응용통계연구 제 **21**권 2호, *2008*년, *pp. 291–299*

반복측정의 이가반응 자료에 대한 로짓 모형

최재성[1)]

요 약

동일 개체가 여러 시점에서 반복되어 측정될 때, 측정값들 간에 종속성을 예상할 수 있다. 본 논문은 한 개체의 이가 반응변수가 g개 시점에서 관측될 때, 종속적인 g개 이가변수들의 다변량 분포로부터 각 시점에서의 주변분포의 동질성을 파악하기 위한 로짓모형을 제시하고 자료분석방법을 제공하고자 한다. 모형과 관련된 가정으로 반복측정이 행해지는 g개 시점은 각기 서로 다른 요인 또는 공변량의 결합수준들로 구성된다고 가정한다. 또한, 모형에서 고려된 처치들이 반복측정에 기인하는 서로 다른 크기의 실험단위들에 행해질 때 모수들을 추정하기 위한 방법으로 가중최소제곱법을 다루고 있다. 여기서 가중최소제곱법은 반응변수들의 종속성으로 인한 공분산 구조에 근거한 모형내 모수들의 효과를 효율적으로 추론하기 위해 이용된다. 제시된 모형은 주변로짓을 이용함으로써 단순히 주변확률분포의 동질성에 대한 검정뿐만 아니라 모형의 타당성 및 요인들의 수준변화에 따른 효과를 파악하기 위한 효과적인 모형임을 보여준다.

주요용어: 주변로짓, 로짓모형, 반복측정, 공분산 구조.

1. 서론

실험 또는 관측조사로부터 한 개체가 여러 시점에서 개체의 반응을 반복하여 측정 하는 경우를 생각해 볼 수 있다. 반복측정의 이가 반응자료란 개체의 한 특성을 나타내는 반응변수가 두 개의 관측값 중 하나로 관측되고 여러 시점 또는 조건에서 동일 반응변수에 대해 관측값들이 주어진 자료를 의미한다. 동일 개체에 대해 반복측정되므로 반응값들 간에 종속성을 갖게 된다. 예를 들면, 관심 모집단에서 확률표본으로 취해진 개체들을 대상으로 여러 시점에서 반복하여 관측할 때, 수집된 자료는 관측값들 간에 상관성을 갖게 되는 종속적인 표본자료가 된다. 여기서는 반응변수가 두 개의 값만을 취하는 이가 반응변수를 가정하므로 이가의 종속적인 표본자료를 가정하게 된다. 반복측정의 이가 반응자료에 대해 자료간의 종속성을 감안한 다양한 자료분석방법들이 많은 문헌에서 논의되고 있다. Liang과 Zeger (1986)는 일반화된 선형모형으로 적합된 종속자료를 분석하는 방법을 소개했다. 이 방법은 이가자료나 Poisson자료와 같은 이산 반응자료에 이상적이다. Koch와 Reinfurt (1971), Koch 등 (1977)은 처음으로 가중최소 제곱 방법을 반복측정의 범주형 자료에 적용했다. 반복측정의 범주형 반응자료의 구조적 특성을 고려한 다양한 모형들 및 분석방법들은 Agresti (1990)에서 논의되고 있으며, Im과 Gianola (1988)는 이원지분계획으

1) (704-701) 대구광역시 달서구 신당동 1000번지, 계명대학교 자연과학대학 통계학과, 교수.
E-mail: jschoi@kmu.ac.kr

로부터 발생하는 분산성분들을 추정하기 위하여 이항자료에 대한 혼합효과 모형을 다루고 있다. Milliken과 Johnson (1984)은 연속변수의 반복측정자료에 대해 다양한 모형과 자료분석 방법을 다루고 있다. 반복측정의 자료에서 일반적인 관심은 각 시점에서 반응변수들의 주변분포의 분석에 있다. 따라서, k개 범주를 갖는 반응변수가 g개 시점에서 측정되는 일반적인 상황에서 $g(k-1)$개 가능한 결과들이 있게 되고 자료분석에서는 $g(k-1)$개의 종속적인 주변비율, 일반화된 로짓, 또는 누적로짓을 생각할 수 있다.

반복측정의 자료분석은 그렇지 않은 자료 즉, g개의 모집단에서 독립적으로 취해진 독립인 표본자료의 분석과는 두 가지 주요 차이점이 있다. 하나는 동일 개체에서 행해진 반복측정들 간의 종속성이고, 두 번째는 연구자가 측정값들을 얻기 위한 실험환경을 조절할 수 없다는 점이다. 따라서 자료가 불균형(unbalanced)이거나 부분적으로 불완전(partially incomplete)일 수 있다. 예로써, 경시적 연구(longitudnal study)에서 개체로 부터의 반응이 하나 또는 여러 시점에서 관심결과와 관련없는 요인들로 인해 관측되지 않을 수 있다. 실험환경에서는 이러한 결측치(missing data)가 발생할 수 있다. 독성학이나 유전학에는 실험동물의 한배의 새끼수가 다를 수 있다. 즉, 반복측정의 수가 실험단위들 마다 일정하지 않을 수 있다.

본 연구는 개체에 대한 반응이 두 개의 값만을 취하는 이가의 반응변수이고 동일 개체가 반복측정요인의 하나인 시간요인 C의 T개 시점에서 측정된다고 가정한다. 앞서의 g개 시점은 아무런 구조도 갖지 않는 단순히 반복측정이 행해지는 시간요인의 수준들의 수 또는 관측시점의 구조를 나타내는 반복측정요인들의 결합수준의 수를 의미할 수 있다. 이와는 달리, 여기서는 구체적으로 반복측정의 한 요인인 시간요인의 T개 개별 관측시점에 종속하는 서로 다른 공변량을 갖는다고 가정한다. 따라서, 종속성을 띄는 이가 반응자료를 분석하기 위한 모형으로 시간종속적인(time-dependent) 공변량의 효과를 포함하는 모형을 제시하고 분석하는 방법을 제공하고자 한다.

2. 모형에 대한 가정

본 연구에서 관심을 갖는 반복측정자료의 분석을 위해 제시하고자 하는 로짓모형과 관련된 가정들을 살펴보기로 한다. 동일한 개체 또는 실험단위들이 연구자가 임의로 배정할 수 없는 시간상의 시점들에서 혹은 다단계의 실험과정으로부터 반복적으로 측정됨을 가정한다. 따라서, 동일 개체의 반복측정으로 인한 반응변수간의 종속성을 가정한다. 반복측정요인으로 간주되는 시간은 반복측정이 행해지는 시점에 종속하는 서로 다른 유형의 공변량이 존재함을 가정한다. 반복측정이 행해지는 시점마다 서로 다른 공변량의 가정이 모형설계의 착안점이다. 이러한 가정과 관련된 모형제시 및 자료분석 예는 찾아보기가 쉽지 않은 점에 유의하고 있다. 반복측정의 시점간에 동일하지 않은 공변량을 가정함으로써 로짓모형내 이들 공변량의 수준효과를 어떻게 포함시킬 것인가를 고려한다. 각 측정시점에서 관측되는 공변량의 수준 수는 동일할 수도 있고 그렇지 않을 수도 있음을 가정한다. 각 시점에서 종속하는 공변량은 어떤 구조적 특성을 갖고 있지 않음을 가정한다. 또한 개체의 반응에 영향을 미칠 수 있는 다수의 범주형 독립변수들이 있음을 가정한다. 그리고 주변확률

의 주변 로짓변환의 주기적 추세변화에 대해 영향을 미칠 것으로 고려된 독립변수들 및 반복측정요인들의 효과를 알아보기 위한 로짓 선형모형을 가정한다. 이러한 가정하에 모형설계를 위해 개체의 반응변수를 Y라 두자. Y는 두 개의 값만을 취하는 이가의 반응변수이다. 반복측정요인으로 시간요인 C와 C의 각 시점 $(t = 1, 2, \ldots, T)$에서 관측되는 공변량이 존재하고 다른 시점에서의 공변량은 서로 다르다고 가정한다. 이들 공변량들은 모두 범주형 변수들로 가정한다. C의 한 시점 t에서의 공변량 X_t는 w_t개의 범주를 갖는 범주형 변수라 두자. 따라서, 반복측정자료는 시간요인 C와 C의 각 수준에서 결부된 공변량의 결합수준으로부터 반복측정된 자료로 주어진다. 반복측정이 행해지는 전체 관측시점의 수를 g라 두면 $g = \sum_{t=1}^{T} w^t$이 된다. 모형설계와 관련된 가정들을 내포하고 있는 반복측정 자료의 구조와 모형구축 과정을 모발치료제의 예로써 구체적으로 논의하고 분석하는 방법을 제시하고자 한다.

3. 자료의 구조

2절에서 논의된 모형가정과 관련된 자료의 구조를 살펴보기 위하여 모발치료제로 개발된 제품의 효과를 알아보기 위한 실험상황을 가정한다. 개체의 반응(Y)은 효과적이다 또는 그렇지 않다의 두가지 반응을 나타낸다고 가정한다. 따라서 치료제품에 대한 반응은 두 개의 관측값만을 취하는 이가 반응이다. 모발치료제를 이용하는 개체는 장기간의 제품사용에서 주기적으로 그 효과를 판단하기 위한 측도로 서로 다른 공변량을 이용한다고 하자. 즉, 치료제품을 1개월, 2개월,..., T개월 사용후 모발의 생성여부에 대한 판단근거로 관측시점에 종속되어 이용되는 공변량의 수준 수를 고려할 때, 개체의 반응은 시간요인 C의 T개 시점과 해당하는 공변량의 수준결합에서 관측된다. 따라서, 개체에 대한 반응은 $g = \sum_{t=1}^{T} w^t$개에서 행해진다. 이들 관측값들은 동일 개체에서의 반응값들 이므로 반응값들 간에 상관성을 가정할 수 있게 된다. 모발치료를 원하는 개체들의 집단에서 일부의 개체들을 확률표본으로 추출하여 이들 개체들을 대상으로 월별로 반응을 조사하게 되면 2^g개 가능한 결과들에 대한 자료표를 얻게 된다. 치료제품의 효과에 영향을 미치는 변수로 탈모정도, 탈모유형, 성별, 연령 등을 고려할 수 있다. 탈모정도(U_g)는 탈모초기증상, 탈모시작, 탈모진행의 세 부류로 분류한다. 변수 U_g는 고려된 다른 독립변수들과는 달리 관측시점 g에서 이용된 공변량임을 나타낸다. 탈모유형(I)은 A형과 B형으로 나눈다. 성별(J)간에 차이가 있을 수 있으므로 남성과 여성으로 구분한다. 탈모가 중장년에서의 질환으로 간주될 때, 연령(H)은 30대, 40대, 50대로 분류할 수 있다. 모발치료제의 효과를 조사하기 위한 목적의 반복측정된 이가반응 자료는 각 시점에서의 주변확률분포에서 효과를 나타내는 주변확률이 시간이 경과함에 따라 어떻게 변화하고 있는 가를 파악하는 데 있다. 그리고 주변확률에 영향을 미치는 변수들로 인해 어떻게 어느 정도 영향을 받고 있는 가를 조사하는 데 있게 된다. 동일 개체로 부터의 반복적인 측정으로 인해 야기되는 변수들 간의 종속성으로 인해 공분산 구조를 감안한 자료분석 모형이 제공되어야 한다. 탈모정도는 각 시점과 관련된 요인으로 간주한다. 다른 독립변수들 과는 달리 모발 치료제를 평가하는 시점에서의 탈모정도를 나타내는 변수이기 때문이다. 모발치료제의 장기간 이용은 초기에는 신체의 부작용을 감지할

표 3.1: 반복측정의 이가반응 자료구조표

성별	탈모유형	연령	Time1 탈모정도 1	2	3	Time2 부작용정도 1	2	3	표본비율
m	A	30	pos	pos	pos	pos	pos	pos	p_{111111}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	B	50	pos	neg	neg	neg	neg	neg	p_{100000}
f	A	30	neg	pos	pos	pos	pos	pos	p_{011111}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	B	50	neg	neg	neg	neg	neg	neg	p_{000000}

수 없으나 일정기간 뒤에는 부작용을 유발할 수도 있다. 이러한 부작용의 정도에 대한 관측이 어느 시점에서 관측된다고 하면 또 다른 요인을 추가할 수 있게 된다.

본 연구에서 논의되는 자료의 구조를 단순하게 파악하기 위하여 다음과 같이 가정한다. 반복측정요인들로 시간요인 C와 C의 두 시점 (T=2)에서 각각 탈모정도 U_g, 부작용의 정도 V_g를 나타내는 공변량을 고려한다. 이들 요인들의 결합수준에서 관측되는 개체의 반응변수는 효과가 있으면 pos로 아니면 neg로 표시한다. 시점 1과 관련된 공변량으로 탈모정도를 나타내는 변수는 세 개의 수준이 순서형이므로 1, 2, 3으로 나타낸다. 탈모유형 I는 두 개의 범주 A와 B를 취한다. 성별 J는 m과 f로 나타낸다. 연령 H는 30, 40, 50으로 표기한다. 시점 2에서 부작용의 정도를 나타내는 변수 V_g도 그 정도에 따라 1, 2, 3의 세 수준을 가정한다. 시간요인의 두 시점에서 서로 다른 요인들을 고려한 이유는 치료제의 사용시점에서 탈모정도에 따른 효과를 조사하고 장기간 사용에 따른 부작용에 의한 효과를 파악하기 위함이다.

많은 변수들을 고려하고 있기 때문에 자료구조의 단순성을 위해 간단한 구조의 반복이가 자료를 먼저 살펴보기로 한다. 두 개의 시점에서 반복측정된 이가반응이 두 그룹에서 관측된 자료의 형태는 표 3.1과 같다.

이 예에서 108개의 반복측정값들이 3개의 반복측정 요인들과 3개의 독립변수들로부터 구해진다. 반복측정요인들의 수준수가 증가할 때 결과로 주어지는 반응 프로필(response profile)의 수는 급격히 증가함을 알 수 있다. 또한, 시점에서 다른 공변량을 고려해야 할 때, 기존의 시간종속적인(time-dependent) 요인들과의 결합수준으로 주어짐에 유의할 필요가 있다.

4. 모형에 관한 논의

가정된 실험환경으로부터 주어진 반복측정된 이가 반응자료를 분석하기 위한 모형을 생각해 보기로 한다. 3절의 예에서, 시간요인의 두 시점과 수준이 3인 두 공변량과의 결합수준들로 주어지는 반복측정의 6개 결합시점 $g = 1, 2, 3, 4, 5, 6$에서의 관측결과들의 확률들은

$$\pi_{111111}(hij) = P(Y_1 = \text{pos}, Y_2 = \text{pos}, Y_3 = \text{pos}, Y_4 = \text{pos}, Y_5 = \text{pos}, Y_6 = \text{pos}; hij),$$
$$\pi_{111110}(hij) = P(Y_1 = \text{pos}, Y_2 = \text{pos}, Y_3 = \text{pos}, Y_4 = \text{pos}, Y_5 = \text{pos}, Y_6 = \text{neg}; hij),$$
$$\vdots$$
$$\pi_{000000}(hij) = P(Y_1 = \text{neg}, Y_2 = \text{neg}, Y_3 = \text{neg}, Y_4 = \text{neg}, Y_5 = \text{neg}, Y_6 = \text{neg}; hij)$$

이다. 이들 칸 확률들을 이용한 주변확률 $\phi_1(g; hij)$을 구하기 위해 $\boldsymbol{\pi}$를 다음과 같이 정의한다.

$$\boldsymbol{\pi} = (\pi_{111111}(hij), \pi_{111110}(hij), \ldots, \pi_{000000}(hij)),$$

이때, $\phi_1(g = 1; h = 30; i = A, j = m) = [\mathbf{1}, \mathbf{0}, \ldots, \mathbf{0}]'\boldsymbol{\pi}$이다.

주변확률벡타는 선형변환행렬 $\mathbf{A}$와 $\boldsymbol{\pi}$를 이용하여 다음과 같이 주어진다.

$$\boldsymbol{\phi} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{72} \end{bmatrix} = \mathbf{A}\boldsymbol{\pi},$$

이때, 공변량의 수준결합에서 관측되는 시점에서의 주변로짓(marginal logit)은 $L(g; hij) = \log[\phi_1(g; hij)/\phi_2(g; hij)]$ 이다.

따라서, 자료분석을 위해 이용가능한 로짓 선형모형의 하나는

$$L(g; hij) = \alpha + \beta a_h + \tau_i + \gamma_j + \delta_g \tag{4.1}$$

로 주어진다. 여기서 α는 평균 주변로짓을 나타내는 절편이고 β는 양적변수인 연령의 효과를 나타내는 회귀변수이다. τ_i와 γ_j은 각기 탈모유형과 성별의 수준효과를 나타내고 있다. δ_g는 시점 g에서의 시기효과를 나타낸다. 모형 (4.1)에서 고려된 미지모수들에 대한 추론은 처치들이 서로 다른 크기의 두 실험단위에 행해졌으므로 구분해서 추론할 필요가 있다. 모형 (4.1)은 그 효과의 추론에 있어 두 부분으로 나눌 수있다.

$$\begin{aligned} L(g; hij) &= \alpha + \beta a_h + \tau_i + \gamma_j \\ &\quad + \delta_g, \end{aligned} \tag{4.2}$$

모형의 위 부분은 전체크기의 실험단위인 개체간의 변동을 나타내는 공분산 행렬에 근거해야 하고 아래 부분의 δ_g에 대한 추론은 작은 크기의 실험단위인 개체내 변동을 나타내는 공분산 행렬에 근거하여 추론함이 옳다.

모형 (4.1)은 시점 g가 두 요인의 결합수준으로 구성되고, 두 요인은 순서를 나타내는 수량범주이므로 양적변화에 따른 효과를 고려할 때 다음과 같이 모형을 변화할 수 있다.

$$L(g; hij) = \alpha + \beta a_h + \tau_i + \gamma_j + \beta_{1g} u_g + \beta_{2g} v_g, \tag{4.3}$$

단, β_{1g}는 탈모정도의 수준(u_g)에 따른 회귀변수이고, β_{2g}는 부작용의 정도를 나타내는 수준(v_g)에 따른 회귀효과를 나타낸다.

모형 (4.3)도 모형(4.2)와 유사하게 두 부분으로 구분한다.

$$\begin{aligned} L(g;hij) &= \alpha + \beta a_h + \tau_i + \gamma_j \\ &\quad + \beta_{1g} u_g + \beta_{2g} v_g, \end{aligned} \tag{4.4}$$

모형 (4.4)에서의 두 계층의 모수들은 각기 타당한 공분산 행렬을 이용하여 추론되어야 한다.

위와 같은 모형을 적합시키기 위하여 최우추정법을 이용하는 것은 단순하지 않다. 개별 범주의 확률을 이용하는 다항우도의 최우추정법은 주변확률을 이용하는 모형에 적합하지 않다. Koch 등 (1977)은 주변확률을 이용하는 모형을 적합시키기 위해 가중최소제곱법(WLS)을 이용했다. 반복측정된 이가 반응자료를 주변로짓 모형하에 분석할 때 유의할 점은 반복측정요인인 시간이 실험에서 개체에 임의로 배정되지 않는다는 점에 주목해야 한다. 확률화에 의한 임의 배정이 실험자에 의해 조정되지 않기 때문에 일반적인 분할구(split-plot) 분석에서 타당한 가정들을 따르지 않는 공분산 구조를 갖게 된다.

일반적으로 반복측정계획들은 적어도 둘 이상의 서로 다른 크기의 실험단위를 포함하는 구조를 갖는다. 여기서 개체는 처치구조의 한 요인인 시간상에서 측정되고 있다. 동일 개체를 서로 다른 시점에서 측정함으로써 개체는 본질적으로 각 시점의 개체로 분할되고 반응은 각 시점에서 측정된다. 따라서, 전체 시간에 노출된 개체는 크기가 큰 실험단위로 간주되고 부분적인 시점에서 관측되는 동일 개체는 작은 크기의 실험단위로 간주된다. 이런 관점에서 모형과 분석은 두 개의 부분으로 나누어 지게 된다. 특히, 시간은 실험자 임의로 조절할 수 없으므로 공분산 구조하에서 반복측정된 자료를 분석해야 한다는 점을 염두에 두어야 한다.

5. 모발치료 자료예

3절에서 논의된 자료구조를 갖는 반복측정의 이가 반응자료를 분석하는 방법을 생각해 보기로 한다. 반복측정 요인들의 수준수가 증가할 때 결과로 주어지는 반응 프로필(response profile)의 수는 급격히 증가함을 알 수 있다. 또한, 시간요인의 각 시점에 종속하는 공변량을 고려해야 할 때, 관측시점들은 시간요인의 수준들과 시간종속적인(time-dependent) 공변량과의 결합수준으로 주어짐에 유의할 필요가 있다. 많은 변수들을 고려하고 있기 때문에 자료구조의 단순성을 위해 간단한 구조의 반복측정의 이가 반응자료를 분석해 보기로 한다. 시간요인의 두 시점에서 반복 측정된 이가반응이 두 그룹에서 관측된 자료가 다음과 같다고 하자.

위 자료를 분석하기 위하여 식 (4.2)를 자료에 적합한 모형으로 변형하여 적용하여 보기로 한다. 식 (4.2)는 다음과 같이 변형될 수 있다.

$$\begin{aligned} L(g;hij) &= \alpha + \gamma_j \\ &\quad + \delta_t + \delta_w + \delta_{tw}, \end{aligned} \tag{5.1}$$

여기서 α는 주변로짓의 평균을 나타내는 절편이고, γ_j는 그룹 j의 효과이다. 반복측정의 시점 $g = \{(t,w);(1,1),(1,0),(2,1),(2,0)\}$의 4개의 결합수준으로 주어진다. δ_t는 시점 t에서

표 5.1: 모발치료의 생성자료

그룹1					그룹2				
Time=1			Time=2		Time=1			Time=2	
탈모정도 (보통=1 심함=0)			부작용유무 (없음=1 있음=0)		탈모정도 (보통=1 심함=0)			부작용유무 (없음=1 있음=0)	
(1,1)	(1,0)	(2,1)	(2,0)	n_g	(1,1)	(2,1)	(2,1)	(2,2)	n_g
1	1	1	1	180	1	1	1	1	160
1	1	1	0	4	1	1	1	0	8
1	1	0	1	17	1	1	0	1	25
1	1	0	0	3	1	1	0	0	3
1	0	1	1	10	1	0	1	1	12
1	0	1	0	7	1	0	1	0	7
1	0	0	1	1	1	0	0	1	1
1	0	0	0	8	1	0	0	0	8
0	1	1	1	220	0	1	1	1	200
0	1	1	0	120	0	1	1	0	100
0	1	0	1	18	0	1	0	1	14
0	1	0	0	7	0	1	0	0	8
0	0	1	1	8	0	0	1	1	6
0	0	1	0	5	0	0	1	0	4
0	0	0	1	9	0	0	0	1	9
0	0	0	0	170	0	0	0	0	180

의 시기효과를 나타내고, δ_w는 시점 t에서의 요인 w의 효과이다. δ_{tw}는 시점 t에서의 시기효과와 요인 w의 효과간의 교호작용을 나타낸다.

모형 (5.1)을 표 5.1의 자료에 적합시켰을 때, 그룹 주효과는 유의하지 않는 것으로 주어졌다. 따라서, 그룹효과를 제외한 축소모형은 다음과 같다.

$$\begin{aligned} L(g;j) &= \alpha \\ &\quad +\delta_t + \delta_w + \delta_{tw}, \end{aligned} \tag{5.2}$$

모형 (5.2)내의 모수들의 추정값과 추정오차는 다음과 같이 주어진다. $\hat{\alpha}$ =1.44 (0.067), $\hat{\delta}_t = -0.55$ (0.034), $\hat{\delta}_w = -0.64$ (0.036)이고 $\hat{\delta}_{tw} = -1.22$ (0.050)이다.

잔차에 대한 자유도는 4이고 p값은 0.7369이므로 모형은 적합한 것으로 간주된다. 또한, 모든 모수들이 유의수준 0.001에서 유의함을 나타내므로 주변확률의 동질성은 없다고 판단한다. α에 대한 추정오차는 개체간의 변이가 독립이라는 가정으로 부터 추정된 표준오차이고 나머지 모수들에 대한 표준오차는 반복측정으로부터 오는 개체내의 종속성에 따른 공분간 구조에 근거하여 추정된 오차이다. 즉, 개체간 요인 (그룹)은 독립인 측정치를 갖게 되고 개체내 요인 (시점)은 종속적인 측정값을 취하게 된다는 가정 때문이다. 종속성을 고려하지 않으면 잘못된 표준오차를 얻게 되고, 양의 상관이 존재할 때 개체간 효과의 표준오차는 작게 추정되고 개체내 효과의 표준오차는 과대추정함으로 인해서 효과들의 추론에 있어 비효율적일 수 있다. 참고로, 자료로부터 구해진 주변로짓 벡타는 $L(g;j)' = [0.292\ 0.723\ 0.704\ 0.588\ 0.301\ 0.695\ 0.678\ 0.573]$이다.

6. 결론

본 논문은 실험 또는 관측조사를 통하여 수집되는 자료가 반복측정의 이가반응 자료의 분석을 위한 로짓모형을 고려하고 있다. 개체의 이가 반응변수가 여러 시점에서 반복적으로 측정되기 때문에 관측치간에 종속성을 갖게 되는 경향이 있다. 그러나 개체들은 확률표본으로 추출되기 때문에 개체간의 변이는 독립으로 간주할 수 있다. 따라서 하나의 개체가 g개의 시점에서 관측될 때, 두 개의 서로 다른 크기의 실험단위를 갖는 실험으로부터 자료가 수집되었음을 의미한다. 또한, g개의 시점은 시간요인과 다른 요인과의 결합수준들로 주어질 수 있음을 보여주고 있다.

반복측정계획으로부터 수집되는 자료들이 일반적으로 서로 다른 크기의 실험단위를 갖는 분할구 실험계획이나 지분계획등의 자료들과는 달리 시간이 반복측정의 한 요인인 경우 관측시점을 연구자 임의로 조절할 수 없다는 점에 있다. 따라서, 시점에 따른 개체내 관측값 또는 측정값들 간의 변이가 일정한 상관성을 보이지 않는다는 점에 유의하고 있다. 그러므로 개체간의 변이에 따른 공분산 구조와 개체내 변이에 따른 공분산 구조를 고려하고 있다. 또한 시간이 반복측정의 요인일 때 시간과 함께 변하는 공변량이 존재하며 서로 다른 시점에 다른 공변량을 가정할 수 있다는 점을 논의하고 있다. 이러한 가정하에 수집된 자료에 대한 분석모형으로 로짓모형을 가정하고 모형내 미지모수들에 대한 추론을 다루었다. 구체적인 자료분석 방법은 모발치료의 자료 예를 통하여 모형설정에 대한 방법을 기술하고 가중최소제곱추정 방법으로 모수들을 추정하는 방법을 설명하고 있다.

참고문헌

Agresti, A. (1990). *Categorical Data Analysis*, John Wiley & Sons, New York.

Im, S. and Gianola, D. (1988). Mixed models for binomial data with an application to lamb mortality. *Applied Statistics*, **37**, 196–204.

Koch, G. G. and Reinfurt, D. W. (1971). The analysis of categorical data from mixed models. *Biometrics*, **27**, 157–173.

Koch, G. G., Landis, J. R., Freeman, J. L., Freeman, D. H. and Lehnen, R. G. (1977). A general methodology for the analysis of experiments with repeated measurement of categorical data, *Biometrics*, **33**, 133–158.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika*, **73**, 13–22.

Milliken, A. G. and Johnson, D. E. (1984). *Analysis of Messy Data*, Van Nostrand Reinhold, New York.

[2008년 1월 접수, 2008년 2월 채택]

A Logit Model for Repeated Binary Response Data

Jaesung Choi[1)]

ABSTRACT

This paper discusses model building for repeated binary response data with different time-dependent covariates each occasion. Since repeated measurements data are having correlated struture, weighed least squares(WLS) methodology is applied. Repeated measures designs are usually having different sizes of experimental units like split-plot designs. However repeated measures designs differ from split-plot designs in that the levels of one or more factors cannot be randomly assigned to one or more of the sizes of experimental units in the experiment. In this case, the levels of time cannot be assigned at random to the time intervals. Because of this nonrandom assignment, the errors corresponding to the respective experimental units may have a covariance matrix. So, the estimates of effects included in a suggested logit model are obtained by using covariance structures.

Keywords: Marginal logit, logit model, repeated measures, covariance matix.

1) Professor, Dept. of Statistics, Keimyung University, 1000 Sindang-Dong, Dalseo Gu, Daegu 704-701, Korea.
E-mail: jschoi@kmu.ac.kr