응용통계연구 (2011)
24(2), 279–292
DOI: 10.5351/KJAS.2011.24.2.279

여러가지 신뢰도에 근거한 자동차 보험료 예측

김영화[1] · 김미정[2] · 김명준[3]

[1]중앙대학교 통계학과, [2]중앙대학교 통계학과, [3]삼성화재 자동차보험팀

(2010년 12월 접수, 2011년 2월 채택)

요 약

합리적인 보험료를 책정하기 위해 사용되는 신뢰도 이론은 보험통계학의 중요한 주요 이론 가운데 하나이다. 본 논문에서는 신뢰도 이론의 기본 개념과 함께 유효대수 법칙, 제곱근 법칙, Bühlmann 신뢰도, Bühlmann 신뢰도, Bühlmann-Straub 신뢰도 등을 소개하였다. 또한 이러한 방법들에 근거하여 새로운 보험료를 실제 자료를 사용하여 예측하였다. 결론적으로, 유효대수 법칙이 가장 정확한 예측력을 보였다.

주요용어: 신뢰도, 유효대수, 제곱근 법칙, Bühlmann 신뢰도, Bühlmann-Straub 신뢰도.

1. 서론

현대인에게 자동차가 생활필수품이 되면서 자연스럽게 자동차보험료 또한 가계지출항목의 필수항목이 되었다. 자동차 보험계약은 보험사업자와 보험계약자 사이에 이루어진 계약으로 피보험자의 위험에 관하여, 미리 정한 조건하에서 보험금을 보험수익자에게 지급하고 그 대가로서 일정한 조건 하에 보험계약자는 보험자에게 보험료를 납입하는 것을 그 내용으로 한다. 이 때 보험료란 보험회사에서 판매하는 "보험"이라는 상품의 가격이다. 보험업계에서는 적정한 보험료를 책정하는 것(보험요율 산정)이 매우 중요하다. 그 이유는 보험료가 너무 낮게 산정되면 보험회사가 손해를 보며, 반대로 너무 높게 산정되면 보험회사 또는 전체 보험업계에 많은 이익을 창출해 시장에서 배척당할 수 있기 때문이다. 또한, 보험료의 인상은 전체적인 물가인상에 영향을 미치므로 보험감독기관에서는 보험료와 보험요율 수준에 대해서 특별한 관심을 가지고 있다. 개별주체(가계 또는 기업)가 납입하는 보험료는 대수의 법칙에 의하여 경제적 손실의 대가로서 산출된다. 즉 이 보험료 총액은 지급되어야 할 보험금의 총액과 같아야 한다(수지균형의 원칙). 이때, 고객분류(customer segmentation)와 분류된 그룹에 대한 가격책정(pricing)은 합리적인 보험료를 산출하기 위한 가장 중요한 요소이다. 첫 번째 요소인 고객분류는 위험도에 따라 고객들을 적절한 그룹으로 분류하는 것이다. 같은 그룹에 속한 고객들끼리는 그 특성이 매우 유사하여 그룹내 편차가 적어야 하며(급내분산(within variance) 최소화), 다른 그룹과는 편차가 일정 수준 이상(급간분산(between variance) 최대화)이어야 적절한 그룹화가 이루어졌다고 할 수 있다. 두 번째 요소인 가격책정은 과거의 데이터를 바탕으로 각 그룹의 위험도를 정확하게 예측하여 수치화하는 것을 의미한다. 이러한 위험도를 예측하는데 있어 가장 문제가 되는 것이 특정 그룹 또는 셀의 관측치가 너무 작게 나타나는 경우이다. 위에서 언급했듯이 보험은 대수의 법칙을 토대로 성립하고 있다. 요컨대 우연이라고 생각되는 사건도 대량으로 관찰하면 일정한 규칙이 발견된다고 하는 원리이며, 개개인에 대해서는 전혀 우연인 사고도 대량 관찰의 결과 일정한 확률로 발생한다는 것을 토대로 손해보험이 성립한다. 역으로 말

[1]교신저자: (156-756) 서울시 동작구 흑석동 221, 중앙대학교 통계학과, 부교수. E-mail: gogators@cau.ac.kr

하자면, 같은 위험에 노출된 사람이 대량 관찰될 정도로 다수가 되지 않거나, 동일 위험에 노출된 사람의 수가 다수이더라도 그로부터 위험발생률을 도출할 수가 없으면 합리적인 보험료를 책정하는 것은 어렵다. 따라서 분석하고자 하는 자료를 얼마나 믿을 수 있는가에 대한 신뢰도(credibility)의 개념이 탄생하게 되었으며, 신뢰도는 고객분류와 가격위험도의 판단이라는 두 가지 큰 주제에서 보다 세밀한 분류의 필요성과 그에 대한 적절한 가격정책의 균형을 잡아주는 필수 불가결한 도구로 활용되고 있는 것이다. 본 논문은 고객분류를 하는 경우 중 특이한 케이스에 대한 연구로서, 그룹화된 셀의 관측치가 적은 경우에 어떤 방식으로 해당 셀에 대한 적정한 위험도를 부여하는가를 살펴보고, 현재까지 제시된 여러 가지 방식들의 적용을 통하여 상황별로 적절한 대안이 어떠한 것인지 실증적인 검증을 통하여 그 대안을 제시하고자 한다. 또한 본 논문의 구성은 다음과 같다.

제 2장에서는 보험료 책정에서 중요한 요소로 사용되는 신뢰도의 기본 개념과 신뢰도를 산출하는 여러 가지 이론을 김영화와 이현수 (2010)의 내용을 바탕으로 소개한다. 제 3장은 실증 자료 분석으로서 자동차보험사의 실제 자료에 대하여 2장에서 소개하였던 여러 가지 신뢰도를 적용하여 보험료를 책정하고, 그 결과의 비교를 통하여 어떠한 방법을 적용하는 것이 가장 합리적인지를 제시한다. 제 4장은 결론 부분으로서 본 연구의 결과를 요약하고 추후 연구 주제에 관하여 논의한다.

2. 신뢰도(Credibility)

2.1. 신뢰도의 기본 개념

손해보험의 그 모든 종류에 있어서 위험의 불균질성 요인이 다수 존재하고 있다. 위험이 불균질한 보험계약 집단에서는 위험의 혼성비율이 변화하면 과거의 보험청구 발생률에 의거한 보험요율로는 전혀 부적절하게 되어 버리는 일이 많다. 따라서 이 위험을 경감하기 위하여 관련하는 여러 가지 위험요소의 그룹마다 보험청구 발생률이나 평균지급금액 또는 보험요율이 계산된다. 그렇지만 이와 같이 그룹으로 나누어져 있는 경우라도 그룹에 아직 상당한 정도의 위험의 불균질성이 남겨져 있는 경우가 많다. 많은 계약에서는 참 보험청구 발생률이 그 그룹의 평균 보험청구 발생률의 근처에 있을 것이지만 그 중에는 꽤 높은 것이나 지나치게 낮은 것도 있을 것이다. 특히, 특정한 그룹 또는 셀의 관측치가 적은 경우에는 이러한 문제가 더욱 심각해진다. 따라서, 동일한 보험료 산출방식을 모든 그룹에 그대로 적용하는 것은 합리적이지 못하며, 과거와 당해년도의 데이터에 근거하여 그 다음 해의 보험료 산출 방식을 수정 보완하는 것이 타당한 접근방법이라 할 수 있다. 즉, 현재까지 수집된 데이터를 얼마나 믿을 수 있는가에 대한 신뢰도(credibility) 개념이 보험료 책정에 중요한 요소로 사용되고 있다. 보험학에서의 신뢰도의 의미는 신뢰(credence)의 척도로 소개되었으며, 신뢰도가 높을수록 현재의 데이터를 믿는 정도가 높아지는 것을 말한다. 신뢰도의 개념은 1914년에 설립된 CAS(미국 손해보험협회)에서 초기 보험료에 현재까지 수집된 데이터를 어떻게 적용할 것인가 하는 문제의 해결책으로 제시되어 보험회사의 가격정책에 적용되기 시작하였다. 신뢰도의 범위는 0과 1사이이며, 데이터가 너무 적어서 보험요율을 정하는데 전혀 사용할 수 없으면 신뢰도는 ‘0’이고, 데이터가 충분하여 완벽한 신뢰도이면 신뢰도는 ‘1’이라도 정하는 것이 일반적인 방법이다. Mowbray (1914)는 현재의 데이터를 믿을 수 있기 전에 얼마나 많은 시행의 결과가 필요한가를 연구하였고, Whitney (1918)는 새로운 추정치를 얻기 위하여 현재의 추정치와 새로운 데이터를 어떻게 조합하여 사용할 것인가를 연구하여 다음과 같은 기본 개념을 정립하였다.

$$\text{New Rate} = \text{Credibility} \times \text{Observed Data} + (1 - \text{Credibility}) \times \text{Old Rate}.$$

위 식에서 새로운 보험요율 New Rate는 현재 데이터의 정보와 과거의 데이터로부터 구하여 현재 적용되고 있는 보험요율 Old Rate의 가중 평균의 형태임을 알 수 있다. 또한 위 식은 현재 관측치의 평균

과 과거 데이터에 근거하여 얻어진 보험계리인의 추정치인 사전평균(prior mean)이 가중 평균되어 있는 형태로 볼 수 있으며, 이러한 관점에서 베이지안 패러다임과 동일하다고 할 수 있다. 즉, Whitney (1918)의 기본 개념은 다음과 같이 표현할 수 있으며 현재 국내에서 사용되는 가장 고전적인 방법으로서 보험업법 감독규정에 정의되어있다.

$$\text{New Estimate} = Z \times \text{Observed Data} + (1 - Z) \times \text{Prior Mean},$$

여기서 Z는 현재 데이터의 양과 질에 의하여 결정되는 신뢰도($0 \leq Z \leq 1$)를 나타내며, Z는 관측된 데이터에 할당된 신뢰도, $(1 - Z)$는 신뢰도의 보완이라는 의미를 갖는다. Bailey (1950)는 우도함수가 이항분포이고 사전분포가 베타분포인 경우에, 새로운 요율 즉, 사후평균이 이러한 신뢰도의 가중평균이 되는 것을 보였다.

신뢰도에 관하여 주로 연구되는 방법은 변동(fluctuation)을 제한하는 방법과 정확도(accuracy)를 최대한 높이는 방법으로 구분할 수 있다. 변동을 제한하는 방법은 데이터에 포함되어 있는 랜덤 변동의 영향을 제한하는 고전적인 방법이며, 정확도를 최대한 높이는 방법은 추정오차를 최소화하는 방법이다. 이러한 방법에 근거하여 고객의 경력에 따라 신뢰도를 부여하는 여러가지 절차가 Bühlmann (1967, 1969), Bühlmann과 Straub (1970), Hachemeister (1975), Frees (2003) 등에 의하여 개발되었다. 본 논문에서는 변동을 제한하는 방법에 따라 신뢰도를 구하는 방법인 유효대수 법칙(the rule of relative exposure volume)과 제곱근 법칙(the square root rule), 정확도를 최대한 높이는 방법인 Bühlmann 신뢰도와 Bühlmann-Straub 신뢰도를 소개하고 실제 자료의 분석을 통하여 비교하고자 한다.

2.2. 완전신뢰도에 필요한 유효대수

완전 신뢰도는 $Z = 1$인 경우를 의미하며, 유효대수(the relative exposure volume)는 경과일수에 따른 가입대수를 의미한다. 예를 들어, 보험자료분석기간 전체에 가입되어 있는 차량은 유효대수 $= 1$이 되고, 보험자료분석기간 절반에 가입되어 있는 차량은 유효대수 $= 0.5$가 된다.

앞서 소개하였던 국내 보험업법 감독규정에서 정의되어있는 내용을 수식화하여 전개하면 다음과 같다.

$$\begin{aligned} E_2 &= ZT + (1 - Z)E_1 \\ &= ZT + Z(E(T) - ZE(T) + (1 - Z)E_1 \\ &= (1 - Z)E_1 + Z(E(T)Z[T - E(T)], \end{aligned}$$

여기서 E_2는 '새로운 추정치(new estimate)', E_1은 '기존의 추정치(previous estimate)', Z는 '신뢰도(credibility)', T는 '현재의 데이터'를 의미한다. 또한 $(1 - Z)E_1$은 안정성(stability), $ZE(T)$는 진실성(truth; 일치성), $Z[T - E(T)]$는 랜덤오차(random error)로 정의된다.

주어진 오차허용한계(error tolerance level)를 k, 신뢰수준(confidence level)을 p, 표준정규분포의 상위 p-분위수를 z_p라 하면 $P\{Z[T - E(T)] < kE(T)\} = p$ 또는 $P\{T - E(T) < kE(T)/Z\} = p$로부터 T가 정규분포를 따르는 것을 가정하면 $kE(T)/Z = z_p\sqrt{\text{Var}(T)}$이다. 따라서 $Z = kE(T)/\{z_p\sqrt{\text{Var}(T)}\}$를 얻게 된다. 여기서 T의 평균과 분산은 데이터를 통해 얻어진 값을 사용한다.

보험료를 결정하는 요소는 심도(severity)와 빈도(frequency)이며, 보험료는 심도와 빈도의 곱으로 설명된다. 여기서 심도는 사고로 인하여 한 사고 당 발생하는 손실의 정도를 나타내며, 빈도는 일정 기간 동안 발생하는 사고건수를 의미한다. 즉, $T = F \times S$로 표현할 수 있으며, 심도를 상수로 취급하여 고려하지 않는 경우 $T \propto F$가 되며, 본 연구에서는 $T = F$를 가정한다.

표 **2.1.** 완전신뢰도에 필요한 유효대수

		k			
		2.5%	5%	7.5%	10%
	90%	4,329	1,082	481	271
p	95%	6,146	1,537	683	384
	99%	10,616	2,654	1,180	664

2.2.1. 심도를 고려하지 않는 경우 심도(severity)를 상수취급하여 고려하지 않는 경우, 일반적으로 $T = F$라 가정하므로 $E(T) = E(F)$이고, 빈도(frequency) F가 평균 보험료 청구건수(number of claims)가 N인 포아송 분포를 따른다고 가정하면 포아송 분포는 평균과 분산이 같은 특성이 있으므로 $E(T) = \mathrm{Var}(T) = N$이다. 따라서

$$Z = \frac{kE(T)}{z_p\sqrt{\mathrm{Var}(T)}} = \frac{kE(T)}{z_p\sqrt{E(T)}} = \frac{k\sqrt{E(T)}}{z_p} = \frac{k\sqrt{N}}{z_p}$$

이다. 따라서 $Z = 1$ 즉, 신뢰도가 1인 완전신뢰도(full credibility)일 때의 유효대수를 N_{full}이라 하면 위 식으로부터 다음을 얻게 된다.

$$N_{full} = \left(\frac{z_p}{k}\right)^2. \tag{2.1}$$

표 2.1은 완전신뢰도인 경우의 오차허용한계 k와 신뢰수준 p에 따른 N_{full}의 값을 정리해 놓은 것이다.

2.2.2. 심도를 고려하는 경우 심도(severity) S를 고려하는 경우, 만약 빈도(frequency) F가 심도와 서로 독립인 것을 가정하면

$$E[T] = E[F]E[S]$$
$$\mathrm{Var}[T] = E[F]\mathrm{Var}[S] + E[S]^2\mathrm{Var}[F]$$

이므로 이를 $Z = kE(T)/\{z_p\sqrt{\mathrm{Var}(T)}\}$에 대입하면 다음을 얻게 된다.

$$Z = \frac{kE(T)}{z_p\sqrt{\mathrm{Var}(T)}} = \frac{kE[F]E[S]}{z_p\{E[F^2]\mathrm{Var}[S] + E[S]^2\mathrm{Var}[F]\}}$$

따라서 $Z = 1$ 즉, 신뢰도가 1인 완전신뢰도(full credibility)의 경우 위 식은 다음과 같이

$$k^2E[S]^2E[F]^2 = {z_p}^2\left(\mathrm{Var}[F]E[S]^2 + \mathrm{Var}[S]E[F]\right)$$

또는

$$E[F] = \left(\frac{{z_p}^2}{k}\right)^2\left(\frac{\mathrm{Var}[F]}{E[F]} + \frac{\mathrm{Var}[S]}{E[S]^2}\right)$$

로 표현할 수 있다. 위 식을 $E[F]$에 관하여 풀었을 때의 해를 $Z = 1$일 때의 유효대수로 사용해야 하는데, 위 식의 우측항에도 $E[F]$가 포함되어 직접적으로 $E[F]$의 해를 구할 수 없다. 이때, 빈도의 분포(예를 들어, 포아송분포)에서 우측식내의 $\mathrm{Var}[F]/E[F]$가 보통 상수로 취급되는 것을 이용하면 좌측의 $E[F]$만 미지수로 처리되므로 그 해를 다음과 같이 $Z = 1$일 때의 유효대수 N_{full}로 정의한다.

$$N_{full} = \left(\frac{z_p}{k}\right)^2\left(\frac{\mathrm{Var}[F]}{E[F]} + \frac{\mathrm{Var}[S]}{E[S]^2}\right). \tag{2.2}$$

위 식에서, 만약 심도를 상수 취급하여 고려하지 않으면 $\text{Var}[S] = 0$이고, 빈도 F가 평균 보험료 청구건수가 N인 포아송 분포를 따른다고 가정하면 포아송 분포는 평균과 분산이 같은 특성이 있으므로 $\text{Var}[F]/E[F] = 1$이므로 식 (2.1)과 같게 된다.

2.3. 부분신뢰도(Partial Credibility)

완전신뢰도(full credibility)를 얻기 위하여 필요한 유효대수가 확보되어있는 경우 $Z = 1$을 부여하며, 그렇지 않은 경우 $0 \le Z < 1$인 부분신뢰도(partial credibility)를 부여한다. $Z = 1$인 경우의 보험료 유효대수를 N_{full}이라고 할 때, 보험료 유효대수 N이 N_{full}보다 작은 경우, 즉 $N < N_{full}$인 경우에 부분신뢰도(partial credibility)를 결정하여 부여하고, $N \ge N_{full}$인 경우에는 완전신뢰도 $Z = 1$을 부여한다. 부분신뢰도를 부여하는 방법은 다음의 네 가지 방법이 주로 사용된다.

2.3.1. 유효대수 법칙(The rule of relative exposure volume) 유효대수 법칙은 전년도의 유효대수와 당해년도의 유효대수를 사용하여 익년도의 보험료 산출에 적용할 신뢰도를 결정하는 방법으로서, 계산과 적용이 매우 쉬운 장점이 있다. 유효대수 법칙과 제곱근 법칙은 미국 손해보험협회에서 보험요율 산정에 사용하고 있는 방법으로서 Whitney (1918)가 처음으로 제안하였고, Perryman (1932), Longley-Cook (1960) 등의 연구에서 주로 사용되기 시작하였다. 유효대수 법칙에서 부분신뢰도 Z는 다음과 같이 결정된다.

$$Z = \frac{N_0}{N_0 + N_1},$$

여기서 N_0는 전년도의 유효대수, N_1은 당해년도의 유효대수이다.

2.3.2. 제곱근 법칙(The square root rule) 제곱근 법칙은 새로운 추정치에 대한 데이터의 기여도의 표준편차(standard deviation)가 완전신뢰도의 값에 대응하는 값을 갖도록 고안되었다. 제곱근 법칙에서 부분신뢰도 Z는 다음과 같이 결정된다.

$$Z = \sqrt{\frac{N}{N_{full}}}\ .$$

2.3.3. Bühlmann 신뢰도 최대 정확도를 갖는 신뢰도로 알려져 있는 Bühlmann 신뢰도는 최소제곱 신뢰도라고도 불리며, Bühlmann (1967, 1969)에 의해 제안되었다. Bühlmann 신뢰도는 서로 독립인 두 추정량의 제곱오차(squared error)를 사용하여 부분신뢰도를 구하는 방법이다.

Bühlmann 신뢰도를 사용하기 위해서는 Bühlmann 신뢰도의 모수 K를 계산하거나 추정하여 사용하는 것이 필요하다. K는 과정분산의 기댓값 EPV(the expected value of the process variance)와 가설평균의 분산 VHM(the hypothetical mean)의 비(ratio)로 정의되며 이를 사용하여 Bühlmann 신뢰도 Z는 다음과 같이 정의된다.

$$Z = \frac{N}{N + K}, \quad K = \frac{\text{EPV}}{\text{VHM}},$$

여기서 EPV와 VHM은 임의의 확률변수 θ에 대하여 확률변수 X의 분산 $\text{Var}[X]$의 특성을 이용하여 다음과 같이 정의할 수 있다.

$$\begin{aligned}\text{Var}[X] &= E[\text{Var}(X|\theta)] + \text{Var}[E(X|\theta)] \\ &= \text{EPV} + \text{VHM}\end{aligned}$$

즉, $\mathrm{Var}[X] = \mathrm{EPV} + \mathrm{VHM}$이며, EPV와 VHM은 각각 다음과 같이 변형하여 표현할 수 있다.

$$\begin{aligned}\mathrm{EPV} &= E[\mathrm{Var}(X|\theta)] \\ &= E\left[E\left(X^2|\theta\right)\right] - E\left\{[E(X|\theta)]^2\right\} \\ &= E\left(X^2\right) - E\left\{[E(X|\theta)]^2\right\} \\ \mathrm{VHM} &= \mathrm{Var}[E(X|\theta)] \\ &= E\left\{[E(X|\theta)]^2\right\} - \{E[E(X|\theta)]\}^2 \\ &= E\left\{[E(X|\theta)]^2\right\} - [E(X)]^2.\end{aligned}$$

Bühlmann 신뢰도의 이해를 위하여 2.1절에서 소개하였던 식 $E_2 = ZT + \{(1-Z)\times E_1\}$을 살펴 보기로 한다. 여기서 E_2는 '새로운 추정치(new estimate)', Z는 신뢰도, T는 '현재의 데이터'로서 제곱오차(squared error)가 U, E_1은 '기존의 추정치(previous estimate)'로서 제곱오차가 V이며 T와 E_1은 서로 독립이다. 새로운 추정치 E_2의 제곱오차를 W라 하면 T와 E_1은 서로 독립이므로 $W = Z^2U + (1-Z)^2V$이며, W를 최소화하는 Z의 해를 구하면 $Z = V/(U+V)$이며, 이것이 Bühlmann 신뢰도의 기본 아이디어이다.

2.3.4. Bühlmann-Straub 신뢰도 Bühlmann-Straub 신뢰도는 Bühlmann 신뢰도의 일반화된 형태로 생각할 수 있으며, Bühlmann 신뢰도의 약점은 지난 보험기간 동안에 발생한 여러 종류의 청구건수 또는 여러 종류의 보상액 분포를 고려하지 않는다는 것이다. 예를 들어, 단체 보험에서 Bühlmann 신뢰도는 2년 또는 그 이상의 보험기간 동안 단체 보험가입자의 수에 대한 변화를 고려할 수 없다. Bühlmann-Straub 신뢰도는 이러한 결함을 보완할 수 있는 신뢰도이다. Bühlmann-Straub 신뢰도의 기본원리는 간단하다. Bühlmann 신뢰도에서는 사전확률을 임의로 지정하여 사용하였지만 Bühlmann-Straub 신뢰도에서는 보상 건수에 대해 가중치를 두어 사전확률을 사용한다. 즉, m_j를 j번째 보험기간 동안의 보상건수라 하면, $E(X_j|\theta)$의 점추정량은 다음과 같이 나타낼 수 있다.

$$E(X_j|\theta) = Z\bar{X} + (1-Z)\mu = Z\frac{\sum_j m_j X_j}{\sum_j m_j} + (1-Z)\mu.$$

본 논문의 실증 자료 분석에서는 Bühlmann 신뢰도에서의 EPV와 VHM을 EPV_1과 VHM_1, Bühlmann-Straub 신뢰도에서의 EPV와 VHM을 EPV_2와 VHM_2라 표기하고, 이에 근거하여 계산되는 얻어지는 Bühlmann 신뢰도의 모수를 K_1, Bühlmann-Straub 신뢰도의 모수를 K_2라 표기하기로 한다.

3. 실증 자료 분석

자동차 보험사에서는 다양한 고객의 정보를 이용하여 사고 발생시 고객에게 지급하는 보험금을 기준으로 고객이 납부하는 보험료를 책정한다. 따라서 고객 정보의 정확한 분석을 통한 고객 집단별 보험료의 정교한 추정은 보험사의 손익에 가장 큰 영향을 준다고 할 수 있다. 본 장에서는 김영화와 이현수 (2010)에서 사용하였던 국내 자동차 보험회사의 실제 자료에 대하여 네 가지의 신뢰도를 적용하여 보험료를 결정하는 방법을 소개하고, 각 방법의 효율성을 비교한다.

3.1. 고객 데이터와 위험도

분석에 사용된 자료는 국내 화재보험사인 S사의 실제 자료로서 전체 932,880명의 고객 데이터를 사용한다. 설명변수로는 고객의 사고유무에 따른 할인할증, 운전경력, 운전자한정특약, 차종, 연령대, 성별이

표 3.1. 설명변수

설명변수명	범주수	범주 내용
할인할증	3	사고(할증), 신규, 무사고(할인)
운전경력	4	1∼2년, 3∼4년, 5∼6년, 7년이상
운전자한정특약	4	부부, 1인, 가족형제, 기타
차종	3	소형, 중대형, 대형(SUV, 다인승)
연령대	3	20∼39세, 40∼59세, 60세이상
성별	2	남, 여

표 3.2. 분류된 고객 데이터 및 요약값

연도	총관측수	평균청구액(원)
year1	311,882	44,148.778
year2	310,333	45,001.463
year3	310,665	43,616.629

사용되었으며, 이러한 설명변수들에 따라 고객 집단을 분류하여 사고 1건당 발생되는 평균 손해액, 즉 보험사에서 지급하는 보험금을 집단별로 계산하고 신뢰도를 적용하여 새로운 보험료를 책정한다. 설명변수의 자세한 내용은 표 3.1과 같으며 종속변수는 보험금이다.

분석에 사용된 실제 데이터는 932,880건이며, 이를 Uniform$(0, 1)$을 따르는 난수를 발생시켜 세 집단으로 분류하였다. 분류된 첫 번째 집단(year1)은 전년도 데이터, 두 번째 집단(year2)은 당해년도의 데이터로 간주하여 익년도 보험료를 산출하는데 필요한 데이터로 사용하였으며, 세 번째 집단(year3)은 여러 가지 신뢰도를 사용하여 산출한 보험료가 적절한 지를 판단하는 가상의 익년도 데이터로 사용하였다. 전체 고객은 사고유무에 따른 할인할증, 운전경력, 운전자한정, 차종, 연령대, 성별에 따라 864개$(3 \times 4 \times 4 \times 3 \times 3 \times 2)$의 집단으로 분류하였다. 다음 표 3.2는 분류된 세 집단에 대한 요약값이며, year1과 year2의 데이터를 사용하여 신뢰도를 구한 후 새로운 보험금을 추정한 후, 이를 year3의 평균 청구액과 비교하여 여러가지 신뢰도를 비교한다

신뢰도의 비교분석을 위하여 여러 가지 신뢰도를 구하기 전에 먼저 6개의 설명변수의 각 수준에 따라 현행상대도와 위험도를 구한다. 현행상대도는 전년도 데이터(year1)를 사용해 각 변수의 수준에 해당하는 고객들의 평균손해액을 총평균손해액으로 나눈 값을 의미한다. 본 논문에서는 i번째 변수의 j번째 수준의 현행상대도를 C_{ij}라 나타내기로 한다. 예를 들어, 첫 번째 설명변수인 할인할증의 첫 번째 수준(사고(할증))에 해당하는 현행상대도 C_{11}은 year1의 고객중 사고를 낸 고객의 평균손해액이 64,521.12이고 year1의 평균손해액이 44,148.778이므로 $C_{11} = 64{,}521.12/44{,}148.778 = 1.4614475$이다.

위험도를 구하는 방법도 현행상대도를 구하는 방법과 동일하지만, year1이 아닌 year2를 사용해서 구한다. 본 논문에서는 i번째 변수의 j번째 수준의 위험도를 D_{ij}라 나타내기로 한다. 예를 들어, 할인할증의 첫 번째 수준(사고(할증))에 해당하는 위험도 D_{11}은 year2의 고객 중 사고를 낸 고객의 평균손해액이 63,430.26이고 year2의 평균손해액이 45,001.463이므로 $D_{11} = 63{,}430.26/45{,}001.463 = 1.4095155$이다. 표 3.3은 이러한 방법으로 6개 설명변수에 대하여 현행상대도와 위험도를 계산한 결과의 일부이다.

3.2. 완전신뢰도의 경우에 필요한 유효대수 추정

현재 국내 자동차 보험회사에서는 오차허용한계 k는 0.05, 신뢰수준 p는 0.9를 주로 사용하므로 여기서도 그 값들을 사용하였다. 완전신뢰도 즉, $Z = 1$인 경우 필요한 유효대수(N_{full})를 심도를 고려하지 않은 경우는 표 2.1의 값을 이용하여 1,082이고, 심도를 고려하는 경우는 식 (2.2)를 이용해서 구할 수 있으며 표 3.4는 그 결과의 일부이다.

표 3.3. 설명변수의 수준별 현행상대도와 위험도

변수	수 준	현행상대도	위험도
할인할증	사고(할증)	1.46144747	1.40951551
	신규	1.44867362	1.51779866
	무사고(할인)	0.90433783	0.89879189
운전경력	1∼2년	1.07765293	1.12336503
	3∼4년	1.18310047	1.25996148
	5∼6년	0.94232280	0.93619801
	7년이상	0.96786258	0.92961244
⋮	⋮	⋮	⋮
성별	남	1.07354704	1.06484138
	여	0.88620324	0.89984652

표 3.4. 심도를 고려하는 경우 N_{full}의 값

변수	범주내용	유효대수
할인할증	사고(할증)	1,535
	신규	1,694
	무사고(할인)	1,604
운전경력	1∼2년	1,366
	3∼4년	1,697
	5∼6년	1,663
	7년이상	1,541
⋮	⋮	⋮
성별	남	1,636
	여	1,573

3.3. 부분신뢰도 추정

완전 신뢰도를 얻기 위하여 필요한 최소한의 유효대수가 확보되어 있는 경우는 100% 신뢰도를 부여하지만 그렇지 않은 경우에는 100%보다 작은 부분 신뢰도를 부여하게 된다. 유효대수 법칙에 따른 부분신뢰도를 Z_1, 심도를 고려하지 않는 경우 제곱근 법칙에 따른 부분신뢰도를 Z_2, 심도를 고려하는 경우 제곱근 법칙에 따른 부분신뢰도를 Z_3, Bühlmann 신뢰도를 Z_4, Bühlmann-Straub 신뢰도를 Z_5라 하자. 즉, N_0를 전년도 유효대수, N_1를 당해년도 유효대수, N을 당해년도 청구건수, N_{full_1}을 심도를 고려하지 않은 경우 완전신뢰도에 필요한 유효대수, N_{full_2}를 심도를 고려하는 경우 완전신뢰도에 필요한 유효대수, K_1을 Bühlmann 신뢰도의 모수, K_2를 Bühlmann-Straub 신뢰도의 모수라고 하면 Z_1, Z_2, Z_3, Z_4, Z_5는 각각 다음과 같이 정의된다.

$$Z_1 = \frac{N_1}{N_0 + N_1}, \qquad Z_2 = \sqrt{\frac{N}{N_{full_1}}}, \qquad Z_3 = \sqrt{\frac{N}{N_{full_2}}}, \qquad Z_4 = \frac{N_1}{N_1 + K_1}, \qquad Z_5 = \frac{N_1}{N_1 + K_2}.$$

표 3.5는 부분신뢰도를 구하기 위하여 6개 변수의 범주에 따라 자료를 분류하여 예시한 것이며, 평균심도는 각 범주의 심도를 표본수로 나눈 값을 의미한다.

Bühlmann 신뢰도를 구하는 과정을 살펴보면 다음과 같다.

X_1을 year1의 평균 보험금 청구건수, X_2를 year2의 평균 보험금 청구건수라 하면 X_1과 X_2의 분포는 각각 $X_1 \sim \text{Poisson}(\lambda_1 = 17.491898), X_2 \sim \text{Poisson}(\lambda_2 = 17.574074)$이며, 여기서 λ_1, λ_2는 각각

표 3.5. 변수의 범주에 따라 분류된 자료

변수	범주내용	year1 표본수	year1 청구건수	year1 평균심도	year2 표본수	year2 청구건수	year2 평균심도
할인할증	사고(할증)	19,027	1,178	64,521.12	18,735	1,242	63,430.26
	신규	35,337	2,343	63,957.17	35,282	2,309	68,303.16
	무사고(할인)	257,518	11,592	39,925.41	256,316	11,633	40,446.95
운전경력	1∼2년	14,662	689	47,577.06	14,419	719	50,553.07
	3∼4년	55,055	3,078	52,232.44	54,997	3,185	56,700.11
	5∼6년	134,557	6,242	41,602.40	133,875	6,298	42,130.28
	7년 이상	107,608	5,104	42,729.95	107,042	4,982	41,833.92
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
성별	남	189,444	9,836	47,395.79	188,375	9,835	47,919.42
	여	12,438	527	39,124.79	121,958	5,349	40,494.41

표 3.6. 연도별 사전확률, 평균, 평균제곱, 분산

연도	사전확률	평균	평균제곱	분산
X_1(year1)	0.5	17.491898	305.966501	17.491898
X_2(year2)	0.5	17.574074	308.848080	17.574074

표 3.7. Bühlmann-Straub 신뢰도에 적용되는 연도별 사전확률

연도	청구건수	사전확률
X_1(year1)	15,113	$\frac{15{,}113}{15{,}113+15{,}184}=0.498828$
X_2(year2)	15,184	$\frac{15{,}184}{15{,}113+15{,}184}=0.501172$

year1과 year2의 총보험금 청구건수를 전체 범주의 수로 나누어 추정한 값이다. 다음 표 3.6은 EPV_1과 VHM_1을 계산하기 위하여 필요한 값들을 정리하여 놓은 것이다. 여기서 적용된 사전확률의 값 θ는 국내 보험회사에서 현재 사용하고 있는 확률값을 사용하였다.

표 3.6의 값들을 사용하여 Bühlmann 신뢰도의 모수 K_1은 다음과 같이 계산된다.

$$\begin{aligned}
\text{EPV}_1 &= E[\text{Var}(X|\theta)] = (0.5\times 17.491898)+(0.5\times 17.574074) = 17.532986\\
E\left[E(X|\theta)^2\right] &= (0.5\times 305.966501)+(0.5\times 308.848080) = 307.407290\\
[E(X)]^2 &= [(0.5\times 17.491898)+(0.5\times 17.574074)]^2 = 307.405602\\
\text{VHM}_1 &= \text{Var}[E(X|\theta)] = E\left[E(X|\theta)^2\right]-[E(X)]^2 = 0.00168822\\
K_1 &= \frac{\text{EPV}_1}{\text{VHM}_1} = \frac{17.532986}{0.001688} = 10{,}385.48.
\end{aligned}$$

다음으로 Bühlmann-Straub 신뢰도를 구하는 과정을 살펴보자. X_1과 X_2, λ_1과 λ_2의 값은 Bühlmann 신뢰도를 구하는 과정에서 사용하였던 값과 동일하게 사용한다. 단, Bühlmann-Straub 신뢰도는 사전확률을 동일하게 주지 않고, 각 연도별 청구건수에 따라 가중치를 두어 계산한다. 다음 표 3.7은 Bühlmann-Straub 신뢰도를 구하기 위한 연도별 사전확률을 나타내고 있는데, 본 연구에서는 Uniform $(0,1)$ 분포를 사용하여 데이터를 year1과 year2로 분류하였기 때문에 표 3.7의 사전확률이 표 3.6의 사전확률과 매우 비슷한 값을 갖게 되는 것을 알 수 있고, 이러한 이유로 인해 Bühlmann 신뢰도와

표 3.8. 신뢰도 추정값

변수	범주내용	Z_1	Z_2	Z_3	Z_4	Z_5
할인할증	사고(할증)	0.503866	1	0.899649	0.643362	0.643359
	신규	0.500389	1	1	0.772585	0.772583
	무사고(할인)	0.501170	1	1	0.961060	0.961059
운전경력	1∼2년	0.504178	0.815093	0.725416	0.581306	0.581304
	3∼4년	0.500264	1	1	0.841158	0.841157
	5∼6년	0.501270	1	1	0.928009	0.928008
	7년이상	0.501318	1	1	0.911558	0.911557
⋮	⋮	⋮	⋮	⋮	⋮	⋮
성별	남	0.501415	1	1	0.947749	0.947748
	여	0.500982	1	1	0.921526	0.921526

Bühlmann-Straub 신뢰도를 사용했을 때 그 결과가 매우 비슷하게 된다. 그러나 Bühlmann 신뢰도에 적용하는 사전확률을 조정하면 그 결과는 달라지게 되며 이는 추후 논의하기로 한다.

또한 표 3.7의 값들을 사용하여 Bühlmann-Straub 신뢰도의 모수 K_2는 다음과 같이 계산된다.

$$\begin{aligned}
\mathrm{EPV}_2 &= E[\mathrm{Var}(X|\theta)] = (0.498828 \times 17.491898) + (0.501172 \times 17.574074) = 17.533082 \\
E\left[E(X|\theta)^2\right] &= (0.498828 \times 305.966501) + (0.501172 \times 308.848080) = 307.410667 \\
[E(X)]^2 &= [(0.498828 \times 17.491898) + (0.501172 \times 17.574074)]^2 = 307.408979 \\
\mathrm{VHM}_2 &= \mathrm{Var}[E(X|\theta)] = E\left[E(X|\theta)^2\right] - [E(X)]^2 = 0.00168819 \\
K_2 &= \frac{\mathrm{EPV}_2}{\mathrm{VHM}_2} = \frac{17.53308}{0.00169} = 10{,}385.60.
\end{aligned}$$

표 3.8은 각 변수의 범주에 따른 4가지 신뢰도의 추정값이다. 표 3.8에 제시한 신뢰도의 값의 계산과정을 표 3.4와 표 3.5를 참조하여 몇 가지 살펴보면 다음과 같다.

$$\begin{aligned}
Z_1:&\ 0.503866 = \frac{19{,}027}{19{,}027 + 18{,}735}, \qquad 0.500389 = \frac{35{,}337}{35{,}337 + 35{,}282}, \ldots \\
Z_2:&\ 1 = \min\left(\sqrt{\frac{1{,}242}{1{,}082}}, 1\right), \qquad 1 = \min\left(\sqrt{\frac{2{,}309}{1{,}082}}, 1\right), \ldots \\
Z_3:&\ 1 = \min\left(\sqrt{\frac{1{,}242}{1{,}535}}, 1\right), \qquad 1 = \min\left(\sqrt{\frac{2{,}309}{1{,}694}}, 1\right), \ldots \\
Z_4:&\ 0.643362 = \frac{18{,}735}{18{,}735 + 10{,}385.48}, \qquad 0.772585 = \frac{35{,}282}{35{,}282 + 10{,}385.48}, \ldots \\
Z_5:&\ 0.643359 = \frac{18{,}735}{18{,}735 + 10{,}385.60}, \qquad 0.772583 = \frac{35{,}282}{35{,}282 + 10{,}385.60}, \ldots.
\end{aligned}$$

3.4. 부과보험료 추정 및 신뢰도 비교

부과보험료를 추정하기 위해서는 반영위험도를 먼저 계산해야 하는데, 반영위험도는 2.1절에서 소개하였던 Whitney (1918)의 제안에 따라 전년도의 자료를 근거로 계산한 위험도와 당해년도의 자료를 근거로 계산한 현행위험도를 신뢰도를 가중치로 하는 가중평균으로 계산한다. 즉, '반영위험도 $= Z \times$ 위험도 $+ (1 - Z) \times$ 현행상대도'로 계산된다.

표 **3.9.** 신뢰도에 따른 부과보험료

집단	표본수	신뢰도				
		Z_1	Z_2	Z_3	Z_4	Z_5
1	85	46,814	49,234	49,234	48,194	48,194
2	118	56,048	58,261	58,261	57,123	57,123
3	410	41,075	41,230	41,230	41,014	41,014
⋮	⋮	⋮	⋮	⋮	⋮	⋮
864	498	49,176	48,790	48,790	48,613	48,613

표 **3.10.** 전체 손해액과 신뢰도에 따른 부과보험료 총합의 차이

	신뢰도				
	Z_1	Z_2	Z_3	Z_4	Z_5
부과보험료 총합	1.4011E10	1.4058E10	1.4059E10	1.3948E10	1.3948E10
차이	+4.6E7	+9.3E7	+9.4E7	−1.7E7	−1.7E7
전체 손해액			1.3965E10		

본 논문에서는 i번째 변수의 j번째 수준의 반영위험도를 R_{ij}라 표현하기로 한다. 예를 들어, 할인할증의 첫 번째 수준(사고(할증))에 해당하는 반영위험도 R_{11}은 표 3.3의 값들을 이용하여 다음과 같이 계산된다.

$$R_{11} = Z \times D_{11} + (1 - Z) \times C_{11} = Z \times 1.40951551 + (1 - Z) \times 1.46144747$$

반영위험도를 구한 후에, 각 변수의 수준에 대한 반영위험도의 곱으로 부과 보험료를 계산한다. 예를 들어, '할인할증 $=1$ (사고(할증)), 운전경력 $=2$ (3∼4년), 운전자한정특약 $=2$ (1인), 차종 $=1$ (소형), 연령대 $=3$ (60세 이상), 성별 $=1$ (남)'인 집단의 부과보험료(New Rate)는 다음과 같이 계산한다.

$$\text{New Rate} = \text{Loss\_Mean} \times R_{11} \times R_{22} \times R_{32} \times R_{41} \times R_{53} \times R_{61},$$

여기서 Loss_Mean은 총평균손해액으로서 당해년도의 전체 손해액을 전체 가입자수로 나눈 값이다. 표 3.9에 전체 864개 집단에 대한 부과보험료를 예시하였으며, 표 3.9에서 보는 것과 같이 신뢰도를 5가지 방법에 의해서 구했기 때문에 부과보험료도 5가지로 계산되었다.

만일 위와 같이 추정된 부과보험료의 총합이 당해년도의 전체 손해액과 같지 않다면 당해년도 부과보험료의 수준을 동일하게 하는 수지균등의 원칙에 따라 전체적인 보험료의 수준에 변동이 없도록 조정하여야 한다. 본 논문에서 사용한 자료의 당해년도에 대하여 전체 손해액, 부과보험료의 총합, 전체 손해액과 부과보험료의 총합의 차이는 표 3.10과 같으므로 이러한 차이를 전체표본수로 나누어 그 값을 가감하는 절차를 거쳐야하며, 최종적으로 수정된 부과보험료가 표 3.11이다. 또한 표 3.12는 전체 864개 범주에 대하여 신뢰도에 따라 추정된 보험료와 year3에 발생한 실제 손해액 사이의 오차제곱합의 평균값을 계산한 결과이다. 결과적으로, 유효대수법칙에 근거하여 구한 부분신뢰도가 가장 정확하게 추정하는 것으로 나타났으며, Bühlmann 신뢰도, Bühlmann-Straub 신뢰도, 심도를 고려한 제곱근 법칙에 근거한 신뢰도, 심도를 고려하지 않은 제곱근 법칙에 근거한 신뢰도 순서로 예측의 정확성을 보였다.

이러한 결과는 다른 방법들과 달리 특정 분포 또는 독립성 등에 대한 가정을 필요로 하지 않는 유효대수법칙의 장점에 기인한 것으로 판단되며 현재 국내 보험사에서도 유효대수법칙을 주로 사용하여 보험료를 산정하고 있다.

표 **3.11.** 신뢰도에 따른 부과보험료의 수정치

집단	표본수	신뢰도				
		Z_1	Z_2	Z_3	Z_4	Z_5
1	85	46,666	48,933	48,931	48,248	48,194
2	118	55,900	57,961	57,958	57,177	57,123
3	410	40,926	40,930	40,927	41,069	41,014
⋮	⋮	⋮	⋮	⋮	⋮	⋮
864	498	49,028	48,490	48,487	48,668	48,668

표 **3.12.** 신뢰도에 따른 평균 오차제곱합

신뢰도	평균 오차제곱합
Z_1	9.0455E13
Z_2	1.0563E14
Z_3	1.0562E14
Z_4	9.8631E14
Z_4	9.8631E14

표 **3.13.** 사전확률에 따른 Bühlmann 신뢰도의 평균오차제곱합

사전확률		평균 오차제곱합
year1	year2	
0.01	0.99	8.892E13
0.10	0.90	9.280E14
0.20	0.80	9.617E14
0.30	0.70	9.771E14

앞서 설명한 바와 같이 본 논문에서는 자료를 Uniform$(0, 1)$ 난수를 이용하여 랜덤하게 세 그룹(year1, year2, year3)으로 나누어 사용했기 때문에, Bühlmann 신뢰도에 적용된 사전확률의 값과 Bühlmann-Straub 신뢰도에 적용된 사전확률의 값이 거의 비슷하여 추정보험료의 평균오차제곱합이 같은 결과를 보이고 있다. 그러나 Bühlmann 신뢰에도 적용되는 사전확률의 값을 조정한다면 표 3.13과 같은 결과를 얻을 수 있다.

4. 결론

자동차보험을 판매하는 보험사와 자동차보험을 구매하는 보험가입자에게 보험료는 매우 중요한 요소이다. 보험가입자의 특성에 따라 부과되는 자동차 보험료는 보험사의 손익에 가장 큰 영향을 미치는 동시에 물가에 큰 영향을 미치기 때문에 이를 책정하는 보험사에게 가입자의 위험도에 따라 적정한 보험료를 책정하는 것은 매우 중요한 의사결정과정이다.

본 연구에서는 적정한 보험료를 산출하는데 사용할 수 있는 다섯 가지 신뢰도의 개념 및 특성을 살펴보고, 실증 자료분석을 통하여 각 신뢰도의 보험료 추정의 정확성을 비교하였다. 전체 보험가입 고객은 사고유무에 따른 할인할증, 운전경력, 운전자한정특약, 차종, 연령대, 성별에 따라 864개$(3 \times 4 \times 4 \times 3 \times 3 \times 2)$의 그룹으로 분류하였으며, 나누어진 그룹의 관측치가 적은 경우에 적절한 위험도를 부여하는 부분신뢰도의 개념을 적용하여 각 그룹별로 적용할 새로운 보험료를 추정하였다. 특히 새로운 보험료를 추정하는 과정에서는 김영화와 이현수 (2010)에서 사용한 추정방법을 개선하여 반영위험도의 개념을 적용하였다.

실증자료 분석 결과, 유효대수법칙에 근거하여 구한 부분신뢰도가 가장 정확하게 추정하는 것으로 나타났으며, Bühlmann 신뢰도, Bühlmann-Straub 신뢰도, 심도를 고려한 제곱근 법칙에 근거한 신뢰도, 심도를 고려하지 않은 제곱근 법칙에 근거한 신뢰도의 순서로 보험료 예측의 정확성을 보였다. 본 논문에서 사용한 자료는 Uniform$(0, 1)$ 난수를 사용하여 랜덤하게 year1, year2, year3으로 나누었기 때문에 year1과 year2에 대하여 각 년도의 청구건수가 거의 비슷하게 되어, Bühlmann 신뢰도와 Bühlmann-Straub 신뢰도는 차이를 보이지 않았다. 만약 과거 데이터를 이용하여 Bühlmann 신뢰도의 사전확률을 조정하는 방법이 개발되어 사전확률의 값을 달리 부여한다면 더욱 정확한 예측 결과를 얻을 수 있을 것으로 판단된다. 또한 본 논문의 연구 결과는 특정 보험사의 자료에 근거한 것이므로 타사에서 본 연구 결과를 직접 적용하는 경우에는 유의하여야 하며 각 보험사 가입자의 특성을 고려하여 적용하여야 할 것으로 판단된다.

참고문헌

김영화, 이현수 (2010). 신뢰도에 근거한 자동차보험 가격산출 비교, <한국통계학회논문집>, **17**, 713–724.

Bailey, A. (1950). A generalized theory of credibility, *Proceedings of the Casualty Actuarial Society*, XXXII, **13**, 1–20.

Bühlmann, H. (1967). Experience rating and credibility I, *JASTIN Bulletin*, **4**, 199–207.

Bühlmann, H. (1969). Experience rating and credibility II, *ASTIN Bulletin*, **5**, 157–165.

Bühlmann, H. and Straub, E. (1970). Glabwürdigkeit fur schandensätze, *Bulletin of the Swiss Association of Actuaries Communications in Statistics, Theory and Methods*, **70**, 111–133.

Frees, E. W. (2003). Multivariate credibility for aggregate loss models, *North American Actuarial Journal*, **7**, 13–37.

Hachemeister, C. A. (1975). Credibility for regression models with application to trend, *Credibility, Theory and Application*, Academic Press, New York, 129–163.

Longley-Cook, H. (1960). An Introduction to Credibility Theory, XLIX PCAS **194**.

Mowbray, A. H. (1914). How extensive a payroll exposure is necessary to give a dependable pure premium, *Proceedings of the Casualty Actuarial Society*, **1**, 24–30.

Perryman, F. S. (1932). *Some Notes on Credibility*, XIX PCAS, **65**.

Whitney, A. W. (1918). The theory of experience rating, *Proceedings of the Casualty Actuarial Society*, **4**, 274–292.

Estimating the Automobile Insurance Premium Based on Credibilities

Yeong-Hwa Kim[1] · Mi Jung Kim[2] · Myung Joon Kim[3]

[1]Department of Statistics, Chung-Ang University
[2]Department of Statistics, Chung-Ang University
[3]Samsung Fire & Marine Insurance Co.

(Received December 2010; accepted February 2011)

Abstract

Credibility theory is one of the most important theories of actuarial science to calculate the proper insurance premium. In this paper, the rule of relative exposure volume, the square root rule, the Bühlmann credibility and Bühlmann-Straub credibility with the basic concept of credibility have been introduced, Also, we estimate new premiums based on these methods for real data. As a result, the rule of relative exposure volume provides the highest accuracy.

Keywords: Bühlmann credibility, Bühlmann-Straub credibility, credibility, relative exposure volume, square root rule.

[1]Corresponding author: Associate Professor, Department of Statistics, Chung-Ang University, Seoul 156-756, Korea. E-mail: gogators@cau.ac.kr