응용통계연구 제 6권 2호
1993년 9월 pp.253-263.

다구찌 실험분석에 있어서 일반화선형모형 대 자료변환[1]

이 영 조[2]

요 약

최근 다구찌 실험에 대한 관심이 고조되어 일반화 선형모형에서 평균과 분산의 동시모형화가 연구되고 있다. 하나의 자료 변환만으로는 자료분석에 필요한 모든 조건들을 만족시킬 수 없기 때문에 다구찌 품질실험의 자료들을 일반화 선형모형으로 분석하는 것이 바람직하다. 본 논문에서는 이 자료변환법과 일반선형모형을 이용한 분석법을 소개, 비교하고 일반화 선형모형을 다구찌 자료에 적용할 수 있는 GLIM 프로그램을 제시한다.

1. 개 요

우리나라에서는 최근 제품개발의 단계에서 품질관리기법을 이용, 공정과정을 개선, 품질 향상을 통해 국제 경쟁력을 높이려는 노력이 활발해지고 있다. 이에 부응하여 다구찌 품질공학에 관한 소개서가 염봉진(1989), 박성현(1990)에 의해 출간되었다.

다구찌 품질공학(Taguchi, 1985, Taguchi 와 Wu, 1985)은 목표치에 대한 제품변이를 최대한 줄이는 적정공정 과정을 찾아 품질을 향상(Quality Improvement)시키는 것을 목적으로 한다. 적정공정 과정을 찾기 위해서는 실험계획법의 부분실시법(Fractional Experiment)을 사용하는데 특히 주효과 계획법(Main Effect Plan)을 활용, 공정과정들에 대한 실험을 한다. 그러나 다구찌 실험에서는 실시할 실험의 처리(Treatment)들, 즉 각 공정과정 조합들이 결정되면 순수한 반복을 통해 제품의 처리내 변이를 측정하는 것이 아니라, 제품의 변이에 영향을 준다고 생각되는 소음인자들에 대해 다시 부분실시법을 실시, 소음인자 조합들을 결정하고 각 처리 내에서 이들 소음인자 조합들에 대한 제품의 특성치를 측정한다. 이와같이 자료가 얻어지면, 분산분석법을 이용 공정과정에 대한 분석을 하게 된다. Logothetis 와 Wynn(1989)의 다구찌 분석법 해설을 따르면 우선 각 공정과정을 실험계획법의 인자(Factor)로 설정하고, 인자들을 제품간 변이조정인자(Variability Control Factor)와 제품 목표치 조정인자(Target Control Factor)로 분산분석법을 통해 분류한다. 그 다음 변이조정인자 수준 결정을 통해 제품간 변이를 최대한 줄인 후 목표치 조정인자 수준 결정을 통해 제품의 평균을 성능 목표치에 맞춘다. 한편 다구찌 방법 소개자들 마다 용어의 차이를 보이는데 Nair 와 Pregibon(1988)은 목표치 조정인자를 조절인자(Adjustment Factor)라고 부르기로 한다. 그러므로 다구찌 분석법은 통계학적으로 보면 제품간 변이를 측정하는 분산측도(Dispersion Measure)와 제품의 평균을 측정하는 평균측도(Mean Measure)를 결정한 후, 이 측도들의 분석을 통해 각 인자들이 평균과 분산에 미치는 영향을 파악, 최적공정 과정을 찾아내는 방법이라 정의할 수 있다. 다구찌는 특히 분산측정치를 소음신호(Signal to Noise)라는 통계량으로 요약하고 이를 분산분석법을 이용하여 분석하였다.

이와 같은 다구찌 방법은 다음의 세 가지 요소로 구성된다고 볼 수 있다.

---

1) 본 연구는 한국과학재단 연구지원의 결과임.

2) (200-702) 강원도 춘천시 옥천동 1번지 춘천 한림대학교 통계학과

(1) 공학적측면 : 중요하다고 생각되는 공정과정, 소음인자 및 이들의 수준 결정.
(2) 실험계획법 : 실험을 실시할 공정과정 조합들과 소음인자 조합들을 결정.
(3) 분석 : 분산측도, 평균측도 결정 및 분산분석법을 통한 분석.

통계학 분야에서 다구찌 방법을 논할 때 처음 두 요소에 대해서는 큰 반대가 없으나 세번째 다구찌 분석에 대해서는 통계적으로 적합하지 않다고 생각되고 있다. 이에 대한 대안으로 Box(1886, 1988), León과 그외(1987), Nair 와 Pregibon(1986)등의 학자들은 측정된 자료를 자료변환(Data Transformation)한 뒤 변환된 자료들의 평균과 분산측도들을 결정, 분석하여 적정공정 과정을 찾아야한다고 주장한다. 그러나 이러한 자료변환을 이용한 통계적분석법도 다음 장에서 밝히다시피 여러가지 문제점을 내포하고 있어서 Nelder 와 Lee(1991)은 일반화선형모형(Generalized Linear Model)에서 평균과 분산의 동시모형화(Joint Modelling of the Mean and Dispersion Parameter)를 이용한 다구찌 자료의 분석방법을 제안하였다.

본 논문에서는 자료변환법과 일반선형모형을 이용한 분석법을 소개, 비교한 후 후자의 방법에 대하여 GLIM 패키지를 이용해서 다구찌 자료를 분석할 수 있는 Macro를 제시하였다.

2. 자 료 변 환

일반회귀 분석은 모평균에 대한 분석 모형으로 종속변수가

(i) 정규분포를 따르고,
(ii) 평균이 독립변수들의 선형함수로 표현되며,
(iii) 분산이 평균과 관계없이 상수라는

세 가지 가정을 바탕으로 하고있다. 그러나 이 세 가지 가정의 일부 또는 모두를 만족하는 실제 자료는 매우 드물다. Box와 Cox(1964)는 실제 자료가 위의 세 조건을 만족시키지 않을 경우에는 이들을 충족시키도록 자료변환을 한 뒤 분석을 해야 한다고 주장하였다. 그러나 Hougaard(1982)가 밝혔듯이 이 세 가지 조건을 동시에 충족시키는 자료변환은 존재하지 않는다.

간단한 예로 포아송 분포를 생각해 보자. 정규분포 가정 (i)을 위하여는 $Y^{2/3}$, 분산이 상수인 가정 (iii)을 위하여는 $Y^{1/2}$, 포아송자료를 분석할 때, 보통 많이 사용되는 곱모형(Multiplicative Model)을 선형모형이 되게하기 위하여는 log Y 변환이 필요하다. 자료변환의 문제점은 이와 같이 세 조건을 충족시키기 위하여 각기 다른 세 가지 자료변환이 요구된다는 점이다. 한편 일반회귀계수 추정치의 최소제곱합(Least Square)성질 등은 평균과 분산의 처음 두 적률(Moment)에만 관련되어 정규분포 성질과는 직접적 관련이 없으므로 특히 유의하여야 할 가정들로 (ii)와 (iii)을 생각할 수 있다.

종래 회귀분석은 종속변수 평균의 모형화에 촛점을 두어 왔다. 그러나 다구찌 방법의 대두로 평균 뿐만 아니라 분산에 대한 모형화가 중요하게 되어 Box(1986, 1988)는 자료변환을 통해 추구하여야 할 성질 (ii), (iii)을 다음과 같이 확장하였다.

(a) 절약성(Parsimony): 평균모형 및 분산모형을 가장 간단한 선형모형으로 표현,
(b) 분리성(Seperation): 평균과 분산 사이에 불필요한 함수적 종속 관계 제거.

이 두 성질을 고찰하기 전에 우선 유념해야 할 점은 앞에서도 지적했다시피 하나의 자료변환으로 이 두 가지 성질 모두를 충족할 수는 없다.

Box(1986, 1988)는 다구찌가 사용하는 주효과 계획들이 교호작용들을 고려할 수 없으므로 이들 교호작용들을 작게하는 즉, 절약성 성질을 만족하는 자료변환을 중시하는 것 같다. 그 밖에 León과 그외(1987), Nair 와 Pregibon(1986) 등의 학자들은 분리성을 강조하여 이를 만족하는 분산측도(Variance Measure)를 PerMIA(Performance Measure Indenpendent of Adjustment)라고 부른다.

다음과 같은 성질을 만족하는 확률변수 Y 를 생각하자.

$$E(Y) = \mu,$$
$$Var(Y) = \phi\mu^2$$

이 성질을 만족하는 분포의 예로 감마(Gamma) 분포를 들 수 있다. 이제 표본평균을 $\overline{Y}$, 표본분산을 S_Y^2 이라 하자. 다구찌의 소음신호비(Signal to Noise Ratio)는 다음과 같다.

$$SN = 10\log(\overline{Y}^2/S_Y^2)$$
$$\approx 10\log(\hat{\mu}^2/\hat{\phi}\hat{\mu}^2)$$
$$\approx -10\log\hat{\phi}.$$

그러므로 소음신호비 SN을 최대로 한다는 것은, 분산성분 ϕ 를 최소로 하는 하는것을 의미한다. León 과 그외(1987)는 ϕ 의 함수들의 측정치를 PerMIA라 부른다. 그러므로 다구찌의 소음신호비도 PerMIA 이다.

Box(1986, 1988)는 상기 다구찌의 SN 이 타당한 경우는 Z = log Y 가 합당한 자료변환일 때만이라고 하였다. 이 경우 Var(Z) $\approx \phi$ 의 성질을 만족하므로 Z 의 표본분산의 함수들이 모두 PerMIA 가 된다. 다구찌의 SN 은 다음과 같이 표현된다.

$$SN \approx -10\log S_Z^2.$$

종합하면 León 과 그외(1987)는 Y 가 감마분포를 할때 다구찌의 SN 은 PerMIA 이고, Box 는 Y 가 Log-Normal 분포를 따를때만 다구찌의 SN 이 합당한 분산측도라 했다. 감마분포와 Log-Normal분포는 $Var(Y) \approx \mu^2$ 일 경우의 자료분석 때 흔히 사용되는 분포들로써, 이중 한 분포가 사실(true)일 때 잘못 다른 분포를 사용해도 효율이 크게 떨어지지 않는다(Firth, 1988 참조). 그러므로 실제 자료 분석 시 이 두 분포들이 서로 비슷한 분석 결과를 제공하므로 두 분포중 어느 한 분포를 사용하여도 대개의 경우 큰 문제를 일으키지 않는다.

종합하면 다구찌의 소음신호비는 자료가 감마분포를 따를 때의 PerMIA이다. 이는 분리성을 만족하는 자료변환이라고 볼 수 있다. 주효과 계획법이 교호작용들을 고려할 수 없으므로 실험의 타당성을 위하여는 교호작용을 없애는, 즉 절약성 성질을 만족시키는 자료변환이 필요하다. 자료변환의 단점은 분리성과 절약성을 동시에 충족시키는 변환이 존재하지 않는다는 것이다. 또한 다구찌 분석에서 유의할 점은

$$-10\log\hat{\phi} = W_1\gamma_1 + W_2\gamma_2 + \cdots$$

이라는 관계로 $\hat{\phi}$ 의 log변환이 선형가법 성질(linear additive property)을 만족한다고 가정하는 것이다.

3. 일반화 선형모형

일반화 선형모형은 정규분포 가정을 하는 일반회귀분석의 확장으로 다음과 같은 로그우도함수를 가진다(McCullagh 와 Nelder, 1989).

$$l(\theta, y) = \{y\theta - b(\theta)\}/a(\phi) + c(y, \phi).$$

여기서 θ 를 canonical 모수라 하고, ϕ 를 산포모수(Dispersion Parameter)라고 부른다. 이러한 성질을 만족하는 분포들로는 정규분포 뿐 아니라 포아송, 이항분포, 감마분포, 음이항(Negative Binomial)분포, 역정규(Inverse Gaussian)분포들을 들 수 있다. 일반화 선형모형에서는 이 분포들의 평균과 분산을 다음과 같이 표현한다.

$$E(Y) = \mu,$$
$$Var(Y) = \phi V(\mu),$$

여기서, $V(\mu)$를 분산함수라고 부른다. 감마 분포의 경우 $V(\mu) = \mu^2$ 으로 전 장에서 다루었다. 그 밖에 이항분포는 $V(\mu) = \mu(1-\mu)$, 포아송분포는 $V(\mu) = \mu$로, 분포에 따라 여러 다른 종류의 분산함수가 존재한다. 상기 평균과 분산 관계식을 보면 일반화 선형모형에서는 분산함수를 정확히 지정함으로써 분리성을 얻을 수 있다. 전 장에서 살핀 바에 의하면 일반화 선형모형의 산포모수(Dispersion Parameter) ϕ의 추정치의 함수가 PerMIA 이다. 일반화선형모형에서는 평균의 연관함수(Link Function)을 통해 다음과 같은 선형모형을 얻는다.

$$\eta = g(\mu) = X_1\beta_1 + X_2\beta_2 + \cdots.$$

즉, 평균모형의 절약성을 연관함수의 지정으로 획득할 수 있다.

이제 일반화 선형모형의 분산 모형을 생각하여 보자. Box(1988)는 분산함수가 $V(\mu) = \mu^{\alpha}$로 주어지는 경우에 대하여 다룬 바 있고, Nair 와 Pregibon(1988a, 1988b)은 일반적으로 분산함수가 $V(\mu) = f(\mu)$ 로 주어지는 경우에 대한 자료변환을 다루었다. 일반화선형모형에서는 일반적인 분산함수를 Wedderburn(1974)의 Quasi-Likelihood 방법에 의해 다룰 수 있다. 이제부터 일반적인 분산함수 $V(\mu)$ 를 가정하자. 이 경우 Deviance는 다음과 같이 정의된다.

$$d = -2\int_y^{\mu} \frac{y-\mu}{V(\mu)}\,d\mu.$$

이를 이용하여 Nelder 와 Pregibon(1987)은 다음과 같이 Extended Quasi-Likelihood를 정의하였다.

$$-2Q = \sum\{\frac{d_i}{\phi_i} + \log 2\pi\phi_i V(y_i)\}.$$

만약 주어진 분산함수를 만족하는 지수군(Exponential Family)이 존재하는 경우 Extended Quasi-Likelihood는 그 지수군 분포의 안부점근사(Saddle-Point Approximation)가 된다. 좀더 자세한 성질은 Nelder 와 Pregibon(1987)을 참조한다. Extended Quasi-Likelihood를 이용하여 스코어함수(score function)를 만들면

$$\frac{\partial(2Q)}{\partial\gamma} = \sum\frac{d_i - \phi_i}{\phi_i^2}\,\frac{\partial\phi_i}{\partial\gamma}$$

을 만족한다. 이는 Deviance d_i 를 감마분포로 적합할 때의 스코어함수와 동일하다.

이제 $h(\phi)$ 를 산포모수 ϕ 의 연관함수라 정의하고, 다음과 같은 선형모형을 생각할 수 있다.

$$\xi = h(\phi) = W_1\gamma_1 + W_2\gamma_2 + \cdots.$$

독립변수 $W_1, W_2, \cdots$ 들은 평균모형에서의 독립변수 $X_1, X_2, \cdots$ 들과 같을 수도 있다.

이상 일반화 선형모형의 평균분산 동시모형은 다음과 같이 정리할 수 있다.

[표 1]

	평 균 모 형	분 산 모 형
종 속 변 수	Y	d
모 평 균	μ	ϕ
선 형 모 형	$\eta = X1\beta 1 + X2\beta 2 \cdots$	$\xi = W1\gamma 1 + W2\gamma 2 + \cdots$
연 관 함 수	$\eta = g(\mu)$	$\xi = h(\phi)$
분 산 함 수	$V(\mu)$	$\phi 2$

일반화 선형모형에서는 분산함수 $V(\mu)$의 선정을 통해 분리성을, 연관함수 $g(\mu)$의 선정을 통해 평균모형의 절약성을, 그리고 연관함수 $h(\phi)$의 선정을 통해 분산모형의 절약성을 얻는다. 일반화 선형모형 관점에서 다구찌 SN은 다음 표와 같이 요약될 수 있다.

[표 2]

종 속 변 수	평균 $\overline{Y}$	산 포 d
분 포	감 마	감 마
연 관 함 수	Log	Log

일반화 선형모형의 평균과 분산모형들의 회귀계수 추정에 대하여 생각해 보자. 평균모형의 계수들을 Quasi-Likelihood 방법으로 추정하면 그 때의 추정치들이 분포 가정이 틀린 경우에도 다른 추정치들 보다 강건(robust)하다(Lee 와 Nelder, 1992 참조). 분산모형의 계수들은 이 장에서 설명한 Extended Quasi-Likelhood 방법을 사용하는 것이 특히 다구찌 실험과 같은 소표본 자료의 분석에 좋다는 사실이 Nelder 와 Lee(1992)에 의하여 밝혀졌다. 이 장에서 설명한 일반화 선형모형 방법을 좀 더 자세히 알기 위해서는 Nelder 와 Lee(1991)을 참조하고 이를 수행하기 위한 GLIM Macro 는 부록을 참조한다.

4. 비 교

Quinlan(1985)은 Speedometer Cable의 수축에 영향을 미치는 15가지 공정 과정에 대해 다구찌 실험을 실시하였다. 사용한 실험계획법은 2^{15-11} = 16회의 주효과 계획법이다. Box(1988) 또한 이 실험에 대한 분석을 하였는데 주효과 계획에서 교호작용을 고려할 수 없으므로 절약성을 중시하여 log변환을 사용하였다. Cable의 축소는 비율이므로 2장에서 밝힌 바와 같이 이항분포의 분산함수 $\mu(1-\mu)$가 log변환의 경우 은연 중에 가정한 분산함수 $V(\mu)=\mu^2$보다 타당하다고 생각된다. 그러므로 Nelder와 Lee(1991)는 다음 표와 같은 모형을 가정하고 분석작업을 수행하였다.

[표 3]

	평 균 모 형	분 산 모 형
종 속 변 수	Y	d
연 관 함 수	log	log
분 산 함 수	$\mu(1-\mu)$	$\phi2$

일반화 선형모형을 사용하면 분산모형의 연관 함수를 log로 택하여 Box(1988)의 절약성을 고려하였을 뿐 아니라, 분산함수를 $\mu(1-\mu)$로 택하여 분리성 또한 고려하였다.

Quinlan(1985)의 자료에는 다구찌 실험전 공정과정에서의 Cable 수축율들의 평균과 표준편차 및 다구찌 실험 후 찾은 적정공정 과정에서의 평균과 표준편차가 조사, 제시되어 있다. 그러므로 이와 같이 실제 실험을 통하여 얻어진 수축율들의 평균과 표준편차를 Box(1988)의 자료변환법에 의한 예측치 및 Nelder와 Lee(1991)의 일반화 선형모형 분석에 의한 예측치와 비교하는 것이 흥미롭다. 그 결과는 다음과 같다.

[표 4]

	다구찌 실험 전		다구찌 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

Nelder와 Lee(1991)는 여러 가능한 다른 평균과 분산 모형에서도 위와 같은 예측치들을 구하였다. 상기값은 그들 논문에서 m_{10}, d_5라고 표현된 모형에서의 예측치이다. 다른 모형에서의 예

측치를 기재하지 않은 이유는 모형이 다르더라도 일반화 선형모형을 사용하면 거의 비슷한 예측치를 제공하기 때문이다. 표4에서 자료변환을 사용한 예측치들이 일반화 선형모형을 사용한 예측치들 보다 매우 나쁜 것을 알 수 있다. 이는 절약성만을 고려한 log 자료변환의 경우 타당치 못한 분산함수 $V(\mu) = \mu^2$ 을 은연 중 사용하게 되어 회귀계수 추정치들의 효율이 매우 떨어지기 때문이다.

다구찌 분석의 통계학 논문들은 거의 모두가 자료변환법에 관해 연구되어 왔었다. 본 논문에서 지적하였다시피 분리성과 절약성을 동시에 만족시킬 수 있는 자료변환이 존재하지 않기 때문에 일반화 선형모형의 사용이 선호된다. 그러나 이에 대한 본격적인 논문으로는 현재까지 Nelder 와 Lee(1991) 논문 한 편 뿐이므로 이에 대한 보다 면밀한 연구가 요구된다. 일반화 선형모형의 단점으로는 엔지니어들이 사용하기에는 난해한 통계적 지식이 필요하기 때문에 이 방법의 대중화가 어렵다는 점을 들 수 있다. 부록에 수록한 GLIM Macro가 이 면에서 도움이 되었으면 한다.

참 고 문 헌

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

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

[3] Box, G. E. P. (1986), "Studies in Quality Improvement: Signal to Noise Ratios, Performance Criteria and Statistical Analysis: Part I," Technical Report 11, University of Wisconsin-Madison, Center for Quality and Productivity Improvement.

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

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

[6] Firth, D. (1988) "Multiplicative errors: lognormal or gamma?" *Journal of the Royal Statistical Society*, **B**, **50**, 266-268.

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

[8] Lee, Y. and Nelder, J. A. (1992), "The Robustness of Estimators from Quasi-Likelihood and Quadratic Estimating Fucntions." A manuscript submitted for publication.

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

[10] Logothetis, N and Wynn, H. P. (1989), *Quality Through Design: Experimental Design, Off-Line Quality Control and Taguchi's Contributions*, Oxford: Clarendon Press.

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

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

[13] Nair, V. N. and Pregibon, D. (1988a), "Analyzing Dispersion Effects from Replicated Factorial Experiments," *Technometrics* , **30**, 247-257.

[14] Nair, V. N. and Pregibon, D. (1988b), "Discussion of the Paper by Box," *Technometrics*, **30**, 24-29.

[15] Nelder, J. A. and Lee, Y. (1991), "Generalized Linear Models for the Analysis of

Taguchi-Type Experiments." *Applied Stochastic Models and Data Analysis*, **7**, 107- 120.
[16] Nelder, J. A. and Lee, Y. (1992) "Likelihood, Quasi-likelihood and Pseudo-likelihood: Some Comparisons." *Journal of the Royal Statistical Society*, **B**, **54**, 273-284.
[17] Nelder, J. A. and Pregibon, D. (1987), "An Extended Quasi-Likelihood Function," *Biometrika*, **74**, 221-231.
[18] Quinlan, J. (1985), "Product Improvement by Application of Taguchi Methods," in *American Supplier Institute News* (special symposium ed.), Dearborn, MI: American Supplier Institute, 11-16.
[19] Taguchi, G. (1985), "Quality Engineering in Japan," *Communications in Statistics*, **A**, 14, 2785-2801.
[20] Taguchi, G. and Wu, Y. (1985), *Introduction to Off-Line Quality Control*, : Central Japan Quality Control Association, Nagoya, Japan
[21] Wedderburn, R. W. M. (1974), "Quasi-Likelihood Functions, Generalized Linear Models and the Gauss-Newton Method," *Biometrika*, **61**, 439-447.

부 록

```
!   Macros for modelling both mean and dispersion parameter
!                   Written by J. A. Nelder and Youngjo Lee
!   ---------------------------------------------------
!                     Outline of model
!                     ----------------
!  The model for the mean is a GLM for which var(Y)=phi*V(mu),
!  where phi is the dispersion parameter and V() the variance
!  function for the error distribution; as usual there is a link
!  function eta=g(mu), where eta is the linear predictor related
!  to covariates x1,x2,...,xk by eta=sum(beta(i)*x(i)).
!
!  The dispersion phi, as well as mu, may vary over runs; the model
!  for the dispersion has a dispersion link function zeta=h(phi),
!  where zeta is the dispersion linear predictor, related to
!  dispersion covariates u1,u2,...,up by zeta=sum(gamma(i)*u(i)).
!
!  The response variate for the dispersion is taken as the square
!  of either the Pearson or the deviance residual from the analysis
!  of the means (option set by %d in the macros) & its variance
!  function is taken as phi**2.  This can be modified for non-Normal
!  errors in the mean to allow for kurtosis by using the Prentice
!  weights as prior weights (option set by %p in macros).
!  Another possible adjustment to the dispersion equations allows for
!  the d.f. lost in fitting the parameters in the model for the
!  means (option set by %h in the macros).
! -----------------------------------------------------------------
!  To use the macros read the instructions in mdhelp.  The two models
!  are set up by putting the error,link, offset & linear predictor for
!  both models as the contents of standard macros.
!  WARNINGS -------------------------------------------------------
!  **** uses scalars %a %c %d %e %h %i %l %o %p %q %z %z1...4,8,9 ****
```

```
!            ****  prior weight must be set in pw__ first  ****
!  ---------------------------------------------------------------
!  Allows 8 forms for dispersion estimating equations viz.
!    all combinations of:
!  (i)   X2 or deviance component as disp. response (%d=0 or 1)
!  (ii)  Prentice weights or extended quasi-likelihood EQL (%p=0 or 1)
!  (iii) deviance adjustment factor off or on (%h=0 or 1)
!
$pr 'use mdhelp for help'!
!
$m mdhelp !
$pr '   For model for mean:- user must set l.p. in macro mlp!
 & change defaults, if needed, for!
     yvar in macro myv, error in merr, link in mlin & offset in moff!
   For model for dispersion:- user must set l.p. in macro dlp.!
 & change defaults, if needed, for!
     error in derr, link in dlin & offset in doff!
Prior weights are initialised to 1 in pw__,!
 so set pw__ to prior weights if not all 1!
Set %o=1 if using own macros for %a power var fn & %l power link fn!
Set %d=0 for X2, %d=1 for deviance, for yvar in fit of dispersion model!
Set %p=0 for Prentice weights, =1 for EQL for dispersion analysis!
Set %h=1 for factor (1-h) in dispersion equations!
Use %z9 to set no. of iterations in fit for mean (default=2)!
Use %z8 to set no. of iterations in fit for dispersion (default=3)!
Use su to set up, then use macro fit for one cycle,!
 f3 for 3 cycles ab initio, or rf3 for 3 cycles from previous fit.!
Use eqd_ to print EQD & prs_ to print summary of model!
Use sqd_ to save current value of EQD & d.f. when using f3 '$$e
!
!   default settings for merr,mlin, myv, moff ,derr, doff & dlin
$m myv y $e $m merr n$e $m derr g$e $m mlin i$e $m dlin l$e
$m moff $e $m doff $e
!   default setting of prior weight & switches for disp. p.w. etc.
!     & default no. of cycles for fitm & fitd
$ca pw__=1 : %c=1/6: %d=%h=%p=1 : %z9=2 : %z8=3!
$var 1 cdf_ ceqd df__ eqdf leqd lqdf pld_ plm_!
$ca cdf_=ceqd=df__=eqdf=leqd=lqdf=pld_=0!
!
$m su !      sets up for fit
$del vfy_ di__ di0_ x2__ dr__ qd__ mwt_ dwt_ vlm_ wm__ wd__ wtm_!
mfv_ dfv_ dyv0 dyv_ yf__ hfac!
$ca mwt_=dwt_=mfv_=dfv_=hfac=1:%z=0$$e
!
$m fitm !    fits model for means
$wa $ca wm__=mwt_*pw__!
$y #myv $err #merr $lin #mlin $off #moff $sca 1 $wei wm__!
$t t %yv m i yvm_ $ca %fv=%if(%z==0,%if(%o==0,%yv,(%yv+yvm_)/2),mfv_)!
$swi %o sown!
$pr'---  model for mean'$rec %z9$f #mlp $ext%vl!
$ca mfv_=%fv:df__=%df: vlm_=%vl:wtm_=%wt:plm_=%pl!
$u dinc!    calculates yvar for dispersion as X2 or deviance component
$ca dyv0=%if(%d==0,x2__,%if(di__>0,di__,x2__))+(pw__==0):dyv_=dyv0/hfac$$e
```

```
!
$m fitd !    fits model to dispersion using EQL (modified if %h=1);
!            assumes model for means fitted
$ca hfac=vlm_*wtm_*mwt_:%z4=%cu hfac:hfac=1-%h*hfac*plm_/%z4!
 :  dyv_=dyv0/hfac!        allows use of (1-h) factor
$u cdw_!    allows prior weight for dispersion if %p=0, none if %p=1
$ca wd__=dwt_*pw__*hfac $wa!
$y dyv_ $off #doff $wei wd__ $err #derr $lin #dlin!
$t t %yv m i yvm_ $ca %fv=%if(%z==0,yvm_,dfv_)!
$pr'---  model for dispersion'$rec %z8 $sca 2 $f #dlp!
$ca pld_=%pl:mwt_=1/%fv:dfv_=%fv:%z=%z+1 $u eqd_ $$e
!
$m cdw_ !   calculates prior weights for dispersion based on dist. of y
$ca dwt_=1 $exit %p $swi %err ndw_ pdw_ bdw_ gdw_ $$e
!
$m ndw_ $$e
$m pdw_ $ca dwt_=1/(1+dfv_/(2**fv))$$e
$m bdw_ $ca dwt_=1/(1+0.5*dfv_*(%bd/(%fv*(%bd-%fv))-6/%bd))$$e
$m gdw_ $ca dwt_=1/(1+3*dfv_)$$e
!
$m sown !    sets macros for power var fn & link
$own m1__ m2__ m3__ m4__!
$ca %lp=%if(%l==0,%log(%fv),%fv**%l)$$e
!
!            macros for own with power variance function (power %a)
!            and power link (%l)
$m m1__ $ca %fv=%if(%l==0,%exp(%lp),%lp**(1/%l))$$e
$m m2__ $ca %dr=%if(%l==0,1/%fv,%l**fv**(%l-1))$$e
$m m3__ $ca %va=%fv**%a$$e
$m m4__ !    calculates %di for var. fn mu**%a
$ca yf__=%yv/%fv:%z3=(1000-%tr(1000**a+.01))/1000!
:%z2=(2000-%tr(1000**a+.01))/1000!
:%di=2*((/%z3)*(yf__**log(yf__)-yf__+1)**fv +(/%z2)*(yf__-%log(yf__)-1)!
+(%z3&%z2)*((yf__**%z2-%z2*yf__+%z3)*%fv**%z2)/(%z3**z2))$$e
!
$m eqd_ !    calculates extended quasi-deviance
$ca %q=%cu(qd__=(pw__/=0)*(dwt_*dyv_/dfv_+%log(2**pi*vfy_*dfv_))*hfac)!
 : eqdf=df__-pld_ $pr'cycle '*i %z: 'EQD ='%q' with '*i eqdf' d.f.':!
'---------------------------'$$e
!
$m sqd_ !    saves current value of EQD & d.f. from %q & eqdf
$ca leqd=%q :lqdf=eqdf$$e
!
$m prs_  !    prints summary
$pr'(1-h) switch (%h) '*i %h,1'  :deviance switch (%d) '*i %d,1!
' :disp. prior weights (%p) '*i %p,1!
$ca cdf_=eqdf-lqdf:ceqd=%q-leqd:%o=%o+1 $swi %o prs1 prs2 $ca%o=%o-1!
$pr' disp ---  link 'dlin ' :error 'derr' :offset 'doff' :l.p. 'dlp!
:'change in EQD ='ceqd' with '*i cdf_ ' d.f.'$$e!
!
$m prs1 !
$pr' mean ---  link 'mlin ' :error 'merr' :offset 'moff' :l.p. 'mlp$$e
$m prs2 !
```

```
$pr ' mean --- link power '%l' :var fn power '%a' :l.p. 'mlp$$e
!
$m fit !      does one cycle of mean and dispersion fit
$u fitm $u fitd$$e
!
$m f3 !       initializes, does three cycles of fit and prints summary
$u su $u fit$u fit$u fit$u prs_$$e!
!
$m rf2 !      refits 2 cycles without initializing & prints summary
$u sqd_$u fitd$u fit$u prs_$$e
!
$m rf3 !      refits 3 cycles without initializing & prints summary
$u sqd_$u fitd$u fit$u fit$u prs_$$e
!
$m dinc !     calculates V(y), deviance & X2 components, & deviance
!             residuals for 4 distributions & own with power link
!             and power var. fn.
$swi %err nvd_ pvd_ bvd_ gvd_ ovd_ ovd_ ovd_ ovd_ ovd_!
$ca dr__=(2*(%yv>%fv)-1)*%sqrt(di__/dfv_)$$e
!
$m nvd_ !    sets up V(y) & X2 and dev components for Normal
$ca vfy_=1:di__=pw__*(%yv-%fv)**2:x2__=di__$$e
!
$m pvd_ !    sets up V'(y) & X2 and dev components for Poisson
$ca vfy_=%yv+%c:di__=%yv/%fv:di__=2*pw__*%fv*(di__*%log(di__)+1-di__)!
:x2__=pw__*((%yv-%fv)**2)/%fv$$e
!
$m bvd_ !    sets up V'(y) & X2 and dev components for binomial
$ca vfy_=(%yv+%c)*(%bd-%yv+%c)/(%bd+%c)!
: x2__=pw__*((%yv-%fv)**2)*%bd/(%fv*(%bd-%fv))!
: di__=2*pw__*(%yv*%log(%yv/%fv))!
: di__=di__+2*pw__*(%bd-%yv)*%log((%bd-%yv)/(%bd-%fv))$$e
!
$m gvd_ !    sets up V(y) & X2 and dev components for gamma
$ca vfy_=%yv**2:di0_=2*pw__*(%log(%fv/%yv)+(%yv-%fv)/%fv)!
  : x2__=pw__*((%yv-%fv)/%fv)**2:di__=%if(di0_<0,x2__,di0_)$$e
!
$m ovd_ !    sets up V(y) &  X2 & dev components for own
$ca vfy_=%yv**%a : di__=pw__*%di : x2__=pw__*((%yv-%fv)**2/%va)$$e
!
$ret
$fin
```

Generalized Linear models versus Data Transformation for the Analysis of Taguchi Experiment[1)]

Youngjo Lee[2)]

Abstract

Recent interest in Taguchi's methods have led to developments of joint modelling of the mean and dispersion in generalized linear models. Since a single data transformation cannot produce all the necessary conditions for an analysis, for the analysis of the Taguchi data, the use of the generalized linear models is preferred to a commonly used data transformation method. In this paper, we will illustrate this point and provide GLIM macros to implement the joint modelling of the mean and dispersion in generalized linear models.

---

1) This research is supported by the Korean Science and Engineering Foundation.

2) Department of Statistics, Hallym University, Chunchon, Korea