

혼합물 반응표면분석에서 디자인 스페이스 구축을 위한 효율적인 실험계획

정중희 · 임용빈[†]

이화여자대학교 통계학과

Efficient Designs to Develop a Design Space in Mixture Response Surface Analysis

Jong Hee Chung · Yong B. Lim[†]

Department of Statistics, Ewha Womans University

ABSTRACT

Purpose: The practical design for experiments with mixtures of q components is consisted in the four types of design points, vertex, center of edge, axial, and center points in a $(q-1)$ -dimensional simplex space. We propose a sequential method for the successful construction of the design space in Quality by Design (QbD) by allowing the different number of replicates at the four types of design points in the practical design when the quadratic canonical polynomial model is assumed.

Methods: To compare the mixture designs efficiency, fraction of design space (FDS) plot is used. We search for the practical mixture designs whose the minimal half-width of the tolerance interval per a standard deviation, which is denoted as d_2 , is less than 4.5 at 0.8 fraction of the design space. They are found by adding the different number of replicates at the four types of the design points in the practical design.

Results: The practical efficient mixture designs for the number of components between three and five are listed. The sequential method to establish a design space is illustrated with the two examples based on the simulated data.

Conclusion: The designs with the center of edge points replications are more efficient than those with the vertex points replication. We propose the sample size of at least 23 for three components, 28 for four components, and 33 for the five components based on the list of efficient mixture designs.

Key Words: Quality by Design(QbD), Design Space, Tolerance Interval, Practical Efficient Mixture Designs

● Received 9 April 2020, 1st revised 22 May 2020, accepted 25 May 2020

† Corresponding Author(yblim@ewha.ac.kr)

© 2020, Korean Society for Quality Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-Commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

* This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education(NRF-2017R1D1A1B03030751).

1. 서론

품질향상을 위한 설계기반 품질고도화(quality by design)는 품질이 제품으로 설계되어야 한다는 최초의 Juran(1992)의 개념이 확대되어 최근에는 실험계획법을 근간으로 하는 실험설계와 수학적 모델에 기반 한 디자인 스페이스(design space)를 개발하게 되었다. 디자인 스페이스 내에서의 제조는 품질 목표 사항(QTPP)과 주요 품질 특성(CQA)이 만족된 제품이 제조 될 수 있도록 보장한다. 즉, 설명변수들의 흔들림이 발생하여 공정조건이 최적조건에서 약간 벗어나도 디자인 스페이스 내에서 제조가 된다면 품질 목표 사항이 만족된 제품이 생산되지만, 디자인 스페이스를 벗어나서 생산된 제품은 품질을 보장받지 못하게 된다. 최근에 의약품 개발에서도 체계적인 의약품 개발을 위한 가이드 라인으로 디자인 스페이스 개념이 적용되고 있다.(ICH Q8)

디자인 스페이스는 품질특성의 허용한계인 공차(tolerance) 개념을 이용하여 생산된 제품의 품질 목표범위 안에 공정 조건에서 제조된 제품의 tolerance 신뢰구간의 포함 여부에 따라서 정의된다. Tolerance 신뢰구간은 신뢰수준 $(1-\alpha)$ 에서 모집단의 $P\%$ 가 신뢰구간에 포함되는 것인데, 보통 $P=0.99$ 로 정의하여 디자인 스페이스에 속한 공정 조건에서 제품이 제조되면 제품의 99%는 품질이 보증 받게 된다. 따라서, 2개 이상의 품질관련 반응 변수들이 있을 경우 이들 tolerance 신뢰구간에 대한 교집합으로 구축된 디자인 스페이스가 공집합이 되지 않아야 하며 그 안에 최적조건이 있어야 품질을 보장하는 공정조건들이 생긴다. 이를 위해 산포가 작은 정밀한 tolerance 신뢰구간이 구축되어야 하는데, 이는 실험설계를 할 때 적절한 실험 계획 및 실험의 크기를 결정하여 예측치의 표준편차를 적정수준으로 줄이는 것과 연관이 있다. 정중희 등(2019)은 반응표면분석에서 디자인 스페이스 구축을 위한 효율적인 실험설계를 찾기 위해 FDS(Fraction of Design Space) 그림을 이용하는 방법을 소개했다. FDS 그림은 Zahran 등(2003)이 제안한 방법으로 총 설명변수 영역에서 지정된 예측치의 표준편차보다 작거나 같은 설명변수 영역에 대한 비율을 그린 그래프이다. 정중희 등(2019)에서는 가정된 이차모형에 대한 반응표면분석으로 실용적인 실험계획인 회전가능 중심합성계획의 요인점, 축점, 중심점을 반복하여 각 반복수마다 FDS 비율이 80%이상인 되는 단위 표준편차 크기 당 tolerance 신뢰구간 기준의 최소한의 폭(d_2)을 찾고, 그 폭이 4.5이하 일 때를 효율적인 실험계획으로 제안하였다. 효율적인 실험계획의 결과를 보면, 축점보다는 요인점을 반복할 때 d_2 값이 작아지고, D-효율성도 좋아지게 됨을 알 수 있었다.

본 논문에서는 혼합물 분석에서 디자인 스페이스 구축을 위해 FDS 그림을 이용한 효율적인 실험을 설계하는 방법에 대해 제안하고자 한다. 혼합물 분석의 설명변수들은 혼합물의 구성 성분의 비율이다. 즉, 반응변수의 평균이 구성 성분의 비율만의 함수가 되고 혼합물의 양에는 영향을 받지 않는 반응표면분석의 특수한 경우이다. q 개 성분의 혼합물에 있어서 x_i 를 i 번째 성분의 혼합비율이라고 하면, 혼합비율들은 수식 (1)과 같은 관계식을 만족해야 하기에 혼합물 분석의 실험 공간은 $q-1$ 차원 심플렉스(simplex) 공간이 된다.

$$x_1 + x_2 + \dots + x_q = 1, \text{ where } x_i \geq 0, i = 1, 2, \dots, q \quad (1)$$

이러한 혼합물 분석을 활용하는 분야는 제약업에서 지효성 알약의 혼합비율이나 석유화학분야에서 가솔린의 혼합비율, 반도체와 LCD 분야에서의 식각용액의 혼합비, 식음료업에서의 과즙음료의 혼합비율의 결정, 표백력과 세탁력을 최적화하는 세제의 혼합비 등 광범위하다.

본 논문에서는 가정된 이차모형에 대한 정준 다항 회귀 모형(canonical polynomial)에 대해서 디자인 스페이스 구축을 위해 심플렉스 공간 영역에서 꼭지점(Vertex), 모서리점(Centroid of Edge), 축점(Axial), 중앙점(Center)을

반복하여 각 반복수마다 FDS 비율이 80% 이상이 되는 단위 표준편차 크기 당 tolerance 신뢰구간 기준의 최소한의 폭인 d_2 를 찾아 비교하여 d_2 를 가장 작게 하는 효율적인 실험설계를 찾고자 한다. FDS 비율이 최소한 80% 이상이 되도록 선택한 이유는 정종희 등(2019)에서도 언급하였듯이 Whitcomb(2008)이 예측치의 표준편차가 주어졌을 때에 FDS 비율이 최소한 80% 이상이 되도록 실험의 설계 및 크기를 결정하라는 실무적 제안에 따라서 이다. 특히, d_2 가 4.5를 상회하는 실험설계에 대해서 축차적 접근방법으로 d_2 를 축소하여 4.5이하가 될 수 있는 실험점별 반복수를 찾고자 한다. $d_2 = 4.5$ 로 선택된 이유는 6 시그마(six sigma) 경영에서 평균이 1.5 시그마 이동이 일어난 경우에 규격 한계 까지의 거리가 4.5 시그마이고, 불량률 0.000679%로 공정능력이 매우 우수함을 나타내는 지표에 해당되기 때문이다.

그래서 2장에서는 혼합물 실험설계에 대해 구체적으로 설명하고, 3장에서는 성분수별 혼합물 실험설계에서 반복에 따른 FDS 비율이 80% 이상이 되는 단위 표준편차 크기 당 mean과 tolerance 신뢰구간 기준의 최소한의 폭을 찾아 제시하며, d_2 가 4.5를 상회하는 경우에 축차적으로 d_2 를 축소하여 4.5이하가 될 수 있는 반복수를 찾아 제시한다. 4장에서는 사례분석을 통해 디자인 스페이스를 구축하여 그 안에 최적조건이 포함되는지를 살펴보는 일련의 과정을 소개한다.

2. 혼합물 실험설계

q 개의 성분으로 구성된 혼합물 실험 공간은 $q-1$ 차원 심플렉스 공간이 된다. 또한, 수식 (1)과 같은 제약 조건에 따라 가정된 이차모형은 수식 (2)와 같이 정준 다항 회귀 모형이 된다.

$$y_j = \sum_{i=1}^q \beta_i x_{ij} + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \beta_{ik} x_{ij} x_{kj} + \epsilon_j \text{ where } j=1, \dots, n, \epsilon_j \sim N(0, \sigma^2) \tag{2}$$

$$\text{단, } x_i \geq 0, i=1, \dots, q, \sum_{i=1}^q x_i = 1$$

예를 들어, $q=3$ 개의 성분으로 구성된 혼합물 실험을 가정해 보자. 실험 공간은 Figure 1의 (a)와 같이 2차원의 심플렉스 공간이고, 가정된 이차모형을 위한 실용적인 실험설계는 Figure 1의 (b)와 같이 {3,2} 심플렉스 격자법 설계에 축점 ($x_i = x_j = 1/6, x_k = 2/3$)과 중앙점 ($x_i = x_j = x_k = 1/3$)이 추가된 경우이다.

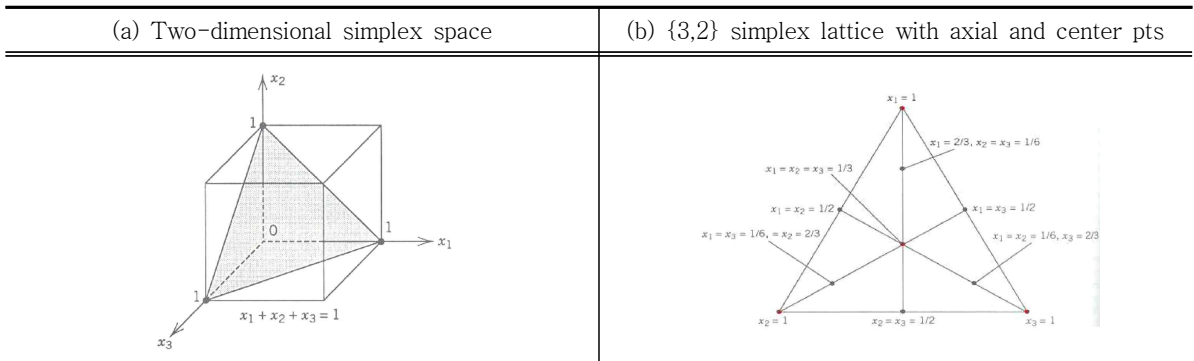


Figure 1. Mixture design for three components

꼭지점, 모서리점, 축점, 중앙점인 4개의 부분으로 이루어진 실험설계로 설계행렬(design matrix)이 결정되고, Figure 1의 (a)에 주어진 2차원 심플렉스 공간 내의 한 점 x_0 를 선택하면 그 점에서 실험의 크기 n 에 대하여 조정되지 않은 예측치 분산(UPV : unscaled prediction variance)의 제곱근 \sqrt{UPV} 가 예측 표준편차가 된다.

$$\sqrt{UPV(x_0)} = \sqrt{\frac{Var[\hat{y}(x_0)]}{\sigma^2}} = \sqrt{x_0'(X'X)^{-1}x_0} \tag{3}$$

단, X 는 설계행렬(design matrix)이다.

정중희 등(2019)에서 \sqrt{UPV} 의 추정치를 신뢰구간의 폭(평균기준 d_1 , tolerance 기준 d_2)과 표준편차 s 를 이용하여 표현하였으며 이는 수식 (4), 수식 (5)와 같다.

- x_0 에서의 모평균 반응값에 대한 $(1-\alpha)\%$ 신뢰구간 기준

$$\sqrt{UPV(x_0)} = \sqrt{\frac{Var[\hat{y}(x_0)]}{s^2}} = \frac{1}{t(\alpha/2,df)} \frac{d_1}{s} \tag{4}$$

- $(1-\alpha)\%$ 신뢰수준에서 모집단의 $P\%$ 가 신뢰구간에 포함되는 tolerance 신뢰구간 기준

$$\sqrt{UPV(x_0)} = \left[\frac{d_2}{s} - \Phi^{-1}(P) \sqrt{\frac{n-p}{\chi^2(1-\alpha/2, n-p)}} \right] / t(\alpha/2, n-p) \tag{5}$$

Φ^{-1} : 정규 역 누적 분포함수, P : 모집단의 비율

전체 실험 영역의 모든 x_0 에서 산출된 예측 표준편차 중에서 기준 예측치의 표준편차보다 작거나 같은 실험 영역의 비율을 FDS 비율로 정의한다. 따라서 수식 (4)와 수식 (5)에서 산출된 예측 표준편차 추정치들을 이용하여 FDS 비율이 80%가 되는 기준 예측치를 구할 수 있고 그때의 d_1, d_2 를 찾을 수 있다. FDS 비율이 80%가 되는 d_1, d_2 의 값이 작을수록 FDS 관점에서 효율적인 실험설계로 제안한다.

3. 효율적인 실험설계

FDS 관점에서 효율적인 실험설계를 제안하기 위해 혼합물 실험설계의 구성성분의 개수인 q 가 3, 4, 5 인 경우에 FDS 비율이 80% 이상이 되는 단위 표준편차 당 mean과 tolerance 신뢰구간 기준의 최소한의 폭인 d_1, d_2 를 찾아 비교하고, 효율적인 실험설계를 추천하고자 한다. 이를 위해 제 1종 오류의 확률 α 는 0.05, 표준편차 s 는 1, tolerance 신뢰구간을 위한 모집단의 비율 P 는 0.99로 가정한다. d_1, d_2 의 값이 작을수록 FDS 관점에서 효율적인 실험설계가 된다. 또한 제시되는 실험설계의 효율성을 비교하기 위해 D-효율(scaled D-optimal efficiency) 값도 제시하여 비교하는데, 이때 D-효율성은 수식 (6)에 정의된 것처럼 관측치 1개당 효율성을 정의하고 있고, 작은 값일수록

효율적인 실험설계에 해당한다.

$$D\text{-효율성(scaled D-opt)} = n(\det(X'X)^{-1})^{1/p} \quad (6)$$

q 개의 성분일 경우 가정된 이차모형을 위한 혼합물 실험설계는 $\{q, 2\}$ 심플렉스 격자법을 기본 설계로 한다. q 개의 꼭지점, qC_2 개의 모서리점, q 개의 축점, 1개의 중앙점을 형성할 수 있으며 각 점에서 완전(또는 부분) 반복으로 실험의 크기를 증가시킬 수 있다. 축점과 중앙점은 이차모형을 가정할 때 포함하지 않아도 되나, 축점과 중앙점이 추가되면 심플렉스 공간 내부의 정보가 추가될 수 있으며 특별 삼차모형으로 모형 확장이 가능하여 실용적인 실험설계가 가능해진다. 예를 들어, $q=3$ 인 경우, 간결한 실용적인 실험설계는 $\{3, 2\}$ 심플렉스 격자법에 중앙점이 추가된 실험설계로 실험의 크기는 꼭지점 3개, 모서리점 3개, 중앙점 1개로 7개가 된다. 여기서 꼭지점 전체를 한번 반복하면 실험의 크기는 10개가 된다. 물론 모서리점을 한번 반복해도 10개가 된다. 이런 방식으로 각 영역들을 반복하면 동일한 실험의 크기에서도 서로 다른 후보 실험설계들을 찾을 수 있다. 동일한 실험의 크기인 경우 어떤 형태의 후보 실험점을 반복하는 것이 더 효율적인지 찾아보고자 Table 1과 같이 축차적인 방법으로 실험점을 추가하여 효율성을 비교한다. 실험설계는 실험계획에 특화된 소프트웨어인 Design Expert 12를 통해 생성하였으며 FDS 비율 및 d_1 , d_2 , D-효율성 지표는 정종희 등(2019)에서 개발한 R 프로그램을 이용하여 산출하였다.

Table 1의 (a)에는 성분의 수 $q=3$ 인 경우에 실험의 크기 n , $18 \leq n \leq 21$ 에 대하여 FDS 비율이 80%가 되는 단위 표준편차 당 d_2 가 가장 작은 실험설계를 찾아 영역별로 반복수를 정리한 실험설계가 'Space Type'으로 표시되어 있다. $n=18$ 인 경우는 꼭지점 6개(2회 반복), 모서리점 9개(3회 반복), 중앙점 3개(3회 반복)인 설계의 d_2 가 4.94로 다른 조합에 비해 가장 작다는 것을 의미한다. $n=19$ 인 경우에는 꼭지점 6개(2회 반복), 모서리점 12개(4회 반복), 중앙점 1개인 실험설계를 추천한다. $n=20, 21$ 인 경우에는 $n=19$ 의 실험설계에서 중앙점의 수를 각각 증가시킨 실험설계이다. 실험의 크기 $n=18$ 인 경우 실험자료의 분석을 생각해 보자. 추천된 d_2 가 4.94인 실험설계로부터 얻어진 실험자료의 분석 결과, 디자인 스페이스가 공집합인 경우에 해결책의 하나는 이 보다 더 작은 d_2 값을 갖게 하기 위해 축차적으로 추가 실험을 실시하는 것이다. 이럴 경우 어떤 영역의 점을 반복해야 하는지 알아보기 위해 추가점 3개를 각 영역별로 반복했을 때의 d_1 , d_2 와 D-효율성을 찾아 정리하였다. 또한, I-optimal 실험설계로 심플렉스 영역의 4가지 형태의 모든 실험점들을 후보실험점으로 하여 3개의 추가점을 찾은 결과도 같이 보여주었다. 찾아진 I-optimal 실험설계는 모서리점을 3개 추가하도록 나와서 결과값들이 모서리점 추가와 동일하다. 이는 실험의 크기 $n=19, 20, 21$ 개 일 때도 마찬가지이다. 결론적으로는 현 실험설계에서 모서리점을 한 번 더 반복하는 것이 다른 영역을 반복하는 것보다 d_2 값이 작아져 FDS 관점에서 더 효율적이라는 것을 알 수 있다. 또한, 꼭지점을 반복하면 d_2 값이 별로 차이가 없는 반면에 D-효율성 관점에서 가장 효율적이 된다. 따라서 실험자의 선호에 따라서 모서리점 혹은 꼭지점을 반복하는 실험설계를 선택할 수 있다. 축점을 반복하면 FDS 지표인 d_2 나 D-효율성도 비효율적이 되나 축점으로 부터 얻고자 하는 심플렉스 내부 정보나 특별 삼차모형 확대 등이 가능하다는 장점이 있다.

Table 1의 (b)는 성분의 수 $q=4$ 인 경우에 해당하는 데, 기본설계가 꼭지점과 축점의 수는 4개이지만 모서리점은 6개로 구성되어 있다. 따라서, 추가 실험점을 4개와 6개일 때로 구분하였고, 추가 실험점 4개일 때의 모서리점은 후보 실험점을 6개의 모서리점만으로 제한한 경우에 I-optimal 기준에서 최적으로 하는 모서리점들을 선택하여 모서리점들의 부분 반복이 되도록 하였다. 마찬가지로 추가점 6개인 경우의 꼭지점과 축점은 후보 실험점들을 각각 꼭지점과 축점으로만 제한한 경우에 I-optimal 기준에서 최적으로 하는 꼭지점과 축점을 선택하여 완전+부분 반복이 되도록 하였다. 또한, I-optimal 설계는 성분의 수 $q=3$ 인 경우와 마찬가지로 모든 후보점들을 대상으로 선택한 결

과이다. 실험의 크기 n , $22 \leq n \leq 25$ 에 대하여 d_2 값이 가장 작은 설계들을 찾은 결과 실험의 크기 $n=22$ 인 실험설계는 꼭지점 4개(1회반복), 모서리점 12개(2회반복), 축점 4개(1회반복), 중앙점 2개이다. 실험의 크기 $n=23$ 인 실험설계는 성분의 수 $q=3$ 인 경우와 유사하게 모서리점 18개(3회반복)로 늘어나는 대신 축점이 없어지고 중앙점은 1개로 선택된다. $n=24, 25$ 인 경우에는 $n=23$ 의 실험설계에서 중앙점의 수를 하나씩 각각 증가시킨 실험설계이다. 성분의 수 $q=4$ 인 경우에 축차적으로 추가 실험점을 찾은 결과, $q=3$ 인 경우와 동일하게 모서리점을 추가하는 것이 FDS 관점에서는 가장 효율적이라는 결과를 얻을 수 있다.

Table 1의 (c)는 성분의 수 $q=5$ 인 경우에 해당한다. 기본설계에서 꼭지점과 축점의 수는 5개이지만 모서리점은 10개로 구성되어 있다. 성분의 수 $q=4$ 인 경우와 마찬가지로 축차적으로 실험점을 추가할 때 추가 실험점의 수가 5개인 경우와 10개인 경우로 나누어 부분 또는 완전+부분 반복으로 추가되는 점에 대한 d_2 를 비교하면 모서리점 추가 실험설계와 I-optimal 설계가 큰 차이가 없음을 알 수 있다. 실험의 크기 n , $26 \leq n \leq 30$ 에 대하여 d_2 값이 가장 작은 설계들을 찾은 결과 실험의 크기 $n=26$ 인 실험설계는 꼭지점 5개(1회반복), 모서리점 20개(2회반복), 중앙점 1개이다. 실험의 크기 n , $27 \leq n \leq 30$ 인 경우에는 $n=26$ 의 실험설계에서 중앙점의 수를 하나씩 각각 증가시킨 실험설계이다. 이들에 대하여 축차적으로 추가 실험점을 찾은 결과, $q=3, 4$ 인 경우와 동일하게 모서리점을 추가하는 것이 FDS 관점에서는 효율적이라는 결과를 얻을 수 있다. 단, 추가 실험점의 수가 5인 경우에는 $q=3, 4$ 인 경우와 유사하게 꼭지점을 반복하면 d_2 값이 별로 차이가 없는 반면에 D-효율성 관점에서 가장 효율적이어서 꼭지점의 반복도 고려할 수 있지만, 추가실험점의 수가 10인 경우에는 모서리점을 추가하는 것이 FDS 관점에서는 효율적이다.

정중희 등(2019)은 반응표면분석에서 $d_2 \leq 4.5$ 인 설계를 최종적으로 효율적인 설계로 추천하였다. 혼합물 분석에서도 동일한 기준을 적용해 보면, 성분 개수 3개인 경우는 실험의 크기 $n \geq 23$, 성분 개수 4개는 $n \geq 28$, 성분 개수 5개는 $n \geq 33$ 인 경우에 해당된다.

Table 1. Minimal d_1, d_2 for mixture designs

(a) Designs with three components

N	Space Type				Index	Value	Additional points (+ 3)				
	Vertex	Centedge	AxialCB	Center			Vertex	Centedge	AxialCB	Center	I-optimal
18	6	9	0	3	d1	1.04	0.95	0.93	0.97	0.99	0.93
					d2	4.94	4.66	4.63	4.67	4.69	4.63
					scaled-D	27.05	25.81	27.76	29.66	29.86	27.76
19	6	12	0	1	d1	0.96	0.89	0.88	0.91	0.91	0.88
					d2	4.79	4.55	4.53	4.56	4.56	4.53
					scaled-D	26.19	24.78	27.24	28.57	28.57	27.24
20	6	12	0	2	d1	0.94	0.87	0.86	0.89	0.90	0.86
					d2	4.70	4.48	4.47	4.50	4.51	4.47
					scaled-D	26.96	25.35	27.94	29.31	29.40	27.94
21	6	12	0	3	d1	0.93	0.86	0.85	0.88	0.89	0.85
					d2	4.63	4.43	4.42	4.45	4.46	4.42
					scaled-D	27.76	25.96	28.66	30.06	30.24	28.66

(2) Designs with four components

- Four Additional points

N	Space Type				Index	Value	Additional points (+ 4)				
	Vertex	Centedge	AxialCB	Center			Vertex	Centedge	AxialCB	Center	I-optimal
22	4	12	4	2	d1	1.09	1.01	0.97	1.02	1.03	0.99
					d2	4.99	4.66	4.63	4.67	4.68	4.65
					scaled-D	69.13	62.71	70.41	77.51	78.28	74.89
23	4	18	0	1	d1	1.00	0.93	0.91	0.93	0.93	0.92
					d2	4.83	4.53	4.52	4.54	4.54	4.53
					scaled-D	61.73	55.06	64.70	68.75	69.00	67.16
24	4	18	0	2	d1	0.98	0.90	0.90	0.92	0.92	0.91
					d2	4.74	4.47	4.46	4.48	4.49	4.47
					scaled-D	63.47	56.38	66.20	70.42	70.90	68.91
25	4	18	0	3	d1	0.96	0.88	0.88	0.90	0.91	0.89
					d2	4.66	4.41	4.41	4.43	4.44	4.42
					scaled-D	65.28	57.74	67.77	72.14	72.81	70.70

- Six Additional points

N	Space Type				Index	Value	Additional points (+ 6)				
	Vertex	Centedge	AxialCB	Center			Vertex	Centedge	AxialCB	Center	I-optimal
22	4	12	4	2	d1	1.09	0.99	0.92	0.99	1.01	0.92
					d2	4.99	4.56	4.48	4.56	4.57	4.48
					scaled-D	69.13	62.41	70.42	81.56	82.96	70.42
23	4	18	0	1	d1	1.00	0.91	0.87	0.91	0.91	0.90
					d2	4.83	4.44	4.40	4.44	4.44	4.43
					scaled-D	61.73	54.56	65.66	72.17	72.81	70.93
24	4	18	0	2	d1	0.98	0.89	0.86	0.89	0.90	0.88
					d2	4.74	4.38	4.35	4.39	4.40	4.38
					scaled-D	63.47	55.75	67.07	73.81	74.74	72.02
25	4	18	0	3	d1	0.96	0.87	0.84	0.89	0.90	0.86
					d2	4.66	4.33	4.31	4.35	4.36	4.33
					scaled-D	65.28	56.98	68.53	75.49	76.67	71.76

(3) Designs with five components
 – Five Additional points

N	Space Type				Index	Value	Additional points (+ 5)				
	Vertex	Centedge	AxialCB	Center			Vertex	Centedge	AxialCB	Center	I-optimal
26	5	20	0	1	d1	1.09	1.01	0.98	1.00	0.98	0.96
					d2	5.09	4.67	4.64	4.65	4.63	4.62
					scaled-D	102.74	97.43	107.14	117.12	117.65	115.12
27	5	20	0	2	d1	1.06	0.99	0.95	0.98	0.97	0.95
					d2	4.96	4.59	4.56	4.58	4.57	4.56
					scaled-D	105.60	99.69	109.56	119.92	120.74	118.08
28	5	20	0	3	d1	1.03	0.97	0.93	0.96	0.96	0.94
					d2	4.86	4.53	4.50	4.53	4.52	4.50
					scaled-D	108.55	102.01	112.06	122.78	123.85	120.81
29	5	20	0	4	d1	1.01	0.95	0.92	0.95	0.95	0.93
					d2	4.77	4.48	4.45	4.48	4.48	4.46
					scaled-D	111.54	104.37	114.61	125.68	126.98	124.23
30	5	20	0	5	d1	0.99	0.94	0.91	0.94	0.94	0.92
					d2	4.70	4.43	4.41	4.43	4.44	4.41
					scaled-D	114.58	106.77	117.21	128.61	130.11	125.66

– Ten Additional points

N	Space Type				Index	Value	Additional points (+ 10)				
	Vertex	Centedge	AxialCB	Center			Vertex	Centedge	AxialCB	Center	I-optimal
26	5	20	0	1	d1	1.09	0.98	0.88	0.94	0.93	0.87
					d2	5.09	4.45	4.35	4.40	4.40	4.34
					scaled-D	102.74	98.91	108.83	130.90	133.26	119.89
27	5	20	0	2	d1	1.06	0.96	0.86	0.93	0.93	0.87
					d2	4.96	4.40	4.30	4.36	4.37	4.31
					scaled-D	105.60	100.82	110.90	133.63	136.40	122.52
28	5	20	0	3	d1	1.03	0.94	0.85	0.92	0.92	0.85
					d2	4.86	4.36	4.26	4.33	4.34	4.27
					scaled-D	108.55	102.78	113.05	136.40	139.56	125.17
29	5	20	0	4	d1	1.01	0.93	0.83	0.91	0.92	0.85
					d2	4.77	4.32	4.22	4.30	4.31	4.24
					scaled-D	111.54	104.79	115.24	139.20	142.72	127.84
30	5	20	0	5	d1	0.99	0.91	0.82	0.90	0.91	0.84
					d2	4.70	4.28	4.19	4.27	4.28	4.21
					scaled-D	114.58	106.83	117.47	142.03	145.88	130.53

4. simulation 사례

4.1 성분 개수 3개 - 이차모형

Koons 등(1985)은 플라스틱 파이프를 만드는데 사용되는 물질인 ABS(acrylonitrile butadiene styrene)의 배합과 관련된 실험을 하였다. ABS는 일반적으로 폴리 부타디엔(x_1)과 스티렌 아크릴로 니트릴(x_2)으로 만들어 지는데, 석탄-타르피치(x_3)가 추가되어 ASTM 사양을 충족하는 물리적 특성을 갖는 파이프를 형성할 수 있는지 여부를 결정하기 위해 혼합물 실험을 실행하였다. 반응변수는 y_1 (Izod impact strength, 범위 > 1, 망대특성), y_2 (deflection temperature under load, 범위 > 190, 망대특성), y_3 (yield strength, 범위 > 5000, 망대특성)로 3개이다. 꼭지점, 모서리점, 축점, 중앙점을 각각 1회 반복한 기본 설계로 하여 총 실험수는 10개이며, 실제 성분비를 0과 1사이의 pseudo-component로 나타내어 표시하면 Table 2와 같이 실험 데이터가 주어진다.

Table 2. Raw data for example 1

Space Type	x_1	x_2	x_3	y_1	y_2	y_3
Vertex	1	0	0	7.4	205	4350
	0	1	0	6.1	213	6080
	0	0	1	0.8	183	5065
Centedge	0.5	0.5	0	7.3	204	5280
	0.5	0	0.5	3.9	188	5080
	0	0.5	0.5	2.1	191	5740
Axial CB	0.667	0.167	0.167	6.1	192	5035
	0.167	0.667	0.167	4.9	202	5615
	0.167	0.167	0.667	2.5	186	5385
Center	0.333	0.333	0.333	4.4	197	5410

Table 2를 이용하여 각각의 반응변수에 대해서 적절한 모형을 찾은 결과가 수식 (7)에 주어지고, 각각의 반응변수에 대한 모형평가 통계량은 y_1 모형은 $Adj-R^2=0.994$, $MSE=0.0284$ y_2 는 $Adj-R^2=0.834$, $MSE=15.35$ y_3 은 $Adj-R^2=0.975$, $MSE=5599.96$ 을 얻었다.

$$\hat{y}_1 = 7.38x_1 + 6.04x_2 + 0.87x_3 + 2.22x_1x_2 - 0.51x_1x_3 - 5.20x_2x_3$$

$$\hat{y}_2 = 199.88x_1 + 209.54x_2 + 178.88x_3 \tag{7}$$

$$\hat{y}_3 = 4374.27x_1 + 6045.18x_2 + 5077.91x_3 + 182.98x_1x_2 + 1508.43x_1x_3 + 570.25x_2x_3$$

반응변수 목표 범위 안에서 동시 최적화를 해보면, 성분 비율이 $x_1 = 0, x_2 = 1, x_3 = 0$ 일 때 $\hat{y}_1 = 6.04, \hat{y}_2 = 209.54, \hat{y}_3 = 6045.18$ 의 예측치를 얻을 수 있다. 또한, Figure 2의 (a)와 같이 반응변수 목표 범위에 대한 tolerance 신뢰구간 (one side)의 교집합인 디자인 스페이스가 생성되고 그 안에 최적의 해가 포함됨을 확인할 수 있다. 그러나, 디자인 스페이스의 영역이 너무 좁아 성분 비율이 최적조건에서 조금만 흔들려도 디자인 스페이스를 벗어나 제품의 품질이 만족되지 않을 수 있다. 각 반응변수의 예측치, 표준편차와 tolerance 신뢰하한 값을 이용하여 단위 표준편차 당 tolerance d_2 값으로 환산해보면, y_1 의 d_2 값은 8.17, y_2 는 6.03, y_3 는 8.17이 된다. 그래서, 디자인 스페이스를 넓게 형성하기 위해 d_2 값을 작게 하는 방향으로 축차적인 설계를 진행하였다. Table 1의 (a)에서 축점이 3개 있으면서 d_2 값이 4.5보다 작은 설계는 꼭지점 6개(2회반복), 모서리점 12개(4회반복), 축점 3개(1회반복), 중앙점 3개(3회반복)인 설계로 FDS 80% 이상이 되는 가장 작은 d_2 값이 4.45임을 확인할 수 있다. 따라서, 원 설계에 꼭지점 1회, 모서리점 3회, 중앙점 2회 반복을 더 추가하고, 시뮬레이션으로 반응값을 생성하기 위해 수식 (7)의 식에 $N(0, MSE)$ 으로 발생된 오차항을 각각 더하였다. 그렇게 생성된 반응값 들이 Table 3에 정리되어 있다.

Table 3. Additional data for sequential design of experiment

Space Type	x_1	x_2	x_3	y_1	y_2	y_3
Vertex	1	0	0	7.432	210.6	4359
	0	1	0	6.488	219.5	6132
	0	0	1	1.007	186.3	4981
Centedge	0.5	0.5	0	7.158	221.9	5334
	0.5	0	0.5	4.335	200.9	5124
	0	0.5	0.5	2.331	201.3	5582
	0.5	0.5	0	7.171	209.3	5274
	0.5	0	0.5	3.941	197.9	5237
	0	0.5	0.5	2.371	200.3	5696
	0.5	0.5	0	7.415	212.2	5114
	0.5	0	0.5	4.271	197.4	5085
	0	0.5	0.5	2.158	202.3	5549
Center	0.333	0.333	0.333	4.351	199.1	5436
	0.333	0.333	0.333	4.437	207.5	5480

앞서 진행된 방법과 동일하게 이차 모형식을 적합하여 반응변수 목표 범위 안에서 최적화를 해보면, 성분 비율이 $x_1 = 0, x_2 = 1, x_3 = 0$ 인 것이 최적의 해가 되고 Figure 2의 (b)와 같이 디자인 스페이스 안에 최적의 해가 포함된다. 또한, 디자인 스페이스 크기가 더 커졌음을 확인할 수 있다. 따라서, 최적의 성분 비율에서 품질을 생성하다가 흔들림이 발생하더라도 디자인 스페이스 내라면 품질의 만족도에는 큰 영향을 주지 않을 것이라 예상할 수 있다.

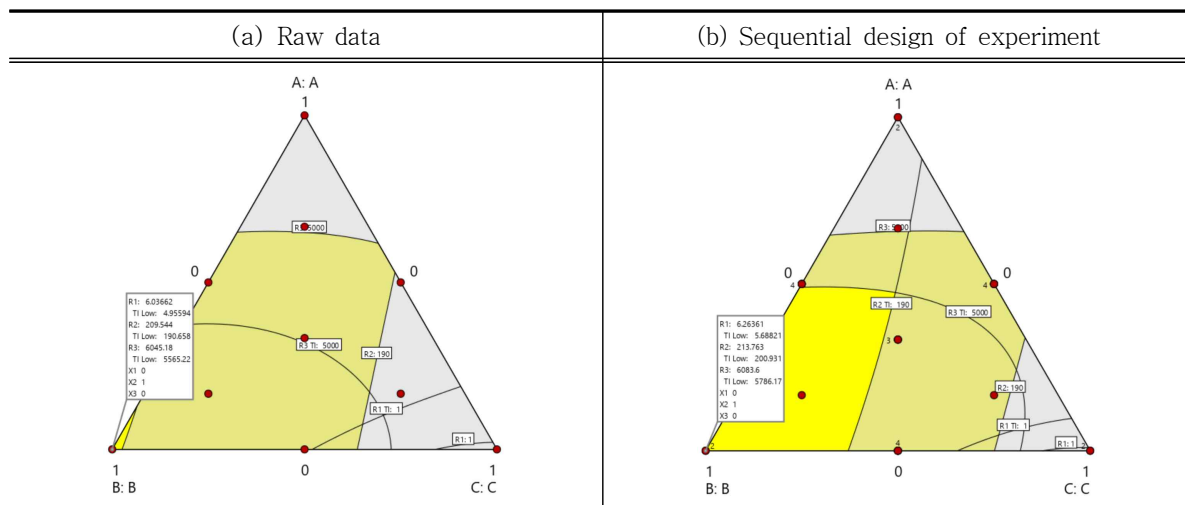


Figure 2. Design space with tolerance interval (Bright yellow : design space)

4.2 성분 개수 3개 - 특수 삼차모형

Design Expert 사용 지침서(Stat-Ease 2019) 예제로 있는 세제의 점도(viscosity, y_1 , 범위 39~48, 망대특성)와 혼탁도(turbidity, y_2 , 범위 800~900, 망소특성)를 반응값으로 하고 물(x_1), 알코올(x_2), 요소(x_3) 3가지 성분 비율을 설명변수로 하는 혼합물 실험의 경우를 살펴보고자 한다. 꼭지점 6개(2회반복), 모서리점 3개(1회반복), 축점 3개(1회반복), 중앙점 2개(2회반복)로 총 실험수 14개이며 실제 성분비를 0과 1사이의 pseudo-component로 나타내어 표시된 실험 데이터는 Table 4에 정리되어 있다.

Table 4. Raw data for example 2

Space Type	x_1	x_2	x_3	y_1	y_2
Vertex	1	0	0	37.2	378
	0	1	0	91.6	546
	0	0	1	144	641
	1	0	0	40.8	436
	0	1	0	87.8	323
	0	0	1	130	786
Centedge	0.5	0.5	0	67.9	436
	0.5	0	0.5	46.5	630
	0	0.5	0.5	45.3	949
Axial CB	0.667	0.167	0.167	35.1	671
	0.167	0.667	0.167	51.7	730
	0.167	0.167	0.667	70.7	874
Center	0.333	0.333	0.333	34.8	984
	0.333	0.333	0.333	46.0	1122

Table 4를 이용하여 적절한 모형을 찾은 결과, y_1 은 이차 모형식에 적합되나, y_2 는 특수 삼차 모형식(special cubic)에 적합되어 수식 (8)과 같은 예측치가 얻어졌다. 이때 y_1 의 모형 평가 통계량은 $Adj-R^2=0.976, MSE=29.70$, y_2 는 $Adj-R^2=0.853, MSE=8751.45$ 을 얻었다.

$$\hat{y}_1 = 38.54x_1 + 89.25x_2 + 138.03x_3 + 6.01x_1x_2 - 165.31x_1x_3 - 271.42x_2x_3$$

$$\hat{y}_2 = 408.67x_1 + 428.10x_2 + 705.15x_3 + 32.66x_1x_2 + 238.94x_1x_3 + 1411.51x_2x_3 + 8445.33x_1x_2x_3 \quad (8)$$

반응변수 목표 범위 안에서 동시 최적화를 해보면, 성분 비율이 $x_1=0.433, x_2=0.433, x_3=0.133$ 일 때 $\hat{y}_1=49.68, \hat{y}_2=769.49$ 의 예측치를 얻을 수 있다. 또한, Figure 3의 (a)와 같이 반응변수 목표 범위에 대한 tolerance 신뢰구간(one side)의 교집합인 디자인 스페이스가 생성되었으나 그 영역이 너무 좁고, 최적조건이 디자인 스페이스 밖에 형성되고 있다. 각 반응변수의 단위 표준편차 당 tolerance d_2 값으로 환산해보면, y_1 의 d_2 값은 5.46, y_2 는 5.73이 된다. 이번에도 디자인 스페이스를 넓게 형성하기 위해 d_2 값을 작게 하는 방향으로 축차적인 설계를 진행하였다. y_2 가 특수 삼차 모형식으로 적합되어 성분 3개에 특수 삼차 모형식을 가정했을 때 FDS 80%이상을 만족하는 가장 작은 d_2 값을 설계별로 비교해 보면, 꼭지점 6개(2회반복), 모서리점 12개(4회반복), 축점 3개(1회반복), 중앙점 4개(4회반복)인 경우가 d_2 값이 4.49로 4.5보다 작게 됨을 알 수 있었다. 모서리점을 반복하는 것이 FDS 관점에서 더 효율적이라는 개념은 이차 모형식을 가정했을 때와 동일하다. 특수 삼차 모형식을 가정하더라도 d_2 값의 변화 면에서 Table 1의 결과와 큰 차이가 없다. 따라서, 원 설계에 모서리점 3회, 중앙점 2회 반복을 더 추가하고, 시물레이션으로 반응값을 생성하기 위해 수식 (8)의 식에 $N(0, MSE)$ 으로 발생된 오차항을 각각 더하였다. 그 결과가 Table 5에 정리되어 있다.

Table 5. Additional data for sequential design of experiment

Space Type	x_1	x_2	x_3	y_1	y_2
Centedge	0.5	0.5	0	67.16	511.4
	0.5	0	0.5	48.54	664.8
	0	0.5	0.5	58.62	1096
	0.5	0.5	0	71.93	505.2
	0.5	0	0.5	42.64	659.6
	0	0.5	0.5	60.51	1044
	0.5	0.5	0	73.17	477.4
	0.5	0	0.5	52.43	762.1
	0	0.5	0.5	41.4	1034
Center	0.333	0.333	0.333	51.63	1094
	0.333	0.333	0.333	35.38	1122

축차적으로 얻어진 데이터를 합하여 분석하면 성분 비율이 $x_1=0.5, x_2=0.5, x_3=0$ 인 것이 최적의 해가 되고 Figure 3의 (b)와 같이 최적의 해가 포함되면서 더 넓어진 디자인 스페이스를 구축할 수 있다.

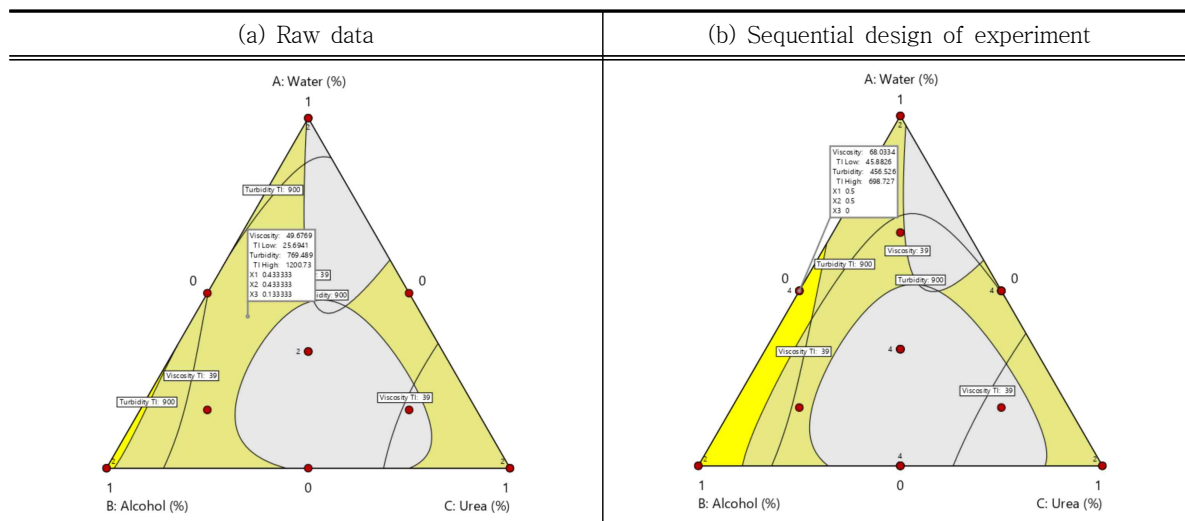


Figure 3. Design space with tolerance interval (Bright yellow : design space)

5. 결론

반응표면분석의 특수한 형태인 혼합물 분석에 있어서 실험의 목적이 품질 고도화에서 디자인 스페이스의 효율적인 구축이라면, 여러 실험설계들의 비교를 위해 FDS 비율을 이용할 것을 제안한다. 잘 알려진 실용적인 혼합물 실험 설계는 q 개의 구성 성분의 혼합물 실험공간인 $q-1$ 차원 심플렉스 공간상의 4가지 형태의 실험점인 꼭지점, 모서리점, 축점, 중앙점으로 구성된 실험 설계이다. 각 형태의 실험점의 반복을 통해 효율적인 실험설계를 선택할 수 있는 여러 후보 실험설계들을 만들어 낼 수 있다. 이들 중에서 가정된 이차모형에 대한 FDS 비율이 80% 이상이 되는 단위 표준편차의 크기 당 tolerance 신뢰구간 기준의 최소한의 폭인 d_2 가 가장 작은 설계를 효율적인 설계라고 제안한다. 다른 영역에 비해 모서리점을 반복할 때 d_2 값이 작아짐을 알 수 있고, 이는 특수 삼차 모형을 가정할 때도 동일함을 예제를 통해 살펴 보았다. 본 논문에서 축차적인 방법으로 디자인 스페이스를 더 넓게 구축하는 과정을 설명한 예제들을 바탕으로 혼합물 분석이 필요한 연구원들도 품질 고도화 관점에서 효율적인 실험설계를 구축할 수 있을 것이라 기대한다.

REFERENCES

- Chung, J. H., Kim, J., Lim, Y. B. 2019. Efficient Designs to Develop a Design Space in Quality by Design. Journal of the Korean Society for Quality Management 47(3):523-535.
- Juran JM. 1992. Juran on Quality by Design: The New Steps for Planning Quality into Goods and Services. New York: Simon & Schuster.
- Koons, G. F., Wilt, M. H. 1985, Design and Analysis of an ABS Pipe Compound Experiment. Experiments in Industry: Design, Analysis, and Interpretation of Results, p 111-117.
- Stat-Ease, Inc. Design-Expert ® Software Version 12. 2019. Tutorials.

<https://www.statease.com/docs/v12/tutorials/>

US Food and Drug Administration. 2009. Guidance for Industry: Q8 (R2) Pharmaceutical Development. Center for Drug Evaluation and Research.

Whitcomb, P.J. 2008. FDS-A Power Tool for Designers of Optimization Experiments. Stateaser Newsletter from Stat-Ease., Inc.

Zahran, A., Anderson-Cook, C. M., Myers, R. H. 2003. Fraction of Design Space to Assess Prediction Capability of Response Surface Designs. Journal of Quality Technology 35(4):377-386.

저자소개

임용빈 서울대학교 계산통계학과를 졸업하고, 동 대학원에서 이학석사 학위를 취득하였으며 미국 퍼듀대학교에서 통계학 박사학위를 취득하였다. 현재 이화여자대학교 통계학과 교수로 재직중이며, 주요 관심분야는 실험계획법, 반응표면분석, QbD에서 Design Space 구축, 데이터마이닝 기법을 활용한 품질향상, 식스시그마 경영 등이다.

정중희 이화여자 대학교 통계학과를 졸업하고, 동 대학원에서 이학석사와 박사 학위를 취득하였다. 주요 관심분야는 실험계획법, 반응표면분석 등이다.