⊠ 응용논문

일반화 선형모형을 통한 품질개선실험 자료분석 *

이영조
서울대학교 통계학과
임용빈
이화여자대학교 통계학과

Generalized Linear Models for the Analysis of Data from the Quality-Improvement Experiments

Youngjo Lee
Dept. of Statistics, Seoul National University
Yong Bin Lim
Dept. of Statistics, Ewha Womans University

Abstract

The advent of the quality-improvement movement caused a great expansion in the use of statistically designed experiments in industry. The regression method is often used for the analysis of data from such experiments. However, the data for a quality characterstic often takes the form of counts or the ratio of counts, e.g. fraction of defectives. For such data the analysis using generalized linear models is preferred to that using the simple regression model. In this paper we introduce the generalized linear model and show how it can be used for the analysis of non-normal data from quality-improvement experiments.

* 이 논문은 1995년도 문교부 기초 과학 연구비 지원(과제번호 BSRI-95-1415)에 의한 연구 결과임.

1. 서론

다구찌 품질개선 실험이 소개됨으로써 통계학적으로 계획된 실험들이 품질개선 및 생산성 향상을 위해 대폭 사용하게 되었다. 우리나라에서도 이에 관련된 소개서가 염봉진(1989), 박성현(1990, 1993)에 의해 출간되었다. 품질개선 실험으로 부터 얻어진 자료를 분석하는 기본적인 수단으로 회귀분석이 사용되는 바 이는 반응변수 y가 다음 세가지 조건을 만족한다고 가정한다.

(1) 정규성(Normality) : 반응변수 y가 정규분포를 따른다.

(2) 가법성(Additivity) : $\mu = E(y) = \beta_0 + X_1\beta_1 + X_2\beta_2 + \cdots + X_k\beta_k$.

(3) 등분산성 (Constant Variance) : $Var(y) = \sigma^2 I$.

품질개선실험에서 불량품의 개수나 불량율등의 변수들은 포아송 분포나 이항 분포를 따르므로 정규성 가정을 만족하지 않으며, 또한 분산이 일정하지 않고 평균에 따라 변한다. 예를 들면 포아송 분포에서는 분산이 평균에 비례한다. 흔히 이런 자료들은 독립 변수들의 효과에 대한 가법모형(2)이 평균척도(μ-scale)에서 성립한다는 보장도 없다. 포아송 분포에서는 가법성이 평균의 로그척도에서 흔히 만족되므로 이 세 가지 가정의 일부 또는 모두를 만족하는 자료는 실제 매우 드물다.

Box와 Cox(1964)는 자료가 위의 세 조건을 일부 또는 모두 만족시키지 않을 경우에는 이들을 충족시키도록 자료변환을 한 뒤 분석을 해야 한다고 주장하였다. 그러나 Hougaard(1982)가 밝혔듯이 이 세 가지 조건을 동시에 충족시킬 수 있는 자료변환은 존재하지 않는다. 간단한 예로 포아송 분포를 생각해 보자. 정규분포 가정을 위하여는 $y^{2/3}$, 등분산성 가정을 위하여는 $y^{1/2}$, 포아송자료에서 많이 발생하는 곱모형(Multiplicative Model)이 가법성을 만족하기 위하여는 $\log y$ 변환이 필요하다. 자료변환의 문제점은 이와 같이 세 조건을 충족시키기 위하여 각기 다른 세 가지 자료변환이 요구된다는 점이다.

이러한 어려움을 극복하기 위하여 Nelder와 Wedderburn(1972)이 일반화 선형모형을 개발하고 McCullagh과 Nelder(1989, 2nd edition)가 1983년 이에 대한 소개 책자를 발간한 후 통계학 분야에서 널리 사용되고 있다. 또한 SAS, SPLUS, GLIM, GENSTAT 등 여러 패키지에서 일반화 선형모형을 위한 소프트웨어가 개발되어 현재 자료분석에 손쉽게 이용될 수 있다.

본 논문에서는 일반화 선형모형이 어떻게 품질개선 실험에서 얻어진 자료들을 분석하는데 이용될 수 있는지에 대해 소개하고, 실제 자료들을 분석해 봄으로써 회귀분석과 비교하고 그 장점을 설명하고자 한다. 한국에서도 일반화 선형모형을 이용하여 보다 다양한 품질개선 자료들에 대한 분석이 효율적으로 이루어졌으면 한다.

2. 일반화 선형모형

일반화 선형모형은 회귀분석의 세가지 가정을 확장하여 좀더 일반적인 자료들을 분석하고자 하는데 목적이 있다.

2.1. 반응변수의 분포와 분산함수

일반화 선형모형에서는 반응변수의 분포가정은 정규분포 뿐 아니라 포아송분포, 이항분포, 감마분포, 음이항(negative binomial)분포, 역정규(inverse Gaussian)분포 등을 포함하는 일모수지수족(one-parameter exponential family)으로 확장될 수 있다. 즉 정규분포를 따르는 연속형 자료의 회귀분석 뿐만 아니라, 결점수등의 개수 자료는 포아송분포, 불량율 등의 비율에 관한 자료는 이항분포, 양의 값을 갖는 연속변수로 등변동계수(constant coefficient of variation)를 갖는 자료는 감마분포를 이용하여 회귀분석할 수 있다

일반화 선형모형에서는 평균과 분산이 다음과 같이 표현된다.

$$E(y) = \mu , \qquad Var(y) = \phi V(\mu) \tag{1}$$

여기서, ϕ 를 산포모수(dispersion parameter)라고 하며 $V(\mu)$를 분산함수라고 부른다. 산포모수 ϕ는 반응변수 y의 분산 중 y의 평균에 상관없는 부분을 나타내며 분산함수 $V(\mu)$ 는 y의 분산이 어떻게 y의 평균에 따라 변하는 가를 나타낸다.

분포에 따라 분산함수들이 달라지는 바 이를 <표 1>로 요약하였다. 분산함수가 성해지면 일반화 선형모형의 지수족 중에서 분포가 지정되므로 일반화 선형모형에서는 분포의 지정과 분산함수의 지정은 동치이다. 정규분포의 경우는 $V(\mu)=1$ 로 분산이 평균에 따라 변하지 않음, 즉 등분산성을 나타낸다. 그러므로 분산함수의 지정은 등분산성 가정의 확장이라고 볼 수 있다. 예를 들면 포아송분포의 경우 $V(\mu)=\mu$ 로 분산이 평균이 증가함에 따라 증가함을 나타낸다. 이항분포에서는 p가 모비율을 나타내며 μ=np의 관계가 있다. 그러므로 이항분포에서는 $V(\mu) = \mu(n-\mu)/n$ = np(1-p)로 p가 0 이나 1 에 가까우면 분산이 작아짐을 알 수 있다. 감마분포에서는 $Var(y)=\phi\mu^2$로 $\phi^{1/2}$는 변동계수를 나타낸다. 분산함수가 주어지면 일반화 선형모형에서는 회귀계수들이 가중최소제곱법(weighted least square)을 이용한 최우 추정법(maximum likelihood estimation)으로 추정되므로 분산함수의 올바른 지정이 회귀계수 추정치들의 효율성에 큰 영향을 미친다. 예를 들어 이항분포를 따르는 비율 자료의 분석을 등분산성을 가정한 회귀분석으로 한다면 매우 비효율적인 분석이 된다.

〈표 1〉 분포에 따른 분산함수

분포	산포모수	분산함수
정규분포	σ^2	1
포아송분포	1	μ
이항분포	1	$\mu(n-\mu)/n$
음이항분포	1	$\mu + \alpha\mu^2$
감마분포	ϕ	μ^2
역정규분포	ϕ	μ^3

2.2. 연관함수와 가법성

회귀분석 모형에서는 가법성이 다음과 같이

$$\mu = \beta_0 + X_1\beta_1 + X_2\beta_2 + \cdots + X_k\beta_k$$

평균(μ)척도에서 성립한다고 가정하는 반면, 일반화 선형모형에서는 가법성이 다음과 같이 연관함수 $g(\ )$를 통해

$$\eta = g(\mu) = \beta_0 + X_1\beta_1 + X_2\beta_2 + \cdots + X_k\beta_k \tag{2}$$

성립한다고 한다. 분포에 따라 자주 사용되는 연관함수들을 표 2 에 정리하였다.

〈표 2〉 분포와 연관함수

분포	연관함수 $g(\ )$	모형
정규분포	μ	Linear model
이항분포	$\log p/(1-p)$	Logit model
	$\Phi^{-1}(p)$	Probit model
포아송분포	$\log\mu$	Log-linear model
음이항분포	$\log\mu$	
감마분포	$\log\mu$ 또는 $1/\mu$	
역정규분포	$\log\mu$ 또는 $1/\mu^2$	

일반화 선형모형에서는 자료의 변환을 통해 가법성을 성취하는 것이 아니라 모수의 변환을 통해 가법성을 충족시킨다. 예를 들어 개수 자료들을 분석하는데 사용되는

포아송분포에서 가법성을 성취하기 위해 로그변환이 필요하다 하자. 이 경우 $\log y$ 변환자료에 대한 회귀분석은 연관함수를 $\log \mu$로 잡고 일반화 선형모형을 사용하는 것과 유사하다. 그러나 로그값이 0 값에서 정의되지 않으므로 자료변환 $\log y$ 에서는 자료의 손실이 발생할 수 있으나 모수변환 $\log \mu$ 를 통해 가법성을 가정하는 것은 아무런 자료의 손실을 초래하지 않는다. 예를 들어 평균 $\mu = 1$ 인 포아송분포를 생각해보자. 이 경우 반응변수 y가 0 일 확률이 0.3679로 $\log \mu = 0$ 은 항상 정의되나 $\log y$ 는 약 37%의 자료에서 정의되지 않는다.

연관함수의 중요한 역할로 다음의 두 가지를 들 수 있다. 포아송 경우를 생각해 보면 평균 μ 는 항상 양수이나 $\eta(=X\beta)$ 가 취할 수 있는 값의 범위는 $(-\infty, \infty)$이다. 그러므로 정규분포를 가정한 회귀분석에서는 $\eta = g(\mu) = \mu$ 로 잡기 때문에 평균 μ의 예측치가 음의 값을 가질 수 있으나 로그 연관함수를 가정하면 $\mu = \exp(\eta) = \exp(X\beta)$의 예측치가 항상 양의 값만을 취한다.

이제 두 수준을 갖는 A, B 두 인자를 가정하고 다음의 가법모형을 생각해 보자.

$$\mu_{ij} = \beta_0 + a_i + b_j . \tag{3}$$

표 3 의 자료 1에서는 $\beta_0 = 25$, $a_1 = -10$, $a_2 = 10$, $b_1 = -5$, $b_2 = 5$ 로 (3)의 관계를 만족하나 자료 2는 (3)과 같이 A, B 두 인자의 주효과만의 가법모형으로는 나타낼 수 없고 A, B의 교호작용이 필요하다. 그러나 자료 2는 곱 모형으로 설명되므로 로그 연관함수를 사용하면 다음과 같이 표현된다.

$$\log \mu_{ij} = \beta_0 + a_i + b_j .$$

이 경우 자료 2는 $\beta_0 = 0.5 \log 600$, $a_1 = -0.5 \log 3$, $a_2 = 0.5 \log 3$, $b_1 = -0.5 \log 2$, $b_2 = 0.5 \log 2$ 의 관계를 만족한다 (임용빈, 이영조 (1995)). 그러므로 연관함수의 적절한 선택은 요인실험법(factorial design)에서 불필요한 교호작용을 없애줄 수 있다.

〈표 3〉 A, B 인자수준에 따른 μ 값의 변화

자료1			
		B	
		1	2
A	1	10	20
	2	30	40

자료 2			
		B	
		1	2
A	1	10	20
	2	30	60

3. 소음신호와 평균과 분산의 동시 모형화

다구찌(1985, 1987) 품질개선실험의 목표는 제품간의 변이를 줄이고 제품의 평균을 성능목표치에 맞추는 적정 공정과정을 찾아내는데 있다. 다구찌(1987)는

$$SN = 10 \log \frac{\mu^2}{Var(y)}$$

을 정의하고 이의 추정치로써 다음과 같은 SN 비를 사용할 것을 주장하였다.

$$SN \text{비} = 10 \log \frac{\bar{y}^2}{s_y^2} .$$

다구찌의 SN 비는 제품간 변이를 반영하는 측도로서 SN 비가 크면, 변이가 작다는 것을 의미한다. 그러므로, 다구찌는 SN 비를 크게 하는 공정과정 조합을 품질개선 실험을 통해 찾아야한다고 하였다.

감마분포에서는 $Var(y) = \phi\mu^2$ 이므로

$$SN = -10 \log \phi$$

가 된다. 그러므로 SN 비를 크게 하는 것은 산포모수 ϕ를 작게 하는 것과 동치이다. 일반적으로 일반화 선형모형에서는 $E(y) = \mu$, $Var(y) = \phi V(\mu)$ 로서 ϕ는 반응변수 y의 분산 중 평균 μ와 상관없는 부분을 나타낸다. 그러므로 SN 비를 크게 하는 것은 ϕ를 줄임으로써 제품의 평균 μ에 영향을 주지않고 제품간 변이를 줄이고자 하는 것이다. Nair와 Pregibon(1986), Leon과 그외(1987) 등은 ϕ의 함수의 추정치를 PerMIA(Performance Measure Independent of Adjustment)라고 하였다. 그러므로 다구찌의 SN 비는 분산함수가 $V(\mu) = \mu^2$인 경우에 한하여 PerMIA가 된다. 일반화 선형모형에서 $Var(y) = \phi V(\mu)$로 정의하는 바 이 식은 일반적인 분산함수에서 PerMIA가 정의될 수 있게 해준다. 또한 회귀분석법을 통해 SN 비를 크게 하는 공정과정을 찾는다는 것은 가법성이 $\log\phi$의 측도에서 만족됨을 의미한다. 즉, 다구찌의 SN 비를 통한 품질 개선 실험은 일반화 선형 모형에서 보면,

(i) $V(\mu) = \mu^2$와

(ii) 가법성이 $\log\phi$의 측도에서 만족된다는

두 가지 가정을 한다고 볼 수 있다.

제품간 변이를 줄이고 제품평균을 성능목표치에 맞추는 적정 공정과정을 찾기 위

하여 Nelder와 Lee(1991)는 다음과 같은 평균과 분산의 동시 모형을 생각하였다.

$$\eta = g(\mu) = \beta_0 + X_1\beta_1 + X_2\beta_2 + \cdots + X_k\beta_k \quad (4)$$

$$\log \phi = \gamma_0 + Z_1\gamma_1 + Z_2\gamma_2 + \cdots + Z_p\gamma_p .$$

$(X_1, X_2, \cdots, X_k)$와 $(Z_1, Z_2, \cdots, Z_p)$의 두 군데 모두 들어가는 설명변수는 평균과 분산에 동시적으로 영향을 주며 $(X_1, X_2, \cdots, X_k)$에만 들어가는 변수는 평균에만, $(Z_1, Z_2, \cdots, Z_p)$에 들어가는 변수는 분산에만 영향을 주는 공정과정들이 된다. 그러므로 평균과 분산에 동시적으로, 또는 분산에만 영향을 주는 공정과정들을 통해 제품간 변이를 줄인 후 평균만 영향을 주는 공정과정을 통해 제품의 평균을 성능 목표치에 맞춘다. 평균과 분산의 동시 모형(4)에서 사용할 연관함수 $g()$로 <표 2>를 이용할 수 있겠다. 즉 반응변수가 이항분포를 따르면 로짓 또는 프라빗등을 사용하고 포아송분포를 따르면 로그함수등을 이용할 수 있겠다.

4. 예제

4.1 타일 실험과 이항분석

다구찌(1987, pp 165-181)는 이나세이또사의 타일 실험을 소개, 분석하였다. 실험은 27-4 부분실시법을 이용하여 일곱 개의 공정과정을 인자(A-G)로 하여 실험하였다. 반응변수는 100개의 타일 중 불량품의 개수로 정규분포 보다는 이항분포를 따른다고 볼 수 있겠다. 실험의 개요는 <표 4> 에 요약되어 있다.

〈표 4〉

A	B	C	D	E	F	G	y
1	1	1	1	1	1	1	16
1	1	1	2	2	2	2	17
1	2	2	1	1	2	2	12
1	2	2	2	2	1	1	6
2	1	2	1	2	1	2	6
2	1	2	2	1	2	1	68
2	2	1	1	2	2	1	42
2	2	1	2	1	1	2	26

다구찌는 가장 불량율을 작게하는 공정과정수준 조합을 적정 공정과정이라 정의

한 후 이를 찾기 위하여 <표 5>를 제시하였다.

〈표 5〉

인자와 수준	총 불량타일 개수	불량타일 퍼센트
A_1	51	12.75
A_2	142	35.50
B_1	107	26.75
B_2	86	21.50
C_1	101	25.25
C_2	92	23.00
D_1	76	19.00
D_2	117	29.25
E_1	122	30.50
E_2	71	17.75
F_1	54	13.50
F_2	139	34.75
G_1	132	33.00
G_2	61	15.25

<표 4>의 자료를 이용하여 분산분석표를 만들면 잔차항이 없으므로 주효과들이 유의한가를 검정할 수 없다. 그러므로 다구찌는 4 페이지(173-176)에 걸친 긴 논의를 통해 B와 C의 두 주효과가 유의하지 않다고 주장하고 이를 잔차항으로 한 분산분석표를 제시하였다. 이제 (A1, *, *, D1, E2, F1, G2)를 적정공정과정이라 하자. 다구찌는 데시벨(db)의 개념을 이용한 오메가법을 통해, 적정공정과정에서 100개의 타일을 만들면 평균 2개의 불량품이 생산된다고 예측하였다. 다구찌의 오메가법은 가법성(2)이 데시벨 척도 (db = $10\log p/(1-p)$)에서 성립한다고 가정하는 것으로, 일반화 선형모형에서 로짓 연관함수($\log p/(1-p)$)를 사용하는 것과 동일하다.

이항분포를 가정한, 다음의 로짓모형

$$\eta_{hijklmn} = \log \frac{p_{hijklmn}}{1-p_{hijklmn}} = \beta_0 + a_h + b_i + c_j + d_k + e_l + f_m + g_n \,. \quad (5)$$

을 이용해 분석한 결과가 <표 6>에 정리되어 있다. 이 로짓모형은 앞에서 설명한 바와 같이 가법성이 다구찌의 데시벨 척도에서 성립한다고 가정하는 것이다. 이항분포에서는 산포모수가 $\phi=1$로 주어져 있으므로, 이의 추정없이 주효과들을 검정할 수 있다. <표 6>을 보면 B의 주효과만이 유의하지 않다.

<표 6>의 결과를 얻기위한 Glim 명령어는 다음과 같다. y=(16, 17, ..., 26)는 반응변수, n=(100, 100, ..., 100)은 이항분포의 실험수, 인자 A-G들을 a-g의 변수로 입력되었다고 하자.

$yvar y <-- 반응변수가 y임을 지정함.
$err b;n <-- 분포가 실험수가 n인 이항분포임을 지정함.

$lin g <-- 연관함수가 로짓(g)임을 지정함.
$fit a+b+c+d+e+f+g <-- 독립변수들이 a-g임을 지정함.
$dis e <-- <표 6>의 회귀계수를 인쇄할 것을 명령함.

이제 적정공정과정 (A1, *, C2, D1, E2, F1, G2)에서 불량품의 개수가 평균 몇 개인지 예측해 보면,

$\eta_{1*21212} = \beta_0 + a_1 + c_2 + d_1 + e_2 + f_1 + g_2$ 로서

이의 추정치가 -3.907(=1.6582-0.5323-0.8672-0.8492)으로

$\mu_{1*21212} = \frac{\exp(\eta_{1*21212})}{1+\exp(\eta_{1*21212})}$ 의 추정치가 0.0197($= \frac{\exp(-3.907)}{1+\exp(-3.907)}$)가 된다. 그러므로 적정 공정과정 조합에서 100개의 타일을 만들면 평균 2개의 불량품이 생산된다고 예측을 하게된다.

다구찌는 적정공정과정을 찾고 공정과정들을 검정할때, 통계학에서 보면 다음과 같은 회귀분석모형를 가정했다고 볼 수 있다.

$\mu_{hijklmn} = \beta_0 + a_h + b_i + c_j + d_k + e_l + f_m + g_n$.

다구찌(1987, page 179)가 지적하였듯이 이 회귀분석 모형을 이용하여, 적정 공정과정 조합(A1, *, *, D1, E2, F1, G2)에서 100개의 타일을 만들때 불량개수를 예측하면 평균 -18.25 개의 불량품이 생산된다는 말도 안되는 예측을 하게 된다. 이러한 모순을 피하기 위해 다구찌는 오메가법을 개발하였는데 이는 일반화 선형모형에서 로짓 모형(5)를 사용하는 것과 유사하다. 즉, 다구찌는 동일한 자료에 대해서조차 검정의 문제이냐 또는 예측의 문제이냐에 따라 다른 모형들을 가정하여 분석하였음을 의미한다. 이에 반하여 일반화 선형모형을 이용하면, 검정의 문제이든 예측의 문제이든 하나의 일관된 모형하에서 분석이 이루어진다.

〈표 6〉

모수	추정치	표준오차	t 값
β_0	-1.6582	.2728	-6.079
a_2	1.1553	.2141	5.369
b_2	-0.2178	.2141	-1.017
c_2	-0.5323	.2141	-2.486
d_2	0.5239	.2141	2.447
e_2	-0.8672	.2141	-4.050
f_2	1.2651	.2141	5.908
g_2	-0.8492	.2141	-3.966

단, Glim에서는 $a_1 = b_1 = c_1 = d_1 = e_1 = f_1 = g_1 = 0$ 의 조건을 사용하여 모형(5)의 모수들을 추정한다.

4.2 Wave-Solder 실험과 포아송 분석

전자회로에서 부품들을 부착시키려고 납땜할 때, 납땜 이음매에 결함들이 발생한다. 전자회로에서 납땜 이음매 결함 개수들을 줄이기 위한 Condra(1993)의 품질 개선 실험 자료 중 일부가 표 7에 있다. 납땜시 7개(A-G)의 공정이 필요하다. 반응변수는 납땜 이음매 결함 개수로 정규분포보다는 포아송분포를 따른다.

〈표 7〉

A	B	C	D	E	F	G	y
1	1	1	1	1	1	1	13
1	1	1	2	2	2	2	4
1	2	2	1	2	2	1	36
1	2	2	2	1	1	2	53
2	1	2	1	2	1	2	29
2	1	2	2	1	2	1	10
2	2	1	1	1	2	2	28
2	2	1	2	2	1	1	100

포아송분포를 가정한, 다음의 로그선형모형 (log-linear model)

$$\eta_{hijklmn} = \log \mu_{hijklmn} = \beta_0 + a_h + b_i + c_j + d_k + e_l + f_m + g_n \; . \quad (6)$$

을 이용해 분석한 결과와 유의하지 않은 주효과들을 하나씩 제거하면서(backward elimination) 유의한 효과들로만 이루어진 최종 모형의 결과가 표 8에 정리되어 있다. <표 8>을 보면 포아송분포에서도 산포모수가 $\phi = 1$이므로 완전 모형에서 회귀계수 추정치의 표준오차와 t 값을 구할 수 있다. 표 8을 보면, 최종 모형에서 C, D, G의 주효과들이 유의하지 않았다. 이제 적정공정과정 (A1, B1, *, *, E1, F2, *)에서 납땜 이음매 결함 개수가 평균 몇 개인지 예측해 보면

$\eta_{11**12*} = \beta_0 + a_1 + b_1 + e_1 + f_2$ 로서 이의 추정치가 1.744(=2.606-0.862) 으로 $\mu_{11**12*} = \exp(\eta_{11**12*})$의 추정치가 5.720(= exp(1.744))가 된다. 그러므로 적정 공정과정 조합에서 평균 5.72개의 납땜 이음매 결함이 발생한다고 예측할 수 있다. 이 예에서 본 바와 같이 일반화선형모형의 포아송로그선형모형을 이용하면, 반응변수가 결함 개수인 자료들을 일관성 있게 분석할 수 있다.

<표 8>의 결과를 얻기위한 Glim 명령어는 다음과 같다. y=(13, 4, ... , 100)는 반응

변수, 인자 A-G들을 a-g의 변수로 입력하였다고 하자.

$yvar y <-- 반응변수가 y임을 지정함.
$err p <-- 분포가 포아송분포임을 지정함.
$lin l <-- 연관함수가 로그(l)임을 지정함.
$fit a+b+c+d+e+f+g <-- 완전모형의 독립변수들이 a-g임을 지정함.
$dis e <-- 완전모형의 회귀계수를 인쇄할 것을 명령함.
$fit -c-d-g <-- 완전모형에서 독립변수 c, d, g를 뺄것을 명령함.
$dis e <-- 최종모형에서의 회귀계수를 인쇄할 것을 명령함.

〈표 8〉

완전 모형 (Full model)				최종 모형			
모수	추정치	표준오차	t 값	모수	추정치	표준오차	t 값
β_0	2.565	.277	9.248	β_0	2.606	.171	15.242
a_2	0.526	.186	2.825	a_2	0.309	.136	2.272
b_2	1.468	.186	7.888	b_2	1.355	.150	9.039
c_2	0.334	.186	1.794				
d_2	-0.146	.186	-0.784				
e_2	0.193	.186	1.038	e_2	0.361	.136	2.645
f_2	-0.976	.186	-5.245	f_2	-0.862	.136	-6.325
g_2	-0.250	.186	-1.344				

단, $a_1 = b_1 = c_1 = d_1 = e_1 = f_1 = g_1 = 0$.

4.3 Speedometer Cable 실험과 평균과 분산의 동시 모형화

Quinlan(1985)은 Speedometer Cable의 수축에 영향을 미치는 15가지 공정과정(A-O)에 대하여 실험을 하였다. 실험의 목표는 케이블 수축을 최소로 하는 적정 공정과정을 찾는 것으로 215-11 부분실시법을 사용하였다. 이 실험에 대한 분석으로는 자료변환을 이용한 Box(1988), 일반화 선형모형을 이용한 Nelder와 Lee(1991) 연구들이 있었다. Nelder와 Lee(1991)는 반응변수가 케이블의 수축으로 비율변수이므로, 수축이 없거나($\mu=0$) 또는 백퍼센트 수축이 일어난($\mu=1$) 경우 반응변수의 분산이 0이 되기 때문에 $Var(y) = \phi\mu(1-\mu)$, 즉 $V(\mu)=\mu(1-\mu)$의 분산함수를 가정하였다. <표 1>에서 보다시피 이항분포에서 비율모수를 p라 할 때 $Var(y) = np(1-p)$인 바, 이 분산함수는 이항분포에서 자주 사용된다. 한편 연관함수 $g(\mu)$

를 log μ 로 잡았는데 이유는 Box(1988)가 지적한 바와 같이 이 연관함수가 가장 유의한 두 주효과 E와 G의 교호작용을 없애주기 때문이다. Nelder와 Lee(1991)의 평균과 분산 동시 모형의 최종 모형은 다음과 같다.

$$\log \mu_{hijklmno} = \beta_0 + a_h + c_i + d_j + e_k + f_l + g_m + h_n + k_0$$

$$\log \phi_{ijklm} = \gamma_0 + d_i + f_j + g_k + h_l + n_m .$$

이 Nelder와 Lee(1991)의 분석에 의하면 {D, F, G, H}의 공정과정은 평균과 분산에 동시적으로 영향을 주며 {N}은 분산에만 {A, C, E, K}는 평균에만 영향을 주는 공정과정이다. Box(1988)는 log y 변환을 사용하여 자료를 분석하였다. 한편 Quinlan(1985)은 실험전 기존의 공정과정에서 생산된 스피도미터 케이블들의 실제 수축율 평균과 표준편차와, 실험 후 그들이 찾아낸 새로운 적정 공정과정에서 생산된 케이블들의 실제 평균과 표준편차를 제시하였다. 한편 Nelder와 Lee(1991)는 수축율의 평균과 표준편차의 실제 값을 Box(1988)의 자료변환법 및 그들의 일반화선형모형법에 의한 예측치들과 비교하였는데, 이는 <표 9>에 제시되어 있다.

〈표 9〉

	실험 전		실험 후	
	평 균	표 준 편 차	평 균	표 준 편 차
실 제	26.0	5.0	5.0	2.5
일반화선형모형법	25.9	6.43	6.22	2.09
자 료 변 환	27.7	3.28	11.5	1.36

<표 9>를 보면 새로운 공정과정이 기존의 공정과정과 비교하여 케이블 평균 수축율을 줄일 뿐만 아니라 표준편차도 줄여 품질을 향상시킴을 알 수 있다. 또한 일반화선형모형을 이용한 Nelder와 Lee(1991)의 방법이 자료변환을 이용한 Box(1988)의 방법보다 실제 평균과 표준편차를 훨씬 정확하게 예측함을 알 수 있다. Nelder와 Lee(1991, 1992)의 방법을 수행하기 위한 Glim 프로시져는 이 영조(1993)에 있고, 보다 개선된 방법에 대한 연구로 Lee와 Nelder(1995) 등이 있으며, 현재 분석에 이용할 수 있는 소프트웨어로 GENSTAT 프로시져가 개발되어 있다. 이 자료에 대한 보다 구체적인 분석 및 방법론은 상기 논문들을 참조하기 바란다.

5. 결론

다구찌의 품질 개선 방법은 적정 공정과정을 찾기 위하여 잘 계획된 실험계획법

들이 사용되어야 하며, 제품의 품질을 개선하기 위하여 제품간 변이를 줄여야 한다는 것으로 근본적인 아이디어들은 매우 중요하며 존중되어야 한다. 그러나, 이러한 아이디어들을 실천하기 위한 다구찌의 자료분석법은 Box(1988)가 지적한 바와 같이 많은 문제점이 있어, 보다 향상된 자료분석법을 이용하여 품질개선 실험자료들을 해석할 필요가 있다. 품질개선 자료들은 거의 모두가 자료변환 후 회귀분석방법을 통해 해석된다; Box(1988). 일반화 선형모형에 의한 분석법은 회귀분석의 확장으로 개수, 비율 등의 품질개선 실험에서 발생하는 자료들을 효율적으로 분석할 수 있게 해준다. 일반화 선형모형에서 연관함수 및 분산함수의 개념은 자료분석의 폭을 넓혀주며, 특히 산포모수 ϕ는 다구찌의 *SN* 비, PerMIA 등의 개념을 이해하기 용이하게 하여준다. 본 논문에서는 일반화 선형모형을 소개하였다. 이를 위한 여러 종류의 소프트웨어가 사용가능하므로 일반화 선형모형을 이용하여 한국에서 다양한 자료들에 대한 분석이 효율적으로 이루어졌으면 한다.

참고문헌

[1] 박성현(1990), 다구찌 방법을 중심으로 한 응용실험계획법, 영지문화사, 서울.

[2] 박성현(1993), 품질공학, 민영사, 서울

[3] 염봉진(1988), 제품 및 공정설계를 위한 다구찌 방법, 산학협동교재, 생산성본부 서울.

[4] 이영조(1993), 다구찌 실험 분석에 있어서 일반화선형모형 대 자료변환, 응용통계연구, 6권, 2호, pp. 253-263.

[5] 임용빈, 이영조 (1995). 몬테칼로 씨뮬레이션을 통한 SN비의 효율성 비교, 품질경영학회지, 23권, 2호, pp. 28-42.

[6] Addelman, S. and Kempthorne, O. (1991), "Some Main Effects Plans and Orthogonal Arrays of Strength Two," Annals of Mathematical Statistics, No. 2, pp. 1167-1176.

[7] Box, G. E. P. (1988), "Signal-to-Noise Ratios, Performance Criteria and Transformations," Technometrics, Vol. 30, pp. 1-17.

[8] Box, G. E. P. and Cox, D. R. (1964), "An Analysis of Transformations," (with discussion), Journal of the Royal Statistical Society, B, Vol. 26, pp 211-252.

[9] Condra, L. W.(1993), Reliability Improvement with Design of Experiments. Marcel Dekker, New York

[10] Hougaard, P. (1982), "Parametrizations of Non-Linear Models," Journal of the Royal Statistical Society, B, Vol. 44, pp. 244-252.

[11] Kempthorne, O. (1952), The Design and Analysis of Experiments. Wiley, New York.

[12] Lee, Y. and Nelder, J. A. (1995), "Generalized Linear Models for the Analysis of Quality-Improvement Experiments," A manuscript submitted for publication.

[13] Leon, R. V., Shoemaker, A. C., and Kackar, R. N. (1987), "Performance Measures Independent of Adjustment," (with discussion), Technometrics, Vol. 29, pp. 253-285.

[14] McCullagh, P. and Nelder, J. A. (1989), Generalized Linear Models, 2nd ed. London: Chapman and Hall.

[15] Nair, V. N. and Pregibon, D. (1986), "A Data Analysis Strategy for Quality Engineering Experiments," AT&T Technical Journal, Vol. 65, pp. 73-84.

[16] Nelder, J. A. and Lee, Y. (1991), "Generalized Linear Models for the Analysis of Taguchi-Type Experiments," Applied Stochastic Models and Data Analysis, Vol. 7, pp. 107-120.

[17] Nelder, J. A. and Lee, Y. (1992), "Likelihood, Quasi-likelihood and Pseudo-likelihood: Some Comparisons," Journal of the Royal Statistical Society, B, pp. 54,

[18] Nelder, J. A. and Wedderburn, R. W. M. (1972), "Generalized Linear Models," Journal of the Royal Statistical Society, A, Vol. 132, pp. 370-384.

[19] Quinlan, J. (1985), "Product Improvement by Application of Taguchi Methods," in American Supplier Institute News (special symposium ed.) Dearborn, MI: American Supplier Institute, pp. 11-16.

[20] Taguchi, G. (1985), "Quality Engineering in Japan," Communications in Statistics, A, Vol. 14, pp. 2785-2801.

[21] Taguchi, G. (1987), System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Cost, White Plains, NY: UNIPUB/ Kr International.