

15

Thiết kế thí nghiệm (Design of experiments)

Cụm từ “thí nghiệm” ở đây không chỉ bao gồm các hoạt động trong phòng thí nghiệm, mà còn bao gồm cả những công trình khảo sát rộng lớn hơn như thử nghiệm lâm sàng đối chứng ngẫu nhiên (randomized clinical trial), các công trình nghiên cứu tiêu biểu một thời điểm (còn gọi là nghiên cứu cắt ngang hay cross-sectional study), thăm dò ý kiến, điều tra về điều tra dân số, v.v... Ngay cả một chính sách kinh tế cũng có thể xem là một thí nghiệm – thí nghiệm xã hội.

Một thí nghiệm đạt tiêu chuẩn khoa học phải là một thí nghiệm được thiết kế có hệ thống và khách quan. Chẳng hạn như để biết tỉ lệ mắc bệnh đái đường trong một quần thể, chúng ta không cần phải khám nghiệm tất cả cá nhân trong quần thể đó, mà chỉ chọn ngẫu nhiên một số cá nhân đại diện. Tuy nhiên nếu số lượng cá nhân đại diện (còn gọi là *mẫu*) quá thấp thì công trình nghiên cứu sẽ không cho kết quả chính xác; ngược lại nếu số lượng mẫu quá lớn, chúng ta sẽ phung phí tiền bạc và cơ sở vật chất một cách không cần thiết. Do đó, mục tiêu của thiết kế nghiên cứu là (i) để phát hiện một ảnh hưởng hay tác dụng của một can thiệp, và (ii) sử dụng cơ sở vật chất và tài lực một cách tối ưu.

Qua các chương trước, chúng ta đã làm quen với một số mô hình phân tích số liệu. Kết quả của các phân tích này chỉ có giá trị khoa học khi số liệu được thu thập đúng phương pháp, và khi công trình nghiên cứu được thiết kế một cách tối ưu. Các mô hình thống kê không thể cung cấp cho chúng ta thông tin về chất lượng của nghiên cứu, vì đây là một khía cạnh cần sự thẩm định cẩn thận của nhà nghiên cứu. Do đó, thiết kế nghiên cứu, đóng một vai trò rất quan trọng cho việc thành bại của một công trình khoa học. Có thể nói rằng một nghiên cứu nếu được thiết kế cẩn thận và đúng phương pháp thì mức độ thành công đã đạt được 50%. Chương này và chương sau sẽ bàn qua một số khái niệm căn bản về thiết kế nghiên cứu và một số mô hình nghiên cứu thông dụng.

15.1 Thuật ngữ

Để thuận tiện cho việc theo dõi và quán triệt các khái niệm nghiên cứu, có lẽ chúng ta phải làm quen và phân biệt được một số thuật ngữ quan trọng trong khi thiết kế một nghiên cứu.

Đơn vị nghiên cứu (experimental unit): Tùy theo lĩnh vực nghiên cứu, đơn vị nghiên cứu có thể là đối tượng (như bệnh nhân hay tình nguyện

viên), mẫu ruộng, sản phẩm, qui trình sản xuất, v.v... Đơn vị nghiên cứu là đối tượng sử dụng trực tiếp cho việc đo lường. Chẳng hạn như, trong nghiên cứu về vị đắng của cà phê, nhà nghiên cứu có thể cho một nhóm người tiêu thụ nếm thử nhiều loại cà phê khác nhau, và các loại cà phê này chính là đơn vị nghiên cứu. Trong các nghiên cứu lâm sàng, nhà nghiên cứu có thể chọn hai nhóm bệnh nhân để so sánh hiệu quả của hai thuật điều trị, và trong trường hợp này, mỗi bệnh nhân là một đơn vị nghiên cứu.

Yếu tố can thiệp (factors): là những can thiệp (intervention) áp dụng trên các đối tượng nghiên cứu. Yếu tố can thiệp còn có khi được gọi là biến độc lập (independent variable) hay biến giải thích (explanatory variable). Trong ví dụ nghiên cứu lâm sàng vừa đề cập trên, hai thuật điều trị là yếu tố can thiệp. Hay trong nghiên cứu về hiệu quả của hai loại giống lúa, thì giống lúa được xem là yếu tố can thiệp.

Mức độ can thiệp (treatment levels): là những “giá trị” của một yếu tố can thiệp. Chẳng hạn như nếu hai thuật điều trị là hai loại thuốc, và mỗi loại thuốc có 3 liều lượng, thì liều lượng là mức độ can thiệp. Hay trong nghiên cứu cảm quan, nhà nghiên cứu có thể cho người tiêu thụ nếm thử vị ngọt của một loại bia, nhưng bia được sản xuất với ba công thức khác nhau, thì công thức chính là mức độ can thiệp.

Nhóm (block): Trong nhiều nghiên cứu, một nhóm yếu tố can thiệp có thể sắp đặt thành từng nhóm (hay khối). Chẳng hạn như trong một nghiên cứu cảm quan về vị đắng của 3 loại cà phê (A, B và C), nhà nghiên cứu có thể chọn một số đối tượng nghiên cứu (người tiêu thụ) và chia đối tượng thành ba nhóm 1, 2 và 3 như sau:

	Nhóm 1	Nhóm 2	Nhóm 3
Can thiệp	A, B, C	A, B, C	A, B, C

Trong phương án này, cá nhân trong mỗi nhóm đều thử nghiệm tất cả 3 loại cà phê, và thứ tự A, B, C không thay đổi giữa các nhóm. Phương án này còn có tên là *balance complete block design* (phương án cân đối nhóm).

Hoặc nhà nghiên cứu có thể chọn 2 loại cà phê cho ba nhóm:

	Nhóm 1	Nhóm 2	Nhóm 3
Can thiệp	A, B	B, C	A, C

Trong phương án này, mỗi nhóm chỉ thử nghiệm 2 loại cà phê, nhưng thứ tự loại cà phê được thay đổi theo từng nhóm. Phương án này còn có tên là *balance incomplete block design* (phương án cân đối nhóm không đầy đủ).

Phương án cân đối nhóm còn được sử dụng khá phổ biến trong các nghiên cứu lâm sàng. Chẳng hạn như nghiên cứu thử nghiệm hiệu quả của hai loại thuốc điều trị bệnh loãng xương, nhà nghiên cứu có thể chọn 100 bệnh nhân, và chia thành 5 nhóm (mỗi nhóm có 20 người). Trong mỗi nhóm, 10 người được điều trị bằng thuốc A và 10 người được điều trị bằng thuốc B. Phân nhóm phải được tiến hành hoàn toàn ngẫu nhiên để đảm bảo tính khách quan của nghiên cứu.

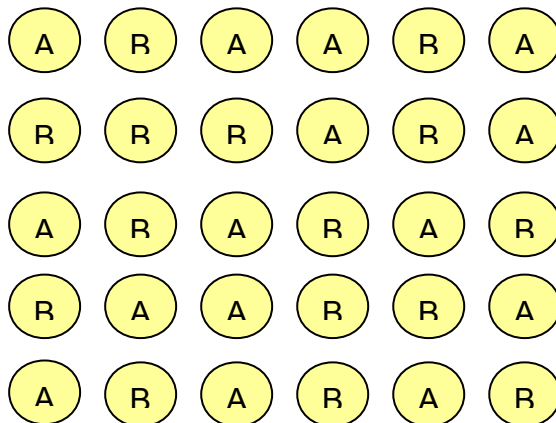
Tiêu chí (response variable): là biến số chịu ảnh hưởng của yếu tố can thiệp. Chẳng hạn như trong nghiên cứu cảm quan về vị đắng của cà phê thì vị đắng là tiêu chí nghiên cứu; hay trong nghiên cứu về hiệu quả của hai thuật điều trị bệnh loãng xương thì mật độ xương (bone mineral density) là tiêu chí.

Ví dụ 1: Một thí nghiệm cảm quan đơn giản. Để biết người tiêu thụ đánh giá độ ngọt của một loại nước ngọt thế nào, các nhà nghiên cứu sản xuất ra hai loại nước ngọt với công thức A và B. Trong thí nghiệm, người tiêu thụ được cho thử nước ngọt và cho điểm độ ngọt (từ 1 = “không ngọt” đến 10 = “quá ngọt”) như sau. Vấn đề đặt ra là tìm một phương án nghiên cứu sao cho lượng thông tin thu thập được tối đa và đảm bảo tiêu chuẩn khoa học.

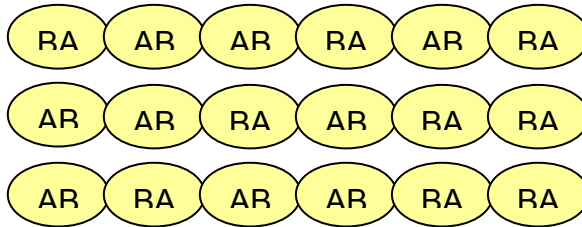
1 (không ngọt)	2	3	4	5	6	7	8	9	10 (quá ngọt)
----------------------	---	---	---	---	---	---	---	---	---------------------

Phương án 1: các nhà nghiên cứu ngẫu nhiên mời n (n có thể là 15) khách hàng và cho mỗi khách hàng uống thử cả hai loại nước ngọt và phân tích khác biệt về độ ngọt giữa hai sản phẩm từ mỗi người.

Phương án 2: ngẫu nhiên chọn $2n$ (hay 30 người), rồi ngẫu nhiên chia thành 2 nhóm. Nhóm 1 uống nước ngọt công thức A, và nhóm 2 uống nước ngọt công thức B như sau:



Phương án 3: ngẫu nhiên chọn n (hay 15 khách hàng); mỗi khách hàng được cho uống hai loại nước ngọt, nhưng thứ tự AB và BA được phân chia một cách ngẫu nhiên như sau. Phương án này có 2 yếu tố can thiệp (A và B) cho mỗi nhóm (block). Nói cách khác, mỗi khách hàng là một “nhóm”.



Mỗi phương án trên đều có lợi thế và bất tiện. Thứ nhất, về mặt cơ sở vật chất và chi phí, phương án 2 đòi hỏi số lượng đối tượng nghiên cứu cao gấp hai lần phương án 1, đắt tiền hơn và tốn nhiều thời gian hơn.

Thứ hai, về mặt khoa học, phương án 2 đòi hỏi nhà nghiên cứu phải so sánh hai nhóm một cách độc lập, và độ nhiễu thông tin (noise) của phương án này chắc chắn phải cao hơn độ nhiễu của phương án 1 và 3. “Độ nhiễu” ở đây có thể đo bằng phương sai (variance). Để hiểu khái niệm quan trọng này, chúng ta cần phải đi qua một khái niệm thống kê học căn bản. Gọi tiêu chí đo độ ngọt của hai nhóm là x_1 và x_2 ; Gọi phương sai của độ ngọt của hai nhóm là s_1^2 và s_2^2 . Bởi vì theo phương án 2, hai nhóm độc lập nhau (tức là khách hàng thử sản phẩm A không phải là khách hàng thử sản phẩm B) cho nên phương sai của độ khác biệt giữa hai sản phẩm $x_1 - x_2$ (kí hiệu $s_{x_1 - x_2}^2$) là:

$$s_{x_1 - x_2}^2 = s_1^2 + s_2^2 \quad [1]$$

Nếu phương sai của hai nhóm bằng nhau $s_1^2 = s_2^2 = s^2$, thì phương sai của độ khác biệt đơn giản là:

$$s_{x_1 - x_2}^2 = 2 s^2.$$

Nhưng với phương án 1, bởi vì mỗi khách hàng thử cả hai sản phẩm, do đó, x_1 và x_2 **không độc lập** với nhau, và phương sai của độ khác biệt là:

$$s_{x_1 - x_2}^2 = s_1^2 + s_2^2 - 2 \text{cov}(x_1, x_2) \quad [2]$$

Trong đó, $\text{cov}(x_1, x_2)$ có nghĩa là hiệp biến (covariance), tức phản ánh độ tương quan giữa x_1 và x_2 . Bởi vì độ tương quan giữa x_1 và x_2 chắc chắn phải là một số dương (lớn hơn 0); Do đó, phương sai trong công thức [2] luôn luôn nhỏ hơn phương sai trong công thức [1].

Nói cách khác, độ nhiễu thông tin của phương án 1 và 3 lúc nào cũng nhỏ hơn độ nhiễu của phương án 2. Do đó, phương án 1 và 3 có ưu thế hơn phương án 2.

Thứ ba, phương án 1 và 3 giống nhau ở điểm mỗi khách hàng uống thử cả hai loại sản phẩm, nhưng phương án 3 thì thứ tự sản phẩm được thay đổi ngẫu nhiên (chứ không cố định như phương án 1). Sự thay đổi ngẫu nhiên như từ A sang B (và B sang A) cũng có thể xem là một cách “blocking” (phân nhóm), do đó, nhà nghiên cứu có thể kiểm soát thêm một nguồn dao động quan trọng.

Vì vậy, trong ba phương án này, có thể nói phương án 3 là tối ưu nhất. Nhưng tất nhiên vấn đề còn tùy thuộc vào đặc tính của sản phẩm và tình hình thực tế. Có nhiều sản phẩm mà phương án 1 và 3 không thể áp dụng vì lí do an toàn hay hiệu ứng “hawthorne” (sẽ bàn trong phần dưới đây).

15.2 Ba nguyên tắc quan trọng của một nghiên cứu

Một nghiên cứu khoa học phải tuân thủ theo ba nguyên tắc: ngẫu nhiên hóa (randomization), lặp lại nhiều lần (replication), và phân nhóm (blocking).

Tại sao phải ngẫu nhiên hóa? Trong nhiều nghiên cứu, chúng ta phải lấy mẫu (sample) từ một quần thể (population). Một trong những yêu cầu quan trọng của lấy mẫu là mẫu phải mang tính đại diện cho quần thể. Chẳng hạn như nếu trong quần thể 1 triệu người có 50% nam và 20% người có trình độ văn hóa cao hơn lớp 12. Nếu chúng ta chọn 100 người từ quần thể này, mẫu được chọn được xem là đại diện khi có khoảng 50 nam và 20 người có học vấn trên lớp 12. Chọn mẫu ngẫu nhiên là phương án tốt nhất để đảm bảo tính đại diện này.

Đối với một nhóm đối tượng, ngẫu nhiên hóa còn có khả năng cân đối các đặc điểm giữa các nhóm can thiệp. Giả dụ chúng ta đã mời được một nhóm gồm 50 tình nguyện viên sẵn sàng tham gia vào một công trình nghiên cứu cảm quan để thử vị chua của 2 loại nước giải khát (nói cách khác, chúng ta có 2 nhóm, và mỗi nhóm có 25 người). Dĩ nhiên 50 người này có nhiều đặc tính cá nhân khác nhau, chẳng hạn như độ tuổi, giới tính, trình độ văn hóa, sở thích cá nhân, v.v... tất cả những đặc tính này có thể có ảnh hưởng đến cảm nhận về sản phẩm. Do đó, để “cân đối” các đặc tính này cho hai nhóm, cách duy nhất và khách quan nhất là phân chia họ thành hai nhóm một cách ngẫu nhiên.

Vì phần lớn các mô hình phân tích thống kê dựa vào giả định rằng đối tượng được chọn ngẫu nhiên từ một quần thể, cho nên ngẫu nhiên hóa còn đảm bảo tính hợp lí của kết quả phân tích.

Một trong những “tiêu chuẩn vàng” của khoa học là kết quả nghiên cứu phải có tính *có thể lặp lại* (repeatability) hay tái xác nhận. Nói một cách khác, nếu có một nghiên cứu đã được công bố bởi một nhà khoa học nào đó; Nếu một nhà nghiên cứu khác lặp lại nghiên cứu đó bằng những phương pháp và với điều kiện đã được mô tả, phải đạt được những kết quả tương tự. Đó là một tiêu chuẩn cực kì quan trọng để phân biệt giữa khoa học và ngụy khoa học (pseudoscience). Một quan sát được lặp lại nhiều lần thì quan sát đó có độ tin cậy cao. Và độ tin cậy cao cho phép kết luận nghiên cứu có giá trị cao.

Ngẫu nhiên hóa có thể làm cân đối các đặc điểm của đối tượng nghiên cứu cho các yếu tố can thiệp, nhưng với điều kiện số lượng đối tượng phải tương đối lớn. Khi số lượng đối tượng nghiên cứu nhỏ, thì ngẫu nhiên hóa không có hiệu quả cao. Chẳng hạn như với 6 đối tượng chia thành 2 nhóm, ngẫu nhiên hóa có thể cho ra kết quả 4 đối tượng thuộc nhóm A và 2 đối tượng thuộc nhóm B. Do đó, một cách khác để đảm bảo tính cân đối là phân nhóm. Trong trường hợp trên, chúng ta có thể chia thành 3 nhóm (mỗi nhóm 2 đối tượng), và ngẫu nhiên hóa được tiến hành cho từng nhóm.

Phân nhóm không ảnh hưởng đến khâu phân tích số liệu, bởi vì chúng ta không có mục đích tìm hiểu tiêu chí cho từng nhóm. Phân nhóm chỉ có ảnh hưởng và giá trị trong khâu thiết kế nghiên cứu.

15.3 Ảnh hưởng giả dược (placebo), Hawthorne, và kín đáo

Trong các thí nghiệm liên quan đến con người và bệnh nhân, hai yếu tố khác có thể ảnh hưởng đến kết quả nghiên cứu, đó là giả dược và sự kín đáo (blinding). Để hiểu rõ hai ảnh hưởng này, chúng ta có thể xem xét một ví dụ sau đây. Để biết thuốc alendronate có hiệu quả ngăn ngừa gãy xương hay không, các nhà nghiên cứu chia 100 bệnh nhân thành hai nhóm can thiệp: nhóm 1 có 50 bệnh nhân được cho uống thuốc alendronate thật, và nhóm 2 cũng gồm 50 bệnh nhân được cho thuốc alendronate giả (còn gọi là giả dược hay placebo), nhưng hai loại thuốc hoàn toàn giống nhau, bệnh nhân và bác sĩ không thể phân biệt được thuốc nào là giả và thuốc nào là thật!

Thí nghiệm như vừa mô tả đặt ra hai vấn đề nan giải. Kinh nghiệm từ nhiều nghiên cứu lâm sàng y khoa cho thấy một xu hướng chung là bệnh nhân thường tự cho rằng sức khỏe họ có cải tiến hay tốt hơn, chỉ vì họ được điều trị (cho dù “điều trị” là giả dược)! Yếu tố tâm lí này thường được gọi là “**placebo effect**” hay *hiệu ứng giả dược*. Hiệu ứng giả dược có thể giải thích khoảng 35% kết quả của các nghiên cứu lâm sàng, đặc biệt là đối với các thuốc giảm đau, xuyên, trầm cảm (depression), bệnh đường ruột, và cao huyết áp. Chính vì lí do này, việc đánh giá hiệu quả của một thuật điều trị thường phải có một nhóm đối

chứng (hay placebo) và độ khác biệt giữa hai nhóm can thiệp có thể xác định là hệ quả của thuốc thật hay do giả dược.

Yếu tố thứ hai là **hiệu ứng Hawthorne**. Con người nói chung có khả năng thích ứng rất cao, và khả năng này gây ra không ít khó khăn cho nghiên cứu khoa học. Chẳng hạn như, khi chúng ta cho một nhóm người tiêu thụ nếm vị đắng của cà phê nhiều lần, thì lần đầu người tiêu thụ vì chưa quen với vị đắng nên họ có thể cảm thấy rất đắng và cho điểm cao, nhưng đến lần 2 hay lần 3 thì vì quen với vị đắng nên họ cho điểm thấp xuống. Hay trong nghiên cứu lâm sàng, nếu bệnh nhân biết mình đang được theo dõi, họ sẽ cố gắng làm hài lòng bác sĩ và sự khách quan của bệnh nhân có thể bị ảnh hưởng. Thuật ngữ cho hiện tượng này là “Hawthorne effect”.

Yếu tố thứ ba là sự chủ quan của nhà nghiên cứu. Nếu bác sĩ biết bệnh nhân sử dụng thuốc thật hay giả dược, cách đánh giá của họ có thể ảnh hưởng đến kết quả nghiên cứu. Vì thế, trong các nghiên cứu lâm sàng nghiêm chỉnh, nhà nghiên cứu không được biết bệnh nhân đang được điều trị bằng thuốc hay giả dược, và phương cách này có tên là “**blinding**” (làm mù), tạm dịch là “kín đáo”. Việc giữ kín này phải được duy trì ở bệnh nhân và bác sĩ. Nói cách khác, cả bệnh nhân và bác sĩ đều không biết bệnh nhân thuộc vào nhóm can thiệp hay nhóm giả dược.

Tuy nhiên, không phải bất cứ nghiên cứu lâm sàng nào cũng có thể duy trì sự kín đáo như thế. Chẳng hạn như nghiên cứu về hiệu quả của một thuật giải phẫu, bệnh nhân chắc chắn biết họ được giải phẫu thật hay giả (vì không có cái gọi là “giải phẫu giả”). Ngoài ra, vì lí do y đức, không phải nghiên cứu nào cũng có thể sử dụng giả dược. Nếu chúng ta biết rằng căn bệnh có nguy hiểm đến tính mạng của bệnh nhân và thuốc có hiệu quả, thì không có lí do gì nhà nghiên cứu cho bệnh nhân dùng giả dược. Trong các trường hợp này, nhà nghiên cứu phải suy nghĩ kĩ và phát triển một phương án nghiên cứu sao cho vừa không vi phạm y đức mà vừa đáp ứng các tiêu chuẩn khoa học.

15.4 Vài ví dụ về nguyên tắc của thiết kế nghiên cứu

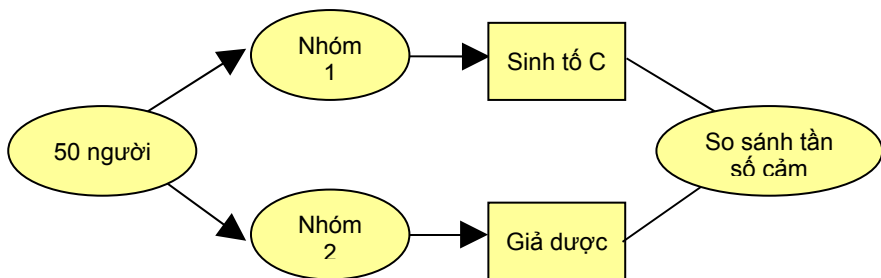
Để quán triệt rõ các nguyên tắc trên, chúng ta thử xem qua một công trình nghiên cứu sau đây về hiệu quả của sinh tố C cho điều trị cảm cúm. Có giả thiết cho rằng sinh tố C có thể ngăn ngừa cảm cúm. Vấn đề đặt ra là chúng ta nên thiết kế nghiên cứu để thử nghiệm giả thiết này như thế nào để đạt tiêu chuẩn khoa học. Giả dụ, chúng ta có 50 người tình nguyện tham gia vào công trình nghiên cứu, chúng ta có thể chọn một trong những phương án sau đây:

Phương án 1. Cho 50 người uống sinh tố C trong vòng 6 tháng, và trong thời gian đó ghi nhận số lần cảm cúm. Kết quả cho thấy sau 6 tháng điều trị, tần số cảm cúm trung bình là 1.4 lần / đối tượng.

Phương án 2. Chia 50 người thành 2 nhóm nam và nữ. Cả hai nhóm được điều trị bằng sinh tố C trong vòng 6 tháng. Kết quả cho thấy sau 6 tháng điều trị, tần số cảm cúm trung bình trong nhóm nam là 1.4 lần / đối tượng, còn nhóm nữ tần số này là 1.9 lần / đối tượng.

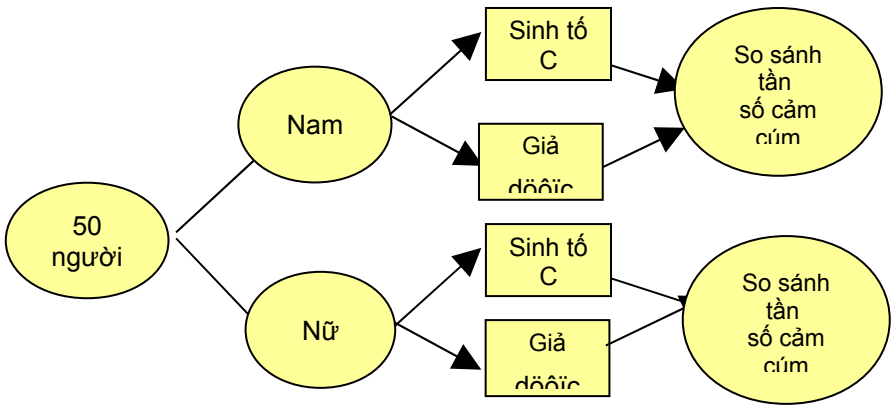
Phương án 3. Chia 50 người thành 2 nhóm một cách ngẫu nhiên. Nhóm 1 gồm 25 người được điều trị bằng sinh tố C trong vòng 6 tháng. Nhóm 2 không được điều trị, nhưng vẫn được theo dõi 6 tháng. Kết quả cho thấy sau 6 tháng điều trị, tần số cảm cúm trung bình trong nhóm 1 là 1.4 lần / đối tượng, còn nhóm đối chứng là 1.9 lần / đối tượng.

Phương án 4. Nhờ một công ti dược sản xuất 50 hộp thuốc sinh tố C, và 50 hộp giả dược sinh tố C. Chia 50 người thành 2 nhóm một cách ngẫu nhiên: nhóm 1 gồm 25 người được điều trị bằng sinh tố C; nhóm 2 nhận giả dược. Cả hai nhóm được theo dõi 6 tháng. Kết quả cho thấy sau 6 tháng điều trị, tần số cảm cúm trung bình trong nhóm 1 là 1.4 lần / đối tượng, còn nhóm đối chứng là 1.4 lần / đối tượng.



Phương án 5. Cũng giống như phương án 4, nhưng chúng ta phân chia (blocking) hai nhóm can thiệp theo giới tính. Giới tính có thể có ảnh hưởng đến nguy cơ cảm cúm (nam thường bắt cần hơn nữ), cho nên chúng ta chia 50 người thành hai nhóm nam và nữ. Mỗi nhóm được ngẫu nhiên hóa thành hai nhóm can thiệp để đảm bảo cân đối nam và nữ cho từng nhóm. Cả hai nhóm được theo dõi 6 tháng. Sau 6 tháng điều trị, tần số cảm cúm trung bình cho từng nhóm can thiệp và giới tính có thể tóm lược như sau:

	Nhóm 1 (sinh tố C)	Nhóm 2 (đối chứng)
Nam	1.4	1.9
Nữ	1.2	1.5



Dựa vào các nguyên lý thiết kế trên, có gì sai lầm trong 4 phương án nghiên cứu trên? Sau đây là vài nhận xét chính:

- Sai lầm của phương án 1 là không có nhóm đối chứng, cho nên kết quả không thể so sánh và cũng rất khó diễn dịch. Tần số trung bình 1.4 lần / đối tượng chẳng có ý nghĩa gì.
- Phương án 2 có nhóm đối chứng, nhưng vì yếu tố can thiệp được phân chia theo giới tính, cho nên độ khác biệt giữa tần số trung bình 1.4 và 1.9 lần / đối tượng không thể nói là do ảnh hưởng của giới tính hay do ảnh hưởng của sinh tố C.
- Phương án 3 có nhóm đối chứng, nhưng sai lầm là không có blinding, vì đối tượng nghiên cứu biết họ nhận hay không nhận thuốc. Nhóm được điều trị có thể sẽ chủ quan không rửa tay (vì nghĩ rằng sinh tố C đã bảo vệ họ), và điều này có thể ảnh hưởng đến kết quả nghiên cứu. Vì thế, kết quả này cũng khó diễn dịch.
- Phương án 4 không có sai lầm nào. Phương án này có nhóm đối chứng và đối tượng nghiên cứu được phân chia một cách ngẫu nhiên và kín đáo, đảm bảo việc so sánh có giá trị khoa học.
- Phương án 5 cũng không có sai lầm nào và tốt hơn phương án 4, vì ảnh hưởng của giới tính được kiểm soát qua cách phân chia ngẫu nhiên.

15.5 Thí nghiệm với một yếu tố (single-factor designs)

Như tên gọi ám chỉ, các thí nghiệm một yếu tố chỉ có một yếu tố can thiệp. Phần lớn các thử nghiệm lâm sàng đối chứng ngẫu nhiên (chia bệnh một cách ngẫu nhiên thành hai nhóm can thiệp) là một dạng của thiết kế này. Tuy

nhiên, có nhiều phương án trong thiết kế này có thể ứng dụng cách phân nhóm (blocking). Ví dụ sau đây sẽ cho chúng ta một ý tưởng về hiệu quả của phân nhóm trong các thí nghiệm một yếu tố.

Ví dụ 2. Một nhóm nghiên cứu nông nghiệp muốn nghiên cứu ảnh hưởng của phân bón đến sự tăng trưởng của lúa. Ba liều lượng urê được sử dụng (thấp, trung bình, và cao – sẽ viết tắt bằng tiếng Anh là low, medium và high). Nhóm nghiên cứu chọn 6 địa điểm (A, B, C, D, E, và F), và mỗi địa điểm có 3 mảnh đất để thí nghiệm (1, 2, 3). Sau đây là vài phương án thí nghiệm mà nhóm nghiên cứu có thể chọn:

Phương án 1 - CRD (completely randomized design): Ở đây, nhóm nghiên cứu có $6 \times 3 = 18$ nơi để thí nghiệm, và 3 yếu tố can thiệp để chia nhóm. Nói cách khác, mỗi yếu tố can thiệp sẽ được áp dụng tại 3 nơi. Với phương án này CRD 3 yếu tố can thiệp được ngẫu nhiên hóa cho tất cả 18 nơi, và kết quả có thể là:

Địa điểm	Mảnh đất 1	Mảnh đất 2	Mảnh đất 3
A	Low	High	Low
B	Medium	Medium	High
C	High	Medium	Low
D	Medium	Low	High
E	Medium	Low	Medium
F	Low	High	High

Trong phương án này, vì cách phân chia ngẫu nhiên, cho nên mỗi địa điểm có khi “nhận” hai liều lượng urê “low” và một loại urê “high” (như địa điểm A). Thành ra, so sánh giữa hai yếu tố can thiệp, như “low” và “high”, phải điều chỉnh độ dao động giữa các địa điểm.

Phương án 2 - RCB (randomized block design): Với phương án này mỗi địa điểm và mỗi mảnh đất sẽ được áp dụng một yếu tố can thiệp; do đó, hoàn toàn cân đối. Nếu xem ba mảnh đất ở mỗi địa điểm thí nghiệm là ba block, thì phương án này đảm bảo tại mỗi địa điểm, mỗi block được phân chia một can thiệp như sau:

Địa điểm	Mảnh đất 1	Mảnh đất 2	Mảnh đất 3
A	Low	High	Medium
B	Medium	Low	High
C	High	Medium	Low
D	Medium	Low	High
E	High	Low	Medium
F	Low	High	Medium

Phương án 3 - IBD (incomplete block design): Với phương án này, nhà nghiên cứu có thể chỉ cần 2 mảnh đất, và mỗi mảnh đất được áp dụng một yếu tố can thiệp như sau:

Địa điểm	Mảnh đất 1	Mảnh đất 2
A	Low	High
B	Medium	Low
C	High	Medium
D	Medium	Low
E	High	Low
F	Low	High

Phương pháp phân tích kết quả từ các nghiên cứu này đã được trình bày trong chương 11.

15.6 Thí nghiệm với hai yếu tố (two-factor designs)

Các thiết kế vừa trình bày trong phần trên nhằm mục đích đánh giá ảnh hưởng của một yếu tố can thiệp. Trong nhiều trường hợp, nhà nghiên cứu muốn đánh giá ảnh hưởng của hai yếu tố can thiệp, và các phương án trên không thể ứng dụng được. Chẳng hạn như khi nhà nghiên cứu muốn phân tích ảnh hưởng của ánh sáng (cao hay thấp) và độ ẩm (khô hay ướt) đến sự tăng trưởng của cây giống trong một nhà kính (greenhouse), thì các phương án thí nghiệm với hai yếu tố cần phải được xem xét cẩn thận.

Phương án 1 - CRD. Chúng ta muốn điều tra ảnh hưởng của nhiệt độ (thấp và cao), vật liệu (A và B), và phương pháp sản xuất (cơ khí và hóa chất) đến độ mạnh của giấy. Các phối hợp yếu tố can thiệp có thể như sau:

Nhóm can thiệp	Nhiệt độ	Vật liệu	Phương pháp
1	Thấp	A	Cơ khí
2	Cao	A	Cơ khí
3	Thấp	B	Cơ khí
4	Thấp	A	Hóa chất

Qua thiết kế này, chúng ta có thể phân tích ảnh hưởng của nhiệt độ bằng cách so sánh độ mạnh giữa nhóm 1 và 2. Ảnh hưởng của vật liệu có thể so sánh giữa nhóm 1 và 3. Ảnh hưởng của phương pháp có thể so sánh giữa nhóm 1 và 4.

Những so sánh trên hợp lí chỉ với điều kiện là ảnh hưởng của các yếu tố can thiệp là **cộng hưởng** (additive effect). Nói cách khác, các so sánh trên chỉ hợp lí nếu ảnh hưởng của một yếu tố không phụ thuộc vào các yếu tố khác, chẳng hạn như ảnh hưởng của nhiệt độ không tùy thuộc vào ảnh hưởng của vật liệu hay phương pháp. Nếu giả định này không đúng thì kết quả so sánh có thể thiếu khách quan và sai.

Phương án 2 – Giai thừa (factorial design). Một phương án khác cho nhiều thông tin hơn phương án trên và cho phép chúng ta phân tích ảnh hưởng tương tác của các yếu tố can thiệp là **factorial design**. Trong trường hợp trên, chúng ta có 3 yếu tố can thiệp, và mỗi yếu tố có 2 bậc, cho nên tổng số là $2^3 = 8$ nhóm như sau:

Nhóm can thiệp	Nhiệt độ	Vật liệu	Phương pháp
1	Thấp	A	Cơ khí
2	Cao	A	Cơ khí
3	Thấp	B	Cơ khí
4	Thấp	B	Cơ khí
5	Thấp	A	Hóa chất
6	Cao	A	Hóa chất
7	Thấp	B	Hóa chất
8	Thấp	B	Hóa chất

Với phương án “cân đối” này, chúng ta có thể ước tính ảnh hưởng của mỗi yếu tố can thiệp dễ dàng:

- Ảnh hưởng của nhiệt độ: so sánh nhóm 1, 3, 5, 7 và 2, 4, 6, 8;
- Ảnh hưởng của vật liệu: so sánh nhóm 1, 2, 5, 6 và 3, 4, 7, 8;
- Ảnh hưởng của phương pháp: so sánh nhóm 1, 2, 3, 4 và 5, 6, 7, 8.

Ngoài ra, **ảnh hưởng tương tác (interaction effects)** cũng có thể ước tính bằng cách so sánh tổng hợp giữa các nhóm. Chẳng hạn như để biết ảnh hưởng của nhiệt độ có tùy thuộc vào phương pháp sản xuất hay không, chúng ta có thể so sánh giữa nhóm 1+3 và 2+4, và 5+7 với 6+8.

Phương án 3 – Phân mẫu (Split-plot design). Để điều tra ảnh hưởng của 3 loại giống đậu nành (A, B và C) và hai loại phân bón (P1 và P2), các nhà nghiên cứu so sánh sản lượng đậu nành được trồng trọt dưới 6 điều kiện can thiệp trên. Nếu mỗi điều kiện can thiệp được lặp lại 2 lần, thì nghiên cứu cần $2 \times 6 = 12$ mảnh đất (plot) cho nghiên cứu.

Một cách thiết kế cho nghiên cứu trên là phương án giai thừa như đề cập phần trên, nhưng phương án này có thể khó khăn trong thực tế. Một phương

án khác dễ dàng hơn và thực tế hơn là phương án phân mẫu. Phương án này cần đến hai lần sắp xếp ngẫu nhiên. Trước hết, hai loại phân bón được phân chia một cách ngẫu nhiên cho 4 nhóm như sau:

1	2	3	4
P1	P2	P2	P1

Bước hai, ba loại giống sẽ được phân ngẫu nhiên cho từng nhóm, và kết quả có thể giống như sau:

1	2	3	4
P1	P2	P2	P1
B	C	A	C
A	B	C	A
C	A	B	B

Phương án 4 – Hình vuông Latin (Latin square). Công ti dầu muốn so sánh hiệu suất (đo bằng cây số - km - trên mỗi lít) của 4 loại dầu (A, B, C và D). Công ti có được 4 tài xế và 4 loại xe. Vì độ dao động hay khác biệt giữa người lái xe và loại xe, hai yếu tố này phải được kiểm soát trong khi thiết kế nghiên cứu. Phương án tốt nhất cho nghiên cứu này là phương án hình vuông Latin. Theo phương án này, 4 loại dầu được phân chia một cách ngẫu nhiên cho từng tài xế và loại xe như sau:

Tài xế	Loại xe			
	Ford	Toyota	Honda	Nissan
1	D	B	C	A
2	B	C	A	D
3	C	A	D	B
4	A	D	B	C

Như vậy, tài xế 1 sẽ điều khiển xe Ford với dầu loại D, sau đó là Toyota với loại dầu B, Honda với dầu C và Nissan với dầu A, v.v... Với phương án này, công ti có thể phân tích ảnh hưởng cộng hưởng của từng loại dầu, hay phân tích ảnh hưởng tương tác giữa loại dầu và loại xe, hay giữa loại dầu và tài xế. Cách phân tích số liệu từ phương án hình vuông Latin đã được mô tả chi tiết ở Chương 11 (11.7).

15.7 Phương pháp ngẫu nhiên hóa

Trong tất cả các phương án trên, một khía cạnh then chốt là sắp xếp các đối tượng nghiên cứu vào các yếu tố can thiệp một cách ngẫu nhiên (tôi sẽ gọi tắt là *ngẫu nhiên hóa* – randomize). Giả dụ, chúng ta có 8 đối tượng (có thể là bệnh nhân) cần phân chia cho ba nhóm can thiệp T1, T2 và T3. Nếu theo yêu cầu, nhóm T1 và T3 mỗi nhóm cần phải có 3 đối tượng và nhóm T2 cần 2 đối tượng:

T1	T2	T3
n=3	n=2	n=3

Vấn đề đặt ra là làm cách nào để ngẫu nhiên hóa? Chúng ta có thể làm một số bước như sau:

- Trước hết, chúng ta lập danh sách 8 đối tượng:

T1 T1 T1 T2 T2 T3 T3 T3

- Dùng hàm `sample` để chọn ngẫu nhiên (`sample(1:8)` có chức năng tạo ra một dãy số ngẫu nhiên từ 1 đến 8):

```
> sample(1:8)
[1] 7 2 5 4 1 8 6 3
```

- Nhập hai dãy số với nhau, chúng ta có:

T1 T1 T1 T2 T2 T3 T3 T3
7 2 5 4 1 8 6 3

Nói cách khác, đối tượng số 7, 2 và 5 sẽ nhận can thiệp T1, đối tượng 1 nhận can thiệp T2, và đối tượng 8, 6 và 3 nhận T3.

Bất cứ công trình nghiên cứu khoa học nào cũng được tiến hành theo một qui trình gần như bất biến: đặt giả thiết, thiết kế nghiên cứu, thu thập dữ liệu, phân tích dữ liệu, và báo cáo kết quả. Do đó, phân tích số liệu là khâu gần cuối cùng của một công trình nghiên cứu (trước khi diễn dịch kết quả phân tích và viết báo cáo khoa học). Các khâu trong qui trình nghiên cứu trên có liên hệ khăng khít với nhau. Nếu một khâu trong qui trình đó có vấn đề thì hệ quả là khâu tiếp theo cũng có vấn đề. Mượn cách nói của người xưa “Vạn sự khởi đầu nan”, cũng có thể nói rằng khi thiết kế nghiên cứu sai thì kết quả phân tích cũng không có ý nghĩa gì. Tất cả các phương pháp phân tích số liệu chỉ cho ra kết quả tốt khi công trình nghiên cứu được thiết kế đúng và thích hợp. Do đó, xem xét

cẩn thận các phương án nghiên cứu và đối chiếu với tình hình thực tế là một nhu cầu rất quan trọng cho nghiên cứu khoa học.