

Survival analysis for contract maintenance period using life insurance data

Dae Geon Yang^a · Il Do Ha^{a,1} · Geon Ho Cho^b

^aDepartment of Statistics, Pukyong National University;

^bDivision of Cosmetic Science and Technology, Industrial Quality Engineering,
Daegu Haany University

(Received September 18, 2018; Revised October 2, 2018; Accepted October 5, 2018)

Abstract

The life insurance industry is interested in various factors that influence the long-term extensions of insurance contracts such as the necessity for the advisors' long-term management of consumers, product consulting, and improvement of the investment aspects. This paper investigates important factors leading to a long-term contract that forms an important part of the life insurance industry in Korea. For this purpose we used the data of contents (i.e., data from Jan 1, 2011 to Dec 31, 2016) of the contracts of xxx insurance company. In this paper, we present how to select important variables to influence the duration of the contract maintenance via a penalized Cox's proportional hazards (PH) modelling approach using insurance life data. As the result of analysis, we found that the selected important factors were the advisor's status, the reward type 2 (annuity insurance) and tendency 4 (safety-pursuing type).

Keywords: Cox's proportional hazards model, life insurance data, penalized variable selection, survival analysis

1. 서론

1.1. 연구의 배경 및 목적

오늘날 보험회사들은 2008년 미국의 서브프라임 모기지 사태 이후, 금융권의 새로운 국제회계기준(International Financial Reporting Standards 9, 17, 19; IFRS 9, 17, 19)변경과 보험회사의 재무건전성 규제인 위험기준자기자본(risk based capital)비율 강화로 인하여 보험계약의 장기유지와 가입자의 해약으로 인한 이탈 방지와 관련된 해결방안에 많은 관심을 두고 있다. 최근 각 생명보험회사들은 경쟁력을 높이기 위한 방법으로 고객의 신규모집과 기존고객의 계약건의 고객 이탈방지를 줄임으로 보험사의 보험계약의 효력 상실로 인한 직·간접 부정적인 영향을 관리하여 보험회사 수익성을 제고하는 방안

This paper is condensed form of the first author's master thesis from the Pukyong National University, Busan, Korea. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No.NRF-2017R1E1A1A03070747).

¹Corresponding author: Department of Statistics, Pukyong National University, 45 Yongso-ro, Nam-gu, Pusan 48513, Korea. E-mail: idha1204@gmail.com

이 중요한 화두가 되고 있다. 이러한 국내·외 보험시장 환경의 변화는 보험회사로 하여금 계약건 장기 유지, 설계사와 보험 모집 중사자의 모집 수수료 그리고 가입자의 조기해지 및 보험계약 효력 상실로 인한 보험사 손실 최소화 등이 향후 보험사 존폐에 필수불가결하게 영향을 미치는 중요한 변수임을 인식하게 하였다.

이와 관련된 국내 연구로써 Kim과 Lee (2012)는 생명보험 계약 민원 데이터를 활용하여 소비자 불만족이 거시경제지표 보다 생명보험 계약의 해지에 더 영향을 미치는 유효한 변수임을 확인하였고 Choi와 Choi (2008)은 보험계약 유지율과 관련하여 실업률, 시장이자율 등의 주요 경제변수들이 생명보험 종목별 및 회사별 해약률에 영향을 미친다는 것을 분석하였다. 그러나 이는 생명보험 계약건 유지에 영향을 미치는 변수로 경제변수 뿐만 아니라 실질적으로 고객의 투자 상품에 대한 회피 수준, 상품유형, 설계사의 고객 관리 유·무 등 생명보험 계약건 내에서 계약유지에 유의미하게 영향을 미칠 수 있는 변수들을 고려하지 않고 분석한 점은 아쉬운 부분이라 하겠다.

이에 본 연구에서는 생명보험회사의 고객들의 계약건 장기 유지에 영향을 미칠 수 있는 변수들 중 실제 현장에서 계약을 담당하고 있는 설계사들의 실무경험을 토대로 중요하다고 판단되는 변수들을 선별하여 이들 중 계약건 장기 유지기간에 가장 영향을 미치는 변수의 선택 문제를 논의하고자 한다. 이러한 측면을 고려하여 본 연구는 실제 생명보험자료를 사용하여 생명보험 계약건의 장기 유지에 유의미한 영향을 미치는 변수를 선택하기 위한 생존모형으로 Cox (1972)의 비례위험모형을 활용하고자 한다. 따라서 이 모형을 적용하여 계약유지와 관련된 회귀변수의 효과를 추정하고 동시에 별점화 가능성(penalized likelihood) (Tibshirani, 1996; Fan과 Li, 2001)를 이용하여 기존의 변수선택 방법 (Tibshirani, 1997; Fan과 Li, 2002; Ha 등, 2014)을 실제 보험자료에 적용하고자 한다.

1.2. 자료의 범위 및 구성

이 연구에 사용된 생명보험자료의 수집방법 및 구성은 다음과 같다. 본 연구에 활용된 자료는 A 생명보험사의 계약건으로 여러 지점의 지점원 전체의 계약건 중 지점원별 기간 조회하여 나타나는 자료만 발취하여 2011년 1월 1일부터 2016년 12월 31일까지의 계약건을 토대로 계약서(청약서) 및 계약건의 청약서 내용만을 활용하여 분석하였다. 여기서 데이터의 형태는 이름, 나이, 주민번호 앞자리, 계약금액, 주소, 연락처, 성별, 보험종류, 투자성향, 가족사항, 등의 보험계약과 관련된 모든 정보가 포함되어 있으나, 개인정보보호법에 의거하여 개인 식별과 관련된 고유 인자에 대한 일체의 내용은 포함하지 않고 진행하였다. 연구 당시 표본 데이터는 무작위로 추출된 674건의 생명보험 계약건과 관련(청약서 및 상품) 내용을 기초로 하였다. 보험 계약건의 특성상 계약자, 피보험자, 수익자로 나누어져 있으나 보험의 계약건의 유지와 기타 변경 등은 계약자가 모든 지위를 가진다. 미성년자 (15세 이하)는 부모(친권자)가 계약자 또는 동의가 있어야만 하고, 배우자의 교차계약(상속세 등의 이유로 계약자를 서로 상대방으로 진행), 법인이 계약자가 되어 진행되는 계약도 있으므로 계약자가 계약과 관련하여 가장 중요하다. 따라서 생명보험 계약건의 계약유지와 관련하여서는 계약자를 기준으로 하여 피보험자와 수익자는 연구에 포함하지 않았다.

이들 생명보험 계약건 중 일인의 계약자가 체결한 동일 다 건의 계약건과 일인의 계약자가 다른 상품군의 다수 계약건 경우는 계약유지기간(즉 계약의 효력 상실이 발생할 때까지의 기간)이 가장 긴 생명보험 계약건만을 기준으로 ID로 구분하여 연구 진행한다. 본 연구의 목적이 계약유지와 관련된 변수들의 확인을 요하는 바, 이러한 변수가 아닌 또 다른 원인(실제 데이터 상 사망 3건, 자살 2건, 이민 1건으로 매우 적음)으로 인한 계약건의 소멸(해지, 해약, 기타 보험 효력 없음)로 인한 부분은 본 연구의 목적과 상이함으로 본 연구에 포함하지 않았다. 이상 수치를 나타내는 계약건인 거치형(일시납) 3억원 계약건, 월납입액 1억원 등의 계약건 및 ID를 제외한 총 476건의 계약 자료를 활용하여 연구하였다.

Table 1.1. Explanation for variables of the life insurance data

변수	설명
ID	476건의 계약건수의 일련번호
Time	보험계약유지기간(월)
Status	중도절단 지시변수(0: 정상, 1: 효력 상실)
Advisor	설계사 변경 유무(0: 변경, 1: 정상)
Sex	성별(0: 남성, 1: 여성)
Age	만 나이
Disease	고지사항(0: 없음, 1: 있음)
Job	직업 위험 분류표(0: 비위험직, 4: 4등급 위험직, 3: 3등급 위험직, 2: 2등급 위험직, 1: 1등급 위험직)
Type	보험 상품 군(1: 종신(변액, 비변액), 2: 연금(적격, 비적격), 3: 저축(변액 및 어린이변액보험), 4: 보장성(건강 및 단독 실비), 5: 거치형(즉시연금)) (4 → 1, 5 → 2)로 3개의 Type으로 그룹화 함
Tendency	투자성향(0: 무, 5: 위험회피형, 4: 안정추구형, 3: 중립형, 2: 적극형, 1: 위험선호형)
Corporate	법인 유무(0: 개인, 1: 법인)
Drive	운전 유무(0: 비운전, 1: 운전)

계약건별 보험효력 발생은 체결일 기준 그리고 해지 및 실효는 일반해약, 반송, 실효계약건의 효력 상실을 기준으로 각각 정리한다. 이를 토대로 계약건 유지기간(보험효력 상실 월-보험효력 발생 월)을 산출하여 계약기간을 정의한다. 연구시점 당시 또는 연구진행 중 이미 사망하였거나, 자살, 지급소멸 등의 이유로 생성된 ID(계약건)에 대해서는 본 연구의 목적과 상이함과 동시에 극히 소수의 ID임으로 자료에 포함하지 않고 연구를 진행한다.

본 자료에 대한 변수의 설명은 다음과 같고 Table 1.1로 요약된다.

- 생존시간(Time)은 계약 후 계약효력 상실까지의 시간(보험계약유지기간)을 월로 나타낸다.
- 중도절단 지시변수(Status; censoring indicator)는 생명보험 계약건의 효력 상실 유·무를 나타낸다(0: 정상, 1: 효력 상실). 여기서 “0”은 중도절단을 나타낸다.
- 설계사(Advisor)와 관련된 정보는 계약 체결 당시부터 설계사가 존속 또는 퇴사, 타사 이직 및 해촉, 기타 사유로 설계사 변경이 한 번이라도 있거나 또는 모집설계사가 현재까지 해당 계약건의 수급설계사로 존속하는 경우로 나눈다(0: 변경, 1: 정상).
- 보험계약 시 고지 사항(Disease)과 관련된 유병, 과거 질병, 및 재해로 인한 치료이력, 질병으로 인한 투약 중 등의 고지 대상(기왕력)과 관련된 정보도 유, 무로 분류한다(0: 없음, 1: 있음).
- 직업(Job)과 관련된 위험 등급은 생명보험 가입 시기 직업군과 위험정도에 따라 생명보험업에서 정해진 직업 분류코드로 위험등급을 5개 등급으로 나눈다(0: 비 위험직, 4: 4등급 위험직, 3: 3등급 위험직, 2: 2등급 위험직, 1: 1등급 위험직).
- 계약건(Type)은 보장성, 연금, 종신, 거치형 연금, 저축성 등 5개 상품 군으로 이루어져 있으나, 계약건수가 적은 거치형 연금은 즉시 연금이므로 연금 상품 군으로, 보장성보험은 생명보험회사의 실손 판매가 적은 단독 실비 또는 단순 건강보험 상품이므로 종신으로 상품군을 통합하여 3가지 상품 군으로 정리한다(1: 종신, 2: 연금, 3: 저축).
- 나이(Age)는 보험업에서 사용되는 만 나이로 표기한다.
- 성별(Sex)은 남성(0) 여성(1)로 나타낸다.

- 보험가입 금액(Cost)은 원 단위로 표기한다.
- 투자성향(Tendency)은 투자를 원치 않거나, 펀드와 같은 어떤 투자유형 및 방법으로 투자를 해도 무관하다고 판단한 계약건의 적합성 진단의 경우는 “무”로 표기되었으며 위험 성향형에서부터 비 위험 선호 형까지 5개 군으로 나눈다(0: 무, 5: 위험회피 형, 4: 안정추구 형, 3: 중립 형, 2: 적극 형, 1: 위험선호 형).
- 운전(Drive)의 유무는 운전함과 운전하지 않음으로 구분한다(0: 비 운전, 1: 운전).
- 계약자는 법인(Corporate)과 개인으로 구분한다(0: 개인, 1: 법인)

위에서 기술한 변수들중 Age와 Cost만 연속형 변수이고 나머지는 범주형 변수이다. 다만 Cost의 값은 금액의 과도한 변동 및 납입방법의 상이 (예: 월납 또는 일시납 계약) 등으로 분석에서 제외하였다.

본 논문의 구성은 다음과 같다. 2절에서는 Cox의 비례위험모형에서 벌점함수(penalty function)에 기초한 가중도 함수를 이용하여 생명보험 계약건 유지에 유의한 영향을 미치는 변수선택 방법을 간략히 리뷰한다. 여기서 우리는 세 가지 벌점화 변수선택법을 사용하여 변수선택과 추정을 동시에 수행한다. 즉 least absolute shrinkage and selection operator (LASSO) (Tibshirani 1996), smoothly clipped absolute deviation (SCAD) (Fan과 Li 2001), 그리고 hierarchical likelihood (HL) (Lee와 Oh, 2014) 변수선택법을 사용한다. 3절에서는 생명보험사의 실 계약건을 토대로 계약서 및 계약건의 청약서 내용만을 활용한 자료를 체계적으로 분석한다. 즉 기초통계분석을 먼저 실시한다. 다음으로, 카플란-마이어(Kaplan-Meier) (Kaplan과 Meier, 1958) 분석법을 통하여 보험계약 효력 상실이 발생한 기간마다 구간 계약유지에 대한 그룹들 간 차이가 있는지를 비교하고, 잘 알려져 있는 세 가지 검정법(Log-rank, Gehan, Tarone-Ware)을 적용하여 그룹 간 생존율에 유의한 차이가 있는지를 확인한다. 나아가 2절에서 언급한 Cox 모형 분석법을 이용하여 모형선택 및 변수선택을 실시한다. 4절에서는 분석결과에 대해 토론 및 제언을 한다. 마지막으로 변수선택에 사용된 frailtyHL R 패키지 (Ha 등, 2012, 2018)의 R 코드를 부록에서 제시한다.

2. 생명보험 자료를 활용한 콕스 비례위험 모형 분석

2.1. 콕스 비례위험 모형의 기본 개념

본 연구에서는 생명보험자료를 이용한 계약유지와 관련된 회귀변수의 효과를 추정하고, 생명보험 계약 유지에 유의미한 영향을 미치는 변수를 찾기 위해 콕스 비례위험 모형을 활용하였다. 콕스 비례위험 모형은 생명보험 계약유지에 영향을 미치는 다양한 공변량(covariate)들을 파악하는데 유용한 분석 방법으로서, 중도절단(censoring)된 자료와 같은 불완전한 자료(incomplete data)에 대한 분석도 가능하다. 특히 이 모형은 세 가지의 큰 특성을 가지고 있다. 첫 번째는 비례위험성(proportional hazard; PH)의 가정과 두 번째는 준모수적(semi-parametric) 모형이다. 마지막으로 무 절편항(no intercept term)이라는 특성을 가진다.

콕스 비례위험 모형은 생명보험 계약유지기간에 대한 확률 분포의 가정 없이 위험함수와 설명변수와의 관계를 설명할 수 있다. 즉 i 번째 생명보험 계약의 생명보험 계약유지기간 T_i ($i = 1, \dots, n$)에 대한 위험률과 공변량의 관계로 나타내며, 이는 다음과 같이 정의된다.

$$\lambda(t; x_i) = \lambda_0(t) \exp\left(x_i^T \beta\right), \quad i = 1, \dots, n,$$

여기서 $\lambda_0(t)$ 는 미지의 기저(unknown baseline) 위험함수로서 p 개의 공변량들의 벡터 $x_i = (x_{i1}, \dots, x_{ip})^T$ 의 모든 값들이 0일 때의 위험함수이며, i 번째 계약의 공변량 x_i 에 대응하는 회귀모수들의 벡터

$\beta = (\beta_1, \dots, \beta_p)^T$ 는 주요한 관심추론의 대상이 된다. 콕스 비례위험 모형은 $\lambda_0(t)$ 에 대한 아무런 분포 가정이 없이 좋은 통계적 성질을 제공한다. 반면에 비례위험이라는 가정이 만족하지 않으면 문제가 발생할 수도 있다.

다음은 이러한 콕스 비례위험 모형의 주요한 특성이다.

- (가) 준 모수적 모형: $\lambda_0(t)$ 의 형태는 알 수 없지만(non-parametric), 공변량의 함수 형태는 $\exp(x^T \beta)$ 으로 아는 형태로 설정한다.
- (나) 비례위험: 두 그룹간의 위험률 비, 즉 위험비(hazard ratio)가 시간에 관계 없이 일정한 상수에 비례(proportional)한다. 따라서 콕스 비례 위험 모형은 PH라는 가정으로 관심회귀모수의 해석이 용이하다.
- (다) 무 절편항: $\lambda_0(t)$ 가 절편항 역할을 하기 때문에 선형 예측식(linear predictor) $x_i^T \beta$ 는 동일성(identifiability)문제로 인해 절편항을 포함하지 않는다.

2.2. 콕스 비례위험 모형의 벌점화 변수선택법

변수 선택을 위한 벌점화 가능성도 l_p 는 다음과 같이 정의된다 (Fan과 Li, 2002; Ha 등, 2014):

$$l_p(\beta) = l(\beta) - n \sum_{j=1}^p J_\gamma(|\beta_j|)$$

여기서 $l(\beta)$ 은 콕스 모형에 대한 β 의 편 로그-가능도 함수(partial log-likelihood function)이며, $(J_\gamma(|\bullet|))$ 는 조율(tuning)모수 γ ($\gamma \geq 0$)를 가지는 벌점함수이다. 여기서 γ 의 값이 커질수록 공변량을 적게 선택하기 때문에 단순한 모형이 된다. 이러한 벌점화 변수선택법의 주요한 장점은 중요한 변수를 선택함과 동시에 그 변수의 회귀계수를 추정하는 것이다. 따라서 중요하지 않은 변수에 대한 회귀계수를 0으로 추정함으로써 중요하지 않은 변수를 삭제한다. 본 연구에서 고려하는 세 가지 벌점함수의 정의는 다음과 같다.

- (가) LASSO (Tibshirani, 1996):

$$J_\gamma(|\beta|) = \gamma|\beta|.$$

- (나) SCAD (Fan과 Li, 2001):

$$J_\gamma = \gamma I(|\beta| \leq \gamma) + \frac{(a\gamma - |\beta|)}{a - 1} + I(|\beta| > \gamma),$$

여기서 $x_+ = xI(x > 0)$ 으로서 x 는 양수부분으로 표시되며, $a = 3.7$ 이다.

- (다) HL (Lee와 Oh, 2014):

$$J_\gamma(|\beta|) \equiv J_{(\alpha, w)}(|\beta|) = \log \Gamma\left(\frac{1}{w}\right) + \frac{\log w}{w} + \frac{\beta^2}{2\alpha u(|\beta|)} + \frac{(w-2) \log u(|\beta|)}{2w} + \frac{u(|\beta|)}{w},$$

여기서 $u(|\beta|) = [\{8w\beta^2/\alpha + (2-w)^2\}^{1/2} + (2-w)]/4$.

특히 HL 벌점함수는 w 의 값에 따라 그 모양이 바뀌며, w 가 0인 경우 ridge가 되며, w 가 2인 경우 LASSO가 되며, w 가 20보다 큰 경우 0에서의 값이 음의 무한대 값을 갖는 형태가 된다. 따라서 계산적인 효율성을 위해 $w = 2.1, 3, 10, 30, 50$ 의 값을 보통 사용 한다 (Lee와 Oh, 2014; Ha 등, 2017). 한

Table 3.1. Basic statistics for the life insurance data

변수	기초통계량
Advisor	0 = 135(28%), 1 = 341(72%)
Sex	0 = 208(44%), 1 = 268(56%)
Age	mean = 40.9, median = 40, min = 17, max = 78
Coporate	0 = 461(97%), 1 = 15(3%)
Drive	0 = 108(23%), 1 = 368(77%)
Disease	0 = 447(94%), 1 = 29(6%)
Type	3 = 39(8%), 2 = 166(35%), 1 = 271(57%)
Job	0 = 305(64%), 4 = 128(27%), 3 = 34(7%), 2 = 6(1%), 1 = 1(1%)
Tendency	0 = 181(38%), 5 = 22(5%), 4 = 79(17%), 3 = 83(17%), 2 = 83(17%), 1 = 27(6%)

편, 조율모수 γ 를 선택하기 위해 Ha 등 (2014)의 제안에 따라 본 논문에서는 $l(\beta)$ 를 이용한 Bayesian information criterion (BIC) 기준을 사용하여 각 별점화 방법에 대해 BIC를 최소로 하는 조율모수 γ 를 선택한다. 위의 세 가지 변수선택법에 대한 성질 등 자세한 설명은 Ha 등 (2014)와 Kim 등 (2015)을 참조할 수 있다.

3. 실제 자료 분석

3.1. 기초 통계 분석

본 연구에 사용된 Table 1.1에 있는 생존자료의 변수에 대한 기초 통계분석은 다음과 같다 (Table 3.1). 표집된 데이터 중 중도절단비율(censoring rate)은 52%를 나타내며, 이와 관련된 나머지 변수에 대한 기초 통계량의 분석은 아래와 같다.

- Advisor의 변경(모집설계사와 수금설계사가 상이한 경우)과 관련하여서는 변경이 되지 않고 유지되는 계약건이 135건으로 28%를 차지하며, 변경이 된 경우는 341건으로 72%를 차지한다.
- 계약자의 성별과 관련하여 남성 계약자가 208건(44%), 여성 계약자가 268건(56%)이며, 청약서 상의 고지 내용 중 기왕력(지금까지 걸렸던 질병이나 외상 등 진찰을 받는 현재에 이르기까지의 병력이다)자와 관련하여서는 기왕력이 없음이 447건(94%), 기왕력이 존재함이 29건(6%)으로 조사되었다.
- 계약자의 나이로서 평균나이는 40세이며 17세에서 78세의 범위로 구성된다.
- 직업군은 무위험직에서 위험직 순으로 305건(64%), 128건(27%), 34건(7%), 6건(1%), 3건(1%) 순으로 나타났으나, 무위험직 또는 4등급의 비교적 안전한 위험등급의 직군이 91%로 가장 많은 직업군으로 나타났다.
- 상품군은 총 5가지의 군(종신, 연금, 저축, 보장성, 거치형)으로 나누었으나 이 중에서 보장성과 거치형 상품의 경우는 보장성은 종신군으로 거치형은 연금군으로 상품군을 세 가지군으로 나누었다.
- 이는 보장성의 경우 단독실비나 건강보장의 상품군으로 종신보장에 속하며, 거치형의 경우 즉시연금 또는 거치형 연금 상품 계약건임으로 연금 상품군으로 포함하였다. 이로 인해 종신의 경우 271건(57%), 연금166건(35%), 저축성 39건(8%)으로 구분된다.
- Tendency(투자성향)의 경우는 무위험(투자성향분석에 임하지 않거나, 투자자산의 구분을 가지지 않음)부터 위험선호형까지 5가지 군으로 나타내었으며, 182건(38%), 22건(5%), 79건(17%), 83건(17%), 83건(17%), 27건(6%)로 나타난다.

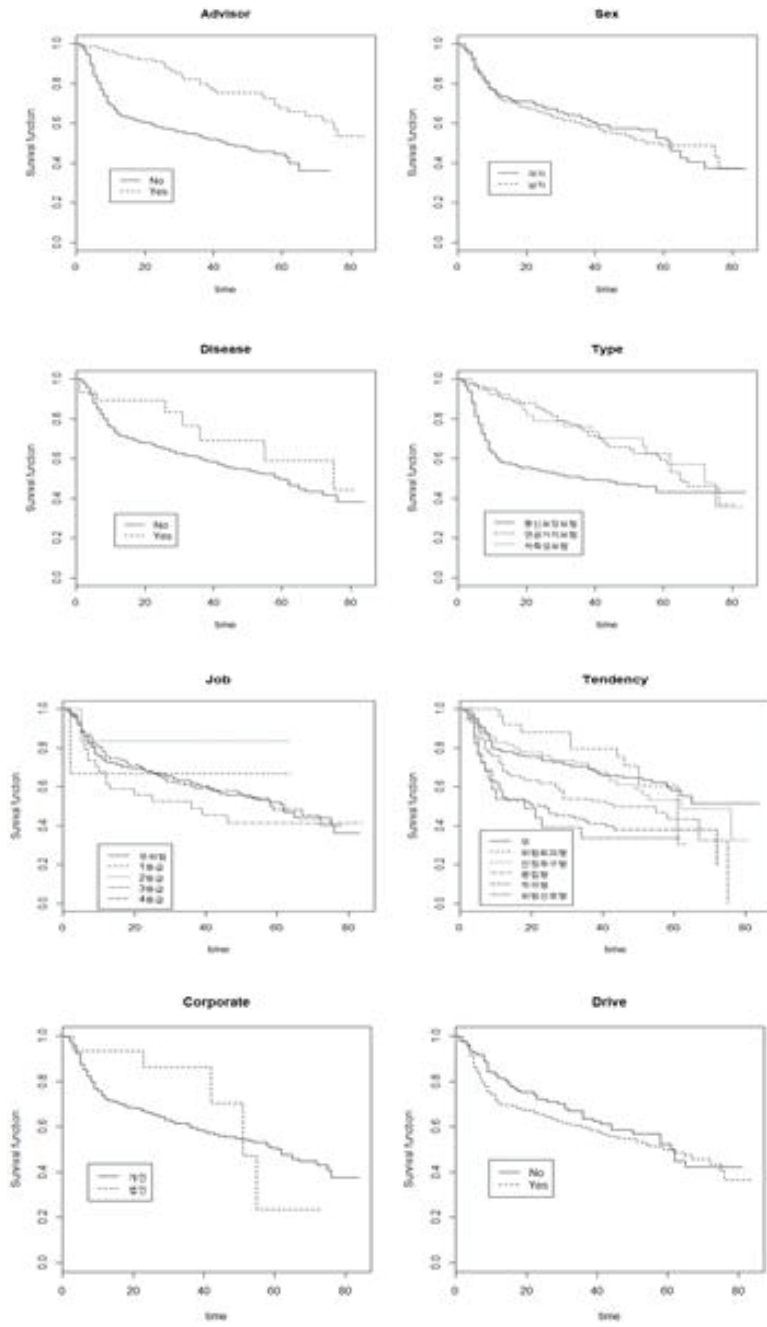


Figure 3.1. Plot of survival curves of Kaplan-Meier.

- 계약자가 일반 개인이거나 법인인 경우는 개인이 461건(97%), 법인 15건(3%)이며, 계약자의 운전 유무와 관련해서는 운전함이 368건(77%), 운전하지 않음이 108건(23%)의 값을 가진 것을 볼 수 있다.

Table 3.2. Comparison of survival rates among groups (P : p -value)

변수	Log-rank	Gehan test	Tarone-Ware
Advisor	$P < 0.001$	$P < 0.001$	$P < 0.001$
Sex	$P = 0.766$	$P = 0.731$	$P = 0.741$
Age	$P = 0.360$	$P = 0.256$	$P = 0.303$
Disease	$P = 0.160$	$P = 0.132$	$P = 0.141$
Job	$P = 0.526$	$P = 0.549$	$P = 0.535$
Type	$P < 0.001$	$P < 0.001$	$P < 0.001$
Tendency	$P < 0.001$	$P < 0.001$	$P < 0.001$
Corporate	$P = 0.516$	$P = 0.331$	$P = 0.409$
Drive	$P = 0.394$	$P = 0.221$	$P = 0.295$

3.2. 카플란 마이어 분석

Figure 3.1의 카플란 마이어의 생존율 그림을 통하여 보험계약 효력 상실이 발생한 기간 마다 구간 계약 유지와 관련된 각 범주형 변수에 대해 그룹들 간을 비교해 본다. 이를 통해 그룹 내 계약유지기간의 유지확률(즉 생존율)을 또한 확인할 수 있다. 부가적으로 이러한 유지확률은 계약유지기간의 보험계약 효력 상실이 발생하지 않을 확률로도 해석할 수 있다.

Figure 3.1에서 주의 깊게 살펴 볼 변수인 Advisor의 경우 수급설계사와 모집설계사가 동일한 계약건과 그렇지 아니한 계약건에 따라 계약유지 생존율에 큰 차이가 있음을 확인 할 수 있으며, 다른 변수(예: Sex, disease 등)에 대해서는 앞으로 진행될 연구를 통하여 재차 확인하기로 한다.

다음은 그룹 간 생존율 차이 검정을 위하여 Log-rank, Gehan, Tarone-Ware 검정법을 통하여 검정하였으며, 그 결과는 Table 3.2와 같다.

3.3. 콕스 비례위험 모형의 분석

Table 3.2에 의하면 Log-rank, Gehan과 Tarone-Ware 검정법에서 나타난 결과로 3가지 검정법 모두에서 동일한 변수 즉 Advisor, Type, Tendency의 변수를 유의한 값으로 그룹간의 차이가 있었다.

Table 3.3의 콕스 비례위험 모형의 적합 결과에 따르면 Advisor의 변수 즉 모집설계사와 수급설계사가 다르거나 같은 경우에 설계사의 변경 유무가 가장 유의한 변수의 값(Z 값 = -4.088 , P -값 < 0.001)을 가진다. 모집 설계사와 수급 설계사가 같은 경우(Advisor1)가 그렇지 못한 경우(Advisor0)에 비해 계약 효력 상실의 위험률이 $\exp(-0.884) = 0.413$ 배, 즉 약 59% 정도 위험이 감소한다.

상품군으로는 연금 상품군(Type2)이 유의한 변수의 값(P -값 < 0.001)을 가지며, 종신 상품군(Type1)에 비해 계약 효력 상실의 위험률이 $\exp(-0.638) = 0.528$ 배, 즉 약 47% 정도 위험이 감소한다. 적합성 진단을 통한 투자 성향지표인 Tendency의 경우는 Tendency4(안정추구형)의 경우 (P -값 = 0.004), Tendency5(위험회피형)의 경우 (P -값 = 0.032)로 둘 다 5% 유의수준에서 유의한 값을 가진다. Tendency0(무등급)에 비해 Tendency4(안정추구형)는 $\exp(0.573) = 1.774$ 배, Tendency5(위험회피형)는 $\exp(0.646) = 1.908$ 배로서 각각 77%와 91%로 만큼 위험이 증가함을 알 수 있다. 여기서 Tendency 0의 경우 무위험 성향이 아니라 등급이 없는 무등급을 의미한다. 특히 위험선호형(Tendency1)에서 안정추구형(Tendency5)으로 변화할수록 회귀계수 추정치가 양수로 커지는 경향이 있어 계약효력 상실의 위험률이 증가함을 알 수 있다. 참고적으로, 실제 현장에서 위험회피성향(Tendency4 또는 Tendency5)을 가진 계약자들은 위험선호형(Tendency1)에 비해 약간의 경기 변화(예: 약간의 금리변화)에도 큰 반응을 보이기 때문에 해약률의 위험이 증가하는 경향이 있다.

Table 3.3. Results of fitting Cox's proportional hazards model for life insurance data

Variable	coef	exp (coef)	SE (coef)	z	$\Pr(> z)$	
Advisor0	0.000	0.000	0.000			
Advisor1	-0.884	0.413	0.216	-4.088	$P < 0.001$	***
Sex0	0.000	0.000	0.000			
Sex1	-0.050	0.951	0.150	-0.335	0.738	
Age	0.055	1.057	0.008	0.744	0.457	
Disease0						
Disease1	0.016	1.016	0.417	0.038	0.970	
Job0	0.000	0.000	0.000			
Job1	0.448	1.565	1.08	0.415	0.678	
Job2	-0.820	0.440	1.026	-0.799	0.424	
Job3	0.296	1.344	0.253	1.167	0.243	
Job4	0.011	1.011	0.159	0.069	0.945	
Type1(종신)	0.000	0.000	0.000			
Type2(연금)	-0.638	0.528	0.155	-4.115	$P < 0.001$	***
Type3(저축)	-0.443	0.642	0.276	-1.604	0.109	
Tendency0	0.000	0.000	0.000			
Tendency1	-0.352	0.703	0.366	-0.962	0.336	
Tendency2	0.031	1.031	0.214	0.143	0.886	
Tendency3	0.361	1.435	0.198	1.827	0.068	
Tendency4(안정추구)	0.573	1.774	0.199	2.884	0.004	**
Tendency5(위험회피형)	0.646	1.908	0.302	2.143	0.032	*
Corporate0	0.000	0.000	0.000			
Corporate1	0.150	1.162	0.439	0.341	0.733	
Drive0	0.000	0.000	0.000			
Drive1	0.040	1.041	0.178	0.227	0.820	

*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

3.4. 콕스 비례위험 모형에서 변수 선택

콕스 비례위험 모형에서 Ha 등 (2014)의 변수 선택 방법을 이용한 결과는 Table 3.4에서 주어진다. 표 작성 및 설명의 편의상 Table 3.4에 있는 No penalty 결과는 Table 3.3과 같다. Table 3.4에 의하면 No penalty에서는 네 가지 변수 즉 Advisor1, Type2, Tendency4와 Tendency5가 유의한 것으로 나타났다. Table 3.4의 변수선택 결과 LASSO에서는 Advisor1, Age, Type2, 그리고 Tendency4를, SCAD에서는 Advisor1과 Type2를 HL에서는 Advisor1과 Type2, 그리고 Tendency4를 각각 선택하였다. 여기서 LASSO는 No-penalty에서 유의하지 않은 변수인 Age를 선택하였지만, SCAD와 HL은 이 변수를 선택하지 않았다.

특히 SCAD는 No-penalty에서 매우 유의한 Tendency4를 선택하지 못했다. 하지만 HL은 No-penalty에서 아주 유의한 세 가지 변수 즉(Advisor1, Type2, 그리고 Tendency4)를 선택하여 중요한 변수를 선택함을 알 수 있다. 첨가적으로, Table 3.4의 결과를 주는 “frailtyHL” R 패키지 (Ha 등, 2018)의 R 코드는 부록에 제시하였다.

4. 결론 및 제언

본 연구의 결과로 확인된 생명보험 계약건(청약서) 및 계약 내용에서 생명보험 계약건 유지에 영향을 미

Table 3.4. Variable selection and estimation using Cox's proportional hazards model for life insurance data: estimates (SE)

Variable	No penalty	LASSO	SCAD	HL
Advisor1	-0.884 (0.216)	-0.708 (0.123)	-1.065 (0.190)	-0.858 (0.153)
Sex1	-0.05 (0.15)	0	0	0
Age	0.055 (0.075)	0.021 (0.022)	0	0
Disease1	0.016 (0.417)	0	0	0
Job1	0.448 (1.08)	0	0	0
Job2	-0.82 (1.026)	0	0	0
Job3	0.296 (0.253)	0	0	0
Job4	0.011 (0.159)	0	0	0
Type2(연금)	-0.638 (0.155)	-0.371 (0.094)	-0.596 (0.149)	-0.457 (0.117)
Type3	-0.443 (0.276)	0	0	0
Tendency1	-0.352 (0.366)	0	0	0
Tendency2	0.031 (0.214)	0	0	0
Tendency3	0.361 (0.198)	0	0	0
Tendency4(안정추구)	0.573 (0.199)	0.184 (0.067)	0	0.252 (0.100)
Tendency5(위험회피형)	0.646 (0.302)	0	0	0
Corporate1	0.150 (0.439)	0	0	0
Drive1	0.040 (0.178)	0	0	0

LASSO = least absolute shrinkage and selection operator; SCAD = smoothly clipped absolute deviation; HL = hierarchical likelihood.

치는 주요한 세 가지 변수들에 대한 제언은 다음과 같이 요약된다.

첫 번째로 Advisor의 변경의 유무로써, 모집 설계사와 수급설계사가 상이하거나 같은 경우가 가장 유의미한 영향력을 가지는 것으로 확인하였다.

이는 모집 설계사의 장기근속이 되지 않음으로써, 기계약건의 관리 및 활용도의 미흡으로 인한 불만과 불편을 야기함으로써 조기에 보험 효력 상실이 이루어지고 있음을 반증한다. 그러므로 보험회사 및 관련금융 기관은 설계사 및 그에 준하는 종사자들로 하여금, 지속적인 교육과 역량을 높여 줌으로써 직업에 대한 만족도 향상과 생명보험 계약건의 계약체결 후 받게 되는 수수료의 선지급을 지향하여 성과가 미흡할 시 경제적 어려움으로 조기 이탈, 탈락을 방지하기 위해 계약체결 후 받게 되는 수당을 장기 분할로 지급함으로써 인하여, 설계사 및 그에 준하는 종사자의 장기간 근속함에 경제적인 뒷받침이 되도록 지향해야 할 것이다.

또한 보험회사 및 이와 연관된 금융기관은 우수 인증 설계사를 우대함으로써 단기간에 많은 성과를 이루는 설계사 및 종사자를 위촉하기 위한 방안으로, 높은 수수료율을 책정하여 지급하기 보단, 동일 회사 3년 이상 장기간 근속과 민원 건수 없음, 동시에 계약건 유지율 90%(13회)이상, 80%(25회)이상 유지하는 우수 인증 설계사를 우대하는 정책 및 회사 방침으로 건전한 영업 환경과 모집, 계약건 장기유지에 힘써야 할 것이다.

위촉직 등의 형태인 특수직으로 분류되어진 설계사 및 모집 종사자의 법정복지후생과 법정외복지후생의 처우 개선으로 장기 근속할 수 있도록 보장 받는 정책 개선이 필요하다.

두 번째는 상품군(Type2)으로써 생명보험 계약건의 주된 상품군인 연금, 보장성 보험인 종신보험, 장기간 투자 저축의 목적인 저축성 보험계약건으로 설계사 및 모집 종사자는 고객과 상담을 통하여 체결되는 계약 진행방식의 장점을 잘 활용하여, 계약자로 하여금 정확한 인지와 활용 방법을 안내 및 통보함으로써

써 각 상품 별로 주어진 혜택을 계약자가 유용하게 활용할 수 있도록 하여야 할 것이다.

계약자는 계약건의 효력 상실로 발생하는 경제적, 기타 부과 이득을 능숙하게 활용 할 수 있도록 해지 또는 해약의 방법을 권유 및 유도하여 계약효력 상실을 발생되지기보다, 중도인출, 약관대출, 특약 부과 및 특약 일부 해지, 감액 완납 등의 방법을 통하여 계약자에게 돌아올 손실을 최소화 할 수 있도록 하는 인식과 금융상품 활용 지식의 함양이 필요하다.

마지막으로 Tendency와 같은 투자 성향과 관련된 변수는 본 연구에 활용되어진 생명보험 계약건의 경우 적합성 원칙 질문에 응함에 있어 진단결과에 따라 계약자가 가입할 수 있는 보험계약 종류와 선택 할 수 있는 Fund와 Fund 투자 비중이 결정되어져 우선시 선택, 안내 확인하는 방법을 취하고 있었으며, 투자 성향 진단결과와 다른 생명보험 계약을 하기 위해서는 부적합 체결확인서를 작성해야함으로 계약 체결함에 추가적인 과정이 포함되어져있다. 이처럼 투자와 저축에 관련해서는 계약자 본인이 선택하는 투자 방법과 투자처, 투자 비중의 설명과 인지, 계약자가 납입 가능하고, 유지 가능한 금액과 기간의 설정으로 계약자가 원하는 생명보험 계약건으로 계약자의 만족도를 높여 줌으로 장기 계약유지에 도움이 되어야한다.

앞서 확인된 3가지(Advisor, Type, Tendency)변수에 중점을 둔다면 생명보험사 및 관련 금융기관은 계약건 장기유지와 설계사와 모집 종사자의 장기 근속으로 인한 생명보험의 이미지 개선과 재무상의 긍정적인 방향으로 이어질 것이다. 설계사 및 종사자들 또한 기존계약건의 지속적인 관리와 안내로 계약자의 만족도를 높여 추가적인 계약과 소개로 인하여 지속적인 활동이 가능할 것으로 보인다.

끝으로 다변량 생존분석을 위한 프레이리티(frailty)모형 등과 관련된 연구에 의하면, LASSO 방법은 중요하지 않은 변수를 많이 선택하는 경향을 보이는 반면, SCAD와 HL의 방법에서는 대부분 적절한 변수를 거의 동일하게 선택함을 보였다 (Ha 등, 2014, 2017). 하지만 3.4절에서는 세 방법 간 뚜렷한 차이를 보이는 것이 아니어서 세 방법의 보다 정밀한 비교를 위해서는 모의실험 및 추가적인 자료 분석 연구가 필요할 것으로 사료된다.

부록

R codes for variable selection in Cox PH model

```
library(frailtyHL)
lifedata=read.table(" result.csv" , header=T, sep=" , " )

formula <- Surv(time,status)~advisor+sex+age+disease+job+type
+tendency+corporate+drive+(1|id)

# No penalty
fit.cox_NO<-frailty.vs(formula, model="lognorm",penalty="lasso",data=lifedata,
  B=c(rep(0,17)),tun1=0,varfixed=TRUE,varinit=0)
z <- as.matrix(round(fit.cox_NO$beta/fit.cox_NO$se,3))
p_value <- as.matrix((1-pnorm(abs(z)))*2)

# LASSO
fit.cox_LASSO<-frailty.vs(formula,model="lognorm",penalty="lasso",data=lifedata,
```

```

B=c(fit.cox_NO$beta),tun1=seq(0,0.1,0.01),varfixed=TRUE, varinit=0)

# SCAD
fit.cox_SCAD<-frailty.vs(formula,model="lognorm",penalty="scad",data=lifedata,
  B=c(fit.cox_LASSO$beta),tun1=seq(0,0.2,0.01),varfixed=TRUE, varinit=0)

# HL
fit.cox_HL<-frailty.vs(formula,model="lognorm",penalty="hl",data=lifedata,
  B=c(fit.cox_LASSO$beta),tun1=c(2.1,3,10,30,50),tun2=seq(0.001,0.1,0.01),
  varfixed=TRUE,varinit=0)

```

References

- Choi, Y. M. and Choi, W. (2008). The effect of economic variables on lapse and surrender rate in life insurance, *Journal of Insurance and Finance*, **55**, 3–36.
- Cox, D. R. (1972). Regression models and life tables (with Discussion), *Journal of the Royal Statistical Society B*, **74**, 187–220.
- Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association*, **96**, 1348–1360.
- Fan, J. and Li, R. (2002). Variable selection for Cox's proportional hazards model and frailty model, *The Annals of Statistics*, **30**, 74–99.
- Ha, I. D., Jeong, J. H., and Lee, Y. (2017). *Statistical Modelling of Survival Data with Random Effects: H-likelihood*, Springer, Singapore.
- Ha, I. D., Noh, M., Kim, J., and Lee, Y. (2018) frailtyHL: frailty models using h-likelihood, From: <http://CRAN.Rproject.org/package=frailtyHL>. R package version 2.1.
- Ha, I. D., Noh, M., and Lee, Y. (2012). frailtyHL: a package for fitting frailty models with h-likelihood, *The R Journal*, **4**, 307–320.
- Ha, I. D., Pan, J., Oh, S., and Lee, Y. (2014). Variable selection in general frailty models using penalized h-likelihood, *Journal of Computational and Graphical Statistics*, **23**, 1044–1060.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association*, **53**, 457–481.
- Kim, B., Ha, I. D., Noh, M., Na, M. H., Song, H. C., and Kim, J. (2015). Variable selection in frailty models using frailtyHL R package: breast cancer survival data, *The Korea Journal of Applied Statistics*, **28**, 965–976.
- Kim, S. Y. and Lee, J. J. (2012). A comparative study on the effects of macroeconomic variables and consumer's dissatisfaction on the lapses of life insurance contracts, *Korean Journal of Insurance*, **93**, 237–269.
- Lee, Y. and Oh, H. S. (2014). A new sparse variable selection via random-effect model, *Journal of Multivariate Analysis*, **125**, 89–99.
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso, *Journal of the Royal Statistical Society Series B*, **58**, 267–288.
- Tibshirani, R. (1997). The Lasso method for variable selection in the Cox model, *Statistics in Medicine*, **16**, 385–395.

생명보험자료를 이용한 계약유지기간에 대한 생존분석

양대건^a · 하일도^{a,1} · 조건호^b

^a부경대학교 통계학과, ^b대구의한대학교 화장품공학부 산업품질공학전공

(2018년 9월 18일 접수, 2018년 10월 2일 수정, 2018년 10월 5일 채택)

요약

최근에 생명보험 산업은 보험계약의 장기 연장에 영향을 미치는 다양한 요인들에 관심을 두고 있다. 예를 들어 모 집 설계사의 장기간 고객관리의 필요성, 상품상담, 투자측면의 개선 등이다. 따라서 본 연구에서는 우리나라 생명보험사의 장기계약을 유지하는 중요한 요인들을 조사하고자 한다. 이를 위해 우리나라의 모 생명보험사의 2011년 1월 1일부터 2016년 12월 31일까지의 계약건의 내용에 대한 데이터를 사용하였다. 본 논문에서는 이러한 데이터를 사용하여 벌점화 콕스 비례위험모형 접근법을 통해 계약유지기간에 중요한 영향을 미치는 변수를 선택하는 방법을 제시한다. 분석결과 설계사의 변경 유무, 연금 상품군, 그리고 안정적 투자성향과 같은 세 가지 변수가 계약건 유지에 주요한 요인으로 선택되었다.

주요용어: 콕스 비례위험모형, 생명보험자료, 벌점화 변수선택, 생존분석

본 연구는 제1저자 양대건의 부경대학교 석사학위논문의 일부를 발췌, 수정한 논문임. 이 논문은 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구 사업임 (No. NRF-2017R1E1 A1A03070747).

¹교신저자: (48513) 부산광역시 남구 용소로 45, 부경대학교 통계학과. E-mail: idha1204@gmail.com