
고령화 연구 패널 조사 자료를 이용한 입원 일수에 미치는 요인 분석

박 수 영

서울시립대학교 통계학과

2018년 12월 24일

초록

제로팽창 개수형 자료는 반응변수가 영의 값이 가정된 분포의 0에 대한 기대 관측치에 비하여 과다하게 나타나는 형태의 자료를 뜻한다. 자료가 가정된 분포의 기댓값에 비해 추정된 분산이 훨씬 큰 경우 포아송 모형에 과대산포 문제가 발생할 수 있다. 이러한 자료의 특성들을 고려하여 적절한 모형을 선택하고 올바른 분석을 하고자 한다. 본 연구에서는 제로 팽창 모형과 허들모형을 각 가정된 분포와 연결함수 및 모수를 이용한 방법을 이용하였으며 예제로 고령화 연구 패널조사의 자료를 이용한다. 본 연구에서는 두 모형의 적합도를 AIC와 로그 가능도 값을 이용하여 비교하였다. 아울러 고령자들의 입원 일수로 환자들을 속성으로 설명할 수 있도록 하며 분류하여 보험 설계를 위한 마케팅 제언을 목표로 한다. 데이터 분석 결과, 음이항 허들 모형이 입원 일수 자료에서 가장 AIC 값이 작게 나타나고 로그 가능도 값이 크게 나타나 잘 적합됨을 알 수 있었다.

keyword : 고령화 연구 패널 조사, 입원 일수, 영과잉 모형, 허들모형

목 차

초록	1
I. 서론	3
가. 연구의 배경 및 필요성	3
나. 선행 연구	3
다. 연구 목적	4
라. 연구 방법	4
II. 이론적 배경 및 모형	4
가. 제로팽창 모형	4
나. 허들 모형	5
III. 실증 분석	6
가. 데이터 소개 및 변수 설명	6
나. 모형 적합	7
(1) 데이터 탐색	7
(2) 모형 적합	8
(3) 모형 개선	8
다. 모형 비교	9
라. 최종 모형에 대한 고찰	9
IV. 결론 및 시사점	11
V. 참고문헌	12

I. 서론

가. 연구의 배경 및 필요성

한국은 OECD국 중 가장 빠른 고령화 속도로 올해 고령사회에 진입하였다. 지난 7월 기준 65세 이상 고령자 수는 738만여 명으로, 처음으로 고령사회 기준인 14%를 넘어섰다. 이와 같이 우리나라는 앞으로 초고령사회로의 진입이 세계적으로 유례없는 속도로 진행될 것으로 예상되며 이에 대한 대비책 마련 등이 사회적 이슈로 부각되고 있다. 그러나 국민건강보험공단의 수술통계 연보에 따르면 수술환자의 평균 입원일수가 감소추세를 보이고 있음에도 불구하고 수술 건당 진료비는 증가해 환자들의 부담이 늘어나고 있다. 즉, 병원 이용률이 높은 고령환자층의 부담은 훨씬 가중될 것으로 예상된다. 흥미로운 점은, 실제로 만성 질병이 대부분인 노인들의 단기 의료서비스 수요가 크게 증가하였다는 점인데, 이는 의료 기술의 발달로 수술 기간 자체는 감소하였지만 진료비는 증가한 배경 때문임을 암시한다. 따라서 이 수요를 효과적으로 예방 및 관리하며 또한 그에 맞는 의료이용서비스 및 보험제도에 대한 연구의 필요성이 제기된다.

이에 한국고용정보원에서는 고령화연구패널조사(KLoSA)를 실시하여 우리나라의 실정을 파악하여고효과적인 사회복지정책을 수립 및 연구하는 데에 활용될 기초 통계자료를 구축하였다. 고령화연구패널조사는 2006년 당시 제주도를 제외한 지역에 거주하는 45세 이상 중고령자 중 일반 가구 거주지를 대상으로 표집 및 조사를 실시하였으며, 이후 짝수 연도에는 동일한 조사 항목을 중심으로 기본조사를 실시하고, 홀수 연도에는 기본조사에 포함되지 않은 내용을 중심으로 특정 주제를 정하여 조사를 실시하고 있다. 2016년 제6차 기본조사를 실시하였으며 본 연구에서는 제6차 기본조사 자료를 이용하였다.

따라서 본 연구는 고령화연구패널조사를 이용하여 노인의 의료서비스 이용에 대하여 입원 일수 측면에서 어떠한 요인들이 영향을 미치는지 살펴보고 그에 따른 의료이용의 차이를 비교 및 예측하여 현재 고령 인구의 의료서비스 행태를 분석하고자 한다. 따라서 이는 노인 복지를 넘어서 고령사회를 대비한 맞춤형 의료서비스 이용체계를 개선하여 효율적인 고령 세대의 의료 서비스 비용 부담을 위한 보험 제도를 수립할 수 있는 근거가 되는 데에 본 연구의 필요성이 제기된다.

나. 선행 연구

의료서비스 이용은 인구학적 요인, 사회경제적 요인, 문화 심리적 요인 등에 영향을 받으며 각 요인 간의 상호작용 속에서 의료서비스이용에 영향을 미친다고 보고되어 왔다(배상수, 1992). 입원 일수는 의료서비스 이용의 한 부분으로 질병으로 인한 치료의 의미와 만성적인 질병에 대한 예방까지의 의미를 포함할 수 있다. 의료서비스 이용에 관한 연구에서 주로 이용되는 Andersen (1978)¹의 모형은 소인성 요인, 가능성요인, 의료수요요인을 포함한다. Andersen에 따르면 인구학적 변수 및 사회 구조적 변수를 포함하는 소인성 요인은 의료이용에 직접적인 영향보다는 가능성 요인과 의료 수요 요인을 통해 간접적으로 영향을 미친다고 제시했다.

한편, Deb, P., & Trivedi, P. K. (1997)²는 노인들의 의료 지출이 젊은 인구보다 훨씬 빠르게 증가하였고 이들의 건강관리에 대하여 상대적으로 더 큰 필요성을 제시하였다. 또한 건강관리 수요의 결정요인으로서 건강상태와 소득의 상대적 중요도를 강조하였는데 이에 ZINB(Zero-Inflated Negative Binomial) 모형을 사용하였다. Deb, P., & Trivedi, P. K.의 연구에 따르면 만성질환의 수

¹ Andersen, R.M. (1978), "Health status indices and access to medical care", American Journal of Public Health, 68(5): 213-230.

² Deb, P., & Trivedi, P. K. (1997). Demand for medical care by the elderly: a finite mixture approach. Journal of applied Econometrics, 12(3), 313-336.

나 장애 여부 등이 의료서비스 이용의 증가에 영향을 미치지만 인종이나 거주 지역 등 인구학적 특성은 크게 영향을 미치지 못한다고 나타났다.

다. 연구 목적

따라서 외국에서는 흔히 일어나는 연구에 대하여 본 연구에서는 국내 데이터로의 연구에 있어서 입원 일수의 요인 분석을 통한 연구의 의의를 두며 다음과 같은 구체적인 연구 목적을 지닌다.

첫째, 노령인구의 의료서비스 이용 중의 한 형태인 입원 일수에 영향을 주는 요인을 인구학적 특성, 의료서비스 필요요인으로 파악한다.

둘째, 노령인구의 입원일수 요인분석 결과에 따른 요인별 의료 서비스 이용 행태를 비교 분석하여 올바른 의료서비스 이용체계를 개선토록 결과를 제언한다.

라. 연구 방법

본 연구에서는 입원 일수에 미치는 변수와 그 유의성을 알아내기 위해 적절한 모형을 선택한다. 고령화 연구 패널 조사에서 고령자 대상으로 지난 조사 이후 혹은 지난 1년간 입원 일수에 미치는 요인들을 분석한다. 여기서 입원 일수 변수는 영(0)의 값이 과다하게 많은 영 과잉 개수형 자료이다. 이에 대하여 본 연구에서는 제로팽창 음이항 회귀모형과 음이항 허들모형을 제안한다. 여기서 모형 비교 통계량은 AIC와 Log-likelihood를 이용한다. 2장에서는 모형에 대하여 기술하고 3장에서는 실제 데이터에 적용하여 각 모형의 모수를 추정하고 각 모형의 적합도를 비교하고자 한다. 마지막으로 6장에서는 본 연구의 결론을 기술한다.

II. 이론적 배경 및 모형

가. 제로팽창 모형

제로팽창 개수형 자료는 반응변수 $Y_i, i = 1, 2, \dots, N$ 가 음이 아닌 정수 값을 가지며 0의 값이 고정된 분포에서 기대되는 관측 도수에 비하여 과다하게 포함될 경우 이를 영과잉 개수형 자료(zero-inflated count data)라 한다.

(i) 제로팽창 포아송 회귀모형

제로팽창 자료에 대하여 0이 될 확률을 π_0 ($0 \leq \pi_0 < 1$), 포아송 모수를 λ 라 할 때 제로팽창 포아송(ZIP) 분포의 확률질량함수는 다음과 같다.

$$P_r(Y = y) = \begin{cases} \pi_0 + (1 - \pi_0)\exp(-\lambda) & , Y = 0 \\ \frac{(1 - \pi_0)\exp(-\lambda)\lambda^y}{y!} & , Y > 0 \end{cases}$$

Lambert(1992)³에 의해 제안된 ZIP 분포에 기반한 회귀 모형을 적합하고자 할 때 다음과 같은 연결함수를 생각할 수 있다. 이때 X 와 Z 는 각각 평균 모수와 제로팽창 발생오즈에 영향을 미

³ Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics, 34, 1-14.

치는 공변량 행렬이며 β 와 γ 는 모수벡터이다.

$$\log(\lambda) = X\beta, \quad \log\left(\frac{\omega}{1-\omega}\right) = Z\gamma$$

(ii) 제로팽창 음이항 회귀모형

0이 과다하게 포함되는 영과잉 자료에서 평균보다 분산이 기대보다 훨씬 크게 나타나는 경우 포아송 모형을 가정하기가 어렵다. 따라서 음이항 분포를 가정한 제로팽창 모형은 아래와 같다.

$$P_r(Y = y) = \begin{cases} \pi_0 + (1 - \pi_0)(1 + \alpha r)^{-\alpha^{-1}} & , Y = 0 \\ (1 - \pi_0) \frac{\Gamma(y + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y + 1)} (1 + \alpha r)^{-\alpha^{-1}} \left(\frac{\alpha r}{1 + \alpha r}\right)^y & , Y > 0 \end{cases}$$

여기서 위의 로짓 연결함수를 사용하여 음이항 회귀 모형을 정의한다.

나. 허들 모형

Mullahy(1986)⁴에 의해 정의된 허들 모형에서는 반응변수 $Y_i, i = 1, 2, \dots, N$ 가 0의 값이 발생하는 확률을 결정한다. 여기서 확률은 0값이 발생하는 확률과 0을 포함하지 않는 분포로부터의 확률의 혼합 분포 형태이다. 0이 발생할 확률을 π_0 , 그렇지 않을 확률을 $\pi_+ = 1 - \pi_0$ 라 할 때, 허들 모형은 다음과 같다.

$$P_r(Y = y) = \begin{cases} \pi_0 & , Y = 0 \\ \frac{\pi_+}{\pi_0} f(Y) & , Y > 0 \end{cases} \quad , \pi_+ = 1 - \pi_0$$

(i) 포아송 허들 모형

위의 모형에 가정된 분포로 포아송 분포를 적용하면 다음과 같이 표기할 수 있다.

$$P_r(Y = y) = \begin{cases} \pi_0 & , Y = 0 \\ \frac{(1 - \pi_0)e^{-\lambda}\lambda^y}{(1 - e^{-\lambda})y!} & , Y > 0 \end{cases}$$

여기서 π_0 는 제로 팽창 확률로 독립변수들에 의한 영향을 고려하여 회귀모형이 제안된다. 이때 연결함수에 따라 모형이 상이하다. 본 연구에서 연결함수는 로짓 연결함수를 고려한다. X 와 Z 가 독립변수의 공변량 벡터이며 β 와 γ 가 모수벡터일 때 로짓 연결함수는 다음과 같이 정의할 수 있다.

$$\log(\lambda) = X'\beta, \quad \text{logit}(\pi_0) = \log\left(\frac{\pi_0}{1 - \pi_0}\right) = Z'\gamma$$

(ii) 음이항 허들 모형

음이항 분포를 가정한 허들 모형은 다음과 같이 표기할 수 있다. 음이항 허들 모형 또한 위치럼 연결함수를 로짓 연결함수와 cloglog 연결함수를 사용할 수 있다.

$$P_r(Y = y) = \begin{cases} \pi_0 & , Y = 0 \\ (1 - \pi_0) \frac{\Gamma(y + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y + 1)} (1 + \alpha r)^{-\alpha^{-1}} \left(\frac{\alpha r}{1 + \alpha r}\right)^y & , Y > 0 \end{cases}$$

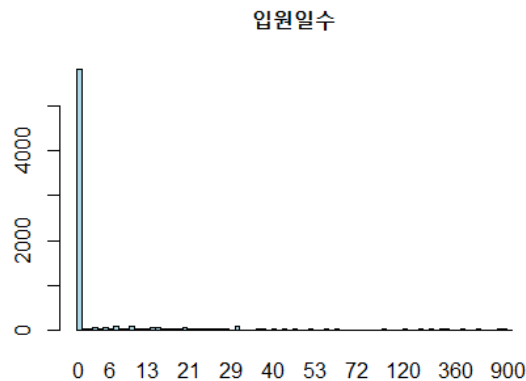
⁴ Mullahy, J. (1986). Specification and testing of some modified count data models, Journal of Econometrics, 33, 341 - 365

III. 실증 분석

가. 데이터 소개 및 변수 설명

본 연구에서는 한국고용정보원의 고령화연구패널조사의 16년도 6차 조사자료의 라이트버전을 사용하였다. 유효표본수는 6618명, 변수의 수는 403개이다. 이때 반응변수는 2년 전 실시했던 지난 5차 조사 이후 입원한 일수를 사용한다. 변수들은 7분류로 나눌 수 있다. 인적속성, 가족, 건강 관련 변수, 근로 상태, 소득 및 소비, 자산, 주관적 기대감의 7분류 중 의미 있는 변수 15개를 주관적으로 선택하였다. 반응 변수인 입원 일수는 0에서 365까지의 값을 가지므로 포아송 모형을 적합시킬 수 있으나 0이 전체 데이터의 88.29%를 차지하므로 영과잉 자료이다. 또한 평균은 3.609, 분산은 413.306으로 과대산포를 보인다고 할 수 있다.

[그림 1 : 입원 일수의 히스토그램]



[표 1: 데이터에 사용된 변수 설명]

변수 설명	변수
지난 기본조사 이후 입원일 수(1 차는 가장 최근 입원에 대한)	days
응답자 학력(1=초등졸업이하,2=중졸업,3=고등졸업,4=대학졸업이상)	edu
응답자 성별(1=남자, 0=여자)	gender
응답자 연령(=조사년도-응답자 출생연도)	age
**기본조사 당시 현재 결혼상태 (1=혼인 중, 0=그 외)	marital
가족(자녀, 부모, 기타가족)으로부터 지원 받은 금전 총액	money_from
응답자(본인)의 건강상태 (1=최상~5=매우나쁨)	health
지난 기본조사 당시와 비교한 주관적 건강상태(2 차 이후)	sub_health
건강상태로 인한 활동(일) 제한(1=매우 그렇다~4=전혀 그렇지않다)	limit_work
만성질환 개수	chronic_sum
규칙적 운동 여부(일주일 1 회 이상) (1=예, 0=아니오)	work_out
**흡연자 구분(1=현재 흡연, 0=현재 비흡연)	smoke
**음주자 구분(1=현재 음주, 0=현재 비음주)	alcohol
민간의료보험 가입여부(1=예, 0=아니오)	ins2
외래 지불한 비용 (응답자가 지불한 비용만 포함)(단위 : 만원)	oopm
** (공통)현재 경활상태(1=취업자, 0=비취업자)	ecoact

** : 의미상 2원화 더미코딩하였음.

나. 모형 적합

(1) 데이터 탐색

각 변수에 대하여 기초 통계량을 구하면 다음과 같다.

[표 2: 각 변수들의 기초통계량]

변수	N	평균	표준편차	최솟값	최댓값
days	6595	3.6092494	20.3299372	0	365.00
edu	6595	2.0315390	1.0449004	1.00	4.00
gender	6595	0.4212282	0.4937935	0	1.00
age	6595	0.2842480	0.1845682	0	1.00
marital	6595	0.7425322	0.4372724	0	1.00
money_from	6595	0.0044469	0.0195281	0	1.00
health	6595	3.7821077	0.8692056	1.00	5.00
sub_health	6595	3.2589841	0.6024990	1.00	5.00
limit_work	6595	2.6705080	0.8046172	1.00	4.00
chronic_sum	6595	1.3519333	1.2387292	0	7.00
work_out	6595	0.3413192	0.4741882	0	1.00
smoke	6595	0.1034117	0.3045189	0	1.00
alcohol	6595	0.3147839	0.4644650	0	1.00
ins2	6595	0.3102350	0.4626248	0	1.00
oopm	6595	0.0296593	0.0643443	0	1.00
ecoact	6595	0.3528431	0.4778906	0	1.00

데이터에서 입원 일수와 각 변수 간의 상관관계를 구해본 결과는 다음과 같다. 입원 일수와 주관적 건강상태의 상관계수가 0.236으로 가장 높게 나타났다.

[표 3: 각 변수들의 반응변수 Y와의 상관관계]

변수	edu	gender	age	marital	money_from
cor	-0.0779921	-0.002197949	0.09797675	-0.04888353	0.07455716
변수	health	sub_health	limit_work	chronic_sum	work_out
cor	0.1761943	0.2361742	-0.1985297	0.1178014	-0.06678593
변수	smoke	alcohol	ins2	oopm	ecotype
cor	-0.02046687	-0.06602199	-0.04652762	0.01601526	-0.077153

입원 일수 자료에 일반 선형회귀모형 적합 후 다중공선성 통계량을 살펴본 결과 10이상의 값이 나타나지 않아 변수들 간 다중공선성은 없다고 할 수 있다.

[표 4: 각 변수들의 다중공선성]

변수	edu	gender	age	marital	money_from
cor	1.545303	1.668693	2.122700	1.310013	1.023559
변수	health	sub_health	limit_work	chronic_sum	work_out
cor	1.925037	1.573105	1.558950	1.342165	1.141298
변수	smoke	alcohol	ins2	oopm	ecotype
cor	1.208378	1.338248	1.451278	1.057332	1.462997

(2) 모형 적합

입원 일수 자료에 대하여 포아송 회귀 모형을 적합 시킨 결과 모든 변수가 모형에 대하여 유의하다고 나타났다. 그러나 AIC 값이 98018.95, BIC 값이 98127.65로 너무 크고 및 Log-Likelihood 값이 자유도가 16인 -48993.47로 너무 작게 나타났다. 또한 이 자료는 과대산포가 존재하므로 다음에서 음이항 회귀모형을 고려한다. 음이항 회귀모형을 적합시키면 AIC는 11435.99, BIC는 11551.49, 로그가능도값은 자유도 17인 -5700.993으로 나타나며 이는 포아송 회귀모형보다는 비교적 안정화된 값이다. 그러나 이 자료는 제로 팽창 데이터이므로 이에 대하여 고려하기 위하여 제로 팽창 모형을 적용한다.

로짓 연결함수를 사용하여 제로팽창 포아송 회귀모형을 적합시킨 결과, AIC값은 35536 BIC가 35753, 로그가능도 값은 -17736으로 포아송 회귀모형보다는 안정화되었으나 위의 음이항 회귀 모형보다는 여전히 좋지 않게 나타났다. 따라서 과대산포를 고려하기 위해 제로팽창 음이항 회귀 모형을 적합시킨다. 로짓 연결함수를 사용한 제로팽창 음이항 회귀모형의 결과, Log(thata)로 나타나는 과대산포가 존재하는 데이터의 모형의 유의성에 대하여 p-value가 유의하다고 나타났다. 이에 AIC 10655.99, BIC 10880, 로그가능도 -5294.995으로 위의 음이항 회귀모형보다 적합도가 더 좋게 나타났다.

제로팽창을 보인 데이터를 고려하는 또다른 모형인 허들 모형을 고려해보자. 입원 일수에 대하여 포아송 분포를 가정한 허들 모형을 적합시킨 결과, AIC는 35536, 로그가능도는 자유도가 32인 -17736으로 나타났다. 위의 제로팽창 음이항 회귀모형보다도 좋지 않은 적합 결과가 나타났다. 따라서 과대산포를 컨트롤하는 허들모형을 적합시키기 위해 포아송이 아닌 음이항 분포를 가정한다. 음이항 허들모형을 적합시킨 결과 AIC는 10656.57, 로그가능도는 자유도가 33인 -5295.284로 나타났다.

(3) 모형 개선

모형 개선에서는 위에서 나타난 모형들의 제로팽창 모수를 추정하여 공변량으로 둔 모형을 적합시킬 것이다. 여기서는 음이항 분포를 가정한 두 모형인 제로팽창 음이항 회귀모형과 음이항 허들모형에 대하여 다뤄 각 모형 적합도가 더 좋은 모형을 찾는다.

첫째, 제로팽창 음이항 회귀모형에서 제로팽창 오즈에 유의하다고 나타난 변수는 총 6개였다. 이에 대하여 제로팽창 오즈의 공변량으로 두고 모형을 적합시킨 결과, AIC는 10651.11, 로그 가능도 값은 자유도 24를 가진 -5301.554로 나타났다. 위의 모든 변수를 공변량으로 두었던 제로 팽창 음이항 회귀모형과 비슷하지만 조금 덜 좋은 적합도를 보였다.

둘째, 음이항 허들모형에서 제로팽창 오즈에 유의한 변수 총 6가지를 제로팽창에 대한 공변량으로 두고 모형을 적합시킨 결과, AIC는 10651.16, 로그 가능도 값은 자유도가 24인 -5301.581로 마찬가지로 모든 변수를 공변량으로 고려한 모형보다 조금 덜 적합된 모습을 보였다.

다. 모형 비교

지금까지 입원 일수 자료에 대하여 모형을 적합시킨 모형 선택 통계량 값들은 다음과 같다.

[표 5: 8가지 모형의 AIC와 로그가능도]

	AIC	Log-Likelihood(df)
포아송 회귀모형	98018.95	-48993.47(16)
음이항 회귀모형	11435.99	-5700.993(17)
ZIP	35536	-17736(32)
ZINB	10655.99	-5294.995(33)
포아송 허들모형	35536	-17736(32)
음이항 허들모형	10656.57	-5295.284(33)
조정된 ZINB	10651.11	-5301.554(24)
조정된 음이항 허들	10651.16	-5301.581(24)

위의 8가지 모형을 비교하기 위하여 LR 검정을 실시한다. 첫째, 위의 두 모형 일반화 선형 회귀모형인 포아송 회귀모형과 음이항 회귀모형에 대하여 LR 검정을 실시한 결과, LR 검정 통계량은 $-2(L_0 - L_1) = -2(-48993.47 - (-5700.993)) = 86584.954$ 로 나타났다. 두 모형의 자유도의 차로 나타나는 LR 통계량은 자유도가 $|v_0 - v_1| = 1$ 인 카이제곱 분포에서 기각역에 속한다. 따라서 자료에는 통계적으로 유의하게 과대산포가 존재한다고 할 수 있다. 둘째, 제로팽창에 대한 검정을 위해 음이항 회귀모형과 제로팽창 음이항 회귀모형에 대하여 LR 검정을 실시한다. LR 검정 통계량은 $-2(L_0 - L_1) = -2(-5700.993 - (-5294.995)) = 811.996$ 으로 카이제곱 분포에서 귀무가설을 통계적으로 유의하게 기각할 수 있다. 따라서 이 모형은 영과잉이 존재한다고 할 수 있다. 또한 제로팽창 음이항 모형과 음이항 허들모형 간 적합도 검정을 위해 LR 검정을 실시하면 통계적으로 유의한 적합도의 차이를 보이지는 않았다. 따라서 BIC를 비교하면

결과적으로, 입원 일수는 과대산포가 큰 자료이므로 포아송 모형보다 음이항 모형의 적합이 더 좋은 결과를 보여주었다. 또한 자료의 영과잉으로 인해 일반 선형회귀모형보다는 제로팽창을 고려한 제로팽창 모형 및 허들모형이 더 좋은 모형선택 통계량 값을 보였다. 따라서 로그가능도 값이 가장 크고 AIC 값이 비교적 낮은 ZINB 모형을 최종적으로 선택한다.

라. 최종 모형에 대한 고찰

최종적으로 선택된 ZINB 모형에 대한 결과가 아래 [표 6]에 나타나 있다. 아래의 결과로부터 어떠한 설명변수들이 입원 일수에 얼마나 영향을 미치는지 살펴볼 수 있다.

먼저 유의수준 0.05내에서 평균 모수 λ 에 유의한 변수들은 상수항, 최종 학력, 성별, 본인의 건강상태, 주관적 건강상태, 건강상태로 인한 활동 제한 수준, 민간의료보험 가입 여부로 총 7개의 변수가 나타났다. 먼저, 최종학력이 한 단위 높을수록 반응변수가 $\exp(-0.13496) = 0.87$ 배 높아진다. 즉, 최종학력이 낮을수록 병원에 입원한 일수가 많아졌다. 둘째, 성별이 남자일수록 입원일수가 $\exp(0.27693) = 1.32$ 배 높아진다. 즉, 남자일수록 병원에 입원한 일수가 많아졌다. 셋째, 건강상태가 한 단위 더 나쁠수록 입원일수가 $\exp(0.15814) = 1.17$ 배 높아지므로 건강상태가 나쁠수록 입원일수가 많아졌다. 넷째, 본인의 주관적 건강상태가 한 단계 나쁠수록 입원일수가 $\exp(0.18) = 1.20$ 배 높아지므로 주관적 건강상태가 나쁠수록 입원일수가 많아졌다. 다섯째, 건강상태로 인한 일(활동) 제한 수준이 한 단위 덜할수록 입원일수가 $\exp(-0.23439) = 0.79$ 배 높아졌다. 즉, 건강상태로 인한 일 제한을 느낄수록 입원일수가 많아졌다. 여섯째, 민간의료보험을 가입했을수록 입원일수가 $\exp(0.22896) = 1.26$ 배 높아지므로 민간의료보험을 가입한 경우 입원일수가 많아졌다.

[표 6 : ZINB 모형 적합 결과]

변수	추정치		표준오차		P-value	
	λ	π_0	λ	π_0	λ	π_0
(Intercept)	2.56932	5.94215	0.43454	0.45861	<.0001 ***	<.0001 ***
edu	-0.13496	0.01713	0.04899	0.05185	0.00587 **	0.741056
gender	0.27693	-0.15204	0.10026	0.10803	0.00574 **	0.159334
age	0.16387	0.43330	0.30558	0.32427	0.59178	0.181469
marital	-0.07218	-0.08510	0.10343	0.10505	0.48527	0.417885
money_from	2.53615	-5.01270	1.34354	2.19705	0.05907	0.022516 *
health	0.15814	-0.37736	0.06905	0.07431	0.02202 *	<.0001 ***
sub_health	0.18103	-0.79626	0.06423	0.07956	0.00483 **	<.0001 ***
limit_work	-0.23439	0.32540	0.06209	0.06506	0.00016 ***	<.0001 ***
chronic_sum	0.00311	-0.25757	0.03245	0.03545	0.92366	<.0001 ***
work_out	-0.15934	0.02492	0.10591	0.10098	0.13247	0.805069
smoke	-0.10510	-0.06761	0.16027	0.15997	0.51197	0.672566
alcohol	-0.21288	0.22167	0.10902	0.11374	0.05086	0.051311 .
ins2	0.22896	-0.44499	0.11579	0.11643	0.04800 *	0.0001 ***
oopm	-0.96704	-0.92961	0.54276	0.58676	0.07480 .	0.113121
ecoact	-0.16809	-0.08363	0.11017	0.11460	0.12708	0.465519
Log(theta)	-0.15922		0.06623		0.01622 *	

다음으로 입원을 하지 않았을 확률 ω 에 유의한 변수들은 상수항, 가족으로부터 지원받은 금액, 건강상태, 주관적 건강상태, 활동 제한수준, 만성질환의 수, 민간의료보험 가입여부로 총 6개의 독립변수가 나타났다. 첫째, 가족으로부터 지원받은 금액이 높을수록 제로팽창의 발생 오즈가 $\exp(-5.01270) = 0.067$ 배 높아졌다. 즉, 금액을 적게 지원받았을수록 입원을 하지 않았을 확률이 높아졌다. 둘째, 건강상태가 한 단위 더 나쁠수록 제로팽창 오즈가 $\exp(-0.37736) = 0.69$ 배 높아졌다고 나타났고 이는 건강상태가 좋을수록 입원하지 않았을 확률이 높음을 의미한다. 셋째, 본인의 주관적 건강상태가 한 단위 더 나쁠수록 제로팽창 오즈가 $\exp(-0.79626) = 0.45$ 배 높아졌으며 이는 주관적 건강상태가 좋을수록 입원하지 않았을 확률이 높음을 의미한다. 넷째, 일(활동) 제한을 느끼지 않을수록 제로팽창 확률의 오즈가 $\exp(0.32540) = 1.38$ 배 높아졌고 이는 건강상태로 인한 활동의 제한을 느끼지 않았을수록 입원을 하지 않았을 확률이 높아졌음을 의미한다. 다섯째, 만성질환의 수가 많을수록 제로팽창 오즈가 $\exp(-0.25757) = 0.77$ 배 높아졌다. 즉, 만성질환의 수가 적을수록 입원하지 않았을 확률이 높아졌다. 마지막으로, 민간의료보험을 가입했을수록 제로팽창 오즈가 $\exp(-0.44499) = 0.64$ 배 높아졌다. 즉, 민간의료보험을 가입하지 않은 경우 입원하지 않았을 확률이 높아졌다고 할 수 있겠다.

[표 7]은 실제 데이터와 모형에서 예측값을 비교한 결과이다. 실제 데이터에서 0이 나타난 경우는 88.3%로 매우 높게 나타났다. 최종 모형인 ZINB로 예측값을 구해본 결과 45.7%로 0이 나타났고 이는 실제 데이터와는 다소 차이를 보였다. 하지만 여기서 과대산포 지수를 보면 113.5에서 10.9로 많이 낮아진 것을 볼 수 있다.

[표 7 : 실제 데이터와 ZINB 모형의 예측값]

days	실제 데이터		ZINB 적합	
	빈도	백분율(%)	빈도	백분율(%)
0	5823	88.3	3011	45.7
1	10	0.2	1291	19.6
2	10	0.2	514	7.8
3	35	0.5	285	4.3
4	24	0.4	204	3.1
5	46	0.7	165	2.5

6	16	0.2	160	2.4
7	89	1.3	107	1.6
8	10	0.2	126	1.9
9	10	0.2	77	1.2
10	71	1.1	81	1.2
11	4	0.1	62	0.9
12	7	0.1	66	1.0
12+	440	6.7	446	6.8
no. of days	6595		6595	
mean	3.61		3.56	
var	413.31		42.37	
overdispersion	113.51		10.90	

IV. 결론 및 시사점

본 연구에서는 0이 과다하게 포함된 입원 일수 자료에 대한 모형 적합에 대해 살펴보았다. 조사된 고령인구에 대하여 지난 조사 이후 입원하지 않은 비율이 전체 중 88%로 매우 높은 편이다. 이 경우 일반화 선형회귀모형에서는 설명이 불가능하여 이에 대한 적합으로 제로팽창 모형 및 허들모형을 적합하였다. 또한 입원 일수에 대하여 과대산포가 존재하여 포아송 모형에서 관측할 수 없는 주정의 오류를 고려하기 위하여 음이항 분포를 가정한 제로팽창 모형 및 허들모형을 최종적으로 제안하였다.

위에서 고려한 총 8가지 모형을 모두 적용시킨 결과, 음이항 분포를 가정한 제로팽창 모형 및 허들모형이 가장 좋은 적합도를 보였고 음이항 회귀모형, 포아송 분포를 가정한 제로팽창 모형 및 허들모형, 포아송 회귀모형 순으로 좋은 결과를 보임을 살펴볼 수 있었다. 특히 이 자료에서는 음이항 허들모형보다 ZINB 모형이 가장 우수한 모형으로 나타나 모형 선택 및 진단 과정에서 최종적으로 선택되었다.

본 연구의 한계점으로는 첫째, 소득을 연구에서 고려하지 않았다. 근본적으로 소득이 적을수록 의료서비스를 이용하지 못할 가능성이 크나 본 연구에서 사용한 데이터에서는 소득에 대한 질문이 포함되었으나, 응답률이 낮아 사용하지 못하였다. 따라서 가장 근본적인 독립변수를 이용하지 못하여 대체적으로 가족으로부터 지원받은 금액과 외래진료비용을 사용하였으나 이는 충분히 대체되지 못하여 크게 유의하지 못한 변수로 나타났다. 둘째, 데이터 내에서는 고령 응답자의 세부 사항 별로 기록이 되어있으나 변수의 단순화로 인해 분석에서는 고려하지 못하였다. 일례로, 만성 질환의 수가 아닌 응답자가 보유한 만성질환의 종류별로 더미변수로 둔 후 모형을 적합시켰다면 의학통계학적으로 더 의미있는 연구가 되었을 것이나 이를 고려하지 못하였다. 따라서 연구가 환자의 질환으로 초점을 맞추고 분석을 확장한다면 더 다양한 설명이 가능했을 것이다. 셋째, 실제 입원 일수에 대하여 0에서 365의 값을 가지므로 연속형 변수로 취급할 수 있는 선택지가 존재하는 것이 사실이다.

그러나 이러한 한계점에도 본 연구의 의의에 맞는 결과가 나타났다. ZINB 모형 적합 결과, 입원 일수에 영향을 미치는 변수들과 그 영향성에 대하여 알 수 있었다. 따라서 분석을 통해 입원에 미치는 요인들을 살펴볼 수 있었으며 나아가 본 연구를 통해 입원 수요가 많은 고령인구에 대하여 의료서비스 이용 형태 및 이용 개선을 위한 제언점을 제시할 수 있다.

V. 참고문헌

- [1] Ridout, M., Hinde, J., Demetrio, C. G. B. (1998). Models for count data with many zeros, Proceedings of the XIXth International Biometrics Conference, Cape Town, Invited Papers, 179–192.
- [2] Andersen, R. M. (1978). Health status indices and access to medical care, American Journal of Public Health, 68(5): 213–230.
- [3] Deb, P., & Trivedi, P. K. (1997). Demand for medical care by the elderly: a finite mixture approach. Journal of applied Econometrics, 12(3), 313–336.
- [4] Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics, 34, 1–14.
- [5] Mullahy, J. (1986). Specification and testing of some modified count data models, Journal of Econometrics, 33, 341–365
- [6] 선지영, 심정숙, 정병철(2014). 허들모형에 대한 베이지안 추론, Journal of The Korean Data Analysis Society: Vol. 16. No. 4 (B). pp. 1837–1847
- [7] 이병란(2007), “노인의 의료서비스 이용에 영향을 주는 요인에 관한 연구”, 가톨릭대학교 석사 학위 논문.