

다범주 순서형 품질특성을 갖는 제품의 최적 공정조건 결정에 관한 연구

김상철*, 윤원영**†, 전영록***

* 포스코, ** 부산대학교 산업공학과, *** 경남대학교 테크노경영학부

Optimal Process Condition for Products with Multi-Categorical Ordinal Quality Characteristic

Sang-Cheol Kim*, Won-Young Yun**†, Young-Rok Chun***

* POSCO, ** Department of Industrial Engineering, Pusan National University,

*** Department of Industrial Engineering, Kyungnam University

Key Words : Logistic regression, ANOVA, Simulation, Optimal process control

Abstract

This paper deals with an optimal process control problem in production of hull structural steel plate with high defective rate. The main quality characteristic(dependent variable) is the internal quality(defect) of plates and is dependent on process parameters(independent variables). The dependent variable(quality characteristics) has three categorical ordinal data and there are 35 independent variables(29 continuous variables and 6 categorical variables).

In this paper, we determine the main factors and to develop the mathematical model between internal quality predicted probabilities and the main factors. Secondly, we find out the optimal process condition of main factors through analysis of variance(ANOVA) using simulation. We consider three models to obtain the main factors and the optimal process condition: linear, quadratic, error models.

1. 서 론

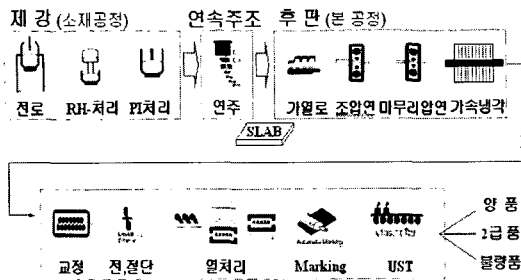
철강산업 분야에서도 강제 제조기술의 발달로 공정과정의 연속화와 공정시간의 단축

이 이루어지고 있다. 이를 통해 제조원가를 감소시켜 생산효율 증대는 물론 제품의 고급화 및 고기능화에 의한 부가가치의 극대화를 활발히 추진하고 있다.

제철소 후판제품의 제조공정을 간략하게 도식화하면 <그림 1>과 같다. 우선 철강을

† 교신저자 wonyun@pusan.ac.kr

생산하기 위해 주원료인 철광석과 원료탄 그리고 부원료인 석회석, 망간 등을 고로에 장입하여 선철을 만든다. 다음 제강공정에서 강을 정련한 후 연속주조를 거쳐서 최초의 형상을 갖는 슬래브(Slab)를 만든다. 생성된 슬래브를 가열로에서 재가열 후 급속의 소성변형을 이용하여 지시된 치수로 만들기 위한 압연공정, 가열·압연공정을 거친다. 그 후 다시 압연 후 수냉각에 의한 강제 냉각속도 제어를 통하여 용접성이 우수한 고인성·고장력강을 제조하는 가속냉각작업이 수행된다. 압연 시 또는 가공 열처리 시 발생된 형상불량을 평탄하게 교정하는 교정 작업, 압연 완료된 날판을 요구치수로 절단하는 전·절단작업, 제품을 라인 밖에서 적정온도로 재가열하고 공냉 또는 수냉을 실시하여 제품조직의 균질화, 미세화, 내부응력제거 및 탄화물 석출을 촉진시켜 강의 내부품질을 개선하기 위해 실시하는 열처리 공정, 제품상에 식별표시를 하는 표시작업 및 초음파 탐상검사에 의한 내부품질결함검사를 거쳐서 후판제품이 만들어진다 [4,5].



<그림 1> 후판제품 제조공정

후판제품의 주요 품질관리 항목으로는 크게 기계적 성질인 인장강도(TS), 항복강도

(YP), 연신율(EL) 그리고 내부품질, 표면품질, 평탄도, Camber, 폭, 길이, 두께 등이 있다. 이러한 품질관리 항목 중 내부품질은 매우 중요하다.

판 내부에 결함이 있게 되면 용접이나 가공을 곤란하게 하는 등 치명적일 수 있으며 보일러나 원자로용으로 사용하는 경우에는 열피로나 취성과파괴의 기점으로 될 가능성이 있다. 이러한 내부결함 유무의 판단은 초음파 탐상에 의한 비파괴 검사로 실시되는 것이 일반적이며, 현재 자동탐상방법에 의하여 라인 상에서 실시되고 있다 [5].

후판제품의 내부품질은 제강공정에서 부터 후판공정까지 다양한 인자들에 영향을 받고 있다. 이 품질특성치의 평가는 현재 "1(합격품)", "2(2급품)", "3(불합격품)"와 같은 세 가지 값인 다범주-순서형 자료로 표현된다. 그러므로 범주형 자료의 특성을 가진 종속변수와 이에 영향을 미치리라고 생각되는 독립변수가 범주형과 연속형이 동시에 존재하는 경우의 분석이 요구된다.

따라서, 본 논문은 변수들 간의 상호관련성을 찾고 종속변수에 영향을 미치는 독립변수 중에서 주요변수를 선정하고 이들의 관계에 대하여 종속변수의 각 값에 대한 누적확률을 수식으로 구현할 수 있는 순서형 다범주 로지스틱 회귀분석(Cumulative Logistic Models for Ordinal Responses)을 실시하고자 한다.

이렇게 해서 도출된 주요공정변수(일차모형, 이차모형)의 최적조건을 탐색하기 위하여 로지스틱 회귀모형에서 도출된 종속변수의 각 값의 예측확률의 수식을 사용하였다. 그리고 도출된 주요공정변수의 작업범위를 일정한 구간(수준)을 나누어 범주화시켜 시뮬레이션의 반복실험을 통해 선정된 주요공정변수의 모든 수준에서 자료를 도출, 이를

수식에 넣어 종속변수 각 값에서의 예측확률을 계산하고 판매단가를 가중치로 부여한 새로운 목적함수를 설정하고 완전 요인실험을 통하여 분산분석을 실시하여 유의한 변수 및 최적공정조건을 결정하고자 한다.

2. 로지스틱 회귀 모형

2.1 로지스틱 회귀모형

이항반응변수 Y 와 양적 설명변수 X 에 대해 $\pi(x)$ 를 $X=x$ 일 때 "성공" 확률($P(Y=1)$)이라 하면 로지스틱회귀 모형은 식 (1)의 선형식으로서 표현될 수 있는데 이를 로지스틱 회귀함수(Logistic Regression Function)라 하며, $\pi(x)$ 는 S-곡선 형태의 함수로서 증가 또는 감소하는 곡선이다.

$$\text{logit}[\pi(x)] = \log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta x \quad (1)$$

여기서,

$$\pi(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} \quad (2)$$

이다. 모수 β 는 S-곡선의 증가 또는 감소율을 나타낸다.

설명변수가 여러 개인 경우 (설명변수들은 연속형, 범주형 변수 또는 두 가지 유형이 모두 가능함[1, 3, 6, 7]) 설명변수들을 X_1, X_2, \dots, X_k 할 때 $Y=1$ 일 확률 π 에 대한 로지스틱모형을 식(3)과 같이 일반화할 수 있다.

$$\text{logit}(\pi(\vec{x})) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (3)$$

범주 1, ..., J 를 갖는 반응변수 Y 가 범주 j 이하에 해당할 확률은 아래 식 (4)와 같으며, 이를 j 번째 누적확률이라 하며,

$$\begin{aligned} P(Y \leq 1) &\leq P(Y \leq 2) \leq \dots \leq P(Y \leq J) = 1 \\ P(Y \leq j) &= \pi_1 + \dots + \pi_j, \quad j = 1, \dots, J \end{aligned} \quad (4)$$

처음 $J-1$ 개의 누적확률의 로짓(Logit)은 식(5)와 같다.

$$\begin{aligned} \text{logit}[P(Y \leq j)] &= \log\left(\frac{P(Y \leq j)}{1 - P(Y \leq j)}\right) \\ &= \log\left(\frac{\pi_1 + \dots + \pi_j}{\pi_{j+1} + \dots + \pi_J}\right), \quad j = 1, \dots, J-1 \end{aligned} \quad (5)$$

식 (5)의 누적 로짓은 설명변수 X 에 대해 아래 식(6)의 모형과 같다[3, 6, 9, 10].

$$\begin{aligned} \text{logit}[P(Y \leq j | X = x_1, \dots, x_k)] \\ = \alpha + \beta_1 x_1 + \dots + \beta_k x_k, \quad j = 1, \dots, J-1 \end{aligned} \quad (6)$$

그리고 설명변수 중에 범주형 변수(이하 질적 변수)가 있을 경우 가변수를 사용하여 모형을 표현할 수가 있다[1].

2.2 로지스틱 모형의 검정

다 범주 로지스틱모형에 대한 검정방법은 다항분포에 근거하여 우도함수가 표현되고 이에 대해 간단히 요약하면 다음과 같다. 보다 상세한 내용을 위해 언급된 참고문헌을 참고하면 된다.

(1) $H_0: \beta_i = 0$ 에 대한 유의성검정

즉 성공(예측)확률이 특정 설명변수에 영향을 받는가를 검정하는 것으로 다음의 검정통계량을 사용한다[6, 9].

$$Z^2 = (\hat{\beta} / ASE)^2 \quad (7)$$

여기서 ASE 는 $\hat{\beta}$ 의 점근적 표준오차이다.

(2) $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ 에 대한 유의성 검정

우도함수의 비를 이용하는 것으로 우도비 검정 통계량은 식(8)으로 정의되며, 여기서 l_0 와 l_1 은 최대로그 우도함수이며, 이 통계량은 근사적으로 자유도 k 인 카이제곱분포를 따른다[3, 5, 9, 11].

$$\begin{aligned} X^2 &= -2 \log(l_0/l_1) = -2[\log(l_0) - \log(l_1)] \\ &= -2(L_0 - L_1) \sim \chi^2_{(k)} \end{aligned} \quad (8)$$

다음은 스코어검정으로 귀무가설을 나타내는 단순모형에서의 최대 우도함수 값(l_0)의 기울기의 크기를 활용한 것으로 식(9)와 같이 표현하며, 이 분포는 점근적으로 자유도 k 인 카이제곱분포를 따른다[7, 9, 11].

$$X^2 = \left[\frac{\delta \log l_0}{\delta \beta} \right] \left[\frac{\delta^2 \log l_0}{\delta \beta^2} \right]^{-1} \left[\frac{\delta \log l_0}{\delta \beta} \right] \sim \chi^2_{(k)} \quad (9)$$

관찰된 결과 값과 예측된 확률간의 순위 상관으로서 둘 이상의 변수들 간에 존재하는 의존성의 강도를 수량화하는 통계측도인 연관성 측도는 일치와 불일치의 개념을 이용한 γ , Kendall의 τ_a , Stuart의 C , Somers의 D 등이 있다[9,11].

잔차와 관련된 가정의 검정을 위해서는 피어슨 잔차($e_i = y_i - \hat{\mu}_i / \sqrt{\hat{\mu}_i}$)를 사용하게 된다[9,11].

3. 최적 공정조건 도출

3.1 분석 절차

본 연구에서 사례로 삼은 P제철소에서 생산하는 제품 중 선급용 TMCP제 중 2개 강종을 중심으로 분석, 간략한 절차는 <그림 2>와 같고, 상세내용은 다음과 같다.

절차1. 후판제품의 내부품질결함에 영향이 있을 것으로 판단되는 설명변수를 아래 <표 1>과 같이 선정하였다. (설명변수들의 변수명과 단위는 모두 약어로 표기)

절차2. 분석에 사용된 자료는 2001.1월부터 4월까지 작업한 실적 자료인 764개이다.

절차3. 종속변수는 3가지 범주로 나타나며 각 범주별 자료 비율은 양품 58%, 2급품 18.1%, 불량품 24%, 독립변수는 35개(연속형 변수 29개, 질적 변수 6개)이다.

절차4. 종속변수와 독립변수간의 로지스틱 회귀모형 설정, 최적공정조건 설정을 위해 회귀모형으로는 일차모형, 이차모형(교호효과 포함), 그리고 특정 독립변수가 다른 독립변수의 작업조건에 영향을 미치는 경우에 대해서이며, 본 사례에서는 주편폭에 따라서 주조 속도의 작업기준이 변하기 때문에 이를 반영하여 분석을 실시 즉, "주조 속도"를 오차(작업기준대비 실적 자료)로 변환하여 분석을 실시하였고, 또 4 가지의 주편폭 중 특정 주편폭(2,150)에 한정하여 분석을 실시하였다.

절차5. 종속변수에 영향을 미치는 주요변수를 도출함에 있어서 다단계 공정에서는 여러 설명변수들이 복잡하게 연관되어 있어 변수 선택 기법인 "Forward Regression" 및

<표 1> 종속변수와 관련된 설명변수

구 분	내 용	
종속변수	UST 검사결과 => 1, 2,	
설명 변수	제 강	소강성분 : C,Mn,Ti,S,P,Si,Cu,Nb,V,Ni,N RH처리시간, RH산소사용량,RH최저진공도
	연 주	Tundish기준온도, 평균온도, 이론응고온도, 탕면변동폭, 주조속도, 2차냉각패턴, 주속변경, 연연주실적, Tundish예열시간, Tundish delta-T(T/D) 실측온도-이론응고온도
	후 판	가열로 재로시간, 압연방법, 마무리 압연온도, 마지막 제어압연 시작온도/종 료온도, 가속냉각 개시온도/종료온도/냉각속도, 마지막 제어압연, 잔압하울, 폭출비, 압하울

"Backward Regression"방법을 이용하여 주요 변수 및 종속변수의 각 값에 대한 예측확률의 수식을 도출하였다.

절차6. 절차5에 의해 1차 로지스틱회귀모형을 변수 간 교호작용, 주요변수와 종속변수의 예측확률과의 관계가 일차식이 아닌 이차식여부를 알아보기 위해 종속변수의 예측확률과 1차 도출된 주요변수와의 관계를 산점도를 통해 확인, 교호작용과 이차효과가 포함된 예측확률의 수식을 2차로 산출하였다.

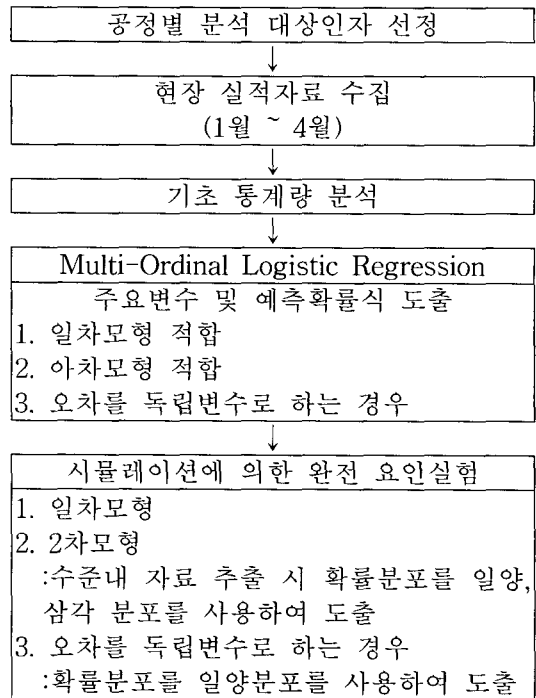
절차7. 오차를 독립변수로 설정하는 경우에 대해서는 Forward Regression방법을 이용하여 주요변수 및 종속변수의 각 값(범주)에 대한 예측확률의 수식을 도출하였다.

절차8. 절차5, 6에서 도출된 주요변수에 대해 최적공정조건을 탐색하기 위하여 주요변수의 조업범위를 3수준으로 범주화하고 시물레이션에 의해 자료를 발생시켜 로지스틱 회귀모형에서 도출된 예측확률의 수식에 대입하여 종속변수의 각 값에 해당하는 예측확률을 산출하였다.

절차9. 종속변수의 각 값으로 인해 판매되는

비용에 가중치를 부여하여 예측확률과의 적(積)으로 계산, 새로운 목적함수(T_p : Total Profit)로 하여 주요변수의 최적조건을 탐색하였다.

통계적 분석을 위해 SAS를 사용하여 분석하였다.



<그림 2> 최적공정조건 분석절차

3.2 로지스틱 회귀모형 선정

3.2.1 1차 모형분석

종속변수에 영향을 끼칠 것으로 생각되는 35개의 설명변수 중 유의한 영향을 주는 변수의 선택을 위해 Backward Regression과 Forward Regression방법을 사용하여 764개의 자료로부터 변수의 유의수준 $\alpha=0.05$ 를 만족하는 조건하의 주요변수를 선정, <표 2>에 Forward Regression결과의 단계별 주요 통계량, 그리고 <표 3>에는 Backward Regression과 Forward Regression결과를 나타내었다.

<표 3> 결과에서 Backward Regression보다 Forward Regression결과가 잔차의 카이제곱 통계량이 다소 크게 나타났으나 관리하여야 할 변수의 수가 일반적으로 적을수록

<표 2> Forward Regression에 의한 진입변수 단계별 통계량

단계	진입변수	모형 적합 통계량			잔차 카이제곱	연관성 C
		우도비	스코어	왈드		
1	X_{29}	88.12	86.67	85.33	144.5	0.678
2	X_{29}, X_{34}	128.64	122.60	118.85	112.5	0.690
3	X_{29}, X_{31}, X_{13}	139.73	129.43	123.76	104.4	0.705
4	$X_{29}, X_{31}, X_{13}, X_{39}$	153.12	138.41	132.20	94.5	0.715
5	$X_{29}, X_{31}, X_{13}, X_{39}, X_{21}$	171.19	152.64	145.00	80.9	0.729
6	$X_{29}, X_{31}, X_{13}, X_{39}, X_{21}, X_{34}$	179.88	158.43	149.49	73.3	0.731
7	$X_{29}, X_{31}, X_{13}, X_{39}, X_{21}, X_{34}, X_{30}$	190.37	166.71	155.37	64.2	0.736

<표 3> 1차 주요변수선정 분석 결과

구분	Backward Reg'	Forward Reg'
설명변수의 개수	35	35
유의인자수(유의수준:0.05)	17	7

모형적합 통계량	우도비	236.7(p=0.0001)	190.4(p=0.0001)
	스코어	194.9(p=0.0001)	166.7(p=0.0001)
	왈드	166.2(p=0.0001)	155.4(p=0.0001)
잔차 카이제곱 검정	카이제곱(χ^2)	22.36	64.24
	유의 확률(p)	0.4384	0.0004
연관성 측도	일치상(%)	75.8	73.4
	C	0.759	0.736

공정제어가 용이하므로 $X_{13}, X_{21}, X_{28}, X_{29}, X_{30}, X_{34}, X_{31}$ 의 7개변수를 1차 주요변수로 선정하였다.

여기서 변수 X_{34} 는 질적 변수로 4개의 가변수를 사용하였다. 1차 분석결과 나타난 통계량을 살펴보면 우선 로지스틱 회귀모형에 관한 SAS분석결과를 보면 절편항, 추정치, 점근적 표준오차, $H_0: \beta_i = 0, H_1: \beta_i \neq 0$ 에 대한 카이제곱통계량 및 유의확률이 <그림 3>에 나타나 있다.

예를 들어 식(1)의 로지스틱회귀모형에서의 모수 α, β 의 추정치와 점근적 표준오차는 X_{13} 변수의 경우 0.0529, 0.0112로 산출 되었으며, 카이제곱 통계량은 $Z^2 = (\hat{\beta})^2 / (ASE)^2 = 0.0529^2 / 0.0112^2 = 22.4607$ 로서 검정결과 극히 유의한 것으로 나타났으며, 7개 설명변수 모두 통계적으로 유의한 것을 알 수 있다. 즉 7개의 독립변수는 유의수준($\alpha=0.05$) 기준으로 로지스틱 회귀모형을 구성함에 있어 필요한 변수임을 의미한다. <그림 3>의 결과의 추정치로부터 식(5)의 로지스틱회귀모형을 설정할 수가 있는데 모형은 아래와 같다.

$$P[Y \leq 1] = 1 + \exp[-(17.5878 + 0.0529(X_{13}) + 0.0066(X_{21})) - 0.0049(X_{28} - 0.1126(X_{29}) - 0.2745(X_{30}) - [0.1249(X_{34_1}) - 0.6544(X_{34_2}) - 0.8436(X_{34_3}) + 2.6198(X_{34_4}) - 0(X_{34_5})] - 0.353(X_{39}))}]^{-1} \tag{10}$$

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	17.5878	3.0007	34.3553	<.0001
Intercept2	1	18.6257	3.0099	38.2929	<.0001
X13	1	0.0529	0.0112	22.4607	<.0001
X21	1	0.00660	0.0017	15.1549	<.0001
X28	1	-0.00488	0.0013	14.0536	0.0002
X29	1	-0.1126	0.0140	64.8821	<.0001
X30	1	-0.2745	0.0871	9.9278	0.0016
X34	1	-0.1249	0.2214	0.3182	0.5727
X34	2	-0.6544	0.2218	8.7061	0.0032
X34	3	-0.8436	0.2231	14.3053	0.0002
X34	4	2.6198	0.4512	33.7107	<.0001
X39	1	-0.3530	0.0842	17.5848	<.0001

<그림 3> 1차 주요변수의 최우추정치

$$\begin{aligned}
 P[Y \leq 2] = & 1 + \exp[-\{18.6257 + 0.0529(X_{13}) + \\
 & 0.0066(X_{21}) - 0.0049(X_{28}) - 0.1126(X_{29}) \\
 & - 0.2745(X_{30}) - [0.1249(X_{34_1}) \\
 & - 0.6544(X_{34_2}) - 0.8436(X_{34_3}) + \\
 & 2.6198(X_{34_4}) - 0(X_{34_5}) - 0.353(X_{39})\}]^{-1}
 \end{aligned}$$

(11)

$$P(Y=3) = 1 - [P(Y \leq 2)]$$

(12)

여기서 X_{34-i} 는 X_{34} 의 가변수이다.

$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ 귀무가설에 대한 우도비, 스코어, 왈드 통계량에 의한 검정결과 <그림 4>와 같이 모두 극히 유의한 것으로 나타났고, 예측확률과 관찰된 반응치와의 연관성측도는 <그림 5>를 보면 일치쌍의 비율이 전체의 73.4%를 차지하고, 통계량은 Stuart의 C가 0.736으로 비교적 상관이 있는 것으로 나타났다. 즉 1차 선형 로지스틱 회귀모형에서 설정된 종속변수의 수식결과 예측확률과 관찰된 반응치 간에 상관이 비교적 큰 것으로 나타났다.

여기서 2차모형의 가능성을 파악하기 위해 도출된 수식의 예측확률과 각 주요변수

간의 이차함수 여부를 알아보기 위해 산점도는 <그림 6>과 같이 나타났으며, X_{39} 변수를 포함한 몇 개의 변수가 예측확률과의 관계가 이차함수 형태를 가진 것을 볼 수 있다.

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	190.3686	10	<.0001

<그림 4> $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ 검정통계량

Association of Predicted Probabilities and Observed Responses

Percent Concordant	73.4	Somers' D	0.472
Percent Discordant	26.3	Gamma	0.473

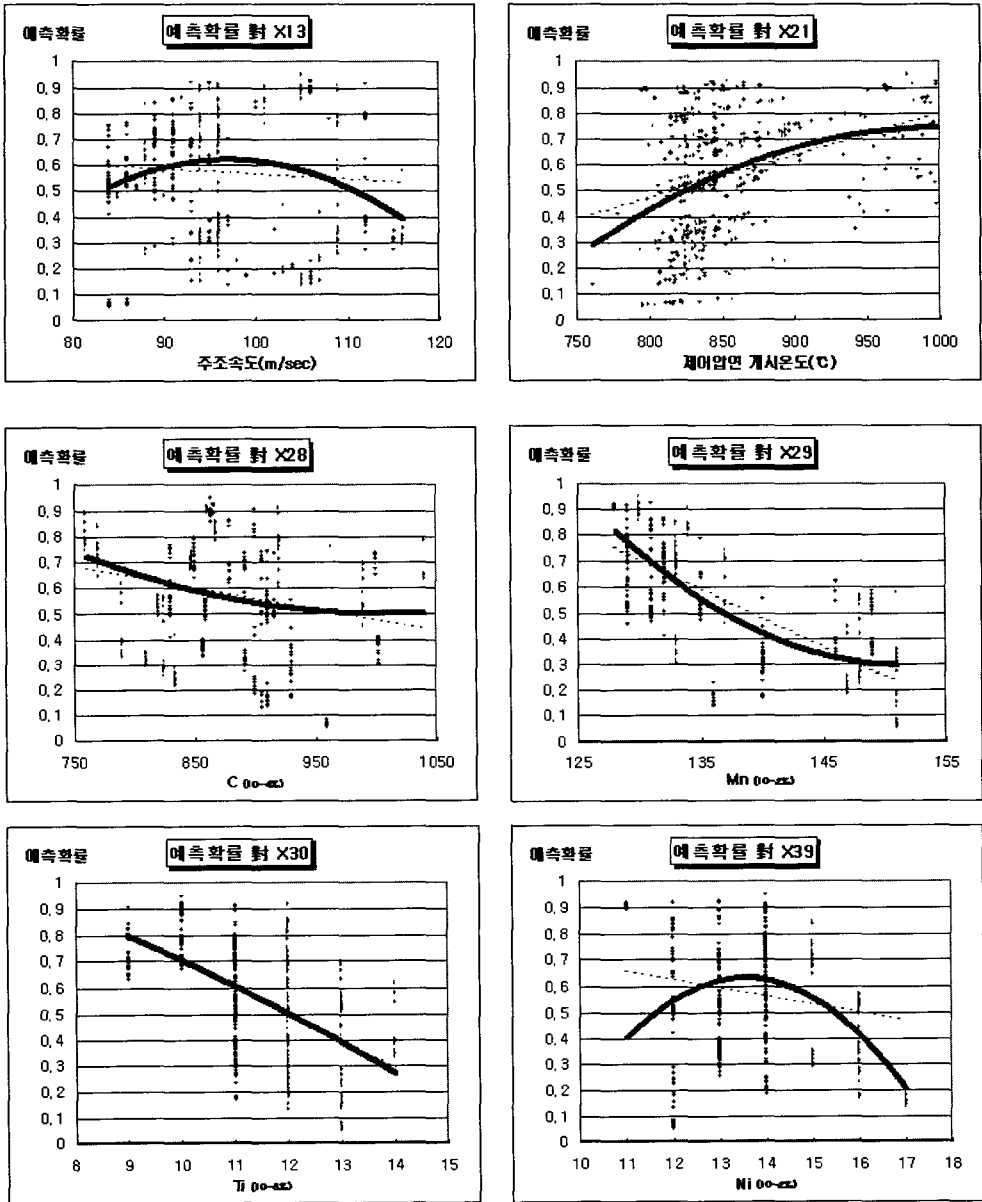
<그림 5> 예측확률과 관측치의 연관성측도

3.2.2 2차 모형분석

1차 선형로지스틱 회귀모형 결과 선정된 7개의 주요설명 변수는 산점도 분석결과 예측확률과 몇몇 주요 변수 간에 2차 함수관계결과를 볼 수 있었으므로 선정된 주요변수들의 이차식과 교호효과를 고려하여 분석을 해보면, 7개의 변수들 중 질적 변수를 제외한 6개 변수의 교호작용 및 이차형태의 변수를 포함시켜 전체 28개 변수를 모형에 넣어서 Forward Regression을 실시한 결과 위의 <표 4>같은 결과를 얻었다.

2차 분석결과 나타난 통계량을 살펴보면 우선 로지스틱회귀모형에 관한 결과가 <그림 7>에 나타나 있다.

<그림 7>의 추정치로부터 로지스틱회귀모형을 표현하면 아래와 같다.



<그림 6> 예측확률과 주요변수간의 산점도

<표 4> 2차 모형에서의 단계별 주요통계량

단계	진입변수	모형적합통계량			잔차 카이 제곱	연관성 C
		우도비	스코어	왈드		
1	X ₂₉	88.12	86.67	85.33	123.1	0.678
2	X ₂₁ , X ₃₄	128.64	122.60	118.85	89.6	0.694
3	X ₂₉ , X ₃₄ , X ₁₃	139.73	129.43	123.76	81.9	0.705
4	X ₂₉ , X ₃₄ , X ₁₃ , X ₃₀ ²	154.73	139.61	133.38	70.0	0.715
5	X ₂₉ , X ₃₄ , X ₁₃ , X ₃₀ ² , X ₂₁	172.91	153.92	146.18	56.3	0.730
6	X ₂₉ , X ₃₄ , X ₁₃ , X ₃₀ ² , X ₂₁ , X ₃₉	185.01	162.98	153.00	46.7	0.732
7	X ₂₉ , X ₃₄ , X ₁₃ , X ₃₀ ² , X ₂₁ , X ₃₉ , X ₂₀	192.19	169.42	157.26	39.3	0.738
8	X ₂₉ , X ₃₄ , X ₁₃ , X ₃₀ ² , X ₂₁ , X ₃₉ , X ₂₀ , X ₂₈	200.76	174.47	159.72	30.5	0.741
9	X ₂₉ , X ₃₄ , X ₁₃ , X ₃₀ ² , X ₂₁ , X ₃₉ , X ₂₀ , X ₂₈ , X ₂₉ × X ₂₉	207.72	178.01	153.28	23.7	0.741
10	X ₂₉ , X ₃₄ , X ₁₃ , X ₃₀ ² , X ₂₁ , X ₃₉ , X ₂₀ , X ₂₈ , X ₂₉ × X ₂₉ , X ₁₃ × X ₂₀	212.89	180.44	152.90	19.2	0.742
11	X ₂₉ , X ₃₄ , X ₁₃ , X ₃₀ ² , X ₂₁ , X ₃₉ , X ₂₀ , X ₂₈ , X ₂₉ × X ₂₉ , X ₁₃ × X ₃₀ , X ₁₃ × X ₂₁	217.92	183.99	155.74	14.7	0.745

종속변수의 각 값에 대한 예측확률은 로지스틱 회귀모형에서 귀무가설 $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ 에 대한 우도비, 스코어, 왈드 통계량에 의한 검정결과 <표 4>의 마지막 단계의 세 통계량 모두 극히 유의한 것으로 나타났다.

$$\begin{aligned}
 P[Y \leq 1] = & [1 + \exp\{-117.2 - 0.0310(X_{13}) - 0.0407(X_{21}) \\
 & + 0.1162(X_{28}) + 0.6947(X_{29}) \\
 & + 2.3775(X_{30})[-0.8151(X_{34_1}) - 1.0627(X_{34_2}) \\
 & - 1.0392(X_{34_3}) + 1.6892(X_{34_4}) - 0(X_{34_5})] \\
 & + 5.0645(X_{39}) - 0.1983(X_{39}^2) + 0.0005(X_{13} \times X_{21}) \\
 & - 0.0292(X_{13} \times X_{30}) - 0.0009(X_{28} \times X_{29})\}]^{-1}
 \end{aligned}$$

(13)

$$\begin{aligned}
 P[Y \leq 2] = & [1 + \exp\{-116.2 - 0.0310(X_{13}) - 0.0407(X_{21}) \\
 & + 0.1162(X_{28}) + 0.6947(X_{29}) \\
 & + 2.3775(X_{30})[-0.8151(X_{34_1}) - 1.0627(X_{34_2}) \\
 & - 1.0392(X_{34_3}) + 1.6892(X_{34_4}) - 0(X_{34_5})] \\
 & + 5.0645(X_{39}) - 0.1983(X_{39}^2) + 0.0005(X_{13} \times X_{21}) \\
 & - 0.0292(X_{13} \times X_{30}) - 0.0009(X_{28} \times X_{29})\}]^{-1}
 \end{aligned}$$

(14)

그리고 잔차의 카이제곱 검정에서는 잔차의 카이제곱 통계량이 14.7382로서 유의(p>0.6143)하지 않는 것으로 나타났다. 예측 확률과 관찰된 반응치와의 연관성측도에 대해 <그림 8>를 보면 일치쌍이 전체의 74.3%를 차지하고, 통계량은 Stuart의 C가 0.745로 비교적 상관이 있는 것으로 나타났다.

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard	Chi-Square	Pr > ChiSq
			Error		
Intercept	1	-117.2	44.4150	6.9678	0.0083
Intercept2	1	-116.2	44.4120	6.8418	0.0089
X13	1	-0.0310	0.2335	0.0176	0.8945
X21	1	-0.0407	0.0213	3.6490	0.0561
X28	1	0.1162	0.0378	9.4224	0.0021
X29	1	0.6947	0.2494	7.7613	0.0053
X30	1	2.3775	1.1495	4.2779	0.0386
X34 1	1	-0.8151	0.3080	7.0016	0.0081
X34 2	1	-1.0627	0.3211	10.9505	0.0009
X34 3	1	-1.0392	0.2564	16.4299	<.0001
X34 4	1	1.6892	0.5646	8.9522	0.0028
X39	1	5.0645	1.6323	9.6267	0.0019
XX39	1	-0.1983	0.0602	10.8516	0.0010
X13*X21	1	0.000497	0.000226	4.8580	0.0275
X13*X30	1	-0.0292	0.0121	5.7858	0.0162
X28*X29	1	-0.00091	0.000284	10.2057	0.0014

<그림 7> 2차 주요변수의 최대우도추정치

Association of Predicted Probabilities and Observed Responses

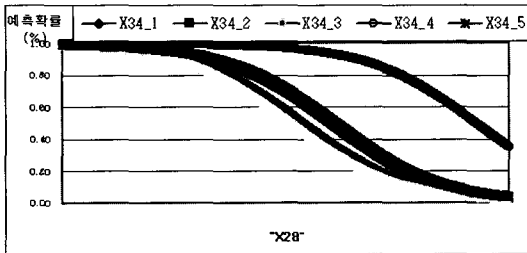
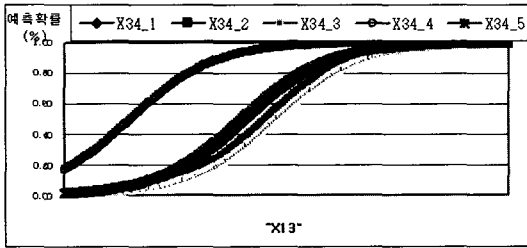
Percent Concordant	74.3	Somers' D	0.489
Percent Discordant	25.3	Gamma	0.491

<그림 8> 예측확률과 관측치의 연관성측도

1,2차에 걸쳐서 다범주-순서형 종속변수에 대한 로지스틱 회귀분석결과 7개의 주요 변수가 종속변수에 영향을 미치는 것으로 나타났으며, 1차 2차 모형에 대한 예측확률의 수식모형이 설정되는 것을 보았다.

2차 로지스틱 회귀모형에서 도출된 주요 변수 중 예측확률(P(Y=I))과 양(+), 음(-)

의 관계를 가지고 있는 X_{13} , X_{28} 변수에 대해 질적 변수인 X_{34} 의 각 값에 대해 로지스틱 회귀곡선을 그려 보면 <그림 9>와 같은 결과를 얻을 수가 있다.



<그림 9> 변수와 예측확률간의 로지스틱 회귀곡선

X_{13} 변수의 경우 예측확률과 단조증가의 곡선형태를 가지고 있으며 작업범위 즉, 변수의 범위내에서는 비교적 선형의 형태를 보이고 있으나, X 축에 대해 작업범위를 넓게 하였을 경우 그림의 형태가 로지스틱 회귀곡선인 S-곡선의 형태를 보이고 있음을 알 수 있다. 그리고 X_{34} 질적 변수에 대해 5개의 그림 형태가 나타나며, $X_{34.4}$ 인 경우 다른 범주값에 비해 예측확률이 높은 것을 알 수 있다.

3.2.3 오차(기준값-실적값)를 독립변수로 하는 모형분석

작업공정에 있어 주편폭에 따른 주조 속

도의 작업기준이 다르므로 이를 감안하여 주조 속도의 실적 자료값을 기준값 대비 실적값의 차이 즉, 오차(변수명: R_{X13})를 주조속도 변수로 대체하여 로지스틱 회귀분석을 실시하였다. 주편폭에 따른 주조 속도의 기준은

주편폭 (mm)	1551~1700	1701~2000	2001~2200
주조속도 (cm/min)	120	110	100

이다.

<표 5> Forward Regression에 의한 진입변수 단계별 통계량

단계	진입변수	모형적합통계량			잔 차	연관성 C
		우도비	스코어	왈드		
1	X_{29}	85.39	84.08	82.92	152.8	0.676
2	X_{29}, X_{34}	126.92	120.88	117.25	120.5	0.692
3	X_{29}, X_{34}, R_{X13}	158.47	143.94	135.67	96.7	0.713
4	$X_{29}, X_{34}, R_{X13}, X_{28}$	179.33	156.97	143.41	80.0	0.728
5	$X_{29}, X_{34}, R_{X13}, X_{29}, X_{21}$	191.47	166.39	149.72	70.4	0.736
6	$X_{29}, X_{34}, R_{X13}, X_{29}, X_{21}, X_{11}$	197.89	168.99	148.96	65.0	0.736
7	$X_{29}, X_{34}, R_{X13}, X_{29}, X_{21}, X_{11}, X_{11}$	206.51	174.37	151.00	57.3	0.741
8	$X_{29}, X_{34}, R_{X13}, X_{29}, X_{21}, X_{11}, X_{14}, X_{25}$	210.94	179.16	156.15	53.0	0.742
9	$X_{29}, X_{34}, R_{X13}, X_{29}, X_{21}, X_{11}, X_{14}, X_{25}, X_{29}$	216.60	184.44	163.36	47.5	0.744
10	$X_{29}, X_{34}, R_{X13}, X_{29}, X_{21}, X_{11}, X_{14}, X_{25}, X_{33}, X_9$	223.05	187.06	161.15	42.0	0.749

Forward Regression 방법을 사용하여 주요변수를 선정하였으며, <표 5>에 단계별 주요 통계량을 나타내었다.

그리고 도출된 주요변수의 최우 추정치 통계량이 아래 <그림 10>에 나타나 있다.

예측확률로 표현하면 식 (15), (16)과 같다.

$$\begin{aligned}
 P[Y \leq 1] = & 1 + \exp\{-[26.1564 - 0.0025(X_9) - 0.099(\text{Residual}_{X_{13}}) \\
 & - [0.610(X_{14_1}) - 0.722(X_{14_2}) - 0.718(X_{14_3}) \\
 & - 0(X_{14_4})] + 0.0044(X_{21}) + 0.4231(X_{25}) - 0.0056(X_{28}) \\
 & - 0.0998(X_{29}) - [0.7946(X_{34_1}) + 0.4966(X_{34_2}) \\
 & - 0.5137(X_{34_3}) - 2.1788(X_{34_4}) - 0(X_{34_5})] \\
 & - 0.2347(X_{39}) - 0.0415(X_{41})\} \}
 \end{aligned}$$

(15)

$$\begin{aligned}
 P[Y \leq 2] = & 1 + \exp\{-[27.2306 - 0.0025(X_9) - \\
 & 0.099(\text{Residual}_{X_{13}}) - [0.610(X_{14_1}) \\
 & - 0.722(X_{14_2}) - 0.718(X_{14_3}) - 0(X_{14_4})] \\
 & + 0.0044(X_{21}) + 0.4231(X_{25}) - 0.0056(X_{28}) \\
 & - 0.0998(X_{29}) - [0.7946(X_{34_1}) + 0.4966(X_{34_2}) \\
 & - 0.5137(X_{34_3}) - 2.1788(X_{34_4}) - 0(X_{34_5})] \\
 & - 0.2347(X_{39}) - 0.0415(X_{41})\} \}
 \end{aligned}$$

(16)

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	26.1564	4.2671	37.5751	<.0001
Intercept2	1	27.2306	4.2755	40.5632	<.0001
X9	1	-0.00249	0.00101	6.1052	0.0135
RESI_X13	1	-0.0990	0.0163	36.9678	<.0001
X14	1	-0.6097	0.4063	2.2519	0.1334
X14	2	0.7217	0.4586	2.4761	0.1156
X14	3	0.7178	0.8419	0.7269	0.3939
X21	1	0.00439	0.00192	5.2102	0.0225
X25	1	0.4231	0.1375	9.4717	0.0021
X28	1	-0.00560	0.00140	15.9306	<.0001
X29	1	-0.0998	0.0162	37.7752	<.0001
X34	1	-0.7946	0.4723	2.8301	0.0925
X34	2	-0.4966	0.2459	4.0799	0.0434
X34	3	0.5137	0.2706	3.6046	0.0576
X34	4	2.1788	0.4761	20.9456	<.0001
X39	1	-0.2347	0.0859	7.4697	0.0063
X41	1	-0.0415	0.0154	7.2715	0.0070

<그림 10> 주요변수의 최우 추정치

Association of Predicted Probabilities and Observed Responses

Percent Concordant 74.8 Somers' D 0.498
 Percent Discordant 25.0 Gamma 0.499

<그림 11> 예측확률과 관측치의 연관성측도

$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ 에 대한 검정결과 <표 5>의 마지막에 나타난 세 통계량 모두 극히 유의한 것으로 나타났다. 가정된 모형의 연관성측도결과가 아래 <그림 11> 과 같이 나타났다.

앞 절에서 언급된 주편폭에 따른 주조 속도의 작업기준 상이(相異)에 대해 특정한 주편폭을 표본으로 하여 해당 주편폭에 대한 자료추출, 로지스틱 회귀모형을 설정하였다. 주편폭이 1600, 1900, 2000, 2150(mm)에 대해 각각 134, 135, 261, 234개의 자료를 대상으로 분석을 실시하였다.

Forward Regression방법을 사용하여 주요 변수를 선정하였으며, 아래 <표 6>에 단계별 주요 통계량을 나타내었다.

Forward Regression 실시결과 도출된 주요 변수의 최우 추정치 통계량이 <그림 12>에 나타나 있다.

중속변수의 각 값에 대해 예측확률로 표현하면 다음과 같다.

<표 6> Forward Regression에 의한 진입변수 단계별 통계량

단계	진입변수	모형적합통계량		잔 차 제곱	연관성 C	
		우도비	스코어			왈드
1	X ₂₁	85.39	84.08	82.92	152.8	0.676
2	X ₂₁ , X ₉	126.92	120.88	117.25	120.5	0.692
3	X ₂₁ , X ₉ , X ₁₃	158.47	143.94	135.67	96.7	0.713
4	X ₂₁ , X ₉ , X ₁₃ , X ₁₁	179.33	156.97	143.41	80.0	0.728
5	X ₂₁ , X ₉ , X ₁₃ , X ₁₁ , X ₃₂	191.47	166.39	149.72	70.4	0.736
6	X ₂₁ , X ₉ , X ₁₃ , X ₁₁ , X ₃₂ , X ₃₇	197.89	168.99	148.96	65.0	0.736

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-2.5149	7.9472	0.1001	0.7517
Intercept2	1	-1.2718	7.9481	0.0256	0.8729
X9	1	-0.0030	0.0009	12.1492	0.0005
X11	1	1.5931	0.4204	14.3586	0.0002
X13	1	0.2620	0.0727	12.9741	0.0003
X29	1	-0.2429	0.0430	31.8812	<.0001
X32	1	0.4622	0.1261	13.4354	0.0002
X37	1	0.5190	0.1797	8.3365	0.0039

<그림 12> 주요변수의 최우 추정치

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	69.6353	6	<.0001

<그림 13> $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ 검정통계량

$$P[Y \leq 1] = 1 + \exp[-\{-2.5149 - 0.00302(X_9) + 1.5931(X_{11_1}) + 0(X_{11_2}) + 0.262(X_{13}) - 0.2429(X_{29}) + 0.4622(X_{32}) + 0.5190(X_{37})\}]^{-1} \tag{17}$$

$$P[Y \leq 2] = 1 + \exp[-\{-1.2718 - 0.00302(X_9) + 1.5931(X_{11_1}) + 0(X_{11_2}) + 0.262(X_{13}) - 0.2429(X_{29}) + 0.4622(X_{32}) + 0.5190(X_{37})\}]^{-1} \tag{18}$$

$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ 귀무가설에 대한 검정결과 <그림 13> 같이 세 통계량 모두 극히 유의한 것으로 나타났다.

Association of Predicted Probabilities and Observed Responses

Percent Concordant	73.2	Somers' D	0.494
Percent Discordant	23.8	Gamma	0.510

<그림 14> 예측확률과 관측치의 연관성 측도

가정된 모형의 적합성에 대한 잔차의 카이제곱검정에서 잔차의 카이제곱통계량은 21.3326 (p 값=0.6739) 으로 유의하지 않으며, 예측확률과 관찰된 반응치와의 연관성에서는 <그림 14>에서와 같이 일치쌍의 비율이 73.2%, Stuart의 C는 0.747로 상관이 큰 것으로 나타났다.

4. 시물레이션을 이용한 최적공정조건

4.1 2차 모형에서의 최적조건

3.2절에서 개발된 로지스틱 회귀모형을 기초로 하여 주요변수의 최적공정조건을 찾고자 한다. 최적공정조건을 찾기 위해서 종속변수별 제품의 판매단가를 기준으로 하여 목적함수식을 세우고, 목적함수를 최대로 하는 공정조건을 찾는 데 목적이 있다. 그러나 반응변수의 값이 3개이고 설명변수 역시 설정된 수준에서 편차가 있으므로 해석적으로 목적함수를 최대로 하는 공정조건을 찾기가 힘들며 따라서 본 논문에서는 시물레이션을 이용하여 요인배치법에 따라 실험 자료를 추출하여 로지스틱 회귀모형에서 도출된 세 가지 경우에 대해 아래 절차대로 분석을 실시하였다.

절차1. 도출된 주요설명변수에 있어 연속형 변수는 작업범위 내에서 일정한 간격(등간격)에 의해 3 수준으로 분류하였다.

절차2. 2차 모형분석에서 도출된 모형에 대해서는 변수의 각 수준 내에서 실제 발생할 수 있는 일양분포와 삼각분포(각 수준의 실제값에 대한 변동을 현장전문가로부터 간단히 예측하기 위함)의 확률분포를 이용하여 모든 수준별로 100회씩 자료를 발생시켰다.

절차3. 발생된 각 주요변수의 수준별 자료를 로지스틱 회귀모형에서의 예측확률모형에 대입시켜 예측확률을 도출하였다.

절차4. 절차 3에서 도출된 예측확률에 각 종속변수가 가지는 판매단가를 가중치로 적용하여 새로운 종속변수(T_p : Total Profit) 값을 목적함수로 두는데 식 (19)와 같다. (관측치는 3연속변수의 개수×범주 질적 변수)

$$T_p = \sum_{k=1}^3 C_k \cdot P(Y=k) \quad (19)$$

$k:1,2,3$ (종속변수의 각 값)

C_k : 종속변수 각 값의 판매단가

$P(Y=k)$: 종속변수 각 값의 성공확률

절차5. 도출된 관측치에 대해 분산분석을 실시하여 유의한 변수를 도출하고, 유의한 변수의 초기 최적공정조건을 선정하였다.

절차6. 선정된 초기 최적수준에 대해 만일 설명변수의 변동단위를 줄일 경우의 효과를 파악하기 위하여 절차 5에서 도출된 각 변수별 최적조건을 다시 일정한 간격으로 3수준으로 분류, 절차 2~5의 방법으로 다시 세부 분석을 실시하여 최적공정조건을 도출하였다.

절차7. 세 번째 모형에 대해서 일양분포의 확률분포를 이용하여 절차대로 분석을 실시하여 최적조건을 도출하였다.

로지스틱 회귀모형에서 도출된 7개 주요 설명변수에 대해 우선 연속형 변수(6개)에 대해 조업조건 범위 내에서 <표 7>과 같이 균등하게 3수준으로 나누었다.

<표 7> 주요변수의 수준 분류

주요 변수	수 준				
	1	2	3	4	5
X ₁₃	84~95	96~105	106~116		
X ₂₁	761~834	835~907	908~980		
X ₂₈	760~853	854~947	948~1040		
X ₂₉	128~135	136~143	144~151		
X ₃₀	9.0~10.6	10.7~12.3	12.4~14.0		
X ₃₉	11.0~13.0	13.1~15.0	15.1~17.0		
X ₃₄	1	2	3	4	5

각 설명변수의 수준에 대해서 발생하는 실제 설명변수 값의 확률분포는 일양분포와 삼각분포를 이용하여 각각 모든 수준에서 100개씩 발생시켰다.

주요설명 변수의 각 수준에서의 실제 설명 변수값을 발생시켜 앞 절에서 개발된 로지스틱 회귀모형에서 도출된 종속변수의 각 값(범주:1,2,3)의 예측확률 수식에 적용하여 나온 확률 값들의 평균(P₁평균, P₂평균, P₃평균)에 종속변수의 각 값에 대한 판매단가(1=>373천원, 2=>289.6천원, 3=>147천원)를 곱하여 수치변환(T_p/백만)시켜 각 관찰 치에 대해 T_p 값을 구한다. 이 경우 7개의 인자에 대한 요인배치법을 통해 분산 분석을 실시하여 최적 조건을 구한다. 먼저 2차 모형에서 일양분포를 가정하는 경우에 대한 분산분석결과를 살펴보면, <표 8>에서 나타난 바와 같이 주요변수인 7개 변수와 교호작용변수 모두 유의한 것으로 나타났으나, 기여율측면에서는 교호작용변수가 다른 주요변수보다도 매우 낮음을 볼 수 있다. 비교적 제곱합이 상대적으로 작은 변수들을 오차항에 풀링(Pooling)시켜서 분산분석을 실시하였다<표 9>.

<표 8> 주요변수별 분산분석표 (2차-일양분포)

변수	제곱합 (S.S)	자유도 (Φ)	평균제곱 (V)	분산비 (F ₀)	순변동 (S')	기여율 (ρ)
X ₁₃	1.1299	2	0.5649	950.1**	1.1287	6.7
X ₂₁	1.0667	2	0.5334	897.0**	1.0656	6.3
X ₂₈	2.4252	2	1.2126	2039.3**	2.4240	14.3
X ₂₉	2.1889	2	1.0945	1840.6**	2.1878	12.9
X ₃₀	1.7491	2	0.8745	1470.7**	1.7479	10.3
X ₃₄	3.4158	4	0.8539	1436.1**	3.4134	20.1
X ₃₉	2.8421	2	1.4211	2389.9**	2.8410	16.7
X ₁₃ × X ₂₁	0.1781	4	0.0445	152.1**	0.1769	1.0
X ₁₃ × X ₃₀	0.2775	4	0.0693	236.9**	0.2763	1.6
X ₂₈ × X ₂₉	0.6428	4	0.1607	548.7**	0.6416	3.8
e	2.1573	3616	0.0003		1.0671	6.3
Total	16.9750	3644			16.9750	100.0

수준별 평방합을 살펴보면 X_{13} , X_{21} 은 3 수준일 때 X_{28} , X_{29} , X_{30} 은 1수준, X_{34} 는 4수준, X_{39} 는 1, 2수준일 때 최적인 것을 알 수 있다.

주요변수별 최적수준에 대해 보다 세밀한 최적조건을 탐색하기 위한 수준은 <표 10>와 같다.

<표 9> 주요변수별 분산분석표 및 수준별 평방합(2차-일양분포)

변수	제곱합 (S.S)	자유도 (f)	평균제곱 (V)	분산비 (Fo)	순변동 (S')	기여율 (p)
X_{28}	2.4252	2	1.2126	808.4**	2.4222	14.3
X_{29}	2.1889	2	1.0945	729.7**	2.1859	12.9
X_{30}	1.7491	2	0.8745	583.0**	1.7461	10.3
X_{34}	3.4158	4	0.8539	569.3**	3.4128	20.1
X_{39}	2.8421	2	1.4211	947.4**	2.8391	16.7
(e)	5.4523	3632	0.0015		4.3689	25.7
Total	16.9750	3644			16.9750	100.0

구분	X_{13}	X_{21}	X_{28}	X_{29}	X_{30}	X_{34}	X_{39}
수준합	1 수준	305.5	306.8	371.5	369.6	364.5	187.4
	2 수준	334.7	333.5	331.7	331.6	334.1	181.4
	3 수준	357.8	357.7	294.7	296.7	299.3	182.1
	4 수준						240.2
	5 수준						206.8
계	998.0	998.0	998.0	998.0	998.0	998.0	

<표 10> 변수의 1차 최적 조건내에서의 3 수준분류

주요변수	수준		
	1	2	3
X_{13}	106~109	110~113	114~116
X_{21}	908~932	933~956	957~980
X_{28}	760~792	793~823	824~853
X_{29}	128.0~130.3	130.4~132.6	132.7~135.0
X_{30}	9.0~9.5	9.6~10.0	10.1~10.6
X_{39}	11.0~11.6	11.7~12.3	12.4~13.0
X_{34}	1	2	3
			4
			5

이 각 변수의 최적조건에서 다시 수준을 3 수준으로 분류하였다. 그리고 동일한 절차를 따라 분석한 결과(<표 11> 참조) X_{28} , X_{29} 및 교호작용변수는 평방합이 작아 오차항에 풀링시켰으며, X_{34} , X_{39} 변수의 기여율이 50%를 차지하는 것으로 나타났으며, 수

준별 평방합을 살펴보면 X_{34} 가 4수준, X_{39} 는 1수준"에서 최적인 것을 알 수 있다.

같은 방법으로 삼각분포를 사용한 경우를 살펴보면, <표 7>과 같이 주요변수별 수준을 일양분포와 같이 나누고, 관측치를 3,645 회 시뮬레이션에 의한 분산분석을 실시한 결과가 <표 12>에 나타나있다. 삼각분포에 의한 분석 결과에서도 일양분포 때와 같이 X_{13} , X_{21} 및 교호작용변수의 평방합이 상대적으로 작아 오차항에 풀링시켰으며, 그 외 5개 변수는 모두 유의한 것으로 나타났고, 변수별 기여율은 일양분포 때와 마찬가지로 거의 비슷하게 나타났다. 수준별 평방합을 살펴보면 X_{13} , X_{21} 은 3수준일 때 X_{28} , X_{29} , X_{30} 은 1수준, X_{34} 는 4수준, X_{39} 는 1, 2수준"일 때 최적인 것을 알 수 있다.

<표 7>에서 나타난 주요변수별 최적수준에 대해 보다 세밀한 최적조건을 탐색하기 위하여 각 변수의 최적조건에서 다시 수준을 3수준으로 분류하여 분산분석을 실시하여 <표 13> 과 같은 결과를 얻었다.

일양분포 때와 마찬가지로 주요변수 X_{28} , X_{29} , 교호작용 변수 3개 변수의 평방합이 상대적으로 작아 오차항에 풀링시켰으며, X_{34} , X_{39} 변수의 기여율이 50%를 차지하는 것으로 나타났으며, 수준별 평방합을 살펴 보면 X_{34} 가 4수준, X_{39} 는 1수준"에서 최적인 것을 알 수 있다.

4.2 오차를 독립변수로 하는 모형에서의 최적조건

로지스틱 회귀모형에서 도출된 10개 주요 변수에 대해 수준별 유의성 여부를 분산분석을 통해 검증, 최적공정조건을 도출하기 위하여 앞에서 언급한 절차대로 분석을 실

시하였다. 우선 연속형 변수(8개)에 대해 작업조건범위 내에서 <표 14>와 같이 균등하게 3수준으로 나누었다.

<표 11> 주요변수별 분산분석표 및 수준별 평방합

변수	제곱합 (S.S)	자유도 (φ)	평균제곱 (V)	분산비 (Fo)	순변동 (S')	기여율 (ρ)
X ₂₈	2.5172	2	1.2586	985.1**	2.5146	14.2
X ₂₉	2.2707	2	1.1353	888.6**	2.2671	12.8
X ₃₀	1.8132	2	0.9066	709.6**	1.8105	10.2
X ₃₄	3.5054	4	0.8764	686.0**	3.5003	19.8
X ₃₉	2.9681	2	1.4841	1161.6**	2.9655	16.8
(e)	4.6402	3632	0.0013		4.6568	26.2
Total	17.7148	3644			17.7148	100.0

구 분	X ₁₃	X ₂₁	X ₂₈	X ₂₉	X ₃₀	X ₃₄	X ₃₉	
수 준 합	1 수준	306.3	307.6	373.5	371.5	366.3	188.0	361.8
	2 수준	336.0	334.9	333.0	332.9	335.5	181.9	354.9
	3 수준	359.5	359.3	295.3	297.3	300.0	182.6	285.0
	4 수준						241.5	
	5 수준						207.8	
계	1001.7	1001.7	1001.7	1001.7	1001.7	1001.7	1001.7	

<표 12> 주요변수별 분산분석표 및 수준별 평방합

변수	제곱합 (S.S)	자유도 (φ)	평균제곱 (V)	분산비 (Fo)	순변동 (S')	기여율 (ρ)
X ₁₃	0.0267	2	0.0134	333.1**	0.0267	6.0
X ₂₁	0.0135	2	0.0067	168.1**	0.0134	3.0
X ₃₀	0.0301	2	0.0151	375.1**	0.0300	6.7
X ₃₄	0.0652	4	0.0163	405.7**	0.0650	14.5
X ₃₉	0.1667	2	0.0834	207.63**	0.1666	37.2
(e)	0.1458	3632	0.0000		0.1463	32.7
Total	0.4480	3644			0.4480	100.0

구 분	X ₁₃	X ₂₁	X ₂₈	X ₂₉	X ₃₀	X ₃₄	X ₃₉	
수 준 합	1 수준	439.6	440.9	444.8	444.7	447.8	265.0	451.1
	2 수준	444.5	444.2	443.9	443.9	444.5	263.4	448.2
	3 수준	447.6	446.6	442.9	443.1	439.4	263.6	432.4
	4 수준						271.3	
	5 수준						268.6	
계	1331.7	1331.7	1331.7	1331.7	1331.7	1331.7	1331.7	

각 변수의 수준에 대해서 발생하는 자료의 확률분포는 앞 절의 결과에서 일양분포 및 삼각분포에 따라 결과가 거의 유사하게 나타났으므로 본 실험에서는 일양분포만 이용하였으며, 각 수준에서 100개씩 발생시켜 총 131,220회(38×41×51) 시뮬레이션에 의한 요인배치법에 의한 조건에서 자료를 발생시

켜 분산분석을 실시한 결과 다음 <표 15>과 같은 결과를 얻었다.

<표 13> 주요변수별-수준별 분산분석표 및 평방합

변수	제곱합 (S.S)	자유도 (φ)	평균제곱 (V)	분산비 (Fo)	순변동 (S')	기여율 (ρ)
X ₁₃	0.0247	2	0.0123	334.2**	0.0246	6.1
X ₂₁	0.0124	2	0.0062	168.6**	0.0124	3.1
X ₃₀	0.0277	2	0.0139	375.5**	0.0276	6.9
X ₃₄	0.0586	4	0.0147	397.2**	0.0585	14.5
X ₃₉	0.1453	2	0.0726	1968.9**	0.1452	36.1
(e)	0.1340	3632	0.0000		0.1345	33.4
Total	0.4027	3644			0.4027	100.0

구 분	X ₁₃	X ₂₁	X ₂₈	X ₂₉	X ₃₀	X ₃₄	X ₃₉	
수 준 합	1 수준	448.3	448.3	449.6	450.1	449.7	269.7	449.9
	2 수준	449.1	449.1	449.1	449.1	449.1	268.2	449.1
	3 수준	449.8	449.7	448.4	447.8	448.3	267.5	448.1
	4 수준						271.5	
	5 수준						270.0	
계	1347.1	1347.1	1347.1	1347.1	1347.1	1347.1	1347.1	

<표 15>에서 보는 바와 같이 분산분석결과 10개 변수 모두 극히 유의(분산비:***)한 것으로 나타났으며, 기여율을 보면 X₃₄(22.3%), Resi_X₁₃(22.2%) 순인 것을 알 수 있다. 변수의 최적조건을 찾기 위해 수준별 평방합의 추이를 살펴보면 X₉, X₂₈, X₂₉, X₃₉, X₄₁변수는 1수준, X₁₄는 2,3수준, X₂₁, X₂₅는 3수준, X₃₄는 4수준, 즉 값이 4일 때 최적인 것을 알 수 있다. Resi_X₁₃변수의 경우 1수준일 때 가장 좋으므로 비교적 잔차가 적을수록, 즉 주변폭에 대한 주조 속도의 실제 값이 기준값에 근접할수록 좋은 결과가 나타나는 것을 알 수 있다.

주변폭(2,150mm) 234개 자료의 로지스틱 회귀모형에서 도출된 6개 주요변수에 대해 수준별 유의성 여부를 분산분석을 통해 검증, 최적공정조건을 도출하기 위하여 연속형변수(5개)에 대해 작업조건 범위 내에서 아래 <표 16>과 같이 균등하게 3수준으로 나누었다.

<표 14> 주요변수의 수준분류

주요변수	수 준				
	1	2	3	4	5
X ₉	1512~1813		1814~2115		2116~2416
Resi_X ₁₃	0~12		13~24		25~36
X ₂₁	761~834		835~907		908~980
X ₂₅	0.0~1.7		1.8~3.4		3.5~5.2
X ₂₈	760~853		854~947		948~1040
X ₂₉	128~135		136~143		144~151
X ₃₉	11.0~13.0		13.1~15.0		15.1~17.0
X ₄₁	34.0~41.6		41.7~49.3		49.4~57.0
X ₁₄	1	2	3	4	5
X ₃₄	1	2	3	4	5

<표 15> 주요변수별-수준별 분산분석표 및 평방향

변수	제공합 (S.S)	자유도 (φ)	평균제공 (V)	분산비 (F ₀)	순변동 (S')	기여율 (ρ)
X ₉	0.42	2	0.21	54152**	0.42	8.5
Resi_X ₁₃	1.09	2	0.54	140144**	1.09	22.2
X ₁₄	0.36	3	0.12	30734**	0.36	7.2
X ₂₁	0.09	2	0.04	11318**	0.09	1.7
X ₂₅	0.40	2	0.20	51379**	0.40	4.9
X ₂₈	0.20	2	0.10	25847**	0.20	8.0
X ₂₉	0.44	2	0.22	56076**	0.44	8.9
X ₃₄	1.10	4	0.28	70983**	1.10	22.3
X ₃₉	0.16	2	0.08	20794**	0.16	3.1
X ₄₁	0.07	2	0.04	9489**	0.07	1.2
e	0.51	131196	0.00		0.59	12.0
Total	4.84	131219			4.84	100.0

구 분	X ₉	R_X ₁₃	X ₁₄	X ₂₁	X ₂₅	X ₂₈	X ₂₉	X ₃₄	X ₃₉	X ₄₁	
수 준 합	1 수준	14.2	14.7	9.1	12.8	12.3	13.9	14.2	7.2	13.9	13.7
	2 수준	13.3	13.4	10.4	13.3	13.3	13.3	7.2	13.3	13.3	
	3 수준	12.3	11.7	10.4	13.7	14.2	12.6	12.3	8.3	12.7	12.9
	4 수준			9.8					9.2		
	5 수준								7.9		
계	39.8	39.8	39.8	39.8	39.8	39.8	39.8	39.8	39.8	39.8	

각 변수의 수준에 대해서 발생하는 자료의 확률분포는 앞 절과 마찬가지로 일양분포를 이용하였으며, 각 수준에서 100개씩 발생시켜 총 486회(35×21) 시뮬레이션에 의해 요인배치법에 의한 조건에서 자료를 발생시켜 분산분석을 실시한 결과 다음 <표 17>와 같은 결과를 얻었다.

<표 17>에서 보는 바와 같이 분산분석결과 모든 주요변수가 유의(분산비:**)한 것으로 나타났으며, 변수의 최적조건을 찾기 위해 수준별 평방향의 추이를 살펴보면 X₉,

X₁₁, X₂₉변수는 1수준, X₁₃, X₃₂, X₃₇은 3수준 일 때 최적인 것을 알 수 있다.

<표 16> 주요변수의 수준분류

주요변수	수 준		
	1	2	3
X ₉	1532~1826	1827~2121	2122~2416
X ₁₃	88.0~92.3	92.4~96.7	96.8~101.0
X ₂₉	128.0~135.6	135.7~143.3	143.4~151.0
X ₃₂	13.0~15.3	15.4~17.6	17.7~20.0
X ₃₇	16.0~17.0	17.1~18.0	18.1~19.0
X ₁₁	1		2

<표 17> 주요변수별-수준별 분산분석표 및 평방향

변수	제공합 (S.S)	자유도 (φ)	평균제공 (V)	분산비 (F ₀)	순변동 (S')	기여율 (ρ)
X ₉	0.21	2	0.11	347**	0.21	8.9
X ₁₁	0.26	1	0.26	838**	0.26	10.8
X ₁₃	0.35	2	0.18	568**	0.35	14.6
X ₂₉	1.04	2	0.52	1686**	1.04	43.4
X ₃₂	0.31	2	0.16	507**	0.31	13.1
X ₃₇	0.07	2	0.04	117**	0.07	3.0
e	0.15	474	0.00		0.15	6.1
Total	2.39	485			2.39	100.0

구 분	X ₉	X ₁₁	X ₁₃	X ₂₉	X ₃₂	X ₃₇	
수 준 합	1 수준	49.5	73.6	40.0	54.2	40.3	42.9
	2 수준	45.4	62.4	45.4	45.8	45.4	45.4
	3 수준	41.1		50.6	35.9	50.3	47.7
	4 수준						
	5 수준						
계	136.0	136.0	136.0	136.0	136.0	136.0	

5. 결론 및 추후 연구방향

본 연구를 통하여 통계적 기법으로 후판 선급 TMCP제품의 내부품질결함의 불량률 저감을 위하여 로지스틱 회귀모형을 통한 예측확률의 수식모형 및 주요변수도출, 시뮬레이션을 통한 자료의 난수발생으로 완전 요인실험을 통한 주요변수의 최적조건 탐색을 하였으며 다음과 같은 결론을 얻었다.

1) 후판 선급 TMCP제품을 생산시 내부 품질에 중요한 영향을 미치고 있는 설명변수를 도출하였으며, 상기 도출된 변수들을 체계적으로 관리함으로써 우수한 제품을 생산할 수 있다.

2) 주요변수의 관리범위 내에서 품질이 우수한 제품을 생산할 수 있는 예측확률이 높은 로지스틱 회귀모형을 개발함으로써 주요변수들의 변화에 따른 예측확률의 변화를 알 수 있으므로 선 공정(제강공정) 불량시 후 공정에서 이를 만회할 수가 있다.

3) 로지스틱 회귀모형으로 인해 도출된 주요변수들의 최적공정조건 도출을 위해 작업범위 내에서 범주화시켜 시뮬레이션을 통한 자료의 난수발생으로 완전 요인실험을 실시함으로써 주요변수의 최적조건을 알아낼 수가 있었다.

4) 시뮬레이션으로 자료를 일양 및 삼각분포로 난수로 발생시켜 분산분석을 실시한 결과 난수발생의 분포에 의한 결과의 변동은 크게 변하지 않았음을 알 수 있었다.

5) 로지스틱 회귀모형 수행시 일차만이 고려된 모형과 이차 모형을 설정하였으나,

6) 분석결과 교호작용의 효과보다 모형에 선정된 일차 효과의 영향이 크게 작용하는 것으로 나타났다.

본 논문에서는 요인배치법을 이용한 최적화를 실시하고 설명변수의 흔들림(변동)을 일양, 삼각분포를 가정하였으나 반응표면분석 방법이나 정규분포를 이용하는 것이 추후 연구가능한 분야로 판단된다.

참고 문헌

- [1] 박성현(1998), 「회귀분석」, 민영사.
- [2] 송문섭, 조신섭(1997), 「Windows용 SAS를 이용한 통계자료분석」, 자유아카데미.
- [3] 정광모, 최용석(2000), 「SAS를 활용한 범주형 자료분석」, 자유아카데미.
- [4] POSCO(1993), 「압연부문 업무메뉴얼 I」.
- [5] POSCO(1993), 「품질관리부문 업무메뉴얼」.
- [6] Agresti, Alan(1996), *An Introduction to Categorical Data Analysis*, John Wiley and Sons, New York.
- [7] Anderson, E. B.(1997), *Introduction to the Statistical Analysis of Categorical Data*, Springer-Verlag Berlin Heidelberg, New York.
- [8] Everitt, Brian S. and Dunn, Graham (1991), *Applied Multivariate Data Analysis*, London.
- [9] Hosmer, D. W. and Lemeshow, S. (1989), *Applied Logistic Regression*, John Wiley and Sons, New York.
- [10] Johnson, J.D.(1992), *Applied Multivariate Data Analysis*, Springer-Verlag, New York.
- [11] Johnson, Valen E(1999), *Ordinal Data Modeling*, Springer-Verlag, New York.