

2010년 한국금융학회
정기학술대회 발표논문

개인신용정보와 생명보험상품 사고율 간의 상관성 연구

2010. 6

정재욱(세종대)
여유희 (KCB)

I. 서론

보험회사가 보험계약의 승인심사 및 보험요율 산정 등에 개인신용정보(personal credit information)를 활용할 수 있는지 여부와 그 효과 등에 대하여 여러 찬반논란이 있다. 우선 찬성하는 측에서는 보험신용평점¹⁾이 개인이 보유한 위험을 보다 정확히 예측하는데 도움을 주며, 역선택 문제를 경감시킴으로써 궁극적으로 개인에게 혜택을 제공할 것이라고 주장하고 있다. 반면에 이를 반대하는 측에는 개인의 신용이 보험리스크를 예측하는데 도움을 준다는 어떠한 확고한 증거도 없으며, 또한 더 나아가 오히려 저소득층, 소수인종그룹 등 사회 특수계층의 보험료를 높이는 등 부정적인 영향을 미칠 수 있다는 주장을 펴고 있다.

이와 관련하여 해외에는 비교적 소수의 실증연구 결과물이 존재하는데, 특히 개인신용정보의 집중 및 이용이 자유로운 미국에서 개인신용정보의 활용이 자동차보험 등 가계성보험(personal lines of insurance)에 어떠한 영향을 미치는 지에 대한 연구가 비교적 활발히 진행되어 왔다.²⁾³⁾ 그럼에도 불구하고 개인신용정보를 활용한 보험신용평점의 개발 및 적용이 대부분의 보험회사에게는 영업비밀인 동시에 시장에서의 경쟁우위를 차지할 수 있는 중요한 요소로 간주되고 있는 관계로, 일반인들에게 공표된 분석결과를 찾아보기란 매우 어려운 게 사실이다.

한편 국내에서는 처음으로 정재욱·손승호(2009)가 개인신용정보와 개인용자동차보험 사고발생확률 및 손해율 사이의 상관관계를 실증 분석하였다. 분석결과에 의하면 첫째, 개인의 신용등급(1~10등급)과 보험사고(사고빈도 및 손해율)와의 관계를 살펴본 결과, 개인용자동차보험의 경우 이들 사이에 매우 밀접한 상관관계가 존재하고 있음을 보여주었다. 둘째, 다변량 로지스틱 회귀분석모형의 비교 분석을 통해 순수보험정보만 고려한 경우보다는 부정적 개인신용정보를 추가로 고려한 모형의 보험사고 예측력이 좀 더 우수하며, 여기에 긍정적 개인신용정보까지 추가로 반영할 경우 모형의 보험사고 예측력이 더욱 우수하다는 것을 보여주었다. 그러나 이 연구는 일정기간 동안 특정 1개 손해보험사의 개인용자동차보험의 신규 및 갱신 계약 건만을 대상으로 분석을 실시하였기 때문에 동 기간 내 개인용자동차보험 신청 고객 중 승인이 거절된 건에 대한 추가적인 분석이 이루어지지 못했다는 한계를 내포하고 있다.⁴⁾

1) 보험신용평점(credit-based insurance score)은 일반적으로 개인의 신용과 관련된 자료(즉, 과거 연체 또는 파산 경험, 신용한도소진율, 신규 신용개설을 위한 문의건수 및 신설계좌수, 신용기록, 기타 신용거래내역 등을 포함)를 사용하여 계산됨.

2) 미국에서는 공정거래보고법(FTRA, 1970)이 도입된 이후 보험계약 승인 심사 시 개인신용보고서의 활용이 법적으로 허용되었고, 해당 주(州)에 따라 신용정보의 활용범위가 상이하지만 보험요율 산정에 적용하는 것도 인정되고 있음. 또한 공명신용거래법(FACTA, 2003)이 발효된 이후 개인 신용정보의 이용이 보다 광범위하게 활성화되었음.

3) 주요 선행연구로는 Fair Trade Commission(2007), Wu&Guszcza(2003), Monaghan(2002) 등을 참조하기 바람.

4) 이러한 분석이 가능해지려면 자동차종합보험 신청시 승인거절 건에 대한 신청정보 및 보험사고정

본 연구의 목적은 국내에서는 처음으로 국내 A생명보험사의 실제 데이터를 활용하여 개인신용정보와 생명보험상품의 사고율 간에 어떠한 상관관계가 있는지를 살펴보는 데 있다. 즉, 현재 생명보험사가 내부적으로 활용하고 있는 기존의 언더라이팅 요소에 개인신용정보를 추가로 고려할 경우 보험사고 예측에 어느 정도의 한계효용 효과가 발생하는지를 계량적으로 검증하고자 한다.

본 연구의 구성은 다음과 같다. 제I장의 서론에 이어, 제II장에서는 데이터 및 분석방법론에 대해 상세히 기술한다. 다음으로 제III장에서는 분석의 결과 및 한계를 기술하고, 마지막으로 제IV장은 연구의 결론이다.

보가 필요하나, 현재로서는 해당 정보의 수집이 매우 어려운 실정임. 특히 승인거절 건에 대한 보험사고정보는 손해보험사 간 보험사고발생 정보가 공유되지 않는 한 정보수집이 거의 불가능하다고 볼 수 있음.

II. 데이터 및 분석방법론

1. 데이터

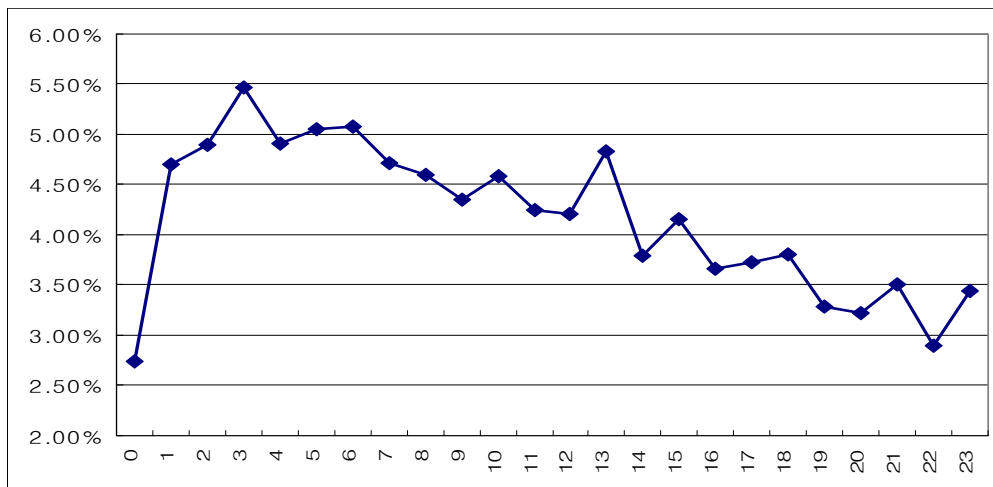
(1) 분석 대상

본 분석에 사용한 data는 국내 A생명보험사의 2007년 1~7월 기간 동안 보험계약 신청자를 대상으로 하였다⁵⁾. 분석대상보험은 건강보장 및 상해보장 상품으로 한정하였는데, 이는 역선택 및 모럴헤저드가 타 상품 대비 상대적으로 빈번히 발생할 수 있는 상품을 대상으로 하는 것이 분석의 목적에 부합한다고 판단하였기 때문이다. 이와 같은 요건에 따라 본 분석에 사용된 대상자수는 총 109,689명이다.

(2) 예측 기간

먼저 보험계약 신청 이후의 사고발생추이를 바탕으로 합리적인 보험사고 예측기간 설정을 위해 Vintage분석을 실시하였다.

<그림 1> 보험계약 신청 이후 기간 경과에 따른 사고 발생률 추이



<그림 1>에서 보듯이 보험계약 신청 이후 기간 경과에 따른 사고발생률의 추이를 살펴보면 처음 6개월까지는 비교적 높은 사고율을 기록하다가, 이후로는 점차 감소하는 것으로 나타났다. 이는 보험계약 신청자의 역선택 문제로 인해 계약초기에 상대적으로 많은 보험사고가 발생함에 따라 나타난 현상으로 추측할 수 있다.

Vintage 분석결과를 바탕으로 보험사고 예측기간은 보험계약 신청일 이후 2년

5) 다만 미성년자, 방카슈랑스, 단체보험, 단순청약테스트건 등은 보험업무의 특성상 각 청약건 평가 및 결과적용이 현실적으로 어렵다는 현업 의견을 반영하여 분석대상에서 제외하였음.

내 보험사고 발생여부로 설정하였다. 이는 무엇보다 보험계약 신청일 이후 2년이 경과하면 사고발생증가율이 일정수준으로 안정화된다는 점을 우선적으로 고려했기 때문이다. 또한 예측기간을 장기로 설정할 경우 분석대상 data가 현재와는 거리가 먼 과거시점으로 회귀되어 분석결과의 최신성에 문제가 있을 수 있다는 현실적인 한계점도 추가로 고려하였다. 이와 같이 예측기간을 설정하였을 때, 전체 분석대상자의 사고발생여부에 대한 구성비는 아래의 <표 1>과 같다.

<표 2> 분석대상자의 사고발생 여부 구성비

구분	대상자수	구성비
미사고	101,767	92.78%
사고	7,922	7.22%
합계	109,689	100.00%

(3) 활용 변수

보험사고 예측을 위해 활용된 변수는 정보출처에 따라 다음과 같이 크게 세 부문으로 나누어지며, 정보출처별 구체적인 주요 활용 변수는 아래의 <표 2>와 같다.

① 보험사 고유정보

현재 보험사에서 언더라이팅 시 고려하는 주요정보로 위험등급, 성별, 가입연령, 현재 보험계약 보유건수, 과거 보험사고 경험유무 등의 5개 변수를 분석에 활용하였다.

② 은행연합회 정보

은행연합회(이하 'KFB')가 수집하여 제공하는 정보로 대출 및 카드 개설정보, 3개월 이상 연체정보, 공공정보 등으로 구성되어 있다. 각 회원사에서는 이러한 정보를 요약된 정보항목의 형태로 가공하여 각종 의사결정에 활용하게 되는데, 대표적으로 활용되는 변수 54개를 선정하여 본 분석에 활용하였다.

③ 코리아크레딧뷰로 정보

신용개설과 연체정보 등의 부정적 정보가 대부분인 은행연합회 및 여타 신용정보회사와는 차별적으로 코리아크레딧뷰로(이하 'KCB')가 회원사⁶⁾로부터 수집하는 정보로 대출실적 및 카드사용정보, 신상정보 등의 상세한 정보를 포함하고 있다. 본 분석에서는 KCB가 관리하는 3,401개의 요약정보⁷⁾ 중 KCB회원사에 의해 대표적으로 활용되는 주요 변수 295개를 활용하였다.

6) 2010.5 현재 KCB 회원사는 112개사이며, 업권별로는 은행업권 14개, 카드업권 5개, 캐피탈업권 22개, 저축은행업권 42개, 보험업권 13개, 증권업권 3개, 기타 13개로 각각 구성되어 있음.

7) KCB가 수집하는 정보를 회원사의 여신의사결정과 고객신용평가 등에 활용할 수 있도록 요약된 항목의 형태로 제공하는 서비스로, CPS(Credit profile service)라 함.

<표 2> 정보출처별 주요 활용 변수 및 개수

유형 구분		주요 변수	총 개수	활용 개수
보험사 고유정보		위험등급, 성별, 가입연령, 현재 보험계약 보유건수, 과거 사고 경험유무	5개	5개
KFB 정보	대출정보	대출일자, 약정금액 등	137개	22개
	보증정보	보증일자, 보증금액 등	56개	3개
	개설정보	신용카드 및 체크카드 개설일자/해지일자 등	35개	9개
	신용도 판단정보 및 공공정보 ⁸⁾	3개월 이상 연체정보, 금융질서문란정보, 공공정보 등	106개	20개
	소계		334개	54개
KCB 정보	대출정보 ⁹⁾	대출건수, 대출일자, 약정금액/잔액, 상환원금/이자 등	770개	47개
	보증정보	연대보증 및 지급보증 건수, 금액 등	141개	4개
	카드정보 ¹⁰⁾	카드한도, 일시불/할부/현금서비스 이용금액, 청구/결제금액 등	783개	104개
	연체정보	5영업일 이상/1만원 이상 연체정보, 대지급정보 등	1,383개	112개
	신상정보	연소득, 직업직위, 주거관련사항 등	50개	13개
	조회정보	조회건수 등	274개	15개
	소계		3,401개	295개
합계			3,740개	354개

8) 신용도 판단정보는 연체 등 정보(3개월 이상 연체, 대위변제, 대지급, 부도, 관련인 정보), 금융질서 문란정보 및 관련인 정보를 의미하며, 공공정보는 재판/결정 정보, 체납정보 등 공공기관 보유 정보를 의미함.

9) 대출업체별 개인합산정보가 관리되는 KFB와는 달리, 개인의 개별대출건에 대한 개설, 실적 및 상환이력 등의 정보가 수집되며, 거래형태(분할상환/일시상환/한도대출 등) 및 자금용도(주택자금/생활안정자금 등), 신용/담보/보증인 여부 등의 상세한 정보가 공유됨.

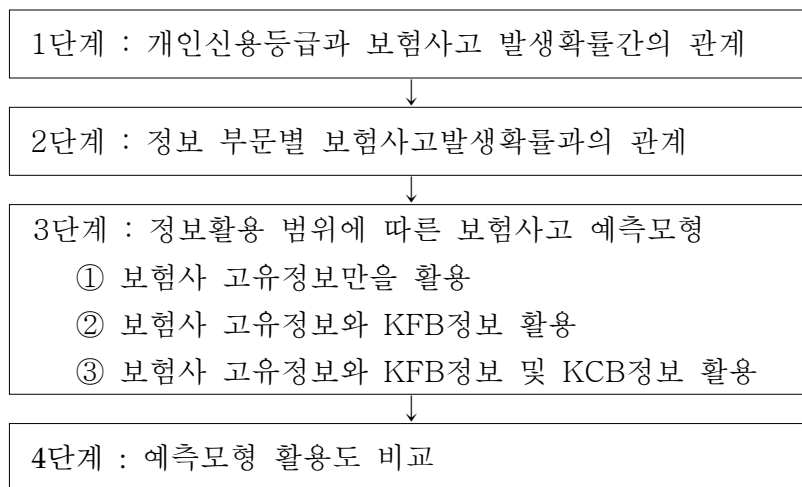
10) 카드개설일 정보만 관리되는 KFB와는 달리, 개설카드의 한도 및 이용실적, 청구 및 결제정보 등의 상세한 정보가 공유됨.

2. 분석 방법론

(1) 분석 개요

개인신용정보와 생명보험사고와의 상관관계 분석을 위해 다음의 <그림 2>에서와 같이 단계별 분석을 실시하였다.

<그림 2> 단계별 분석 프로세스



먼저 CB사에서 제공하는 개인신용등급과 보험사고발생확률간의 상관관계를 개략적으로 살펴본 뒤, 정보 부문별 분석대상변수와 보험사고발생확률 간의 관계분석을 통해 사고율 예측에 유의미한 변수의 패턴을 찾아내고 이를 바탕으로 1차적인 변수 선정하였다. 다음으로 이렇게 선정된 변수를 활용하여 보험사고 예측모형을 도출하되, 정보 활용의 단계적 확장에 따른 보험사고 예측에의 기여도를 각각 분석하였다. 마지막으로 3개의 서로 다른 예측모형의 활용도 비교를 통해 보험사고 예측에 있어 개인신용정보 활용효과를 분석하였다.

(2) 개별변수 분석

각 변수별 보험사고 발생확률과의 상관관계 분석을 위해 information value¹¹⁾라는 개념을 도입하였다. 여기서 information value 분석이란 각 설명변수가 목표변수를 설명하는데 어떠한 pattern과 어느 정도의 power를 가지고 있는지를 알아보는데 주요하게 쓰이는 분석으로, information value의 값이 클수록 예측력이 높다는 것을 의미한다. 본 분석에서는 편의상 information value가 0.002이상인 항목을 1차

11) $information\ value = \sum_{i=1}^k \{(\%nontarget_i - \%target_i) * \ln(\%nontarget_i / \%nontarget_i)\}$

적으로 선정하여 사고예측모형 도출에 활용하였다.¹²⁾

(3) 예측모형 도출

보험사고 예측모형 도출을 위해 로지스틱 회귀모형(logistic regression)을 채택하였다. 로지스틱 선형모형은 반응변수가 이항변수일 때 자주 사용하는 일반화 선형모형의 하나이다. 즉, Y 가 0 또는 1(즉, 미사고 또는 사고) 중 하나로 결정되는 이항변수이고 이를 설명하기 위한 k 개의 설명변수를 $X=(x_1, x_2, \dots, x_k)$ 로 표현할 때 예측하고자 하는 사고확률 $p(X) = \text{prob}(Y=1|X)$ 을 로지스틱 회귀모형으로 표현하면 다음의 식 (1)과 같다.

$$\text{logit}[p(X)] = \log\left(\frac{p(X)}{1-p(X)}\right) = \alpha + \beta'X^{13)} \quad (1)$$

위 식 (1)의 사고확률에 대해서 다음과 같이 나타낼 수 있다.

$$p(X) = \frac{\exp(\alpha + \beta'X)}{1 + \exp(\alpha + \beta'X)} \quad (2)$$

로지스틱의 회귀계수의 추정방법은 MLE(Maximum likelihood estimation)를 사용하였으며, 이를 위해 SAS package에서 제공하는 procedure를 활용하였다. 또한 보험사고를 설명하기 위한 설명변수의 선택방법으로는 stepwise selection¹⁴⁾을 이용하였다.

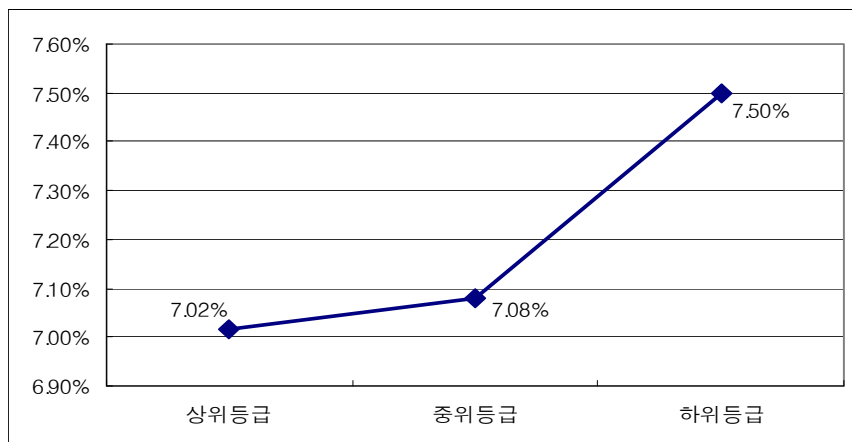
12) 변수별 information value 산출은 Infocentricity사의 XENO(모델링 패키지)를 이용하였음.
13) 이때 β 는 $(\beta_1, \beta_2, \dots, \beta_k)$ 로 구성된 회귀계수의 벡터표현임.
14) stepwise selection은 통계적으로 유의하며 카이제곱 통계량이 가장 큰(즉, Y 값 예측에의 기여도가 가장 큰) 변수를 선택해 나가는 동시에, 기 선택된 변수군 중 통계적 유의성을 상실한 변수를 제거해 나가는 과정을 반복해 나가며 최적 설명변수군을 찾아나가는 방법론임. stepwise selection은 더 이상 유의한 변수가 모형에 들어오거나 제거되지 않을 때 선택을 종료하게 됨.

III. 분석 결과

1. 개인신용등급과 보험사고와의 관계

먼저 분석의 단순화를 위해 10단계의 개인신용등급을 3개(상중하)의 등급수준으로 재구성하였다¹⁵⁾. <그림 3>에서 보듯이 개인신용등급 수준이 낮을수록 보험사고율이 상대적으로 확연히 높게 나타났다.

<그림 3> 개인신용등급 수준에 따른 보험사고율



또한 개인신용등급과 보험사고율의 상관관계에 대한 통계가설검정을 위해 Cochran-Armitage trend test¹⁶⁾를 실시하였다. 테스트 결과에 의하면 유의수준 1%내에서 귀무가설이 기각되었는데, 이는 유의수준 1%내에서 개인신용등급이 낮아지면 보험사고율이 높아진다는 것을 의미한다.

2. 정보부문별 보험사고와의 상관관계

본 절에서는 <표 2>의 각 정보출처별 각 변수에 대해 보험사고와의 상관관계를 분석하였다. 이를 통해 보험사고를 예측하는 데에 유의미한 정보는 부문별로 어떠한 것들이 있는지 살펴보고, 이를 바탕으로 보험사고 예측모형 도출에 활용할 설명 변수를 1차적으로 선정하였다.¹⁷⁾

15) 개인신용등급 중 1~3등급을 상위등급, 4~6등급을 중위등급, 7~10등급을 하위등급으로 재구성하였다.

16) Cochran-Armitage trend test는 순서형 설명변수의 수준에 따른 이항형 종속변수의 추세를 판단하기 위한 대표적인 test 방법론임.

17) 각 정보출처별 주요변수의 분석결과는 <부록>을 참조하기 바람.

(1) 보험사 고유정보

생명보험사가 현재 언더라이팅 시 고려하는 대표적인 변수는 성별, 가입연령, 위험등급, 현재 정상유지중인 보험건수¹⁸⁾, 과거 지급보험 건수 등 총 5개이다. 각 변수에 대해 사고율 및 information value 분석을 실시한 결과, 5개 변수 모두 information value 기준을 만족하여 사고율과 높은 상관관계를 지닌 것으로 나타났다(<부록>의 <표 A-1>~<표 A-5>). 따라서 위의 5개 변수는 언더라이팅 시 반드시 고려되어야 하는 정보임을 확인하였다.

(2) KFB 정보

분석에 활용된 KFB 정보변수는 총 54개이며, 이중 information value 기준으로 보험사고율을 예측하는 데에 유의하다고 판단되어 1차적으로 선택된 변수는 총 27개였다.

먼저 KFB 대출정보와 관련하여 업권별 대출건수 및 금액관련 6개 변수가 선택된 반면, 사고와의 상관관계는 대출의 업권에 따라 그 패턴이 다르게 나타났다(<부록>의 <표 A-6> 및 <표 A-7>). 예컨대, 주로 은행업권에서 대출실행규모가 큰 고객의 경우에는 사고율이 낮게 나타난 반면에 상대적으로 카드론 및 제2금융권에서의 대출의존도가 높은 고객의 경우에는 사고율이 높게 나타났는데, 이와 같은 현상은 업권별 고객의 특징이 반영된 것으로 볼 수 있다. 즉, 카드론이나 제2금융권에서 대출을 실행하는 고객은 주로 은행권에서 대출을 거절당했거나 이미 일정수준 이상의 대출 보유한 상태에서 추가적인 대출 또는 급전이 필요하여 고금리의 이자 부담을 안은 채 대출을 실행한 경우가 대부분으로 은행대출고객에 비해 평균적으로 신용등급이 낮은 고객으로 간주할 수 있다. 분석의 결과에서 보듯이 이러한 고객군이 실제로 보험사고율도 상대적으로 높게 나타났다.

KFB 개설관련 변수 9개는 모두 information value 기준을 만족하여 1차적으로 선택되었는데, 주로 신용거래를 오랜 기간 한 고객인 경우 사고율도 낮게 나타났다(<부록>의 <표 A-8>).

KFB 신용도 판단정보 및 공공정보와 관련한 20개 변수 중에는 information value 기준으로 총 12개가 선택되었는데, 관련 정보를 많이 보유한 고객일수록 사고율이 높게 나타났다(<부록>의 <표 A-9>). 이는 연체나 체납 등 신용거래 상황에 문제를 가진 고객이 보험사고율도 높게 나타남을 의미하는 것으로, 금융거래에서의 부실이 보험사고와 연관성이 있음을 알 수 있다.

마지막으로 KFB 보증정보관련 변수는 모두 사고율과의 분석결과에서 information value 기준을 만족하지 못하여 변수로 채택되지 못했다.

18) 해약이나 해지, 감소, 실효 등이 아닌 정상적인 보험 보유 상태를 의미함.

(3) KCB 정보

분석에 활용된 KCB 정보변수는 총 295개이며, 이중 information value 기준으로 보험사고율을 예측하는 데에 유의하다고 판단되어 1차적으로 선택된 변수는 총 154개였다.

먼저 KCB 대출정보관련 변수 47개 중 information value 기준으로 선택된 변수는 총 19개였다. KFB 대출정보에서와 같이 대출업권에 따른 특징이 유사하게 나타났으며, 추가적인 특징 중 하나로 담보대출 규모가 클수록 사고율이 낮게 나타났다(<부록>의 <표 A-10>).

KCB 보증정보관련 변수 4개 중 information value 기준으로 2개의 변수가 1차적으로 선택되었는데, 지급보증건수가 많을수록 보험사고율이 높게 나타났다(<부록>의 <표 A-11>). 즉, 신용거래 또는 계약 등의 목적으로 보증보험사의 보증을 받은 고객인 경우 그렇지 않은 고객에 비해 사고율이 높게 나타났으며, 이는 주로 저신용/고위험인 고객이 보증보험사의 지급보증을 받게 되는 것에 기인한다.

KCB 카드실적정보관련 변수 104개 중 2개를 제외한 102개의 변수가 모두 information value 기준을 만족하였는데, 특히 신용카드 사용패턴과 관련한 정보가 보험사고와 밀접한 관계가 있는 것으로 나타났다(<부록>의 <표 A-12> 및 <표 A-13>). 즉, 할부나 현금서비스에 비해 일시불의 사용비중이 높은 고객일수록 사고율이 낮게 나타난 반면에 현금서비스의 한도소진율이 높은 고객일수록 사고율이 높게 나타났다. 또한 고금리인 신용카드사의 현금서비스를 한도에 가까이 사용하는 고객은 주로 급전이 필요하고 신용상태가 좋지 않은 고객으로, 이러한 고객은 보험사고율도 높게 나타났다.

KCB 연체정보관련 변수 112개 중 information value 기준으로 선택된 변수는 총 22개였다. 금융기관에 장기간 연체가 있었거나 대지급 경험이 많을수록 보험 사고율이 높은 것으로 나타났다((<부록>의 <표 A-14> 및 <표 A-15>). 이는 지급능력이 부족하여 대출금이나 카드의 연체가 장기로 이어지거나 보증보험사로부터 대지급을 받은 사람은 이후 보험사고율도 높게 나타남을 의미한다.

KCB 신상정보관련 변수 13개 중 information value 기준으로 선택된 변수는 총 9개였으며, 대표적 특징으로는 개인의 연소득이 높을수록 보험사고율이 낮게 나타났다(<부록>의 <표 A-16> 참조).

마지막으로 KCB 조회정보관련 변수 15개 중 information value 기준을 만족하는 변수는 없었다. 여기서 신용조회관련 정보는 사회적 이슈로 인해 신용평가 시에도 주로 반영되지 않는 민감한 정보로, 보험사고율 분석 대상변수로는 적절치 않을 것으로 판단된다.

3. 정보활용 범위에 따른 보험사고 예측모형

본 절에서는 앞에서 언급한 3단계 분석절차에 따라 보험사고 예측모형을 도출한 후, 최종적으로 모형에 선택된 변수의 조합과 변수별 특성을 살펴보았다.

(1) 모형1 : 보험사 고유정보만을 활용

먼저 A생명보험사 내부의 고유정보만을 활용하여 보험사고 예측모형을 도출한 결과는 <표 3>과 같다. 보험사 고유정보의 경우 예상대로 5개의 모든 변수가 매우 유의한 수준으로 사고예측모형에 포함되었다.

<표 3> 모형1 분석결과

정보출처	변수	beta	방향성	p-value ¹⁹⁾
보험사 고유정보	성별(남:0, 여:1)	0.1239	+	<.0001
	가입연령	0.1816	+	<.0001
	위험등급	0.0327	+	0.001
	현재 정상 유지중인 보험 총건수	0.2760	-	<.0001
	과거 총 지급건수	0.2752	+	<.0001

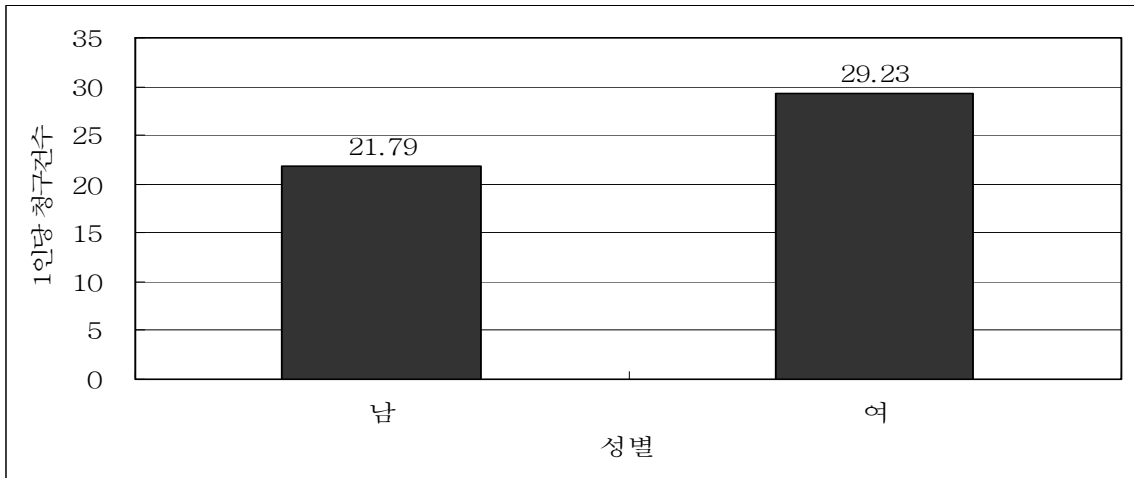
각 변수에 대해 세부적으로 사고율과의 관계 pattern을 살펴보면, 먼저 성별의 경우 여성의 사고율이 남성의 사고율보다 높게 나타났다. <그림 4>에서와 같이 실제 건강보험심사평가원 통계자료²⁰⁾에서도 여성의 보험청구가 남성보다 상대적으로 많은 것으로 나타나는데, 이는 여성의 경우 30/40대에 출산전후 부인과 질환에 대한 보험수요가 많기 때문인 것으로 추측된다.

가입연령 분석결과 연령대가 높아질수록 사고율이 높아지는 것으로 나타났는데, 이와 같은 추세 역시 건강보험심사평가원 통계자료를 통해 객관적으로 확인이 가능하다(<그림 5>).

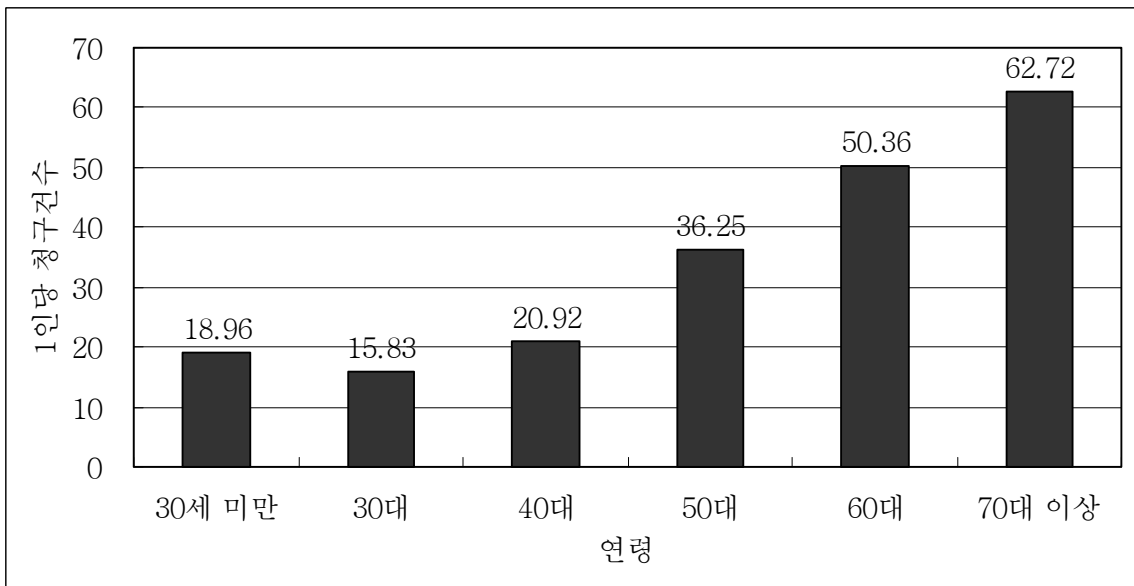
19) 선택된 변수에 의해 도출된 logistic regression 모형의 각 변수별 p-value값으로, 그 값이 작을 수록 사고율 예측에 유의한 변수임(유의수준 α 는 10%로 설정하였음).

20) 2009년 진료비 통계지표 기준임.

<그림 4> 성별에 따른 1인당 청구건수



<그림 5> 연령대별 1인당 청구건수



위험등급 또한 모형에 유의한 변수로 선택되었다. 생명보험사에서 사용하는 위험 등급은 주로 보험개발원에서 직업군별 재해사고분석을 통해 제시되는 자료와 보험사 내부 분석자료를 기초로 하여 정해지는 기준으로, 최저위험등급의 사고율을 제외하면 전반적으로 내부 위험등급이 높아짐에 따라 사고율이 높게 나타났다.

과거 보험실적과 관련하여 현재 정상유지 중인 보험이 많을수록 사고율이 현저히 감소하는 것으로 나타났으며, 특히 5개 분석대상 변수 중 가장 뛰어난 사고 변별력을 기록하였다. 이는 이미 해당 보험사의 기존 고객으로서의 실적을 보유하고 loyalty를 지닌 고객일수록 신규보험가입에 따른 사고발생률도 낮다는 것을 보여주고 있다.

마지막으로 청약시점 이전 과거에 이미 보험금을 지급받은 경험이 있는 고객일수

록 본 보험계약에 대한 사고발생률도 높게 나타났는데, 이는 과거 잦은 사고경험으로 보험 상품의 혜택을 체험한 고객들이 이후로도 계속적으로 보험을 납용 또는 악용하는 사례가 적지 않음을 보여주고 있다.

결론적으로 보험사에서 언더라이팅 시 고려하는 5개 변수는 서로간의 공선성(collinearity) 문제없이 모두 보험사고 예측에 유의미함이 통계적으로 증명되었다.

(2) 모형2 : 보험사 고유정보와 KFB정보 활용

다음으로 기존의 생명보험사 고유정보에 KFB 정보를 추가하여 사고예측모형을 도출하였다. 분석에 활용한 주요 KFB정보는 <표 2>에 제시된 총 54개 변수 중 information value 기준을 만족한 27개이며, 최종적으로 도출된 분석결과는 <표 4>와 같다. 즉, 로지스틱 stepwise selection에 따라 최종모형에서는 총 7개 변수가 선택되었으며, 모형1의 5개 보험사 내부정보의 유의성과 방향성이 그대로 유지된 채 KFB정보 2개가 추가적으로 선택되었다.

<표 4> 모형2 분석결과

정보출처	변수	beta	방향성	p-value
보험사 고유정보	위험등급	0.0288	+	0.0007
	성별(남:0, 여:1)	0.1181	+	<.0001
	가입연령	0.1892	+	<.0001
	현재 정상 유지중인 보험 총건수	0.2686	-	<.0001
	과거 총 지급건수	0.2790	+	<.0001
KFB정보	은행업권 대출총금액(KFB)	0.1597	-	0.0270
	최초 신용개설일로부터의 기간(KFB)	0.1017	-	<.0001

추가적으로 선택된 KFB 정보부문 변수에 대해 사고율과의 상관관계를 세부적으로 살펴보면, 먼저 은행업권 대출총금액의 경우에는 은행업권에 대출총금액이 높은, 즉 대출실행의 규모가 큰 상위 고객군의 사고율은 그렇지 않은 고객군에 비해 사고율이 상대적으로 낮게 나타났다. 이는 은행에서의 대출심사 시 정밀한 신용평가 및 소득, 재직확인, 재산세 등 각종자료증빙을 통해 고객의 상환능력 및 의지를 면밀히 검토한 후, 이를 바탕으로 대출승인여부와 승인금액 수준이 결정되기 때문인 것으로 보인다. 또한 이는 실제로 은행권 개인대출고객의 신용등급과 연체율이 타업권에 비해 가장 우량한 현상과 일맥상통하는 것으로, 대출의 상환의지와 능력이 일정 수준 이상 확보되는 고객의 경우에는 보험사고율도 상대적으로 낮음을 의미한다.

다음으로 최초 신용개설일로부터의 기간 변수가 추가적으로 선택되었는데, 최초 신용개설일로부터의 기간이 길수록, 즉 신용거래를 오래한 고객일수록 보험사고율이 낮게 나타났다. 이는 최근에 신용거래를 하기 시작한 고객들(주로 사회초년생)에 비해 오랜 기간 신용거래를 해온 고객들(주로 안정적인 직업을 유지해온 고객들)이 상대적으로 보험사고율이 낮음을 의미한다.

(3) 모형3 : 보험사 고유정보와 KFB정보 및 KCB정보 활용

마지막으로 보험사 고유정보와 27개의 KFB정보 외에 KCB정보를 추가적으로 활용하여 예측모형을 도출하였다. 분석에 활용한 주요 KFB정보는 <표 2>에 제시된 총 295개 변수 중 information value기준을 만족한 154개이며, 최종 도출된 분석 결과는 <표 5>와 같다.

<표 5> 모형3 분석결과

정보출처	변수	beta	방향성	p-value
보험사 고유정보	위험등급	0.0290	+	0.0007
	성별(남:0, 여:1)	0.0935	+	0.0028
	가입연령	0.1912	+	<.0001
	현재 정상 유지중인 보험 총건수	0.2661	-	<.0001
	과거 총 지급건수	0.2756	+	<.0001
KCB정보	담보대출총금액	0.2452	-	0.0012
	최근1년 신용카드 일시불 이용금액비율	0.0304	-	0.0691
	대지급 총건수	0.1962	+	0.0754
	KCB 결정연소득	0.0595	-	0.0057

분석 결과에서 나타났듯이 로지스틱 stepwise selection에 의해 최종적으로 모형에 포함된 변수는 총 9개로, 모형1의 보험사 내부정보 변수(5개)의 유의성은 그대로 유지된 채 4개의 KCB정보가 추가적으로 선택되었다. 여기서 특이한 점은 기존의 모형2에서 선택되었던 KFB정보는 모두 탈락되었다는 점이며, 이는 보험사고 예측에 있어 금융거래실적 위주의 상세한 정보(긍정적 정보)의 기여도가 상대적으로 높음을 간접적으로 보여주는 사례이다. 이는 금융권 연체예측에 있어서 연체정보 등 부정적 정보의 역할이 상대적으로 높은 특징과는 상당한 차이가 있음을 의미하는 것으로, 보험사고가 신용정보와는 분명한 상관관계가 존재하지만 기존의 금융연체 예측과는 차별적인 특징이 존재함을 나타내는 것이기도 하다.

보험사 정보 이외에 선택된 KCB관련 변수와 사고율과의 상관관계를 살펴보면, 먼저 담보대출금액이 높은 경우 사고율이 상당히 낮게 나타났는데, 담보대출금액이 높다는 것은 보유한 담보물(주로 아파트 등 부동산)의 시가가 높다는 것을 의미하며, 따라서 본인소유의 고가 부동산 자산을 가진 경우 이후 보험 사고율이 낮게 나타날 확률이 높다고 추론할 수 있다.

신용카드관련 정보로는 최근1년 신용카드 일시불 이용금액비율이 선택되었는데, 신용카드 이용 시 일시불 사용비중이 높은 사람일수록 보험사고율이 낮은 것으로 나타났다. 즉, 신용카드 이용 시 할부나 현금서비스를 주로 사용하는 사람에 비해 일시불을 사용하는 사람이 상대적으로 보험사고율이 낮는데, 이는 지불능력이 우수한 사람의 경우 할부나 현금서비스를 주로 이용하는 사람에 비해 보험사고를 일으킬 확률이 상대적으로 낮다고 추정할 수 있다.

연체관련 정보로는 대지급 총건수가 선택되었는데, 보증보험의 대지급을 받은 경우 이후 보험사고율도 높게 나타났다. 이는 고객의 신용거래에서의 상환능력과 보험사고 간에 밀접한 연관성이 있음을 보여주고 있다.

마지막으로 신상정보 중에서는 KCB 결정연소득이 선택되었는데, 소득금액이 높을수록 사고율이 낮게 나타났다. 특히 4천 5백만원 이상 최상위 집단의 사고율이 현저히 낮았다. 모형3에 선택된 KCB정보의 분석결과를 종합하면, 일정규모 이상의 자산을 보유하고 지불능력이 우수한 고객일수록 이후 보험사고율이 낮게 나타남을 알 수 있다.

4. 예측모형 활용도 비교

앞에서 기술한 3개의 보험사고 예측모형 간의 활용도 비교 분석을 위해 분석대상을 개인신용등급이 하위 10%에 속하는 고위험군 고객으로 한정하였다. 이는 실제로 몇몇 생명보험사에서 개인신용등급이 낮은 10% 이하의 고객을 대상으로 보험인수조건을 강화하기 위해 개인신용정보를 암묵적으로 활용하고 있는 상황을 충분히 감안하고자 했기 때문이다.

아래 <표 6>에서 보듯이 사고개선율의 효과가 모형1<모형2<모형3 순이었으며, 이는 하위 2%, 5%, 10%군에 대해 모두 일관되게 나타났다. 즉, 고위험 고객군으로 규정할 수 있는 하위 2%, 5%, 10%의 고객에 대해 보험인수조건을 강화할 경우 이로 인한 전체 사고율 개선효과가 모형1<모형2<모형3의 순으로 나타났으며, 이는 모형3이 고위험군을 선별해내는 데 상대적으로 우수함을 의미한다. 예컨대, 모형3에서 하위 10% 고위험군 고객에 대해 보험인수를 강화할 경우 생명보험사 입장에서 9.694%의 사고 개선율 효과를 기대할 수 있다.

<표 6> 고위험군에 대한 예측모형별 사고 개선율 비교

고위험군 정의	모형구분	고위험군 사고율	비위험군 ²¹⁾ 사고율	사고개선율 ²²⁾
하위2%	모형1	12.766%	7.114%	1.462%
	모형2	13.489%	7.095%	1.728%
	모형3	13.572%	7.085%	1.863%
하위5%	모형1	11.817%	6.968%	3.493%
	모형2	11.823%	6.947%	3.778%
	모형3	12.303%	6.941%	3.867%
하위10%	모형1	11.189%	6.583%	8.823%
	모형2	11.412%	6.578%	8.892%
	모형3	11.619%	6.520%	9.694%

결론적으로 언더라이팅 시 보험사 내부정보 외에 개인신용정보를 활용하여 보험사고를 예측하고 이를 바탕으로 고위험군으로 선정된 고객에 대해 보험인수를 아예 거절하거나 또는 인수조건을 일부 제한할 경우 생명보험사 입장에서는 전반적인 사고율 감소에 따른 경영개선 효과를 기대할 수 있을 것으로 보인다. 특히 이러한 효과는 부정적 신용정보 뿐 아니라 긍정적 신용정보까지 모두를 활용하였을 때 가장 뛰어난 것으로 나타났다.

5. 분석의 한계

먼저 본 분석에 사용된 자료는 A생명보험사의 과거 7개월간의 특정 보험상품에 대한 보험계약 신청과 연관된 정보(총 109,689건)로, 이는 계량분석을 진행하기에 무리 없는 규모이긴 하나, 보다 안정적이고 신뢰성 있는 분석결과를 위해서는 적어도 1년 이상의 기간선정과 더불어 이에 따른 충분한 분석샘플 확보가 필요할 것으로 판단된다.

다음으로 본 분석에서는 보험사고 빈도(frequency)를 예측의 목적으로 설정하였으나, 실제 상황에서는 보험금 지급금액과 관련된 사고의 심도(severity)가 더욱 중요한 평가요인으로 고려된다. 특히 사고의 빈도와 심도는 반드시 동일한 pattern을 따르지 않기 때문에 분석을 위한 통계모형이 본 분석에서 사용한 판별모형보다는 심도예측모형이 오히려 더 적합한 것으로 판단되며, 이러한 부분의 보완은 추후 연구과제로 남겨두고자 한다.

21) 비위험군이란 고위험군을 제외한 나머지 보험고객을 의미함.

22) 사고개선율 = (전체 사고율 - 비위험군 사고율)/전체사고율

IV. 결론

본 연구에서는 개인신용정보와 특정 생명보험상품 사고율간의 상관관계를 분석하였다. 현재 국내에서는 보험계약 승인 심사시 개인의 신용정보를 조회하여 의사결정에 활용하는 데 대해 찬반의 논란이 있으나, 본 연구에서는 정책적 판단을 배제한 채 순수 데이터 분석을 통해 그 상관관계를 규명하는 데에 초점을 맞추었다. 이를 위해 국내 A 생명보험사의 2007년 1월에서 7월까지 상해보험과 건강보험 청약자를 대상으로 단계적 분석을 실시하였다.

먼저 개인신용정보와 생명보험 사고율과의 관계를 개략적으로 살펴보기 위해 개인신용등급과 사고율과의 상관관계를 분석해 보았다. 분석의 단순화를 위해 10등급으로 구분되는 개인신용등급을 상·중·하로 재구성하여 분석한 결과, 개인신용등급 수준이 낮을수록 보험사고율이 높아진다는 사실이 통계적으로 검증되었다.

다음으로 정보 출처별 주요항목 354개에 대해 보험사고율과의 상관관계를 분석하였다. 그 결과, 부문별로 다소의 차이는 있었지만 대체적으로 사고율과 밀접한 상관관계를 가진 것으로 나타났다. 또한 이와 같은 개별항목 분석을 통해 사고율 예측에 통계적으로 유의한 186개의 변수를 1차적으로 선별하였다.

이렇게 선정된 변수를 활용하여 보험사고모형을 도출하였는데, 정보활용 범위에 따라 3개의 모형으로 각각 구분하였다. 즉, 보험사 고유정보만을 활용한 모형1, 보험사 고유정보와 KFB정보를 활용한 모형2, 보험사 고유정보와 KFB정보 및 KCB정보를 활용한 모형3을 로지스틱 회귀모형을 활용하여 각각 도출하였다.

마지막으로 3개 예측모형의 비교를 통해 보험사고 예측에 있어 개인신용정보 활용효과를 비교 분석하였다. 분석결과에 의하면, 사고개선율에 의한 효과가 모형1 < 모형2 < 모형3의 순으로 나타났다. 즉, 보험사 고유정보만을 고려한 경우보다는 KFB정보를 추가했을 때 모형의 사고예측력이 좀 더 우수하였으며, 보험사 고유정보와 KFB정보뿐만 아니라 KCB정보를 모두 감안한 모형의 예측효과가 가장 높게 나타났다. 이는 생명보험사가 모형을 통해 고위험군을 대상으로 보험인수조건을 강화하고자 할 경우, 보험사 고유정보만 고려한 경우보다는 외부의 개인신용정보를 활용했을 때 보험사고율 개선효과 전반적으로 향상될 수 있음을 의미한다. 결론적으로 생명보험사가 언더라이팅 시 개인신용정보를 적극적으로 활용할 경우 전반적인 사고율 감소에 따른 경영개선 효과를 기대할 수 있을 것으로 보인다.

<참고문헌>

정재욱·손승호(2009), “개인신용정보와 개인용자동차보험 사고발생확률 및 손해율 간의 상관관계 연구”, 리스크관리연구 제20권, 제2호

Fair Trade Commission(2007), “*Credit-based Insurance Score: Impacts on Consumers of Automobile Insurance*”, United States Congress

Monaghan, James E.(2002), “*The Impact of Personal Credit History on Loss Performance in Personal Lines*”, *The Use of Credit History for Personal Lines of Insurance: Report to the NAIC*, American Academy of Actuaries

S. Kulback(1959), “*Information Theory and Statistics*”, John Wiley & Sons

Wu, Cheng-Sheng Peter and James Guszczka(2003), “*Does Credit Score Really Explain Insurance Losses? : Multivariate Analysis from Data Mining Point of View*”, Deloitte & Touche LLP

<부 록>

<표 A-1> 성별별 사고율 분석

성별	미사고%	사고%	전체%	사고율	IV ²³⁾
여성	64.65%	65.79%	64.73%	7.34%	0.000
남성	35.35%	34.21%	35.27%	7.01%	0.000
합계	100.00%	100.00%	100.00%	7.22%	0.001

<표 A-2> 가입연령대별 사고율 분석

가입연령	미사고%	사고%	전체%	사고율	IV
30세 미만	16.51%	14.04%	16.33%	6.21%	0.004
30세 ~ 39세	25.58%	19.53%	25.15%	5.61%	0.016
40세 ~ 49세	34.35%	32.34%	34.20%	6.83%	0.001
50세 ~ 59세	21.91%	30.96%	22.57%	9.91%	0.031
60세 이상	1.65%	3.13%	1.76%	12.86%	0.010
합계	100.00%	100.00%	100.00%	7.22%	0.062

<표 A-3> 위험등급별 사고율 분석

위험등급	미사고%	사고%	전체%	사고율	IV
위험등급4	18.66%	20.02%	18.76%	7.71%	0.001
위험등급3	11.28%	13.37%	11.43%	8.44%	0.004
위험등급2	2.76%	3.22%	2.80%	8.32%	0.001
위험등급1	0.53%	0.50%	0.53%	6.87%	0.000
비위험직	66.77%	62.89%	66.49%	6.83%	0.002
합계	100.00%	100.00%	100.00%	7.22%	0.008

<표 A-4> 현재 정상유지중인 보험건수별 사고율 분석

현재 정상유지중인 보험건수	미사고%	사고%	전체%	사고율	IV
없음	56.02%	65.48%	56.70%	8.34%	0.015
1건	19.32%	18.29%	19.25%	6.86%	0.001
2건	9.49%	7.13%	9.32%	5.53%	0.007
3건	5.57%	3.71%	5.44%	4.93%	0.008
4건 이상	9.59%	5.39%	9.29%	4.19%	0.024
합계	100.00%	100.00%	100.00%	7.22%	0.054

23) IV는 Information vale를 의미함.

<표 A-5> 과거 총 지급건수별 사고율 분석

과거 총 지급건수	미사고%	사고%	전체%	사고율	IV
없음	84.60%	83.02%	84.48%	7.10%	0.000
1건 ~ 2건	9.29%	9.98%	9.34%	7.72%	0.001
3건 이상	6.11%	6.99%	6.18%	8.18%	0.001
합계	100.00%	100.00%	100.00%	7.22%	0.002

<표 A-6> 은행업권 대출총금액대별 사고율 분석

은행업권 대출총금액 (KFB)	미사고%	사고%	전체%	사고율	IV
7천만원 미만	95.56%	96.38%	95.62%	7.28%	0.000
7천만원 이상	4.44%	3.62%	4.38%	5.97%	0.002
합계	100.00%	100.00%	100.00%	7.22%	0.002

<표 A-7> 제2금융권(카드론 포함) 대출금액 비율별²⁴⁾ 사고율 분석

제2금융권(카드론 포함)대 출금액 비율	미사고%	사고%	전체%	사고율	IV
30% 미만	24.60%	26.55%	24.74%	7.73%	0.001
30% 이상 80% 미만 또는 대출없음	51.56%	50.59%	51.49%	7.08%	0.000
80% 이상	23.85%	22.86%	23.78%	6.93%	0.000
합계	100.00%	100.00%	100.00%	7.20%	0.002

<표 A-8> 최초 신용개설일로부터의 기간(KFB)별 사고율 분석

최초 신용개설일로부터 의 기간(KFB)	미사고%	사고%	전체%	사고율	IV
120일 미만	28.57%	30.98%	28.75%	7.78%	0.002
120일 이상~ 10년 미만	50.96%	51.54%	51.00%	7.30%	0.000
10년 이상	20.47%	17.48%	20.25%	6.23%	0.005
합계	100.00%	100.00%	100.00%	7.22%	0.007

24) 총 대출금액 중 카드론 및 제2금융권(할부금융/저축은행/보험/조합/새마을금고/신협에서 실행한 대출금액의 비율임.

〈표 A-9〉 KFB 신용도 판단정보 및 공공정보 총건수별 사고율 분석

KFB 신용도 판단정보 및 공공정보 총건수	미사고%	사고%	전체%	사고율	IV
2건 이상	9.52%	10.65%	9.61%	7.98%	0.001
1건	6.42%	6.92%	6.46%	7.72%	0.000
없음	84.05%	82.43%	83.94%	7.08%	0.000
합계	100.00%	100.00%	100.00%	7.20%	0.002

〈표 A-10〉 담보대출총금액대별 사고율 분석

담보대출 총금액	미사고%	사고%	전체%	사고율	IV
1억 미만	97.56%	98.41%	97.63%	7.28%	0.000
1억 이상 ~ 2억 미만	1.52%	1.19%	1.50%	5.72%	0.001
2억 이상	0.91%	0.40%	0.88%	3.32%	0.004
합계	100.00%	100.00%	100.00%	7.22%	0.005

〈표 A-11〉 지급보증 총건수별 사고율 분석

지급보증 총건수	미사고%	사고%	전체%	사고율	IV
2건 이상	1.44%	2.01%	1.48%	9.75%	0.002
1건 또는 없음	98.56%	97.99%	98.52%	7.17%	0.000
합계	100.00%	100.00%	100.00%	7.20%	0.002

〈표 A-12〉 최근1년 카드일시불 이용금액비율²⁵⁾별 사고율 분석

최근1년간 신용카드 일시불 이용금액 비율	미사고%	사고%	전체%	사고율	IV
일시불 이용금액 없음 또는 총 이용금액 없음 또는 신용카드 없음	51.96%	55.29%	52.20%	7.65%	0.002
50% 미만 일시불 사용	18.79%	18.37%	18.76%	7.07%	0.000
50% 이상 일시불 사용	29.26%	26.34%	29.05%	6.55%	0.003
합계	100.00%	100.00%	100.00%	7.22%	0.005

25) 최근1년간 카드일시불 이용금액비율 = 최근1년간 (일시불 총이용금액)/(총이용금액)*100%

<표 A-13> 최근3개월 신용카드 유효현금서비스 한도소진율²⁶⁾별 사고율 분석

최근3개월내 신용카드 유효현금서비스한도소진율	미사고%	사고%	전체%	사고율	IV
90% 이상 또는 현금서비스한도 없음 또는 신용카드 없음	73.47%	75.85%	73.64%	7.42%	0.001
10% 이상 90% 미만	5.19%	4.37%	5.13%	6.13%	0.001
10% 미만	21.34%	19.78%	21.23%	6.71%	0.001
합계	100.00%	100.00%	100.00%	7.20%	0.003

<표 A-14> 최근1년내 유지연체 최장연체일수²⁷⁾별 사고율 분석

최근1년내 유지연체 최장연체일수	미사고%	사고%	전체%	사고율	IV
9개월 이상	5.94%	6.97%	6.01%	8.35%	0.002
6개월 이상 9개월 미만	2.42%	2.44%	2.42%	7.26%	0.000
2개월 이상 6개월 미만	2.22%	2.10%	2.21%	6.85%	0.000
2개월 미만 또는 무연체	89.43%	88.49%	89.36%	7.13%	0.000
합계	100.00%	100.00%	100.00%	7.20%	0.002

<표 A-15> 대지급 총건수별²⁸⁾ 사고율 분석

대지급 총건수	미사고%	사고%	전체%	사고율	IV
2건 이상	1.25%	1.74%	1.28%	9.79%	0.002
1건 이하	98.75%	98.26%	98.72%	7.19%	0.000
합계	100.00%	100.00%	100.00%	7.22%	0.002

<표 A-16> KCB 결정연소득금액대별²⁹⁾ 사고율 분석

KCB 결정연소득	미사고%	사고%	전체%	사고율	IV
2천 6백만원 미만	60.02%	60.93%	60.09%	7.32%	0.000
2천 6백만원 이상~ 3천 5백만원 미만	30.18%	30.54%	30.20%	7.30%	0.000
3천 5백만원 이상~ 4천 5백만원 미만	6.70%	6.48%	6.68%	7.00%	0.000
4천 5백만원 이상	3.10%	2.06%	3.03%	4.91%	0.004
합계	100.00%	100.00%	100.00%	7.22%	0.005

26) 현금서비스한도가 있는 신용카드의 (3개월 현금서비스 총이용금액)/(3개월 현금서비스 총한도)

27) 최근 1년 이내에 있었던 금융사 연체(대지급 포함) 중 가장 오랜 기간의 연체지속일수

28) 개인이 보증보험사를 통해 통신료/단말기, 보증서대출 등의 대지급을 받은 건수

29) KCB 결정연소득은 KCB가 보유한 등록연소득 및 소득추정보형을 통해 결정된 연소득금액임.