

역학 III(분석방법)				번호: III - H - 5	
제 목	국문	군집 이분형 자료분석을 위한 조건부로지스틱모형 추정량과 일반화선형모형 추정량 비교			
	영문	HGLMs versus conditional logistic models for the analysis of clustered binary data			
저 자 및 소 속	국문	강위창 ¹⁾ , 이무송 ²⁾ , 윤성철 ³⁾ , 노맹석 ³⁾ , 하일도 ⁴⁾ 이영조 ³⁾ 1) 대전대학교 정보통계학과, 2) 울산의대 예방의학교실, 3) 서울대학교 통계학과, 4) 경산대학교 통계학과			
	영문	Weechang Kang ¹⁾ , Moo-Song Lee ²⁾ , Sungchul Yoon ³⁾ , Maengseok Noh ³⁾ , Youngjo Lee ³⁾ 1) Department of Information and Statistics, Daejeon University, 2) Department of Preventive Medicine, University of Ulsan, 3) Department of Statistics, Seoul National University, 4) Department of Statistics, Kyungsan University			
분 야	역 학 기타	발 표 자	강위창	발표형식	구 연
			일반회원		
진행상황	연구완료				
<p>1. 연구목적</p> <p>군집 이분형 자료는 쌍둥이 연구, 코호트 연구, 교차임상시험 등에서 자주 생성된다. 이러한 자료에 대한 분석법으로는 변량효과모형법과 조건부로지스틱모형법이 있다. 특히, 조건부로지스틱모형은 일반적으로 효율적인 모수 추정량을 제공하는 것으로 믿어져 군집 이분형 자료분석에 그 사용이 권장되었다. 하지만 이 분석법은 결과변수 쌍들 중에서 서로 일치하지 않는 쌍만의 자료를 이용하기 때문에 분석시 자료의 손실을 피할 수 없다. 본 연구에서는 환자와 같은 각 개체의 효과를 변량인자로 모형화할 수 있는 다단계일반화선형모형을 이용하여 군집 이분형 자료를 분석할 수 있음을 보이고 수치적 연구를 통하여 다단계일반화선형모형 분석법이 조건부 로지스틱회귀모형 분석법보다 더 적절한 분석법임을 보이고자 한다.</p> <p>2. 연구방법</p> <p>랜덤하게 선택된 79개의 각 병원에서 피부질환 환자를 대상으로 수행된 치료약과 대조약에 대한 피부질환 호전여부 자료(Stokes et al, 1995)와, 대조약과 두 치료약의 선호도 평가를 위해 환자를 두 나이 그룹으로 층화한 후 각 나이 그룹에서 150명 씩을 대상으로 수행한 two-period 교차임상시험 자료(Koch et al, 1977)를 사용한다. 또한 이들 자료에 근거하여 생성한 가상자료를 두 방법으로 분석하여 결과를 비교한다.</p> <p>3. 연구결과</p> <p>피부질환 호전여부 자료에 대한 분석결과를 간략히 정리하면 다음과 같다.</p>					

공변량	다단계일반화선형모형			조건부로지스틱모형		
	추정값	표준오차	t-값	추정값	표준오차	t-값
처리약	0.939	0.382	0.013	0.703	0.360	0.051
초기질환정도						
병원 내	1.740	0.366	<0.001	1.092	0.335	<0.001
병원 간	0.418	0.216	0.054			
나이						
병원 내	0.018	0.023	0.441	0.025	0.022	0.256
병원 간	0.014	0.019	0.459			
성별						
병원 내	0.183	0.533	0.734	0.531	0.555	0.337
병원 간	0.052	0.391	0.896			

조건부로지스틱모형 결과에서 처리약의 효과는 유의수준 5%에서 경계수준의 유의성(p-value=0.051)을 갖지만 다단계일반화선형모형에서는 유의하다(p-value=0.013). 또한 초기질환정도의 유의성에서도 다단계일반화선형모형의 결과가 더 강하다.

다음표는 상기자료에 근거하여 300회 가상자료를 생성하여 수행한 수치분석결과이다.

공변량	참값	다단계일반화선형모형			조건부로지스틱모형		
		MSE	Bias	S.D.	MSE	Bias	S.D.
처리약	0.9	0.271	0.181	0.489	0.391	0.195	0.596
초기질환정도							
병원 내	1.7	0.376	0.338	0.512	0.602	0.372	0.682
병원 간	0.4	0.079	0.081	0.269			
나이							
병원 내	0.02	0.001	0.005	0.027	0.002	0.008	0.038
병원 간	0.01	0.0005	-0.0002	0.021			
성별							
병원 내	0.2	0.366	0.051	0.604	0.772	0.058	0.878
병원 간	-0.1	0.867	-0.072	0.930			

다단계일반화선형모형의 추정치가 조건부로지스틱모형의 그것보다 평균제곱오차, 편의, 표준편차의 측면에서 보다 우수함을 확인할 수 있다.

4. 고찰

조건부로지스틱회귀모형이 군집 이분형 자료분석에 일반적으로 사용되는 이유는 (1) 신뢰할 만한 추정량이 없었다는 것과 (2) 이 모형이 효율적인 모수 추정량을 제공하는 것으로 믿어져 왔기 때문이다. 하지만 본 연구를 통하여 다단계일반화선형모형으로 보다 효율이 좋은 신뢰할 수 추정량을 구현할 수 있다 사실을 발견하였다. 군집 이분형 자료분석에 다단계일반화선형모형을 적용함으로써 보다 정밀한 분석을 수행할 수 있겠다.