일반화 선형모형의 디자인 행렬을 이용한 품질 실험 설계
- Design of Experiment Using Design Matrix in Terms of Generalized Linear Model -

최 성 운*
Sungwoon Choi*

Abstract

This study proposes the generation mechanism of various design matrix using generalized linear model for design of experiment. Design generation method of GLM analysis, factorial design(FD) with center points, ANOVA design with lack-of-fit test, and response surface design are introduced. In central composite(CC) design, orthogonal blocking and fractional factorial design(FFD) are presented. We compare the design of Box-Benhken(BB) and face-centred central compsite design.
Keywords: GLM, Design Matrix, CC, BB, FD, FFD

1. 서론

소비자와 시장이 요구하는 주관적이고 추상적인 요구품질을 QFD(Quality Function Deployment)의 HOQ(House of Quality)를 이용하여 객관적이고 구체적인 설계품질로 변환한다. 설계품질은 스펙(Specification)을 양쪽스펙, 상한스펙, 하한스펙 등의 계량스펙과 한도견본, 표본견본 등의 계수스펙으로 구분하여 구성된다.

품질실험설계는 고객이 요구하는 스펙을 만족시키기 위해 망목특성치, 망소특성치, 망대특성치의 목표 실험 반응(Response) 데이터에 영향을 주는 작업조건, 설비조건, 재료의 종류, 측정조건, 치공구조건, 금형조건, 외주생산조건, 환경조건 등의 변수(Variable)인자에 대한 최적 수준 처리조건을 설정하는 것이다.

* 경원대학교 산업공학과

이 경우 ANOVA(Analysis of Variance)가 유용하게 사용되나 이 분석은 1차항을 모두 주효과로, 변수의 곱이 2차항은 2차 교호작용으로, 변수의 제곱은 곡면성(Curvature)으로 묶어서 분석한다. MINITAB에서는 축차 *SS*(Sequential Sum of Squares)을 제공하나 개별항에 대한 정보가 미약하다.

이런 ANOVA의 단점을 보완하는 방식으로 개별항을 직교분해(Orthogonal Decomposition)할 수 있는 회귀(Regression)형태의 일반화 선형 모형(GLM : Generalized Linear Model)[4]이 있다. GLM은 직교성(Orthogonality), 블록화(Blocking), 회전성(Rotationality), 교락(Confounding, Aliasing, Resolution), 일양 정밀도(Uniform Precision), 새로운 실험점 추가 등이 용이할 뿐 아니라 직교 블록화(Orthogonal Blocking)[3]의 분해와 합성이 가능해져 실험의 실시, 분석, 평가에 있어 구조적인 파악과 더불어 효율적이고 효과적인 실현이 가능해 진다.

따라서 본 연구에서는 GLM을 이용하여 품질실험설계에서 유용한 디자인 행렬에 대한 적용단계와 분석방법을 제안한다. 본 연구의 목적은 다음 세가지와 같다. 첫째 중심점을 갖는 2^2 FD(Factorial Design)[1]의 디자인 행렬 생성과 GLM, ANOVA 작성 절차를 단계별로 제시한다. 둘째 중심점을 축점과 더불어 직교성, 회전성을 이용하여 3수준의 실험을 실시하는 중심합성(Central Composite : CC)[1] 디자인 행렬 CC(2^3)과 FFD(Fractional Factorial Design) CC(2_V^{5-1}), CC(2_V^{8-2}) 디자인 행렬을 제시한다. 끝으로 극단점의 반응값이 요구되지 않는 경우의 Box-Behnken(RB) BB(3)[1]과 α=1인 내접설계 CC(2^3), 3^3FD의 디자인 행렬을 비교분석한다.

2. 2^k FD GLM 디자인 행렬

디자인 행렬을 위한 GLM을 1차식 $\widehat{y}=\widehat{\beta}_o+\widehat{\beta}_1x_1+\widehat{\beta}_2x_2+\widehat{\beta_{12}}x_1x_2$를 사용하며 생성, 분석 절차는 다음과 같다.

단계 1 : 고객이 요구하는 스펙에 대한 반응특성값(Response Characteristic Value) y의 망목, 망대, 망소의 실험 목표를 설정한다.

단계 2 : 특성요인도(Cause and Effect Diagram)를 이용하여 반응값에 영향을 주는 인자변수(Factor Variable)를 선정한다.

단계 3 : 변수 X의 구체적이고 실현가능한 수준처리(Level Treatment)를 선정한다.

단계 4 : GLM에 맞는 변수의 디자인 행렬 X, 반응값 y를 추가되는 중심점을 고려한 실험횟수(Runs)에 따라 작성한다.

단계 5 : GLM에서는 항(Term), 효과(Effect), 계수(Coefficient), 계수의 SE(Standard Error), T통계량, P-Value가 분석된다. 항은 GLM에서 상수, x_1, x_2, $x_1^2+x_2^2$ (Center Point :중심점)에 해당되며 효과는 수준간의 평균차이다. 계수는 $y=x\beta+\epsilon$에서 $\widehat{\beta}=(x'x)^{-1}x'y$이고 계수 $SE=(x'x)^{-1}$의 대각원소(Diagonal

Element)×MSE)$^{1/2}$, $T=\beta/$계수 SE, P-Value는 T값이 갖는 우측 또는 좌측 면적이다.

단계 6 : 회귀 ANOVA에서는 Source, DF, SS, MS, F, P가 분석된다. Source는 주효과 x_1+x_2, 2차 교호작용 $x_1 \times x_2$, 곡면성(Curvature), $x_1^2+x_2^2$, 잔차오차(RE, Residual Error), 순수오차(PE : Pure Error), 적합결여 오차(LOFE : Lack of Fit Error), 총계로 분석된다. DF는 2수준인 경우 자유도 1, 3수준인 경우 자유도 2가 된다. SS는 수준별 실험횟수(Runs)가 일정할 경우 수준별 합의 제곱을 수준의 개수로 나누고, 반응값의 제곱을 전체 런의 수로 나눈값으로 빼주면 된다. 수준별 런의 수가 다른 경우 $SS=\sum_{i=1}^{r_i} T_{r_i}^2/r_i-\frac{T^2}{r}$, $r_i=$각 수준별 런의 수, $r=\sum r_i$, $T_{r_i}=$각 수준별 합, $T=\sum r_i$로 계산된다. 회귀 ANOVA는 Total = Error + Regression에서 $(y_i-\overline{\overline{y}})=(y_i-\widehat{y_i})+(\widehat{y_i}-\overline{\overline{y}})$의 직교식 형태가 된다. 여기서 $(y_i-\widehat{y_i})=(y_i-\overline{y_i})+(\overline{y_i}-\widehat{y_i})$로 RE = PE + LOFE으로 직교분해되어 분석된다.

단계 7 : GLM, ANOVA에서 Term, Source의 P-Value > 유의수준 α로 유의적이 아닐 경우 그 항을 제거하고 NPP(Normal Probability Plot)의 Sparsity Test와 PSE(Pseudo Standard Error)에 의한 검정을 실시한다. 이 검정은 포화실험(Saturated Design)에 의한 검정으로 많은 변수의 유의성을 선별(Screening)하는 효과적인 검정이다.[2]

단계 8 : 반응표면(Response Surface)분석

1) $y=\beta_0-\beta_1x_1+\beta_2x_2$인 경우 x_1에서는 -1수준, x_2에서는 +1수준의 기울기가 높다($\because \beta_1<0, \beta_2>0$)
2) $y=\beta_0-\beta_1x_1+\beta_2x_2+\beta_{12}x_1x_2$인 경우 x_1의 -1수준, x_2의 1수준의 비틀림(Distortion)이 발생한다. ($\because \beta_{12}$가 있어서)
3) $y=\beta_0-\beta_1x_1+\beta_2x_2+\beta_{12}x_1x_2-\beta_{11}x_1^2$인 경우 x_1에서 Cap(Concave)모양이다. ($\because \beta_{11}<0$)
4) $y=\beta_0-\beta_1x_1+\beta_2x_2+\beta_{12}x_1x_2-\beta_{11}x_1^2-\beta_{22}x_2^2$인 경우 x_2에서 Cap(Concave) 모양이다. ($\because \beta_{22}>0$)
5) $y=\beta_0-\beta_1x_1+\beta_2x_2+\beta_{12}x_1x_2-\beta_{11}x_1^2+\beta_{22}x_2^2$인 경우 x_1에서 Cap(Concave), x_2에서 Cup(Convex)으로 Saddle Point 이다. ($\because \beta_{11}<0, \beta_{22}>0$)
6) $y=\beta_0-\beta_1x_1+\beta_2x_2+\beta_{12}x_1x_2+\beta_{11}x_1^2+\beta_{22}x_2^2$인 경우 x_1, x_2에서 Cup(Convex) 모양이다. ($\because \beta_{11}>0, \beta_{22}>0$)[4]

예 : 4개의 중심점이 있는 2^2디자인 행렬

	$\widehat{y}$	$\widehat{\beta_0}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_{12}}$	$\widehat{\beta_{11}}$		
	Runs	상수	x_1	x_2	x_1x_2	x_1^2	x_2^2	y
실험점(FP)	1	1	-1	1	-1	1	1	
: 극단점(EP)	2	1	-1	-1	1	1	1	
: 꼭지점(VP)	3	1	1	-1	-1	1	1	
	4	1	1	1	1	1	1	
중심점(CP)	5	1	0	0	0	0	0	
	6	1	0	0	0	0	0	

3. CC(2^k) GLM 디자인 행렬

CC(2^k) 중심합성계획은 2^k의 요인점(FP : Factorial Point, Extreme Point, Vertices Point)과 $2k$(수준의 수)의 축점(SP : Star Point, Axial Point)과 $4\sqrt{2^k}-2k+4$의 중심점(CP : Center Point)으로 구성된다.

회전성을 위한 $\alpha=(2^k)^{1/4}$로 일양 정밀도(Uniform Precision)의 방향이 아닌 거리의 중심원이 예측분산과 분석에 용이하게 사용된다.

중심합성계획은 Block으로 직교분해(Orthogonal Decomposition)가 가능하고 직교분해된 경우 각 블록내의 변수간의 곱은 영(Zero)이 되고 블록내의 실험점 수와 전체 런 수의 비를 고려한 변수제곱의 비율도 균형있게 층별하여야 한다. 예를 들어 $CC(2_V^{5-1})$ [4]를 1개의 Block으로 하는 경우 2^{5-1}FP+10SP + 6SP로 2개의 직교블럭으로 하는 경우 첫 번째 블럭은 2^{5-1}FP + 6CP, 두 번째 블럭은 10SP + 1SP로 분해된다.

예 : $CC(2^3)$, $CC(2_V^{5-1})$, $CC(2_V^{8-2})$

$CC(2^3)$에서 ±1의 x_1, x_2, x_3가 8런, 0의 x_1, x_2, x_3가 6런, ±1.68이 자신의 변수에 있는 2런씩 3번으로 총 20번의 런의 수로 실험을 실시한다.

$CC(2_V^{5-1})$은 CC에 FFD(Fractional Factorial Design)을 합성한 방법으로 V는 Design Resolution을 나타내며 $x_5=x_1x_2x_3x_4$로 새로운 변수의 생성자(Generator)가 되며 실험을 32런을 한다.

$CC(2_V^{8-2})$는 생성자로 $x_7=x_1x_2x_3x_4$, $x_8=x_1x_2x_3x_4$을 사용하며 90의 런으로 시험된다.[4]

GLM의 적용절차는 지면상 생략하고 단계별 적용은 2와 같다.

4. BB(k) GLM 디자인 행렬

Box-Behnken 디자인 행렬은 중심합성계획 CC에서 변수의 극단값이 의미가 없는 경우 변점(Edge)을 선정하는 방법이다. BB(3)는 (x_1, x_2, x_3)변수에 4런의 $(\pm 1, \pm 1, 0)$, 4런의 $(\pm 1,\ 0,\ \pm 1)$, 4런의 $(0, \pm 1, \pm 1)$, 3런의 $(0, 0, 0)$ 등 총 15회 실험을 실시한다. 이는 27번을 실험하는 3^3 FD(Factorial Design)보다 실험횟수를 절약할 수 있는 장점이 있다.

내접설계 CC(2^3)는 CC(2^3)에 $\alpha = 1$인 면중심(Face Centered)의 계획으로 수정한 것으로 축점(SP)이 없어져 디자인 영역(Design Area)이 그 수준의 영역으로 축소되며 20회의 실험을 실시한다.

디자인 영역관점에서 3^3, BB(3), $CC(2^3)$, 내접설계 2^3순으로 좁아져 변수의 제약조건의 위험도에 따라 적절한 GLM 디자인 행렬을 선택해야 한다. 실험횟수 관점으로는 BB(3), $CC(2^3)$, 3^3순으로 커져 BB(3)가 가장 경제적인 디자인 방식이다.

작성절차는 2와 같으며 구체적인 디자인 3^3에 의한 분석이론은 지면상 생략한다.

5. 결 론

본 연구에서는 품질실험설계에 적용할 수 있는 일반화 선형모형을 이용한 다양한 디자인 행렬의 생성방법과 적용방안을 제시하였다. 중심점을 갖는 2^k요인배치법(FD)의 디자인 행렬 생성방법, GLM분석, 적합성 결여가 포착된 ANOVA 분석 방법, 반응표면 분석방법을 제시하였다. 중심합성(CC)계획에서는 직교블럭화, 해상도를 이용한 부분실시법(FFD)의 방안을 제안하였고 박스벤켄(BR)과 내접설계의 중심합성(CC)계획법, 3^k 요인배치법(FD)의 디자인을 비교하였다.

6. 참 고 문 헌

[1] 최성운, “안전 및 환경적용을 위한 최소실험계획”, 대한안전경영과학회지, 7 (5) (2005) : 69-84.

[2] 최성운, “k^n요인배치법에서 포화실험에 의한 요인효과의 검정”, 대한안전경영과학회 춘계학술대회 발표 논문집, (2008) : 295-299.

[3] Mathews P.G., Design of Experiments with Minitab, ASQ Quality Press, 2005.

[4] Neter J.M., Kutner M.H., Nachtsheim C.J., Wasserman W., Applied Linear Statistical Models, 4th Ed., McGraw-Hill, 1996.