi cấy mô (Invitro)

Hầu hết các thí nghiệm thuộc nhóm nghiên cứu trong phòng. Còn một số ít sẽ là thí nghiệm trong nhà lưới (nhà polytylen) và thí nghiệm đồng ruộng là bước sau cùng.

Loại thí nghiệm này có những đặc trưng riêng khác với thí nghiệm khác.

Khái niệm về nuôi cấy mô (Explantace) được hiểu là: sử dụng một bộ phận sống bất kỳ nào đó của cơ thể như: một tế bào, một mô, một nhóm các mô của một cơ thể hoặc một nhóm cơ thể được tách cố ý từ cây và được nuôi cấy trong ống nghiệm thường gọi là invitro. Đây là thuật ngữ duy nhất trong thí nghiệm.

Thường các mô nuôi cấy được lấy từ rễ, thân, phần, chồi, lá, phôi tế bào và các bộ phận khác của cây.

Điều kiện của thí nghiệm là hoàn toàn nhân tạo gồm: các nhân tố hoá học, vật lý và nhân tố sinh hoá học. Các nhân tố đó cần cố gắng tạo ra được điều kiện vô trùng để kết quả sau sẽ tạo ra một sự sống mới (Cá thể mới sau sẽ sinh trưởng và phát triển).

Vì vậy, điều kiện thí nghiệm quan trọng nhất là: kỹ thuật và các điều kiện nuôi cấy phải sạch và vô trùng. Thí nghiệm thuộc nhóm nuôi cấy mô có các ý nghĩa sau:

- Thí nghiệm nghiên cứu có điều kiện hầu như độc lập với môi trường bên ngoài.
- Thí nghiệm có khả năng thực hiện liên tục.
- Dễ dàng tạo ra số lượng cá thể lớn (hệ số nhân cao).
- Tạo ra các giống sạch bệnh, các nguồn vật liệu sạch cho công tác chọn tạo giống cây trồng.

2.5.5. Thí nghiệm trong điều kiện sản xuất

Loại này thường được gọi là: "Làm thí nghiệm khoa học ở cơ sở sản xuất".

Thực tế sản xuất nông nghiệp liên quan mật thiết với điều kiện tự nhiên và điều kiện kinh tế xã hội. Do đó, có sự chênh lệch về năng suất cây trồng giữa các vùng sinh thái khác nhau, thậm chí có thể ở trong cùng điều kiện sinh thái nhưng điều kiện - xã hội khác nhau cũng làm cho năng suất cây trồng khác nhau.

Một thực trạng khác nữa là: năng suất cây trồng trong các thí nghiệm của các nhà khoa học thường cao hơn năng suất cây trồng do nông dân thực hiện. Bởi vì, làm thí nghiệm nên quy mô, diện tích thường nhỏ (chỉ tính trên mỗi ô thí nghiệm), các yếu tố nghiên cứu có khối lượng ít nên có độ đồng đều cao. Song các nhà khoa học đều nhận thức được là: tác dụng của các biện pháp kỹ thuật mới đều phải được đánh giá bằng mức độ thích ứng trên đồng ruộng của nông dân chứ không phải chỉ ở trên ô thí nghiệm đạt được là đầy đủ. Vì vậy tiến hành thí nghiệm trong điều kiện sản xuất là bộ phận nghiên cứu cần thiết, trước khi chuyển giao kết quả nghiên cứu (ứng dụng trong sản xuất) cho nông dân với mục tiêu phát triển nông nghiệp sinh thái bền vững. Có hai loại thí nghiệm thuộc nhóm này là:

Loại 1: Nghiên cứu phát triển kỹ thuật (nhằm nghiên cứu để xây dựng và phát triển kỹ thuật mới).

Loại 2: Nghiên cứu tiếp cận kỹ thuật (nhằm thử nghiệm các kỹ thuật mới) xem mức độ tiếp thu của nông dân như thế nào?

Cả hai loại này đều tồn tại song song.

Loại nghiên cứu phát triển kỹ thuật là nhà khoa học làm trong điều kiện sản xuất thực tế tại trang trại sản xuất (nó rộng hơn trại thí nghiệm) của cơ quan nghiên cứu.

Còn loại nghiên cứu tiếp cận kỹ thuật là mô hình thử nghiệm được tự người nông dân làm trên đồng ruộng của chính mình, có sự cố vấn của các nhà khoa học. Vì vậy, đôi khi phải có sự điều chỉnh cho hợp lý với điều kiện của nông dân.

Hiện nay đa phần các nghiên cứu được tập trung cho loại thí nghiệm tiếp cận kỹ thuật. Các luận cứ của loại này bao gồm:

- a/ Tăng diện tích thử nghiệm (tăng diện tích ô và tăng vị trí thử nghiệm).
- b/ Chọn điểm thử nghiệm phải đại diện
- c/ Chọn yếu tố đưa vào thử nghiệm
- d/ Mở rộng sự tham gia của nông dân (có sự giám sát của các nhà khoa học).
- e/ Cần phải có một số các thông tin tối thiểu của các nhân tố về môi trường trong khu vực chọn địa điểm thử nghiệm.
- f/ Tiêu chuẩn thử nghiệm

Tiêu chuẩn này phải xác định được 2 câu hỏi là:

- (1) Lấy gì làm cơ sở để so sánh?
- (2) Những chỉ số dùng để so sánh cụ thể là gì?

Với câu hỏi 1 có 2 cách trả lời là: hoặc so sánh với hệ thống canh tác hiện tại của nông dân, hoặc là so sánh với một hệ thống canh tác được ấn định trước và coi nó là tiêu chuẩn. Nhưng thường chọn cách trả lời 1 là đơn giản và dễ được người nông dân chấp nhận.

Câu hỏi 2 là vấn đề phức tạp và khó: thực tế cần phải căn cứ vào các chỉ tiêu như sức sản xuất, lợi nhuận, tính ổn định, giá trị dinh dưỡng, lưu thông phân phối và sử dụng tài nguyên (đất, sức lao động...).

h/ Thu thập các tài liệu và thông tin khác có liên quan

Trong quá trình thí nghiệm các số liệu được thu thập gồm các loại như:

- Các số liệu về điều kiện tự nhiên bao gồm: đất, nước, địa hình, khí hậu... (hay còn gọi là các số liệu về môi trường của đồng ruộng).
- Các số liệu về điều kiện kinh tế - xã hội của nông dân bao gồm: lực lượng sản xuất, động lực sản xuất.
- Các số liệu về nông học (thu ở các ô thử nghiệm như năng suất và một số đặc trưng khác).
- Yêu cầu tài nguyên (các nguồn cần) cho mô hình thử nghiệm, nhu cầu về nhân lực (lao động) và đầu tư.

Hai loại số liệu đầu không phụ thuộc vào kỹ thuật thử nghiệm áp dụng mà nó được lấy ngay tại cơ sở thử nghiệm.

Còn số liệu về nông học và nhu cầu tài nguyên phụ thuộc nhiều vào kỹ thuật thử nghiệm. Cách thu thập số liệu nông học khác hẳn cách thu thập số liệu về nhu cầu tài nguyên. Sự khác nhau thể hiện như sau:

+ Số liệu nhu cầu tài nguyên không thể đánh giá ở những ô thử nghiệm nhỏ dùng trong thử nghiệm nông học. Theo Viện nghiên cứu lúa quốc tế tại Philippine (International Rice Research Institute - IRRI) để thấy được nhu cầu về nhân lực với cây lúa có thể được dự đoán trên diện tích gieo trồng từ 800 - 1000 m². Trong đó ô thử nghiệm nông học chỉ có diện tích 20 - 60 m².

+ Số liệu nhu cầu tài nguyên không thay đổi nhiều theo môi trường tự nhiên như đất và số liệu về nông học. Do vậy với thí nghiệm về số liệu nông học cần nhắc lại. Còn số liệu nhu cầu tài nguyên không cần nhắc lại. Vậy kỹ thuật làm thử nghiệm được lưu ý ở những nội dung sau đây:

2.5.5.1. Chọn địa điểm thử nghiệm

Căn cứ vào tình hình biến động của điều kiện tự nhiên trong địa bàn nghiên cứu thử nghiệm trong giai đoạn mô tả, có thể phân chia địa bàn thành 2 - 8 nhóm phụ. Tiêu chuẩn chia nhóm dựa vào dự kiến về hiệu quả nông học (như làm mây vụ trong năm, trồng cây gì?) quyết định về số nhóm chia là điều quan trọng vì chia quá ít sẽ không có đủ các thông tin cần thiết. Còn chia quá nhiều sẽ phức tạp tốn công và có thể cũng tạo nên biến động lớn của các thông tin.

Ở mỗi nhóm phụ chọn 2 đến 5 cánh đồng bằng phương pháp ngẫu nhiên

2.5.5.2. Thiết kế và bố trí lô thử nghiệm

* Loại 1: Dạng cải tiến quản lý mô hình canh tác hiện đang sử dụng thì phải tiến hành như sau:

- Xác định xem hợp phần (tổ hợp) kỹ thuật cải tiến khác với kỹ thuật của nông dân ở những điểm nào?

- Nên ghép các tổ hợp có liên quan mật thiết với nhau lại là một tổ hợp. Thí dụ như biện pháp dùng thuốc trừ cỏ dại và xới xáo nên ghép lại làm một.

- Số công thức phải nhiều hơn số hợp phần là 2 đơn vị.

Thí dụ: có 4 hợp phần nghiên cứu (giống, sâu bệnh, phân bón và chăm sóc cộng với làm cỏ) thì sẽ có 6 công thức xử lý thử nghiệm, cụ thể là:

Công thức I (xử lý 1): kỹ thuật đề nghị (nghĩa là: giống đề nghị, phương pháp trừ sâu bệnh đề nghị, phương pháp bón phân đề nghị và cuối cùng là biện pháp chăm sóc và làm cỏ cũng đề nghị).

Công thức II (xử lý 2): gồm kỹ thuật của nông dân, nghĩa là: giống của nông dân, phương pháp trừ sâu bệnh của nông dân, phân bón cũng của nông dân và biện pháp xới xáo cũng làm cỏ cũng do dân.

Công thức III (xử lý 3) gồm: giống, phương pháp trừ sâu bệnh và cách chăm sóc là do nông dân, chỉ có chế độ bón phân đề nghị.

Công thức IV (xử lý 4) gồm: phương pháp trừ sâu bệnh, phương pháp chăm sóc của nông dân, còn giống và chế độ bón phân là đề nghị.

Công thức V (xử lý 5) gồm: giống, phương pháp trừ sâu bệnh và phân bón đề nghị. Còn duy chỉ phương pháp chăm sóc làm cỏ là của nông dân.

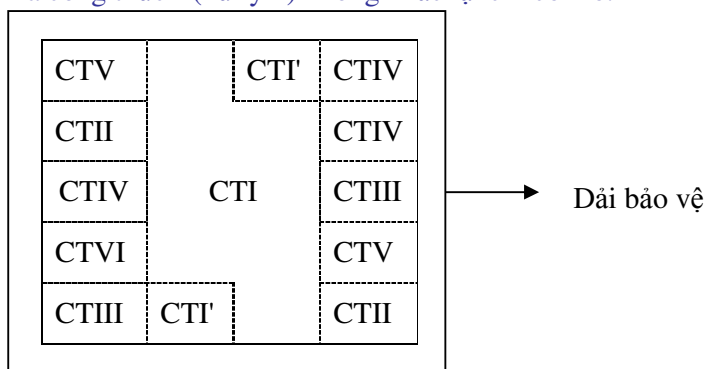
Công thức VI (xử lý 6) gồm: giống, phương pháp trừ sâu bệnh và cách chăm sóc đề nghị. Chỉ còn chế độ phân bón là của nông dân.

- Diện tích lô và cách sắp xếp lô thử nghiệm gồm:

Có hai diện tích lô thử nghiệm gồm:

Ô nhỏ dùng thử nghiệm để có số liệu nông học (có nhắc lại) sẽ có số công thức như dự kiến. Thí dụ nêu trên sẽ có 6 công thức.

Ô lớn dùng lấy số liệu về nhu cầu tài nguyên (dùng cho các hợp phần thử nghiệm đề nghị ở thí dụ trên là công thức I (xử lý 1) không nhắc lại chỉ có 1 ô. Minh hoạ như sơ đồ 2.3:



Sơ đồ 2.3. Thiết kế sắp xếp các lô thí nghiệm

Như vậy, tổng diện tích khoảng 1000 m², mô hình cho 4 tổ hợp thay đổi kỹ thuật. Trong đó có 2 ô phụ là (công thức I) nằm trong ô lớn kỹ thuật đề nghị.

* Loại 2 có 2 dạng là:

+ Đối với loại (2a): cải tiến mô hình canh tác bằng đưa thêm một hoặc nhiều cây mới vào mô hình hiện đang sử dụng. Đối với mỗi mô hình, cần xác định xem trồng thêm cây bổ

sung có đòi hỏi thay đổi công tác quản lý những cây trồng hiện có không? nếu có nên đưa cây trồng mới vào công thức xử lý thử nghiệm. Cách thiết lập lô thử nghiệm làm như loại 1.

Nếu là thêm cây thì đề nghị ra một hệ canh tác để thử nghiệm xem kỹ xem là mô hình thử nghiệm có đòi hỏi thay đổi công tác quản lý không?

Nếu có thì những thay đổi phải được đưa vào thử nghiệm. Cách thiết lập làm như loại 1.

Thí dụ: Trước đây trên chân đất vùn cao chỉ sử dụng cây 2 vụ lúa. Nay muốn đưa thêm vào vụ đông 1 cây mới là đậu tương như vậy có thể phải thay đổi giống lúa có thời gian sinh trưởng ngắn hơn và cũng có thể chuyển đổi từ cây sang gieo vãi hoặc thay đổi vụ trồng và phân bón cho phù hợp. Công thức luân canh sẽ là lúa - lúa - đậu tương đông). Như vậy cách thiết kế sẽ như (sơ đồ 2.3) giống như loại 1.

+ Với loại (2b) là thay đổi cấu trúc mô hình khác hoàn toàn so với mô hình của nông dân hiện đang sử dụng. Với loại này, cần so sánh tổng quát toàn bộ mô hình, không so sánh riêng từng hợp phần kỹ thuật với nhau. Kiểu này ít làm hơn kiểu 2a.

2.5.5.3. Quản lý lô thử nghiệm

Các biện pháp đều do nông dân quản lý, trừ những hợp phần đưa vào thử nghiệm do cán bộ nghiên cứu quản lý. Còn nếu thấy có thể được thì đề nghị giao cho nông dân quản lý ô to.

2.5.5.4. Thu thập số liệu

a) Điều kiện tự nhiên

Chọn một số các chỉ tiêu (thông số) tối thiểu để có thể đo được điều kiện tự nhiên (yếu tố môi trường) ở cánh đồng thử nghiệm. Nhân tố này được chọn tùy thuộc vào mức độ ảnh hưởng của nó tới hiệu quả nông học của mô hình thử nghiệm. Các nhân tố đó là: lượng mưa (chế độ mưa), bức xạ, kết cấu đất, vị trí địa lý, địa hình, mực nước ngầm.

b) Điều kiện kinh tế - xã hội: chọn một số chỉ tiêu tối thiểu về điều kiện kinh tế - xã hội của nông dân mà những thông số này có ảnh hưởng tới khả năng thích ứng của nông dân với mô hình chọn. Các chỉ tiêu thường là khả năng về nhân lực, vốn, điều kiện về đất đai và khả năng về quản lý.

c) Số liệu nông học

Chọn ấn định các chỉ tiêu thông thường là năng suất kinh tế, tình hình sâu, bệnh, cỏ dại ... các số liệu ở hai ô nhỏ đặt trong ô to.

d) Số liệu về nhu cầu tài nguyên gồm các chỉ tiêu sau: nhu cầu nhân lực, đất đai và nhu cầu về quản lý (giống, phân bón...).

e) Số liệu về kinh tế: Thu ở mô hình của nông dân (định giá tại thời điểm khi mô hình nông dân đang sử dụng).

Còn các công thức xử lý dùng nghiên cứu thì ghi trực tiếp ngay trên đồng ruộng.

f) Chọn các phương pháp đo lường các loại số liệu đã nêu trên cho phù hợp và đảm bảo độ chính xác và tin cậy (với các số liệu về nông học có thể áp dụng các biện pháp thông thường như ở các trại thí nghiệm, số liệu về nhu cầu tài nguyên cần có quy định cụ thể để thống nhất sự đo lường).

2.5.5.5. Phân tích số liệu

a) So sánh hiệu quả nông học của mô hình thử nghiệm với mô hình của nông dân.

b) So sánh việc sử dụng tài nguyên của mô hình thử nghiệm với mô hình nông dân.

c) Kết hợp số liệu nông học với số liệu về nhu cầu tài nguyên để tính ra một số thông số khác, tạo điều kiện so sánh mô hình được cụ thể và rõ hơn (như lợi nhuận và giá trị dinh dưỡng).

- d) Sau khi có đánh giá mô hình thử nghiệm nếu có hiệu quả sẽ triển khai rộng mô hình.
- e) So sánh nhu cầu tài nguyên của mô hình với việc đáp ứng nhu cầu đó của nông dân, từ đó tìm ra mặt hạn chế nếu có.

Bài tập: Giáo viên sẽ ra bài tập cho sinh viên sau khi đã thực tập và quan sát các thí nghiệm trên đồng ruộng. Có thể nêu một số tên thí nghiệm cho sinh viên tự thiết kế và xây dựng đề cương cho thí nghiệm (có thể làm theo nhóm 5 sinh viên / nhóm)

CHƯƠNG III - TIẾN HÀNH THÍ NGHIỆM TRÊN ĐỒNG RUỘNG

Yêu cầu của chương là: Người học phải bắt tay triển khai cụ thể một thí nghiệm sau khi có đề cương nghiên cứu. Cách lấy mẫu để theo dõi các chỉ tiêu đặt ra.

3.1. Chia ô thí nghiệm

Đây là việc chuyển thiết kế thí nghiệm trên sơ đồ đã xây dựng ra ngoài thực địa nơi làm thí nghiệm. Vì vậy, phải thực hiện chính xác để tạo ra các ô thí nghiệm có diện tích và kích thước giống nhau, đảm bảo đúng các vị trí đã sắp xếp như thiết kế. Tuy nhiên, mức chênh lệch cho phép về chiều dài mảnh có thể dao động từ 5 – 10 cm/100 m dài, đặc biệt 4 góc ruộng, góc ô phải vuông.

Tóm tắt việc chia ô khái quát như sau:

Dụng cụ cần có:

- La bàn (giúp định hướng của khu thí nghiệm)
- Thước vuông góc (êke vuông) giúp định các góc mảnh thí nghiệm và ô thí nghiệm.
- Các sào tiêu để xác định vị trí các đường thẳng để cắm mốc và sau đó đắp bờ ngăn hoặc vị trí ranh giới giữa các lần nhắc lại và các ô với nhau.
- Thước (có thể thước dây hay thước thép) để đo cắm định vị trí.

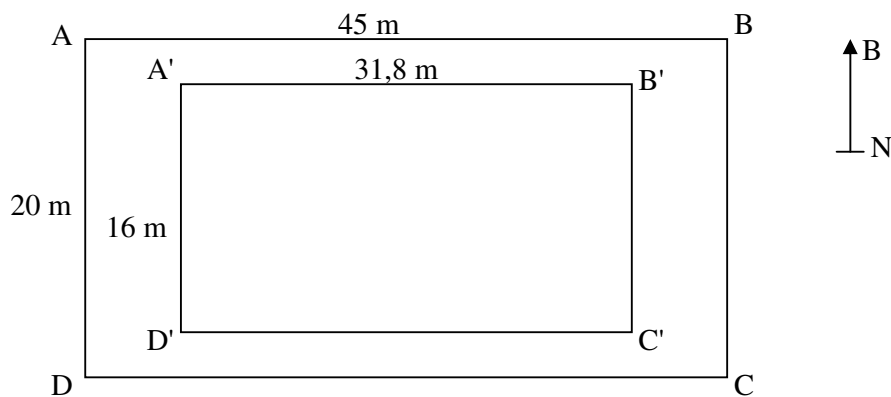
Nguyên tắc chia ô thí nghiệm thường xác định từ ô to về ô nhỏ (có nghĩa là từ mảnh thí nghiệm chia thành từng lần nhắc lại sau đó đến chia các ô trong lần nhắc lại). Có thể lần lượt xem xét qua thí dụ sau đây: So sánh 5 mức dinh dưỡng tới các chỉ tiêu sinh trưởng, phát triển, sâu bệnh và năng suất lúa xuân tại Gia Lâm - Hà Nội.

Như vậy, chọn địa điểm là khu thí nghiệm lúa của khoa Nông học, trường Đại học Nông nghiệp I (5 mức dinh dưỡng là 5 công thức có ký hiệu I, II, III, IV, V). Thí nghiệm được nhắc lại 3 lần với diện tích mỗi ô là 30 m^2 có kích thước (5 m x 6 m). Các bước chia ô cụ thể làm như sau:

3.1.1. Xác định ranh giới khu thí nghiệm

Sơ đồ 3.1. Định khu thí nghiệm

Căn cứ vào thiết kế đã nêu và địa điểm chọn cụ thể của đặt ruộng thí nghiệm (giả sử như sơ đồ 3.1) sau:

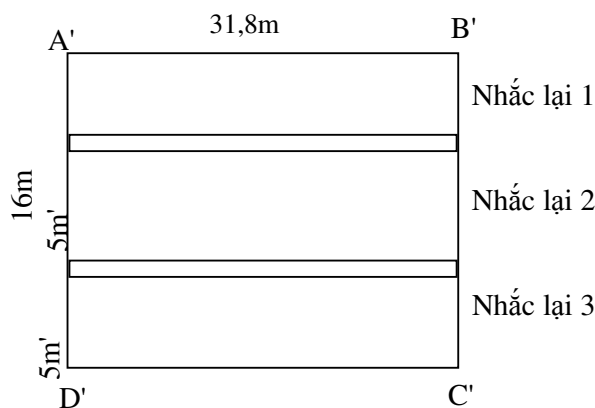


Sơ đồ 3.1. Định vị khu thí nghiệm

Chiều dài mảnh thí nghiệm tính là 31,8 m (kể cả khoảng cách giữa các ô trong lần nhắc lại cách nhau 0,4m). Dải bảo vệ giữa hai đầu ruộng sẽ có độ rộng 6,0 - 7,0 m. Tùy điều kiện đất cụ thể, chiều ngang mảnh thí nghiệm sẽ là 16m (kể cả khoảng cách giữa hai lần nhắc lại là 0,5 m). Trong thí nghiệm này ô thí nghiệm sẽ có dạng gần vuông nhưng mảnh thí nghiệm có hình chữ nhật.

3.1.2. Xác định ranh giới lần nhắc lại

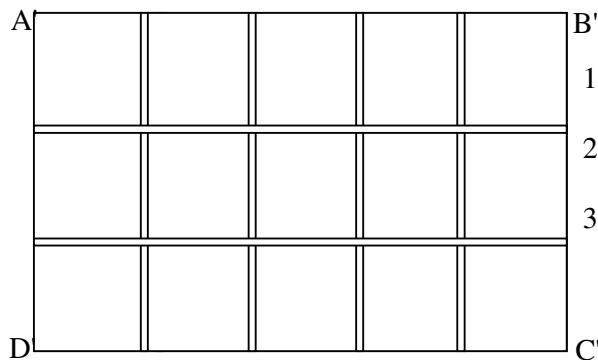
Khi đã có vị trí chính xác của mảnh thí nghiệm sẽ xác định được ranh giới giữa các lần nhắc lại (3 lần) như đã nêu ở trên giữa các lần nhắc lại chúng tôi để khoảng cách đắp bờ 0,5 m rộng hơn so với khoảng cách giữa các ô trong cùng lần nhắc lại (song khoảng cách giữa các lần nhắc phải có độ rộng tối thiểu bằng khoảng cách các ô). Cách chia như sơ đồ 3.2.



Sơ đồ 3.2. Xác định lần nhắc lại của thí nghiệm

3.1.3. Xác định ranh giới các ô thí nghiệm

Như vậy, phải chia mỗi lần nhắc lại làm 5 phần bằng nhau (mỗi phần 6 m và giữa các phần có khoảng cách 0,4 m) cho 5 công thức nghiên cứu (sơ đồ 3.3).



Sơ đồ 3.3. Chia hoàn toàn mảnh thí nghiệm gồm 5 công thức nhắc lại 3 lần. Tổng số ô thí nghiệm là 15 ô = 5 x 3

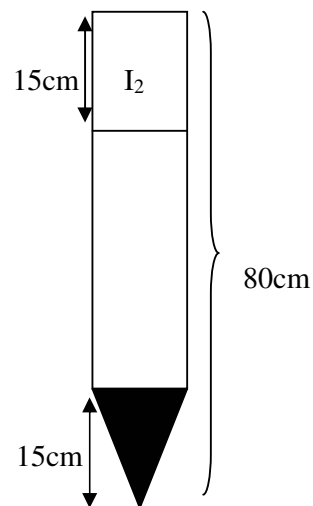
3.1.4. Cắm biển tên các công thức và tên thí nghiệm

Sau khi đã hoàn thành việc chia ô thí nghiệm xong ngoài đồng ruộng, người chủ trì thí nghiệm phải đánh dấu (cắm thẻ tên các công thức ở các ô theo như cách sắp xếp đã xác định

(mục 2.4.1.) cho phù hợp với điều kiện cụ thể nơi đặt thí nghiệm, nội dung nghiên cứu để tránh sự nhầm lẫn trong các công việc cụ thể khi làm thí nghiệm.

Thẻ đánh dấu (ghi) tên công thức có thể được làm bằng gỗ, tre hoặc kim loại không gỉ. Tuy nhiên, với điều kiện của Việt Nam thì phần lớn được làm bằng gỗ hoặc tre.

Thẻ có chiều cao từ 70 - 80 cm, chiều rộng của thẻ từ 8 - 10 cm. Trong đó phần nhọn được cắm xuống đất (có thể từ 15 - 20 cm). Đầu trên của thẻ khoảng 15cm dùng để viết công thức, tên ký hiệu công thức ghi theo số la mã (I, II, III...). Lần nhắc lại ghi số thường (chỉ số dưới như hình 3.4) (đây là thẻ ghi công thức I lần nhắc lại 2).

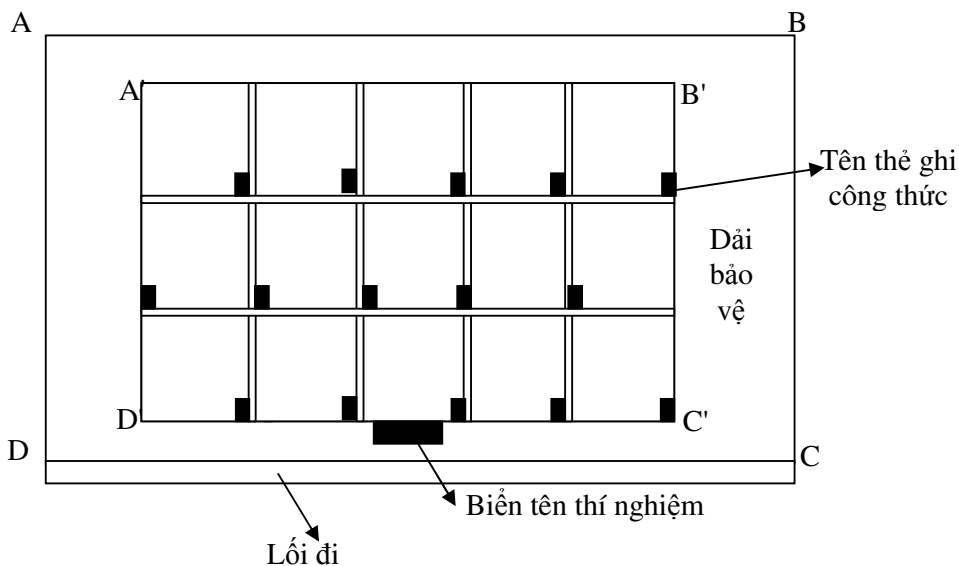


Thẻ sơn toàn bộ màu trắng còn chữ viết có thể màu đỏ hoặc đen là tốt nhất. Các thẻ nên quy định thống nhất cắm ở một vị trí giữa hàng biên và hàng nghiên cứu trong ô về một phía.

Các mặt có ghi chữ số ký hiệu quay đầu về một hướng (sơ đồ 3.5).

Còn biểu ghi tên thí nghiệm nên có dạng hình chữ nhật kích thước nên (50 cm x 30 cm) là vừa. Trên đó có viết tên thí nghiệm, nền bảng sơn trắng còn chữ tên thí nghiệm nên viết màu sơn đỏ. Biểu tên thí nghiệm được cắm ở phía trước thí nghiệm (giữa dải bảo vệ và hàng biên của ô thí nghiệm)

Sơ đồ 3.4. Thẻ ghi công thức



Sơ đồ 3.5.

3.2. Làm đất, bón phân và gieo cây thí nghiệm

3.2.1. Làm đất

Công việc này có thể được chia làm 2 dạng.

- Làm đất giống nhau trên cả ruộng thí nghiệm. Thường hay được sử dụng cho nghiên cứu các biện pháp kỹ thuật canh tác như (so sánh giống, bón phân, mật độ và một số kỹ thuật gieo cây khác...).

- Loại làm đất riêng biệt cho từng ô (hay từng công thức). Thường được áp dụng cho thí nghiệm về thời vụ, biện pháp làm đất, theo dõi tác động của phân bón liên tục đến sự thay đổi của đất.

Khi đã cơ bản hoàn thành việc làm đất tùy chọn một trong 2 dạng như trên, trước khi gieo cây phải cào lại hoặc san lại bằng cào tay để đảm bảo mặt bằng đồng đều của ô thí nghiệm, đạt càng phẳng càng tốt.

3.2.2. Bón phân cho thí nghiệm

Yêu cầu trước hết của công tác này là đảm bảo số lượng và chất lượng (chủng loại) như đúng quy hoạch thí nghiệm đã quy định. Để đạt được yêu cầu trên cần phải tính toán chính xác lượng phân bón cho mỗi ô thí nghiệm, nhất là các dạng phân có lượng ít như phân công nghiệp. Nếu không sẽ tạo sai số rất lớn. Các loại phân công nghiệp đều được quy ra lượng nguyên chất thí dụ đạm, lân, kali (N; P₂O₅; K₂O)/ha.

$$X = \frac{a \cdot c}{100 \cdot b} \quad (3.1)$$

Trong đó:

X: lượng phân bón cho 1 ô (kg)

a: lượng phân nguyên chất cho 1ha (kg/ha)

b: tỷ lệ chất dinh dưỡng (nguyên chất có trong loại phân nghiên cứu (%)).

c: diện tích ô thí nghiệm (m²).

Nếu như làm trong chậu vại hoặc ô xi măng (kể cả diện tích ô trên ruộng quá bé), lượng phân bón được tính theo công thức:

$$X = \frac{100a \times c}{b} \quad (3.2)$$

(trong đó X lượng phân bón cho ô đó tính bằng gam).

Tùy lượng phân bón mà dùng loại cân có độ chính xác cho phù hợp. Sau khi cân phải cho vào túi riêng có đánh dấu tên của công thức để tránh nhầm lẫn (khâu này dễ gây nhầm lẫn nhất). Khi đem ra ruộng nên để phân vào vị trí của ô sau đó kiểm tra cẩn thận lại toàn bộ các ô theo công thức rồi mới bón. Còn phân hữu cơ tùy loại mà có thể quy ra chất lượng dinh dưỡng của phân. Cách tính cũng như trên song lượng nhiều hơn nên đơn giản hơn.

Nhìn chung tất cả các dạng phân đều phải tơi, trộn đều trước khi bón.

Cách bón, yêu cầu phải thật đều trên ô thí nghiệm. Nếu không thực hiện tốt vấn đề này dễ tạo sai số ngay trong công thức và sẽ rất lớn. Nếu phân có lượng quá ít sợ khó bón đều nên trộn với đất bột mịn (không phải cát) để dễ bón đều.

Với thí nghiệm cây gieo hạt nên lưu ý không để phân tiếp xúc trực tiếp với hạt giống.

3.2.3. Gieo, cấy thí nghiệm

Đây là khâu cuối cùng của bố trí xong thí nghiệm. Tùy theo yêu cầu của thí nghiệm mà có thể đưa ra:

- Gieo cấy giống nhau và đồng thời. Kiểu này phải đảm bảo đồng thời (cùng thời gian) và đồng đều.

- Gieo cấy khác nhau và không đồng thời:

Đây là kiểu gieo cấy của thí nghiệm về thời vụ, mật độ khi gieo cấy cần chú ý: với kiểu cấy đồng thời và đồng đều thì nên làm càng gọn càng tốt. Làm lần lượt từng lần nhắc lại (tất

cả các công thức) sau sang làm tiếp theo hoặc nếu phải có một số người làm công việc gieo cấy thì mỗi người nên làm trọn vẹn một số lần nhắc lại nhất định. Tránh làm 1 hay 2 công thức nào đó ở các lần nhắc lại chỉ 1 người.

Phải đảm bảo gieo cấy đúng quy định thí nghiệm, khoảng cách hàng và cây đồng đều, phải thẳng hàng ngang và hàng dọc (nhất là với cây trồng hàng rộng và lâu năm).

Đảm bảo mật độ sau gieo cấy cũng như lấp hạt, đồng đều về số lượng dành hay hạt gieo cấy. Nếu thí nghiệm về giống không được để lẫn giống. Vì vậy, trước khi gieo cấy cần xem lại tên giống cho đúng vị trí ô.

Trong gieo hạt có thể tính khối lượng hạt gieo theo công thức sau

$$P = \frac{100M \times A}{B} \quad (3.3)$$

Trong đó:

P: là khối lượng hạt gieo (kg/ha)

M: số cây cần có trên 1 ha

A: khối lượng 1000 hạt (gram)

B: sức sống của hạt ngoài đồng (tỷ lệ nảy mầm %).

Với cây hàng rộng (nhất là cây lâu năm) nên lưu ý có thể trồng dặm bổ sung khi có cây chết ngay từ những ngày đầu sau gieo cấy trong thời gian cho phép.

Với một số cây hàng năm nên tính toán thận trọng để có thể dặm hoặc tỉa định cây có kết hợp với chăm sóc càng sớm càng tốt.

3.2.4. Chăm sóc thí nghiệm

Đây là công việc thường xuyên gồm: tỉa cành, làm cỏ, xới xáo, vun gốc, bấm ngọn, bón thúc, phòng chống sâu, bệnh ... Quá trình này yêu cầu làm đúng quy định như thiết kế tránh làm sai lệch để kết quả thí nghiệm được chính xác.

Phải lưu tâm và làm đầy đủ các công việc này cho cả dải bảo vệ.

3.3. Theo dõi thí nghiệm

Đây là phần công việc hết sức phức tạp và cần thiết, nếu không thực hiện đúng sẽ dẫn đến kết quả không đại diện và không chính xác.

3.3.1. Quan sát thí nghiệm

Mục đích chính của thí nghiệm đồng ruộng là thu hoạch năng suất (năng suất sinh học, năng suất lý thuyết, năng suất kinh tế...), song cũng phải có các thông tin để giải thích cho sự khác nhau giữa các công thức nghiên cứu của thí nghiệm. Như vậy cần phải có các thông tin quan sát được về cây, về điều kiện thời tiết nhằm trả lời các câu hỏi sau:

1/ Cây trồng có phản ứng với điều kiện ngoại cảnh và các biện pháp kỹ thuật nghiên cứu như thế nào?

2/ Sự khác nhau giữa các công thức thí nghiệm với nhau, với công thức đối chứng là do nguyên nhân nào? (ở các chỉ tiêu nghiên cứu).

Các thông tin quan sát bên cạnh việc trả lời cho 2 câu hỏi lớn trên còn có ý nghĩa khác nữa là:

- Sẽ làm cho thí nghiệm mang tính đại diện cao, giúp các nhà khoa học dự đoán hướng có thể áp dụng kết quả nghiên cứu vào điều kiện cụ thể về khí hậu, thời tiết, về đất đai, về điều kiện kinh tế - xã hội phù hợp.

- Giúp các nhà khoa học đánh giá độ tin cậy của kết quả thí nghiệm ở mức độ cho phép đã được quy định.

Nội dung của các vấn đề quan sát bao gồm

- Quan sát cây trồng gồm: quan sát các thời kỳ sinh trưởng, tình hình sinh trưởng, phân tích diễn biến các quá trình sinh lý xảy ra trong các quá trình sinh trưởng (chu kỳ sống) của cây, những thay đổi về kích thước, cấu tạo và giải phẫu của cây (nếu thấy cần thiết).

- Quan sát ngoại cảnh gồm: quan trắc các diễn biến về thời tiết, các nhân tố khí hậu nông nghiệp và quan sát diễn biến tình hình sâu, bệnh, cỏ dại (chủ yếu là sâu bệnh). Phân tích đất thí nghiệm nếu có điều kiện là rất tốt.

Quan sát các thời kỳ sinh trưởng của cây được bắt đầu ngay sau khi gieo cây ít ngày như: theo dõi cây mọc, cây sống, cây chết, lá, thân, cành, ra hoa, đậu quả... Các số liệu này được đối chiếu với số liệu về khí tượng, từ đó giúp cho các nhà khoa học có nhận định toàn diện mối quan hệ giữa cây với môi trường sống của chúng trong thí nghiệm.

Từ kết quả quan sát đưa ra những nhận xét cho các công thức một cách logic và khoa học.

Tình trạng sinh trưởng của cây còn thể hiện ở sự hình thành các cơ quan phát dục, tính lớp vỏ, các yếu tố cấu thành năng suất và năng suất.

Phân tích quá trình sinh lý, trao đổi chất trong cây gồm: khả năng hút nước, thoát nước của cây (rễ và lá...), hiệu suất quang hợp thuần, sự tích lũy chất dinh dưỡng như N, P, K...

Quan sát các số liệu về khí tượng nông nghiệp (lấy từ Đài khí tượng gần nhất). Tùy điều kiện có thể lấy theo tần suất trung bình của một số năm hay chỉ lấy trong thời gian làm thí nghiệm.

Song các số liệu thường được lấy theo tháng (từ tháng 1 đến tháng 12) của năm, các nhân tố thường là: nhiệt độ bình quân tháng, tổng số giờ nắng trong tháng, tổng lượng mưa của tháng sau đó tính tổng lượng mưa cả năm (mm), ẩm độ không khí bình quân các tháng.

Có thể lấy theo bình quân 10 ngày như vậy thông tin khá chi tiết. Điều này có liên quan tới các thí nghiệm về thời vụ, thí nghiệm về BVTV.

Quan sát tình hình sâu, bệnh phải thường xuyên để có thể biết được sự ảnh hưởng của nó tới năng suất.

Quan sát đất đai hay được chú ý ở thí nghiệm về phân bón hoặc biện pháp làm đất, tưới nước, luân canh... Tuy nhiên, phải làm liên tục nhiều vụ, nhiều năm (nhóm thí nghiệm lâu năm hay dài hạn) như vậy trước và sau thí nghiệm phải phân tích đất mới có cơ sở để giải thích chính xác.

Nhưng không thể quan sát tất cả các nhân tố được, tùy thuộc vào nhân lực, kinh phí của đề tài để lựa chọn cho thực sự có ý nghĩa khoa học và ý nghĩa kinh tế với mục tiêu đề tài đặt ra.

Tóm lại, các thí nghiệm về sinh lý và sinh hoá cần quan sát về cây là chủ yếu, các thí nghiệm có liên quan đến thổ nhưỡng thì nặng phân tích về đất. Các thí nghiệm về phân bón (dinh dưỡng) cần quan sát hài hoà cả đất và cây. Các chỉ tiêu nghiên cứu phải dựa hoàn toàn và thực hiện đúng như đề cương đã xây dựng cho thí nghiệm. Tránh làm tùy tiện.

3.3.2. Phương pháp chọn mẫu nghiên cứu

Mục đích là mẫu phải đảm bảo tính đại diện, khách quan, chính xác và dựa trên quan điểm toán học xác suất thống kê. Tuy nhiên, để đạt mục đích trên còn phải kết hợp với cả độ lớn của mẫu nghiên cứu mới đầy đủ.

3.3.2.1. Chọn mẫu ngẫu nhiên

Đây là phương pháp chọn mẫu mà các cá thể được lấy ra quan sát, đo đếm là hoàn toàn ngẫu nhiên.

Cách thực hiện: toàn bộ các cá thể trong ô thí nghiệm được đánh số sau đó gấp thăm hoặc tra bảng ngẫu nhiên để chọn ra được các cá thể của mẫu (loại trừ các cá thể ở hàng biên).

* Ưu điểm: mẫu nghiên cứu mang tính khách quan và các giá trị thu được tuân theo quy định của đại lượng ngẫu nhiên, do đó các tham số của mẫu mang tính đại diện, nhưng các cá thể trong mẫu mang tính biến động (không đồng đều). Song đó là hiện trạng của thí nghiệm (tính chân thực) độ chính xác của kết quả nghiên cứu cao.

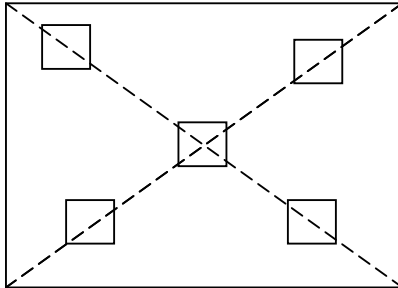
* Tuy nhiên phương pháp chọn mẫu ngẫu nhiên cũng có các tồn tại đó là: khi dung lượng mẫu không đủ lớn có thể dẫn đến kết quả không chính xác (tính đại diện thấp). Bên cạnh đó việc thực hiện lấy mẫu phức tạp và tốn thời gian.

Thực tế các nhà khoa học thực nghiệm Việt Nam hiện nay còn đang ít áp dụng phương pháp chọn mẫu này.

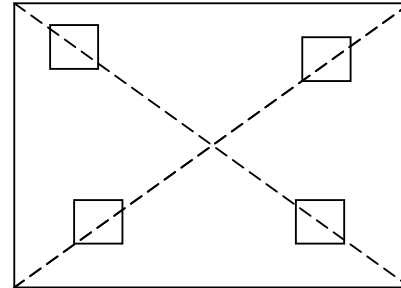
3.3.2.2. Chọn mẫu phân phối đều

Chọn phân phối đều ở đây có thể thực hiện trên ô thí nghiệm, hoặc trên khu vực điều tra. Phân phối đều có hai dạng sau:

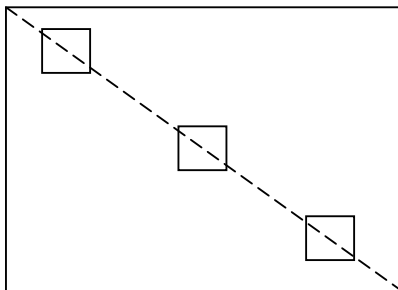
Dạng thứ nhất: Phân phối đều theo đường chéo, minh họa trên sơ đồ 3.6.



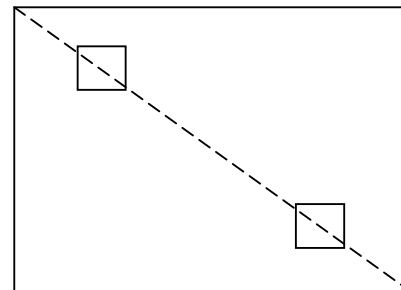
Sơ đồ (3.6)a phân phối đều theo đường chéo chéo 5 điểm



Sơ đồ (3.6)b đường chéo 4 điểm



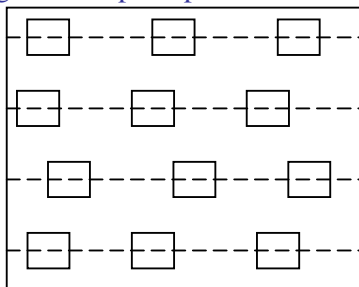
Sơ đồ (3.6)c đường chéo 3 điểm



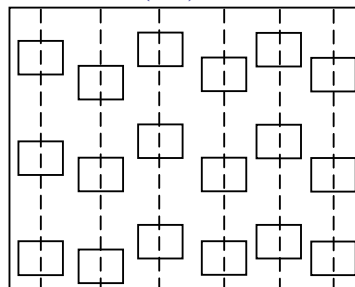
Sơ đồ (3.6)d đường chéo 2 điểm

Trong dạng này, việc lấy mẫu theo đường chéo phân phối đều là tốt hơn cả, vì nó đại diện cho nhiều vị trí trên mảnh thí nghiệm do đó tính chính xác sẽ cao và hiện nay hầu hết các nhà khoa học Nông học đang áp dụng cách chọn mẫu này.

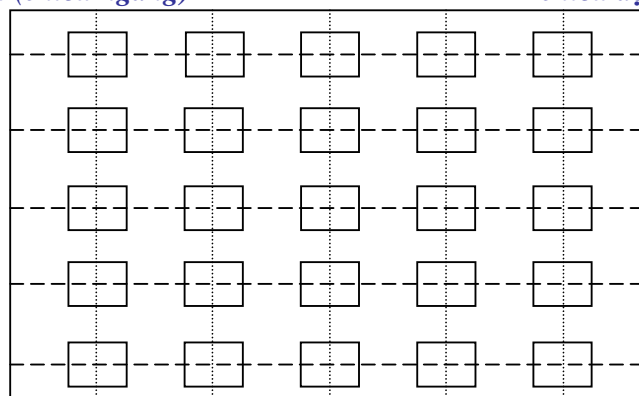
Dạng thứ hai: phân phối đều theo tuyến. Như hình ở sơ đồ (3.7) sau



Sơ đồ (3.7)a phân phối đều theo tuyến một chiều (chiều ngang)



Sơ đồ (3.7)b phân phối đều theo tuyến một chiều dọc



Sơ đồ (3.7)c phân phối đều theo mạng lưới

Trong dạng lấy phân phối đều, kiểu mạng lưới thường hay được áp dụng hơn cả nhất trong lĩnh vực điều tra.

Ưu điểm: các tham số thống kê thu được theo cách lấy mẫu này có độ chính xác cao và khách quan. Cách này không phức tạp, vẫn đảm bảo tính ngẫu nhiên và tùy thuộc vào số lượng cá thể của thí nghiệm mà chọn mẫu có độ lớn khác nhau. Nếu mẫu nhỏ vẫn ít tạo ra sự biến động của mẫu với hiện trạng thực trong thí nghiệm.

3.3.3. Độ lớn của mẫu

Hay còn gọi là "dung lượng" mẫu, độ lớn của mẫu (n) lớn hay nhỏ phụ thuộc vào các điều kiện sau:

- Loại chỉ tiêu nghiên cứu.
- Độ chính xác của thí nghiệm.
- Số lượng cá thể (cây) có trong mỗi công thức thí nghiệm.

Theo kết quả nghiên cứu của Nguyễn Thị Lan (1995, 1996 và 2002, 2005) cho biết:

Trên đối tượng là cây lạc, cây đậu tương và lúa cho thấy: dung lượng mẫu phụ thuộc vào loại chỉ tiêu nghiên cứu vì có chỉ tiêu biến động ít (chiều cao cây, số nhánh đẽ hay số cành quả/cây, số quả hay số hạt/bông...), có chỉ tiêu biến động nhiều. Vì vậy, việc lấy một dung lượng mẫu thống nhất như nhau là không hợp lý, cụ thể chấp nhận sai số $\leq 5\%$.

Với lạc và đậu tương ở các chỉ tiêu như: chiều cao cây, lấy 30 là đủ, số quả trên cây, số thân/cây (với lạc) và khối lượng quả/cây nên lấy theo đôi 70 - 80 cây cho mỗi công thức.

Theo dõi về khối lượng 1000 hạt của đậu tương và lúa Nguyễn Thị Lan (2002, 2005) đề nghị nên cân mỗi mẫu có số hạt là 200 hoặc 250 hạt cân 3 mẫu hoặc 5 mẫu và đã đảm bảo cho sai số < 5 %.

Cũng có thể dựa vào công thức sau để xác định dung lượng mẫu nghiên cứu

$$n_{ct} = \frac{40.000 \times s^2}{(\bar{x})^2 \times (\Delta\%)^2} \quad (3.4)$$

Ở đây n_{ct} : dung lượng mẫu cần thiết nghiên cứu

s^2 : phương sai mẫu lấy thử

\bar{x} : trung bình của mẫu lấy thử

$\Delta\%$: sai số cho phép của chỉ tiêu nghiên cứu

Mẫu lấy phải được phân phối đều ở tất cả các lần nhắc lại. Về thời điểm quan sát được chia làm 2 loại:

Loại 1: Theo dõi cố định, nghĩa là chọn cố định các cây theo dõi dựa vào mẫu đã xác định và sau đó theo dõi liên tiếp cả quá trình (lần theo dõi này cách lần theo dõi khác ở một số ngày nhất định). Loại này thường áp dụng ở các chỉ tiêu thuộc quá trình sinh trưởng (chiều cao cây, tốc độ ra lá, số nhánh đẻ...).

Loại 2: Theo dõi định kỳ có nghĩa là tùy theo từng thời kỳ cần thiết mà quan sát hoặc lấy mẫu phân tích, loại này thường đặt ra với các chỉ tiêu như: xác định khối lượng tích lũy chất khô; xác định năng suất sinh học, hàm lượng sắc tố, chỉ số diện tích lá; hiệu suất quang hợp; hàm lượng các chất N, P, K.... Đây là những chỉ tiêu khi lấy mẫu theo dõi phải phá vỡ các đối tượng nghiên cứu.

3.4. Thu hoạch thí nghiệm

Đây là công việc cuối cùng của quá trình thực hiện thí nghiệm trên ruộng nhưng 1 hết sức có ý nghĩa.

3.4.1. Công tác chuẩn bị

Chuẩn bị đầy đủ các dụng cụ phục vụ cho việc thu hoạch thí nghiệm như: các dụng cụ bao gói để đựng, phơi sấy nông sản và lưu giữ năng suất.

Thu dải bảo vệ: có thể tiến hành trước hoặc sau khi thu thí nghiệm hoặc thu sau và phải để riêng một khu vực không để gần chỗ để phân thu các công thức tránh nhầm lẫn.

3.4.2. Lấy mẫu định yếu tố cấu thành năng suất và năng suất cá thể

Công việc này phải làm trước thu hoạch toàn thí nghiệm và phải làm riêng rẽ từng cây từng lần nhắc lại. Tùy thuộc vào mỗi loại cây mà có các chỉ tiêu năng suất khác nhau.

3.4.3. Thu hoạch năng suất

Việc này phải làm riêng từng ô ở từng lần nhắc lại. Từ việc thu hoạch, tuốt đập và cân đong, có thể cân năng suất tươi hoặc khô ở độ ẩm tiêu chuẩn, hoặc phơi thật khô (sấy khô) để cân năng suất.

3.4.4. Phương pháp điều chỉnh năng suất ô thí nghiệm

Chỉ đặt ra khi số cây trên ô bị khuyết do yếu tố khách quan, nếu không xem xét để điều chỉnh sẽ có thể làm cho kết quả nghiên cứu kém chính xác. Vì vậy, điều chỉnh trước khi đánh giá kết quả nghiên cứu.

Hiện này có nhiều cách để giải quyết vấn đề này:

- Điều chỉnh năng suất của ô theo năng suất cá thể trung bình.

$$Y' = \bar{X} \times N \quad (3.5)$$

Ở đây

\bar{X} : năng suất cá thể trung bình của ô

Y': năng suất ô đã hiệu chỉnh

N: số cây trong ô

- Theo Zalemski

$$Y' = \frac{1}{2} [Y + (\bar{X} \times N)] \quad (3.6)$$

Ở đây: \bar{X} : năng suất cá thể trung bình của ô

Y: năng suất thực của ô

Y': năng suất ô đã hiệu chỉnh

N: số cây trong ô

- Theo Derevitski thì

$$Y' = Y \frac{x + \bar{x}}{2x} \quad (3.7)$$

Trong đó: x : số cây thực tế còn trong ô

\bar{x} : số cây trung bình của ô

Y' và Y như công thức (3.6) đã nêu

Bài tập: Sinh viên thực tập trên ruộng

CHƯƠNG IV - TỔNG KẾT SỐ LIỆU QUAN SÁT

Chương này đề cập đến tham số cơ bản của mẫu thường được dùng để phân tích các kết quả nghiên cứu sinh học nói chung cũng như ngành nông nghiệp nói riêng. Một số tham số thường được sử dụng để trình bày kết quả nghiên cứu khoa học trong các ấn phẩm.

4.1. Chinh lý số liệu quan sát

4.1.1. Khái niệm về số liệu thô và số liệu tinh

4.1.1.1. Số liệu thô

Số liệu thô là loại số liệu thu được trực tiếp ngay trên đồng ruộng (ghi chép).

Thí dụ: theo dõi số hạt/bông cho lúa ở 3 công thức thí nghiệm như sau:

Bảng 4.1. Số hạt lúa /bông

Bông	CTI	CTII	CTIII	Bông	CTI	CTII	CTII
1	156	192	199	16	177	205	194
2	172	205	196	17	169	194	245
3	196	213	223	18	162	215	256
4	152	178	178	19	225	205	217
5	168	196	157	20	216	214	198
6	152	157	199	21	198	224	205
7	176	199	214	22	205	224	250
8	161	252	230	23	212	234	217
9	145	186	214	24	224	218	219
10	121	200	189	25	198	242	234
11	172	218	192	26	204	235	244
12	119	234	190	27	205	196	198
13	153	246	205	28	192	225	202
14	162	196	223	29	249	205	215
15	179	185	216	30	243	217	237

Từ số liệu trong bảng 4.1 ta rất khó xác định số hạt bình quân /bông của từng công thức một các đại diện.

Để giúp giải thích các vấn đề nêu trên cần phải chuyển số liệu thô sang số liệu tinh

4.1.1.2. Số liệu tinh

Số liệu tinh là số liệu được tính trực tiếp từ số liệu thô theo những tham số thống kê cần thiết.

Từ số liệu tinh mới có thể đánh giá sự khác nhau ở các chỉ tiêu nghiên cứu trong các công thức thí nghiệm. Ví dụ số liệu tinh của kết quả nghiêm cứu được thể hiện qua bảng sau (bảng 4.2)

Bảng 4.2. Năng suất cà phê chè trong 3 năm ở các công thức trên đất Daklak
(kg quả tươi/cây)

Công thức	1984	1985	1986	Bình quân	% so đối chứng
Đối chứng	4.2	0.2	4.	3.0	100
Sicarol	5.2	0.9	7.2	4.4	146.6
Bayleton bột	4.8	2.9	8.8	5.5	188.3
Bayleton nước	5.0	3.1	9.5	5.9	196.6
Bordeaux	4.5	1.7	7.1	4.4	146.6

4.2. Phân loại số liệu

Các số liệu trong nghiên cứu thu được có thể được chia thành 2 loại là: Số liệu định lượng (hay còn gọi là số lượng) và số liệu định tính.

4.2.1. Số liệu định lượng

Là những số liệu mà người quan sát có thể cân, đong, đo, đếm được dễ dàng và chính xác.

Thí dụ như: chiều cao cây, khối lượng hạt, năng suất...

4.2.2. Số liệu định tính

Số liệu này không, đo, đếm, cân đong mà chỉ chia thành một số loại (lớp, cấp)

Thí dụ như: Màu sắc hạt, tình hình sâu, bệnh, hình dạng hạt...

4.3. Kiểm tra số liệu nghi ngờ

Trong quá trình thực hiện theo dõi thí nghiệm khó có thể đoán hết được sai sót, nhất là sai sót do sai số thô gây ra (như đã trình bày trong chương 2 mục 2.1.3). Để có thể khẳng định số liệu trước khi đưa vào phân tích nên loại bỏ ngay hay tạm để ra ngoài. Phương pháp xử lý như sau:

$$w_{tn} = \frac{x_i^* - \bar{x}}{s} \quad (4.1)$$

Ở đây

x_i^* : số liệu nghi ngờ

\bar{x} : trung bình mẫu (có tính cả số liệu nghi ngờ)

s : độ lệch chuẩn mẫu

ω_{tn} : tiêu chuẩn kiểm tra (giá trị thực tính) giá trị này sẽ được so sánh với giá trị lý thuyết là:

ω_{α} : với bậc tự do $df = n - 1$ ở mức ý nghĩa α (tra bảng 13 phần phụ lục)

α : ở mức ý nghĩa 0,01; df là bậc tự do; n là dung lượng mẫu

Nếu $\omega_{tn} < \omega_{t}$ chấp nhận (để lại) số nghi ngờ.

Nếu $\omega_{tn} > \omega_{t}$ loại bỏ số nghi ngờ (loại ra khỏi dãy số liệu quan sát).

Thí dụ theo dõi chiều cao cây lúa ở $n = 10$ khóm có các giá trị sau (cm): 98,2; 92,0; 82,7; 92,5; 89,0; 87,9; 99,2; 99,5; 97,0; 100,5.

Trong đó có giá trị 82,7 cm là nghi ngờ nên cần được kiểm tra. Vậy, có $\bar{X} = 93,9$ cm; $s = 6,0$ cm.

$$\omega_{t,n} = \frac{82,7 - 93,9}{6,0} = |1,87|$$

$$\omega_{(0,01,df=9)} = 2,33 \quad (\text{bảng 13 phụ lục})$$

Vậy $\omega_{t,n} < \omega_{(0,01;9)}$ ($\omega_{tn} < \omega_{t}$)

Nên chấp nhận giá trị $x_i^* = 82,7$ cm trong dãy đo 10 khóm lúa theo dõi chiều cao cây.

4.4. Cách sắp xếp số liệu

4.4.1. Với số liệu định lượng

Nếu như dãy số liệu quan sát $n < 30$ thì vấn đề trình bày số liệu đơn giản, nhưng nếu như dãy số liệu nhiều nên tiến hành phân nhóm (hay tổ) số liệu. Vậy phân bao nhiêu nhóm (tổ) là vừa. Khó có thể có câu trả lời hoàn toàn chính xác cho vấn đề này. Kinh nghiệm cho thấy số tổ tùy theo giá trị của chỉ tiêu theo dõi mà chia từ 5 – 20. Có thể tham khảo một vài công thức sau đây.

- Theo Dospekhov (1979) thì

Khi có 30 - 60 số liệu thì chia 6 - 7 tổ.

60 - 100 số liệu thì chia 7 - 8 tổ.

Trên 100 số liệu thì nên chia 8 - 15 tổ.

- Theo nhóm thống kê Trường Đại học Nông nghiệp Bắc Kinh thì

Khi có 40 - 60 số liệu nên chia 6 - 8 tổ.

60 - 100 số liệu nên chia 7 - 10 tổ

100 - 200 số liệu nên chia 10 - 12 tổ.

200 - 500 số liệu nên chia 12 - 17 tổ.

Trên 500 số liệu nên chia 17 - 20 tổ.

Để xác định số tổ cần phân chia, theo B.Rooke và Carruther có thể tính theo công thức sau:

$$k = 5 \lg.n \quad (4.2)$$

Ở đây, k : số tổ phân chia

n : số lượng cá thể quan sát (số số liệu).

Sau khi xác định được số tổ (là số nguyên), ta sẽ tính được khoảng cách tổ. Nếu gọi khoảng cách tổ là h thì:

$$h = \frac{x_{\max} - x_{\min}}{K} \quad (4.3)$$

Ở đây: h là khoảng cách tổ

x_{\max} : giá trị lớn nhất của dãy số liệu

x_{\min} : giá trị nhỏ nhất của dãy số liệu

Sau đó sẽ sắp xếp các tổ và lập bảng phân phối tần số.

Bảng 4.3. Kết quả đo chiều cao của 100 cây cà chua (cm)

76	73	75	73	74	74	74	74	74	77
74	72	75	76	73	71	73	80	75	75
68	72	78	74	75	74	69	77	77	72
72	76	76	77	70	77	72	74	77	76
78	72	70	74	76	72	73	71	74	74
75	79	75	74	75	74	71	73	75	73
75	70	73	75	70	72	72	71	76	73
74	76	74	75	74	76	75	75	73	73
78	74	73	75	74	73	72	76	73	76
74	71	72	71	79	78	69	77	73	71

Nếu sử dụng việc xác định số tổ bằng công thức: $k = 5 \cdot \lg n$

Thay số vào có $k = 5 \times \lg 100 = 10$ tổ

$x_{\min} = 68$ cm và $x_{\max} = 80$ cm

$$\text{Khoảng cách tổ } h = \frac{80 - 68}{10} = 1,2 \text{ cm}$$

Như vậy, sắp xếp các tổ sẽ gặp khó khăn vì khoảng cách lẻ nếu làm tròn là 1,0 thì không được chấp nhận, nếu làm tròn là 2,0 lại quá lớn

Vậy nên chia dãy số liệu làm 12 tổ. Như vậy sẽ có khoảng cách tổ

$$h = \frac{80 - 68}{12} = 1,0 \text{ cm}$$

Bảng phân phối tần số được xác định như sau

Số tổ	Giới hạn tổ	Giá trị giữa tổ (x_i)	Kết quả quan sát	Tần số tổ (m_i)
1	68 - 69	68,5		1
2	69 - 70	69,5	┌	2
3	70 - 71	70,5	□	4
4	71 - 72	71,5	□ □	7
5	72 - 73	72,5	□ □ □	11
6	73 - 74	73,5	□ □ □ □	15
7	74 - 75	74,5	□ □ □ □ □	20
8	75 - 76	75,5	□ □ □ □	15
9	76 - 77	76,5	□ □ □	11
10	77 - 78	77,5	□ □	7
11	78 - 79	78,5	□	4
12	79 - 80	79,5	└	3

4.4.2. Số liệu định tính

Đây là loại số liệu rất khó có khả năng định lượng chính xác nên việc chỉnh lý cũng không đơn giản. Vì vậy, phải định ra được các tiêu chuẩn, trên cơ sở đó sẽ sắp xếp vào nhóm.

Thí dụ: Tìm hiểu mức độ hại của bệnh bạc thau dâu

Chúng tôi xác định cấp lá bệnh như sau

Cấp 0: không có lá bệnh

Cấp 1: tỷ lệ bệnh xuất hiện 0 - 10 %

Cấp 2: tỷ lệ bệnh xuất hiện 11 - 20 %

Cấp 3: tỷ lệ bệnh xuất hiện 21 - 30 %

Cấp 4: tỷ lệ bệnh xuất hiện 31 - 40 %

Cấp 5: tỷ lệ bệnh xuất hiện > 40 %

Nếu gọi m_i là tần số của các cấp bệnh quan sát được và n là dung lượng mẫu (số cá thể lấy quan sát) thì

$$n = m_0 + m_1 + m_2 + m_3 + m_4 + m_5 = \sum_{i=1}^{\sigma} m_i \quad (4.4)$$

Xác suất của mỗi cấp bệnh là

$$f_i = \frac{m_i}{n} \quad (4.5)$$

Bảng 4.5. Số lá dâu bị bệnh bậc thau các cấp bệnh và tỷ lệ cấp bệnh

TT	Cấp bệnh	Tần số (m_i)	Tần suất (f_i)	Tỷ lệ %
1	0	0	0	0
2	1	45	0,450	45,0
3	2	40	0,400	40,0
4	3	10	0,100	10,0
5	4	5	0,050	5,0
6	5	0	0	0
	Σ	100	1,000	100

Như vậy, bệnh chủ yếu là mức cấp 1 và cấp 2

4.5. Các tham số đặc trưng của mẫu

Để có những nhận định khái quát cho tổng thể nghiên cứu về độ lớn hay biến động của chỉ tiêu theo dõi cần phải tính được các tham số đại diện.

4.5.1. Các tham số đại diện về vị trí của mẫu

4.5.1.1. Trung bình (còn gọi là trung bình cộng - \bar{x})

Số trung bình còn có tên gọi là trung bình số học, đây là tham số được sử dụng nhiều nhất. Số bình quân là tham số đại diện cho độ lớn trung bình của mẫu nghiên cứu.

Giá trị này được tính như sau:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (4.6)$$

Trong đó: \bar{x} trung bình mẫu

x_i : giá trị quan sát thứ i

n : dung lượng mẫu

$$\text{hoặc: } \bar{x} = \frac{\sum_{i=1}^k x_i m_i}{n}; \quad n = \sum_{i=1}^k m_i \quad (4.7)$$

Ở đây: x_i : giá trị đại diện (giữa) nhóm thứ i

m_i : tần số nhóm thứ i

k : nhóm phân chia

Số trung bình theo công thức này gọi là trung bình có trọng lượng hay trung bình có trọng số.

Thí dụ: Đo chiều cao của 20 khóm lúa giống P₄ lấy ngẫu nhiên có kết quả sau (đv: cm)

95	102	100	99	91	95	95	97	101	102
92	93	93	94	91	96	97	100	92	95

Chiều cao trung bình tính được là

$$\bar{x} = \frac{95 + 102 + \dots + 102 + \dots + 92 + 95}{20} = 96 \text{ (cm)}$$

Hoặc tính theo trọng số

$$\bar{x} = \frac{(91 \times 2) + (92 \times 2) + (93 \times 2) + 94 + (95 \times 4) + 96 + (97 \times 2) + 99 + (100 \times 2) + 101 + (102 \times 2)}{20} = 96 \text{ (cm)}$$

Như vậy 2 cách tính cho kết quả như nhau

4.5.1.2. Số một (Mode)

Một là số liệu có tần số hay số lần xuất hiện nhiều nhất trong dãy số quan sát, thí dụ nêu trên số một sẽ là 95cm

4.5.1.3. Số trung vị (Median - M_e)

Nếu sắp xếp số liệu theo trật tự từ nhỏ đến to thì trung vị là số đứng ở vị trí trung gian chia dãy số liệu làm 2 nửa bằng nhau.

Thí dụ: Khảo sát số quả trên cây cà chua của 7 cây giống số 6 có kết quả sau (đv: quả/cây)

22	23	25	26	28	29	30
----	----	----	----	----	----	----

Các giá trị đã được sắp xếp thứ tự tăng dần, nên số trung vị sẽ là cây thứ 4 có số quả là 26 quả/cây.

Trong dãy số này n = 7 (lẻ) số trung vị có vị trí thứ 4 và giá trị trung vị là 26. Do đó, công thức tổng quát tìm giá trị trung vị sẽ là:

$$M_e = x_{\left(\frac{n+1}{2}\right)} \tag{4.8}$$

Nhưng nếu n là số chẵn thì giá trị trung vị sẽ được tính theo công thức tổng quát sau

$$M_e = \frac{x_{[n/2]} + x_{[n/2] + 1}}{2} \tag{4.9}$$

$x_{[n/2]}$ với $[n/2]$ là phần nguyên của $n/2$

Trường hợp này có số trung vị, song không có vị trí của số trung vị.

4.5.1.4. Số trung bình nhân (trung bình hình học - \bar{x}_g)

Đây là giá trị dùng biểu thị nhịp điệu tăng trưởng một chỉ tiêu nào đó trong thời gian nghiên cứu. Trung bình nhân được tính theo công thức sau:

$$\bar{x}_g = \sqrt[n]{x_1 \times x_2 \times \dots \times x_i \times \dots \times x_n} = \sqrt[n]{\prod_{i=1}^n x_i} \quad (4.10)$$

(\bar{x}_g): trung bình nhân
n: dung lượng mẫu
 x_i : giá trị quan sát thứ i

4.5.2. Các tham số đại diện cho sự phân tán của mẫu

Các tham số chỉ sự phân tán hay biến động của mẫu gồm:

4.5.2.1. Phương sai mẫu (s^2)

Phương sai được coi là tham số cơ bản nhất đại diện cho tính phân tán của dãy số liệu quan sát. Phương sai được tính bằng công thức:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad \text{Ở đây: } n - 1 \text{ gọi là bậc tự do mẫu} \quad (4.11)$$

$$\text{hoặc: } s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot m_i}{n - 1} \quad (4.12)$$

Phương sai này còn có tên gọi là phương sai thống kê hay phương sai mẫu đã hiệu chỉnh.

4.5.2.2. Độ lệch chuẩn mẫu (s)

Độ lệch chuẩn là tham số được tính từ phương sai mẫu qua công thức:

$$s = \sqrt{s^2} \quad (4.13)$$

4.5.2.3. Độ lệch chuẩn của trung bình \bar{x} hay sai số chuẩn ($s_{\bar{x}}$)

Độ lệch chuẩn của trung bình \bar{x} tính bằng công thức:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \quad \text{hay} \quad s_{\bar{x}} = \sqrt{\frac{s^2}{n}} \quad (4.14)$$

4.5.2.4. Biên độ dao động của dãy số liệu (Range)

Biên độ dao động R là chênh lệch giữa giá trị nhỏ nhất và giá trị lớn nhất trong dãy số liệu quan sát

$$R = x_{\max} - x_{\min} \quad (4.15)$$

4.5.2.5. Hệ số biến động (CV%)

Đây là tham số thống kê cho phép so sánh mức độ biến động của nhiều mẫu khác nhau ở các chỉ tiêu nghiên cứu khác nhau. Do đó, hệ số biến động được sử dụng phổ biến trong đánh giá kết quả nghiên cứu. Hệ số biến động được tính như sau:

$$CV\% = \frac{s}{\bar{x}} \times 100 \quad (4.16)$$

Ở đây, s : độ lệch chuẩn

\bar{x} : trung bình mẫu

Thí dụ: Đã tính được chiều cao trung bình của 20 khóm với giống lúa P₄ là 96 cm. Song chưa biết được mức độ biến động về chiều cao cụ thể. Vì vậy, có thể tính được các tham số biến động của chiều cao cây P₄ theo công thức minh họa sau:

$$\begin{aligned} s^2 &= \frac{\sum_{i=1}^{20} (x_i - \bar{x})^2}{n - 1} \\ &= \frac{(95 - 96)^2 + (102 - 96)^2 + \dots + (95 - 96)^2}{19} = \frac{251}{19} = 13,21053 \text{ (cm)}^2 \end{aligned}$$

Như vậy, phương sai về chiều cao 20 cây giống lúa P₄ là 13,21053 (cm)².

Độ lệch chuẩn của giống P₄

$$s = \sqrt{s^2} = \sqrt{13,21053 \text{ (cm)}^2} = 3,635 \text{ cm} \approx 3,6 \text{ cm}$$

- Độ lệch chuẩn của số trung bình \bar{x}

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{3,6}{\sqrt{20}} = 0,805 \text{ cm} \approx 0,80 \text{ cm}$$

$$\text{Hệ số biến động } CV\% = \frac{3,6}{96} \times 100 = 3,75\%$$

4.6. Các tham số đặc trưng của số liệu định tính

Các số liệu định tính (đặc trưng chất lượng) thường được biểu thị dưới dạng tần số (số nguyên) hoặc biểu thị dưới dạng tần suất hay tỷ lệ (%).

Trong mẫu n cá thể được chia thành các lớp (nhóm) A; B; C... với các tần số tương ứng $m_1; m_2; m_3 \dots$. Các tần suất

$$f_1 = \frac{m_1}{n}; f_2 = \frac{m_2}{n}; f_3 = \frac{m_3}{n} \dots f_k = \frac{m_k}{n}$$

Hay viết tổng quát số liệu định tính thứ i có tần suất f_i và

$$f_i = \frac{m_i}{n} \quad (4.17)$$

Để đánh giá mức độ xuất hiện nhiều hay ít của một đặc trưng định tính nào đó, các nhà thực nghiệm thường sử dụng tần suất

$$f_i = \frac{m_i}{n}$$

Hoặc tỷ lệ

$$f_i \% = \frac{m_i}{n} \times 100 \quad (4.18)$$

Đánh giá mức độ biến động của số liệu định tính, có thể sử dụng tham số độ lệch chuẩn (s_p). s_p gọi là độ lệch chuẩn của số liệu định tính, tham số này tính bằng công thức sau:

$$s_p = \sqrt[k]{f_1 \times f_2 \times \dots \times f_k} = \sqrt[k]{\prod_{i=1}^k f_i} \quad (4.19)$$

Ở đây: f_i : là tần suất của nhóm định tính thứ i

k : là số nhóm định tính; $i = 1, \dots, k$

Để thuận tiện có thể chuyển công thức tính độ lệch chuẩn của số liệu định tính như sau

$$\begin{aligned} \lg s_p &= \frac{1}{k} (\lg f_1 + \dots + \lg f_i + \dots + \lg f_k) \\ &= \frac{1}{k} \sum_{i=1}^k \lg f_i \end{aligned} \quad (4.20)$$

Dựa vào số nhóm định tính đã phân chia có thể tính được giá trị s_p cực đại (s_{pmax}) như sau:

Bảng 4.6. Độ lệch chuẩn cực đại trong số liệu định tính

Số nhóm k	Giá trị s_{pmax}	Số nhóm k	Giá trị s_{pmax}
2	0,500 (50,0 %)	5	0,200 (20,0 %)
3	0,333 (33,3 %)	6	0,167 (16,7 %)
4	0,250 (25,0 %)	7	0,143 (14,3 %)

Giá trị s_{pmax} phụ thuộc vào số lớp (nhóm) phân chia và sự biến động của chúng.

Với số liệu định tính cũng có thể tính được hệ số biến động theo công thức sau

$$CV \% = \frac{s_p}{s_{pmax}} \times 100 \quad (4.21)$$

Trong trường hợp dung lượng mẫu n đủ lớn ($n \geq 120$) có thể dùng độ lệch chuẩn của trung bình số liệu định tính (s_p^-)

$$s_p^- = \frac{s_p}{\sqrt{n}} \quad (4.22)$$

4.7. Một số quy tắc về làm tròn số trong tính toán

Kết quả nghiên cứu từ thực nghiệm là những giá trị ngẫu nhiên và độc lập. Vì vậy, khi tính toán cần thiết phải có những nguyên tắc vừa đảm bảo tính chính xác vừa đảm bảo ý nghĩa của các giá trị ở mẫu đại diện.

4.7.1. Con số có nghĩa

Nghiên cứu thực nghiệm chỉ có thể thực hiện ở một mẫu với độ lớn n , trong đó các giá trị x_i là độc lập và ngẫu nhiên. Do đó, khi tính toán các tham số thống kê cần thiết, kết quả cuối cùng sẽ có những giá trị hết sức lẻ (nhiều số thập phân). Song kết quả cuối cùng cũng nên chỉ chấp nhận con số có nghĩa (lưu ý ở phần chữ số thập phân) bằng với các giá trị quan sát x_i hay các giá trị trong phép tính. Điều này vừa đảm bảo tính chính xác vừa đảm bảo ý nghĩa các chỉ tiêu nghiên cứu trong thực tế.

Thí dụ: Theo dõi một mẫu có $n = 12$ cây cà chua vụ xuân hè với giống số 48 tại Từ Liêm - Hà Nội năm 2002

Các kết quả qua quan sát chiều cao cây sau trồng 45 ngày như sau (cm)

59,0 59,3 61,0 55,1 61,5 63,7
68,5 62,7 57,8 60,1 61,2 62,0

Như vậy chiều cao trung bình $\bar{x} = \frac{731,9}{12} = 60,99167 \text{ cm}$

Tuy nhiên, các x_i quan sát chỉ lấy một số lẻ (chính xác 1/10 cm). Vì vậy, nếu lấy 3 con số có nghĩa là $\bar{x} = 61,0 \text{ cm}$

Thí dụ: Theo dõi số hạt trên bông lúa vụ xuân của 10 bông lấy mẫu, các giá trị quan sát là

102 115 129 105 101 100 95 108 102 104

Vậy khi tính số hạt bình quân của một bông ta được giá trị tính toán

$$\bar{x} = \frac{1061}{10} = 106,1 \text{ hạt/bông}$$

Song số hạt của một bông lại là số nguyên, không có số lẻ khi quan sát. Do đó, chỉ nên lấy giá trị bình quân là số nguyên sẽ có ý nghĩa, như vậy số hạt bình quân của một bông là $\bar{x} \approx 106$ hạt. Tuy nhiên cũng có thể giữ nguyên $\bar{x} = 106,1$ hạt/bông vì khi tính trung bình có thể lấy thêm 1 số lẻ và độ lệch chuẩn s lấy 2 số lẻ.

4.7.2. Cách làm tròn số (quy tắc xấp xỉ)

Sau khi đã xác định được số chữ số có nghĩa phải tiến hành làm tròn số. Điều này hầu như luôn xảy ra trong tính toán các kết quả thực nghiệm.

Quy định chiều cao cây lấy chính xác tới 1/10 (cm), do đó, kết quả cuối cùng sẽ lấy thêm một số thập phân.

Giả sử $\bar{x} = 125,543 \text{ cm}$, chỉ quy định chỉ lấy một số lẻ, vì vậy $\bar{x} \approx 125,5 \text{ cm}$ hoặc nếu có trung bình $\bar{x} = 106,876 \text{ cm} \rightarrow \bar{x} \approx 106,9 \text{ cm}$.

Bài tập:

4.1. Theo dõi chiều cao lúa kang dân 18 vụ xuân giai đoạn đẻ nhánh ta có các số liệu sau (cm):

21; 20; 23; 20; 19; 20; 18; 23; 24; 22; 26; 24; 22; 25; 21; 23; 23; 26; 22; 22; 26; 28; 20; 21; 26; 21; 20; 24; 23; 23; 23; 22; 22; 18; 19; 19.

Hỏi:

a) Tính trung bình (\bar{x}) của chiều cao cây với giống Kang dân 18 và vẽ đồ thị phân phối tần suất của chỉ tiêu.

b) Hãy tính tham số khác như (số mod, trung vị, phương sai, độ lệch chuẩn, độ lệch chuẩn của số bình quân và hệ số biến động)

4.2 Điều tra bệnh đạo ôn hại lúa ở 105 khóm lúa có kết quả sau:

Không bị bệnh: 25 khóm

Bệnh hại nhẹ: 40 khóm

Bệnh hại trung bình: 25 khóm

Bệnh hại nặng: 15 khóm

Hỏi

a) Hãy tính tần suất (tỷ lệ) bị bệnh ở các mức khác nhau trong mẫu nghiên cứu.

b) Hãy tính các tham số như: độ lệch chuẩn, hệ số biến động của dãy số bên trên.

CHƯƠNG V ƯỚC LƯỢNG

Chương này sẽ giới thiệu các dạng ước lượng cụ thể đối với số trung bình của một đặc trưng định lượng và xác suất của một đặc trưng định tính nào đó (tỷ lệ) trong một quần thể (hay công thức). Đây là hai đặc trưng mà các nhà nghiên cứu luôn dùng khi công bố các kết quả nghiên cứu khoa học.

5.1. Đặt vấn đề

Như chúng ta đã biết, đối tượng nghiên cứu trong nông nghiệp khá phức tạp, trong quá trình nghiên cứu không thể quan sát và đo đếm tất cả các cá thể có của quần thể (công thức) với những lý do sau:

- Không có điều kiện về nhân lực, vật lực và thời gian để theo dõi.
- Phải bảo vệ đối tượng nghiên cứu.

Do đó phải tiến hành lấy mẫu ngẫu nhiên (n) cá thể mang tính đại diện để tiến hành nghiên cứu (quan sát hay đo đếm). Từ kết quả quan sát của mẫu đưa ra kết luận (đánh giá) cho toàn bộ quần thể (công thức). Kết luận đưa ra được gọi là kết luận thống kê. Nên từ quần thể quan sát đưa ra một kết luận (đánh giá) đối với độ lớn của trung bình (hay xác suất) thì ta có một ước lượng.

Từ kết quả của mẫu suy ra kết quả của cả đám đông thì không tránh khỏi sai số chỉ có điều là khả năng và mức độ mắc sai số là như thế nào? Nội dung của chương này sẽ nghiên cứu sai số và khả năng hạn chế sai số đó khi tiến hành ước lượng để đạt tới mong muốn cho phép mà thôi.

5.2. Các phương pháp ước lượng

5.2.1. Ước lượng điểm

Ước lượng điểm của một tham số thống kê nào đó là dạng ước lượng mà từ kết quả quan sát của một mẫu lấy ngẫu nhiên mang tính đại diện của tổng thể, đưa ra một con số và cho rằng con số đó là giá trị gần đúng tốt nhất cho tham số muốn biết.

Thí dụ: Biến ngẫu nhiên X (định lượng hoặc định tính) có phân phối xác suất phụ thuộc vào một tham số θ chưa biết. Từ biến ngẫu nhiên này lấy một mẫu ngẫu nhiên n quan sát.

Gọi x_i là quan sát thứ i , còn x_i là giá trị cụ thể của X_i . Trong mẫu quan sát hàm $f(X_1, X_2, \dots, X_n)$ được dùng để ước lượng θ . Vấn đề đặt ra là chọn hàm nào?

Ký hiệu $Q_n = f(x_1, x_2, \dots, x_n)$ là hàm ước lượng của θ . Q_n là một biến ngẫu nhiên có giá trị cụ thể $q = f(x_1, x_2, \dots, x_n)$. Vậy q là ước lượng điểm của θ .

$$\theta = q \quad (5.1)$$

Có thể tính được độ lệch chuẩn của Q_n và ước lượng điểm lúc này sẽ là:

$$\theta = q \pm \sqrt{D(Q_n)} \quad (5.2)$$

Trong đó $\sqrt{D(Q_n)}$ là độ lệch chuẩn của Q_n .

Thí dụ: Tổng thể có phân phối chuẩn ($\mu; \sigma^2$) μ là trung bình (kỳ vọng) chưa biết cần ước lượng. Lấy n quan sát $x_1, x_2, x_i, \dots, x_n$ tính $\bar{x} = \frac{\sum x_i}{n}$ và $s_x = \frac{s}{\sqrt{n}}$ như vậy có thể đưa ra được ước lượng điểm của kỳ vọng μ .

$$\mu \approx \bar{x} \quad \text{hoặc} \quad \mu \approx \bar{x} \pm \frac{s}{\sqrt{n}} \quad (5.3)$$

5.2.2. Ước lượng khoảng

Ước lượng khoảng của một tham số thống kê nào đó là từ kết quả quan sát của mẫu, đưa ra được một tương ứng với một độ tin cậy nhất định. Mọi giá trị nằm trong khoảng đó đều được coi là giá trị gần đúng tốt nhất của tham số.

Giả sử θ là tham số cần ước lượng. Nếu gọi q_1 là giới hạn dưới và q_2 là giới hạn trên, α là xác suất để mắc sai lầm thì ước lượng khoảng của θ được viết như sau:

$$P(q_1 \leq \theta \leq q_2) = 1 - \alpha = P \quad (5.4)$$

Trong đó:

$[q_1; q_2]$ là khoảng tin cậy của tham số θ

P: Gọi là độ tin cậy (thường lấy với xác suất lớn 0,95; 0,99 và 0,999).

$\alpha = 1 - P$ (thường lấy xác suất nhỏ 0,05; 0,01 và 0,001).

Có thể cấu tạo khoảng tin cậy bằng 3 phương pháp đó là dựa vào phân phối chính xác của hàm ước lượng, dựa vào bất đẳng thức Tsêbusep và phương pháp gần đúng.

Thực tế thì trong 3 phương pháp, phương pháp dựa vào bất đẳng thức Tsêbusep ít sử dụng hơn cả.

5.3. Ước lượng giá trị trung bình của tổng thể (khi đặc trưng nghiên cứu có phân phối chuẩn)

Do chỉ quan sát được n cá thể trong mẫu mà lại mong muốn đánh giá được của toàn công thức (cần biết trung bình của công thức hay còn gọi là kỳ vọng). Cho nên có thể xem xét cụ thể như sau:

5.3.1. Ước lượng trị số trung bình của tổng thể khi dung lượng mẫu n đủ lớn ($n \geq 30$).

Giả sử X có phân phối chuẩn $N(\mu, \sigma^2)$, trong thực tế thì hầu như chúng ta không biết phương sai σ^2 mà chỉ tính được phương sai thống kê của mẫu s^2 . Vì vậy, khi dung lượng mẫu đủ lớn thì có thể coi $\sigma^2 = s^2$. Theo tính chất của phân phối chuẩn chúng ta có:

$$s_x = \frac{s}{\sqrt{n}} \quad \text{hay} \quad \sqrt{\frac{s^2}{n}}$$

Vì vậy, khi phân phối của \bar{x} là tiệm cận với phân phối chuẩn thì kỳ vọng hay trung bình tổng thể μ sẽ được xác định qua ước lượng điểm hoặc ước lượng khoảng như sau:

$$\text{Ước lượng điểm } \mu = \bar{x} \pm s_x \quad \text{và} \quad \mu = \bar{x}$$

Ước lượng khoảng như sau

$$P(\bar{x} - u_{\alpha} s_x \leq \mu \leq \bar{x} + u_{\alpha} s_x) = 1 - \alpha \quad (5.5)$$

U_{α} là giá trị tra ở bảng Φ (phụ lục)

Nếu lấy độ tin cậy P là 0,95 thì $U_{\alpha} = 1,96$; P = 0,99 thì $U_{\alpha} = 2,58$ và P = 0,999 thì $U_{\alpha} = 3,29$.

Tương tự sẽ suy ra khoảng tin cậy cụ thể như sau:

$$P(\bar{x} - 1,96 s_x \leq \mu \leq \bar{x} + 1,96 s_x) = 1 - 0,05 = 0,95$$

$$P(\bar{x} - 2,58 s_x \leq \mu \leq \bar{x} + 2,58 s_x) = 1 - 0,01 = 0,99$$

$$P(\bar{x} - 3,29 s_x \leq \mu \leq \bar{x} + 3,29 s_x) = 1 - 0,001 = 0,999$$

Thí dụ: Chọn một mẫu $n = 50$, điều tra năng suất cá thể của một số giống cà chua xuân hè (kg/cây). Từ đó có năng suất cá thể trung bình $\bar{x} = 1,48$ kg; độ lệch chuẩn của năng suất là 0,35 kg/cây. Hãy đưa ra ước lượng cho năng suất cá thể của cà chua điều tra nêu trên.

Trước hết ta đưa ra ước lượng điểm có năng suất như sau

$$\mu \approx 1,48 \pm \frac{0,35}{\sqrt{50}} \approx (1,48 \pm 0,05) \text{ kg/cây hoặc } \mu = 1,48 \text{ kg/cây.}$$

Ước lượng khoảng ở độ tin cậy P = 0,95 gọi tắt là khoảng tin cậy sẽ là

$$P\left(1,48 - 1,96 \frac{0,35}{\sqrt{50}} \leq \mu \leq 1,48 + 1,96 \frac{0,35}{\sqrt{50}}\right) = 1 - 0,05 = 0,95$$

$$P(1,48 - 0,10 \leq \mu \leq 1,48 + 0,10) = 1 - 0,05 = 0,95$$

Điều này có nghĩa với độ tin cậy 95%, năng suất cá thể của cà chua từ 1,38 đến 1,58 kg/cây.

Nếu như $\alpha = 0,01$ thì khoảng tin cậy được xác định là:

$$P\left(1,48 - 2,58 \frac{0,35}{\sqrt{50}} \leq \mu \leq 1,48 + 2,58 \frac{0,35}{\sqrt{50}}\right) = 1 - 0,01$$

$$P(1,48 - 0,13 \leq \mu \leq 1,48 + 0,13) = 1 - 0,01$$

$$P(1,35 \leq \mu \leq 1,61) = 1 - 0,01$$

Năng suất từ 1,35 đến 1,61 kg/cây

5.3.2. Ước lượng số trung bình quần thể khi dung lượng mẫu $n < 30$

Lúc này không thể coi phương sai chưa biết σ^2 là s^2 được do đó phải dùng đến phân phối t (Student). Khoảng tin cậy của trị số trung bình có dạng như sau:

$$P(\bar{x} - t_{(\alpha, df)} s_x \leq \mu \leq \bar{x} + t_{(\alpha, df)} s_x) = 1 - \alpha \quad (5.6)$$

Ở đây giá trị $t_{(\alpha,df)}$ với $df = n - 1$ tra ở bảng phân phối t (phụ lục)

Thí dụ: Theo dõi năng suất của bắp cải trong thí nghiệm vụ đông tại Đông Anh Hà Nội, dung lượng mẫu điều tra $n = 25$, năng suất bình quân $\bar{x} = 175,5$ tạ/ha với độ lệch chuẩn $s = 20,5$ tạ/ha. Hãy đưa ra khoảng tin cậy 95 % cho năng suất bắp cải vụ đông tại điểm nghiên cứu ở Đông Anh Hà Nội.

Trước hết tra bảng t ở mức $\alpha = 0,05$ với số bậc tự do

$df = n - 1$ và $df = 25 - 1 = 24$. Như vậy, giá trị $t_{(0,05,df=24)} = 2,06$

Khoảng sẽ được xác định như sau

$$P\left(175,5 - 2,06 \frac{20,5}{\sqrt{25}} \leq \mu \leq 175,5 + 2,06 \frac{20,5}{\sqrt{25}}\right) = 1 - 0,05 = 0,95$$

$$P(167,1 \leq \mu \leq 183,9) = 1 - 0,05$$

Hay viết gọn lại $P(\mu = 175,5 \pm 8,4)$ tạ/ha với mức ý nghĩa $\alpha = 0,05$

5.4. Xác định dung lượng mẫu

Như đã biết khoảng tin cậy với trung bình của quần thể phụ thuộc vào độ tin cậy và dung lượng mẫu. Khi dung lượng mẫu lớn khoảng tin cậy trung bình có dạng

$$\mu = \bar{x} \pm \Delta \quad (5.7)$$

Như vậy Δ là sai số ước lượng và chúng ta muốn $\Delta \leq \varepsilon$ với ε càng nhỏ càng tốt để khoảng tin cậy hẹp.

$$\Delta \leq \frac{(U_{\alpha}) \times s}{\sqrt{n}} \quad \text{hoặc} \quad \Delta \leq \frac{t_{\alpha,df} \times s}{\sqrt{n}}$$

$$\text{Như vậy } n \geq \frac{(U_{\alpha})^2 s^2}{\Delta^2} \quad \text{hoặc} \quad n \geq \frac{t_{(\alpha,df)}^2 \times s^2}{\Delta^2} \quad (\text{khi dung lượng mẫu nhỏ})$$

Khi $\alpha = 0,05$ thì $U_{(0,05)} = 1,96$ và có thể lấy ≈ 2

Còn giá trị $t_{(0,05,df)} \geq 1,96$ phụ thuộc vào độ tự do có thể tra trong bảng 4 phần

$$\text{phụ lục. Vậy } n_{ct} \geq \frac{4 \times s^2}{\Delta^2} \quad (5.8)$$

Ở đây Δ có giá trị chứa đơn vị đo như các quan sát x_i hoặc \bar{x} . Ta còn có thể tính được độ lớn n cần thiết khi trước một sai số ước lượng $\Delta\%$ qua công thức sau

$$n_{ct} \geq \frac{4 \times s^2}{(\bar{x})^2 \times (\Delta\%)^2} \times 10000 = \frac{40.000 \times s^2}{(\bar{x})^2 \times (\Delta\%)^2} \quad (5.9)$$

Thí dụ: Quan sát 10 cành cà phê chè Catimor trồng 2 năm. Đếm số quả trên cành có trung bình $\bar{x} = 121$ quả/cành. Độ lệch chuẩn $s = 25$ quả/cành. Để số quả bình quân của vườn

cà phê mong muốn $\mu = (121 \pm 10)$ quả/cành ($\Delta \leq 10$ quả/cành) thì dung lượng $n = 10$ như đã lấy thử đủ đảm bảo sai số đưa ra hay chưa?

n cần thiết cho $\Delta = 10$ tính như sau

$$\text{Do } \Delta \text{ là giá trị số lượng nên } n_{ct} \geq \frac{4 \times 25^2}{10^2} = \frac{625 \times 4}{100} = 25 \text{ cành}$$

Vậy để cho sai số của số quả/cành là 10 quả/cành thì dung lượng $n = 10$ như đã lấy thử là còn chưa đủ lớn mà phải lấy thêm ít nhất là 15 cành nữa để tổng số cành quan sát $n \geq 25$.

Nếu lại đưa ra $\Delta\%$ mong muốn là 5% thì

$$n_{ct} \geq \frac{40.000 \times (25)^2}{(121)^2 \times 5^2} = 68,3 \text{ hay } \approx 68 \text{ cành hoặc } 69 \text{ cành}$$

Như vậy, $n = 10$ còn quá nhỏ so với mong muốn để sai số ước lượng $\Delta = 5\%$. Phải lấy thêm rất nhiều cành nữa mới đủ chấp nhận sai số ước lượng nêu trên.

5.5. Ước lượng xác suất của tổng thể (hay ước lượng tỷ lệ)

Trong thực nghiệm sinh học, rất nhiều trường hợp phải nghiên cứu các xác suất hay tỷ lệ, như tỷ lệ sống của cây con sau khi đem từ vườn ươm trồng ra lô sản xuất, tỷ lệ bệnh, sâu hoặc tỷ lệ mọc mầm của hạt...

Thí dụ: Trong một quần thể có N cá thể (N rất lớn) và giả sử có M cá thể có đặc tính A. Như vậy, xác suất của A là $p = M/N$ (đây là theo lý thuyết). Song ta không thể có điều kiện để tính p trực tiếp. Vì vậy, phải lấy một mẫu ngẫu nhiên từ quần thể ấy. Trong n phần tử của mẫu đếm được m phần tử có đặc tính A. Vậy tần suất của đặc tính A trong mẫu sẽ là $f = m/n$.

Để ước lượng xác suất p của các cá thể có đặc tính A cần phải xem xét các điều kiện cụ thể sau:

5.5.1. Khi sự kiện A có xác suất không gần 0 và 1

5.5.1.1. Khi dung lượng n đủ lớn ($n > 100$)

Lúc này luật phân phối nhị thức của xác suất của A sẽ tiệm cận với luật phân phối chuẩn và như vậy $s_p^- = \sqrt{\frac{f(1-f)}{n}}$ và biểu thức ước lượng điểm có thể viết như sau:

$$p = f \pm s_p^- \quad (5.10)$$

$$\text{Hoặc } p \approx f \quad (5.11)$$

Khoảng tin cậy của sự kiện A có xác suất p trong quần thể sẽ có dạng sau

$$P(f - u_\alpha s_p^- \leq p \leq f + u_\alpha s_p^-) = 1 - \alpha \quad (5.12)$$

Hoặc viết gọn như sau :

$$P(p = f \pm u_\alpha s_p^-) = 1 - \alpha \quad (5.13)$$

Cụ thể:

$$p = f \pm 1,96s_p^- \text{ là khoảng tin cậy } 95\%$$

$$p = f \pm 2,58s_p \text{ là khoảng tin cậy } 99\%$$

$$p = f \pm 3,29s_p \text{ là khoảng tin cậy } 99,9\%$$

Thí dụ: Để dự đoán tỷ lệ sâu đục quả cà chua vụ xuân hè 2002 tại Gia Lâm, Hà Nội, tiến hành lấy ngẫu nhiên một mẫu $n = 630$ quả, trong đó có 82 quả bị sâu đục. Hãy đưa ra các ước lượng cho tỷ lệ sâu đục quả cà chua trong nghiên cứu trên.

Do độ lớn $n = 630$ là lớn nên:

* *Ước lượng điểm.*

Gọi p là xác suất bị sâu đục quả của quần thể, f là tần suất của mẫu có quả bị sâu

$$f = \frac{82}{630} = 0,130$$

$$p = 0,130 \text{ hay } 13,0 \%$$

$$\text{Hoặc } p = 0,130 \pm \sqrt{\frac{0,130(1-0,130)}{630}}$$

$$\text{Hoặc } p = 0,130 \pm \sqrt{\frac{0,130(1-0,130)}{630}}$$

$$= 0,130 \pm 0,0134 \text{ hay } p = (13,0 \pm 1,34)\%$$

* *Ước lượng khoảng:*

Nếu chọn mức ý nghĩa $\alpha = 0.05$ thì tỷ lệ sâu đục quả cà chua nghiên cứu sẽ được xác định như sau:

$$p = f \pm 1,96s_p \Rightarrow 0,130 \pm (1,96 \times 0,0134)$$

$$= 0,130 \pm 0,0262 \text{ hay } p = (13,0 \pm 2,62)\%$$

Với độ tin cậy 95% thì tỷ lệ cà chua bị sâu đục quả vụ xuân hè 2002 tại Gia Lâm, Hà Nội nằm trong khoảng từ 10,38 % đến 15,62%.

- Nếu chọn $\alpha = 0,01$ thì khoảng tin cậy lúc này là:

$$P = 0,130 \pm 2,58 \cdot S_p \Rightarrow 0,130 \pm 0,0346 \text{ hay từ } 9,54 \% \text{ đến } 16,46 \%$$

- Nếu $\alpha \leq 0,001$ thì khoảng sẽ thay đổi từ 8,59 % đến 17,4 %

5.5.1.2. Khi dung lượng $n < 100$ (không đủ lớn)

Do mẫu nhỏ nên không thể áp dụng hàm tiệm cận để ước lượng được mà phải dùng phân phối nhị thức. Nhưng việc tính toán sẽ phức tạp nên các nhà toán học thống kê xác suất đã lập bảng tính sẵn cho độ lớn n từ 4 đến 100 (chỉ áp dụng cho khoảng 95% độ tin cậy). Khoảng này sẽ được tìm ở các bảng 6 (a, b, c) phần phụ lục.

Bảng 6(a) áp dụng cho khoảng 95% của tỷ lệ mẫu bé ($x = m$)

$$\text{Với } 4 \leq n \leq 10$$

Bảng 6 (b) với tỷ lệ của mẫu khi $10 \leq n \leq 100$ và $0 \leq m \leq 25$

Bảng 6 (c) với tỷ lệ khi $60 \leq n \leq 100$ và $26 \leq m \leq 50$

Thí dụ: Áp dụng một biện pháp điều trị bằng thuốc kháng sinh cho bệnh vàng đầu của tôm sú bố mẹ.

Tiến hành xử lý ở $n = 20$ tôm bố mẹ; sau xử lý quan sát thấy có 5 tôm khỏi bệnh và 15 tôm không khỏi bệnh. Vậy khoảng tin cậy 95% của khỏi bệnh là bao nhiêu?

* Nếu lấy ước lượng điểm thì ở đây xác suất (tỷ lệ) khỏi bệnh vàng đầu của tôm sú bố mẹ sẽ là

$$p = f = \frac{m}{n} = \frac{5}{20} = 0,250 \text{ hay } 25,0\%$$

* Nếu tìm khoảng tin cậy 95% cho tỷ lệ khỏi bệnh sẽ dùng trong bảng 6 (b) tra tại cột 5 hàng 20

Hàng trên là $p_1\% = 8,7\%$

Hàng dưới là $p_2\% = 49,1\%$

Như vậy $P(p_1 \leq p \leq p_2) = 1 - 0,05$ sẽ từ 8,7% đến 49,1% là tỷ lệ khỏi bệnh vàng đầu của tôm sú bố mẹ.

5.5.2. Khi sự kiện A có xác suất gần 0 hoặc gần 1

Trong trường hợp này xác suất của A tuân theo luật Poisson (hay còn gọi là hàm phân phối xác suất của sự kiện hiếm). Dựa theo luật Poisson người ta đã lập một bảng tính sẵn để có ước lượng khoảng cho sự kiện A này. Tuy nhiên, chỉ ứng với độ tin cậy 95 % (bảng 7 phụ lục).

Còn với ước lượng điểm thì cũng chỉ lấy gần đúng tốt nhất cho xác suất của tổng thể là xác suất của A trong mẫu quan sát.

Thí dụ: Nghiên cứu ảnh hưởng của chiếu xạ lên hạt giống đến hiện tượng dị hình của cây sau xử lý. Mẫu xử lý có độ lớn $n = 12500$ hạt táo, sau đó đem gieo và theo dõi cây con. Gọi A là hiện tượng dị hình, quan sát thấy có Axitamin = 105 cây. Hãy đưa ra các dạng ước lượng cho kết quả xử lý trên về hiện tượng đột biến kiểu hình.

Gọi p là xác suất hay tỷ lệ đột biến kiểu hình của liều lượng xử lý trên, kết quả thống kê mẫu có tần suất:

$$f = \frac{105}{12500} = 0,0084 \text{ hay } 0,84\%$$

* Vậy ước lượng điểm của hiện tượng đột biến kiểu hình của liều lượng xử lý trên $p \approx f\%$ là 0,84 %

* Ước lượng khoảng được xác định sẵn qua bảng 7 phụ lục. Song, bảng chỉ cho hai giá trị np_1 và np_2 ứng với 95% độ tin cậy.

$P(p_1, p_2) = 1 - 0,05$ với p_1, p_2 tính từ

$$p_1 = \frac{np_1}{n} \text{ và } p_2 = \frac{np_2}{n}$$

Trong trường hợp ở đây np_1 và np_2 phải được tra từ giá trị gần đúng sau

Trong bảng (7) chỉ có tới $x = m$ nhiều nhất là 100. Từ giá trị $m = 105$ không có trong bảng. Nên phải giảm (lùi 10 lần); $m = 10,5$ lấy gần đúng $m = 11$.

Tra ở $m = 11$ (hàng 10 cột 1) có $np_1 = 5,5$; $np_2 = 19,7$

Muốn có p_1 và p_2 thì $p_1 = \frac{np_1}{n}$. Nhưng vì các giá trị np_1 và np_2 đều được tính lùi 10 lần nên lúc này n chỉ còn $n = 1250$. Từ đó

$$np_1 = 5,5 \text{ và } np_2 = 19,7$$

$$p_1 = \frac{5,5}{1250} = 0,0044 \text{ hay } 0,44 \%$$

$$p_2 = \frac{19,7}{1250} = 0,01576 \text{ hay } 1,576 \% \text{ lấy gần đúng } 1,58 \%$$

Vậy tỷ lệ đột biến kiểu hình của liều lượng xử lý này sẽ dao động từ 0,44 % đến 1,58 % với độ tin cậy 95%.

Bài tập:

5.1. Điều tra năng suất ngô trên địa bàn xã Phú Linh thị xã Hà Giang (tạ/ha) của 45 hộ dân tộc ta có kết quả sau:

41; 38; 35; 42; 42; 36; 40; 36; 34; 36; 35; 36; 34; 42; 39; 39; 44; 37; 44; 36; 41; 43; 42; 42; 42; 43; 39; 43; 39; 44; 40; 43; 43; 35; 38; 39; 39; 42; 43; 37; 44; 40; 39; 43; 43.

Hãy đưa ra các dạng ước lượng cho năng suất ngô của vùng điều tra nói trên (ước lượng điểm và ước lượng khoảng)

5.2. Đếm số hạt trên bông lúa của một giống ta có các số liệu sau”

120; 119; 116; 110; 121; 118; 106; 133; 123; 115; 112; 126; 109; 128; 123; 107; 132; 125; 106; 124.

Hãy đưa ra các dạng ước lượng cho năng suất ngô của vùng điều tra nói trên (ước lượng điểm và ước lượng khoảng)

5.3. Theo dõi tỷ lệ bật mầm các mắt ghép, người ta đã tiến hành theo dõi ở 200 cây ghép đã cho thấy kết có 148 cây đã bật mầm. Hãy đưa ra ước lượng điểm và ước lượng khoảng của hiện tượng bật mầm của mắt ghép nêu trên

Ghi chú: Cán bộ giảng dạy cho thêm một số bài tập cụ thể khác cho sinh viên.

CHƯƠNG VI KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

Đây là bài toán thống kê rất hay gặp trong thực nghiệm. Ở đây đề cập đến việc kiểm định (so sánh) hai số trung bình của một đặc tính định lượng hay hai xác suất (hai tỷ lệ) của một đặc tính định tính của quần thể.

6.1. Những khái niệm chung và ý nghĩa

Trong nghiên cứu thường phải so sánh các tham số thống kê như số trung bình, phương sai, xác suất của một mẫu với một tiêu chuẩn cho trước nào đó, hoặc 2 mẫu với nhau hay nhiều mẫu với nhau. Thông thường các tham số có sự khác nhau (khác nhau về số học), nhưng ta lại cần xem xét sự sai khác này có rõ ràng hay không? ở mức độ nào?

Nếu chúng chỉ khác nhau trong phạm vi ngẫu nhiên thì sự khác nhau này được coi như không đáng kể (không có ý nghĩa). Nếu chúng khác nhau ngoài phạm vi ngẫu nhiên thì kết luận sự khác nhau ấy là do tác động của nhân tố thí nghiệm (ngoài tác động của nhân tố ngẫu nhiên). Đây là vấn đề mà các nhà sinh học nông nghiệp cần quan tâm.

Để kiểm định người ta dùng các kết quả thực nghiệm quan sát ở mẫu với việc vận dụng công cụ toán học là lý thuyết xác suất để kiểm tra những giả thiết đã cho. Nếu tài liệu thực nghiệm phù hợp với giả thiết thì giả thiết được chấp nhận. Ngược lại thì giả thiết bị bác bỏ. Sự phù hợp mà ta nói ở đây không phải là tuyệt đối, mà chỉ là nói phù hợp theo một tiêu chuẩn nào đó xác định trước đủ thoả mãn những yêu cầu của thực tiễn.

Thí dụ: Trong nông học người ta thường so sánh (hay kiểm định) sự sinh trưởng, phát triển, diễn biến sâu bệnh hại cây trồng cũng như các chỉ tiêu năng suất được gieo trồng bằng những biện pháp kỹ thuật khác nhau để xem chúng có ảnh hưởng thực sự đến các chỉ tiêu nghiên cứu hay không?

Trong chương này chúng tôi sẽ tập trung đề cập đến một số tiêu chuẩn thống kê được dùng để kiểm định trong các trường hợp cụ thể.

6.2. Trường hợp hai mẫu độc lập

Mẫu độc lập hay thí nghiệm độc lập là những khái niệm tương đối. Theo nghĩa rộng người ta gọi mẫu độc lập hay thí nghiệm độc lập nếu một quá trình thực nghiệm nào đó được thiết kế một cách độc lập với những thí nghiệm khác.

6.2.1. Tiêu chuẩn U của phân phối tiêu chuẩn

Nếu trong trường hợp kiểu phân phối lý thuyết đặc trưng cho 2 kết quả (2 mẫu) nghiên cứu chưa biết thì yêu cầu dung lượng mẫu lấy phải được coi là đủ lớn ($n_1 > 30$ và $n_2 > 30$). Theo luật số lớn thì trong trường hợp mẫu lớn, phân phối xác suất của số trung bình mẫu \bar{X}

xấp xỉ luật chuẩn với kỳ vọng $M\bar{X} = \mu$ và phương sai $D\bar{X} = \frac{s^2}{n}$

$$\text{Nhu vậy } \bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$$

$$\bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

Giả thiết H_0 : $\mu_1 = \mu_2$ hay $\mu_1 - \mu_2 = 0$
 Đối thiết H_1 : $\mu_1 \neq \mu_2$ hay $\mu_1 - \mu_2 \neq 0$ (6.1)

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right) \quad (6.2)$$

Được kiểm định bằng tiêu chuẩn u của phân phối tiêu chuẩn với mức ý nghĩa α , tính giá trị thực nghiệm như sau:

$$u_m = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (6.3)$$

Nếu phương sai của 2 tổng thể không được biết trước và dung lượng mẫu đủ lớn thì có thể thay một cách gần đúng phương sai tổng thể bằng phương sai mẫu. Có nghĩa là

$$\sigma_1^2 \approx s_1^2 \quad \text{và} \quad \sigma_2^2 \approx s_2^2$$

Lúc này tiêu chuẩn phù hợp như sau:

$$u_{t.n} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (6.4)$$

Nếu như $|u_{t.n}| < u_\alpha$ tra ở bảng Φ (bảng 1 phụ lục) với mức ý nghĩa α thì giả thiết H_0 được chấp nhận nghĩa là hai trung bình của hai mẫu bằng nhau. Ngược lại nếu $|u_{t.n}| \geq u_\alpha$ thì giả thiết bị bác bỏ nghĩa là hai trung bình của hai mẫu là khác nhau.

Thí dụ

Đo chiều cao cây cuối cùng của 2 giống lúa mới có kết quả như sau:

Giống I: Đo $n = 42$ khóm có chiều cao trung bình $\bar{x}_1 = 95,2cm$

Độ lệch chuẩn về chiều cao là $s_1 = 3,2$ cm.

Giống II: Đo $n = 40$ khóm có chiều cao trung bình $\bar{x}_1 = 98,5cm$

Độ lệch chuẩn tương ứng $s_2 = 3,4$ cm.

Hỏi chiều cao cây của hai giống có khác nhau hay không với mức ý nghĩa $\alpha = 0,05$

Vì hai dung lượng mẫu n_1 và n_2 lấy từ hai giống lúa nghiên cứu là đủ lớn, và không biết trước được hai phương sai tổng thể. Nên ta có thể dùng tiêu chuẩn u của phân phối chuẩn để kiểm định.

Giả thiết H_0 : $\mu_1 = \mu_2$ hay $\mu_1 - \mu_2 = 0$

Đối thiết H_1 : $\mu_1 \neq \mu_2$ hay $\mu_1 - \mu_2 \neq 0$

Để kiểm định giả thiết H_0 ta áp dụng biểu thức (6.4).

$$u_{t,n} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{|95,2 - 98,5|}{\sqrt{\frac{3,2^2}{42} + \frac{3,4^2}{40}}} = |4,52|$$

Với $\alpha = 0,05$ $u_\alpha = 1,96$

Ở đây $|u_m| > u_{0,05}$ ($|4,52| > 1,96$)

Nên ta bác bỏ giả thiết H_0 và chấp nhận đối thiết H_1 là giống khác nhau thì chiều cao khác nhau rõ rệt.

6.2.2. Tiêu chuẩn t của phân phối Student

Tiêu chuẩn này được áp dụng trong trường hợp luật phân phối của hai tổng thể mà đại diện là 2 mẫu có phân phối chuẩn và phương sai của hai tổng thể được coi là bằng nhau. Nếu thoả mãn hai điều kiện này thì có thể kiểm tra.

Giả thiết H_0 : $\mu_1 = \mu_2$ hay $\mu_1 - \mu_2 = 0$

Đối thiết H_1 : $\mu_1 \neq \mu_2$ hay $\mu_1 - \mu_2 \neq 0$ bằng tiêu chuẩn T của Student như sau:

$$T_{t,n} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (6.5)$$

Trong đó: \bar{x}_1 và \bar{x}_2 là trung bình của 2 mẫu là mẫu 1 và mẫu 2.

s_1^2 và s_2^2 là phương sai của mẫu 1 và mẫu 2.

n_1 và n_2 là dung lượng quan sát của 2 mẫu.

Thường trường hợp này được áp dụng khi n_1 và n_2 là không đủ lớn $n_1 < 30$ và $n_2 < 30$ hoặc $n_1 > 30$ và $n_2 < 30$ hoặc $n_2 > 30$ và $n_1 < 30$

Nếu $|T_{t,n}| < t_{(\alpha, n_1 + n_2 - 2)}$ tra bảng t với $n_1 + n_2 - 2$ bậc tự do (phụ lục bảng 4) thì giả thiết H_0 được chấp nhận, nghĩa là trung bình của hai mẫu bằng nhau. Ngược lại nếu như $|T_{t,n}| > t_{(\alpha, n_1 + n_2 - 2)}$ tra bảng với bậc tự do = $n_1 + n_2 - 2$ thì giả thiết H_0 bị bác bỏ. Nghĩa là chấp nhận đối thuyết H_1 trung bình hai mẫu là khác nhau.

Thí dụ

So sánh năng suất của hai giống cà chua vụ xuân hè

Giống số 6: theo dõi 16 địa điểm năng suất trung bình hạt 30,6 tấn/ha.

Độ lệch chuẩn về năng suất 4,5 tấn/ha.

Giống số 204 A theo dõi ở 19 địa điểm, năng suất bình quân đạt 27,0 tấn/ha. Độ lệch chuẩn năng suất là 4,0 tấn/ha.

Biết rằng phân phối về năng suất của cà chua là phân phối chuẩn và hai phương sai lý thuyết được coi là bằng nhau.

Hãy cho biết là trung bình của hai giống trên có khác nhau hay không ở mức ý nghĩa $\alpha = 0,05$ và 0,01.

Vì hai tổng thể đã đáp ứng các điều kiện nêu ra nên có thể áp dụng tiêu chuẩn t của phân phối Student để kiểm định giả thiết

Giả thiết H_0 : $\mu_1 = \mu_2$ hay $\mu_1 - \mu_2 = 0$

Đối thiết H_1 : $\mu_1 \neq \mu_2$ hay $\mu_1 - \mu_2 \neq 0$

Theo (6.5)

$$T_{t,n} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Thay vào được

$$T_{t,n} = \frac{|30,6 - 27,0|}{\sqrt{\frac{(16 - 1) \cdot 4,5^2 + (19 - 1) \cdot 4,0^2}{16 + 19 - 2} \left(\frac{1}{16} + \frac{1}{19} \right)}} = 2,064$$

$t_{(0,05;33)} = 2,04$, như vậy ($|T_{t,n}| > t_{0,05;33}$) nên giả thiết H_0 bị bác bỏ và chấp nhận H_1 với câu trả lời "hai giống cà chua nói trên có năng suất trung bình khác nhau".

Với $\alpha = 0,01$ thì $t_{(0,01;33)} = 2,75$

Vì $2,064 < 2,75$ nên chấp nhận giả thiết H_0 có nghĩa là "hai giống cà chua nói trên có năng suất trung bình bằng nhau".

Trong một số trường hợp khi áp dụng tiêu chuẩn t của Student để so sánh 2 mẫu độc lập, điều kiện về luật phân phối chuẩn của tổng thể được thoả mãn, nhưng 2 phương sai của tổng thể không được biết trước.

Khi đó phải kiểm định bằng tiêu chuẩn F của Fisher về sự bằng nhau của hai phương sai (sẽ nêu ở mục 6.5).

Nếu hai phương sai bằng nhau thì áp dụng kiểm định như biểu thức (6.5).

Nếu hai phương sai được coi là khác nhau thì biểu thức áp dụng cho tiêu chuẩn t của Student như sau:

$$T_{t,n} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{6.6}$$

Nhưng $|T_{t,n}|$ sẽ được so sánh với giá trị t lý thuyết ở mức ý nghĩa α tính như sau:

$$t_{\alpha}^* = \frac{\left[t_{(\alpha, df_1 = n_1 - 1)} \frac{s_1^2}{n_1} \right] + \left[t_{(\alpha, df_2 = n_2 - 1)} \frac{s_2^2}{n_2} \right]}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \tag{6.7}$$

Nếu như $|T_{t,n} < t^*|$ thì chấp nhận H_0 (hai trung bình là bằng nhau)

Nếu như $|T_{t,n} > t^*|$ bác bỏ giả thiết H_0 thì chấp nhận H_1 (hai trung bình là khác nhau).

Thí dụ

Phân tích hàm lượng đường tổng số (%) của hai giống cà chua vụ xuân hè.

- Giống MV1 phân tích ở 6 mẫu có hàm lượng đường tổng số đạt 3,09% và phương sai về đường tổng số là 0,75 %.

- Giống mới phân tích ở $n_2 = 8$ mẫu có giá trị trung bình của đường tổng số đạt 2,79%. Phương sai có giá trị 0,10 %.

Hãy cho hàm lượng đường tổng số của hai giống nêu trên có khác nhau hay không ?

Cho hai phương sai là không bằng nhau với mức ý nghĩa $\alpha = 0,05$.

Do hai phương sai không bằng nhau nên áp dụng biểu thức (6.6)

$$T_{t,n} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \text{ thay số vào ta có } T_{t,n} = \frac{|3,09 - 2,79|}{\sqrt{\frac{0,75}{6} + \frac{0,16}{8}}} = |0,798|$$

Áp dụng biểu thức (6.7) tính $t_{0,05}^*$ như sau:

$$t_{\alpha}^* = \frac{\left[t_{(\alpha, df_1=n_1-1)} \frac{s_1^2}{n_1} \right] + \left[t_{(\alpha, df_2=n_2-1)} \frac{s_2^2}{n_2} \right]}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Thay số vào

$$t_{0,05}^* = \frac{\left(2,75 \times \frac{0,75}{6} \right) + \left(2,36 \times \frac{0,10}{8} \right)}{\frac{0,75}{6} + \frac{0,10}{8}} = 2,54$$

Như vậy, $|T_{t,n} < t^*|$ ($0,798 < 2,54$) nên chấp nhận giả thiết H_0 "hai giống cà chua nói trên có hàm lượng đường tổng số như nhau"

6.3. Trường hợp hai mẫu theo cặp

6.3.1. Khái niệm về mẫu theo cặp

Thực tiễn của nghiên cứu nông học nói riêng và nghiên cứu trong nông nghiệp cũng như sinh học nói chung, ngoài những mẫu độc lập như đã nêu ở trên, nhà khoa học khi nghiên cứu còn gặp một loạt các mẫu quan sát không độc lập mà lại có xếp theo cặp (đôi)

Thí dụ: khi đo chiều cao cây ta dùng hai loại thước đo để đo cho cùng một mẫu. Hoặc xác định khối lượng 1000 hạt của mẫu bằng hai chiếc cân khác nhau. Hay hai kỹ thuật viên A và B đếm khuẩn lạc của n hộp petri. Như vậy, các mẫu trên đều được thực hiện 2 lần quan sát, kết quả được xác định theo cặp.

Những kết quả như vậy sẽ được kiểm định theo phương pháp cặp. Do vậy, người làm công tác thống kê sẽ vận dụng các tiêu chuẩn khác cho phù hợp. Dưới đây chúng tôi trình bày tiêu chuẩn thông thường nhất qua thí dụ sau:

6.3.2. Tiêu chuẩn t của Student

Giả thiết H_0 : $\mu_x = \mu_y$ hay $\mu_x - \mu_y = 0$

Đối thiết H_1 : $\mu_x \neq \mu_y$ hay $\mu_x - \mu_y \neq 0$

Tính các hiệu sai $d_i = y_i - x_i$ (hiệu của hai mẫu) sau đó tính độ lệch chuẩn của hiệu sai theo biểu thức sau:

$$|T_{t,n}| = \frac{\bar{d}}{s_{\bar{d}}} = \frac{|\bar{d}|}{\sqrt{\frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{n(n-1)}}} \quad (6.8)$$

$$\text{với } \bar{d} = \frac{\sum d_i}{n} \quad (6.9)$$

$$s_d^2 = \frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{(n-1)} \quad (6.10)$$

Sai số chuẩn $s_{\bar{d}} = \frac{s_d}{\sqrt{n}}$ (Độ lệch chuẩn hiệu sai bình quân)

$|T_{t,n}| < t_{(\alpha,n-1)}$ chấp nhận H_0

$|T_{t,n}| > t_{(\alpha,n-1)}$ chấp nhận H_1

Thí dụ

Đo đường vanh của 18 cây cao su bằng hai loại thước khác nhau kết quả được ghi lại trong bảng 6.1 sau. Có sự sai khác rõ rệt giữa 2 loại thước đo hay không ?

Bảng 6.1. So sánh hai mẫu theo cặp bằng tiêu chuẩn t

TT	x_i của thước 1	y_i của thước 2	$d_i = x_i - y_i$	d_i^2
1	29,7	29,0	0,7	0,49
2	34,9	34,5	0,4	0,16
3	37,0	37,5	- 0,5	0,25
4	28,5	28,0	0,5	0,25
5	26,9	26,5	0,4	0,16
6	27,0	27,1	- 0,1	0,01
7	28,7	28,5	0,2	0,04
8	31,2	31,0	0,2	0,04
9	31,5	31,0	0,5	0,25
10	33,7	33,5	0,2	0,04

11	34,0	33,7	0,3	0,09
12	35,9	36,0	-0,1	0,01
13	41,2	41,5	-0,3	0,09
14	29,5	29,0	0,5	0,25
15	30,7	30,5	0,2	0,04
16	33,5	33,7	-0,2	0,04
17	36,2	36,0	0,2	0,04
18	37,5	37,5	0	0
Tổng			$\sum d_i = 3,1$	$\sum d_i^2 = 2,25$

$$s_{\bar{d}} = \sqrt{\frac{\sum di^2 - (\sum di)^2}{n(n-1)}} = \sqrt{\frac{2,25 - \frac{(3,1)^2}{18}}{18(18-1)}} = 0,0749$$

Thay số vào biểu thức (6.8)

$$|T_{t,n}| = \frac{0,01722}{0,0749} = 2,30$$

Tra bảng t của Student với $\alpha = 0,05$ bậc tự do $df = 17$ ta được giá trị $t_{(0,05;17)} = 2,11$

Vì $|T_{t,n}| > t_{(0,05;17)}$ nên hai thước đo khác nhau thì kết quả trung bình khác nhau rõ rệt.

6.4. Kiểm định tính độc lập (kiểm định tính thuần nhất của các mẫu về chất)

Đây là trường hợp áp dụng cho số liệu nghiên cứu mang tính định tính, hay còn gọi là kiểm định phi tham số.

6.4.1. So sánh 2 tỷ lệ (hay gọi là kiểm định 2 xác suất)

Giả sử có 2 tổng thể (2 công thức), công thức I có N_1 phân tử, trong đó có M_1 phân tử có đặc tính A và $N_1 - M_1$ phân tử không có đặc tính A.

Công thức II có N_2 phân tử, trong đó có M_2 phân tử có đặc tính A và tương tự $N_2 - M_2$ phân tử không có đặc tính A.

Từ 2 công thức rút mẫu ngẫu nhiên với dung lượng n_1 và n_2 . Như vậy, tần suất của A ở các mẫu lần lượt là

$$f_1 = \frac{m_1}{n_1} \quad \text{và} \quad f_2 = \frac{m_2}{n_2}$$

Dựa vào 2 kết quả này để kiểm định giả thiết H_0 với nội dung là

$$H_0: p_1 = p_2 \quad (p_1 - p_2 = 0) \tag{6.11}$$

Ở đây p_1 là xác suất của các cá thể A trong công thức I

p_2 là xác suất của các cá thể A trong công thức II

Để có thể chấp nhận hay bác bỏ giả thiết H_0 ta phải xem xét cụ thể qua các trường hợp sau:

6.4.1.1. Trường hợp 2 dung lượng mẫu đều lớn

$n_1 > 100$ và $n_2 > 100$ và P_0 xa 0 và 1

Trường hợp này áp dụng tiêu chuẩn U của hàm phân phối chuẩn như sau:

$$|u_{t,n}| = \frac{|f_1 - f_2|}{\sqrt{\frac{f_1(1-f_1)}{n_1} + \frac{f_2(1-f_2)}{n_2}}} \tag{6.12}$$

$|u_{t,n}| < u_\alpha$ chấp nhận giả thiết có nghĩa là coi hai xác suất p_1 và p_2 như nhau.

$|u_{t,n}| \geq u_\alpha$ bác bỏ giả thiết H_0 nghĩa là coi $p_1 \neq p_2$

Thí dụ

Nghiên cứu hiện tượng sâu đục quả với 2 giống cà chua được trồng vụ xuân hè tại đồng bằng sông Hồng. Giống cà chua số 6 trong $n_1 = 300$ quả có 36 quả bị sâu đục.

Giống số 48 trong $n_2 = 240$ quả có 38 quả bị sâu đục.

Hỏi: tỷ lệ sâu đục quả của 2 giống cà chua có khác nhau hay không? (cho $\alpha = 0,05$)

Tỷ lệ bị sâu ở giống số 6 là $f_1 = \frac{36}{300} = 0,120$ hay 12,0%

Tỷ lệ bị sâu ở giống số 48 là $f_2 = \frac{38}{240} = 0,158$ hay 15,8%

Thay số vào biểu thức (6.12)

$$|u_{t,n}| = \frac{f_1 - f_2}{\sqrt{\frac{f_1(1-f_1)}{n_1} + \frac{f_2(1-f_2)}{n_2}}} = \frac{0,120 - 0,158}{\sqrt{\frac{0,12 \times 0,88}{300} + \frac{0,158}{240}}} = 1,27$$

$u_{0,05} = 1,96$

Như vậy $|u_{t,n}| < 1,96$ chấp nhận giả thiết H_0 có nghĩa là ở mức ý nghĩa 0,05 "tỷ lệ sâu đục quả của 2 giống cà chua như nhau".

6.4.1.2. Trường hợp n_1 hoặc $n_2 < 100$ hoặc cả n_1 và n_2 nhỏ hơn 100)

Trường hợp này ta phải sử dụng hàm phân phối χ^2 (khi bình phương) để kiểm định. Có thể tính được $\chi_{t,n}^2$ bằng một biểu thức sau:

$$\chi_{t,n}^2 = \sum_{i=1}^{2 \times 2} \frac{|(m_{ij} - M_{ij}) - 0,5|^2}{M_{ij}} \tag{6.13}$$

Ở đây m_{ij} là tần số thực nghiệm của đặc tính j trong công thức thứ i
 M_{ij} là tần số lý thuyết của đặc tính j ở công thức thứ i
 0,5 là hệ số hiệu chỉnh YATES.

Nếu $\chi_{t,n}^2 < \chi_{(\alpha,1)}$ giá trị χ^2 ở mức nghĩa α và bậc tự do $df = 1$ thì chấp nhận giả thiết H_0 $p_1 = p_2$ (2 tỷ lệ đem kiểm định là như nhau).

Ngược lại nếu $\chi_{t,n}^2 \geq \chi_{(\alpha,1)}$ bác bỏ giả thiết H_0 và coi hai tỷ lệ đem kiểm định là khác nhau.

Thí dụ

Khảo sát hai phương pháp trừ bệnh sương mai cho cà chua có kết quả như sau:

Phương pháp I: Phun cho $n_1 = 70$ cây có 52 cây không có bệnh.

Phương pháp II: Phun cho $n_2 = 50$ cây có 30 cây không có bệnh.

Vậy phương pháp trừ sương mai cho cà chua khác nhau thì hiệu lực có khác nhau hay không?

Các tần số lý thuyết tính như sau

Nếu giả thiết $H_0: p_1 = p_2$ là đúng trong tổng số $70 + 50 = 120$ cây có $52 + 30 = 82$ cây không bị bệnh, vậy có thể coi số cây không bị bệnh của phương pháp I là

$$\frac{82}{120} \times 70 = 47,83 \approx 48 \text{ cây}$$

Số cây vẫn bị sương mai của phương pháp I là $70 - 48 \text{ cây} = 22 \text{ cây}$.

Phương pháp II có số cây không bệnh là

$$\frac{82}{120} \times 50 = 34,16 \approx 34 \text{ cây}$$

Số cây vẫn bị sương mai của phương pháp II là $50 - 34 \text{ cây} = 16 \text{ cây}$.

Như vậy, thay vào công thức (6.13) để tính $\chi_{t,n}^2$ như sau:

$$\chi_{t,n}^2 = \frac{|(52-48)-0,5|^2}{48} + \frac{|(18-22)-0,5|^2}{22} + \frac{|(30-34)-0,5|^2}{34} + \frac{|(20-16)-0,5|^2}{16} = 1,94$$

Tra bảng χ^2 với $\alpha = 0,05$ và $df = 1$ được $\chi_{(0,05;1)}^2 = 3,84$ (Bảng 3 phụ lục)

Như vậy, $\chi_{y,n}^2 < \chi_{(0,05;1)}$ nên chấp nhận giả thiết H_0 là hai phương pháp trừ bệnh sương mai khác nhau thì kết quả như nhau ở mức $\alpha = 0,05$.

6.4.1.3. Trường hợp hai sự kiện hiếm (hay xác suất bé)

Trường hợp này biến đổi để đưa về biến có phân phối chuẩn

Công thức biến đổi như sau:

$$\varphi = 2 \frac{\pi}{180} \arcsin \sqrt{f}$$

Ở đây $\pi = 3,1416$; f là tần suất của mẫu

Có bảng tính sẵn đối với $f = 0,00001$ đến $0,0099$

(Bảng 9 phụ lục)

Sau khi đã biến đổi ta áp dụng công thức sau

$$|u_m| = \frac{\varphi_1 - \varphi_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (6.14)$$

$|u_{t,n}|$ được so sánh với u_α nếu $|u_{t,n}| < u_\alpha$ chấp nhận giả thiết H_0 (hai tỷ lệ so sánh là như nhau)

Ngược lại nếu $|u_{t,n}| \geq u_\alpha$ bác bỏ H_0 (hai tỷ lệ đó là khác nhau).

Thí dụ: Có 2 liều lượng xử lý gây đột biến kiểu hình ở đậu tương bằng phương pháp vật lý. Liều lượng I xử lý 5000 hạt thấy có 6 cây dị hình. Liều lượng II xử lý 10.000 hạt thấy có 9 cây dị hình.

Hãy cho biết liều lượng xử lý khác nhau có dẫn tới phản ứng gây đột biến kiểu hình khác nhau hay không với mức ý nghĩa $\alpha = 0,05$.

$n_1 = 5000$ và như vậy xác suất của đột biến là

$$f_1 = \frac{6}{5000} = 0,0012$$

$n_2 = 10.000$ xác suất gây đột biến là

$$f_2 = \frac{9}{10.000} = 0,0009$$

Tương tự có các giá trị tra bảng (9) phần phụ lục là:

$$\varphi_1 = 0,0693 \text{ và } \varphi_2 = 0,0600$$

Áp dụng biểu thức (6.14)

$$|u_m| = \frac{\kappa_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{0,0693 - 0,0600}{\sqrt{\frac{1}{5000} + \frac{1}{10000}}} = 0,537$$

$$u_{0,05} = 1,96$$

$$|u_{t,n}| < u_{0,05}$$

Chấp nhận giả thiết H_0 "hai liều lượng xử lý khác nhau thì hiệu quả gây đột biến kiểu hình như nhau"

6.4.2. Kiểm tra tính thuần nhất của nhiều mẫu về chất (kiểm định nhiều xác suất)

Trường hợp có k công thức và các kết quả quan sát ở mẫu được chia thành r cấp (chất lượng) khác nhau thì giả thiết H_0 được đưa ra là:

H_0 : các mẫu về chất là thuần nhất và tiêu chuẩn kiểm định phù hợp là χ^2 , trong đó giá trị $\chi_{t,n}^2$ (khi bình phương thực nghiệm) tính như sau:

$$\chi_m^2 = N \left[\sum_{i,j} \frac{m_{ij}^2}{t_j \times n_i} - 1 \right] \quad (6.15)$$

Trong đó:

k: là số công thức

r : là số cấp chất lượng

m_{ij} : là tần số quan sát thực nghiệm ở công thức thứ i và đặc trưng định tính (tính chất) thứ j .

t_j : tổng tần số quan sát của các công thức đối với cấp j

n_i : dung lượng mẫu của công thức thứ i

N là tổng số các quan sát; $N = n_1 + n_2 + \dots + n_k = t_1 + t_2 + \dots + t_r$

Nếu $\chi_{i,n}^2 < \chi_{(\alpha,df)}^2$ với bậc tự do $df = (k - 1)(r - 1)$ thì chấp nhận giả thiết H_0 có nghĩa là các công thức khác nhau (các mẫu) đồng nhất về chất. Còn nếu $\chi_{i,n}^2 \geq \chi_{(\alpha,df)}^2$ thì bác bỏ giả thiết H_0 . Như vậy phải chấp nhận các công thức khác nhau thì chất lượng khác nhau (không thuần nhất).

Số liệu tổng quát được trình bày trong bảng sau (bảng 6.2)

Bảng 6.2. Sắp xếp các số liệu thực nghiệm (tần số m_{ij})

Công thức Cấp	I	II	...	i	...	k	T_i
1	$m_{1,1}$	$m_{2,1}$...	$m_{i,1}$...	$m_{k,1}$	T_1
2	$m_{1,2}$	$m_{2,2}$...	$m_{i,2}$...	$m_{k,2}$	T_2
·	·	·	·	·	·	·	·
·	·	·	·	·	·	·	·
·	·	·	·	·	·	·	·
j	$m_{1,j}$	$m_{2,j}$...	$m_{i,j}$...	$m_{k,j}$	T_j
·	·	·	·	·	·	·	·
·	·	·	·	·	·	·	·
·	·	·	·	·	·	·	·
r	$m_{1,r}$	$m_{2,r}$...	$m_{i,r}$...	$m_{k,r}$	T_r
n_i	n_1	n_2	...	n_i	...	n_k	N

Thí dụ

Có 3 mật độ trồng cà phê với I, II, III với cùng biện pháp kỹ thuật, cùng tuổi cây. Theo theo dõi mức độ bệnh gỉ sắt hại cà phê theo 3 cấp (nhẹ, trung bình và nặng). Kết quả quan sát như sau:

Bảng 6.3. Các tần số thực nghiệm

Mật độ Chất lượng	I	II	III	Tổng số
Nhẹ	80	65	60	205
Trung bình	60	75	130	265
Nặng	70	80	70	220
Tổng số	210	220	260	690

Hỏi mật độ khác nhau thì mức độ nhiễm bệnh gỉ sắt có khác nhau hay không ở mức ý nghĩa $\alpha = 0,05$.

Giả thiết H_0 : mật độ trồng cà phê khác nhau nhưng mức độ nhiễm gỉ sắt là như nhau

Để chấp nhận giả thiết H_0 hoặc bác bỏ H_0 ta áp dụng biểu thức (6.15) qua tiêu chuẩn χ^2

$$\chi_m^2 = N \left[\sum_{i,j=1}^{3 \times 3} \frac{m_{ij}^2}{t_j \times n_i} - 1 \right]$$

thay số vào ta có:

$$\chi_{t.n}^2 = 690 \left(\frac{80^2}{205 \times 210} + \frac{65^2}{205 \times 220} + \frac{60^2}{205 \times 260} + \frac{60^2}{265 \times 210} + \frac{75^2}{265 \times 220} + \frac{130^2}{265 \times 210} + \frac{70^2}{220 \times 210} + \frac{80^2}{220 \times 220} + \frac{70^2}{220 \times 260} \right) - 1 = 2,8708$$

Tra bảng 3 phần phụ lục $\chi_{0,05}^2$ với $df = (3- 1) (3-1)$ bậc tự do ($df = 4$) được giá trị $\chi_{(0,05;4)}^2 = 9,488$

$2,7807 < 9,488$ chấp nhận giả thiết H_0 "mật độ trồng cà phê khác nhau nhưng mức độ nhiễm bệnh gỉ sắt như nhau".

Trường hợp nếu có k công thức (k mẫu) nhưng số cặp $r = 2$ (số đặc trưng định tính là 2) A và \bar{A} số liệu ghi như bảng (6.4)

$$\text{Thì: } \chi_m^2 = \frac{N^2}{Ta \times Tb} \left[\sum_{i=1}^k \frac{a_i^2}{n_i} - \frac{T_a^2}{N} \right] \quad (6.16)$$

$\chi_{t.n}^2$ được so sánh với χ^2 mức ý nghĩa α và bậc tự do $df = (k - 1)$ tương ứng. Nếu $\chi_{t.n}^2 < \chi_{(\alpha,k-1)}^2$ chấp nhận giả thiết H_0 có nghĩa là hai đặc trưng định tính của các công thức là như nhau. Ngược lại $\chi_{t.n}^2 \geq \chi_{(\alpha,k-1)}^2$ bác bỏ giả thiết H_0 nghĩa là "các đặc trưng định tính của các cây trồng nghiên cứu là khác nhau" ở mức ý nghĩa.

Thí dụ: Trong nhà máy chế biến thực phẩm, người ta đã tiến hành điều tra chất lượng sản phẩm được sản xuất ra từ các dây chuyền khác nhau (1, 2, 3, 4) của cùng một doanh nghiệp chất lượng được phân làm 2 loại là thành phẩm (A) và phế phẩm ($B = \bar{A}$). Kết quả cụ thể như sau:

Bảng 6.4. Các tần số thực nghiệm

Dây chuyền \ Chỉ tiêu	1	2	3	4	T_i
A	75	82	65	82	304
B	15	12	10	9	46
n_i	90	94	75	91	350

Giả thiết H_0 đặt ra là: chất lượng sản phẩm của các dây chuyền sản xuất có thuần nhất không.

Giả thiết H_0 này được kiểm tra bằng tiêu chuẩn χ^2 như biểu thức (6.16)

$$\chi_m^2 = \frac{N^2}{Ta \times Tb} \left[\sum_{i=1}^K \frac{a_i^2}{n_i} - \frac{T_a^2}{N} \right]$$

Thay số vào:

$$\chi_m^2 = \frac{350^2}{304 \times 46} \left[\left(\frac{75^2}{90} + \frac{82^2}{94} + \frac{65^2}{75} + \frac{82^2}{91} \right) - \frac{304^2}{350} \right] = 1,804$$

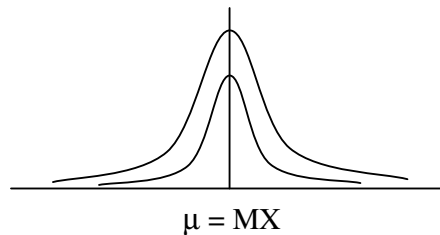
$\chi_{(\alpha, df)}^2$ ở mức ý nghĩa $\alpha = 0,05$ bậc tự do $df = 4-1 = 3$ là 7,815.

$\chi_{t,n}^2 < \chi_{(0,05;3)}^2$ ($1.804 < 7,815$) nên ta chấp nhận giả thiết H_0 nghĩa là "các dây chuyền sản xuất khác nhau của doanh nghiệp có chất lượng sản phẩm được sản xuất ra là như nhau" ở mức 5% (hay $\alpha = 0,05$).

6.5. Kiểm tra giả thiết về hai phương sai

Hay còn gọi là so sánh 2 phương sai

Phương sai là một trong hai đặc trưng chính của tổng thể. Phương sai đại diện cho mức độ phân tán của các giá trị ngẫu nhiên của tổng thể nó biểu trưng một đặc điểm căn bản của các hiện tượng biến thiên nội tại trên một bản chất đồng nhất. Có nhiều trường hợp gặp hai đám đông có trung bình như nhau, nhưng phương sai của chúng lại khác nhau nên hai đám đông đó cũng khác nhau (hình 6.1).



Vì vậy, muốn xem xét 2 phương sai cần phải kiểm định để có thể đưa ra được nhận định hai phương sai ấy có đồng nhất (như nhau) hay không?

Thí dụ: Có hai tổng thể thực nghiệm (hai công thức) là công thức I gồm có n_1 cá thể có phương sai mẫu s_1^2 , công thức II có n_2 cá thể với phương sai mẫu s_2^2 . Vấn đề cần đặt ra là sự chênh lệch hay khác nhau giữa hai phương sai s_1^2 và s_2^2 là do một nguyên nhân nào đấy hay chỉ là sự khác nhau ngẫu nhiên (do rút mẫu ngẫu nhiên, còn 2 phương sai lý thuyết bằng nhau). Để giải quyết vấn đề này phải kiểm định giả thiết.

$$H_0: \sigma_1^2 = \sigma_2^2 \text{ hay } \frac{\sigma_1^2}{\sigma_2^2} = \frac{\sigma_1^1}{\sigma_1^2} = 1 \quad (6.17)$$

Đối thiết H_1 : $\sigma_1^2 \neq \sigma_2^2$ bằng kiểm định F (Fisher - Snedecor).

Phương sai s^2 là một đại lượng ngẫu nhiên có phân phối χ^2 với $n - 1$ bậc tự do. Như vậy, ta sẽ có 2 đại lượng s_1^2 và s_2^2 với 2 đại lượng χ^2 như sau:

Đại lượng χ_1^2 với bậc tự do $df_1 = n_1 - 1$

Đại lượng χ_2^2 với bậc tự do $df_2 = n_2 - 1$

Giả sử 2 phương sai lý thuyết bằng nhau $\sigma_1^2 = \sigma_2^2$. Thiết lập tỷ số giữa 2 phương sai mẫu là $F_{t,n} = \frac{s_1^2}{s_2^2}$. Tỷ số này cũng chính là tỷ số $\frac{\chi_1^2 \times df_1}{\chi_2^2 \times df_2} = \frac{\chi_2^2 \times df_2}{\chi_1^2 \times df_1}$ (6.18)

Tỷ số giữa 2 phương sai là đại lượng ngẫu nhiên tuân theo luật phân phối Fisher (F). Như vậy dựa vào bảng tính sẵn F (bảng 5) phân phụ lục có thể tìm được ngưỡng so sánh $F_{(\alpha; df_1, df_2)}$ ứng với mức ý nghĩa α và 2 bậc tự do df_1 và df_2 .

df_1 là bậc tự do của phương sai tử số (cột trong bảng F)

df_2 là bậc tự do của phương sai mẫu số (hàng trong bảng F)

Nếu $F_{t,n} < F_{tt}$ thì hai phương sai bằng nhau hay sự khác nhau giữa hai phương sai là ngẫu nhiên. Nếu $F_{t,n} \geq F_{tt}$ thì hai phương sai khác nhau (không bằng nhau).

(Cần lưu ý khi tính $F_{t,n}$ thì chọn phương sai tử số có giá trị số học lớn hơn phương sai mẫu số).

Thí dụ: Để tìm hiểu sự khác nhau giữa năng suất mủ cao su (năng suất cá thể: kg/cây) của hai điều kiện địa hình trồng khác nhau là đất bằng và đất dốc 5%, người ta tiến hành khảo sát trên đất bằng $n_1 = 29$ cây, có $s_1^2 = 0,35$ (kg/cây); trên đất dốc $n_2 = 21$ cây có $s_2^2 = 0,65$ (kg/cây). Hai điều kiện trồng cao su khác nhau thì phương sai của năng suất cá thể có bằng nhau với mức ý nghĩa $\alpha = 0,05$.

Xuất phát từ giả thiết H_0 là hai phương sai bằng nhau, như vậy áp dụng biểu thức (6.18) cụ thể là giá trị $F_{t,n}$

$$F_{t,n} = \frac{0,65}{0,35} = 1,857$$

Với bậc tự do $df_1 = n_2 - 1 = 20$ (cột 20) và bậc tự do $df_2 = 28$ (hàng 28) ta tìm được giá trị $F_{(0,05;20,28)} = 1,96$. Như vậy $F_{t,n} < F_{(0,05;20,28)}$ chúng ta chấp nhận giả thiết H_0 nghĩa là trồng cao su trên đất bằng và đất dốc 5% thì phương sai của năng suất cá thể (lượng mủ cao su cao) là bằng nhau.

CHƯƠNG VII PHƯƠNG PHÁP SẮP XẾP CÔNG THỨC THÍ NGHIỆM VÀ PHÂN TÍCH KẾT QUẢ.

Mục đích của chương này là

Giúp cho sinh viên nắm vững phương pháp thiết kế một thí nghiệm nghiên cứu cơ bản, phương pháp bố trí thí nghiệm theo các kiểu khác nhau, biết cách phân tích phương sai (ANOVA) kết quả của thí nghiệm. Từ các kiến thức lý thuyết biết vận dụng vào thực tiễn việc bố trí thí nghiệm trong phòng cũng như trên đồng ruộng, biết cách phân tích kết quả thí nghiệm sau khi kết thúc và cách công bố kết quả nghiên cứu trên tạp chí khoa học hoặc trong các báo cáo khoa học.

Nội dung của chương:

1. Giới thiệu các kiểu thiết kế, sắp xếp thí nghiệm
 - Nhóm thí nghiệm một nhân tố:
 - i. Thiết kế kiểu hoàn toàn ngẫu nhiên (CRD)
 - ii. Thiết kế kiểu khối ngẫu nhiên đầy đủ (RCB)
 - iii. Thiết kế kiểu ô vuông la tinh (LS)
 - Nhóm thí nghiệm hai nhân tố:
 - i. Kiểu tổ hợp các mức của hai nhân tố có kiểu thiết kế hoàn toàn ngẫu nhiên (CRD) và khối ngẫu nhiên đầy đủ (RCB)
 - ii. Kiểu Chia ô (Split-Plot)
2. Phương pháp phân tích phương sai (ANOVA) tương ứng với các kiểu thiết kế
3. Cách so sánh các trung bình của các công thức và công bố kết quả phân tích, so sánh trên các báo cáo khoa học, luận văn tốt nghiệp hoặc tạp chí khoa học.

7.1. Các thí nghiệm một nhân tố.

7.1.1. Khái niệm:

Thí nghiệm một nhân tố là những thí nghiệm chỉ có một nhân tố thay đổi trong lúc tất cả các nhân tố khác được giữ nguyên. Như vậy, trong thí nghiệm này các công thức sẽ chứa các mức độ khác nhau của nhân tố đó, tất cả các nhân tố khác được áp dụng chung như nhau (ở cùng 1 mức bắt buộc) cho tất cả các công thức.

Ví dụ các thí nghiệm thử nghiệm giống cây trồng là thí nghiệm một nhân tố, ở đây giống là nhân tố thay đổi với các mức khác nhau (các giống khác nhau), có nghĩa là ở các công thức khác nhau, trên các mảnh thí nghiệm khác nhau, còn các nhân tố khác như quản lý, phân bón, phòng trừ sâu, bệnh... phải được áp dụng như nhau cho các mảnh thí nghiệm. Hay một ví dụ khác: nghiên cứu lượng phân đạm ảnh hưởng tới cây trồng thì một số lượng đạm thay đổi qua các công thức cho các mảnh thí nghiệm, còn lại tất cả các nhân tố khác được giữ cố định ở mức như nhau.

Hoặc một thí nghiệm so sánh mật độ cây trồng chẳng hạn thì mật độ thay đổi qua các công thức, còn các yếu tố khác được giữ cố định ở mức độ như nhau cho các mảnh thí nghiệm.

Đối với thí nghiệm một nhân tố, có thể thiết kế hoàn toàn ngẫu nhiên (CRD), thường dùng khi có ít công thức và khi có nhiều đơn vị thí nghiệm đồng nhất. Kiểu thiết kế thứ hai là

khối ngẫu nhiên bao gồm khối ngẫu nhiên đầy đủ (RCBD) và kiểu khối không đầy đủ (RIBD) sẽ được giới thiệu ở phần sau, kiểu thiết kế thứ ba là ô vuông la tinh (LS) sẽ được giới thiệu ở phần sau.

7.1.2. Các phương pháp sắp xếp và phân tích kết quả.

7.1.2.1. Thí nghiệm thiết kế kiểu hoàn toàn ngẫu nhiên (CRD)

Khái niệm

Một thí nghiệm được thiết kế kiểu hoàn toàn ngẫu nhiên khi các công thức được chỉ định một cách hoàn toàn ngẫu nhiên vào các ô thí nghiệm sao cho mỗi mảnh (ô) thí nghiệm đều có cơ hội như nhau để nhận được bất kỳ một công thức nào. Theo kiểu sắp xếp này, bất kỳ sự khác nhau nào (ngoài nhân tố thí nghiệm) giữa các ô thí nghiệm đều được coi là sai số thí nghiệm. Kiểu sắp xếp này chỉ phù hợp khi các đơn vị (ô, mảnh) thí nghiệm hoàn toàn đồng nhất (thường là thí nghiệm trong phòng), còn thí nghiệm ngoài đồng, thường có sự biến động lớn giữa các đơn vị thí nghiệm nên không hay áp dụng kiểu sắp xếp này.

Quá trình sắp xếp

Để minh họa qui trình sắp xếp và tiến tới vẽ sơ đồ của việc sắp xếp, ta dùng một ví dụ: thí nghiệm có 4 công thức A, B, C, D mỗi công thức được nhắc lại 5 lần, việc sắp xếp tiến hành theo các trình tự sau:

+ Bước 1. Xác định tổng số ô thí nghiệm cần có N.

$$N = r * t$$

Trong đó: N: tổng số ô thí nghiệm

t: số công thức cho mỗi lần nhắc lại

r: số lần nhắc lại của mỗi công thức

ở ví dụ này ta có: $N = r * t = 5 * 4 = 20$ ô như trong hình 7.1

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16
17	18	19	20

Hình 7.1. Sơ đồ mẫu cho thí nghiệm thiết kế kiểu hoàn toàn ngẫu nhiên có 4 công và thức và 5 lần nhắc lại

Ấn định các công thức vào các ô thí nghiệm bằng một trong các cách ngẫu nhiên sau:

Cách 1: Lấy bảng số ngẫu nhiên (bảng 1 phụ lục)

- a. Từ điểm xuất phát bất kỳ trong bảng, đọc thẳng xuống dưới lấy 20 số ngẫu nhiên liên tục có 3 chữ số. (Ta lấy 3 chữ số để cho không có số nào trùng với số thứ tự từ 1 đến 20 là các số có 1 và 2 chữ số đã được ấn định cho các ô ở bước 2). Trong ví dụ này, 20 số ngẫu nhiên có 3 chữ số cùng thứ tự xuất hiện của chúng được ghi lại như bảng sau:

Thứ tự xuất hiện	Số ngẫu nhiên	Thứ tự xuất hiện	Số ngẫu nhiên
1	937	11	918
2	149	12	772
3	908	13	243
4	361	14	494
5	953	15	704
6	749	16	549
7	180	17	957
8	951	18	157
9	018	19	571
10	427	20	226

b. Xếp thứ tự theo hạng tăng dần hoặc giảm dần 20 số ngẫu nhiên ở bước 2. Trong ví dụ này xếp hạng thứ tự từ nhỏ nhất (số 1) đến lớn nhất (số 20) như sau:

Thứ tự xuất hiện	Số ngẫu nhiên	Xếp hạng	Thứ tự xuất hiện	Số ngẫu nhiên	Xếp hạng
1	937	17	11	918	16
2	149	2	12	772	14
3	908	15	13	243	6
4	361	7	14	494	9
5	953	19	15	704	12
6	749	13	16	549	10
7	180	4	17	957	20
8	951	18	18	157	3
9	018	1	19	571	11
10	427	8	20	226	5

c. Chia N số xếp hạng ở bước 3 thành t nhóm, mỗi nhóm chứa r số thứ tự mà các số ngẫu nhiên xuất hiện. Trong ví dụ này, 20 số xếp hạng được chia thành 4 nhóm, mỗi nhóm chứa 5 số như sau:

Nhóm	Thứ tự xếp hạng				
1	17	2	15	7	19
2	13	4	18	1	8
3	16	14	6	9	12
4	10	20	3	11	5

d. Cho mỗi nhóm mang tên một công thức, nhóm 1 mang tên công thức A, nhóm 2 mang tên công thức B, nhóm 3 mang tên công thức C và nhóm 4 mang tên công thức D. Thứ tự xếp hạng được coi như thứ tự của các ô trong sơ đồ hình 7.1. Như vậy, kết quả là công thức A

nằm ở các vị trí ô số 17, 2, 15, 7 và 19. Công thức B nằm ở các vị trí ô số 13, 4, 18, 1, và 8. Cứ tiếp tục như vậy để chỉ định vị trí của các công thức còn lại.

Cách 2. Lấy một cỗ bài

a. Rút ra N con bài, mỗi lần rút một con sau đó trộn lại cỗ bài và rút con khác (không bỏ con đã rút vào). Phương pháp này áp dụng khi $N < 52$. Trong ví dụ này 20 con được rút ra cùng thứ tự của chúng như sau.

Thứ tự xuất hiện	1	2	3	4	5	6	7	8	9	10
Tên các con bài	J	3	A	R	Q	5	6	9	9	8
	♥	♦	♦	♥	♣	♥	♣	♠	♥	♠

Thứ tự xuất hiện	11	12	13	14	15	16	17	18	19	20
Tên các con bài	A	4	3	2	7	10	A	4	6	10
	♥	♦	♥	♣	♣	♠	♣	♠	♥	♣

b. Xếp hạng 20 con bài rút ra từ bước 1 theo thứ tự số từ 2 đến Át, nếu cùng số thì theo thứ tự của chất bài..

Trong ví dụ này, 20 con được xếp hạng tương ứng với thứ tự xuất hiện như sau:

Thứ tự xuất hiện và xếp hạng

1	2	3	4	5	6	7	8	9	10
14	7	9	15	5	11	2	19	13	18

Thứ tự xuất hiện và xếp hạng

11	12	13	14	15	16	17	18	19	20
16	8	10	1	3	20	6	17	12	4

c. Chia các số xếp hạng thành 4 nhóm, gán công thức và xếp các công thức vào các ô theo như cách 1.

Trong ví dụ này được kết quả thu được như sau:

Công thức	Án định vào các ô số				
A	14	7	9	15	5
B	11	2	19	13	18
C	16	8	10	1	3
D	20	6	17	12	4

Cách 3. Rút thăm.

a. Chuẩn bị N mẫu giấy, chia mẫu giấy thành t nhóm, các mẫu giấy trong mỗi nhóm có cùng ký hiệu của một công thức. Trong ví dụ này sẽ có 5 mẫu mang chữ A, 5 mẫu chữ B, ... Trộn lẫn 20 mẫu giấy trong một hộp (các mẫu giấy được gấp kín).

b. Rút mỗi lần một mẫu giấy, đặt vào các ô theo trật tự từ đầu đến cuối. Mở mảnh giấy ra, ta có công thức được chỉ định vào các ô như sau:

Thứ tự xuất hiện (ô) công thức theo các số như sau:

1	2	3	4	5	6	7	8	9	10
D	B	A	B	C	A	D	C	B	D
11	12	13	14	15	16	17	18	19	20
D	A	A	B	B	C	D	C	C	A

Trong ví dụ này như công thức A nằm ở các ô 3, 6, 12, 13, và 20.

Phân tích phương sai

Có 2 nguồn biến động trong N quan sát thu được từ thí nghiệm sắp xếp kiểu CRD. Một nguồn là biến động do công thức, nguồn thứ 2 là sai số thí nghiệm (do các yếu tố ngẫu nhiên tác động). Độ lớn của 2 nguồn biến động này được so sánh để biết liệu sự khác nhau của các quan sát từ các công thức có thực sự hay chỉ là sự khác nhau ngẫu nhiên. Sự khác nhau của các công thức là thực sự nếu biến động của các công thức đủ lớn hơn biến động của sai số thí nghiệm.

Sắp xếp thí nghiệm kiểu này có thuận lợi là đơn giản cho quá trình phân tích phương sai, đặc biệt là khi số lần nhắc lại của các công thức không bằng nhau.

1) Số lần nhắc lại bằng nhau.

Các bước phân tích phương sai cho thí nghiệm sắp xếp kiểu CRD khi số lần nhắc lại bằng nhau được minh họa bằng thí nghiệm dùng thuốc hoá học trừ rầy nâu và sâu đục thân lúa, với 4 lần nhắc lại và 7 công thức (6 loại thuốc mới và 1 đối chứng), số liệu được ghi trong bảng 7.1. Bảng 7.1. Năng suất lúa ở các công thức và các lần nhắc lại khác nhau.

Thứ tự công thức	Năng xuất hạt (kg/ ha)				Tổng công thức(T_i)	Trung bình công thức(\bar{x}_i)
	1	2	3	4		
T ₁	2.537	2.069	2.104	1.797	8.507	2.127
T ₂	3.366	2.591	2.211	2.544	10.712	2.678
T ₃	2.536	2.459	2.827	2.385	10.207	2.552
T ₄	2.387	2.453	1.556	2.116	8.512	2.128
T ₅	1.997	1.679	1.649	1.859	7.184	1.796
T ₆	1.796	1.704	1.904	1.320	6.724	1.681
T ₇	1.401	1.516	1.270	1.077	5.264	1.316
Tổng toàn bộ (G)					57.110	2.040
Trung bình toàn bộ						

Ghi chú: T₁: Dol- Mix (1kg)
 T₂: Dol- Mix (2kg)
 T₃: DDT+γ - BHC
 T₄: Azodrin

T₅:Dimecron- Boom
 T₆: Dimecron- knap
 T₇: Đối chứng
 r: Số nhắc lại = 4
 t: số công thức= 7

Các bước tính toán như sau:

Bước 1. Tính các tổng T_i của các công thức; tổng toàn bộ G và các \bar{x}_i (trung bình của các công thức) như trong bảng 7.1.

Bước 2. Đưa ra một bảng phân tích phương sai như sau:

Nguồn biến động	Bậc tự do (df)	Tổng bình phương (SS)	Bình phương Trung bình (MS)	F _{tn}	Fbảng	
					5%	1%
Công thức (T)						
Sai số (E)						
Toàn bộ (To)						

Bước 3. Tính các bậc tự do (df).

- Bậc tự do của toàn bộ dfTo = N-1 = r * t - 1 = 4 * 7 - 1 = 27
- Bậc tự do của công thức dfT = t - 1 = 7 - 1 = 6
- Bậc tự do của sai số dfE = t (r - 1) = 7(4 - 1) = 21 hoặc bằng bậc tự do toàn bộ - bậc tự do công thức = 27- 6 = 21

Bước 4. x_{ij} là số liệu của công thức i, lần lặp j

T_i: tổng các x_{ij} của công thức thứ i

N là tổng số mảnh thí nghiệm: N = t * r = 7 * 4 = 28

Tính số điều chỉnh (CF) và các loại tổng bình phương (SS) như sau:

$$CF = \frac{G^2}{N} = 116484004$$

$$SSTo = \sum_{i=1}^r x_{ij}^2 - CF = (2537^2 + 2069^2 + \dots + 1270^2) - 116484004 = 7577412$$

$$SST = \sum_{i=1}^t \frac{T_i^2}{r} - CF = \frac{8507^2 + 10712^2 + \dots + 5264^2}{4} - 116484004 = 5587174$$

$$SSE = SSTo - SST = 1990238$$

Bước 5. Tính các bình phương trung bình (MS) cho mỗi nguồn biến động.

$$MST = \frac{SST}{t-1} = 931196$$

$$MSE = \frac{SSE}{t(r-1)} = 94773$$

Bước 6. Tính trị số F_m để kiểm tra mức ý nghĩa khác nhau của các công thức

$$F_m = \frac{MST}{MSE} = \frac{931196}{94773} = 9,83$$

Chú ý: Giá trị F chỉ được tính khi bậc tự do của sai số đủ lớn cụ thể là khi $dfE \geq 6$

Bước 7. Tìm giá trị F (α, dfT, dfE) trong bảng F (bảng 5 phụ lục)

Trong ví dụ này ta có

$$F_{(0,05, 6,21)} = 2,57$$

$$F_{(0,01, 6, 21)} = 3,81$$

Bước 8. Đưa tất cả các giá trị tính được từ bước 3 đến bước 7 vào bảng phân tích phương sai xây dựng ở bước 2 như bảng 7.2.

Bảng 7.2. Bảng phân tích phương sai

Nguồn biến động	Bậc tự do (df)	SS	MS	$F_m^{(b)}$	F bảng	
					5%	1%
Công thức	6	5587174	931196	9,83**	2,57	3,81
Sai số	21	1990238				
Toàn bộ	27	7577412				

(b) ** Tại mức ý nghĩa 1%

cv% = 15.51%

Bước 9. So sánh các giá trị F_m và F trong bảng (F_{lt}) ở các mức ý nghĩa khác nhau theo nguyên tắc:

1. Nếu $F_m > F_{lt}$ tại mức ý nghĩa 1%, thì sự khác nhau giữa các công thức ở mức ý nghĩa cao (đánh dấu ** trên giá trị của F_m).
2. Nếu $F_m > F_{lt}$ tại mức ý nghĩa 5% nhưng $\leq F_{lt}$ tại mức ý nghĩa 1% thì sự khác nhau giữa các công thức là có ý nghĩa ở mức 5%, ta ghi một dấu * trên giá trị của F_m .
3. Nếu $F_m \leq F_{lt}$ tại mức ý nghĩa 5% thì sự khác nhau giữa các công thức là không có ý nghĩa, kết quả chỉ bằng chữ ns trên giá trị của F_m .

Chú ý khi kết luận sự khác nhau giữa các công thức là không có ý nghĩa nhưng không phải tất cả các công thức là như nhau vì phép thử F không có ý nghĩa vẫn có thể có sự khác nhau giữa các công thức nhưng rất nhỏ hoặc sai số thí nghiệm quá lớn. Vì vậy bất kỳ khi nào F tỏ ra không có ý nghĩa, người nghiên cứu cần xem lại độ lớn của sai số thí nghiệm và sự khác nhau giữa các trung bình công thức. Nếu cả 2 giá trị đều lớn thì nên làm lại thí nghiệm và tìm cách làm cẩn thận để giảm sai số thí nghiệm để thấy rõ sự khác nhau giữa các công thức. Mặt khác nếu cả hai giá trị đều nhỏ chứng tỏ sự khác nhau giữa các công thức là quá nhỏ nên không cần thí nghiệm thêm nữa.

Trong thí dụ này F_m là 9,83 > giá trị F bảng ở mức ý nghĩa 1% là 3,81 chứng tỏ sự khác nhau giữa các công thức ở ý nghĩa cao. Nói cách khác là trong 100 trường hợp thì có 99 trường hợp người ta luôn thấy sự khác nhau giữa các công thức khi thí nghiệm. Cũng cần nói thêm là kiểm tra F chỉ cho biết sự khác nhau chung giữa các công thức chứ không cho biết sự khác nhau hay không của từng đôi công thức, muốn biết phải tiến hành so sánh các trung bình của từng đôi một.

Bước 10. Tính trung bình toàn bộ và hệ số biến động CV% như sau.

$$\text{Trung bình toàn bộ } \bar{X} = \frac{G}{N}$$

$$CV\% = \frac{\sqrt{MSE}}{\bar{X}} * 100$$

Trong ví dụ này:

$$\bar{X} = \frac{57110}{28} = 2040$$

$$CV\% = \frac{\sqrt{94773}}{2040} * 100 = 15,1\%$$

CV% chỉ độ chính xác của các công thức, nó cho biết sai số thí nghiệm lớn hay nhỏ. Vậy CV% (sai số thí nghiệm) lớn, nhỏ là tùy thuộc vào từng thí nghiệm. Người ta thường hay ghi giá trị CV% ở dưới bảng phân tích phương sai.

Sai số thí nghiệm được chấp nhận ở mức nào, tùy theo kiểu thí nghiệm, loại cây trồng và đặc trưng quan sát. Bảng kinh nghiệm cho thấy ở IRRRI có thể chấp nhận trong thí nghiệm trồng lúa có CV% từ 6 - 8% cho thí nghiệm giống, 10 - 12% cho thí nghiệm phân bón và 13 - 15% cho thí nghiệm bảo vệ thực vật, CV% của các đặc trưng quan sát khác nhau là khác so với năng suất. Ví dụ CV% cho năng suất 10% thì số nhánh đẻ khoảng 20% và với chiều cao là 3%.

Bước 11. So sánh trung bình của các công thức theo tiêu chuẩn t.

Tính giới hạn sai khác nhỏ nhất có ý nghĩa tại mức ý nghĩa α . Theo công thức.

$$LSD_{\alpha} = t_{\alpha} * s_{\bar{d}}$$

Trong đó: t_{α} : giá trị t lý thuyết tra từ bảng t với bậc tự do bằng bậc tự do của sai số ký hiệu là $t_{\alpha, dfE}$, (bảng 4 phụ lục).

$s_{\bar{d}}$ là sai số chuẩn (hay sai số của trung bình)

$$s_{\bar{d}} = \sqrt{\frac{2MSE}{r}}$$

Trong đó: r: số nhắc lại

MSE bình phương trung bình của sai số ở bước 5.

Trong ví dụ này ta có

$$LSD_{0,05} = 2,080 \sqrt{\frac{2 * 94773}{4}} = 453Kg / ha$$

$$LSD_{0,01} = 2,831 \sqrt{\frac{2 * 94773}{4}} = 616Kg / ha$$

- Lập bảng so sánh năng suất giữa các công thức với nhau như sau.

Thứ tự	Năng suất (kg)	T ₁ -	T ₂ -	T ₃ -	T ₄ -	T ₅ -	T ₆ -
T ₁	2127	-	-	-	-	-	-
T ₂	2678	-551*	-	-	-	-	-
T ₃	2552	-425 ^{ns}	126 ^{ns}	-	-	-	-
T ₄	2128	-1 ^{ns}	550*	424 ^{ns}	-	-	-
T ₅	1796	331 ^{ns}	882**	756**	332 ^{ns}	-	-
T ₆	1681	446 ^{ns}	997*	871**	447 ^{ns}	115 ^{ns}	-
T ₇	1316	811**	1.362*	1236**	812**	480*	365 ^{ns}

Ghi chú **: Khác nhau ở mức ý nghĩa 1%

*: Khác nhau ở mức ý nghĩa 5%

ns: Không khác nhau có ý nghĩa.

Kết luận: - Trừ công thức 6 còn lại tất cả các công thức xử lý thuốc đều có năng suất cao hơn đối chứng.

- Công thức 2 và 3 có năng suất cao nhất.

Bước 12. Thể hiện kết quả so sánh.

Từ kết quả so sánh ở bước 11, xếp trung bình của các công thức theo thứ tự giảm dần. Biểu diễn kết quả ở một mức xác suất ý nghĩa nào đó (ví dụ 5%) dưới dạng hình học và gán cho chúng các chỉ số bằng các chữ khác nhau tương ứng với các mức khác nhau để cho biết các công thức khác nhau thì mang các chỉ số bằng chữ khác nhau.

Trong ví dụ này, kết quả được biểu diễn như sau.

Năng suất các công thức được sắp xếp:

Năng suất: 2678 2552 2128 2127 1796 1681 1316

Mức : a _____

b _____

c _____

d _____

Bước 13. Công bố kết quả.

Thứ tự công thức	Năng suất (kg/ha)	Chỉ số đánh giá	Sai số chuẩn(ES)
T ₁	2127	bc	$ES = \frac{\sqrt{MSE}}{\sqrt{r}} = 153$
T ₂	2678	a	
T ₃	2552	b	
T ₄	2128	bc	
T ₅	1796	c	
T ₆	1681	cd	
T ₇	1316	d	

LSD_{0.05} = 453kg/ha; CV% = 15,51%.

Ghi chú: những công thức mang chữ ở cột chỉ số đánh giá giống nhau là giống nhau, ngược lại khác chữ là khác nhau có ý nghĩa ở mức tin cậy 95%.

2) *Số nhắc lại công thức không bằng nhau.*

Nói chung, phương pháp phân tích phương sai thí nghiệm sắp xếp kiểu CRD không có gì phức tạp khi số lần nhắc lại không bằng nhau nên kiểu sắp xếp CRD thường được áp dụng cho các thí nghiệm mà ít có đủ vật liệu để tạo ra số nhắc lại bằng nhau cho mỗi công thức.

Ví dụ:

- Số con cho mỗi động vật không đủ bằng nhau cho mỗi công thức về thức ăn.
- Thí nghiệm so sánh chiều dài thân của côn trùng bắt được trong bẫy đèn.
- Thí nghiệm lúc đầu bố trí nhắc lại bằng nhau nhưng trong quá trình thí nghiệm có thể bị mất hoặc bị phá hoại.

Các phương pháp phân tích phương sai cho thí nghiệm sắp xếp kiểu CRD khi số nhắc lại không bằng nhau như sau:

Ta có một thí nghiệm so sánh hiệu quả của các cách sử dụng thuốc trừ cỏ khác nhau sắp xếp kiểu CRD, kết quả ghi lại như bảng 7.3.

Bảng 7.3. Năng suất lúa thu được trên các cách sử dụng thuốc trừ cỏ khác nhau khi số nhắc lại không bằng nhau.

Công thức	Năng suất (kg/ ha)/ nhắc lại				Tổng công thức(T _i)	Trung bình công thức
	1	2	3	4		
1	3.187	4.610	3.562	3.217	14.576	3.644
2	3.390	2.875	2.775		9.040	3.013
3	2.797	3.001	2.505	3.490	11.793	2.948
4	2.832	3.103	3.448	2.255	11.638	2.910
5	2.233	2.743	2.727		7.703	2.568
6	2.952	2.272	2.470		7.694	2.565
7	2.858	2.895	2.458	1.723	9.934	2.484
8	2.308	2.335	1.975		6.618	2.206
9	2.013	1.788	2.248	2.115	8.164	2.041
10	3.202	3.060	2.240	2.690	11.192	2.798
11(đ/c)	1.192	1.652	1.075	1.030	4.949	1.237
Tổng toàn bộ (G)					103.301	2.583
Trung bình toàn bộ						

Bước 1. Tính tương tự bước 1, 2 của ví dụ trên

Bước 2. Ta có t là số công thức, N là tổng số quan sát, xác định độ tự do của mỗi nguồn biến động như sau:

$$\begin{aligned}dfT_o &= N - 1 = 40 - 1 = 39 \\dfT &= t - 1 = 11 - 1 = 10 \\dfE &= dfT_o - dfT = 39 - 10 = 29\end{aligned}$$

Bước 3. Tính số hiệu chỉnh CF và các tổng bình phương:

$$\begin{aligned}CF &= \frac{G^2}{N} = \frac{(103301)^2}{40} = 266777415 \\SST_o &= \sum_{i=1}^n x^2_i - CF = 20209724 \\SST &= \sum_{i=1}^t \frac{T^2_i}{r_i} - CF = \frac{14576^2}{4} + \frac{9040^2}{3} + \dots + \frac{4949^2}{4} - CF = 15090304 \\SSE &= SST_o - SST = 5119420\end{aligned}$$

Bước 4. Theo như bước 5 đến 10 của ví dụ trên.

Phân tích phương sai tổng hợp cho ví dụ này ở bảng 7.4. kết quả kiểm tra F cho thấy sự khác nhau có ý nghĩa cao giữa các công thức.

Bảng 7.4. Phân tích phương sai của năng suất trong bảng 7.3^(a).

Nguồn biến động	Bậc tự do	SS	MS	F _m ^(b)	F bảng	
					5%	1%
Công thức	10	15090304	1509030	8,55 ^{**}	2,18	3,00
Sai số	29	5119420	176532			
Toàn bộ	39	20209724				

(a) cv% = 16,3%

(b) ^{**} tại mức ý nghĩa 1%

Bước 5. So sánh trung bình các công thức.

1- Tính giới hạn sai khác nhỏ nhất có ý nghĩa LSD_{0.05}

+ Cho việc so sánh giữa các công thức có 4 lần nhắc

$$LSD_{0.05} = t_{(0.05,29)} * \sqrt{\frac{2MSE}{r}} = 2,045 \sqrt{\frac{2 * 176532}{4}} = 608 \text{ kg/ha}$$

+ Cho việc so sánh giữa các công thức có 3 lần nhắc với 4 lần nhắc.

$$LSD_{0.05} = t_{(0.05,29)} \sqrt{MSE \left(\frac{1}{r_i} + \frac{1}{r_j} \right)} = 2,045 \sqrt{176532 \left(\frac{1}{4} + \frac{1}{3} \right)} = 656 \text{ kg/ha}$$

+ Cho việc so sánh giữa các công thức có 3 lần nhắc

$$LSD_{0.05} = t_{0.05,29} \sqrt{\frac{2MSE}{r}} = 2,045 \sqrt{\frac{2 * 176532}{3}} = 429,61$$

Vì ở đây ta có một đối chứng không dùng thuốc nên tốt nhất là so sánh các công thức với đối chứng như bảng sau:

Số công thức	Số nhắc lại	T.B công thức	Sai khác so với đối chứng	LSD _{0,05}
1	4	3.644	2.407*	608
2	3	3.013	1.776*	656
3	4	2.948	1.761*	608
4	4	2.910	1.673*	608
5	3	2.568	1.331*	656
6	3	2.565	1.328*	656
7	4	2.484	1.247*	608
8	3	2.206	969*	656
9	4	2.041	804*	608
10	4	2.798	1516*	608
11(đ/c)	4	1.237	-	-

Hoặc ta cũng lập bảng so sánh các công thức với nhau như trường hợp trên để tìm công thức nào cao hơn xong thể hiện so sánh kết quả dưới dạng hình học để xếp hạng a, b, c,

7.1.2.2. Thí nghiệm sắp xếp theo kiểu khối ngẫu nhiên đầy đủ (RCBD hay RCB).

Khái niệm.

Kiểu sắp xếp RCB là một kiểu sắp xếp được sử dụng rộng rãi nhất trong nghiên cứu nông nghiệp. Nó hoàn toàn phù hợp với thí nghiệm ngoài đồng khi số công thức không quá lớn và khu thí nghiệm có thể biết trước được chiều hướng biến đổi độ phì nhiêu của đất. Đặc trưng chủ yếu của kiểu sắp xếp này là tạo ra các khối có kích thước bằng nhau và mỗi khối chứa tất cả các công thức của một lần nhắc lại.

Kỹ thuật tạo khối.

Mục tiêu chính của việc tạo khối là giảm sai số thí nghiệm bằng việc hạn chế sự đóng góp của nguồn biến động đã biết trong các đơn vị thí nghiệm. Điều này được làm bằng cách nhóm các đơn vị thí nghiệm vào các khối mà sự khác nhau trong mỗi khối là tối thiểu và sự khác nhau giữa các khối là tối đa. Bởi vì chỉ có biến động trong một khối mới có thể trở thành thành phần của sai số thí nghiệm, việc tạo khối là hiệu quả nhất để giảm sai số thí nghiệm khi biết trước được sự biến động của đơn vị thí nghiệm. Hình dạng mảnh thí nghiệm và hướng khối cũng cần quan tâm sắp đặt, sắp đặt sao cho sự khác nhau giữa các khối càng lớn còn sự khác nhau trong khối càng nhỏ.

Có 2 điều đáng chú ý để giúp cho việc tạo khối có hiệu quả và thích hợp là:

- + Lựa chọn nguồn biến động được sử dụng như là cơ sở của việc tạo khối.
- + Lựa chọn dạng và hướng khối cho thích hợp.
- + Những nguồn biến động thường được dùng làm cơ sở cho việc tạo khối luôn được dự đoán là:

- + Đất đai không đồng nhất mà đối với thí nghiệm phân bón hoặc giống thì người ta quan tâm trước hết tới năng suất.
- + Hướng di chuyển của côn trùng, mà thường là sự lan tràn của côn trùng trong các thí nghiệm thuốc hoá học.
- + Độ dốc của đồng ruộng, đặc biệt trong thí nghiệm nghiên cứu phản ứng của cây với sự điều tiết nước.

Sau khi đã biết hướng của nguồn biến động là cơ sở cho việc tạo khối rồi, phải lựa chọn kích thước và dạng khối cho thích hợp để làm tối đa biến động giữa các khối.

Có một số gợi ý cho việc lựa chọn kích thước và dạng khối như sau:

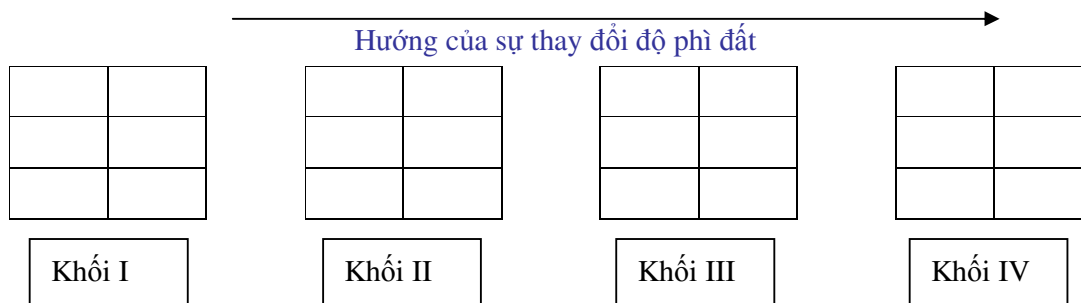
- 1- Khi sự thay đổi đồng nhất theo một hướng thì để khối dài và hẹp và để chiều dài của khối vuông góc với hướng của sự thay đổi.
- 2- Khi có sự thay đổi độ phì theo 2 hướng thì chọn hướng có sự thay đổi mạnh hơn làm cơ sở để tạo khối theo gợi ý 1.
- 3- Khi sự thay đổi độ phì theo 2 hướng ngang bằng nhau thì cần phải:
 - Sử dụng khối càng vuông càng tốt
 - Sử dụng khối dài và hẹp với chiều dài của khối trục giao với một hướng thay đổi và sử dụng kỹ thuật Covariance để quan tâm tới hướng kia.
 - Sử dụng kiểu sắp xếp ô vuông la tinh.
- 4- Khi có sự biến đổi mà không dự đoán được chiều hướng thì nên làm khối vuông.

Bất cứ khi nào dùng kỹ thuật khối thì phải duy trì và điều khiển sao cho biến động giữa các khối lớn hơn biến động trong mỗi khối. Ví dụ, nếu ta không đủ điều kiện làm cả thí nghiệm trong một ngày thì phải làm xong gọn cho từng khối một trong một thời gian càng ngắn càng tốt. Như vậy nếu có biến động xảy ra theo thời gian thì nó sẽ được đưa vào biến động giữa các khối nên có thể loại trừ ra được khỏi sai số thí nghiệm. Nếu có nhiều người quan sát, theo dõi thí nghiệm thì mỗi người phải được chỉ định quan sát cho tất cả các mảnh của cùng một khối. Bằng cách này ta có thể loại sự biến động giữa các người quan sát ra khỏi sai số thí nghiệm.

Tiến hành ngẫu nhiên và vẽ sơ đồ sắp xếp.

Quá trình ngẫu nhiên hoá cho sắp xếp kiểu RCB được tiến hành một cách tách biệt và độc lập với mỗi khối. Sử dụng một thí nghiệm có 6 công thức A, B, C, D, E, F và 4 lần nhắc lại để minh hoạ phương pháp.

Bước 1. Chia khu vực thí nghiệm thành r khối bằng nhau, r là số nhắc lại theo kỹ thuật khối như miêu tả ở 7.1.2.1. Trong ví dụ, diện tích thí nghiệm được chia làm 4 khối như hình 7.2. Giả thiết hướng thay đổi độ phì dọc theo chiều dài của khu thí nghiệm, dạng khối là hình chữ nhật trục giao với hướng của sự thay đổi.



Hình 7.2. Chia thí nghiệm thành 4 khối, mỗi khối 6 mảnh.

Bước 2. Chia mỗi khối thành t mảnh thí nghiệm, t là số công thức. Xong chỉ định t công thức vào t mảnh trong mỗi khối theo cách hoàn toàn ngẫu nhiên.

Ví dụ chỉ định t công thức vào t mảnh của khối 1 như sau:

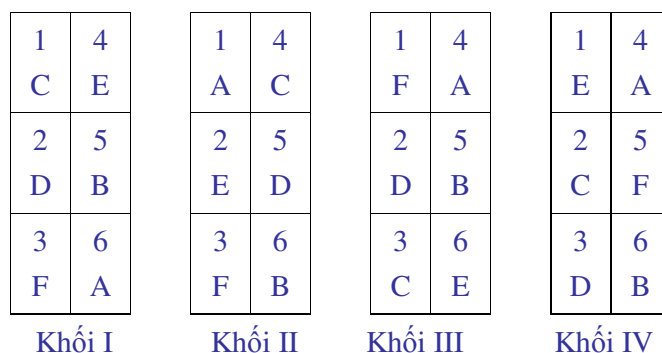
Đánh số liên tiếp từ 1 đến 6 vào 6 mảnh của khối I từ trên xuống dưới và từ trái sang phải như hình 7.3.

Xong chỉ định 6 công thức vào 6 mảnh 1 cách ngẫu nhiên bằng cách dùng bảng ngẫu nhiên như giới thiệu ở trên

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

Hình 7.3 Đánh số mảnh và chỉ định ngẫu nhiên 6 công thức vào 6 mảnh

Bước 3. Làm lại bước 2 cho toàn bộ các khối còn lại. Như vậy, trong ví dụ này ta được sơ đồ cuối cùng như hình 7.4.



Hình 7.4 Sơ đồ mẫu của thí nghiệm sắp xếp kiểu khối ngẫu nhiên với 6 công thức

A, B, C, D, E, F và 4 nhắc lại.

Phân tích phương sai

Có ba nguồn biến động trong kiểu sắp xếp RCB, đó là công thức, nhắc lại (khối) và sai số thí nghiệm.

Để minh họa các bước cho phân tích phương sai của kiểu sắp xếp RCB ta dùng số liệu của một thí nghiệm so sánh 6 loại mật độ gieo lúa NN8 như bảng 7.5.

Bảng 7.5. Năng suất lúa NN8 với 6 mật độ gieo khác nhau trong kiểu sắp xếp RCB với 4 lần nhắc lại.

Công thức (kg hạt / ha)	Sản lượng hạt thu được (kg/ ha)				Tổng công thức (T _i)	Trung bình công thức x _i
	N. lại I	N.lại II	N. lại III	N. lại IV		
25	5.113	5.398	5.307	4.678	20.496	5.124
50	5.346	5.952	4.719	4.624	20.281	5.070
75	5.272	5.713	5.483	4.749	21.217	5.304
100	5.164	4.831	4.986	4.410	19.391	4.848
125	4.804	4.848	4.432	4.748	18.832	4.708
150	5.254	4.542	4.919	4.098	18.813	4.703
Tổng nhắc lại (R)	30.953	31.284	29.846	26.947		
Tổng toàn bộ(G)					119.030	
Trung bình toàn bộ						4.960

Bước 1. Đưa số liệu vào bảng theo các công thức và nhắc lại, tính tổng các công thức (T_i), tổng các nhắc lại (R_j) và tổng toàn bộ (G) như bảng 7.5.

Bước 2. Xây dựng bảng phân tích phương sai sau.

Nguồn biến động	Bậc tự do (df)	Tổng bình phương (SS)	Bình phương Trung bình (MS)	F _{in}	F bảng	
					5%	1%
Nhắc lại(R) (khối)						
Công thức (T)						
Sai số (E)						
Toàn bộ (To)						

Bước 3. r là số nhắc lại, t là số công thức, bậc tự do cho mỗi nguồn biến động được tính như sau:

$$\text{Bậc tự do toàn bộ } df_{To} = r * t - 1 = 24 - 1 = 23$$

$$\text{Bậc tự do nhắc lại } df_R = r - 1 = 4 - 1 = 3$$

Bậc tự do của công thức dfT = t - 1 = 6 - 1 = 5

Bậc tự do của sai số dfE = (r - 1)(t - 1) = 3 * 5 = 15

Hoặc dfE = dfTo - dfR - dfT = 23 - 3 - 5 = 15

Bước 4. Tính số hiệu chỉnh (CF) và các loại tổng bình phương

$$CF = \frac{G^2}{rt} = \frac{119030^2}{4 * 6} = 590339204$$

$$SSTo = \sum_{i=1}^t \sum_{j=1}^r x^2_{ij} - CF = (5113^2 + 5398^2 + \dots + 4098^2) - CF = 4801068$$

$$SSR = \sum_{j=1}^r R^2_{j/} - CF = \frac{30953^2 + 31284^2 + 29846^2 + 26947^2}{6} = 1944361$$

$$SST = \sum_{i=1}^t T^2_{i/} - CF = \frac{20496^2 + \dots + 18813^2}{4} - CF = 1198331$$

$$SSE = SSTo - SSR - SSCt = 1568376$$

Bước 5. Tính các bình phương trung bình cho mỗi nguồn biến động.

$$MSR = \frac{SSR}{r - 1} = 648120$$

$$MST = \frac{SST}{r - 1} = 239666$$

$$MSE = \frac{SSE}{(r - 1)(t - 1)} = 110558$$

Bước 6. Tính giá trị F để kiểm tra sự khác nhau giữa các công thức.

$$F_m = \frac{MST}{MSE} = 2,17$$

Bước 7. So sánh F_m với F bảng để kết luận như bước 9 của 7.1.2.1

Bước 8. Tính hệ số biến động cv%.

$$cv\% = \frac{\sqrt{MSE}}{T_{binh\ toan\ bo}} * 100 = 6.7\%$$

Bước 9. Điền tất cả các giá trị tính được từ bước 3 đến bước 8 vào bảng phân tích phương sai ở bước 2. Kết quả được ghi trên bảng 7.6.

Bảng 7.6. Phân tích phương sai của RCB về năng suất hạt ở bảng 7.5

Nguồn biến động	Bậc tự do	Tổng bình phương	Bình phương Trung bình	F _m	F bảng	
					5%	1%
Nhắc lại	3	1944361	648120		2,90	4,50
Công thức	5	1198331	239666	2,17 ^{ns}		
Sai số	15	1658376	110558			
Toàn bộ	23	4801068				

$$cv\% = 6,7\%$$

ns : không ý nghĩa.

Tính hiệu quả của khối

Mục đích của việc tạo khối là làm tối đa sự khác nhau giữa các khối và sự khác nhau giữa các mảnh trong khối càng nhỏ càng tốt. Vậy cần kiểm tra xem mục đích này có đạt được không trong mỗi thí nghiệm kiểu RCB. Phương pháp tính toán như sau (dựa trên số liệu của bảng 7.5 và 7.6)

Bước 1. Xác định mức ý nghĩa của biến động nhắc lại qua trị số F_R

$$F_R = \frac{MSR}{MSE}$$

Và kiểm tra ý nghĩa của nó bằng cách so với F bảng với bậc tự do $df_1 = r - 1$ và $df_2 = (r - 1)(t - 1)$. Hiệu quả của việc tạo khối có ý nghĩa nếu $F_R > F$ bảng.

Ở ví dụ này,

$$F_R = \frac{648 \cdot 120}{110 \cdot 558} = 5,86 \quad F_{0,05, 3, 15} = 3,29$$

Vậy sự khác nhau giữa các khối là có ý nghĩa cao vì $F_R > F_{0,05}$

Bước 2. Xác định độ lớn của sự giảm sai số thí nghiệm do việc tạo khối qua việc tính hiệu quả tương đối R.E như sau:

$$RE = \frac{(r - 1)MSR + r(t - 1)MSE}{(rt - 1)MSE}$$

Trong đó. MSR là bình phương trung bình của nhắc lại.

MSE là bình phương trung bình của sai số trong phân tích phương sai của RCB.

Nếu bậc tự do của sai số nhỏ hơn 20, R.E sẽ được nhân với số hiệu chỉnh **K** được xác định như sau:

$$K = \frac{[(r - 1)(t - 1) + 1][t(r - 1) + 3]}{[(r - 1)(t - 1) + 3][t(r - 1) + 1]}$$

Trong ví dụ này, ta có.

$$RE = \frac{(3)(648120) + 4(5)(110558)}{(24 - 1)(110558)} = 1,63$$

Vì bậc tự do của sai số là 15, số điều chỉnh K là

$$K = \frac{[(3)(5) + 1][6(3) + 3]}{[(3)(5) + 3][6(3) + 1]} = 0,982$$

Và RE điều chỉnh là $RE = K(RE) = (0,982)(1,63) = 1,60$

Vậy sử dụng kiểu RCB làm tăng độ chính xác của thí nghiệm lên 60%.

7.1.2.3. Thí nghiệm sắp xếp kiểu ô vuông la tinh (Latin Square)

Khái niệm

Đặc trưng cơ bản của sắp xếp ô vuông la tinh (LS) là khả năng xử lý cùng một lúc hai nguồn biến động đã biết giữa các đơn vị thí nghiệm. Coi đó như hai khối độc lập, thay cho chỉ có một khối trong thiết kế kiểu RCB. Trong kiểu LS hai khối đó là khối hàng và khối cột

vuông góc nhau và được sắp xếp để đảm bảo cho mỗi công thức chỉ được có mặt một lần trong khối hàng và một lần trong khối cột. Ta có thể ước lượng được biến động giữa các khối hàng cũng như khối cột và tách nó ra khỏi sai số thí nghiệm.

Các trường hợp thích ứng cho kiểu LS.

- + Thí nghiệm ngoài đồng khi khu thí nghiệm có sự thay đổi độ phì theo 2 hướng vuông góc nhau hoặc có một hướng nhưng lại có hiệu quả dư thừa từ những đợt trước.
- + Thí nghiệm về thuốc hoá học mà sự di chuyển của côn trùng theo hướng dự báo được trực giao với hướng thay đổi độ phì nhiều của khu thí nghiệm.
- + Thử nghiệm trong nhà kính mà các mảnh thí nghiệm được sắp xếp theo đường thẳng trực giao với tường nhà (hai tường tạo một góc thì sẽ tạo hai nguồn biến động).
- + Thử nghiệm trong phòng với số lần nhắc lại qua thời gian cũng tạo hai nguồn biến động.

Việc thể hiện khối hàng và khối cột trong thiết kế LS là có ích vì quan tâm được hai nguồn biến động đặc biệt nhưng nó lại là hạn chế vì yêu cầu mỗi công thức chỉ được một lần trên hàng hoặc trên cột chỉ được thoả mãn khi số nhắc lại bằng số công thức. Khi số công thức lớn thì thiết kế kiểu LS không đáp ứng được. Mặt khác khi công thức nhỏ thì bậc tự do của sai số thí nghiệm trở nên quá nhỏ để có thể sử dụng khi tính F_{tn} .

Vậy thực tế thiết kế kiểu LS chỉ được áp dụng cho những thí nghiệm có số công thức không ít hơn 4 và không nhiều hơn 8. Vì hạn chế này mà thiết kế kiểu LS không được sử dụng rộng rãi trong các thí nghiệm nông nghiệp. Mặc dầu nó có tiềm năng lớn để điều khiển sai số thí nghiệm.

Quá trình ngẫu nhiên và vẽ sơ đồ thí nghiệm.

Quá trình ngẫu nhiên hoá và vẽ sơ đồ sắp xếp cho thí nghiệm kiểu ô vuông latin được thể hiện qua một thí nghiệm có 5 công thức A, B, C, D, E như sau:

Bước 1. Lựa chọn một sơ đồ ô vuông latin mẫu với 5 công thức (bảng 13 phụ lục) như bảng dưới đây:

A	B	C	D	E
B	A	E	C	D
C	D	A	E	B
D	E	B	A	C
E	C	D	B	A

Bước 2: Ngẫu nhiên việc sắp xếp theo hàng của sơ đồ lựa chọn từ bước 1 theo sơ đồ ngẫu nhiên bằng phương pháp bảng ngẫu nhiên:

+ Lựa chọn 5 số ngẫu nhiên có 3 chữ số: ví dụ 628, 846, 475, 902 và 425 xong xếp hạng các số được chọn từ thấp đến cao nhất.

Số ngẫu nhiên	Thứ tự xuất hiện	Xếp hạng
628	1	3
846	2	4
475	3	2
902	4	5
452	5	1

+ Sử dụng thứ tự xếp hạng thể hiện số hàng của sơ đồ đã chọn và thứ tự xuất hiện thể hiện số hàng của sơ đồ mới.

Trong ví dụ này thì hàng số 3 của sơ đồ đã chọn trở thành hàng đầu tiên của sơ đồ mới, hàng thứ 4 của sơ đồ đã chọn thành hàng thứ 2 của sơ đồ mới... kết quả của sơ đồ mới sau khi đã ngẫu nhiên hoá số hàng là:

C	D	A	E	B
D	E	B	A	C
B	A	E	C	A
E	C	D	B	A
A	B	C	D	E

Bước 3. Ngẫu nhiên việc sắp xếp số cột.

Sử dụng cùng phương pháp cho sắp xếp hàng ở bước 2. Trong ví dụ này, 5 số ngẫu nhiên được chọn và thứ bậc của chúng như sau:

Số ngẫu nhiên	Thứ tự xuất hiện	Xếp hạng
792	1	4
032	2	1
947	3	5
293	4	3
196	5	2

Số xếp hạng được thể hiện cho số cột thu được của sơ đồ bước 2 và số thứ tự xuất hiện dùng làm số cột của sơ đồ cuối.

Trong ví dụ này, cột thứ tư của sơ đồ thu được trong sơ đồ bước 2 trở thành cột thứ nhất của sơ đồ cuối, cột thứ nhất của sơ đồ bước 2 trở thành cột thứ 2 của sơ đồ cuối... Đây là sơ đồ cuối cùng cho thí nghiệm.

Số hàng	Số cột				
	1	2	3	4	5
1	E	C	B	A	D
2	A	D	C	B	E
3	C	B	D	E	A
4	B	E	A	D	C
5	D	A	E	C	B

Phân tích phương sai.

Có 4 nguồn biến động trong thiết kế LS đó là hàng, cột, công thức và sai số thí nghiệm.

Minh họa phương pháp tính, ta dùng số liệu về năng suất hạt của 3 giống ngô lai(A, B, D) và giống đối chứng (C) từ một thí nghiệm so sánh giống với kiểu thiết kế LS, 4 x 4 trong bảng 7.7.

Bảng 7.7. Sản lượng hạt của 3 giống có triển vọng (A, B và D) và giống đối chứng (C).

Số hàng	Năng suất hạt (tấn/ha)				Tổng hàng (R)
	Cột 1	Cột 2	Cột3	Cột 4	

1	1,640 ^(B)	1,210 ^(D)	1,425 ^(C)	1,345 ^(A)	5,620
2	1,457 ^(C)	1,185 ^(A)	1,400 ^(D)	1,290 ^(B)	5,350
3	1,670 ^(A)	0,710 ^(C)	1,665 ^(B)	1,180 ^(D)	5,225
4	1,565 ^(D)	1,290 ^(B)	1,655 ^(A)	0,660 ^(C)	5,170
Tổng cột (C)	4,395	4,475	6,145	6,350	
Tổng toàn bộ(G)					21,365

Các bước phân tích phương sai như sau:

Bước 1. Cộng số liệu quan sát được theo hàng và theo cột như bảng 7.7.

Bước 2. Tính các tổng của mỗi công thức và trung bình công thức như sau.

<u>Công thức</u>	<u>Tổng số</u>	<u>Trung bình</u>
A	5,855	1,464
B	5,885	1,471
C	4,270	1,068
D	5,355	1,399

Bước 3. Xây dựng đề cương bảng phân tích phương sai như sau.

Nguồn biến động	Bậc tự do	Tổng bình phương	Bình phương Trung bình	F _{tn}	F	
					5%	1%
Hàng(R)						
Cột(C)						
Công thức (T)						
Sai số (E)						
Toàn bộ (To)						

Bước 4. Ký hiệu t là số công thức, xác định độ tự do cho các nguồn biến động như sau.

$$df_{To} = t^2 - 1 = 16 - 1 = 15$$

$$df_R = df_C = df_T = t - 1 = 4 - 1 = 3$$

$$df_E = (t - 1)(t - 2) = (4 - 1)(4 - 2) = 6$$

$$\text{Hoặc } df_E = df_{To} - df_R - df_C - df_T = 6$$

Bước 5. Tính số hiệu chỉnh (CF) và các loại tổng bình phương:

$$CF = \frac{G^2}{t^2} = 28,528952$$

$$SSTo = \sum x^2 - CF = 1,413923$$

$$SSR = \frac{\sum R^2}{t} - CF = 0,030154$$

$$SSC = \frac{\sum C^2}{t} - CF = 0,827342$$

$$SST = \frac{\sum T^2}{t} - CF = 0,426842$$

$$SSE = SST_0 - SSR - SSC - SST = 0,129585$$

Bước 6. Tính bình phương trung bình cho mỗi nguồn biến động

$$MSC = \frac{SSC}{t-1} = 0,275781$$

$$MSR = \frac{SSR}{t-1} = 0,010051$$

$$MST = \frac{SST}{t-1} = 0,142281$$

$$MSE = \frac{SSE}{(t-1)(t-2)} = 0,21598$$

Bước 7. Tính F_{tn} để kiểm tra hiệu quả của công thức.

$$F_{tn} = \frac{MST}{MSE} = 6,59$$

Bước 8. So sánh F_{tn} với F bảng tại 5% = 4,76 hoặc F bảng tại 1% = 9,78

Vậy các công thức khác nhau đã dẫn đến kết quả khác nhau tại mức ý nghĩa 5%.

Bước 9. Tính cv%

$$cv\% = \frac{\sqrt{MSE}}{\text{trung bình toán bo}} * 100 = 11,0\%$$

Bước 10. Điền tất cả các giá trị tính được bước 4 đến bước 9 vào bảng phân tích phương sai ở bước 3 như bảng 7.8.

Bảng 7.8. Phân tích phương sai của tài liệu bảng 7.7

Nguồn biến động	Bậc tự do	Tổng bình phương	Bình phương Trung bình	F_{tn}	F bảng	
					0.05	0.01
Hàng (R)	3	0,030154	0,010051			
Cột (C)	3	0,827342	0,275781			
Công thức	3	0,426842	0,142281	6,59*	4,76	9,87
Sai số	6	0,129585	0,021598			
Toàn bộ	15	1,413923				

CV% = 11,0%

*ý nghĩa tại mức 5%.

Tính hiệu quả của hàng, cột và khối

Bước 1. Kiểm tra mức ý nghĩa khác nhau giữa khối hàng và cột.

A. Tính F để kiểm tra sự khác nhau của hàng và cột.

$$F_{(hang)} = \frac{MSR}{MSE} = \frac{0,010051}{0,021598} < 1$$

$$F_{(cot)} = \frac{MSC}{MSE} = \frac{0,275781}{0,021598} = 12,71$$

B. So sánh F_{tn} với F_{lt} .

- Với $F_{(hang)} < 1$ nên không có ý nghĩa của hàng.
- Với $F_{(cot)}$ bằng 12,77 lớn hơn F bảng với độ tự do $f_1 = 3$ và $f_2 = 6$ là 4,76 tại mức ý nghĩa 5% và 9,78 tại mức 1%. Nên sự khác nhau giữa các khối cột là có ý nghĩa ở mức 1%. Đó là sự thành công của việc tạo khối cột trong thí nghiệm này.

Bước 2. Tính hiệu quả tương đối của thiết kế LS so với CRD và RCB.

- So với CRD ta có.

$$R.E(CRD) = \frac{MSR + MSC + (t-1)MSE}{(t+1)MSE}$$

$$\text{Vậy } R.E(CRD) = \frac{0,010051 + 0,275781 + (4-1)(0,021598)}{(4+1)(0,021598)} = 3,25$$

Vậy tác dụng của kiểu LS là đã làm tăng độ chính xác của thí nghiệm lên 225%. Chứng tỏ nếu dùng kiểu CRD thì phải có 2,25 lần nhắc nữa mới đủ nói lên sự khác nhau giữa các công thức với cùng độ lớn như kiểu LS đã tìm ra.

- + So với RCB ta có 2 cách tính.

- Khi các hàng được coi là khối.

$$R.E(RCB, hang) = \frac{MSR + (t-1)MSE}{tMSE}$$

- Khi các cột được coi là khối

$$R.E(RCB, cot) = \frac{MSC + (t-1)MSE}{tMSE}$$

Khi độ tự do của sai số trong phân tích phương sai LS nhỏ hơn 20 giá trị R.E sẽ được nhân với số điều chỉnh K.

$$K = \frac{[(t-1)(t-2)+1][(t-1)^2+3]}{[(t-1)(t-2)+3][(t-1)^2+1]}$$

Theo ví dụ ta có

$$R.E(RCB, hang) = \frac{0,010051 + (4-1)(0,021598)}{4(0,021598)} = 0,87$$

$$R.E(RCB, cot) = \frac{0,275781 + (4-1)(0,021598)}{4(0,021598)} = 3,94$$

Do độ tự do của sai số trong phân tích phương sai LS chỉ bằng 6 nên ta có K như sau:

$$K = \frac{[(4-1)(4-2)+1][(4-1)^2+3]}{[(4-1)(4-2)+3][(4-1)^2+1]} = 0,93$$

và có các R.E điều chỉnh như sau:

$$R.E (RCB, hàng) = (0,87)(0,93) = 0,81$$

$$R.E (RCB, cột) = (3,94) (0,93) = 3,66.$$

Vậy khối cột đã làm tăng độ chính xác của thí nghiệm lên 266% khi đó khối hàng không làm tăng so với RCB. Vậy trong thí nghiệm này nếu dùng kiểu RCB với cột là khối thì hiệu quả cũng như kiểu LS.

7. 2. Thí nghiệm hai nhân tố

Các cơ thể sinh vật cùng một lúc chịu ảnh hưởng của nhiều yếu tố môi trường khác nhau. Phản ứng của cơ thể chúng với bất kỳ một yếu tố đơn lẻ nào cũng sẽ khác nhau trong các mức của nhân tố khác. Vậy thí nghiệm đơn yếu tố thường bị hạn chế khi áp dụng vào thực tế trong sự ảnh hưởng của các yếu tố khác. Hay nói cách khác kết quả của thí nghiệm một yếu tố chỉ có thể được áp dụng trong trường hợp các yếu tố khác được duy trì ở mức độ nhất định. Để giảm bớt khó khăn này, các thí nghiệm 2 hay nhiều yếu tố rất cần thiết.

Có nhiều kiểu sắp xếp thí nghiệm 2 hay nhiều yếu tố khác nhau.

7.2.1. Thiết kế theo kiểu khối ngẫu nhiên đầy đủ (RCB)

Thường sử dụng kiểu thiết kế này khi phối hợp toàn bộ các nhân tố thành các công thức tổ hợp.

Ví dụ có một thí nghiệm 2 nhân tố trong thiết kế kiểu RCB với 5 tỷ lệ đạm, 3 giống lúa và 4 lần nhắc lại.

Bước 1: Ký hiệu số nhắc lại là r, mức của nhân tố A(giống) bằng a, mức của nhân tố B(đạm) bằng b. Xây dựng bảng tổng quát cho phân tích phương sai như bảng 7.9

Bảng 7.9. Đề cương phân tích phương sai

Nguồn biến động	Độ tự do	Tổng bình phương	Trung bình bình phương	F tính	F bảng	
					5%	1%
Nhắc lại	r-1=3					
Công thức	ab-1=14					
Giống (A)	a-1=2					
Đạm(B)	b-1=4					
Tương tác A*B	(a-1)(b-1)=8					
Sai số	(r-1)(ab-1)=42					
Toàn bộ	rab-1=59					

Xây dựng sơ đồ phối hợp các nhân tố (bảng 7.10)

Bảng 7.10. Các tổ hợp công thức của 3 giống lúa và 5 mức đạm

Mức đạm (kg/ ha)	Giống lúa		
	V ₁	V ₂	V ₃
0 (N ₀)	N ₀ V ₁	N ₀ V ₂	N ₀ V ₃
40(N ₁)	N ₁ V ₁	N ₁ V ₂	N ₁ V ₃
70 (N ₂)	N ₂ V ₁	N ₂ V ₂	N ₂ V ₃
100 (N ₃)	N ₃ V ₁	N ₃ V ₂	N ₃ V ₃
130 (N ₄)	N ₄ V ₁	N ₄ V ₂	N ₄ V ₃

Sơ đồ sắp xếp các công thức ở hình 7.5 và năng suất các ô được ghi ở bảng 7.11

Bảng 7.11. Năng suất của 3 giống lúa thí nghiệm kết hợp với 5 mức đạm, thiết kế RCB

Năng suất hạt (tấn/ha)					
Mức đạm kg/ha	Nlai I	Nlai II	Nlai III	Nlai IV	Tổng các công thức (T _i)
V ₁					
N ₀	3.852	2.606	3.144	2.894	12.496
N ₁	4.788	4.936	4.562	4.608	18.894
N ₂	4.576	4.454	4.884	3.924	17.838
N ₃	6.034	5.276	5.906	5.652	22.868
N ₄	5.874	5.916	5.984	5.518	23.292
V ₂					
N ₀	2.846	3.794	4.108	3.444	14.192
N ₁	4.956	5.128	4.150	4.990	19.224
N ₂	5.928	5.698	5.810	4.308	21.744
N ₃	5.664	5.362	6.458	5.474	22.958
N ₄	5.458	5.546	5.786	5.932	22.722
V ₃					
N ₀	4.192	3.754	3.738	3.428	15.112
N ₁	5.250	4.582	4.896	4.286	19.014
N ₂	5.822	4.848	5.678	4.932	21.280
N ₃	5.888	5.524	6.042	4.756	22.210
N ₄	5.864	6.264	6.056	5.362	23.546
Tổng NLai (R)	76.992	73.688	77.202	69.508	
Tổng Toàn bộ					297.390

N Lại I				
V ₃ N ₂	V ₂ N ₁	V ₁ N ₄	V ₁ N ₁	V ₂ N ₃
V ₃ N ₀	V ₁ N ₃	V ₃ N ₄	V ₁ N ₂	V ₃ N ₃
V ₂ N ₄	V ₃ N ₁	V ₂ N ₀	V ₁ N ₀	V ₂ N ₂

N Lại II				
V ₂ N ₃	V ₃ N ₃	V ₁ N ₁	V ₂ N ₀	V ₂ N ₁
V ₁ N ₃	V ₃ N ₂	V ₁ N ₂	V ₁ N ₄	V ₂ N ₄
V ₁ N ₀	V ₃ N ₄	V ₂ N ₂	V ₃ N ₁	V ₃ N ₁

N Lại III				
V ₁ N ₁	V ₃ N ₀	V ₁ N ₀	V ₃ N ₁	V ₁ N ₄
V ₂ N ₂	V ₁ N ₂	V ₁ N ₃	V ₂ N ₄	V ₃ N ₄
V ₂ N ₀	V ₃ N ₂	V ₂ N ₁	V ₂ N ₃	V ₃ N ₃

N Lại IV				
V ₁ N ₂	V ₂ N ₂	V ₂ N ₄	V ₁ N ₀	V ₂ N ₀
V ₁ N ₃	V ₃ N ₁	V ₁ N ₄	V ₁ N ₁	V ₂ N ₃
V ₃ N ₀	V ₂ N ₁	V ₃ N ₂	V ₃ N ₃	V ₃ N ₄

Hình 7.5. Sơ đồ mẫu của thí nghiệm tổ hợp 5 mức đạm và 3 giống lúa, 4 lần nhắc lại trong thiết kế RCB

Bước 2. Tính tổng các công thức (T) các tổng nhắc lại (R) và tổng toàn bộ (G), như trong bảng 7.11. và tính các loại tổng bình phương: toàn bộ, nhắc lại, công thức và sai số theo phương pháp chung của thí nghiệm 1 nhân tổ bố trí kiểu RCB.

$$CF = \frac{G^2}{r.a.b} = \frac{297,390^2}{(4)(3)(5)} = 1474,014$$

$$SSTo = \sum x^2 - CF = [3,825^2 + 2,606^2 + \dots + 5,362^2] - 1474,014 = 53,530$$

$$SSR = \frac{\sum R^2}{a.b} - CF = \frac{76,992^2 + \dots + 69,508^2}{(3)(5)} - 1474,014 = 2,599$$

$$SST = \frac{\sum T^2}{r} - CF = \frac{12,496^2 + \dots + 23,546^2}{4} - 1747,140 = 44,578$$

$$SSE = SSTo - SSR - SST = 53,530 - 2,599 - 44,578 = 6,353$$

Các kết quả phân tích phương sai cơ bản ghi trên bảng 7.12

Bảng 7.12 Bảng phân tích phương sai cơ bản từ số liệu bảng 7.11

Nguồn biến động	Bậc tự do	Tổng bình phương	Bình phương Trung bình	F _{tn}	F bảng	
					5%	1%
Nhắc lại	3	2,599	0,866	5,74 ^{**}	2,83	4,21
Công thức	14	44,578	3,184	21,09 ^{**}	1,94	2,54
Sai số	42	6,353	0,511			
Toàn bộ	59	52,530				

Bước 3: Xây dựng bảng 2 chiều cho các tổng nhân tố A*B được ghi trong bảng 7.12

Bảng 7.13 Năng suất thể hiện qua giống và đạm (tính từ bảng 7.11)

Đạm	Tổng năng suất (AB)			Tổng đạm (B)
	V ₁	V ₂	V ₃	
N ₀	12,496	14,192	15,112	41,800
N ₁	18,894	19,224	19,014	57,132
N ₂	17,838	21,744	21,280	60,862
N ₃	22,868	22,985	22,210	68,036
N ₄	23,292	22,722	23,546	69,560
Tổng giống(A)	95,388	100,840	101,162	297,390

Bước 4: Tính 3 thành phần tổng bình phương như sau:

$$SSA = \frac{\sum A^2}{rb} - CF = \frac{95,388^2 + 100,840^2 + 101,162^2}{4*5} - 1474,014 = 1,052$$

$$SSB = \frac{\sum B^2}{r.a} - CF = \frac{41,800^2 + \dots + 69,560^2}{4*3} - 1474,014 = 41,234$$

$$SAB = SST - SSA - SSB = 44,578 - 1,052 - 41,234 = 2,292$$

Bước 5: Tính bình phương trung bình cho mỗi nguồn biến động

$$MSA = \frac{SSA}{a-1} = \frac{1,052}{2} = 0,526$$

$$MSB = \frac{SSB}{b-1} = \frac{41,234}{4} = 10,308$$

$$MSAB = \frac{SSAB}{(a-1)(b-1)} = \frac{2,292}{(2)*(4)} = 0,286$$

$$MSE = \frac{SSE}{(r-1)(ab-1)} = \frac{6,353}{(3)[(3)*(5)-1]} = 0,151$$

Bước 6: Tính giá trị F cho mỗi một trong 3 thành phần nhân tố như sau

$$F(A) = \frac{MSA}{MSE} = \frac{0,526}{0,151} = 3,48$$

$$F(B) = \frac{MSB}{MSE} = \frac{10,308}{0,151} = 68,26$$

$$F(AB) = \frac{MSAB}{MSE} = \frac{0,286}{0,151} = 1,89$$

Bảng 7.14. Bảng phân tích phương sai

Nguồn biến động	Độ tự do	SS	MS	F tính	F bảng	
Nhắc lại	3	2.599	0.866	5.74**	2.83	4.29
Công thức	14	44.578	3.184	21.09**	1.94	2.54
A	(2)	1.052	0.526	3.48*	3.22	5.15
B	(4)	41.234	10.308	68.26**	2.59	3.80
A*B	(8)	2.292	0.286	1.89 ^{ns}	2.17	2.96
Error	42	6.353	0.151			
Tổng	59	53.530				

$$CV\% = 7.8\%$$

Bước 7: Tính hệ số biến động

$$cv\% = \frac{\sqrt{MSE}}{TBT\text{toan}Bo} \times 100 = \frac{\sqrt{0,151}}{4,956} \times 100 = 7,8\%$$

Bước 8: Kết luận của bảng 6 cho thấy công thức nhân tố B khác nhau đã cho năng suất khác nhau ở độ tin 99%. Còn nhân tố A khác nhau chỉ khác nhau ở độ tin cậy 95%.

Bước 9: Tính giá trị sai khác nhỏ nhất có ý nghĩa: $LSD_{0,05}$

- Cho so sánh trung bình các công thức tổ hợp hai nhân tố: lấy các giá trị cột thứ 6 bảng 7.11 chia cho r

$$LSD_{0,05} = t_{0,05,42} * \sqrt{\frac{2MSE}{r}} = 2,021 * \sqrt{\frac{2*0,151}{4}} = 0,555$$

- Cho so sánh trung bình các giống qua các mức phân: lấy các giá trị dòng cuối cùng bảng 7.13 chia cho $r \cdot t_b$

$$LSD_{0,05} = t_{0,05,42} * \sqrt{\frac{2MSE}{r * t_b}} = 2,021 * \sqrt{\frac{2 * 0,151}{4 * 5}} = 0,248$$

- Cho so sánh trung bình các mức đạm qua các giống: lấy các giá trị cột 5 bảng 7.13 chia cho $r \cdot t_a$

$$LSD_{0,05} = t_{0,05,42} * \sqrt{\frac{2MSE}{r * t_a}} = 2,021 * \sqrt{\frac{2 * 0,151}{4 * 3}} = 0,321$$

7.2.2. Thiết kế kiểu chia ô lớn ô nhỏ (Split-plot)

Khái niệm

Kiểu thiết kế này cho thí nghiệm 2 nhân tố là phù hợp hơn kiểu khối ngẫu nhiên đầy đủ. Theo kiểu thiết kế Split- plot, một trong hai nhân tố được chỉ định vào ô chính hay ô lớn và được gọi là nhân tố ô chính. Ô chính được chia thành ô phụ cho nhân tố thứ hai gọi là nhân tố ô phụ.

Trong thiết kế này, độ chính xác số đo của nhân tố ô chính bị xem nhẹ để làm tăng độ chính xác của nhân tố ô phụ. Số đo hiệu quả của nhân tố ô phụ và tương tác của nó với nhân tố ô chính có độ chính xác cao hơn so với kết quả thu được khi thiết kế khối ngẫu nhiên đầy đủ. Mặt khác số đo hiệu quả của các công thức ô chính lại kém chính xác hơn so với kết quả thu được bằng cách thiết kế khối ngẫu nhiên đầy đủ.

Việc chỉ định một nhân tố nào đấy vào ô chính hoặc ô phụ là rất quan trọng. Một vài gợi ý sau đây có thể làm cơ sở cho việc tham khảo khi thiết kế.

(1) Độ chính xác:

Nếu nhân tố B yêu cầu độ chính xác cao hơn thì hãy chỉ định nhân tố B vào ô phụ và nhân tố A vào ô chính.

(2) Vấn đề hiệu quả:

Nhân tố nào có hiệu quả rõ hơn và dễ phát hiện ra hơn thì để ở ô chính còn nhân tố có hiệu quả không rõ bằng thì ấn định vào ô phụ. Như vậy nó sẽ tăng cơ hội để phát hiện hiệu quả khác nhau của nhân tố có hiệu quả kém hơn.

(3) Thực tiễn sản xuất yêu cầu:

Để thực hiện một biện pháp kỹ thuật nào đó phải tạo điều kiện cho nó được thực hiện ở ô lớn hơn để giảm sai số lân cận (tưới nước, làm đất..)

Trong thiết kế Split- plot cả hai phương pháp ngẫu nhiên hoá và phân tích phương sai đều được thực hiện theo hai giai đoạn: 1 cho mức ô chính và 1 cho mức ô phụ.

Có thể chỉ định nhân tố ô chính theo bất kỳ thiết kế khối ngẫu nhiên đầy đủ nào.

Quá trình ngẫu nhiên hoá và sơ đồ sắp xếp

Có 2 quá trình ngẫu nhiên hoá 1 cách tách biệt trong thiết kế Split- plot. Một quá trình cho ô chính và một cho ô phụ. Trong mỗi lần nhắc lại, các công thức ô chính được chỉ định ngẫu nhiên cho các mảnh chính trước, sau đó chỉ định ngẫu nhiên các công thức ô phụ trong mỗi mảnh chính.

Thí dụ có một thí nghiệm 2 nhân tố gồm 6 mức đạm (công thức ô chính) và 4 giống lúa (công thức ô phụ) với 3 lần nhắc lại.

Ký hiệu ô chính (nhân tố A) có 6 mức: a = 6; N₀ N₁ N₂ N₃ N₄ N₅

Ký hiệu ô phụ (nhân tố B) có 4 giống: b = 4; V₁ V₂ V₃ V₄

Ta có sơ đồ thiết kế như hình 7.6

N_4	N_3	N_1	N_0	N_5	N_2	N_1	N_0	N_5	N_2	N_4	N_3	N_0	N_1	N_4	N_5	N_3	N_2
V ₂	V ₁	V ₁	V ₂	V ₄	V ₃	V ₁	V ₄	V ₃	V ₁	V ₁	V ₃	V ₄	V ₃	V ₁	V ₂	V ₂	V ₁
V ₁	V ₄	V ₂	V ₃	V ₃	V ₂	V ₃	V ₁	V ₄	V ₂	V ₄	V ₂	V ₂	V ₄	V ₂	V ₃	V ₃	V ₄
V ₃	V ₂	V ₄	V ₁	V ₂	V ₁	V ₂	V ₂	V ₁	V ₄	V ₂	V ₄	V ₁	V ₁	V ₄	V ₂	V ₄	V ₂
V ₄	V ₃	V ₄	V ₄	V ₁	V ₄	V ₄	V ₄	V ₃	V ₂	V ₃	V ₁	V ₃	V ₂	V ₁	V ₄	V ₁	V ₃
Nhắc lại I						Nhắc lại II						Nhắc lại III					

Hình 7.6. Sơ đồ thiết kế thí nghiệm.

Phân tích phương sai

Phân tích được tiến hành theo ô chính và ô phụ từ một thí nghiệm 2 nhân tố (6 mức đạm và 4 giống lúa) theo hình trên với đạm là ô chính, giống là ô phụ. Kết quả năng suất được ghi trong bảng 7.15

Bảng 7.15. Năng suất của 4 giống lúa ở các mức đạm và nhắc lại khác nhau

Giống	Năng suất hạt kg/ha		
	N.lai I	N.lai II	N.lai III
	No (0kg N/ha)		
V ₁ (IR ₈)	4430	4478	3358
V ₂ (IR ₅)	3944	5314	3660
V ₃ (C ₄ -63)	3464	2944	3142
V ₄ (Peta)	4126	4482	4836
	N ₁ (60kgN/ha)		
V ₁	5418	5166	6423
V ₂	6502	5858	5586
V ₃	4768	6004	5556
V ₄	5192	4604	4652
	N ₂ (90kg N/ha)		
V ₁	6076	6420	6704
V ₂	6008	6127	6642
V ₃	6244	5724	6014
V ₄	4564	5744	4146
	N ₃ (120kgN/ha)		
V ₁	6462	7056	6680
V ₂	7139	6982	6564
V ₃	5792	5880	6730
V ₄	2774	5036	3638
	N ₄ (180kgN/ha)		

V ₁	7290	7548	7552
V ₂	7682	6659	6576
V ₃	7080	6662	6006
V ₄	1414	1960	5480
N ₅ (180kgN/ha)			
V ₁	8452	8832	8818
V ₂	6228	7387	6006
V ₃	5594	7122	5480
V ₄	2248	1380	2014

Bước 1: Xây dựng đề cương phân tích phương sai theo bảng 7.16.

Bảng 7.16. Đề cương phân tích phương sai của thí nghiệm kiểu Split-plot.

Nguồn biến động	Bậc tự do	Tổng bình phương	Bình phương trung bình	F _{tn}	F bảng	
					5%	1%
Nhắc lại	r-1=2					
Nhân tố ô chính (A)	a-1=5					
Sai số (a)	(r-1)(a-1)=10					
Nhân tố ô phụ (B)	b-1=3					
A*B	(a-1)(b-1)=15					
Sai số (b)	(r-1)(b-1)a=36					
Toàn bộ	rab-1=71					

Bước 2: Xây dựng 2 bảng của các tổng.

A. Bảng 2 chiều của các tổng: - nhắc lại x nhân tố A (RA)

- các tổng nhắc lại (R)
- các tổng nhân tố A (A)
- tổng toàn bộ (G)

được thể hiện ở bảng 7.17

Bảng 7.17. Bảng năng suất của đạm và nhắc lại cho các giống

Đạm	Tổng năng suất (RA)			Tổng đạm (A)
	Nhắc lại I	Nhắc lại II	Nhắc lại III	
N ₀	15964	17218	15488	48670
N ₁	21880	21623	22226	65738
N ₂	22874	24015	23506	70395

N ₃	22167	24954	23252	70373
N ₄	23466	23064	23214	69744
N ₅	22522	24721	22318	69561
Tổng nhắc lại (R)	128873	135604	130004	
Tổng toàn bộ (G)				394481

B. Bảng 2 chiều của các tổng cho nhân tố A và B, các tổng nhân tố B được tính trong bảng 7.18

Bảng 7.18. Bảng năng suất của đạm với giống (AB)

Đạm	Tổng năng suất (AB)			
	V ₁	V ₂	V ₃	V ₄
N ₀	12785	12918	9550	13444
N ₁	17016	17946	16328	14448
N ₂	19200	18777	17982	14436
N ₃	20198	20685	18042	11448
N ₄	22690	20852	20062	6140
N ₅	26102	19621	18196	5642
Tổng của giống (B)	117964	110799	100160	65558

Bước 3: Tính số hiệu chỉnh CF và các tổng bình phương cho các nhân tố ô chính như sau:

$$CF = \frac{G^2}{rab} = \frac{394481^2}{(3)(6)(4)} = 2161323047$$

$$SSTo = \sum x^2 - CF = (4430^2 + \dots + 2014^2) - 2161323047 = 204747916$$

$$SSR = \frac{\sum R^2}{ab} - CF = \frac{128873^2 + 135604^2 + 130004^2}{(6)(4)} - CF = 1082577$$

$$SSA = \frac{\sum A^2}{rb} - CF$$

$$= \frac{48670^2 + \dots + 69561^2}{(3)(4)} - CF = 30429200$$

$$SSE_{(a)} = \frac{\sum RA^2}{b} - CF - SSR - SSA$$

$$= \frac{15964^2 + \dots + 22318^2}{4} - CF - SSR - SSA = 1419678$$

Bước 4: Tính các tổng bình phương cho phân tích ô phụ như sau:

$$SSB = \frac{\sum B^2}{r \cdot a} - CF = \frac{117964^2 + \dots + 65558^2}{3 \cdot 6} - CF = 89888101$$

$$SSAB = \frac{\sum (AB)^2}{r} - CF - SSB - SSA$$

$$= \frac{12758^2 + \dots + 5642^2}{3} - CF - SSB - SSA = 69343487$$

Error(b) = SSTo - (SSR + SSA + SSE_(a) + SSB + SSAB)

$$= 204747916 - (1082577 + 30429200 + 1419678 + 89888101 + 69343487) = 12584873$$

Bước 5: Với mỗi nguồn biến động, tính bình phương trung bình bằng cách chia các SS cho độ tự do tương ứng của nó như sau:

$$MSR = \frac{SSR}{r - 1} = \frac{1082577}{2} = 541288$$

$$MSA = \frac{SSA}{a - 1} = \frac{30429200}{5} = 6085840$$

$$MSE(a) = \frac{SSE(a)}{(r - 1)(a - 1)} = \frac{1419678}{10} = 141967,8$$

$$MSB = \frac{SSB}{b - 1} = \frac{89888101}{3} = 29962700,3$$

$$MSAB = \frac{SSAB}{(a - 1)(b - 1)} = \frac{69343487}{15} = 4622899,1$$

$$MSE(b) = \frac{SSE(b)}{a(r - 1)(b - 1)} = \frac{12584873}{36} = 349580,1$$

Bước 6: Tính giá trị F_{tn} cho mỗi hiệu quả cần được thử nghiệm bằng cách chia mỗi trung bình bình phương cho trung bình bình phương sai số tương ứng nó.

$$F_m(A) = \frac{MSA}{MSE(a)} = \frac{6085840}{141967,8} = 42,87$$

$$F_m(B) = \frac{MSB}{MSE(b)} = \frac{29962700,3}{349580,1} = 85,71$$

$$F_m(A \times B) = \frac{MSAB}{MSE(b)} = \frac{4622899,1}{349580,1} = 13,22$$

So sánh $F_{tn}(A)$, $F_{tn}(B)$, $F_{tn}(A \times B)$ với F lý thuyết ở các mức xác suất để kết luận.

Bước 7: Tính 2 hệ số biến động: 1 tương ứng với ô chính, 1 tương ứng với ô phụ và tương tác của 2 nhân tố.

$$CV(a) = \frac{\sqrt{MSE(a)}}{TBT_0} \times 100 = \frac{\sqrt{141968}}{5479} \times 100 = 6,9\%$$

$$CV(b) = \frac{\sqrt{MSE(b)}}{TBT_0} \times 100 = \frac{\sqrt{3495808}}{5479} \times 100 = 10,8\%$$

Bước 8: Điền kết quả tính toán được vào đề cương phân tích phương sai đưa ra ở bước 1. Kết quả ghi trong bảng 7.19

Bảng 7.19. Kết quả phân tích phương sai:

Nguồn biến động	Bậc tự do	Tổng bình phương	Bình phương Trung bình	F _{tn}	F bảng	
					5%	1%
Nhắc lại	2	1082577	541228			
A	5	30429200	6085840	42,87**	3,33	5,64
Sai số (a)	10	1419678	141968			
B	3	89888101	29962700	85,71**	2,68	4,38
A*B	15	69343487	4622899	13,22**	1,96	2,58
Sai số (b)	36	12584873	349580			
Toàn bộ	71	204747916				

$$CV(a)=6,9\%, CV(b)= 10,8\%$$

Bước 9: So sánh trung bình các công thức

Ta phải tính các giá trị $LSD_{0,05}$ để làm căn cứ cho việc so sánh:

- So sánh các trung bình của các mức ô chính (đạm) được tính trung bình qua các mức của ô phụ (giống)

$$LSD_{0,05}(A) = t_{0,05,df(E_a)} * \sqrt{\frac{2 * MSE_a}{r * b}} = 2,228 * \sqrt{\frac{2 * 141968}{3 * 4}} = 342,7$$

Các trung bình lấy từ cột cuối cùng của bảng 7.17 chia cho (r*b)

- So sánh các trung bình của các mức ô phụ (giống) được tính qua tất cả các mức ô chính (đạm)

$$LSD_{0,05}(B) = t_{0,05,df(E_b)} * \sqrt{\frac{2 * MSE_{(b)}}{r * a}} = 2,021 * \sqrt{\frac{2 * 349580}{3 * 6}} = 398$$

Các trung bình lấy từ dòng cuối của bảng 7.18 chia cho (r*a)

- So sánh các trung bình của các tổ hợp cả hai nhân tố (AxB)

$$LSD_{0,05}(AxB) = t_{0,05,df(E_b)} * \sqrt{\frac{2 * MSE_b}{r}} = 2,021 * \sqrt{\frac{2 * 349580}{3}} = 975$$

Các trung bình lấy từ mỗi giá trị trong bảng 7.18 chia cho r để so sánh

7.3. Một số điều chú ý trước khi tiến hành phân tích phương sai

7.3.1. Khi số liệu thí nghiệm bị mất

Trong quá trình thí nghiệm, có thể có ô thí nghiệm nào đó bị mất năng suất hoặc vì lý do nào đó phải loại bỏ. Trong trường hợp phân tích phương sai một nhân tố bố trí kiểu hoàn toàn ngẫu nhiên, việc mất số liệu không ảnh hưởng gì vì ta có thể phân tích phương sai cho thí nghiệm 1 nhân tố khi số lần nhắc lại không bằng nhau. Nhưng khi phân tích phương sai hai nhân tố thì mất một số liệu là mất tính trực giao, lúc đó kết quả phân tích sẽ rất kém chính xác. Vì vậy, nếu mất một hoặc một số số liệu ta phải thay chúng bằng các số khác sao cho tổng bình phương sai số đạt nhỏ nhất.

7.3.1.1 Khi bị mất một số liệu

Ví dụ: Có một thí nghiệm bố trí kiểu khối ngẫu nhiên đầy đủ gồm 3 công thức và 4 lần nhắc lại, kết quả ghi lại như sau (bảng 7.20)

Bảng 7.20. Năng suất thí nghiệm qua các lần nhắc lại

Công thức	Các lần nhắc lại				Tổng T_i
	1	2	3	4	
1	362	311	205	165	1043
2	550	489	x	341	1380
3	562	508	375	289	1734
Tổng của nhắc lại			580		$\Sigma x=4157$

Kết quả cho thấy ở công thức 2, lần nhắc lại 3 có năng suất bị mất được ghi bằng chữ x.

Để cho việc tính toán được thuận tiện, ta thay các giá trị từ bảng 7.20 sang bảng 7.21 với $X_1 = x - 377$ (377 là trung bình thực tế toàn thí nghiệm)

Bảng 7.21

Công thức	$X_1 = x - 377$				Tổng T_i
	1	2	3	4	
1					
2	-15	-66	-172	-212	-465
3	173	112	X_1	-36	$249 + X_1$
	185	131	-2	-88	226
Tổng Nhắc lại (R)	343	177	$X_1 - 174$	-336	$X_1 + 10$

Vì ta có:

$$SSE = SST_o - SST - SSR$$

$$= \sum X_1^2 - \frac{\sum T_i^2}{r} - \frac{\sum R_j^2}{t} + \frac{(\sum X)^2}{t * r}$$

Vậy:

$$SSE = ((-15)^2 + (-66)^2 + \dots + (-88)^2) - ((-465)^2 + (249 + X_1)^2 + 226^2) : 4 -$$

$$(343^2 + 177^2 + (X_1 - 174)^2 + (-336)^2) : 3 + \frac{(X_1 + 10)^2}{4 * 3}$$

$$= X_1^2 + 182012 - \frac{(267301 + (249 + X_1)^2)}{4} - \frac{(262561 + (X_1 - 174)^2)}{3} + \frac{(X_1 + 10)^2}{12}$$

Để tìm giá trị của X_1 cho SSE nhỏ nhất ta lấy đạo hàm theo X_1 rồi cho bằng 0

$$\frac{\partial SSE}{\partial X_1} = 2X_1 - \frac{2(X_1 + 249)}{4} - \frac{2(X_1 - 174)}{3} + \frac{2(X_1 + 10)}{12} = 0$$

$$12X_1 - 82 = 0$$

$$X_1 = 6,8 = 7$$

Như vậy năng suất ô bị mất $X = 377 + 7 = 384$.

Thay thế năng suất bị mất vào bảng 7.20 hoặc 7.21 để phân tích phương sai thông thường. Nhưng khi tính độ tự do cho sai số ngẫu nhiên phải bớt đi 1 vì tổng số ô quan sát chỉ có 11 ô.

Ta cũng có thể tìm giá trị của số liệu bị mất bằng công thức 7.1 sau:

$$X = \frac{t * T + r * R - \sum x}{(t - 1) * (r - 1)} \quad (7.1)$$

Trong đó:

X: Năng suất ô bị mất cần tìm

t: số công thức

r: số lần nhắc lại

T: tổng năng suất của các lần nhắc lại ở công thức bị mất

R: tổng năng suất của các công thức ở lần nhắc lại bị mất

$\sum x$: tổng năng suất thí nghiệm trừ ô bị mất.

áp dụng công thức trên để tìm năng suất ô bị mất của bảng 7.20 ta có:

$$X = \frac{3 * 1380 + 4 * 580 - 4157}{(3 - 1) * (4 - 1)} = 383,8 = 384$$

Kết quả tính cũng phù hợp với cách tính trên đây.

7.3.1.2 Khi bị mất 2 ô số liệu

Ví dụ: có một thí nghiệm gồm 4 công thức và 4 lần nhắc lại được bố trí kiểu RCB, kết quả năng suất được ghi trong bảng 7.22

Bảng 7.22 Năng suất của các công thức qua các lần nhắc lại

Công thức	Năng suất của các lần nhắc lại				Tổng T_i
	1	2	3	4	
1	60,4	64,4	65,2	59,6	249,6
2	66,8	67,5	70,0	Y	204,3
3	56,3	61,5	66,7	51,1	253,6
4	42,9	X	50,1	45,2	138,2
Tổng Rep_j	226,4	193,4	252,0	155,9	827,7

Trước khi tính sản lượng của 2 ô bị mất, ta phải giả định rằng: thí nghiệm chỉ mất một ô và khi tính lần thứ nhất: bằng cách lấy sản lượng trung bình toàn thí nghiệm thay vào một ô bị mất để tính ô còn lại.

Bước 1:

$$\text{Ta giả định thay: } Y = \frac{\sum X}{N - 2} = \frac{827,7}{16 - 2} = 59,1$$

áp dụng công thức (7.1) để tính X cho bước một ta có:

$$X_1 = \frac{t * T + r * R - \sum x}{(t - 1) * (r - 1)} = \frac{4 * 138,2 + 4 * 193,4 - (827,7 + 59,1)}{(4 - 1) * (4 - 1)} = 48,8$$

Thay giá trị X_1 vào để tính Y theo công thức (7.1):

$$Y_1 = \frac{t * T + r * R - \sum x}{(t - 1) * (r - 1)} = \frac{4 * 204,3 + 4 * 155,9 - (827,7 + 48,8)}{(4 - 1) * (4 - 1)} = 62,7$$

Bước 2: Tính X_2 và Y_2 thông qua X_1 và Y_1 bằng công thức (7.1) ta có:

$$X_2 = \frac{4 * 138,2 + 4 * 193,4 - (827,7 + 62,2)}{(4 - 1) * (4 - 1)} = 48,5$$

$$Y_2 = \frac{4 * 204,3 + 4 * 155,9 - (827,7 + 48,5)}{(4 - 1) * (4 - 1)} = 62,7$$

Nếu như cứ tiếp tục tính X, Y trong nhiều lần nữa thì ta sẽ được các trị số X và Y tiến dần đến một số ổn định, số ổn định đó chính là trị số cần tìm. Nhưng ở đây không phải là tìm đến một số hoàn toàn chính xác. Qua kinh nghiệm cho thấy có thể dùng các giá trị X_2 và Y_2 thay thế cho số liệu bị mất để tiến hành tính phân tích phương sai. Cũng như đã nói ở trên, trong trường hợp thí nghiệm bị mất 2 ô thì độ tự do của sai số ngẫu nhiên phải trừ đi 2.

Một điều cần chú ý là các năng suất phục hồi được dùng để phân tích phương sai chứ không dùng để tính năng suất bình quân.

7.3.2. Chuyển đổi số liệu trước khi phân tích phương sai

Toàn bộ phân tích phương sai dựa trên giả thiết các sai số ngẫu nhiên phân phối chuẩn $N(0, \sigma^2)$ với phương sai bằng nhau. Để đảm bảo giả thiết này, trong một số trường hợp phải thực hiện việc đổi biến. Có 4 cách đổi biến thường dùng:

- Khi trong dãy các số liệu quan sát có những trị số bằng 0, thì giá trị quan sát thực x phải được đưa về dạng lg x hoặc lg (x+1).
- Khi trong dãy các số liệu quan sát có những trị số gần bằng 0 hay là không có ý nghĩa, thì giá trị quan sát thực X phải đưa về dạng \sqrt{x} hoặc $\sqrt{x+1}$
- Khi dãy số liệu quan sát là những tỷ lệ như tỷ lệ sâu bệnh, tỷ lệ chết, thì giá trị quan sát thực phải chuyển về dạng góc $\arcsin \sqrt{Tyle}$ như bảng 8 phần phụ lục.

Ví dụ: Nghiên cứu tỷ lệ chết của hạt cao lương khi hạt phần bị bệnh ở 8 giống, kết quả được ghi lại trong bảng (7.23).

Bảng 7.23. Tỷ lệ chết của hạt cao lương khi hạt phần bị bệnh

Công thức (giống)	Giá trị quan sát (x)			Biến mới (X ₁)			Tổng T _i	Trung bình	Tỷ lệ T.bình	Đánh giá
	I	II	III	I	II	III				
1(Đ/c)	64,6	66,7	69,4	53,5	54,8	56,4	164,7	54,9	67,0	Đ/c
2	67,0	64,2	68,0	54,9	53,2	55,6	163,7	54,6	66,5	ns
3	1,0	0,4	0,5	5,7	3,6	4,0	13,3	4,4	0,6	*
4	12,2	10,6	13,3	20,4	19,0	21,4	60,8	20,3	12,0	*
5	2,7	1,4	2,4	9,5	6,8	8,9	25,2	8,4	2,1	*
6	63,8	62,2	59,6	53,0	52,1	50,5	155,6	51,9	62,0	*
7	1,2	1,0	0,8	6,3	5,7	5,1	17,1	5,7	1,0	*
8	55,8	55,4	52,8	48,3	48,1	46,6	145,0	47,7	54,7	*
Tổng Nhắc lại				251,6	243,3	248,5	743,4= ΣX ₁	31,2= X ₁	27,0= X	

Các giá trị quan sát thực x được đổi sang $X_1 = \arcsin \sqrt{Tyle}$. Khi biết tỷ lệ P căn cứ vào bảng tính sẵn ta có giá trị x₁ như bảng (7.23). Các bước tính toán tiếp theo ứng với các giá trị X₁ như đã trình bày trong các ví dụ trên.

Phương pháp tính cụ thể:

$$N = t \cdot r = 8 \cdot 3 = 24$$

$$CF = (\Sigma X_1)^2 : N = (743,4)^2 : 24 = 23028,82;$$

$$SSTo = \Sigma X_1^2 - CF = (53,5^2 + 54,8^2 + \dots + 46,6^2) - 23028,82 = 11457,22$$

$$SSR = \Sigma Rep^2 : t - CF = (251,6^2 + 243,3^2 + 248,5^2) : 3 - 23028,82 = 4,40$$

$$SST = \Sigma T_i^2 : r - CF = (164,7^2 + 163,7^2 + \dots + 143,0^2) : 8 - 23028,82 = 11.434,89$$

$$SSE = SSTo - SSR - SST = 11457,22 - 11434,89 - 4,40 = 17,93$$

Lập bảng phân tích phương sai

Nguồn B.động	Tổng bình phương (SS)	Bậc tự do (df)	Phương sai (MS)	F _{tn}	F _{.05}
Toàn bộ	11457,22	23			
Nhắc lại	4,40	2			

Công thức	11434,89	7	1633	1276	2,77
Sai số	17,93	14	1,28		

$$SE = \sqrt{\frac{MSE}{r}} = \sqrt{\frac{1,28}{3}} = 0,65$$

$$LSD_{.05} = t_{.05;df=14} * \sqrt{\frac{2MSE}{r}} = 2,15 * \sqrt{\frac{2 * 1,28}{3}} = 2,0$$

Kết luận: Tỷ lệ chết của hạt cao lương khi hạt phần bị bệnh của các giống 3, 4, 5, 6, 7, 8 khác với đôi chứng một cách có ý nghĩa ở độ tin 95%(đánh dấu bằng một dấu sao). Giống số 2 khác không có nghĩa (đánh dấu bằng ns).

Chú ý:

Trong quá trình phân tích phương sai ở trên ta đã dùng một sai số chung cho tất cả các công thức. Việc này chỉ được thực hiện khi phương sai của các công thức tương tự nhau. Vậy để áp dụng phương pháp phân tích phương sai được chính xác ta cần kiểm tra xem các phương sai của các công thức có tương tự nhau không?

Để kiểm tra ta dùng tiêu chuẩn Bartlet theo công thức sau:

$$\chi^2 = 2,3026 \left[\log \bar{S}^2 * \sum_{i=1}^t f_i - \sum_{i=1}^t (f_i \log S_i^2) \right]$$

Trong đó: 2,3026 là một hệ số không đổi

f_i là bậc tự do tương ứng từng công thức

$f_i = (r_i - 1)$

r_i là số lần nhắc lại

\bar{S}^2 là phương sai trung bình

$$\bar{S}^2 = \frac{\sum_{i=1}^t S_i^2}{t} \text{ Trong đó: } t \text{ là số công thức thí nghiệm}$$

S_i^2 là phương sai của từng công thức

$$S_i^2 = \frac{\sum (x_i - \bar{x})^2}{r_i - 1}$$

χ^2_{tn} được so với χ^2 lý luận (bảng 3 phụ lục) ứng với mức xác suất ý nghĩa α và bậc tự do $df = t-1$.

Nếu χ^2_{tn} lớn hơn χ^2 lý luận, ta kết luận là phương sai giữa các công thức khác nhau không cho phép áp dụng phương pháp phân tích phương sai

Nếu χ^2 nhỏ hơn hoặc bằng χ^2 lý luận, ta kết luận phương sai giữa các công thức khác nhau là tương tự và cho phép áp dụng được phương pháp phân tích phương sai để đánh giá kết quả thí nghiệm.

Ví dụ: dùng tiêu chuẩn Bartlett để kiểm tra xem có thể dùng phương pháp phân tích phương sai để đánh giá thí nghiệm sau đây được không?

Bảng 7.24. Năng suất các ô thí nghiệm so sánh hiệu lực các dạng phân lân

Công thức	Năng suất ở các lần nhắc lại(kg/ô)				
	1	2	3	4	5
Nền (I)	20,7	20,2	22,1	21,0	21,3
Nền + Supe lân (II)	31,6	36,8	25,3	21,7	36,5
Nền + Apatit (III)	27,5	31,2	30,5	29,6	31,7
Nền + Photphorit(IV)	28,7	29,3	30,6	28,5	29,1

Để kiểm tra phải lập bảng tính phương sai cho từng công thức. Kết quả tính toán được ghi lại trong bảng (7.25) để tính tiêu chuẩn Bartlett.

Bảng 7.25 Dùng tiêu chuẩn Bartlett để đánh giá sự không đồng đều của các phương sai trong thí nghiệm so sánh hiệu lực của các dạng phân lân

Công thức	Phương sai S_i^2	Độ tự do	Log của phương sai ($\log S_i^2$)	$f_i \log S_i^2$
Nền (I)	0,505	4	1,7033	2,8132
Nền + Supe lân (II)	45,330	4	1,6561	6,6244
Nền + Apatit (III)	4,180	4	0,6212	2,4848
Nền + Photphorit(IV)	0,680	4	1,8325	1,3300
Cộng	50,695	16		6,5524

$$\bar{S}^2 = 12,67$$

$$\log 12,67 = 1,1038$$

$$\chi^2 = 2,3026 (16 * 1,1038 - 6,5524)$$

$$= 2,3026 * 11,1084$$

$$= 25,578$$

$$\chi^2_{0,05, 3} = 7,82$$

Vậy, phương sai giữa các công thức là khác nhau rõ rệt, do vậy không thể dùng phương pháp phân tích phương sai để đánh giá kết quả thí nghiệm trên được.

Trường hợp phương sai giữa các công thức không tương tự nhau ta có thể dùng một trong các phương pháp sau đây:

Dùng phương pháp phân đoạn để phân tích kết quả thí nghiệm.

Cơ sở lý luận của phương pháp này là so sánh các trị số bình quân của các công thức. Ta có thể tìm hiểu nội dung của phương pháp thông qua ví dụ về so sánh hiệu lực của các dạng phân lân (bảng 7.26)

Các trị số t tính ở bảng được so sánh với trị số t lý luận:

$$t_{\alpha, df = r1+r2 - 2} = t_{\alpha, df = 2r - 2}$$

Trong đó r là số lần nhắc lại

$$t_{0,05, df=8} = 2,306$$

- Qua kết quả tính trên bảng ta có một số nhận xét:

Công thức II hệ số biến động qua các lần nhắc quá lớn so với các công thức khác vì vậy cần xem xét kỹ số liệu của công thức này.

- Các công thức II, III, và IV có năng suất hơn hẳn công thức I
- Giữa các dạng lần năng suất không khác nhau.

Bảng 7.26. quá trình phân tích kết quả thí nghiệm

Công thức	x_i	\bar{x}	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$S^2_{\bar{x}}$	CV%	T_{in}
I. Nền	20,7	21,1	0,4	0,16	0,710	3,4	$t_{I/II}=3,07$ $t_{I/III}=9,45$ $t_{I/IV}=17,6$
	20,2		0,9	0,81			
	22,1		1,0	1,00			
	21,0		0,1	0,01	0,101		
	21,3		0,2	0,04			
Cộng	105,3			2,02			
II. Nền + S _p	31,6	30,3	1,30	1,69	6,70	22	$t_{II/III}=0,01$ $t_{II/IV}=0,37$
	36,8		6,50	42,25	9,06		
	25,3		5,00	25,00			
	21,7		8,60	73,96			
	36,5		6,20	38,44			
Cộng	151,9			181,34			
III Nền + A _p	27,5	30,1	3,40	11,56	2,044	6,8	$t_{III/IV}=1,0$
	31,2		1,10	1,21	0,830		
	30,5		0,40	0,16			
	29,6		0,50	0,25			
	31,7		1,60	2,56			
Cộng	150,5			16,74			
IV Nền + P _p	28,7	29,3	0,50	0,25	0,824	2,8	
	29,3		0,10	0,01	0,136		
	30,6		1,40	1,96			
	28,5		0,70	0,49			
	29,1		0,10	0,01			
Cộng	146,2			2,72			

7.3.3. Phương pháp so sánh đa biên độ của Dun can

Để khắc phục một số nhược điểm của phương pháp so sánh các trung bình theo tiêu chuẩn LSD, D.B. Duncan đề nghị dùng phương pháp đa biên độ để đánh giá sự sai khác của các trung bình của công thức.

Các bước phân tích kết quả thí nghiệm theo phương pháp đa biên độ về căn bản là giống với phương pháp phân tích phương sai như đã trình bày, chỉ có khác là xác định giới hạn sai khác nhỏ nhất có ý nghĩa không phải chỉ có một giới hạn chung cho mỗi cặp như LSD mà dùng nhiều giới hạn dựa trên biên độ đáng kể trong bảng tính sẵn của Duncan theo các bước cụ thể sau:

Sắp xếp năng suất bình quân của các công thức theo thứ tự tăng dần.

Dùng phương pháp phân tích phương sai để tính phương sai ngẫu nhiên toàn thí nghiệm (MSE).

- Tính độ lệch chuẩn của các hiệu sai bình quân

$$S_{\bar{d}} = \sqrt{\frac{2MSE}{r}} \text{ hoặc } S_{\bar{d}} = \sqrt{MSE \frac{r_1 + r_2}{r_1 * r_2}}$$

- Căn cứ vào mức tin cậy α và bậc tự do ứng với bậc tự do của phương sai ngẫu nhiên và khoảng cách $p = 2, 3, t$ để tìm $(t-1)$ biên độ đáng kể (r_p) trong bảng D. B. Duncan bảng 11 phụ lục.
- Tính biên độ nhỏ nhất có ý nghĩa theo Duncan R_p (*shortest significant ranges*) như sau:

$$R_p = \frac{(r_p)(S_{\bar{d}})}{\sqrt{2}} \text{ với } p = 2, 3, t$$

Để minh họa cho cách tính này, ta dùng ví dụ đã nêu trên đây (bảng 7.24) có năng suất bình quân được xếp theo thứ tự tăng dần như sau:

Công thức	1	4	3	2
Năng suất	21,1	29,2	30,1	30,3

Các giá trị từ bảng 7.24. được chuyển sang biến mới như bảng 7.27 như sau

Bảng 7.27 Năng suất theo biến mới $X_i = x_i - 27,6$

Công thức	Nhắc lại					T_i
	1	2	3	4	5	
I	-6,9	-7,4	-5,5	-6,6	-6,3	-32,7
II	4,0	9,2	-2,1	-5,9	8,9	14,1
III	-0,1	3,6	2,9	2,0	4,1	12,5
IV	1,1	1,9	3,0	0,9	1,5	8,4
Tổng N.Lại	-1,9	7,3	-1,8	-9,6	8,2	$\sum X_{ij}=2,3$

Tổng số ô thí nghiệm: $N = t * r = 4 * 5 = 20$

$CF = (\sum X_{ij})^2 : N = (2,3)^2 : 20 = 0,264$

$$SSTo = \sum x_{ij}^2 - CF = (6,9^2 + 7,4^2 + \dots + 1,5^2) - 0,264 = 493,736$$

$$SST = \sum T_i^2 : r - CF = \frac{32,72^2 + 14,1^2 + 12,5^2 + 8,4^2}{5} - 0,264 = 298,60$$

$$SSR = \sum R_j^2 : t - CF = \frac{1,9^2 + 7,3^2 + \dots + 8,2^2}{4} - 0,264 = 54,62$$

$$SSE = 493,736 - 298,6 - 54,62 = 140,52$$

Lập bảng phân tích phương sai

Nguồn B.động	Tổng bình phương (SS)	Bậc tự do (df)	MS	F _{tn}	F _{0,05}
Toàn bộ	493,74	19			
Nhắc lại	54,62	4			
Công thức	298,60	3	99,5	8,5	3,49
Sai số	140,52	12	11,7		

Kết luận: ảnh hưởng của các công thức khác nhau đến năng suất là rõ

Tính độ lệch chuẩn của các hiệu sai

$$S_{\bar{a}} = \sqrt{\frac{2MSE}{r}} = \sqrt{\frac{2 * 11,7}{5}} = 1,673$$

ở mức tin cậy $\alpha = 0,05$ và $df_E = 12$ (Bậc tự do của phương sai ngẫu nhiên), theo bảng Duncan ta có:

p	2	3	4
Biên độ đáng kể (r _p)	3,08	3,23	3,31

Tìm biên độ nhỏ nhất có ý nghĩa (R_p) bằng cách:

Lấy biên độ đáng kể nhân với S_đ

p	2	3	4
R _p	3,6	3,8	3,9

Kiểm tra sự sai khác giữa các công thức

- So sánh giữa 2 với 1: $30,3 - 21,1 = 9,2 > 3,9$
- So sánh giữa 2 với 4: $30,3 - 29,2 = 1,1 < 3,8$
- So sánh giữa 2 với 3: $30,3 - 30,1 = 0,2 < 3,6$
- So sánh giữa 3 với 1: $30,1 - 21,1 = 9 > 3,8$
- So sánh giữa 3 với 4: $30,1 - 29,2 = 0,9 < 3,6$

Kết luận:

- Ta thấy có sự khác nhau đáng kể giữa các công thức 2,3,4 với công thức 1 nghĩa là bốn các dạng lân khác nhau giữa các công thức đều cho kết quả tốt khác nhau.
- Không thấy có sự sai khác giữa các dạng lân với nhau.

Điều này phù hợp với kết luận ở phương pháp LSD và đem lại kết quả nhanh hơn, chính xác hơn.

Đến đây ta cũng có thể biểu diễn hình học kết quả so sánh và công bố kết quả như các ví dụ trên đã làm.

Các câu hỏi ôn tập:

1. Hãy nêu các cách thiết kế thí nghiệm và cho biết trong trường hợp nào cần thiết kế kiểu nào? Cho ví dụ minh họa?
2. Bằng các ví dụ cụ thể hãy vẽ sơ đồ thiết kế đầy đủ của mỗi kiểu thiết kế?
3. Nêu cách phân tích phương sai tổng quát cho mỗi trường hợp thiết kế?
4. Nêu các cách so sánh các trung bình các công thức cho mỗi trường hợp thiết kế?
5. Nêu cách công bố kết quả cho mỗi trường hợp thiết kế?
6. Tự tìm trong thực tiễn (các kết quả đã nghiên cứu trước) cho mỗi trường hợp thiết kế một ví dụ cụ thể và hãy phân tích phương sai chúng.

CHƯƠNG VIII PHÂN TÍCH TƯƠNG QUAN VÀ HỒI QUI

Mục tiêu

Giúp cho học sinh thấy được ý nghĩa cần thiết của phân tích tương quan và hồi qui về mặt lý luận cũng như thực tiễn. Sau khi nghiên cứu chương này, người học phải biết được cách tính hệ số tương quan, biết cách đánh giá ý nghĩa của hệ số tương quan, cách lập phương trình hồi qui tuyến tính đơn (một biến số), biết cách kiểm tra độ tin cậy của phương trình và cách ứng dụng vào thực tiễn sản xuất nông nghiệp như thế nào.

Nội dung bao gồm:

1. Khái niệm về tương quan và hồi qui
2. Cách xác định dạng tương quan
3. Các cách tính hệ số tương quan, cách đánh giá hệ số tương quan
4. Cách xây dựng phương trình hồi qui tuyến tính đơn, kiểm tra độ tin cậy của phương trình mới xây dựng được và minh họa ứng dụng của chúng trong thực tiễn sản xuất nông nghiệp.

8.1. Đặt vấn đề

Trong thiên nhiên mọi hiện tượng và sự vật không phải độc lập mà liên quan với nhau khiến cho cả vũ trụ là một chỉnh thể thống nhất. Trong lĩnh vực sinh học cũng vậy giữa đám đông sinh vật và trong từng cá thể sinh vật trong quá trình phát sinh phát triển và tồn tại luôn luôn có sự liên quan và ràng buộc lẫn nhau giữa sinh vật và môi trường, giữa cá thể sinh vật với nhau và giữa các bộ phận trong từng cá thể.

Nhiệm vụ của con người là phát hiện ra mối liên hệ khách quan ấy. C.Mác đã nói "không nên tạo sự liên hệ để vận dụng nó, mà phát hiện sự liên hệ từ bên trong sự vật rồi chứng minh sự liên hệ đó trong thực tiễn, vì đó là điều có thể chứng minh".

Tính quy luật trong thế giới khách quan là do sự liên hệ và tác động qua lại giữa các sinh vật và hiện tượng mà có và chính sự tác động ấy đã dẫn đến sự vận động và phát triển không ngừng của thế giới tự nhiên. Bằng phương pháp toán học ta cũng có thể tìm ra hình thức liên hệ giữa các hiện tượng và đánh giá được mức độ của sự liên hệ đó. Vì vậy, phân tích tương quan có thể giúp chúng ta dựa vào một đặc trưng hoặc một số đặc trưng nào đó để đoán ra một đặc trưng khác và cũng nhờ phân tích tương quan như vậy giúp ta phát hiện ra được quy luật của sinh vật để hướng sự phát triển của chúng theo chiều hướng có lợi cho con người.

Vậy quan hệ tương quan là gì?

Như trên đã nói: trong thiên nhiên có nhiều hiện tượng có liên quan lẫn nhau, song sự liên hệ đó không giống như sự liên hệ hàm số. Trong liên hệ hàm số: với một giá trị của biến số độc lập ta có thể xác định được một trị số của biến số phụ thuộc tương ứng.

Thí dụ: từ đường kính của đường tròn ta có thể xác định được diện tích và chu vi của nó. Trái lại trong quan hệ tương quan ứng với một trị số của biến số độc lập có thể có nhiều trị số của biến số phụ thuộc mà ta chỉ có thể lấy trị số bình quân của những trị số của biến số phụ thuộc làm đại diện tương ứng với biến số độc lập.

Vì vậy thực chất của quan hệ tương quan là quan hệ giữa một bên là biến số độc lập và một bên là số trung bình của những trị số của biến số phụ thuộc.

Quan hệ giữa biến số độc lập và số trung bình của những trị số của biến số phụ thuộc là quan hệ tương quan.

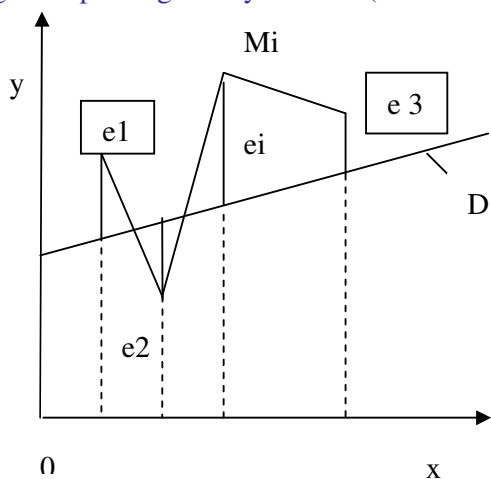
Phương trình toán học biểu thị mối quan hệ đó gọi là phương trình hồi quy; cho nên nhiệm vụ đầu tiên của phân tích tương quan là xác định các tham số của phương trình hồi quy.

Như trên đã nói: mỗi biến số độc lập có thể có nhiều trị số của biến số phụ thuộc mà đại diện là số trung bình của chúng. Nếu các trị số đó phân bố càng tập trung quanh trị số trung bình thì biểu thị mức độ liên hệ các biến số càng chặt chẽ. Do đó nhiệm vụ thứ hai của phân tích tương quan là xác định mức độ liên hệ giữa các hiện tượng

8.2. Tương quan tuyến tính đơn.

8.2.1. Khái niệm và các đặc trưng của tương quan

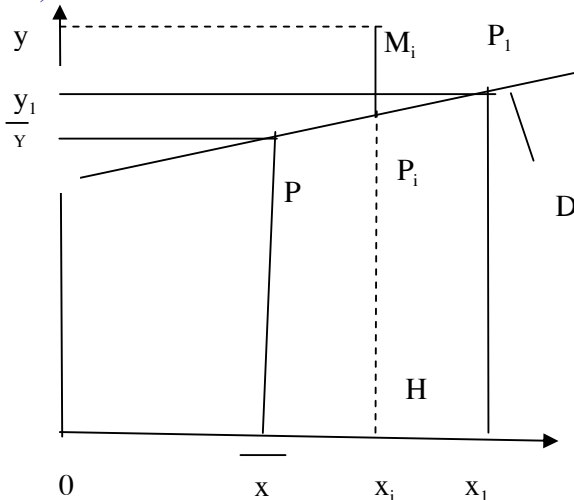
Giả sử ta có một đám mây toạ độ $M_i(X_i, y_i)$, đám mây có thể được đại diện bằng đường thẳng D có phương trình $y = ax + b$ (hình 8.1 và 8.2).



Hình 8.1

Trên hình 8.1. Mỗi điểm M_i có độ lệch e_i đối với D ;

Trên hình 8.2. Độ lệch e_i là đoạn M_iP_i .



Hình 8.2

$$M_iP_i = M_iH - P_iH$$

$$e_i = Y_i - (ax_i + b)$$

Vấn đề đặt ra là xác định đường D (nghĩa là ta tính a và b) thế nào cho tổng các bình phương độ lệch e_i là nhỏ nhất. Đường tìm ra được là đường thẳng các bình phương tối thiểu và phương pháp tính toán gọi là phương pháp bình phương tối thiểu.

Nếu ta gọi Q_{yx} là tổng bình phương các độ lệch từ các điểm toạ độ đến đường D theo hướng trục y thì:

$$Q_{yx} = \sum_{i=1}^n \{y_i - (ax_i + b)\}^2$$

Trong đó: n là dung lượng mẫu quan sát.

Như vậy Qyx là một hàm số của a và b.

$$Qyx = f(a, b)$$

Để cho đường D đại diện cho các điểm tọa độ thì phải làm cho

$$Qyx = \sum_{i=1}^n \{y_i - (ax_i + b)\}^2 \quad \text{đạt giá trị cực tiểu.}$$

Muốn cho $yx=f(a,b)$ là cực tiểu thì điều kiện cần là cho đạo hàm riêng theo a, b bằng không

$$\frac{\partial Qyx}{\partial a} = 0 \quad \text{và} \quad \frac{\partial Qyx}{\partial b} = 0$$

Như vậy:

$$\frac{\partial Qyx}{\partial a} = -2 \sum_{i=1}^n (y_i - ax_i - b)x_i = 0 \quad (1)$$

$$\frac{\partial Qyx}{\partial b} = -2 \sum_{i=1}^n (y_i - ax_i - b) = 0 \quad (2)$$

Từ (1) và (2) ta lập được hệ phương trình tuyến tính đối với hai ẩn số a và b

$$\sum x_i y_i = a \sum X_i^2 + b \sum X_i \quad (3)$$

$$\sum y_i = a \sum X_i + nb \quad (4)$$

Sau khi rút gọn hệ phương trình trên ta có:

$$a = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \quad (8-1)a$$

$$a = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} \quad (8-1)b$$

Sau khi đã tính được a theo công thức (8-1) ta có thể tính b theo công thức:

$$b = \bar{y} - \bar{x}a \quad (8-2)$$

Đường D đi qua điểm p có tọa độ (\bar{x}, \bar{y}) trên hình (8-2) hệ số góc a của đường D bằng:

$$a = \frac{P_1N}{PN} = \frac{y_1 - \bar{y}}{x_1 - \bar{x}}$$

Do đó phương trình đường D được thiết lập qua công thức:

$$y - \bar{y} = a(x - \bar{x}) \quad (8-3)$$

Sau khi thiết lập được phương trình, ta cần kiểm tra để xác định giới hạn tin cậy của phương trình hồi quy tuyến tính.

Như đã trình bày ở trên phương trình hồi quy tuyến tính là một ước lượng của một hồi quy lý thuyết y trên x , đặc thù của đám đông lý thuyết mà đám đông thực tế quan sát chỉ là một mẫu bất kỳ. Vậy ta phải xác định giới hạn tin cậy của ước lượng ấy, để tránh những sai số lớn nếu ta sử dụng một ước lượng không đúng. Trung bình của bình phương những chênh lệch giữa những trị số \hat{y}_i ước lượng và y_i thực tế quan trắc là:

$$s^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n-2} \quad (8-4)$$

Bậc tự do ở đây bằng $n - 2$, vì trong n cặp so sánh $(y_i - \hat{y}_i)$ ta bị hai liên hệ ràng buộc là những phương trình tính \bar{y} và \bar{x} .

Người ta đã chứng minh rằng: phương sai của hệ số hồi quy thực nghiệm được tính theo công thức.

$$S_a^2 = \frac{\sum (Y_i - \hat{Y}_i)^2}{(n-2) \sum (x_i - \bar{x})^2} \quad (8-5)$$

và phương sai của $b = (\bar{y} - a\bar{x})$ là:

$$S_b^2 = \frac{\sum (Y_i - \hat{Y}_i)^2}{n(n-2)} \quad (8-6)$$

Trong hai công thức (8-5) và (8-6) số hạng mới cần phải tính là:

$$\begin{aligned} & \sum (Y_i - \hat{Y}_i)^2 \\ & \sum (y_i - \hat{y}_i)^2 = \sum \left[y_i - \bar{y} - a(x_i - \bar{x}) \right]^2 = \sum (y_i - \bar{y})^2 + a^2 \sum (x_i - \bar{x})^2 - 2a \sum (x_i - \bar{x})(y_i - \bar{y}) \\ & = \sum (y_i - \bar{y})^2 - a^2 \sum (x_i - \bar{x})^2 = \sum y_i^2 - n\bar{y}^2 - a^2 \left(\sum x_i^2 - n\bar{x}^2 \right) \end{aligned}$$

Biết được S_a^2 và S_b^2 ta sẽ tính được S_a và S_b , sau đó tính các t thực nghiệm để kiểm định các giả thiết $a = 0$ và $b = 0$

$$t_{(a)} = \frac{a}{S_a} \quad .t(b) = \frac{b}{S_b} \quad (8-7)$$

Cuối cùng đối chiếu t thực nghiệm này với các trị số lý thuyết $(t_{\alpha, n-2})$ ở bảng phụ lục để đưa ra kết luận

Vấn đề tiếp theo là tính hệ số tương quan.

Trên hình (8.1) đường gấp khúc đi qua các điểm tọa độ $M(x, y)$ là D_1 . Để đánh giá D_1 chênh lệch nhiều hay ít so với đường D nghĩa là cần biết mức độ tương quan giữa y và x ta không xét đến hệ số góc a của đường D mà nghiên cứu hệ số tương quan r .

$$r = a \frac{S_x}{S_y} \quad (8-8)$$

Trong đó S_x là độ lệch chuẩn của x .

S_y là độ lệch chuẩn của y .

Giá trị của r bằng a chia cho S_y/S_x nghĩa là: r là giá trị của hệ số góc khi ta lấy S_x và S_y làm đơn vị đo lường x và y .

Như vậy, r không phụ thuộc vào các đơn vị đo lường của x và y như a , nhờ đó ta có thể lập được bảng r chung cho các trường hợp.

Trong công thức (8- 8) ta có:

$$S_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \quad \text{va} \quad S_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}}$$

Nên:
$$\frac{S_x}{S_y} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{\sum (y_i - \bar{y})^2}}$$

Và thay thế a bằng giá trị của nó ta có:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \times \sqrt{\frac{\sum (x_i - \bar{x})^2}{\sum (y_i - \bar{y})^2}}$$

Chia tử số và mẫu số cho $\sqrt{\sum (x_i - \bar{x})^2}$ thì công thức (8- 8) có dạng (8- 9)

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}} \quad (8-9)a$$

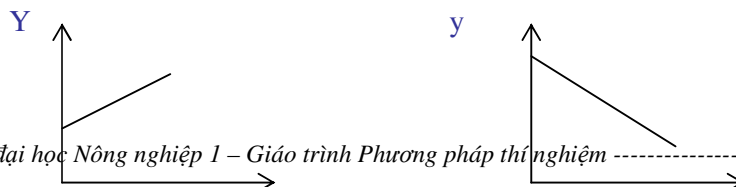
$$r = \frac{\sum xy - (\sum x \cdot \sum y) : n}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n} \right]} \sqrt{\left[\sum y^2 - \frac{(\sum y)^2}{n} \right]}} \quad (8-9)b$$

Căn cứ vào công thức (8-9) ta thấy: giá trị của r luôn luôn nằm trong khoảng từ -1 đến +1.

Người ta đã lập bảng hệ số tương quan trong đó có những giá trị tuyệt đối của r ứng với bậc tự do bằng $n- 2$ (n là số mẫu quan sát) với các mức xác suất khác nhau (xem bảng 10 phụ lục). Bảng hệ số tương quan chỉ cho ta hai giá trị P là : 0,05; 0,01 và ứng với bậc tự do $df < 100$. Trong trường hợp bậc tự do $df > 100$ và $P < 0,01$ có thể kiểm định

$$t_m = \frac{|r|}{\sqrt{1-r^2}} \times \sqrt{n-2} \quad (8-10)$$

và đọc ở bảng t ứng với mức xác suất P và bậc tự do bằng $n-2$ ta sẽ được trị số t lý thuyết để so sánh trị số t_m từ công thức (8-10), nếu t_m lớn hơn t lý thuyết chứng tỏ giữa x và y tương quan khác không, nếu $t_m \leq t$ lý thuyết thì giữa x và y không có tương quan.



X

x

Hình 8.3. Tương quan thuận**Hình 8.4. Tương quan nghịch**

Từ công thức (8-9) ta thấy rằng r có thể là dương(+), có thể là(-). Nếu r là (+) thì quan hệ giữa x và y là tương quan thuận, xem hình (8.3). Nếu r là (-) thì quan hệ giữa x và y nghịch, xem hình (8. 4).

8.2.2. Các ví dụ

8.2.2.1. Trường hợp dung lượng mẫu nhỏ ($n < 30$)

Thí dụ (8-1): tìm hiểu mối quan hệ giữa hàm lượng lân tổng số và năng suất lúa. Kết quả phân tích 21 mẫu được ghi ở bảng (8.1).

Quá trình tính toán như sau:

- Lập bảng tính các giá trị Σx ; Σy ; Σx^2 ; Σy^2 ; Σxy như bảng (8.1).

$$N= 21$$

$$\bar{x} = (\Sigma x) : n = 1,101 : 21 = 0,052$$

$$\bar{y} = (\Sigma y) : n = 1148,2 : 21 = 54,0$$

$$\Sigma (x - \bar{x})^2 = \Sigma x^2 - (\Sigma x)^2 : n = 0,058133 - (1,101)^2 : 21 = 0,00041$$

$$\Sigma (y - \bar{y})^2 = \Sigma y^2 - (\Sigma y)^2 : n = 63738,34 - (1148,2)^2 : 21 = 959,14$$

$$\Sigma (x - \bar{x})(y - \bar{y}) = \Sigma xy - (\Sigma x \Sigma y) : n = 60,7747 - (1,101 \times 1148,2) : 21 = 0,5847$$

Tính hệ số tương quan và phương trình hồi quy theo công thức (8- 1), (8-3),và (8-9)

$$r = \frac{\Sigma (x - \bar{x})(y - \bar{y})}{\sqrt{\Sigma (x - \bar{x})^2 \Sigma (y - \bar{y})^2}} = \frac{0,5847}{\sqrt{0,00041 \times 959,14}} = 0,832$$

$$a = \frac{\Sigma (x - \bar{x})(y - \bar{y})}{\Sigma (x - \bar{x})^2} = \frac{0,5847}{0,00041} = 1426$$

Bảng 8.1. Quan hệ giữa lân tổng số và năng suất lúa của 21 mẫu phân tích

Thứ tự	Những trị số quan sát		Bình phương		x.y
	P ₂ O ₅ %	Năng suất y (tạ/ ha)	X ²	y ²	
1	0,058	63,7	0,003364	4057,69	3,6946
2	0,057	62,3	0,003249	3881,29	3,5511
3	0,035	56,2	0,002809	3203,56	2,9998
4	0,054	58,6	0,002916	3433,96	3,1144
5	0,046	48,1	0,002116	2313,61	2,2126
6	0,048	45,3	0,002304	2052,09	2,1744
7	0,051	53,8	0,002601	2894,44	2,7438
8	0,045	45,3	0,002025	2052,09	2,0385
9	0,050	52,4	0,002500	2745,76	2,6200
10	0,056	60,9	0,003136	3708,81	3,4104
11	0,056	60,0	0,003136	3600,00	3,3600
12	0,055	60,0	0,003025	3600,00	3,3000
13	0,049	46,7	0,002401	2180,89	2,2883
14	0,050	51,6	0,002500	2662,56	2,5800
15	0,052	53,8	0,002704	2894,44	2,7976
16	0,045	43,9	0,002025	1927,21	1,9755
17	0,057	55,0	0,003249	3025,00	3,1350
18	0,056	62,3	0,003136	3881,29	3,4888
19	0,058	56,0	0,003364	3136,00	3,2480
20	0,058	66,6	0,003364	4435,56	3,8628
21	0,047	45,3	0,002209	2052,09	2,1291
	1,101	1.148,2	0,058133	63.738,34	60,7747

$$y = \bar{y} + a(x - \bar{x}) = 54 + 1426(x - 0,052)$$

Ta có: $y = 1426x - 20$

Kiểm tra mức độ tương quan và giới hạn tin cậy của phương trình.

Theo các công thức (8-5); (8- 6); (8- 7) ta có:

$$S_a^2 = \frac{\sum y_i^2 - n\bar{y}^2 - a^2(\sum x_i^2 - n\bar{x}^2)}{(n-2)(\sum x_i^2 - n\bar{x}^2)} = \frac{63738,34 - 21 \times 54^2 - 1426^2 \times 0,00041}{19 \times 0,00041}$$

$$= \frac{1668,6}{0,0078} = 213.923,0$$

Và $S_a = 425,49$

$$S_b^2 = \frac{\sum y_i^2 - n\bar{y}^2 - a^2(\sum x_i^2 - n\bar{x}^2)}{n(n-2)} = \frac{1668,6}{21 \times 19} = 4,18$$

Và $S_b = 2,044$

$$t_a = \frac{a}{S_a} = \frac{1426}{425,49} = 3,35$$

$$t_b = \frac{b}{S_b} = \frac{20}{2,044} = 9,78$$

$t_{0,01, 19} = 2,861$ (bảng 4 phụ lục.)

Kết luận:

Phương trình hồi quy giữa năng suất lúa và hàm lượng lân ở trong đất tin cậy ở mức xác suất $P = 0,99$.

Hệ số r tra bảng ứng với độ tự do $df = n-2 = 19$ và mức ý nghĩa $\alpha = 0,01$ bằng 0,5487.

Như vậy, hệ số r tính > r lý luận, ta có thể kết luận rằng tương quan giữa lân ở trong đất và năng suất lúa là rõ với mức tin cậy 95%.

8.2.2.2. Trường hợp dung lượng mẫu lớn ($n > 30$)

Trong trường hợp nhiều số liệu (n lớn), ta lập một bảng hai chiều gọi là bảng tương quan. Cách lập bảng tương quan như trong bảng (8- 2), bảng được chia thành nhiều ô, mỗi ô chứa tần số n_{ij} có hai giá trị x_i và y_j của hai đặc tính x và y .

Bảng 8 -2.

Y	X					Tổng f_y
			x_i			
y_j			n_{ij}			
Tổng f_x						n

Các giá trị của x và y trình bày trong bảng là các giá trị giữa của từng tổ. Cách tính hệ số tương quan trong trường hợp n lớn mà số liệu được phân thành từng tổ, cũng như khi tính số trung bình và độ lệch chuẩn, ta đổi gốc tọa độ để tính các phép tính trên biến số mới X_i và Y_i .

$$X_i = \frac{x_i - A_x}{C_x} ; Y_i = \frac{y_i - A_y}{C_y}$$

Do đó:

$$\bar{x} = A_x + C_x \left(\sum f_x X_i \right) : n$$

$$\bar{y} = A_y + C_y \left(\sum f_y Y_i \right) : n$$

$$\sum (x - \bar{x})^2 = C_x^2 \left[\sum f_x X_i^2 - \left(\sum f_x X_i \right)^2 : n \right]$$

$$\sum (y - \bar{y})^2 = C_y^2 \left[\sum f_y Y_i^2 - \left(\sum f_y Y_i \right)^2 : n \right]$$

$$\sum (x - \bar{x})(y - \bar{y}) = C_x C_y \left[\sum f X_i Y_i - \left(\sum f_x X_i \times \sum f_y Y_i \right) : n \right]$$

Thí dụ (8- 2):

Nghiên cứu mối quan hệ giữa tỷ lệ mùn x (%) và hàm lượng lân y (miligam) trên 100 gam đất. Kết quả phân tích 64 mẫu được ghi trong bảng (8-3).

Các bước tính toán như sau:

- Lập bảng phân tổ hai chiều:

+ Chia tổ cho từng dãy biến số, ta thấy n = 64 như vậy ta có thể chia các dãy số liệu trên thành 6 đến 8 tổ. Để cho các số liệu thực tế nằm gọn trong các tổ ta lấy số tổ ứng với biến số x là 7 và biến số y là 6. Nên khoảng cách tổ như sau:

Bảng 8-3.

Mẫu đất	mùn x%	Lân y	Mẫu đất	mùn x%	Lân y	Mẫu đất	Mùnx%	Lân y	Mẫu đất	mùnx%	Lân y
1	1,57	30	17	1,35	17	33	0,96	6	49	1,42	27
2	1,58	28	18	1,31	17	34	1,08	9	50	1,36	25
3	1,1	25	19	1,29	16	35	1,16	19	51	1,55	24
4	1,21	27	20	1,38	17	36	1,12	17	52	1,36	22
5	1,44	25	21	1,38	16	37	1,01	11	53	1,46	28
6	1,37	24	22	1,36	14	38	1,07	11	54	1,39	28
7	1,45	25	23	1,36	16	39	1,10	16	55	1,63	36
8	1,49	27	24	1,20	17	40	1,22	17	56	1,57	36
9	1,38	24	25	1,36	16	41	1,22	16	57	1,37	27
10	1,41	25	26	1,29	14	42	1,12	19	58	1,48	25
11	1,55	25	27	1,30	12	43	0,86	20	59	1,61	28
12	1,45	25	28	1,32	12	44	0,79	19	60	1,61	30
13	1,30	22	29	1,17	11	45	1,19	23	61	1,70	28
14	1,30	22	30	1,22	11	46	1,15	22	62	1,61	28
15	1,39	20	31	1,09	9	47	1,13	18	63	1,04	9
16	1,46	22	32	1,13	9	48	1,34	20	64	1,12	10

$$C_x = \frac{X_{\max} - X_{\min}}{6 \rightarrow 8} = \frac{1,70 - 0,79}{6 \rightarrow 8} = \frac{0,91}{7} = 0,13\%$$

$$C_y = \frac{Y_{\max} - Y_{\min}}{6 \rightarrow 8} = \frac{36 - 6}{6 \rightarrow 8} = \frac{30}{6} = 5mg$$

Căn cứ vào các số tổ và khoảng cách tổ ta lập bảng phân tổ hai chiều bằng cách định giới hạn tổ, tính các trị số giữa tổ, tần số của từng tổ ứng với các biến số và tổng tần số như bảng 8-4

Bảng 8-4. Bảng phân tổ hai chiều cho các đại lượng x và y tính từ bảng 8-3

y \ x		x							
		0,79-0,91	0,92-1,04	1,05-1,17	1,18-1,30	1,31-1,43	1,44-1,56	1,57-1,70	Tổng fy
		Trị số giữa tổ							
		0,58	0,98	1,11	1,24	1,37	1,50	1,64	
36-31	Trị số giữa a tổ	33						2	2
30-26		28			1	3	2	6	12
25-21		23			1	4	5	7	17
20-16		18	2	5	4	8			20
15-11		13		2	3	2			9
10-6		8		3	1				4
Tổng fx		2	3	11	13	18	9	8	64=n

Từ số liệu ở bảng (8-4) ta lập bảng phân tổ hai chiều theo biến số mới và lập bảng tính các tổng theo công thức tính r và a, b của phương trình $y = ax + b$ như bảng (8-5)

Bảng 8-5. Bảng tính hệ số tương quan và phương trình hồi quy theo biến số mới X_i, Y_i .

$\frac{Y_i - 18}{5}$	X_i	X-1,24	-3	-2	-1	0	1	2	3	Fy	fyY ₁	fyY ₁ ²
		0,13										
	Y	X	0,85	0,98	1,11	Ax=1,24	1,37	1,50	1,63			
3	33								2	2	6	18
2	28					1	3	2	6	12	24	48
1	23			1		4	5	7		17	17	17
0	Ay=18	2	1	5	4	4	8			20	9	0
-1	13		2	2	3	3	2			9	-9	9
-2	8			3	1	1				4	-8	16
fx			2	3	11	13	18	9	8	n=64	30=	110=
fxX _i			-6	-6	-11	0	18	18	24	ΣfxX _i	ΣfyY _i	ΣfyY _i ²
fxX _i ²			18	12	11	0	18	36	72	167=		
fX _i Y _i			0	4	7	0	9	22	54	96=		
										ΣfX _i Y _i		

Chú ý fx, fy hoặc f ký hiệu ở chương này tương tự với ký hiệu n_i (tần số) ở các chương khác.

Trong bảng (8-5) chọn A_(x) = 1,24 và A_(y) = 18

Tính biến số mới theo công thức:

$$X_i = \frac{x_i - 1,24}{0,13} ; Y_i = \frac{y_i - 18}{5}$$

- Tính các tích số fxX_i và fyY_i và các tổng.

$$\Sigma fxX_i = 37; \Sigma fyY_i = 30$$

- Tính các tích số fxX_i² và fyY_i² và các tổng

$$\Sigma fxX_i^2 = 167; \Sigma fyY_i^2 = 110$$

- Tính các tích số fX_iY_i và tổng ΣfX_iY_i = 96.

Trong đó f là tần số ở từng ô ứng với từng giá trị X_i và Y_i.

Cách tính như sau: Tổ 1: fX_iY_i = 2 x (-3) x 0 = 0

$$\text{Tổ 2: } fX_i Y_i = 1 x (-1) x 0 + 2 x (-2) x (-1) = 4$$

$$\text{Tổ 3: } fX_i Y_i = 1 x (-1) x 1 + 5 x (-1) x 0 + 2 x (-1) x (-1) + 3 x (-1) x (-2) = 7$$

.....

$$\text{Tổ 7: } fX_i Y_i = 2 x 3 x 3 + 6 x 3 x 2 = 54$$

- Tính \bar{x} và \bar{y}

$$\bar{y} = A_y + C_y \sum f_y Y_i : n \quad \bar{y} = 18 + 5 \times 30 : 64 = 20,3 \text{ mgr}$$

$$\bar{x} = A_x + C_x \sum f_x X_i : n \quad \bar{x} = 1,24 + 0,13 \times 37 : 64 = 1,32\%$$

• Tính các tổng:

$$\Sigma(x - \bar{x})^2 = C_x^2 [\Sigma f_x X_i^2 - (\Sigma f_x X_i)^2 : n] = 0,13^2 (167 - 37^2 : 64) = 2,46$$

$$\Sigma(y - \bar{y})^2 = C_y^2 [\Sigma f_y Y_i^2 - (\Sigma f_y Y_i)^2 : n] = 5^2 (110 - 30^2 : 64) = 2398,5$$

$$\Sigma(x - \bar{x})(y - \bar{y}) = C_y C_x [(\Sigma f_x X_i Y_i - \Sigma f_x X_i \Sigma f_y Y_i) : n] = 0,13 \times 5 (96 - 37 \times 30 : 64) = 51,13$$

• Tính hệ số tương quan và phương trình hồi quy:

$$r = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\sqrt{\Sigma(x - \bar{x})^2 \Sigma(y - \bar{y})^2}} = \frac{51,13}{\sqrt{2,46 \times 2398,5}} = 0,67$$

$$a_{y/x} = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\Sigma(x - \bar{x})^2} = \frac{51,13}{2,46} = 20,8 \text{ mgr}$$

$$y = \bar{y} + a_{y/x}(x - \bar{x}) = 20,3 + 20,8(x - 1,32) = 20,8x - 7,2.$$

• Kiểm tra mức độ tương quan và giới hạn tin cậy của phương trình theo các công thức (8-5); (8-6); và (8-7).

$$S_a^2 = \frac{C_y^2 [\Sigma f_y Y_i^2 - (\Sigma f_y Y_i)^2 : n] - a^2 C_x^2 [\Sigma f_x X_i^2 - (\Sigma f_x X_i)^2 : n]}{(n-2) C_x^2 [\Sigma f_x X_i^2 - (\Sigma f_x X_i)^2 : n]}$$

$$= \frac{2398,5 - 20,8^2 * 2,46}{62 * 2,46} = \frac{1334,2}{152,5} = 8,74$$

$$S_a = 2,956$$

$$S_b^2 = \frac{C_x^2 [\Sigma f_x X_i^2 - (\Sigma f_x X_i)^2 : n] - a^2 C_y^2 [\Sigma f_y Y_i^2 - (\Sigma f_y Y_i)^2 : n]}{(n-2) * n}$$

$$= \frac{2398,5 - 20,8^2 * 2,46}{62 * 64} = \frac{1334,2}{62 * 64} = 0,336$$

$$S_b = 0,58$$

$$t_a = \frac{20,8}{2,956} = 7,03 \quad t_b = \frac{7,2}{0,58} = 12,41$$

$t_{0,01, 62} = 2,576$ tra bảng 4 phụ lục.

Như vậy ; t_a và t_b lớn hơn t lý luận.

Kết luận:

Phương trình hồi quy giữa tỷ lệ mùn và hàm lượng lân ở trong đất tin cậy được ở mức xác suất $P = 0,99$

Hệ số r tra bảng phụ lục 10 ứng với độ tự do $(n - 2) = 62$ và mức xác suất $\alpha = 0,01$ bằng 0,3248.

Như vậy: r tính > r lý luận, ta có thể kết luận rằng tương quan giữa lân ở trong đất và tỷ lệ mùn là rất chặt.

Chú ý:

Ta cũng có thể tính gần đúng hệ số tương quan theo công thức tính tương quan thứ tự của Spearman

$$r = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

Trong đó: d là hiệu số các trị số thứ tự của các cặp tương quan.

n là số cặp tương quan.

Tính hệ số tương quan theo phương pháp của Spearman tương đối đơn giản, rút ngắn được thời gian. Nhưng nó chỉ có lợi khi ta không cần phải xây dựng phương trình hồi quy. Còn trong tính toán cần thiết phải xây dựng phương trình hồi quy thì không nên áp dụng phương pháp này. Trị số r tính được ở trên được so sánh với trị số r lý luận ở bảng phụ lục để đánh giá mức độ quan hệ.

Ta có thể tìm hiểu nội dung chi tiết của phương pháp thông qua thí dụ sau:

Thí dụ 8-3.

Nghiên cứu mối tương quan giữa lượng mưa và năng suất kết quả ghi lại như bảng (8 - 6) (theo tài liệu phương pháp thí nghiệm đồng ruộng của A. Jacob 1965).

Từ số liệu quan sát được, ta đem sắp xếp theo thứ tự từ nhỏ đến lớn và ghi số liệu thứ tự của x và y vào bảng.

Trường hợp khi số thứ tự trùng nhau thì ký hiệu số thứ tự là số trung bình của thứ tự các số liệu trùng nhau.

Thí dụ năm 1909 và 1913 đều có lượng mưa là 108 ly. Theo thứ tự thì số 108 sẽ lần lượt chiếm hai số thứ tự là 7 và 8 vấn đề đặt ra là thứ tự 7 đặt ở năm nào? vì vậy ta phải xếp thứ tự trung bình là 7,5 để xếp.

Thay các trị số đã tính được vào công thức ta được:

$$r = 1 - \frac{6 * 813}{26(26^2 - 1)} = 0,72$$

Tra bảng phụ lục 10 khi độ tự do $df = n - 2 = 26 - 2 = 24$ ta có $r_{01} = 0,487$ như vậy r tính lớn hơn r_{01} .

Do vậy ta có thể kết luận chắc chắn rằng giữa lượng mưa và năng suất có quan hệ chặt.

Bảng 8-6. Nghiên cứu mối quan hệ giữa lượng mưa X (tính bằng mm) và năng suất y (tính bằng tạ / ha)

Năm	Trị số quan sát		Tứ tự		D	d ²
	X	Y	X	y		
1911	71	16,6	1	2	-1	1
1915	89	16,4	2	1	+1	1
1901	96	25,0	3	10	-7	49
1917	98	19,2	4	3	+1	1
1903	105	26,2	5	13,5	-8,5	72
1918	106	20,2	6	6	0	0
1913	108	19,4	7,5	4	+3,5	12
1909	108	22,6	7,5	8	- 0,5	0
1925	110	30,2	9,0	17	-8	64
1906	111	19,6	10	5	+ 5	25
1912	119	29,6	11	16	- 5	25
1919	123	25,6	12	11	+ 1	1
1914	132	30,6	13	20	- 7	49
1905	135	20,4	14	7	+ 7	49
1910	137	24,2	15	9	+ 6	36
1902	144	32,6	16	23	- 7	49
1916	147	30,4	17,5	18,5	- 1	1
1924	147	30,4	17,5	18,5	- 1	1
1920	156	31,0	19	21	- 2	4
1907	161	33,8	20	25	- 5	25
1922	162	31,6	21	22	- 1	1
1900	177	26,2	22	12	+ 10	100
1921	191	35,8	23	26	- 3	9
1906	209	29,2	24	15	+ 9	81
1923	235	33,6	25	24	+ 1	1
1908	246	26,6	25	24	+ 1	1

Chú ý: Trong một số trường hợp chúng ta phải so sánh mức độ tương quan giữa hai đặc trưng xem có giống nhau hay khác nhau. Vì r không phân phối chuẩn nên phải biến đổi

$$Z = \frac{1}{2} \ln \frac{1+r}{1-r}$$
 trước khi so sánh, giá trị Z được cho trong bảng 12 phụ lục,
 từ giá trị của r_1 ta tìm được Z_1 .
 từ giá trị của r_2 ta tìm được Z_2 .

Độ lệch chuẩn S_{z_1} và S_{z_2} được tính theo công thức:

$$S_{z_1} = \frac{1}{\sqrt{n_1 - 3}} ; \quad S_{z_2} = \frac{1}{\sqrt{n_2 - 3}}$$

Sở dĩ bậc tự do ở đây bằng $n - 3$ vì trong quá trình tính Z ta phải trải qua 3 bước, mất một độ tự do khi tính trung bình, một khi tính hệ số tương quan và một khi đổi giá trị từ r và Z.

Độ lệch chuẩn của hiệu $Z_1 - Z_2$ được tính theo công thức:

$$S_{(z_1 - z_2)} = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}$$

Trong đó: n_1 số cặp tương quan r_1

n_2 số cặp tương quan r_2 .

Trị số t thực nghiệm được tính theo công thức.

$$t_m = \frac{d}{S_{(z_1 - z_2)}} = \frac{Z_1 - Z_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$$

Trị số t_m được so sánh với trị số t trong bảng ở mức ý nghĩa nhỏ (0,05 hoặc 0,01) và bậc tự do tương ứng bằng $df = n_1 + n_2 - 6$.

Nếu : $t_m > t_{\text{bảng}}$ thì ta kết luận hai hệ số tương quan nghiên cứu khác nhau rõ rệt.

$T_m \leq t_{\text{bảng}}$ thì ta kết luận hai hệ số tương quan nghiên cứu khác nhau không có ý nghĩa.

Thí dụ 8-4 : nghiên cứu mối tương quan giữa hàm lượng mùn tổng số ở trong đất và năng suất lúa trên đất bạc màu huyện Gia Lương với số điều tra $n_1 = 63$ được $r_1 = 0,738$ và cũng ở các điểm điều tra trên phân tích tương quan giữa hàm lượng lân tổng số trong đất với năng suất $n_2 = 61$ (có hai mẫu số liệu không thu được) được $r_2 = 0,808$.

Như vậy giữa hai chỉ tiêu nghiên cứu chỉ tiêu nào có quan hệ với năng suất chặt chẽ hơn.

Để trả lời câu hỏi đặt ra ở trên chúng ta xuất phát từ giả thiết cho rằng: giữa hai chỉ tiêu nghiên cứu có mức độ quan hệ với năng suất là như nhau, và kiểm tra giả thuyết trên theo phương pháp so sánh hai trị số Z.

Tra bảng ứng với $r_1 = 0,738$ ta có $Z_1 = 0,9505$
 $r_2 = 0,808$ ta có $Z_2 = 1,1270$
 $d = Z_2 - Z_1 = 1,1270 - 0,9505 = 0,1765$

$$S_{z_1-z_2} = \sqrt{\frac{1}{63-3} + \frac{1}{61-3}} = 0,1841$$

$$t_m = \frac{d}{S_{z_1-z_2}} = \frac{0,1765}{0,1841} = 0,958$$

$$t_{0,05, 118} = 1,96$$

$t_m < t_{\text{bảng}}$ như vậy ta có thể kết luận giả thuyết nêu trên là đúng. Hay nói cách khác là: quan hệ giữa mùn và lân với năng suất lúa trên đất bạc màu huyện Gia Lương không khác nhau (hoặc khác nhau không có ý nghĩa)

8.3. Tương quan tuyến tính nhiều biến

8.3.1. Ý nghĩa của nghiên cứu tương quan nhiều yếu tố

Tương quan và hồi quy mà ta xét tới trên đây chỉ là mối tương quan tương hỗ giữa hai yếu tố. Trong thực tế không phải chỉ phát sinh quan hệ tương hỗ giữa hai sinh vật, hai hiện tượng. Ví dụ: năng suất cây trồng không phải chỉ liên quan mật thiết với lượng mưa mà ôn độ cũng có ảnh hưởng rất lớn. Do đó, khi xét tới những nhân tố ảnh hưởng đến năng suất ta không chỉ chú ý đến lượng mưa mà còn phải chú ý cả đến mối quan hệ của ôn độ và các nhân tố khác. Đồng thời khi nghiên cứu sâu mối quan hệ năng suất với lượng mưa mà không cô lập hoặc loại trừ những ảnh hưởng của các nhân tố khác đến sản lượng cũng là không hợp lý.

Nếu thí nghiệm chỉ nghiên cứu quan hệ của hai nhân tố mà không đề cập tới các nhân tố khác, nhưng các nhân tố đó lại thường có ảnh hưởng đến năng suất, thì ta phải tìm cách cố định, cô lập các yếu tố đó lại. Thí dụ: khi nghiên cứu mối quan hệ giữa năng suất và lượng mưa ta cần cố định cô lập các yếu tố khác như: ôn độ để tìm mối tương quan đơn thuần giữa năng suất với lượng mưa mà thôi.

Phương pháp thống kê nhằm nghiên cứu mối quan hệ thuần túy giữa hai yếu tố được gọi là: hồi quy riêng và tương quan riêng.

Một phương pháp khác nhằm tìm hiểu mối quan hệ giữa một biến số, một yếu tố này với một số yếu tố khác hay biến số khác. Thí dụ: nghiên cứu mối quan hệ giữa năng suất cây trồng với các yếu tố khí hậu. Phương pháp thống kê nghiên cứu mối quan hệ của một yếu tố với hai yếu tố khác trở lên về hồi quy gọi là hồi quy bội và tương quan được gọi là tương quan bội.

8.3.2. Hệ số hồi quy riêng và phương trình hồi quy bội.

Phương trình hồi quy bội có dạng $Y = f(a, b, c, \dots)$

Nếu các biến số a, b, c, ... tương đối độc lập và quan hệ với y theo dạng tuyến tính thì:

$$Y = A_0 + Aa + Bb + Cc + \dots$$

Trong đó: A, B, C, ... được gọi là hệ số hồi quy riêng. Bằng phương pháp bình phương nhỏ nhất ta có thể xây dựng được hệ ma trận mẫu:

$$\Sigma y = nA_0 + A\Sigma a + B\Sigma b + C\Sigma c + \dots \quad (1)$$

$$\Sigma ay = A_0\Sigma a + A\Sigma a^2 + B\Sigma ba + C\Sigma ca + \dots \quad (2)$$

$$\Sigma by = A_0\Sigma b + A\Sigma ab + B\Sigma b^2 + C\Sigma cb + \dots \quad (3)$$

$$\Sigma cy = A_0\Sigma c + A\Sigma ac + B\Sigma c^2 + C\Sigma c^2 + \dots \quad (4)$$

Trong đó: Σy ; n; Σa ; Σb ; Σc ; Σay ; Σa^2 ; Σba ; Σac ; Σby ; Σb^2 ; Σcy ; Σc^2 là các giá trị thực nghiệm.

Giải hệ ma trận mẫu ta được A_0, A, B, C, \dots và thiết lập được phương trình.

Trường hợp các biến a, b, c, ... quan hệ với y không theo dạng tuyến tính thì phải tuyến tính hoá các quan hệ phi tuyến để áp dụng hệ ma trận mẫu nêu ở trên.

Trường hợp các biến a, b, c có quan hệ lẫn nhau thì phương hướng giải quyết phức tạp hơn nhiều do đây chúng ta không đề cập tới trong phạm vi giáo trình này.

Bước tiếp theo là xác định độ tin cậy của phương trình và hệ số hồi quy riêng.

Xác định độ tin cậy của phương trình.

Chúng ta có thể dễ dàng thấy sự sai khác giữa các trị số y quan sát so

với trị số \bar{y} là do hai nguyên nhân :

y khác với \bar{y} là do các yếu tố nghiên cứu khác nhau a, b, c, ... dẫn tới

y khác với \bar{y} là do các yếu tố ngoài các yếu tố nghiên cứu dẫn tới.

Nếu ta ký hiệu : SST₀ là biến động toàn bộ thì:

$$SST_0 = \sum (y_i - \bar{y})^2$$

SST là biến động do các yếu tố nghiên cứu thì:

$$SST = A \sum (a - \bar{a}) (y - \bar{y}) + B \sum (b - \bar{b}) (y - \bar{y}) + C \sum (c - \bar{c}) (y - \bar{y}) + \dots$$

SS_E là biến động do các yếu tố không nghiên cứu thì:

$$SS_E = SST_0 - SST$$

Nếu biến động do các yếu tố nghiên cứu gây ra xấp xỉ bằng với biến động do các yếu tố không nghiên cứu, thì chúng ta vai trò của các yếu tố nghiên cứu a, b, c, ... đối với y là chưa rõ. Trường hợp biến động do các yếu tố nghiên cứu gây ra lớn hơn biến động do các yếu tố không nghiên cứu gây ra, thì chúng ta vai trò của các yếu tố nghiên cứu a, b, c, ... đối với y là rất rõ.

Để tiện cho việc so sánh người ta dựa vào phương pháp phân tích phương sai của các nguyên nhân biến động.

Phương sai của yếu tố nghiên cứu:

$$MST = \frac{SST}{dfT}$$

Trong đó: dfT bằng số nhân tố nghiên cứu.

Phương sai của các yếu tố không nghiên cứu:

$$MSE = \frac{SSE}{dfE}$$

Trong đó: bậc tự do của các yếu tố không nghiên cứu dfE = n-1-dfT

Tính
$$F_m = \frac{MST}{MSE}$$

Nếu $F_m \leq F_\alpha$ trong bảng với mức ý nghĩa α , dfT, dfE thì ta kết luận vai trò của các yếu tố nghiên cứu đối với y là không rõ.

Nếu $F_m > F_\alpha$ trong bảng với mức ý nghĩa α , dfT, dfE, thì ta kết luận vai trò của các yếu tố nghiên cứu đối với y là rõ ràng, có tác dụng quyết định.

Trị số F_m ở trên là kết quả tổng hợp của nhiều yếu tố nghiên cứu.

Dựa vào phương sai của hệ số hồi quy riêng để xác định độ tin cậy của hệ số hồi quy theo phương pháp so sánh của Student

$$\text{Với: } t_a = \frac{a}{S_{a.bcd \dots}}; \quad t_b = \frac{b}{S_{b.acd \dots}}; \quad t_c = \frac{c}{S_{c.abd \dots}}$$

Trong đó: $S_{a.bcd \dots}$ là độ lệch chuẩn của hệ số a, khi cố định b, c, d,...

$S_{b.acd \dots}$ là độ lệch chuẩn của hệ số b, khi cố định a, c, d,...

$S_{c.abd \dots}$ là độ lệch chuẩn của hệ số c, khi cố định a, b, d,...

Việc xác định các độ lệch chuẩn của một hệ số khi cố định các hệ số khác được tiến hành theo nguyên tắc sau:

Từ phương trình:

$$Y = A_0 + A_a + B_b + C_c + D_d + \dots$$

Khi ta cố định b, c, d, ... thì phương trình trên có dạng:

$$Y_{a.bcd} = A_0 + A_a + Q(*)$$

Trong đó Q là tổng của các yếu tố cố định.

Phương trình (*) nếu ta thay thế các giá trị a_i thì thu được kết quả:

$$a_1 \text{ ta sẽ có } \hat{Y}'_{a1.bcd}$$

$$a_2 \text{ ta sẽ có } \hat{Y}'_{a2.bcd}$$

$$\dots$$

$$a_n \text{ ta sẽ có } \hat{Y}'_{an.bcd}$$

Các trị số $\hat{Y}'_{a1.bcd \dots}$; $\hat{Y}'_{a2.bcd \dots}$ khác với $Y_{a1.bcd \dots}$; $Y_{a2.bcd \dots}$; $Y_{an.bcd \dots}$ nhiều hay ít nói lên mức độ tin cậy của hệ số hồi quy riêng.

$$S^2_{a.bcd \dots} = \frac{\sum_{i=1}^n (y_{a_i.bcd \dots} - \hat{y}_{a_i.bcd})^2}{n-1-p} \quad (**)$$

Trong đó p là số yếu tố cố định.

Trị số t ở trên được so sánh với $t_{(\alpha, n-1-p)}$

Nếu $t > t_{(\alpha, n-1-p)}$ thì ta khẳng định được rằng hệ số hồi quy riêng nói trên tin cậy được.

Nếu $t \leq t_{(\alpha, n-1-p)}$ thì ta khẳng định được rằng hệ số hồi quy riêng nói trên là không thể tin cậy được.

Thí dụ 8-5: nghiên cứu mối quan hệ giữa lân vô cơ (a) và lân hữu cơ (b) với hàm lượng lân dễ tiêu (y) ở trong đất. Kết quả phân tích 18 mẫu đất được ghi lại ở bảng (8 -8).

Bảng 8 -8.

Mẫu đất	a	b	y	\hat{Y}	y- \hat{Y}
1	0,4	53	64	61,6	2,4
2	0,4	23	60	59,0	1,0
3	3,1	19	71	63,4	7,6
4	0,6	34	61	60,3	0,7
5	4,7	24	54	66,7	-12,7
6	1,7	65	77	64,9	12,1
7	9,4	44	81	76,9	4,1
8	10,1	31	93	77,0	16,0
9	11,6	29	93	79,6	13,4
10	12,6	58	51	83,8	-32,8
11	10,6	37	76	79,0	-3,0
12	23,1	46	96	101,6	-5,6
13	23,1	50	77	101,9	-24,9
14	21,6	44	93	98,7	-5,7
15	23,1	56	95	102,4	-7,4
16	1,9	36	54	62,8	-8,8
17	26,8	58	68	109,2	58,8
18	29,9	51	99	114,2	-15,2
Tổng trung bình	215 11,94	758 42,11	1463 81,28	1463	0,00

Căn cứ vào kết quả phân tích về mặt nông học cho thấy giữa a, b tương quan độc lập và quan hệ của nó đối với y có dạng tuyến tính. Do vậy phương trình quan hệ giữa lân vô cơ (a) lân hữu cơ (b) với lân dễ tiêu (y) có dạng:

$$Y = A_0 + Aa + Bb.$$

Để xác định A_0 , A, B ta có thể dựa vào hệ ma trận sau:

$$\Sigma y = nA_0 + A\Sigma a + B\Sigma b \tag{1}$$

$$\Sigma ay = A_0\Sigma a + A\Sigma a^2 + B\Sigma ba \tag{2}$$

$$\Sigma by = A_0\Sigma b + A\Sigma ab + B\Sigma b^2 \tag{3}$$

Để tiện cho việc tính, ta có thể dựa vào các công thức rút gọn sau:

$$A = \frac{\sum (b - \bar{b})^2 \sum (a - \bar{a})(y - \bar{y}) - \sum (b - \bar{b}) * (a - \bar{a}) \sum (b - \bar{b})(y - \bar{y})}{D}$$

$$B = \frac{\sum (a - \bar{a})^2 \sum (b - \bar{b})(y - \bar{y}) - \sum (b - \bar{b}) * (a - \bar{a}) \sum (a - \bar{a})(y - \bar{y})}{D}$$

Trong đó: $D = \sum (a - \bar{a})^2 \sum (b - \bar{b})^2 - [\sum (a - \bar{a})(b - \bar{b})]^2$

$$Y = \bar{y} + A(a - \bar{a}) + B(b - \bar{b})$$

Kết quả tính ta có:

$$\sum (a - \bar{a})^2 = 1.752,96$$

$$\sum (b - \bar{b})^2 = 3.155,78$$

$$\sum (a - \bar{a})(b - \bar{b}) = 1.085,61$$

$$\sum (a - \bar{a})(y - \bar{y}) = 3.231,48$$

$$\sum (b - \bar{b})(y - \bar{y}) = 2.216,44$$

$$\sum (y - \bar{y})^2 = 12.389,61$$

$$D = 1752,96 * 3155,78 - (1085,61)^2 = 4353,4$$

$$A = \frac{3155,78 * 3231,48 - 1085,61 * 2216,44}{4353,4} = 1,7898$$

$$B = \frac{1752,96 * 2216,44 - 1085,61 * 2231,48}{4353,4} = 0,0866$$

Phương trình hồi quy có dạng:

$$\hat{y} = 81,28 + 1,7898(a - 11,94) + 0,0866(b - 42,11)$$

$$\hat{y} = 56,26 + 1,789a + 0,0866b$$

Từ phương trình trên chúng ta thấy rằng: cứ tăng 1ppm lân vô cơ trong đất có thể tăng thêm 1,7898 ppm lân dễ tiêu, nhưng khi tăng 1ppm lân hữu cơ thì chỉ tăng thêm được 0,0866 ppm lân dễ tiêu.

Chúng ta có thể dùng phương trình hồi quy trên để tính ra lượng lân dễ tiêu ở mỗi mẫu đất.

Thí dụ, ở mẫu đất một ta có:

$$\hat{Y}_1 = 56,26 + 1,7898 * 0,4 + 0,0866 * 53 = 61,6 \text{ ppm.}$$

\hat{Y}_1 so với Y_1 thực có sự sai khác:

$$Y_1 - \hat{Y}_1 = 64 - 61,6 = 2,4 \text{ ppm}$$

Xác định độ tin cậy của phương trình:

Kết quả tính toán ta có:

$$SSTo = \sum (y - \bar{y})^2 = 12389,61$$

$$SST = A \sum (a - \bar{a})(y - \bar{y}) + B \sum (b - \bar{b})(y - \bar{y}) = 1,7898 * 3231,48 + 0,0866 * 2216,44 = 5976,6$$

$$SS_E = SSTo - SST = 12389,61 - 5976,6 = 6414,0$$

$$MST = \frac{SST}{dfT} = \frac{5976,6}{2} = 2987,8$$

$$MSE = \frac{SSE}{dfe} = \frac{6414,0}{18-1-2} = 427,6$$

$$F = \frac{MST}{MSE} = \frac{2987,8}{427,6} = 6,99$$

Tra bảng $F_{0,01, 2, 15} = 3,67$

Kết luận: vai trò của lân vô cơ và lân hữu cơ có tác dụng quyết định đến lượng lân dễ tiêu ở trong đất.

Xác định độ tin cậy của hệ số hồi quy riêng:

Công thức (**) sau khi rút gọn ta có:

$$S_{a.b}^2 = MS_E^2 * \frac{\sum (b - \bar{b})^2}{D} = 427,6 * \frac{3.155,78}{4.353,4} = 0,31$$

$$S_{b.a}^2 = MS_E^2 * \frac{\sum (a - \bar{a})^2}{D} = 427,6 * \frac{1.752,96}{4.353,4} = 0,1723$$

$$S_{b.a} = 0,415 ; S_{a.b} = 0,557$$

$$t_a = \frac{a}{S_{a.b}} = \frac{1,7898}{0,557} = 3,21$$

$$t_b = \frac{b}{S_{b.a}} = \frac{0,0866}{0,415} = 0,21$$

$$t_{0,01/f=16} = 2,921$$

Kết luận: hệ số hồi quy riêng của lân vô cơ và lân dễ tiêu ở trong đất tin cậy được ở mức xác suất 0,01. Vì $t_a > t_{0,01}$. Còn hệ số hồi quy riêng của lân hữu cơ và lân dễ tiêu ở trong đất là không tin cậy được vì $t_b < t_{0,01}$.

Hay nói một cách khác là mối tương quan giữa lân vô cơ và lân dễ tiêu ở trong đất là chặt chẽ, còn lân hữu cơ và lân dễ tiêu ở trong đất là không chặt.

Thí dụ 8-6. Nghiên cứu ảnh hưởng của một số yếu tố thời tiết đến năng suất bông, kết quả số liệu quan sát ghi lại trong bảng (8 - 9)(theo Vũ Công Hậu).

a là lượng mưa (mm) từ gieo đến ra hoa

b là lượng mưa (mm) từ ra hoa đến nở quả.

c là độ ẩm tương đối của không khí (%) từ ra hoa đến nở quả.

y là sản lượng bông hạt (tạ/ha)

\hat{Y}_1 là sản lượng tính được từ công thức (tạ/ha).

Trong thí dụ (8-5) cho thấy sự sai khác năng suất giữa các năm chủ yếu là do yếu tố thời tiết quyết định; vì ở đây các yếu tố khác là cố định.

Bảng 8-9

Năm	a	b	c	y	\hat{Y}	y- \hat{Y}
1957	52	40	81	5,5	5,58	-0,08
1958	33	37	90	2,1	2,07	0,03
1959	72	95	66	20,5	20,45	0,05
1960	15	58	40	9,6	10,02	-0,42
1961	40	20	75	1,7	1,36	0,34
1962	32	41	80	3,8	3,87	-0,07
1963	76	54	83	10,3	10,63	-0,33
1964	10	85	70	11,7	11,71	-0,01
1965	68	70	65	15,2	15,07	0,13
1966	57	109	45	24,4	24,10	0,30
1967	24	62	64	9,3	9,10	0,20
1968	46	75	71	13,0	15,2	-0,20
1969	35	55	82	6,5	6,4	0,10
1970	54	68	63	13,8	13,47	0,33
1971	66	82	68	16,6	16,98	-0,38
Cộng	648	949	1043	164,0	164,01	-0,01

Từ số liệu thực nghiệm kết hợp với việc phân tích về quan hệ giữa các yếu tố thời tiết ta có thể chọn phương trình mẫu:

$$Y = A_0 + Aa + Bb + Cc.$$

Và hệ ma trận để xác định các hệ số A_0 , A, B, C như sau:

$$\Sigma y = nA_0 + A\Sigma a + B\Sigma b + C\Sigma c \quad (1)$$

$$\Sigma ay = A_0\Sigma a + A\Sigma a^2 + B\Sigma ab + C\Sigma ac \quad (2)$$

$$\Sigma by = A_0\Sigma b + A\Sigma ab + B\Sigma b^2 + C\Sigma bc \quad (3)$$

$$\Sigma cy = A_0\Sigma c + A\Sigma ac + B\Sigma bc + C\Sigma c^2 \quad (4)$$

Thay các trị số thực nghiệm vào hệ ma trận trên ta có:

$$164 = 15A_0 + 684A + 949B + 194C \quad (1)$$

$$8385,4 = 684A_0 + 37100A + 45.007B + 47949C \quad (2)$$

$$12.461,8 = 949A_0 + 45.007A + 67989B + 6326C \quad (3)$$

$$10.612,3 = 1043A_0 + 47.949A + 63.326B + 75.135C \quad (4)$$

Giải hệ ma trận ta được:

$$A_0 = 1,0950; \quad A = 0,1010; \quad B = 0,2037; \quad C = -0,1101.$$

Phương trình quan hệ giữa sản lượng bông và các yếu tố thời tiết được thiết lập.

$$y = 1,095 + 0,1010a + 0,2037b - 0,1101c$$

Kiểm tra độ tin cậy của phương trình:

$$\text{Biến động toàn bộ } SST_0 = \sum (y_i - \bar{y})^2 = 815,5$$

Biến động do các yếu tố nghiên cứu:

$$\begin{aligned} SST &= A \sum (a - \bar{a})(y_i - \bar{y}) + B \sum (b - \bar{b})(y_i - \bar{y}) + C \sum (c - \bar{c})(y_i - \bar{y}) \\ &= 0,101 \times 929,8 + 0,2037 \times 2128,6 \times (-0,1101) \times (-941,9) = 631,2 \end{aligned}$$

Biến động do các yếu tố không nghiên cứu:

$$SSE = SST_0 - SST = 815,5 - 631,2 = 184,3.$$

Phương sai của yếu tố nghiên cứu:

$$MST = \frac{SST}{dfT} = \frac{631,2}{3} = 210$$

Phương sai của yếu tố không nghiên cứu

$$MSE = \frac{SSE}{dfE} = \frac{184,3}{11} = 16,7$$

$$F_m = \frac{MST}{MSE} = \frac{210}{16,7} = 12,6$$

$$F_{0,01, 3, 11} = 6,22$$

Như vậy chúng ta thấy mối quan hệ giữa các yếu tố khí tượng nghiên cứu có quan hệ chặt chẽ với sản lượng bông hạt. Trong điều kiện đất đai và biện pháp kỹ thuật tương tự như nhau thì 3 yếu tố nghiên cứu trên có tác dụng quyết định lớn đến năng suất

- Thời gian từ gieo đến nở hoa lượng mưa tăng 1mm thì năng suất tăng được 0,1010 tạ/ha.
- Thời gian từ ra hoa đến nở quả lượng mưa tăng 1mm thì năng suất tăng được 0,2037 tạ/ha.
- Thời gian từ ra hoa đến nở quả độ ẩm không khí tăng 1% thì năng suất bông giảm 0,1101 tạ/ha.

8.4. Phương pháp lập phương trình hồi qui cho các đặc trưng chất lượng.

Đặc điểm của các loại tài liệu này là không đo đếm cụ thể được, mà chỉ có thể được biểu hiện bằng chữ (có) hay (không). Ta cũng có thể thay chữ có bằng dấu (+) hay bằng số (1), chữ không bằng dấu (-) hay số (0). Ta tìm hiểu nội dung phương pháp này thông qua thí dụ sau:

Thí dụ 8-7.

Điều tra kỹ thuật trồng lạc ở 70 hợp tác xã trên một số mặt sau đây:

- Chọn hạt và xử lý cho nứt mầm (a)
- Cày để ải trước khi gieo hạt ít nhất 2 tháng (b).
- Đào mương thoát nước quanh ruộng lạc (c).
- Bón phân chuồng ải (d).
- Bón lân và vôi (e).
- Phun thuốc boócđô trừ bệnh đốm lá (f).

- Gieo đúng thời vụ vào tháng hai (g).

Kết quả điều tra được ghi lại như bảng (8 - 10)

Năng suất lạc tùy theo biện pháp kỹ thuật khác nhau, theo Vũ Công Hậu (phương pháp thống kê trong thí nghiệm nông nghiệp. Nhà xuất bản nông thôn 1974)

Bảng 8 - 10.

Chỉ tiêu	Số liệu hệ thống biện pháp kỹ thuật áp dụng									
	1	2	3	4	5	6	7	8	9	10
Chọn hạt và sử lý (a)	+	+	+	-	+	-	+	-	-	-
Cày ải (b)	+	+	+	-	+	+	-	+	-	+
Thoát nước (c)	+	+	+	+	-	-	+	+	+	-
Phân chuồng (d)	+	+	-	+	-	-	-	-	-	-
Phân lân + vôi (e)	+	+	-	+	-	+	+	+	+	-
Phun bécđô (f)	+	-	-	+	+	+	-	-	-	+
Thời vụ gieo hạt (g)	-	-	+	-	+	-	-	-	-	-
Số hợp tác xã điều tra	4	4	8	3	9	15	4	7	2	11
Năng suất trung bình	25,3	22,4	20,5	19,0	18,0	16,5	15,1	13,8	12,0	11,2

Dòng đầu cho mười cột đánh số từ 1 đến 10 là số hiệu biện pháp kỹ thuật áp dụng, biện pháp nào áp dụng thì đánh dấu +, không áp dụng thì đánh dấu -, còn hai dòng cuối cùng là số lượng hợp tác xã đã áp dụng hệ thống biện pháp kỹ thuật và năng suất trung bình của các hợp tác xã đó tính bằng tạ lạc vỏ trên 1ha.

Để thấy rõ ảnh hưởng của từng biện pháp kỹ thuật tới năng suất ta có thể dùng phương trình sau để biểu diễn:

$$S = Q + Aa + Bb + Cc + Dd + Ee + Ff + Gg$$

Trong đó : S là năng suất ở mỗi nhóm hợp tác xã.

Q là trị số nói lên ảnh hưởng của các yếu tố khác mà trong điều tra không đề cập tới A, B, C...,G là những hệ số nói lên vai trò của từng yếu tố điều tra.

Để tìm được ẩn số của phương trình trên chúng ta áp dụng phương pháp bình phương nhỏ nhất để thiết lập phương trình mẫu sau.

$$\Sigma S = nQ + A\Sigma a + B\Sigma b + C\Sigma c + D\Sigma d + E\Sigma e + F\Sigma f + G\Sigma g$$

$$\Sigma aS = Q\Sigma a + A\Sigma a^2 + B\Sigma ba + C\Sigma ca + D\Sigma da + E\Sigma ea + F\Sigma fa + G\Sigma ga$$

$$\Sigma bS = Q\Sigma b + A\Sigma ab + B\Sigma b^2 + C\Sigma cb + D\Sigma db + E\Sigma eb + F\Sigma fb + G\Sigma gb$$

$$\Sigma cS = Q\Sigma c + A\Sigma ac + B\Sigma b^2 + C\Sigma c^2 + D\Sigma dc + E\Sigma ec + F\Sigma fc + G\Sigma gc$$

$$\Sigma dS = Q\Sigma d + A\Sigma ad + B\Sigma bd + C\Sigma cd + D\Sigma d^2 + E\Sigma ed + F\Sigma fd + G\Sigma gd$$

$$\Sigma eS = Q\Sigma e + A\Sigma ae + B\Sigma be + C\Sigma ce + D\Sigma de + E\Sigma e^2 + F\Sigma fe + G\Sigma ge$$

$$\Sigma fS = Q\Sigma f + A\Sigma af + B\Sigma bf + C\Sigma cf + D\Sigma df + E\Sigma ef + F\Sigma f^2 + G\Sigma gf$$

$$\Sigma gS = Q\Sigma g + A\Sigma ag + B\Sigma bg + C\Sigma cg + D\Sigma dg + E\Sigma eg + F\Sigma fg + G\Sigma g^2$$

Để tính được các giá trị thực nghiệm n, Σa , Σb ... các dấu ở bảng được thay bằng số 1 và số 0:

1 ứng với dấu (+), 0 ứng với dấu (-)... và các yếu tố không nghiên cứu được coi là như nhau ở tất cả các điểm điều tra ký hiệu bằng số 1 như bảng 8-11

Phương trình xuất phát về năng suất lạc tùy theo hệ thống biện pháp áp dụng.

Bảng 8 - 11.

Số liệu hệ thống biện pháp	Biện pháp kỹ thuật								Hợp tác xã trong nhóm	S	S'
	Q	A	B	C	D	E	F	G			
1	1	1	1	1	1	1	1	0	4	25,3	25,35
2	1	1	1	1	1	1	0	0	4	22,4	22,64
3	1	1	1	1	0	0	0	1	8	20,5	20,31
4	1	0	0	1	1	1	1	0	3	19,0	18,63
5	1	1	1	0	0	0	1	1	9	18,0	18,17
6	1	0	1	0	0	1	1	0	15	16,4	16,43
7	1	1	0	1	0	1	0	0	4	15,1	14,81
8	1	0	1	1	0	1	0	0	7	13,8	13,72
9	1	0	0	1	0	1	0	0	2	12,0	13,33
10	1	0	1	0	0	0	1	0	14	11,2	11,22
Cộng									70		

Trong bảng (8 - 11): S' được tính từ phương trình lý luận.

Thay n, Σa , Σb , Σc ,..., ΣS , Σa^2 , Σab ,..., ΣgS bằng các trị số điều tra được ta có hệ phương trình:

$$1159,1 = 70Q + 29A + 61B + 25C + 11D + 39E + 45F + 17G \quad (1)$$

$$577,2 = 29Q + 29A + 25B + 20C + 8D + 12E + 13F + 17G \quad (2)$$

$$1017,7 = 61Q + 25A + 61B + 16C + 8D + 30E + 42F + 17G \quad (3)$$

$$496,2 = 25Q + 20A + 16B + 25C + 11D + 17E + 7F + 8G \quad (4)$$

$$247,8 = 11Q + 8A + 8B + 11C + 11E + 7F + 0G \quad (5)$$

$$676,3 = 39Q + 12A + 30B + 17C + 11D + 39E + 22F + 0G \quad (6)$$

$$724,5 = 45Q + 13A + 42B + 7C + 7D + 22E + 45F + 9G \quad (7)$$

$$326,0 = 17Q + 17A + 17B + 8C + 0D + 0E + 9F + 17G \quad (8)$$

Chú ý: $\Sigma S = 1159,1 = 4 \times 25,3 + 4 \times 22,4 + \dots + 14 \times 11,2$

$$\Sigma a = 29 = 1 \times 4 + 1 \times 4 + 1 \times 8 + 1 \times 9 + 1 \times 4$$

$$\Sigma ab = 25 = 1 \times 1 \times 4 + 1 \times 1 \times 4 + 1 \times 1 \times 8 + 1 \times 1 \times 9.$$

Giải hệ phương trình trên ta thiết lập được phương trình lý luận, qua đó nói lên vai trò của các yếu tố điều tra tới năng suất lạc.

$$\hat{Y}_{abc} = 3,26 + 1,480a + 5,244b + 4,849c + 2,589d + 5,214e + 2,711f + 5,471g.$$

Qua phương trình thiết lập trên cho ta thấy: tất cả các yếu tố điều tra đều góp phần làm tăng năng suất lạc. Nhưng mỗi biện pháp có tầm quan trọng khác nhau như: cày ải, bón lân và vôi, thời vụ là quan trọng bậc nhất.

Tiêu nước trên ruộng là thứ nhì; bón phân chuồng, phun bécđô là thứ yếu còn chọn việc xử lý hạt giống tuy có làm tăng năng suất nhưng không đáng kể.

- Công thức và cách tính hệ số tương quan riêng.

Hệ số tương quan được tính từ ba biến số được gọi là hệ số tương quan cấp 1. Đó là trường hợp tính hệ số tương quan riêng của hai biến số bất kỳ chỉ có một biến số cô lập. Khi tính hệ số tương quan riêng cấp I trước hết ta phải tính hệ số tương quan cấp 0 tức là tương quan đơn r_{12} , r_{13} và r_{23} .

Công thức tính hệ số tương quan riêng cấp I.

$$R_{12.3} = \frac{r_{12} - r_{13} * r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

$$R_{13.2} = \frac{r_{13} - r_{12} * r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{23}^2)}}$$

$$R_{23.1} = \frac{r_{23} - r_{12} * r_{13}}{\sqrt{(1 - r_{12}^2)(1 - r_{13}^2)}}$$

$R_{12.3}$ là hệ số tương quan riêng giữa x_1 và x_2 khi x_3 cố định.

$R_{13.2}$ là hệ số tương quan riêng giữa x_1 và x_3 khi x_2 cố định.

$R_{23.1}$ là hệ số tương quan riêng giữa x_2 và x_3 khi x_1 cố định.

Hệ số tương quan riêng được tính ra từ bốn biến số gọi là hệ số tương quan riêng cấp 2. Khi tính toán hệ số tương quan riêng cấp 2 ta cô lập hai biến số. Khi tính R_{12-34} ; R_{13-24} trước tiên ta phải tính $R_{12.3}$; $R_{24.3}$; $R_{13.2}$; $R_{34.2}$ tức là hệ số tương quan riêng cấp 1.

Công thức tính hệ số tương quan riêng cấp 2.

$$R_{12.34} = \frac{R_{12.3} - R_{14.3} * R_{24.3}}{\sqrt{(1 - R_{14.3}^2)(1 - R_{24.3}^2)}}$$

$$R_{13.24} = \frac{R_{13.2} - R_{14.2} * R_{34.2}}{\sqrt{(1 - R_{14.2}^2)(1 - R_{34.2}^2)}}$$

$$R_{14.23} = \frac{R_{14.2} - R_{13.2} * R_{43.2}}{\sqrt{(1 - R_{13.2}^2)(1 - R_{43.2}^2)}}$$

Thí dụ 8 - 8: nghiên cứu tương quan năng suất với chiều cao và số quả của 30 cây bông.

Hệ số tương quan giữa năng suất (1) với số quả (2) $r_{12} = 0,86$. Hệ số tương quan giữa năng suất (1) với chiều cao (3): $r_{13} = 0,65$. Hệ số tương quan giữa số quả (2) và chiều cao (3): $r_{23} = 0,72$.

Khi chiều cao cây không thay đổi thì tương quan giữa năng suất và số quả được tính theo công thức sau.

$$R_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} = \frac{0,86 - 0,65 \cdot 0,72}{\sqrt{(1-0,65^2)(1-0,72^2)}} = 0,743$$

Còn khi cố định số quả thì tương quan giữa năng suất và chiều cao cây được tính theo công thức .

$$R_{13.2} = \frac{r_{13} - r_{12} \cdot r_{23}}{\sqrt{(1-r_{12}^2)(1-r_{23}^2)}} = \frac{0,65 - 0,86 \cdot 0,72}{\sqrt{(1-0,86^2)(1-0,72^2)}} = 0,087$$

Xác định độ tin cậy của hệ số tương quan riêng.

Khi dùng tiêu chuẩn t để xác định độ tin cậy của hệ số tương quan riêng, độ tự do dùng không giống như xác định t cho hệ số tương quan đơn giản. (khi xác định t cho hệ số tương quan đơn giản, độ tự do của nó là: N - 2). Nhưng khi xác định t cho hệ số tương quan riêng độ tự do của nó là N - P - 2. Trong đó P là số biến số bị cô lập.

Ở hệ số tương quan riêng cấp 1, chỉ có một biến số bị cô lập nên P = 1.

Ở hệ số tương quan riêng cấp 2 thì P = 2. Công thức xác định t của hệ số tương quan riêng như sau:

$$t = \frac{R}{\sqrt{1-R^2}} \cdot x\sqrt{N-P-2}$$

Sau khi tính t ta tra bảng t lý luận ở độ tự do là N - P - 2. Nếu $t_m > t_{0,5}$ chứng tỏ hai yếu tố đó có tương quan rõ, như thí dụ (8 - 8).

$$\begin{cases} R_{12.3} = 0,743 \\ N = 30 \\ P = 1 \end{cases}$$

Thay vào công thức trên ta được:

$$t_2 = \frac{0,743}{\sqrt{1-(0,743)^2}} \cdot x\sqrt{30-1-2} = 5,77$$

Và $R_{13.2} = 0,087$

$$t_3 = \frac{0,087}{\sqrt{1-(0,087)^2}} \cdot x\sqrt{30-1-2} = 0,45$$

Tra bảng t ở độ tự do bằng 30 - 1 - 2 = 27 ta được:

$$t_{0,5} = 2,05$$

Ta thấy $t_2 > t_{.05}$ còn $t_3 < t_{.05}$.

Như vậy ta có thể kết luận được rằng năng suất của bông có tương quan chặt chẽ với số quả chứ không có quan hệ với chiều cao. Điều này chứng tỏ mối quan hệ $R_{13} = 0,65$ là tương quan giả tạo

- Hệ số tương quan bội.

Ta có 3 biến số x_1, x_2, x_3 . Hệ số tương quan bội giữa x_1 và 2 biến số x_2, x_3 có thể ký hiệu là R_{1-23} hay R_{1-32} hệ số tương quan bội giữa x_2 với x_1 và x_3 ký hiệu là R_{2-13} hay R_{2-31} .

Công thức tính hệ số tương quan bội 3 biến như sau:

$$R_{1,23} = \sqrt{1 - (1 - r_{12}^2)(1 - R_{13,2}^2)}$$

Công thức tính hệ số tương quan bội giữa 4 biến. Hệ số tương quan bội của X_1 với 3 biến X_2, X_3, X_4 như sau:

$$R_{1,234} = \sqrt{1 - (1 - r_{12}^2)(1 - R_{13,2}^2)(1 - R_{14,23}^2)}$$

Thí dụ 8 - 9: tìm hiểu mối quan hệ tổng hợp giữa năng suất lúa với hàm lượng mùn và hàm lượng lân ở trong đất hợp tác xã Đại Lai - Gia Lương - Hà Bắc với số mẫu phân tích $n = 21$ ta có:

- + Hệ số tương quan giữa năng suất (1) và mùn (2) $r_{12} = 0,764$.
- + Hệ số tương quan giữa năng suất (1) và lân (3) $r_{13} = 0,871$.
- + Hệ số tương quan giữa năng suất và lân khi cô lập mùn:

$$R_{13,2} = 0,548$$

Thay các trị số vào công thức trên ta có:

$$R_{1-23} = \sqrt{1 - (1 - 0,764^2)(1 - 0,548^2)}$$

$$R_{1-23} = 0,8268$$

Xác định độ tin cậy của hệ số tương quan bội.

Để xác định độ tin cậy của hệ số tương quan bội ta có thể dùng tiêu chuẩn F.

$$F = \frac{R^2}{(1 - R^2)} \times \frac{(N - P - 1)}{P}$$

Trong đó: R: hệ số tương quan bội.

N: số cặp biến số

P: số biến số trừ đi 1.

Sau khi tính F_m , tra bảng F ở bậc tự do bằng $\frac{P}{N - P - 1}$

Nếu $F_m > F_{.05}$ ta kết luận rằng, tương quan chặt và ngược lại.

$F_m \leq F_{.05}$ kết luận là chúng không có tương quan, hay tương quan không chặt.

Từ thí dụ (8 - 9) ta thay các số đã biết vào công thức tính F_m sẽ có:

$$F_m = \frac{0,8268^2}{(1 - 0,8268^2)} \times \frac{(21 - 2 - 1)}{2} = 19,44$$

Tra bảng $F_{(0,05, 2, 18)} = 3,55$

$F_m > F_{.05}$ chứng tỏ mối tương quan này rất rõ. Có nghĩa là năng suất lúa trong điều kiện nghiên cứu phụ thuộc rất lớn vào hàm lượng lân và mùn trong đất.

Câu hỏi ôn tập

1. Hãy cho biết ý nghĩa của phân tích tương quan - Hồi qui
2. Hãy tìm một ví dụ cụ thể (từ các nghiên cứu trước) cho bài tập phân tích tương quan hồi qui đường thẳng một biến.

CHƯƠNG IX TỔNG KẾT THÍ NGHIỆM

Đây là công việc cuối cùng của các nhà khoa học sau khi đã có kết quả nghiên cứu khoa học. Các kết quả này sẽ được báo cáo trong các hội nghị khoa học và gửi đi đăng báo cũng như các tạp chí khoa học chuyên ngành. Song các kết quả này trước khi báo cáo tại hội nghị khoa học cần phải được trình bày trong bản báo cáo chi tiết. Chương này trang bị cho người học biết cách viết báo cáo khoa học:

- Biết trình bày số liệu trong báo cáo khoa học (Bảng biểu, đồ thị hoặc biểu đồ minh họa)
- Biết cách trình bày một báo cáo khoa học.

9.1. Cách trình bày số liệu trong báo cáo khoa học

Thông thường các số liệu được trình bày bằng bảng số liệu và có thể được minh họa qua đồ thị hoặc biểu đồ.

9.1.1. Phương pháp trình bày bằng bảng số liệu

Đây được coi là phương pháp cơ bản cho việc trình bày các số liệu nghiên cứu khoa học. Cấu tạo của một bảng số liệu gồm các phần sau đây:

1. Số thứ tự của bảng (thường được đánh giá số theo thứ tự 1, 2, 3, 4...). Đôi khi có kèm theo các chỉ số như 2a, 3b... trong đó số thể hiện nội dung chính còn chữ thể hiện nội dung chi tiết của nội dung chính đó.

2. Tên bảng: tên bảng ghi ngắn gọn, tóm tắt, nhưng phải thể hiện đầy đủ nội dung các số liệu trong bảng.

Chú ý: Tên bảng có thể ghi đơn vị đo của số liệu thể hiện trong bảng nếu là thống nhất trong toàn bảng (bảng chữ trình bày một loại chỉ tiêu).

3. Thân bảng gồm các hàng ngang và các cột dọc. Các hàng ngang nên thể hiện nguyên nhân ảnh hưởng, còn các cột dọc thể hiện kết quả của ảnh hưởng. Không nên để quá nhiều cột dọc chỉ nên từ 6 - 8 cột là vừa. Còn nếu nhiều nên tách thành các bảng giả sử 2a, 2 b...

4. Phần nội dung ghi trong bảng: khi ghi chép số liệu cần chú ý

Mỗi chỉ tiêu (kết quả) tùy theo hàng ngang hay cột dọc trong cách trình bày nhưng phải thống nhất phần số lẻ (số thập phân), phần số thập phân lấy nhiều hay ít tùy thuộc vào loại chỉ tiêu nghiên cứu và độ chính xác của các chỉ tiêu ấy.

Thí dụ về một bảng số liệu

Bảng 9.1. Ảnh hưởng của phân bón bón (phân hữu cơ) đến các yếu tố năng suất và năng suất lúa đặc sản vụ mùa

Chi tiêu Công thức	Số bông / khóm	Số hạt/bông	Số hạt chắc /bông	TL (%) hạt chắc	P 1000 hạt (gam)	Năng suất (tạ/ha)	
						Lý tuyết	Thực thu
I (Đ/c)	10,0	154	105	67,9	19,20	42,34	32,4
II	11,2	152	99	65,4	19,27	45,05	33,2
III	11,3	179	128	71,3	19,29	58,59	33,3
IV	12,2	181	129	71,6	19,34	63,69	34,5
V	11,6	181	122	67,5	19,27	57,22	31,7

Chú ý: khi đã quyết định lấy phần số lẻ thống nhất nếu là số 0 cũng phải viết đủ.

9.1.2. Phương pháp dùng đồ thị và biểu đồ

Trong báo cáo khoa học người ta thường dùng đồ thị hoặc biểu đồ để mô tả những hiện tượng hoặc quá trình của hiện tượng. Đồ thị có tác dụng làm sáng tỏ hình tượng tương đối phức tạp mà bảng thống kê khó nhận biết, nhưng đồ thị lại không thể hiện được độ chính xác cụ thể của số liệu mà chỉ cho ta một xu hướng.

Nhiệm vụ chủ yếu của đồ thị và biểu đồ là để so sánh những số liệu thống kê, biểu thị kết cấu của tổng thể biểu thị quá trình phát triển của hiện tượng hoặc diễn đạt tình hình phân phối giữa các đơn vị trong một tổng thể. Có nhiều loại biểu đồ như biểu đồ hình cột, biểu đồ khối chữ nhật, biểu đồ hình bán tròn và biểu đồ phân phối.

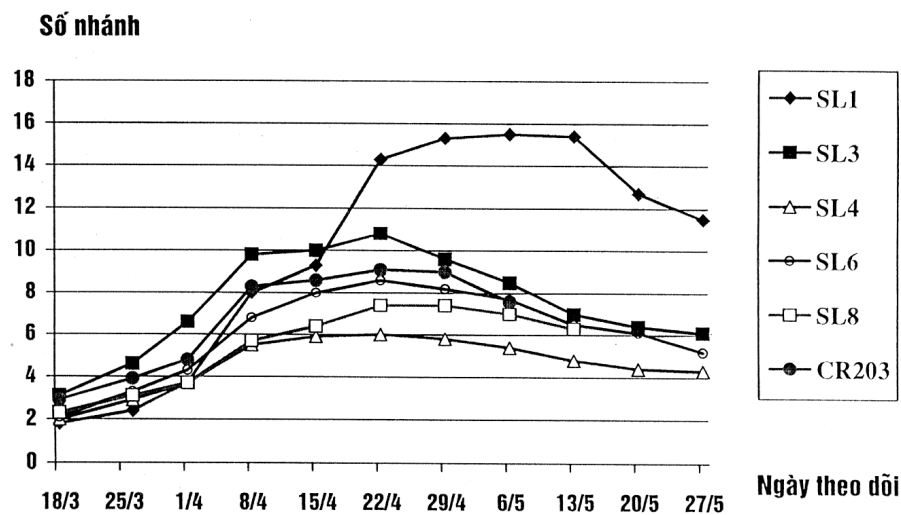
Còn đồ thị thì có dạng ta thường gọi là là đồ thị đường biểu diễn.

Cách vẽ đồ thị và biểu đồ nên chú ý như sau:

* Đồ thị đường biểu diễn là đồ thị dùng đường biểu diễn lên xuống để biểu thị sự diễn biến thay đổi của hiện tượng. Nó có thể phản ánh quá trình phát triển của hiện tượng (động thái). Loại đồ thị này thường dùng để biểu thị tình hình biến đổi về số lượng của các hiện tượng thay đổi theo thời gian. Điều này thể hiện qua đồ thị (9.1)

Đồ thị đường biểu diễn thì trục hoành biểu diễn thời gian, còn trục tung biểu diễn kết quả của chỉ tiêu nghiên cứu theo thời gian (xem đồ thị 9.1).

Trên đồ thị biểu diễn nhiều đường có thể dùng ký hiệu hoặc màu sắc khác nhau để vẽ mô tả giúp người đọc dễ dàng nhận biết.



Đồ thị 9.1: Động thái đẻ nhánh của các dòng, giống vật liệu siêu lúa có triển vọng đại diện vụ xuân 2002

Đồ thị đường biểu diễn thì trục hoành biểu diễn thời gian, còn trục tung biểu diễn kết quả của chỉ tiêu nghiên cứu theo thời gian xem đồ thị 9.1

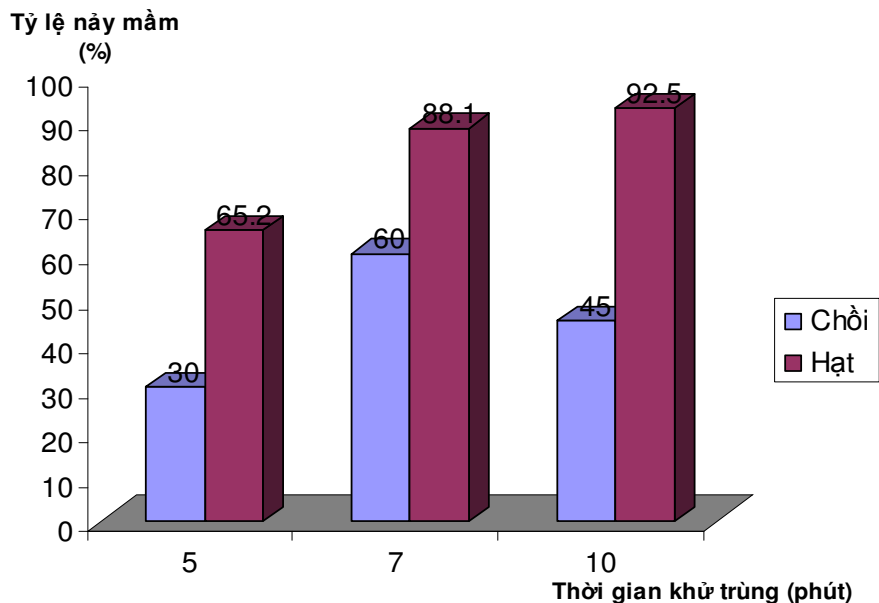
Trên đồ thị biểu diễn nhiều đường có thể dùng ký hiệu hoặc màu sắc khác nhau để vẽ mô tả giúp người đọc dễ dàng nhận biết.

* Biểu đồ:

Tùy thuộc vào loại chỉ tiêu biểu diễn mà có thể dùng cho phù hợp với biểu đồ sau:

+ Biểu đồ hình khối chữ nhật:

Đây là dạng biểu đồ hay dùng biểu diễn kết quả cụ thể của một chỉ tiêu tại một thời điểm nào đó. Thí dụ như biểu diễn năng suất của cây trồng theo các công thức nghiên cứu khác nhau.

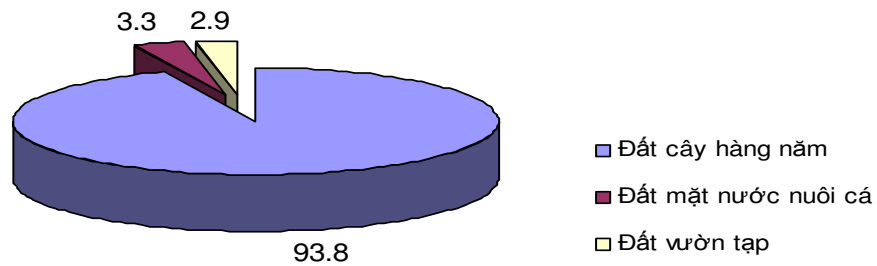


Biểu đồ 9.2. Ảnh hưởng của thời gian khử trùng mẫu đến tỷ lệ mẫu sống

Biểu đồ này chỉ đánh giá kết quả qua độ cao của khối chữ nhật nên khi vẽ lưu ý cho bằng nhau về chiều ngang của khối và khoảng cách giữa các công thức cũng phải đều.

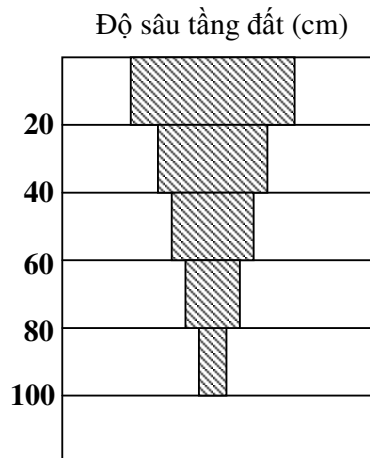
+ Biểu đồ hình bán tròn:

Dạng này hay được dùng biểu diễn cơ cấu của một tổng thể (tính theo % tương đối). Ta có thể xem qua mô tả của biểu đồ (9.3) sau:



Biểu đồ 9.3. Hiện trạng sử dụng đất nông nghiệp của huyện Từ Sơn - Bắc Ninh năm 1998

+ Biểu đồ phân bố thực ra cũng có thể biểu diễn qua biểu đồ hình bánh tròn. Song nếu như dùng biểu đồ hình bánh tròn thì hình dạng không cụ thể, còn nếu dùng biểu đồ phân phối sẽ cụ thể và rõ ràng hơn. Thí dụ như biểu diễn trọng lượng hay tỷ lệ % rễ cây phân bố ở các độ sâu tầng đất khác nhau (biểu đồ 9.4).



Biểu đồ 9.4. Khối lượng rễ cây ngô phân phối ở các độ sâu tầng đất trồng

9.2. Phương pháp tổng kết và viết báo cáo kết quả thí nghiệm (kết quả nghiên cứu khoa học)

Tổng kết thí nghiệm là một khâu quan trọng trong công tác nghiên cứu khoa học. Nó phản ánh toàn bộ kết quả nghiên cứu thí nghiệm, chỉ cần số liệu xác thực, phương pháp nghiên cứu chính xác thì dù kết quả nhiều hay ít đều khẳng định được ý nghĩa khoa học. Công việc tổng kết thí nghiệm gồm các nội dung sau:

- Chính lý số liệu, tính toán thống kê các tham số đại diện và dùng các tiêu chuẩn thống kê để xử lý và từ đó rút ra kết luận.
- Viết báo cáo khoa học.
- Giám định kết quả thí nghiệm.

9.2.1. Chính lý số liệu, tính các tham số thống kê và xử lý thống kê kết quả thí nghiệm

Sau khi các số liệu đã được tính toán từ số liệu thô ra số liệu tinh theo các tham số thống kê cần thiết, các số liệu này phải được sắp xếp vào các bảng và cần thiết có thể dùng đồ thị và biểu đồ để minh họa. Quá trình sắp xếp ta dễ dàng đánh giá và so sánh kết quả. Vì vậy khi tính toán cần hết sức cẩn thận tránh sự nhầm lẫn.

Ngoài việc so sánh các kết quả trong bảng còn cần so sánh các số liệu trong bảng khác nhau. Có nhận xét trên nhiều số liệu, phân tích một cách toàn diện các số liệu thì kết luận rút ra mới khách quan và chính xác.

Riêng số liệu về năng suất thí nghiệm thì cần đặc biệt quan tâm nhiều hơn. Vì đây là mục tiêu cuối cùng đánh giá thành công của thí nghiệm. Do đó cần phân tích kỹ lưỡng qua phân tích phương sai để có độ tin cậy cao.

9.2.2. Viết báo cáo khoa học

Viết báo cáo khoa học là trình bày một cách thật sự khoa học và logic. Song ngắn gọn kết quả của công trình nghiên cứu.

Nội dung của một báo cáo khoa học có thể bao gồm các vấn đề như sau:

+ Tên đề tài nghiên cứu (hay tên thí nghiệm) như đã trình bày trong đề cương nghiên cứu.

+ Tên tác giả, ghi rõ tên người chủ trì và các cộng tác viên.

+ Mở đầu:

Phần này nêu tính cấp thiết của vấn đề nghiên cứu, mục đích và yêu cầu. Chủ yếu là nói rõ lý do dẫn đến việc thực hiện đề tài này và những vấn đề gì cần giải quyết ý nghĩa khoa học và thực tiễn. Viết ngắn gọn.

+ Chương I: Tổng quan về vấn đề nghiên cứu

Phần này nêu bật được cơ sở khoa học của đề tài nghiên cứu, cũng như các kết quả nghiên cứu có liên quan mà các nhà khoa học trên thế giới và trong nước đã làm được.

+ Chương II: Nội dung và phương pháp nghiên cứu

Trong chương này nêu được địa điểm, đối tượng, thời gian cũng như các phương pháp nghiên cứu đã được sử dụng trong đề tài cùng các nội dung và các chỉ tiêu nghiên cứu mà đề tài đặt ra.

+ Chương III : Kết quả nghiên cứu và thảo luận

Việc sắp xếp các kết quả tùy thuộc vào phương pháp trình bày song cần đảm bảo tính logic khoa học và khách quan, đây được coi là phần chính của báo cáo khoa học.

+ Phần kết luận và đề nghị

Kết luận nêu tóm tắt những nhận xét về kết quả cuối cùng của đề tài từ đó có thể đề xuất hướng giải quyết cũng như chủ trương phục vụ sản xuất. Vì vậy kết luận phải khách quan, chính xác. Chỉ kết luận những gì đã làm được có kết quả có các số liệu chứng minh. Không suy luận, không kết luận chung chung và mượn kết quả ở chỗ khác để đưa thêm vào báo cáo. Tuy nhiên, kết luận không nói chi tiết quá và cô đọng quá cần tìm ra những điểm mới, điểm hay để đề xuất.

Đề nghị ghi ngắn gọn những đề xuất được cần tiếp tục thực hiện để đề tài hoàn thiện hơn và khi áp dụng vào sản xuất đạt hiệu quả cao hơn thiết thực hơn.

+ Phần phụ lục và tài liệu tham khảo

Mục này gồm có 2 nội dung sau:

- Phụ lục

Nếu có các số liệu thống kê không cần thiết trình bày ở chương III (kết quả nghiên cứu) thì trình bày ở phụ lục.

- Tài liệu tham khảo

Cuối bản báo cáo khoa học cần phải có mục lục về tài liệu đã tham khảo sử dụng cho báo cáo của mình. Các tài liệu này phải được sắp xếp theo thứ tự A, B, C... của tên tác giả, mỗi tài liệu phải có tên tác giả rồi đến tên sách hoặc tên bài báo trong tạp chí, nhà xuất bản và năm xuất bản. Nếu là bài báo phải nêu trang của tạp chí, tên tạp chí và số tạp chí.

Vì vậy báo cáo khoa học cần phải đạt một số điều kiện sau:

- Tính chính xác: tất cả các tài liệu, số liệu, các bảng số liệu, đồ thị, phương pháp nghiên cứu đều chuẩn và khách quan. Phải có các tiêu chuẩn (test) đánh giá cụ thể và cuối cùng là sai số thí nghiệm có đáng tin cậy hay không.

- Tính khách quan: kết quả nghiên cứu cần phải phản ánh tính chất thực tế khách quan, không theo ý chủ quan của mình.

- Tính thực tế: kết quả nghiên cứu khoa học phải được ứng dụng trong sản xuất của người dân hoặc sử dụng về mặt khoa học tùy thuộc vào vấn đề nghiên cứu.

- Tính ổn định: điều này có nghĩa là kết quả phải mang tính quy luật có như vậy thì tính thực tế sẽ hiệu quả và thiết thực.

- Tính sáng tạo: đây là điều mà ở tất cả các vấn đề đã nghiên cứu đặt ra. Bởi vì khoa học là luôn tìm ra cái mới, cái sáng tạo hơn so với cái cũ. Tùy thuộc vào vấn đề nghiên cứu và kinh phí cũng như thời gian mà tính sáng tạo ở những mức độ khác nhau.

PHỤ LỤC
MỘT SỐ THUẬT NGỮ DÙNG TRONG GIÁO TRÌNH

Thuật ngữ	Anh
Bậc tự do	Degree of freedom
Các số đặc trưng của mẫu	Statistics ; Statistical measures; Characteristics of a sample
Công thức (mức)	Treatment
Chênh lệch nhỏ nhất	Least significant difference (LSD)
Dung lượng mẫu (kích thước mẫu; cỡ mẫu)	Size of sample
Độ lệch chuẩn	Standart diviation
Độ tin cậy	Degree of confidence
Giả thiết thống kê	Statistical hypothesis
Giả thiết không (H_0)	Null hypothesis
Đôi thiết	Alternative hypothesis
Hàm phân phối	Distribution function
Hồi quy tuyến tính	Linear regression
Hệ số biến động	Coefficient of variation
Hệ số tương quan	Correlation coefficient
Khối	Block
Kiểm định giả thiết	Tests of hypothesis; Testing hypothesis
Kỳ vọng toán	Mathematical expectation
Mẫu	Sample
Mod	Mode
Nguyên tắc bình phương bé nhất	Method (principle) of least square
Nhắc lại	Replication
Phân phối chuẩn	Normal distribution; Gaussian distribution
Phân phối khi bình phương	Khi-square distribution
Phân phối Fisher (F)	Fisher-Snedercordistribution; F distribution
Phân phối nhị thức	Binomial distribution
Phân phối Stiuđon	Student distribution; T distribution
Phân phối	Poatxon Poisson distribution
Phép thử (thí nghiệm)	Experiment
Phương sai; trung bình bình phương	Variance (dispersion MS)
Phân tích phương sai	Analysis of variance (ANOVA)
Sai số chuẩn	Standart error
Sai số ngẫu nhiên	Error
Số liệu định lượng	Quantitative data
Số liệu định tính	Qualitative data
Sắp xếp (thiết kế)	Design
Sắp xếp hoàn toàn ngẫu nhiên	Completely randomized design (CRD)
Sắp xếp khối ngẫu nhiên đầy đủ	Radomized complete Block design (RCB)
Sắp xếp ô vuông latin	Latin square design (LS)
Sắp xếp chia ô lớn-ô nhỏ	Split-plot design

Tổng bình phương	Sum of squares (SS)
Tổng bình phương toàn phần	Total sum of square
Tổng bình phương công thức	Treatment sum of square
Tổng bình phương nhắc lại	Replication sum of square
Tổng bình phương ngẫu nhiên	Error sum of square
Tần số	Frequency
Tần suất (xác suất)	Probability
Tổng thể	Population
Trung bình cộng (Bình quân mẫu)	Mean; Sample mean; Arthmatic mean; Average
Trung vị	Median
Tương quan	Correlation
Ước lượng	Estimatie; Estimation of parameters
Ước lượng điểm	Point estimate
Ước lượng khoảng (khoảng tin cậy)	Interval estimate
Ước lượng khoảng của kỳ vọng	Interval estimtion of mean (Confidence for mean)
Ước lượng khoảng của xác suất	Interval estimation of probability (Confidence interval for probability)

Bảng 13. Sơ đồ Ô vuông la tinh mẫu

3 x 3										4 x 4															
				A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D			
A	B	C		B	A	D	C		B	C	D	A		B	D	A	C		B	A	D	C			
B	C	A		C	D	B	A		C	D	A	B		C	A	D	B		C	D	A	B			
C	A	B		D	C	A	B		D	A	B	C		D	C	B	A		D	C	B	A			
5 x 5										6 x 6									7 x 7						
A	B	C	D	E				A	B	C	D	E	F					A	B	C	D	E	F	G	
B	A	E	C	D				B	F	D	C	A	E					B	C	D	E	F	G	A	
C	D	A	E	B				C	D	E	F	B	A					C	D	E	F	G	A	B	
D	E	B	A	C				D	A	F	E	C	B					D	E	F	G	A	B	C	
E	C	D	B	A				E	C	A	B	F	D					E	F	G	A	B	C	D	
								F	E	B	A	D	C					F	G	A	B	C	D	E	
																		G	A	B	C	D	E	F	
			8 x 8												9 x 9										
A	B	C	D	E	F	G	H							A	B	C	D	E	F	G	H	I			
B	C	D	E	F	G	H	A							B	C	D	E	F	G	H	I	A			
C	D	E	F	G	H	A	B							C	D	E	F	G	H	I	A	B			
D	E	F	G	H	A	B	C							D	E	F	G	H	I	A	B	C			
E	F	G	H	A	B	C	D							E	F	G	H	I	A	B	C	D			
F	G	H	A	B	C	D	E							F	G	H	I	A	B	C	D	E			
G	H	A	B	C	D	E	F							G	H	I	A	B	C	D	E	F			
H	A	B	C	D	E	F	G							H	I	A	B	C	D	E	F	G			
														I	A	B	C	D	E	F	G	H			
						10 x 10									11 x 11										
A	B	C	D	E	F	G	H	I	J					A	B	C	D	E	F	G	H	I	J	K	
B	C	D	E	F	G	H	I	J	A					B	C	D	E	F	G	H	I	J	K	A	
C	D	E	F	G	H	I	J	A	B					C	D	E	F	G	H	I	J	K	A	B	
D	E	F	G	H	I	J	A	B	C					D	E	F	G	H	I	J	K	A	B	C	
E	F	G	H	I	J	A	B	C	D					E	F	G	H	I	J	K	A	B	C	D	
F	G	H	I	J	A	B	C	D	E					F	G	H	I	J	K	A	B	C	D	E	
G	H	I	J	A	B	C	D	E	F					G	H	I	J	K	A	B	C	D	E	F	
H	I	J	A	B	C	D	E	F	G					H	I	J	K	A	B	C	D	E	F	G	
I	J	A	B	C	D	E	F	G	H					I	J	K	A	B	C	D	E	F	G	H	
J	A	B	C	D	E	F	G	H	I					J	K	A	B	C	D	E	F	G	H	I	
														K	A	B	C	D	E	F	G	H	I	J	

BẢNG 1: BẢNG SỐ NGẪU NHIÊN

1315	0589	5267	2568	0831	9298	6606	4322
0847	7475	0010	3211	0261	5729	6190	3653
1558	8179	1741	2727	8380	1630	0577	8616
1309	3207	9422	2798	2805	5366	0512	2540
1838	2452	3947	9562	2012	2890	1282	6308
3809	9806	1750	4635	2946	6082	5511	9170
0709	8892	0381	2248	4926	3282	5329	2156
1493	8313	3579	7717	5046	3306	1670	0515
7957	0824	9036	9547	4680	4788	1396	4107
9978	3121	3513	5937	8495	6928	8392	9602
7080	9054	8006	8918	9160	9327	0400	9293
3001	7407	8863	1036	7507	1568	2368	8086
8743	4626	4930	9968	2943	8886	2575	5051
1840	8035	1916	8496	2252	0552	0906	2851
1322	1539	0350	6806	3005	4306	1688	0817
5245	3194	3314	7899	9745	8818	6413	7634
4353	5145	3642	8013	8326	0011	5353	3169
4308	4603	9722	0503	2366	6419	8689	3393
3653	0387	9349	9351	3680	4833	0226	8015
3118	6412	3244	9862	4167	0208	5643	9612
4409	0046	6797	2411	0767	9375	5025	2702
7438	2531	0633	1719	4312	5930	0905	1064
3321	5121	7816	8705	2912	9578	4504	2094
1193	0280	3442	8068	9102	5867	4560	7118
0037	4400	8517	5827	3146	8637	8202	8321

TIẾP BẢNG 1

1801	0157	1705	3710	4600	1279	4910	3420
1009	0350	7841	0790	9934	8130	2749	1412
5112	9327	2302	7691	1281	9274	0882	1155
6815	6388	7188	3716	2415	6188	4077	6542
1134	1855	6484	9573	7740	6879	9307	1716
1394	2583	7871	8591	1076	9901	5837	8712
5995	3206	4455	7684	0271	4899	0114	6202
0328	4967	7540	5130	4989	7841	0937	4645
3681	9898	3300	6927	3901	8086	0990	5244
6106	3547	7487	9811	0890	2180	3222	5268
6929	8891	2880	4726	1294	0183	0813	1858
9775	6808	6168	3089	1864	2000	3674	4524
6625	5372	5006	5698	3925	9745	6307	9939
7987	6323	4323	8079	8090	2533	5986	5021
7444	0468	7148	6181	3963	3263	6543	7284
5238	2402	5002	6960	9864	4319	3066	9618
3371	6637	2783	1269	7487	3748	6285	7657
4830	9538	7548	9131	1906	5765	5248	5449
2734	2856	5435	3671	5129	4925	9732	3357
7376	5583	0920	1289	4345	2255	6127	2350
4110	1570	9188	9214	8398	7591	1597	5787
3468	7511	6498	1046	1165	9876	2749	4563
5402	9339	3509	5653	0241	5676	5097	7172
3349	8231	9465	1319	8722	3300	8777	3127
9850	9076	4408	5713	2005	1876	7396	1709

BẢNG 2: HÀM PHÂN PHỐI CHUẨN $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{x^2}{2}} dx$ (t = 0 đến +3,9)

t	0	1	2	3	4	5	6	7	8	9
0,0	0,5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
1	5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
2	5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
3	6179	6217	6265	6293	6331	6368	6406	6443	6480	6517
4	6554	6591	6628	6664	6700	6766	6772	6808	6844	6879
0,5	0,6915	6950	6985	7019	7051	7088	7123	7157	7190	7224
6	7257	7290	7324	7357	7389	7422	7454	7486	7517	7549
7	7580	7611	7642	7673	7704	7734	7764	7794	7823	7852
8	7881	7910	7939	7967	7995	8023	8051	8078	8106	8133
9	8159	8186	8212	8238	8264	8389	8315	8340	8365	8389
1,0	0,8413	8438	8461	8488	8508	8531	8554	8577	8599	8621
1	8663	8665	8686	8708	8729	8749	8770	8790	8810	8830
2	8849	8869	8888	8907	8925	8944	8962	8980	8997	9015
3	9032	9049	9066	9082	9099	9115	9131	9147	9162	9177
4	9192	9207	9222	9236	9251	9265	9279	9292	9306	9319
1,5	0,9332	9345	9357	9370	9382	9394	9406	9418	9429	9441
6	9452	9463	9474	9484	9495	9505	9515	9525	9535	9545
7	9554	9564	9573	9582	9591	9599	9608	9616	9625	9633
8	9641	9649	9656	9664	9671	9678	9686	9693	9699	9706
9	9713	9719	9726	9732	9738	9744	9750	9756	9764	9767
2,0	0,9773	9778	9783	9788	9793	9798	9803	9808	9812	9817
1	9821	9826	9830	9834	9838	9842	9846	9850	9854	9857
2	9861	9864	9868	9871	9875	9878	9881	9884	9887	9890
3	9893	9896	9898	9901	9904	9906	9909	9911	9913	9916
4	9918	9920	9922	9925	9927	9929	9931	9932	9934	9936
2,5	0,9938	9940	9941	9943	9945	9946	9948	9949	9951	9952
6	9953	9955	9956	9957	9959	9960	9961	9962	9963	9964
7	9965	9966	9967	9968	9969	9970	9971	9972	9973	9974
8	9974	9975	9976	9977	9977	9978	9979	9979	9980	9981
9	9981	9982	9982	9883	9984	9984	9985	9985	9686	9986
t	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9
$\Phi(t)$	0,9987	9990	9993	9995	9996	9997	9998	9999	9999	9999

Z là biến phân phối chuẩn tắc N (0,1)

$$\Phi(t) = p(Z < t)$$

BẢNG 3: CÁC GIÁ TRỊ TỚI HẠN CỦA HÀM
 χ^2 (Khi bình phương)

α df	0,995	0,990	0,975	0,950	0,050	0,025	0,010	0,005
1	0,000	0,000	0,001	0,004	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	5,991	7,387	9,210	10,597
3	0,072	0,115	0,216	0,352	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	11,070	12,832	15,086	16,750
6	0,676	0,872	1,237	1,635	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	14,067	16,013	18,457	20,278
8	1,344	1,647	2,180	2,733	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,892	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	23,685	26,119	29,141	31,319
15	4,601	5,229	6,262	7,261	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,390	28,869	31,526	34,805	37,156
19	6,844	7,633	8,907	10,117	30,144	32,852	36,191	38,582
20	7,434	8,260	9,591	10,851	31,410	34,170	37,566	39,997
21	8,034	8,897	10,283	11,591	32,671	35,479	38,932	41,401
22	8,643	9,542	10,982	12,338	33,924	36,781	40,289	42,796
23	9,260	10,196	11,689	13,091	33,172	38,076	41,638	44,181
24	9,886	10,856	12,401	13,848	36,415	39,364	42,980	45,558
25	10,520	11,524	13,120	14,611	37,652	40,646	44,314	46,928
26	11,160	12,198	13,844	15,379	38,885	41,923	45,642	48,290
27	11,808	12,878	14,573	16,151	40,113	43,195	46,963	49,645
28	12,461	13,565	15,308	16,928	41,337	44,461	48,278	50,994
29	13,121	14,256	16,047	17,708	42,557	45,722	49,588	52,335
30	13,787	14,953	16,791	18,493	43,773	46,979	50,892	53,672

Bảng các giá trị tới hạn của $\chi^2(\alpha, df)$
 α - mức ý nghĩa; df - bậc tự do

df \ α	0,955	0,990	0,975	0,950	0,050	0,025	0,010	0,005
31	14,458	15,655	17,539	19,281	44,985	48,232	52,191	55,002
32	15,134	16,362	18,291	20,072	46,194	49,480	53,486	56,328
33	15,815	17,073	19,047	20,867	47,400	50,725	54,775	57,648
34	16,501	17,789	19,806	21,664	48,602	51,966	56,061	58,964
35	17,192	18,509	20,569	22,465	49,802	53,203	57,342	60,275
36	17,887	19,233	21,336	23,269	50,998	54,437	58,619	61,581
37	18,586	19,960	22,106	24,075	52,192	55,668	59,893	62,883
38	19,289	20,691	22,878	24,884	53,384	56,895	61,162	64,181
39	19,996	21,426	23,654	25,695	54,572	58,120	62,428	65,475
40	20,707	22,164	24,433	26,509	55,758	59,342	63,691	66,766
41	21,421	22,906	25,215	27,326	56,942	60,561	64,950	68,053
42	22,138	23,650	25,999	28,144	58,124	61,777	66,206	69,336
43	22,860	24,398	26,785	28,965	59,304	62,990	67,459	70,616
44	23,584	25,148	27,575	29,787	60,481	64,201	68,710	71,892
45	24,311	25,901	28,366	30,612	61,656	65,410	69,957	73,166
46	25,041	26,657	29,160	31,439	62,830	66,616	71,201	74,437
47	25,775	27,416	29,956	32,268	64,001	67,821	72,443	75,704
48	26,511	28,177	30,754	33,098	65,171	69,023	73,683	76,969
49	27,249	28,941	31,555	33,930	66,339	70,222	74,919	78,231
50	27,991	29,707	32,357	34,764	67,505	71,420	76,154	79,490
51	28,735	30,475	33,162	35,600	68,669	72,616	77,386	80,746
52	29,481	31,246	33,968	36,437	69,832	73,810	78,616	82,001
53	30,230	32,019	34,776	37,276	70,993	75,002	79,843	83,253
54	30,981	32,793	35,586	38,116	72,153	76,192	81,069	84,502
55	31,735	33,571	36,398	38,958	73,311	77,380	82,292	85,749
56	32,491	34,350	37,212	39,801	74,468	78,567	83,514	86,994
57	33,248	35,131	38,027	40,646	75,624	79,752	84,733	88,237
58	34,008	35,914	38,844	41,492	76,778	80,936	85,950	89,477
59	34,770	36,698	39,662	42,339	77,930	82,117	87,166	90,715
60	35,534	37,485	40,482	43,188	79,082	83,298	88,379	91,952
61	36,300	38,273	41,303	44,038	80,232	84,476	89,591	93,186
62	37,068	39,063	42,126	44,889	81,381	85,654	90,802	94,419
63	37,838	39,855	42,950	45,741	82,529	86,830	92,010	95,649
64	38,610	40,649	43,776	46,595	83,675	88,004	93,217	96,878
65	39,383	41,444	44,603	47,450	84,821	89,177	94,422	98,105

α df	0,955	0,990	0,975	0,950	0,050	0,025	0,010	0,005
66	40,158	42,240	45,431	48,305	85,965	90,349	95,626	99,330
67	40,935	43,038	46,261	49,162	87,108	92,638	96,828	100,554
68	41,714	43,838	47,092	50,020	88,250	92,688	98,028	101,776
69	42,493	44,639	47,924	50,879	89,391	93,856	99,227	102,996
70	43,275	45,442	48,758	51,739	90,531	95,023	100,425	104,215
71	44,058	46,246	49,592	52,600	91,670	96,189	101,621	105,432
72	44,843	47,051	50,428	53,462	92,808	97,353	102,816	106,647
73	45,629	47,858	51,265	54,325	93,945	98,516	104,010	107,862
74	46,417	48,666	52,103	55,189	95,081	99,678	105,202	109,074
75	47,206	49,475	52,942	56,054	96,217	100,839	106,393	110,285
76	47,996	50,286	53,782	56,920	97,351	101,999	107,582	111,495
77	48,788	51,097	54,623	57,786	98,484	103,158	108,771	112,704
78	49,581	51,910	55,466	58,654	99,617	104,316	109,958	113,911
79	50,376	52,725	56,309	59,522	100,749	105,473	111,144	115,116
80	51,172	53,540	57,153	60,391	101,897	106,629	112,329	116,321
81	51,969	54,357	57,998	61,262	103,010	107,783	113,512	117,524
82	52,767	55,174	58,845	62,132	104,139	108,937	114,695	118,726
83	53,567	55,993	59,692	63,004	105,267	110,090	115,876	119,927
84	57,368	56,813	60,540	63,876	106,395	111,242	117,057	121,126
85	55,170	57,634	61,389	64,749	107,522	112,393	118,236	122,324
86	55,973	58,456	62,239	65,623	108,648	113,544	119,414	123,522
87	56,777	59,279	63,089	66,498	109,773	114,693	120,591	124,718
88	57,582	60,103	63,941	67,373	110,898	115,841	121,767	125,912
89	58,389	60,928	64,793	68,249	112,022	116,989	122,942	127,106
90	59,196	61,754	65,647	69,126	113,145	118,136	124,116	128,299
91	60,005	62,581	66,501	70,003	114,268	119,282	125,289	129,490
92	60,815	63,409	67,356	70,882	115,390	120,427	126,462	130,681
93	61,625	64,238	68,211	71,760	116,511	121,571	127,633	131,871
94	62,437	65,068	69,068	72,640	117,632	122,715	128,803	133,059
95	63,250	65,898	69,925	73,520	118,752	123,858	129,973	134,247
96	64,063	66,730	70,783	74,401	119,871	125,000	131,141	135,433
97	64,878	67,562	71,642	75,282	120,990	126,141	132,309	136,619
98	65,693	68,396	72,501	76,164	122,108	127,282	133,476	137,803
99	66,510	69,230	73,361	77,046	123,225	128,422	134,641	138,987
100	67,328	70,065	74,222	77,929	124,342	129,561	135,807	140,170

α df	0,995	0,990	0,975	0,950	0,050	0,025	0,010	0,005
102	68,965	71,737	75,946	79,697	126,574	131,838	138,134	142,532
104	70,606	73,413	77,672	81,468	128,804	134,111	140,459	144,891
106	72,251	75,092	79,401	83,240	131,031	136,382	142,780	147,247
108	73,899	76,774	81,133	85,015	133,257	138,651	145,099	149,599
110	75,550	78,458	82,867	86,792	135,480	140,916	147,414	151,948
112	77,204	80,146	84,604	88,570	137,701	143,180	149,727	154,295
114	78,862	81,836	86,342	90,351	139,921	145,441	152,037	156,637
116	80,522	83,529	88,084	92,134	142,138	147,700	154,344	158,977
118	82,185	85,225	89,827	93,918	144,354	149,957	156,648	161,314
120	83,852	86,923	91,573	95,705	146,567	152,211	158,950	163,648
122	85,521	88,624	93,320	97,493	148,779	154,464	161,249	165,980
124	87,192	90,327	95,070	99,283	150,989	156,714	163,546	168,308
126	88,866	92,033	96,822	101,074	153,198	158,962	165,841	170,634
128	90,543	93,741	98,576	102,867	155,405	161,209	168,133	172,957
130	92,223	95,451	100,331	104,662	157,610	163,453	170,423	175,278
132	93,904	97,163	102,089	106,459	159,814	165,696	172,711	177,596
134	95,588	98,878	103,848	108,257	162,016	167,936	174,996	179,913
136	97,275	100,595	105,690	110,056	164,216	170,175	177,280	182,227
138	98,964	102,314	107,372	111,857	166,415	172,412	179,561	184,538
140	100,655	104,034	109,137	113,659	168,613	174,648	181,841	186,847
142	102,348	105,757	110,903	115,463	170,809	176,881	184,117	189,153
144	104,043	107,482	112,671	117,268	173,004	179,114	186,393	191,459
146	105,741	109,209	114,441	119,075	175,198	181,344	188,666	193,761
148	107,441	110,937	116,212	120,883	177,390	183,573	190,938	196,062
150	109,142	112,668	117,985	122,692	179,581	185,800	193,207	198,360
200	152,241	156,432	162,728	168,279	233,994	241,058	249,445	255,264
250	196,160	200,939	208,098	214,392	287,882	295,689	304,939	311,346
300	240,663	245,973	253,912	260,878	341,395	349,874	359,906	366,844
400	330,903	337,155	346,482	354,641	447,632	457,306	468,724	476,607
500	422,303	426,387	439,936	449,147	553,127	563,851	576,493	585,206
600	514,529	522,365	543,019	544,180	658,094	669,769	683,515	692,981
700	607,379	615,907	628,577	639,613	762,661	775,210	789,974	800,131
800	700,725	709,897	723,513	735,362	866,911	880,275	895,984	906,786
900	794,475	804,251	818,756	831,370	970,904	985,032	1001,629	1013,036
1000	888,563	898,921	914,257	927,594	1074,679	1089,531	1106,969	1118,948

BẢNG 4: BẢNG CÁC GIÁ TRỊ t_α CỦA PHÂN PHỐI t (phân phối Student)

Độ tự do (df)	Mức ý nghĩa (α)		
	0,05	0,01	0,001
1	12,71	63,66	636,58
2	4,30	9,92	31,60
3	3,18	5,84	12,92
4	2,78	4,60	8,61
5	2,57	4,03	6,87
6	2,45	3,71	5,96
7	2,36	3,50	5,41
8	2,31	3,36	5,04
9	2,26	3,25	4,78
10	2,23	3,17	4,59
11	2,20	3,11	4,44
12	2,18	3,05	4,32
13	2,16	3,01	4,22
14	2,14	2,98	4,14
15	2,13	2,95	4,07
16	2,12	2,92	4,01
17	2,11	2,90	3,97
18	2,10	2,88	3,92
19	2,09	2,86	3,88
20	2,09	2,85	3,85
21	2,08	2,83	3,82
22	2,07	2,82	3,79
23	2,07	2,81	3,77
24	2,06	2,80	3,75
25	2,06	2,79	3,73
26	2,06	2,78	3,71
27	2,05	2,77	3,69
28	2,05	2,76	3,67
29	2,05	2,76	3,66
30	2,04	2,75	3,65
40	2,02	2,70	3,55
60	2,00	2,66	3,46
120	1,98	2,62	3,37
∞	1,96	2,58	3,29

Kiểm định hai phía ($|T| > t$) = α

BẢNG 5: CÁC GIÁ TRỊ TỚI HẠN CỦA HÀM PHÂN PHỐI F(FISHER)
($\alpha \leq 0,05$ (hàng tròn) và $\alpha \leq 0,01$ (hàng dưới))

$\frac{df_1}{df_2}$	1	2	3	4	5	6	7	8	9	10	11	12
1	161,45 4052,18	199,50 4999,34	215,71 5403,53	224,58 5624,26	230,16 5763,96	233,99 5858,95	236,77 5928,33	238,88 5980,95	240,54 6022,40	241,88 6055,93	242,98 6083,40	243,90 6106,68
2	18,51 98,50	19,00 99,00	19,16 99,16	19,25 99,25	19,30 99,30	19,33 99,33	19,35 99,36	19,37 99,38	19,38 99,39	19,40 99,40	19,40 99,41	19,41 99,42
3	10,13 34,12	9,55 30,82	9,28 29,46	9,12 28,71	9,01 28,24	8,94 27,91	8,89 27,67	8,85 27,49	8,81 27,34	8,79 27,23	8,76 27,13	8,74 27,05
4	7,71 21,20	6,94 18,00	6,59 16,69	6,39 15,98	6,26 15,52	6,16 15,21	6,09 14,98	6,04 14,80	6,00 14,66	5,96 14,55	5,94 14,45	5,91 14,37
5	6,61 16,26	5,79 13,27	5,41 12,06	5,19 11,39	5,05 10,97	4,95 10,67	4,88 10,46	4,82 10,29	4,77 10,16	4,74 10,05	4,70 9,96	4,68 9,89
6	5,99 13,75	5,14 10,92	4,76 9,78	4,53 9,15	4,39 8,75	4,28 8,47	4,21 8,26	4,15 8,10	4,10 7,98	4,06 7,87	4,03 7,79	4,00 7,72
7	5,59 12,25	4,74 9,55	4,35 8,45	4,12 7,85	3,97 7,46	3,87 7,19	3,79 6,99	3,73 6,84	3,68 6,72	3,64 6,62	3,60 6,54	3,57 6,47
8	5,32 11,26	4,46 8,65	4,07 7,59	3,84 7,01	3,69 6,63	3,58 6,37	3,50 6,18	3,44 6,03	3,39 5,91	3,35 5,81	3,31 5,73	3,28 5,67
9	5,22 10,56	4,26 8,02	3,86 6,99	3,63 6,42	3,48 6,06	3,37 5,80	3,29 5,61	3,23 5,47	3,18 5,35	3,14 5,26	3,10 5,18	3,07 5,11
10	40,96 10,04	4,10 7,56	3,71 6,55	3,48 5,99	3,33 5,64	3,22 5,39	3,14 5,20	3,07 5,06	3,02 4,94	2,98 4,85	2,94 4,77	2,91 4,71

Bậc tự do ứng với tỷ số là L_A , ứng với mẫu số là L_B
 Bảng các giá trị tới hạn $F(\alpha, L_A, L_B)$

$df_1 \backslash df_2$	1	2	3	4	5	6	7	8	9	10	11	12
11	4,84 9,65	3,98 7,21	3,59 6,22	3,36 5,67	3,20 5,32	3,09 5,07	3,01 4,89	2,95 4,74	2,90 4,63	2,85 4,54	2,82 4,46	2,79 4,40
12	4,75 9,33	3,89 6,93	3,49 5,95	3,26 5,41	3,11 5,06	3,00 4,82	2,91 4,64	2,85 4,50	2,80 4,39	2,75 4,30	2,72 4,22	2,69 4,16
13	4,67 9,07	3,81 6,70	3,41 5,74	3,18 5,21	3,03 4,86	2,92 4,62	2,83 4,44	2,77 4,30	2,71 4,19	2,67 4,10	2,63 4,02	2,60 3,96
14	4,60 8,86	3,74 6,51	3,34 5,56	3,11 5,04	2,96 4,69	2,85 4,46	2,76 4,28	2,70 4,14	2,65 4,03	2,60 3,94	2,57 3,86	2,53 3,80
15	4,54 8,68	3,68 6,36	3,29 5,42	3,06 4,89	2,90 4,56	2,79 4,32	2,71 4,14	2,64 4,00	2,59 3,89	2,54 3,80	2,51 3,73	2,48 3,67
16	4,49 8,53	3,63 6,23	3,24 5,29	3,01 4,77	2,85 4,44	2,74 4,20	2,66 4,03	2,59 3,89	2,54 3,78	2,49 3,69	2,46 3,62	2,42 3,55
17	4,45 8,40	3,59 6,11	3,20 5,19	2,96 4,67	2,81 4,34	2,70 4,10	2,61 3,93	2,55 3,79	2,49 3,68	2,45 3,59	2,41 3,52	2,38 3,46
18	4,41 8,29	3,55 6,01	3,16 5,09	2,93 4,58	2,77 4,25	2,66 4,01	2,58 3,84	2,51 3,71	2,46 3,60	2,41 3,51	2,37 3,43	2,34 3,37
19	4,38 8,18	3,52 5,93	3,13 5,01	2,90 4,50	2,74 4,17	2,63 3,94	2,54 3,77	2,48 3,63	2,42 3,52	2,38 3,43	2,34 3,36	2,31 3,30
20	4,35 8,10	3,49 5,85	3,10 4,94	2,87 4,43	2,71 4,10	2,60 3,87	2,51 3,70	2,45 3,56	2,39 3,46	2,35 3,37	2,31 3,29	2,28 3,23

$\frac{df_1}{df_2}$	1	2	3	4	5	6	7	8	9	10	11	12
21	4,32 8,02	3,47 5,78	3,07 4,87	2,84 4,37	2,68 4,04	2,57 3,81	2,49 3,64	2,42 3,51	2,37 3,40	2,32 3,31	2,28 3,24	2,25 3,17
22	4,30 7,95	3,42 5,66	3,05 4,82	2,82 4,31	2,66 3,99	2,55 3,76	2,46 3,59	2,40 3,45	2,34 3,35	2,30 3,26	2,26 3,18	2,23 3,12
23	4,28 7,88	3,42 5,66	3,03 4,76	2,80 4,26	2,64 3,94	2,53 3,71	2,44 3,54	2,37 3,41	2,32 3,30	2,27 3,21	2,24 3,14	2,20 3,07
24	4,26 7,82	3,40 5,61	3,01 4,72	2,78 4,22	2,62 3,90	2,51 3,67	2,42 3,50	2,36 3,36	2,30 3,26	2,25 3,17	2,22 3,09	2,18 3,03
25	4,24 7,77	3,37 5,57	2,99 4,68	2,76 4,18	2,60 3,85	2,49 3,63	2,40 3,46	2,34 3,32	2,28 3,22	2,24 3,13	2,20 3,06	2,16 2,99
26	4,23 7,72	3,37 5,53	2,98 4,64	2,74 4,14	2,59 3,82	2,47 3,59	2,39 3,42	2,32 3,29	2,27 3,18	2,22 3,09	2,18 3,02	2,15 2,96
27	4,21 7,68	,35 5,49	2,96 4,60	2,73 4,11	2,57 3,78	2,46 3,56	2,37 3,39	2,31 3,26	2,25 3,15	2,20 3,06	2,17 2,99	2,13 2,93
28	4,20 7,64	3,34 5,45	2,95 4,57	2,71 4,07	2,56 3,75	2,45 3,53	2,36 3,36	2,29 3,23	2,24 3,12	2,19 3,03	2,15 2,96	2,12 2,90
29	4,18 7,60	3,33 5,42	2,93 4,54	2,70 4,04	2,55 3,73	2,43 3,50	2,35 3,33	2,28 3,20	2,22 3,09	2,18 3,00	2,14 2,93	2,10 2,87
30	4,17 7,50	3,32 5,39	2,92 4,51	2,69 4,02	2,53 3,70	2,42 3,47	2,33 3,30	2,27 3,17	2,21 3,07	2,16 2,98	2,13 2,91	2,09 2,84

$\frac{df_1}{df_2}$	1	2	3	4	5	6	7	8	9	10	11	12
32	4,15 7,50	3,29 5,34	2,90 4,46	2,67 3,97	2,51 3,65	2,40 3,43	2,31 3,26	2,24 3,13	2,19 3,02	2,14 2,93	2,10 2,86	2,07 2,80
34	4,13 7,44	3,28 5,29	2,88 4,42	2,65 3,93	2,49 3,61	2,38 3,39	2,29 3,22	2,23 3,09	2,17 2,98	2,12 2,89	2,08 2,82	2,05 2,76
36	4,11 7,40	3,26 5,25	2,87 4,38	2,63 3,89	2,48 3,57	2,36 3,35	2,28 3,18	2,21 3,05	2,15 2,95	2,11 2,86	2,07 2,79	2,03 2,72
38	4,10 7,35	3,24 5,21	2,85 4,34	2,62 3,86	2,46 3,54	2,35 3,32	2,26 3,15	2,19 3,02	2,14 2,92	2,09 2,83	2,05 2,75	2,02 2,69
40	4,08 7,31	3,23 5,18	2,84 4,31	2,61 3,83	2,45 3,51	2,34 3,29	2,25 3,12	2,18 2,99	2,12 2,89	2,08 2,80	2,04 2,73	2,00 2,66
42	4,07 7,28	3,22 5,15	2,83 4,29	2,59 3,80	2,44 3,49	2,32 3,27	2,24 3,10	2,17 2,97	2,11 2,86	2,06 2,78	2,03 2,70	1,99 2,64
44	4,06 7,25	3,21 5,12	2,82 4,26	2,58 3,78	2,43 3,47	2,31 3,24	2,23 3,08	2,16 2,95	2,10 2,84	2,05 2,75	2,01 2,68	1,98 2,62
46	4,05 7,22	3,20 5,10	2,81 4,24	2,57 3,76	2,42 3,44	2,30 3,22	2,22 3,06	2,15 2,93	2,09 2,82	2,04 2,73	2,00 2,66	1,97 2,60
48	4,04 7,19	3,19 5,08	2,80 4,22	2,57 3,74	2,41 3,43	2,29 3,20	2,21 3,04	2,14 2,91	2,08 2,80	2,03 2,71	1,99 2,64	1,96 2,58
50	4,03 7,17	3,18 5,06	2,79 4,20	2,56 3,72	2,40 3,41	2,29 3,19	2,20 3,02	2,13 2,89	2,07 2,78	2,03 2,70	1,99 2,63	1,95 2,56

$\frac{df_1}{df_2}$	1	2	3	4	5	6	7	8	9	10	11	12
55	4,02 7,12	3,16 5,01	2,77 4,16	2,54 3,68	2,38 3,37	2,27 3,15	2,18 2,98	2,11 2,85	2,06 2,75	2,01 2,66	1,97 2,59	1,93 2,53
60	4,00 7,08	3,15 4,98	2,76 4,13	2,53 3,65	2,37 3,34	2,25 3,12	2,17 2,95	2,10 2,82	2,04 2,72	1,99 2,63	1,95 2,56	1,92 2,50
65	3,99 7,04	3,14 4,95	2,75 4,10	2,51 3,62	2,36 3,31	2,24 3,09	2,15 2,93	2,08 2,80	2,03 2,69	1,98 2,61	1,94 2,53	1,90 2,47
70	3,98 7,01	3,13 4,92	2,74 4,07	2,50 3,60	2,35 3,29	2,23 3,07	2,14 2,91	2,07 2,78	2,02 2,67	1,97 2,59	1,93 2,51	1,89 2,45
75	3,97 6,99	3,12 4,90	2,73 4,05	2,49 3,58	2,34 3,27	2,22 3,05	2,13 2,89	2,06 2,76	2,01 2,65	1,96 2,57	1,92 2,49	1,88 2,43
80	3,96 6,96	3,11 4,88	2,72 4,04	2,49 3,56	2,33 3,26	2,21 3,04	2,13 2,87	2,06 2,74	2,00 2,64	1,95 2,55	1,91 2,48	1,88 2,42
85	3,95 6,94	3,10 4,86	2,71 4,02	2,48 3,55	2,32 3,24	2,21 3,02	2,12 2,86	2,05 2,73	1,99 2,62	1,94 2,54	1,90 2,46	1,87 2,40
90	3,95 6,93	3,10 4,85	2,71 4,01	2,47 3,53	2,32 3,23	2,20 3,01	2,11 2,84	2,04 2,72	1,99 2,61	1,94 2,52	1,90 2,45	1,86 2,39
95	3,94 6,91	3,09 4,84	2,70 3,99	2,47 3,52	2,31 3,22	2,20 3,00	2,11 2,83	2,04 2,70	1,98 2,60	1,93 2,51	1,89 2,44	1,86 2,38
100	3,94 6,90	3,09 4,82	2,70 3,98	2,46 3,51	2,31 3,21	2,19 2,99	2,10 2,82	2,03 2,69	1,97 2,59	1,93 2,50	1,89 2,43	1,85 2,37

$df_1 \backslash df_2$	1	2	3	4	5	6	7	8	9	10	11	12
	3,92	3,07	2,68	2,44	2,29	2,17	2,08	2,01	1,96	1,91	1,87	1,83
125	6,84	4,78	3,94	3,47	3,17	2,95	2,79	2,66	2,55	2,47	2,39	2,33
	3,90	3,06	2,66	2,43	2,27	2,16	2,07	2,00	1,94	1,89	1,85	1,82
150	6,81	4,75	3,91	3,45	3,14	2,92	2,76	2,63	2,53	2,44	2,37	2,31
	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,84	1,80
200	6,76	4,71	3,88	3,41	3,11	2,89	2,73	2,60	2,50	2,41	2,34	2,27
	3,86	3,02	2,63	2,39	2,24	2,12	2,03	1,96	1,90	1,85	1,81	1,78
400	6,70	4,66	3,83	3,37	3,06	2,85	2,68	2,56	2,45	2,37	2,29	2,23
	3,85	3,00	2,61	2,38	2,22	2,11	2,02	1,95	1,89	1,84	1,80	1,76
1000	6,66	4,63	3,80	3,34	3,04	2,82	2,66	2,53	2,43	2,34	2,27	2,20
	3,84	3,00	2,61	2,37	2,22	2,10	2,01	1,94	1,88	1,83	1,79	1,75
∞	6,64	4,61	3,79	3,32	3,02	2,81	2,64	2,51	2,41	2,32	2,25	2,19

$\frac{df_1}{df_2}$	14	16	20	24	30	40	50	75	100	200	500	∞
1	245,36 6143,00	246,47 6170,01	248,02 6208,66	249,05 6234,27	250,10 6260,35	251,14 6286,43	251,77 6302,26	252,62 6323,68	253,04 6333,92	253,68 6349,76	254,06 6359,54	254,29 6365,59
2	19,42 99,43	19,43 99,44	19,45 99,45	19,45 99,46	19,46 99,47	19,47 99,48	19,48 99,48	19,48 99,48	19,49 99,49	19,49 99,49	19,49 99,50	19,50 99,50
3	8,71 26,92	8,69 26,83	8,66 26,69	8,64 26,60	8,62 26,50	8,59 26,41	8,58 26,35	8,56 26,28	8,55 26,24	8,54 26,18	8,53 26,15	8,53 26,13
4	5,87 14,25	5,84 14,15	5,80 14,02	5,77 13,93	5,75 13,84	5,72 13,75	5,70 13,69	5,68 13,61	5,66 13,58	5,65 13,52	5,64 13,49	5,63 13,47
5	4,64 9,77	4,60 9,68	4,56 9,55	4,53 9,47	4,50 9,38	4,46 9,29	4,44 9,24	4,42 9,17	4,41 9,13	4,39 9,08	4,37 9,04	4,37 9,02
6	3,96 7,60	3,92 7,52	3,87 7,40	3,84 7,31	3,81 7,23	3,77 7,14	3,75 7,09	3,73 7,02	3,71 6,99	3,69 6,93	3,68 6,90	3,67 6,88
7	3,53 6,36	3,49 6,28	3,44 6,16	3,41 6,07	3,38 5,99	3,34 5,91	3,32 5,86	3,29 5,79	3,27 5,75	3,25 5,70	3,24 5,67	3,23 5,65
8	3,24 5,56	3,20 5,48	3,15 5,36	3,12 5,28	3,08 5,20	3,04 5,12	3,02 5,07	2,99 5,00	2,97 4,96	2,95 4,91	2,94 4,88	2,93 4,86
9	3,03 5,01	2,99 4,92	2,94 4,81	2,90 4,73	2,86 4,65	2,83 4,57	2,80 4,52	2,77 4,45	2,76 4,41	2,73 4,36	2,72 4,33	2,71 4,31
10	2,86 4,60	2,83 4,52	2,77 4,41	2,74 4,33	2,70 4,25	2,66 4,17	2,64 4,12	2,60 4,05	2,59 4,01	2,56 3,96	2,55 3,93	2,54 3,91

$df_1 \backslash df_2$	14	16	20	24	30	40	50	75	100	200	500	∞
11	2,74 4,29	2,70 4,21	2,65 4,10	2,61 4,02	2,57 3,94	2,53 3,86	2,51 3,81	2,47 3,74	2,46 3,71	2,43 3,66	2,42 3,62	2,41 3,60
12	2,64 4,05	2,60 3,97	2,54 3,86	2,51 3,78	2,47 3,70	2,43 3,62	2,40 3,57	2,37 3,50	2,35 3,47	2,32 3,41	2,31 3,38	2,30 3,36
13	2,55 3,86	2,51 3,78	2,46 3,66	2,42 3,59	2,38 3,51	2,34 3,43	2,31 3,38	2,28 3,31	2,26 3,27	2,23 3,22	2,22 3,19	2,21 3,17
14	2,48 3,70	2,44 3,62	2,39 3,51	2,35 3,43	2,31 3,35	2,27 3,27	2,24 3,22	2,21 3,15	2,19 3,11	2,16 3,06	2,14 3,03	2,13 3,01
15	2,42 3,56	2,38 3,49	2,33 3,37	2,29 3,29	2,25 3,21	2,20 3,13	2,18 3,08	2,14 3,01	2,12 2,98	2,10 2,92	2,08 2,89	2,07 2,87
16	2,37 3,45	2,33 3,37	2,28 3,26	2,24 3,18	2,19 3,10	2,15 3,02	2,12 2,97	2,09 2,90	2,07 2,86	2,04 2,81	2,02 2,78	2,01 2,76
17	2,33 3,35	2,29 3,27	2,23 3,16	2,19 3,08	2,15 3,00	2,10 2,92	2,08 2,87	2,04 2,80	2,02 2,76	1,99 2,71	1,97 2,68	1,96 2,66
18	2,29 3,27	2,25 3,19	2,19 3,08	2,15 3,00	2,11 2,92	2,06 2,84	2,04 2,78	2,00 2,71	1,98 2,68	1,95 2,62	1,93 2,59	1,92 2,57
19	2,26 3,19	2,21 3,12	2,16 3,00	2,11 2,92	2,07 2,84	2,03 2,76	2,00 2,71	1,96 2,64	1,94 2,60	1,91 2,55	1,89 2,51	1,88 2,49
20	2,22 3,13	2,18 3,05	2,12 2,94	2,08 2,86	2,04 2,78	1,99 2,69	1,97 2,64	1,93 2,57	1,91 2,54	1,88 2,48	1,86 2,44	1,84 2,42

$\frac{df_1}{df_2}$	14	16	20	24	30	40	50	75	100	200	500	∞
21	2,20 3,07	2,16 2,99	2,10 2,88	2,05 2,80	2,01 2,72	1,96 2,64	1,94 2,58	1,90 2,51	1,88 2,48	1,84 2,42	1,83 2,38	1,81 2,36
22	2,17 3,02	2,13 2,94	2,07 2,83	2,03 2,75	1,98 2,67	1,94 2,58	1,91 2,53	1,87 2,46	1,85 2,42	1,82 2,36	1,80 2,33	1,78 2,31
23	2,15 2,97	2,11 2,89	2,05 2,78	2,01 2,70	1,96 2,62	1,91 2,54	1,88 2,48	1,84 2,41	1,82 2,37	1,79 2,32	1,77 2,28	1,76 2,26
24	2,13 2,93	2,09 2,85	2,03 2,74	1,98 2,66	1,94 2,58	1,89 2,49	1,86 2,44	1,82 2,37	1,80 2,33	1,77 2,27	1,75 2,24	1,73 2,21
25	2,11 2,89	2,07 2,81	2,01 2,70	1,96 2,62	1,92 2,54	1,87 2,45	1,84 2,40	1,80 2,33	1,78 2,29	1,75 2,23	1,73 2,19	1,71 2,17
26	2,09 2,86	2,05 2,78	1,99 2,66	1,95 2,58	1,90 2,50	1,85 2,42	1,82 2,36	1,78 2,29	1,76 2,25	1,73 2,19	1,71 2,16	1,69 2,13
27	2,08 2,82	2,04 2,75	1,97 2,63	1,93 2,55	1,88 2,47	1,84 2,38	1,81 2,33	1,76 2,26	1,74 2,22	1,71 2,16	1,69 2,12	1,67 2,10
28	2,06 2,79	2,02 2,72	1,96 2,60	1,91 2,52	1,87 2,44	1,82 2,35	1,79 2,30	1,75 2,23	1,73 2,19	1,69 2,13	1,67 2,09	1,66 2,07
29	2,05 2,77	2,01 2,69	1,94 2,57	1,90 2,49	1,85 2,41	1,81 2,33	1,77 2,27	1,73 2,20	1,71 2,16	1,67 2,10	1,65 2,06	1,64 2,04
30	2,04 2,74	1,99 2,66	1,93 2,55	1,89 2,47	1,84 2,39	1,79 2,30	1,76 2,25	1,72 2,17	1,70 2,13	1,66 2,07	1,64 2,03	1,62 2,01

$\frac{df_1}{df_2}$	14	16	20	24	30	40	50	75	100	200	500	∞
32	2,01 2,70	1,97 2,62	1,91 2,50	1,86 2,42	1,82 2,34	1,77 2,25	1,74 2,20	1,69 2,12	1,67 2,08	1,63 2,02	1,61 1,98	1,60 1,96
34	1,99 2,66	1,95 2,58	1,89 2,46	1,84 2,38	1,80 2,30	1,75 2,21	1,71 2,16	1,67 2,08	1,65 2,04	1,61 1,98	1,59 1,94	1,57 1,91
36	1,98 2,62	1,93 2,54	1,87 2,43	1,82 2,35	1,78 2,26	1,73 2,18	1,69 2,12	1,65 2,04	1,62 2,00	1,59 1,94	1,56 1,90	1,55 1,87
38	1,96 2,59	1,92 2,51	1,85 2,40	1,81 2,32	1,76 2,23	1,71 2,14	1,68 2,09	1,63 2,01	1,61 1,97	1,57 1,90	1,54 1,86	1,53 1,84
40	1,95 2,56	1,90 2,48	1,84 2,37	1,79 2,29	1,74 2,20	1,69 2,11	1,66 2,06	1,61 1,98	1,59 1,94	1,55 1,87	1,53 1,83	1,51 1,81
42	1,94 2,54	1,89 2,46	1,83 2,34	1,78 2,26	1,73 2,18	1,68 2,09	1,65 2,03	1,60 1,95	1,57 1,91	1,53 1,85	1,51 1,80	1,49 1,78
44	1,92 2,52	1,88 2,44	1,81 2,32	1,77 2,24	1,72 2,15	1,67 2,07	1,63 2,01	1,59 1,93	1,56 1,89	1,52 1,82	1,49 1,78	1,48 1,75
46	1,91 2,50	1,87 2,42	1,80 2,30	1,76 2,22	1,71 2,13	1,65 2,04	1,62 1,99	1,57 1,91	1,55 1,86	1,51 1,80	1,48 1,76	1,46 1,73
48	1,90 2,48	1,86 2,40	1,79 2,28	1,75 2,20	1,70 2,12	1,64 2,02	1,61 1,97	1,56 1,89	1,54 1,84	1,49 1,78	1,47 1,73	1,45 1,71
50	1,89 2,46	1,85 2,38	1,78 2,27	1,74 2,18	1,69 2,10	1,63 2,01	1,60 1,95	1,55 1,87	1,52 1,82	1,48 1,76	1,46 1,71	1,44 1,69

$\frac{df_1}{df_2}$	14	16	20	24	30	40	50	75	100	200	500	∞
55	1,88 2,42	1,83 2,34	1,76 2,23	1,72 2,15	1,67 2,06	1,61 1,97	1,58 1,91	1,53 1,83	1,50 1,78	1,46 1,71	1,43 1,67	1,41 1,64
60	1,86 2,39	1,82 2,31	1,75 2,20	1,70 2,12	1,65 2,03	1,59 1,94	1,56 1,88	1,51 1,79	1,48 1,75	1,44 1,68	1,41 1,63	1,39 1,60
65	1,85 2,37	1,80 2,29	1,73 2,17	1,69 2,09	1,63 2,00	1,58 1,91	1,54 1,85	1,49 1,77	1,46 1,72	1,42 1,65	1,39 1,60	1,37 1,57
70	1,84 2,35	1,79 2,27	1,72 2,15	1,67 2,07	1,62 1,98	1,57 1,89	1,53 1,83	1,48 1,74	1,45 1,70	1,40 1,62	1,37 1,57	1,36 1,54
75	1,83 2,33	1,78 2,25	1,71 2,13	1,66 2,05	1,61 1,96	1,55 1,87	1,52 1,81	1,47 1,72	1,44 1,67	1,39 1,60	1,36 1,55	1,34 1,52
80	1,82 2,31	1,77 2,23	1,70 2,12	1,65 2,03	1,60 1,94	1,54 1,85	1,51 1,79	1,45 1,70	1,43 1,65	1,38 1,58	1,35 1,53	1,33 1,50
85	1,81 2,30	1,76 2,22	1,70 2,10	1,65 2,02	1,59 1,93	1,54 1,83	1,50 1,77	1,45 1,69	1,42 1,64	1,37 1,56	1,34 1,51	1,32 1,48
90	1,80 2,29	1,76 2,21	1,69 2,09	1,64 2,00	1,59 1,92	1,53 1,82	1,49 1,76	1,44 1,67	1,41 1,62	1,36 1,55	1,33 1,49	1,30 1,46
95	1,80 2,28	1,75 2,20	1,68 2,08	1,63 1,99	1,58 1,90	1,52 1,81	1,48 1,75	1,43 1,66	1,40 1,61	1,35 1,53	1,32 1,48	1,29 1,45
100	1,79 2,27	1,75 2,19	1,68 2,07	1,63 1,98	1,57 1,89	1,52 1,80	1,48 1,74	1,42 1,65	1,39 1,60	1,34 1,52	1,31 1,47	1,29 1,43

$df_1 \backslash df_2$	14	16	20	24	30	40	50	75	100	200	500	∞
125	1,77 2,23	1,73 2,15	1,66 2,03	1,60 1,94	1,55 1,85	1,49 1,76	1,45 1,69	1,40 1,60	1,36 1,55	1,31 1,47	1,27 1,41	1,25 1,38
150	1,76 2,20	1,71 2,12	1,64 2,00	1,59 1,92	1,54 1,83	1,48 1,73	1,44 1,66	1,38 1,57	1,34 1,52	1,29 1,43	1,25 1,38	1,23 1,34
200	1,74 2,17	1,69 2,09	1,62 1,97	1,57 1,89	1,52 1,79	1,46 1,69	1,41 1,63	1,35 1,53	1,32 1,48	1,26 1,39	1,22 1,33	1,19 1,28
400	1,72 2,13	1,67 2,05	1,60 1,92	1,54 1,84	1,49 1,75	1,42 1,64	1,38 1,58	1,32 1,48	1,28 1,42	1,22 1,32	1,17 1,25	1,13 1,19
1000	1,70 2,10	1,65 2,02	1,58 1,90	1,53 1,81	1,47 1,72	1,41 1,61	1,36 1,54	1,30 1,44	1,26 1,38	1,19 1,28	1,13 1,19	1,09 1,12
∞	1,69 2,09	1,65 2,00	1,57 1,88	1,52 1,79	1,46 1,70	1,40 1,60	1,35 1,53	1,29 1,42	1,25 1,36	1,17 1,25	1,11 1,16	1,05 1,07

BẢNG 6a: KHOẢNG TIN CẬY CỦA ƯỚC LƯỢNG TỶ LỆ ($f=m/n$) VỚI ĐỘ TIN CẬY $P \geq 0,95$ KHI DUNG LƯỢNG MẪU BÉ ($4 \leq n \leq 10$)

n \ m	0	1	2	3	4	5	6	7	8	9	10
4	0	0,6	6,8	19,4	39,9						
	60,2	80,6	93,2	99,4	100,0						
5	0	0,5	5,3	14,7	28,4	47,8					
	52,2	71,16	85,3	94,7	99,5	100,0					
6	0	0,4	4,3	11,8	22,3	35,9	54,1				
	45,7	64,1	77,7	88,2	95,7	99,6	100,0				
7	0	0,4	3,7	9,9	18,4	29,0	42,1	59,0			
	41,9	57,9	71,0	80,6	90,1	96,3	99,6	1000,0			
8	0	0,3	3,2	8,5	15,7	24,5	34,8	47,3	63,1		
	36,9	52,7	65,2	75,5	84,3	91,5	96,8	99,7	100,0		
9	0	0,3	2,8	7,5	13,7	21,2	29,9	40,0	51,7	66,4	
	33,6	48,3	60,0	70,0	78,8	86,3	92,5	97,2	99,7	100,0	
10	0	0,3	2,5	6,7	12,2	18,7	26,2	34,8	44,4	55,5	69,2
	30,8	44,5	55,6	65,2	73,8	81,3	87,8	93,3	97,5	97,5	100,0

BẢNG 6b: KHOẢNG TIN CẬY CỦA ƯỚC LƯỢNG TỶ LỆ $f = \frac{m}{n}$ MẪU BÉ ($10 \leq n \leq 100$) VÀ m ($10 \leq x \leq 25$)

$\frac{m}{n}$	0	1	2	3	4	5	6	7	8	9	10	11	12
10	0 30,8	0,25 44,5	2,5 55,6	6,7 65,2	12,2 73,8	18,7 84,3							
20	0 16,8	0,13 24,9	1,2 31,7	3,2 37,9	5,7 43,7	8,7 49,1	11,9 54,3	15,4 59,4	19,1 63,9	23,1 68,5	27,2 72,8		
30	0 11,6	0,08 17,2	0,8 22,4	2,1 26,5	3,8 30,7	5,6 34,7	7,7 38,6	9,9 42,3	12,3 45,9	14,7 49,4	17,3 52,9	19,9 56,2	22,7 59,4
40	0 8,8	0,06 13,1	0,6 16,9	1,6 20,4	2,8 23,7	4,2 26,8	5,7 29,8	7,4 32,8	9,0 35,7	10,8 38,5	12,7 41,2	14,6 43,9	16,6 46,5
50	0 7,1	0,05 10,7	0,5 13,7	1,2 16,5	2,2 19,2	3,3 21,8	4,5 23,3	5,8 28,8	7,2 24,2	8,6 31,5	10,0 33,7	11,5 36,0	13,0 38,2
60	0 6,0	0,04 8,9	0,4 10,5	1,0 13,9	1,8 16,2	2,8 18,4	3,8 20,5	4,8 22,6	5,9 24,6	7,1 26,6	8,3 25,8	9,5 30,5	10,8 32,4
70	0 5,1	0,036 7,7	0,3 10,6	0,9 12,0	1,6 14,0	2,4 15,9	3,2 17,7	4,1 19,5	5,1 21,3	6,1 23,0	7,1 24,7	8,1 26,4	9,2 28,0
80	0 4,5	0,031 6,8	0,3 8,7	0,8 10,6	1,4 12,3	2,4 14,0	2,8 15,6	3,6 17,2	4,4 18,8	5,3 20,3	6,2 21,8	7,1 23,3	8,0 24,7
90	0 4,0	0,028 6,0	0,3 7,8	0,7 9,4	1,2 11,0	1,8 12,5	2,5 14,0	3,2 15,4	3,9 16,7	4,7 18,1	5,5 19,5	6,3 20,8	7,1 22,1
120	0 3,6	0,025 5,4	0,2 7,0	0,6 8,5	1,1 9,9	1,6 11,3	2,2 12,6	2,9 13,9	3,5 15,2	4,2 16,4	4,9 17,6	5,6 18,8	6,3 20,0

TIẾP BẢNG 6b

$\frac{m}{n}$	13	14	15	16	17	18	19	20	21	22	23	24	25
10													
20													
30	25,5 62,5	28,4 65,6	31,3 68,7										
40	18,6 49,0	20,6 51,5	22,7 54,2	24,9 56,5	27,1 58,9	29,3 61,4	31,5 63,8	33,8 66,2					
50	14,6 48,3	16,2 42,5	17,9 44,6	19,5 46,7	21,2 48,7	22,9 50,7	24,7 52,7	26,4 54,6	28,2 56,6	30,0 58,6	31,8 60,6	33,6 62,5	35,5 64,3
60	12,1 34,3	13,4 36,1	14,7 37,9	16,1 39,7	17,5 41,5	18,9 43,2	20,3 45,0	21,7 46,7	26,6 48,4	24,7 50,1	26,1 51,8	27,5 53,4	29,1 55,1
70	10,2 29,6	11,4 31,2	12,5 32,8	13,7 34,4	14,8 36,0	16,0 37,5	17,2 39,1	18,4 40,6	19,6 42,1	20,8 43,6	22,1 45,1	23,3 46,6	24,6 48,1
80	8,9 26,2	9,9 27,6	10,9 29,0	11,9 30,4	12,9 31,8	13,9 33,2	14,9 34,6	16,0 35,9	17,0 37,3	18,1 38,6	19,1 40,0	20,3 41,3	21,3 42,6
90	7,9 23,4	8,8 24,7	9,6 26,0	10,5 27,3	11,4 28,5	12,3 29,8	13,2 31,0	14,1 32,2	15,1 33,4	16,0 34,6	16,9 35,8	17,9 37,0	18,8 38,2
120	7,1 21,2	7,9 22,4	8,6 23,5	9,4 24,7	10,2 25,9	11,0 27,0	11,8 28,1	12,7 29,2	13,5 30,3	14,3 31,4	15,2 32,5	16,0 33,6	16,9 34,7

BẢNG 6c: KHOẢNG TIN CẬY CỦA ƯỚC LƯỢNG TỶ LỆ $f = \frac{m}{n}$ VỚI ĐỘ TIN CẬY 0,95

KHI DUNG DUNG LƯỢNG MẪU n (60 ≤ n ≤ 100) và (26 ≤ n ≤ 56)

m \ n	26	27	28	29	30	31	32	33	34	35	36
60	30,6 56,7	32,4 58,4	33,7 60,0	35,2 61,2	36,8 63,2						
70	25,9 49,5	27,2 51,0	28,5 52,4	29,8 53,8	31,1 55,3	32,4 56,7	33,7 58,1	35,1 59,5	36,4 60,8	37,8 62,2	
80	22,4 43,9	23,6 45,2	24,7 46,5	25,8 47,8	26,9 49,0	28,1 50,3	29,2 51,6	30,3 52,8	31,5 54,1	32,7 55,3	33,8 56,5
90	19,8 39,4	20,8 40,6	21,8 41,8	22,7 42,9	23,7 44,1	24,7 45,2	25,7 46,3	26,7 47,5	27,8 48,6	28,8 49,7	29,8 50,9
100	17,7 3	18,6 36,8	19,5 37,8	20,4 38,9	21,2 40,0	22,1 41,0	23,0 42,1	23,9 43,2	24,8 44,2	25,7 45,2	26,6 46,3

TIẾP BẢNG 6c

m n	37	38	39	40	41	42	43	44	45	46	47	48	49	50
60														
70														
80	35,0 57,8	36,2 59,0	37,4 60,2	38,6 61,4										
90	30,8 52,0	31,9 53,1	32,9 54,2	34,0 55,3	35,0 56,4	36,1 57,5	37,1 58,6	38,2 59,6	39,3 60,7					
100	27,6 4,73	28,5 48,3	29,4 49,3	30,3 50,3	31,3 51,3	32,2 52,3	33,1 53,3	34,1 54,3	35,0 55,3	36,0 56,3	36,9 57,2	37,9 58,2	38,9 59,2	39,8 60,2

BẢNG 7: KHOẢNG TIN CẬY CỦA ƯỚC LƯỢNG TẦN SUẤT p RẤT BÉ HOẶC RẤT LỚN Ở ĐỘ TIN CẬY $P \geq 0,95$ (các giá trị np_1 và np_2)

m → ↓	0	1	2	3	4	5	6	7	8	9	10
00		0,025	0,24	0,62	1,09	1,62	2,20	2,81	3,45	4,12	
		5,57	7,22	8,16	10,24	11,61	13,06	14,42	15,76	17,08	
10	4,8	5,5	6,2	6,9	7,6	8,3	9,0	9,8	10,6	11,4	
	18,4	19,7	21,0	22,3	23,6	24,9	26,1	27,3	28,5	29,7	
20	12,2	13,0	13,8	14,6	15,4	16,2	17,0	17,8	18,6	19,9	
	30,9	32,1	33,3	34,5	35,7	36,9	38,1	39,3	40,5	41,1	
30	20,2	21,0	21,3	22,7	23,5	24,4	25,2	26,1	26,9	27,8	
	42,8	40,0	45,2	46,3	47,5	48,6	49,8	50,9	52,1	55,3	
40	28,6	29,5	30,3	31,2	32,0	32,9	33,7	34,6	35,4	36,3	
	54,4	55,5	56,1	57,8	59,0	60,1	61,3	62,5	63,6	64,7	
50	37,1	38,0	38,8	39,7	40,5	41,4	42,3	43,1	40,0	44,9	
	56,9	67,0	68,2	69,3	70,8	71,6	72,7	73,9	75,0	76,1	
60	45,8	46,6	47,5	48,4	49,3	50,2	51,0	51,9	52,8	53,7	
	77,2	78,4	79,5	80,6	81,8	82,9	84,0	85,1	86,2	87,3	
70	54,6	55,3	55,5	57,2	58,1	59,0	59,9	60,8	61,7	62,6	
	88,4	89,5	90,7	91,8	92,9	94,0	95,1	96,2	97,3	98,4	
80	63,4	64,3	65,5	66,1	67,0	67,9	68,8	69,7	70,8	71,5	
	99,6	100,7	101,8	102,9	104,0	105,1	106,2	107,3	108,4	109,5	
90	72,4	73,3	74,2	75,1	76,0	76,9	77,8	78,7	79,6	80,5	81,4
	110,6	111,7	112,8	113,9	115,0	116,1	117,2	118,3	119,4	120,5	121,6

BẢNG 8 : ĐÔI CÁC GIÁ TRỊ f % RA GÓC φ VỚI GÓC $\varphi = \arcsin \sqrt{f\%}$

%	Phần lẻ									
	0	1	2	3	4	5	6	7	8	9
0	0,0	1,8	2,6	3,1	3,6	4,0	4,4	4,8	5,1	5,4
1	5,7	6,0	6,3	6,6	6,8	7,0	7,3	7,5	7,7	7,9
2	8,1	8,3	8,5	8,7	8,9	9,1	9,3	9,5	9,6	9,8
3	10,0	10,1	10,3	10,6	10,6	10,8	10,9	11,1	11,2	11,4
4	11,5	11,7	11,8	12,0	12,1	12,2	12,4	12,5	12,7	12,8
5	12,9	13,0	13,2	13,3	13,4	13,6	13,7	13,8	13,9	14,1
6	14,2	14,3	14,4	14,5	14,6	14,8	14,9	15,0	15,1	15,2
7	15,3	15,4	15,6	15,7	15,8	15,9	16,0	16,1	16,2	16,3
8	16,4	16,5	16,6	16,7	16,8	17,0	17,1	17,2	17,3	17,4
9	17,5	17,6	17,7	17,8	18,0	18,0	18,2	18,2	18,2	18,3
10	18,4	18,5	18,6	18,7	18,8	18,9	19,0	19,1	19,2	19,3
11	19,4	19,5	19,6	19,6	19,7	19,8	19,9	20,0	20,1	20,2
12	20,3	20,4	20,4	20,5	20,6	20,7	20,8	20,9	21,0	21,0
13	21,1	21,2	21,3	21,4	21,5	21,6	21,6	21,7	21,8	22,0
14	22,1	22,2	22,1	22,2	22,3	22,4	22,5	22,6	22,6	22,7
15	22,8	22,9	23,0	23,0	23,1	23,2	23,3	23,3	23,4	23,5
16	23,6	23,7	23,7	23,8	23,9	24,0	24,0	24,1	24,2	24,3
17	24,4	24,4	24,5	24,6	24,6	24,7	24,8	24,9	25,0	25,0
18	25,1	25,2	25,2	25,3	25,3	25,5	25,6	25,6	25,7	25,8
19	25,8	25,9	26,0	26,1	26,1	26,2	26,3	26,4	26,4	26,5
20	26,6	26,6	26,7	26,8	26,9	26,9	27,0	27,1	27,1	27,2
21	27,3	27,4	27,4	27,5	27,6	27,6	27,7	27,8	27,8	27,9
22	28,0	28,0	28,1	28,2	28,2	28,8	28,4	28,4	28,5	28,6
23	28,7	28,7	28,8	28,9	28,9	29,0	29,1	29,1	29,2	29,3
24	29,3	29,4	29,5	29,5	29,6	29,7	29,8	29,8	29,9	29,9
25	30,0	30,1	30,1	30,2	30,3	30,3	30,4	30,5	30,5	30,6
26	30,7	30,7	30,8	30,9	30,9	31,0	31,0	31,0	31,2	31,2
27	31,2	31,3	31,4	32,5	31,6	31,6	31,7	31,8	31,8	31,9
28	32,0	32,0	32,1	32,1	32,2	32,3	32,3	32,4	32,5	32,5
29	32,6	32,6	32,7	32,8	32,8	32,9	33,0	33,0	33,1	33,2
30	33,2	33,3	33,3	33,4	33,5	33,5	33,6	33,6	33,7	33,8
31	33,8	33,9	34,0	34,0	34,1	34,1	34,2	34,3	43,3	34,4
32	34,4	34,5	34,6	34,6	34,7	34,8	34,8	34,9	35,0	35,0
33	35,1	35,1	35,2	35,2	35,3	35,4	35,4	35,5	35,6	35,6
34	35,7	35,7	35,8	35,9	35,9	36,0	36,0	36,1	36,2	36,2

%	Phần lễ									
	0	1	2	3	4	5	6	7	8	9
35	36,3	36,3	36,4	36,5	36,5	36,6	36,6	36,7	36,8	36,8
36	36,9	36,9	37,0	37,0	37,1	37,2	37,2	37,3	37,4	37,4
37	37,5	37,5	37,6	37,6	37,7	37,8	37,8	37,9	37,9	38,0
38	38,1	38,1	38,2	38,2	38,3	38,4	38,4	38,5	38,5	38,6
39	38,6	38,7	38,8	38,8	38,9	38,9	39,0	39,1	39,1	39,2
40	39,2	39,3	39,4	39,4	39,5	39,5	39,6	39,6	39,7	39,8
41	39,8	39,9	39,9	40,0	40,0	40,1	40,2	40,2	40,3	40,3
42	40,4	40,5	40,5	40,6	40,6	40,7	40,7	40,8	40,9	40,9
43	41,0	41,0	41,1	41,2	41,2	41,2	41,3	41,4	41,4	41,5
44	41,6	41,6	41,7	41,7	41,8	41,8	41,9	42,0	42,0	42,L
45	42,1	42,2	42,2	42,3	42,4	42,4	42,5	42,5	42,6	42,6
46	42,7	42,8	42,8	42,9	42,9	43,0	43,1	43,1	43,2	43,2
47	43,1	43,3	43,4	43,4	43,5	43,6	43,6	43,7	43,7	43,8
48	43,8	43,9	44,0	44,0	44,1	44,1	44,2	44,3	44,3	44,4
49	44,4	44,5	44,5	44,5	44,7	44,7	44,8	44,8	44,9	44,9
50	45,0	45,0	45,1	45,2	45,2	45,3	45,3	45,5	45,5	45,5
51	45,6	45,6	45,7	45,8	45,8	45,9	45,9	46,0	46,0	46,1
52	46,2	46,2	46,3	46,3	46,4	46,5	46,5	46,6	46,6	46,7
53	46,7	46,8	46,8	46,9	47,0	47,0	47,1	47,1	47,2	47,2
54	47,3	47,4	47,4	47,5	47,5	47,6	47,6	47,7	47,8	47,8
55	47,9	47,9	48,0	48,0	48,1	48,2	48,2	48,3	48,3	48,4
56	48,4	48,5	48,6	48,7	48,7	48,7	48,8	48,8	48,9	49,0
57	49,0	49,1	49,2	49,2	49,3	49,3	49,4	49,4	49,5	49,5
58	49,6	49,7	49,8	49,8	49,8	49,9	49,9	50,0	50,1	50,1
59	50,2	50,2	50,3	50,4	50,4	50,5	50,5	50,6	50,6	50,7
60	50,8	50,8	50,9	50,9	51,0	51,0	51,1	51,2	51,2	51,3
61	51,4	51,4	51,5	51,5	51,6	52,6	51,7	51,8	51,8	51,9
62	51,9	52,0	52,1	52,1	52,2	52,2	52,3	52,3	52,4	52,5
63	52,5	52,6	52,6	52,7	52,8	52,8	52,9	53,0	53,0	53,1
64	53,1	53,2	53,3	53,3	53,4	53,4	53,5	53,6	53,6	53,7

65	53,7	53,8	53,8	53,9	54,0	54,1	54,1	54,2	54,2	54,3
66	54,3	54,4	54,4	54,5	54,6	54,6	54,7	54,8	54,8	54,9
67	54,9	55,0	55,1	55,1	55,2	55,2	55,3	55,4	55,4	55,5
68	55,6	55,6	55,7	55,7	55,8	55,9	55,9	56,0	56,0	56,1
69	56,2	56,2	56,3	56,4	56,4	56,5	56,5	56,6	56,6	56,7
70	56,8	56,8	56,9	57,0	57,0	57,1	57,2	57,2	57,2	57,4
71	57,4	57,5	57,5	57,6	57,7	57,7	57,8	57,9	57,9	58,0
72	58,0	58,1	58,2	58,2	58,3	58,4	58,4	58,5	58,6	58,6
73	58,7	58,8	58,8	58,9	59,0	59,0	59,1	59,2	59,2	59,3
75	59,3	59,4	59,4	59,5	59,6	59,7	59,7	59,8	59,9	59,9

%	Phân lễ									
	0	1	2	3	4	5	6	7	8	9
75	60,0	60,1	60,1	60,2	60,3	60,3	60,4	60,5	60,5	60,6
76	60,7	60,7	60,8	60,9	60,9	61,0	61,1	61,2	61,2	61,3
77	61,3	61,4	61,5	61,6	61,6	61,7	61,8	61,8	61,9	62,0
78	62,0	62,1	62,2	62,2	62,3	62,4	62,4	62,5	62,6	62,6
79	62,7	62,8	62,9	62,9	63,0	63,1	63,2	63,2	63,3	63,4
80	63,4	63,5	63,6	63,6	63,7	63,8	63,9	63,9	64,0	64,1
81	64,1	64,2	64,3	64,4	64,4	64,5	64,6	64,7	64,8	64,8
82	64,9	65,0	65,0	65,1	65,2	65,3	65,4	65,4	65,5	65,6
83	65,6	65,7	65,8	65,9	66,0	66,0	66,1	66,2	66,3	66,3
84	66,4	66,5	66,6	66,7	66,7	66,8	66,9	67,0	67,0	67,1
85	67,2	67,3	67,4	67,4	67,5	67,6	67,7	67,8	67,9	67,9
86	68,0	68,1	68,2	68,3	68,4	68,4	68,5	68,6	68,7	68,8
87	68,9	69,0	69,0	69,1	69,2	69,3	69,4	69,5	69,6	69,6
88	69,7	69,8	69,9	70,0	70,1	70,2	70,3	70,4	70,4	70,5
89	70,6	70,7	70,8	70,9	71,1	71,1	71,2	71,3	71,4	71,5
90	71,6	71,7	71,8	71,8	72,0	72,0	72,2	72,2	72,3	72,4
91	72,5	72,6	72,7	72,8	73,0	73,0	73,2	73,3	73,4	73,5
92	73,6	73,7	73,8	73,9	74,0	74,1	74,2	74,3	74,4	74,6
93	74,7	74,8	74,9	75,0	75,1	75,2	75,4	75,5	75,6	75,7
94	75,8	75,9	76,1	76,2	76,3	76,4	76,6	76,7	76,8	77,0
95	77,1	77,2	77,3	77,5	77,6	77,8	77,9	78,0	78,2	78,3
96	78,5	78,6	78,8	78,9	79,1	79,2	79,4	79,5	79,7	79,9
97	80,0	80,2	80,4	80,5	80,7	80,9	81,1	81,3	81,5	81,7
98	81,9	82,1	82,3	82,5	82,7	83,0	83,1	83,4	83,7	84,0
99	84,3	84,6	84,9	85,2	85,6	86,0	86,4	86,9	87,4	84,2
100	90,0									

BẢNG 9. ĐỔI CÁC GIÁ TRỊ TẦN SUẤT f RA GÓC φ với $\varphi = 2\frac{\Pi}{180} \arcsin \sqrt{f}$

f	0	1	2	3	4	5	6	7	8	9
0,0000	0,0000	0,0063	0,0089	0,0110	0,0126	0,0442	0,0155	0,0167	0,0179	0,0190
0,0001	0,0200	0,0210	0,0220	0,0228	0,0236	0,0245	0,0253	0,0261	0,0269	0,0276
0,0002	0,0283	0,0290	0,0297	0,0303	0,0310	0,0317	0,0323	0,0329	0,0335	0,0341
0,0003	0,0346	0,0352	0,0358	0,0363	0,0369	0,0374	0,0379	0,0385	0,0390	0,0394
0,0004	0,0400	0,0405	0,0410	0,0415	0,0420	0,0425	0,0429	0,0434	0,0438	0,0443
0,0005	0,11448	0,11452	0,0456	0,11460	0,11465	0,11469	0,0173	0,11478	0,11482	0,11486
0,0006	0,0490	0,11494	0,11498	0,0502	0,0506	0,0510	0,0514	0,0518	0,0522	0,0525
0,0007	0,0529	0,0533	0,0537	0,0540	0,0543	0,0547	0,0551	0,0555	0,0559	0,0562
0,0008	0,0566	0,0569	0,0573	0,0576	0,0580	0,0583	0,0587	0,0590	0,0593	0,0597
0,0009	0,0600	0,06114	0,0607	0,0610	0,0613	0,0616	0,0620	0,0623	0,0626	0,0629
0,001	0,0632	0,0664	0,0693	0,0721	0,0748	0,0775	0,0800	0,0824	0,0849	0,0372
0,002	0,0895	0,0917	0,0939	0,0959	0,0979	0,1000	0,1020	0,1039	0,1059	0,1078
0,003	0,1096	0,1114	0,1132	0,1149	0,1167	0,1184	0,1200	0,1218	0,1234	0,1250
0,004	0,1266	0,1282	0,1297	0,1312	0,1328	0,1343	0,1358	0,1373	0,1387	0,14114
0,005	0,1415	0,1429	0,1443	0,1457	0,1471	0,1488	0,1498	0,1511	0,1525	0,1538
0,006	0,1551	0,1563	0,1576	0,1589	0,1602	0,1614	0,1626	0,1639	0,1651	0,1664
0,007	0,1676	0,1687	0,1699	0,1711	0,1723	0,1734	0,1746	0,1757	0,1769	0,1780
0,008	0,1791	0,1803	0,1814	0,1825	0,1836	0,1847	0,1857	0,1868	0,1879	0,1890
0,009	0,1900	0,1911	0,1921	0,1932	0,1942	0,1953	0,1963	0,1973	0,1983	0,1993
0,010	0,2003	0,2102	0,2165	0,2286	0,2372	0,2456	0,2537	0,2615	0,2691	0,2765

BẢNG 10. HỆ SỐ TƯƠNG QUAN LÝ THUYẾT r_α

α df	0,05	0,01	α df	0,05	0,01
1	0,997	1,000	24	0,388	0,496
2	0,950	0,990	25	0,378	0,487
3	0,878	0,959	26	0,374	0,478
4	0,811	0,917	27	0,367	0,470
5	0,754	0,874	28	0,361	0,463
6	0,707	0,834	29	0,355	0,456
7	0,666	0,798	30	0,349	0,449
8	0,632	0,765	35	0,325	0,418
9	0,602	0,735	40	0,304	0,383
10	0,576	0,708	45	0,288	0,372
11	0,553	0,684	50	0,273	0,354
12	0,532	0,661	60	0,250	0,325
13	0,514	0,644	70	0,232	0,302
14	0,479	0,623	80	0,217	0,283
15	0,482	0,606	90	0,205	0,267
16	0,458	0,590	100	0,195	0,254
17	0,456	0,575	125	0,174	0,228
18	0,444	0,561	150	0,159	0,208
19	0,433	0,549	200	0,138	0,181
20	0,423	0,537	300	0,113	0,148
21	0,413	0,526	400	0,098	0,128
22	0,404	0,515	500	0,088	0,115
23	0,396	0,505	1000	0,062	0,081

**BẢNG 11. CÁC GIÁ TRỊ ỨNG VỚI BIÊN ĐỘ ĐÁNG KÊ
CỦA DUNCAN ($\alpha \leq 0,05$)**

P df	2	3	4	5	6	7	8	9
1	17,97	17,97	17,97	17,97	17,97	17,97	17,97	17,97
2	6,08	6,08	6,08	6,08	6,08	6,08	6,08	6,08
3	4,50	4,52	4,52	4,52	4,52	4,52	4,52	4,52
4	3,93	4,01	4,03	4,30	4,03	4,03	4,03	4,03
5	3,64	3,57	3,80	3,81	3,81	3,81	3,81	3,81
6	3,46	3,59	3,65	3,68	3,69	3,70	3,70	3,70
7	3,34	3,48	3,55	3,59	3,61	3,62	3,63	3,63
8	3,26	3,40	3,48	3,52	3,55	3,57	3,58	3,58
9	3,20	3,34	3,42	3,47	3,50	3,52	3,54	3,54
10	3,15	3,29	3,38	3,43	3,46	3,49	3,50	3,52
11	3,11	3,26	3,34	3,40	3,44	3,46	3,48	3,48
12	3,08	3,22	3,31	3,37	3,41	3,44	3,46	3,47
13	3,06	3,20	3,29	3,35	3,39	3,42	3,44	3,46
14	3,03	3,18	3,27	3,33	3,37	3,40	3,43	3,44
15	3,01	3,16	3,25	3,31	3,36	3,39	3,41	3,43
16	3,00	3,14	3,24	3,30	3,34	3,38	3,40	3,42
17	2,98	3,13	3,22	3,27	3,33	3,37	3,39	3,41
18	2,97	3,12	3,21	3,27	3,32	3,36	3,38	3,40
19	2,96	3,11	3,20	3,26	3,31	3,35	3,38	3,40
20	2,95	3,10	3,19	3,26	3,30	3,34	3,37	3,39
24	2,92	3,07	3,16	3,23	3,28	3,32	3,34	3,37
30	2,89	3,04	3,13	3,20	3,25	3,29	3,32	3,35
40	2,86	3,01	3,10	3,17	3,22	3,27	3,30	3,33
60	2,83	2,98	3,07	3,14	3,20	3,24	3,28	3,31
120	2,80	2,95	3,04	3,12	3,17	3,22	3,25	3,29
∞	2,77	2,92	3,02	3,09	3,15	3,19	3,23	3,26

TIẾP BẢNG 11 VỚI Ý NGHĨA $\alpha \leq 0,05$

df \ P	10	12	14	16	18	20	50	100
1	17,97	17,97	17,97	17,97	17,97	17,97	17,97	17,97
2	6,08	6,08	6,08	6,08	6,08	6,08	6,08	6,08
3	4,52	4,52	4,52	4,52	4,52	4,52	4,52	4,52
4	4,03	4,03	4,03	4,03	4,03	4,03	4,03	4,03
5	3,81	3,81	3,81	3,81	3,81	3,81	3,81	3,81
6	3,70	3,70	3,70	3,70	3,70	3,70	3,70	3,70
7	3,63	3,63	3,63	3,63	3,63	3,63	3,63	3,63
8	3,58	3,58	3,58	3,58	3,58	3,58	3,58	3,58
9	3,55	3,55	3,55	3,55	3,55	3,55	3,55	3,55
10	3,52	3,52	3,52	3,52	3,52	3,52	3,52	3,52
11	3,50	3,51	3,51	3,51	3,51	3,51	3,51	3,51
12	3,48	3,50	3,50	3,50	3,50	3,50	3,50	3,50
13	3,47	3,48	3,49	3,49	3,49	3,49	3,49	3,49
14	3,46	3,47	3,48	3,48	3,48	3,48	3,48	3,48
15	3,45	3,46	3,48	3,48	3,48	3,48	3,48	3,48
16	3,44	3,46	3,47	3,48	3,48	3,48	3,48	3,48
17	3,43	3,45	3,46	3,47	3,48	3,48	3,48	3,48
18	3,42	3,44	3,46	3,47	3,47	3,47	3,47	3,47
19	3,42	3,44	3,46	3,47	3,47	3,47	3,47	3,47
20	3,41	3,44	3,45	3,46	3,47	3,47	3,47	3,47
24	3,39	3,42	3,44	3,46	3,46	3,47	3,48	3,48
30	3,37	3,40	3,43	3,45	3,46	3,47	3,49	3,49
40	3,35	3,39	3,42	3,44	3,46	3,47	3,50	3,50
60	3,33	3,37	3,41	3,43	3,45	3,47	3,54	3,54
120	3,31	3,36	3,39	3,42	3,45	3,47	3,58	3,60
∞	3,29	3,34	3,38	3,41	3,44	3,47	3,64	3,74

TIẾP BẢNG 11 VỚI Ý NGHĨA $\alpha \leq 0,01$

P df	1	2	3	4	5	6	7	8
1	90,03	90,03	90,03	90,03	90,03	90,03	90,03	90,03
2	14,04	14,04	14,04	14,04	14,04	14,04	14,04	14,04
3	8,26	8,32	8,32	8,32	8,32	8,32	8,32	8,32
4	6,51	6,68	6,74	6,76	6,76	6,76	6,76	6,76
5	5,70	5,89	5,99	6,04	6,06	6,07	6,07	6,07
6	5,34	5,44	5,55	5,61	5,66	5,68	5,69	5,70
7	4,95	5,14	5,26	5,33	5,38	5,42	5,44	5,45
8	4,75	4,94	5,06	5,14	5,19	5,23	5,26	5,28
9	4,60	4,79	4,91	4,99	5,04	5,09	5,12	5,14
10	4,48	4,67	4,79	4,87	4,93	4,98	5,01	5,04
11	4,39	4,58	4,70	4,78	4,84	4,89	4,92	4,95
12	4,32	4,50	4,62	4,71	4,77	4,82	4,85	4,88
13	4,26	4,44	4,56	4,64	4,71	4,76	4,79	4,82
14	4,21	4,39	4,51	4,59	4,65	4,70	4,74	4,78
15	4,17	4,35	4,46	4,55	4,61	4,66	4,70	4,73
16	4,13	4,31	4,42	4,51	4,57	4,62	4,66	4,70
17	4,10	4,28	4,39	4,48	4,54	4,59	4,63	4,66
18	4,07	4,25	4,36	4,44	4,51	4,56	4,60	4,64
19	4,05	4,22	4,34	4,42	4,48	4,53	4,58	4,61
20	4,02	4,20	4,31	4,40	4,46	4,51	4,55	4,59
24	3,96	4,13	4,24	4,32	4,39	4,44	4,48	4,52
30	3,89	4,06	4,17	4,25	4,31	4,37	4,41	4,44
40	3,82	3,99	4,10	4,18	4,24	4,30	4,34	4,38
60	3,76	3,92	4,03	4,11	4,17	4,23	4,27	4,31
120	3,70	3,86	3,96	4,04	4,11	4,16	4,20	4,24
∞	3,64	3,80	3,90	3,98	4,04	4,09	4,14	4,17

TIẾP BẢNG 11 VỚI Ý NGHĨA $\alpha \leq 0,01$

df \ P	10	12	14	16	18	20	50	100
1	90,03	90,03	90,03	90,03	90,03	90,03	90,03	90,03
2	14,04	14,04	14,04	14,04	14,04	14,04	14,04	14,04
3	8,32	8,32	8,32	8,32	8,32	8,32	8,32	8,32
4	6,76	6,76	6,76	6,76	6,76	6,76	6,76	6,76
5	6,07	6,07	6,07	6,07	6,07	6,07	6,07	6,07
6	5,70	5,70	5,70	5,70	5,70	5,70	5,70	5,70
7	5,46	5,47	5,47	5,47	5,47	5,47	5,47	5,47
8	5,29	5,31	5,32	5,32	5,32	5,32	5,32	5,32
9	5,16	5,18	5,20	5,20	5,21	5,21	5,21	5,21
10	5,06	5,09	5,11	5,12	5,12	5,12	5,12	5,12
11	4,98	5,01	5,03	5,04	5,05	5,06	5,06	5,06
12	4,91	4,94	4,97	4,99	5,00	5,01	5,01	5,01
13	4,85	4,89	4,92	4,94	4,95	4,96	4,97	4,97
14	4,80	4,84	4,87	4,89	4,91	4,92	4,94	4,94
15	4,76	4,80	4,83	4,86	4,87	4,89	4,91	4,91
16	4,72	4,77	4,80	4,82	4,84	4,86	4,89	4,89
17	4,69	4,74	4,77	4,80	4,82	4,83	4,87	4,87
18	4,66	4,71	4,74	4,77	4,97	4,81	4,86	4,86
19	4,64	4,69	4,72	4,75	4,77	4,79	4,84	4,84
20	4,62	4,66	4,70	4,73	4,75	4,77	4,83	4,83
24	4,55	4,60	4,63	4,66	4,69	4,71	4,80	4,80
30	4,48	4,53	4,57	4,60	4,63	4,65	4,77	4,78
40	4,41	4,46	4,50	4,54	4,57	4,59	4,74	4,76
60	4,34	4,39	4,44	4,47	4,50	4,53	4,71	4,76
120	4,27	4,33	4,37	4,41	4,44	4,47	4,67	4,77
∞	4,20	4,26	4,31	4,34	4,38	4,41	4,64	4,78

BẢNG 12: BẢNG ĐỔI r THÀNH Z = $\frac{1}{2} \text{Ln} \frac{1+r}{1-r}$

r	z	R	z	r	z	r	z
0,000	0,125	0,126	0,250	0,255	0,375	0,394	0,394
0,005	0,130	0,131	0,255	0,261	0,380	0,400	0,400
0,010	0,135	0,136	0,260	0,266	0,385	0,406	0,406
0,015	0,140	0,141	0,265	0,271	0,390	0,412	0,412
0,020	0,145	0,146	0,270	0,277	0,395	0,418	0,418
0,025	0,025	0,150	0,151	0,275	0,282	0,400	0,424
0,030	0,030	0,155	0,156	0,280	0,288	0,405	0,430
0,035	0,035	0,160	0,161	0,285	0,293	0,410	0,436
0,040	0,040	0,165	0,167	0,290	0,299	0,415	0,442
0,045	0,045	0,170	0,172	0,295	0,304	0,420	0,448
0,050	0,050	0,175	0,177	0,300	0,310	0,425	0,454
0,055	0,055	0,180	0,182	0,305	0,315	0,430	0,460
0,060	0,060	0,185	0,187	0,310	0,321	0,435	0,466
0,065	0,065	0,190	0,192	0,315	0,326	0,440	0,472
0,070	0,070	0,195	0,198	0,320	0,332	0,445	0,478
0,075	0,075	0,200	0,203	0,325	0,337	0,450	0,485
0,080	0,080	0,205	0,208	0,330	0,343	0,455	0,491
0,085	0,085	0,210	0,213	0,335	0,348	0,460	0,497
0,090	0,090	0,215	0,218	0,340	0,354	0,465	0,504
0,095	0,095	0,220	0,224	0,345	0,360	0,470	0,510
0,100	0,100	0,225	0,229	0,350	0,365	0,475	0,517
0,105	0,105	0,230	0,234	0,355	0,371	0,480	0,523
0,110	0,110	0,235	0,239	0,360	0,377	0,485	0,530
0,115	0,116	0,240	0,245	0,365	0,383	0,490	0,536
0,120	0,121	0,245	0,250	0,370	0,388	0,495	0,543

TIẾP BẢNG 12

r	z	R	z	r	z	r	z
0,500	0,549	0,625	0,733	0,750	0,973	0,875	1,354
0,505	0,556	0,630	0,741	0,755	0,984	0,880	1,376
0,510	0,563	0,635	0,750	0,760	0,996	0,885	1,398
0,515	0,570	0,640	0,758	0,765	1,008	0,890	1,422
0,520	0,576	0,645	0,767	0,770	1,020	0,895	1,447
0,525	0,583	0,650	0,775	0,775	1,113	0,930	1,658
0,530	0,590	0,655	0,784	0,780	1,127	0,935	1,697
0,535	0,597	0,660	0,793	0,785	1,142	0,940	1,738
0,540	0,604	0,665	0,802	0,790	1,157	0,945	1,783
0,545	0,611	0,670	0,811	0,795	1,172	0,950	1,832
0,550	0,618	0,675	0,820	0,800	1,188	0,950	1,886
0,555	0,626	0,680	0,829	0,805	1,113	0,930	1,658
0,560	0,633	0,685	0,838	0,810	1,127	0,935	1,697
0,565	0,640	0,690	0,848	0,815	1,142	0,940	1,738
0,570	0,648	0,695	0,858	0,820	1,157	0,945	1,783
0,575	0,655	0,700	0,867	0,825	1,172	0,950	1,832
0,580	0,662	0,705	0,877	0,830	1,188	0,955	1,886
0,585	0,670	0,710	0,877	0,835	1,204	0,960	1,946
0,590	0,678	0,715	0,897	0,840	1,221	0,965	2,014
0,595	0,685	0,720	0,908	0,845	1,238	0,970	2,092
0,600	0,693	0,725	0,918	0,850	1,256	0,975	2,815
0,605	0,701	0,730	0,929	0,855	1,274	0,980	2,298
0,610	0,709	0,735	0,940	0,860	1,293	0,985	2,443
0,615	0,717	0,740	0,950	0,865	1,313	0,990	2,647
0,620	0,125	0,745	0,962	0,870	1,333	0,995	2,994

BẢNG 13: CÁC GIÁ TRỊ LÝ THUYẾT CỦA TIÊU CHUẨN ω

df	$\alpha \leq 0,01$	df	$\alpha \leq 0,01$
3	1,41	15	2,39
4	1,70	16	2,40
5	1,87	17	2,41
6	2,00	18	2,42
7	2,10	20	2,43
8	2,17	22	2,45
9	2,23	25	2,48
10	2,26	30	2,49
11	2,30	40	2,53
12	2,32	60	2,54
13	2,35	120	2,56
14	2,37	210	2,57
		∞	2,578

TÀI LIỆU THAM KHẢO

1. Dương Văn Ban: Luận văn Thạc sĩ nông nghiệp, 1998
2. Bohumi Kaba: Statistika – Giáo trình của trường Đại học nông nghiệp Praha (nguyên bản tiếng Séc), NXB Giáo dục Praha, 1980.
3. Clair.A. Collins and Frances M. Seeny: Statistical Experimental Design and Interpretation. Copyright 1999 by John Wiley and Sons Ltd.
4. Nguyễn Văn Cương: Luận văn Tiến sĩ nông nghiệp, 2004.
5. Kiều Xuân Đàm: Luận văn Tiến sĩ nông nghiệp, 2002.
6. Nguyễn Văn Đức: Mô hình thí nghiệm nông nghiệp. NXB Nông nghiệp Hà Nội, 2002.
7. Goger Piteresen: Agricultural Field Experiments. Copyright by Merel Dekker Inc.
8. Nguyễn Đình Hiền: Giáo trình xác suất thống kê. NXB Đại học Sư phạm, 2004.
9. Jana Stavkova: Biomemtrika – Giáo trình của trường đại học BRNO (nguyên bản tiếng Séc), 1988.
10. Jaroslav Kubisec and Jaroslav Dufek: Statistika – Giáo trình của trường Đại học nông nghiệp BRNO (nguyên bản tiếng Séc), 1978.
11. Jim Foler, Lou Cohen and Phil Jarvis: Pratical Statistical for Field Biology. Second Edition. Copyright 1988 by John & Wiley.
12. Ngô Kim Khôi. Thống kê toán học trong lâm nghiệp. Nhà xuất bản Nông nghiệp, 1998.
13. Kwanchai A. Gomez & Arturo A. Gomez. Statistical Procedures for Agricultural Research. 2 nd Edition. Copyright 1984 by John Wiley & Sons, Inc.
14. Nguyễn Thị Lan: Xác định dung lượng mẫu cần thiết ở một số chỉ tiêu nghiên cứu đối với cây đậu tương. Tạp chí KHKTNN, Trường Đại học Nông nghiệp I, tập I số 4/2003.
15. Nguyễn Thị Lan: Xác định dung lượng mẫu cần thiết ở một số chỉ tiêu nghiên cứu đối với cây lúa. Tạp chí KHKTNN, Trường Đại học Nông nghiệp I, tập II số 4/2005.
16. Nguyễn Thị Lan: Xác định dung lượng mẫu cần thiết ở một số chỉ tiêu nghiên cứu trong thí nghiệm đồng ruộng với cây trồng cạn. Kết quả NCKH, Khoa Nông học 1995 –1996, Trường Đại học nông nghiệp I, NXB Nông nghiệp, 1997.
17. Trần Kim Loang: Luận văn Tiến sĩ nông nghiệp, 1997.
18. Chu Văn Mẫn. Ứng dụng tin học trong sinh học. Nhà xuất bản Đại học Quốc gia Hà nội, 2001
19. Rober G.D. Steed and Jame H. Torrie: Priciplea and Procedures of statistics. Copyright 1960 by the Mc Graw – Hill Book Company.
20. Lê Văn Tiến. Giáo trình lý thuyết xác suất và thống kê toán học. Nhà xuất bản Nông nghiệp, 1999.
21. Phạm Chí Thành, Phạm Tiến Dũng và Nguyễn Thị Lan: Biện pháp nâng cao độ chính xác của thí nghiệm đồng ruộng Việt Nam. Tuyển tập các công trình nghiên cứu KHKTNN, Trường Đại học Nông nghiệp I. NXB Nông nghiệp, 1996.
22. Phạm Chí Thành: Giáo trình Phương pháp thí nghiệm đồng ruộng. NXB Nông nghiệp, 1998
23. Phạm Chí Thành, Phạm Tiến Dũng: Biện pháp nâng cao độ chính xác của thí nghiệm trên đất dốc. Tuyển tập các công trình nghiên cứu KHKTNN, Trường Đại học Nông nghiệp I. NXB Nông nghiệp, 1996.

24. Lê Văn Tiến. Giáo trình lý thuyết xác suất và thống kê toán học. Nhà xuất bản Nông nghiệp, 1999.
25. Ngô Kim Khôi. Thống kê toán học trong lâm nghiệp. Nhà xuất bản Nông nghiệp, 1998
26. Chu Văn Mẫn. Ứng dụng tin học trong sinh học. Nhà xuất bản Đại học Quốc gia Hà nội, 2001.
27. Nguyễn Văn Đức. Mô hình thí nghiệm trong nông nghiệp. Nhà xuất bản Nông nghiệp, Hà nội, 2002.

MỤC LỤC

MỞ ĐẦU

CHƯƠNG I - ĐẠI CƯƠNG VỀ CÔNG TÁC NGHIÊN CỨU TRONG KHOA HỌC NÔNG NGHIỆP

- 1.1. Vai trò của công tác nghiên cứu khoa học nông nghiệp.....
- 1.2. Các bước trong quá trình nghiên cứu khoa học nông nghiệp
 - 1.2.1. Thu thập thông tin (Bước 1)
 - 1.2.2. Xây dựng giả thuyết khoa học (Bước 2)
 - 1.2.3. Chứng minh giả thuyết khoa học (Bước 3)
 - 1.2.4. Biện luận để rút ra kết luận và xây dựng lý thuyết khoa học (Bước 4)
- 1.3. Các nhóm phương pháp thí nghiệm trong nông học
- 1.3.1. Nhóm của các nghiên cứu trong phòng.....
- 1.3.2. Nhóm các thí nghiệm trong nhà lưới
- 1.3.3. Nhóm phương pháp nghiên cứu trên đồng ruộng.....

CHƯƠNG II - THIẾT KẾ THÍ NGHIỆM TRÊN ĐỒNG RUỘNG

- 2.1. Các yêu cầu của thí nghiệm đồng ruộng.....
 - 2.1.1. Thí nghiệm đồng ruộng phải mang tính đại diện.....
 - 2.1.2. Yêu cầu về nguyên tắc sai khác duy nhất.....
 - 2.1.3. Thí nghiệm phải đạt độ chính xác quy định
 - 2.1.4. Thí nghiệm phải có khả năng diễn lại
 - 2.1.5. Thí nghiệm phải được đặt trên những mảnh ruộng có lịch sử canh tác rõ ràng
 - 2.2. Các loại thí nghiệm ngoài đồng ruộng
 - 2.2.1. Thí nghiệm thăm dò
 - 2.2.2. Thí nghiệm chính thức
 - 2.2.2.1. Theo số lượng nhân tố thí nghiệm.....
 - 2.2.2.2. Chia theo thời gian nghiên cứu.....
 - 2.2.2.3. Theo khối lượng nghiên cứu
 - 2.2.3. Thí nghiệm làm trong điều kiện sản xuất.....
- 2.3. Xây dựng chương trình thực hiện thí nghiệm.....
 - 2.3.1. Một số vấn đề liên quan đến xây dựng chương trình thí nghiệm
 - 2.3.1.1. Các loại công thức trong thí nghiệm.....

2.3.1.2. Xác định số lượng công thức trong một thí nghiệm	
2.3.1.3. Một số chỉ tiêu kỹ thuật có liên quan khi thiết kế thí nghiệm	
2.3.2. Xây dựng nền thí nghiệm	
2.3.3. Chọn đất thí nghiệm	
2.3.3.1. Điều tra về địa hình	
2.3.3.2. Điều tra lý tính đất và hoá tính đất	
2.4. Xây dựng đề cương nghiên cứu	
2.4.1. Cơ sở để xây dựng đề cương	
2.4.2. Nội dung của đề cương nghiên cứu	

CHƯƠNG III - TIẾN HÀNH THÍ NGHIỆM TRÊN ĐỒNG RUỘNG Error! Bookmark not defined.

3.1. Chia ô thí nghiệm	
3.1.1. Xác định ranh giới khu thí nghiệm	
3.1.2. Xác định ranh giới lần nhắc lại	
3.1.3. Xác định ranh giới các ô thí nghiệm	
3.1.4. Cắm biển tên các công thức và tên thí nghiệm	
3.2. Làm đất, bón phân và gieo cấy thí nghiệm	
3.2.1. Làm đất	
3.2.2. Bón phân cho thí nghiệm	
3.2.3. Gieo, cấy thí nghiệm	
3.2.4. Chăm sóc thí nghiệm	
3.3. Theo dõi thí nghiệm	
3.3.1. Quan sát thí nghiệm	
3.3.2. Phương pháp chọn mẫu nghiên cứu	
3.3.2.1. Chọn mẫu ngẫu nhiên	
3.3.2.2. Chọn mẫu phân bố đều	
3.3.3. Độ lớn của mẫu	
3.4. Thu hoạch thí nghiệm	
3.4.1. Công tác chuẩn bị	
3.4.2. Lấy mẫu định yếu tố năng suất và năng suất cá thể	
3.4.3. Thu hoạch năng suất	
3.4.4. Phương pháp điều chỉnh năng suất ô thí nghiệm	

CHƯƠNG IV - TỔNG KẾT SỐ LIỆU QUAN SÁT

4.1. Chinh lý số liệu quan sát	4.1.1. Khái niệm về số liệu thô và số liệu tinh.....
4.1.1.1. Số liệu thô	
4.1.1.2. Số liệu tinh	
4.2. Phân loại số liệu	
4.2.1. Số liệu mang tính chất định lượng	
4.2.2. Số liệu mang tính định tính.....	
4.3. Kiểm tra số liệu nghi ngờ.....	
4.4. Cách sắp xếp số liệu	
4.4.1. Với số liệu định lượng.....	
4.4.2. Số liệu mang tính định tính.....	
4.5. Các tham số đặc trưng của mẫu	
4.5.1. Các tham số đại diện về độ lớn của mẫu	
4.5.1.1. Số bình quân (\bar{x}).....	
4.5.1.2. Số một (Mode).....	
4.5.1.3. Số trung vị (Median - x_{Me})	
4.5.1.4. Số trung bình nhân (trung bình hình học) (\bar{x}_g).....	
4.5.2. Các tham số đại diện cho sự phân tán của mẫu	
4.5.2.1. Phương sai mẫu	
4.5.2.2. Độ lệch tiêu chuẩn mẫu (s).....	
4.5.2.3. Độ lệch tiêu chuẩn của số bình quân (s_x^-).....	
4.5.2.4. Biên độ dao động của dãy số liệu (Range).....	
4.5.2.5. Hệ số biến động (CV%)	
4.6. Các tham số đặc trưng của đại lượng định tính.....	
4.7. Một số quy tắc về làm tròn số trong tính toán	
4.7.1. Con số có nghĩa.....	
4.7.2. Cách làm tròn số (quy tắc xấp xỉ)	

CHƯƠNG V - ƯỚC LƯỢNG

5.1. Đặt vấn đề
5.2. Các phương pháp ước lượng
5.2.1. Ước lượng điểm
5.2.2. Ước lượng khoảng
5.3. Ước lượng giá trị trung bình của tổng thể

5.3.1. Ước lượng trị số trung bình tổng thể khi dung lượng mẫu n là đủ lớn ($n \geq 30$).....	
5.3.2. Ước lượng số trung bình quần thể khi $n < 30$	
5.4. Xác định dung lượng mẫu.....	
5.5. Ước lượng tần số của tổng thể (hay ước lượng tỷ lệ).....	
5.5.1. Khi sự kiện A có xác suất không gần 0 và 1	
5.5.1.1. Khi dung lượng n đủ lớn ($n > 100$).....	
5.5.1.2. Khi dung lượng $n < 100$ (không đủ lớn).....	
5.5.2. Khi sự kiện A có xác suất gần 0 hoặc gần 1	

CHƯƠNG VI - KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

6.1. Những khái niệm chung và ý nghĩa.....	
6.2. Trường hợp hai mẫu độc lập.....	
6.2.1. Tiêu chuẩn U của phân bố tiêu chuẩn	
6.2.2. Tiêu chuẩn t của phân bố Student	
6.3. Trường hợp hai mẫu liên hệ (không độc lập)	
6.3.1. Khái niệm về mẫu liên hệ.....	
6.3.2. Tiêu chuẩn t của Student	
6.4. Kiểm định tính độc lập (kiểm định tính thuần nhất của các mẫu về chất)	
6.4.1. So sánh 2 tỷ lệ (hay gọi là kiểm định 2 xác suất)	
6.4.1.1. Trường hợp 2 dung lượng mẫu đều đủ lớn	
6.4.1.2. Trường hợp $n < 100$ (n_1 hoặc $n_2 < 100$).....	
6.4.1.3. Trường hợp hai sự kiện hiếm (hay xác suất bé)	
6.4.2. Kiểm tra tính thuần nhất của nhiều mẫu về chất (kiểm định nhiều xác suất)	
6.5. Kiểm tra giả thuyết về hai phương sai.....	

CHƯƠNG VII - PHƯƠNG PHÁP SẮP XẾP CÔNG THỨC THÍ NGHIỆM VÀ PHÂN TÍCH KẾT QUẢ.

79

Mục tiêu.....	
Nội dung của chương:	
7.1. Các thí nghiệm một nhân tố.	
7.1.1. Khái niệm:	
7.1.2. Các phương pháp sắp xếp và phân tích kết quả.....	
7.1.2.1. Thí nghiệm sắp xếp kiểu hoàn toàn ngẫu nhiên (CRD).....	
7.1.2.2. Thí nghiệm sắp xếp theo kiểu khối ngẫu nhiên đầy đủ (RCB).	
7.1.2.3. Thí nghiệm sắp xếp kiểu ô vòng la tin (Latin Square).....	

7. 2. Thí nghiệm hai nhân tố	
7.2.1. Thiết kế theo kiểu khối ngẫu nhiên đầy đủ (RCB)	
7.2.2. Thiết kế kiểu chia ô lớn ô nhỏ (Split-plot)	
7.3. Một số điều chú ý trước khi tiến hành phân tích phương sai.....	
7.3.1. Khi số liệu thí nghiệm bị mất	
7.3.1.1 Khi bị mất một số liệu.....	
7.3.1.2 Khi bị mất 2 ô số liệu.....	
7.3.2. Chuyển đổi số liệu trước khi phân tích phương sai	
Các câu hỏi ôn tập:.....	
CHƯƠNG VIII - PHÂN TÍCH TƯƠNG QUAN	
Mục tiêu.....	
Nội dung bao gồm:.....	
8.1. Đặt vấn đề	
8.2. Tương quan đường thẳng một biến số.....	
8.2.1. Khái niệm và các đặc trưng của tương quan	
8.2.2. Các ví dụ.....	
8.2.2.1. Trường hợp dung lượng mẫu nhỏ ($n < 30$)	
8.2.2.2. Trường hợp dung lượng mẫu lớn ($n > 30$).....	
8.3. Tương quan tuyến tính nhiều biến.....	
8.3.1. Ý nghĩa của nghiên cứu tương quan nhiều yếu tố	
8.3.2. Hệ số hồi quy riêng và phương trình hồi quy phức.	
8.4. Phương pháp lập phương trình tương quan cho các đặc trưng chất lượng.....	
CHƯƠNG IX - TỔNG KẾT THÍ NGHIỆM	149
9.1. Cách trình bày số liệu trong báo cáo khoa học	149
9.1.1. Phương pháp trình bày bằng bảng số liệu	149
9.1.2. Phương pháp dùng đồ thị và biểu đồ.....	150
9.2. Phương pháp tổng kết và viết báo cáo kết quả thí nghiệm (kết quả nghiên cứu khoa học).....	152
9.2.1. Chinh lý số liệu, tính các tham số thống kê và xử lý thống kê kết quả thí nghiệm	153
9.2.2. Viết báo cáo khoa học	153
PHỤ LỤC	155