R

VIỆN NĂNG LƯỢNG NGUYÊN TỬ VIỆT NAM
VIỆN CÔNG NGHỆ XẠ HIẾM

KHOA HỌC VÀ CÔNG NGHỆ VỀ NHIÊN VẬT LIỆU HẠT NHÂN

Huỳnh Văn Trung (Chủ biên), Thái Bá Cầu
Đỗ ngọc Liên, Lê Bá Thuận, Đỗ qúy Sơn, Cao Hùng Thái,
Nguyễn Bá Tiến, Cao Đình Thanh, Nguyễn Lanh

QUY HOẠCH HÓA THỰC NGHIỆM

NGUYỄN LANH

HÀ NỘI -2005

ĐỀ CƯƠNG BÀI GIẢNG CHUYÊN ĐỀ

PHÂN TÍCH THỐNG KÊ ỨNG DỤNG VÀ QUY HOẠCH THỰC NGHIỆM

Nguyễn Lanh, Viện CNXH

Mục đích môn học

Sau khi học xong, các học viên có thể:

- Biết cách phân tích các số liệu thực nghiệm một cách khoa học để đưa ra những đánh giá đúng đắn và khách quan;
- Có khả năng xây dựng nên các mô hình hồi quy từ các số liệu thực nghiệm thu thập được trong quá khứ để khảo sát và dự đoán trước các tính chất của đối tượng cần nghiên cứu;
- Biết cách tổ chức các thí nghiệm để nghiên cứu các đối tượng một cách hợp lý và khoa học, cho phép thu được nhiều thông tin nhất với số lượng thí nghiệm tối thiểu.

MỤC LỤC

1. MỘT SỐ KHÁI NIỆM CƠ BẢN VỀ THỐNG KÊ

1.1 Khái niệm chung về phân tích thống kê và mô hình thống kê

1.1.1 Một số khái niệm cơ bản về thống kê

Thống kê là một ngành khoa học, cụ thể là một ngành của toán học, nhằm thu thập số các liệu thực tế, phân tích các số liệu đó, diễn giải ý nghĩa và rút ra các kết luận từ quá trình phân tích các số liệu thu được. Như vậy, phương pháp thống kê sử dụng những kỹ thuật liên quan đến thu thập số liệu, đo đạc (cả định tính và định lượng), và phân tích toán học.

Đối tượng chủ yếu của phân tích thống kê là các biến ngẫu nhiên, cụ thể ở đây là các thông số của các quá trình, đối tượng được nghiên cứu. Thông qua việc phân tích này mà ta có thể xác định được những mối quan hệ nhất định giữa các thông số đầu vào và ra này. Có thể nêu ra một số dạng và các đặc trưng cơ bản của biến ngẫu nhiên như sau.

Các *đại lượng ngẫu nhiên* (biến ngẫu nhiên) có thể được chia ra thành hai loại là loại gián đoạn và loại liên tục. Quan hệ giữa giá trị của một đại lượng ngẫu nhiên với xác suất xảy ra nó được thể hiện qua *hàm phân bố f(x)* của biến ngẫu nhiên đó, hàm này còn được gọi là *hàm phân bố vi phân* để phân biệt với *hàm phân bố tích phân* được định nghĩa là hàm biểu diễn xác suất để đại lượng ngẫu nhiên có giá trị không vượt quá x_i.

Các biến ngẫu nhiên còn được chia ra thành hai loại là *biến độc lập* và *biến phụ thuộc* tùy theo điều kiện xác suất xuất hiện của biến này có phụ thuộc vào sự xuất hiện của biến khác hay không.

Khi sắp xếp các biến ngẫu nhiên theo thứ tự giá trị tăng dần hay giảm dần thì giá trị trung bình của dãy số đó được gọi là *Median*.

Giá trị của xác suất lớn nhất của đại lượng ngẫu nhiên được gọi là *Mod*.

Kỳ vọng toán học (hay còn gọi là gái trị trung bình) của một đại lượng ngẫu nhiên được tính theo công thức:

$$M(x) = \int_{-\infty}^{+\infty} x.f(x).dx \qquad (1\text{-}1) \quad \text{(đối với biến liên tục), và}$$

$$= \sum_{i=1}^{n} x_i.f(x_i) \qquad (1\text{-}2) \quad \text{(đối với biến gián đoạn).}$$

Kỳ vọng toán học có tính chất chung là:

1. Kỳ vọng của tổng các biến ngẫu nhiên thì đúng bằng tổng các kỳ vọng của chúng:
2. Kỳ vọng của tích các biến ngẫu nhiên độc lập bằng tích các kỳ vọng của chúng

Phương sai: là đại lượng đặc trưng cho mức độ phân tán của các giá trị biến ngẫu nhiên so với kỳ vọng của chúng. Theo lý thuyết xác suất, phương sai được định nghĩa theo công thức:

$$D(x) = \int_{-\infty}^{\infty} [x - M(x)]^2.f(x).dx \qquad (1\text{-}3) \quad \text{(đối với biến liên tục),}$$

và

$$= M[x - M(x)]^2 \qquad (1\text{-}4) \quad \text{(đối với biến gián đoạn)}$$

Phương sai có các tính chất:

1. Phương sai của hằng số bằng 0
2. Đưa hằng số ra ngoài dấu phương sai: $D(c.x) = c^2.D(x)$
3. Phương sai của tổng các biến ngẫu nhiên độc lập bằng tổng các phương sai. Thí dụ với 2 biến x và y, theo định nghĩa:

$$D(x+y) = M[(x+y) - M(x+y)]^2$$

$$= M[x-M(x)]^2 + M[y-M(y)]^2 + 2M\{[x-M(x)][y-M(y)]\}$$

Từ đây: $D(x+y) = D(x) + D(y) + 2cov(xy)$

Trong đó: $cov(xy) = M\{[x-M(x)][y-M(y)]\}$ (1-5)

Đại lượng $cov(xy)$ được gọi là giá trị tương quan, đặc trưng cho mối quan hệ giữa các biến ngẫu nhiên x và y. Nếu x và y độc lập với nhau thì $cov(xy) = 0$. Trong thực tế, để thuận tiện người ta thường dùng khái niệm *hệ số tương quan* được định nghĩa như sau:

$$r(x,y) = \frac{\operatorname{cov}(x,y)}{\sqrt{D(x).D(y)}} \qquad (1\text{-}6)$$

Hệ số tương quan trên có giá trị thay đổi từ -1 đến $+1$, nó đặc trưng cho mức độ "quan hệ chặt chẽ" giữa các biến. Nếu hệ số tương quan $r(x,y) = 0$ thì hai biến x, y đó được coi là không có quan hệ gì với nhau. Ngược lại, nếu $r(x,y) \to 1$ thì x và y được coi là có mối quan hệ rất "rõ ràng".

Trong thực tế triển khai các thí nghiệm hóa và công nghệ hóa học thường chỉ có thể tiến hành một số lượng hữu hạn các thí nghiệm, trong đó để lấy được số liệu ở một điểm thường phải tiến hành làm vài ba thí nghiệm hay phép đo song song. Các phép đo này thường phải có tính độc lập với nhau và được tiến hành trong những điều kiện như nhau. Các đánh giá của ta sau đó cũng phải dựa trên cơ sở các số liệu thực nghiệm thu được. Do đây cũng chỉ là các đánh giá gần đúng nên được gọi là các "đánh giá chọn lọc".

Kỳ vọng toán học của một bộ các phép đo song song được tính theo giá trị trung bình số học $\bar{x}$. Toàn bộ các đại lượng ngẫu nhiên được gọi là *tập gốc*.

Khi thiết lập các thí nghiệm để đo các đại lượng ngẫu nhiên thường có những giới hạn nhất định do không thể làm một số thực nghiệm vô cùng lớn. Do vậy, các đánh giá về thống kê sẽ có những khác biệt so với cách định nghĩa theo quan điểm lý thuyết xác suất. Cụ thể: các đánh giá về các tham số thống kê khi tiến hành n thí nghiệm song song gồm:

a/ Giá trị trung bình: $$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (1\text{-}7)$$

b/ Phương sai, còn được gọi là *phương sai chọn* và được tính theo:

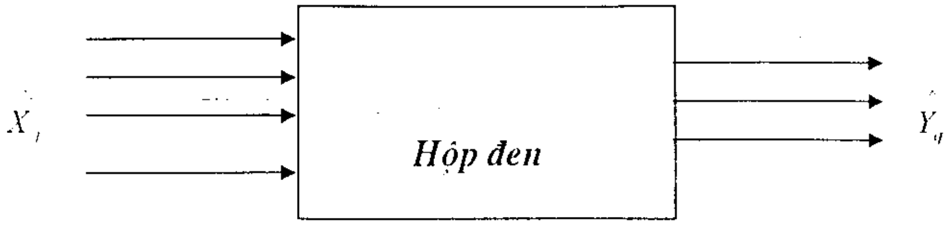
$$s^2(x) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{f} \qquad (1\text{-}8)$$

trong đó, f được gọi là bậc tự do và bằng n-1.

1.1.2 Khái niệm chung về mô hình thống kê

Mô hình thống kê, trong nhiều trường hợp còn được gọi là mô hình "hộp đen", được hình thành trên cơ sở vận dụng toán học thống kê vào thực nghiệm để tìm ra các mối quan hệ giữa các thông số đầu vào và đầu ra mà không cho biết bản chất và cấu trúc của hệ. Do đó, việc sử dụng mô hình thống kê không cho phép hiểu rõ được quy luật bảo toàn cũng như quy luật động học trong vận động của hệ mà chỉ hiểu được tương tác giữa các yếu tố cần quan tâm với một mức độ xác suất nào đó của hệ cụ thể trong phạm vi nghiên cứu. Như vậy, mô hình thống kê chỉ cho phép xác định các đại lượng ra theo các đại lượng vào với một phân bố xác suất nhất định.

Mô tả về nguyên tắc của một mô hình hộp đen cùng các quan hệ giữa các thông số vào – ra được đưa ra trên Hình 1.



Hình 1: Sơ đồ một mô hình hộp đen, X_i : các biến vào. Y_q : các biến ra

Với một sơ đồ nguyên lý về mô hình hộp đen như trên, người ta có thể sử dụng một số cách để tìm ra các quan hệ phụ thuộc của các thông số đầu ra vào các thông số đầu vào. Có hai phương pháp thường được dùng nhiều nhất là:

1. Phương pháp xây dựng các phương trình hồi quy thực nghiệm dựa trên cơ sở phân tích thụ động các số liệu có được trong một thời gian làm việc đủ dài của hệ. Phương pháp này còn thường được gọi là phân tích hồi quy và được dùng rất phổ biến từ khoảng giữa thế kỷ 20 và cho đến ngay cả hiện nay. Ưu điểm chính của phương pháp này là:

   a/ Có thể lấy số liệu trực tiếp từ quá trình sản xuất thực tế, không cần phải bố trí các thông số đầu vào theo một quy hoạch bắt buộc được định trước;

   b/ Đơn giản trong cách tính toán (có thể tính toán được các hệ số của mô hình bằng tính tay),

   c/ Các kết quả thu được (thường ở dạng các đa thức bậc một hoặc bậc hai) thuận tiện cho các công việc phân tích và tính toán tiếp theo. Tuy nhiên, để có thể sử dụng được các phương trình này cần có các hiểu biết nhất định về bản chất lý-hóa của quá trình, không nên chỉ nhìn vào các con số thu được và "*ngoại suy ra một cách mù quáng*". Cách suy nghĩ này trong nhiều trường hợp sẽ dẫn đến kết luận sai lầm hoàn toàn!

Nhược điểm chính của phương pháp là không thể khảo sát tính chất của hệ ở những vùng nằm ngoài miền làm việc, những vùng mà ta không có khả năng thu thập được số liệu.

2. Phương pháp mô hình hóa thứ hai dựa trên cơ sở chủ động bố trí các thí nghiệm để đo các thông số đầu ra theo các thông số đầu vào đã định trước. Do chủ động bố trí các thông số đầu vào nên có thể chủ động mở rộng được miền cần khảo sát ra lân cận vùng làm việc thông thường của hệ. Từ đó có thể giúp tìm ra miền làm việc tối ưu của hệ. Do vậy, đây là phương pháp rất quan trọng được dùng phổ biến trong nghiên cứu để xây dựng các mô hình theo kiểu hộp đen nhằm mô tả những hệ có cấu trúc phức tạp mà không thể mô tả được bằng những phương trình toán học thông thường.

Khó khăn lớn nhất khi sử dụng phương pháp này là người sử dụng cần phải có những hiểu biết tương đối kỹ về đối tượng cần được mô tả để có thể xây dựng nên mô hình theo các mối quan hệ thể hiện đúng bản chất của hệ, nếu không sẽ đưa ra những mô hình không phù hợp và các thông tin/ kết luận rút ra từ đó sẽ không chính xác. Hoặc trong một số trường hợp khác, tuy mô hình tương hợp (về mặt thống kê) nhưng do sự kém hiểu biết về bản chất của quá trình thực của người xây dựng mô hình nên từ mô hình (thống kê) thu được lại rút ra những kết luận sai lệch. Vì vậy, hiểu biết về đối tượng được nghiên cứu vẫn là yêu cầu đầu tiên đối với người làm mô hình để từ đó có thể xây dựng nên những mô hình hợp lý cũng như biết cách phân tích và rút ra những thông tin chính xác về đối tượng được nghiên cứu.

1.2 Các phương pháp kiểm định thống kê ứng dụng

1.2.1 Đánh giá khoảng tin cậy

Một trong những ứng dụng quan trọng của tính toán thống kê là xác định khả năng làm việc ổn định của một quá trình công nghệ nào đó trong các điều kiện làm việc như nhau, hay tính lặp lại của những thí nghiệm song song. Mặc dù các điều kiện tiến hành các quá trình này được giữ như nhau nhưng do thực tế còn có rất nhiều những yếu tố khác cũng ảnh hưởng đến các kết quả đo được nên bao giờ cũng có các sai khác giữa các kết quả thu được.

Để tính toán khoảng tin cậy, trong thực tế thường tiến hành một số thí nghiệm song song để thu lấy một bộ kết quả ngẫu nhiên từ tập gốc. Khi này, bài toán của nhà thực nghiệm là: từ bộ chọn thu được đó hãy xác định kỳ vọng của biến ngẫu nhiên này. Tất nhiên sẽ có sai số giữa giá trị kỳ vọng thực của biến ngẫu nhiên với giá trị trung bình tính được từ các số liệu thực nghiệm.

Theo cách đặt vấn đề như trên, ước lượng của kỳ vọng sẽ là $\bar{x}$ với độ lệch s^2 được xác định theo công thức (1-8).

Khi thay đổi số các thí nghiệm song song, giá trị của $\bar{x}$ và s^2 cũng sẽ thay đổi theo do chúng chỉ là các đánh giá của các giá trị kỳ vọng và phương sai thực, và khi số các thực nghiệm n tăng càng lớn thì độ chính xác của chúng sẽ càng lớn hơn.

Nếu lấy một số bộ chọn khác nhau từ một tập gốc, mỗi bộ chọn sẽ có một cặp giá trị $\bar{x}$ và s^2 riêng và khác nhau. Do các bộ chọn này là ngẫu nhiên nên các sai sai số này cũng là ngẫu nhiên. Từ đây, vấn đề được quan tâm là: Có thể đánh giá sự sai khác giữa $M(x) - \bar{x}$ chính xác đến mức nào. Nếu gọi sai số là δ thì cần phải xác định giá trị của δ sao cho thỏa mãn

$$\left|M(x) - \bar{x}\right| \le \delta \qquad (1\text{-}9)$$

Do $\bar{x}$ là đại lượng ngẫu nhiên có giá trị phụ thuộc n nên δ cũng phụ thuộc n. Nếu gọi γ là độ tin cậy, có thể viết lại thành:

$$\phi(\left|M(x) - \bar{x}\right| \le \delta) = 1 - \gamma \qquad (1\text{-}10)$$

Kết quả biểu diễn cuối cùng của bài toán xác định khoảng tin cậy còn thường được viết ở dạng:

$$(\bar{x} - \delta) < M(x) < (\bar{x} + \delta)$$

hay $$M(x) \in \bar{x} \pm \delta \qquad (1\text{-}11)$$

Trong đó, $(\bar{x} - \delta) - (\bar{x} + \delta)$: khoảng tin cậy.

Để có thể đánh giá δ thường giả thiết các biến ngẫu nhiên này tuân theo hàm phân bố chính quy. Tuy nhiên khi đó cần phải biết được phương sai gốc δ và số điểm thí nghiệm n thường cần phải đủ lớn (từ vài chục trở lên).Trong thực tế thường xảy ra trường hợp ngược lại: không biết được phương sai gốc và số các điểm thí nghiệm song song cũng thường là ít: khi này phải dùng đến hàm phân bố Student. Đây là hàm phân bố không phụ thuộc vào phương sai gốc với biến ngẫu nhiên là:

$$t = \frac{\left[M(x) - \bar{x}\right]\sqrt{n}}{s(x)} \qquad (1\text{-}12)$$

Như vậy, hàm phân bố của biến ngẫu nhiên này phụ thuộc số thí nghiệm song song hay chính là thể tích của bộ chọn. Giá trị của đại lượng t được đưa ra trong các bảng tính sẵn và được xác định theo 2 tham số là mức ý nghĩa và số bậc tự do. Sau khi dùng bảng tính tra được t, công thức tính toán khoảng tin cậy cho biến ngẫu nhiên x sẽ là:

$$M(x) = \bar{x} \pm \frac{t.s(x)}{\sqrt{n}} \qquad (1\text{-}13)$$

1.2.2 Kiểm tra một số giả thiết thống kê

Nguyên tắc chung của việc kiểm tra một giả thiết thống kê thường là so sánh một đại lượng θ nào đó với một giá trị tới hạn $\theta_{t.h.}$ tương ứng và giả thiết rằng 2 đại lượng này là như nhau. Các kết luận có thể được chia thành 2 loại:

1. Bác bỏ giả thiết trong khi nó đúng: mắc sai lầm loại 1 với xác suất α
2. Công nhận giả thiết là đúng trong khi nó sai: sai lầm loại 2 với xác suất phạm phải là β.

Thường người ta cố tìm cách làm giảm sai lầm loại 1, do đó xác suất α thường lấy rất nhỏ và khi đó β sẽ là rất lớn. Trong các thí nghiệm công nghệ thường lấy $\alpha = 0{,}05$ (mức ý nghĩa). Các giả thiết thống kê thường được kiểm tra theo các bước sau:

1. Trên cơ sở các số liệu thực nghiệm, tính ra đại lượng thực nghiệm cần kiểm tra θ_{tn} tương ứng với mức ý nghĩa cho trước
2. So sánh đại lượng θ_{tn} trên với giá trị tới hạn $\theta_{t.h}$ tương ứng với mức ý nghĩa cho trước
3. Từ kết quả so sánh mà rút ra kết luận công nhận hay bác bỏ giả thiết.

Một số dạng so sánh thường gặp trong thực tế là:

<u>So sánh hai giá trị trung bình số học và hai phương sai chọn:</u>

Các số liệu được đưa ra so sánh thường được lấy từ hai dãy thí nghiệm song song. Nhiệm vụ đặt ra là xác định xem hai giá trị trung bình $\bar{x}_1$ và $\bar{x}_2$ có phải là các đánh giá của kỳ vọng toán học của biến ngẫu nhiên x không, và các giá trị s_1^2 và s_2^2 có phải là phương sai chọn và là các đánh giá của phương sai gốc không.

<u>So sánh hai phương sai:</u>

Dùng tiêu chuẩn Fisher (F). Khi có hai phương sai s_1^2 và s_2^2, giả thiết $s_1^2 > s_2^2$, khi đó giá trị thực nghiệm $F_{tn} = s_1^2 / s_2^2$

Từ bảng 2 (phụ lục ...) tra giá trị tiêu chuẩn $F_p(f_1, f_2)$, trong đó p là mức ý nghĩa, f_1 và f_2 là số bậc tự do của tử và mẫu số tương ứng.

Nếu $F_{tn} < F_p(f_1, f_2)$: công nhận giả thiết: s_1^2 và s_2^2 là các đánh giá của cùng một phương sai gốc.

1.3 Một số thí dụ ứng dụng về kiểm tra giả thiết thống kê

a. Thí dụ 1:

Giả sử có các kết quả đo lấy từ hai mẫu thí nghiệm khác nhau với các giá trị bằng số là như sau:

Mẫu 1	Mẫu 2
$\bar{x}_1 = 20{,}5$	$\bar{x}_2 = 21.3$
$s_1^2 = 0{,}09$	$s_2^2 = 0.16$
$n_1 = 5$	$n_2 = 6$

Cần làm rõ xem: liệu 2 nhóm thí nghiệm này có thể được coi là cùng được lấy ra từ một tập gốc hay không?

<u>*Giải*</u>:

Kiểm tra xem hai bộ số liệu này có phải từ cùng một tập gốc không?

Trước hết xác định giá trị F_{tn}: $F_{tn} = \dfrac{0{,}16}{0{,}09} = 1{,}8$

Từ bảng 2, tra giá trị $F_{0.05}(5,4) = 6{,}3$ $\Rightarrow$ công nhận giả thiết: hai bộ số liệu này được rút ra từ cùng một tập gốc.

Kiểm tra xem hai giá trị trung bình trên có thể coi là bằng nhau (về mặt thống kê) hay không, trước hết xác định phương sai chung:

$$\bar{s}^2 = \frac{0{,}16 \times 5 + 0{,}09 \times 4}{9} = 0{,}13 \quad \Rightarrow \quad \bar{s} = \sqrt{0{,}13} = 0{,}36$$

Tính giá trị thực nghiệm:

$$t_{tn} = \frac{21{,}3 - 20{,}5}{0{,}36}\sqrt{\frac{5 \times 6}{5 + 6}} = 3{,}68$$

Tra bảng giá trị $t_{0,05}(9) = 2{,}26 < 3{,}68$

$\Rightarrow$ Bác bỏ giả thiết về sự bằng nhau của hai giá trị trung bình này

2. QUY HOẠCH THỰC NGHIỆM (TRỰC GIAO)

Trước đây, để nghiên cứu ảnh hưởng của các yếu tố công nghệ khác nhau đến hiệu quả của một quá trình, người ta thường tiến hành các nghiên cứu thực nghiệm theo kiểu cố định các yếu tố và chỉ cho một yếu tố thay đổi để làm thành một dãy thí nghiệm. Sau đó, làm lại lần lượt các dãy thí nghiệm tương tự với các yếu tố khác. Với cách làm này, số thí nghiệm phải làm thường là rất lớn, và trong trường hợp có nhiều yếu tố cùng có tác động đồng thời (tương tác kép) thì cách làm trên không thể mang lại kết quả chính xác.

Quy hoạch thực nghiệm là phương pháp tổ chức các thí nghiệm sao cho chỉ mất một số ít thí nghiệm nhất nhưng có thể thu nhận được lượng thông tin nhiều nhất. Đây là phương pháp thực nghiệm cho phép nghiên cứu ảnh hưởng đồng thời của nhiều yếu tố công nghệ tới một chỉ tiêu nào đó của quá trình mà không cần phải cố định và thay đổi lần lượt từng biến số. Ngoài ra, bằng phương pháp quy hoạch thực nghiệm còn có thể phát hiện ra những hiệu ứng tương tác kép mà bằng các cách làm thực nghiệm cổ điển hầu như không thể phát hiện ra được, hoặc nếu có nhận biết thấy thì cũng không thể định lượng được.

Kết quả cuối cùng của quy hoạch thực nghiệm là xây dựng nên một mô hình toán học ở dạng phương trình hồi quy biểu thị mối quan hệ giữa các thông số đầu ra với các thông số đầu vào. Trên cơ sở mô hình toán học này có thể tiến hành tính toán, xác định trước được các đặc tính của đầu ra khi các yếu tố đầu vào thay đổi, từ đó định hướng cho việc tìm ra miền hoạt động tối ưu của quá trình công nghệ.

2.1 Các nguyên lý cơ bản của quy hoạch thực nghiệm

Để xác lập mô hình thống kê cho một quá trình hóa học, thường phải thực hiện các bước: xác định hệ; xác định cấu trúc hệ; xác định hàm toán mô tả hệ; xác định các thông số của mô hình mô tả hệ và kiểm tra tính tương hợp của các mô tả đó.

2.1.1 Xác định hệ

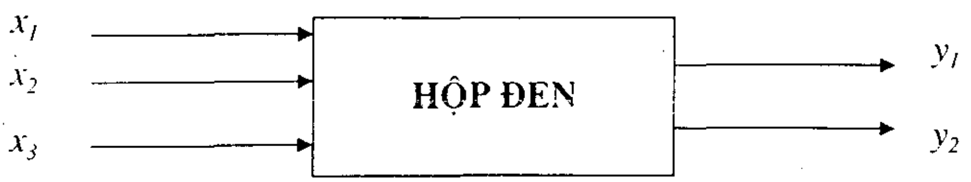
Nếu gọi $F_{ĐK}$ là bậc tự do điều khiển và F_{HH} là bậc tự do hình học của hệ thì số các yếu tố độc lập tối đa ảnh hưởng lên hệ được xác định theo công thức:

$$F = F_{ĐK} + F_{HH} \tag{2-1}$$

Tuỳ theo yêu cầu nghiên cứu mà có thể chọn k yếu tố ($k \le F$) ảnh hưởng lên một hàm mục tiêu y nào đó hoặc nhiều hàm mục tiêu. Hàm mục tiêu có thể là các chỉ tiêu công nghệ như hiệu suất quá trình, năng suất quá trình.... hoặc có thể là các chỉ tiêu kinh tế kỹ thuật như giá thành sản phẩm, lợi nhuận thu được... . Điều quan trọng nhất là hàm mục tiêu cần thể hiện được đồng thời mối quan hệ của nó với các biến độc lập và giúp cho người nghiên cứu đánh giá được một cách đúng đắn các chỉ tiêu mà mình quan tâm.

2.1.2 Xác định cấu trúc hệ

Phương pháp mô tả thống kê quan niệm hệ là một hộp đen, trong đó cấu trúc và tính chất bên trong là không biết rõ. Nói cách khác, mô hình thống kê không cho biết bản chất bảo toàn và bản chất động học của hệ, mà chỉ mô tả mối quan hệ giữa các thông số đầu vào và các thông số đầu ra của hệ trên cơ sở các phương pháp của toán học thống kê.



Minh họa mô hình hộp đen

Như vậy, mối quan *đầu vào - đầu ra* ở đây thuần túy là quan sát thực nghiệm chứ không được dựa trên một cơ sở lý thuyết nào. Trên cơ sở các quan sát thực nghiệm này mà người làm thực nghiệm mới suy nghĩ, ước đoán để giả thiết ra các dạng quan hệ có thể có giữa các biến đầu vào với các thông số đầu ra (quan hệ $x_i - y_j$), ví dụ đơn giản là có thể giả thiết chúng có quan hệ tuyến tính với nhau.

Chính vì lý do trên nên trên thực tế việc xây dựng thành công một mô hình hồi quy thực nghiệm phụ thuộc chủ yếu vào hiểu biết của người nghiên cứu thực nghiệm về quá trình cần nghiên cứu. Các hiểu biết này sẽ giúp định hướng trong việc xác định dạng của quan hệ $x_i - y_j$ cũng như các khả năng biểu diễn gần đúng các đường quan hệ này bằng những dạng phương trình đơn giản hơn (các đa thức bậc 1 và 2).

2.1.3 Xác định các hàm toán mô tả hệ

Trong trường hợp cấu trúc hộp đen, các hàm toán mô tả hệ là các hàm nhiều biến $y = \varphi(x_1, x_2, ..., x_k)$ được phân tích thành dãy Taylo tức là hàm hồi quy lý thuyết:

$$y_q = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{j,u=1}^{k} \beta_{ju} x_j x_u + \sum_{j=1}^{k} \beta_{jj} x_j^2 + ... \quad \text{với } \forall q=(1,n) \qquad (2\text{-}2)$$

Để xác định được các hệ số hồi quy lý thuyết β_0, β_j, β_{ju}, β_{jj},... cần phải có vô hạn số thực nghiệm mà trong thực tế số thực nghiệm chỉ có thể là hữu hạn, nên chỉ xuất hiện các hệ số hồi quy thực nghiệm (còn gọi là các thông số của mô hình thống kê) b_0, b_j, b_{ju}, b_{jj},... và vì vậy hàm toán mô tả hệ là hàm hồi quy thực nghiệm [1]:

$$Y_q = b_0 + \sum_{j=1}^{k} b_j x_j + \sum_{j,u=1}^{k} b_{ju} x_j x_u + \sum_{j=1}^{k} b_{jj} x_j^2 + ... \qquad (2\text{-}3)$$

Phương trình (2-3) là dạng tổng quát của mô hình thống kê mô tả đối tượng nghiên cứu.

2.1.4 *Xác định các thông số và kiểm tra tính tương hợp của mô hình thống kê*

Các thông số của mô hình thống kê (các hệ số hồi quy) từ N thực nghiệm (số thực nghiệm N phải lớn hơn số thông số) được xác định theo công thức sau [1]:

$$b_j = \frac{\sum_{i=1}^{N} x_{ji} y_i}{\sum_{i=1}^{N} x_{ji}^2} \quad \text{với } \forall j=(0,k) \qquad (2\text{-}4)$$

$$b_{ju} = \frac{\sum_{i=1}^{N} x_{ji}^2 x_{ui} y_i}{\sum_{i=1}^{N} x_{ji}^2 x_{ui}^2} \quad \text{với } \forall j,u=(1.k)\ (j \neq u) \qquad (2\text{-}5)$$

Trong trường hợp kế hoạch bậc một hai mức tối ưu, hệ số bất kỳ của phương trình hồi quy b_j được xác định bằng tích không vô hướng của cột y với cột tương ứng x_j chia cho số thí nghiệm N trong ma trận kế hoạch hóa, và công thức (2-4) trở thành:

$$b_j = \frac{1}{N} \sum_{i=1}^{N} x_{ji} y_i \qquad (2\text{-}6)$$

Tính tương hợp của các hệ số b_j được kiểm tra bằng công thức:

$$t_{bj} \geq t_p(f) \qquad (2\text{-}7a)$$

trong đó:

$t_p(f)$ là giá trị tra bảng của chuẩn số Student ở mức có nghĩa p và bậc tự do lặp $f = m-1$ (với m là số thí nghiệm tại tâm kế hoạch);

t_{bj} là chuẩn số Student của hệ số b_j được xác định theo công thức:

$$t_{bj} = \frac{|b_j|}{S_{bj}} \qquad (2\text{-}7)$$

Giá trị của độ lệch tiêu chuẩn S_{b_j} của phân bố b_j được xác định theo công thức:

$$S_{b_j} = \sqrt{\frac{S_{ll}^2}{\sum_{i=1}^{N} x_{ji}^2}} \qquad (2\text{-}8)$$

trong đó S^2_{ll} là phương sai lặp được xác định theo công thức:

$$S_{ll}^2 = \frac{\sum_{a=1}^{m} (y_{0a} - \overline{y}_0)^2}{m-1} \qquad (2\text{-}9)$$

trong đó:

y_{0a} là giá trị của hàm mục tiêu ở thực nghiệm thứ a tại tâm kế hoạch thực nghiệm:

y_0 là giá trị trung bình của m thực nghiệm tại tâm kế hoạch, được tính theo công thức:

$$y_0 = \frac{1}{m} \sum_{a=1}^{m} y_{0a} \qquad (2\text{-}10)$$

Tính tương hợp của mô hình với kết quả thực nghiệm được kiểm tra theo công thức sau:

$$F \le F_{1\text{-}p}(f1.f2) \qquad (2\text{-}11)$$

trong đó:

$F_{1\text{-}p}(f1.f2)$ là giá trị tra bảng của chuẩn số Fisher với mức ý nghĩa p, bậc tự do lặp f_2 và bậc tự do dư $f_1 = N - l$ (với l là hệ số có nghĩa trong phương trình hồi quy);

F là chuẩn số Fisher được xác định theo công thức:

$$F = \frac{S_d^2}{S_{ll}^2} \qquad (2\text{-}12)$$

trong đó:

S^2_{ll} là phương sai lặp được tính theo công thức (2-9).

S^2_d là phương sai dư được tính theo công thức:

$$S_d^2 = \frac{\sum_{i=1}^{N}(y_i - Y_i)^2}{N - l} \qquad (2\text{-}13)$$

với y_i, Y_i tương ứng là các giá trị thực nghiệm đo được và tính toán được của hàm mục tiêu.

Nếu công thức (2-11) thoả mãn, thì mô hình thống kê (hoặc hàm hồi quy) là tương hợp với kết quả thực nghiệm. Trường hợp ngược lại, phải cải tiến mô hình bằng cách chuyển sang mô tả hệ bằng hàm hồi quy bậc hai (mô hình thống kê phi tuyến bậc hai) vì khi đó các mối quan hệ giữa các thông số đầu vào và ra đã phức tạp hơn, không thể coi gần đúng là các quan hệ tuyến tính được nữa.

Nếu mục đích của người nghiên cứu là tìm chế độ hoạt động tối ưu của hệ công nghệ thì mặc dù các điều kiện về tương hợp (2-11) có thể đã được thỏa mãn, ta vẫn phải tiếp tục cải tiến mô hình để tiến về vùng dừng và cuối cùng phải xây dựng một quy hoạch bậc 2 tại vùng đó để tìm ra bộ giá trị cực trị của hệ (cực đại hay cực tiểu tùy theo mục đích nghiên cứu). Khi dùng kế hoạch bậc 1 đơn hình, ta tiến về vùng dừng theo thuật toán đơn hình và đường tiến lên là đường zig-zắc nhưng luôn hướng về vùng dừng, còn khi dùng kế hoạch bậc 1 hai mức tối ưu thì ta tiến về vùng dừng theo phương pháp gradient với thuật toán leo dốc hoặc xuống dốc theo đường dốc nhất (theo hướng gradient của bề mặt biểu diễn). Tại vùng dừng, mô hình tuyến tính sẽ không còn áp dụng được nữa (điều kiện về tính tương hợp (2-11) không được thỏa mãn) nên để mô tả được vùng này ta phải dùng mô hình phi tuyến bậc 2 với dạng điển hình chung là:

$$y = b_0 + \sum_{j=1}^{k} b_j x_j + \sum_{\substack{j,u=1 \\ j \neq u}}^{k} b_{ju} x_j x_u + \sum_{j=1}^{k} b_{jj} x_j^2 \qquad (2\text{-}14)$$

Để xác định các thông số của mô hình phi tuyến (2-14) cần làm thực nghiệm theo quy hoạch bậc 2 mà phương pháp thường được dùng nhất là quy hoạch bậc 2 Box-Wilson. Một số phương pháp quy hoạch bậc 2 khác cũng thường được ứng dụng là quy hoạch bậc 2 Box-Hunter, hay quy hoạch bậc 2 tối ưu D của Kiefer v.v.. Nội dung các phương pháp quy hoạch thường được dùng trong thực tế sẽ được nêu ra trong các phần tiếp theo.

2.2 Các phương pháp kế hoạch hóa thực nghiệm

Mục đích của kế hoạch hóa thực nghiệm (hay còn gọi là quy hoạch thực nghiệm) là tìm cách xây dựng một kế hoạch làm thực nghiệm sao cho chỉ cần số thí nghiệm ít nhất nhưng vẫn có khả năng mô tả mối quan hệ giữa các thông số đầu vào với thông số đầu ra một cách chính xác và đáng tin cậy theo quan điểm của xác suất thống kê.

Thông thường, khi tiến hành các khảo sát thực nghiệm người ta cần xác định ảnh hưởng của một số điều kiện thí nghiệm đối với một (hay một vài) thông số cần khảo sát. Khi kế hoạch hóa thực nghiệm, số các điều kiện thí nghiệm được coi là số các mức xác định đối với thông số cần khảo sát.

2.2.1 *Quy hoạch thực nghiệm bậc một*

Trong kế hoạch hóa thực nghiệm hai mức bậc 1, khi khảo sát ảnh hưởng của *k* yếu tố lên một thông số nào đó, số thí nghiệm cần làm sẽ bằng tổ hợp từ 2 mức của *k* yếu tố đó (mỗi yếu tố sẽ được khảo sát tại 2 giá trị cao và thấp), do vậy, tổng số thí nghiệm cần làm sẽ bằng 2^k. Do vậy, loại kế hoạch hóa thực nghiệm này còn được gọi là kế hoạch 2^k hay kế hoạch hóa thực nghiệm toàn phần. Vùng nằm trong giới hạn của các mức cao và thấp của mỗi yếu tố được nghiên cứu được gọi là vùng nghiên cứu theo các thông số công nghệ đã cho.

Thí dụ, giả sử nghiên cứu ảnh hưởng của ba yếu tố là: nhiệt độ (trong khoảng 100 – 200°C), áp suất (trong khoảng 20 – 60 kp/cm^2) và thời gian lưu (từ 10 đến 30 phút) lên hiệu suất tạo thành sản phẩm, ta sẽ biểu diễn các thông số này như sau:

Ký hiệu z_1 là nhiệt độ, z_2 là áp suất, z_3 là thời gian lưu, theo các điều kiện trên ta có: $z_1^{max} = 200°C$, $z_1^{min} = 100°C$, $\Delta z_1 = 50°C$, hay ở dạng tổng quát ta có:

$$z_j = \frac{z_j^{max} + z_j^{min}}{2} \;;\quad \Delta z_j = \frac{z_j^{max} - z_j^{min}}{2} \tag{2-15}$$

trong đó Δz_j là đơn vị thay đổi hay khoảng thay đổi theo trục z_j. Điểm nằm giữa hai mức cao và thấp được ký hiệu là z_j^0, đối với biến z_1 ở trên ta có $z_1^0 = 150°C$.

Điểm có tọa độ z_1^0, z_2^0, z_3^0... z_k^0 được gọi là tâm của kế hoạch hay mức cơ sở.

Để thuận tiện cho tính toán, ta cần chuyển sang hệ tọa độ không thứ nguyên $x_1, x_2, ..., x_k$ theo công thức:

$$x_j = \frac{z_j - z_j^0}{\Delta z_j} \quad \forall j = 1,2,..,k \qquad (3\text{-}16)$$

Trong hệ tọa độ không thứ nguyên mới này, tọa độ của mức trên là +1 còn mức dưới là -1; tọa độ tại điểm tâm là 0 và trùng với gốc tọa độ. Trong thí dụ trên của ta, $k = 3$. Số điểm thí nghiệm cần làm sẽ là $N = 2^3 = 8$ thí nghiệm. Bảng kế hoạch thí nghiệm tương ứng cùng các kết quả đầu ra (y_j) được thể hiện ở bảng dưới như sau:

Bảng 2.1 Các số liệu thí nghiệm theo quy hoạch 2^3

Số thứ tự các TN	Các giá trị thực của các biến			Các giá trị của các biến trong hệ tọa độ không thứ nguyên			Thông số đầu ra
	Z_1	Z_2	Z_3	x_1	x_2	x_3	y
1	100	20	10	-1	-1	-1	2
2	200	20	10	+1	-1	-1	6
3	100	60	10	-1	+1	-1	4
4	200	60	10	+1	+1	-1	8
5	100	20	30	-1	-1	+1	18
6	200	20	30	+1	-1	+1	10
7	100	60	30	-1	+1	+1	8
8	200	60	30	+1	+1	+1	12

Nếu hình dung ba biến x_1, x_2, x_3 như ba trục tọa độ trong không gian biến không thứ nguyên thì 8 điểm thực nghiệm của quy hoạch sẽ tạo thành một hình hộp lập phương mà các điểm thí nghiệm chính là các đỉnh của hình hộp đó. Tâm của quy hoạch cũng chính là tâm của hình hộp.

Ta viết lại ma trận kế hoạch thí nghiệm 2^3 và các kết quả thực nghiệm từ bảng 2.1 thành dạng như ở bảng 2.2, trong đó có đưa thêm vào cột biến hằng số $x_0 = 1$.

Bảng 2.2 Ma trận kế hoạch thực nghiệm 2^3

Số TT	x_0	x_1	x_2	x_3	y
1	+1	-	-	-	y_1
2	+1	+	-	-	y_2
3	+1	-	+	-	y_3
4	+1	+	+	-	y_4
5	+1	-	-	+	y_5
6	+1	+	-	+	y_6
7	+1	-	+	+	y_7
8	+1	+	+	+	y_8

Các số liệu (điểm thực nghiệm) được bố trí trong ma trận kế hoạch thực nghiệm 2.2 có các tính chất sau:

$$\sum_{i=1}^{N} x_{ui} x_{ji} = 0 \quad u \neq j; u, j = 1,2,...,k \tag{2-17}$$

$$\sum_{i=1}^{N} x_{ji} = 0 \qquad j = 1,2,...,k \tag{2-18}$$

$$\sum_{i=1}^{N} x_i^2 = N \qquad j = 0,1,...,k \tag{2-19}$$

trong đó k – số yếu tố độc lập,

N – số thí nghiệm

Tính trực giao (tính chất (2-18)) của ma trận kế hoạch hóa thực nghiệm giúp cho đơn giản hóa nhiều trong khâu tính toán các hệ số của phương trình hồi quy, bởi vì ma trận hệ số của phương trình chuẩn (X*X) trở thành ma trận đường chéo và các phân tử của nó chính bằng số thí nghiệm N trong ma trận kế hoạch hóa. Các phần tử đường chéo của ma trận ngược $(X^*X)^{-1}$ là $C_{jj} - 1/N$. Từ đó, các hệ số của phương trình hồi quy $b_0, b_1,..,b_k$ được tính theo công thức đơn giản như sau (ký hiệu B là véc tơ cột chứa các hệ số $b_0, b_1,..,b_k$):

$$B = \begin{bmatrix} b_0 \\ b_1 \\ .. \\ .. \\ b_k \end{bmatrix} = (X^*X)^{-1}X^*Y = \begin{bmatrix} 1/N & & 0 \\ & 1/N & \\ & & \\ 0 & & 1/N \end{bmatrix} \begin{bmatrix} \sum x_{0i}y_i \\ \sum x_{1i}y_i \\ \\ \sum x_{ki}y_i \end{bmatrix} \tag{2-20}$$

$$B = \begin{bmatrix} \dfrac{\sum x_{0i}y_i}{N} \\ \dfrac{\sum x_{1i}y_i}{N} \\ \\ \dfrac{\sum x_{ki}y_i}{N} \end{bmatrix} \tag{3-21}$$

Như vậy, hệ số bất kỳ của phương trình hồi quy b_j được xác định bằng cách nhân cột y với cột tương ứng x_j, sau đó đem chia cho số thí nghiệm N trong ma trận kế hoạch hóa thực nghiệm:

$$b_j = \frac{1}{N}\sum_{i=1}^{N} x_{ji}y_i \tag{2-22}$$

Phương trình hồi quy tuyến tính bậc 1 cho 3 biến như trong thí dụ này có thể được viết ở dạng đơn giản chỉ gồm các hệ số đơn:

$$y = b_o + b_1x_1 + b_2x_2 + b_3x_3 \tag{2-23}$$

hay ở dạng đầy đủ hơn, bao gồm cả các hệ số tương tác tương hỗ:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3 \quad (2\text{-}24)$$

Để xác định các hệ số tương tác kép (b_{12}, b_{13}, b_{23}) và hệ số tác dụng ba (b_{123}) cần phải mở rộng ma trận ở bảng 2.2 thành ma trận ở dạng như bảng 2.3. Ma trận trong bảng 2.3 có thêm các cột chứa các hệ số tương tác với giá trị được tính bằng cách nhân theo hàng ngang các giá trị trên các cột x_i, x_j tương ứng. Tính toán các hệ số tương tác bội cũng tương tự như tính toán các hệ số đơn trong phương trình hồi quy.

Bảng 2.3: Ma trận mở rộng để tính các hệ số tương tác bội

Số TT	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	y
1	+1	-1	-1	-1	+1	+1	+1	-1	2
2	+1	+1	-1	-1	-1	-1	+1	+1	6
3	+1	-1	+1	-1	-1	+1	-1	+1	4
4	+1	+1	+1	-1	+1	-1	-1	-1	8
5	+1	-1	-1	+1	+1	-1	-1	+1	10
6	+1	+1	-1	+1	-1	+1	-1	-1	18
7	+1	-1	+1	+1	-1	-1	+1	-1	8
8	+1	+1	+1	+1	+1	+1	+1	+1	12

Nếu tiến hành các thí nghiệm song song thì có thể xác định phương sai lặp S^2_{ll} theo công thức (2-9), kiểm tra tính có nghĩa của các hệ số hồi quy và tính tương hợp của mô hình.

Ta nhận thấy rằng ma trận tương quan $(X^*X)^{-1}$ đối với thực nghiệm được kế hoạch hóa trên bảng 2.3 là ma trận đường chéo, có nghĩa là các hệ số của phương trình hồi quy không tương quan với nhau. Từ đây, ta có thể xác định mức độ có nghĩa của các hệ số của phương trình hồi quy theo chuẩn số Student và việc loại bỏ những hệ số không có nghĩa không ảnh hưởng đến giá trị của các hệ số còn lại.

Sau khi được tính toán từ các số liệu thực nghiệm, các giá trị của các hệ số b_j chính là các đánh giá của các hệ số lý thuyết tương ứng β_j (tức hệ số hồi quy lý thuyết), hay có thể ký hiệu $b_j \rightarrow \beta_j$. Từ đây có thể suy ra rằng. giá trị độ lớn của từng hệ số (b_j) đặc trưng cho mức độ đóng góp của yếu tố tương ứng (x_j) đối với đại lượng được nghiên cứu (y) như được thể hiện trong phương trình hồi quy.

Từ các giá trị trên bảng 2.3, tính toán được giá trị các hệ số b_j như sau:

$$b_0 = \frac{\sum_{i=1}^{8} x_{0i} y_i}{8} = \frac{1.2 + 1.6 + 1.4 + 1.8 + 1.10 + 1.18 + 1.8 + 1.12}{8} = 8.5$$

$$b_1 = \frac{\sum_{i=1}^{8} x_{1i} y_i}{8} = \frac{-1.2 + 1.6 - 1.4 + 1.8 - 1.10 + 1.18 - 1.8 + 1.12}{8} = 2.5$$

Tiến hành tương tự cho các hệ số còn lại, ta được các kết quả tiếp theo như sau:

$b_2 = -0,5$; $b_3 = 3,5$; $b_{12} = -0,5$; $b_{13} = 0,5$; $b_{23} = -1,5$; $b_{123} = -0,5$

Do phần tử đường chéo của ma trận tương quan là bằng nhau nên tất cả các hệ số của phương trình (2-22) và (2-23) được xác định với độ chính xác như nhau:

$$S_{b_j} = \frac{S_{ll}}{\sqrt{N}}$$

Để tính phương sai lặp lại, tiến hành 3 thí nghiệm tại tâm ($x_1 = 0$, $x_2 = 0$, $x_3 = 0$. tương đương với điều kiện: $z_1^0 = 150^oC$, $z_2^0 = 40at$, $z_3^0 = 20$ phút) thu được các kết quả là: $y_1^0 = 8$; $y_2^0 = 9$; $y_3^0 = 8,8$. Từ đó tính ra được giá trị trung bình $\overline{y_0} = 8,6$. Áp dụng công thức (3-9) tính được phương sai lặp là $S_{ll}^2 = 0,28$.

Đồng thời tính được: $S_{b_j} = \frac{S_{ll}}{\sqrt{N}} = 0,19$

Tra bảng Giá trị chuẩn số Student trong *Phụ lục 1* với $f_2 = 3 - 1 = 2$ và mức ý nghĩa $p = 0,05$ được $t_{0,05;\,2} = 4,303$. Đối chiếu với điều kiện có nghĩa của hệ số hồi quy thực nghiệm (2-7a) sẽ xác định được những hệ số có nghĩa còn lại của phương trình. Cụ thể như sau:

Từ điều kiện (2-7a) và (2-7), để hệ số có nghĩa cần thỏa mãn các điều kiện:

$$t_{bj} = \frac{|b_j|}{S_{bj}} \geq t_{p.f2} \quad \text{hay} \quad |b_j| \geq S_{bj}.t_{p.f2}$$

Theo các số liệu ở trên tính ra được: $t_{0,05;\,2}.S_{bj} = 4,303 \times 0,19 = 0,818$.

So sánh giá trị này với giá trị các hệ số b_j tìm được ở trên nhận thấy: chỉ có các hệ số b_0, b_1, b_3 *và* b_{23} là có nghĩa (có giá trị $|b| \geq 0,818$). Như vậy, phương trình hồi quy thực nghiệm thu được cuối cùng có dạng:

$$y = 8,5 + 2,5\, x_1 + 3,5\, x_3 - 1,5\, x_2x_3 \qquad (2\text{-}25)$$

Để đánh giá mức độ tương hợp của mô hình trong việc mô tả bức tranh thực, cần kiểm tra tỷ số giữa phương sai dư và phương sai lặp lại chuẩn số theo Fisher. Các bước thứ tự như sau:

Phương sai dư được tính theo công thức:

$$S^2_{du} = \frac{1}{n-l}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \qquad (2\text{-}26)$$

trong đó: l là số hệ số có nghĩa còn lại trong phương trình hồi quy thực nghiệm thu được. Cụ thể, trong thí dụ này ta có:

$$S_{du}^2 = \frac{1}{8-4}\{[2-(8,5+2,5.(-1)+3,5.(-1)-1,5.(1))]^2 +$$

$$[6-(8,5+2,5.(1)+3,5.(-1)+1,5.(1))]^2 + \ldots\} = \frac{8}{4} = 2$$

Phương sai lặp lại được tính theo các thí nghiệm tại tâm như đã làm ở trên và đã thu được kết quả là $S_{ll}^2 = 0,28$. Từ đây tính được ra:

$$F = \frac{2}{0,28} = 7,14$$

Tra bảng giá trị chuẩn số Fisher trên bảng Phụ lục 2 ứng với bậc tự do lặp $f_2 = 3 - 1 = 2$ và bậc tự do dư $f_1 = N - l = 8 - 4 = 4$ theo mức ý nghĩa 5% ta có $F_{2;4;0,05} = 19,2$

Do giá trị F tính được là $7,14 < 19,2$ nên mô hình hồi quy thu được ở đây được coi là tương hợp với bức tranh mô tả thực nghiệm.

Quy hoạch riêng phần:

Theo phương pháp kế hoạch hóa toàn phần, số điểm thí nghiệm cần làm là 2^k trong đó k là số yếu tố cần khảo sát. Hiển nhiên là khi số yếu tố k này tăng lên thì số thí nghiệm phải làm cũng sẽ tăng lên theo quy luật hàm mũ và sẽ là rất lớn khi k đủ lớn (thí dụ $k \geq 5$). Trong những trường hợp này, để có thể vẫn sử dụng được kế hoạch hóa thực nghiệm ta có thể áp dụng phương pháp kế hoạch hóa riêng phần 2^{k-p} để xác định các tham số của phương trình hồi quy thực nghiệm. Trong quy hoạch loại này, p là số quan hệ phát sinh và cũng chính là số hiệu ứng tuyến tính được thay bằng các hiệu ứng tương hỗ.

Trở lại với thí dụ trên, dùng quy hoạch riêng phần để xác định các hệ số của phương trình (2-23) với $p = 1$, khi đó, số thí nghiệm phải làm sẽ là $2^{3-1} = 4$ thí nghiệm theo kế hoạch ở bảng 2.4 dưới đây, trong đó thay cột giá trị x_3 vào cột hiệu ứng kép x_1x_2 (vì thông thường các hiệu ứng kép chỉ đóng vai trò thứ yếu).

Bảng 2-4: Bảng kế hoạch hóa riêng phần 2^{3-1}

i \ j	x_0	x_1	x_2	x_3	y
1	1	-1	-1	+1	10
2	1	+1	-1	-1	6
3	1	-1	+1	-1	4
4	1	+1	+1	+1	12

Khi dùng kế hoạch hóa riêng phần, các hệ số tuyến tính chính là các hệ số hỗn hợp, nó cũng bao gồm trong nó cả các hệ số tương hỗ, cụ thể:

$$b_1 \to \beta_1 + \beta_{23}\,; \qquad b_2 \to \beta_2 + \beta_{13}\,; \qquad b_3 \to \beta_3 + \beta_{12}$$

Khi sử dụng kế hoạch hóa riêng phần, cần phải xác định quan hệ nào sẽ được dùng làm quan hệ phát sinh để đưa ra các thông tin lớn nhất. Về nguyên tắc, nên chọn đúng quan hệ phát sinh tương ứng (các cột x_j có được trong bảng ma trận, trong thí dụ trên là các cột tương ứng các biến x_1 và x_2); còn nếu không có thì dùng tích tương hỗ bậc cao nhất (trong thí dụ này dùng cột tích x_1x_2 cho biến x_3).

Trong các bài toán thực, các tương tác ba thường có ít ý nghĩa hơn so với các tác dụng kép, do vậy, khi cần khảo sát 4 thông số đầu vào (x_1 – x_4), để giảm bớt số thí nghiệm cần thiết, người ta thường chỉ làm quy hoạch riêng phần 2^{4-1} = 8 thí nghiệm, trong đó cột x_4 được đặt vào cột tác dụng ba:

$$x_4 = x_1x_2x_3 \qquad (2\text{-}26)$$

Khi dùng quan hệ trên để lập kế hoạch thực nghiệm, ma trận kế hoạch hóa sẽ có dạng như ở bảng 2.5a dưới đây:

Bảng 2.5a: Ma trận kế hoạch hóa riêng phần 2^{4-1} với $x_4 = x_1x_2x_3$

No.	x_0	x_1	x_2	x_3	x_4
1	+1	+1	+1	+1	+1
2	+1	-1	-1	+1	+1
3	+1	-1	+1	+1	-1
4	+1	+1	-1	+1	-1
5	+1	+1	+1	-1	-1
6	+1	-1	-1	-1	-1
7	+1	-1	+1	-1	+1
8	+1	+1	-1	-1	+1

Khi dùng giao thoa xác định $x_1x_2x_3x_4 = 1$, ta có hệ thống đánh giá hỗn hợp:

$$\left.\begin{array}{ll}
x_0 = x_1x_2x_3x_4 = x_{jj}^2 & b_0 \to \beta_0 + \beta_{1234} + \sum_{j=1}^{k} \beta_{jj} \\
x_1 = x_2x_3x_4 & b_1 \to \beta_1 + \beta_{234} \\
x_2 = x_1x_3x_4 & b_2 \to \beta_2 + \beta_{134} \\
x_3 = x_1x_2x_4 & b_3 \to \beta_3 + \beta_{124} \\
x_4 = x_1x_2x_3 & b_4 \to \beta_4 + \beta_{123} \\
x_1x_2 = x_3x_4 & b_{12} \to \beta_{12} + \beta_{34} \\
x_1x_3 = x_2x_4 & b_{13} \to \beta_{13} + \beta_{24} \\
x_1x_4 = x_2x_3 & b_{14} \to \beta_{14} + \beta_{23}
\end{array}\right\} \qquad (2\text{-}27)$$

Ta cũng có thể dùng một quan hệ phát sinh khác, thí dụ:

$$x_4 = x_1x_2 \qquad (2\text{-}28)$$

Khi đó, ma trận ứng với tương quan trên sẽ được viết như trong bảng 2.5b, trong đó các cột x_1, x_2 và x_3 vẫn như trước, chỉ khác cột x_4 được tính theo tích của hai cột x_1 và x_2.

Bảng 2.5b: Ma trận kế hoạch hóa riêng phần 2^{4-1} với $x_4 = x_1x_2$

No.	x_0	x_1	x_2	x_3	x_4
1	+1	+1	+1	+1	+1
2	+1	-1	-1	+1	+1
3	+1	-1	+1	+1	-1
4	+1	+1	-1	+1	-1
5	+1	+1	+1	-1	+1
6	+1	-1	-1	-1	+1
7	+1	-1	+1	-1	-1
8	+1	+1	-1	-1	-1

Giao thoa xác định tương ứng sẽ là: $x_1x_2x_4 = 1$

Hệ thống đánh giá khi này sẽ là:

$$\left.\begin{array}{ll}
x_0 = x_1x_2x_4 = x_{ii}^2 & b_0 \to \beta_0 + \beta_{124} + \sum_{i=1}^{k}\beta_{ii} \\
x_1 = x_2x_4 & b_1 \to \beta_1 + \beta_{24} \\
x_2 = x_1x_4 & b_2 \to \beta_2 + \beta_{14} \\
x_3 = x_1x_2x_3x_4 & b_3 \to \beta_3 + \beta_{1234} \\
x_4 = x_1x_2 & b_4 \to \beta_4 + \beta_{12} \\
x_1x_3 = x_2x_3x_4 & b_{13} \to \beta_{13} + \beta_{234} \\
x_2x_3 = x_1x_3x_4 & b_{23} \to \beta_{23} + \beta_{134} \\
x_3x_4 = x_1x_2x_3 & b_{34} \to \beta_{34} + \beta_{123}
\end{array}\right\} \qquad (2\text{-}29)$$

Như vậy, sử dùng kế hoạch hóa riêng phần theo quan hệ $x_4 = x_1x_2$ sẽ mang ngụ ý rằng ta đang quan tâm đến các hệ số β_{13}, β_{23} và β_{34}. Như vậy, loại kế hoạch riêng phần mà trong đó p hiệu ứng tuyến tính (bậc 1) được thay bằng các hiệu ứng tương hỗ sẽ được ký hiệu là kế hoạch hóa riêng phần 2^{k-p}.

Tóm lại, các kế hoạch trực giao hai mức tối ưu 2^k và 2^{k-p} có các ưu điểm sau:

- Do là kế hoạch trực giao nên tính toán đơn giản,
- Tất cả các hệ số đều được xác định không phụ thuộc vào nhau,
- Mỗi hệ số đều được xác định theo kết quả của toàn bộ N thí nghiệm,
- Tất cả các hệ số hồi quy đều được xác định với phương sai bé nhất và là như nhau (kế hoạch có tính tối ưu D).

Ngoài ra, cần lưu ý rằng các kế hoạch 2^k và 2^{k-p} có tính chất tâm xoay. Do không có tương quan giữa các hệ số, theo định luật cộng phương sai trong trường hợp phương trình tuyến tính k yếu tố, ta sẽ có:

$$S_y^2 = S_{b_0}^2 + x_1^2 S_{b_1}^2 + \dots + x_k^2 S_{b_k}^2$$

Vì $S_b^2 = \dfrac{S_{II}^2}{N}$ nên ta có:

$$S_y^2 = \frac{S_{II}^2}{N}(1 + x_1^2 + \dots + x_k^2) = \frac{S_{II}^2}{N}(1 + \rho^2) \qquad (2\text{-}30)$$

với $\rho^2 = \sum_{i=1}^{k} x_i^2$

trong đó ρ là bán kính cầu trong không gian k chiều.

Đại lượng nghịch đảo của S_y^2 có thể coi như mức độ thông tin chứa trong phương trình hồi quy.

Theo phương trình (2-30), lượng thông tin giảm tỷ lệ thuận với bình phương bán kính cầu ρ^2 và là đồng nhất với tất cả các điểm có khoảng cách như nhau. Do vậy, loại kế hoạch này được gọi là kế hoạch tâm xoay.

2.2.2 Tiến đến vùng tối ưu theo phương pháp leo dốc

Giả sử ta cần tìm giá trị cực đại (hay cực tiểu) của một hàm số ở dạng tổng quát $y = f(x_1, x_2, \ldots, x_n)$, theo các nguyên tắc của toán giải tích, đường đi ngắn nhất từ một điểm bất kỳ trên bề mặt đồng mức của hàm số đó sẽ là đường đi theo hướng gradient. Nếu hàm có k biến số, gradient trong không gian k chiều sẽ được biểu diễn theo công thức chung:

$$grady = \frac{df}{dx_1}\vec{i} + \frac{df}{dx_2}\vec{j} + \ldots + \frac{df}{dx_k}\vec{k} \qquad (2\text{-}31)$$

trong đó $\vec{i}, \vec{j}, \ldots, \vec{k}$ là các véc tơ đơn vị trên các trục tọa độ.

Đối với bài toán quy hoạch thực nghiệm, giả sử tại một thời điểm nào đó, ta đã tìm ra được phương trình hồi quy mô tả quan hệ giữa các thông số đầu vào (2 biến số x_1 và x_2) và đầu ra (y) là một phương trình hồi quy bậc nhất:

$$Y = b_0 + b_1 x_1 + b_2 x_2 \qquad (2\text{-}32)$$

Khi đó, hướng gradient của hàm hồi quy Y sẽ chính là hướng theo hai hệ số b_1 và b_2 vì

$$\frac{dY}{dx_1} = b_1 \quad ; \quad \frac{dY}{dx_2} = b_2 \qquad (2\text{-}33)$$

Lưu ý rằng, nghiệm của bài toán quy hoạch tuyến tính luôn luôn nằm trên biên, nên đối với trường hợp hàm hồi quy là bậc một, có thể xác định ngay điểm tối ưu (hàm y đạt giá trị lớn nhất hay nhỏ nhất) theo giá trị tối đa (+1) hay tối thiểu (-1) của các biến x_1 và x_2 tùy theo tùy theo giá trị của các hệ số b_1, b_2 là dương hay âm. Để tìm tối ưu tiếp, có thể lấy điểm này làm điểm khởi đầu rồi xây dựng một kế hoạch thực nghiệm mới để khảo sát tiếp ở vùng lân cận theo các biến x_1 và x_2. Tại miền có chứa cực trị, hay còn gọi là miền hầu như ổn định, hàm hồi quy bậc một sẽ không còn phù hợp để mô tả các mối quan hệ giữa các biến *X-Y*. Khi đó sẽ phải xây dựng hàm hồi quy bậc hai để có thể mô tả chính xác hơn mối quan hệ phi tuyến này.

2.2.3 Quy hoạch thực nghiệm bậc 2 (trực giao)

Trong vùng gần với cực trị, các hiệu ứng tương hỗ và các hiệu ứng bình phương trở nên có ý nghĩa rõ rệt. Hiện tại, quy hoạch bậc hai là phương pháp duy nhất được dùng để mô tả các mối quan hệ phi tuyến giữa các thông số đầu vào và đầu ra trong quy hoạch thực nghiệm.

Có một số phương pháp xây dựng kế hoạch thực nghiệm bậc hai khác nhau. Tuy nhiên, được dùng phổ biến nhất vẫn là kế hoạch hóa bậc hai trực giao do có ưu điểm cơ bản là giúp cho đơn giản hóa nhiều trong quá trình tính toán. Nguyên tắc chung khi xây dựng các ma trận kế hoạch thực nghiệm trong quy hoạch bậc hai là:

Ma trận kế hoạch thực nghiệm bậc hai có thể coi là gồm hai phần: phần kế hoạch hóa thực nghiệm của quy hoạch bậc một và thêm một phần mở rộng để phục vụ cho quy hoạch bậc hai

- Các điểm thực nghiệm trong phần mở rộng của ma trận kế hoạch hóa thực nghiệm cho k thông số bao gồm $2k$ điểm bổ sung (còn gọi là điểm sao – ký hiệu dấu *) có tọa độ là $+\alpha$ và $-\alpha$, và n_o điểm thực nghiệm tại tâm, trong đó giá trị của α và số điểm thực nghiệm tại tâm n_o sẽ khác nhau và phụ thuộc vào loại kế hoạch hóa thực nghiệm được sử dụng.

- Tổng số điểm thực nghiệm cần thiết khi lập kế hoạch hóa cho k thông số sẽ là:

$N = 2^k + 2k + n_0$ khi dùng quy hoạch toàn phần 2^k. hay

$N = 2^{k-p} + 2k + n_0$ khi dùng quy hoạch bán phần 2^{k-p}

Đối với quy hoạch trực giao bậc hai, số điểm thực nghiệm tại tâm n_o có thể chọn tùy ý, vì vậy thường lấy $n_o = 1$. Để có được tính trực giao cho ma trận thực nghiệm, toạ độ cho các điểm sao cần xác định thích hợp. Cụ thể như sau:

Xét trường hợp quy hoạch cho 2 yếu tố ($k = 2$), bảng ma trận thực nghiệm khi này sẽ được biểu diễn như trong Bảng 2-6 dưới đây:

Bảng 2-6: Ma trận thực nghiệm của quy hoạch bậc 2 cho 2 biến

N	x_0	x_1	x_2	x_1x_2	x_1^2	x_2^2
1	+1	+1	+1	+1	+1	+1
2	+1	+1	-1	-1	+1	+1
3	+1	-1	-1	+1	+1	+1
4	+1	-1	+1	-1	+1	+1
5	+1	$+\alpha$	0	0	α^2	0
6	+1	$-\alpha$	0	0	α^2	0
7	+1	0	$+\alpha$	0	0	α^2
8	+1	0	$-\alpha$	0	0	α^2
9	+1	0	0	0	0	0

Ma trận này không thể là trực giao do không thỏa mãn được các điều kiện trực giao đối với các biến trên các cột x_1^2 và x_2^2, cụ thể là:

$$\sum_{i=1}^{N} x_{0i} x_{ji}^2 \neq 0$$

$$\sum_{i=1}^{N} x_{ji}^2 x_{ui}^2 \neq 0$$

Để đưa ma trận trên về dạng trực giao cần tiến hành một số phép biến đổi, thay các biến trực tiếp x_1^2 và x_2^2 bằng các biến x'_j mới:

$$x'_j = x_j^2 - \frac{\sum_{i=1}^{N} x_{ji}^2}{N} = x_j^2 - \overline{x_j^2} \tag{2-34}$$

Khi đó, các cột biến mới x'_j trong ma trận kế hoạch hóa thực nghiệm cũng sẽ thỏa mãn điều kiện trực giao:

$$\begin{aligned} &\sum_{i=1}^{N} x_{oi} x'_{ji} = \sum_{i=1}^{N} (x_{ji}^2 - \overline{x_j^2}) = 0 \\ &\sum_{i=1}^{N} x'_{ji} . x_{ui} = 0 \end{aligned} \tag{2-35}$$

Cuối cùng, để ma trận có thể trực giao hoàn toàn, người ta chọn khoang cách α sao cho thỏa mãn điều kiện đẳng thức bằng không của các số hạng không đường chéo của ma trận tương quan $(XX^*)^{-1}$

$\alpha^4 + 2^k\alpha^2 - 2^{k-1}.(k + 0{,}5n_0) = 0.$ đối với quy hoạch toàn phần, hay

$\alpha^4 + 2^{k-p}\alpha^2 - 2^{k-1-p}.(k + 0{,}5n_0) = 0$ đối với quy hoạch bán phần

Các kết quả tính toán giá trị cụ thể của α theo các giá trị k khác nhau được đưa ra trong bảng 2-7 dưới đây:

Bảng 2-7: Giá trị của α theo số biến k và loại quy hoạch

Số yếu tố độc lập k	2	3	4	5
Quy hoạch loại	2^2	2^3	2^4	2^{5-1}
Giá trị α	1	1,215	1,414	1,547

Ma trận dùng để tính toán các hệ số của phương trình hồi quy có sử dụng các biến mới x'_1 và x'_2 được đưa ra trong bảng 2-8 ở dưới:

N^0	x_0	x_1	x_2	$x_1 x_2$	x'_1	x'_2
1	+1	+1	+1	+1	+1/3	+1/3
2	+1	+1	-1	-1	+1/3	+1/3
3	+1	-1	-1	+1	+1/3	+1/3
4	+1	-1	+1	-1	+1/3	+1/3
5	+1	+1	0	0	+1/3	-2/3
6	+1	-1	0	0	+1/3	-2/3
7	+1	0	+1	0	-2/3	+1/3
8	+1	0	-1	0	-2/3	+1/3
9	+1	0	0	0	-2/3	-2/3

Do tính trực giao của ma trận nên các hệ số hồi quy sẽ độc lập với nhau và được xác định theo các công thức sau:

$$\left.\begin{aligned} b'_0 &= \frac{\sum_{i=1}^{N} y_i}{N} \\ b_i &= \frac{\sum_{i=1}^{N} x_{ji} \cdot y_i}{\sum_{i=1}^{N} x_{ji}^2} \\ b_{ii} &= \frac{\sum_{i=1}^{N} x'_{ji} \cdot y_i}{\sum_{i=1}^{N} (x'_{ji})^2} \end{aligned}\right\} \qquad (2\text{-}36)$$

Phương sai của các hệ số được tính theo công thức:

$$S_{b_i}^2 = \frac{S_{th}^2}{\sum_{i=1}^{N} x_{ji}^2} \qquad (2\text{-}37)$$

Kết quả tính toán sẽ cho ta phương trình hồi quy bậc 1 của các biến độc lập (x_j) và các biến mới (x'_j):

$$Y = b'_0 + \sum_{j=1}^{k} b_j.x'_j + \sum_{\substack{j,u=1 \\ j\neq u}}^{k} b_{ju}.x'_j.x'_u + \sum_{j=1}^{k} b_{jj}.x'_j \quad (2\text{-}38)$$

Đổi ngược trở lại các x'_j thành các biến ban đầu x_j theo công thức biến đổi (2-34) ta sẽ thu được dạng cuối cùng của phương trình hồi quy bậc hai theo các biến nguyên gốc x_j. Từ phương trình hồi quy (2-38), giá trị của hệ số b_0 được xác định lại theo công thức:

$$b_0 = b'_0 - b_{11}.\overline{x}_1^2 - \ldots - b_{kk}.\overline{x}_k^2 \quad (2\text{-}39)$$

và phương sai của hệ số này được đánh giá theo:

$$S_{b_0}^2 = S_{b'_0}^2 - \sum_{j=1}^{k} S_{b_{jj}}^2 (\overline{x}_j^2)^2 \quad (2\text{-}40)$$

Phương trình hồi quy bậc hai cuối cùng sẽ được viết lại ở dạng:

$$Y = b_0 + \sum_{j=1}^{k} b_j.x_j + \sum_{\substack{j,u=1 \\ j\neq u}}^{k} b_{ju}.x_j.x_u + \sum_{j=1}^{k} b_{jj}.x_j^2 + \ldots \quad (2\text{-}41)$$

Sau khi có được phương trình hồi quy bậc hai, các bước tiếp theo cũng tương tự khi xử lý phương trình hồi quy bậc một, bao gồm:

Kiểm tra tính có nghĩa của các hệ số theo tiêu chuẩn Student dựa trên giá trị của phương sai lặp; loại bỏ các hệ số vô nghĩa:

1. Kiểm tra tính tương hợp của phương trình hồi quy theo chuẩn số Fisher theo tỷ số của các phương sai.

Phương trình sẽ là tương hợp nếu tỷ số của các phương sai nhỏ hơn giá trị ứng với mức ý nghĩa p (thường lấy bằng 0,05) và số bậc tự do của phương sai dư và phương sai lặp.

$$F < F_p(f_1, f_2) \qquad (2\text{-}42)$$

trong đó: $f_1 = N - l$: bậc tự do của phương sai dư,

f_2 : bậc tự do của phượng sai lặp

N : số thí nghiệm trong ma trận kế hoạch thực nghiệm

l : số hệ số trong phương trình hồi quy bậc 2 đối với k yếu tố

Tính ý nghĩa của hệ số được kiểm tra theo chuẩn số Student theo giá trị $t_{bj} = |b_j| / S_{bj}$. Hệ số sẽ có nghĩa nếu $t_{bj} > t_{p,f2}$. Cần lưu ý rằng trong trường hợp quy hoạch trực giao bậc hai, các hệ số của phương trình hồi quy được xác định với độ chính xác khác nhau, trong khi quy hoạch trực giao bậc một lại cho phép xác định các hệ số với độ chính xác như nhau.

2.3 Thí dụ minh họa

Xây dựng mô hình toán học cho quá trình chuyển hóa sinh hóa học

Trong thí dụ này sẽ xem xét việc xây dựng mô hình hồi quy để khảo sát một quá trình chuyển hóa hóa học trên cơ sở xử lý các số liệu thực nghiệm được tổ chức theo các phương pháp quy hoạch thực nghiệm khoa học. Động học của quá trình trong trường hợp này được đánh giá là phụ thuộc vào hai yếu tố chính là:

- lưu lượng thổi gió: $Z_1 = 38{,}4 - 97{,}7$ l/h
- thời gian phản ứng: $Z_2 = 2 - 6$ ngày

Mô hình được xây dựng theo các bước sau:

Mã hóa các biến theo công thức:

$$X_i = \frac{Z_i - Z_i^0}{\Delta Z_i}$$

Dạng tổng quát của mô hình này là:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2$$

trong đó: $\hat{y} = \frac{N}{N_0}$: mức độ tăng trưởng của khối sinh hóa;

N_0, N : lượng chất ban đầu và lượng chất chuyển hóa được tại thời điểm t

x_1, x_2 : biến mã hóa của hai thông số cần khảo sát

Ma trận kế hoạch hóa thực nghiệm hai mức tối ưu được lập ra như ở bảng dưới (các cột chứa biến x_j). Sau khi tiến hành các thí nghiệm, thu được các kết quả thực nghiệm tương ứng (cột y).

j / i	Biến thực		Biến mã hóa				
	Z_1	Z_2	x_0	x_1	x_2	x_1x_2	y
1	38.4	6	+	-	+	-	3.28
2	97.7	6	+	+	+	+	3,25
3	38.4	2	+	-	-	+	1,685
4	97.7	2	+	+	-	-	6,91

Xác định các hệ số của phương trình hồi quy: Các hệ số b_j của phương trình hồi quy thực nghiệm được tính từ bảng số liệu thí nghiệm trên theo công thức sau:

$$b_j = \frac{1}{N}\sum x_{ji}.y_i$$

và thu được các giá trị:

$b_0 = 3{,}556 \quad b_1 = -1{,}074 \quad b_2 = -0{,}291 \quad b_{12} = -1{,}089$

Xác định mức ý nghĩa của các hệ số b_j: thực hiện theo các bước sau:

- Tính phương sai lặp lại của các thí nghiệm: thường được xác định bằng cách làm một số thí nghiệm tại tâm ($x_1 = 0$. $x_2 = 0$). Ở đây làm 2 thí nghiệm và thu được kết quả như sau: $y^0_1 = 3.250$; $y^0_2 = 3{,}185$. Giá trị trung bình của 2 giá trị trên: $y^0 = 3.2175$. Phương sai lặp được tính theo công thức:

$$S_{tt}^2 = \frac{1}{m-1}\sum_{u=1}^{m}(y_u^0 - y^0)^2 = \frac{1}{2-1}[(3.25-3.2175)^2 + (3.185-3.2175)^2] = 0{,}002125$$

- Tính phương sai cho các hệ số b_j:

$$S_{b_j}^2 = \frac{S_{tt}^2}{4} = \frac{0.002125}{4} = 0.000531$$; từ đây tính ra $S^2_{bj} = 0{,}0230$

- Xác định giá trị tra bảng chuẩn số Student:

Với bậc tự do lặp $f_2 = m\text{-}1 = 2\text{-}1 = 1$. và mức ý nghĩa $p = 0{,}05$, tra được giá trị $t_{1;0,05} = 12.706$.

- Xác định sự có nghĩa của các hệ số b_j và dạng của phương trình hồi quy cuối cùng:

Các hệ số b_j sẽ có nghĩa khi:

$$|b_j| \geq S_{bj}t_{1;0,05} = 0.023.12.706 = 0{,}293$$

So sánh giá trị 0,293 ở trên với các giá trị b_j tính được trước đó, thấy chỉ có giá trị của b_2 là 0,291 < 0,293. vì vậy hệ số b_2 là không có nghĩa và bị bỏ đi khỏi phương trình hồi quy thực nghiệm. Phương trình hồi quy cuối cùng sẽ có dạng:

$$y = 3{,}556 - 1{,}074.x_1 - 1{,}089.x_1.x_2$$

Tính tương hợp của mô hình thu được: được xác định nhờ chuẩn số Fisher. Giá trị này được xác định theo công thức:

$$F = \frac{S^2_{du}}{S^2_{ll}}$$ trong đó S^2_{du} được tính theo công thức (2-26)

Thay số vào tính được: $S^2_{du} = 0,579$ và $F = 0,579/0,002125=272,45$

Tra bảng Phụ lục 2 của chuẩn số Fisher ứng với bậc tự do lặp $f_2 = 1$, bậc tự do dư $f_l = N - 1 = 4 - 3 = 1$ và mức ý nghĩa $p = 0,05$ tìm được giá trị: $F_{1:1:0.05} = 164$. Giá trị này nhỏ hơn giá trị F tính được (272,45): điều này dẫn đến kết luận là mô hình tuyến tính xác định được ở trên không tương hợp với bức tranh thực nghiệm. Một nhận xét khác là hai giá trị thực nghiệm và tính toán tại tâm (y_0 và b_0) chênh lệch nhau rất đáng kể:

$$|b_0 - y_0| = 3,556 - 3,2175 = 0,3385$$

Từ đó, ta có thể kết luận là mô hình tuyến tính không có khả năng mô tả chính xác quá trình này. Vì vậy cần xây dựng mô hình phi tuyến thích hợp. Mô hình đơn giản nhất là đa thức bậc 2 có dạng thức đã được mã hóa theo quy hoạch bậc 2 trực giao như sau:

$$y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 - b_{11}x'_1 + b_{22}x'_2$$

Ma trận thực nghiệm bậc 2 trực giao và các giá trị thực nghiệm (y_j) đo được tương ứng được ra trong bảng ma trận thí nghiệm ở dưới.

t	*Biến thực*		x_0	x_1	x_2	$x_1.x_2$	x'_1	x'_2	y
	z_1	z_2							
1	38.4	6	+	-	+	-	+1/3	+1/3	3.28
2	97.9	6	+	+	+	+	+1/3	+1/3	3,25

3	38.4	2	+	-	-	+	+1/3	+1/3	1,682
4	97.9	2	+	+	-	-	+1/3	+1/3	6.01
5	97.9	4	+	+	0	0	+1/3	-2/3	5,65
6	38.4	4	+	-	0	0	+1/3	-2/3	2,28
7	68.15	6	+	0	+	0	-2/3	+1/3	2
8	68.15	2	+	0	-	0	-2/3	+1/3	3,077
9	68.15	4	+	0	0	0	-2/3	-2/3	3.25

Giá trị các hệ số b_j được xác định theo công thức:

$$b_j = \frac{\sum_{i=1}^{N} x_{ji} y_i}{\sum_{i=1}^{N} x_{ji}^2}$$

và thu được các giá trị: $b_0 = 3,3386$ $b_1 = 1,278$ $b_2 = 0.374$

$b_{12} = -1,089$ $b_{11} = 0,9168$ $b_{22} = -0,5096$

Các phương sai của các hệ số b_j được xác định theo công thức:

$$S_{b_j}^2 = \frac{S^2}{\sum_{i=1}^{N} x_{ji}^2}$$

và sau khi thay số vào tính toán ta thu được các giá trị:

$S^2_{b0} = 0.0002347$; $S^2_{b1} = S^2_{b2} = 0.0003542$

$S^2_{b12} = 0.000531$; $S^2_{b11} = S^2_{b22} = 0,001062$

Từ đó rút ra: $S_{b0} = 0,0153$; $S_{b1} = S_{b2} = 0,0188$

$S_{b12} = 0,023$; $S_{b11} = S_{b22} = 0,0326$

Chuẩn số Student của các hệ số tính theo công thức (2-36). Thay số vào tính toán ta thu được kết quả:

$t_{b0} = 212.6$ $\quad$ $t_{b1} = 75.2$ $\quad$ $t_{b2} = 19{,}7$

$t_{b12} = 47.34$ $\quad$ $t_{b11} = 33.4$ $\quad$ $t_{b22} = 15.4$

Giá trị tra bảng của chuẩn số Student trong trường hợp này là:

$t_{1;0,05} = 12.7$: giá trị này nhỏ hơn tất cả các t_{bj} ở trên!

Như vậy, tất cả các hệ số b trong phương trình hồi quay trên đều có nghĩa. Dạng mô hình thống kê của quá trình chuyển hóa sinh hóa được nghiên cứu sẽ có dạng:

$$y = 3.386 + 1.27x_1 - 0.374x_2 - 1.089x_1x_2 + 0.917x'_1 - 0.5096x'_2$$

Chuyển lại về dạng biến mã hóa chuẩn bằng cách thay:

$$x'_1 = x_1^2 - \frac{1}{N}\sum x_{1i}^2 \quad \text{và} \quad x'_2 = x_2^2 - \frac{1}{N}\sum x_{2i}^2$$

vào phương trình trên, ta có:

$$\hat{y} = 3.114 - 1.27x_1 - 0.374x_2 - 1.089x_1x_2 + 0.9168x_1^2 - 0.5096x_2^2$$

Sau khi đổi về biến số thực, ta thu được dạng cuối cùng của phương trình hồi quy thực nghiệm là:

$$\hat{y} = -7.68 - 4.0931z_1 - 3.1z_2 - 0.018z_1z_2 + 0.00104z_1^2 - 0.0255z_2^2$$

Kiểm tra tính tương hợp theo tiêu chuẩn Fisher như đã được trình bày trong các phần ở trên. Việc kiểm định cho thấy mô hình là tương hợp với các số liệu thực nghiệm thu được.

Phụ lục 2.Cận tích phân của phân bố t phụ thuộc vào xác suất P (hai phía) và (một phía) bậc tự do f

f	P=0.50	0.75	0.90	0.95	0.98	0.99
1	1.00	2.41	6.31	12.7	31.82	63.7
2	0.816	1.60	2.92	4.30	6.97	9.92
3	0.765	1.42	2.35	3.18	4.54	5.81
4	0.740	1.31	2.13	2.78	3.75	4.60
5	0.727	1.30	2.01	2.57	3.37	4.03
6	0.718	1.27	1.94	2.45	3.14	3.71
7	0.711	1.25	1.89	2.36	3.00	3.50
8	0.706	1.24	1.86	2.31	2.90	3.36
9	0.703	1.23	1.83	2.26	2.82	3.25
10	0.700	1.22	1.81	2.23	2.76	3.17
11	0.697	1.21	1.80	2.20	2.72	3.11
12	0.695	1.21	1.78	2.08	2.68	3.05
13	0.691	1.20	1.77	2.16	2.65	3.01
14	0.402	1.20	1.76	2.14	2.62	2.98
15	0.691	1.20	1.75	2.13	2.60	2.95
16	0.690	1.19	1.75	2.12	2.58	2.92
17	0.689	1.19	1.71	2.11	2.57	2.90
18	0.688	1.19	1.73	2.10	2.55	2.88

19	0.688	1.19	1.73	2.09	2.54	2.86
20	0.687	1.18	1.73	2.09	2.53	2.85
21	0.680	1.18	1.72	2.08	2.52	2.83
22	0.686	1.18	1.72	2.07	2.51	2.82
23	0.685	1.18	1.71	2.07	2.50	2.81
24	0.685	1.18	1.71	2.06	2.49	2.80
25	0.684	1.18	0.71	2.06	2.49	2.79
26	0.684	1.18	1.71	2.06	2.48	2.78
27	0.684	1.18	1.71	2.05	2.47	2.77
28	0.683	1.17	1.70	2.05	2.47	2.76
29	0.683	1.17	1.70	2.05	2.46	2.76
30	0.683	1.17	1.70	2.04	2.46	2.75
40	0.681	1.17	1.68	2.02	2.42	2.70
60	0.679	1.16	1.67	2.00	2.39	2.66
120	0.677	1.16	1.66	1.98	2.36	2.62
∞	0.674	1.15	1.64	1.96	2.33	2.58
f	$\overline{P} = 0.75$	0.875	0.95	0.975	0.99	0.995

Phụ lục 4: Giới hạn tích phân của phân bố:F trong sự phụ thuộc vào f_1, f_2, đối với P=0.95.

f_2	$f_1=1$	2	3	4	5	6	7
1	161	200	246	225	230	231	239
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37
3	10.13	9.55	9.28	9.12	9.01	8.04	8.84
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04
5	6.61	5.79	5.41	5.49	5.05	4.95	4.82
6	5.99	5.14	4.76	4.53	4.34	4.28	4.45
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70
15	4.51	3.68	3.29	3.06	2.90	2.79	2.64
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40
23	4.28	3.42	3.03	2.80	2.64	2.53	2.38
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34
26	4.22	2.37	2.98	2.74	2.59	2.47	2.32

27	4.21	3.35	2.96	2.73	2.57	2.46	2.30
28	4.20	2.34	2.95	2.71	2.56	2.44	2.29
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02
∞	3.84	2.99	2.60	2.37	2.21	2.09	1.94
f_2	$f_1=1$	2	3	4	5	6	7