보 건 행 정 학 회 지 : 제 3 권 제 2 호
Korean J. of Health Policy & Administration
Vol.3, No.2, 159-176, November 1993

외래이용빈도 분석의 모형과 기법

한림대학교 사회의학연구소
이영조 · 한달선 · 배상수

Abstract

A Poisson Regression Aanlysis of Physician Visits

Youngjo Lee, Dalsun Han, Sangsoo Bae
Institute of Health Services and Management, Hallym University

The utilization of outpatient care services involves two steps of sequential decisions. The first step decision is about whether to initiate the utilization and the second one is about how many more visits to make after the initiation. Presumably, the initiation decision is largely made by the patient and his or her family, while the number of additional visits is decided under a strong influence of the physician. Implication is that the analysis of the outpatient care utilization requires to specify each of the two decisions underlying the utilization as a distinct stochastic process.

This paper is concerned with the number of physician visits, which is, by definition, a discrete variable that can take only non-negative integer values. Since the initial visit is considered in the analysis of whether or not having made any physician visit, the focus on the number of visits made in addition to the initial one must be enough. The number of additional visits, being a kind of count data, could be assumed to exhibit a Poisson distribution. However, it is likely that the distribution is over-dispersed since the number of physician visits tends to cluster around a few values but still vary widely. A recently reported study of outpatient care utilization employed an analysis based upon the assumption of a negative binomial distribution which is a type of over-dispersed Poisson distribution. But there is an indication that the use of Poisson distribution making adjustments for over-dispersion results in less loss of efficiency in parameter estimation

* 이 논문은 한림대학교 교비연구비 및 독일 대외협력기금(GTZ)의 재정지원으로 실시되었음.

compared to the use of a certain type of distribution like a negative binomial distribution.
An analysis of the data for outpatient care utilization was performed focusing on an assessment of appropriateness of available techniques. The data used in the analysis were collected by a community survey in Hwachon Gun, Kangwon Do in 1990. It was observed that a Poisson regression with adjustments for over-dispersion is superior to either an ordinary regression or a Poisson regression without adjustments for over-dispersion. In conclusion, it seems the most approprite to assume that the number of physician visits made in addition to the initial visist exhibits an over-dispersed Poisson distribution when outpatient care utilization is studied based upon a model which embodies the two-part character of the decision process underlying the utilization.

Ⅰ. 서 론

의료체계는 사회의 구성원들이 의료를 이용함으로써 건강을 누리게 하는 것을 기능으로 한다. 그러므로 의료이용의 질적, 양적 적정성과 연관요인의 이해는 의료체계의 성과를 파악하고, 발전방안을 강구하는데 필수적이다. 의료이용양상의 분석은 이와 같은 이해를 높이기 위한 노력의 하나인 것이다. 의료이용에 대하여는 여러가지 측면에서 분석되어야 하지만, 기본적인 과제의 하나는 의료이용량을 결정해 주는 요인들과 기전을 밝히는 것이다. 의료이용의 양은 의료서비스의 가격과 더불어 의료정책에서 핵심적 관심사의 하나일 수밖에 없는 국민의료비를 결정하는 주된 요소라는 점이 그러한 분석의 중요성을 말해주는 단적인 예가 되겠다.

모든 분석에서 그러하듯이 의료이용도에 대한 분석에 있어서도 자료의 질, 이론적 모형등과 아울러서 실증적 분석모형의 설정과 모형의 추정기법이 결과의 타당성을 크게 좌우함은 물론이다. 이 논문은 외래이용 결정과정의 특성에 대한 이론적 이해를 토대로 이용빈도의 실증적 분석에 적절한 기법을 제시하려는 시도이다. Duan 등(1983)은 의료수요의 결정은 의료이용여부, 이용한다면 입원할 것인지 외래진료를 받을 것인지, 그리고 양적으로는 얼마나 이용할 것인지 등에 관한 일련의 결정과정을 포함하는 데, 각 결정이 서로 이질적이므로 별도로 분석함이 타당하다고 지적하였다. 이와 같은 관점에서 생각할 때 외래이용빈도는 2단계의 결정과정을 거쳐서 정해진다고 볼 수 있다. 첫째는 이용여부 또는 최초방문의 결정으로서, 이 결정에는 환자 자신이나 가족이 주로 작용할 것이다. 두번째 단계인 추후방문에 관한 결정에는 의사의 권고나 최초방문기관에서의 체험이 크게 영향을 미칠 것이다. Pholmeier와 Ulrich (1992)는 의료이용과정의 이와 같은 특성등을 고려하여 의사방문수에 대한 분석에서 Negative Binomial 분

포에 입각한 모형을 사용한 바 있다. 우리나라에서는 아직도 이러한 점을 명시적으로 다룬 분석이 없었고, 외래의료이용을 이용경험의 유무와 이용경험자의 방문횟수의 두 단계로 구분해서 분석한 연구들도 그 이유와 기타 관련된 방법론적 사항들을 분명히 다루지는 않고 있다(한달선외,1986;송건용외,1983;권순호외,1987). 이러한 문제의 인식에서 출발하여 먼저 분석의 대상인 외래의료 이용과정의 본질과 확률기전을 간략히 기술하고, 이어서 분석기법의 선택에서 고려해야 할 외래이용빈도 자료의 속성을 살핀 후에, Negative Binomial 분포 가정에 의한 모형보다 더 적합하다고 생각되는 과분산 포아송분석을 실제로 조사된 자료에 적용하여 그 타당성을 검토할 것이다.

Ⅱ. 외래이용과정의 2단계 모형과 확률기전

외래의료이용은 환자가 의료서비스를 받기로 결정하고, 실제로 어떤 의료기관을 방문함으로써 시작된다. 일단 의료이용이 시작된 후에 1회 방문으로 끝낼 것인지, 추가적으로 다시 방문할 것인지, 의료를 더 이용하되 방문할 의료기관을 변경할 것인지, 입원의료나 약국이용 등으로 전환할 것인지를 환자가 결정함에 있어서는 의사의 권고를 포함하여 최초 방문기관에서의 체험이 강력하게 영향을 미친다. 다시 말해서 외래의료이용의 최초방문은 주로 환자의 결정에 의해서 이루어지지만 추가적 방문횟수는 환자의 의사뿐만 아니라 다른 요인들의 영향을 받아서 결정됨을 뜻한다. 이는 외래의료이용의 과정을 이용여부 또는 최초이용에 관한 결정과 추가이용횟수에 대한 결정의 2단계로 볼 수 있음을 뜻하며, 각 단계의 결정에 작용하는 요인들이 다를 것임을 시사한다.

외래이용빈도의 2단계 결정과정의 확률기전은 다음과 같이 기술할 수 있겠다. 어느 지역주민들의 특정 기간동안의 외래이용도에 대한 자료에서 Yi를 i번째 주민의 외래이용빈도라 하면, 이는 그 주민의 외래이용여부 Ui와 외래이용을 시작한 이후의 추가방문횟수 Vi의 두가지 요소로 이루어지며, Ui와 Vi의 정의는 다음과 같다.

$$Ui = \begin{cases} 0 & \text{만약 } Yi = 0 \\ 1 & \text{만약 } Yi \geq 1, \end{cases}$$

$$Vi = \quad Yi - 1 \text{ 만약 } Yi \geq 1.$$

이제 $Po = P(Ui = 0)$를 주민이 그 기간동안에 외래의료를 이용하지 않을 확률이라고 하면,

외래이용빈도 Yi의 확률은 다음과 같이 정의된다.

$$P(Yi=yi) = \begin{cases} Po & \text{만약 } Ui=0 \\ (1-Po)\ P(Vi=yi-1 \mid Ui=1) & \text{만약 } Ui=1 \end{cases}$$

주민의 총수를 n, 외래이용 경험이 없는 주민의 총수를 d라고 하면, 외래이용의 경험이 있는 주민의 총수는 n－d가 된다. 여기서 외래이용자의 집합을 W라고 하면,

$$\prod P(Yi=yi)=Po^{d}(1-Po)^{n-d}\prod_{i\in W}P(Vi=yi-1 \mid Ui=1)=\prod P(Ui)\prod_{i\in W}P(Vi=yi-1 \mid Ui=1)$$

가 성립한다. 그러므로 외래이용빈도 Yi에 대한 분석은 외래이용여부 Ui에 대한 분석과 외래이용환자의 추가방문횟수 Vi에 대한 분석의 2단계로 나누어서 다루어질 수 있다. Ui는 상기 확률기전에서 보다시피 이항분포를 따르므로 logit 분석을 이용해 분석할 수 있다. 본 논문에서는 외래이용 결정기전의 두번째 단계인 외래이용빈도에 대한 분석에 촛점을 두고자 한다. 그런데 Vi의 분포는 명백하질 못하므로 그 분포를 파악하는 일이 선결과제이다.

Ⅲ. 외래이용빈도 자료의 속성과 분석기법

추가이용횟수를 포함한 외래이용빈도는 정의상 그 값이 항상 양의 정수인 갯수변수이므로 일단 포아송(Poisson)분포를 따른다고 가정해 볼 수 있다. 포아송 분포란 확률변수값이 0,1,2, ……의 갯수일 때 자주 사용되는 분포로서 다음의 평균과 분산관계를 만족시킨다.

$$E(Y) = \mu,\ Var(Y) = \phi\mu \quad 단\ \phi = 1 \qquad (1)$$

포아송 분포와 같이 갯수변수의 확률분포로서 $\phi>1$인 분포들의 군을 총칭하여 과분산 포아송분포라고 한다. 사람당 기생충 수, 어느특정 면적의 벌레수, 절제된 악성종양의 특정부위의 종양세포수 등의 갯수변수들을 보면 갯수들이 적은 경우 아주 적고, 큰 경우에는 아주 큰 경향이 있는 데, 이러한 경우 포아송분포로 적합되지 않고, 과분산 포아송분포로 적합하여야 하는 것이다.[1] 한편 확률변수의 분포가 포아송분포라 하더라도 분석시 필요한 설명변수가 누락된 경우에는 과분산이 일어난다(Aitkin외, 1989, 4장 참조).

1) 이 성질을 만족하는 분포로는 특수한 Negative Binomial 분포가 있고, 이에 대한 연구로는 사회과학분야에서 Hausman외(1984), King(1988), 통계학분야에서는 Nelder와 Lee(1992), Lee와 Nelder(in press)가 있다. 과분산 포아송분포가 생성되는 확률과정에 대하여는 McCullaugh와 Nelder(1989)에 기술되어 있다.

요약하면 과분산은 대개의 경우 (i) 종속변수의 분포자체가 과분산 포아송인 경우 (ii) 필요한 독립변수가 누락되어 있는 경우 발생하게 된다. 어떠한 이유에서든지 과분산이 발생하면 과분산을 보정하여 분석을 하여야 한다. 과분산이 발생하면 대개의 경우 상기 두 요인중 어떤 요인에 의해 과분산이 발생했는지 알기 어렵다. 그런데 중요한 설명변수의 누락으로 과분산이 발생했을 경우에는 회귀계수값의 추정에도 오류가 있을 가능성이 크므로 특히 주의해야 한다. 그러므로 실험자료가 아닌 관측자료를 주로 사용하는 사회과학 분야에서는 필요한 모든 설명변수를 분석에 포함시켰다는 보장이 없으므로 과분산 발생시의 분석기법에 대한 이해가 매우 중요하다.

강원도 화천군에서 1990년에 수집한 자료에서 외래이용도의 실제 분포를 살펴보면 다음과 같다. 15일 간에 상병이 있었던 주민은 2,213명이었는데 이들의 외래이용횟수(Yi)별 분포와 이들중 외래이용의 경험이 있었던 712명의 추가방문횟수(Vi)별 분포는 표1에서 보는 바와 같다. 또, $Y_i=0$, $V_i=0$항의 빈도수가 화천자료와 같게 구한 Poisson분포도 같은 표에 제시하였다. 이들 두 분포를 비교해 보면 실제 화천자료의 Yi, Vi값의 분포들이 포아송 분포보다 꼬리가 길어서 과분산이 있음을 알 수 있다.[2] 물론 설명변수들을 추가하면 과분산을 어느 정도 줄일 수는 있으나 완전히 없앨 수는 없다. 따라서 추가이용횟수를 포함한 외래이용빈도는 과분산포아송분포에 입각하여 분석함이 타당하다는 시사를 얻게 된다. Pholmeier와 Ulrich(1992)는 과분산 포아송 분포군에 속하는 Negative Binomial분포를 사용하여 외래이용빈도에 대한 자료를 분석하였으나, Lee와 Nelder(in press)는 평균 모수측정시 특정한 분포, 예를 들어 Negative Binomial분포 등을 사용하여 적합하는 것 보다는 과분산 포아송분포를 사용, 모수값을 추정하는 것이 분포에 대한 가정이 틀릴 경우에도 효율성의 상실이 적다는 것을 보였다. 분포에 대한 가정은 검증하기가 어려우므로 Lee와 Nelder의 과분산 포아송 분석을 사용하는 것이 바람직하다 하겠다.

요약하면 외래이용도에 대한 연구는 거의 모든 경우에 관측자료를 사용하게 되고, 이용빈도가 갯수변수이며, 앞에서 논의한 바와 같이 포아송분포에 비하여 꼬리가 긴 과분산 포아송분포를 보이고 있다는 점을 고려할 때, 분석에 과분산포아송분포 가정에 의한 기법을 사용하는 것이 타당하리라 추론할 수 있다.

---

2) 산술평균이 같도록 한 Poisson분포와 비교하여 과분산의 경향을 관찰할 수도 있음은 물론이다.

<표 1> 상병자의 외래이용횟수(Yi)별 분포

Yi값	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
화천자료	1511	281	174	88	51	38	21	26	4	3	7	2	3	2	4	8
Poisson	1511	583.4	112.6	14.5	1.4	0.1	0	0	0	0	0	0	0	0	0	0

외래이용경험자의 추가방문횟수(Vi = Yi − 1)별분포

Yi값	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
화천자료	281	174	88	51	38	21	26	4	3	7	2	3	2	4	8
Poisson	281	261.3	121.4	37.6	8.7	1.6	0.3	0	0	0	0	0	0	0	0

Ⅳ. 외래 추가방문횟수에 대한 과분산포아송 분석

1. 자료와 변수

외래이용빈도는 이용여부와 이용경험자의 추가방문횟수로 나누어서 분석하는 것이 타당함은 이미 설명하였다. 앞에서 살핀 바 있는 화천주민에 대한 1990년의 외래이용도 조사자료 중 15일간 급성 또는 만성질환으로 병의원 또는 보건소의 외래를 방문한 경험이 있는 화천주민 712명을 대상으로 추후방문횟수 Vi를 조사하였다. 외래이용을 시작한 이후에 추가로 병원, 의원, 보건소 또는 보건지소를 방문한 횟수(Vi= 외래이용횟수−1)를 종속변수로 하였고, 이를 설명하기 위하여 분석에 포함된 독립변수와 그 측정방법은 표 2와 같다.

외래 추가방문횟수는 의료이용의 개시여부와 달리 환자의 특성뿐만 아니라 의료제공자의 특성에 의해서도 크게 영향을 받을 것이다. 따라서 외래방문횟수의 분석에는 의료제공자의 특성에 관한 다양한 변수들을 포함하여야만 한다. 그러나 본 연구에서는 자료의 제약상 이용의료기관의 형태와 이용의료기관이 단골 이용처인가를 제외하고는 나머지 의료제공자 변수는 포함시키지 못하였다. 이용의료기관의 형태는 보건기관이냐, 민간의료기관이냐로 구분하였는 데, 두개의 가변수를 사용하여 판별되었다. 표 2에서 병의원을 이용하지 않고(H=1), 병의원과 보건지소를 동시에 이용하지 않은(D=1)주민은 조사기간 동안 보건소만 이용한 경우라 할 수 있다.

우리나라에서는 개원의들은 행위별 수가제의 적용을 받고, 병원이나 공공부문의 의사들은 봉급을 받는다. 의료서비스의 제공양을 늘리고자 하는 경제적 유인은 봉급을 받는 의사가 행위별 수가제의 적용을 받는 의사보다 적다. 특히 보건기관에 근무하는 의사들은 대부분 병역의무를 대체하기 위하여 한시적으로 배치된 공중보건의이므로 병원에 근무하는 의사보다도 진료의 양을 늘릴 동기가 더 약하다. 따라서 동일의료기관에 대한 방문횟수는 민간의료기관이 보건기관보다 많을 것이라 예측할 수 있다. 한편 보건기관의 경우 필요한 경우 환자를 의뢰할 수 있는 동료의사도 드물며, 의료장비나 시설 등도 민간 의료기관에 비해 매우 열악하다. 따라서 보건기관 의사들은 환자의 증상이 심하면 민간병원을 방문하도록 권유할 가능성이 높다. 이는 보건기관을 최초 방문기관으로 선택한 심한 증상의 환자들은 처음부터 병의원을 이용했을 경우보다도 방문횟수가 더 많을 가능성이 있음을 뜻한다. 따라서 방문의료기관의 형태와 이용빈도의 관계는 환자의 상태와 최초방문 의료기관의 상호작용에 의해 결정될 것이다.

환자들은 단골의사를 방문할 때 안정감을 느낀다. 또한 진료결과에 대한 만족도도 높아 진료의 지속성이 유지될 것이다. 그러나 본 연구의 조사기간이 15일이라는 점을 감안하면, 진료의 지속성이 의료이용에 미치는 영향은 미미할 것이다. 오히려 환자가 동일 질병으로 의료기관을 전전하는 우리나라의 현실을 감안하면 단골의사의 존재는 외래 추가방문횟수를 감소시키는 방향으로 작용하리라 보여진다. 또한 단골의사의 경우 환자를 잘 알므로 환자의 상태를 관찰하기 위한 재진을 덜 필요로 할 수도 있다.

성별로는 생리학적, 사회심리학적 요인때문에 남자보다 여자의 의료이용경험율이 높고, 의료이용경험자당 방문횟수도 많은 것으로 알려져 있다. 다시 말해 여성들의 외래추가방문횟수는 남성보다 많을 것으로 예견된다. 연령별 의료이용경험은 청장년층보다 소아와 노인에게서 빈번한 것으로 알려져 있다. 이는 추가방문횟수도 마찬가지일 것이다. 왜냐하면 소아와 노인과 같은 취약계층에 대해서는 의사들의 재진빈도가 잦으며, 노인들은 만성질환이 많기 때문이다.

가계수입과 환자의 교육수준은 서로 상관관계가 매우 높은 변수로서 두 변수 모두 추가방문횟수에 정의 영향을 미친다. 의사들은 재진의 결정시 환자의 소득을 참작할 것이고, 소득과 교육수준이 높을 수록 경제적 부담에 관계없이 충분한 진료를 받으려는 경향이 있을 것으로 생각되기 때문이다. 그러나 이 두 변수가 외래 추가방문에 미치는 영향의 크기는 의료이용의 결정에 미치는 영향에 비해 작을 것이다. 물론 이 두변수가 미치는 영향의 크기는 환자의 건강상태를 보정하여야 정확히 알 수 있다. 왜냐하면 저소득층은 고소득층에 비해 건강수준이 나쁘

기 때문에 유사한 상병에 이환되더라도 환자에게 나타나는 증상은 훨씬 위중할 수 있기 때문이다.

가구주 소득원은 소득의 크기, 의료보장의 종류, 그리고 의료이용에 대한 인식과 시간비용을 반영하는 변수이다. 이중 소득의 크기나 의료보장의 종류는 따로 독립변수로 선정되어 있으므로 가구주 소득원이 외래 이용빈도에 미치는 영향을 의료이용 인식과 결부하여 살펴보면 다음과 같다. 농업종사자들은 상업이나, 봉급생활자보다 전통적인 삶의 방식을 유지하는 경우가 많다. 또한 봉급생활자나 상업종사자들보다 벽지에 사는 경우가 많으므로 의료이용에 따른 교통비나 시간비용이 더 든다. 이에 따라 농업종사자들은 다른 사람들보다 의료이용을 꺼려 할 가능성이 높다. 반면 이들이 의료이용을 하는 경우는 위중한 질병일 경우가 많아 추가방문횟수가 많을 수도 있다. 즉 가구주의 직업이 외래 추가방문횟수에 미칠 방향을 선험적으로 예측하기는 어렵다.

가구원수는 두가지 기전으로 의료이용에 영향을 미친다. 가구원 수가 많으면 소득이 있는 가구원도 많아 가구의 총소득을 증가시킨다. 반면 일단 가구소득이 결정되고나면 가구원수는 일인당 가처분소득을 감소시키는 방향으로 작용한다. 즉 가구원수는 소득을 통해 간접적으로 의료이용에 영향을 미친다. 또 간병해 줄 사람이 가내에 있는지 여부에 의해서도 의료이용에 영향을 미친다. 물론 단독가구나 노인 단독세대의 경우 병의원을 이용하기 위하여 집을 비운다는 것이 쉽지 않아 외래 추가이용이 줄어들 가능성도 있다.

거주지는 교통편의도나 의료기관까지의 거리를 반영하는 변수로서, 읍지역 거주자가 다른 지역에 비해 외래 추가방문횟수가 많을 것으로 짐작된다.

의료보장의 종류에 따라 급여범위와 본인 부담이 다르면 당연히 재진빈도도 다를 것이다. 그러나 우리나라에서는 지역, 공교, 직장의료보험간에 급여의 차이를 볼 수 없다. 하지만 의료보호의 경우에는 보험환자에 비해 급여수준이나 수가에 차이가 있을 뿐만 아니라 적용자의 연령, 소득수준, 건강상태 등 많은 측면에서 다르다. 그러므로 의료보장변수는 의료보호적용자인가 아닌가로 분류하였다.

평소의 건강상태가 나쁜 경우 유사한 질환에 이환되더라도 환자가 느끼는 증상은 훨씬 심각할 수 있고, 의사들도 이들 환자에 대해서는 보다 주의깊게 진료를 하려 할 것이므로 의료이용이 많은 경향을 보일 것이다.

현재 본인이 앓고 있는 질병의 종류와 증세가 심한 정도 역시 의료이용에 영향을 미친다. 조사대상기간이 15일인 점을 감안하면 만성질환자들 보다 급성질환자의 환자당 방문횟수가 더

많을 것이다. 동일 질환에 이환되더라도 유병기간이 긴 경우에 의료이용이 잦으리라는 것은 당연하다. 그러나 질병이 매우 심한 경우에는 오히려 의료기관을 다시 방문하거나, 전전하기가 어려울 수도 있다. 이에 활동제한일수를 질병의 위중도를 파악하는 또 하나의 변수로 선정하였다.

<표 2> 외래추가방문횟수를 설명하기 위하여 사용된 독립변수

인구학적 변수

SEX : 성별(남자 1, 여자 2)

AGE : 연령(10세 간격으로 가변수 처리, 80세이상 한 군으로 처리)

사회경제적 변수

I : 가계수입

EDU : 가구주 교육수준(국졸미만 1, 중졸미만 2, 고졸미만 3, 고졸이상 4)

JOB : 가구주 소득원(농축산업 1, 상업 2, 봉급 3, 기타 4)

LNOF : log 가구원수

RES : 거주지(읍내 및 보건소 지역 1, 기타 2)

MS : 의료보장 유형(지역,공교,직장 1, 보호 2)

의료필요변수

S : 평소건강상태(건강한 편 1, 허약한 편 2)

PQ : 급만성 유병상태(만성질환 1, 급성질환 2, 급만성질환 동시이환 3)

P1 : 보름간 만성유병일수

P2 : 보름간 만성활동제한일수

Q1 : 보름간 급성유병일수

Q2 : 보름간 급성활동제한일수

의료제공자의 특성

FD : 단골의사 유무(없음 1, 있음 2)

H : 병의원이용 여부(없음 1, 있음 2)

D : 병의원, 보건소 동시 이용여부(없음 1, 있음 2)

2. 분석의 결과와 타당성

GLIM 패키지를 사용하여 포아송분포를 적합하여 회귀계수의 추정치와 표준편차를 구하였다. Scaled Deviance는 1523.8이고, 자유도는 654이다. 과분산이 일어나지 않았다면 Scaled Deviance는 자유도 654인 Chi-square 분포를 따르므로 값이 654에 가까워야 한다. 그런데 앞서 설명한 식 (1)의 ϕ의 추정치는 1523.8/654 = 2.330으로서 과분산이 발생하였음을 알 수 있으므로, 과분산을 보정한 표준편차를 얻었다.[3] <표 3>은 회귀분석에 포함된 독립변수 중에서 인구학적 변수인 연령과 성을 제외한 다른 변수들의 회귀계수와 표준편차, 그리고 과분산을 보정한 표준편차를 보여주고 있다. 과분산을 보정하지 않은 상태에서는 연구모형에 포함된 독립변수 중에서 가계수입과 가구주 소득원을 제외한 나머지 변수들의 회귀계수가 모두 t-검정에서 유의하였다. 이에 <표 3>에는 가계수입과 가구원수의 회귀계수와 표준편차는 제시하지 않았으며, 제시된 독립변수들의 회귀계수에 대한 유의수준은 과분산을 보정한 상태만 표시하였다. FD×RES변수는 t-검정에서는 유의하지 않았으나 로그우도차에 의한 검정에서는 유의수준 0.05에서 유의하여 포함시켰다. t-통계량에 의한 검정결과와 로그우도차에 의한 검정결과가 다른 경우에는 로그우도차에 의한 검정결과를 선호하기 때문이다.

분석결과를 상세히 살피기 전에 우선 그 타당성을 가늠하기 위하여 잔차의 분포를 검토하기로 한다. 과포아송 분석시 잔차로는 Deviance Residual이 추천된다. Deviance Residual의 분포는 일반적으로 정규분포에 가깝다(McCullaugh & Nelder, 1989). (그림 1)의 Normal Plot을 보면 직선에 가까워 분석상 특별한 하자가 있다고 생각되지는 않는다.[4] McCullaugh과 Nelder (1989)는 잔차 그림을 예측치의 제곱근에 대하여 그리라고 추천한다. 이유는 이 경우에 아래 잔차 그림에서 보다시피 아래 쪽에 감소하는 직선들이 생긴다. 맨 아래 직선은 Vi = 0인 관측치들에 대한 잔차이고, 두번째 직선은 Vi = 1인 관측치들에 대한 잔차이다. 포아송 분포를 따르는 변수의 값이 0,1,2,……로 하방한계는 있으나 상방한계는 없다. 그러므로 상반부 잔차그림이 타당성 검사시 의미가 있는데, 가운데 부분의 잔차 변이는 이 부분에 자료가 많기 때문이라고 생각되므로 잔차그림에서 과분산 포아송 적합이 타당하지 않은 점을 찾을 수 없고, 잔차히스토그램도 과포아송군이 평균이 작을 때 한쪽으로 치우쳐 있으므로 문제가 없는 것 같다.

---

3) GLIM 패키지에서 $ca %s = 1523.8/654 $sca %s $f.$ 명령문을 수행하면 과분산을 보정한 표준편차를 구할 수 있다.

4) 과분산 포아송분석시 잔차는 포아송 분석시 잔차에서 ϕ 추정치의 제곱근으로 나누어 주어 계산하기 때문에 타당성 검토를 포아송 분석잔차를 이용해도 결론이 동일하다.

일반회귀분석을 사용한 외래이용도에 대한 분석이 적지 않은 데, 이 분석방법은 다음의 평균분산관계를 갖는 정규분포를 가정한다.

$$E(Vi) = \mu = X\beta\ ,\ Var(Vi) = \Phi\ ,$$

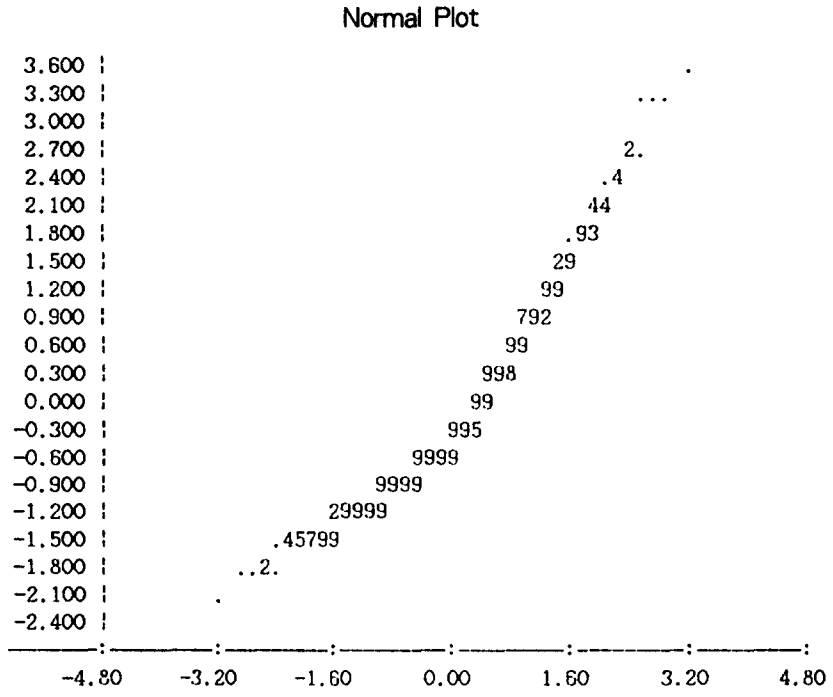
<표 3> 연구에 사용된 변수들의 회귀계수, 표준편차 및 과분산보정 표준편차

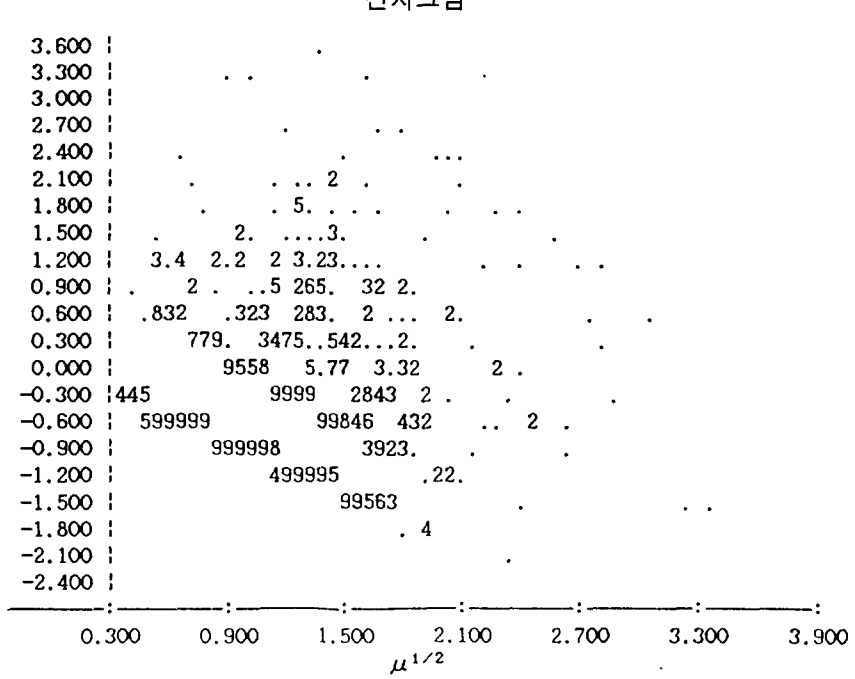
변 수	회귀계수[1)]	표준편차	과분산보정 표준편차
1	-2.560**	0.5448	0.8316
D(2)	0.5579**	0.08264	0.1261
S(2)	-0.1898	0.07177	0.1096
PQ(2)	0.05410	0.3632	0.5544
PQ(3)	-1.524**	0.2283	0.3485
P1	0.2079*	0.06695	0.1022
P2	-0.01059*	0.003435	0.005244
Q1	0.1884**	0.03920	0.05983
Q2	-0.006949*	0.002156	0.003291
FD(2)	1.176	0.6399	0.9768
LNOF	0.3138	0.1253	0.1912
RES(2)	-0.05053	0.09651	0.1473
H(2)	1.815**	0.4316	0.6587
EDU(2)	0.9014	0.3944	0.6021
EDU(3)	1.296	0.4783	0.7300
EDU(4)	0.5599	0.4932	0.7529
MS(2)	1.148**	0.2621	0.4000
MS(3)	0.6127	0.3267	0.4988
FD(2).LNOF	0.4614**	0.1160	0.1770
FD(2).RES(2)	0.1965	0.1256	0.1917
FD(2).PQ(2)	-1.032*	0.3355	0.5121
FD(2).PQ(3)	1.167**	0.2153	0.3287
FD(2).H(2)	-0.6025	0.5863	0.8949
FD(2).EDU(2)	-1.178	0.5127	0.7827
FD(2).EDU(3)	-1.669	0.6509	0.9935
FD(2).EDU(4)	-1.233	0.6698	1.022
H(2).EDU(2)	-1.235	0.4149	0.6333
H(2).EDU(3)	-1.304	0.5054	0.7714
H(2).EDU(4)	-0.8201	0.5184	0.7913
FD(2).P1	-0.001545	0.03327	0.05078
H(2).P1	0.04002	0.01796	0.02742
FD(2).MS(2)	-0.9364	0.4520	0.6899
FD(2).MS(3)	-0.1848	0.4534	0.6921
H(2).MS(2)	-1.064*	0.2910	0.4442
H(2).MS(3)	-0.6336	0.3492	0.5330
FD(2).H(2).EDU(2)	1.530	0.5407	0.8253
FD(2).H(2).EDU(3)	1.148	0.6855	1.046
FD(2).H(2).EDU(4)	1.415	0.7021	1.072
FD(2).H(2).P1	-0.09825*	0.02760	0.04213
FD(2).H(2).MS(2)	1.055	0.4754	0.7257
FD(2).H(2).MS(3)	0.7062	0.4795	0.7319

1) 과분산을 보정하지 않은 상태에서는 모든 변수의 회귀계수가 통계적으로 유의하므로 과분산을 보정한 상태에서의 유의수준만 제시되었음.

* $P<0.05$, ** $P<0.01$

(그림 1) 과포아송분석의 잔차





잔차 히스토그램

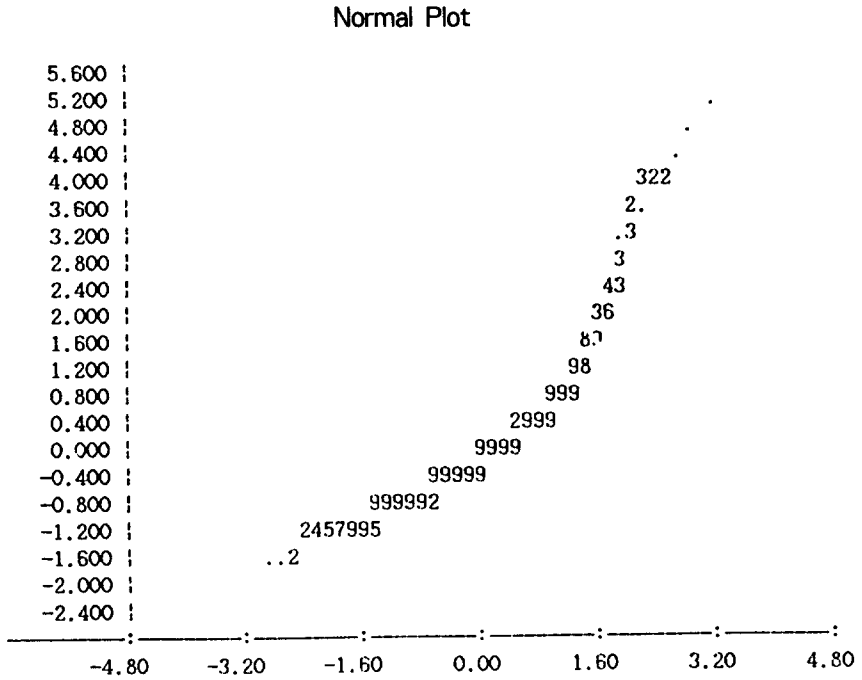
```
[-2.40,-2.10) 1
[-2.10,-1.80) 1
[-1.80,-1.50) 18  DDDDDD
[-1.50,-1.20) 58  DDDDDDDDDDDDDDDDDDD
[-1.20,-0.90) 90  DDDDDDDDDDDDDDDDDDDDDDDDDDDDDD
[-0.90,-0.60) 122 DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD
[-0.60,-0.30) 102 DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD
[-0.30, 0.00) 68  DDDDDDDDDDDDDDDDDDDDDDD
[ 0.00, 0.30) 62  DDDDDDDDDDDDDDDDDDDDD
[ 0.30, 0.60) 57  DDDDDDDDDDDDDDDDDDD
[ 0.60, 0.90) 40  DDDDDDDDDDDDD
[ 0.90, 1.20) 28  DDDDDDDDD
[ 1.20, 1.50) 25  DDDDDDDD
[ 1.50, 1.80) 13  DDDD
[ 1.80, 2.10) 14  DDDDD
[ 2.10, 2.40) 1
[ 2.40, 2.70) 7   DD
[ 2.70, 3.00) 1
[ 3.00, 3.30) 2   D
[ 3.30, 3.60) 1
[ 3.60, 3.90) 1
[ 3.90, 4.20) 0
[ 4.20, 4.50] 0
```

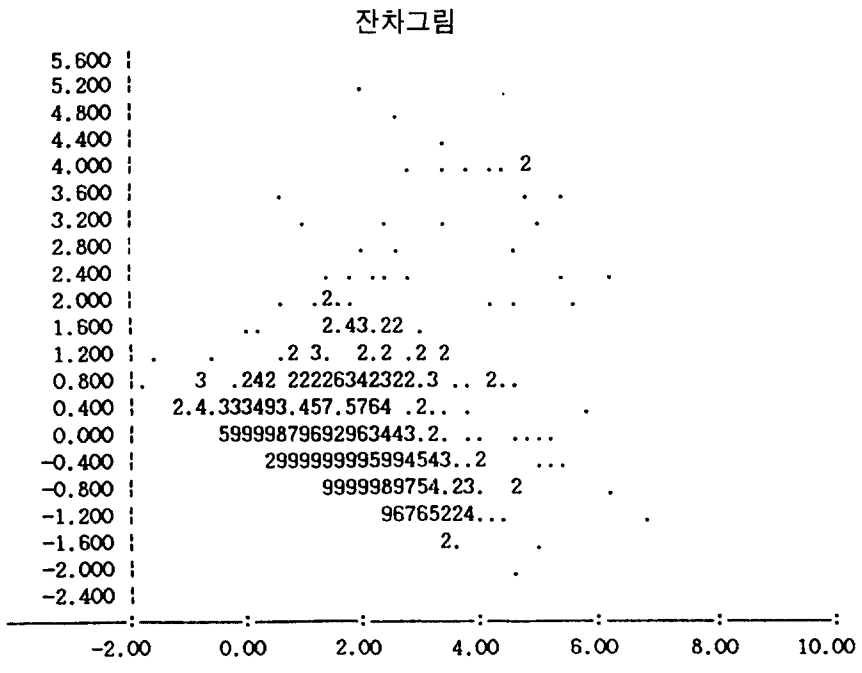
이를 살펴보면 분산과 평균은 서로 관련이 없고, 평균 μ가 선형모형을 만족한다고 가정한다. 전 절의 포아송분석에서 사용한 모형으로 일반회귀분석을 하고 그 잔차에 대해 조사하였다. (그림 2)의 Normal Plot에서 그어진 선이 직선이 아니므로 이는 잔차들이 정규분포를 따르지 않음을 의미한다. 예측치 μ에 대한 잔차그림을 보면 잔차의 분산이 μ값이 증가함에 따라 증가하고 있어서 분산과 평균이 서로 상관이 없다는 일반회귀 분석가정이 성립하지 않음을 알 수 있다. 잔차 그림에서 보다시피 일반회귀분석의 또 하나의 단점은 주민에 따라 외래이용빈도의 기대값이 음수가 될 수도 있다는 점이다. 실제로 잔차 히스토그램에서 잔차분포가 음의 값쪽으로 편향되어 있어 잔차분포가 대칭인 정규분포가정을 만족하지 않음을 보인다.

자료분석상 이와 같은 일반회귀분석 가정이 성립되지 않는 갯수변수를 정규분포를 사용한 일반회귀분석을 사용, 적합하면 분산함수를 감안하지 않아 분석이 매우 비효율적이다. 예시하면 포아송분석에서는 유의한 모든 회귀계수가 일반회귀분석으로 분석한 경우에는 모두 유의하지 않게 나왔다. 그러므로 Normal Plot, 잔차그림등을 그려보고 일반회귀분석가정이 성립하지 않을 때에는 회귀분석을 하여서는 안된다.

이러한 검토의 결과를 종합적으로 고려할 때 외래이용빈도에 대하여는 과분산 포아송분포 가정에 입각한 분석기법의 사용이 분명히 더 타당하다.

(그림 2) 일반회귀분석의 잔차





잔차 히스토그램

```
[-2.00,-1.60) 2   D
[-1.60,-1.20) 18  DDDDD
[-1.20,-0.80) 76  DDDDDDDDDDDDDDDDDDD
[-0.80,-0.40) 169 DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD
[-0.40, 0.00) 176 DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD
[ 0.00, 0.40) 114 DDDDDDDDDDDDDDDDDDDDDDDDDDDDDD
[ 0.40, 0.80) 62  DDDDDDDDDDDDDDDD
[ 0.80, 1.20) 34  DDDDDDDDD
[ 1.20, 1.60) 16  DDDD
[ 1.60, 2.00) 17  DDDD
[ 2.00, 2.40) 4   D
[ 2.40, 2.80) 6   DD
[ 2.80, 3.20) 3   D
[ 3.20, 3.60) 4   D
[ 3.60, 4.00) 5   D
[ 4.00, 4.40) 3   D
[ 4.40, 4.80) 2   D
[ 4.80, 5.20) 1
[ 5.20, 5.60) 0
```

3. 외래 추가방문횟수의 연관요인

포아송분포로 적합한 외래 추가방문횟수의 결정모형을 <표 3>에 제시된 변수들을 이용하여 구하면 다음과 같다.

$$\log\mu = AGE\times SEX + D + S + Q1 + Q2 + P2 + FD\times(LNOF + RES + PQ) + H\times FD \times (EDU + P1 + MS)^{*}$$

그러나 앞에서 살펴 보았듯이 포아송분포를 이용하여 외래 추가방문을 적합하는 것은 자료의 성격상 타당하지 않다. 이에 따라 과분산을 보정하면 과분산에 대한 보정이 없었을 때 유의한 것으로 나타났던 변수중 S, RES, FD×RES, FD×EDU, H×EDU, H×FD×EDU, FD×MS, H×FD×MS 등의 효과들이 유의하지 않은 것으로 나타난다. 그러므로 과분산을 보정하지 않고 포아송분석에 의해 상기 효과들이 있다고 하는 것은 분석상 오류이므로 과분산이 발생한 경우 앞에서 밝혔듯이 이에 대한 보정을 항상하여야 한다. 과분산을 보정한 외래 추가방문횟수 결정모형은 아래와 같이 정리될 수 있다.

$$\log\mu = AGE\times SEX + D + P2 + Q1 + Q2 + FD\times(LNOF + PQ) + H\times FD\times P1 + H\times MS^{*}$$

---

* 유의한 변수로 모형에서 제시되지 않았더라도 변수들간에 higher interaction이 있으면 해당 lower interaction들 및 main effects가 있는 것으로 보아야 함(McCullaugh & Nelder,1989).

변수의 선정에서 언급하였듯이 병의원과 보건소를 동시에 이용한 경우(D)에 외래 추가방문횟수가 이들 의료기관중 하나를 이용한 경우보다 많았다. 이는 본 연구의 추가방문횟수가(총이용횟수−1)로 정의되었기 때문에 의료기관을 이동하게 되면 그만큼 추가방문이 많게 나타나기 때문이다. 그러나 활동에 제한이 있을 정도로 질환이 심한 경우에는(P2, Q2) 오히려 추가방문이 적었는 데, 이는 환자의 상태가 심하여 의료기관을 전전하거나, 자주 방문할 수 없었기 때문이거나, 질환의 종류가 사고나 손상 등과 같이 장애는 있으나 계속적인 치료는 요하지 않기 때문으로 보여진다. 물론 아래에서 설명하는 바와 같이 활동제한일수가 유병일수 등 본 연구에서 사용된 다른 변수들과 상당한 공중다공성을 보이므로 위에서 제시된 결과가 통계적 오류일 가능성도 배제할 수 없다.

본 연구와 같이 관측자료를 이용하는 경우는 변수들간에 공중 다공성이 있으므로 개별 변수의 효과를 정확하게 알기는 어렵다. 그러나 변수들간의 교호작용(interaction)을 통하여 독립변수들이 종속변수에 어떻게 영향을 미치는 가를 가늠할 수는 있다.

단골의사의 존재는 가구원 수와 상호작용(FD×LNDF)하면서 추가방문횟수에 영향을 미치는 것으로 나타났는 데, 이는 단골의사라야 가구의 형편을 이해하고, 진료시에도 이를 고려해 주기 때문일 것이다. 가구원 수가 많은 환자가 단골의사를 방문한 경우에 추가방문횟수가 많았는 데, 이는 단독가구에 대해서는 의사들이 투약기간을 늘려 재방문까지의 기간을 가능한 연장하려 하기 때문일 것이다. 환자의 질병상태가 의료이용에 미치는 영향은 의료이용처가 단골의사인가 아닌가에 따라 상이한 영향을 미치는 것으로 나타났다. 이는 급만성질환을 동시에 앓고 있는 환자(PQ)가 단골의사를 방문(FD)하게 되면 의사가 환자의 만성질환에 대해서 알고 있으므로 환자의 평소건강상태를 모르는 의사보다 더욱 세심하게 환자를 진찰하게 되어 추가방문횟수를 증가시키기 때문일 것이다.

의료공급자의 특성을 반영하는 변수의 하나인 이용의료기관의 형태는 가정한 바와 같이 외래이용에 큰 영향을 미쳤다. 민간병의원을 이용한 환자(H)의 추가방문횟수는 보건기관이용자보다 유의하게 많았다. 이는 양 기관의 의료공급자에 대한 보수지불방식과 의사의 근무동기에 차이가 있기 때문일 것이다. 이러한 사실은 민간병의원을 이용한 의료보호환자(H×MS)의 추가방문이 의료보험적용자 보다 적다는 사실과 함께 의료공급자의 행태가 환자들의 의료이용에 큰 영향을 미친다는 가정을 뒷받침해 주는 좋은 증거라 하겠다. 다만 단골 병의원을 방문한 환자는 다른 의료기관을 방문한 환자들에 비해 만성유병일수(P1)가 외래 추가이용에 유의한 영향을 미쳤다. 이는 민간병의원이라 할지라도 자주 이용하는 환자들이 만성적으로 심한 질병을 앓고 있으면 투약기간을 늘려 주기 때문으로 보여진다.

V. 결 론

의료이용에 관한 분석은 의료체계의 성과를 판정하고 발전방안을 강구하는 데 필수적이다. 따라서 의료이용에 관한 연구는 보건의료분야의 핵심적 연구과제의 하나라 할 수 있다.

의료이용에 관한 분석시 결과의 타당성은 자료의 질이나 이론적 모형 뿐만 아니라 실증적 분석 모형의 설정과 모형의 추정기법에 의해 영향을 받게 된다. 그러므로 의료이용에 관한 연구의 가치를 높이기 위해서는 상기요인들을 충분히 고려하여야 한다.

외래의료이용의 결정과정은 본질적으로 2단계로 구성되어 있다. 즉 환자의 특성에 의해 주로 결정되는 최초방문과 의사의 권고가 크게 영향을 미치는 추후방문으로 구별된다. 또한 외래의료이용은 자료의 성질이 갯수변수이며, 그 분포가 정규분포를 하지 않는다는 특성을 지니고 있다.

그러나 아직도 외래이용에 관한 연구시 결정과정의 본질이나, 자료의 성격 및 분포를 고려하지 않는 경우가 많다. 본 연구는 이러한 문제인식하에 외래이용빈도 자료의 속성을 실증적으로 검토하고, 이의 분석에 적합하다고 생각되는 기법을 이용하여 실제자료를 분석하고, 결과의 타당성을 검토하는 데 그 목적이 있다. 연구자료는 1990년 강원도 화천군에서 수집된 의료이용에 관한 자료중 조사기간 15일간중 외래이용의 경험이 있었던 712명이다.

연구대상자들의 추가외래방문횟수를 일반회귀분석하여 잔차분석을 한 결과 외래추가방문횟수는 일반회귀분석으로 적합하는 것이 타당하지 않음을 알 수 있었다. 또한 실제 이용자료를 가상적 포아송분포와 비교한 결과 과분산이 있음도 알 수 있었다. 따라서 외래 의료이용은 과분산포아송분포를 이용하여 분석하는 것이 타당하다는 결론을 얻었다.

과분산포아송분석 결과 예상한 대로 환자개인의 특성을 반영하는 사회경제적 변수나 의료필요변수와 더불어 이용의료기관의 단골의사여부, 또는 민간의료기관여부 등의 의료공급자 특성이 외래추가방문횟수에 크게 영향을 미치는 것으로 나타났다.

그러나 의료공급자변수들도 단독으로 추가방문에 영향을 미치기보다는 질병의 중증도, 의료보장의 성격, 가구원 수 등 환자의 특성과 상호작용하면서 영향을 미쳤다. 이는 외래추가방문여부는 의사가 주로 결정하는 분야이긴 하지만, 의료이용의 속성상 본질적으로는 의사와 환자가 서로 조우한다는 사실을 감안하면 당연하다 하겠다.

본 연구와 같이 복잡한 사회현상에 관한 자료는 필요한 독립변수를 모두 분석에 포함하기 어려우므로 과분산이 발생하게 된다. 종속변수의 분포자체가 과분산을 하고 있는 경우도 많다.

따라시 적절한 분석기법을 찾아 분석결과의 타당성을 높이려는 노력이 반드시 필요하다. 또한 관측자료를 이용하게 되면 주요한 설명변수가 누락되는 경우가 많아, 이로 인한 회귀계수 분석에 오류가 있는지를 면밀히 검토하여야 한다.

참 고 문 헌

권순호, 한달선, 이규식 ; 의료수요분석의 방법과 결과에 대한 고찰, 한림대학교 사회의학연구소, 1987.

송건용, 김영숙, 김영임 ; 농어촌벽지 보건의료에 관한 연구, 한국인구보건연구원, 1983.

한달선, 권순호, 권순원, 황성주 ; 춘천시민의 의료이용양상과 관련요인, 한림대학 사회의학연구소, 1986.

Aitkin M., D. Anderson, B. Francis and J. Hinde(1989);Statistical Modelling in GLIM, Oxford Science Press.

Duan, N., W. Manning, C.Morris and J.Newhause(1983);"A Comparison of Alternative Models for the Demand for Medical Care", J. of Bussiness and Economic Statistics, 1:2, 115-26.

Hausman, J., B.H.Hall and Z. Griliches(1984);"Econometric Models for Count Data with an Application to the Patents-R & D Relationship". Econometrica, 52, 909~938.

King, G.(1988);Statistical Models for Political Science Event Counts: Bias in Conventional Procedures and Evidence for the Exponential Poisson Regression Model. Am. J. Pol. Sci. 32, 838~868.

Lee, Y. and J.A.Nelder(in press); "The Robustness of Estimators from Quasi-Likelihood and Quadratic Estimating Functions", A manuscriptsubmitted for publication.

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

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.

Pohlmeir, W. and V.Ulrich(1992);"An Econometric Model of the Two Part Decision Process of the Demand for Health". Paper presented at the 48th Congress of the International Institute of Public Finance, Seoul.