



**BÀI GIẢNG PHƯƠNG PHÁP  
NGHIÊN CỨU KHOA HỌC**

# PHẦN I

## TỔNG QUAN VỀ PHƯƠNG PHÁP NGHIÊN CỨU KHOA HỌC

### Chương 1

#### KHÁI NIỆM VỀ NGHIÊN CỨU KHOA HỌC VÀ PHƯƠNG PHÁP KHOA HỌC

##### I. Nghiên cứu khoa học

Nghiên cứu khoa học là tìm hiểu, xem xét, điều tra có phương pháp khoa học để từ những dữ liệu đã có muốn đạt đến một kết quả nghiên cứu mới hơn, cao hơn, giá trị hơn. Mục đích của nghiên cứu khoa học xét về thực chất là nhận thức thế giới và cải tạo thế giới.

##### II. Đề tài nghiên cứu khoa học

###### 1. Khái niệm đề tài

Đề tài là một hình thức tổ chức NCKH do một người hoặc một nhóm người thực hiện. Một số hình thức tổ chức nghiên cứu khác không hoàn toàn mang tính chất nghiên cứu khoa học, chẳng hạn như: Chương trình, dự án, đề án. Sự khác biệt giữa các hình thức NCKH này như sau:

*Đề tài:* được thực hiện để trả lời những câu hỏi mang tính học thuật, có thể chưa đề ý đến việc ứng dụng trong hoạt động thực tế.

*Dự án:* được thực hiện nhằm vào mục đích ứng dụng, có xác định cụ thể hiệu quả về kinh tế và xã hội. Dự án có tính ứng dụng cao, có ràng buộc thời gian và nguồn lực.

*Đề án:* là loại văn kiện, được xây dựng để trình cấp quản lý cao hơn, hoặc gửi cho một cơ quan tài trợ để xin thực hiện một công việc nào đó như: thành lập một tổ chức; tài trợ cho một hoạt động xã hội, ... Sau khi đề án được phê chuẩn, sẽ hình thành những dự án, chương trình, đề tài theo yêu cầu của đề án.

*Chương trình:* là một nhóm đề tài hoặc dự án được tập hợp theo một mục đích xác định. Giữa chúng có tính độc lập tương đối cao. Tiến độ thực hiện đề tài, dự án trong chương trình không nhất thiết phải giống nhau, nhưng nội dung của chương trình thì phải đồng bộ.

###### 2. Đối tượng nghiên cứu và phạm vi nghiên cứu

*Đối tượng nghiên cứu:* là bản chất của sự vật hay hiện tượng cần xem xét và làm rõ trong nhiệm vụ nghiên cứu.

*Phạm vi nghiên cứu:* đối tượng nghiên cứu được khảo sát trong phạm vi nhất định về mặt thời gian, không gian và lĩnh vực nghiên cứu.

###### 3. Mục đích và mục tiêu nghiên cứu

Khi viết đề cương nghiên cứu, một điều rất quan trọng là làm sao thể hiện được mục tiêu và mục đích nghiên cứu mà không có sự trùng lặp lẫn nhau. Vì vậy, cần thiết để phân biệt sự khác nhau giữa mục đích và mục tiêu.

*Mục đích:* là hướng đến một điều gì hay một công việc nào đó trong nghiên cứu mà người nghiên cứu mong muốn đề hoàn thành, nhưng thường thì mục đích khó có thể đo lường hay định lượng. Nói cách khác, mục đích là sự sắp đặt công việc hay điều gì đó được đưa ra trong nghiên cứu. Mục đích trả lời câu hỏi “nhằm vào việc gì?”, hoặc “để phục vụ cho điều gì?” và mang ý nghĩa thực tiễn của nghiên cứu, nhắm đến đối tượng phục vụ sản xuất, nghiên cứu.

*Mục tiêu:* là thực hiện điều gì hoặc hoạt động nào đó cụ thể, rõ ràng mà người nghiên cứu sẽ hoàn thành theo kế hoạch đã đặt ra trong nghiên cứu. Mục tiêu có thể đo lường hay định lượng được. Nói cách khác, mục tiêu là nền tảng hoạt động của đề tài và làm cơ sở cho việc đánh giá kế hoạch nghiên cứu đã đưa ra, và là điều mà kết quả phải đạt được. Mục tiêu trả lời câu hỏi “làm cái gì?”.

### **III. Cấu trúc của phương pháp luận nghiên cứu khoa học**

Nghiên cứu khoa học phải sử dụng PPKH: bao gồm chọn phương pháp thích hợp (luận chứng) để chứng minh mối quan hệ giữa các luận cứ và giữa toàn bộ luận cứ với luận đề; cách đặt giả thuyết hay phán đoán sử dụng các luận cứ và phương pháp thu thập thông tin và xử lý thông tin (luận cứ) để xây dựng luận đề.

#### **1. Luận đề**

Luận đề trả lời câu hỏi “cần chứng minh điều gì?” trong nghiên cứu. Luận đề là một “phán đoán” hay một “giả thuyết” cần được chứng minh.

#### **2. Luận cứ**

Để chứng minh một luận đề thì nhà khoa học cần đưa ra các bằng chứng hay luận cứ khoa học. Luận cứ bao gồm thu thập các thông tin, tài liệu tham khảo; quan sát và thực nghiệm. Luận cứ trả lời câu hỏi “Chứng minh bằng cái gì?”. Các nhà khoa học sử dụng luận cứ làm cơ sở để chứng minh một luận đề. Có hai loại luận cứ được sử dụng trong nghiên cứu khoa học:

*Luận cứ lý thuyết:* bao gồm các lý thuyết, luận điểm, tiền đề, định lý, định luật, qui luật đã được khoa học chứng minh và xác nhận là đúng. Luận cứ lý thuyết cũng được xem là cơ sở lý luận.

*Luận cứ thực tiễn:* dựa trên cơ sở số liệu thu thập, quan sát và làm thí nghiệm.

#### **3. Luận chứng**

Để chứng minh một luận đề, nhà nghiên cứu khoa học phải đưa ra phương pháp để xác định mối liên hệ giữa các luận cứ và giữa luận cứ với luận đề. Luận chứng trả lời câu hỏi “Chứng minh bằng cách nào?”. Trong nghiên cứu khoa học, để chứng minh một luận đề, một giả thuyết hay sự tiên đoán thì nhà nghiên cứu sử dụng luận chứng, chẳng hạn kết hợp các phép suy luận, giữa suy luận suy diễn, suy luận qui nạp và loại suy. Một cách sử dụng luận chứng khác, đó là phương pháp tiếp cận và thu thập thông tin làm luận cứ khoa học, thu thập số liệu thống kê trong thực nghiệm hay trong các loại nghiên cứu điều tra.

### **IV. Phương pháp khoa học**

Là hệ thống cách thức, quy tắc được đúc kết lại nhằm chỉ dẫn cho ta đạt được mục đích một cách tốt nhất với sự tổn kém (sức lực, thời gian, tiền bạc...) ít nhất.

Có ba phương pháp chung trong nghiên cứu khoa học. Đó là: phương pháp nghiên cứu lý thuyết, phương pháp nghiên cứu thực nghiệm và phương pháp nghiên cứu phi thực nghiệm.

*Phương pháp nghiên cứu lý thuyết* được sử dụng trong cả khoa học tự nhiên, khoa học xã hội và các khoa học khác. Đây chính là phương pháp nghiên cứu tại bàn giấy mà chất liệu cho nghiên cứu chỉ gồm những khái niệm, quy luật, tư liệu, số liệu... đã có sẵn trước đó. Nghiên cứu lý thuyết là thuần túy dựa trên khái niệm, phán đoán và suy luận để đưa ra những giải pháp cho vấn đề.

*Phương pháp nghiên cứu thực nghiệm* được thực hiện bởi những quan sát sự vật hoặc hiện tượng diễn ra trong thực tế và trong điều kiện có sự tác động theo chủ định của người nghiên cứu. Nói một cách khác nghiên cứu thực nghiệm là quan sát tại hiện trường hoặc trên mô hình do người nghiên cứu tạo ra với những tham số đã được khống chế trước. Phương pháp này được sử dụng nhiều trong khoa học tự nhiên như vật lý, hoá học, nông nghiệp, tiến hành bố trí thí nghiệm để thu thập số liệu, để giải thích và kết luận.

*Phương pháp nghiên cứu phi thực nghiệm* cũng dựa vào những quan sát các sự vật hoặc hiện tượng đang diễn ra, nhưng không có bất cứ sự can thiệp hay tác động nào gây biến đổi trạng thái thực sự của đối tượng nghiên cứu. Đây là phương pháp được áp dụng trong cuộc phỏng vấn, hội thảo, điều tra bằng bản câu hỏi.

Các bước cơ bản trong phương pháp khoa học

Quan sát sự vật, hiện tượng và xác định vấn đề nghiên cứu

Thiết lập giả thuyết hay sự tiên đoán

Thu thập thông tin, số liệu thí nghiệm

Xử lý, phân tích dữ liệu

Kết luận xác nhận hay phủ nhận giả thuyết.

## Chương 2

# “VẤN ĐỀ” NGHIÊN CỨU KHOA HỌC

### I. Bản chất của quan sát

Trước đây, con người dựa vào niềm tin để giải thích những gì thấy được xảy ra trong thế giới xung quanh mà không có kiểm chứng hay thực nghiệm để chứng minh tính vững chắc của những quan niệm, tư tưởng, học thuyết mà họ đưa ra. Ngoài ra, con người cũng không sử dụng phương pháp khoa học để có câu trả lời cho câu hỏi.

Ngày nay, các nhà khoa học không ngừng quan sát, theo dõi sự vật, hiện tượng, quy luật của sự vận động, mối quan hệ, ... trong thế giới xung quanh và dựa vào kiến thức, kinh nghiệm hay các nghiên cứu có trước để khám phá, tìm ra kiến thức mới, giải thích các quy luật vận động, mối quan hệ giữa các sự vật một cách khoa học. Quan sát để cảm nhận sự kiện (tự xảy ra hoặc do chủ động bố trí) là bước đầu tiên để nhận ra vấn đề cần giải quyết.

### II. “Vấn đề” nghiên cứu khoa học

#### 1. Đặt câu hỏi

Bản chất của quan sát thường đặt ra những câu hỏi, từ đó đặt ra “vấn đề” nghiên cứu cho nhà khoa học và người nghiên cứu. Câu hỏi đặt ra phải đơn giản, cụ thể, rõ ràng (xác định giới hạn, phạm vi nghiên cứu) và làm sao có thể thực hiện thí nghiệm để kiểm chứng, trả lời. Cách đặt câu hỏi thường bắt đầu như sau: Làm thế nào, bao nhiêu, xảy ra ở đâu, nơi nào, khi nào, ai, tại sao, cái gì...? Đặt câu hỏi hay đặt “vấn đề” nghiên cứu là cơ sở giúp nhà khoa học chọn chủ đề nghiên cứu thích hợp. Sau khi chọn chủ đề nghiên cứu, một công việc rất quan trọng trong phương pháp nghiên cứu là thu thập tài liệu tham khảo (tùy theo loại nghiên cứu mà có phương pháp thu thập thông tin khác nhau).

#### 2. Phân loại “vấn đề” nghiên cứu khoa học

Sau khi đặt câu hỏi và “vấn đề” nghiên cứu khoa học đã được xác định, công việc tiếp theo cần biết là “vấn đề” đó thuộc loại câu hỏi nào. Nhìn chung, “vấn đề” được thể hiện trong 3 loại câu hỏi như sau:

Câu hỏi thuộc loại thực nghiệm.

Câu hỏi thuộc loại quan niệm hay nhận thức.

Câu hỏi thuộc loại đánh giá.

#### **Câu hỏi thuộc loại thực nghiệm**

Câu hỏi thuộc loại thực nghiệm là những câu hỏi có liên quan tới các sự kiện đã xảy ra hoặc các quá trình có mối quan hệ nhân quả về thế giới của chúng ta. Để trả lời câu hỏi loại này, chúng ta cần phải tiến hành quan sát hoặc làm thí nghiệm. Tất cả các kết luận phải dựa trên độ tin cậy của số liệu thu thập trong quan sát và thí nghiệm.

#### **Câu hỏi thuộc loại quan niệm hay nhận thức**

Loại câu hỏi này có thể được trả lời bằng những nhận thức một cách logic, hoặc chỉ là những suy nghĩ đơn giản cũng đủ để trả lời mà không cần tiến hành thực nghiệm hay quan sát. Suy nghĩ đơn giản ở đây được hiểu là có sự phân tích nhận thức và lý lẽ hay lý do, nghĩa là sử dụng các nguyên tắc, qui luật, pháp lý trong xã hội và những cơ sở khoa học có trước. Cần chú ý sử dụng các qui luật, luật lệ trong xã hội đã được áp dụng một cách ổn định và phù hợp với “vấn đề” nghiên cứu.

#### **Câu hỏi thuộc loại đánh giá**

Câu hỏi thuộc loại đánh giá là câu hỏi thể hiện giá trị và tiêu chuẩn. Câu hỏi này có liên quan tới việc đánh giá các giá trị về đạo đức hoặc giá trị thẩm mỹ. Để trả lời các câu hỏi loại này, cần hiểu biết nét đặc trưng giữa giá trị thực chất và giá trị sử dụng. Giá trị thực chất là giá trị hiện hữu

riêng của sự vật mà không lệ thuộc vào cách sử dụng. Giá trị sử dụng là sự vật chỉ có giá trị khi nó đáp ứng được nhu cầu sử dụng và nó bị đánh giá không còn giá trị khi nó không còn đáp ứng được nhu cầu sử dụng nữa.

### **3. Cách phát hiện “vấn đề” nghiên cứu khoa học**

Các “vấn đề” nghiên cứu khoa học thường được hình thành trong các tình huống sau:

Quá trình nghiên cứu, đọc và thu thập tài liệu nghiên cứu giúp cho nhà khoa học phát hiện hoặc nhận ra các “vấn đề” và đặt ra nhiều câu hỏi cần nghiên cứu (phát triển “vấn đề” rộng hơn để nghiên cứu). Đôi khi người nghiên cứu thấy một điều gì đó chưa rõ trong những nghiên cứu trước và muốn chứng minh lại. Đây là tình huống quan trọng nhất để xác định “vấn đề” nghiên cứu.

Trong các hội nghị chuyên đề, báo cáo khoa học, kỹ thuật, ... đôi khi có những bất đồng, tranh cãi và tranh luận khoa học đã giúp cho các nhà khoa học nhận thấy được những mặt yếu, mặt hạn chế của “vấn đề” tranh cãi và từ đó người nghiên cứu nhận định, phân tích lại và chọn lọc rút ra “vấn đề” cần nghiên cứu.

Trong mối quan hệ giữa con người với con người, con người với tự nhiên, qua hoạt động thực tế lao động sản xuất, yêu cầu kỹ thuật, mối quan hệ trong xã hội, cư xử... làm cho con người không ngừng tìm tòi, sáng tạo ra những sản phẩm tốt hơn nhằm phục vụ cho nhu cầu đời sống con người trong xã hội. Những hoạt động thực tế này đã đặt ra cho người nghiên cứu các câu hỏi hay người nghiên cứu phát hiện ra các “vấn đề” cần nghiên cứu.

“Vấn đề” nghiên cứu cũng được hình thành qua những thông tin bức xúc, lời nói phàn nàn nghe được qua các cuộc nói chuyện từ những người xung quanh mà chưa giải thích, giải quyết được “vấn đề” nào đó.

Các “vấn đề” hay các câu hỏi nghiên cứu chợt xuất hiện trong suy nghĩ của các nhà khoa học, các nhà nghiên cứu qua tình cờ quan sát các hiện tượng của tự nhiên, các hoạt động xảy ra trong xã hội hàng ngày.

Tính tò mò của nhà khoa học về điều gì đó cũng đặt ra các câu hỏi hay “vấn đề” nghiên cứu.

## **Chương 3**

# **THU THẬP TÀI LIỆU VÀ ĐẶT GIẢ THUYẾT**

### **I. Tài liệu**

#### **1. Mục đích thu thập tài liệu**

Thu thập và nghiên cứu tài liệu là một công việc quan trọng cần thiết cho bất kỳ hoạt động nghiên cứu khoa học nào. Các nhà nghiên cứu khoa học luôn đọc và tra cứu tài liệu có trước để làm nền tảng cho NCKH. Đây là nguồn kiến thức quý giá được tích lũy qua quá trình nghiên cứu mang tính lịch sử lâu dài. Vì vậy, mục đích của việc thu thập và nghiên cứu tài liệu nhằm:

Giúp cho người nghiên cứu nắm được phương pháp của các nghiên cứu đã thực hiện trước đây. Tránh trùng lặp với các nghiên cứu trước đây.

Giúp người nghiên cứu có phương pháp luận hay luận cứ chặt chẽ hơn.

Có thêm kiến thức rộng, sâu về lĩnh vực đang nghiên cứu.

Xem xét tính khả thi để từ đó hình thành hướng nghiên cứu thích hợp

#### **2. Phân loại tài liệu nghiên cứu**

Phân loại tài liệu để giúp cho người nghiên cứu chọn lọc, đánh giá và sử dụng tài liệu đúng với lãnh vực chuyên môn hay đối tượng muốn nghiên cứu. Có thể chia ra 2 loại tài liệu: tài sơ cấp (hay tài liệu liệu gốc) và tài liệu thứ cấp.

##### **2.1. Tài liệu sơ cấp**

Tài liệu sơ cấp là tài liệu mà người nghiên cứu tự thu thập, phỏng vấn trực tiếp, hoặc nguồn tài liệu cơ bản, còn ít hoặc chưa được chú giải. Một số vấn đề nghiên cứu có rất ít tài liệu, vì vậy cần phải điều tra để tìm và khám phá ra các nguồn tài liệu chưa được biết. Người nghiên cứu cần phải tổ chức, thiết lập phương pháp để ghi chép, thu thập số liệu.

##### **2.2. Tài liệu thứ cấp**

Loại tài liệu này có nguồn gốc từ tài liệu sơ cấp đã được phân tích, giải thích và thảo luận, diễn giải. Các nguồn tài liệu thứ cấp như: Sách giáo khoa, báo chí, bài báo, tập san chuyên đề, tạp chí, biên bản hội nghị, báo cáo khoa học, internet, sách tham khảo, luận văn, luận án, thông tin thống kê, hình ảnh, video, băng cassette, tài liệu-văn thư, bản thảo viết tay...

#### **3. Nguồn thu thập tài liệu**

Thông tin thu thập để làm nghiên cứu được tìm thấy từ các nguồn tài liệu sau:

Luận cứ khoa học, định lý, quy luật, định luật, khái niệm... có thể thu thập được từ sách giáo khoa, tài liệu chuyên ngành...

Các số liệu, tài liệu đã công bố được tham khảo từ các bài báo trong tạp chí khoa học, tập san, báo cáo chuyên đề khoa học...

Số liệu thống kê được thu thập từ các Niên Giám Thống Kê: Chi cục thống kê, Tổng cục thống kê... Tài liệu lưu trữ, văn kiện, hồ sơ, văn bản về luật, chính sách... thu thập từ các cơ quan quản lý Nhà nước, tổ chức chính trị - xã hội.

Thông tin trên truyền hình, truyền thanh, báo chí... mang tính đại chúng cũng được thu thập và được xử lý để làm luận cứ khoa học chứng minh cho vấn đề khoa học.

### **II. Giả thuyết**

## 1. Khái niệm “giả thuyết nghiên cứu”

Giả thuyết”, hoặc “giả thuyết khoa học”, hoặc đơn giản hơn, “giả thuyết nghiên cứu” (Hypothese) là gì? “Giả thuyết là nhận định sơ bộ, là kết luận giả định của nghiên cứu”, hoặc “Giả thuyết là luận điểm cần chứng minh của tác giả”, hoặc “Giả thuyết là câu trả lời sơ bộ, cần chứng minh, vào câu hỏi nghiên cứu của đề tài”.

## 2. Mối quan hệ giữa giả thuyết và “vấn đề” khoa học

Sau khi xác định câu hỏi hay “vấn đề” nghiên cứu khoa học, người nghiên cứu hình thành ý tưởng khoa học, tìm ra câu trả lời hoặc sự giải thích tới vấn đề chưa biết (đặt giả thuyết). Ý tưởng khoa học này còn gọi là sự tiên đoán khoa học hay giả thuyết giúp cho người nghiên cứu có động cơ, hướng đi đúng hay tiếp cận tới mục tiêu cần nghiên cứu. Trên cơ sở những quan sát bước đầu, những tình huống đặt ra (câu hỏi hay vấn đề), những cơ sở lý thuyết (tham khảo tài liệu, kiến thức đã có,...), sự tiên đoán và những dự kiến tiến hành thực nghiệm sẽ giúp cho người nghiên cứu hình thành một cơ sở lý luận khoa học để xây dựng giả thuyết khoa học.

## 3. Cấu trúc của một “giả thuyết”

### Cấu trúc có mối quan hệ “nhân-quả”

Mối quan hệ trong giả thuyết là nguyên nhân này có thể ảnh hưởng đến kết quả nghiên cứu.

### Cấu trúc “Nếu - thì”

“Nếu” (hệ quả hoặc nguyên nhân)... có liên quan tới (nguyên nhân hoặc hệ quả)..., “thì” nguyên nhân đó có thể hay ảnh hưởng đến hệ quả.

Một số nhà khoa học đặt cấu trúc này như là sự tiên đoán và dựa trên đó để xây dựng thí nghiệm kiểm chứng giả thuyết.

## 4. Cách đặt giả thuyết

Điều quan trọng trong cách đặt giả thuyết là phải đặt như thế nào để có thể thực hiện thí nghiệm kiểm chứng “đúng” hay “sai” giả thuyết đó. Vì vậy, trong việc xây dựng một giả thuyết cần trả lời các câu hỏi sau:

Giả thuyết này có thể tiến hành thực nghiệm được không?

Các biến hay các yếu tố nào cần được nghiên cứu?

Phương pháp thí nghiệm nào (trong phòng, khảo sát, điều tra, bảng câu hỏi, phỏng vấn...) được sử dụng trong nghiên cứu?

Các chỉ tiêu nào cần đo đạt trong suốt thí nghiệm?

Phương pháp xử lý số liệu nào mà người nghiên cứu dùng để bác bỏ hay chấp nhận giả thuyết?

Tóm lại, giả thuyết đặt ra dựa trên sự quan sát, kiến thức vốn có, các nguyên lý, kinh nghiệm trước đây hoặc dựa vào nguồn tài liệu tham khảo, kết quả nghiên cứu tương tự trước đây để phát triển nguyên lý chung hay bằng chứng để giải thích, chứng minh câu hỏi nghiên cứu. Xét về bản chất logic, giả thuyết được đặt ra từ việc xem xét bản chất riêng, chung của sự vật và mối quan hệ của chúng hay gọi là quá trình suy luận. Quá trình suy luận là cơ sở hình thành giả thuyết khoa học.

## 5. Chứng minh giả thuyết khoa học

Chứng minh giả thuyết khoa học là quá trình quan sát, quá trình là 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 kết luận giả thuyết (một phần giả thuyết) “sai” (nghĩa là bác bỏ giả thuyết hay chứng minh giả thuyết sai) hoặc kết luận giả thuyết “đúng”.



Thường thì các nhà khoa học vận dụng kiến thức để tiên đoán mối quan hệ giữa biến độc lập và biến phụ thuộc.

Chứng minh giả thuyết khoa học có hai cách, đó là: Quan sát hay điều tra và làm thí nghiệm thực nghiệm.

### **5.1. Quan sát hay điều tra**

Là việc tìm hiểu theo dõi thực tế, giúp ta phân biệt được 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 từ đó suy ra bản chất của chúng dựa trên nhận thức của người nghiên cứu. Trên cơ sở đó phân tích, đánh giá để tổng hợp lại thành nhận thức hiểu biết của con người về sự việc hay hiện tượng đó. Như vậy, quan sát là đi từ bên ngoài sự việc vào trong nhận thức. 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.

### **5.2. Thí nghiệm**

Là những công việc mà người nghiên cứu tự xây dựng để quan sát các chỉ tiêu trên đối tượng thí nghiệm nhằm kiểm chứng giả thuyết.

Một số vấn đề liên quan đến thí nghiệm

#### **5.2.1. Các biến trong thí nghiệm**

Trong nghiên cứu thực nghiệm, có 2 loại biến thường gặp trong thí nghiệm, đó là biến độc lập (independent variable) và biến phụ thuộc (dependent variable).

**Biến độc lập** là các yếu tố, điều kiện khi bị thay đổi trên đối tượng nghiên cứu sẽ ảnh hưởng đến kết quả thí nghiệm. Như vậy, đối tượng nghiên cứu chứa một hoặc nhiều yếu tố, điều kiện thay đổi. Nói cách khác kết quả số liệu của biến phụ thuộc thu thập được thay đổi theo biến độc lập.

Trong biến độc lập, thường có một mức độ đối chứng hay nghiệm thức đối chứng (chứa các yếu tố, điều kiện ở mức độ thông thường) hoặc nghiệm thức đã được xác định mà người nghiên cứu không cần tiên đoán ảnh hưởng của chúng. Các nghiệm thức còn lại sẽ được so sánh với nghiệm thức đối chứng hoặc so sánh giữa các cặp nghiệm thức với nhau.

**Biến phụ thuộc** (còn gọi là chỉ tiêu thu thập) là những chỉ tiêu đo đạc và bị ảnh hưởng trong suốt quá trình thí nghiệm, hay có thể nói kết quả đo đạc phụ thuộc vào sự thay đổi của biến độc lập.

#### **5.2.2. Các loại công thức trong thí nghiệm**

**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 nghiên cứu.

Trong 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.

**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 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.

## **Chương 4**

### **CÁCH TRÌNH BÀY ĐỀ TÀI NGHIÊN CỨU**

#### **I. Các phần chính trong một luận văn/chuyên đề tốt nghiệp**

##### **Phần 1 MỞ ĐẦU**

Nêu lên tính cấp thiết của đề tài, mục đích và mục tiêu của đề tài.

##### **Phần 2 TỔNG QUAN VỀ VẤN ĐỀ NGHIÊN CỨU**

Phần này viết sơ lược những cơ sở lý luận liên quan đến đề tài nghiên cứu. Tóm tắt ngắn gọn về phương pháp và kết quả đạt được cũng như vấn đề còn hạn chế của các nhà nghiên cứu khác liên quan đến vấn đề nghiên cứu.

##### **Phần 3 PHƯƠNG PHÁP NGHIÊN CỨU**

Phần này nêu lên phương pháp cụ thể để thực hiện đề tài.

##### **Phần 4 KẾT QUẢ VÀ THẢO LUẬN**

Trình bày kết quả nghiên cứu và thảo luận với các nghiên cứu khác.

##### **Phần 5 KẾT LUẬN VÀ KIẾN NGHỊ**

Kết luận phải khẳng định được những kết quả đạt được, những đóng góp mới. Kết luận cần ngắn gọn, không có lời bàn và bình luận thêm. Chỉ kết luận những vấn đề đã thực hiện.

Phần đề nghị phải xuất phát từ nội dung nghiên cứu. Đề nghị phải cụ thể, rõ ràng, thiết thực và có thể áp dụng được.

#### **TÀI LIỆU THAM KHẢO**

##### **PHỤ LỤC**

Mục đích của phụ lục là trữ thông tin và liệt kê những bảng số liệu liên quan để người đọc quan tâm có thể kiểm tra và tra cứu. Có thể phụ lục chứa các số liệu tính toán thống kê (bảng ANOVA, bảng hồi quy) hoặc mô tả các phương pháp phân tích, phương pháp thực hiện còn tương đối mới mà người đọc chưa hoàn toàn quen thuộc.

Nếu tác giả sử dụng bảng điều tra, bảng điều tra phải được trình bày trong phụ lục theo đúng hình thức đã được sử dụng, không nên kết cấu hay hiệu đính lại. Các tính toán từ mẫu điều tra được trình bày tóm tắt trong các bảng biểu của bài viết và có thể trình bày trong phần Phụ lục.

#### **II. Cách trình bày kết quả số liệu nghiên cứu**

##### **1. Trình bày dạng văn viết**

Đối với những số liệu đơn giản, tốt nhất nên trình bày, giải thích ở dạng câu văn viết và các số liệu được cho vào trong ngoặc đơn không nên trình bày ở dạng bảng và hình.

## 2. Trình bày bảng

### 2.1. Cấu trúc bảng số liệu

Cấu trúc bảng chứa các thành phần sau đây:

- Số và tựa bảng
- Tựa cột
- Tựa hàng
- Phần thân chính của bảng là vùng chứa số liệu
- Chú thích cuối bảng
- Các đường ranh giới giữa các phần.

**Bảng 2.1:** Thời gian trung bày của hoa hồng bảo quản lạnh và xử lý sau khi bảo quản trong phòng lạnh

Thời gian	Có xử lý		Không xử lý		Đối chứng
	Bảo quản ướt	Bảo quản khô	Bảo quản ướt	Bảo quản khô	
Sau 7 ngày	17 <sup>ac</sup> ± 0	15 <sup>cg</sup> ± 0	12 <sup>ac</sup> + 0	11 <sup>df</sup> ± 0	12 ± 0
Sau 14 ngày	15,7 <sup>ab</sup> ± 0,3	8,7 <sup>ce</sup> ± 0,3	11 <sup>ab</sup> + 0	6,7 <sup>d</sup> ± 0,3	11 ± 0
Sau 21 ngày	13,3 <sup>bc</sup> ± 0,3	3 <sup>cg</sup> ± 1,4	8,7 <sup>bc</sup> + 0	2 <sup>df</sup> ± 1,2	11 ± 0

(Giá trị trung bình của 3 lần lặp lại, mean ± SE, giá trị trung bình với các ký hiệu giống nhau trong cùng một cột chỉ sự khác biệt ở mức ý nghĩa 5%)

### 3. Trình bày biểu đồ, hình

Sử dụng hình nhằm minh họa các kết quả và mối quan hệ giữa các biến cho đọc giả dễ thấy hơn khi trình bày bằng bảng số liệu hoặc văn bản. Sử dụng hình có thuận lợi là đọc giả hiểu nhanh chóng các số liệu mà không mất nhiều thời gian khi nhìn bảng. Các dạng hình được sử dụng gồm biểu đồ cột (colume chart), biểu đồ thanh (bar chart), biểu đồ tần suất (frequency histogram), biểu đồ phân tán (scatterplot), biểu đồ đường biểu diễn (line chart), biểu đồ hình bánh (pie chart), biểu đồ diện tích (area chart), sơ đồ chuỗi (flow chart), sơ đồ phân cấp tổ chức (organization chart), hình ảnh (photos) ...

#### Biểu đồ sử dụng cho số liệu phân tích thống kê

Sử dụng số liệu ở Bảng 2.1

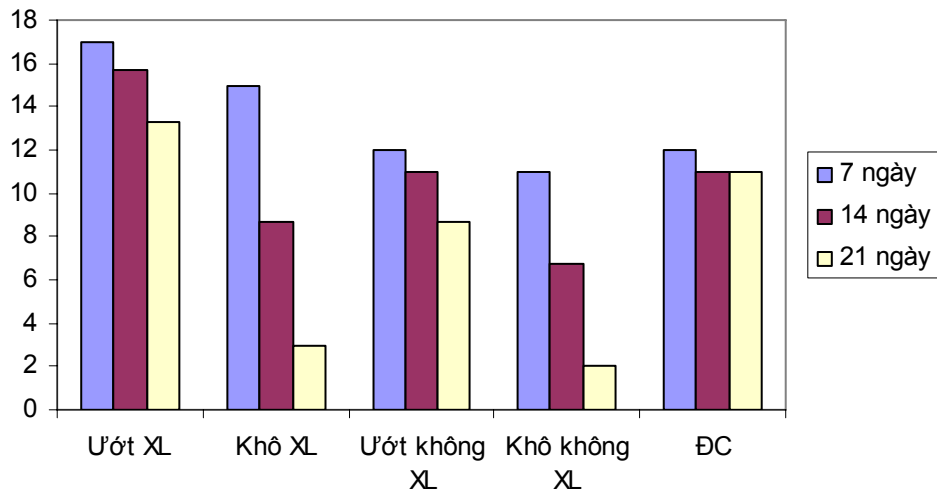
**Bước 1:** Sắp xếp lại số liệu

	A	B	C	D	E	F	G	H
1		<b>7 ngày</b>	<b>SE</b>	<b>14 ngày</b>	<b>SE</b>	<b>21 ngày</b>	<b>SE</b>	
2	<b>Ướt XL</b>	17	0	15,7	0,3	13,3	0,3	
3	<b>Khô XL</b>	15	0	8,7	0,3	3	1,4	
4	<b>Ướt không XL</b>	12	0	11	0	8,7	0	
5	<b>Khô không XL</b>	11	0	6,7	0,3	2	1,2	
6	<b>ĐC</b>	12	0	11	0	11	0	
7								

**Bước 2:** Vẽ đồ thị cột

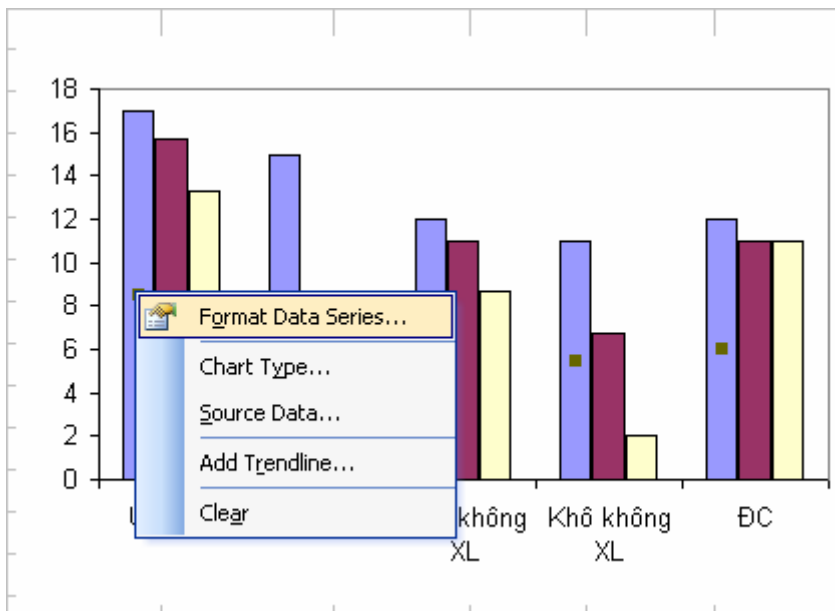
Chọn các cột dữ liệu để vẽ đồ thị cột. Chọn A1:A6 và B1:B6 và D1:D6 và F1:F6

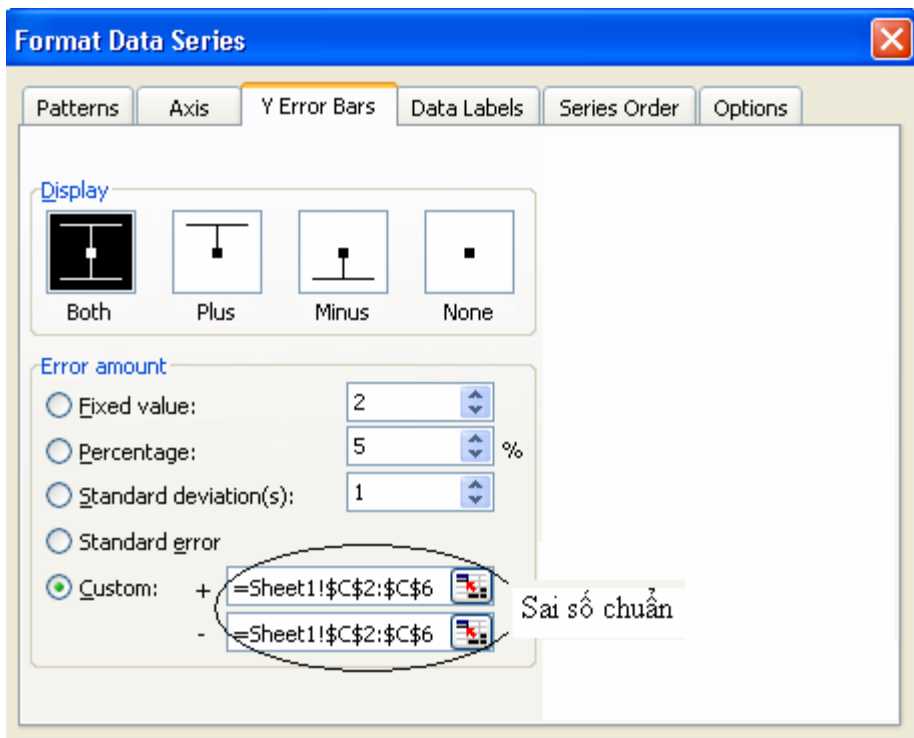
Vào **Insert/Chart.../Column**



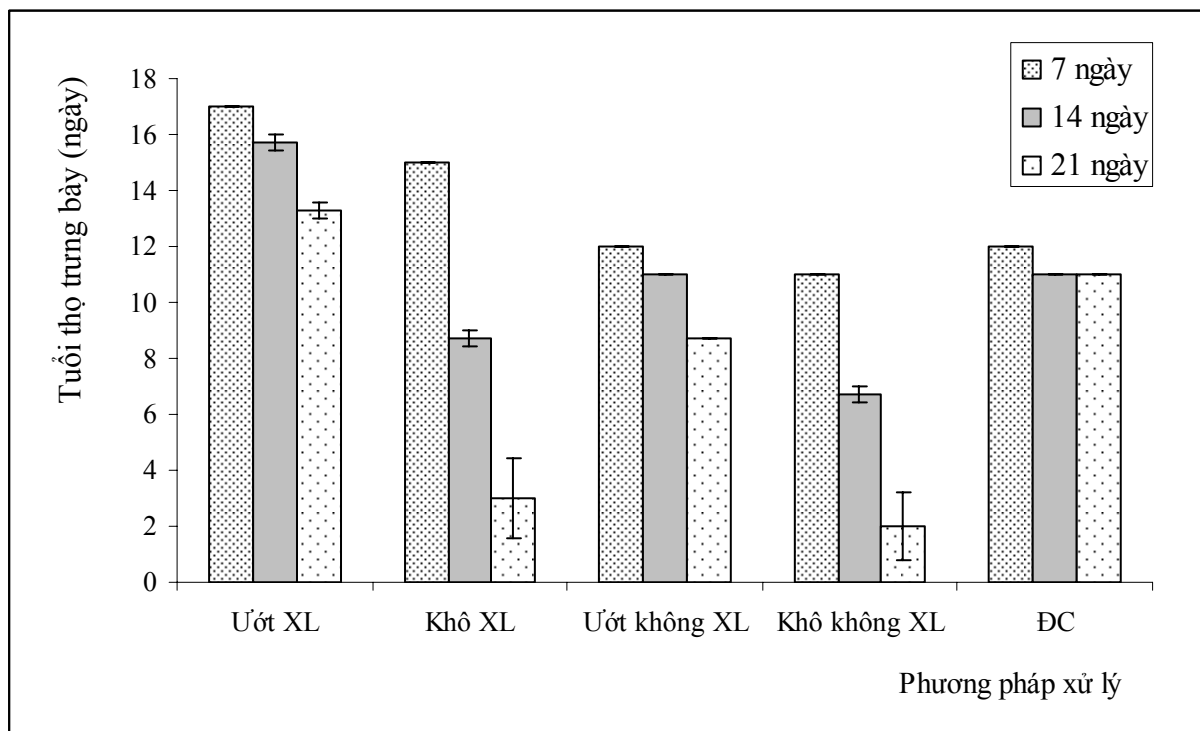
**Bước 3:** Đưa các giá trị SE vào mỗi giá trị trung bình

Nhấp chọn cột 7 ngày





Thực hiện tương tự cho cột 14 ngày và 21 ngày.



**Biểu đồ 2.1:** Tuổi thọ trung bình của hoa hồng sau xử lý và bảo quản lạnh 7 ngày, 14 ngày, 21 ngày.

# PHẦN II

## XỬ LÝ VÀ PHÂN TÍCH SỐ LIỆU

### Chương 1

## THỐNG KÊ MÔ TẢ VÀ MỘT SỐ KHÁI NIỆM CƠ BẢN

### I. Thống kê mô tả

Thống kê mô tả là phương pháp nhằm tóm tắt, tổng kết về kết quả của dữ liệu hay của thí nghiệm để nêu bật những thông tin quan trọng cần tìm hiểu. Nó bao gồm các tính toán cơ bản mang tính chất mô tả như số bình quân, độ lệch chuẩn, phương sai, hệ số biến động...

VD 1.1: sấy 50 mẫu xoài và cần ghi nhận độ dai, độ ẩm sau khi sấy. Một báo cáo cho biết độ dai bình quân, độ ẩm bình quân với độ lệch chuẩn và hệ số biến động của thí nghiệm sẽ dễ hiểu hơn là báo cáo số liệu thô của cả 50 mẫu xoài. Đó là bản chất của thống kê mô tả. Thống kê mô tả cung cấp một phương tiện để giảm một số lớn các số liệu phức tạp thành những thông tin có giá trị tóm tắt.

### II. Khái niệm về dân số và mẫu

**Dân số (population)** là tập hợp các quan sát có chung một số đặc tính mà ta quan tâm nghiên cứu. Tổng số quan sát trong dân số được ký hiệu là  $N$ .

VD 1.2:

Chiều cao của học sinh lớp 10 của Việt Nam thì dân số (tổng thể) là tập hợp tất cả chiều cao của học sinh lớp 10 ở Việt Nam.

Chiều cao của nữ sinh viên khoa Nông Lâm là tập hợp tất cả chiều cao của các nữ sinh viên ở khoa Nông Lâm.

Dân số thường có số quan sát rất lớn khó thu thập được toàn bộ số liệu. Vì thế để có được thông tin phản ánh về vấn đề cần quan tâm thì có thể thu thập số liệu thông qua mẫu.

**Mẫu (sample)** là một tập hợp con của dân số. Số quan sát trong mẫu được ký hiệu là  $n$ .

Việc phân tích số liệu trên mẫu có thể suy ra các đặc tính cho toàn bộ dân số với một mức độ tin cậy nào đó được xác định trước.

VD 1.3: Chiều cao của học sinh lớp 10 của Việt Nam thì dân số là tập hợp tất cả chiều cao của học sinh lớp 10 ở Việt Nam. Tuy nhiên, để đo được chiều cao của tất cả học sinh lớp 10 của cả nước thì rất tốn kém và mất nhiều thời gian. Do đó có thể chọn đo một số học sinh lớp 10, vậy chiều cao của số học sinh lớp 10 được chọn để đo là mẫu.

Việc chọn mẫu như thế nào, cỡ mẫu bao nhiêu đều có ảnh hưởng đến kết quả nghiên cứu, chọn mẫu phù hợp sẽ phản ánh đúng đặc tính của tổng thể (dân số). Nếu chỉ đo chiều cao của học sinh lớp 10 tại Hà Nội và TP. Hồ Chí Minh thì sẽ có sự thiên lệch rất lớn.

### III. Phương pháp lấy mẫu

#### 1. Chọn mẫu ngẫu nhiên đơn giản

Chọn mẫu ngẫu nhiên đơn giản (hay chọn mẫu hoàn toàn ngẫu nhiên) là phương pháp chọn mẫu sao cho khả năng được chọn của tất cả các đơn vị được chọn là như nhau. Mỗi đơn vị được chọn đều không có dụng ý trước mà chỉ là sự ngẫu nhiên.

Việc lấy mẫu ngẫu nhiên có thể tiến hành theo cách lấy mẫu không hoàn trả lại (sampling without replacement) hay theo cách lấy mẫu có hoàn trả lại (sampling with replacement).

## **2. Chọn mẫu ngẫu nhiên phân tầng**

Chọn mẫu ngẫu nhiên phân tầng (hay còn gọi là chọn mẫu phân loại điển hình) là phương pháp chọn mẫu dựa trên việc phân chia tổng thể thành nhiều nhóm khác nhau, sau đó lấy mẫu một cách ngẫu nhiên trong từng nhóm.

VD 1.4: điều tra kích thước của giống xoài A, ta có thể tiến hành ở vùng trọng điểm X sản xuất nhiều loại xoài A nhất của tỉnh đó.

## **IV. Phân loại số liệu**

### **1. Số liệu định lượng**

Là số liệu có thể cân, đong, đo, đếm được dễ dàng chính xác. Dữ liệu định lượng bao gồm những giá trị trả lời cho câu hỏi “bao nhiêu”?

### **2. Số liệu định tính**

Số liệu này không cân, đong, đo đếm được, dùng để xác định thuộc tính. Dữ liệu định tính sử dụng thang đo danh nghĩa hay thang đo thứ tự.

VD 1.5: Giới tính, màu sắc hạt, bệnh, hình dạng hạt...

## **V. Cách sắp xếp và trình bày số liệu**

### **1. Phân tổ**

Số liệu thống kê thường được trình bày dưới dạng bảng và đồ thị. Khi có số liệu thô, cần phải sắp xếp theo tần số hay nhóm để dễ quan sát và phân tích.

**Bảng 1.1:** 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

Bằng cách nhóm các chiều cao ta sẽ có thông tin dễ đánh giá hơn

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

$$k = 5 \lg(n)$$

Trong đó

k      là số tổ phân chia  
n      số quan sát

Xác định khoảng cách tổ (là số nguyên)

$$h = \frac{x_{\max} - x_{\min}}{k}$$

Trong đó

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

Với số liệu bảng 1.1, ta có số tổ  $k = 5 \times \lg 100 = 10$  tổ

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

Làm tròn 1cm không được chấp nhận vì chưa nhóm thành tổ. Như vậy khoảng cách tổ nên là 2cm. Tính lại số tổ  $k = 6$  (tổ).



## 2. Phân bố tần số

**Bảng 1.2:** Phân phối tần số về chiều cao cây cà chua

Chiều cao cây cà chua (cm)	Số cây cà chua (Tần số)
68-70	7
70-72	18
72-74	35
74-76	26
76-78	11
78-80	3

Sử dụng trong Excel

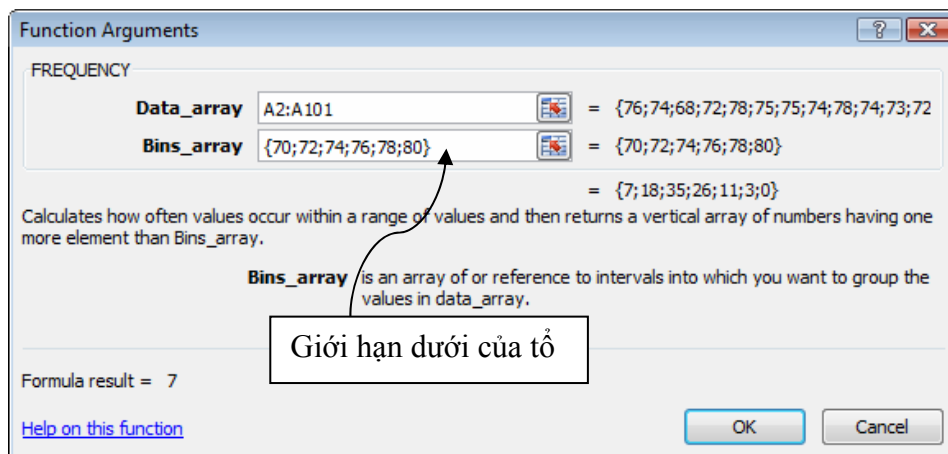
	A	B	C	D
1	Chiều cao cây cà chua (cm)	Chiều cao cây cà chua (cm)	Số cây cà chua (Tần số)	
2	76	68-70		
3	74	70-72		
4	68	72-74		
5	72	74-76		
6	78	76-78		
7	75	78-80		
8	75			
9	74			

Sử dụng hàm FREQUENCY để tính tần số trong mỗi tổ

Lưu ý hàm FREQUENCY trả về nhiều giá trị cùng một lúc hay trả số liệu khối. Do đó phải chọn khối mà hàm FREQUENCY trả về.

**Bước 1:** Chọn các ô từ C2:C7 (tần số sẽ xuất hiện tại các ô này)

**Bước 2:** Insert/Function/FREQUENCY



**Bước 3:** Không nhấn OK. Ấn tổ hợp phím CTRL+SHIFT+ENTER

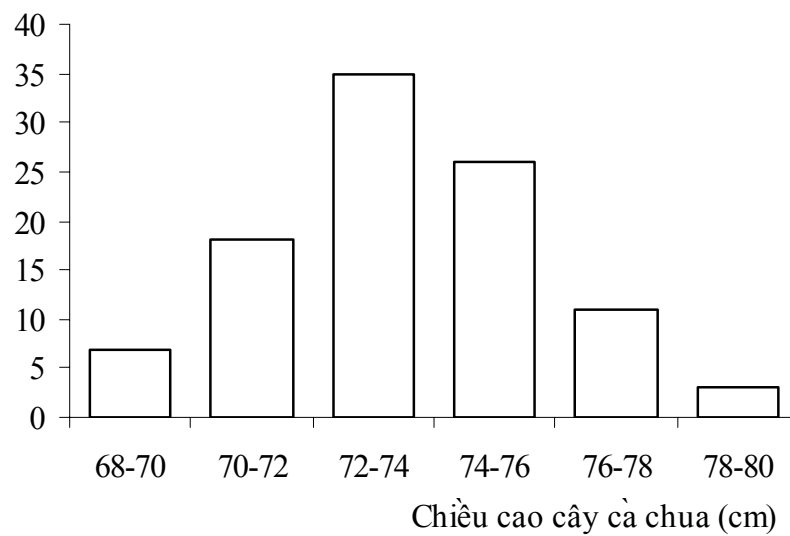
Kết quả bảng tính

	A	B	C
1	Chiều cao cây cà chua (cm)	Chiều cao cây cà chua (cm)	Số cây cà chua (Tần số)
2	76	68-70	7
3	74	70-72	18
4	68	72-74	35
5	72	74-76	26
6	78	76-78	11
7	75	78-80	3
8	75		

### Vẽ biểu đồ

Thực hiện trong Excel: **Insert/Chart/Column/Next**

Số cây cà chua (Tần số)



**Biểu đồ 1.1:** Phân bố tần số về chiều cao của cà chua

## 2.1. Phân bố tần số tích lũy

VD 1.6: Xác định hàm lượng phospho có trong lá cây, ta có một bảng phân bố tần số và phân bố tần số tích lũy của số liệu như sau:

**Bảng 1.3:** Hàm lượng phospho trong lá cây

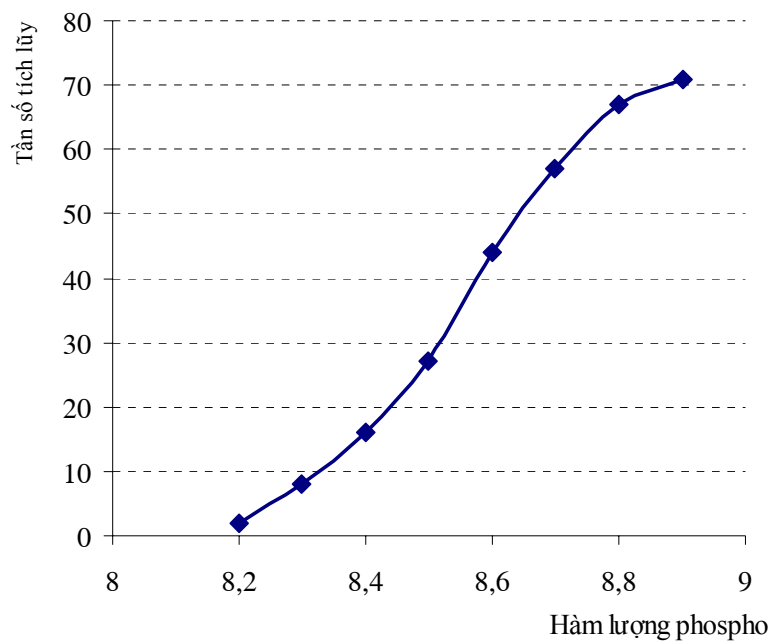
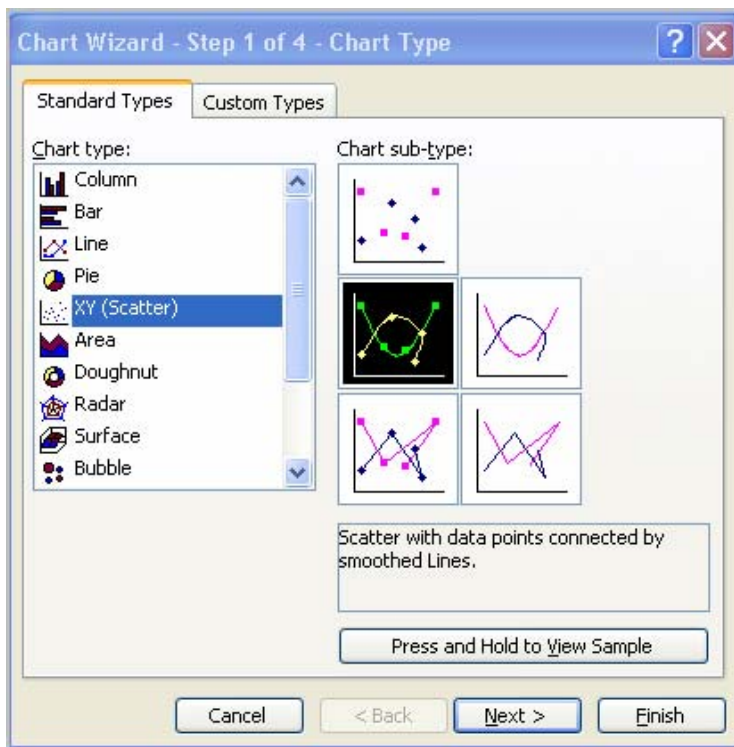
Lượng phospho (mg/g lá cây)	Tần số	Tần số tích lũy		Tần số tương đối tích lũy (%)	
		Bắt đầu từ thấp	Bắt đầu từ cao	Bắt đầu từ thấp	Bắt đầu từ cao
8,15-8,25	2	2	71	2,82%	100%
8,25-8,35	6	8	69	11,27%	97,18%
8,35-8,45	8	16	63	22,54%	88,73%
8,45-8,55	11	27	55	38,03%	77,46%
8,55-8,65	17	44	44	61,97%	61,97%
8,65-8,75	13	57	27	80,28%	38,03%
8,75-8,85	10	67	14	94,37%	19,72%
8,85-8,95	4	71	4	100%	5,63%

Giá trị tần số tích lũy có thể tính từ thấp đến cao hay từ cao đến thấp đều cần thiết. VD: Số lá có hàm lượng phospho ít hơn 8,55 mg/g là 27 tương ứng là 38,03%. Số lá có hàm lượng phospho lớn hơn 8,55 mg/g là 44 tương ứng 61,97%.

Vẽ biểu đồ tần số tích lũy

	A	B	C	D
1	Hàm lượng phospho mg/g lá cây	Giá trị giữa của tổ	Tần số tích lũy	
2	8,15-8,25	8,2	2	
3	8,25-8,35	8,3	8	
4	8,35-8,45	8,4	16	
5	8,45-8,55	8,5	27	
6	8,55-8,65	8,6	44	
7	8,65-8,75	8,7	57	
8	8,75-8,85	8,8	67	
9	8,85-8,95	8,9	71	
10				

Thực hiện trong Excel: **Insert/Chart/XY (Scatter)/Next**



**Biểu đồ 1.2:** Tần số tích lũy về hàm lượng phospho

## VI. Các tham số đặc trưng của mẫu

### 1. Đo sự tập trung

Các số trong một mẫu có khuynh hướng tập trung về một số nào đó. Để đo độ tập trung của các số người ta đưa ra các khái niệm trung bình, trung vị, số thường xuyên xuất hiện.

#### 1.1. Số trung bình

##### 1.1.1. Trung bình cộng

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

Trong đó  $\bar{X}$  là trung bình mẫu  
 $x_i$  giá trị quan sát thứ  $i$   
 $n$  số quan sát hay cỡ mẫu

##### 1.1.2. Bình quân gia quyền hay bình quân cộng có trọng số

Số bình quân gia quyền không chỉ phụ thuộc vào các giá trị của quan sát  $x_i$  mà còn phụ thuộc vào vai trò của các quan sát đó trong tổng thể.

$$\bar{X} = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i} \quad f_i \quad \text{là trọng số hay quyền số}$$

VD 1.7: Tính năng suất lúa bình quân vụ mùa của một số xã từ số liệu sau:

Năng suất (tạ/ha) $x_i$	Giá trị giữa của tổ	Diện tích (ha) $f_i$	$x_i f_i$
<30	27,5	150	4.125
30-35	32,5	100	3.250
35-40	37,5	200	7.500
40-45	42,5	400	17.000
45-50	47,5	250	11.875
>50	52,5	50	2.625
	Tổng	1.150	46.375

$$\text{Năng suất trung bình} = \frac{46.375}{1.150} = 40,3 \text{ tạ/ha}$$

#### 1.2. Số trung vị

Là số nằm giữa dãy số khi dãy số được sắp xếp từ nhỏ đến lớn. Số trung vị cho kết quả nhanh được về ước lượng trung bình.

VD 1.8: Khảo sát số quả cà chua của 7 cây giống có kết quả sau (ĐVT: quả/cây)

22    23    25    26    28    29    30

Giá trị đã sắp xếp theo 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. Do đó, công thức tổng quát tìm giá trị trung vị sẽ là

Nếu n lẻ thì số trung vị là số có thứ tự  $\frac{(n+1)}{2}$

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

$$M_e = \frac{x_{n/2} + x_{n/2+1}}{2}$$

VD 1.9: Số chiều dài trái xoài như sau

10    12    12    13    14    14    15    15    16    17

$$\text{Số trung vị sẽ là} = \frac{14+14}{2} = 14$$

### 1.3. Số Mode

Mode là số có tần số xuất hiện nhiều nhất trong dãy số quan sát.

Có thể có một hay nhiều số mode

VD 1.10: Số liệu trong VD 1.9 có 3 số mode là 12, 14, 15.

## 2. Đo độ phân tán

### 2.1. Khoảng biến thiên (Range)

Là chênh lệch giữa giá trị nhỏ nhất và giá trị lớn nhất trong dãy số liệu

$$R = x_{\max} - x_{\min}$$

### 2.2. Phương sai (Variance)

Phương sai là giá trị trung bình của bình phương các độ lệch giữa các giá trị của dữ liệu và giá trị trung bình.

**Phương sai mẫu** (Sample Variance)

$$S^2 = \frac{\sum (x_i - \bar{X})^2}{n-1}$$

Với  $n$  số quan sát trong mẫu (cỡ mẫu)  
 $\bar{X}$  trung bình mẫu.

**Phương sai tổng thể** (dân số)

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

Với  $N$  số quan sát trong tổng thể  
 $\mu$  trung bình tổng thể

### 2.3. Độ lệch chuẩn (Standard Deviation)

Độ lệch chuẩn là căn bậc 2 của phương sai

**Độ lệch chuẩn của mẫu** ( $S_x$ )

$$S_x = \sqrt{S^2}$$

**Độ lệch chuẩn của tổng thể**

$$\sigma = \sqrt{\sigma^2}$$

### 2.4. Sai số chuẩn (Standard Error)

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

## 2.5. Hệ số biến thiên (Coefficient of Variation)

Nó cho biết sai số của thí nghiệm, tham số thống kê này cho phép so sánh mức độ biến động của nhiều mẫu khác nhau ở các chỉ tiêu khác nhau.

Hệ số biến động được dùng để so sánh các đại lượng có độ lớn khác nhau. Thường vật lớn thì sai biệt giữa các quan sát cũng lớn, vật nhỏ thì sai biệt nhỏ. Ví dụ dễ hình dung là tai voi và tai chuột. Nhưng so sánh với số trung bình của chúng thì độ biến động không khác nhau lắm. Một ứng dụng khác là so sánh sự biến thiên giữa hai đại lượng khác đơn vị, ví dụ Chiều cao và Trọng lượng.

Tùy theo dữ liệu là mẫu hay tổng thể

$$CV\% = \frac{S}{\bar{X}} \times 100 \quad \text{hay} \quad CV\% = \frac{\sigma}{\mu} \times 100$$

VD 1.11: Có hai máy đóng gói A và B, kết quả

Máy A:  $\bar{X}_A = 21 \text{ g}$  và  $S_A = 3,2 \text{ g}$

Máy B:  $\bar{X}_B = 15 \text{ g}$  và  $S_B = 3 \text{ g}$

Tính  $CV_A = \frac{3,2}{21} \times 100 = 15,24\%$

$$CV_B = \frac{3}{15} \times 100 = 20\%$$

Vậy máy A đóng gói ổn định về trọng lượng hơn máy B

VD 1.12: Đo chiều cao của 20 khóm lúa giống P<sub>4</sub> lấy ngẫu nhiên có kết quả sau (ĐVT: cm)

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

	A	B	C	D	E	F
1	95		Trung bình	=AVERAGE(A1:A20)	96	
2	102		Khoảng biến thiên	=VALUE(MAX(A1:A20)-MIN(A1:A20))	11	
3	100		Phương sai	=VAR(A1:A20)	13,0526	
4	99		Độ lệch chuẩn	=STDEV(A1:A20)	3,6128	
5	91		Hệ số biến thiên	=(E4/E1)*100	3,7634	
6	95					
7	95					
8	97					
9	101					
10	102					
11	92					
12	93					
13	93					
14	94					
15	91					
16	96					
17	97					
18	100					
19	92					
20	95					
21						

Hãy sử dụng thống kê mô tả trong Excel

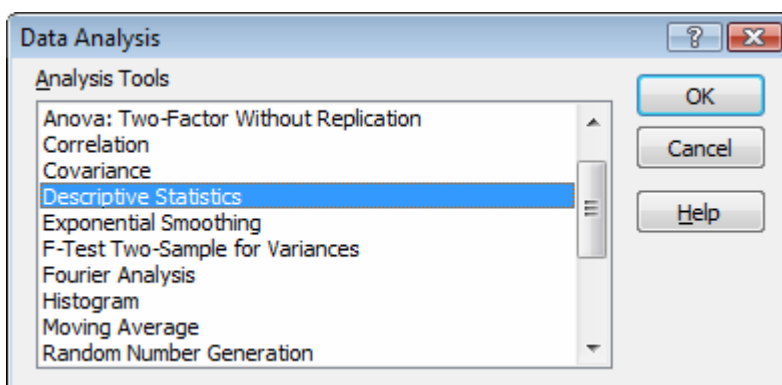
**Bước 1:** Chọn **Tool/Data Analysis**

**Bước 2:**

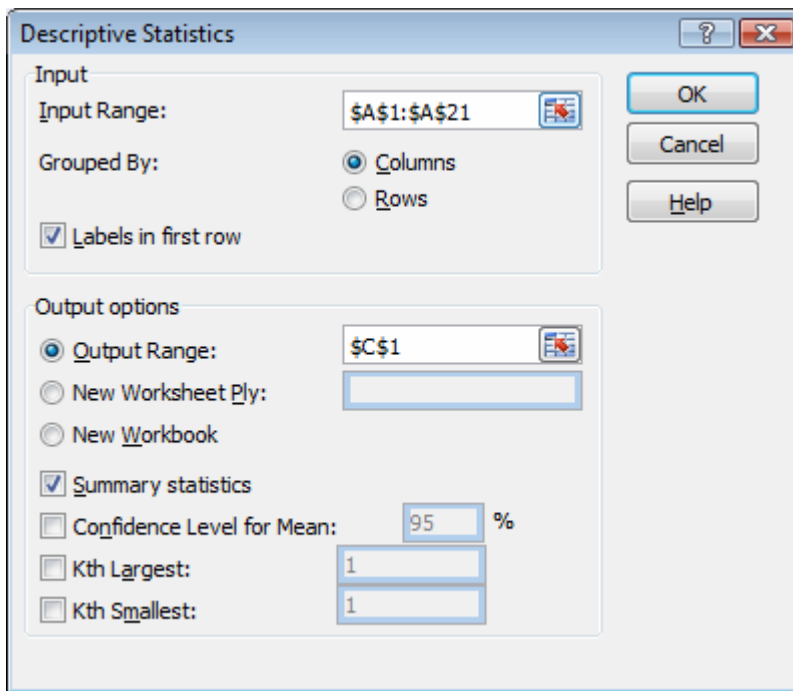
Nếu trong **Tool** chưa có **Data Analysis** thì nhập **Tool/Add-Ins...** xuất hiện hộp thoại chọn **Analysis ToolPak**

Nếu trong **Tool** đã có **Data Analysis** thì không thực hiện bước 2

**Bước 3:** Sau khi đã thực hiện **Tool/Data Analysis** xuất hiện hộp thoại chọn **Descriptive Statistics** nhập **OK**.








---

*Chieu cao khom lua*

---

Mean		96
Standard Error	0,807856162	
Median		95
Mode		95
Standard Deviation	3,61284259	
Sample Variance	13,05263158	
Kurtosis	-1,098602865	
Skewness	0,334827048	
Range		11
Minimum		91
Maximum		102
Sum		1920
Count		20

---

## Chương 2

# XÁC SUẤT VÀ PHÂN PHỐI XÁC SUẤT

### I. Xác suất

#### 1. Thí nghiệm ngẫu nhiên, không gian mẫu, biến cố

##### 1.1. Thí nghiệm ngẫu nhiên (Random Experiment)

Thí nghiệm ngẫu nhiên là một thí nghiệm có hai đặc tính:

- Không biết chắc hậu quả nào sẽ xảy ra.
- Nhưng biết được các hậu quả có thể xảy ra

VD 2.1: Thảy một con xúc sắc là một thí nghiệm ngẫu nhiên vì:

- Ta không biết chắc mặt nào sẽ xuất hiện
- Nhưng biết được có 6 trường hợp xảy ra. (Xúc sắc có 6 mặt 1, 2, 3, 4, 5, 6)

##### 1.2. Không gian mẫu (Sample Space)

Tập hợp các hậu quả có thể xảy ra trong thí nghiệm ngẫu nhiên gọi là không gian mẫu của thí nghiệm đó.

VD 2.2: Không gian mẫu của thí nghiệm thả một con xúc sắc là:

$$E = \{1, 2, 3, 4, 5, 6\}$$

VD 2.3: Không gian mẫu của thí nghiệm thả cùng một lúc hai đồng xu là:

$$E = \{SS, SN, NS, NN\} \text{ với S: Sấp, N: Ngửa}$$

##### 1.3. Biến cố (Event)

###### 1.3.1. Biến cố

- Mỗi tập hợp con của không gian mẫu là một biến cố
- Biến cố chứa một phần tử gọi là biến cố sơ đẳng

VD 2.4: Trong thí nghiệm thả 1 con xúc sắc :

- Biến cố các mặt chẵn xuất hiện là :  $\{2, 4, 6\}$
- Biến cố các mặt lẻ xuất hiện là :  $\{1, 3, 5\}$
- Các biến cố sơ đẳng là :  $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$

###### 1.3.2. Biến cố xảy ra (hay thực hiện)

Gọi  $r$  là một gọi hậu quả xảy ra và  $A$  là một biến cố

Nếu  $r \in A$  ta nói biến cố  $A$  xảy ra

Nếu  $r \notin A$  ta nói biến cố  $A$  không xảy ra

VD 2.5: Trong thí nghiệm thả một con xúc sắc nếu mặt 4 xuất hiện thì:

- Biến cố  $\{2,4,6\}$  xảy ra vì  $4 \in \{2, 4, 6\}$
- Biến cố  $\{1,3,5\}$  không xảy ra vì  $4 \notin \{1, 3, 5\}$

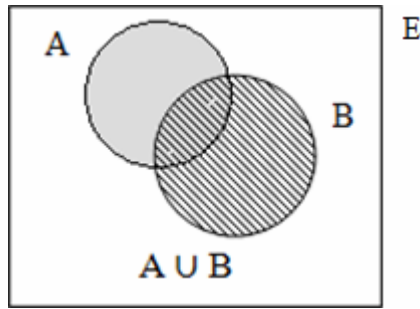
#### 1.4. Các phép tính về biến cố

Cho 2 biến cố  $A, B$  với  $A \subseteq E$  và  $B \subseteq E$

##### 1.4.1. Biến cố hội $A \cup B$ (Union)

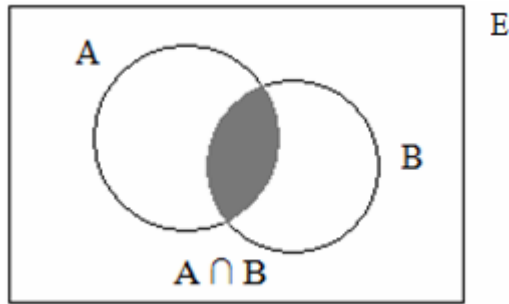
Biến cố hội của 2 biến cố  $A$  và  $B$  được ký hiệu là  $A \cup B$ .

$$A \cup B \text{ xảy ra} \Leftrightarrow (A \text{ xảy ra HAY } B \text{ xảy ra})$$



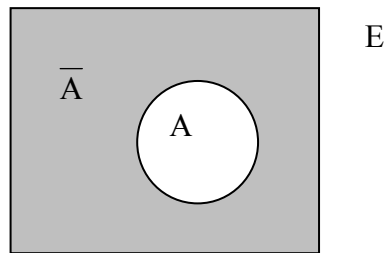
**1.4.2. Biến cố giao  $A \cap B$  (Intersection)**

$$A \cap B \text{ xảy ra} \Leftrightarrow (A \text{ xảy ra} \vee B \text{ xảy ra})$$



**1.4.3. Biến cố đối**

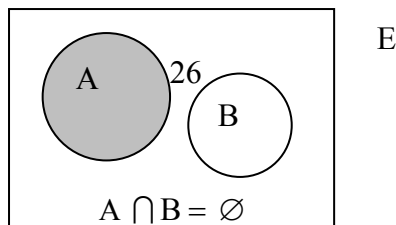
$$\bar{A} \text{ xảy ra} \Leftrightarrow A \text{ không xảy ra}$$



**1.4.4. Biến cố xung khắc**

Hai biến cố được gọi là xung khắc với nhau nếu một biến cố xảy ra thì biến cố kia không thể xảy ra.

$$A \text{ xung khắc với } B \Leftrightarrow A \cap B = \emptyset$$



VD 2.6: Trong thí nghiệm thả một con xúc sắc, ta có không gian mẫu:

$$E = \{1, 2, 3, 4, 5, 6\}$$

Gọi A là biến cố mặt lẻ xuất hiện  $\Rightarrow A = \{1, 3, 5\}$

Gọi B là biến cố khi bội số của 3 xuất hiện  $\Rightarrow B = \{3, 6\}$

Ta có:

$$A \cup B = \{1, 3, 5, 6\}$$

$$A \cap B = \{3\}$$

## 2. Xác suất (Probability)

### 2.1. Định nghĩa :

Nếu không gian mẫu E có N biến cố sơ đẳng và biến cố A có n biến cố sơ đẳng thì xác suất của biến cố A là :

$$P(A) = \frac{n(A)}{N}$$

Một cách khác ta có thể viết :

$$P(A) = \frac{\text{Số trường hợp A xảy ra}}{\text{Số trường hợp có thể xảy ra}}$$

VD 2.7: Trong thí nghiệm thả một con xúc sắc, xác suất của biến cố các mặt chẵn xuất hiện là

$$P(A) = \frac{n(A)}{N} = \frac{3}{6} = \frac{1}{2}$$

### 2.2. Tính chất :

Gọi A là một biến cố bất kỳ trong không gian mẫu E

$$0 \leq P(A) \leq 1$$

### 2.3. Công thức về xác suất :

#### 2.3.1. Xác suất của biến cố hội:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

#### Ghi chú

Nếu A và B là hai biến cố xung khắc ta có: thì định lý cộng xác suất trở thành

$$A \cap B = \emptyset \Rightarrow P(A \cap B) = P(\emptyset) = 0$$

thì định lý cộng xác suất trở thành  $P(A \cup B) = P(A) + P(B)$

#### 2.3.2. Xác suất của biến cố đối

$$P(A) + P(\bar{A}) = 1$$

#### 2.3.3. Xác suất có điều kiện

### Xác suất có điều kiện :

Gọi  $P(B/A)$  là xác suất có điều kiện của biến cố B sau khi biến cố A đã thực hiện.

$P(B/A) = P(A \cap B) / P(A)$	Với $P(A) > 0 ; P(B) > 0$
hay	
$P(A/B) = P(A \cap B) / P(B)$	

### Công thức nhân về xác suất:

Cho hai biến cố A và B trong không gian mẫu E, xác suất của biến cố giao được tính theo công thức:

$P(A \cap B) = P(B/A) \times P(A) \quad \text{hay} \quad P(A \cap B) = P(A/B) \times P(B)$
--

### Biến cố độc lập :

Biến cố gọi là độc lập với biến cố A về phương diện xác suất nếu xác suất của biến cố B không thay đổi cho dù biến cố A đã xảy ra, nghĩa là:

$$P(B/A) = P(B)$$

Ngược lại

$$P(A/B) = P(A)$$

Trong trường hợp hai biến cố độc lập, công thức nhân trở thành:

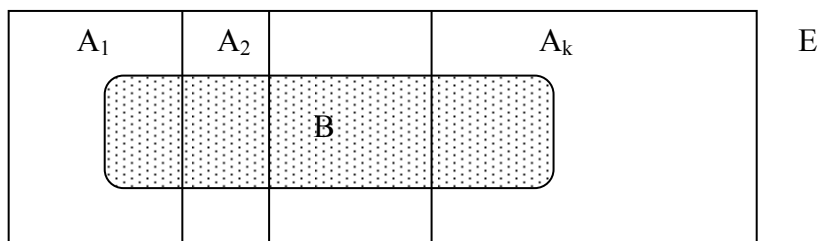
$$P(A \cap B) = P(A) \times P(B)$$

## 2.4. Công thức xác suất đầy đủ – Công thức Bayes

### Công thức xác suất đầy đủ

Giả sử biến cố B xảy ra khi và chỉ khi một trong các biến cố của hệ đầy đủ xung khắc nhau từng đôi một  $A_1, A_2, \dots, A_k$  xảy ra.

Biết xác suất  $P(A_i)$  và  $P(B/A_i)$  hãy tìm  $P(B)$



$$B = (B \cap A_1) \sqcup (B \cap A_2) \sqcup \dots \sqcup (B \cap A_k)$$

$$\Rightarrow P(B) = P[(B \cap A_1) \sqcup (B \cap A_2) \sqcup \dots \sqcup (B \cap A_k)] = P(B \cap A_1) + P(B \cap A_2) + \dots + P(B \cap A_k)$$

Vi

$$P(B \cap A_i) = P(B/A_i) \times P(A_i)$$

$P(B) = \sum_{i=1}^k P(B/A_i) \times P(A_i)$
--

## Công thức Bayes

Giải bài toán ngược của bài toán trên, tức là biết các  $P(A_i)$ ,  $P(B/A_i)$  và biến cố B đã xảy ra, tìm  $P(A_i/B)$

Ta có :

$$B = (B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3) \cup (B \cap A_4)$$

$$\text{và } P(A_i \cap B) = P(A_i/B) \times P(B) = P(B/A_i) \times P(A_i)$$

$$P(A_i/B) = \frac{P(B/A_i) \times P(A_i)}{P(B)}$$

$$P(A_i/B) = \frac{P(B/A_i) \times P(A_i)}{\sum_{i=1}^k P(B/A_i) \times P(A_i)}$$

## II. PHÂN PHỐI XÁC SUẤT

### 1. Biến ngẫu nhiên

**Biến ngẫu nhiên rời rạc** (Discrete Random Variable)

Nếu giá trị của biến ngẫu nhiên X có thể lập thành dãy rời rạc các số  $x_1, x_2, \dots, x_n$  (dãy hữu hạn hay vô hạn) thì X được gọi là biến ngẫu nhiên rời rạc.

**Biến ngẫu nhiên liên tục** (Continuous Random Variable)

Nếu giá trị của biến ngẫu nhiên X có thể lấp đầy toàn bộ khoảng hữu hạn hay vô hạn (a,b) thì biến ngẫu nhiên X được gọi là biến ngẫu nhiên liên tục.

VD 2.8: Số hạt nảy mầm là biến ngẫu nhiên rời rạc

Chiều cao của cây, nhiệt độ nấu, độ mềm của bánh là biến ngẫu nhiên liên tục.

### 2. Phân phối xác suất đối với biến ngẫu nhiên rời rạc

Phân phối xác suất của biến ngẫu nhiên rời rạc X được ký hiệu bằng hàm xác suất  $f(x)$ . Hàm xác suất của một biến ngẫu nhiên rời rạc thể hiện sự tương quan giữa xác suất và các giá trị của biến X.

Chúng ta có thể trình bày phân phối xác suất bằng phương pháp bảng, đồ thị hay biểu thức.

$$f(x) = P(X=x)$$

VD 2.9: Tung xúc sắc, biến ngẫu nhiên rời rạc có các giá trị  $X = 1, 2, 3, 4, 5, 6$

$$\text{Xác suất xuất hiện mặt 1 (xác suất khi giá trị biến ngẫu nhiên } X = 1) \quad P(X=1) = \frac{1}{6}$$

$$\text{Xác suất xuất hiện mặt 2 (xác suất khi giá trị biến ngẫu nhiên } X = 2) \quad P(X=2) = \frac{1}{6}$$

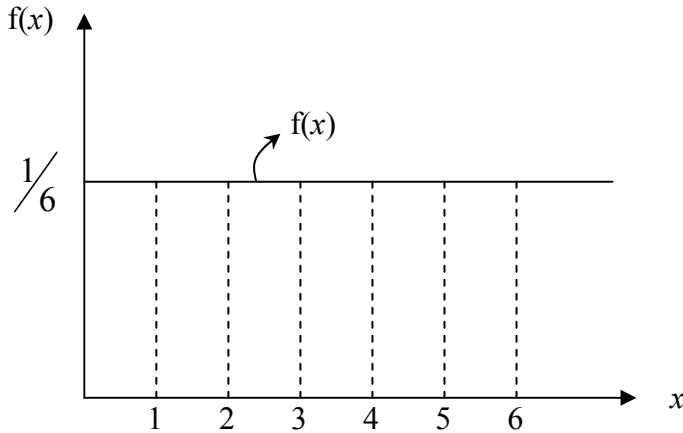
...

$$\text{Xác suất xuất hiện mặt 6 (xác suất khi giá trị biến ngẫu nhiên } X = 6) \quad P(X=6) = \frac{1}{6}$$

$$\Rightarrow P(X=1) = P(X=2) = P(X=3) = P(X=4) = P(X=5) = P(X=6) = \frac{1}{6}$$

Hàm phân phối xác suất đối với biến ngẫu nhiên X là  $f(x) = \frac{1}{6}$

$x$	1	2	3	4	5	6
$f(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$



### 3. Hàm xác suất tích lũy (Cumulative Probability Function)

Hàm xác suất tích lũy  $F_X(x_0)$  của biến ngẫu nhiên X thể hiện xác suất để X không vượt quá giới hạn  $x_0$ .

$$F_X(x_0) = P(X \leq x_0)$$

Tính chất

- $F_X(x_0) = \sum_{x \leq x_0} P_X(x)$
- $0 \leq F_X(x_0) \leq 1 \quad \forall x_0$
- Nếu  $x_1 \leq x_2$  thì  $F_X(x_1) \leq F_X(x_2)$

VD 2.10: Trong thí nghiệm thả 1 con xúc sắc

Tìm  $F_X(2,5)$ . Tìm xác suất tích lũy của biến ngẫu nhiên X đến giá trị 2,5.

Vì X là biến ngẫu nhiên rời rạc nên X có các giá trị 1, 2, 3, 4, 5, 6

$$F_X(2,5) = P(X \leq 2,5) = P(X=1) + P(X=2) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

### 4. Phân phối xác suất đối với biến ngẫu nhiên liên tục (Probability Distributions For Continuous Random Variables)

Phân phối xác suất của biến ngẫu nhiên liên tục được xác định bởi hàm mật độ xác suất.

#### Hàm mật độ xác suất (Probability Density Function)

Gọi X là biến ngẫu nhiên liên tục, gọi x là giá trị bất kỳ nằm trong miền các giá trị có thể có của X.

Hàm mật độ xác suất  $f(x)$  của biến ngẫu nhiên liên tục là hàm có những tính chất sau:

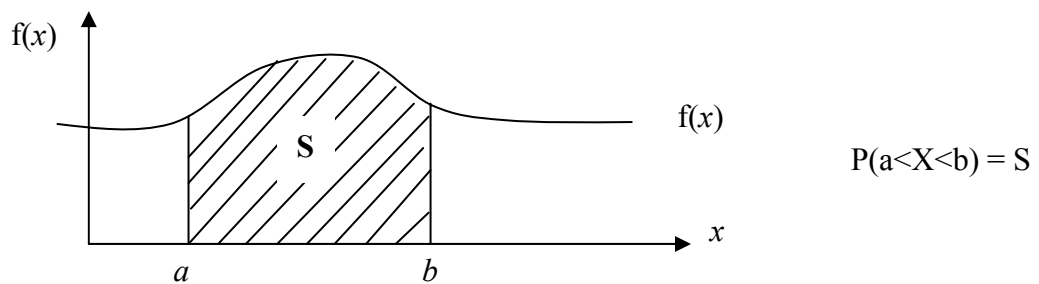
- $f(x) \geq 0, \forall x$
- Xác suất  $P(a < X < b)$  để giá trị của biến ngẫu nhiên X rơi vào khoảng (a,b) được xác định bởi đẳng thức.

$$P(a < X < b) = \int_a^b f(x) dx$$

### Ghi chú

Đồ thị của hàm mật độ xác suất  $f(x)$  được gọi là đường cong mật độ xác suất (Probability Density Curve) hay đường cong tần số (Frequency Curve) hay cũng còn được gọi đường cong phân phối xác suất đối với biến ngẫu nhiên liên tục. Tung độ của mỗi điểm trên đường cong gọi là mật độ xác suất.

Về mặt hình học xác suất để biến ngẫu nhiên rơi vào khoảng  $(a, b)$  bằng diện tích hình thang cong giới hạn bởi đường cong phân phối xác suất, trục  $Ox$ ,  $x = a$ ,  $x = b$ .



$$\int_{-\infty}^{+\infty} f(x) dx = 1 \Rightarrow \text{Toàn bộ diện tích hình thang cong là 1}$$

### Hàm phân phối tích lũy (Cumulative Distribution Function)

Hàm phân phối tích lũy  $F_X(x)$  của biến ngẫu nhiên liên tục  $X$  thể hiện xác suất để  $X$  không vượt quá giá trị  $x$ .

$$F_X(x) = P(X \leq x)$$

### Tính chất

- $F_X(x) = \int_{-\infty}^x f(x) dx$  với  $f(x)$  là hàm mật độ xác suất
- $0 \leq F_X(x) \leq 1$
- $F_X(-\infty) = 0$
- $F_X(+\infty) = 1$
- $P(a < X < b) = F_X(b) - F_X(a)$

## 5. Phân phối xác suất chuẩn (The Normal Distribution)

### Hàm mật độ xác suất của phân phối chuẩn

Nếu hàm mật độ xác suất của biến ngẫu nhiên  $X$  có dạng

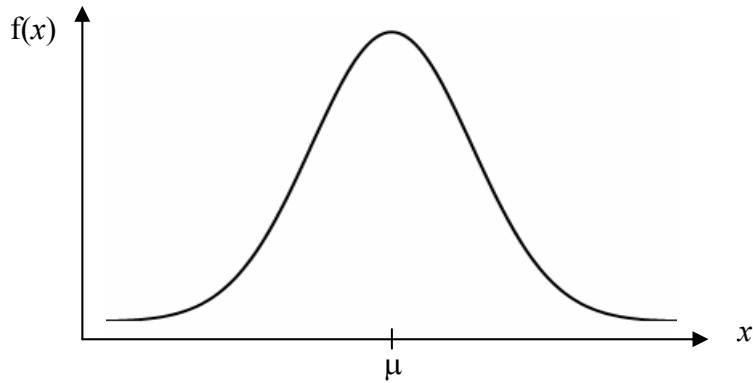


$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

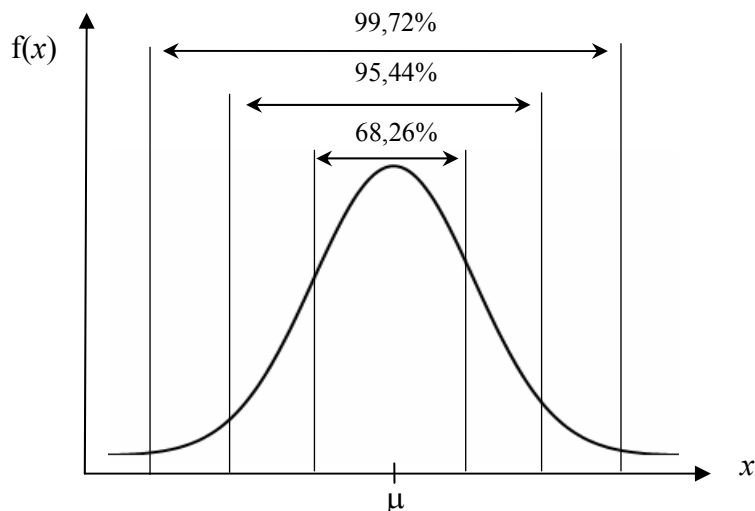
thì biến ngẫu nhiên  $X$  được gọi là tuân theo luật phân phối chuẩn.

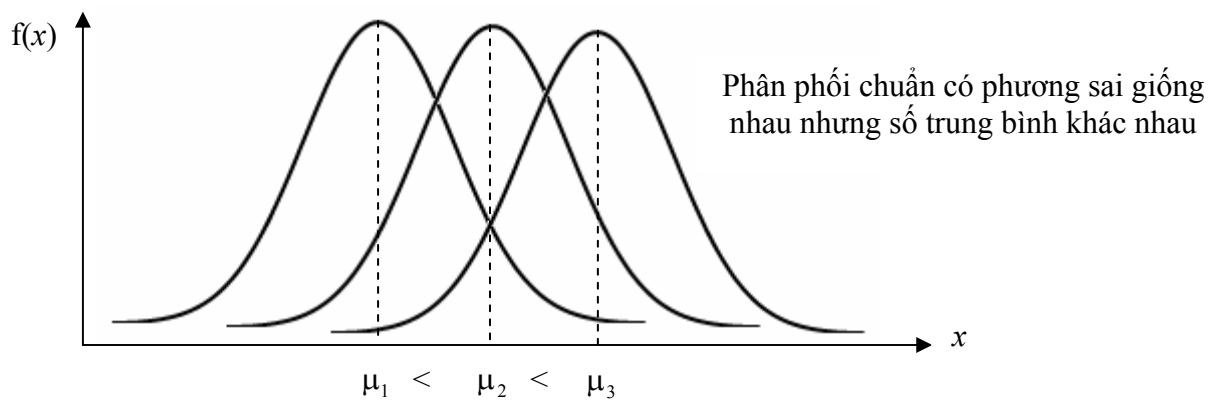
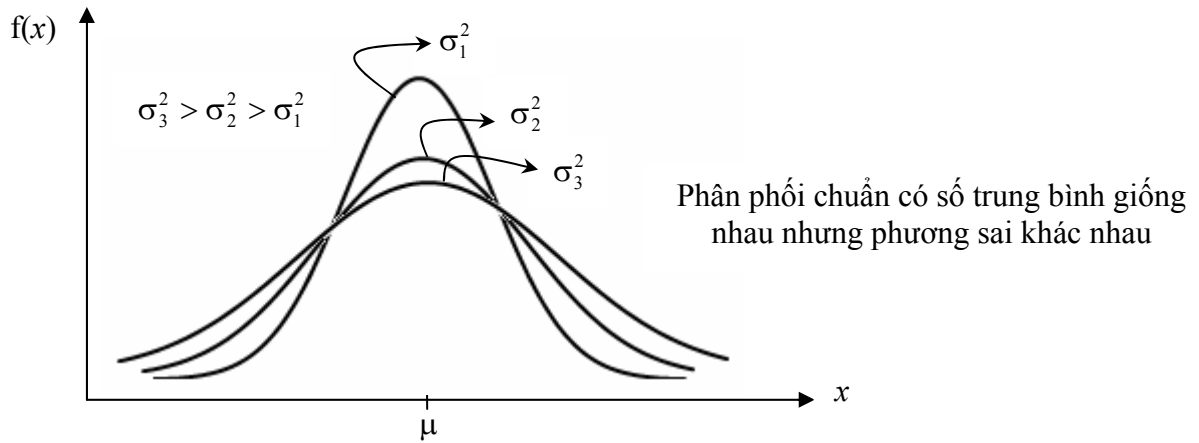
### Tính chất của phân phối chuẩn

Hàm mật độ xác suất của phân phối chuẩn có dạng hình chuông, đối xứng qua trị số trung bình  $\mu$ .



- Hai thông số  $\mu$  (giá trị trung bình) và  $\sigma$  (độ lệch chuẩn) xác định vị trí và hình dạng phân phối xác suất chuẩn.
- Điểm cao nhất của đường cong ở tại giá trị trung bình  $\mu$ , đồng thời cũng là số trung vị và yếu vị (số mode).
- Giá trị trung bình có thể âm, dương hay bằng 0.
- Độ lệch chuẩn xác định độ rộng của đường cong, nó càng lớn thì đường cong càng rộng.
- Tổng diện tích dưới đường cong là 1 (0,5 bên trái giá trị trung bình và 0,5 bên phải).
- Xác suất của biến ngẫu nhiên chuẩn chính là diện tích dưới đường cong.
- Với phân phối chuẩn
  - 68,26% giá trị của biến nằm trong khoảng  $\mu \pm \sigma$
  - 95,44% giá trị của biến nằm trong khoảng  $\mu \pm 2\sigma$
  - 99,72% giá trị của biến nằm trong khoảng  $\mu \pm 3\sigma$





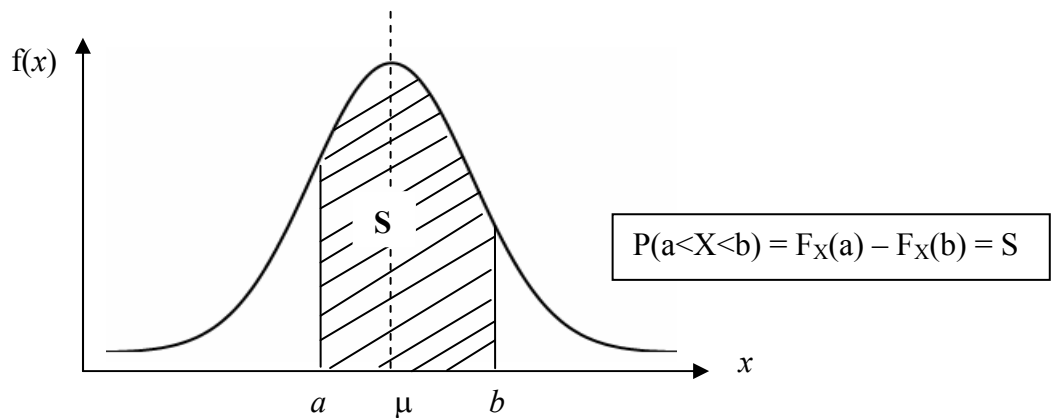
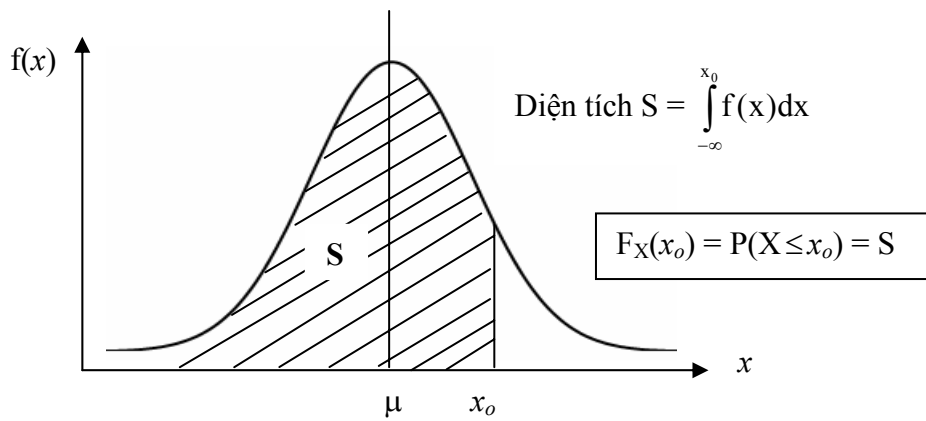
Nếu biến ngẫu nhiên  $X$  tuân theo phân phối chuẩn có số trung bình là  $\mu$  và phương sai là  $\sigma^2$ , ta ký hiệu:

$$X \sim N(\mu, \sigma^2)$$

Hàm phân phối tích lũy của phân phối chuẩn

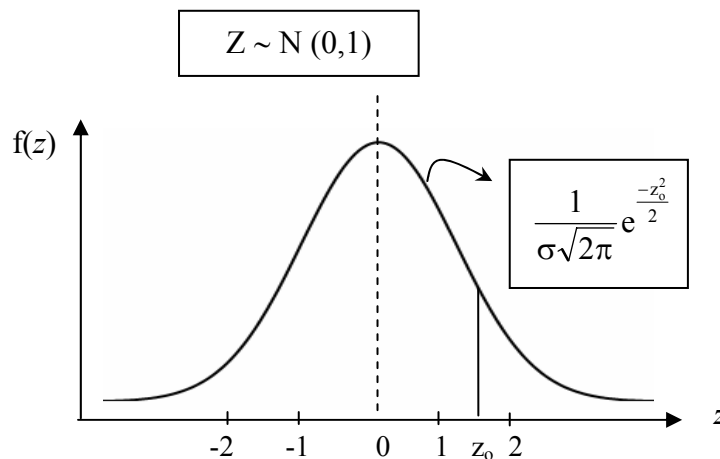
Cho  $X \sim N(\mu, \sigma^2)$ . Hàm phân phối tích lũy của biến ngẫu nhiên  $X$  tuân theo phân phối chuẩn được định nghĩa như sau:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$



## 6. Phân phối chuẩn chuẩn hóa (Standard Normal Distribution)

Phân phối chuẩn chuẩn hóa là phân phối chuẩn có số trung bình là 0 và phương sai là 1. Biến ngẫu nhiên tuân theo phân phối chuẩn chuẩn hóa được gọi là biến ngẫu nhiên chuẩn hóa (Standard Normal Variable) và được ký hiệu là  $Z$ .

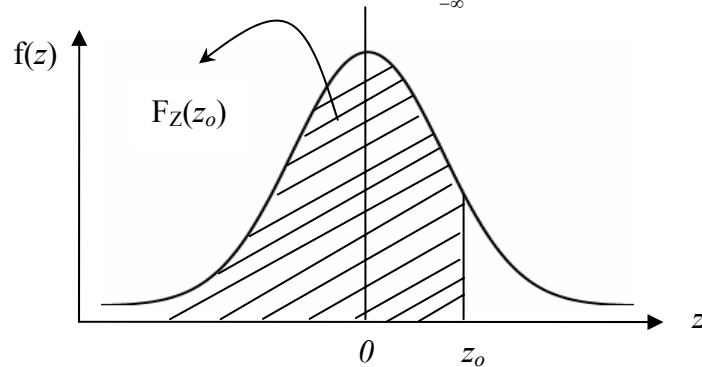


Tung độ của một điểm bất kỳ trên đường cong chuẩn sẽ được xác định từ phương trình của hàm mật độ xác suất của phân phối chuẩn.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{Với } \mu = 0, \sigma = 1 \text{ và } x = z$$

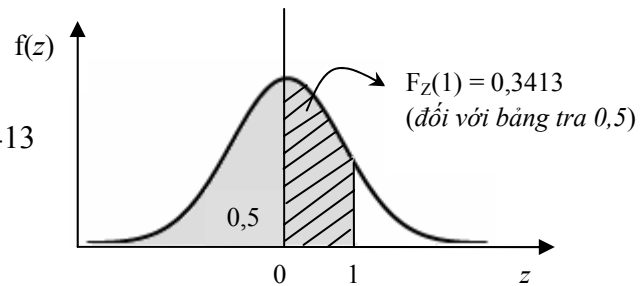
Giá trị của hàm phân phối tích lũy của phân phối chuẩn chuẩn hóa (cũng bằng diện tích nằm dưới đường cong chuẩn) được lập thành bảng.

Các bảng này cho các giá trị của  $F_Z(z_0) = P(Z \leq z_0) = \int_{-\infty}^{z_0} f(z) dz$



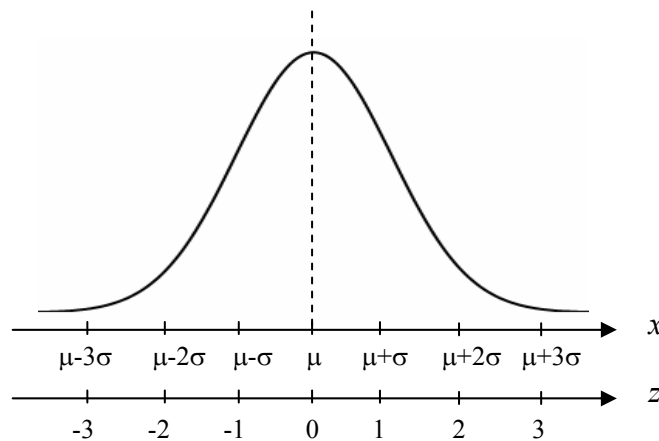
Một số bảng lập sẵn (phần phụ lục), chỉ cho ta diện tích nằm dưới đường cong chuẩn từ 0 đến z.

VD 2.11:  $F_Z(1) = P(Z \leq 1) = 0,5 + 0,3413$



### Chuẩn hóa biến ngẫu nhiên

Nếu biến ngẫu nhiên X có số trung bình là  $\mu$  và phương sai là  $\sigma^2$  hay  $X \sim N(\mu, \sigma^2)$ , thì biến ngẫu nhiên  $Z = \frac{X - \mu}{\sigma}$  sẽ có số trung bình là 0 và phương sai là 1. Z được gọi là biến ngẫu nhiên được chuẩn hóa.



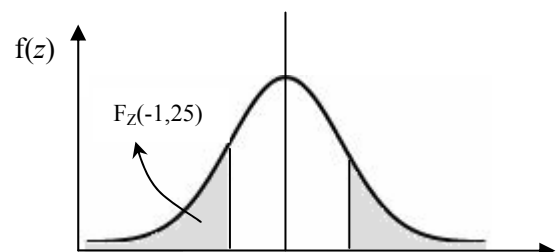
VD 2.12:

Cho  $Z \sim N(0,1)$ . Tìm xác suất để giá trị của Z

- Nhỏ hơn -1,25
- Nằm trong khoảng (-0,5; 0,75)
- Lớn hơn 1

a. Tìm xác suất để giá trị của  $Z \leq -1,25$

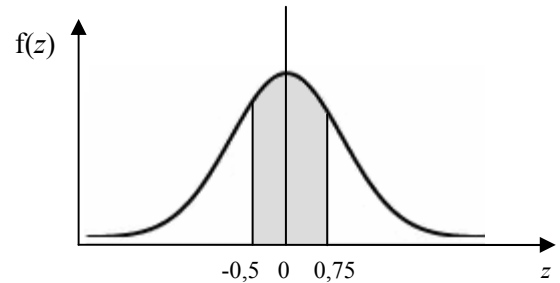
35



$$\begin{aligned}
 \text{Tìm } P(Z \leq -1,25) &= P(Z \geq 1,25) = F_Z(-1,25) \\
 F_Z(-1,25) &= 0,5 - F_Z(1,25) \\
 &= 0,5 - 0,3944 \\
 &= 0,1056
 \end{aligned}$$

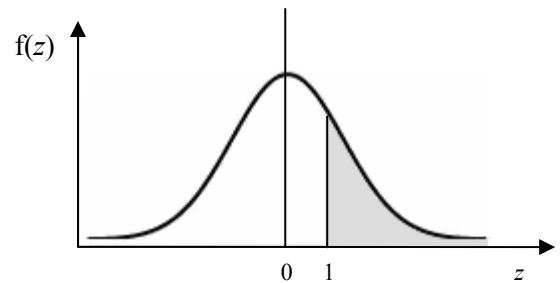
**b. Tìm xác suất để giá trị của Z nằm trong khoảng (-0,5;0,75)**

$$\begin{aligned}
 \text{Tìm } P(-0,5 \leq Z \leq 0,75) &= F_Z(0,75) - F_Z(-0,5) \\
 &= 0,7734 - 0,3085 \\
 &= 0,4649
 \end{aligned}$$



**c. Tìm xác suất để giá trị của Z ≥ 1**

$$\begin{aligned}
 \text{Tìm } P(Z \geq 1) &= 0,5 - F_Z(1) \\
 &= 0,5 - 0,3413 \\
 &= 0,1587
 \end{aligned}$$



**Sử dụng Excel để tính xác suất của phân phối chuẩn chuẩn hóa**

**NORMSDIST** để tính xác suất tích lũy với một giá trị z cho trước.

	A	B	C	D
1	<b>Tính xác suất của phân phối chuẩn chuẩn hóa</b>			
2			<b>Kết quả</b>	
3	$P(Z \leq -1,25)$	<code>=NORMSDIST(-1,25)</code>	0,1056	
4	$P(-0,5 \leq Z \leq 0,75)$	<code>=NORMSDIST(0,75)-NORMSDIST(-0,5)</code>	0,4648	
5	$P(Z \geq 1)$	<code>=1-NORMSDIST(1)</code>	0,1587	
6				

Hãy **Insert/Function/NORMSDIST**

	A	B	C	D
1	<b>Tính xác suất của phân phối chuẩn chuẩn hóa</b>			
2				
3	P(Z ≤ -1,25)	=NORMSDIST(-1,25)		
4				
5	<div data-bbox="368 360 1286 815" style="border: 1px solid gray; padding: 5px;"> <p>Function Arguments</p> <p>NORMSDIST</p> <p>Z: [-1,25] = -1,25</p> <p>= 0,105649774</p> <p>Returns the standard normal cumulative distribution (has a mean of zero and a standard deviation of one).</p> <p>Z is the value for which you want the distribution.</p> <p>Formula result = 0,1056</p> <p><a href="#">Help on this function</a></p> <p>OK Cancel</p> </div>			
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				

VD 2.13: Cho  $X \sim N(15, 16)$ . Tìm xác suất  $X$  có giá trị lớn hơn 18  
 $X$  là biến ngẫu nhiên có phân phối chuẩn với  $\mu = 15$  và  $\sigma^2 = 16$

$$x = 18 \Rightarrow Z = \frac{x - \mu}{\sigma} = \frac{18 - 15}{4} = 0,75$$

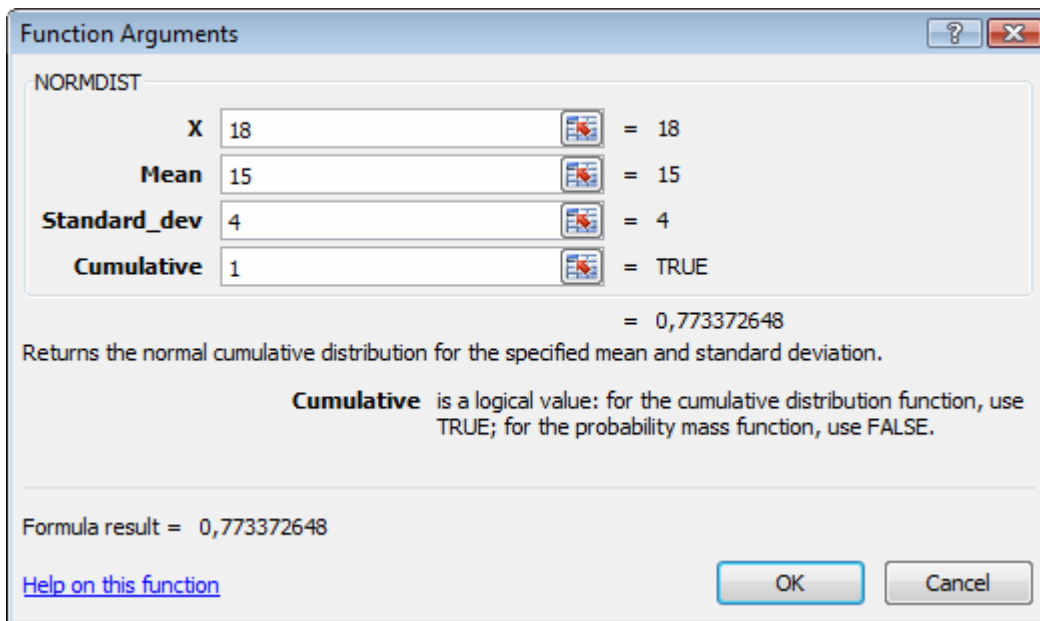
$$P(X > 18) = P(Z > 0,75) = 0,5 - F_Z(0,75) = 0,5 - 0,2734 = 0,2266$$

### Sử dụng Excel để tính xác suất của phân phối chuẩn

**NORMDIST (giá trị x, trung bình, độ lệch chuẩn, 1)** để tính xác suất tích lũy với một giá trị  $x$  cho trước. (1 là để tính xác suất tích lũy với một giá trị  $x$  cho trước, 0 là để tính xác suất tại giá trị  $x$ ).

	A	B	C	D
1			<b>Kết quả</b>	
2	P(X > 18)	=1-NORMDIST(18;15;4;1)	<b>0,22663</b>	
3				

Hãy **Insert/Function/NORMDIST**



Từ kết quả trên cho xác suất tích lũy đến giá trị  $x = 18$ ,  $P(X \leq 18) = 0,7734$   
 $\Rightarrow P(X > 18) = 1 - P(X \leq 18) = 1 - 0,7733 = 0,2266$

**VD 2.14:**

Nếu  $X$  là biến ngẫu nhiên tuân theo phân phối chuẩn có số trung bình là 3 và độ lệch chuẩn là 2. Tìm  $P(4 < X < 6)$ .

$$x = 4 \Rightarrow Z = \frac{x - \mu}{\sigma} = \frac{4 - 3}{2} = 0,5$$

$$x = 6 \Rightarrow Z = \frac{x - \mu}{\sigma} = \frac{6 - 3}{2} = 1,5$$

$$P(4 < X < 6) = P(0,5 < Z < 1,5) = F_Z(1,5) - F_Z(0,5) = 0,4332 - 0,1915 = 0,2417$$

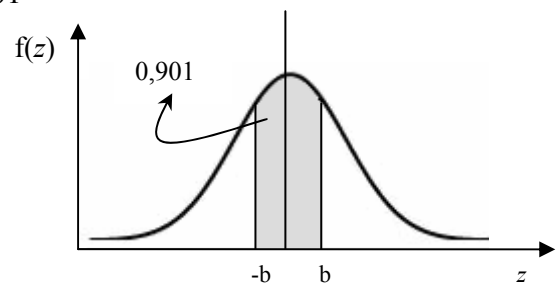
	A	B	C	D
1			<b>Kết quả</b>	
2	$P(4 < X < 6)$	<code>=NORMDIST(6;3;2;1)-NORMDIST(4;3;2;1)</code>	<b>0,2417</b>	
3				

**VD 2.15:** Tìm giá trị của  $b$  biết rằng  $P(-b < Z < b) = 0,901$

$$2F_Z(b) = 0,901$$

$$\Rightarrow F_Z(b) = 0,4505$$

Tra bảng, với xác suất tích lũy là 0,4505 thì giá trị của biến  $Z$  là  $b = 1,65$ .



**Sử dụng Excel để tính giá trị của phân phối chuẩn chuẩn hóa**

**NORMSINV** để tính giá trị  $z$  với một xác suất tích lũy cho trước.

	A	B	C	D
1	Tính xác suất của phân phối chuẩn chuẩn hóa			
2			<b>Kết quả</b>	
3	1. Tìm b với $P(Z \leq b) = 0,9452$	<code>=NORMSINV(0,9452)</code>	<b>1,60</b>	
4				
5	2. Tìm b biết $P(-b < Z < b) = 0,901$			
6	2.1. Tìm -b	<code>=NORMSINV(0,5-(0,901/2))</code>	<b>-1,65</b>	
7	2.2. Tìm b	<code>=NORMSINV(0,5+(0,901/2))</code>	<b>1,65</b>	
8				

## 7. Phân phối Student (Phân phối t)

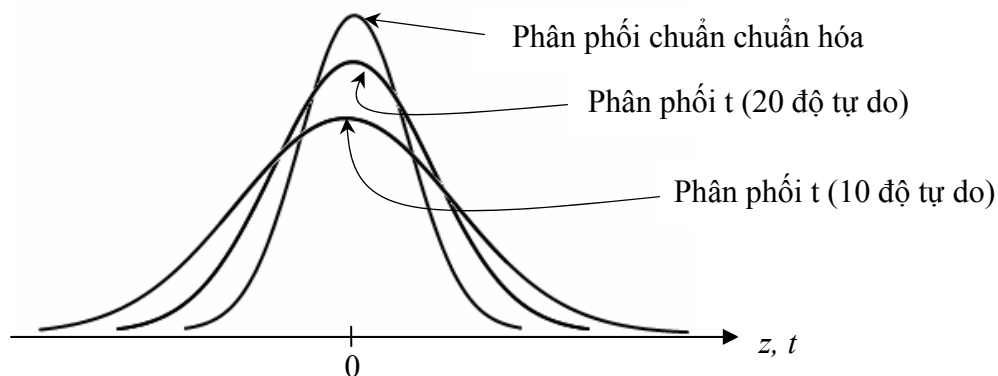
Phân phối t là một họ các phân phối xác suất tương tự được phân biệt với nhau bằng độ tự do.

Biến ngẫu nhiên  $t = \frac{\bar{X} - \mu}{S_x / \sqrt{n}}$ , t tuân theo phân phối Student với độ tự do  $n - 1$ .

Khi độ tự do tăng lên thì sự khác biệt giữa phân phối t và phân phối chuẩn chuẩn hóa càng giảm. Nói cách khác, phân phối t sẽ tiệm cận đến phân phối chuẩn chuẩn hóa.

Giá trị trung bình của phân phối t bằng 0. Đối xứng qua  $t = 0$  và cực đại ở  $t = 0$ .

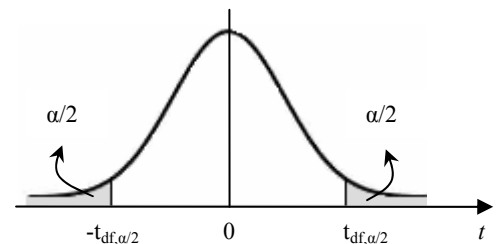
Tìm giá trị t với một xác suất ( $\alpha$ ) và độ tự do (df) cho trước (phụ lục)



### Tra bảng t trong Excel

**TINV( $\alpha$ , df)** để tính giá trị t với một xác suất ( $\alpha$ ) và độ tự do (df) cho trước.

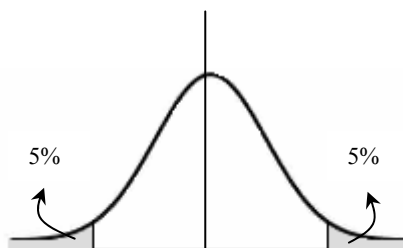
Trong đó  $\alpha$  là diện tích của 2 đuôi.



**VD 2.16:** Tra  $t_{5,5\%} = ?$

Vì  $\alpha$  là diện tích của cả hai đuôi, do đó  $\alpha = 10\%$

	A	B	C	D
1	Tra $t_{5,5\%} = ?$	<code>=TINV(10%;5)</code>	<b>2,015</b>	
2				





**TDIST(giá trị t, bậc tự do, đuôi)** tìm xác suất khi biết giá trị t và bậc tự do.

Tìm diện tích (xác suất) một đuôi

TDIST(giá trị t, bậc tự do, 1)

Tìm diện tích (xác suất) hai phía

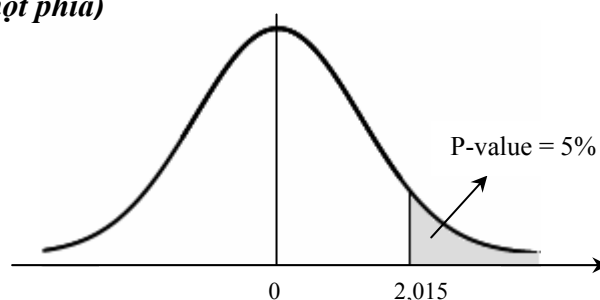
TDIST(giá trị t, bậc tự do, 2)

**VD 2.17:**

Cho giá trị  $t_c = 2,015$  với bậc tự do là 5

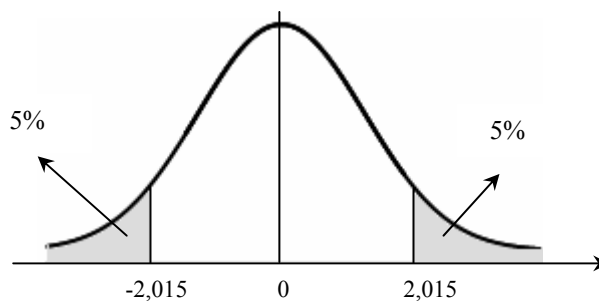
*Tìm xác suất tương ứng với giá trị  $t = 2,015$  (ở một phía)*

$$\begin{aligned} &= \text{TDIST}(2,015;5;1) \\ &= 0,05 \end{aligned}$$



*Tìm xác suất tương ứng với giá trị  $t = 2,015$  (ở hai phía)*

$$\begin{aligned} &= \text{TDIST}(2,015;5;2) \\ &= 0,1 \end{aligned}$$



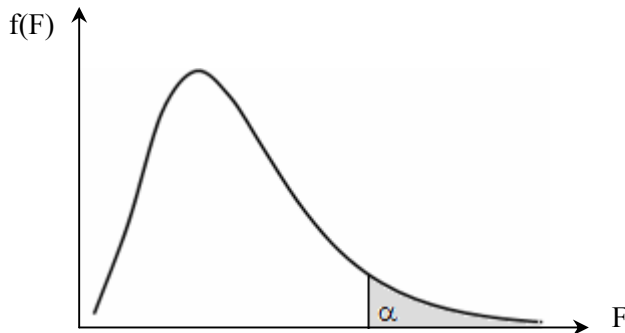
## 8. Phân phối F (Fisher)

Cho tổng thể  $X_1 \sim N(\mu_1, \sigma_1^2)$  từ tổng thể 1 lấy mẫu với cỡ mẫu  $n_1$  và tính được  $S_1^2$

$X_2 \sim N(\mu_2, \sigma_2^2)$  từ tổng thể 1 lấy mẫu với cỡ mẫu  $n_1$  và tính được  $S_2^2$

Biến số  $F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$  có xác suất phân phối theo quy luật nhất định gọi là phân phối F.

$F(\alpha, df_1, df_2)$  là giá trị F với xác suất ( $\alpha$ ), độ tự do ở tử ( $df_1$ ) và độ tự do ở mẫu ( $df_2$ ) cho trước (Phụ lục)



### Tra bảng F trong Excel

=**FINV**( $\alpha$ ;  $df_1$ ;  $df_2$ ) tìm giá trị F với xác suất ( $\alpha$ ), độ tự do ở tử ( $df_1$ ) và độ tự do ở mẫu ( $df_2$ ) cho trước.

=**FDIST**(giá trị F;  $df_1$ ;  $df_2$ ) tìm xác suất với giá trị F, độ tự do ở tử ( $df_1$ ) và độ tự do ở mẫu ( $df_2$ ) cho trước.

## Chương 3

### ƯỚC LƯỢNG

#### I. 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 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.

#### Ước lượng điểm

$\bar{X}$  cho giá trị trung bình  $\mu$

S cho độ lệch chuẩn  $\sigma$

$\bar{p}$  cho tỷ lệ tổng thể p

#### Sai số của việc lấy mẫu

Sự khác biệt tuyệt đối giữa một ước lượng điểm và thông số tương ứng của tổng thể gọi là sai số lấy mẫu.

**Sai số lấy mẫu**  $|\bar{X} - \mu|$  đối với trung bình của mẫu

$|S - \sigma|$  đối với độ lệch chuẩn của mẫu

$|\bar{p} - p|$  đối với tỷ lệ của mẫu

#### II. 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 giá trị 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  $\theta$  là tham số cần ước lượng. Nếu gọi  $q_1$  là giới hạn dưới và  $q_2$  là giới hạn trên,  $\alpha$  là xác suất để mắc sai lầm thì ước lượng khoảng của  $\theta$  được viết như sau:

$$P(a \leq \theta \leq b) = 1 - \alpha$$

Trong đó

[a,b] là khoảng tin cậy của tham số  $\theta$

$1 - \alpha$  là độ tin cậy (thường được chọn theo yêu cầu của người nghiên cứu thông thường độ tin cậy được chọn 0,95; 0,99; 0,999).

$\alpha$  xác suất sai lầm khi chọn khoảng tin cậy [a,b].

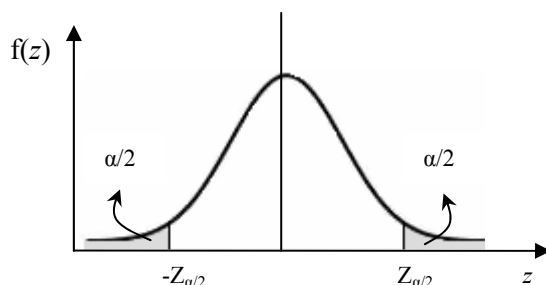
# 1. Ước lượng giá trị trung bình của tổng thể trường hợp mẫu lớn ( $n \geq 30$ )

Giả sử  $X$  có phân phối chuẩn  $N(\mu, \sigma^2)$ , trung bình của tổng thể được tính bởi

$$\bar{X} - \frac{Z_{\alpha/2} \sigma_x}{\sqrt{n}} < \mu < \bar{X} + \frac{Z_{\alpha/2} \sigma_x}{\sqrt{n}}$$

Trong đó

- $\bar{X}$  là trung bình của mẫu
- $\sigma_x$  độ lệch chuẩn của tổng thể
- $n$  cỡ mẫu
- $1 - \alpha$  độ tin cậy
- $Z_{\alpha/2}$  là số có  $P(Z > Z_{\alpha/2}) = \alpha/2$



Trong thực tế thì hầu như chúng ta không thể tính được phương sai của tổng thể ( $\sigma^2$ ) mà chỉ tính được phương sai của mẫu.

$$\bar{X} - \frac{Z_{\alpha/2} S_x}{\sqrt{n}} < \mu < \bar{X} + \frac{Z_{\alpha/2} S_x}{\sqrt{n}}$$

**VD 3.1:** Chọn mẫu  $n=50$ , điều tra năng suất của giống cà chua xuân hè (kg/cây). Từ đó có năng suất trung bình 1,48 kg; độ lệch chuẩn của năng suất 0,35 kg/cây. Hãy đưa ra ước lượng cho năng suất trung bình của loại cà chua nói trên. Với độ tin cậy 95%.

- $n = 50$
- $S = 0,35 \text{ kg}$
- $\bar{X} = 1,48 \text{ kg}$
- $\alpha = 5\%$

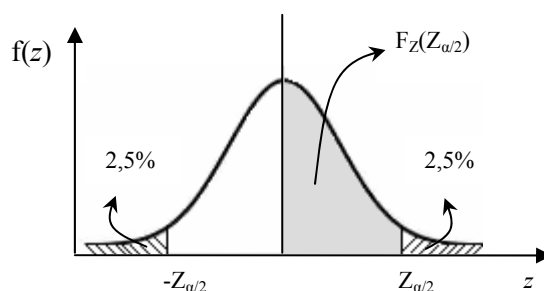
Ta có

$$\bar{X} - \frac{Z_{\alpha/2} S_x}{\sqrt{n}} < \mu < \bar{X} + \frac{Z_{\alpha/2} S_x}{\sqrt{n}}$$

$$P(Z > Z_{\alpha/2}) = \alpha/2 = 2,5\%$$

$$\Leftrightarrow 0,5 - F_Z(Z_{\alpha/2}) = 0,025$$

$$\Leftrightarrow F_Z(Z_{\alpha/2}) = 0,475$$



Tra bảng tìm  $Z_{\alpha/2} = 1,96$

Khoảng ước lượng năng suất trung bình của cà chua

$$\Leftrightarrow 1,48 - \frac{1,96 \times 0,35}{\sqrt{50}} < \mu < 1,48 + \frac{1,96 \times 0,35}{\sqrt{50}}$$

$$\Leftrightarrow 1,382 < \mu < 1,577 \text{ 5\%}$$

Vậy năng suất của cây cà chua xuân hè từ 1,382 kg/cây đến 1,577 kg/cây, với độ tin cậy 95%.

## 2. Ước lượng trung bình của tổng thể: trường hợp mẫu nhỏ ( $n < 30$ )

$$\boxed{\bar{X} - \frac{t_{n-1, \alpha/2} S_X}{\sqrt{n}} < \mu < \bar{X} + \frac{t_{n-1, \alpha/2} S_X}{\sqrt{n}}}$$

**VD 3.2:** Mẫu ngẫu nhiên về chiều dài của 6 quả xoài cát Hòa Lộc (cm):

18,6      18,4      19,2      20,8      19,4      20,5

Tìm khoảng tin cậy 90% đối với trung bình chiều dài của xoài cát Hòa Lộc. Giả sử rằng phân phối chiều dài của tất cả xoài cát Hòa Lộc là phân phối chuẩn.

Tìm trung bình mẫu và phương sai mẫu

	$X_i$	$(X_i - \bar{X})^2$
1	18,6	0,780
2	18,4	1,174
3	19,2	0,080
4	20,8	1,734
5	19,4	0,007
6	20,5	1,034
<b>Tổng</b>	<b>116,9</b>	<b>4,808</b>

$$\bar{X} = \frac{\sum X_i}{n} = \frac{116,9}{6} = 19,48$$

$$\text{Phương sai của mẫu } S_X^2 = \frac{\sum (X_i - \bar{X})^2}{n-1} = \frac{4,808}{5} = 0,9616$$

$$t_{n-1, \alpha/2} = t_{5, 5\%} = 2,015$$

Khoảng ước lượng về chiều dài trung bình của xoài cát Hòa Lộc (độ tin cậy 90%)

$$19,48 - \frac{2,015 \sqrt{0,9616}}{\sqrt{5}} < \mu < 19,48 + \frac{2,015 \sqrt{0,9616}}{\sqrt{5}}$$

$$20,36387 < \mu < 18,59612$$

### Ghi chú

Khoảng ước lượng càng hẹp thì độ chính xác càng cao

$\alpha$  càng bé, độ tin cậy càng cao thì khoảng ước lượng càng rộng

Cỡ mẫu lớn ( $n$  lớn) càng chính xác, khoảng ước lượng càng hẹp.

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

$$\bar{p} - Z_{\alpha/2} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} < p < \bar{p} + Z_{\alpha/2} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Trong đó  $\bar{p}$  là tỷ lệ của mẫu

$P$  là tỷ lệ của tổng thể

**VD 3.4:** Tiến hành kiểm tra 126 gói trái cây sấy còn 2 ngày sử dụng, phát hiện có 42 gói kém chất lượng. Hãy ước lượng khoảng tỷ lệ sản phẩm kém chất lượng với độ tin cậy 95% và 90%.

$$\bar{p} = \frac{42}{126} = 0,3$$

$$Z_{5\%} = 1,645$$

$$Z_{2,5\%} = 1,96$$

Khoảng ước lượng tỷ lệ sản phẩm kém chất lượng với độ tin cậy 95%

$$0,3 - 1,96 \sqrt{\frac{0,3(1-0,3)}{126}} < p < 0,3 + 1,96 \sqrt{\frac{0,3(1-0,3)}{126}}$$

$$0,22 < p < 0,38$$

Khoảng ước lượng tỷ lệ sản phẩm kém chất lượng với độ tin cậy 90%

$$0,3 - 1,645 \sqrt{\frac{0,3(1-0,3)}{126}} < p < 0,3 + 1,645 \sqrt{\frac{0,3(1-0,3)}{126}}$$

$$0,2328 < p < 0,3672$$

## Chương 4

### KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

- Kiểm định liên quan đến một tổng thể
  - + Kiểm định giả thuyết về tham số thống kê  $\theta$  ( $\mu, p, \sigma^2$ )
    - Kiểm định giả thuyết về số trung bình
    - Kiểm định giả thuyết về tỷ lệ
    - Kiểm định giả thuyết về phương sai
- Kiểm định liên quan đến hai tổng thể
  - + Kiểm định giả thuyết về sự khác biệt của tham số thống kê giữa hai tổng thể
  - + Kiểm định giả thuyết về sự khác biệt giữa giá trị trung bình của hai tổng thể: các mẫu độc lập.

Kiểm định giả thuyết là một quá trình thống kê nhằm cung cấp chứng cứ để ủng hộ hay bác bỏ một giả thuyết nào đó. Vì giả thuyết thống kê có thể đúng hoặc sai nên cần kiểm định, việc kiểm định này gọi là kiểm định thống kê vì nó dựa vào thông tin thực nghiệm của mẫu để kết luận.

#### I. Xây dựng giả thuyết không và giả thuyết thay thế

Trong thực tế, không phải lúc nào cũng dễ dàng đưa ra các giả thuyết không và giả thuyết đối. Vì vậy, chúng ta cần phải hết sức thận trọng để đảm bảo rằng các giả thuyết được xây dựng một cách phù hợp và kết luận kiểm định phải cung cấp các thông tin mà người đưa ra quyết định cần.

#### Kiểm định giả thuyết nghiên cứu

Thông thường dễ dàng chứng minh một điều gì đó là không thật (bác bỏ nó), hơn là chứng minh điều gì đó có thật. Vì vậy, muốn chứng minh điều gì, ta đề ra một giả thuyết ngược lại để bác bỏ. Hay giả thuyết nghiên cứu được phát biểu dưới dạng giả thuyết thay thế.

#### Kiểm định tính chính xác của lời phát biểu

Trong các tình huống liên quan đến việc đánh giá tính chính xác của một lời phát biểu hay tuyên bố nào đó, giả thuyết không thường được dựa trên giả định là lời phát biểu đó đúng.

Các giả thuyết không và giả thuyết thay thế là các giả thuyết trái ngược nhau. Chỉ có thể là giả thuyết không đúng hoặc giả thuyết thay thế đúng, không thể xảy ra trường hợp cả hai giả thuyết cùng đúng.

#### 1. Sai lầm loại I và sai lầm loại II

Ở trường hợp lý tưởng là quyết định đúng được thực hiện là

Chấp nhận giả thuyết đúng

Bác bỏ giả thuyết sai

Tuy nhiên, không phải lúc nào cũng quyết định chính xác chẳng hạn như bác bỏ một giả thuyết sai hay chấp nhận một giả thuyết đúng. Vì kiểm định của chúng ta dựa trên mẫu nên phải chấp nhận sai lầm, đó là bác bỏ một giả thuyết đúng hoặc chấp nhận một giả thuyết sai.

Giả thuyết		
Kết luận	Giả thuyết $H_0$ đúng	Giả thuyết $H_0$ sai

Chấp nhận $H_0$	Kết luận đúng $P = 1 - \alpha$	<b>Sai lầm loại II</b> Xác suất phạm sai lầm loại II là $P(II) = \beta$ $\beta$ là năng lực kiểm định
Bác bỏ $H_0$	<b>Sai lầm loại I</b> Xác suất phạm sai lầm loại I là $P(I) = \alpha$ $\alpha$ là mức ý nghĩa của kiểm định	Kết luận đúng $P = 1 - \beta$

Một giả thuyết chưa biết được là đúng hay sai. Một cách lý tưởng, chúng ta muốn giữ cho cả  $P(I)$  và  $P(II)$  càng bé càng tốt. Nhưng càng giảm  $P(I)$  thì làm tăng  $P(II)$ .

Thủ tục kiểm định giả thuyết là ấn định xác suất phạm sai lầm loại I tối đa trong mức độ cho phép (tùy theo tầm quan trọng của sai lầm loại I).

## 2. Miền bác bỏ và miền chấp nhận

Miền bác bỏ là miền chứa các giá trị làm cho giả thuyết  $H_0$  bị bác bỏ

Miền chấp nhận là miền chứa các giá trị giúp cho giả thuyết  $H_0$  không bị bác bỏ.

### Kiểm định một phía (One – Tailed Test)

Khi giả thuyết  $H_1$  có tính chất 1 phía thì việc kiểm định được gọi là kiểm định 1 phía (1 đuôi, 1 đầu).

$$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta > \theta_0 \end{cases} \quad \begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta < \theta_0 \end{cases} \quad \begin{cases} H_0: \theta \geq \theta_0 \\ H_1: \theta < \theta_0 \end{cases} \quad \begin{cases} H_0: \theta \leq \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$$

### Kiểm định hai phía (Two – Tailed Test)

Khi giả thuyết  $H_1$  có tính chất 2 phía thì việc kiểm định được gọi là kiểm định 2 phía.

$$\begin{cases} H_0: \theta > \theta_0 \\ H_1: \theta \neq \theta_0 \end{cases} \quad \begin{cases} H_0: \theta < \theta_0 \\ H_1: \theta \neq \theta_0 \end{cases} \quad \begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta \neq \theta_0 \end{cases}$$



### 3. Các bước kiểm định

Bước 1: Xây dựng giả thuyết

Bước 2: Định ra mức  $\alpha$

Bước 3: Xác định miền bác bỏ, miền chấp nhận

$$\left[ \begin{array}{l} \alpha \\ Z_{\alpha}, Z_{\alpha/2}, t_{n-1,\alpha}, t_{n-1,\alpha/2} \end{array} \right]$$

Bước 4: Tính toán các tham số thống kê

P-value

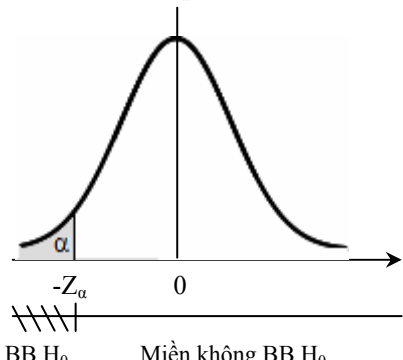
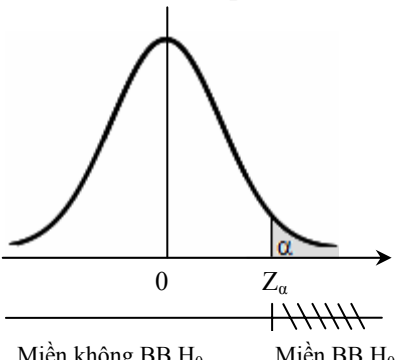
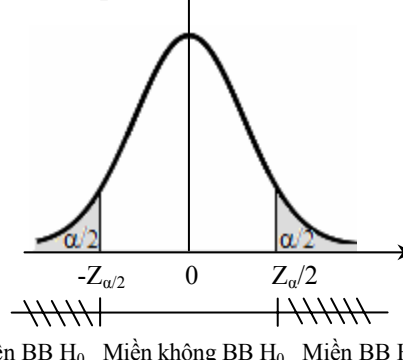
Giá trị thống kê  $Z_c, t_c$

Bước 5: Ra quyết định, nếu giá trị tính toán rơi vào miền bác bỏ  $H_0$  thì ra quyết định bác bỏ  $H_0$ . Ngược lại là không bác bỏ  $H_0$ .

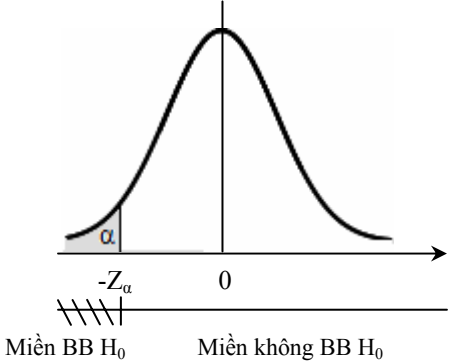
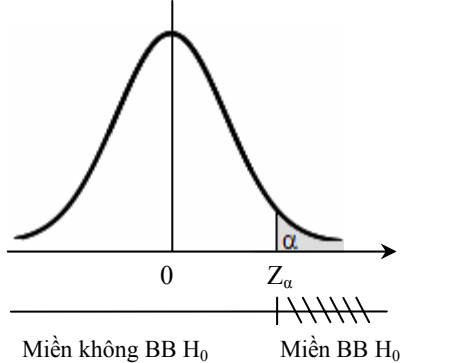
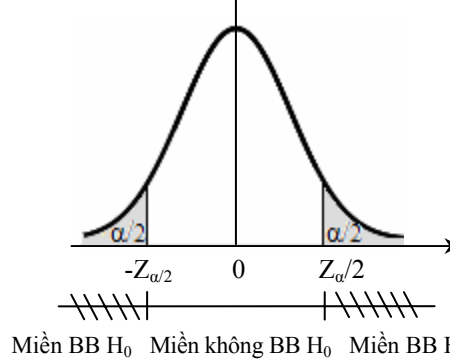
### 4. Kiểm định giả thuyết liên quan đến một tổng thể

#### 4.1. Kiểm định giá trị trung bình

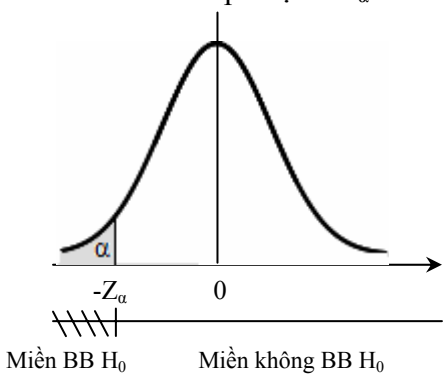
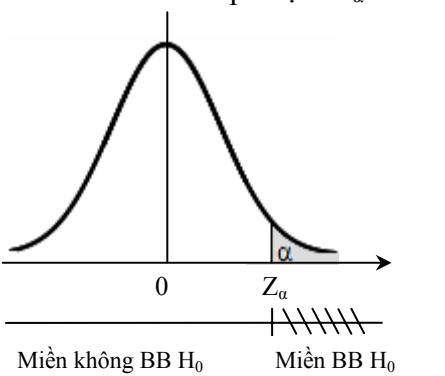
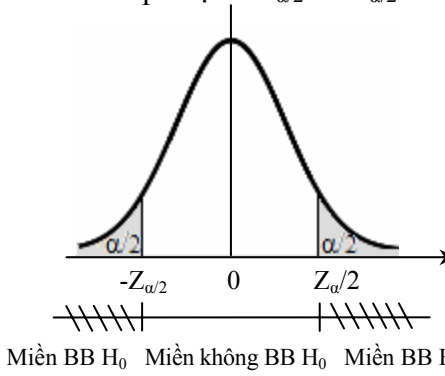
4.1.1. Kiểm định giá trị trung bình của tổng thể: trường hợp mẫu lớn và đã biết phương sai của tổng thể ( $\sigma^2$ )

	<b>Kiểm định 1 phía</b>		<b>Kiểm định 2 phía</b>
	Kiểm định 1 phía trái	Kiểm định 1 phía phải	
Bước 1	Thiết lập giả thuyết $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases} \quad \begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$	Thiết lập giả thuyết $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases} \quad \begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$	Thiết lập giả thuyết $\begin{cases} H_0: \mu > \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases} \quad \begin{cases} H_0: \mu < \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases} \quad \begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$
Bước 2	Xác định mức $\alpha$ (Xác suất phạm sai lầm khi bác bỏ giả thuyết $H_0$ )	Xác định mức $\alpha$	Xác định mức $\alpha$
Bước 3	Xác định giá trị giới hạn giữa miền bác bỏ và miền chấp nhận: $-Z_\alpha$ 	Xác định giá trị giới hạn giữa miền bác bỏ và miền chấp nhận: $Z_\alpha$ 	Xác định giá trị giới hạn giữa miền bác bỏ và miền chấp nhận: $-Z_{\alpha/2}$ và $Z_{\alpha/2}$ 
Bước 4	Tính toán giá trị thống kê $Z_c = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ Hay từ $Z_c$ tính P-value	Tính toán giá trị thống kê $Z_c = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ Hay từ $Z_c$ tính P-value	Tính toán giá trị thống kê $Z_c = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ Hay từ $Z_c$ tính P-value
Bước 5	Bác bỏ $H_0$ nếu $Z_c < -Z_\alpha$ Hay Bác bỏ $H_0$ nếu P-value $< \alpha$	Bác bỏ $H_0$ nếu $Z_c > Z_\alpha$ Hay Bác bỏ $H_0$ nếu P-value $< \alpha$	Bác bỏ $H_0$ nếu $Z_c < -Z_{\alpha/2}$ hay $Z_c > Z_{\alpha/2}$ Hay Bác bỏ $H_0$ nếu P-value $< \alpha$

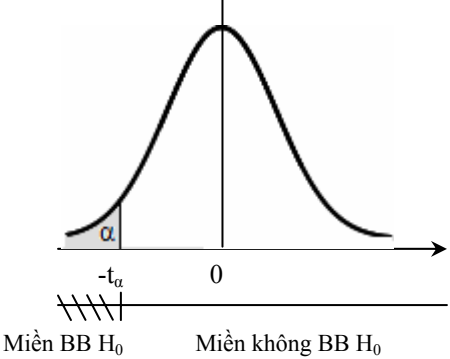
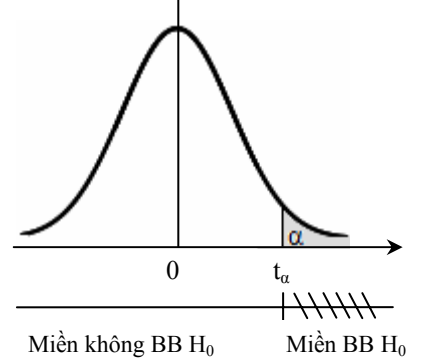
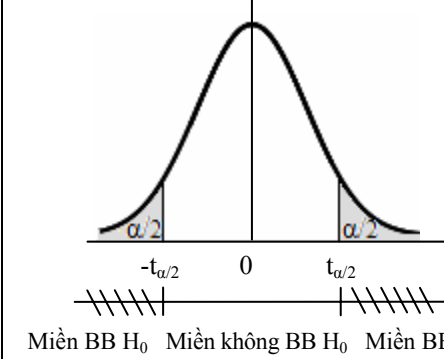
4.1.2. Kiểm định giá trị trung bình của tổng thể: trường hợp mẫu lớn và không biết phương sai của tổng thể ( $\sigma^2$ )

	<b>Kiểm định 1 phía</b>		<b>Kiểm định 2 phía</b>
	Kiểm định 1 phía trái	Kiểm định 1 phía phải	
Bước 1	Thiết lập giả thuyết $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases} \quad \begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$	Thiết lập giả thuyết $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases} \quad \begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$	Thiết lập giả thuyết $\begin{cases} H_0: \mu > \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases} \quad \begin{cases} H_0: \mu < \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases} \quad \begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$
Bước 2	Xác định mức $\alpha$	Xác định mức $\alpha$	Xác định mức $\alpha$
Bước 3	Xác định giá trị giới hạn giữa miền bác bỏ và miền chấp nhận: $-Z_\alpha$ 	Xác định giá trị giới hạn giữa miền bác bỏ và miền chấp nhận: $Z_\alpha$ 	Xác định giá trị giới hạn giữa miền bác bỏ và miền chấp nhận: $-Z_{\alpha/2}$ và $Z_{\alpha/2}$ 
Bước 4	Tính toán giá trị thống kê $Z_c = \frac{\bar{X} - \mu_0}{S_x / \sqrt{n}}$ Hay từ $Z_c$ tính P-value	Tính toán giá trị thống kê $Z_c = \frac{\bar{X} - \mu_0}{S_x / \sqrt{n}}$ Hay từ $Z_c$ tính P-value	Tính toán giá trị thống kê $Z_c = \frac{\bar{X} - \mu_0}{S_x / \sqrt{n}}$ Hay từ $Z_c$ tính P-value
Bước 5	Bác bỏ $H_0$ nếu $Z_c < -Z_\alpha$ Hay Bác bỏ $H_0$ nếu P-value $< \alpha$	Bác bỏ $H_0$ nếu $Z_c > Z_\alpha$ Hay Bác bỏ $H_0$ nếu P-value $< \alpha$	Bác bỏ $H_0$ nếu $Z_c < -Z_{\alpha/2}$ hay $Z_c > Z_{\alpha/2}$ Hay Bác bỏ $H_0$ nếu P-value $< \alpha$

### 4.1.3. Kiểm định giá trị trung bình của tổng thể: trường hợp mẫu nhỏ

	<b>Kiểm định 1 phía</b>		<b>Kiểm định 2 phía</b>
	Kiểm định 1 phía trái	Kiểm định 1 phía phải	
Bước 1	Thiết lập giả thuyết $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases} \quad \begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$	Thiết lập giả thuyết $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases} \quad \begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$	Thiết lập giả thuyết $\begin{cases} H_0: \mu > \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases} \quad \begin{cases} H_0: \mu < \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases} \quad \begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$
Bước 2	Xác định mức $\alpha$	Xác định mức $\alpha$	Xác định mức $\alpha$
Bước 3	Xác định giá trị giới hạn giữa miền bác bỏ và miền chấp nhận: $-Z_\alpha$ 	Xác định giá trị giới hạn giữa miền bác bỏ và miền chấp nhận: $Z_\alpha$ 	Xác định giá trị giới hạn giữa miền bác bỏ và miền chấp nhận: $-Z_{\alpha/2}$ và $Z_{\alpha/2}$ 
Bước 4	Tính toán giá trị thống kê $t_c = \frac{\bar{X} - \mu_0}{S_x / \sqrt{n}}$ Hay từ $t_c$ tính P-value	Tính toán giá trị thống kê $t_c = \frac{\bar{X} - \mu_0}{S_x / \sqrt{n}}$ Hay từ $t_c$ tính P-value	Tính toán giá trị thống kê $t_c = \frac{\bar{X} - \mu_0}{S_x / \sqrt{n}}$ Hay từ $t_c$ tính P-value
Bước 5	Bác bỏ $H_0$ nếu $t_c < -t_\alpha$ Hay Bác bỏ $H_0$ nếu P-value $< \alpha$	Bác bỏ $H_0$ nếu $t_c > t_\alpha$ Hay Bác bỏ $H_0$ nếu P-value $< \alpha$	Bác bỏ $H_0$ nếu $t_c < -t_{\alpha/2}$ hay $t_c > t_{\alpha/2}$ Hay Bác bỏ $H_0$ nếu P-value $< \alpha$

#### 4.2. Kiểm định về tỷ lệ của tổng thể

	<b>Kiểm định 1 phía</b>		<b>Kiểm định 2 phía</b>
	Kiểm định 1 phía trái	Kiểm định 1 phía phải	
Bước 1	Thiết lập giả thuyết $\begin{cases} H_0: p = p_0 \\ H_1: p < p_0 \end{cases} \quad \begin{cases} H_0: p \geq p_0 \\ H_1: p < p_0 \end{cases}$	Thiết lập giả thuyết $\begin{cases} H_0: p = p_0 \\ H_1: p > p_0 \end{cases} \quad \begin{cases} H_0: p \geq p_0 \\ H_1: p > p_0 \end{cases}$	Thiết lập giả thuyết $\begin{cases} H_0: p > p_0 \\ H_1: p \neq p_0 \end{cases} \quad \begin{cases} H_0: p < p_0 \\ H_1: p \neq p_0 \end{cases} \quad \begin{cases} H_0: p = p_0 \\ H_1: p \neq p_0 \end{cases}$
Bước 2	Xác định mức $\alpha$	Xác định mức $\alpha$	Xác định mức $\alpha$
Bước 3	Xác định giá trị giới hạn giữa miền bác bỏ và miền chấp nhận: $-t_\alpha$ 	Xác định giá trị giới hạn giữa miền bác bỏ và miền chấp nhận: $t_\alpha$ 	Xác định giá trị giới hạn giữa miền bác bỏ và miền chấp nhận: $-t_{\alpha/2}$ và $t_{\alpha/2}$ 
Bước 4	Tính toán giá trị thống kê $Z_c = \frac{(\bar{p} - p_0)\sqrt{n}}{p_0(1 - p_0)}$ Hay từ $Z_c$ tính P-value	Tính toán giá trị thống kê $Z_c = \frac{(\bar{p} - p_0)\sqrt{n}}{p_0(1 - p_0)}$ Hay từ $Z_c$ tính P-value	Tính toán giá trị thống kê $Z_c = \frac{(\bar{p} - p_0)\sqrt{n}}{p_0(1 - p_0)}$ Hay từ $Z_c$ tính P-value
Bước 5	Bác bỏ $H_0$ nếu $Z_c < -Z_\alpha$ Hay Bác bỏ $H_0$ nếu P-value $< \alpha$	Bác bỏ $H_0$ nếu $Z_c > Z_\alpha$ Hay Bác bỏ $H_0$ nếu P-value $< \alpha$	Bác bỏ $H_0$ nếu $Z_c < -Z_{\alpha/2}$ hay $Z_c > Z_{\alpha/2}$ Hay Bác bỏ $H_0$ nếu P-value $< \alpha$

**VD 4.1:** Trọng lượng đóng bao của các bao gạo là đại lượng phân phối chuẩn với trọng lượng trung bình theo quy định là 50kg. Có sự phản ánh của khách hàng là gạo bị đóng thiếu, người ta đem cân thử ngẫu nhiên 25 bao và thu được số liệu sau:

Trọng lượng bao (kg)	Số bao
48,0 – 48,5	2
48,5 – 49,0	5
49,0 – 49,5	10
49,5 – 50,0	6
50,0 – 50,5	2

Với mức ý nghĩa  $\alpha = 0,01$  hãy kết luận về điều phản ánh trên

**Bước 1:** 
$$\begin{cases} H_0: \mu = 50 \\ H_1: \mu < 50 \end{cases}$$

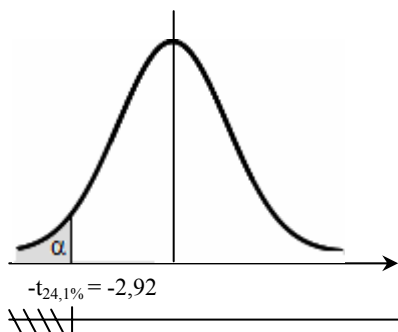
**Bước 2:** Xác suất phạm sai lầm tối đa khi bác bỏ giả thuyết  $H_0$  là  $\alpha = 1\%$

**Bước 3:** Xác định miền bác bỏ và miền chấp nhận

Kiểm định 1 phía trái

$$t_{24,1\%} = \text{TINV}(2\%, 24) = 2,492$$

$$\Rightarrow -t_{24,1\%} = -2,492$$



Miền BB  $H_0$

Miền không BB  $H_0$

**Bước 4:** Tính toán giá trị thống kê

$X_i$	$n$	$nX_i$	$n(X_i - \bar{X})^2$
48,25	2	96,5	2,0808
48,75	5	243,75	1,352
49,25	10	492,5	0,004
49,75	6	298,5	1,3824
50,25	2	100,5	1,9208
Tổng	25	1231,75	6,74

$$\bar{X} = \frac{1231,75}{25} = 49,27$$

$$S_x = \sqrt{\frac{\sum n(X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{6,74}{24}} = 0,52993$$

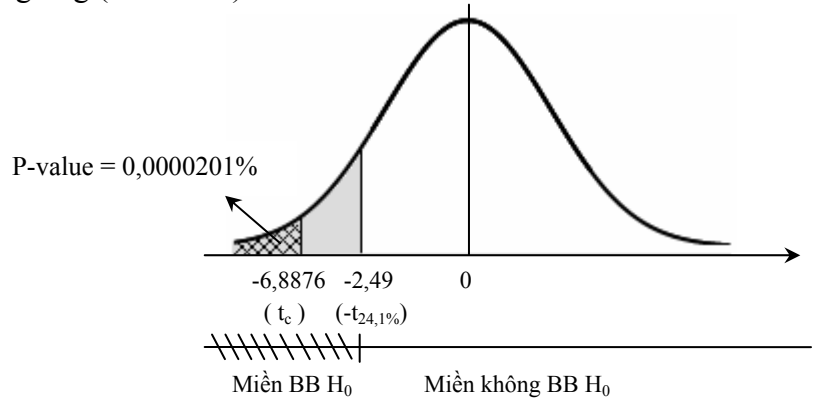
$$t_c = \frac{\bar{X} - \mu_0}{S_x / \sqrt{n}} = \frac{49,27 - 50}{0,52993 / \sqrt{25}} = -6,8876$$

Hay tính P-value

Với giá trị  $t = 6,8876$  tìm xác suất tương ứng (= P-value)

$$= \text{TDIST}(6,8876;24;1)$$

$$= 0,000000201$$



**Bước 5:** Ra quyết định

$t_c < -t_{24,1\%} \Rightarrow$  Bác bỏ giả thuyết  $H_0$

Hay  $P\text{-value} < \alpha \Rightarrow$  Bác bỏ giả thuyết  $H_0$

Khách hàng phản ánh đúng.

**VD 4.2:** Gieo 300 hạt đậu tương. Kết quả là 261 hạt nảy mầm. Người ta nói rằng tỷ lệ nảy mầm của đậu tương là 0,9. Điều nhận định đó có đúng không? Cho mức kiểm định  $\alpha=5\%$ .

**VD 4.3:** Một rừng chuẩn tuổi 10 có chiều cao trung bình là 7,8m với độ lệch chuẩn là 1,2m. Khảo sát chiều cao của 20 cây của một khu rừng ở tuổi 10 cho kết quả như sau:

7,1	7,1	7,0	7,2
7,3	8,1	7,0	7,4
8,5	8,2	7,3	8,6
7,1	6,9	8,2	6,9
6,3	8,0	7,1	7,3

Hỏi chiều cao của khu rừng này có đạt đến chiều cao của rừng chuẩn chưa?

## 5. Kiểm định giả thuyết liên quan đến hai tổng thể

Các giả thuyết

Kiểm định một phía

$$\begin{cases} H_0 & \mu_1 - \mu_2 \leq 0 \\ H_1 & \mu_1 - \mu_2 > 0 \end{cases} \quad \begin{cases} H_0 & \mu_1 - \mu_2 \geq 0 \\ H_1 & \mu_1 - \mu_2 < 0 \end{cases}$$

Kiểm định hai phía

$$\begin{cases} H_0 & \mu_1 - \mu_2 = 0 \\ H_1 & \mu_1 - \mu_2 \neq 0 \end{cases}$$

## 5.1. Kiểm định giả thuyết về sự khác biệt giữa giá trị trung bình của hai tổng thể: Các mẫu độc lập

Lấy mẫu độc lập là cách lấy mẫu trong đó tất cả số liệu của các mẫu (thí nghiệm) đều được lấy cùng một điều kiện hoàn cảnh như nhau; do vậy mà số liệu trong hai mẫu (thí nghiệm) có thể bằng nhau ( $n_1 = n_2$ ) hoặc khác nhau ( $n_1 \neq n_2$ ). Cách lấy mẫu độc lập (không bắt cặp) là trường hợp đặc biệt (chỉ có hai mẫu hay hai thí nghiệm) của cách lấy mẫu theo kiểu hoàn toàn ngẫu nhiên – CRD.

Trong kiểm định có thể phân biệt thành hai trường hợp

Biết phương sai của tổng thể

Không biết phương sai của tổng thể hay cỡ mẫu nhỏ.

### 5.1.1. Biết phương sai của tổng thể ( $\sigma_1^2$ và $\sigma_2^2$ )

Thông kê kiểm định

$$Z_c = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Ra quyết định

Nếu  $|Z_c| > Z_\alpha \Rightarrow$  Bác bỏ  $H_0$  (Kiểm định một phía)

Nếu  $|Z_c| > Z_{\alpha/2} \Rightarrow$  Bác bỏ  $H_0$  (Kiểm định hai phía)

Hay P-value  $< \alpha \Rightarrow$  Bác bỏ  $H_0$

Bác bỏ  $H_0$  nghĩa là có sự khác biệt về giá trị trung bình của hai tổng thể với mức ý nghĩa  $\alpha$ .

### 5.1.2. Trường hợp không biết phương sai của tổng thể hay cỡ mẫu nhỏ

#### a) Trường hợp giả thuyết $H_0: \sigma_1^2 = \sigma_2^2$ được chấp nhận

$H_0: \sigma_1^2 = \sigma_2^2$  được chấp nhận khi  $F_c < F_{\text{bậc tự do ở tử, bậc tự do ở mẫu}, \alpha}$  ( $F_{\text{bảng}}$ )

$$F_c = \frac{S_1^2}{S_2^2} \quad \text{nếu } S_1^2 > S_2^2 \quad (F_{\text{bậc tự do ở tử, bậc tự do ở mẫu}, \alpha} = F_{n_1-1, n_2-1, \alpha})$$

$$F_c = \frac{S_2^2}{S_1^2} \quad \text{nếu } S_2^2 > S_1^2 \quad (F_{\text{bậc tự do ở tử, bậc tự do ở mẫu}, \alpha} = F_{n_2-1, n_1-1, \alpha})$$

Trong trường hợp  $H_0: \sigma_1^2 = \sigma_2^2$  được chấp nhận chúng ta dùng trắc nghiệm t với  $(n_1 + n_2 - 2)$  độ tự do và  $t_c$  như sau:

$$t_c = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}}$$

Ra quyết định



Nếu  $|t_c| > t_{n_1+n_2-2, \alpha} \Rightarrow$  Bác bỏ  $H_0$  (Kiểm định một phía)

Nếu  $|t_c| > t_{n_1+n_2-2, \alpha/2} \Rightarrow$  Bác bỏ  $H_0$  (Kiểm định hai phía)

Hay  $P\text{-value} < \alpha \Rightarrow$  Bác bỏ  $H_0$

Bác bỏ  $H_0$  nghĩa là có sự khác biệt về giá trị trung bình của hai tổng thể với mức ý nghĩa  $\alpha$ .

**b) Trường hợp giả thuyết  $H_0: \sigma_1^2 = \sigma_2^2$  bị bác bỏ**

$H_0: \sigma_1^2 = \sigma_2^2$  bị bác bỏ khi  $F_c > F_{\text{bảng}}$  do ở từ, bậc tự do ở mẫu,  $\alpha$  ( $F_{\text{bảng}}$ )

$$F_c = \frac{S_1^2}{S_2^2} \quad \text{nếu } S_1^2 > S_2^2 \quad (F_{\text{bảng}} = F_{n_1-1, n_2-1, \alpha})$$

$$F_c = \frac{S_2^2}{S_1^2} \quad \text{nếu } S_2^2 > S_1^2 \quad (F_{\text{bảng}} = F_{n_2-1, n_1-1, \alpha})$$

Trong trường hợp  $H_0: \sigma_1^2 = \sigma_2^2$  bị bác bỏ chúng ta dùng trắc nghiệm t với K độ tự do và  $t_c$  như sau:

$$t_c = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad K = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}} \quad (\text{Bậc tự do})$$

Ra quyết định

Nếu  $|t_c| > t_{n_1+n_2-2, \alpha} \Rightarrow$  Bác bỏ  $H_0$  (Kiểm định một phía)

Nếu  $|t_c| > t_{n_1+n_2-2, \alpha/2} \Rightarrow$  Bác bỏ  $H_0$  (Kiểm định hai phía)

Hay  $P\text{-value} < \alpha \Rightarrow$  Bác bỏ  $H_0$

Bác bỏ  $H_0$  nghĩa là có sự khác biệt về giá trị trung bình của hai tổng thể với mức ý nghĩa  $\alpha$ .

**5.2. Kiểm định giả thuyết về sự khác biệt giữa giá trị trung bình của hai tổng thể: lấy mẫu theo cặp (bắt cặp, tạo khối)**

Số quan sát trong hai mẫu (thí nghiệm thức) luôn luôn bằng nhau ( $n_1 = n_2 = n$  cặp)

Hoàn cảnh	1	2	3	...	n
Mẫu 1 (Thí nghiệm thức 1)	$x_{11}$	$x_{12}$	$x_{13}$	...	$x_{1n}$
Mẫu 2 (Thí nghiệm thức 2)	$x_{21}$	$x_{22}$	$x_{23}$	...	$x_{2n}$

Cách lấy mẫu theo cặp là trường hợp đặc biệt (chỉ có 2 mẫu hay 2 thí nghiệm thức) của cách lấy mẫu theo khối (tạo khối, tạo khối nhất phương, khối đầy đủ, RCBD).

Thống kê kiểm định

$$t_c = \frac{|\bar{X}|}{\sqrt{S^2/n}} = \frac{|\bar{X}|\sqrt{n}}{S}$$

Trong đó

$$X_i = x_{1i} - x_{2i}$$

$$\bar{X} = \frac{\sum_{i=1}^n (x_{1i} - x_{2i})}{n} = \frac{\sum_{i=1}^n X_i}{n}$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Ra quyết định

Nếu  $|t_c| > t_{n_1+n_2-2, \alpha} \Rightarrow$  Bác bỏ  $H_0$  (Kiểm định một phía)

Nếu  $|t_c| > t_{n_1+n_2-2, \alpha/2} \Rightarrow$  Bác bỏ  $H_0$  (Kiểm định hai phía)

Hay  $P\text{-value} < \alpha \Rightarrow$  Bác bỏ  $H_0$

Bác bỏ  $H_0$  nghĩa là có sự khác biệt về giá trị trung bình của hai tổng thể với mức ý nghĩa  $\alpha$ .

## 6. Kiểm định sự khác biệt về tỷ lệ giữa hai tổng thể

Các giả thuyết

Kiểm định một phía

$$\begin{cases} H_0 & p_1 - p_2 \leq 0 \\ H_1 & p_1 - p_2 > 0 \end{cases} \quad \begin{cases} H_0 & p_1 - p_2 \geq 0 \\ H_1 & p_1 - p_2 < 0 \end{cases}$$

Kiểm định hai phía

$$\begin{cases} H_0 & p_1 - p_2 = 0 \\ H_1 & p_1 - p_2 \neq 0 \end{cases}$$

Thống kê kiểm định

$$Z_c = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{với} \quad \bar{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2}$$

Ra quyết định

Nếu  $|Z_c| > Z_\alpha \Rightarrow$  Bác bỏ  $H_0$  (Kiểm định một phía)

Nếu  $|Z_c| > Z_{\alpha/2} \Rightarrow$  Bác bỏ  $H_0$  (Kiểm định hai phía)

Hay  $P\text{-value} < \alpha \Rightarrow$  Bác bỏ  $H_0$

Bác bỏ  $H_0$  nghĩa là có sự khác biệt về tỷ lệ của hai tổng thể với mức ý nghĩa  $\alpha$ .

## Chương 5

### BỐ TRÍ THÍ NGHIỆM

#### I. Đại cương về bố trí thí nghiệm

##### 1. Yếu tố (Factor)

Yếu tố là biến số độc lập cần nghiên cứu, có thể là biến định lượng hay định tính.

**VD 5.1:** Nghiên cứu ảnh hưởng của thời gian bảo quản (Yếu tố A), xử lý hóa chất (Yếu tố B) và cách bảo quản (Yếu tố C) đến thời gian trung bày của hoa hồng.

Thời gian trung bày =  $f(\text{Thời gian bảo quản, Xử lý hóa chất, Cách bảo quản}) = f(\text{Yếu tố A, Yếu tố B, Yếu tố C})$

Biến phụ thuộc: Thời gian trung bày của hoa hồng (ngày)

Biến độc lập: Thời gian bảo quản  
Xử lý hóa chất  
Cách bảo quản

##### 2. Mức (Level)

Mức là một loại hình hay một trị số của biến độc lập

**VD 5.2:** So sánh thời gian trung bày của hoa hồng

Biến số A	Thời gian bảo quản		
Biến số B	Xử lý hóa chất		
Biến số C	Cách bảo quản		
Yếu tố A, có 3 mức	Mức 1	7 ngày	Ký hiệu $a_1$
	Mức 2	14 ngày	Ký hiệu $a_2$
	Mức 3	21 ngày	Ký hiệu $a_3$
Yếu tố B, có 3 mức	Mức 1	Công thức 1	Ký hiệu $b_1$
	Mức 2	Công thức 2	Ký hiệu $b_2$
	Mức 3	Công thức 3	Ký hiệu $b_3$
Yếu tố C, có 2 mức	Mức 1	Ướt lạnh	Ký hiệu $c_1$
	Mức 2	Khô lạnh	Ký hiệu $c_2$

##### 3. Nghiệm thức (Treatment)

Nghiệm thức là tổ hợp các mức yếu tố

**VD 5.3:** Thí nghiệm về thời gian trung bày của hoa hồng có 3 yếu tố thì nghiệm thức là tổ hợp các mức yếu tố A, B, C.

$$\begin{aligned} \text{Số nghiệm thức} &= \text{Số mức yếu tố A} \times \text{Số mức yếu tố B} \times \text{Số yếu tố C} \\ &= 3 \times 3 \times 2 = 18 \text{ (nghiệm thức)} \end{aligned}$$

Chẳng hạn

(7 ngày  $\times$  Công thức 1  $\times$  Ướt lạnh) hay  $(a_1b_1c_1)$

.....

(21 ngày  $\times$  Công thức 3  $\times$  Khô lạnh) hay  $(a_3b_3c_2)$

#### 4. Đơn vị thí nghiệm

Đơn vị thí nghiệm là một lần lặp lại của một nghiệm thức

Tổng số đơn vị thí nghiệm = Số nghiệm thức  $\times$  Số lần lặp lại

Đơn vị thí nghiệm có thể mang ý nghĩa trong không gian như các ô ruộng trồng lúa hay thời gian các lần nấu, các lần đo...

#### 5. Nguyên tắc cơ bản về bố trí thí nghiệm

Nguyên tắc 1: Lặp lại (Replication)

Thí nghiệm phải được lặp lại nhiều lần. Lần lặp lại có nghĩa về thời gian hay không gian.

Nguyên tắc 2: Ngẫu nhiên hóa (Randomization)

Chọn mẫu ngẫu nhiên để tránh thành kiến của người làm thí nghiệm, để các tính toán có giá trị vì bản chất của xác suất là sự ngẫu nhiên.

### II. Bố trí thí nghiệm một yếu tố

Là thí nghiệm mà trong đó chỉ có một yếu tố thay đổi, trong lúc các yếu tố khác được giữ bằng cố định. Bằng cách thay đổi các mức khác nhau của yếu tố, ta có các nghiệm thức khác nhau (Treatment).

Có 3 kiểu bố trí thí nghiệm thông dụng

Kiểu hoàn toàn ngẫu nhiên (CRD – Complete Random Design)

Kiểu khối đầy đủ (RCBD – Randomized Complete Block Design)

Kiểu ô vuông Latinh (LS – Latin Squared Design)

#### 1. Kiểu hoàn toàn ngẫu nhiên

Đây là kiểu thí nghiệm cơ bản, dễ phân tích và bố trí. Nó được ứng dụng khi điều kiện ngoại cảnh thật đồng nhất với nhau. Đây là điều ít xảy ra trong thực tế sản xuất. Vì vậy CRD thích hợp trong các phòng thí nghiệm, nơi có những điều kiện có thể kiểm soát được.

**VD 5.4:** Muốn kiểm tra ảnh hưởng của thời gian nấu đến hàm lượng Vitamin C có trong thực phẩm. Biết rằng thời gian nấu từ 15 phút đến 35 phút thì sản phẩm chính đủ để ăn được với một quy trình nấu như nhau (nhiệt độ, tỷ lệ nước...). Nhà nghiên cứu quyết định bố trí một thí nghiệm kiểu CRD với 5 thời gian nấu (5 nghiệm thức) là 15, 20, 25, 30 và 35 phút. Mỗi nghiệm thức có 5 lần lặp lại. Như vậy tổng cộng có  $5 \times 5$  (đơn vị thí nghiệm). Các thí nghiệm này phải tiến hành theo thứ tự ngẫu nhiên. Chúng ta lập bảng 5.1 như sau:

**Bảng 5.1:** Gán đặt số cho thí nghiệm CRD

Thời gian nấu (phút)	Số của thí nghiệm (Thí nghiệm số)
----------------------	-----------------------------------

15	1	2	3	4	5
20	6	7	8	9	10
25	11	12	13	14	15
30	16	17	18	19	20
35	21	22	23	24	25

Để tiến hành một cách ngẫu nhiên, ta sẽ phát số ngẫu nhiên từ 1 đến 25 và tiến hành trình tự theo thứ tự phát số ngẫu nhiên đó.

Sử dụng hàm =**RANDBETWEEN(1;25)** trong Excel

**Bảng 5.2:** Bố trí thí nghiệm CRD

Thứ tự của thí nghiệm	Số của thí nghiệm	Thời gian nấu
1	8	20
2	18	30
3	10	20
4	23	35
5	17	30
6	5	15
7	14	25
8	6	20
9	15	25
10	20	30
11	9	20
12	4	15
13	12	25
14	7	20
15	1	15
16	24	35
17	21	35
18	11	25
19	2	15
20	13	25
21	22	35
22	16	30
23	25	35
24	19	30
25	3	15

Như vậy thí nghiệm đầu tiên là thí nghiệm số 8, thời gian nấu 20 phút. Quá trình này cứ tiếp diễn cho đến khi hết tất cả các thí nghiệm.

## 2. Kiểu khối ngẫu nhiên đầy đủ

Trong nhiều trường hợp, do điều kiện ngoại cảnh tác động làm sai số của thí nghiệm tăng lên. Lúc này phải bố trí làm sao giảm được ảnh hưởng của yếu tố ngoại cảnh này.

Trong bố trí CRD yêu cầu khu thí nghiệm phải đồng nhất, điều này khó thực hiện trong thực tế sản xuất.

“Khối” có thể có ý nghĩa theo không gian hoặc thời gian.

**VD 5.5:** Khi thí nghiệm ngoài đồng ruộng, muốn tìm ảnh hưởng của 6 loại phân bón lên năng suất lúa ở đám ruộng gần bờ sông, nhưng do bề mặt đồng ruộng không bằng phẳng (có độ dốc) nên hàm lượng nước trong đất sẽ khác nhau. Điều kiện ngoại cảnh này sẽ ảnh hưởng đến kết quả năng suất làm cho thí nghiệm không chính xác. Trong trường hợp này ta bố trí kiểu khối, mỗi khối chứa đầy đủ tất cả các loại phân bón (thí nghiệm  $A_1$  đến  $A_6$ ) và có cùng khoảng cách từ cây đến bờ sông. Các thí nghiệm này được bố trí một cách ngẫu nhiên trong khối. Đây là thí nghiệm một yếu tố (phân bón) nhưng có 2 ảnh hưởng đến kết quả năng suất, ảnh hưởng thứ nhất (thí nghiệm phân bón), ảnh hưởng thứ hai (độ dốc mặt ruộng được bố trí thành khối). Sau này phân tích phương sai, ta sẽ xét ảnh hưởng cả thí nghiệm và khối lên kết quả.

**Hình 5.1:** Bố trí kiểu khối

Độ dốc ↓	Khối ↓	1	A1	A2	A3	A4	A5	A6
		2	A2	A1	A4	A3	A6	A5
		3	A4	A6	A3	A1	A5	A2
		4	A6	A5	A2	A4	A3	A1

**VD 5.6:** Kiểm tra ảnh hưởng của kích thước đầu nén lên kết quả đo độ cứng của bánh nướng. Ta tiến hành như sau: dùng 4 cỡ kích thước, đầu nén để đo, mỗi đầu nén sẽ kiểm tra cho bánh ở 4 vị trí trong lò nướng (hoặc 4 mẫu bánh). Tổng cộng  $4 \times 4 = 16$  thí nghiệm. Tuy nhiên, nếu dùng CRD sẽ có sai số do ảnh hưởng của nhiệt độ lò nướng phân bố không đều lên bánh nướng. Ta dùng bố trí kiểu RCBD để giảm bớt ảnh hưởng của sự biến động đo được đo bởi một đầu nén. Ta có bảng kết quả đo như sau:

**Bảng 5.3:** Độ cứng đo được qua thí nghiệm RCBD

		Loại đầu đo			
		A	B	C	D
Khối ↓ (Lặp lại)	Vị trí bánh nướng				
	1	9,3	9,4	9,2	9,7
	2	9,4	9,3	9,4	9,6
	3	9,6	9,8	9,5	10,0
	4	10,0	9,9	9,7	10,2

Như vậy mỗi khối đều có đầy đủ (Complete) tất cả các thí nghiệm (mỗi thí nghiệm chính là lặp lại trong khối). Mỗi thí nghiệm được lặp lại 4 lần ở 4 không gian khác nhau. Bằng cách này các khối (Bánh) tạo thành một đơn vị thí nghiệm đồng nhất trên đó có sự so sánh của 4 đầu đo. Như vậy cách bố trí RCBD sẽ hoàn thiện độ chính xác về so sánh giữa các đầu đo vì đã loại đi sự biến động giữa vị trí các bánh nướng. Tương tự sự khác biệt giữa các mẻ nướng công nhân và thời gian cũng có thể ảnh hưởng đến kết quả và sẽ được kiểm soát bằng cách bố trí khối.

**VD 5.7:** Khối theo thời gian

Thí nghiệm đo năng suất máy đập lúa ở 6 tốc độ khác nhau (6 nghiệm thức). Mỗi buổi (sáng và chiều) có thể thử được 6 lần. Vậy coi khối là 6 lần thử trong mỗi buổi vì lúa cắt cùng buổi sẽ có độ ẩm đồng đều hơn.

Năng suất đập lúa =  $f(\text{tốc độ máy})$ . Chỉ có một yếu tố tác động tuy nhiên buổi là yếu tố ngoại cảnh không giống nhau nên buổi cũng có thể xem là yếu tố thứ hai tác động đến yếu tố đầu ra.

Trong mỗi buổi thứ tự thực hiện các nghiệm thức được bố trí ngẫu nhiên. Một kết quả bốc thăm với 4 lần lặp lại có thể là:

(3 4 6 1 5 2)	(2 6 1 5 4 3)	(6 3 1 2 5 4)	(5 1 6 4 3 2)
Sáng ngày I	Chiều ngày I	Sáng ngày II	Chiều ngày II

So sánh với CRD kết quả có thể là

<b>(6 6 3 2 1 3)</b>	<b>(2 1 4 5 6 2)</b>	(3 2 4 5 4 1)	<b>(5 4 6 5 1)</b>
Sáng ngày I	Chiều ngày I	Sáng ngày II	Chiều ngày II

Bố trí theo kiểu khối để giảm sai số ngẫu nhiên giữa các khối tạo cơ hội đồng đều hơn khi so sánh các nghiệm thức.

Đây là một trong những phương pháp thí nghiệm áp dụng rộng rãi trong nghiên cứu.

### 3. Kiểu ô vuông La tinh (LS – Latin Squared Design)

Bố trí này được áp dụng trong trường hợp có 2 yếu tố ngoại cảnh ảnh hưởng đến kết quả của thí nghiệm.

Trong VD 5.6 đo độ cứng của bánh nướng ngoài tác động của phân bố nhiệt còn có tác động của công nhân nướng bánh. Lúc này thiết kế thí nghiệm phù hợp nhất là mỗi đầu đo thực hiện cho một vị trí bánh của cả 4 người công nhân đó là bố trí kiểu bình phương Latinh.

Độ cứng bánh =  $f(\text{Kích thước đầu đo})$ . Có 2 yếu tố ngoại cảnh phân bố nhiệt và công nhân nướng bánh.

Do đó phải bố trí ngẫu nhiên theo vị trí nướng và theo công nhân.

**Bảng 5.4:** Độ cứng bánh đo được qua thí nghiệm Latinh bình phương

Phân phối đầy đủ theo vị trí	Phân phối đầy đủ theo công nhân				
	Vị trí bánh	Công nhân nướng			
		1	2	3	4
	1	A = 9,7	B = 9,5	C = 9,6	D = 10,2
	2	B = 9,2	C = 9,1	D = 9,2	A = 9,8
3	C = 9,6	D = 9,8	A = 9,1	B = 9,4	
4	D = 10,4	A = 10,0	B = 9,6	C = 10,2	

Nói khác hơn bố trí bình phương Latinh là bố trí khối ngẫu nhiên đầy đủ hai chiều. Một bố trí Latinh bình phương cho p yếu tố là một hình vuông chứa p hàng và p cột.

**Bảng 5.5:** Các bố trí Latinh bình phương

4×4	5×5	6×6
ABDC	ADBEC	ADCEBF
BCAD	DACBE	BAECFD
CDBA	CBEDA	CEDFAB
DACB	BEACD	DCFBEA
	ECDAB	FBADCE
		EFBADC

#### 4. Quy trình bố trí thí nghiệm 1 yếu tố kiểu khối ngẫu nhiên đầy đủ và kiểu ô vuông Latinh

##### 4.1. Kiểu khối ngẫu nhiên đầy đủ

Xét lại ví dụ ảnh hưởng của kích thước đầu nén lên kết quả đo độ cứng của bánh nướng trình bày ở VD 5.6 có thể tóm tắt như sau:

Yếu tố tác động (đang nghiên cứu) có thể kiểm soát được: kích thước đầu nén.

Yếu tố ảnh hưởng không thể (hoặc khó) có thể kiểm soát được – yếu tố ngoại cảnh: sự phân bố nhiệt của lò nướng.

Sự phân bố nhiệt này làm cho độ cứng của bánh không đều sau khi nướng và gây sai số khi thí nghiệm về ảnh hưởng của kích thước đầu nén. Nếu ta bố trí theo kiểu CRD thì sẽ có một số đầu nén bị “thiên vị” (bias). Để tránh hiện tượng thiên vị này mỗi kích thước đầu nén đều được bố trí để đo ở tất cả các vị trí của bánh. Mỗi vị trí của bánh được xem là một khối.

Cách bố trí:

**Bước 1:** Bố trí khối, số khối bằng số lần lặp lại của nghiệm thức



**Bước 2:** Ngẫu nhiên hóa các thí nghiệm trong một khối

Vị trí (khối)	1	2	3	4
	↓	↓	↓	↓
	NT1	NT1	NT1	NT1
	NT2	NT2	NT2	NT2
	NT3	NT3	NT3	NT3
	NT4	NT4	NT4	NT4

Bước 1: Bố trí khối

Vị trí (khối)	1	2	3	4
	↓	↓	↓	↓
	NT1	NT2	NT4	NT1
	NT3	NT1	NT2	NT2
	NT2	NT4	NT3	NT3
	NT4	NT3	NT1	NT4

Bước 2: Ngẫu nhiên hóa

#### 4.2. Kiểu Latinh bình phương

Trong trường hợp có hai yếu tố ngoại cảnh tác động lên kết quả của thí nghiệm. Giảm sự thiên lệch thì hai yếu tố ngoại cảnh đó được loại trừ bằng cách bố trí khối theo 2 chiều.

Cách bố trí:

Bước 1: Bố trí khối theo hai yếu tố ngoại cảnh

Bước 2: Ngẫu nhiên hóa theo hàng

Bước 3: Ngẫu nhiên hóa theo cột

#### VD 5.8:

Hai yếu tố ngoại cảnh là vị trí bánh nướng và thao tác của công nhân.

Bước 1: Bố trí khối theo hai yếu tố ngoại cảnh

Vị trí	1	2	3	4
	↓	↓	↓	↓
CN 1 →	A	B	C	D
2 →	B	C	D	A
3 →	C	D	A	B
4 →	D	A	B	C

Bước 2: Ngẫu nhiên hóa theo hàng

Vị trí	1	2	3	4	<b>Hàng</b>
CN 1 →	C	D	A	B	<b>1</b>
2 →	D	A	B	C	<b>2</b>
3 →	B	C	D	A	<b>3</b>
4 →	A	B	C	D	<b>4</b>

Ngẫu nhiên hóa theo hàng sắp xếp lại các hàng một cách một ngẫu nhiên.

Phát 4 số ngẫu nhiên sau đó xếp hạng.

	A	B	C	D	E	F	G	H
1	<b>NGẪU NHIÊN HÓA THEO HÀNG</b>							
2	Hàm phát số ngẫu nhiên	Phát số ngẫu nhiên	Xếp hạng		=RANK(B3:\$B\$3:\$B\$6;1)			
3		431	399	3	Hàng thứ 3 trở thành hàng đầu tiên			
4	=RANDBETWEEN(1;999)	21	218	1	Hàng thứ 1 trở thành hàng thứ 2			
5		903	271	2	Hàng thứ 2 trở thành hàng thứ 3			
6		962	805	4	Hàng thứ 4 không thay đổi			
7								
8								
9								
10								
11								
12								
13								
14								
15	Lưu ý: Sử dụng hàm <b>Randbetween</b> thì vào Tool/Add-Ins chọn Analysis ToolPak							

Để cố định các số ngẫu nhiên vừa phát ra ở cột A, ta dùng lệnh Copy/Paste Special/Values để dán giá trị sang cột khác.

Xếp hạng

**Bước 3: Ngẫu nhiên hóa theo cột**

Vị trí	1	2	3	4
	↓	↓	↓	↓
CN 1	C	D	A	B
2	D	A	B	C
3	B	C	D	A
4	A	B	C	D

	A	B	C	D	E	F
1	<b>NGẪU NHIÊN HÓA THEO CỘT</b>					
2	Phát số ngẫu nhiên	454	438	281	933	
3	Copy số ngẫu nhiên đã phát Paste Special	857	117	893	824	
4	Xếp hạng		3	1	4	2
5		=RANK(B3;\$B\$3:\$E\$3;1)				
6						Cột 3 trở thành cột đầu tiên
7						Cột 1 trở thành cột thứ 2
8						Cột 4 trở thành cột thứ 3
9						Cột 2 trở thành cột thứ 4
10						

## Chương 6

### PHÂN TÍCH PHƯƠNG SAI MỘT CHIỀU CỦA THÍ NGHIỆM MỘT YẾU TỐ

Khi so sánh 2 tổng thể ta dùng trắc nghiệm t hay Z. Tuy nhiên khi chúng ta bố trí thí nghiệm theo các kiểu CRD, RCBD hay ô vuông Latinh thì phải so sánh nhiều hơn hai tổng thể (số nghiệm thức thường lớn hơn 2). Trong trường hợp này trắc nghiệm F được áp dụng.

Đối với thí nghiệm một yếu tố phân tích phương sai một chiều sẽ áp dụng cho kiểu CRD và phân tích phương sai hai chiều sẽ áp dụng cho kiểu bố trí RCBD và phương sai ba chiều cho kiểu ô vuông Latinh.

Phân tích phương sai một chiều

Bước 1: Lập bảng số liệu

Bước 2: Đặt giả thuyết

Bước 3: Tính toán và lập bảng ANOVA và LSD

#### I. Sắp xếp số liệu

Sắp xếp số liệu theo kiểu CRD

**Bảng 6.1:** Hàm lượng Vitamin C (mg/kg) có trong thực phẩm ở các thời gian nấu khác nhau (thí nghiệm CRD)

Số lần lặp lại	Thời gian nấu (phút)					
	15	20	25	30	35	
1	14	19	12	7	7	
2	18	25	17	10	7	
3	18	22	12	11	15	
4	19	19	18	15	11	
5	19	23	18	11	9	
Tổng	88	108	77	54	49	$\sum T_i = 376$
Trung bình	17,6	21,6	15,4	10,8	9,8	$\bar{Y} = 15,04$

Như vậy bình quân toàn bộ thí nghiệm  $\bar{Y} = 15,04$  mg/kg. Bình quân từng nghiệm thức biến động từ 9,8 mg/kg đến 21,6 mg/kg.

**Bảng 6.2:** Bảng số liệu thí nghiệm một yếu tố kiểu CRD

Số lần lặp lại	Các nghiệm thức					
	1	2	...	j	k	
1	$Y_{11}$	$Y_{12}$	...	$Y_{1j}$	$Y_{1k}$	
2	$Y_{21}$	$Y_{22}$	...	$Y_{2j}$	$Y_{2k}$	
...	...	...	...	...	...	
i	$Y_{i1}$	$Y_{i2}$	...	$Y_{ij}$	$Y_{ik}$	
...	...	...	...	...	...	
n	$Y_{n1}$	$Y_{n2}$	...	$Y_{nj}$	$Y_{nk}$	
Tổng	$T_1$	$T_2$	...	$T_j$	$T_k$	T
Trung bình	$\bar{Y}_1$	$\bar{Y}_2$	...	$\bar{Y}_j$	$\bar{Y}_k$	$\bar{Y}$

Mỗi giá trị trong bảng 6.2 là tổng của các thành phần sau:

$$Y_{ij} = \mu + v_{ij} + \varepsilon_{ij}$$

Trong đó

$\mu$  Trung bình thực của tổng thể

$v_{ij}$  Ảnh hưởng của nghiệm thức. Sự khác biệt giữa trung bình của nghiệm thức j so với trung bình toàn bộ ( $v_{ij} = \bar{Y}_j - \bar{Y}$ ).

$\varepsilon_{ij}$  Sai số ngẫu nhiên. Sai khác giữa các quan sát trong một nghiệm thức với trung bình của nghiệm thức đó ( $\varepsilon_{ij} = Y_{ij} - \bar{Y}_j$ )

$$\sum \sum (Y_{ij} - \bar{Y})^2 = \sum \sum (Y_{ij} - \bar{Y}_j)^2 + \sum \sum (\bar{Y}_j - \bar{Y})^2 \quad \text{Với} \left( \sum \sum = \sum_{j=1}^k \sum_{i=1}^n \right)$$

$SST_0$	=	SSE	+	SST
---------	---	-----	---	-----

$SST_0$  Tổng bình phương toàn bộ

SSE Tổng bình phương sai số ngẫu nhiên

SST Tổng bình phương nghiệm thức

Nếu tất cả trung bình nghiệm thức bằng nhau  $SST=0$

Nếu các trung bình càng khác nhau thì SST càng lớn.

## II. Bảng phân tích phương sai (ANOVA)

Nguồn biến động <i>Source of variation</i>	Độ tự do <i>Degree of freedom</i>	Tổng bình phương <i>Sum of Square</i>	Trung bình bình phương <i>Mean Square</i>	$F_c$	$F_{k-1, n-k, \alpha}^*$ ( $F_{\text{bảng}}$ )
Nghiệm thức (Treatment)	$k - 1$	SST	MST	$\frac{MST}{MSE}$	Tra bảng
Sai số (Error)	$N - k$	SSE	MSE		
Tổng (Total)	$N - 1$	$SST_0$			

Với  $F_{\text{bảng}} = F_{k-1, N-k, \alpha}^* = \text{FINV}(\alpha, k-1, N-k)$

k số nghiệm thức

N số thí nghiệm = số lần lặp lại  $\times$  số nghiệm thức

Phân tích phương sai (trắc nghiệm F) cho biết trong số các trung bình nghiệm thức có khác nhau không? Vấn đề suy diễn kế tiếp là nghiệm thức nào khác với nghiệm thức nào?

Giả thuyết

$$H_0 \quad \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_1 \quad \mu_i \neq \mu_j \text{ có ít nhất 1 cặp (ij) khác nhau}$$

Nếu  $F_c > F_{k-1, N-k, \alpha}^*$  thì bác bỏ giả thuyết  $H_0$ . Có ít nhất trung bình của hai nghiệm thức khác nhau ở mức ý nghĩa  $\alpha$ .

Nếu  $F_c < F_{k-1, N-k, \alpha}^*$  thì không bác bỏ giả thuyết  $H_0$ . Trung bình của các nghiệm thức đều bằng nhau ở mức ý nghĩa  $\alpha$ . Hay các yếu tố không ảnh hưởng đến thí nghiệm.

**VD 6.1:** Sử dụng số liệu của Bảng 6.1

Hàm lượng vitamin C = f(Thời gian nấu)

$$\begin{aligned} SST &= \sum_{j=1}^k \sum_{i=1}^n (\bar{Y}_j - \bar{Y})^2 = \sum_{j=1}^k n_j (\bar{Y}_j - \bar{Y})^2 \\ &= 5(17,6 - 15,04)^2 + 5(21,6 - 15,04)^2 + 5(15,4 - 15,04)^2 + 5(10,8 - 15,04)^2 + 5(9,8 - 15,04)^2 \\ &= 475,76 \end{aligned}$$

$$MST = \frac{SST}{k-1} = \frac{475,76}{4} = 118,94$$

$$\begin{aligned} SSE &= \sum \sum (Y_{ij} - \bar{Y}_j)^2 \\ &= (14 - 17,6)^2 + (18 - 17,6)^2 + (18 - 17,6)^2 + (19 - 17,6)^2 + (19 - 17,6)^2 \\ &\quad + (19 - 21,6)^2 + (25 - 21,6)^2 + (22 - 21,6)^2 + (19 - 21,6)^2 + (23 - 21,6)^2 \\ &\quad + (12 - 15,4)^2 + (17 - 15,4)^2 + (12 - 15,4)^2 + (18 - 15,4)^2 + (18 - 15,4)^2 \\ &\quad + (7 - 10,8)^2 + (10 - 10,8)^2 + (11 - 10,8)^2 + (15 - 10,8)^2 + (11 - 10,8)^2 \\ &\quad + (7 - 9,8)^2 + (7 - 9,8)^2 + (15 - 9,8)^2 + (11 - 9,8)^2 + (9 - 9,8)^2 \\ &= 161,2 \end{aligned}$$

$$MSE = \frac{SSE}{N-k} = \frac{161,2}{20} = 8,06$$

$$F_c = \frac{MST}{MSE} = \frac{118,94}{8,06} = 14,75682 \quad P\text{-value} = 9,12795E-06 = 0,00000912795 \approx 0,000913\%$$

$$= \text{FDIST}(14,75682; 4; 20)$$

$$F_{4,20,1\%}^* = \text{FINV}(1\%, 4, 20) = 4,43$$

$$F_{4,20,5\%}^* = \text{FINV}(5\%, 4, 20) = 2,866$$

$F_c > F_{\text{bảng}}$ . Bác bỏ giả thuyết  $H_0$ . Trung bình giữa các nghiệm thức khác biệt có ý nghĩa ở mức 99% nghĩa là thời gian nấu (từ 15 phút đến 35 phút) ảnh hưởng rất lớn đến hàm lượng Vitamin C có trong sản phẩm.

## Thực hiện trong Excel

### Tool/Data Analysis/Anova: Single Factor

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	Số lần lặp lại	Thời gian nấu (phút)								Anova: Single Factor						
2		15	20	25	30	35										
3	1	14	19	12	7	7			SUMMARY							
4	2	18	25	17	10	7			<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>			
5	3	18	22	12	11	15			15	5	88	17,6	4,3			
6	4	19	19	18	15	11			20	5	108	21,6	6,8			
7	5	19	23	18	11	9			25	5	77	15,4	9,8			
8	Tổng	88	108	77	54	49	376		30	5	54	10,8	8,2			
9	Trung bình	17,6	21,6	15,4	10,8	9,8	15,04		35	5	49	9,8	11,2			
10																
11																
12									ANOVA							
13									<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>	
14									Between Groups	475,8	4	118,94	14,7568	9,1E-06	2,8660	
15									Within Groups	161,2	20	8,06				
16									Total	637	24					
17																
18																
19																
20																
21																

Trường hợp các nghiệm thức **lặp lại khác nhau**

**Bảng 6.3:** Hàm lượng Vitamin C (mg/kg) có trong thực phẩm ở các thời gian nấu khác nhau (thí nghiệm CRD)

Số lần lặp lại	Thời gian nấu (phút)					
	15	20	25	30	35	
1	14	19	12	7	7	
2	18	25	17	10	7	
3	18	22	12	11	15	
4	19	22	18	15	11	
5			18	11		
Tổng	69	66	77	54	40	$\sum T_i = 306$
Trung bình	17,25	22	15,4	10,8	10	$\bar{Y} = 14,57$

$$\begin{aligned}
 SST &= \sum_{j=1}^k \sum_{i=1}^n (\bar{Y}_j - \bar{Y})^2 = \sum_{j=1}^k n_j (\bar{Y}_j - \bar{Y})^2 \\
 &= 4(17,25 - 14,57)^2 + 3(22 - 14,57)^2 + 5(15,4 - 14,57)^2 + 5(10,8 - 14,57)^2 + 4(10 - 14,57)^2 \\
 &= 352,3929
 \end{aligned}$$

$$MST = \frac{SST}{k-1} = \frac{352,3929}{4} = 88,09821$$

$$\begin{aligned}
 SSE &= \sum \sum (Y_{ij} - \bar{Y}_j)^2 \\
 &= (14 - 17,25)^2 + (18 - 17,25)^2 + (18 - 17,25)^2 + (19 - 17,25)^2 \\
 &\quad + (19 - 22)^2 + (25 - 22)^2 + (22 - 22)^2 \\
 &\quad + (12 - 15,4)^2 + (17 - 15,4)^2 + (12 - 15,4)^2 + (18 - 15,4)^2 + (18 - 15,4)^2
 \end{aligned}$$

$$\begin{aligned}
& + (7 - 10,8)^2 + (10 - 10,8)^2 + (11 - 10,8)^2 + (15 - 10,8)^2 + (11 - 10,8)^2 \\
& + (7 - 10)^2 + (7 - 10)^2 + (15 - 10)^2 + (11 - 10)^2 \\
& = 148,75
\end{aligned}$$

$$MSE = \frac{SSE}{N - k} = \frac{148,75}{16} = 9,296875$$

$$F_c = \frac{MST}{MSE} = \frac{88,09821}{9,296875} = 9,47611 \quad P\text{-value} = FDIST(9.47611,4,16) = 0,000399 = 0,0399\%$$

$$F_{4,16,1\%}^* = FINV(1\%,4,16) = 4,772578$$

$$F_{4,16,5\%}^* = FINV(5\%,4,16) = 3,0069$$

$F_c > F_{b\grave{a}ng}$ . Bác bỏ giả thuyết  $H_0$ . Trung bình giữa các nghiệm thức khác biệt có ý nghĩa ở mức 99% nghĩa là thời gian nấu (từ 15 phút đến 35 phút) ảnh hưởng rất lớn đến hàm lượng Vitamin C có trong sản phẩm.

### Thực hiện trong Excel

Tool/Data Analysis/Anova: Single Factor

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	Số lần lặp lại	Thời gian nấu (phút)							Anova: Single Factor						
2		15	20	25	30	35									
3	1	14	19	12	7	7									
4	2	18	25	17	10	7									
5	3	18	22	12	11	15									
6	4	19		18	15	11									
7	5			18	11										
8	Tổng	69	66	77	54	40	306								
9	Trung bình	17,25	22	15,4	10,8	10	14,57								
10															
11	Anova: Single Factor														
12	SUMMARY														
13															
14															
15															
16															
17															
18															
19															
20															
21															
22															
23															
24															
25															
26															
27															
28															
29															
30															
31															
32															
33															
34															
35															
36															
37															
38															
39															
40															
41															
42															
43															
44															
45															
46															
47															
48															
49															
50															
51															
52															
53															
54															
55															
56															
57															
58															
59															
60															
61															
62															
63															
64															
65															
66															
67															
68															
69															
70															

### III. So sánh các cặp trung bình của nghiệm thức

Giả thuyết

$$H_0 \quad \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_1 \quad \mu_i \neq \mu_j \text{ có ít nhất 1 cặp (ij) khác nhau}$$

Kiểm định hai phía

#### 1. Phương pháp LSD (Giới hạn sai khác nhỏ nhất – Least Significant Difference)

Khi phân tích phương sai dùng trắc nghiệm F cho kết quả là bác bỏ  $H_0$  nghĩa là tồn tại ít nhất một cặp có bình quân khác nhau. Vấn đề ở chỗ là các cặp nào khác nhau có ý nghĩa thống kê? Phân tích ANOVA chỉ đánh giá chung ảnh hưởng của nghiệm thức mà không cho biết cặp nào khác biệt có ý nghĩa? Điều này chỉ có thể thực hiện bằng trắc nghiệm t.

Trường hợp các lần lặp lại khác nhau

$$\text{LSD} = t_{v, \alpha/2}^* \sqrt{\left(\frac{1}{n_j} + \frac{1}{n_{j'}}\right) \text{MSE}}$$

$j$  là nghiệm thức  $j$  và  $j'$  là nghiệm thức  $j'$   
 $n_j$  số lần lặp lại của nghiệm thức  $j$   
 $n_{j'}$  số lần lặp lại của nghiệm thức  $j'$   
 $v = N - k$  độ tự do của MSE

Trường hợp các lần lặp lại như nhau

$$\text{LSD} = t_{v, \alpha/2}^* \sqrt{\frac{2\text{MSE}}{n}}$$

$n$  là số lần lặp lại

Nếu  $|\bar{Y}_j - \bar{Y}_{j'}| > \text{LSD}$  thì trung bình của nghiệm thức  $j$  và  $j'$  sai khác ở mức ý nghĩa  $\alpha$ .

**VD 6.2:** Sử dụng bảng số liệu 6.1 so sánh giữa các nghiệm thức

Trường hợp các lần lặp lại giống nhau

$$\text{LSD} = t_{N-k, \alpha/2}^* \sqrt{\frac{2\text{MSE}}{n}}$$

$$t_{N-k, \alpha/2}^* = t_{20; 2, 5\%}^* = \text{TINV}(5\%, 20) = 2,086$$

$$\text{LSD}_{0,05} = 2,086 \sqrt{\frac{2 \times 8,06}{5}} = 3,75$$

$$\text{LSD}_{0,01} = t_{20; 0,5\%}^* \sqrt{\frac{2 \times 8,06}{5}} = \text{TINV}(1\%, 20) \sqrt{\frac{2 \times 8,06}{5}} = 5,1$$

**Bảng 6.4:** Bảng so sánh hàm lượng Vitamin C giữa các thời gian nấu khác nhau.

Thời gian nấu	Hàm lượng trung bình	15	20	25	30	35
15	17,6	-				
20	21,6	-4*	-			
25	15,4	2,2	6,2**	-		
30	10,8	6,8**	10,8**	4,6*	-	
35	9,8	7,8**	11,8**	5,6**	1	-

Qua kết quả cho thấy

- Có 2 cặp không khác biệt đó là nghiệm thức nấu ở (15' – 25') và (30' – 35')
- Các cặp có (\*) đều khác biệt có ý nghĩa với mức  $\alpha = 5\%$  (khác biệt có ý nghĩa ở mức tin cậy 95%).
- Các cặp có (\*\*) đều khác biệt có ý nghĩa với mức  $\alpha = 1\%$  (khác biệt có ý nghĩa ở mức tin cậy 99%).
- Hàm lượng Vitamin C ở nghiệm thức nấu 20 phút là 216 mg/kg khác biệt có ý nghĩa với tất cả các nghiệm thức còn lại. Nghiệm thức nấu ở 20 phút cho giá trị cao nhất của hàm lượng Vitamin C vậy thời gian nấu 20 phút là tốt nhất.

## 2. Phương pháp Duncan



Phải có số lần lặp lại bằng nhau

Bước 1: Sắp xếp các số trung bình của nghiệm thức theo thứ tự tăng dần

Bước 2: Tính sai số chuẩn của trung bình

$$SE = s_{\bar{Y}_j} = \sqrt{\frac{MSE}{n}} \quad n \text{ là số lần lặp lại}$$

Bước 3: Tính khoảng sai biệt có ý nghĩa

$$R_p = r_p(df, \alpha) \times SE$$

$r_p(df, \alpha)$  được tra bảng cho trắc nghiệm Duncan (phụ lục)

$p$  là vị trí tương đối trong thứ tự đã sắp xếp (Vd:  $p=2$  giữa hai số kế nhau)

$df$  bậc tự do của MSE ( $df = N - k = \text{số thí nghiệm} - \text{số nghiệm thức}$ )

Bước 4: Lập bảng tính sự khác biệt bình quân giữa hai nghiệm thức

Tính sự khác biệt bình quân giữa hai nghiệm thức lần lượt bắt đầu từ số lớn nhất tương ứng với số bé nhất.

Nếu  $\bar{Y}_j - \bar{Y}_j' > R_p$  thì hai số trung bình này khác biệt ở mức ý nghĩa  $\alpha$ .

Bước 5: Tập hợp trung bình thành từng nhóm không khác nhau.

**VD 6.3:** Sử dụng bảng số liệu 6.1 so sánh giữa các nghiệm thức bằng phương pháp Duncan

Bước 1: Sắp xếp các số trung bình theo thứ tự tăng dần

Thứ tự nghiệm thức (k)	T5	T4	T3	T1	T2
$\bar{Y}_k$	9,8	10,8	15,4	17,6	21,6

Bước 2: Tính sai số chuẩn của trung bình

$$SE = s_{\bar{Y}_j} = \sqrt{\frac{MSE}{n}} = \sqrt{\frac{8,06}{5}} = 1,27$$

Bước 3: Tính khoảng sai biệt có ý nghĩa

Tra bảng Duncan với  $p=2, 3, 4, 5$  và  $df=25-5$

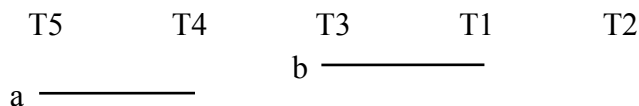
<b>p</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
$r_p(20, 5\%)$	2,95	3,1	3,18	3,25
$R_p$	3,75	3,94	4,04	4,13

Bước 4: Lập bảng tính sự khác biệt bình quân giữa hai nghiệm thức

Hiệu số giữa các cặp nghiệm thức ( $\bar{Y}_j - \bar{Y}_j'$ )									
T2-T5	T2-T4	T2-T3	T2-T1	T1-T5	T1-T4	T1-T3	T3-T5	T3-T4	T4-T5
11,8*	10,8*	6,2*	4,0*	7,8*	6,8*	2,2	5,6*	4,6*	1
$R_5$	$R_4$	$R_3$	$R_2$	$R_4$	$R_3$	$R_2$	$R_3$	$R_2$	$R_2$

Bước 5: Tập hợp trung bình thành từng nhóm không khác nhau.

Nhóm T1-T3 và nhóm T4-T5



Nghiệm thức	Hàm lượng Vitamin C	Chỉ số đánh giá
T1	17,6	b
T2	21,6	
T3	15,4	b
T4	10,8	a
T5	9,8	a

**VD 6.4:** Trong một thí nghiệm so sánh 7 nghiệm thức với 5 lần lặp lại, trung bình các nghiệm thức như sau:

<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>	<b>G</b>
49,6	71,2	67,6	61,5	71,3	58,1	61,0

Và MSE = 66,358

Bước 1: Sắp xếp các số trung bình theo thứ tự tăng dần

<b>A</b>	<b>F</b>	<b>G</b>	<b>D</b>	<b>C</b>	<b>B</b>	<b>E</b>
49,6	58,1	61,0	61,5	67,6	71,2	71,3

Bước 2: Tính sai số chuẩn của trung bình

$$SE = s_{\bar{Y}_j} = \sqrt{\frac{MSE}{n}} = \sqrt{\frac{66,358}{5}} = 3,643$$

Bước 3: Tính khoảng sai biệt có ý nghĩa

Tra bảng Duncan với p=2, 3, 4, 5, 6, 7 và df=35-7

<b>p</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
r <sub>p</sub> (28, 5%)	2,9	3,04	3,13	3,2	3,26	3,3
R <sub>p</sub>	10,6	11,1	11,4	11,7	11,9	12,02

Bước 4: Lập bảng tính sự khác biệt bình quân giữa hai nghiệm thức

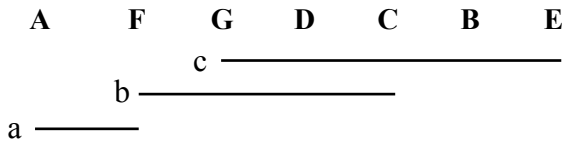
Hiệu số giữa các cặp nghiệm thức ( $\bar{Y}_j - \bar{Y}_j$ )

E-A	E-F	E-G	E-D	E-C	E-B	B-A	B-F	B-G	B-D	B-C	C-A	C-F	C-G	C-D	D-A	D-F	D-G	G-A	G-F	F-A
21,7*	13,2*	10,3	9,8	3,7	0,1	21,6*	13,1*	10,2	9,7	3,6	18*	9,5	6,6	6,1	11,9*	3,4	0,5	11,4*	2,9	8,5
R <sub>7</sub>	R <sub>6</sub>	R <sub>5</sub>	R <sub>4</sub>	R <sub>3</sub>	R <sub>2</sub>	R <sub>6</sub>	R <sub>5</sub>	R <sub>4</sub>	R <sub>3</sub>	R <sub>2</sub>	R <sub>5</sub>	R <sub>4</sub>	R <sub>3</sub>	R <sub>2</sub>	R <sub>4</sub>	R <sub>3</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>2</sub>	R <sub>2</sub>

Bước 5: Tập hợp trung bình thành từng nhóm không khác nhau.

Các cặp nghiệm thức không sai khác ở mức ý nghĩa 5%

E-G    B-G    C-F    D-F    G-F    F-A  
 E-D    B-D    C-G    D-G  
 E-C    B-C    C-D  
 E-B



Các nghiệm thức có cùng gạch dưới không sai khác ở mức ý nghĩa 5%

A    F    G    D    C    B    E  
 a    ab    bc    bc    bc    c    c

Các nghiệm thức có cùng chữ (a, b, c) không sai khác ở mức ý nghĩa 5%

#### IV. Hệ số biến động

$$CV\% = \frac{\sqrt{MSE}}{\bar{Y}} \times 100$$

CV% cho biết sai số của thí nghiệm

#### V. Xử lý bằng phần mềm SPSS cho ví dụ của bảng 6.1

Yêu cầu

- (a) Lập bảng ANOVA, để kiểm định các yếu tố có ảnh hưởng đến thí nghiệm không
- (b) So sánh sự khác biệt bằng LSD và Duncan

##### 1. Nhập số liệu

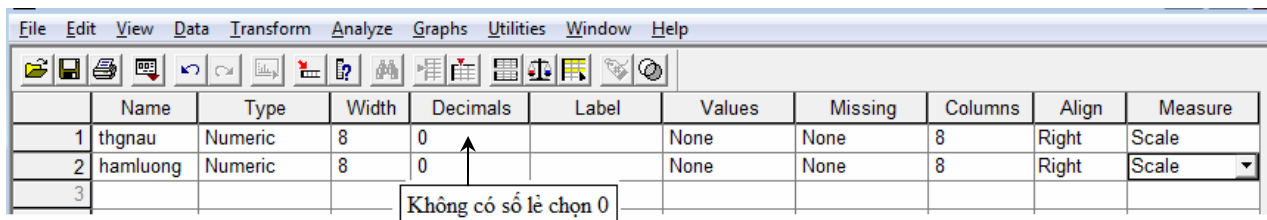
Trước tiên khai báo biến.

Hàm lượng Vitamin C = f(Thời gian nấu)

Biến phụ thuộc      Hàm lượng Vitamin C, đặt tên biến là **hamluong**

Biến độc lập      Thời gian nấu, đặt tên biến là **thgnau**

Nhấp chọn **Variable View** (ở góc dưới bên trái)



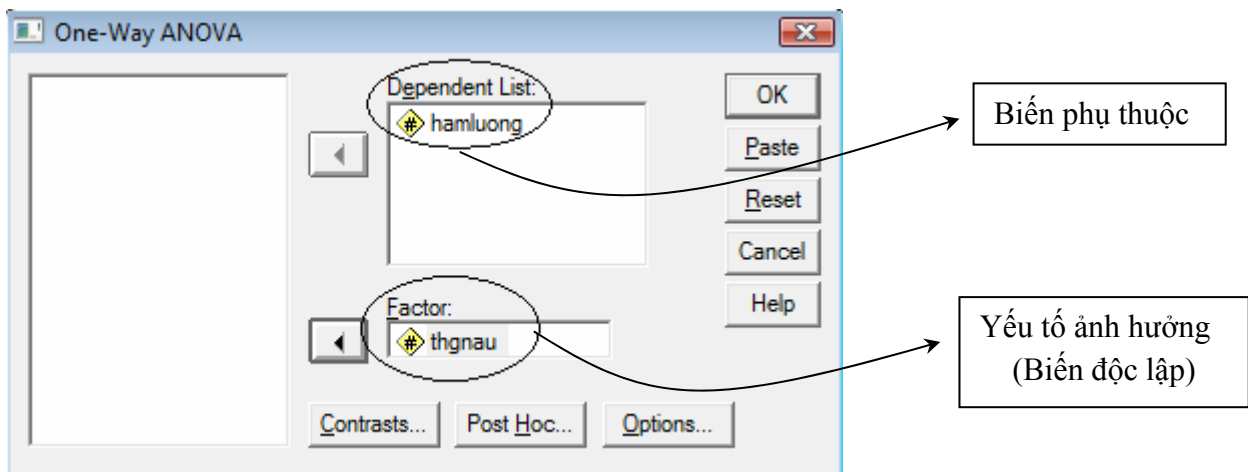
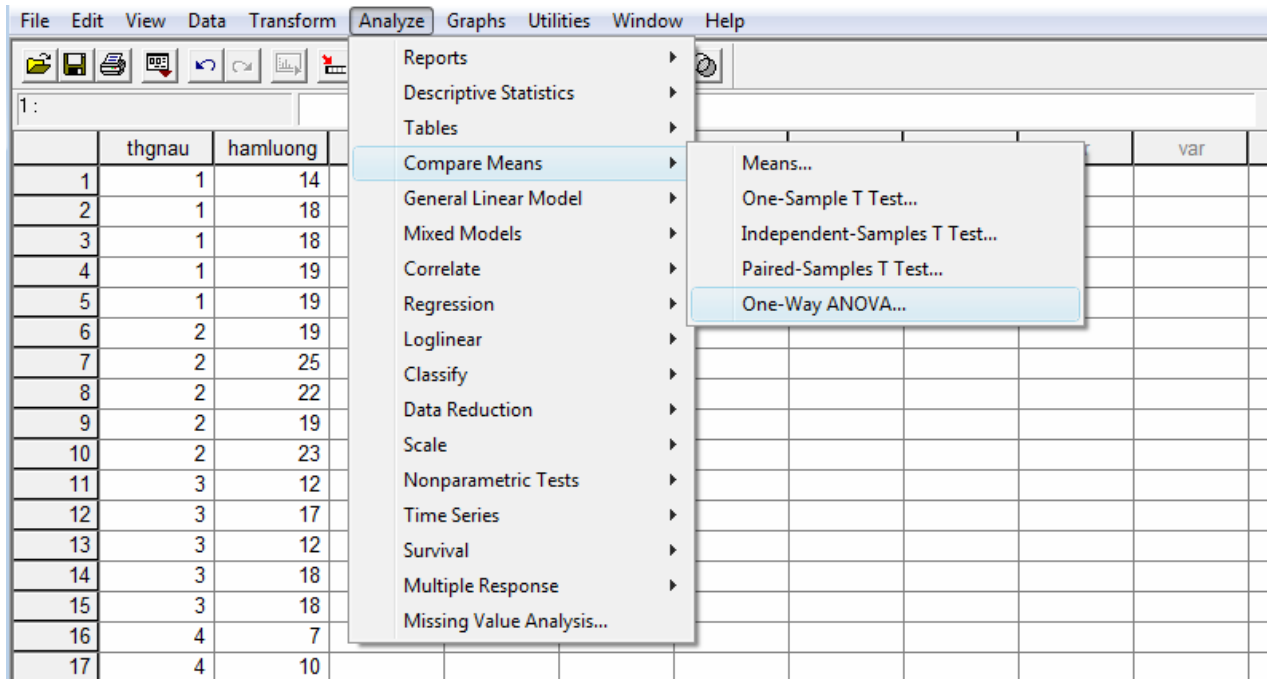
Vào **Data View** để nhập số liệu

	thgnau	hamluong	var	var
1	1	14		
2	1	18		
3	1	18		
4	1	19		
5	1	19		
6	2	19		
7	2	25		
8	2	22		
9	2	19		
10	2	23		
11	3	12		
12	3	17		
13	3	12		
14	3	18		
15	3	18		
16	4	7		
17	4	10		
18	4	11		
19	4	15		
20	4	11		
21	5	7		
22	5	7		
23	5	15		
24	5	11		
25	5	9		

Số 1 chỉ nghiệm thức thứ 1.  
 Nghiệm thức thứ 1 được lặp lại  
 5 lần. Do đó lặp lại 5 lần số 1

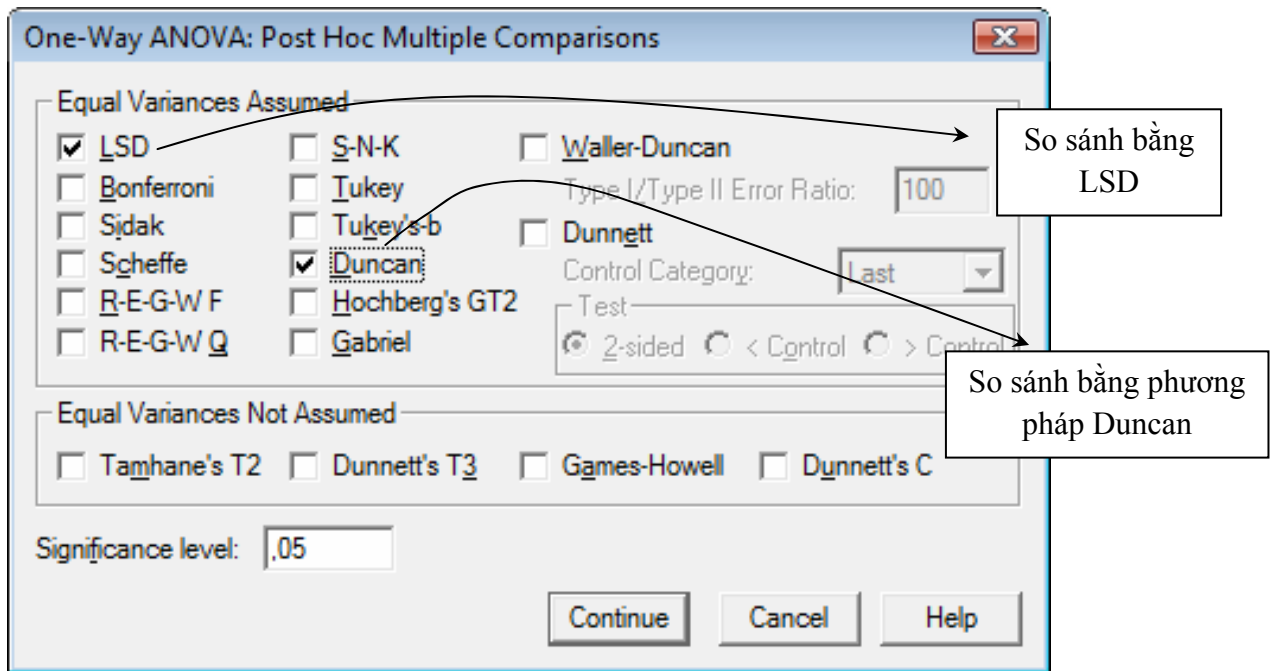
## 2. Lập bảng ANOVA trong thí nghiệm CRD với 1 yếu tố (bảng ANOVA một chiều)

### Analyze/Compare Means/One-Way ANOVA



### 3. So sánh sự khác biệt giữa các nghiệm thức bằng LSD và Duncan

Để cho kết quả của bảng ANOVA và so sánh sự khác biệt của nghiệm thức. Từ hộp thoại trên chọn **Post Hoc...** Xuất hiện hộp thoại sau:



## Kết quả xử lý

### ANOVA

HAMLUONG

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	475,760	4	118,940	14,757	,000
Within Groups	161,200	20	8,060		
Total	636,960	24			

### Post Hoc Tests

#### Multiple Comparisons

Dependent Variable: HAMLUONG

	(I) THGNAU	(J) THGNAU	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
LSD	1	2	-4,00(*)	1,796	,038	-7,75	-,25
		3	2,20	1,796	,235	-1,55	5,95
		4	6,80(*)	1,796	,001	3,05	10,55
		5	7,80(*)	1,796	,000	4,05	11,55
	2	1	4,00(*)	1,796	,038	,25	7,75
		3	6,20(*)	1,796	,003	2,45	9,95
		4	10,80(*)	1,796	,000	7,05	14,55
		5	11,80(*)	1,796	,000	8,05	15,55
	3	1	-2,20	1,796	,235	-5,95	1,55
		2	-6,20(*)	1,796	,003	-9,95	-2,45
		4	4,60(*)	1,796	,019	,85	8,35
		5	5,60(*)	1,796	,005	1,85	9,35
	4	1	-6,80(*)	1,796	,001	-10,55	-3,05
		2	-10,80(*)	1,796	,000	-14,55	-7,05
		3	-4,60(*)	1,796	,019	-8,35	-,85
		5	1,00	1,796	,584	-2,75	4,75
	5	1	-7,80(*)	1,796	,000	-11,55	-4,05
		2	-11,80(*)	1,796	,000	-15,55	-8,05
		3	-5,60(*)	1,796	,005	-9,35	-1,85
		4	-1,00	1,796	,584	-4,75	2,75

\* The mean difference is significant at the .05 level.

### HAMLUONG

	THGNAU	N	Subset for alpha = .05		
			1	2	3
Duncan(a)	5	5	9,80		
	4	5	10,80		
	3	5		15,40	
	1	5		17,60	
	2	5			21,60
Sig.			,584	,235	1,000

Means for groups in homogeneous subsets are displayed.

a Uses Harmonic Mean Sample Size = 5,000.

#### 4. Giải thích kết quả xử lý

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups (Nghiệm thức)	SST	Bậc	MST	$F_c$	P-value (Từ $F_c$ suy ngược ra xác suất, P-value)
Within Groups (Sai số)	SSE	tự	MSE	( $F_{tính}$ )	
Total	$SST_0$	do			

<b>Mean Difference (I-J)</b>	Sai biệt giữa trung bình nghiệm thức (I) và (J)	$\bar{Y}_I - \bar{Y}_J$
<b>Std. Error</b>	Sai số chuẩn của sai biệt các số trung bình	<p>Các lần lặp lại của nghiệm thức không bằng nhau</p> $s_{\bar{Y}_I - \bar{Y}_J} = \sqrt{\left(\frac{1}{n_i} + \frac{1}{n_j}\right) MSE}$ <p>Các lần lặp lại của nghiệm thức như nhau</p> $s_{\bar{Y}_I - \bar{Y}_J} = \sqrt{\frac{2MSE}{n}}$
<b>Sig.</b>	Khác biệt giữa hai nghiệm thức ở mức ý nghĩa Sig.	<p>Giả thuyết <math>H_0 \mu_i = \mu_j</math> <math>H_1 \mu_i \neq \mu_j</math></p> <p>Nếu P-value &lt; <math>\alpha</math>. Bác bỏ <math>H_0</math></p> <p>Có sự khác biệt giữa hai nghiệm thức I và J ở mức ý nghĩa <math>\alpha</math>.</p> <p>VD: Sig.=0,235 (P-value). Có sự khác biệt giữa nghiệm thức (1) và (3) ở mức ý nghĩa 23,5%. Hay cặp (1 và 3) khác biệt ở mức tin cậy là 76,5%.</p> <p>Trong trường hợp này P-value &gt; <math>\alpha = 5\%</math></p> <p>Không bác bỏ <math>H_0</math>. Hay không có sự khác biệt giữa nghiệm thức (1) và (3) ở mức ý nghĩa 5%.</p>
<b>Confidence Interval</b>	Khoảng tin cậy của sự khác biệt	$(\bar{Y}_I - \bar{Y}_J) \pm t_{\alpha/2, df(MSE)} SE$ <p>Mean Diffence(I-J) <math>\pm t_{\alpha/2, df(MSE)} Std.Error</math></p>



## Chương 7

### PHÂN TÍCH PHƯƠNG SAI NHIỀU CHIỀU CỦA THÍ NGHIỆM MỘT YẾU TỐ

Trong thí nghiệm CRD không có sự tác động của ngoại cảnh. Trong chương này chúng ta sẽ xét các bố trí thí nghiệm có hơn 1 nguồn tác động của ngoại cảnh. Do vậy phân tích phương sai xem như “phương sai nhiều chiều”.

#### I. Kiểu khối đầy đủ (RCBD)

Trong trường hợp này không có sự tương tác giữa nghiệm thức và khối nên đây vẫn là thí nghiệm một yếu tố. Vì đối với thí nghiệm hai yếu tố có khả năng xảy ra tương tác giữa hai yếu tố.

#### 1. Sắp xếp số liệu

**Bảng 7.1:** Bảng số liệu thí nghiệm một yếu tố RCBD

Khối (Số lần lặp lại)	Các nghiệm thức					Tổng
	1	2	...	j	k	
1	$Y_{11}$	$Y_{12}$	...	$Y_{1j}$	$Y_{1k}$	$T_{b1}$
2	$Y_{21}$	$Y_{22}$	...	$Y_{2j}$	$Y_{2k}$	$T_{b2}$
...	...	...	...	...	...	
i	$Y_{i1}$	$Y_{i2}$	...	$Y_{ij}$	$Y_{ik}$	$T_{bi}$
...	...	...	...	...	...	
n	$Y_{n1}$	$Y_{n2}$	...	$Y_{nj}$	$Y_{nk}$	$T_{bn}$
<b>Tổng</b>	<b><math>T_1</math></b>	<b><math>T_2</math></b>	...	<b><math>T_j</math></b>	<b><math>T_k</math></b>	<b><math>T</math></b>
<b>Trung bình</b>	<b><math>\bar{Y}_1</math></b>	<b><math>\bar{Y}_2</math></b>	...	<b><math>\bar{Y}_j</math></b>	<b><math>\bar{Y}_k</math></b>	<b><math>\bar{Y}</math></b>

Mỗi giá trị trong bảng 7.1 là tổng của các thành phần sau:

$$Y_{ij} = \mu + \beta_i + \nu_{ij} + \varepsilon_{ij}$$

Trong đó

$\mu$  Trung bình thực của tổng thể

$\beta_i$  Ảnh hưởng của khối (Ảnh hưởng của yếu tố ngoại cảnh)

$\nu_{ij}$  Ảnh hưởng của nghiệm thức. Sự khác biệt giữa trung bình của nghiệm thức j so với trung bình toàn bộ ( $\nu_{ij} = \bar{Y}_j - \bar{Y}$ ).

$\varepsilon_{ij}$  Sai số ngẫu nhiên.

$\sum \sum (Y_{ij} - \bar{Y})^2 = \sum \sum (\bar{Y}_i - \bar{Y})^2 + \sum \sum (\bar{Y}_j - \bar{Y})^2 + \sum \sum (Y_{ij} - \bar{Y}_i - \bar{Y}_j + \bar{Y})^2$	Với $\left( \sum \sum = \sum_{j=1}^k \sum_{i=1}^n \right)$
$SST_0 = SSB + SST + SSE$	

$SST_0$  Tổng bình phương toàn bộ

$SSB$  Tổng bình phương khối

$SSE$  Tổng bình phương sai số ngẫu nhiên

$SST$  Tổng bình phương nghiệm thức

#### 2. Bảng phân tích phương sai (ANOVA)

<b>Nguồn biến động</b> <i>Source of variation</i>	<b>Độ tự do</b> <i>Degree of freedom</i>	<b>Tổng bình phương</b> <i>Sum of Square</i>	<b>Trung bình bình phương</b> <i>Mean Square</i>	$F_c$	$F_{k-1;(b-1)(k-1);\alpha}^*$ <b>(<math>F_{bảng}</math>)</b>
Khối (Block)	b - 1	SSB	MSB	$\frac{MST}{MSE}$	Tra bảng
Nghiệm thức (Treatment)	k - 1	SST	MST		
Sai số (Error)	(b - 1)(k - 1)	SSE	MSE		
Tổng (Total)	bk - 1 = N - 1	SST <sub>0</sub>			

Với  $F_{bảng} = F_{k-1;(b-1)(k-1),\alpha}^* = FINV(\alpha, (k-1), (b-1)(k-1))$

k số nghiệm thức

b số khối

Giả thuyết

$$H_0 \quad \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_1 \quad \mu_i \neq \mu_j \text{ có ít nhất 1 cặp (ij) khác nhau}$$

Nếu  $F_c > F_{k-1;(b-1)(k-1),\alpha}^*$  thì bác bỏ giả thuyết  $H_0$ . Có ít nhất trung bình của hai nghiệm thức khác nhau ở mức ý nghĩa  $\alpha$ .

Nếu  $F_c < F_{k-1;(b-1)(k-1),\alpha}^*$  thì không bác bỏ giả thuyết  $H_0$ . Trung bình của các nghiệm thức đều bằng nhau ở mức ý nghĩa  $\alpha$ . Hay các yếu tố không ảnh hưởng đến thí nghiệm.

### **Xử lý bằng phần mềm Excel**

Sử dụng số liệu ở VD5.6:

Độ cứng của bánh = f(Kích thước đầu nén) {yếu tố ngoại cảnh: vị trí bánh nướng}

**Tool/Data Analysis.../Anova: Two Factor Without Replication**

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Thí nghiệm					Anova: Two-Factor Without Replication						
2		A	B	C	D		SUMMARY		Count	Sum	Average	Variance	
3	Vị trí 1	9,3	9,4	9,2	9,7		Vị trí 1	4	37,6	9,4	0,04667		
4	Vị trí 2	9,4	9,3	9,4	9,6		Vị trí 2	4	37,7	9,425	0,01583		
5	Vị trí 3	9,6	9,8	9,5	10		Vị trí 3	4	38,9	9,725	0,04917		
6	Vị trí 4	10	9,9	9,7	10,2		Vị trí 4	4	39,8	9,95	0,04333		
7													
8	Anova: Two-Factor Without Replication												
9	Input												
10	Input Range: \$A\$2:\$E\$6												
11	<input checked="" type="checkbox"/> Labels												
12	Alpha: 0,05												
13	Output options												
14	<input checked="" type="radio"/> Output Range: \$G\$1												
15	<input type="radio"/> New Worksheet Ply:												
16	<input type="radio"/> New Workbook												
17							ANOVA						
18							Source of Variation	SS	df	MS	F	P-value	F crit
19							Rows	0,825	3	0,275	30,9375	4,5E-05	3,86255
20							Columns	0,385	3	0,12833	14,4375	0,00087	3,86255
21							Error	0,08	9	0,00889			
22							Total	1,29	15				

Giả thuyết

$$H_0 \quad \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_1 \quad \mu_i \neq \mu_j \text{ có ít nhất 1 cặp (ij) khác nhau}$$

Kết quả xử lý từ Excel ta có  $F_c = 14,4$  tương ứng với  $P\text{-value} = 0,00087 < \alpha$

$\Rightarrow$  Bác bỏ giả thuyết  $H_0$

**Hay**  $F_{\text{bảng}} = F_{3,9,5\%}^* = \text{FINV}(5\%, 3, 9) = 3,86255$

$F_c > F_{\text{bảng}} \Rightarrow$  Bác bỏ giả thuyết  $H_0$

$\Rightarrow$  Kích thước đầu đo có ảnh hưởng đến độ cứng của bánh.

### 3. Phương pháp LSD

Tương tự như cách bố trí CRD

$$\text{LSD} = t_{v, \alpha/2}^* \sqrt{\frac{2\text{MSE}}{b}}$$

$b$  là số khối (số lần lặp lại)

$v = (b - 1)(k - 1)$  độ tự do của MSE

$t_{9, 2,5\%}^* = \text{TINV}(5\%; 9) = 2,262$

$\text{LSD}_{0,05} = 2,262 \sqrt{\frac{2 \times 0,008889}{4}} = 0,1508$

**Bảng 7.2:** Bảng so sánh độ cứng với các đầu đo khác nhau (RCBD)

Nghiệm thức	Độ cứng trung bình	A	B	C	D
A	9,575	-			
B	9,6	-0,025	-		
C	9,45	0,125	0,15	-	
D	9,875	-0,3*	0,27*	-0,425*	-

Qua kết quả cho thấy

- Các cặp có (\*) đều khác biệt có ý nghĩa với mức  $\alpha=5\%$  (khác biệt có ý nghĩa ở mức tin cậy 95%).
- Độ cứng bánh ở kích thước đầu đo D khác biệt có ý nghĩa với tất cả các nghiệm thức còn lại. Nghiệm thức này cho giá trị về độ cứng cao nhất, vậy kích thước đầu đo D là tốt nhất.

#### 4. Phương pháp Duncan

Bước 1: Sắp xếp các số trung bình của nghiệm thức theo thứ tự tăng dần

Bước 2: Tính sai số chuẩn của trung bình

$$SE = s_{\bar{Y}_j} = \sqrt{\frac{MSE}{b}} \quad b \text{ là số lần lặp lại}$$

Bước 3: Tính khoảng sai biệt có ý nghĩa

$$R_p = r_p(df, \alpha) \times SE$$

$r_p(df, \alpha)$  được tra bảng cho trắc nghiệm Duncan  
 $p$  là vị trí tương đối trong thứ tự đã sắp xếp  
 $df$  bậc tự do của MSE,  $df = (b - 1)(k - 1)$

Bước 4: Lập bảng tính sự khác biệt bình quân giữa hai nghiệm thức

Bước 5: Tập hợp trung bình thành từng nhóm không khác nhau.

##### Bước 1

<b>C</b>	<b>A</b>	<b>B</b>	<b>D</b>
9,45	9,575	9,6	9,875

##### Bước 2 Tính sai số chuẩn

$$SE = s_{\bar{Y}_j} = \sqrt{\frac{MSE}{b}} = 0,047$$

##### Bước 3

p	2	3	4	5
$r_p(9,5\%)$	3,2	3,34	3,41	3,47
$R_p$	0,1509	0,1575	0,1607	0,1636

##### Bước 4

<b>D-C</b>	<b>D-A</b>	<b>D-B</b>	<b>B-C</b>	<b>B-A</b>	<b>A-C</b>
0,425*	0,3*	0,275*	0,15	0,025	0,125
$R_4$	$R_3$	$R_2$	$R_3$	$R_2$	$R_2$

##### Bước 5

Các cặp nghiệm thức không sai khác ở mức ý nghĩa 5%

**B-C**      **A-C**  
**B-A**

<b>C</b>	<b>A</b>	<b>B</b>	<b>D</b>
----------	----------	----------	----------

a _____			
Các nghiệm thức có cùng gạch dưới không sai khác với mức ý nghĩa 5%			
<b>C</b>	<b>A</b>	<b>B</b>	<b>D</b>
a	a	a	
Các nghiệm thức có cùng chữ (a) không sai khác với mức ý nghĩa 5%			

## 5. Xử lý bằng SPSS

Sử dụng ví dụ 5.6

Độ cứng của bánh = f(Kích thước đầu nén) {yếu tố ngoại cảnh: vị trí bánh nướng}

### 5.1. Nhập số liệu

Biến phụ thuộc:

Độ cứng của bánh, đặt tên biến là **docung**

Yếu tố ảnh hưởng

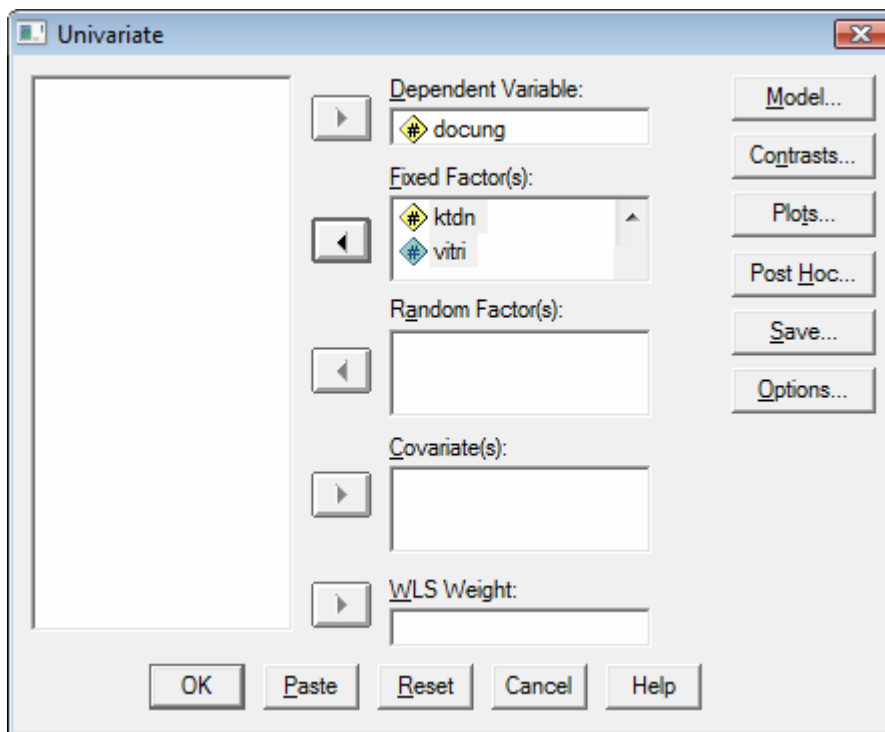
Biến độc lập (Yếu tố cần nghiên cứu) là kích thước đầu nén, đặt tên biến **ktdn**

Yếu tố ngoại cảnh vị trí, đặt tên cho yếu tố ngoại cảnh này là **vitri**

	ktdn	vitri	docung	va
1	1	1	9,3	
2	2	1	9,4	
3	3	1	9,2	
4	4	1	9,7	
5	1	2	9,4	
6	2	2	9,3	
7	3	2	9,4	
8	4	2	9,6	
9	1	3	9,6	
10	2	3	9,8	
11	3	3	9,5	
12	4	3	10,0	
13	1	4	10,0	
14	2	4	9,9	
15	3	4	9,7	
16	4	4	10,2	
17				

## 5.2. Phân tích phương sai (ANOVA)


Analyze/General Linear Model/Univariate...

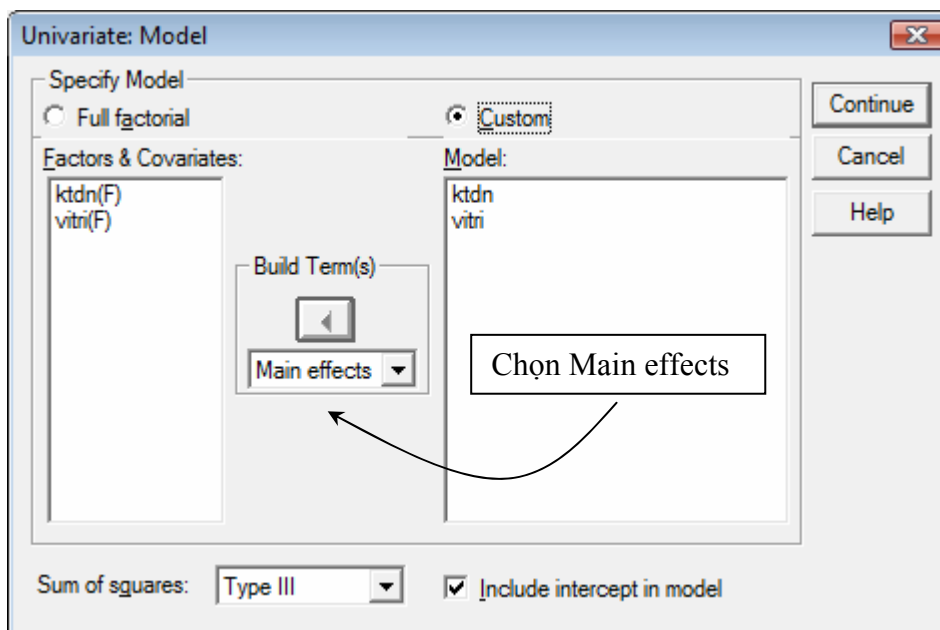


Phần mềm luôn mặc định  $\alpha=5\%$ . Do đó muốn thay đổi  $\alpha$  thì vào **Option...**

Nhấp **M**odel...

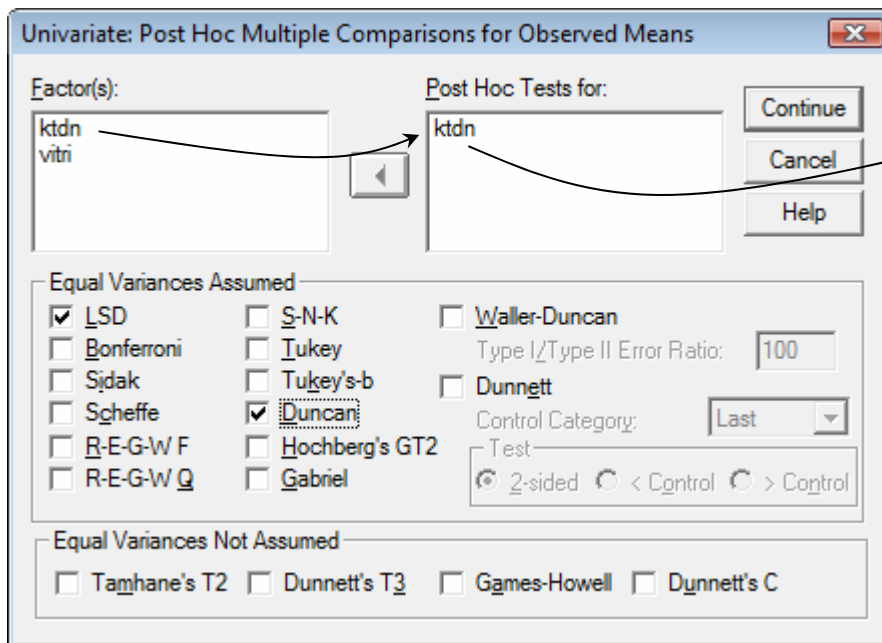
Chọn **C**ustom

Nhấp chọn **ktdn**, rồi nhấp  để đưa biến **ktdn** vào **M**odel. Thực hiện tương tự để đưa yếu tố **vtri** vào **M**odel.



### 5.3. So sánh sự khác biệt của nghiệm thức bằng phương pháp LSD và Duncan

Nhấp **Continue/Post Hoc...**



Chỉ đưa yếu tố **ktdn** vào ô:  
**Post Hoc Test for:**  
*Vì chỉ so sánh sự khác biệt của nghiệm thức.  
 Không so sánh sự khác biệt của block (vị trí)*

Nhấp **Continue/OK**

#### Tests of Between-Subjects Effects

Dependent Variable: DOCUNG

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1,210(a)	6	,202	22,687	,000
Intercept	1482,250	1	1482,250	166753,125	,000
KTDN	,385	3	,128	<b>14,437</b>	<b>,001</b>
VITRI	,825	3	,275	30,937	,000
Error	,080	9	,009		
Total	1483,540	16			
Corrected Total	1,290	15			

a R Squared = ,938 (Adjusted R Squared = ,897)

## Post Hoc Tests

### KTDN

#### Multiple Comparisons

Dependent Variable: DOCUNG

	(I) KTDN	(J) KTDN	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
LSD	1	2	-,025	,0667	,716	-,176	,126
		3	,125	,0667	,094	-,026	,276
		4	-,300(*)	,0667	,001	-,451	-,149
	2	1	,025	,0667	,716	-,126	,176
		3	,150	,0667	,051	-,001	,301
		4	-,275(*)	,0667	,003	-,426	-,124
	3	1	-,125	,0667	,094	-,276	,026
		2	-,150	,0667	,051	-,301	,001
		4	-,425(*)	,0667	,000	-,576	-,274
	4	1	,300(*)	,0667	,001	,149	,451
		2	,275(*)	,0667	,003	,124	,426
		3	,425(*)	,0667	,000	,274	,576

Based on observed means.

\* The mean difference is significant at the ,05 level.

#### DOCUNG

	KTDN	N	Subset	
			1	2
Duncan(a,b)	3	4	9,450	
	1	4	9,575	
	2	4	9,600	
	4	4		9,875
	Sig.			,060

Means for groups in homogeneous subsets are displayed. Based on Type III Sum of Squares The error term is Mean Square(Error) = ,009.

a Uses Harmonic Mean Sample Size = 4,000.

b Alpha = ,05.



## II. Kiểu ô vuông Latinh

### 1. Sắp xếp số liệu

Trường hợp này xem như thí nghiệm được tạo khối 2 chiều. Bảng số liệu được xếp như sau.

Số chiều bảng ANOVA = 1 yếu tố + 3 chiều khối = 3 chiều

**Bảng 7.3:** Bảng số liệu thí nghiệm một yếu tố bố trí theo kiểu ô vuông Latinh

Khối	CỘT				Tổng hàng	Tổng nghiệm thức
	k=1	2	k	n		
Hàng i=1	A=Y <sub>A11</sub>	B=Y <sub>B12</sub>			H <sub>1</sub>	T <sub>A</sub>
2					H <sub>2</sub>	T <sub>B</sub>
i			Y <sub>j ik</sub>		H <sub>i</sub>	T <sub>C</sub>
n				A=Y <sub>Ann</sub>	H <sub>n</sub>	T <sub>D</sub>
Tổng cột	C <sub>1</sub>	C <sub>2</sub>	C <sub>k</sub>	C <sub>n</sub>	T	

Mỗi giá trị trong bảng 7.3 là tổng của các thành phần sau:

$$Y_{j ik} = \mu + v_{ij} + \beta_i + \gamma_k + \varepsilon_{ij}$$

Trong đó

$\mu$  Trung bình thực của tổng thể

$v_{ij}$  Ảnh hưởng của nghiệm thức.

$\beta_i$  Ảnh hưởng của hàng

$\gamma_k$  Ảnh hưởng của khối

$\varepsilon_{ij}$  Sai số ngẫu nhiên.

$$SSE = \sum_{j=1}^n \sum_{i=1}^n \sum_{k=1}^n (Y_{jik})^2 - T^2/n^2 - \sum_{j=1}^n T_j^2/n - T^2/n^2 - \sum_{i=1}^n H_i^2/n - T^2/n^2 - \sum_{i=1}^n C_k^2/n - T^2/n^2$$

$$SSE = SST_0 - SST - SSR - SSC$$

SST<sub>0</sub> Tổng bình phương toàn bộ

SSR Tổng bình phương hàng

SSC Tổng bình phương cột

SSE Tổng bình phương sai số ngẫu nhiên

SST Tổng bình phương nghiệm thức

### 2. Bảng phân tích phương sai (ANOVA)

Nguồn biến động <i>Source of variation</i>	Độ tự do <i>Degree of freedom</i>	Tổng bình phương <i>Sum of Square</i>	Trung bình bình phương <i>Mean Square</i>	F <sub>c</sub>	F <sub>n-1;(n-2)(n-1);α</sub> (F <sub>bảng</sub> )
Nghiệm thức ( <i>Treatment</i> )	n - 1	SST	MST	MST MSE	Tra bảng
Hàng ( <i>Row</i> )	n - 1	SSR	MSR		
Cột ( <i>Column</i> )	n - 1	SSC	MSC		
Sai số ( <i>Error</i> )	(n - 2)(n - 1)	SSE	MSE		
Tổng ( <i>Total</i> )	n <sup>2</sup> - 1	SST <sub>0</sub>			

Với F<sub>bảng</sub> = F<sub>n-1;(n-2)(n-1);α</sub> = FINV(α, (n - 1), (n - 2)(n - 1))

n số nghiệm thức (= số hàng = số cột)

Giả thuyết

$$H_0 \quad \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_1 \quad \mu_i \neq \mu_j \text{ có ít nhất 1 cặp (ij) khác nhau}$$

Nếu  $F_c > F_{n-1; (n-2)(n-1), \alpha}^*$  thì bác bỏ giả thuyết  $H_0$ . Có ít nhất trung bình của hai nghiệm thức khác nhau ở mức ý nghĩa  $\alpha$ .

Nếu  $F_c < F_{n-1; (n-2)(n-1), \alpha}^*$  thì không bác bỏ giả thuyết  $H_0$ . Trung bình của các nghiệm thức đều bằng nhau ở mức ý nghĩa  $\alpha$ . Hay các yếu tố không ảnh hưởng đến thí nghiệm.

### 3. Xử lý bằng SPSS

Sử dụng bảng số liệu 5.4

Độ cứng của bánh = f(Kích thước đầu nén)  $\left[ \begin{array}{l} \text{Yếu tố ngoại cảnh} \left\{ \begin{array}{l} \text{Vị trí bánh nướng} \\ \text{Công nhân nướng} \end{array} \right. \end{array} \right.$

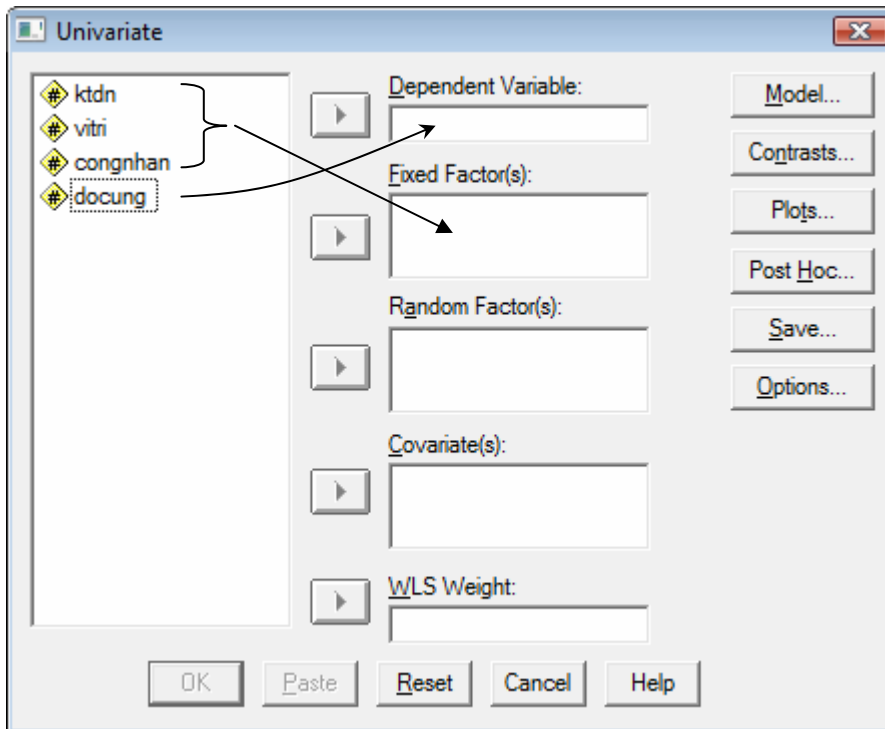
- Biến phụ thuộc:
  - Độ cứng của bánh, đặt tên biến là **docung**
- Yếu tố ảnh hưởng
  - Biến độc lập (Yếu tố cần nghiên cứu) là kích thước đầu nén, đặt tên biến **ktđn**
  - Yếu tố ngoại cảnh
    - Vị trí nướng, đặt tên cho yếu tố ngoại cảnh này là **vitri**
    - Công nhân nướng, đặt tên cho yếu tố ngoại cảnh này là **congphan**

#### 3.1. Nhập số liệu

	ktdn	vitri	congphan	docung	
1	1	1	1	9,7	
2	2	1	2	9,5	
3	3	1	3	9,6	
4	4	1	4	10,2	
5	1	2	4	9,8	
6	2	2	1	9,2	
7	3	2	2	9,1	
8	4	2	3	9,2	
9	1	3	3	9,1	
10	2	3	4	9,4	
11	3	3	1	9,6	
12	4	3	2	9,8	
13	1	4	2	10,0	
14	2	4	3	9,6	
15	3	4	4	10,3	
16	4	4	1	10,4	
17					

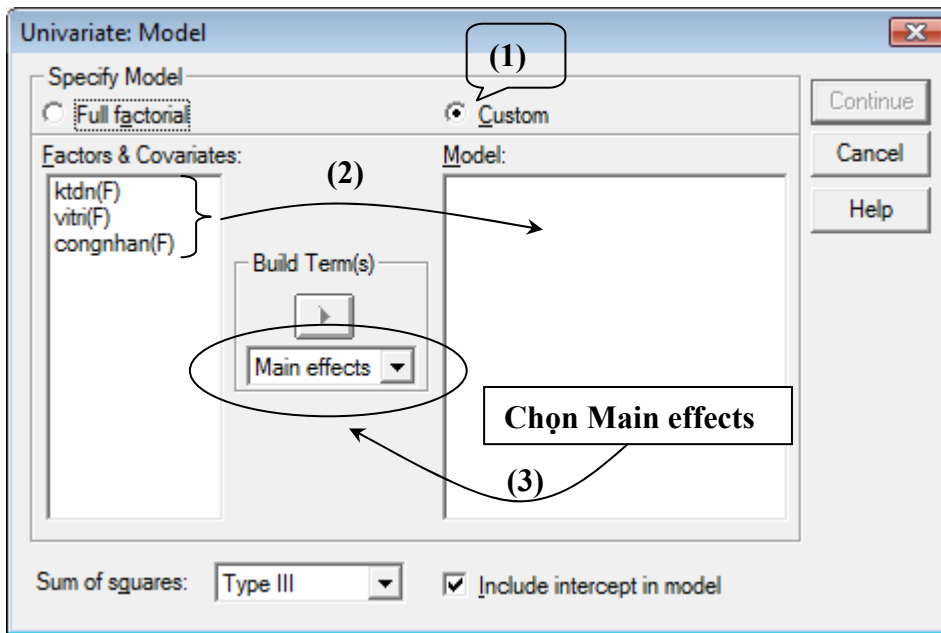
### 3.2. Phân tích ANOVA

Analyze/General Linear Model/Univariate...

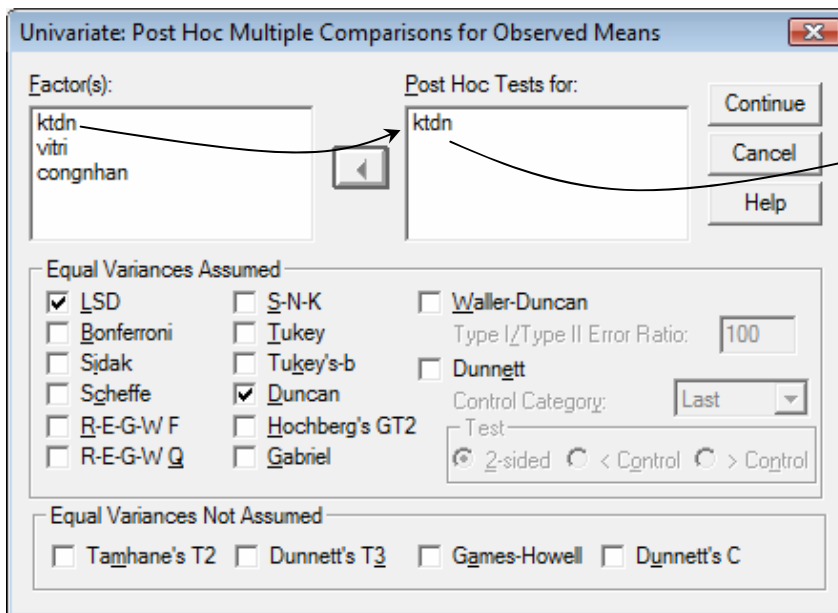


Chọn **Model...** xuất hiện hộp thoại

### 3.3. So sánh sự khác biệt giữa các nghiệm thức bằng LSD và Duncan



Thực hiện xong (1), (2) và (3) nhấp **Continue/Chọn Post Hoc...**



Chỉ đưa yếu tố **ktdn** vào ô:  
**Post Hoc Test for:**  
*Vì chỉ so sánh sự khác biệt của nghiệm thức.  
 Không so sánh sự khác biệt của block (vutri và congphan)*

### Tests of Between-Subjects Effects

Dependent Variable: DOCUNG

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2,265(a)	9	,252	8,629	,008
Intercept	1489,960	1	1489,960	51084,343	,000
KTDN	,455	3	,152	<b>5,200</b>	<b>,042</b>
VITRI	1,225	3	,408	14,000	,004
CONGNHAN	,585	3	,195	6,686	,024
Error	,175	6	,029		
Total	1492,400	16			
Corrected Total	2,440	15			

a. R Squared = ,928 (Adjusted R Squared = ,821)

Giả thuyết

$$H_0 \quad \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_1 \quad \mu_i \neq \mu_j \text{ có ít nhất 1 cặp (ij) khác nhau}$$

Kết quả xử lý từ Excel ta có  $F_c = 5,2$  tương ứng với  $P\text{-value} = 0,042 < \alpha$

⇒ Bác bỏ giả thuyết  $H_0$

**Hay**  $F_{\text{bảng}} = F_{3,6,5\%}^* = \text{FINV}(5\%,3,6) = 4,75$

$F_c > F_{\text{bảng}} \Rightarrow$  Bác bỏ giả thuyết  $H_0$

⇒ Kích thước đầu đo có ảnh hưởng đến độ cứng của bánh.

### Post Hoc Tests

#### KTDN

#### Multiple Comparisons

Dependent Variable: DOCUNG

	(I) KTDN	(J) KTDN	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
LSD	1	2	,225	,1208	,112	-,070	,520
		3	,025	,1208	,843	-,270	,320
		4	-,250	,1208	,084	-,545	,045
	2	1	-,225	,1208	,112	-,520	,070
		3	-,200	,1208	,149	-,495	,095
		4	-,475(*)	,1208	,008	-,770	-,180
3	1	-,025	,1208	,843	-,320	,270	
	2	,200	,1208	,149	-,095	,495	
	4	-,275	,1208	,063	-,570	,020	
4	1	,250	,1208	,084	-,045	,545	
	2	,475(*)	,1208	,008	,180	,770	
	3	,275	,1208	,063	-,020	,570	

Based on observed means.

\* The mean difference is significant at the ,05 level.

**DOCUNG**



## Chương 8

### THÍ NGHIỆM NHIỀU YẾU TỐ

#### I. Tác dụng và tương tác giữa các yếu tố

Có nhiều thí nghiệm liên quan đến hai yếu tố hay nhiều yếu tố. VD: khi nghiên cứu ảnh hưởng của thời gian tan chảy của thịt động và thời gian nấu bằng microwave lên chất lượng thịt, ta có thí nghiệm 2 yếu tố. Khi đối mặt với vấn đề này, trước đây người ta thường cố định tất cả các yếu tố còn lại và chỉ cho 1 yếu tố thay đổi. Quá trình này cứ tiếp tục cho đến khi hết tất cả các yếu tố. Cách làm này gọi là “phương pháp một yếu tố ở một thời gian”. Việc phân tích và giải thích của thí nghiệm như vậy rất đơn giản. Tuy nhiên khi nhà nghiên cứu muốn tổng quát hóa kết quả của thí nghiệm họ gặp phải khó khăn và dễ bị nhầm lẫn. Lý do là khi thay đổi các mức của yếu tố này, sự thay đổi của kết quả khi các mức của yếu tố còn lại thay đổi là hoàn toàn không giống nhau. Sự đáp ứng của một yếu tố phụ thuộc vào các mức của yếu tố thứ hai gọi là sự “tương tác”.

Trong nghiên cứu, đặc biệt ở lĩnh vực có liên quan đến sinh học, trắc nghiệm và giải thích các tương tác là quan trọng nhất. Ví dụ khi đang thực hiện thí nghiệm ba yếu tố A, B và C thì ngoài ảnh hưởng của từng yếu tố, chúng ta sẽ kiểm tra xem có sự tương tác giữa các yếu tố (AB, AC, BC và ABC) hay không? Nếu giữ hai yếu tố cố định và cho thay đổi một yếu tố chúng ta chỉ nói được ảnh hưởng của từng yếu tố riêng lẻ mà thôi. Đối với các nhà sinh học, kết quả của sự phân tích các liên hợp nghiệm thức yếu tố có tính thực tiễn hơn kết quả khảo sát một yếu tố.

Yếu tố B	Yếu tố A	
	a <sub>1</sub>	a <sub>2</sub>
b <sub>1</sub>	X	X + c
b <sub>2</sub>	X + k	?

$\Delta = c$

Khác X + k + c  
có tương tác

X + k + c  
không tương tác

Yếu tố A có ảnh hưởng “tùy theo” mức yếu tố B. “Tùy theo” ngụ ý có sự tương tác.

**VD 8.1:** Độ mềm của thịt = f(nhiệt độ, áp suất)

#### II. Phân tích phương sai (ANOVA) cho thí nghiệm nhiều yếu tố.

##### RCBD

$$Y_{jik} = \mu + \gamma_k + \nu_i + \beta_j + (\nu\beta)_{ij} + \varepsilon_{ijk}$$

Trong đó

- $\mu$  Trung bình thực của tổng thể
- $\gamma_k$  Ảnh hưởng của cột k
- $\nu_i$  Ảnh hưởng của nghiệm thức A
- $\beta_j$  Ảnh hưởng của nghiệm thức B
- $(\nu\beta)_{ij}$  Tương tác giữa A và B
- $\varepsilon_{ijk}$  Sai số ngẫu nhiên

##### CRD

$$Y_{jik} = \mu + \nu_i + \beta_j + (\nu\beta)_{ij} + \varepsilon_{ijk}$$

Trong đó

- $\mu$  Trung bình thực của tổng thể
- $\nu_i$  Ảnh hưởng của nghiệm thức A
- $\beta_j$  Ảnh hưởng của nghiệm thức B
- $(\nu\beta)_{ij}$  Tương tác giữa A và B
- $\varepsilon_{ijk}$  Sai số ngẫu nhiên

**Bảng 8.1:** Bảng ANOVA của thí nghiệm hai yếu tố hoàn toàn ngẫu nhiên (CRD)

<b>Nguồn biến động</b> <i>Source of variation</i>	<b>Độ tự do</b> <i>Degree of freedom</i>	<b>Tổng bình phương</b> <i>Sum of Square</i>	<b>Trung bình bình phương</b> <i>Mean Square</i>	<b>F<sub>c</sub></b>	<b>F<sub>bảng</sub></b>
Yếu tố A	a - 1	SST <sub>A</sub>	MST <sub>A</sub>	MST <sub>A</sub> /MSE	F <sub>dfA,dfE</sub>
Yếu tố B	b - 1	SST <sub>B</sub>	MST <sub>B</sub>	MST <sub>B</sub> /MSE	F <sub>dfB,dfE</sub>
Tương tác A*B	(a - 1)(b - 1)	SST <sub>AB</sub>	MST <sub>AB</sub>	MST <sub>AB</sub> /MSE	F <sub>dfAB,dfE</sub>
Sai số ( <i>Error</i> )	ab(n - 1)	SSE	MSE		
Tổng ( <i>Total</i> )	abn - 1	SST <sub>0</sub>			

Trong đó

n số lần lặp lại

a số nghiệm thức A

b số nghiệm thức B

**Bảng 8.2:** Bảng ANOVA của thí nghiệm hai yếu tố theo kiểu khối ngẫu nhiên đầy đủ (RCBD)

<b>Nguồn biến động</b> <i>Source of variation</i>	<b>Độ tự do</b> <i>Degree of freedom</i>	<b>Tổng bình phương</b> <i>Sum of Square</i>	<b>Trung bình bình phương</b> <i>Mean Square</i>	<b>F<sub>c</sub></b>	<b>F<sub>bảng</sub></b>
Khối	n - 1				
Yếu tố A	a - 1	SST <sub>A</sub>	MST <sub>A</sub>	MST <sub>A</sub> /MSE	F <sub>dfA,dfE</sub>
Yếu tố B	b - 1	SST <sub>B</sub>	MST <sub>B</sub>	MST <sub>B</sub> /MSE	F <sub>dfB,dfE</sub>
Tương tác A*B	(a - 1)(b - 1)	SST <sub>AB</sub>	MST <sub>AB</sub>	MST <sub>AB</sub> /MSE	F <sub>dfAB,dfE</sub>
Sai số ( <i>Error</i> )	(ab - 1)(n - 1)	SSE	MSE		
Tổng ( <i>Total</i> )	abn - 1	SST <sub>0</sub>			

Trong đó

n số khối

a số mức yếu tố A

b số mức yếu tố B

### III. So sánh các số trung bình

#### 1. Đối với yếu tố A

Giả thuyết

$$H_0 \quad \mu_1 = \mu_2 = \dots = \mu_a$$

$$H_1 \quad \mu_i \neq \mu_{i'} \text{ có ít nhất 1 cặp } (i, i') \text{ khác nhau}$$

Nếu  $F_c = MST_A/MSE > F_{dfA,dfE}$ .

**Hay**  $P\text{-value}_A < \alpha$

⇒ Bác bỏ giả thuyết  $H_0$ , yếu tố A có ảnh hưởng đến thí nghiệm.

#### 2. Đối với yếu tố B



Giả thuyết

$$H_0 \quad \mu_1 = \mu_2 = \dots = \mu_b$$

$$H_1 \quad \mu_j \neq \mu_{j'} \text{ có ít nhất 1 cặp } (j, j') \text{ khác nhau}$$

Nếu  $F_c = MST_B/MSE > F_{dfB, dfE}$ .

**Hay**  $P\text{-value}_B < \alpha$

$\Rightarrow$  Bác bỏ giả thuyết  $H_0$ , yếu tố B có ảnh hưởng đến thí nghiệm.

### 3. Đối với yếu tố tương tác

Giả thuyết

$$H_0 \quad \mu_1 = \mu_2 = \dots = \mu_{a,b}$$

$$H_1 \quad \mu_t \neq \mu_{t'} \text{ có ít nhất 1 cặp } (t, t') \text{ khác nhau}$$

#### 3.1. Trường hợp tương tác không có ý nghĩa

Nếu  $F_c = MST_{AB}/MSE < F_{dfAB, dfE}$ .

**Hay**  $P\text{-value}_{(A*B)} > \alpha$

$\Rightarrow$  Bác bỏ giả thuyết  $H_0$ , tương tác không có ý nghĩa.

So sánh  $\mu_i$  và  $\mu_{i'}$  trong a số trung bình của yếu tố A

So sánh  $\mu_j$  và  $\mu_{j'}$  trong b số trung bình của yếu tố B

#### 3.2. Trường hợp tương tác có ý nghĩa

Nếu  $F_c = MST_{AB}/MSE > F_{dfAB, dfE}$

**Hay**  $P\text{-value}_{(A*B)} < \alpha$

$\Rightarrow$  Không bác bỏ giả thuyết  $H_0$ , tương tác có ý nghĩa.

So sánh các số trung bình trong ab nghiệm thức.

#### IV. Xử lý bằng phần mềm SPSS

##### 1. Thí nghiệm 2 yếu tố, bố trí kiểu CRD

##### 1.1. Thí nghiệm 2 yếu tố, bố trí kiểu CRD. Trường hợp tương tác không có ý nghĩa.

VD 8.2: Số liệu bảng 8.3

**Bảng 8.3**

CRD			
Yếu tố A	Yếu tố B		
	b <sub>1</sub>	b <sub>2</sub>	b <sub>3</sub>
a <sub>1</sub>	a <sub>1</sub> b <sub>1</sub>	a <sub>1</sub> b <sub>2</sub>	a <sub>1</sub> b <sub>3</sub>
	5,5	4,5	3,5
	5,5	4,5	4,0
a <sub>2</sub>	a <sub>2</sub> b <sub>1</sub>	a <sub>2</sub> b <sub>2</sub>	a <sub>2</sub> b <sub>3</sub>
	6,0	4,0	3,0
	6,5	5,0	4,0
a <sub>2</sub>	7,0	5,5	5,0
	7,0	5,0	4,5

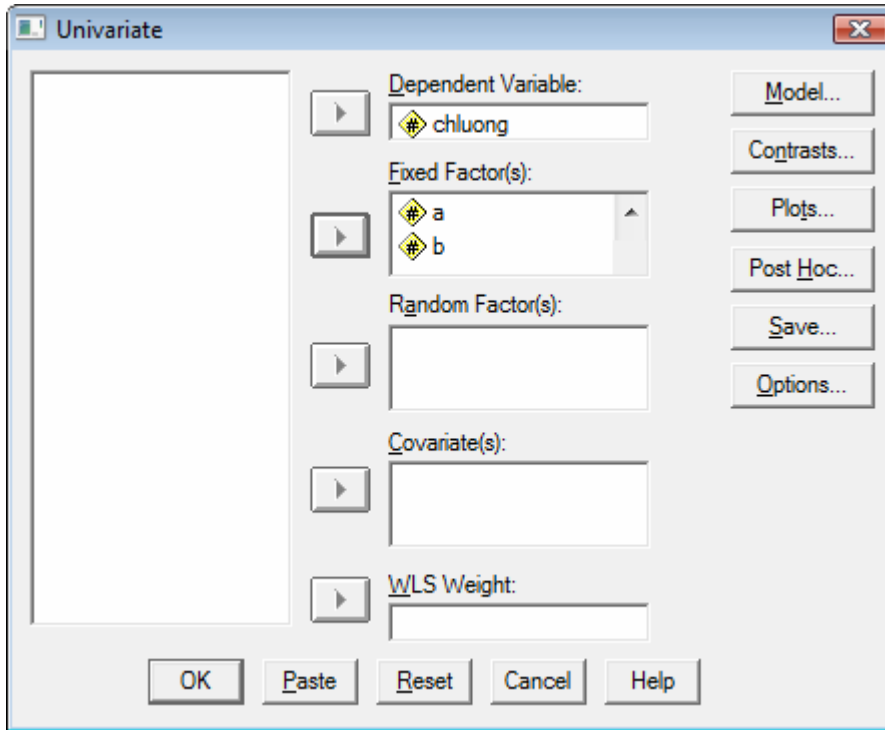
Có hai yếu tố A và B. Trong đó yếu tố A có hai mức yếu tố và yếu tố B có 3 mức yếu tố do đó có 6 nghiệm thức, mỗi nghiệm thức lặp lại 3 lần, có 18 đơn vị thí nghiệm.

##### 1.1.1. Nhập số liệu

	a	b	chluong	var
1	1	1	5,5	
2	1	2	4,5	
3	1	3	3,5	
4	1	1	5,5	
5	1	2	4,5	
6	1	3	4,0	
7	1	1	6,0	
8	1	2	4,0	
9	1	3	3,0	
10	2	1	6,5	
11	2	2	5,0	
12	2	3	4,0	
13	2	1	7,0	
14	2	2	5,5	
15	2	3	5,0	
16	2	1	7,0	
17	2	2	5,0	
18	2	3	4,5	
19				

### 1.1.2. Phân tích phương sai (ANOVA)

Analyze/General Linear Model/Univariate...



#### Tests of Between-Subjects Effects

Dependent Variable: CHLUONG

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	20,333(a)	5	4,067	29,280	,000
Intercept	450,000	1	450,000	3240,000	,000
A	4,500	1	4,500	32,400	,000
B	15,750	2	7,875	56,700	,000
A * B	,083	2	,042	,300	<b>,746</b>
Error	1,667	12	,139		
Total	472,000	18			
Corrected Total	22,000	17			

a R Squared = ,924 (Adjusted R Squared = ,893)

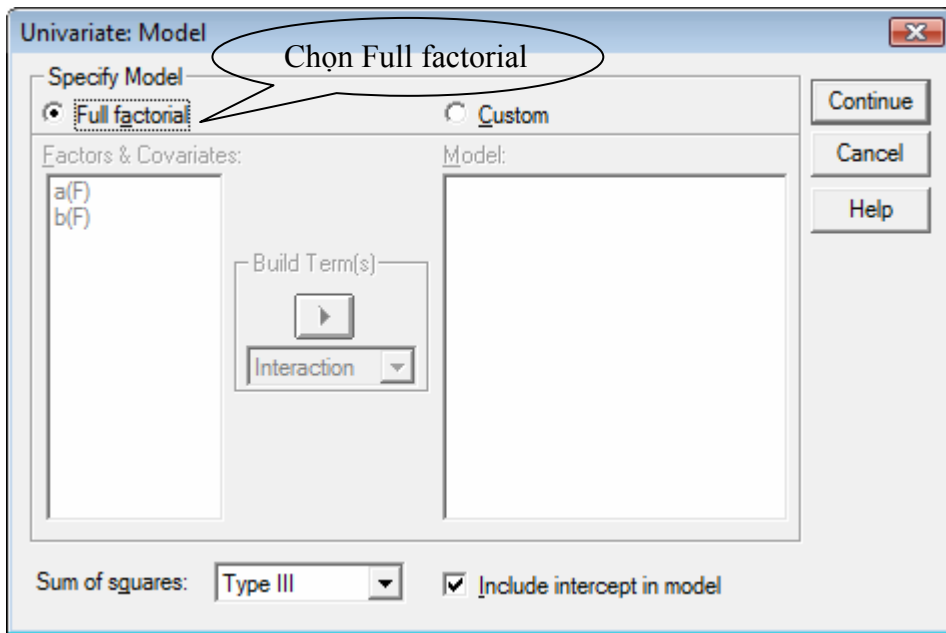
P-value = 0,746 >  $\alpha$ . Tương tác không có ý nghĩa.

Yếu tố A và B đều có ảnh hưởng đến thí nghiệm.

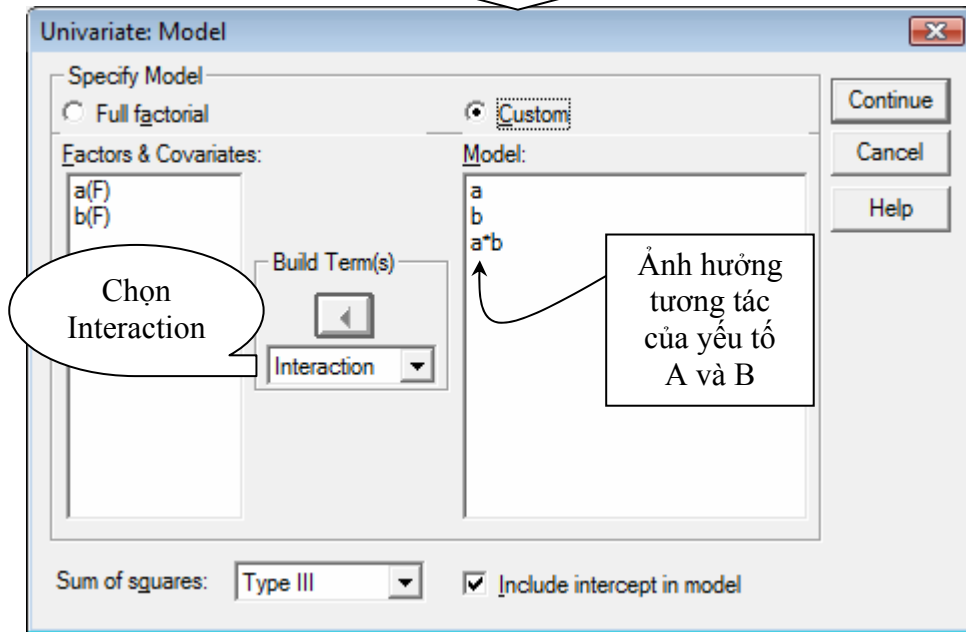
Vì tương tác không có ý nghĩa do đó so sánh các số trung bình của yếu tố A và các số trung bình của yếu tố B.

### 1.1.3. So sánh sự khác biệt của nghiệm thức bằng phương pháp LSD và Duncan

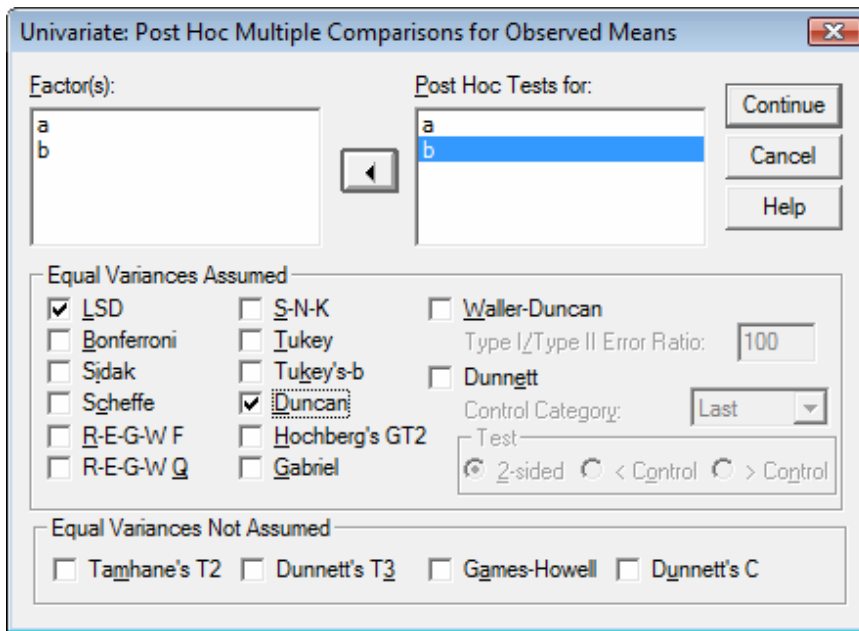
Chọn Model...



Hoặc  
chọn Custom



Nhấp **C**ontinue/Post **H**oc...



### Warnings

Post hoc tests are not performed for A because there are fewer than three groups.

## Post Hoc Tests

### B

#### Multiple Comparisons

Dependent Variable: CHLUONG

	(I) B	(J) B	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
LSD	1	2	1,500(*)	,2152	,000	1,031	1,969
		3	2,250(*)	,2152	,000	1,781	2,719
	2	1	-1,500(*)	,2152	,000	-1,969	-1,031
		3	,750(*)	,2152	,004	,281	1,219
	3	1	-2,250(*)	,2152	,000	-2,719	-1,781
		2	-,750(*)	,2152	,004	-1,219	-,281

Based on observed means.

\* The mean difference is significant at the ,05 level.

## Homogeneous Subsets

CHLUONG

	B	N	Subset		
			1	2	3
Duncan(a,b)	3	6	4,000		
	2	6		4,750	
	1	6			6,250
	Sig.			1,000	1,000

a Uses Harmonic Mean Sample Size = 6,000.

b Alpha = ,05.

## 1.2. Thí nghiệm 2 yếu tố, bố trí kiểu CRD. Trường hợp tương tác có ý nghĩa.

**VD 8.3:** Thí nghiệm 2 yếu tố để tìm ảnh hưởng của nhiệt độ sấy và xử lý ngâm đường đến độ mềm của sản phẩm sấy. Quả được cắt lát với bề dày như nhau rồi đem sấy (không xử lý ngâm đường), hoặc có xử lý ngâm ở dịch đường nồng độ 25% (yếu tố A). Yếu tố B là nhiệt độ sấy: 50°C và 60°C. Mỗi thí nghiệm được lặp lại 3 lần. Các yếu tố khác được giữ không đổi. Thí nghiệm được bố trí hoàn toàn ngẫu nhiên (CRD). Độ mềm sản phẩm được đánh giá qua lực cắt (Newton). Kết quả trong bảng sau:

**Bảng 8.4:** Lực cắt đo từ sản phẩm sấy (N)

Yếu tố B (°C)	Yếu tố A (%)	
	0	25
50	<b>a<sub>1</sub>b<sub>1</sub></b>	<b>a<sub>2</sub>b<sub>1</sub></b>
	150,6	173,8
	128,9	161,3
60	<b>a<sub>1</sub>b<sub>2</sub></b>	<b>a<sub>2</sub>b<sub>2</sub></b>
	137,2	155,6
	225,1	320,4
60	210,6	301,9
	218,5	296,4

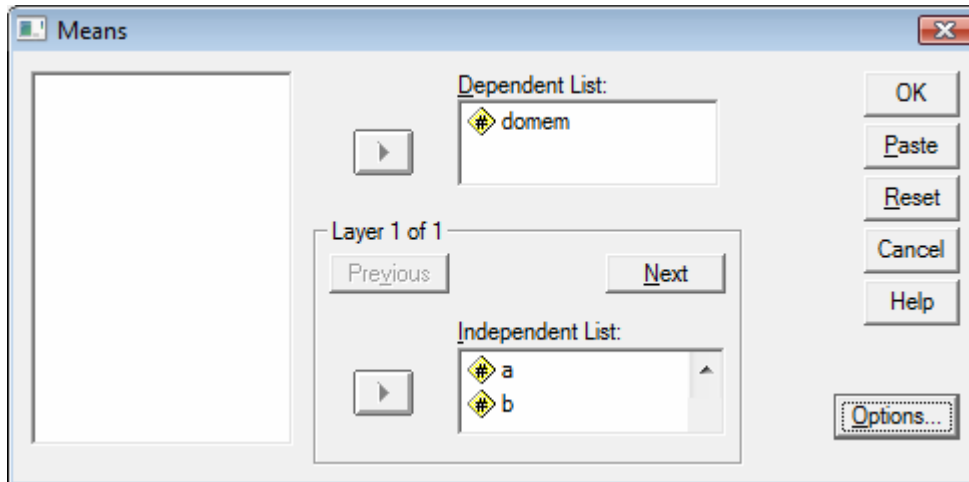
**Bảng 8.5:** Trung bình độ mềm theo từng yếu tố tác động

B	A		Tổng	Trung bình
	0%	25%		
50°C	416,7	490,7	907,4	<b>151,2</b>
60°C	654,2	918,7	1572,9	<b>262,2</b>
Tổng	1070,9	1409,4	2480,3	
<b>Trung bình</b>	<b>178,5</b>	<b>234,9</b>		<b>206,69</b>

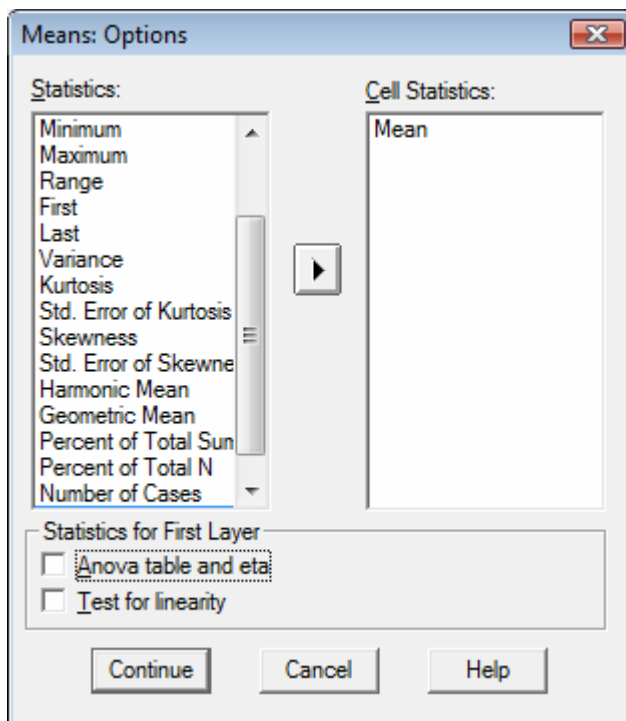
### 1.2.1. Nhập số liệu

	a	b	domem
1	1	1	150,6
2	1	1	128,9
3	1	1	137,2
4	1	2	225,1
5	1	2	210,6
6	1	2	218,5
7	2	1	173,8
8	2	1	161,3
9	2	1	155,6
10	2	2	320,4
11	2	2	301,9
12	2	2	296,4
13			

## Analyze/Compare Means/Means...



## Chọn Options...



## Kết quả

DOMEM \* A

Mean

A	DOMEM
1	178,483
2	234,900
Total	206,692

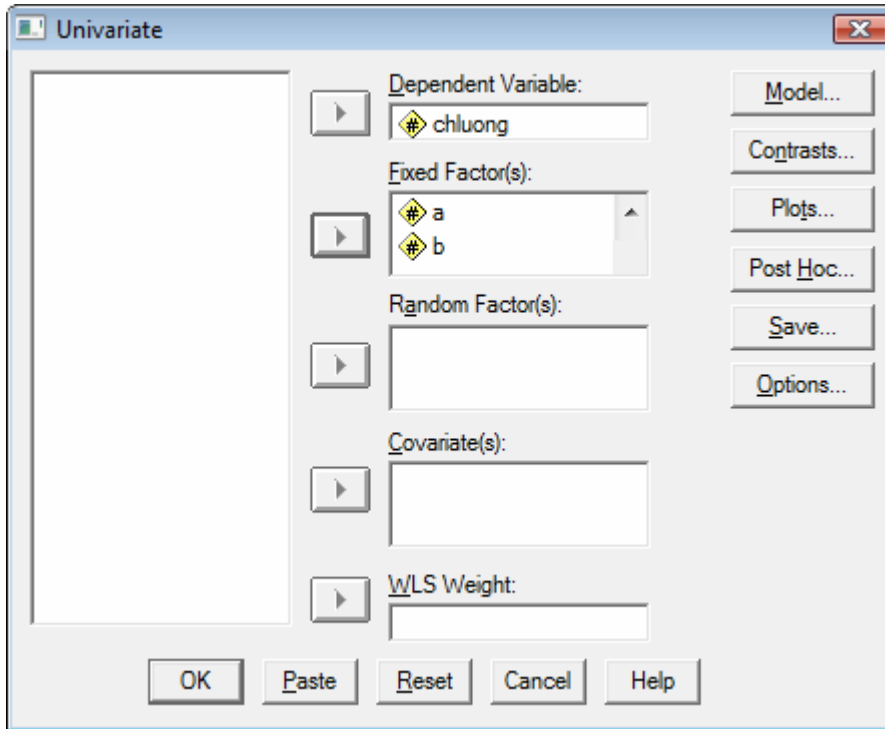
DOMEM \* B

Mean

B	DOMEM
1	151,233
2	262,150
Total	206,692

### 1.2.2. Phân tích phương sai (ANOVA)

Analyze/General Linear Model/Univariate...



#### Tests of Between-Subjects Effects

Dependent Variable: DOMEM

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	49480,229(a)	3	16493,410	158,081	,000
Intercept	512657,341	1	512657,341	4913,570	,000
A	9548,521	1	9548,521	91,518	,000
B	36907,521	1	36907,521	353,741	,000
A * B	3024,188	1	3024,188	28,985	,001
Error	834,680	8	104,335		
Total	562972,250	12			
Corrected Total	50314,909	11			

<  $\alpha$   
<  $\alpha$   
<  $\alpha$

a R Squared = ,983 (Adjusted R Squared = ,977)



## Kiểm định F

Tất cả các Sig. <  $\alpha$ .

Ảnh hưởng của các yếu tố và tương tác lên kết quả thí nghiệm là có ý nghĩa.

### Nhận xét:

Việc xử lý ngâm đường có ảnh hưởng đáng kể đến độ mềm của sản phẩm quả cắt lát sau khi sấy ở độ tin cậy 95%. Xử lý ngâm đường ở nồng độ 25% là giảm độ mềm của sản phẩm (trung bình 234,9 N) so với không xử lý (trung bình 178,5 N).

Nhiệt độ sấy ảnh hưởng có ý nghĩa đến độ mềm của sản phẩm sau khi sấy ở độ tin cậy 95%. Nhiệt độ 60°C làm giảm độ mềm của sản phẩm (trung bình 262,2 N) sau khi sấy so với ở nhiệt độ 50°C (trung bình 151,2 N).

Hai yếu tố nhiệt độ và xử lý ngâm đường có tương tác với nhau.

### 1.2.3. So sánh sự khác biệt của nghiệm thức bằng phương pháp LSD và Duncan

Đối với VD này không cần sử dụng phương pháp LSD và Duncan để so sánh sự khác biệt về trung bình của yếu tố A và B, chỉ cần sử dụng kiểm định F. Vì mỗi yếu tố A và B đều chỉ có hai mức yếu tố.

#### 1.2.3.1. So sánh sự khác biệt về trung bình của yếu tố A và B

Đối với VD này để so sánh sự khác biệt về trung bình của yếu tố A và B, chỉ cần sử dụng kiểm định F.

#### 1.2.3.2. So sánh sự khác biệt về trung bình của yếu tố tương tác (a\*b)

Tương tác có ý nghĩa khi đó thực hiện việc so sánh trung bình của 4 nghiệm thức

$a_1b_1$	$a_1b_2$	$a_2b_1$	$a_2b_2$
1	2	3	4

### Nhập lại số liệu

	ab	domem	va
1	1	150,6	
2	1	128,9	
3	1	137,2	
4	2	225,1	
5	2	210,6	
6	2	218,5	
7	3	173,8	
8	3	161,3	
9	3	155,6	
10	4	320,4	
11	4	301,9	
12	4	296,4	

## Analyze/Compare Means/One Way ANOVA...

### Post Hoc Tests

#### Multiple Comparisons

Dependent Variable: DOMEM

	(I) AB	(J) AB	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
LSD	1	2	-79,167(*)	8,3401	,000	-98,399	-59,934
		3	-24,667(*)	8,3401	,018	-43,899	-5,434
		4	-167,333(*)	8,3401	,000	-186,566	-148,101
	2	1	79,167(*)	8,3401	,000	59,934	98,399
		3	54,500(*)	8,3401	,000	35,268	73,732
		4	-88,167(*)	8,3401	,000	-107,399	-68,934
	3	1	24,667(*)	8,3401	,018	5,434	43,899
		2	-54,500(*)	8,3401	,000	-73,732	-35,268
		4	-142,667(*)	8,3401	,000	-161,899	-123,434
	4	1	167,333(*)	8,3401	,000	148,101	186,566
		2	88,167(*)	8,3401	,000	68,934	107,399
		3	142,667(*)	8,3401	,000	123,434	161,899

\* The mean difference is significant at the .05 level.

#### DOMEM

	AB	N	Subset for alpha = .05			
			1	2	3	4
Duncan(a)	1	3	138,900			
	3	3		163,567		
	2	3			218,067	
	4	3				306,233
	Sig.			1,000	1,000	1,000

Means for groups in homogeneous subsets are displayed.

a Uses Harmonic Mean Sample Size = 3,000.

Trung bình các nghiệm thức  $a_1b_1$ ,  $a_1b_2$ ,  $a_2b_1$ ,  $a_2b_2$  đều khác biệt ở mức ý nghĩa  $\alpha = 5\%$ .

## 2. Thí nghiệm 2 yếu tố, bố trí kiểu RCBD

VD 8.4: Sử dụng số liệu bảng 8.6

**Bảng 8.6**

RCBD				
Yếu tố A	Yếu tố B			Khối
	b <sub>1</sub>	b <sub>2</sub>	b <sub>3</sub>	
a <sub>1</sub>	a <sub>1</sub> b <sub>1</sub>	a <sub>1</sub> b <sub>2</sub>	a <sub>1</sub> b <sub>3</sub>	
	5,5	4,5	3,5	<b>1</b>
	5,5	4,5	4,0	<b>2</b>
	6,0	4,0	3,0	<b>3</b>
a <sub>2</sub>	a <sub>2</sub> b <sub>1</sub>	a <sub>2</sub> b <sub>2</sub>	a <sub>2</sub> b <sub>3</sub>	
	6,5	5,0	4,0	<b>1</b>
	7,0	5,5	5,0	<b>2</b>
	7,0	5,0	4,5	<b>3</b>

Chất lượng = f(Yếu tố A, yếu tố B)

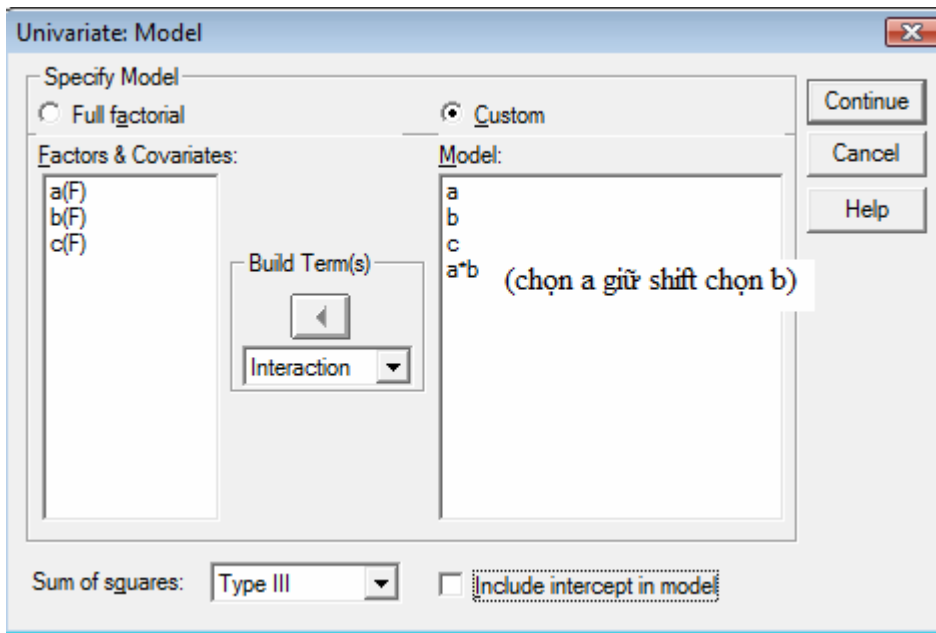
{yếu tố ngoại cảnh: yếu tố C}

### 2.1. Nhập số liệu

	a	b	c	chluong
1	1	1	1	5,5
2	1	1	2	5,5
3	1	1	3	6,0
4	1	2	1	4,5
5	1	2	2	4,5
6	1	2	3	4,0
7	1	3	1	3,5
8	1	3	2	4,0
9	1	3	3	3,0
10	2	1	1	6,5
11	2	1	2	7,0
12	2	1	3	7,0
13	2	2	1	5,0
14	2	2	2	5,5
15	2	2	3	5,0
16	2	3	1	4,0
17	2	3	2	5,0
18	2	3	3	4,5

### 2.2. Phân tích phương sai (ANOVA)

Analyze/General Linear Model/Univariate...



### Tests of Between-Subjects Effects

Dependent Variable: CHLUONG

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Model	470,917(a)	8	58,865	543,365	,000
A	4,500	1	4,500	41,538	,000
B	15,750	2	7,875	72,692	,000
C	,583	2	,292	2,692	,116
A * B	,083	2	,042	,385	,690
Error	1,083	10	,108		
Total	472,000	18			

$> \alpha$

a R Squared = ,998 (Adjusted R Squared = ,996)

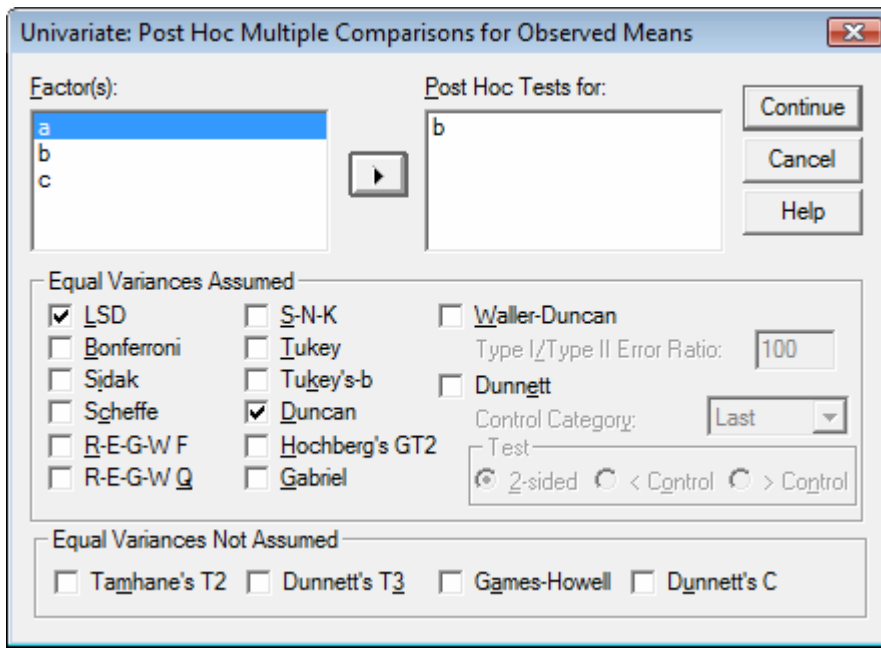
Sig. = 0,69  $> \alpha$

⇒ Tương tác không có ý nghĩa.

Vậy ta so sánh sự khác biệt trung bình của yếu tố A và B.

### 2.3. So sánh sự khác biệt của nghiệm thức bằng phương pháp LSD và Duncan

So sánh sự khác biệt trung bình của yếu tố A và B. Trong VD này chỉ cần so sánh sự khác biệt trung bình của yếu tố B bằng phương pháp LSD và Duncan, đối với yếu tố A chỉ cần sử dụng kiểm định F. Vì yếu tố A chỉ có hai mức yếu tố.



### Multiple Comparisons

Dependent Variable: CHLUONG

	(I) B	(J) B	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
LSD	1	2	1,500(*)	,1900	,000	1,077	1,923
		3	2,250(*)	,1900	,000	1,827	2,673
	2	1	-1,500(*)	,1900	,000	-1,923	-1,077
		3	,750(*)	,1900	,003	,327	1,173
	3	1	-2,250(*)	,1900	,000	-2,673	-1,827
		2	-,750(*)	,1900	,003	-1,173	-,327

Based on observed means.

\* The mean difference is significant at the ,05 level.

### Homogeneous Subsets

CHLUONG

	B	N	Subset		
			1	2	3
Duncan(a,b)	3	6	4,000		
	2	6		4,750	
	1	6			6,250
	Sig.		1,000	1,000	1,000

Means for groups in homogeneous subsets are displayed. Based on Type III Sum of Squares The error term is Mean Square(Error) = ,108.

a Uses Harmonic Mean Sample Size = 6,000.

b Alpha = ,05.