화학공업 발전을 위한 통계의 기여

朴 聖 炫
서울대학교 자연과학대학 계산통계학과

Contribution of Statistics to the Development of Chemical Industry

Sung-Hyun Park

Department of Computer Science and Statistics, College of Natural Sciences
Seoul National University, Seoul 151, Korea

要 約

化學工場이나 硏究所에서 새로운 기술을 개발하거나 생산품의 품질을 관리 및 향상시키는 과정에서 統計의 諸 技法이 어떻게 이용되고 어떠한 寄與를 할 수 있는가에 대하여 說明하였다. 統計의 化學工業的 應用을 크게 두가지로 분류되는데, 하나는 實驗計劃法을 통한 응용이고, 또 하나는 資料分析을 통한 응용이다. 이 두가지 응용방법에 대하여 자세히 기술하고 마지막으로 실제적인 예를 들어서 통계가 화학공장에서 어떻게 응용되는가를 보이고 있다.

Abstract

This article shows what contributions statistics can make for the development of new technology and the quality control and improvement of products in chemical plants or laboratories. The statistical applications in chemical industries can be divided into two parts; one is through experimental designs and the other is through data analyses. These two applications are explained in detail and, lastly, a practical example is illustrated to show the applicability of statistics in a chemical plant.

目 次

1. 序　　論

統計學이 무엇이며 統計가 化學工業의 발전에 어떻게 쓰여질 수 있는가? 위와 같은 질문을 化學工業界에 종사하는 이들에게 할때 정확한 대답을 하는 경우는 흔하지 않다. 筆者도 化學工學科를 졸업하였으나 졸업당시에는 위의 질문에 대하여 거의 아는 바가 없었다. 그후 統計學을 전공하면서 筆者가 가진 큰 관심중에 하나는 統計에서 쓰여지는 諸技法이 化學工業의 발전을 위하여 새로운 기술을 개발하거나 생산품의 품질관리 및 향상을 위하여 어떻게 적용될 수 있는가 하는 문제였다.

이 총설에서는 이 문제에 대하여 그동안 연구집한 내용을 論하려 한다. 지면관계상 全內容을 상세히 다루지는 못하고 몇가지를 중점적으로 다루려고 한다.

먼저 統計學을 간단히 定義해 보면, 統計學은 意思決定科學의 큰 몫을 차지하는 학문으로 사회 및 자연과학의 온갖 현상을 연구하기 위하여 實驗計劃法의 設計, 不確性이 내포된 資料의 抽出, 分析 및 推定을 통하여 의사결정에 필요한 정보의 획득 및 처리하는 방법을 연구하는 학문이라고 말할 수 있을 것이다.

자료의 획득 및 분석을 통한 의사결정과학으로서 통계학의 기여는 인문, 사회, 보건, 경제, 의학 등 광범위한 영역에 걸쳐 공헌하고 있지만 이 총설에서는 化學工業部門에 국한하여 통계가 어떻게 쓰여지고 있는가를 두가지로 나누어 토의하려한다. 첫째로 자료획득을 위한 實驗計劃法을 통한 통계의 응용이고, 둘째로 자료처리의 諸方法을 통한 통계의 응용이다. 위의 두가지 응용을 실제로 보여주기 위하여 자료처리방법중의 하나인 反應曲面分析(Response Surface Analysis)에 대하여 자세히 기술하고 실제 例題를 보일려고 한다.

2. 實驗計劃法(Designs of Experiments)

實驗資料를 통계적으로 分析하여 쓸모있는 情報를 얻기 위해서는 자료획득을 위한 合理的인 방안이 모색되어야 하며 이에 준하여 자료가 획득되어야 한다. 자료획득을 위한 방안이 통계적으로 아무런 合理性이 없이 무작정으로 얻어진 자료라면 그 자료는 가치가 없다. 경제적인 면을 고려하여 적은 양의 자료로부터 많은 情報를 얻을려면 자료획득을 위한 올바른 實驗計劃法이 수립되고 이에 따라서 자료의 획득 및 분석이 이루어 져야 한다. 분석방법도 실험계획에 따라 달라지기 때문에 처음부터 어떠한 목적을 위하여 자료를 구할려는가를 뚜렷이 설정하고 분석방법을 구상한 후에 이들을 만족시킬 수 있는 실험계획법을 設計하여야 한다.

특별히 다른 어느 분야보다도 화학공업실험을 통한 자료의 획득은 일반적으로 實驗裝置가 방대하고 복잡하기 때문에 실험비용이 많이 드는 관계로 최소의 실험으로 최대의 효과를 내는 실험계획법의 설계는 매우 중요하다.

실험계획법을 두가지 형태로 분류할 수 있는데 한가지는 實驗單位(Experimental unit)의 배치에 따른 분류이고 다른하나는 處理組合(Treatment combination)의 선택에 따른 분류이다. 이들 어느 경우에도 統計的 分析方法이 이들과 같이 論議되어야 하는 것은 물론이다.

배치에 따른 분류로는 흔히 쓰이는 것이 完全任意配例法(Completely random design), 亂塊法(Randomized complete block design), 라틴方格法(Latin square design), 不完備블럭法(Incomplete block design) 등을 들을 수 있으며, 처리조합에 의한 분류로는 要因實驗(Factorial experiments)이나 部分要因實驗(Fractional factorial experiments)을 주로 들 수 있다.

예를들어 어떤 품목을 생산하는데 두개의 상이한 생산과정이 개발되었다고 하고 이 두개의 생산과정의 우열비교를 실험을 통하여 검토한다고 하자. 만약 두개의 상이한 환경조건에서 각 생산과정에 대하여 세가지 反復實驗을 한다면 이것은 要因實驗의 입장에서도 각要因의 水準이 2인 二要因實驗이 될 것이다. 좀더 자세히 설명하기 위해서 첫번째 생산과정을 要因 A로 표시하고 이 과정을 실험하기 위한 두개의 환경조건

을 각각 a_0와 a_1으로 표시하자. 두번째 생산과정에 해당하는 조건을 b_0와 b_1으로 표시하면 결국 4가지의 處理組合(a_0b_0, a_0b_1, a_1b_0, a_1b_1)의 表現을 얻을 수 있으며 이와 같은 要因實驗을 2^2 factorial design이라고 부른다. 각 處理組合에 대하여 세가지 反復실험한 것을 表로 만들어 보면 表 1과 같다. 表 1에서 O_{ijk}는 要因 A의 i條件과 要因 B의 j條件의 組合에 대해서 k번째의 反復實驗에서 나온 觀察值를 나타낸다.

表 1. 2^2 Factorial Design

		要因 A	
		條件 a_0	條件 a_1
要因 B	條件 b_0	O_{001} O_{102} O_{003}	O_{101} O_{102} O_{103}
	條件 b_1	O_{011} O_{012} O_{013}	O_{111} O_{112} O_{113}

그러나 위의 실험을 처리조합에 따른 분류가 아니라 배치에 따른 분류로 보고 실험단위에 중점을 두어서 생각하면 4개의 處理方法에 대하여 3反復實驗을 한 完全任意配例法으로 처리할 수 있다. 만약 3反復實驗을 세개의 블럭으로 간주하여 亂塊法으로 처리한다면 表 2와 같은 경우가 될 것이다.

表 2. 亂 塊 法

		處 理 方 法			
		a_0b_0	a_0b_1	a_1b_0	a_1b_1
블 럭	1	O_{001}	O_{011}	O_{101}	O_{111}
	2	O_{002}	O_{012}	O_{102}	O_{112}
	3	O_{003}	O_{013}	O_{103}	O_{113}

물론 어떤 실험계획법을 택할 것인가 하는 것은 소기의 실험목적을 달성하려는 분석방법등을 고려한 후에 정해져야 할 것이다.

위에서 열거된 실험계획법들은 오랜시일에 걸쳐 널리 애용되었으며, 공업뿐 아니라 농업, 보건, 생물, 사회문제 등에 적용되어 왔다. 이에 대한 자세한 설명은 대부분의 통계 교과서에서 찾아볼 수 있으므로 생략하고 최근에 美國을 중심으로 化學反應實驗에 각광을 받고 있는 몇가지 실험법을 자세히 소개하려 한다. 이들은 中心合成計劃(Central composite design), 回轉計劃(Rotatable design), 回轉中心合成計劃(Rotatable central composite design) 등인데 이 계획들을 反應曲面計劃(Response surface design)이라고 통칭하기도 한다. 먼저 反應曲面의 概念을 설명하고 위의 實驗計劃法들을 論議 하려고 한다.

3. 反應曲面概念과 그 實驗計劃法 (Concepts of response surfaces and its experimental designs)

어떤 化工反應에 있어서 몇 개의 독립 변수가 복합적인 작용을 함으로서 어떤 반응을 일으키고 독립변수들의 값이 변함에 따라서 반응량이 달라진다고 하면 이 반응량들의 변화하는 자취를 연결하여 이루어지는 曲面을 우리는 반응곡면이라고 한다. 이때에 반응량을 나타내는 변수를 종속변수라고 한다.

예를들면 어떤 화학반응에 있어서 화학반응량이 온도, 습도 및 압력의 변화에 따라서 달라진다고 하자. 그러면 온도, 습도 및 압력이 독립변수가 되고 이들의 변화에 따라서 영향을 받는 반응량은 종속변수가 될 것이다.

反應曲面을 推定(estimation)함으로서 우리가 통계적으로 얻을 수 있는 것중 중요한 것을 적어보면 다음과 같다.

(1) 독립변수와 종속변수간의 함수관계를 규명하여 독립변수 값의 변화에 따라서 반응량(종속변수)이 어떻게 달라지는 가를 예측할 수 있다.

(2) 독립변수가 어떠한 값을 취할 때 반응량이 最大化(maximize) 또는 最小化(minimize) 하는가를 알 수 있다.

(3) 독립변수와 종속변수간의 함수관계를 규명하고자 할 때에 어떠한 實驗計劃法이 좋은

精度(precision)를 줄 것인가를 알 수 있다.

위의 세가지를 성취할 수 있다면, 이는 化學反應을 연구하는데 있어서 統計學이 큰 寄與를 할 수 있다는 좋은 본보기가 될 것이다. 또한 化學工業生産品의 品質管理면에서도 유용하게 응용될 수 있다.

어떤 化學反應이 k개 $(\xi_1, \xi_2, \cdots, \xi_k)$의 독립변수로 이루어져 있고, 이들의 변화에 따라서 달라지는 종속변수를 η라고 하자. 그러면 그 반응관계를 함수 f로 표시하면

$$\eta = f(\xi_1, \xi_2, \cdots\cdots, \xi_k)$$

로 쓰여질 것이다. 위의 함수 f를 統計的으로 推定할 때에 널리 쓰이는 假定은 이 함수 f가 그려주는 반응량 η의 曲面이 독립변수들의 多項回歸曲面(polynomial regression surface)으로 近似하게 표현될 수 있다는 것이다.

위의 가정은 독립변수의 모든 가능한 값에 대하여 만족될 필요는 없으며, 단지 實驗者(Experimenter)가 관심을 가지고 있는 독립변수의 어떤 領域(Regicn)에서만 만족하면 될 것이다. 독립변수 ξ를 선형변환(Linear transform)시켜서 새로이 정의된 독립변수를 x라고 하고 관심영역(Region of interest)의 중앙이 $(x_1, x_2, \cdots\cdots, x_k) = (0, 0, \cdots\cdots, 0)$라고 하면 편리하다. 관심영역안에서 반응량 η는 第二次多項모델(Second order, polynomial regression model)을 사용하면 실질적으로 좋은 精度를 가진 曲面을 抽定할 수 있는데 그 표현은

$$\eta = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i \leq j}^{k} \beta_{ij} x_i x_j \qquad (1)$$

으로 쓰여진다. 예로서 독립변수가 세개($k=3$)이면

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x\ x_3$$

라고 표현될 수 있을 것이다. 위에서 β_i나 β_{ij}는 미지수인 母數(unknown parameter)이며 이들은 실험을 통하여 얻어지는 資料로부터 推定(Estimate) 되어져야 한다.

이제 식 (1)을 推定하기 위한 資料를 획득하기 위한 實驗計劃法을 알아보자.

3-1. 中心合成計劃(Central composite design)

이 총설에서는 자세히 논의 되지는 않았지만 工業實驗을 위한 計劃중에서 널리 쓰이는 것은 2^k要因實驗計劃(2^k factorial design)일 것이다. 그러나 이 실험계획은 식 (1)의 β_{ij}를 推定하기 힘들다는 단점이 있다. 이를 보완해서 2^k要因實驗에 다음과 같은 實驗點을 첨가시킨 것을 中心合成計劃이라 한다.

$$\begin{array}{ccc} x_1 & x_2 \cdots\cdots\cdots & x_k \end{array}$$
$$\begin{pmatrix} 0 & 0 \cdots\cdots\cdots & 0 \\ -\alpha & 0 \cdots\cdots\cdots & 0 \\ \alpha & 0 \cdots\cdots\cdots & 0 \\ 0 & -\alpha \cdots\cdots\cdots & 0 \\ 0 & \alpha \cdots\cdots\cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 \cdots\cdots\cdots & -\alpha \\ 0 & 0 \cdots\cdots\cdots & \alpha \end{pmatrix}$$

여기에서 중점 $(x_1, x_2, \cdots\cdots, x_k) = (0, 0, \cdots\cdots, 0)$에는 한개 이상의 觀察點을 가질 수 있다. 그리고 α의 값은 실험자가 임의로 선택할 수 있는 값이다.

예를들어서 만약 독립변수가 세개($k=3$)이면 計劃行列(Design matrix)은 다음과 같다. 中心點은 한개라고 가정하기로 하자.

$$\begin{array}{cccc} & x_1 & x_2 & x_3 \\ D= & \multicolumn{3}{c}{\begin{pmatrix} -1 & -1 & -1 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \\ -1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \\ -\alpha & 0 & 0 \\ \alpha & 0 & 0 \\ 0 & -\alpha & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & -\alpha \\ 0 & 0 & \alpha \end{pmatrix}} \end{array}$$

처음의 8개의 點은 2^3 要因實驗計劃에서 얻어지는 것이며 9번째는 中心點이며 나머지 6개의 點은 軸點(Axial points)으로 中心合成計劃이 가지는 특유의 점들이다. 이 計劃을 도표로 그려보면 도표 1.이 된다. 만약 軸點들이 中心點으로부터의 거리를 나타내는 α가 1이 아니라면 각 독립변수는 5개의 다른 水準, $(-\alpha, -1, 0, 1, \alpha)$에서 관찰된 결과가 된다. 그러므로 식 (1)의 제이차곡면을 最小自乘法(Least squares method)을 사용하여 統計的으로 推定하기에는 충분하다.

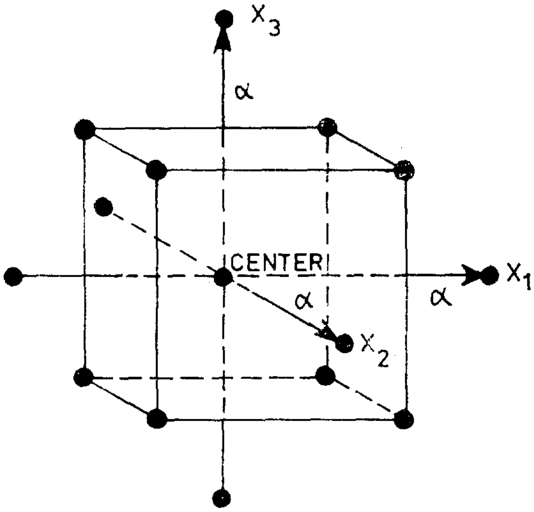

도표 1. 中心合成計劃($k=3$)

만약 $k>2$이면 3^k 要劃計劃法보다 훨씬 적은 량의 관찰점을 이 計劃은 가지면서 각 독립변수는 더많은 水準에서 관찰되는 利點을 가지고 있다. 이는 표 3.을 보면 명백하다.

軸點의 거리 α를 선택하는데는 여러가지 방법이 알려져 있으나 흔히 쓰이는 것중에 하나는 統計理論에서 많이 거론되는 모우먼트行列(Moment matrix) XX'를 diagonalize 시키는 α의 값이다. 이러한 α의 값을 사용하면 抽定되는 β_i와 β_{ij} 사이의 共分散(Covariance)이 전부 영이 되며 이 計劃을 直交中心合成計劃(Orthogonal central composite design)이라고 한다. 直交計劃이 되게하는 α의 값은 k와 중심점의 수의 함수가 되며 이를 표 4.에 제시하여 놓았다. 재미있는 발견은 만약 $k=2$이고 중심점이 하나이면 이 直交中心合成計劃은 3^2 要因實驗計劃과 동일한 것이 된다.

표 3. 中心合成計劃과 3^k 要因計劃의 비교

	中心合成計劃	3^k 要因計劃
水 準	$(-\alpha, -1, 0, 1, \alpha)$	$(-1, 0, 1)$
全觀察點의 수	2^k+2k+(중심점의 수)	3^k

표 4. 直交中心合成計劃을 위한 α의 값

k	중심점의 수			
	1	2	3	4
2	1.000	1.078	1.147	1.210
3	1.216	1.287	1.353	1.414
4	1.414	1.482	1.547	1.607
5	1.596	1.662	1.724	1.784
6	1.761	1.824	1.885	1.943

이 計劃에 대하여 더많은 理論的인 배경을 알기 원하는 독자는 Myers[3]나 Box and Wilson[2]을 참고로 하면 좋을 것이다.

3-2. 回轉計劃(Rotatable design)

앞에서 論한 中心合成計劃과 비슷하게 反應曲面實驗에서 많이 쓰여지는 것은 어쩌면 回轉計劃(Rotatable design)일 것이다. 回轉(Rotatability)의 概念을 설명하여 보자. 식 (1)을 얻어진 資料로부터 抽定하여 適合된 反應曲面(Fitted response surface)을 구하면 다음과 같이 쓸 수 있다.

$$\hat{y}=b_0+\sum_{i=1}^{k} b_i x_i+\sum_{i\le j}^{k} b_{ij}x_i x_j \tag{2}$$

위에서 $\hat{y}$는 η의 推定量(Estimator)이고 b_0, b_i와 b_{ij}는 식 (1)의 β_0, β_i와 β_{ij}의 最小自乘法에 의한 抽定量이다. 식 (2)를 간편하게 벡타로 표시하면

$$\hat{y}=\underline{x}'\underline{b} \tag{3}$$

이 되며 $\underline{x}'=(1, x_1, x_2, \cdots, x_k, x_1^2, x_1x_2, \cdots, x_{k-1}x_k)$이며 $\underline{b}$는 $\underline{x}'$에 상응되는 b_i, b_{ij} 등의 벡타이다.

일반 回歸分析(Regression analysis)에서 알려

진 바와 같이 抽定된 반응 $\hat{y}$의 $\underline{x}$에서의 分散은

$$\mathrm{Var}[\hat{y}]=\underline{x}'(X'X)^{-1}\underline{x}\sigma^2$$

이며 말할 필요도 없이 $\hat{y}$의 分散은 $\underline{x}$의 함수이다. 여기에서 만약 $\hat{y}$의 分散이 中心點(0, 0, …, 0)으로부터의 거리만의 함수이고 $\underline{x}$의 방향의 함수가 아니라면 이와 같은 결과를 주는 計劃을 回轉計劃이라고 부른다. 이 計劃에 대한 좀더 자세한 이론적인 배경을 알고자 희망하는 독자는 Box and Hunter[1]나 筆者가 쓴 Park[4]을 보면 좋을 것이다.

실제로 많이 쓰이는 回轉計劃의 예를 들어보자. 독립변수가 두개 있을 경우에 5개 또는 그 이상의 計劃點(Design point)을 거리반경이 9인 원상에 균등하게 배치하고 中心點에 1개이상의 계획점을 배치하면 回轉計劃이 된다. 원상에 5개의 點을 두면 五角形實驗計劃이라 부르고 6개의 點을 두면 六角形實驗計劃이라 한다. 실제로 五角形에서 八角形까지의 計劃은 널리 쓰여지고 있다.

이들 계획의 長點은 관심영역 전반에 걸쳐서 반응량 η를 추정하는 $\hat{y}$이 균등한 精度를 가진다는 점이다.

3-3. 回轉中心合成計劃(Rotatable central composite design)

위에서 論議된 中心合成計劃과 回轉計劃은 각기 특유한 장점을 가지고 美國을 중심으로 先進國들에서 工業技術의 개발 및 품질관리 등에 흔히 이용되고 있다. 그런데 재미있는 사실은 中心合成計劃이 어떤 특정한 α의 값에 대해서는 回轉性을 갖는다는 것이다. 中心合成計劃은 2^k 要因計劃部分과 中心點, 그리고 軸點들로 이루어져 있는데 軸點의 거리 α가 다음식을 만족시키면 回轉性을 가지게 된다.

$$\alpha=(2^k)^{\frac{1}{4}} \tag{4}$$

이 이론적인 배경은 Park[4]을 보면 설명되어 있다. 예를 들어서 $k=2$이면 $\alpha=1,414$로 하면 이 中心合成計劃은 回轉性을 갖게되며 이런 計劃을 도표로 그려보면 도표 2.가 된다. 여기에서 計劃點은 별표로 표시하였다. 中心點에는 한개 이상의 點이면 되며 하나의 위상에 8개의 點이 균등하게 배치되어 있으므로 八角形計劃이며 이는 물론 回轉計劃인 동시에 中心合成計劃이다.

위에서 열거된 計劃이외에도 여러가지 計劃이 소개되었으나 여기서는 논하지 않겠다.

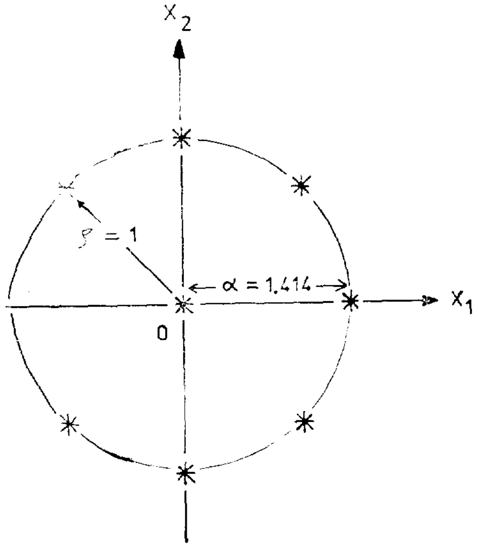

도표 2. 回轉中心合成計劃($k=2$)

4. 資料分析方法

앞에서는 주로 實驗計劃法에 대해서 기술하였는데 여기에서 얻어진 資料를 소기의 목적에 맞게 정확한 分析方法을 사용하지 않으면 애써서 얻은 실험자료에서 有用한 情報를 획득할 수 없다. 실험에서 얻은 자료뿐만 아니라 실제로 工場에 가보면 적은 양이든 많은 양이든 간에 여러종류의 자료를 계속해서 획득하고 있다. 특히 생산과정중에서 품질관리, 안전관리 등에 관한 자료와 화학반응여건 등에 관한 자료들이다. 이런 자료들은 대부분 무시할 수 없는 誤差(Error)를 내포하고 있으며 또 눈에 잘 띄지 않는 여러 要因들의 함수관계를 가지고 있으므로 이들 자료가 줄 수 있는 최대의 情報를 얻기 위해서 올바른 자료처리방법을 선택하는데 세심한 주의가 필요하다. 실제로 비용을 많이 들여 자료를 얻어놓고 올바른 분석방법을 몰라 유용한 자료를 무용하게 방치시키는 예가 허다하다. 다음에서

보통 많이 쓰이는 몇가지 統計的 技法을 간략하게 설명하고자 한다.

(1) 記述的인 方法

가장 기초적인 것으로 주어진 자료의 度數分布表(Frefuency table), 柱狀圖表(Histogram), 평균치 등을 구하거나 표준편차 등을 계산하는 방법이다. 간단명료하게 자료를 분류할 수 있으면 간혹 새로운 사실을 발견할 때가 있다.

(2) 分散分析 및 假說檢定

誤差의 소재를 밝혀내고 변수들이 주는 영향을 알아내기 위해서는 分散分析(Analysis of variance)이 필요하며, 어떤 주어진 假定이 옳은 것인지를 자료로부터 판정하기 위해서는 假說檢定(Hypothesis testing) 方法이 쓰여진다. 새로운 공업기술을 개발할 때에 아직까지 써오던 기술보다 우수한가 아닌가 하는 것은 철저히 檢定해 볼 가치가 있다.

(3) 回歸分析 및 相關分析

여러 변수간의 상관관계 및 반응관계를 규명하기 위해서 回歸 및 相關分析(Regression and Correlation analysis)은 매우 적절한 방법이다. 이 분석방법은 널리 사용되고 있는 기법으로 여러 要因이 상호작용하여 어떤 함수관계를 가질 때에 이를 효과적으로 측정하는 방법을 제시해 준다. 만약 실험을 통하여 통계적으로 이 함수관계를 규명한다면 품질의 관리 및 향상을 위하여 유익한 자료를 얻을 수 있을 것이다.

(4) 管理圖

제품을 생산하고 있는 과정에서 제품의 품질을 계속적으로 검토하기 위하여 여러종류의 품질관리도(Quality control chart)를 사용하고 있는데 예로서 Shewhait 관리도나 Cumulative sum 관리도 같은 것이다. 이 도표들은 제품의 생산과정이 안정상태에 있는가 없는가를 보여주기 때문에 유용하게 이용된다.

(5) 標本抽出法

標本(Sample)을 사용하여 자료를 추출할 때에 어떻게 표본을 선택할 것인가에 대하여 세심한 주의를 하지 않으면 이 표본에 의한 분석이 精度가 좋지 못하고 가치가 없을 수 있다. 따라서 유용한 통계적 분석을 얻고 경제적인 抽出(Sampling)을 하기 위하여 올바른 표본추출법을 사용하여야 한다.

(6) 統計的 씨뮤레이숀

새로운 기술을 개발하거나 또는 당면한 어떤 문제를 풀기 위하여 工場의 全 System을 상대로 실험하는 것이 시간과 경비관계로 힘들 경우가 많다. 이럴 때에 실제의 System을 흉내해서 수치적으로 가상적인 모델을 만들고 관련된 변수들의 통계적인 분포에 따라서 亂數(Random number)를 발생시켜 이 모델이 어떤 결과를 주는가를 관찰하므로서 실제 System의 결과를 가상적으로 분석하는 방법이 統計的 씨뮤레숀이다. 최근에 이용되기 시작하는 방법으로 거의 電子計算機(Electronic computer)가 필수적으로 이용되고 있다.

위에서 일반적으로 쓰이는 統計的 分析方法을 열거하였는데 이와 별도로 앞에서 論議된 反應曲面을 分析하기 위해서는 特有의 技法이 필요하다. 이제 이에 대하여 고찰하여 보자.

5. 反應曲面分析(Response surface analysis)

만약 종속변수인 식 (1)의 η가 우리가 관심을 가지고 있는 어떤 化工反應量이라 한다면 이를 最大 또는 最小化시키는 독립변수들의 反應條件에 우리는 흥미가 있을 것이다. 여기에서는 식 (2)와 같은 反應曲面이 資料의 統計的인 처리방법을 통해서 결정되었을 때에 어떻게 하면 독립변수들의 最適反應條件을 決定하고 反應曲面의 성질을 이해할 수 있는가에 대하여 論하려고 한다.

식 (2)를 간략하게 行例을 사용하여 표현하면

$$\hat{y}=b_0+x'b+x'Bx \tag{5}$$

라고 쓰여지는데 여기에서 각각

$$x=\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_k \end{pmatrix} \quad b=\begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_k \end{pmatrix} \quad B=\begin{pmatrix} b_{11} & b_{12}/2 & b_{13}/2 \cdots b_{1k}/2 \\ & b_{22} & b_{23}/2 \cdots b_{2k}/2 \\ & & b_{33} \ \cdots b_{3k}/2 \\ (\mathrm{Sym}) & & \ddots \quad \vdots \\ & & b_{kk} \end{pmatrix}$$

을 나타낸다. $\hat{y}$을 最大 또는 最小化하는 x의 값을 구하기 위하여 $\hat{y}$을 x로 편미분하면

$$\frac{\partial \hat{y}}{\partial x}=\frac{\partial}{\partial x}[b_0+x'b+x'Bx]$$
$$=b+2Bx \tag{6}$$

만약 어떤 點 x_0이 $\hat{y}$을 最適化한다면 이 點에서의 편미분은 영이 되므로

$$b+2Bx_0=0$$

이 되고 이 x_0을 定常點(Stationary point)이라 부르며 식 (6)으로부터

$$x_0=-B^{-1}b/2 \tag{7}$$

가 얻어진다. 이 定常點은 適合된 反應曲面上에서 다음의 세가지 중에 하나가 될 것이다.

(1) $\hat{y}$이 最大値를 얻게되는 x의 點.

(2) $\hat{y}$이 最小値를 얻게되는 x의 點.

(3) $\hat{y}$이 最大도 아니고 最小도 아닌 x의 鞍部點(Saddle point).

위의 세가지 경우에 定常點을 x들의 軸으로 된 좌표에 위치시키고 y들이 같은 값을 갖는 x들의 자취를 연결시키면 等高線表(Contour chart)를 얻을 수 있다. 이 등고선표는 反應曲面을 이해하는데 매우 유용하다. 이제 反應曲面을 좀더 定量的으로 이해하기 위하여 正準分析(Canonical analysis)에 대하여 연구하여 보자.

定常點 x_0에서의 $\hat{y}$의 값을 $\hat{y}_0$이라고 하자. 그러면 식 (5)로부터

$$\hat{y}_0=b_0+x_0'b+x_0'Bx_0$$

이 되고 좌표의 原點을 x_0으로 옮기기 위하여

$$z=x-x_0$$

이라고 하면 식 (5)로부터

$$\hat{y}=b_0+(z'+x_0')b+(z'+x_0')B(z+x_0)$$
$$=(b_0+x_0'b+x_0'Bx_0)+z'(b+2Bx_0)+z'Bz$$
$$=\hat{y}_0+z'Bz \tag{8}$$

이 된다. 식 (8)의 $z'Bz$에 대하여 어떤 直交變換(Orthogonal transformation)

$$z=Mw \tag{9}$$

을 하면 식 (8)은

$$\hat{y}=\hat{y}_0+z'Bt$$
$$=\hat{y}_0+w'M'BMw$$
$$=\hat{y}_0+\lambda_1 w_1^2+\lambda_2 w_2^2+\cdots\cdots+\lambda_k w_k^2 \tag{10}$$

으로 표시되며 여기에서 $\lambda_1, \lambda_2, \cdots, \lambda_k$는 行例 B의 特性根(Characteristic root)이고 M은 $k\times k$ 直交行例로서 $M'M=MM'=I_k$이다. 行例 M은 어떤 경우에는 중요한 의미를 갖는데 왜냐하면 식 (9)에 M'을 곱하면

$$w=M'z \tag{11}$$

가 되는데 이 관계식은 새로운 독립변수 w가 어떻게 z와 관련되어 있는가를 설명해 주고 있기 때문이다. 이 行例 M은 k개의 벡타로 구성되어 있고

$$M=(M_1M_2\cdots\cdots M_k)$$

M_i를 구하기 위하여 다음의 두개의 방정식을 풀면된다.

$$(B-\lambda_i I_k)M_i=0 \tag{12}$$
$$M_i'M_i=1$$

식 (10)이 식 (5)의 표현보다 반응곡면의 성질을 이해하기 쉬운 점은 다음에 있다. 만약 λ_i들이 모두 陽이면 定常點 x_0는 最小點이 되며 반응량 $\hat{y}$이 w_i의 방향으로 어느정도 급격히 변화하는가 하는 정도를 알 수 있으며 λ_i들이 모두 陰이면 x_0는 最大點이라는 것을 알 수 있다. 물론 일부의 λ_i는 陽이고 일부는 陰이면 x_0는 鞍部點이 된다.

다음에 실제 예를 들어서 위의 反應曲面分析을 고찰하여 보자.

6. 例 題

어떤 工場에서 생산되고 있는 化工製品이 마지막 생산과정에서 두개의 독립변수(하나는 압력, 다른하나는 온도)에 의하여 그 제품의 强度(grams/inch)가 영향을 받는다고 한다. 이 실험은 어떤 압력과 온도를 주는 것이 强度를 세게할 것인가, 또 압력과 온도의 변화에 따라서 제품의 强度에 어떤 변화를 가져오는가를 알아보

기 위해서 다음의 실험자료를 얻었다. 이 실험에서 변화된 독립변수의 값은 다음의 공식에 의하여 얻어졌다.

$$x_1=\frac{\text{압력 (PSI)}-30}{20}$$
$$x_2=\frac{\text{온도}(^\circ\text{C})-205}{10} \qquad (13)$$

표 5.에 실려있는 자료는 回轉計劃의 하나인 六角形實驗計劃法(Hexagonal experimental design)에 의하여 얻어진 것이며 이를 그림으로 그려보면 도표 3.과 같다. 여기에서 별표는 資料

표 5. 例題의 資料

관측번호	변화된 독립변수 x_1	x_2	실제의 독립변수 x_1	x_2	실제의 반응량 y	추정된 반응량 y
1	-1	0	10	205	54.6	56.3
2	$-\frac{1}{2}$	$\sqrt{0.75}$	20	213.7	71.7	70.0
3	$-\frac{1}{2}$	$-\sqrt{0.75}$	20	196.3	59.8	58.1
4	$\frac{1}{2}$	$\sqrt{0.75}$	40	213.7	78.7	80.4
5	$\frac{1}{2}$	$-\sqrt{0.75}$	40	196.7	78.9	80.6
6	1	0	50	205	91.0	89.3
7	0	0	30	205	90.4	89.3
8	0	0	30	205	87.5	89.3
9	0	0	30	205	91.0	89.3
10	0	0	30	205	88.3	89.3

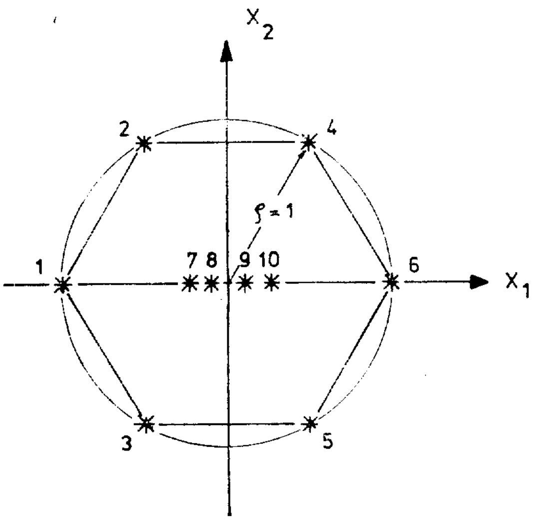

도표 3. 六角形實驗計劃法

點(Data point)을 표시한다.

위의 실험에서 변환된 독립변수 x_1과 x_2의 원점을 중심으로한 부분이 관심영역이라 하고 이 영역근처에서 第二次回歸 모델이 반응관계를 나타내는데 적합하다고 생각하면 다음의 모델을 고려하게 된다.

$$\eta=\beta_0+\beta_0x_1+\beta_0x_2+\beta_{11}x_1^2+\beta_{22}x_2^2+\beta_{12}x_1x_2$$

이 모델을 最小自乘法에 의하여 抽定하면 計劃行列 X와 反應벡타 Y는 다음과 같으므로

$$X=\begin{array}{c} \begin{array}{cccccc} & x_1 & x_2 & x_1^2 & x_2^2 & x_1x_2 \end{array} \\ \begin{pmatrix} 1 & -1 & 0 & 1 & 0 & 0 \\ 1 & -\frac{1}{2} & \sqrt{0.75} & \frac{1}{4} & 0.75 & -\frac{1}{2}\sqrt{0.75} \\ 1 & -\frac{1}{2} & -\sqrt{0.75} & \frac{1}{4} & 0.75 & \frac{1}{2}\sqrt{0.75} \\ 1 & \frac{1}{2} & \sqrt{0.75} & \frac{1}{4} & 0.75 & \frac{1}{2}\sqrt{0.75} \\ 1 & \frac{1}{2} & -\sqrt{0.75} & \frac{1}{4} & 0.75 & -\frac{1}{2}\sqrt{0.75} \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

$$Y=\begin{pmatrix} 54.6 \\ 71.7 \\ 59.8 \\ 78.7 \\ 78.9 \\ 91.0 \\ 90.4 \\ 87.5 \\ 91.0 \\ 88.3 \end{pmatrix}$$

이 모델의 係數는 다음에 의하여 抽定된다.

$$\begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_{11} \\ b_{22} \\ b_{12} \end{pmatrix}=(X'X)^{-1}X'Y=\begin{pmatrix} 89.30 \\ 16.48 \\ 3.38 \\ -16.50 \\ -17.20 \\ -6.99 \end{pmatrix}$$

그럼으로 反應量(强度)의 抽定은

$$\hat{y}=89.30+16.48x_1+3.38x_2$$

$$-16.50x_1^2-17.20x_2^2-6.99x_1x_2$$

식 (6)에 있는 벡타 b와 行列 B는

$$b=\begin{pmatrix}16.48\\3.28\end{pmatrix}\qquad B=\begin{pmatrix}-16.50 & -3.49\\-3.49 & -17.20\end{pmatrix}$$

이고 식 (7)의 定常點 x_0은

$$x_0=-B^{-1}b/2=\begin{pmatrix}0.5002\\-0.0034\end{pmatrix}$$

이다. 正準分析을 하기 위하여 行列 B의 特性根을 구하면 $\lambda_1=-13.34$이고 $\lambda_2=-20.36$이며 x_0에서 $\hat{y}$의 값은 $\hat{y}_0=93.42$이므로 식 (10)의 표현은 다음과 같다.

$$\hat{y}=93.42-13.34w_1^2-20.36w_2^2 \tag{14}$$

그럼으로 우리는 定常點 x_0가 最大點이라는 것을 알 수 있으며 새로이 얻어진 독립변수 w_1과 w_2가 변함에 따라서 y의 값이 감소되는 것은 알 수 있다. 식 (13)으로부터 강도를 최대로 하는 실제 독립변수의 값은

$$0.5002=\frac{\text{압력}-30}{20}$$

$$-0.0034=\frac{\text{온도}-205}{10}$$

으로부터 압력은 약 40 PSI이고 온도는 약 205°C이며 이 最適條件에서 强度는 93.42 grams/inch까지 도달할 수 있다.

식 (14)에 있는 w_1과 w_2와 그리고 x_1과 x_2 사이의 관계를 규명하기 위하여 식 (12)을 이용하여 行例 M을 구해 보자. M_1과 M_2를 각각

$$M_1=\begin{pmatrix}m_{11}\\m_{21}\end{pmatrix}\qquad M_2=\begin{pmatrix}m_{12}\\m_{22}\end{pmatrix}$$

라 놓고 식 (12)을 사용하여 구하면 변환 행렬 M은

$$M=(M_1M_2)=\begin{pmatrix}0.742 & 0.671\\-0.671 & 0.742\end{pmatrix}$$

이고 식 (11)로부터 w와 x와의 관계는

$$w=M'(x-x_0)$$

이므로

$$\begin{pmatrix}w_1\\w_2\end{pmatrix}=\begin{pmatrix}0.742 & -0.671\\0.671 & 0.742\end{pmatrix}\begin{pmatrix}x_1-0.5002\\x_2+0.0034\end{pmatrix}$$

이 된다. 이제 이들 변수의 관계를 $\hat{y}$의 等高線表(Contour chart)와 같이 그려보면 도표 4.를 얻을 수 있다.

도표 4.를 관찰해 보면 w의 중심점 $(w_1, w_2)=(0,0)$에서 w가 움직임에 따라 $\hat{y}$의 값이 감소되는 것을 알 수 있고 그 감소의 정도는 w_2 쪽이 더 빠르다는 것을 볼 수 있다.

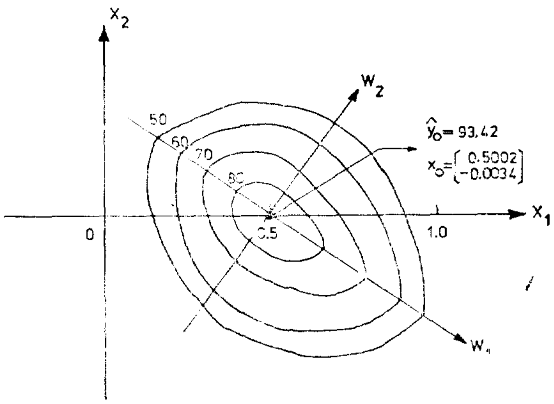

도표 4. x와 w의 관계 및 $\hat{y}$의 등고선표

7. 結 論

아직까지 化學工場 및 硏究所를 중심으로 자료처리를 위한 여러가지 分析方法과 자료의 효율적인 획득을 위한 實驗計劃法에 관하여 論하였다. 특히 자세히 설명된 것은 反應曲面實驗計劃法과 그의 分析이며 실제 예를 들어서 그 응용면을 설명하였다. 실제로 자료의 획득과 분석을 통한 統計의 化學工業的 응용과 그의 역할은 매우 크다고 하겠다. 품질의 엄격한 통계적관리를 통한 질적향상 및 생산과정중에 관련된 여러 변수간에 함수관계를 통계적으로 推定하여 반응의 예측, 最適反應條件의 발견, 그리고 공정기술의 개량이 가능할 것이다. 기타 수없이 많은 응용범위가 있으며 일단 자료를 획득, 처리, 분석하는 업무는 통계업무로 간주되어야 할 것이다.

마지막으로 한가지 첨부하고 싶은 것은 과학과 기술이 급성장함에 따라서 자연히 대량정보의 처리와 분석이 불가피해졌는데, 이는 대부분의 통계적 자료분석이 전자계산기의 도움이 필요함을 의미한다. 이처럼 통계업무처리와 전자계산기를 효율적으로 사용하는 기술은 분리하여 생각할 수 없게끔 되어가고 있다.

바람직스러운 일은 각종의 화학공업 기업체에서 품질의 관리 및 향상, 기술개발 등을 원활히 하기 위하여 필요한 자료의 획득계획, 분석 등을 효율적으로 할 수 있는 統計人을 가지고 있어야 하며 성격상으로 電算업무와 밀접한 관계를 유지하여야 할 것이다.

참고 문헌

1. G.E.P. Box and J.S. Hunter, "Multifactor Experimental Designs for Exploring Response Surfaces," *Annals of Mathematical Statistics*, **28**(1957), 195.
2. G.E.P. Box and K.B. Wilson, "On the Experimental Attainment of Optimum Conditions" *Journals of Royal Statistical Society*, **B13** (1951), 1.
3. R.H. Myers, Response surface Methodology Allyn and Bacon, Inc. Boston(1971).
4. S.H. Park, "反應曲面의 分析 및 實驗計劃法과 그 應用" 統計 **3**(1977), 112.

著者 紹介

著者 朴聖炫 博士는 68년 서울大 화공과 졸업후 1972년에 North Carolina 주립대학에서 석사과정을 마치고 이어 1974년 통계학을 전공하여 박사학위를 획득하였다.

1975년부터 2년간 Mississippi 주립대학 Business school의 조교수로 근무하다가 1977년 귀국하여 현재 서울대 자연대학 계산통계학과 조교수로 근무중이며 겸하여 서울대 전자계산소 부소장을 맡고 있다.