

Estimating home fire severity with statistical distributions

Yunjung Park^a, Inha Song^a, Soyoun Lee^a, Kwang Hyun Nam^b,
Rosy Oh^{1,c}, Jaeyoun Ahn^{2,a}

^aDepartment of Statistics, Ewha Womans University; ^bUnderwriter Inc.; ^cKorea Military Academy

Abstract

This paper evaluates the performance of various distribution assumptions in regression settings for estimating insurance loss. The gamma distribution is commonly used to handle the asymmetry property of loss distribution. However, recent studies highlight the significance of heavy-tailedness in loss distribution. Through an analysis of real home fire insurance data, we compare the effectiveness of different distribution assumptions in regression methods. Our findings show that the choice of parametric distributional assumption is crucial in determining premiums for various insurance products, including “excess of loss insurance” and “limit insurance”. Additionally, we discuss practical considerations for applying our results in home fire insurance.

Keywords: home fire insurance, statistical learning, generalized linear model, gamma distribution

1. 서론

보험 손실 데이터에 관한 여러 연구들을 통해 알 수 있듯이 보험 손실은 일반적으로 비대칭성과 꼬리가 두꺼운 특징을 가지고 있다 (Ahn 등, 2012). 특히 보험손실 중 개별심도(individual severity)는, 즉 사고한건당 피해액, 일반적으로 손실은 음수가 아닌 값을 가지며, 오른쪽으로 치우치고(right skewed), 정규 분포보다 평균치를 중심으로 보다 높고 좁게 집중되어(leptokurtic) 꼬리가 두꺼운 분포를 가지는 것이 잘 알려져 있다 (Punzo 등, 2018). 이러한 특징은 Figure 1의 개별심도의 히스토그램을 통해서 확인 할 수 있다. 여기서 사용된 자료는 화재보험에서 건물의 피해액(왼쪽)과 가재도구 피해액(오른쪽)이다.

전통적으로 보험연구에서는 이러한 개별심도의 특성을 잘 반영할 수 있는 분포로 와이불(Weibull), 파레토(Pareto), 로그정규(log normal), 감마(gamma) 분포 등이 많이 사용되고 있지만 이러한 분포들의 실제 보험 데이터에서 적합도에 관한 연구는 많지 않고, 최근에는 새로운 손실 데이터를 잘 적합할 수 있는 분포에 관한 필요성이 제기되고 있는 상황이다 (Sarabia 등, 2020). 위에서 기술한 개별심도 분포의 오른쪽 꼬리 부분은 결국 고액 사고를 의미한다. 고액사고의 확률 분포를 이해하는 것은 보험자에게 중요하다. 왜냐하면 한 번의 고액사고가 보험자의 재무건전성을 급격히 악화할 수 있기 때문이다. 이러한 고액사고의 확률 분포를 잘 이해하기 위해, 본 논문은 두 가지 개념을 활용하고자 한다. 첫 번째 개념은 보상 한도이다. 보상 한도는 보험사고가 발생하였을 경우 보험회사가 피보험자에게 지급할 금액의 최고한도로서, 보험가입금액이라고도 한다. 일반적으로, 한도가 높을수록 보험료가 충분히 커져야 한다. 예를 들어, 보상 한도가 1억 원일 때 대비해서

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00217022).

¹ Corresponding authors: Department of Mathematics, Korea Military Academy, PO Box 77-2 Nowon-gu, Seoul 01805, Korea. E-mail: rosy.oh5@gmail.com

² Corresponding authors: Department of Statistics, Ewha Womans University, 52 Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Korea. E-mail: jaeyahn@ewha.ac.kr

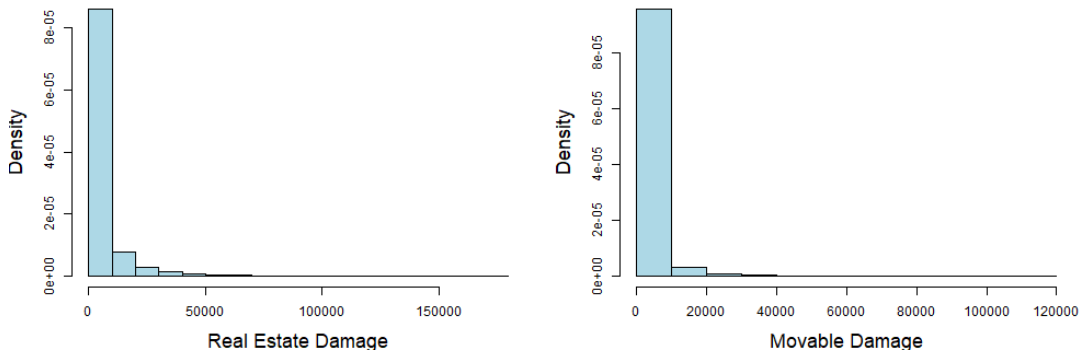


Figure 1: Distribution of individual severity.

보상 한도가 3억 원일 때의 보험료가 더 커야 한다. 보험료가 충분히 커지기 위해서는, 고액 사고의 확률이 실제보다 작게 예측되어서는 안된다.

다만, 한도가 높아짐에 따라 부과되는 보험료가 지나치게 크다면, 이는 보험 가입자에게 과도한 부담이 될 것이다. 이에 따라 고려해야 하는 두 번째 개념이 고액사고 초과점(excess point)이다. 통상적으로 손해보험의 사고는 특정 고액사고 초과점을 기준으로 해당 초과점보다 작은 금액의 사고 빈도가 전체 사고 빈도의 대부분을 차지한다. 따라서 해당 고액사고 초과점 미만의 사고를 보상하기 위해 더 많은 보험료가 부과되어야 하며, 초과점을 초과하는 사고를 보상하기 위해서는 상대적으로 적은 보험료가 부과되어야 한다.

본 논문은 화재 보험 데이터를 이용하여 개별심도 피해 금액에 대해 다양한 분포를 가정해 본 후, 분포 적합도를 살펴봄으로써 현행 화재 보험 제도 하에서 적합한 분포에 관한 연구를 진행하고자 한다. 논문은 다음의 순서로 구성된다. 먼저 제 2장에서 꼬리가 두꺼운 분포와 꼬리가 얇은 분포의 정의를 기술한 후, 제 3장에서 본 논문에서 사용한 분포와 데이터 및 통계적 모형을 소개한다. 그 후 제 4장에서 초과손해액 보험과 보상 한도가 존재하는 보험의 보상 한도 및 고액사고 초과점 별 보험료를 구하는 수식을 정의하고, 제 5장에서 실제 회귀 분석 결과와 분포 별 적합도 비교 결과를 제공한다. 제 6장에서 실제 사고 정보를 예시를 이용하여 초과손해액 보험과 보상 한도 및 고액사고 초과점 별 보험료를 분포별로 살펴본 후, 제 7장에서 실무적 고려 사항을 다룬다. 마지막으로 제 8장에서 연구 결과를 요약하고 시사점을 제시한다.

2. 꼬리가 두꺼운 분포와 꼬리가 얇은 분포

본 장에서는 꼬리가 두꺼운 분포와 꼬리가 얇은 분포의 정의를 설명함으로써 논문의 이해를 돕는다.

2.1. 꼬리가 두꺼운 분포

꼬리가 두꺼운 분포(heavy-tailed distribution)는 지수분포보다 더 두꺼운 꼬리를 가지고 있는 분포이다 (Asmussen, 2003). 대부분의 경우 분포의 오른쪽 꼬리가 중요하지만, 왼쪽 꼬리가 무겁거나 양쪽 꼬리가 무거울 수 있다. 수학적으로는 확률 변수 X 가 분포 함수 $F(x)$ 를 가질 때, 적률 생성 함수 $M_X(t)$ 가 모든 $t > 0$ 에 대하여

무한대의 값을 가지면 $F(x)$ 는 꼬리가 두꺼운 분포를 가진다고 정의한다. 이를 수식으로 표현하면

$$\int_{-\infty}^{\infty} e^{tx} dF(x) = \infty, \quad \text{for all } t > 0.$$

과 같다. 여기서 생존함수 \bar{F} 를 $\bar{F}(x) \equiv \Pr[X > x], x > 0$ 로 정의하면, 꼬리가 두꺼운 분포 함수는

$$\lim_{x \rightarrow \infty} e^{tx} \bar{F}(x) = \infty, \quad \text{for all } t > 0$$

와 같이 표현될 수 있다. 특히 확률 변수 X 가 위와 같은 정의를 만족할 경우

$$P(X > x) = \lim_{x \rightarrow \infty} \frac{\bar{F}(x+y)}{\bar{F}(x)} = 1, \quad \text{for all } y \geq 0$$

과 같은 성질을 가지는데 이는 꼬리가 두꺼운 분포의 경우, 확률변수 X 의 값이 임의로 주어진 값 x 를 초과하면, 다른 임의의 양수 $y > 0$ 에 대하여 X 가 $x + y$ 를 초과할 확률이 1에 가까워지는 성질을 가짐을 뜻한다 (Foss 등, 2013). 로그정규분포와 와이불분포 등이 두터운 꼬리 분포의 대표적인 예이다.

2.2. 꼬리가 얇은 분포

꼬리가 얇은 분포(light-tailed distribution)는 지수분포보다 더 얇은 꼬리를 가지고 있는 분포이다 (Asmussen, 2003). 수학적으로는 상수 $c_0 > 0$ 이 존재하여 적률생성함수(moment generating function)이 유한한 값

$$E[e^{tX}] < \infty$$

을 가지면 꼬리가 얇은 분포로 정의한다. 보험에서 가장 많이 쓰이는 대표적인 꼬리가 얇은 분포로는 감마분포와 정규분포가 있다. 이러한 꼬리가 얇은 분포들은 일반적으로 보험 손실 분포를 잘 반영하지 못함이 알려져 있지만, 예를 들어 감마분포의 경우, 통계적 편리함을 이유로, 보험에서 개별심도의 모형설정을 하는데 많이 쓰이고 있다 (Oh 등, 2021).

3. 분포 및 모형

본 장에서는 분석의 대상이 되는 화재의 심도에 대한 분포 가정과 사용 모형에 대해서 살펴보고자 한다. 먼저 자료의 형식과 각 분포에 대한 정의를 제시하고 개별심도에 대한 일반화 선형 모형을 살펴본 후 총합 손해에 대한 모형을 설명하도록 한다.

본 연구의 관심사는 보험손실의 분포적합성에 있으므로 우선 일반적인 결합위험모형(collective risk model) 중에서 개별심도에 관한 다음과 같은 데이터를 고려하자. 자료의 형태는 보험가입자 $i \in \Omega$ 에 대하여

$$(Y_i, x_i), \quad i \in \Omega \tag{3.1}$$

과 같이 보험가입자 i 의 화재로 인한 개별손해(individual severity) Y_i 와 관련된 설명변수 $x_i \in \mathbb{R}^p$ 의 (행)벡터이다. Ω 는 화재로 보험청구 이력이 있는 보험가입자들의 색인집합(index set)이고 m 명의 보험가입자들을 가진다고 하자. 기호의 간결성을 위하여 이제부터 보험가입자 $i \in \Omega$ 의 설명변수 $x_i \in \mathbb{R}^p$ 는 생략하지만 항상 주어진 것으로 가정한다. 예를 들어 $E[Y_i | x_i]$ 는 $E[Y_i]$ 으로 간결히 나타낼 수 있다.

3.1. 개별 심도 분포 가정

2절에서 설명한 대로 개별 심도 자료는 꼬리가 긴 형태를 갖는다는 것이 잘 알려져 있다. 따라서 본 연구에서 적용하는 실제 자료의 특성을 반영하기 위해 꼬리가 얇은 분포인 정규분포와 감마분포, 그리고 꼬리가 두꺼운 분포인 로그정규분포와 와이불분포를 고려하여 심도를 추정할 후 분포 별 적합도를 비교하기로 한다.

3.1.1. 정규분포

정규분포는 가우시안(Gaussian) 분포라고도 불리며 실수 값의 임의의 변수에 대한 연속확률분포의 한 유형인 꼬리가 얇은 분포이다. 정규분포의 매개변수는 평균 혹은 기대치를 나타내는 μ 와 표준편차를 나타내는 σ 이며, 확률밀도함수는 다음과 같다.

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2\right], \quad x \in \mathbb{R} \quad (3.2)$$

이때, 화재 심도는

$$Y_i \sim N(\text{mean} = \mu(x_i), \text{variance} = \sigma^2), \quad i \in \Omega \quad (3.3)$$

과 같이 가정하며 $\sigma^2 \in \mathbb{R}_{>0}$ 이고, 함수 $\mu : \mathbb{R}^p \rightarrow \mathbb{R}$ 는 사용된 통계 모형에 따라 정의한다.

3.1.2. 감마분포

감마분포는 연속 확률 분포의 두 매개변수 계열로서 일반적으로 사용되는 꼬리가 얇은 분포다. 매개변수는 분포의 모양을 결정하는 형상(shape) 매개변수 $k > 0$ 와 척도 단위를 결정하는 척도(scale) 매개변수 $\theta > 0$ 이며, 확률밀도함수의 일반적인 형식은 다음과 같다.

$$f(x; k, \theta) = \frac{x^{k-1} \exp[-x/\theta]}{\theta^k \Gamma(k)}, \quad \text{for } x > 0. \quad (3.4)$$

이때, 화재 심도는

$$Y_i \sim \text{Gamma}\left(\text{shape} = k, \text{scale} = \frac{\mu(x_i)}{k}\right), \quad i \in \Omega \quad (3.5)$$

과 같이 가정한다. 함수 $\mu : \mathbb{R}^p \rightarrow \mathbb{R}_{>0}$ 는 사용된 통계 모형에 따라 정의 할 수 있고, 평균은 $E[Y_i] = \mu(x_i)$ 이고 분산은 $\text{Var}(Y_i) = (\mu(x_i)^2)/k$ 이다. 여기서 $\mathbb{R}_{>0} := \{r \in \mathbb{R} \mid r > 0\}$ 을 나타낸다.

3.1.3. 로그-정규분포

로그정규분포는 임의의 변수의 로그(log)가 정규 분포를 따르는 꼬리가 두꺼운 분포이다. 따라서 임의의 변수 X 가 로그정규분포를 따르는 경우 $Y = \ln(X)$ 은 정규분포를 갖는다. 모수는 변수의 위치(location) 매개변수 $\mu \in \mathbb{R}$ 와 형상 매개변수 $\sigma > 0$ 이며, 확률밀도함수의 일반적인 형식은 다음과 같다.

$$f(x) = \frac{1}{x\sigma \sqrt{2\pi}} \exp\left[-\frac{(\log x - \mu)^2}{2\sigma^2}\right], \quad x > 0. \quad (3.6)$$

이때, 화재 심도는

$$Y_i \sim \text{Lognormal}(\text{location} = \mu(x_i), \text{scale} = \sigma), \quad i \in \Omega \quad (3.7)$$

과 같이 가정한다. 여기서 함수 $\mu : \mathbb{R}^p \rightarrow \mathbb{R}$ 는 사용된 통계 모형에 따라 정의된다.

3.1.4. 와이불 분포

와이불분포는 실패율이 시간의 거듭제곱에 비례하는 경우 지정된 실패 횟수 간의 시간을 측정하는 연속확률 분포인 꼬리가 두꺼운 분포이다. 와이불분포는 형상모수 $k > 0$ 와 척도모수 $\theta > 0$ 를 가지고 있으며, 확률밀도 함수의 일반적인 형식은

$$f(x; \lambda, k) = \frac{k}{\theta^k} x^{k-1} \exp\left[-\left(\frac{x}{\theta}\right)^k\right], \quad x > 0$$

과 같고 이때, 화재 심도는

$$Y_i \sim \text{Weibull}\left(\text{shape} = k, \text{scale} = \frac{\mu(x_i)}{\Gamma(1 + (1/k))}\right), \quad i \in \Omega \tag{3.8}$$

과 같이 가정한다. 여기서 함수 $\mu: \mathbb{R}^p \rightarrow \mathbb{R}_{>0}$ 는 사용된 통계 모형에 따라 정의된다.

3.2. 개별 심도를 위한 일반화 선형 모형

본 절에서는 (3.3), (3.5), (3.7), 그리고 (3.8)에서 나타난 $\mu(x_i)$ 에 관한 일반화 선형 회귀모형(*generalized linear regression model*)에 기반한 회귀 분석모형 기법을 기술한다 (Nelder와 Wedderburn, 1972). 여기서 $x_i \in \mathbb{R}^p$ 는 (3.1)에서 정의된 설명변수 (행)벡터이다. 논문의 간결성을 위해, 지금부터는 아래 첨자 i 를 생략하여도 의미가 명확한 상황에서는, 아래 첨자 i 를 생략하기로 한다. 우선 심도의 평균 $\mu(x)$ 는

$$\mu(x) = \exp(\beta_0 + x\beta)$$

와 같이 모형 설정한다. 여기서 $\beta_0 \in \mathbb{R}$ 과 $\beta \in \mathbb{R}^p$ 는 열 벡터이고 데이터를 통해 추정해야 할 모수이고, 추정은 최대가능도추정법(*maximum likelihood estimation*)으로 추정하기로 하고, 이때 추정된 $\beta_0 \in \mathbb{R}$ 과 $\beta \in \mathbb{R}^p$ 모수를 $\hat{\beta}_0$ 과 $\hat{\beta}$ 으로 나타내고 이에 해당하는 심도의 예측치를

$$\hat{\mu}(x) = \exp(\hat{\beta}_0 + x\hat{\beta})$$

과 같이 나타내기로 한다.

4. 초과손해액 보험과 보상 한도가 존재하는 보험

본 장에서는 초과손해액(*excess of loss*) 보험과 보상 한도가 존재하는 보험의 보험료 변화에 대해 살펴보고자 한다.

4.1. 초과손해액 보험: 초과점의 변화에 따른 보험료의 변화

초과손해액 보험이란 손해액을 기준으로 피보험자와 보험사 간의 책임을 분할하여 미리 초과점을 약정해 두는 방법으로 초과손해액 재보험은 비비례적재보험(*non-proportional reinsurance*)의 일종이다. 이러한 방법은 초과손해액 재보험에서 흔히 사용되는 방법이다 (Gray와 Pitts, 2012). 본 논문에서는 초과손해액 보험의 개념을 활용하여 통계 분포가 특정 초과점 이상의 확률 분포를 적절하게 추정하는지 확인하고자 한다. 예를 들어, 화재 사고에 대해 두 개의 보험이 존재한다고 가정해 보자. 첫 번째 보험은 보장 한도가 1억 원인 보험이고, 두 번째 보험은 발생손해액의 1억 초과 2억 원을 담보하는 초과손해액 재보험이다. 한도가 1억 원인 보험은 사고가 활발하게 일어나는 작동층(*working layer*)으로서, 많은 보험료가 할당되어야 한다. 반면, 1억 초과 2억 원을 담보하는 보험에 부과되는 보험료는 작동층에 부과되는 보험료보다 적어야 한다.

만일 초과점을 임의의 양수 $d > 0$ 라 한다면 초과점이 주어졌을 때의 초과손해액 보험의 보험금은 아래와 같이 정의된다.

$$(Y - d)_+ = \begin{cases} 0, & Y < d; \\ Y - d, & Y \geq d. \end{cases} \quad (4.1)$$

임의의 변수 Y 의 확률밀도함수와 누적분포함수를 각각 $f_Y(y)$ 와 $F_Y(y)$ 라고 하고, 초과점이 d 로 주어진 초과손해액 보험에서 다음의 통계값을

$$E[(Y - d)_+] = \int_d^{\infty} (y - d) f_Y(y) dy \quad (4.2)$$

기대 보험금이라고 정의 하자. $F(\cdot | \theta)$ 와 $f(\cdot | \theta)$ 를 각각 Y_i 의 누적분포함수와 확률밀도함수로 정의하면, 초과손해액 보험의 보험금 $(Y - d)_+$ 에 해당하는 로그가능도함수(log-likelihood function)는

$$l(\theta) = \sum_{i=1}^n \left(\ln(f(y_i | \theta)) I_{[y_i > d]} + \ln(F(y_i | \theta)) I_{[y_i \leq d]} \right)$$

을 이용하여 계산할 수 있다. 여기서 $I_{[]}$ 는 지시함수(indicator function)이다.

4.2. 보상 한도가 존재하는 보험: 한도의 변화에 따른 보험료의 변화

보상 한도(policy limit)를 임의의 양수 $u > 0$ 라 한다면 보상 한도가 존재하는 보험에서 보험금은

$$Y \wedge u = \begin{cases} Y, & Y < u; \\ u, & Y \geq u \end{cases} \quad (4.3)$$

과 같이 정의된다. 무작위의 변수 Y 의 확률밀도함수와 누적분포함수를 각각 $f_Y(y)$ 와 $F_Y(y)$ 라고 하고, 보상 한도가 $u > 0$ 로 주어진 보험에서 통계값

$$E[Y \wedge u] = \int_0^u y f_Y(y) dy + \int_u^{\infty} u f_Y(y) dy = \int_0^u y f_Y(y) dy + u(1 - F_Y(u)) \quad (4.4)$$

을 기대 심도라고 정의하자. 보상 한도가 존재하는 보험의 보험금 $Y \wedge u$ 에 해당하는 로그가능도는

$$l(\theta) = \sum_{i=1}^n \left(\ln(f(y_i | \theta)) I_{[y_i \leq u]} + \ln(1 - F(y_i | \theta)) I_{[y_i > u]} \right) \quad (4.5)$$

이며 보상한도가 존재하는 보험에서의 기대심도와 초과손해액 보험에서 기대심도는 다음과 같은 성질을 만족한다.

$$E[Y] = E[Y \wedge u] + E[(Y - u)_+].$$

5. 실증분석 결과

본 연구는 2017년도 소방청 자료를 활용하여 실증분석을 수행한다. 해당 자료에 대한 자세한 설명은 Lee 등 (2023)의 내용을 따르며 Table 1은 자료 중에서 실증 분석에 사용한 변수들과 정의를 정리하여 보여준다. 본 장에서는 이 자료에 대한 연구의 주요 결과를 제시한다. 실제 자료를 적용한 일반화 선형 모형 결과로 심도에 영향을 미치는 여러 가지 요인들을 살펴본 후 그래프와 로그가능도를 이용하여 각 분포별 적합도를 비교해 본다.

Table 1: Variable description

Variable	Description	Type	Category
부동산 피해금액	부동산 재산 피해액	Integer	-
동산 피해금액	동산 재산 피해액	Integer	-
지역	주택 시도 구분	Categorical	강원도, 경기도, 경상남도, 경상북도, 광주광역시, 대구광역시, 대전광역시, 부산광역시, 서울특별시, 세종특별자치시, 울산광역시, 인천광역시, 전라남도, 전라북도, 제주특별자치도, 충청남도, 충청북도
장소소분류	주택 구분	Categorical	다가구주택, 다세대주택, 단독주택, 상가주택, 아파트, 연립주택
발화층	발화 층수	Integer	-
소방서거리	소방서와의 거리	Real	-
건물구조조	건물의 뼈대 형식	Categorical	기타조, 목조, 벽돌조, 블록조, 블록조, 비닐하우스파이프조, 샌드위치패널조, 석조, 철골조, 철골철근콘크리트조, 철근콘크리트조, 철조, 치장벽돌조, 컨테이너조
바닥면적	각 층의 면적	Real	-
스프링클러	스프링클러 설비 여부	Binary	설치, 미설치
소화기구	소화기구 사용 여부	Categorical	사용, 기타 소화기구 활용, 미상, 사용법 미숙지, 결측
단독경보형감지기	단독경보형감지기 작동 여부	Binary	작동,미작동

자료: 국토교통부, 「건축구조기준」, 건축구조기준 총칙; 소방청, 「화재통계」

5.1. 화재 심도 분포 시각화

모형을 추정하기에 앞서 본 연구에서 활용하는 실증 자료의 시각화를 통해 어떤 분포를 사용하는 것이 적합한지 살펴보고자 한다. 분석에서 건물 및 가재도구의 화재피해금액 중 0인 값을 갖는 보험가입자의 자료를 제외하는 전처리 작업을 수행했다. 이는 5.3장에서 사용하는 분포 중 감마분포 및 로그정규분포의 지지집합 (support)이 0을 포함하지 않기 때문이다. 해당 전처리 작업은 5.3장에서도 유효하다.

Figure 2는 (a) 원자료 기준 히스토그램과 (b) 로그 변환된 화재피해금액 기준 히스토그램을 나타낸다. Figure 2(a)에서 보이듯이, 건물과 가재도구의 화재피해금액 모두 비대칭적인 분포를 보이고 있다. 특히, 우측으로 꼬리가 긴 분포를 갖고 있고, 이를 통해 국내에서 발생한 화재피해금액은 특정 금액 미만에 집중되어 있다는 사실을 알 수 있다. 즉, 특정 금액을 초과한 화재 사고는 거의 발생하지 않는데, 이 특정 금액을 초과점이라고 할 수 있다. 4.1장에서 설명한 누적분포함수가 이 초과점을 초과하는 피해금액에 대한 보험료 산출하기 위해 사용될 수 있다. 건물과 가재도구의 화재피해금액에 로그 변환 할 경우, 그 분포는 Figure 2(b)에서 보여지는 것과 같다. 로그 변환 시에는 본래 금액에 1을 더한 후 로그 변환을 취했다.

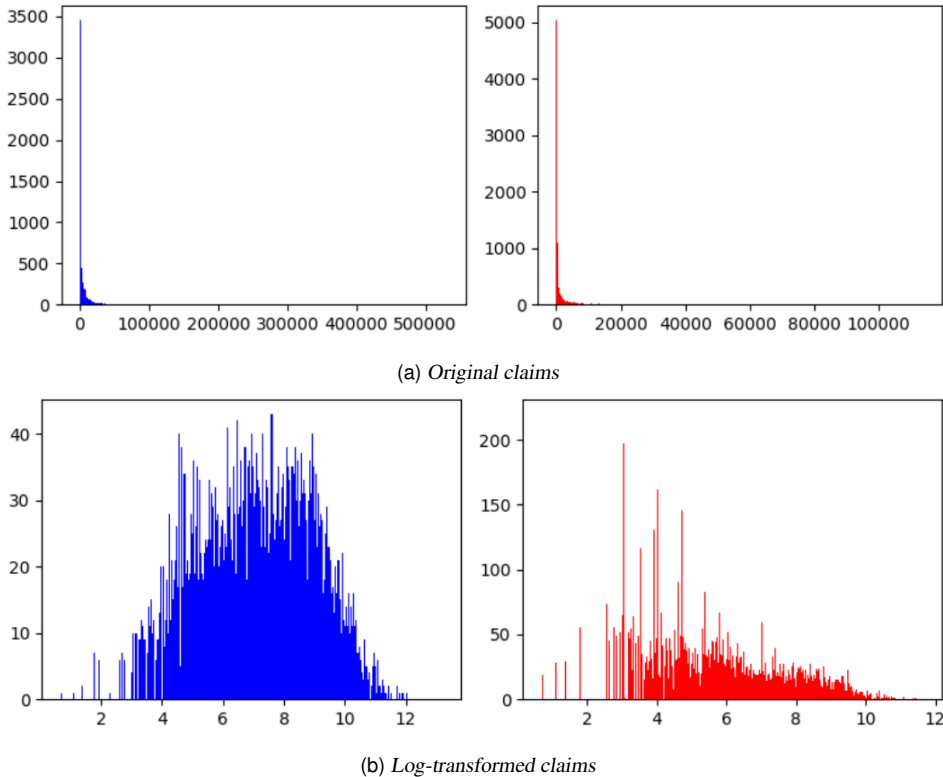


Figure 2: Histogram of fire damage costs for buildings (left) and equipment (right). (Unit: thousand won).

5.2. 일반화 선형 모형 분석 결과

이 장에서는 건물과 가재도구 각각에 대해 정규분포, 감마분포, 로그정규분포, 와이불분포 등 여러가지 분포를 사용하여 적합하고, 각 분포 별로 설명 변수가 피해 금액에 미치는 영향에 대해 알아보려고 한다. 이를 위해 회귀 분석 결과를 Table 2와 Table 3에 제공하되, 지면의 한계로 유의확률이 0.05 미만인 계수 만을 보고하려고 한다. 다만, 화재경보기, 소화기구, 스프링클러 사용 여부의 경우, 화재 피해에 대하여 의미 있는 요인이라는 판단 하에 유의확률이 0.05 이상이라고 해도 표에 포함하였다. 이 장에서는 가장 일반적인 감마분포를 사용한 회귀 분석 결과 만을 보여주고, 나머지 분포에 대한 결과는 Appendix A의 Table을 참고하기로 한다. 아울러, 감마분포에서는 여러 변수들의 교호 작용 조합들을 고려 시 로그 가능도 측면에서 성능이 가장 좋은 회귀식을 학습 후 도출하였다.

5.2.1. 건물

Table 2는 건물 화재 피해 금액에 대하여 감마분포를 가정하고 교호작용을 추가했을 때의 회귀분석 결과를 보여주고 있다. 분석 결과, 지역, 주택 종류, 건물 구조 등 다양한 세분화 기준에 따라 건물의 화재 피해금액에 미치는 영향이 통계적으로 유의미하게 상이한 것을 알 수 있다. 교호작용이 고려된 변수는 주택 종류와 건물 구조, 그리고 주택 종류와 면적이다.

먼저, 지역별로 건물 화재 피해금액에 미치는 영향이 상이하게 나타났다. 가장 감소폭이 큰 세종특별자치

Table 2: Results of gamma generalized linear model analysis for predicting building fire damage costs

Variable	Parameter	Estimate	p-value	
지역(강원도 대비)	경상남도	-0.3337	0.0130	
	광주광역시	-0.6797	0.0000	
	대구광역시	-0.7620	0.0000	
	대전광역시	-0.6606	0.0000	
	부산광역시	-0.7650	0.0000	
	서울특별시	-0.7600	0.0000	
	세종특별자치시	-0.9628	0.0020	
	울산광역시	-0.7741	0.0000	
	인천광역시	-0.4170	0.0110	
	전라북도	-0.3789	0.0080	
	제주특별자치도	-0.4652	0.0330	
	장소소분류*건물 구조(기타조 대비)	단독주택*목조	-1.9939	0.0100
		아파트*목조	-4.4515	0.0360
아파트*철골조		-3.3518	0.0370	
아파트*철근콘크리트조		0.6073	0.0210	
연립주택*샌드위치패널조		0.0000	0.0310	
연립주택*석조		0.0000	0.0100	
장소소분류*면적(m ²)	다가구주택*면적	0.0037	0.0300	
	단독주택*면적	0.0088	0.0000	
단독경보형감지기	화재경보기 작동	-0.0236	0.9000	
스프링클러	스프링클러 설치	-0.2000	0.1390	
	기타소화기구 활용	-0.0684	0.4970	
	소화기구 (결측 대비)	미상	0.2572	0.0570
사용		-0.3546	0.0000	
사용법 미숙지		0.5502	0.0490	
소방서거리 (km)	소방서거리 (km)	0.0080	0.0020	

시에서는 건물 화재피해금액이 대조 지역인 강원도에서 보다 약 62% 감소했으며, 다른 지역 대비해서도 낮은 값을 가졌다. 가장 감소폭이 작은 경상남도의 건물 화재 피해금액은 강원도에서 보다 29% 감소했고, 강원도 보다는 낮지만 다른 지역에 비해서는 높은 값을 가졌다. 주택 종류와 건물 구조의 조합에 따라 피해 금액에 있어서 차이를 보이는 것으로 나타났다. 단독주택의 경우에는 목조로 지어진 경우에 기준이 되는 기타조로 지어진 다가구 주택에 비해서 건물 화재피해금액이 약 87%이 감소함을 보였다. 아파트가 철근콘크리트조로 지어진 경우에는 약 2배에 가까운 피해 금액을 가짐을 확인할 수 있다. 또한 같은 아파트여도 철골조로 지어진 경우에는 약 97%가 감소하며 현저히 낮은 값을 보였다. 주택 면적은 다가구주택, 단독주택에 한해 건물 화재 피해 금액에 통계적으로 유의미한 영향을 미치는 것으로 나타났다. 다가구주택의 경우 주택 면적이 1m² 늘어날 때 건물 화재 피해 금액이 0.3% 늘어났고, 단독주택의 경우 주택 면적이 1m² 늘어날 때 건물 화재 피해 금액이 0.8% 늘어났다.

화재경보기가 작동하고 스프링클러가 설치됐을 때는 그렇지 않은 경우보다 각각 2%, 19% 정도 피해 금액이 감소했는데, 이는 화재경보기와 스프링클러가 작동할 경우 신속한 대피 및 화재 진압이 가능하므로 화재 피해를 줄일 수 있을 것이라는 예상과 일치했다. 소화기구를 사용할 경우는 소화기구에 대한 정보가 없을 때 보다 화재 피해금액을 약 30% 낮출 수 있었다. 반대로 거주자가 소화기구 사용법을 숙지하지 못한 경우는 피해금액이 73% 증가했다. 또한 건물이 소방서로부터 1km 멀리 떨어져 있을수록 건물의 화재 피해 금액은 약 1% 씩 미세하게 증가했다. 소방서가 가까이 위치할수록 화재 시 신속한 대응이 가능하여 피해를 최소화할

Table 3: Results of gamma generalized linear model analysis for predicting equipment fire damage costs

Variable	Parameter	Estimate	p-value
지역(강원도 대비)	대전광역시	-0.5134	0.0400
	부산광역시	-0.5628	0.0100
	서울특별시	-0.6143	0.0010
	울산광역시	-0.5928	0.0320
건물 구조(기타조 대비)	단독주택*벽돌조	0.7011	0.0230
	단독주택*비닐하우스스파이프조	3.1656	0.0300
	단독주택*컨테이너조	1.4123	0.0050
	아파트*철근콘크리트조	0.6735	0.0230
바닥면적(m^2)	단독주택*면적	0.0051	0.0010
발화층*장소소분류	발화층*다가구주택	0.1656	0.0090
	발화층*단독주택	-0.2381	0.0170
단독경보형감지기	화재경보기 작동	-0.5910	0.0060
스프링클러	스프링클러 설치	0.0429	0.8050
	기타소화기구 활용	0.0834	0.5290
소화기구(결측 대비)	미상	0.5818	0.0020
	사용	-0.0177	0.8490
	사용법 미숙지	0.6174	0.1190
소방서거리 (km)	소방서거리 (km)	0.0080	0.0020

수 있기 때문에 계수의 부호는 예상과 일치했다.

위 결과들을 종합해보면 주택 지역, 종류, 면적, 구조, 화재경보기 여부, 소화기구 여부, 그리고 소방서와의 거리가 감마분포를 가정한 건물 화재 피해금액 추정에 중요한 요소로 사용될 수 있음을 알 수 있다.

5.2.2. 가재도구

Table 3은 가재도구 화재 피해 금액에 대하여 감마분포를 가정하고 교호작용을 추가했을 때의 회귀분석 결과를 보여주고 있다. 분석 결과, 지역, 주택 종류, 건물 구조 등 다양한 세분화 기준에 따라 건물의 화재 피해금액에 미치는 영향이 통계적으로 유의미하게 상이한 것을 알 수 있다.

이 중, 지역별로 건물 화재 피해금액에 미치는 영향을 살펴봤을 때 가장 감소폭이 큰 서울특별시의 가재도구 피해금액은 기준 지역인 강원도의 건물 피해금액보다 약 46% 감소한 값을 가졌고, 대전광역시, 부산광역시, 울산광역시 모두 비슷한 값을 가짐을 확인할 수 있었다. 주택 종류와 건물 구조의 조합에 따라 피해 금액에 있어서 차이를 보이는 것으로 나타났다. 단독주택의 경우, 비닐하우스스파이프조로 지어진 단독주택의 경우에는 약 24배의 피해금액을 가졌고, 벽돌조와 컨테이너조는 각각 2배와 4배의 피해금액을 가졌다. 아파트가 철근콘크리트조로 지어진 경우에는 기타조 대비 약 2배에 가까운 피해 금액을 가짐을 확인할 수 있다. 주택 면적은 단독주택에 한해 가재도구 화재 피해금액에 통계적으로 유의미한 영향을 미치는 것으로 나타났다. 단독주택의 경우 주택 면적이 $1m^2$ 늘어날 때 건물 화재 피해금액이 0.5% 늘어났다. 주택종류 별로 층수에 따른 피해금액 또한 차이를 보였다. 화재가 발생하면 불이 아래 층에서 위층으로 번지기 때문에 층수가 낮을수록 화재피해금액이 증가하는 것이 일반적인데, 실제로 본 논문의 분석 결과에서도 단독주택의 경우 발화층이 한 층 증가할수록 피해금액이 21% 감소하는 결과를 보였다. 다만, 다가구주택의 경우 반대로 발화층이 한 층 증가할수록 피해금액이 18% 증가했다. 이는 다가구주택의 다락의 경우 기초소방시설 설치 의무 대상에서 제외되기 때문에 소화기, 소방관 진입창, 피난기구 등 화재 피해를 줄일 수 있는 장비들이 설치되지 않은 경우가 많기 때문인 것으로 판단된다 (Park, 2022).

Table 4: Log-likelihood of $(Y - e)_+$ in excess loss insurance for buildings

Excess point	Normal	Gamma	Log normal	Weibull
0	-76,435	-65,175	-64,550	-64,815
30,000	-31,967	-29,316	-29,249	-29,167
50,000	-25,194	-22,536	-22,520	-22,424
100,000	-15,668	-13,326	-13,333	-13,251

* the unit of excess point : thousand won

화재경보기가 작동했을 때는 그렇지 않은 경우보다 약 45% 피해금액이 감소했는데, 이는 화재경보가 작동할 경우 신속한 대피 및 화재 진압이 가능하므로 화재 피해를 줄일 수 있을 것이라는 예상과 일치했다. 스프링클러의 설치 여부는 유의확률이 0.805로 유의수준 5%에서 유의하지 않은 요인이었다. 화재 발생 시 스프링클러로 인한 수손 또한 피해금액으로 처리되는데, 특히 가재도구의 경우에는 손상된 가재도구를 버리고 새 제품으로 교환하는 경우가 많기 때문에 100%의 손해율까지 적용되기도 한다 (National Fire Service Academy, 2019). 이 같은 사실로 인해 가재도구의 경우 스프링클러 설치가 화재피해금액과 통계적으로 유의미한 관계를 갖지 않았고, 오히려 스프링클러가 설치됐을 때가 그렇지 않은 경우보다 4% 정도 미세하게 높은 피해금액을 가지는 것으로 판단된다. 소화기구를 사용할 경우는 소화기구에 대한 정보가 없을 때보다 화재 피해금액을 2% 정도 낮출 수 있었다. 반대로 거주자가 소화기구 사용법을 숙지하지 못한 경우는 피해금액이 85% 증가했다. 또한 건물과 마찬가지로 가재도구가 소방서로부터 1km 떨어져 있을수록 건물의 화재 피해금액은 약 1%씩 미세하게 증가했다. 소방서가 가까이 위치할수록 화재 시 신속한 대응이 가능하여 피해를 최소화할 수 있기 때문에 계수의 부호는 예상과 일치했으나 그 효과의 크기는 크지 않음을 확인할 수 있다.

위 결과들을 종합해보면 주택이 위치한 지역과 주택 종류, 건물 구조, 주택 면적, 화재경보기 작동 여부, 소화기구 사용 여부, 그리고 소방서와의 거리가 감마분포를 가정한 가재도구의 화재 피해금액 추정에 중요한 요소로 사용될 수 있음을 알 수 있다.

5.3. 분포 별 적합도

본 절에서는 화재 피해금액의 분포 가정에 따른 적합도와 다양한 초과점 및 한도가 보험료에 미치는 영향에 대해 알아보려고 한다. 먼저 모형 추정에 대한 시각적 진단으로 선형 예측값(linear predictor) $\hat{\eta} = \hat{\beta}_0 + x\hat{\beta}$ 과 잔차이탈도(deviance residual) 그림을 활용한다 (Agresti, 2015). Figure A.1을 보면 4가지 분포를 각각 건물과 가재도구의 피해금액에 대해 적합한 결과 정규분포를 제외하고 나머지 분포를 가정한 일반화 선형모형이 화재 피해금액 분석에 적합한 것으로 보인다.

5.3.1. 로그 가능도 함수

앞서 시각적 그래프를 통해 확인한 분포 별 적합도를 로그 가능도 함수를 이용하여 비교해보고자 한다. 초과손해액 보험과 보상 한도가 존재하는 보험에서 각각 초과점과 보상 한도를 설정한 후, 초과점 이상의 피해금액 및 보상 한도 이하의 피해금액에 대한 적합도를 살펴본다.

건물: Table 4는 초과손해액 보험에서 초과점이 0원, 300만원, 500만원, 1000만원 일 때의 건물 피해금액의 분포 별 로그 가능도를 계산한 표이고, Figure 3(a)는 초과점의 변화에 따른 로그 가능도의 변화 양상을 나타낸 그래프이다. 정규분포를 제외한 감마분포, 로그정규분포, 와이불분포는 초과점의 수준에 관계없이 서로 유사한 적합도를 보였다. 초과점에 따른 분포별 적합도를 보다 구체적으로 살펴보면, 초과점이 0원인 경우는 전체 피해금액 데이터에 대한 로그 가능도를 뜻한다. 전체 데이터에서는 로그정규분포, 와이불분포, 감마분포, 정

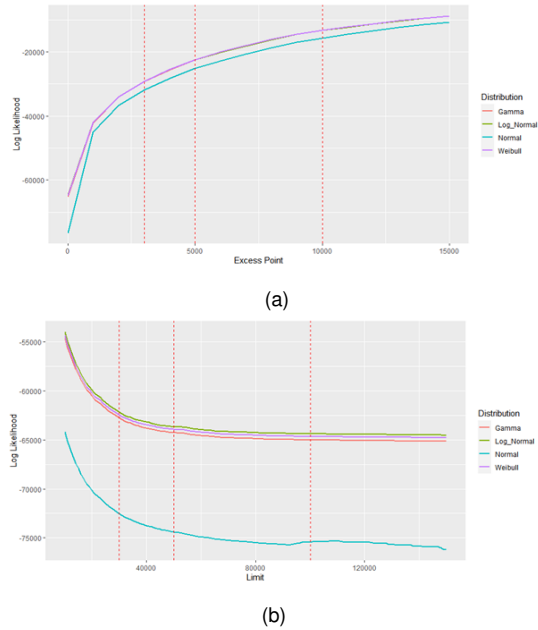


Figure 3: (a) Log-likelihood by distribution according to excess point in excess loss insurance for buildings, (b) Log-likelihood by distribution according to limit in insurance with a compensation limit for buildings.

Table 5: Log-likelihood of $(Y \wedge u)$ in insurance with a compensation limit for buildings

Policy limit	Normal	Gamma	Log normal	Weibull
30,000	-72,535	-62,745	-62,142	-62,432
50,000	-74,407	-64,254	-63,645	-63,921
100,000	-75,448	-65,018	-64,395	-64,663
None	-76,435	-65,175	-64,550	-64,815

* the unit of policy limit : thousand won

규분포 순으로 적합도가 우수했다. 초과점이 300만원, 500만원일 때는 와이불분포의 적합도가 가장 좋았으며 특히 1,000만 원 이상인 데이터에 대해서는 와이불분포가 정규분포에 비해서 약 15% 우수하였다. 꼬리가 두꺼운 분포인 베이불 분포와 로그정규분포가 일관적으로 가장 좋은 성능을 보였지만, 감마분포와 큰 차이는 없었다. 이는 Figure 3(a)를 통해서도 확인할 수 있다. 건물 화재 피해금액이 0인 데이터를 제외한 표본 크기가 약 7,000개인데, 향후 표본 크기가 늘어나면 적합도 성능이 더 향상될 것으로 기대된다.

보상 한도가 존재하는 보험에서 한도의 변화에 따른 건물 피해금액의 분포 별 로그 가능도는 Table 5과 Figure 3(b)에 정리되어있다. 정규분포를 제외한 감마분포, 로그정규분포, 와이불 분포는 한도의 수준에 관계없이 서로 유사한 적합도를 보였다. 한도에 따른 분포별 적합도를 보다 구체적으로 살펴보면, 모든 한도에 대해서 로그정규분포, 와이불분포, 감마분포, 정규분포 순으로 적합도가 좋았다. 한도가 3,000만원일 때와 5,000만원일 때 모두 감마분포가 정규분포에 비해 약 14% 우수하였고, 로그정규분포가 정규분포에 비해서 약 14% 우수하였다. Figure 3(b)를 통해서도 이 같은 사실을 확인할 수 있다.

가계도구: 먼저 초과손해액 보험에서 초과점의 변화에 따른 가계도구 피해금액의 분포 별 로그 가능도를 나

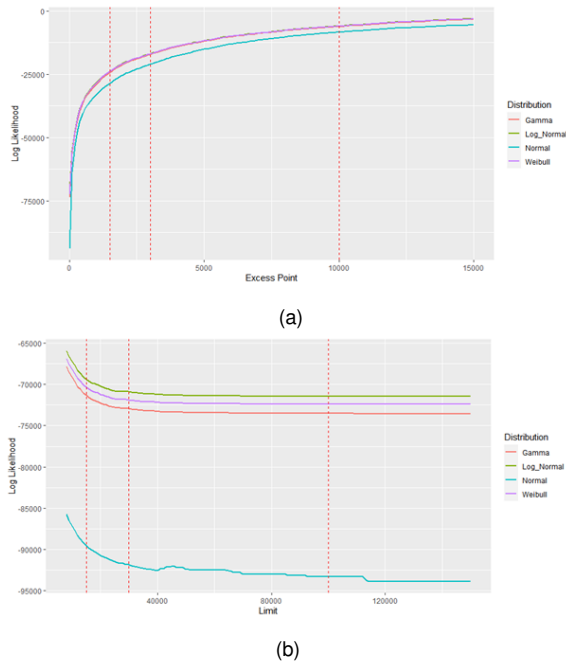


Figure 4: (a) Log-likelihood by distribution according to excess point in excess loss insurance for equipment, (b) Log-likelihood by distribution according to limit in insurance with a compensation limit for equipment.

Table 6: Log-likelihood of $(Y - e)_+$ in excess loss insurance for fire extinguishing equipment

Excess point	Normal	Gamma	Log normal	Weibull
0	-93,897	-73,554	-71,460	-72,421
1,500	-28,488	-24,344	-23,874	-23,946
3,000	-21,061	-17,165	-16,897	-16,931
10,000	-8,368	-6,152	-5,934	-5,993

* the unit of excess point : thousand won

타내는 Table 6와 Figure 4(a)에 따르면 정규분포를 제외한 감마분포, 로그정규분포, 와이불분포는 초과점의 수준에 관계없이 서로 유사한 적합도를 보였다. 초과점에 따른 분포별 적합도를 보다 구체적으로 살펴보면, 가재도구는 모든 초과점에 대해서 로그정규분포, 와이불분포, 감마분포, 정규분포 순으로 우수한 적합도를 가졌다. 이는 건물의 경우 와이불분포가 모든 경우에서 우수했던 것과 다소 다른 결과이다. 1,000만원 이상인 데이터에 대해서는 감마분포가 정규분포에 비해 약 26% 우수하였고, 로그정규분포가 정규분포에 비해서 약 29% 우수하였다. 건물과 마찬가지로, 와이불분포와 로그정규분포가 제일 좋았지만 감마분포와의 큰 차이는 없었다. Figure 4(a)에서도 정규분포의 로그 가능도가 가장 작은 것을 확인할 수 있다.

Table 7은 한도가 1500만 원, 3천만 원, 1억 원, 그리고 한도가 없는 경우 가재도구의 분포 별 로그 가능도를 나타내고, Figure 4(b)은 한도의 변화에 따른 로그 가능도의 변화 양상을 나타낸다. 건물과 마찬가지로 모든 한도에 대해서 로그정규분포의 적합도가 가장 적절한것으로 나타났으며 한도가 1,500만 원일 때와 3,000만 원일 때 모두 로그정규분포가 정규분포에 비해서 약 23% 우수하였다. Figure 4(b)은 모든 한도에 대해 로그정규분포, 와이불분포, 감마분포, 정규분포 순으로 로그 가능도가 큰 양상을 가짐을 보여준다.

Table 7: Log-likelihood of $(Y \wedge u)$ in insurance with a compensation limit for equipment

Policy limit	Normal	Gamma	Log normal	Weibull
15,000	-89,559	-71,329	-69,357	-70,319
30,000	-91,881	-72,929	-70,888	-71,852
100,000	-93,338	-73,531	-71,438	-72,400
None	-93,897	-73,554	-71,460	-72,421

* the unit of policy limit : thousand won

Table 8: Root mean squared error of fire damage costs for buildings

Distribution	Model 1	Model 2
Normal	11,888	11,593
Gamma	11,888	11,606
Log normal	11,880	11,684
Weibull	11,902	11,613

Table 9: Root mean squared error of fire damage costs for equipment

Distribution	Model 1	Model 2
Normal	5,549	5,462
Gamma	5,549	5,449
Log normal	5,547	5,519
Weibull	5,560	5,488

5.3.2. 표준제공근오차

분포별 예측력을 비교하기 위해 화재 피해 금액의 표준제공근오차(root mean squared error)를 계산하였다. 이때 모형 1은 심도의 추정치로 심도의 평균을 사용한다. 모형 2는 심도의 추정치로 분포 별 일반화 선형 모형 추정값을 사용한다. 본 연구는 7 : 3의 비율에 따라 무작위로 훈련 데이터(training data)와 평가 데이터(test data)를 구분하였으며, 표준제공근오차는 평가 데이터에 대한 수치이다. Table 8는 건물화재 피해 금액에 대한 분석 결과이며, Table 9는 가재도구의 화재 피해 금액에 대한 분석결과를 나타낸다.

건물: 모형 1에 비해 모형 2에서 정규분포는 2.5%, 감마분포는 2.4%, 로그정규분포는 2%, 와이불분포는 2.4% 예측력이 향상되었다. 모형 1과 모형 2 모두 분포 별로는 표준제공근오차에 큰 차이가 없는 것을 볼 수 있는데, 회귀 분석에서는 분포와 상관없이 평균을 잘 추정하고, 이 차이는 데이터의 크기가 커질수록 줄어드는 것이 알려져 있다. 본 논문도 이러한 사실에 따라 분포 별로 평균에 대한 예측력에는 큰 차이가 없음을 보이고 있다.

가재도구: 모형 1에 비해 모형 2에서 정규분포는 1.6%, 감마분포는 1.6%, 로그정규분포는 0.5%, 와이불분포는 1.3% 예측력이 향상되었다. 건물과 마찬가지로 모형 1과 모형 2 모두 분포 별로는 표준제공근오차에 큰 차이를 보이지 않는다.

5.4. 주요 발견

초과손해액 보험의 건물 화재 피해금액에서 정규분포보다는 감마분포에서, 감마분포보다는 로그정규분포, 와이불분포에서 좋은 적합도를 보였다. 그리고 본 논문에서 사용한 표본 크기로는 로그정규분포, 와이불분포와 감마분포 간 성능 차이가 크지 않았다. 1천만 원 이상인 데이터에 대해서는 감마분포가 정규분포에 비해 약 15% 우수하였고, 와이불분포가 정규분포에 비해서 약 15% 우수하였다. 보상 한도가 존재하는 보험의

Table 10: Insurance premium for excess loss insurance for buildings (unit: thousand won)

Excess point	Normal	Gamma	Log normal	Weibull
None	2,180	2,479	2,483	2,131
3,000	3,887	1,086	428	926
5,000	3,021	675	129	612
10,000	1,463	222	11	258

Table 11: Insurance premium for insurance with a compensation limit for buildings (unit: thousand won)

Policy limit	Normal	Gamma	Log normal	Weibull
30,000	2,164	2,475	2,483	2,111
50,000	2,180	2,479	2,483	2,128
100,000	2,180	2,479	2,483	2,131
None	2,180	2,479	2,483	2,131

건물 화재 피해금액에서는 로그정규분포, 와이불분포, 감마분포, 정규분포 순으로 적합도가 좋았다. 3,000만원이하 일 때와 5,000만원 이하일 때 모두 감마분포가 정규분포에 비해 약 14% 우수했고, 로그정규분포가 정규분포에 비해서 약 14% 우수하였다.

초과손해액 보험의 가재도구 피해금액에서는 로그정규분포, 와이불분포, 감마분포, 정규분포 순으로 우수한 적합도를 가졌다. 1,000만원 이상인 데이터에 대해서는 감마분포가 정규분포에 비해 약 26% 우수하였고, 로그정규분포가 정규분포에 비해서 약 29% 우수하였다. 보상 한도가 존재하는 보험의 가재도구 피해금액에서는 로그정규분포, 와이불분포, 감마분포, 정규분포 순으로 적합도가 좋았고, 이 순위는 건물 화재 피해액에서의 순위와 동일했다. 1,500만원 이하 일 때와 3,000만원 이하 일 때 모두 감마분포가 정규분포에 비해 약 20% 우수했고, 로그정규분포가 정규분포에 비해서 약 23% 우수하였다. 표준제곱근오차에서는 모든 분포에서 모형 1에 비해 모형 2가 우수했으나, 분포별로는 큰 차이를 보이지 않았다.

6. 초과손해액 보험과 보상 한도가 존재하는 보험의 보험료

본 장에서는 5장에서 구축한 모형을 토대로 4.1장 및 4.2장의 내용을 이용해 건물과 가재도구의 화재 피해에 대한 초과손해액 보험 및 보상 한도가 존재하는 보험의 보험료를 계산한다.

건물: 서울특별시 소재의 철근콘크리트조 아파트 17층에서 화재가 발생한 경우 각 분포 별로 초과점에 따른 초과손해액 보험의 보험료와 보상 한도에 따른 보험료를 계산해 보고자 한다. 이 건물의 바닥면적은 384m² 이고 소방서와 거리는 4.7km이며 소화기구와 스프링클러가 없고 화재경보기가 작동하지 않았다. Table 10 은 각 분포 별로 초과점에 따른 초과손해액 보험의 보험료를 계산한 결과이다. 초과점이 300만원, 500만원, 1,000만원으로 갈수록 초과손해액 보험의 보험료가 줄어드는 것을 확인할 수 있다. 이는 초과점이 커질수록 보험료가 작아짐을 뜻한다. 다만 정규분포에서 초과점이 300만원, 500만원일 때가 초과점이 없을 때보다 보험료가 커져 상식과는 상반되는 결과가 나온 것을 확인할 수 있다. 이를 통해 초과손해액 보험에서 피해 금액에 정규분포를 가정하는 것은 올바른 보험료 산정에 치명적임을 판단할 수 있다.

각 분포 별로 보상 한도 별 보험료를 계산한 결과는 Table 11에 있다. 한도가 3천만 원에서 5천만 원으로 증가할 때 보험료가 증가하는 것을 확인할 수 있다. 정규분포의 경우 1%가 증가하였고 로그정규분포의 경우 거의 증가하지 않았다. 화재 피해금액 분포는 소액에 대부분 몰려있기 때문에, 위와 같이 상대적으로 큰 값의 한도를 설정할 경우 보험료의 증가폭이 크지 않다.

Table 12: Insurance premium for excess loss insurance for equipment (unit: thousand won)

Excess point	Normal	Gamma	Log normal	Weibull
None	2,597	2,623	2,666	2,430
1,500	2,506	1,796	1,294	1,672
3,000	1,713	1,308	562	1,272
10,000	126	361	24	502

Table 13: Insurance premium for insurance with a compensation limit for equipment (unit: thousand won)

Policy limit	Normal	Gamma	Log normal	Weibull
15,000	2,589	2,468	2,661	2,131
30,000	2,597	2,609	2,666	2,344
100,000	2,597	2,623	2,666	2,428
None	2,597	2,623	2,666	2,430

가계도구: Table 12는 광주광역시 소재의 벽돌조 단독주택 1층에서 화재가 발생한 경우 각 분포 별로 초과점에 따른 초과손해액 보험의 보험료를 계산한 결과이다. 이 건물의 바닥면적은 $76.43m^2$ 이고 소방서와 거리는 $1.7km$ 이며 소화기구와 스프링클러가 없고 화재경보기가 작동하지 않았다. 건물과 마찬가지로 초과점이 150만원, 300만원, 1,000만원으로 갈수록 초과손해액 보험의 보험료가 줄어드는 것을 확인할 수 있다. Table 13는 각 분포 별로 보상 한도 별 보험료를 계산한 결과이다. 건물과 마찬가지로 한도가 증가할 때 보험료가 증가하는 것을 확인할 수 있다. 1,500만원에서 3,000만원으로 증가할 때 정규분포의 경우 0.3%, 감마분포의 경우 5.4%, 로그정규분포의 경우 0.2%, 와이불분포의 경우 9.1% 증가하였다. 특히 로그정규분포의 경우 건물에서와 마찬가지로 증가폭이 아주 작음을 확인할 수 있다.

7. 실무적 고려 사항

보험업 종사자가 실무를 할 때에는 다양한 고려 사항이 존재한다. 본 장에서는 본 연구의 내용이 실무에도 적용될 수 있도록 실무적 고려 사항을 추가로 다룬다.

7.1. 데이터 가공

범주형 변수 내에 변수의 개수가 너무 많거나 연속형 변수가 존재하는 등의 경우는 데이터 가공이 필요하다. 고려 가능한 데이터 가공 방법으로는 범주의 재분류와 연속형 변수의 범주화가 있다.

7.1.1. 범주 재분류

하나의 범주형 변수 내에 변수의 개수가 너무 많은 경우에는 범주를 통합 재분류했다. 소화 기구의 경우, 사용한 주택과 그렇지 않은 주택으로 구분했고, 표본 크기가 부족한 범주(고장, 노후, 약제 부족 등)는 데이터에서 제외했다. 주택 종류의 경우, 실무에서 효율을 적용하기 어려운 기타 주택, 기타공동주택, 기타단독주택은 데이터에서 제외했다. 건물 기둥 구조의 경우 유사 종류의 구조로 묶어 강구조, 조적조, 기타조 세개 범주로 재분류했다. 재분류 시 보험개발원의 화재보험 참조순보험요율 건물 규정을 활용했으며, 동 규정에 분류 기준이 나와있지 않은 건물 구조는 포스코 기술연구원의 자문 내용을 참고하여 재분류했다. 구체적으로 철골조, 철조, 샌드위치패널조는 강구조, 석조, 치장벽돌조, 벽돌조는 조적조, 그리고 블럭조, 컨테이너조, 비닐하우스 스파이프조는 기타조로 재분류했다. 건물 지붕 형태의 경우에도, 유사 종류의 지붕 형태끼리 재분류했다.

Table 14: Relative risk of buildings compared to average

Variable	Parameter	Relative risk	p-value
바닥면적	우량체	0.7558	0.000
	표준체	1.0110	0.490
	표준미달체	1.2279	0.000
소방서거리	우량체	0.8674	0.000
	표준체	1.0599	0.087
	표준미달체	1.1131	0.000
소화기구	사용	0.9303	0.000
	미사용	1.3605	0.000
단독경보형감지기	작동	0.8062	0.000
	미작동	1.1726	0.000

블럭조와 초가는 기타층으로, 한식기와와 시멘트기와는 와가로 재분류했다.

7.1.2. 연속형 변수의 범주화

일반화선형모형에서 연속형 변수를 독립 변수로 활용할 때, 연속형 변수가 극단적으로 크거나 작은 값을 지닐 경우 종속 변수에 미치는 영향이 지나치게 커질 수 있다. 이에 대한 해결책은 연속형 변수에 대한 범주화이다. 본 연구에서 범주화를 고려할 수 있는 연속형 변수는 주택 면적과 소방서와의 거리이다. 본 연구는 연속형 변수별로 각각 3개의 범주를 갖도록 범주화하고 위험 집단 별 상대 위험도의 가중 평균이 1이 되도록 범주 간 경계 값을 설정하였다. 즉, 표준적인 위험도를 가진 집단 대비, 위험도가 낮은 집단의 상대적 위험도와 위험도가 높은 집단의 상대적 위험도의 가중 평균이 1이 되도록 범주화했다. 본 논문에서는 지면의 한계로 변수 별로 각각 한 개의 경계값 조합을 사례로 사용한다. 주택 면적의 세 가지 범주는 $80m^2$ 미만, $80m^2$ 이상 $150m^2$ 미만, 그리고 $150m^2$ 이상을 사용한다. 또한, 소방서와의 거리에 대해서는 5.0km 미만, 5.0km 이상 11.0km 미만, 11.0km 이상을 사용한다. 이러한 범주 간 경계 값은 실무 상의 필요에 따라 바뀔 수 있다.

7.2. 결과

7.2.1. 누적분포함수 활용

본 연구는 보험 사고 심도 추정 시 보장 한도를 고려한다. 그리고 보장 한도를 고려하기 위해 4.2장의 사고 심도 확률밀도함수와 누적분포함수를 활용한다. 4.2장의 방법을 통해 다양한 한도에 대한 보험료를 개별적으로 산출할 수 있으며, 특정 몇 개의 한도에 대한 추정 결과를 변수 조합 별 테이블로도 보일 수 있다. 각 변수의 조합을 고려할 경우 테이블의 정보 양이 방대하므로, 본 논문에 테이블의 내용을 포함하지는 않는다.

위험률 테이블을 만들 때에는, 정보 양이 지나치게 방대해지는 것을 피하기 위해 지역, 주택 층, 스프링클러 변수는 보험료 산출 시 제외할 수 있다. 또한 본 연구에서 사용한 데이터는 소방청에서 집계한 화재 피해 금액이므로, 실제 보험 발생손해액은 이와 차이가 날 가능성이 존재한다. 따라서, 본 연구에서 예측한 화재 피해 금액을 실제 보험 발생 손해액 수준으로 맞춰주는 조정(Onlevel) 과정을 수행하기로 한다. 구체적인 방식은 아래와 같다.

- 소방청 연도 별 화재 건당 평균 피해 금액: (A).
- 업계 연도 별 화재보험 건당 평균 발생 손해액: (B).
- 조정 후 보험료: 식 4.8에 따른 보험료 ÷ (A) × (B).

Table 15: Relative risk of equipment compared to average

Variable	Parameter	Relative risk	p-value
바닥면적	우량체	0.7964	0.001
	표준체	0.9482	0.388
	표준미달체	1.1916	0.000
소방서거리	우량체	0.8903	0.003
	표준체	1.0231	0.533
	표준미달체	1.2165	0.000
소화기구	사용	0.9373	0.000
	미사용	1.3551	0.000
단독경보형감지기	작동	0.8392	0.000
	미작동	1.2108	0.000

7.2.2. 다른 기초 통계에 상대 위험도 적용

본 연구가 다른 위험 세분화 방법은 다른 기초 통계에도 적용될 수 있다. 예를 들어, 보험개발원의 참조순보험 효율에 본 연구의 일반화선형모형이 도출한 회귀 계수를 지수화(exponentiate) 하여 상대 위험도로 사용할 수 있다. 이 때, 지수화 방법은 3.2장의 정의를 따른다. 아울러, 보험 실무자로 하여금 상대 위험도의 의미를 보다 쉽게 이해할 수 있도록, 4개 이상의 범주를 가진 범주형 변수를 2개 혹은 3개의 범주로 재분류 할 수 있다. 본 장은 소화기구의 변수에 대해 소화약제 미충진, 소화약제 부족, 고장, 노후, 사용법 미숙지의 경우는 미사용으로 분류했으며, 기타 소화기구활용은 사용으로 분류했다. 단독경보형감지기에 대해서는 건전지방전, 건전지없음, 기타의 경우는 미작동으로 분류했다. 또한, 소화기구와 단독경보형감지기의 미상 값은 제거했으며, 결측 값은 나머지 범주로 재분류했다 (Rubin, 1986; Little, 1988). 이러한 재분류 방식이 변경될 경우, 상대 위험도는 바뀔 수 있다. 위 재분류 결과 각 변수 범주별 및 총 화재 건수는 Table A.7 및 Table A.8에 있다.

7.1.2장이 기술한 바와 같이, 연속형 변수를 범주화 할 때 범주 간 경계 값은 실무 상의 필요에 따라 바뀔 수 있으며, 이 경계값에 따라 상대 위험도 역시 달라질 수 있다. 그리고 다양한 경계 값 중 일부는 각 범주 별 상대 위험도와 각 비중을 곱한 가중 평균을 1로 맞추는 결과를 도출할 수 있다. 특히, 3.2장에 기술된 모형에 따라, 각 변수의 상대 위험도는 서로 곱해질 수 있으며, 곱해진 상대 위험도에 대해서도 가중 평균을 1로 맞추는 결과를 도출할 수 있다. 특히, 곱해진 상대 위험도의 가중 평균을 1로 맞추기 위해, 각 변수의 상대 위험도에 가중 평균의 거듭제곱근의 역수를 곱할 수 있다.

본 장에서 기술한 데이터 전처리 및 범주화 과정을 거쳤을 때, 건물 및 가재도구 피해금액에 대한 상대 위험도는 각각 Table 14 및 Table 15와 같다. 다만, 이 경우 7.2.1장의 방식과 달리 보상 한도 별로 차등적인 보험 효율을 적용하는 것이 제한된다. 또한 7.2.1장에서는 면적과 주택 종류 간 교호작용을 고려한 반면, 상대 위험도 적용시에는 해당 교호작용을 고려하지 않았다. 이는 상대 위험도를 단순화해, 보험 실무자나 일반인 으로 하여금 그 의미가 쉽게 와닿게 하기 위함이다. 마지막으로, 본 방법은 현행 참조순보험효율 방식을 따라, 지역을 효율 변수로 고려하지 않았다.

7.2.3. 실무적 적용: 위험의 등급화

7.2.2의 상대 위험도 적용 방법은 기존에 동일한 보험료를 부담하던 보험 계약자에게 차등적인 보험료를 선택할 기회를 제공한다는 순기능이 있다. 다만, 상대 위험도 조합이 가장 낮은 경우와 가장 높은 경우의 보험료 차이가 상당히 클 수 있으므로, 상대 위험도 조합을 일정 수의 등급으로 묶는 방법을 고려할 수 있다.

등급을 설정할 때에는 유사한 상대 위험도를 가진 위험 집단을 같은 등급으로 묶을 수 있다. 이러한 군집화 방법은 K-평균 군집화(K-means clustering) 방법에서 볼 수 있듯이 통계학 분야에서 빈번하게 사용된다. 또한,

각 등급이 유사한 크기의 모집단을 대표할 수 있도록 함으로써 전체 보험 계약자 집단이 보험료 차등화의 효과를 고르게 누리도록 할 수 있다. 예를 들어, 위험 집단을 총 5개의 등급으로 구분 시, 각 등급이 대표하는 모집단의 크기가 전체인 1을 등급의 개수인 5로 나눈 0.2가 되도록 할 수 있다. 이 경우, 각 등급 내 표본의 크기에 따라 0.2와 다소 차이가 발생할 수 있다.

등급화의 구체적인 과정은 다음과 같다. 본 논문은 대표 목적물인 건물의 피해금액을 기준으로 동 과정을 설명하지만, 유사 목적물인 가재도구에 대해서도 동일한 과정을 적용할 수 있다. 먼저, 7.2.2의 방법에 따라 변수 조합별 상대 위험도인 단순 상대도를 계산한 다음 유사한 단순 상대도를 가진 변수 조합들을 묶어 등급을 구성하고, 각 등급별로 전체 화재 건수 대비 비중을 계산한다. 다음으로, 각 변수 조합이 등급 내에서 차지하는 비중을 계산하고 이를 각 변수 조합의 단순 상대도와 곱해 각 변수 조합이 등급별 상대 위험도에 기여하는 정도를 계산한다. 마지막으로, 등급별 상대 위험도에 대한 기여 정도를 합산하여 등급별 상대 위험도를 계산한다. 본 등급화의 결과 및 과정에 대한 수치는 Table A.9, Table A.10, Table A.11에 있다.

8. 결론

본 연구에서는 주택 화재 데이터를 이용하여 초과손해액 보험과 보상 한도가 존재하는 보험에서의 분포 적합도에 관한 분석을 진행함으로써 분포의 선택이 각각의 보험료에 미치는 영향에 대해 알아보았다. 그 결과 꼬리가 두꺼운 분포를 사용하는 것이 초과손해액 보험과 보상 한도가 존재하는 보험 모두에서 더 좋은 적합도를 가지는 것으로 나타났으나, 데이터의 한계로 인해 감마분포와의 차이는 그리 크지 않은 것을 확인할 수 있었다.

초과손해액 보험의 경우, 1,000만원 이상의 건물 화재 피해금액 데이터에 대해 정규분포보다는 감마분포에서 약 15%, 와이불분포에서 약 15% 우수한 적합도를 보였다. 1,000만원 이상의 가재도구 피해금액 데이터에 대해서는 정규분포에 비해 감마분포가 약 26%, 로그정규분포가 약 29% 우수하였다. 보상 한도가 존재하는 보험의 경우, 3,000만원 이하의 건물 화재 피해금액 데이터에 대해 정규분포에 비해 감마분포와 로그정규분포가 약 14% 좋은 적합도를 가졌다. 1,500만원 이하의 가재도구 피해금액 데이터에 대해 정규분포보다 감마분포가 약 20%, 로그정규분포가 약 23% 우수하였다. 표준제공근오차에서는 모든 분포에서 모형 1에 비해 모형 2가 우수했으나, 분포별로는 큰 차이를 보이지 않았다.

본 논문의 연구 결과를 고려했을 때, 실제 보험 업계에서 보험료 산정 시 보험 손실 데이터의 특성을 반영할 수 있는 분포를 고려한다면 보험 회사와 가입자 양측의 손실을 줄이는 데 기여할 수 있을 것이라 기대된다. 특히 초과손해액 보험과 보상 한도가 존재하는 보험의 보험료를 책정함에 있어서는 사용한 분포에 따라 초과점 및 보상 한도에 따른 보험료의 변화 양상이 상이했으므로 데이터의 특성과 보험 제도를 고려한 분포를 사용하는 것이 중요하리라 판단된다. 마지막으로, 본 연구는 연구 내용이 실무에 적용되기 위한 고려 사항을 제시한다. 먼저 데이터 가공 방법으로 범주 재분류 및 연속형 변수의 범주화 방법을 설명한다. 다음으로, 누적분포함수 적용 방법 및 타 기초 통계에 적용될 수 있는 주요 변수 별 상대 위험도 적용 방법을 제안한다.

Appendix A: Figures and tables

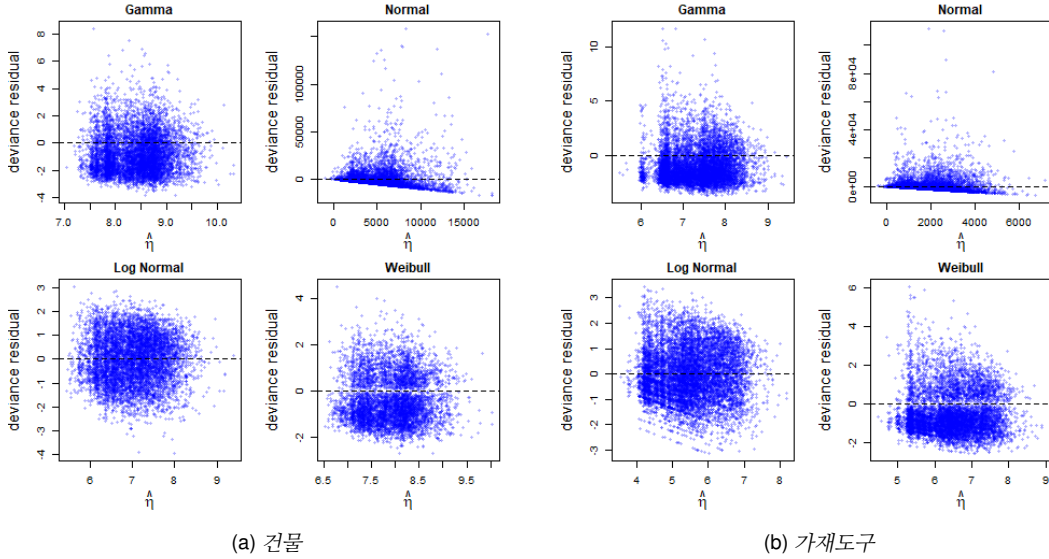


Figure A.1: Deviance residual plot by distribution.

Table A.1: Results of normal linear model analysis for building fire damage costs

Variable	Parameter	Estimate	p-value
지역(강원도 대비)	경상남도	-1922	0.0060
	광주광역시	-3068	0.0011
	대구광역시	-3074	0.0006
	대전광역시	-2833	0.0021
	부산광역시	-3095	0.0002
	서울특별시	-3175	0.0000
	세종특별자치시	-3405	0.0314
	울산광역시	-3273	0.0037
	인천광역시	-2214	0.0090
	전라북도	-2373	0.0037
	제주특별자치도	-2348	0.0411
건물 구조 (기타조 대비)	샌드위치패널조	3084	0.0019
	철골조	6728	0.0000
단독경보형감지기	화재경보기 작동	-397.1	0.6787
스프링클러	스프링클러 작동	-902.5	0.2015
소화기구 (결측 대비)	기타소화기구 활용	50.68	0.9239
	미상	2387	0.0006
	사용	-1113	0.0012
	사용법 미숙지	4403	0.0027
소방서거리 (km)	소방서거리 (km)	62.46	0.0000

Table A.2: Results of normal linear model analysis for fire damage costs of equipment

Variable	Parameter	Estimate	p-value
지역(강원도 대비)	대전광역시	-934.9	0.0163
	부산광역시	-1136	0.0008
	서울특별시	-998.5	0.0004
	울산광역시	-995.8	0.0206
주택 종류(다가구주택 대비)	연립주택	-1185	0.0167
건물 구조(기타조 대비)	샌드위치패널조	1022	0.0295
단독경보형감지기	화재경보기 작동	-233.1	0.4809
스프링클러	스프링클러 작동	-31.19	0.9090
소화기구 (결측 대비)	기타소화기구 활용	107.1	0.6092
	미상	1446	0.0000
	사용	-48.06	0.7413
	사용법 미숙지	1306	0.0373
소방서거리 (km)	소방서거리 (km)	34.6	0.0000

Table A.3: Results of log-normal generalized linear model analysis for building fire damage costs

Variable	Parameter	Estimate	p-value
지역(강원도 대비)	경상남도	-0.5123	0.0000
	광주광역시	-0.4958	0.0023
	대구광역시	-0.5169	0.0008
	대전광역시	-0.5005	0.0020
	부산광역시	-0.9198	0.0000
	서울특별시	-0.7467	0.0000
	울산광역시	-0.6796	0.0004
	전라북도	-0.2948	0.0201
주택 종류(다가구주택 대비)	다세대주택	-0.4120	0.0432
건물 구조(기타조 대비)	철골조	0.5096	0.0054
	철근콘크리트조	-0.3018	0.0487
	컨테이너조	-0.5726	0.0241
단독경보형감지기	화재경보기 작동	-0.0527	0.7500
스프링클러	스프링클러 작동	0.1392	0.2535
소화기구 (결측 대비)	기타소화기구 활용	-0.1530	0.0946
	미상	0.3597	0.0028
	사용	-0.3721	0.0000
	사용법 미숙지	0.6853	0.0068
소방서거리 (km)	소방서거리 (km)	0.0161	0.0000

Table A.4: Results of log-normal generalized linear model analysis for fire damage costs of equipment

Variable	Parameter	Estimate	<i>p</i> -value
지역(강원도 대비)	경기도	0.3009	0.0056
	경상북도	0.3969	0.0017
	부산광역시	-0.3008	0.0024
	서울특별시	-0.7025	0.0000
	울산광역시	-0.5422	0.0013
	인천광역시	0.2733	0.0541
	전라남도	0.3649	0.0063
	충청남도	0.5594	0.0000
주택 종류(다가구주택 대비)	충청북도	0.3649	0.0000
	아파트	-0.4318	0.0107
건물 구조(기타조 대비)	연립주택	-0.5487	0.0047
	벽돌조	-0.3087	0.0421
	철근콘크리트조	-0.4529	0.0029
단독경보형감지기	컨테이너조	0.6029	0.0262
	화재경보기 작동	-0.5140	0.0000
스프링클러	스프링클러 작동	0.3626	0.0007
	기타소화기구 활용	-0.1006	0.2212
소화기구 (결측 대비)	미상	0.3902	0.0008
	사용	0.2144	0.0002
	사용법 미숙지	0.8676	0.0004
소방서거리 (km)	소방서거리 (km)	0.0175	0.0000
바닥면적(m^2)	바닥면적(m^2)	0.0001	0.0076

Table A.5: Results of Weibull distribution generalized linear model analysis for building fire damage costs

Variable	Parameter	Estimate	p-value
지역(강원도 대비)	경상남도	-0.4359	0.0000
	광주광역시	-0.6678	0.0000
	대구광역시	-0.6729	0.0000
	대전광역시	-0.6337	0.0000
	부산광역시	-0.9189	0.0000
	서울특별시	-0.8259	0.0000
	세종특별자치시	-0.7970	0.0016
	울산광역시	-0.7616	0.0000
	인천광역시	-0.4239	0.0016
	전라북도	-0.4027	0.0006
건물 구조(기타조 대비)	샌드위치패널조	0.4373	0.0058
	철골조	0.7328	0.0000
	컨테이너조	-0.7548	0.0012
단독경보형감지기	화재경보기 작동	0.0074	0.9611
스프링클러	스프링클러 작동	-0.1267	0.2601
소화기구 (결측 대비)	기타소화기구 활용	-0.0930	0.2712
	미상	0.3718	0.0008
	사용	-0.3635	0.0000
	사용법 미숙지	0.6711	0.0041
소방서거리(km)	소방서거리(km)	0.0127	0.0000

Table A.6: Results of Weibull distribution generalized linear model analysis for fire damage costs of equipment

Variable	Parameter	Estimate	p-value
지역(강원도 대비)	대전광역시	-0.3733	0.0165
	부산광역시	-0.5909	0.0000
	서울특별시	-0.8679	0.0000
	울산광역시	-0.6635	0.0001
	충청남도	0.3235	0.0165
	충청북도	0.3930	0.0104
주택 종류(다가구주택 대비)	아파트	-0.3804	0.0273
	연립주택	-0.6760	0.0006
건물 구조(기타조 대비)	샌드위치패널조	0.4209	0.0250
단독경보형감지기	화재경보기 작동	-0.5217	0.0001
스프링클러	스프링클러 작동	0.1515	0.1652
소화기구 (결측 대비)	기타소화기구 활용	0.0075	0.9282
	미상	0.6753	0.0000
	사용	0.1013	0.0818
	사용법 미숙지	0.8910	0.0003
소방서거리(km)	소방서거리(km)	0.0166	0.0000
바닥면적(m ²)	바닥면적(m ²)	0.0001	0.0019

Table A.7: Number of fires by category within variables for building fire damage costs

Variable	Parameter	Number of fires	Weight
바닥면적	우량체	2,066	0.289
	표준체	2,837	0.397
	표준미달체	2,242	0.314
Total		7,145	1.000
소방서거리	우량체	3,073	0.430
	표준체	1,758	0.226
	표준미달체	2,314	0.324
Total		7,145	1.000
소화기구	사용	6,016	0.842
	미사용	1,129	0.158
Total		7,145	1.000
단독경보형감지기	작동	3,269	0.458
	미작동	3,876	0.542
Total		7,145	1.000

Table A.8: Number of fires by category within variables for fire damage costs of equipment

Variable	Parameter	Number of fires	Weight
바닥면적	우량체	2,394	0.251
	표준체	3,580	0.376
	표준미달체	3,550	0.373
Total		9,524	1.000
소방서거리	우량체	5,076	0.533
	표준체	2,280	0.239
	표준미달체	2,168	0.228
Total		9,524	1.000
소화기구	사용	8,157	0.856
	미사용	1,367	0.144
Total		9,524	1.000
단독경보형감지기	작동	5,279	0.554
	미작동	4,245	0.446
Total		9,524	1.000

Table A.9: Relative risk by grade for building fire damage costs

Grade	Relative risk	Number of fires	Weight
1	0.602	1,242	0.174
2	0.794	1,549	0.217
3	0.944	1,659	0.232
4	1.194	1,618	0.226
5	1.551	1,077	0.151
Total		7,145	1.000

Table A.10: Relative risk by grade for fire damage costs of equipment

Grade	Relative risk	Number of fires	Weight
1	0.626	1,907	0.200
2	0.813	2,266	0.238
3	0.955	1,649	0.173
4	1.152	1,835	0.193
5	1.499	1,867	0.196
Total		9,524	1.000

Table A.11: Calculation of relative risk by grade for building fire damage costs

Floor* area	Distance*	Fire* extinguisher	Alarm*	Number of fires	Weight	Relative risk	Grade	Weight within grade	Relative risk by grade
T1	T1	1	1	342	0.0479	0.4917	1	0.2754	0.1354
T1	T2	1	1	144	0.0202	0.6008	1	0.1159	0.0697
T1	T3	1	1	178	0.0249	0.6310	1	0.1433	0.0904
T2	T1	1	1	578	0.0809	0.6577	1	0.4654	0.3061
T1	T1	0	1	52	0.0073	0.7191	2	0.0336	0.0241
T1	T1	1	0	222	0.0311	0.7152	2	0.1433	0.1025
T3	T1	1	1	707	0.0990	0.7988	2	0.4564	0.3646
T2	T2	1	1	284	0.0397	0.8037	2	0.1833	0.1473
T1	T2	0	1	30	0.0042	0.8786	2	0.0194	0.0170
T2	T3	1	1	254	0.0355	0.8440	2	0.1640	0.1384
T1	T2	1	0	231	0.0323	0.8739	3	0.1392	0.1217
T1	T3	0	1	50	0.0070	0.9227	3	0.0301	0.0278
T1	T3	1	0	517	0.0724	0.9177	3	0.3116	0.2860
T2	T1	0	1	73	0.0102	0.9619	3	0.0440	0.0423
T2	T1	1	0	324	0.0453	0.9566	3	0.1953	0.1868
T1	T1	0	0	79	0.0111	1.0459	3	0.0476	0.0498
T3	T2	1	1	297	0.0416	0.9761	3	0.1790	0.1747
T3	T3	1	1	88	0.0123	1.0251	3	0.0530	0.0544
T3	T1	0	1	62	0.0087	1.1682	4	0.0383	0.0448
T2	T2	0	1	36	0.0050	1.1753	4	0.0222	0.0262
T3	T1	1	0	494	0.0691	1.1619	4	0.3053	0.3547
T2	T2	1	0	319	0.0446	1.1689	4	0.1972	0.2305
T2	T3	0	1	48	0.0067	1.2343	4	0.0297	0.0366
T1	T2	0	0	44	0.0062	1.2780	4	0.0272	0.0348
T2	T3	1	0	615	0.0861	1.2276	4	0.3801	0.4666
T1	T3	0	0	177	0.0248	1.3421	5	0.1643	0.2206
T2	T1	0	0	78	0.0109	1.3990	5	0.0724	0.1013
T3	T2	0	1	24	0.0034	1.4275	5	0.0223	0.0318
T3	T2	1	0	252	0.0353	1.4197	5	0.2340	0.3322
T3	T3	0	1	22	0.0031	1.4991	5	0.0204	0.0306
T3	T3	1	0	170	0.0238	1.4910	5	0.1578	0.2353
T3	T1	0	0	62	0.0087	1.6991	5	0.0576	0.0978
T2	T2	0	0	65	0.0091	1.7095	5	0.0604	0.1032
T2	T3	0	0	163	0.0228	1.7953	5	0.1513	0.2717
T3	T2	0	0	32	0.0045	2.0762	5	0.0297	0.0617
T3	T3	0	0	32	0.0045	2.1804	5	0.0297	0.0648
Total				7,145					

*Fire area, Distance) T1: 우량체, T2: 표준체, T3: 표준미달체

*Fire extinguisher) 1: 사용, 0: 미사용

*Alarm) 1: 작동, 0: 미작동

References

- Ahn S, Kim JH, and Ramaswami V (2012). A new class of models for heavy tailed distributions in finance and insurance risk, *Insurance: Mathematics and Economics*, **51**, 43–52.
- Agresti A (2015). *Foundations of Linear and Generalized Linear Models*, John Wiley & Sons, New Jersey.
- Asmussen SR (2003). Applied probability and Queues, *Stochastic Modelling and Applied Probability*, **51**, 266–301.
- Foss S, Korshunov D, and Zachary S (2013). *An Introduction to Heavy-Tailed and Subexponential Distributions*, Springer Science & Business Media, New York.
- Gray RJ and Pitts SM (2012). *Risk Modelling in General Insurance: From Principles to Practice*, Cambridge University Press, New York.
- Klugman SA, Panjer HH, and Willmot GE (2012). *Loss Models: From Data to Decisions*, John Wiley & Sons, New York.
- Lee A, Ahn J, Mun H, Nam K, Park Y, Song I, and Park S (in press). Introducing new rate factors and statistical learning to improve fire loss prediction accuracy, *The Korean Journal of Applied Statistics*.
- Little RJA (1988). Missing-data adjustments in large surveys (with discussion), *Journal of Business Economics and Statistics*, **6**, 287–301.
- Nelder JA and Wedderburn RW (1972). Generalized linear models, *Journal of the Royal Statistical Society: Series A (General)*, **135**, 370–384.
- National Fire Service Academy (2019). Fire investigation practice, Technical Paper, Retrieved Mar. 10, 2023, Available from: <https://www.nfsa.go.kr/nfsa/releaseinformation/archive/materials/>
- Oh R, Jeong H, Ahn JY, and Valdez EA (2021). A multi-year microlevel collective risk model, *Insurance: Mathematics and Economics*, **100**, 309–328.
- Punzo A, Bagnato L, and Maruotti A (2018). Compound unimodal distributions for insurance losses, *Insurance: Mathematics and Economics*, **81**, 95–107.
- Park H (2022). Case studies on the fire safety in attic space of multi-family housing (Master's thesis), Seoul National University of Science and Technology, Seoul.
- Rubin DB (1986). Statistical matching using file concatenation with adjusted weights and multiple imputations, *Journal of Business Economics and Statistics*, **4**, 87–94.
- Sarabia JM, Jordá V, Prieto F, and Guillén M (2020). Multivariate classes of GB2 distributions with applications, *Mathematics* **2021**, **9**, 72.
- Sigman K (1999). Appendix: A primer on heavy-tailed distributions, *Queueing Systems*, **33**, 261–275.

Received May 1, 2023; Revised June 7, 2023; Accepted July 5, 2023

통계적 분포를 통한 주택 화재 심도 추정

박윤정^a, 송인아^a, 이소연^a, 남광현^b, 오로지^{1,c}, 안재윤^{2,a}

^a이화여자대학교 통계학과; ^b주식회사 언더라이터; ^c육군사관학교 수학과

요약

본 논문은 보험 손실 추정을 위한 회귀 설정에서 다양한 분포 가정의 성능을 실제 데이터를 사용하여 비교 분석합니다. 감마 분포는 일반적으로 보험의 손실 분포의 비대칭성을 처리하는 데 사용됩니다. 그러나 최근 연구는 보험자료의 분석에 있어서 손실 분포의 두꺼운 꼬리의 중요성을 강조합니다. 실제 주택 화재 보험 데이터 분석을 통해 우리는 회귀 방법에서 다양한 분포 가정의 효과를 비교합니다. 우리의 결과는 보험손실에 대한 분포 가정의 선택이 “초과 손해 보험” 및 “한도 보험”을 포함한 다양한 보험 상품의 보험료 결정에 중요하다는 것을 보여줍니다. 또한 주택 화재 보험의 통계적 모형 설정에 있어서 실제 고려 사항에 대해 논의합니다.

주요용어: 주택화재보험, 통계적 학습, 일반화 선형모형, 감마 분포

이 성과는 서울대학교 경영대학 경영연구소 연구비 지원과 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원 (No. RS-2023-00217022) 을 받아 수행된 연구임.

¹교신저자: (01805) 서울특별시 노원구 사서함 77-2, 육군사관학교 수학과. E-mail: rosy.oh5@gmail.com

²교신저자: (03760) 서울특별시 서대문구 이화여대길 52, 이화여자대학교 통계학과. Email: jaeyahn@ewha.ac.kr