

생존분석기법을 이용한 해지시점 예측 (Cox PHM)

순간 해지확률 $h(t|x_1, \dots, x_p)$ 에 대한 최적 Cox PHM (Proportional Hazard Model)을 찾아보시오.

$$h(t|x) = h(t) \cdot \exp\left(\sum_{j=1}^p \beta_j x_j\right)$$

T = 보험 가입고객이 보험을 해지할 때까지 걸리는 시간

Y = 최종납입기간 = 최종납입횟수*납입간격 = x9 (월 단위로 환산한 최종 납입횟수)

delta = 2001년 6월 기준 3개월 이내 해지여부

Summary of Cox PH Model		
AIC = 1454.91 (df=12)		
variable	coef	p-value
log(가입연령)	-0.8122	0.0031 **
납입방법-3월납	0.1455	0.8981
납입방법-6월납	-6.4804	0.9468
납입방법-연납	0.3892	0.6882
log(월보험료)	0.2497	0.0028 **
계약일자	0.0221	4.04e-12 ***
log(차월)	-22.0125	6.35e-13 ***
연체여부-유	0.1739	0.8440
부활유무-유	-2.5962	0.6537
연체율	-39.8312	< 2e-16 ***
log(차월)*연체여부_유	0.6799	0.0089 **
(부활유무-유)*연체율	2.5965	0.6592

x1~x15 까지의 변수들 중 수치형 변수들에 대한 로그 변환과, 변수들 사이의 교호작용을 고려하고 stepwise regression 을 통해 최종 선택된 모형은 위와 같다. 가입연령과 월 납입 보험료, 계약시점으로부터 현재시점까지의 차의 로그 변환 변수들과 계약일자, 연체율 변수가 주로 중요 변수로 선택되었으며 차월과 연체여부의 교호작용 또한 유의하게 나타났다. 최적 AIC 값은 1454.91 로 part1 의 GLM 과 GAM 모형에서 계산된 AIC 보다는 다소 높다. 특히 현재까지 총 납입해야 할 횟수 중 실제 납입한 횟수의 비율을 계산하여 새롭게 생성한 연체율 변수를 모형에 포함했을 때 AIC 가 눈에 띄게 낮아지는 것을 확인할 수 있었다.