



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

碩士學位 請求論文
指導教授 李 恒 碩

실손의료보험
실적변수 도입 유의성 검증

成均館大學校 一般大學院

保險計理學科

李 加 恩

碩士學位請求論文

실손의료보험
실적변수
도입
유의성
검증

2
0
1
7

李
加
恩

碩士學位 請求論文
指導教授 李 恒 碩

실손의료보험
실적변수 도입 유의성 검증

Verifying statistical significance of private health
insurance's past performance variables using GLM

成均館大學校 一般大學院

保險計理學科

李 加 恩

碩士學位 請求論文
指導教授 李 恒 碩

실손의료보험 실적변수 도입 유의성 검증

Verifying statistical significance of private health
insurance's past performance variables using GLM

이 論文을 保險計理學 碩士學位請求論文으로
提出합니다.

2017 年 10 月 日

成均館大學校 一般大學院

保險計理學科

李 加 恩

이 論文을 李加恩의 保險計理學
碩士學位 論文으로 認定함

2017 年 12 月 日

審査委員長

審査委員

審査委員

목차

제1장 서론	1
제1절. 연구목적	1
제2절. 선행연구	2
제3절. 논문의 구성	3
제2장 일반화선형모형 이론	4
제1절. 일반화선형모형 방법론	4
제3장 현행요율체계에 대한 GLM 분석	8
제1절. 빈도	8
제2절. 심도	12
제4장 GLM을 활용한 사후적 요율 산정	16
제1절. 빈도	16
제2절. 심도	21
제5장 결론	27
제1절 연구의 결과	27
제2절 연구의 의의	28
참고문헌	30

표목차

<표 1-1> 상위 3개사 보험료 인상률, 손해율 합계 평균	1
<표 1-2> 보험금 미청구 기간에 따른 환급률 (홍콩 AIA)	2
<표 1-3> 연속 미청구 보험연도 횟수에 따른 할인율 (말레이시아 AXA)	3
<표 3-1> 사전적 요율산정 사고건수 빈도 GLM 결과값 50대	9
<표 3-2> 사전적 요율산정 사고건수 상대도 50대	10
<표 3-3> 사전적 요율산정 사고건수 빈도 50세 예시	11
<표 3-4> 사전적 요율산정 사고건수 빈도 59세 예시	11
<표 3-5> 사전적 요율산정 지급금액 심도 결과값 50대	13
<표 3-6> 사전적 요율산정 심도 상대도 50대	14
<표 3-7> 사전적 요율산정 지급금액 심도 예시 값 50세	15
<표 3-8> 사전적 요율산정 지급금액 심도 예시 값 59세	15
<표 4-1> 과거실적 표시 변수 예시	17
<표 4-2> 사후적 요율산정 사고건수 빈도 GLM모형 결과값 50대	18
<표 4-3> 사후적 요율산정 사고건수 빈도 GLM모형 결과값 50대	19
<표 4-4> 사후적 요율산정 빈도 사고건수 GLM모형 50세 상해급수 1급 예시값 · 20	
<표 4-5> 사후적 요율산정 빈도 사고건수 GLM모형 59세 상해급수 1급 예시 값 21	
<표 4-6> 사후적 요율산정 지급금액 심도 GLM모형 결과값 50대	23
<표 4-7> 사후적 요율산정 심도 지급금액 상대도 50대	24
<표 4-8> 사후적 요율산정 지급금액 심도 GLM 50세 상해급수1 예시값	25
<표 4-9> 사후적 요율산정 지급금액 심도 GLM 59세 상해급수 1 예시값	25
<표 4-10> AIC값 비교	26

그림목차

<그림 5-1> 현행요율과 사후적 요율 사고건수 결과값 비교 50세 상해급수 1급 .. 27

<그림 5-2> 현행요율과 사후적 요율 지급금액 결과값 비교 50세 상해급수 1급 .. 28

논문 요약

GLM을 활용한 실손의료보험 실적변수 유의성 검증

실손의료보험은 가입자가 질병, 상해로 입원하거나 통원치료를 받은 경우 실제 부담한 의료비를 보험사가 보상하는 상품이다. 2016년말 기준 계약건수가 약 3300만건으로 많은 국민들이 가입하고 있다. 하지만 성장과 더불어 손해율상승과 보험료 인상이 지속되고 있는 실정이고 여전히 단일보험료를 부과하는 사전적 요율산정을 유지하고 있다. 이에 대하여 본 논문에서는 GLM을 사용하여 보험료차등화를 위한 빈도실적변수 도입의 유의성을 검증했다. 현행요율산정에서 쓰이는 연령, 성별, 담보, 상해급수의 유의성을 검증했고 현행요율산정모형에 실적변수를 더하여 사후적 요율산정의 유의성을 실증분석 하였다. GLM모형을 통해서 빈도 실적변수가 통계적으로 유의하다는 결과를 얻었다.

주제어 : 일반화 선형모형 , 실손의료보험 , 보험료차등화 ,
실적변수 , 요율산정

제1장 서론

제1절. 연구목적

실손의료보험은 가입자가 질병, 상해로 입원하거나 통원치료를 받은 경우 실제 부담한 의료비를 보험사가 보상하는 상품이다. 2016년말 기준 계약건수가 약 3300만건으로 많은 국민들이 가입하고 있다. 하지만 성장과 더불어 손해율상승과 보험료 인상이 지속되고 있는 실정이다. 이경아·이항석(2016)은 손해율 상승의 주요 원인은 정보비대칭으로 인한 역선택 유인이 높은 실손의료보험의 상품적 특성에 기인한다고 보고 있다. 이로 인하여 <표 1-1>과 같이 보험사들의 손해율이 증가하고 보험료가 인상된다. 실손의료보험시장에서 저위험군은 나가고 고위험군만 남는 악순환이 반복되어 시장실패의 위험도 있다(Einav·Finkelstein, 2011). 하지만 여전히 단일보험료를 부과하는 요율산정을 유지하고 있는 실정이다. 이에 대하여 본 논문에서는 GLM을 사용하여 보험료차등화를 위한 빈도실적변수 도입의 유의성을 검증했다.

<표 1-1> 상위 3개사 보험료 인상률, 손해율 합계 평균

(단위: 년, %)

구분	직전 3년 간 보험료 인상률 (직전년도 대비)			직전 3년간 손해율		
	2015	2016	2017	2014	2015	2016
합계	18.23	24.9	25.5	127.83	130.1	129.1

자료: 손해보험협회

제2절. 선행연구

본고의 선행연구로는 보험시장의 역선택과 시장실패에 대한 경제학적인 논문, 방법론적인 측면에서는 glm을 활용한 논문과 보험가입자의 과거 지급빈도실적을 바탕으로 요율산정을 한 논문과 해외보험사 예시가 있다.

(Denuit et.al, 2008)은 경험요율에 대해서 성별, 연령등과 같은 정보만을 고려한 사전적 요율산정의 불충분함은 미래의 보험료를 계산할 때 과거 지급빈도를 사용하여 수정될 수 있다고 한다.

(이항석·이수빈·백혜연, 2016)은 베이지언 신뢰도기법을 통하여 계약자의 개별 과거경험을 토대로 보험료를 산정하고 계약자별 위험특성을 반영하는 보험료 차등화 방법에 대해 논의한다. 해외 건강보험의 사례를 살펴보면 과거 빈도실적을 바탕으로 보험 미청구 할인제도 (No Claims Discount, NCD)를 적용하고 있다. 홍콩 AIA사는 보험금 미청구 기간에 따라 보험료의 환급률을 달리 책정한다. 말레이시아 AXA사의 건강보험상품에서는 연속 미청구 보험연도 횟수에 따라 상이한 보험료 할인율을 적용한다.

<표 1-2> 보험금 미청구 기간에 따른 환급률 (홍콩 AIA)

보험금 미청구 기간	환급률
3년	5%
4년	10%
5년 이상	15%

<표 1-3> 연속 미청구 보험연도 횟수에 따른 할인율 (말레이시아 AXA)

연속 미청구 보험연도 횟수 (Number of consecutive policy yeasers with no claim)	보험료 할인율(%)
2년 이하	0
2년	8
3년	10
4년	15
5년 이상	20

제3절. 논문의 구성

3장에서는 과거실적을 미반영한 GLM을 분석한다. 기존의 요율변수인 성별, 연령별, 담보별, 상해급수별 설명변수들의 유의성을 확인한다. 4장에서는 3장의 설명변수에 추가로 계약자의 과거실적을 반영한 설명변수들의 유의성을 확인한다. 각3, 4장의 1절은 반응변수를 기대빈도(사고건수)로 하여 2절은 기대심도(지급금액)로 하여 진행한다.

제2장 일반화선형모형 이론

제1절. 일반화선형모형 방법론

정규 회귀모형에서 반응변수와 설명변수간 다음의 관계를 가정한다.

$$y_i = x_i' \beta + \epsilon_i \quad (2.1)$$

여기서 x_i' 는 설명변수로 이루어진 행 벡터이고 β 는 설명변수들의 계수로 이루어진 벡터로 최소제곱법으로 추정된다. ϵ_i 는 오차항으로 상호 독립이고, 평균이 0, 분산이 상수 σ^2 인 정규분포를 따르는 정규성을 가정한다. 따라서 전통적 선형모형은 i 번째 관측치의 설명변수 행 벡터 x_i' 가 주어졌을 때 반응변수의 조건부기댓값을 다음과 같이 나타낼 수 있다.

$$E[y_i | x_i] = \mu_i = x_i' \beta \quad (2.2)$$

나아가 오차항 ϵ 에 대한 분포 가정은 반응변수의 분포를 결정한다. 즉, 정규회귀 모형에서는 설명변수가 주어졌을 때 반응변수 y 의 분포는 정규분포를 따른다는 것이다.

$$y_i | x_i \sim N(x_i' \beta, \sigma^2) \quad (2.3)$$

이러한 전통적 선형모형은 다양한 분야의 통계분석에서 사용되어 왔지만 정규분

포에 대한 가정과 등분산 (homoscedasticity) 가정은 전통적 선형모형이 적용될 수 있는 경우를 제한하게 된다. 한 예로 빈도분석에서 반응변수가 어떤 사건의 발생건수를 나타내는 이산형 변수에 이용하기에는 어려움이 있다. 일반화 선형모형은 전통적 선형모형을 확장시킨 것으로 비정규성 데이터 분석에 적합하다. 일반화 선형모형은 크게 세 가지 요소로 구성되어 있는데 첫째, 전통적 선형모형에서와 마찬가지로

$$\eta_i = x_i' \beta \quad (2.4)$$

와 같이 선형관계로 표현되는 부분 (systematic component)이다. 둘째, 연결함수 (link function)라 불리는 미분가능한 단조함수 g 는 반응변수의 평균 μ_i 가 설명변수들과 어떠한 관계를 갖는지를 결정한다. 즉,

$$g(\mu_i) = x_i' \beta \quad (2.5)$$

으로 표현된다. 전통적 선형모형은 $g(x) = x$ 인 항등함수로 연결되는 일반화 선형모형의 한 특수한 형태로 볼 수 있다. 연결함수는 반응변수의 분포에 따라 로그함수, 역수함수 등 다양하게 나타난다. 예를 들어서 반응변수가 포아송분포인 경우 연결함수는 로그함수이다. 일반화 선형모형의 마지막 요소는 확률적 요소 (random component)로 반응변수는 지수족 (exponential family)의 분포에 속한다는 것이다. 반응분포가 지수족에 속하면 대개 반응변수가 이분산 (heteroscedastic)인 경우이다. 그래서 분산은 설명변수들로 표현되는 평균과 함께 변하며, 이것은 일반적인 회귀분석의 등분산 (homoscedastic) 가정에 위배되는 것이다. 지수족 분포는 확률밀도함수가

$$f(y) = c(y, \phi) \exp\left\{\frac{y\theta - a(\theta)}{\phi}\right\} \quad (2.6)$$

의 형태를 갖는 분포를 일컫는데 여기서 모수 θ 와 ϕ 는 각각 정준모수 (canonical parameter)와 산포모수(dispersion parameter)로 불리우는 상수이다. $a(\theta)$ 와 $c(y, \theta)$ 의 형태에 따라 반응변수는 정규분포 이외에도 포아송분포, 음이항분포, 감마분포, 이항분포, 역가우시안분포 등 다양한 분포로 적합할 수 있다. 또한 이 경우 반응변수의 분산은

$$Var(y_i) = \frac{\phi V(\mu_i)}{w_i} \quad \left(\text{단, } V(\mu_i) = \frac{\partial^2 a(\theta)}{\partial \theta^2} \right) \quad (2.7)$$

로 표현되는데 여기서 알 수 있는 것은 반응변수의 분산이 전통적 선형모형의 등분산성 가정과 달리 지수족 분포의 분산함수 V 와 반응변수의 기대값 μ_i 에 의해 결정된다는 점이다.

일반화 선형모형에서는 최대우도법 (maximum likelihood estimation: MLE)을 통해 다음의 로그우도함수를 최대화하는 모수 β 와 ϕ 를 추정한다.

$$l(\beta, \phi) = \sum_{i=1}^n \ln f(y_i; \beta, \phi) = \sum_{i=1}^n \left\{ \ln c(y_i, \phi) + \frac{y_i \theta_i - a(\theta_i)}{\phi} \right\} \quad (2.8)$$

항등함수를 연결함수로 갖는 정규분포와 같은 경우를 제외하고는 보통 식 (2.8)의 최대화를 위한 제1계조건 (first-order condition)을 직접 풀기는 매우 어렵다. 따라서 SAS 혹은 R과 같은 통계 패키지에서는 뉴튼-랩슨 방법 (Newton-Raphson method)을 사용하여 식 (2.8)의 테일러 급수의 2차 근사식이 최대값을 갖도록 모수

β 를 변경해 가는 과정을 반복하게 되는데 이때 모수 β 가 수렴하는 값이 MLE의 추정치가 되는 것이다.

이탈도 (deviance)와 잔차 (residual)는 일반화 선형모형에서 추가적으로 알아둬야 할 사항이다. 이탈도는 모델에 어떤 변수를 포함시킬지의 여부와 모델간 우위를 판정하기 위해 사용되는 통계량이다. 두 비교모델의 이탈도의 차를 모수 ϕ 로 나눈 값이 카이제곱 분포를 따른다는 것을 이용하여 두 비교모델에 차이가 있는지의 여부를 통계적으로 확인할 수 있는 것이다. 잔차는 선택된 통계모형이 데이터에 적합한지의 여부와 이상치 (outlier)의 존재여부를 통계적으로 판별하는데 이용되는 통계량으로 전통적 선형회귀 모델의 잔차와 동일한 개념이라 할 수 있다.

제3장 현행요율체계에 대한 GLM 분석

제1절. 빈도

1) 데이터

3장과 4장의 쓰일 데이터를 본 소절에서 설명한다. 실손의료보험의 계약 및 사고 데이터를 사용하였으며, 2015년 이전에 판매된 계약들 중 2016년까지 유지된 계약만을 포함하였다. 2016년 연령 50세에 해당하는 약 10만 건의 계약데이터 중 2015년과 2016년의 보상데이터와 계약자번호가 일치하는 데이터를 사용하였다. 설명변수로 연령, 성별, 상해급수, 담보분류1: 상해, 질병, 종합, 담보분류2: 외래, 입원, 처방과 계약자별 지급건수를 계약자실적변수를 사용하였다. 3장 현행 요율변수에 대한 검증에서는 계약자실적변수를 제외하였다.

2) 계수추정 결과 값

아래의 식의 β 계수들을 추정한다. 연령변수 같은 경우는 있는 그대로의 숫자가 들어가는 연속형 변수이고, 이를 제외한 나머지 변수들은 표시변수로 x 자리에 0 또는 1을 넣는다. 또한 각 변수들 마다의 기준변수가 존재한다. 기준변수는 성별에서는 남성, 상해급수는 1급, 분류1에서는 상해 그리고 분류2에서는 외래로 정한다.

$$\ln(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_8 x_8 \quad (3.1)$$

위의 식의 β 값들을 추정한 값이 <표3-1>의 Estimate 열이다. 연결함수는 log를 사용하며 0건을 포함한 사고건수 빈도 데이터이기 때문에 Poisson 분포를 사용한다. Std. Error는 표준오차를 나타낸다. 세 번째, 네 번째 열을 보면 계수추정치에 대한 검정결과를 나타내고 있다. 여기에서 귀무가설은 나머지 예측값들이 모형에 존재하고 해당 계수는 0이라는 것이다. 유의확률이 통상적으로 유의수준이 0.05보다 작으면 귀무가설이 기각될 수 있고 모수 추정값이 유의수준에서 통계적으로 유의하다고 고려된다. z-value는 모형에서 이 모수 추정값이 필요한지 여부 판단을 위해 유의성 검정을 제공하는 검정통계량이다. z-value 절대값이 클수록 유의하다고 판단한다. 표에 있는 모든 계수들은 0.001 유의수준 하에서 유의하다.

<표 3-1> 사전적 요율산정 사고건수 빈도 GLM 결과값 50대

반응변수	사고건수				
반응분포	Poisson				
연결함수	log				
구분	Estimate	Std. Error	z value	Pr(> z)	유의수준
(Intercept)	-5.42321	0.046916	-115.59	<2e-16	***
성별 여	0.508111	0.005416	93.82	<2e-16	***
연령	0.031153	0.000845	36.88	<2e-16	***
상해급수2급	-0.08277	0.005866	-14.11	<2e-16	***
상해급수3급	-0.09984	0.008124	-12.29	<2e-16	***
분류1종합	2.315234	0.009557	242.26	<2e-16	***
분류1질병	2.16832	0.008368	259.13	<2e-16	***
분류2입원	-1.43728	0.006853	-209.74	<2e-16	***
분류2처방	-1.3266	0.006344	-209.12	<2e-16	***

3) 상대도

상대도는 각 변수들이 반응변수에 상대적으로 어떻게 반응하는지에 대한 지표이다. 상대도는 (3.1) 식의 양변에 exponential을 취해서 구한다. 50대의 빈도 GLM의 상대도를 해석하면 남자보다 여자가 약 1.66배정도 사고가 날 것으로 추정된다. 연령변수의 상대도를 살펴보면 연령이 1세 늘어나면 사고건수가 1.03배 높게 나올 것을 의미한다. 빈도모형의 상대도는 분류1에서 상해와 종합, 질병 사이에 극명한 차이를 보인다. 종합과 질병에서 더 많은 사고건수가 일어난다고 해석가능하다.

<표 3-2> 사전적 요율산정 사고건수 상대도 50대

설명변수		상대도	설명변수		상대도
성별	남	1	연령	+1세	1.03164
	여	1.66215			-
분류1	상해	1	상해급수	1급	1
	질병	8.74358			
	종합	10.12729		2급	0.92056
분류2	외래	1			
	입원	0.23757		3급	0.90499
	처방	0.26537			

4) 예시값

<표 3-3>는 사전적 요율산정 빈도 예시 값이다. 성별, 상해급수 그리고 담보에 따른 그룹들의 기대 사고건수이다. <표3-4>이 <표3-3>와 다른 점은 59세연령을 사용한 것이다. 연령이 높아짐에 따라 기대 사고건수가 증가한다. 분류1과 분류2의 계수들을 고려하여 외래일 때 높은 기대 사고건수를 나타낸다.

<표 3-3> 사전적 요율산정 사고건수 빈도 50세 예시

(단위: 건)

구 분		성별	남자			여자		
		상해급수	1	2	3	1	2	3
담보명	상해	상해입원	0.0050	0.0046	0.0045	0.0083	0.0076	0.0075
		상해외래	0.0210	0.0193	0.0190	0.0348	0.0321	0.0315
		상해처방	0.0056	0.0051	0.0050	0.0092	0.0085	0.0084
	질병	질병입원	0.0435	0.0401	0.0394	0.0723	0.0666	0.0655
		질병외래	0.1832	0.1686	0.1658	0.3045	0.2803	0.2756
		질병처방	0.0486	0.0448	0.0440	0.0808	0.0744	0.0731
	종합	종합입원	0.0504	0.0464	0.0456	0.0838	0.0771	0.0758
		종합외래	0.2122	0.1953	0.1920	0.3527	0.3247	0.3192
		종합처방	0.0563	0.0518	0.0510	0.0936	0.0862	0.0847

<표 3-4> 사전적 요율산정 사고건수 빈도 59세 예시

(단위: 건)

구 분		성별	남자			여자		
		상해급수	1	2	3	1	2	3
담보명	상해	상해입원	0.0066	0.0061	0.0060	0.0110	0.0101	0.0099
		상해외래	0.0277	0.0255	0.0251	0.0461	0.0424	0.0417
		상해처방	0.0074	0.0068	0.0067	0.0122	0.0113	0.0111
	질병	질병입원	0.0576	0.0530	0.0521	0.0957	0.0881	0.0866
		질병외래	0.2425	0.2232	0.2194	0.4030	0.3710	0.3647
		질병처방	0.0643	0.0592	0.0582	0.1070	0.0985	0.0968
	종합	종합입원	0.0667	0.0614	0.0604	0.1109	0.1021	0.1004
		종합외래	0.2808	0.2585	0.2542	0.4668	0.4297	0.4224
		종합처방	0.0745	0.0686	0.0674	0.1239	0.1140	0.1121

제2절. 심도

1) 계수추정 결과값

심도 glm모형은 (3.2)와 같다. $\ln(\text{가입금액})$ 이 오프셋항으로 삽입되어 있다. 오프셋 항이 존재한다는 것이 빈도 GLM모형과 다르다. 가입금액은 입원, 외래, 처방에 대하여 상이하고 이에 따른 지급금액의 범위가 달라질 수 있기 때문에 오프셋이 존재하는 모형이 없는 모형보다 현실적인 값을 만들어낸다.

$$\ln(\mu) = \ln(\text{가입금액}) + \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_8 x_8 \quad (3.2)$$

다음 표의 Estimate은 β 계수들의 추정값이다. Std. Error는 표준오차를 나타낸다. 세 번째, 네 번째 열을 보면 계수추정치에 대한 검정결과를 나타내고 있다. 여기에서 귀무가설은 나머지 예측값들이 모형에 존재하고 해당 계수는 0이라는 것이다. 유의확률이 통상적으로 유의수준이 0.05보다 작으면 귀무가설이 기각될 수 있고 모수 추정값이 유의수준에서 통계적으로 유의하다고 고려된다. t-value는 모형에서 이 모수 추정값이 필요한지 여부 판단을 위해 유의성 검정을 제공하는 검정통계량이다. t-value 절대값이 클수록 유의하다고 판단한다. 연결함수는 log로 양변에 exponential을 취하면 기대지급금액이 상대도의 곱으로 표현할 수 있다.

<표 3-5> 사전적 효율산정 지급금액 심도 결과값 50대

반응변수	지급금액				
반응분포	Gamma				
오프셋항	가입금액				
연결함수	log				
구분	Estimate	Std. Error	t value	Pr(> t)	유의수준
(Intercept)	-0.717693	0.083132	-8.633	< 2e-16	***
성별 여	-0.069294	0.009617	-7.205	5.81E-13	***
연령	0.003233	0.001499	2.157	0.03102	*
상해급수2급	-0.027074	0.01035	-2.616	0.0089	**
상해급수3급	-0.047278	0.014411	-3.281	0.00104	**
분류1종합	0.246817	0.016794	14.697	< 2e-16	***
분류1질병	0.260583	0.014975	17.402	< 2e-16	***
분류2입원	-3.279204	0.01213	-270.331	< 2e-16	***
분류2처방	0.033005	0.011273	2.928	0.00341	**

2) 상대도

분류2는 입원, 외래, 처방으로 나뉘지는데 분류2항목은 얻은 결과값을 그대로 쓴 것이다. 오프셋반영 분류2는 해당가입금액 대비 기준인 외래 가입금액을 곱해준 것이다. 예를 들면 오프셋반영 분류2의 입원은 $0.03766 * \left(\frac{50000000}{250000} \right) = 7.53164$ 이다. 오프셋 반영 분류2가 실질적인 지급금액의 상대도를 보여주고 있다.

<표 3-6> 사전적 요율산정 심도 상대도 50대

설명변수		상대도	설명변수		상대도
성별	남	1	연령	+1세	1.00328
	여	0.93305			-
분류1	상해	1	상해급수	1급	1
	질병	1.29769		2급	0.97329
	종합	1.27995		3급	0.95382
분류2	외래	1	오프셋반 영 분류2	외래	1
	입원	0.03766		입원	7.53164
	처방	1.03356		처방	0.20671

3) 예시값

아래 예시의 오프셋항으로는 입원, 외래, 처방 가입금액의 최빈값인 5000만원, 25만원, 5만원을 사용하여 계산한다. 예를 들어서 남자, 상해급수1급인 계약자의 상해입원과 상해외래의 차이를 살펴보면 다음 표와 같다. 절편과 연령의 효과는 같지만 담보에 따라 상대도와 그의 가입금액이 다르다.

<표 3-7> 사전적 요율산정 지급금액 심도 예시 값 50세

(단위: 원)

구 분		성별	남자			여자		
		상해급수	1	2	3	1	2	3
담 보 명	상해	상해입원	1,079,800	1,050,958	1,029,937	1,007,510	980,599	960,985
		상해외래	143,368	139,539	136,748	133,770	130,197	127,593
		상해처방	29,636	28,844	28,267	27,652	26,913	26,375
	질병	질병입원	1,401,242	1,363,814	1,336,536	1,307,432	1,272,510	1,247,058
		질병외래	186,047	181,078	177,456	173,592	168,955	165,576
		질병처방	38,458	37,431	36,682	35,883	34,925	34,226
	종합	종합입원	1,382,085	1,345,168	1,318,263	1,289,557	1,255,112	1,230,008
		종합외래	183,504	178,602	175,030	171,219	166,645	163,312
		종합처방	37,932	36,919	36,181	35,393	34,447	33,758

<표 3-8> 사전적 요율산정 지급금액 심도 예시 값 59세

(단위: 원)

구 분		성별	남자			여자		
		상해급수	1	2	3	1	2	3
담 보 명	상해	상해입원	1,111,681	1,081,987	1,060,346	1,037,256	1,009,550	989,358
		상해외래	147,601	143,659	140,785	137,720	134,041	131,360
		상해처방	30,511	29,696	29,102	28,468	27,708	27,154
	질병	질병입원	1,442,613	1,404,080	1,375,996	1,346,033	1,310,080	1,283,876
		질병외래	191,540	186,424	182,695	178,717	173,943	170,464
		질병처방	39,594	38,536	37,765	36,943	35,956	35,237
	종합	종합입원	1,422,890	1,384,883	1,357,184	1,327,631	1,292,169	1,266,324
		종합외래	188,922	183,875	180,198	176,274	171,565	168,134
		종합처방	39,052	38,009	37,249	36,438	35,464	34,755

제4장 GLM을 활용한 사후적 요율 산정

제1절. 빈도

1) 계수 추정 결과값

사후적 요율산정에서는 설명변수에 2015년도 사고건수가 추가된다. 해당 고객 번호의 2015년의 사고건수를 과거실적으로 사용한다. 실적년도 2015년도와 2016년도 계약데이터에 모두 해당하는 고객번호를 데이터로 사용한다. 사고연도가 2015인 것은 2015년도의 사고건수로 인식하고 사고연도가 2016인 것은 2016년도의 사고건수로 인식한다.

$$\ln(\mu_{2016}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{14} x_{14} \quad (4.1)$$

위의 식에서 β_0 부터 β_8 까지의 계수는 앞서 했던 사전적 요율산정의 계수와 같고 β_9 부터 β_{14} 계수는 과거실적 설명변수이다. 과거실적이 0건인 데이터가 기준이 된다.

<표 4-1> 과거실적 표시 변수 예시

분류	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
0건	0	0	0	0	0	0
1건	1	0	0	0	0	0
2건	0	1	0	0	0	0
3건	0	0	1	0	0	0
4건	0	0	0	1	0	0
5건	0	0	0	0	1	0
6건 이상	0	0	0	0	0	1

다음 표의 Estimate은 β 계수들의 추정값이다. Std. Error는 표준오차를 나타낸다. 세 번째, 네 번째 열을 보면 계수추정치에 대한 검정결과를 나타내고 있다. 여기에서 귀무가설은 나머지 예측값들이 모형에 존재하고 해당 계수는 0이라는 것이다. 유의확률이 통상적으로 유의수준이 0.05보다 작으면 귀무가설이 기각될 수 있고 모수 추정값이 유의수준에서 통계적으로 유의하다고 고려된다. z-value는 모형에서 이 모수 추정값이 필요한지 여부 판단을 위해 유의성 검정을 제공하는 검정통계량이다. z-value 절대값이 클수록 유의하다고 판단한다. 표에 있는 모든 계수들은 0.001 유의수준 하에서 유의하다. 과거실적의 계수를 보면 과거 사고건수가 많아질수록 2016년도 사고건수가 많아질 것이라고 모형화하고 있다. 연결함수는 log로 양변에 exponential을 취하면 기대지급금액이 상대도의 곱으로 표현할 수 있다.

<표 4-2> 사후적 요율산정 사고건수 빈도 GLM모형 결과값 50대

반응변수	사고건수				
반응분포	Poisson				
연결함수	log				
구분	Estimate	Std. Error	z value	Pr(> z)	유의수준
(Intercept)	-4.851819	0.0470652	-103.087	< 2e-16	***
성별 여	0.2748577	0.0054992	49.981	< 2e-16	***
연령	0.0175537	0.0008489	20.679	< 2e-16	***
상해급수2급	-0.053592	0.0058659	-9.136	< 2e-16	***
상해급수3급	-0.059511	0.0081326	-7.318	2.52E-13	***
분류1종합	1.6754794	0.0098034	170.907	< 2e-16	***
분류1질병	1.6532724	0.0086219	191.754	< 2e-16	***
분류2입원	-0.886466	0.0071223	-124.463	< 2e-16	***
분류2처방	-0.87755	0.0065536	-133.904	< 2e-16	***
15년도 사고건수 1건	1.7649856	0.0059967	294.328	< 2e-16	***
15년도 사고건수 2건	2.2148593	0.0078167	283.348	< 2e-16	***
15년도 사고건수 3건	2.4943763	0.0102885	242.443	< 2e-16	***
15년도 사고건수 4건	2.7215373	0.0139541	195.035	< 2e-16	***
15년도 사고건수 5건	2.8655317	0.0203127	141.071	< 2e-16	***
15년도 사고건수 6건 이상	3.1797444	0.0212813	149.415	< 2e-16	***

2) 상대도

기준수준에 대한 상대적인 비를 나타내고 있다. 사전적 요율산정의 상대도와 비교해보면 상해급수, 분류, 연령, 성별의 상대도가 감소하고 그 대신 과거실적에 대한 설명변수의 상대도가 과거실적에 비례하여 증가하는 추세이다.

<표 4-3> 사후적 효율산정 사고건수 빈도 GLM모형 결과값 50대

설명변수		상대도	설명변수		상대도
성별	남	1	연령	+1세	1.01771
	여	1.31634			
분류1	상해	1	상해급수	1급	1
	질병	5.22405		2급	0.94782
	종합	5.34135		3급	0.94223
분류2	외래	1	15년도 사고건수	0건	1
				1건	5.84149
	입원	0.41211		2건	9.16012
				3건	12.11418
				4건	15.20368
	처방	0.41580		5건	17.55839
				6건 이상	24.04061

3) 예시값

2016년에 연령이 50세이고 상해급수가 1급인 계약자들의 성별, 과거실적별 그리고 담보별 그룹 기대 사고건수이다. 과거실적은 0건, 1건, 2건으로 제한하여 표를 작성한다. 아래 표에서 가장 높은 기대 사고건수를 보이는 그룹은 여자이고 2015년 사고건수가 2건, 담보가 종합외래인 그룹이다. 가장 적은 기대 빈도를 보이는 그룹은 남자이고 2015년 사고건수가 0건, 그리고 담보 종류가 상해입원인 그룹이다.

<표 4-4> 사후적 효율산정 빈도 사고건수 GLM모형 50세 상해급수 1급 예시값

(단위: 건)

		성별	남자			여자		
		15년도 사고건수	0	1	2	0	1	2
담보명	상해	상해입원	0.0077	0.0452	0.0710	0.0102	0.0596	0.0934
		상해외래	0.0188	0.1098	0.1722	0.0247	0.1445	0.2266
		상해처방	0.0078	0.0457	0.0716	0.0103	0.0601	0.0942
	질병	질병입원	0.0405	0.2364	0.3707	0.0533	0.3111	0.4879
		질병외래	0.0982	0.5736	0.8994	0.1293	0.7550	1.1839
		질병처방	0.0408	0.2385	0.3740	0.0537	0.3139	0.4923
	종합	종합입원	0.0414	0.2417	0.3790	0.0545	0.3181	0.4989
		종합외래	0.1004	0.5864	0.9196	0.1322	0.7720	1.2105
		종합처방	0.0417	0.2438	0.3824	0.0549	0.3210	0.5033

<표 4-5>는 <표 4-4>에서 연령만 59세로 바뀐 것으로 <표 4-4>의 각 그룹들에
연령 설명변수를 고려하여 구성요소들에 $e^{0.01755 \times 9} = 1.171107$ 배 해준 것이다.

<표 4-5> 사후적 요율산정 빈도 사고건수 GLM모형 59세 상해급수 1급 예시 값

(단위: 건)

구 분		성별	남자			여자		
		15년도 사고건수	0	1	2	0	1	2
담보명	상해	상해입원	0.0091	0.0530	0.0831	0.0119	0.0698	0.1094
		상해외래	0.0220	0.1286	0.2016	0.0290	0.1693	0.2654
		상해처방	0.0092	0.0535	0.0838	0.0120	0.0704	0.1104
	질병	질병입원	0.0474	0.2768	0.4341	0.0624	0.3644	0.5714
		질병외래	0.1150	0.6717	1.0534	0.1514	0.8842	1.3866
		질병처방	0.0478	0.2793	0.4380	0.0629	0.3677	0.5765
	종합	종합입원	0.0485	0.2830	0.4438	0.0638	0.3726	0.5843
		종합외래	0.1176	0.6868	1.0770	0.1548	0.9041	1.4177
		종합처방	0.0489	0.2856	0.4478	0.0644	0.3759	0.5895

제2절. 심도

1) 계수 추정 결과값

사후적 요율산정의 심도에서는 과거실적을 설명변수로 넣어서 glm의 계수들을 추정한다.

$$\ln(\mu) = \ln(\text{가입금액}) + \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{14} x_{14} \quad (4.2)$$

β_9 부터 β_{14} 까지는 과거실적을 나타내는 계수들이다. 오프셋항으로는 가입금액에 로그를 취한 값을 사용한다. 다음 표의 Estimate은 β 계수들의 추정값이다. Std. Error는 표준오차를 나타낸다. 세 번째, 네 번째 열을 보면 계수추정치에 대한 검정

결과를 나타내고 있다. 여기에서 귀무가설은 나머지 예측값들이 모형에 존재하고 해당 계수는 0이라는 것이다. 유의확률이 통상적으로 유의수준이 0.05보다 작으면 귀무가설이 기각될 수 있고 모수 추정값이 유의수준에서 통계적으로 유의하다고 고려된다. t-value값은 추정한 β 값에 0을 뺀 값을 표준오차로 나눈 것이다. t-value는 모형에서 이 모수 추정값이 필요한지 여부 판단을 위해 유의성 검정을 제공하는 검정통계량이다. t-value 절대값이 클수록 유의하다고 판단한다. 사후적요을 심도 glm은 빈도 glm과는 다르게 과거실적과 지급금액이 비례하지 않는다.

<표 4-6> 사후적 요율산정 지급금액 심도 GLM모형 결과값 50대

반응변수	지급금액				
반응분포	Gamma				
오프셋항	가입금액				
연결함수	log				
구분	Estimate	Std. Error	t value	Pr(> t)	유의수준
(Intercept)	-0.73931	0.083274	-8.878	< 2e-16	***
성별여	-0.06065	0.009722	-6.239	4.41E-10	***
연령16	0.003619	0.001502	2.409	0.01598	*
상해급수2급	-0.02864	0.010366	-2.763	0.00574	**
상해급수3급	-0.04834	0.014431	-3.35	0.00081	***
분류1종합	0.262723	0.017233	15.245	< 2e-16	***
분류1질병	0.273388	0.015336	17.826	< 2e-16	***
분류2입원	-3.29543	0.012669	-260.116	< 2e-16	***
분류2처방	0.022631	0.011412	1.983	0.04736	*
15년도 사고건수 1건	0.014518	0.01058	1.372	0.16999	
15년도 사고건수 2건	-0.04268	0.013661	-3.124	1.78E-03	**
15년도 사고건수 3건	-0.0719	0.017978	-3.999	6.35E-05	***
15년도 사고건수 4건	-0.12935	0.02451	-5.277	0.000000131	***
15년도 사고건수 5건	-0.02807	0.035783	-0.784	0.43285	
15년도 사고건수 6건 이상	-0.12445	0.037552	-3.314	9.20E-04	***

연결함수는 log로 양변에 exponential을 취하면 기대지급금액이 상대도의 곱으로 표현할 수 있다.

2) 상대도

기준이 되는 변수는 $e^0 = 1$ 이 된다. 기준변수에 대한 상대적인 값을 상대도를 통하여 나타내고 있다. 예를 들면 e^{β_1} 는 기준이 되는 남자에 대한 여자의 상대도이다.

분류2는 입원, 외래, 처방으로 나뉘지는데 분류2항목은 얻은 결과값을 그대로 쓴 것이다. 오프셋반영 분류2는 해당가입금액 대비 기준인 외래 가입금액을 곱해준 것이다. 예를 들면 오프셋반영 분류2의 입원은 $0.037052 * \left(\frac{50000000}{250000} \right) = 7.410437$ 이다. 오프셋반영 분류2가 실질적인 지급금액의 상대도를 보여주고 있다.

<표 4-7> 사후적 요율산정 심도 지급금액 상대도 50대

설명변수		상대도	설명변수		상대도
성별	남	1	연령	+1세	1.00362556
	여	0.941149			
분류1	상해	1	상해급수	1급	1
	질병	1.31441		2급	0.97176915
	종합	1.300466		3급	0.95281359
분류2	외래	1	15년도 사고건수	0건	1
	입원	0.037052		1건	1.0146239
	처방	1.022889		2건	0.95821414
오프셋반영	외래	1		3건	0.93062023
분류2	입원	7.410437		4건	0.87866814
	처방	0.204578		5건	0.97232516
				6건 이상	0.88298682

3) 예시값

50세의 상해급수 1급인 사람들의 성별, 과거실적별 그리고 담보별 그룹들의 기대 지급금액이다. 심도는 과거실적에 따른 차이가 빈도에 비해 미미하다는 것을 예시 값을 통해 볼 수있다. 모형에서는 여자의 평균 지급금액이 남자에 비해 작게 나타난다. 아래 예시의 오프셋항으로는 입원, 외래, 처방의 가입금액의 최빈값인 5000만원, 25만원, 그리고 5만원을 사용하여 계산한다.

<표 4-8> 사후적 요율산정 지급금액 심도 GLM 50세 상해급수1 예시값

(단위: 원)

구 분		성별	남자			여자		
		15년도 사고건수	0	1	2	0	1	2
담 보 명	상해	상해입원	1,059,963	1,075,464	1,015,672	997,583	1,012,171	955,898
		상해외래	143,037	145,128	137,060	134,619	136,587	128,993
		상해처방	29,262	29,690	28,039	27,540	27,943	26,389
	질병	질병입원	1,393,226	1,413,601	1,335,009	1,311,233	1,330,408	1,256,442
		질병외래	188,009	190,758	180,153	176,944	179,532	169,550
		질병처방	38,462	39,025	36,855	36,199	36,728	34,686
	종합	종합입원	1,378,446	1,398,605	1,320,847	1,297,323	1,316,295	1,243,113
		종합외래	186,014	188,734	178,241	175,067	177,627	167,752
		종합처방	38,054	38,611	36,464	35,815	36,339	34,318

<표 4-9> 사후적 요율산정 지급금액 심도 GLM 59세 상해급수 1 예시값

(단위: 원)

구 분		성별	남자			여자		
		15년도 사고건수	0	1	2	0	1	2
담 보 명	상해	상해입원	1,095,055	1,111,069	1,049,298	1,030,610	1,045,682	987,545
		상해외래	147,772	149,933	141,597	139,075	141,109	133,264
		상해처방	30,231	30,673	28,968	28,452	28,868	27,263
	질병	질병입원	1,439,352	1,460,401	1,379,207	1,354,644	1,374,455	1,298,039
		질병외래	194,233	197,074	186,117	182,802	185,476	175,164
		질병처방	39,736	40,317	38,075	37,397	37,944	35,835
	종합	종합입원	1,424,083	1,444,909	1,364,576	1,340,274	1,359,874	1,284,269
		종합외래	192,173	194,983	184,143	180,863	183,508	173,305
		종합처방	39,314	39,889	37,671	37,001	37,542	35,454

4) 모형 AIC값 비교

AIC값은 다음 수식과 같이 나타낸다.

$$AIC = -2\ln(L) + 2k \quad (4.3)$$

k는 모형에서 추정된 모수의 개수이고 L은 우도함수 (Likelihood Function)값이다. AIC값은 낮을수록 좋은 모형이라고 해석할 수 있다. 추정된 모수의 개수를 패널티로 부여한다.

<표 4-10> AIC값 비교

구분	현행 효율산정 모형	과거실적 반영 효율산정 모형
빈도	1130648	976050
심도	4602872	4602717

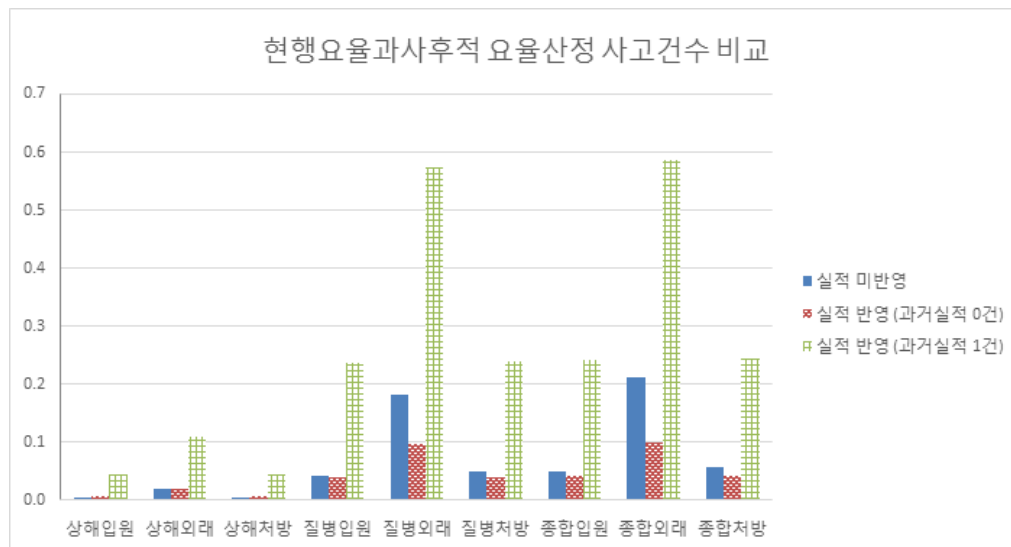
3장의 모형과 4장의 모형을 AIC값을 통하여 비교해보면 과거실적을 반영한 모형의 값이 더 낮았다. 이는 현행효율산정 모형보다 통계적으로 예측하기에 더 우수한 모형이라고 해석할 수 있다.

제5장 결 론

제1절 연구의 결과

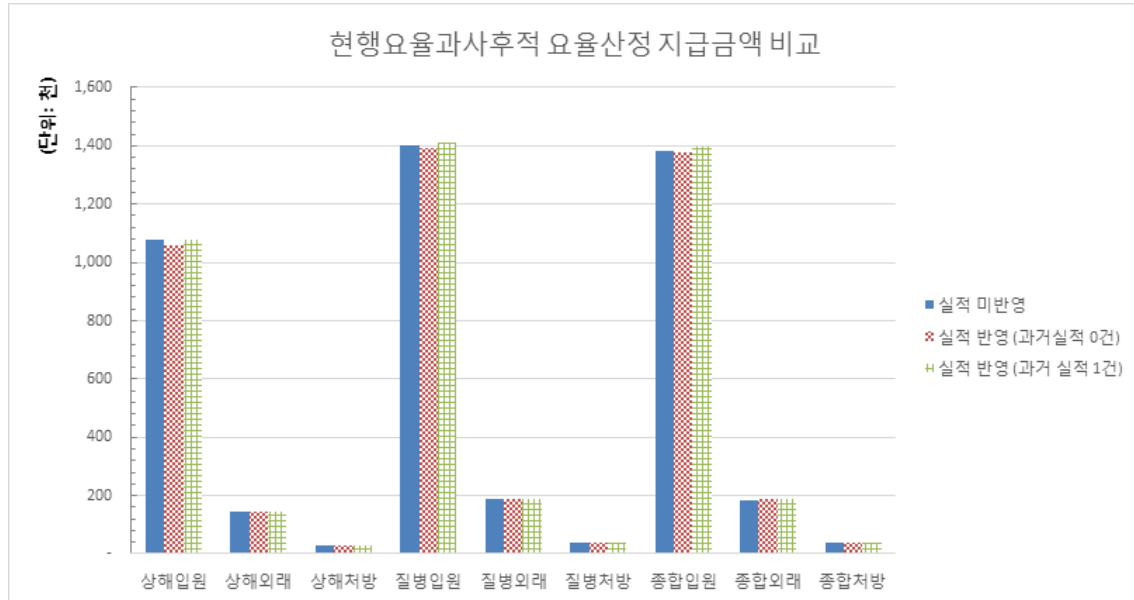
연구의 결과는 <그림 5-1>과 같이 모형들의 결과 값 예시 비교를 통해 보여줬다. 아래의 그림은 현행요율산정을 바탕으로 한 실적미반영 모형과 과거실적을 반영한 사후적 요율산정 모형의 기대사고건수 비교 그래프이다. y축은 기대사고건수이고 x축은 담보를 나타낸다.

<그림 5-1> 현행요율과 사후적 요율산정 사고건수 결과값 비교 50세 상해급수 1급



현행요율산정에 의한 모형의 기대빈도, 사후적 요율산정에서 과거실적이 0건일 때의 기대 빈도, 1건일 때의 기대 빈도를 비교한다. 사전적 요율산정이 과거실적이 0건일 때와 1건일 때의 기대빈도 사이에 위치한다. 심도에서는 과거실적에 상관없이 비슷한 결과를 보여주고 있다.

<그림 5-2> 현행요율과 사후적 요율산정 지급금액 결과값 비교 50세 상해급수 1급



또한 AIC값을 통하여 모형을 비교하였을 때 과거실적을 반영한 모형이 더 우수하였다.

제2절 연구의 의의

본 연구는 실손의료보험시장의 역선택 문제에 따른 보험료 인상률과 손해를 증가를 인식하고 보험료차등화방안으로 실적변수 도입을 주장한다. 역선택의 근본적인 원인인 정보비대칭을 해결하기 위한 방안으로 실적변수로 과거 지급빈도 도입 타당성을 검증한다. 실손의료보험의 실적변수 도입에 대한 실증분석한 것에 의의를 둘 수 있다. 자동차보험에서는 이미 GLM기법을 토대로 운전자의 실적에 따른 보험료차등화를 실시하고 있다. 실손의료보험에서도 요율산정에 있어서 실적변수도입이 필요하다는 것을 GLM분석을 통하여 유의성이 검증되었다.

본 연구는 실적변수도입에 대한 통계적 유의성 검증을 하였다. 향후에는 실적변수를 도입하여 보험료 산출하는 방안을 마련해야 할 것이다. 연구결과를 토대로 하면 2015년 사고건수가 0건인 계약자에 비해 2015년 사고건수가 6건 이상인 계약자의 2016년도 기대빈도는 약 24배이다. 하지만 현실적으로는 보험료를 위 모형에서 산출된 상대도를 바탕으로 하기에는 어려움이 있다고 생각한다.

참고문헌

- De jong, P., Heller, G. Z. (2008). Generalized linear Models for Insurance Data. Cambridge University Press, Cambridge.
- Denuit, M., Marechal, X., Pitrebois, S., Walhin, J. (2007). Actuarial Modelling of Claim Counts. Wiley, Chichester.
- Einav, L., Finkelstein, A. (2011). Selection in Insurance Markets: Teory and Empirics in Pictures. Journal of Economic Perspectives Vol.25 No.1, p.115-138.
- Goldburd, M., Khare, A., Tevet, D. (2016). Generalized Linear Models for Insurance Rating. CAS, Arlington.
- 금융감독원. (2017). 실손의료보험 길라잡이.
- 이경아, 이항석. (2016). 실손의료보험의 역선택과 보험료 차등화. 한국리스크관리학회 Vol.27 No.3, p.93-128.
- 이항석, 이수빈, 백혜연. (2017). 실손의료보험에서 신뢰도기법을 반영한 보험료 산정. 한국보험학회지 Vol.111, p.41-73.
- 주호찬, 이항석. (2013). Estimation of the Expected Loss per Exposure of Export Insurance using GLM. 한국응용통계학회지 Vol.26 No.6, p.857-871.
- 홍종선., 김효중. (2016). SAS를 이용한 회귀모형 자료분석. 서울: 탐진. p.378-406.

ABSTRACT

Verifying statistical significance of private health insurance's past performance variables using GLM

Lee, Ga Eun

Department of Actuarial Science

Sungkyunkwan University

The private health insurance is a product that the insurer compensates for the medical expenses actually incurred when the patient is admitted to the hospital or outpatient due to disease or accident. As of the end of 2016, the number of contracts is about 33 million, and many people are joining. However, with the growth, the loss ratio and the premium increase have been continuing, and we still maintain the a priori ratemaking that imposes a single premium. In this paper, we use the GLM to verify the statistical significance of introducing frequency performance variables for differentiating premiums. We verified the significance of the age, gender, collateral, and class rates used in the current ratemaking, and empirically analyzed the significance of a posteriori calculation by adding the performance variables to the current ratemaking model. As a result, the GLM model shows that the frequency performance variables are statistically significant..