

[Research Paper]

로지스틱회귀분석 모델을 활용한 화학사고 사상사고 예측모형 개발 연구

이태형 · 박춘화* · 박효현** · 곽대훈****

화학물질안전원 공업연구사/충남대학교 과학수사학과 대학원생, *화학물질안전원 공업연구관,
화학물질안전원 주무관, *충남대학교 국가안보융합학부 & 과학수사학과 부교수

A Study on Accident Prediction Models for Chemical Accidents Using the Logistic Regression Analysis Model

Tae-Hyung Lee · Choon-Hwa Park* · Hyo-Hyeon Park** · Dae-Hoon Kwak****

Researcher, National Institute of Chemical Safety / Graduate Student, Dept. of Crime & Forensic Science, Chungnam National Univ.,

*Senior Researcher, National Institute of Chemical Safety,

**Researcher, National Institute of Chemical Safety,

***Associate Professor, School of Integrated National Security & Dept. of Crime & Forensic Science, Chungnam National Univ.

(Received August 26, 2019; Revised October 20, 2019; Accepted November 1, 2019)

요 약

본 연구를 통해 화학사고 사상사고 예측모형을 개발하였다. 모형은 로지스틱회귀분석 모델을 활용하여 사상사고에 영향을 주는 변수를 도출하여 적용하였고, 통계적 검증방법과 오즈비를 활용하여 모형의 신뢰성 및 정확성을 검증하였다. 모형에 활용한 사고 자료는 과거 발생했던 화학사고 통계를 분석하여 활용하였으며, 사고의 유형, 원인, 발생 장소, 사상자 현황 및 사상자를 발생시킨 화학사고 등의 자료 분석을 통해 통계적으로 유의하게 나타난 독립변수($p < 0.05$)를 적용하였다. 본 연구에서 개발한 모형은 사업장에서 화학사고로 인해 발생하는 사상사고의 예방 및 안전시스템 구축을 위한 연구로서 의미가 있다고 할 수 있다. 모형에 의한 분석결과 사상사고 발생에 가장 크게 영향을 미치는 변수는 폭발에 의한 화학사고인 것으로 조사되었다. 따라서 사업장에서 발생하는 폭발 유형의 화학사고를 예방하기 위한 대책마련이 시급하다고 판단된다.

ABSTRACT

Through this study, we developed a model for predicting chemical accidents lead to casualties. The model was derived from the logistic regression analysis model and applied to the variables affecting the accident. The accident data used in the model was analyzed by studying the statistics of past chemical accidents, and applying independent variables that were statistically significant through data analysis, such as the type of accident, cause, place of occurrence, status of casualties, and type of chemical accident that caused the casualties. A significance of $p < 0.05$ was applied. The model developed in this study is meaningful for the prevention of casualties caused by chemical accidents and the establishment of safety systems in the workplace. The analysis using the model found that the most influential factor in the occurrence of casualty in accidents was chemical explosions. Therefore, there is an urgent need to prepare countermeasures to prevent chemical accidents, specifically explosions, from occurring in the workplace.

Keywords : Chemical accident, Accident prediction, Logistic regression, Casualty accident

1. 서 론

2012년 9월 27일 경북 구미에서는 불화수소 화학제품 생산 업체인 (주)휴브글로벌 구미 공장에서 탱크로리에 실린 불화수소 가스를 공장 내 설비에 주입하던 중 근로자의 실

수로 탱크로리의 밸브가 열리면서 가스가 유출되어 공장 근로자 5명이 사망하고 18명이 부상을 당했다. 불화수소 가스 누출 사고의 직접적인 원인은 작업 노동자의 실수 때문이었지만, 안전한 작업을 위한 개인 보호장구 미착용과 안전절차의 규정 미준수 등 작업자의 안일한 생각과 이를

†Corresponding Author, E-Mail: dkwak@cnu.ac.kr TEL: +82-42-821-5298, FAX: +82-42-822-5236

© 2019 Korean Institute of Fire Science & Engineering. All right reserved.

감독해야 할 사업장 관리자의 허술한 감독 그리고 사고 예방 및 대응을 위한 법과 제도적 장치 미흡 또한 본 사고가 발생하게 된 책임적인 측면에서 자유로울 수 없다. 구미 불화수소 가스 누출사고 이후 화학물질의 잠재 위험성에 대한 국민의식이 높아지게 되었고, 2013년 5월에는 「화학물질관리법」을 제정하고 2015년에 이르러 시행하게 되었다. 이에 따라 화학물질의 체계적인 관리와 화학사고 예방을 통해 국민 건강 및 환경을 보호하기 위한 목적으로 화학물질에 대한 통계조사 및 정보체계구축, 유해화학물질 취급 및 설치·운영기준 구체화 등의 안전관리를 강화하였다. 또한, 화학사고 장외영향평가제도, 위해관리계획 및 영업허가제 신설 등을 통해 유해화학물질 예방관리 체계가 강화되었고, 화학사고의 발생 시 즉시 15분 이내 신고의무, 사고수습을 위한 현장조정관 파견 등 화학사고의 대비·대응이 대폭 강화되었다. 이러한 제도들의 시행으로 2015년에는 화학사고 발생건수가 113건에서, 2018년에는 66건이 발생하여 점차 감소하고 있는 추세이다. 또한, 인명피해가 발생한 화학사고는 2015년에 48건이었고, 2018년에는 17건으로 감소하였다⁽¹⁾. 하지만, 화학사고는 사고관련 통계 수치의 감소만으로 안심할 수 있는 사안이 아니다. 화학물질은 폭발성, 인화성, 부식성, 급성독성 등의 특성을 지니고 있어 사고의 발생 가능성이 높거나 사고가 발생한 경우 그 피해 규모가 크게 발생할 수 있으며, 물질 특성에 따라 잔류성이 높고, 확산성이 있어 사고가 발생할 경우 환경, 인명, 재산 등에 심각한 손실 및 피해를 유발하는 경우가 많다⁽²⁾. 「화학물질관리법」에 따르면, 화학사고는 시설의 교체 등 작업 시 작업자의 과실, 시설결함·노후화, 자연재해, 운송사고 등으로 인하여 화학물질이 사람이나 환경에 유출·누출되어 발생하는 일체의 상황을 말하는데⁽³⁾, 이러한 사고는 위험물을 취급하는 화학공장 등의 사업장에서는 화재, 폭발 및 독성물질 누출 등에 의한 화학사고 발생가능성이 상존하고 있다⁽⁴⁾. 20세기 최악의 참사로 기록되고 있는 1984년 인도 보팔참사에서도 알 수 있듯이 화학사고는 단 한 번의 사고로 수십만명의 사상자를 발생시킬 수 있다. 그렇기 때문에 사고의 발생 빈도와 인명피해 사고 빈도가 감소했다고 안심해서는 안된다. 따라서, 여러 화학사고 예방과 대응을 위한 제도의 시행으로 사고의 발생이 감소추세에 있는 현 시점에서 중요하게 생각해야 할 것은 단순히 사고 통계치의 감소보다 문제점을 확인하고 개선하는 작업

이 필요하다. 즉, 과거 발생된 사고들에 대하여 구체적으로 규명하고 분석하여 핵심 원인을 찾아 이를 개선할 수 있는 방안을 도출하고, 정부차원에서 개선 방안을 우리의 안전 및 예방 시스템에 적용하는 방안에 대한 연구와 제도적 보완 등 다각적인 정책적인 노력이 필요한 시점이다.

Table 1은 2013년부터 2018년까지 화학사고 및 사상자 현황 통계를 나타낸다. 사고 현황 통계를 살펴보면, 사업장이 기록한 통계치가 화학사고, 화학사고 사상사고, 사망자, 부상자 등 모든 지표에서 가장 높은 통계치를 보이고 있다. 화학사고 발생건수의 경우 두 번째로 많이 발생한 장소인 도로상에서 보다 약 3배가량 많은 것으로 조사되었으며, 사상사고, 사망, 부상자 등의 지표에서는 각각 약 6배, 8배, 8배가량 높게 발생한 것으로 조사되어, 사업장에서 발생하는 화학사고에 대한 대책마련의 필요성을 보여준다.

이에, 본 연구에서는 사업장에서 발생한 화학사고의 사상사고 중 최근 2013년부터 2018년까지의 6개년 동안의 사고 발생수와 사고자 수가 가장 많았던 사고 장소 중 사업장을 표본(Sample)으로 사고 유발 요인에 대한 변수(Variable)를 도출하여 변수 간 사고에 미치는 특성 분석과 사고 요인별 상관관계 데이터마이닝(Data mining) 기법 중 로지스틱 회귀분석 모델을 통해 사고예측 모형을 개발함으로써, 화학사고 예방 및 효과적인 안전대책에 기여 하고자 한다.

2. 사고 예측모형 선행연구

예측모형을 이용한 사고 발생 확률에 대한 연구사례로 Jin과 Lee⁽⁵⁾는 로지스틱회귀분석 방법에 서울지하철 역사 내 에스컬레이터 전도사고 발생에 영향을 주는 변수를 적용하여 도시철도 사상사고 예측모형을 제시하였다. Kang 등⁽⁶⁾은 도시부 신호교차로에서의 유턴 사고모형 개발을 위해 창원시 3지, 4지 신호교차로의 유턴 교통사고 자료에 대한 특성을 분석하고 로지스틱회귀분석에 의해 사고예측모형을 추정하였다. Park 등⁽⁷⁾은 충북지방경찰청의 사고자료와 현장조사 자료를 근거로, 교통사고와 관련된 기하구조 요소, 환경 요소 등을 분석하여 청주시 4지 신호교차로의 사고위치별 로지스틱 회귀 교통사고 모형을 개발하였다. 본 모형의 Hosmer & Lomeshow 검정은 유입부를 제외한 모형들은 p값이 0.05보다 크기 때문에 통계적으로 적합한 것으로 평가된다. Yan 등⁽⁸⁾은 후미추돌사고와 운전자의 특

Table 1. Status of Chemical Accidents and Casualty by Types of Places (From 2013 to 2018)

Accident place	Chemical accidents	Chemical accident with casualty	Death	Injury
Workplace	331	119	30	398
Road	112	20	4	50
Educational institutions	47	16	0	44
Laboratory	10	7	0	12
Etc	30	7	1	34

성, 도로환경 및 차량 유형으로 구분된 일련의 잠재적인 사고요인과의 관계를 조사하여 설명하였으며, 분석방법에는 Quasi-induced exposure technique와 다중 로지스틱회귀모형을 사용하였다. Park 등⁹⁾은 예측모형을 활용하여 철도사상사고 시나리오 모형 구성 및 위험사건별 등가 사망지수로 환산된 정량적인 위험도 평가모형을 제시하였다.

본 연구에서 화학사고 사상사고의 예측을 위해 모델에 적용된 독립변수는 사업장 규모, 사고 유형, 사고 원인 등이고, 이는 양적인 척도가 아닌 질적인(Qualitative) 범주형 척도로서 이를 크게 원인-결과 이분형 변수(Binary variable)로 나타낼 수 있다. 따라서 원인에 해당되는 독립변수와 결과에 해당되는 종속변수 간 관계를 나타내는 이분형 변수 형태인 로지스틱회귀분석을 사용하여 사고를 분석할 필요가 있다. 독립변수를 이용한 종속변수의 예측은 선형회귀분석을 통해서도 가능하지만 종속변수가 이분형 변수인 경우에는 선형회귀분석을 이용하여 독립변수와 종속변수간의 관계를 분석할 수 없다. 이는 선형회귀분석을 적용하기 위해서는 종속변수가 정규분포, 등분산성, 독립성을 따르는 연속형 변수이어야 하기 때문이다¹⁰⁾.

3. 화학사고 사상사고의 로지스틱회귀분석 모델

본 연구에서는 로지스틱회귀분석 모델을 활용하여 사업장에서 발생하는 화학사고의 사상사고 예측모형을 개발하였으며, 이를 적용해 사상사고에 영향을 미치는 요인을 분석하였다(Figure 1). 사업장에서 발생하는 화학사고 사상사고에 영향을 줄 것으로 판단한 사업장 규모, 사고원인, 사고유형, 장소 등을 주요 요인으로 한정하여 선정하였고 이들 사상사고 유발 요인을 로지스틱회귀분석의 독립변수로 적용하였다. 또한, 화학사고 발생 시 이들 독립변수들에 따른 사상자 발생 여부를 종속변수로 적용하였다. 이는 독립변수인 사상사고 요인들이 종속변수인 사상사고 발생 여부

에 어떤 영향을 주는지 분석함으로써 사상사고 예방을 위한 대책마련을 연구하는데 기여할 수 있을 것으로 사료된다.

화학사고의 사상사고 관련 자료는 화학물질안전원에서 제공하는 “화학안전정보 공유시스템”의 2013년부터 2015년까지 최근 6년간 발생한 530개의 사고 자료를 분석하여 활용하였고, IBM SPSS Statistics 25.0 Software를 이용하여 로지스틱회귀분석을 수행하였다. 통계치의 유의수준(p-Value)은 0.05를 기준으로 분석하였다.

3.1 로지스틱회귀분석(Logistic regression analysis)

로지스틱회귀분석은 특정 사고의 발생 여부를 직접 예측하는 것이 아니라 그 사고가 발생할 확률을 추정하는 분석 방법의 일종이다. 이항 로지스틱 회귀분석의 경우 종속변수가 0, 1만의 값을 갖는 가변수(Dummy variable)인 경우에 y의 기댓값을 나타내는 반응함수의 모양이 다음의 Figure 2와 같이 S형 곡선을 그리는 경우가 실제로 많이 나타난다. 종속변수의 예측 값은 항상 0과 1사이 확률 값을 갖게 되며, 따라서, 본 연구에서는 값이 0.5보다 크면 종속변수인 “사상사고 발생”의 발생 확률이 높고 0.5보다 작으면 사상사고 발생확률이 낮다고 예측할 수 있다. 선형회귀분석에서는 독립변수와 종속변수간의 선형관계를 가정하지만 로지스틱회귀분석에서는 사고 원인이 되는 독립변수와 사고 결과인 종속변수 간 비선형인 S곡선의 형태를 갖는다⁹⁾.

로지스틱회귀분석은 독립변수와 이분형 종속변수 간의 관계를 분석하는데 있어서 다른 회귀분석 방법에 비해 매우 유연하게 사용할 수 있다. 이에, 본 연구에서 화학사고의 사상사고 예측을 위해 로지스틱회귀분석 모델을 개발하여 사용하였다.

본 연구의 로지스틱회귀분석 모델은 다음의 식(1)과 같으며, 화학사고 사상사고 발생 확률을 추정한다. 로지스틱함수는 x가 증가함에 따라 E(y)의 값이 1로 서서히 수렴하는 양상을 보인다.

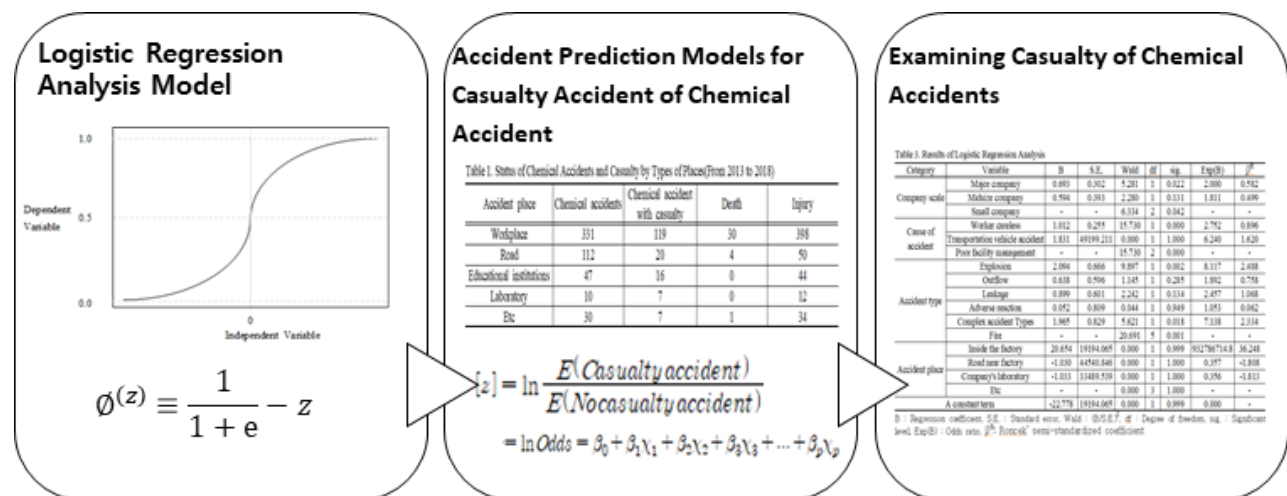


Figure 1. Study method schematic.

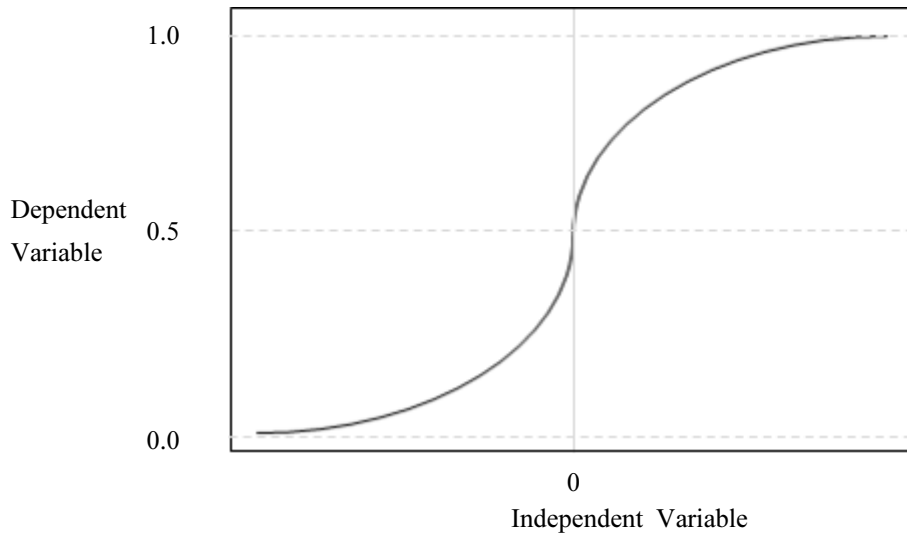


Figure 2. Plot of logistic.

$$E(y) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \quad (1)$$

여기서, 기대반응 $E(y)$ 는 확률을 의미하므로 본 연구에서는 화학사고 사상사고 발생 확률을 의미하며, β_0 와 β_1 은 로지스틱회귀계수로 양(+)의 값 또는 음(-)의 값을 취한다. x 는 독립변수, e 는 자연로그의 밑수로서 약 2.718을 나타낸다.

로지스틱회귀분석이란 단지 두 개의 값만을 가지는 종속 변수와 독립변수들 간의 인과관계를 로지스틱함수를 이용하여 추정하는 통계기법이다. 식(1)의 로지스틱함수는 β_0 와 β_1 에 대하여 비선형함수이나 이를 선형으로 변환시킬 수 있다. 기대반응 $E(y)$ 는 확률을 의미하므로, $E(y) = P_x$ 으로 표현할 수 있으며, 식(2)와 같이 변환을 통해 선형화된다.

$$\ln\left(\frac{pX}{1-pX}\right) = z = \beta_0 + \beta_1 x_1 \quad (2)$$

식(2)와 같은 변환을 로지스틱 변환(Logistic transformation)이라고 부며, 좌변의 $\frac{pX}{1-pX}$ 를 오즈(Odds)라 한다. 독립변수의 수가 두 개 이상인 경우에도 로지스틱 회귀모형이 가능하며 이 경우에는 $z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_p x_p$ 으로 나타내고, 로지스틱 회귀모형은 $\ln\left(\frac{pX}{1-pX}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_p x_p$ 가 된다.

본 연구에서는 독립변수가 다수이므로 로지스틱회귀 모델은 식(3)과 같이 표현될 수 있다.

$$E(y) = \frac{1}{1 + e^{-z}} \quad (3)$$

여기서, z (독립변수의 선형결과 값) = $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$ p 는 독립변수의 개수로서 화학사고 사상사고(Casualty accident)에 영향을 미치는 변수를 나타낸다.

$$e^{-z} = \frac{1 - E(\text{Casualty accident})}{E(\text{Casualty accident})} = \frac{E(\text{No casualty accident})}{E(\text{Casualty accident})} \quad (4)$$

$$[z] = \frac{E(\text{Casualty accident})}{E(\text{No casualty accident})} \quad (5)$$

또한 식(5)와 같이 사상사고가 발생할 확률에 대한 비사상사고(No casualty accident) 발생확률 비율의 로그 값으로 표현되며 본 연구에서는 이 식의 우변 $\frac{E(\text{Casualty accident})}{E(\text{No casualty accident})}$ 을 오즈로 하고 이 로그 값을 취한 값을 식(6)과 같이 표현할 수 있다.

$$[z] = \ln\left(\frac{E(\text{Casualty accident})}{E(\text{No casualty accident})}\right) = \ln Odds = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_p x_p \quad (6)$$

따라서 식(6)은 본 연구의 화학사고 사상사고 예측모형으로 활용할 수 있으며, 회귀계수가 0보다 크면 오즈비는 1보다 작게 된다. 회귀계수가 0이면 오즈비는 1이 된다. 특정 독립변수에 대한 오즈비가 1보다 크다는 것은 그 독립변수를 한 단위 증가시키면 오즈, 즉 심각한 사고 발생 비율이 증가한다는 뜻이며 이는 특정 독립변수에 대한 오즈비 값이 큰 값 일수록 사상사고가 발생할 가능성이 높아진다는 의미이다. 반대로 특정 독립변수에 대한 오즈비가 1보다 작다는 의미는 그 독립변수를 한 단위 증가시키면 오즈비 즉 사상사고가 발생할 가능성이 낮아진다는 의미이다.

3.2 변수의 선정

환경부 화학물질안전원에서 운영 중인 “화학안전정보

Table 2. Variables Used in Accident Prediction Models and Frequency Analysis

Category		Variable	Data coding
Independent variable	Company scale	Major company	1
		Midsized company	2
		Small company	3
	Cause of accident	Worker careless	1
		Transportation vehicle accident	2
		Poor facility management	3
	Accident type	Explosion	1
		Outflow	2
		Leakage	3
		Adverse reaction	4
		Complex accident Types	5
		Fire	6
	Accident place	Inside the factory	1
		Road near factory	2
		Company's laboratory	3
		Etc	4
Dependent variable	Human damage	No casualty accident	0
		Casualty accident	1

공유시스템”(www.csc.me.go.kr)에서는 국내에서 발생한 화학사고에 대하여 발생 요인 특성에 따라 분류하고 이에 대한 사고 통계정보를 제공하고 있으며⁽¹¹⁾, 본 연구에서는 “화학안전정보 공유시스템”의 사업장 화학사고 발생 요인 특성, 즉, 사업장 규모, 사고원인, 사고유형, 사고장소 등의 카테고리화 관련된 변수를 화학사고 사상사고 예측모형을 위한 독립변수로 선정하였다.

Table 2는 본 연구에서 화학사고 사상사고 발생 요인으로 선정한 변수이다. 각 카테고리화 관련된 변수 16개를 독립변수로 선정하였다. 사업장규모의 경우 사고 사업장이 대기업, 중견기업, 중소기업 등에 따라, 사고원인 변수를 통해서는 작업자 부주의, 운송차량사고, 시설관리미흡 등에 따라 사상사고 발생 차이를 검증하고자 하였다. 사고유형의 경우는 화재, 폭발, 유출, 누출, 이상반응, 복합 등이며, 장소의 경우 사고가 발생한 장소로서 사업장 내부, 사업장 인근 도로, 사업장에 소속된 실험실 등에 따라 사상사고 발생에 대한 차이를 검증하고자 하였다.

3.3 사상사고 모형 개발 및 적합성 분석

본 연구에서는 화학사고 사상사고 예측모형을 개발하기 위해 사업장 화학사고와 관련된 변수를 활용하여 로지스틱회귀분석 후 유의성 있는 변수들을 추출하였다(Table 3). 독립변수 16개 중 최종적으로 유의수준이 0.05보다 작아 사상사고 예측모형 구축에 유의하다고 분석된 변수는 대기업, 작업자 부주의, 폭발, 복합사고 발생 등이며, 이 변수들이 화학사고 사

상사고 예측모형의 독립변수이다(Table 3). 통계적으로 유의한 독립변수들 간의 상대적 중요도(영향력)를 측정하기 위해 준표준화계수(Ronck's semi-standardized coefficient)를 계산하였고 그 결과를 살펴보면, 상대적 중요도가 가장 높은 독립변수는 폭발(Explosion)인 것으로 확인되었다^(12,13).

유의한($p < 0.05$) 결과를 나타낸 4개의 독립변수는 사상사고 예측모형에 적용되었으며, 사상사고는 2013년부터 2018년까지 발생한 화학사고 중 사업장에서 사망 또는 부상 등의 인명피해를 발생시킨 사고를 적용하였고, 인명피해가 없는 화학사고를 비사상사고로 구분하였다.

로지스틱 회귀모형의 가설검정은 전체 로지스틱회귀 모델의 유의성 검정, 개별 로지스틱회귀 계수에 대한 유의성 검정의 두 가지 측면에서 수행한다. 전체 로지스틱회귀 모델의 유의성 검정은 우도비(Likelihood ratio)라는 통계량으로 수행한다. 또한 최대우도법(Maximum-likelihood method)라는 방법을 이용하여 모수(로지스틱회귀 계수)를 추정한다. 여기서 모수 추정치가 주어졌을 때 관측값이 나타날 확률을 우도(Likelihood)라고 하고 이 값의 로그 값에 -2배를 한 지표(-2LL, $-2\log\text{-likelihood}$)가 추정된 모델이 데이터를 얼마나 적합 시키는지 척도로 사용한다⁽⁵⁾.

Table 4의 우도비 검정은 모형에 변수가 추가되거나 제거될 때의 모형 적합도의 변화량을 나타내며 -2LL값의 차이로 표현한다. 즉, 독립변수들이 포함되었을 때의 로지스틱회귀 모형의 유용성을 보여준다. 모형은 독립변수를 모두 넣은 것으로, 여기서 귀무가설은 ‘모든 독립변수들의 계

Table 3. Results of Logistic Regression Analysis

Category	Variable	B	S.E.	Wald	df	sig.	Exp (B)	β^R
Company scale	Major company	0.693	0.302	5.281	1	0.022	2.000	0.582
	Midsize company	0.594	0.393	2.280	1	0.131	1.811	0.499
	Small company	-	-	6.334	2	0.042	-	-
Cause of accident	Worker careless	1.012	0.255	15.730	1	0.000	2.752	0.896
	Transportation vehicle accident	1.831	49199.211	0.000	1	1.000	6.240	1.620
	Poor facility management	-	-	15.730	2	0.000	-	-
Accident type	Explosion	2.094	0.666	9.897	1	0.002	8.117	2.488
	Outflow	0.638	0.596	1.145	1	0.285	1.892	0.758
	Leakage	0.899	0.601	2.242	1	0.134	2.457	1.068
	Adverse reaction	0.052	0.809	0.044	1	0.949	1.053	0.062
	Complex accident Types	1.965	0.829	5.621	1	0.018	7.138	2.334
	Fire	-	-	20.691	5	0.001	-	-
Accident place	Inside the factory	20.654	19194.065	0.000	1	0.999	932786714.8	36.248
	Road near factory	-1.030	44540.846	0.000	1	1.000	0.357	-1.808
	Company's laboratory	-1.033	33489.539	0.000	1	1.000	0.356	-1.813
	Etc	-	-	0.000	3	1.000	-	-
A constant term		-22.778	19194.065	0.000	1	0.999	0.000	-

B : Regression coefficient, S.E. : Standard error, Wald : $(B/S.E.)^2$, df : Degree of freedom, sig. : Significant level, Exp (B) : Odds ratio, β^R : Roncek' semi-standardized coefficient

Table 4. Verification of Likelihood

	χ^2	df	sig.
Step	63.541	15	0.000
Block	63.541	15	0.000
Model	63.541	15	0.000

Table 5. Verification of Hosmer and Lemeshow

χ^2	df	sig.
12.446	8	0.132

수가 0이다.'이며 카이제곱 값이 63.541이고, $p = 0.000$ 으로 귀무가설이 기각되므로 모형은 유의한 것으로 나타났다.

Table 5의 Hosmer와 Lemeshow 검정도 모형의 적합도를 검증하며, 카이제곱값은 종속변수의 실제치와 모형에 의한 예측치 간의 일치 정도를 나타낸다. 모형의 적합도 분석을 위해서는 귀무가설 기각 여부와 모형의 유의수준을 확인한다. 귀무가설은 화학사고 사상사고 발생에 대한 '관측빈도와 예측빈도는 일치한다'이며, 유의수준은 0.132으로 귀무가설을 기각하지 않기 때문에 본 연구의 모형은 통계적으로 적합한 것으로 분석되었다.

Table 6에서 -2LogLikelihood (-2LL)은 모수 추정치가 주어졌을 때 관측값이 나타날 확률(우도, Likelihood) 값의 로그 값에 -2배를 한 지표(-2LL, -2log-likelihood)이며, 이 값은

Table 6. Verification of Cox @Snell R² & Nagelkerke R²

-2LogLikelihood (-2LL)	Cox @ Snell R ²	Nagelkerke R ²
373.515	0.163	0.223

모형의 적합도를 나타내는 값으로 이 값이 낮을수록 적합도가 높다고 할 수 있으며. Cox와 Snell의 R제곱과 Nagelkerke의 R제곱은 로그우도함수값을 이용해 계산한 결정계수로 선형회귀분석의 R²에 대응하는 종속변수의 분산비율을 나타내는 유사결정계수 값이 0.163, 0.223이다. 이는 독립변수들로 이루어진 회귀식이 종속변수를 예측할 수 있는 설명력을 나타낸다. 하지만 이 값은 로지스틱회귀모형과 같은 일반화선모형의 결정계수를 일반화하는데 있어 이산모형의 R²보다 작은 값의 최대값을 갖도록 조정해주는 의미이다. 그러므로 모형을 평가할 때 참고하는 수준이지 결정계수에 의한 절대적 평가를 피해야 한다⁶⁾.

4. 사고예측 모형 및 적합성 검증

16개의 독립변수 중 신뢰수준 95%에서 유의한 것으로 분석된 대기업, 작업자 부주의, 폭발, 복합사고 발생 등 4개의 변수를 적용한 화학사고 사상사고 예측모형은 식(7)과 같다. 회귀계수 β 값은 모두 (+)의 부호로 나타내어 다른 독립변수의 변화가 없다는 전제하에 특정 독립변수가 1단위 증가할 경우 사상사고가 발생할 확률이 높아지는 정도를 나

Table 7. Data Analysis Resources for Casualty Accidents of Chemical Accident

Casualty accident prediction model with explosion	Casualty accident prediction model without explosion
$Z = 0.693+1.012+2.094+1.965 = 5.764$	$Z = 0.693+1.012+2.094 \times 0 + 1.965 = 3.67$
$E(y) = \frac{1}{1+e^{-5.764}} = 0.9969 = 99.69\%$	$E(y) = \frac{1}{1+e^{-3.670}} = 0.9752 = 97.52\%$
$z = \ln \frac{0.9969}{0.0031} = \ln 321.5806 = 5.773$	$z = \ln \frac{0.9752}{0.0248} = \ln 39.3226 = 3.671$

Table 8. Classification Accuracy

Observation	Predicted value			Accuracy(%)
	No casualty accident	A casualty accident	Total	
No casualty accident	191	21	212	90.1
A casualty accident	67	52	119	43.7
Total	258	73	331	73.4

타내는 것으로 해석된다. 사상사고가 발생하는데 가장 큰 영향을 미치는 독립변수는 “폭발”인 것으로 분석되었다.

$$z = \ln \frac{E(\text{Casualty accident})}{E(\text{No casualty accident})} = \ln \text{Odds} \quad (7)$$

$$= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p$$

$$= [0.693 \times \text{Major company}(X1)] + [1.012 \times \text{Worker care less}(X2)] + [2.094 \times \text{Explosion}(X3)] + [1.965 \times \text{Complex accident types}(X4)]$$

Jin과 Lee 등⁽⁵⁾는 철도사고 예측모형의 적합성 검증을 위해 오즈비를 활용하여 도시철도 사상사고 사고예측모형의 결과를 해석하고 오즈비 변화율을 분석하여 적합성을 검증하였다. 본 연구에서 화학사고 사상사고 예측모형의 해석과 적합성 검증을 위해 Table 3의 오즈비 값을 분석하였다.

Table 7은 폭발의 발생 여부에 따른 화학사고 사상사고 발생 확률 차이를 분석한 결과를 나타낸다. 즉, 폭발사고를 동반한 화학사고 사상사고 독립변수의 선형 결정 값과 다른 변수들은 변화가 없는 상태에서 폭발사고가 없는 조건으로 변화를 주었을 경우 독립변수의 선형 결정 값을 나타낸다.

Table 7의 분석 결과를 Table 3의 오즈비(Exp(B))와 비교하여 살펴보면, 폭발 발생 없는 화학사고에서 폭발 발생 있는 화학사고로 독립변수가 변할 때 Z값이 3.7852에서 1.6898으로 변한 차이 B값은 2.094로 Explosion 회귀계수와 일치한다(Table 3). 또한, 오즈비 변화율에 대하여 살펴보면, 폭발이 없었던 화학사고에 대한 폭발이 있었던 화학사고의 사상사고 발생 확률은 약 8배가량 높게 나타났으며 이는 Table 3의 오즈비 값 8.117과 일치하며 상대적 중요도를 측정하는 준표준화계수의 결과와도 일치한다($\beta^R = 2.488$).

Table 8은 예측모형에 의한 화학사고 비사상사고와 사상사고 분류 정확도를 나타낸다. 분류표에서 화학사고의 비사상사고와 사상사고에 대한 관찰치와 예측치 간의 차이를

보면 비사상사고의 경우 191건, 사상사고의 경우 52건이 정확히 분류되었으며, 분류정확도는 73.4%의 정확도를 나타냈다.

5. 결론

본 연구를 통해 화학사고 사상사고 예측모형을 개발하였으며, 이를 위해 “화학안전정보 공유시스템”의 사업장 화학사고 발생 요인 카테고리화 관련된 변수를 독립변수로 적용하였다. 또한, 예측모형은 로지스틱회귀분석 모델을 활용하여 사상사고에 영향을 주는 통계적으로 유의한($p < 0.05$) 변수를 도출하여 적용하였고, 모형의 신뢰성 및 정확성을 검증하기 위해 통계적 검증방법과 오즈비를 활용해 모형을 검증하였다. 예측모형에 활용한 사고자료는 과거 발생했던 화학사고 통계 DB⁽¹⁾를 분석하여 활용하였으며, 사고의 유형, 원인, 발생 장소, 사상자 현황 및 사상자 발생 화학사고 등의 자료 분석을 통해 통계적으로 유의하게 나타난 독립변수($p < 0.05$)를 적용하였다. 본 연구에서 개발한 모형은 사업장에서 화학사고로 인해 발생하는 사상사고의 예방 및 안전시스템 구축을 위한 연구로 의의가 있다고 할 수 있다.

예측모형에 의한 화학사고의 분석결과 사상사고 발생에 유의하게 영향을 미치는 변수($p < 0.05$)는 대기업, 작업자 부주의, 대기업, 복합사고 발생 등이었으며, 변수 중 가장 크게 영향을 미치는 변수는 폭발에 의한 화학사고인 것으로 조사되었다. 따라서 사업장에서 발생하는 폭발 유형의 화학사고를 예방하기 위한 대책마련이 시급하다고 판단된다.

현재 환경부에서는 화학사고 예방을 위한 제도로서 장외영향평가제도를 시행하고 있다. 이 제도에서는 유해화학물질의 취급시설을 설치·운영하려는 자는 사전에 화학사고 발생에 따른 사업장 주변 지역의 사람이나 환경 등에 영향을 평가한 화학사고 장외영향평가서를 작성하여 환경부장관에게 제출하도록 하고 있다. 아직 설치되지 않은 취급시설을 대상으로 위험도 산정을 통해 장외영향평가를 수행하

여 화학사고의 발생 가능성을 확인하고 사업장으로 스스로 이에 대한 사고 영향을 줄이거나 사고발생빈도를 줄이는 위험도 감소대책을 적용하도록 유도하는 것이다. 즉, 신규 시설의 사고 피해를 사전에 예측하고 예방대책을 수립한다. 하지만, 장외영향평가의 사고발생 위험도 산정은 기존 가동되고 있는 취급시설을 대상으로 한 화학사고의 발생 가능성과 피해 등은 예측하지 못하고 있는 실정이다. 이는 사업장 화학사고의 발생 유형, 원인, 가능성 등을 알 수 없어 사고 대응과 예방대책수립에도 어려움이 있다. 본 연구에서 개발한 화학사고 사상사고 예측모형은 사업장에서 발생하는 화학사고 유형, 원인, 장소, 피해 등에 대한 예측을 가능하게 한다. 따라서, 본 연구에서 개발한 화학사고 사상사고 예측모형은 향후 사업장 화학사고의 예방 및 피해 감소 등의 안전성 제고를 위한 연구 및 정책 수립 등의 활동에 기여할 수 있을 것이다.

본 연구의 한계점으로서, 독립변수의 선정은 “화학안전정보공유시스템”에서 제공하는 화학사고 발생 요인 카테고리 한정하였다. 향후 화학사고 사상사고와 인과관계가 높은 독립변수 발굴 및 이를 적용한 추가연구가 필요하다. 또한, 화학사고로 인한 사상자피해 뿐만 아니라 환경피해에도 영향을 미칠 수 있는 요인들의 발굴에 대한 지속적인 연구와 예방을 위한 정부의 제도적, 정책적 뒷받침이 필요하다고 사료된다.

References

1. Chemistry Safety Clearing-house (csc.me.go.kr).
2. Deputy Director General of Safety Environment in the Office for Government Policy Coordination, “Accidents of Harmful Substance’s Crisis Management and Direction of Policy” (2013).
3. Ministry of Environment, “Chemicals control Act” (2015).
4. H. S. Lee and J. P. Yim, “A Study on Prevention Measure Establishment through Cause Analysis of Chemical Accidents”, Journal of the Korean Society Safety, Vol. 32, No. 3, pp. 21-27 (2017).
5. S. B. Jin and J. W. Lee, “Study on Accident Prediction Models in Urban Railway Casualty Accidents Using Logistic Regression Analysis Model”, Journal of the Korean Society for Railway, Vol. 20, No. 4, pp. 482-490 (2017).
6. J. H. Kang, K. W. Kim and S. M. Kim, “Development of the U-turn Accident Model at Signalized Intersection in Urban Areas by Logistic Regression Analysis”, Journal of the Korean Society on Civil Engineers, Vol. 34, No. 4, pp. 1279-1287 (2014).
7. B. H. Park, J. M. Yang and J. Y. Kim, “Logistic Regression Accident Models by Location in the Case of Cheong-ju 4-Legged Signalized Intersections”, Journal of Korea Society of Road Engineers, Vol. 11, No. 2, pp. 17-25 (2009).
8. X. Yan, E. Radwan and M. Abdelaty, “Characteristics of Rear-end Accidents at Signalized Intersections Using Multiple Logistic Regression Model”, Accident Analysis & Prevention, Vol. 37, No. 2, pp. 253-259 (2005).
9. C. W. Park, J. B. Wang, M. S. Kim and D. B. Choi, “Development of Risk Assessment for Railway Casualty Accidents”, Journal of Korean Society for Railway, Vol. 12, No. 2, pp. 190-198 (2009).
10. K. Y. Kim, M. S. Jeon, H. C. Kang and S. K. Lee “Regression Analysis by Example”, Free Academy, Seoul, pp. 334-357 (2009).
11. National Institute of Chemical Safety, “Chemistry Safety Clearing-house (csc.me.go.kr)” (2015).
12. R. D. Hartley, S. Maddan and J. T. Walker, “Sentencing Practices under the Arkansas Sentencing Guideline Structure”, Journal of Criminal Justice, Vol. 34, No. 2, pp. 493-506.
13. J. T. Walker and S. Maddan, “Statistics in Criminology and Criminal Justice: Analysis and Interpretation”, Jones & Bartlett Learning (2013).