

범주형 분석에 의한 산업재해사례 요인의 고찰 A Study on Causes of Industrial Accident Cases by a Categorical Analysis

지 경 택* , 송 영 호** , 정 국 삼**
한라그룹* , 충북대학교 안전공학과**

1. 서 론

1.1 연구목적

우리나라의 산업재해통계는 산업재해의 규모 및 원인 등의 분포상태와 근로자에 대한 특성 등을 파악하여 산업재해 예방정책 및 산업재해 보상 보험 운용 방침 수립의 기초 자료로 사용되고 있다.

그런데, 우리나라의 현행 산업재해 통계 산출 방법은 산업재해보험 가입 사업장의 재해자가 제출한 요양신청서 중 업무상 재해로 인정된 재해만을 대상으로 통계를 산출하는 것이다. 따라서, 현재의 산업재해 통계는 다분히 표면적이다, 또한 산업재해 발생에 대한 원인분석도 산업재해조사표에 의해서 재해발생형태, 기인물, 관리적 원인, 불안전 행동, 불안전상태 등의 재해발생 원인에 대해서 행해지고 있지만, 이것은 전체 재해건수에 대한 비율 등으로 나타내어질 뿐이다²⁾. 이러한 산업재해 통계분석만으로는 그 목적을 만족시키기에는 미흡하고, 근본적인 재해발생 원인 규명에도 한계가 있다.

따라서, 본 연구에서는 산업재해 사례를 대상으로 통계적 기법을 이용하여 재해 요인별 상관분석, 재해 요인 간의 영향 분석 및 재해발생의 중요 요인을 분석하고자 하였다.

1.2 연구방법

본 연구에 사용된 통계학적 기법을 설명하면 다음과 같다.

첫째, 산업재해 사례의 data를 분석하여 근로손실일수 및 재해형태와 기타 변수들간의 상관분석을 통하여 상관도를 도출하였다.

둘째, 명목 및 서열척도의 범주형 변수(Categorical Variables)들을 분석하여 변수간의 교차분석을 통하여 독립성과 관련성을 분석하였다.

셋째, 반응변수(종속변수)와 설계변수(독립변수)들간의 인과관계와 독립변수들의 변동에 따라 반응변수의 변화를 예측하기 위하여, Logistic 회귀분석(Logistic Regression Analysis)을 적용하였다.

위와 같은 분석을 위하여 사용된 산업재해 사례 data는 모두 2,133건으로, 충북지역의 사업장을 대상으로 1997년 한 해 동안 수집한 data로 하였고, 분석에 사용된 통계 package는 SPSS (Statistical Package for the Social Science ver. 8.0)을 사용하였다. 그리고, 통계분석을 용이하게 하기 위하여 data 의 coding 을 변경하거나 변환을 시도하였다.

2. 이론적 배경

2.1 상관도 분석

산업 재해 사례에서 요인들 간의 상관도 계산 방법의 기본적인 개념은 다음과 같다. 인자 1 또는 인자 m이 동시에 일어나는 재해 사례는 인자 1 또는 인자 m이 각각 일어난 사례에 대하여 어느 정도의 비율을 갖는가? 하는 문제로 인자 1이 일어난 재해사례 집합을 A, 인자 m이 발생한 재해 사례 집합을 B라 하고 각각의 요소 수를 n(.)로 표시하면 인자 1, 인자 m 사이의 상관도 ρ_{lm} 는

$$\rho_{lm} = \frac{n(A \cap B)}{n(A \cup B)} \dots\dots\dots (2.1)$$

로 표시된다.

이 때, 상관도의 실제 계산을 위해 재해사례 Data를 X_i 로 하면,

$$X_i = [0, 1, 0, 1, \dots\dots\dots, 0, 1]$$

로 재해 사례에 1 : 1로 대응하고 있다.

재해 사례의 분류 인자를 순서대로 번호를 부여하면, i 번째 사례의 j 번째 인자에 있어서의 data는 다음과 같은 것이 된다.

$$X_{ij} = 1 \text{ (i 번째 사고의 j 번째 인자가 일어난 경우)}$$

$$0 \text{ (i 번째 사고의 j 번째 인자가 일어나지 않은 경우)}$$

여기서, $i = 1, 2, 3, \dots\dots, N$ (재해 사례의 총 건수)

$j = 1, 2, 3, \dots\dots, K$ (인자의 총 수)

이다.

변수 X_{ij} 를 위와 같이 정의하기로 하면, 인자 1 과 m 이 재해 사례 총 건수 가운데 함께 일어나고 있는 경우를 A 로 하면,

$$A = \sum_{i=1}^n X_{ij} X_{im} \dots\dots\dots (2.2)$$

그리고, 인자 1 이 일어나고, 인자 m 이 일어나지 않는 경우의 수를 B라고 하면,

$$B = \sum_{i=1}^n X_{ij} (1 - X_{im}) \dots\dots\dots (2.3)$$

로 앞서 정의되었던 상관도 $\rho_{l,m}$ 은 다음 식으로 간단히 표시된다.

$$\rho_{l,m} = \frac{A}{(A+B+C)} \dots\dots\dots (2.4)$$

여기에서 정의된 $\rho_{l,m}$ 은 인자 1 과 인자 m 이 어느 정도의 관련성을 갖고 있는가를 나타내는 양이 된다.¹⁾

2.2 독립성 검정

본 연구에서는 두 변수간의 관련성의 여부와 검정 결과 만일 상호 관련성이 있다면, 그 정도는 얼마나 되는 지를 알아보기 위하여 독립성 검정 방법을 채택하였다.

2.2.1 χ^2 통계량

만일 두 변수가 각각 m, n개의 범주를 갖는다면 교차표는 (m × n)개의 셀(cell)을 이룬다. 이들 두 변수가 갖는 빈도분포에 의하여 (m × n)개의 각 셀에 두 변수가 결합된 기대빈도를 구하고 실제 측정된 빈도와 비교함으로써 χ^2 값을 산출해 내게 되는데, 독립성 검정에서 가장 많이 이용되는 것이 바로 Pearson의 카이제곱 통계량 (Pearson's Chi-square Statistics)이다.³⁾ Pearson의 카이제곱 통계량은

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{(E_i)} \quad \dots\dots\dots (2.5)$$

여기서, O_i 는 i 번째 범주에 대한 관측빈도를 말하고, E_i 는 i 번째 범주에 대한 기대빈도, k는 범주수로 정의되는데, 귀무가설이 맞을 때에는 자유도(degree of freedom ; df)가 (m-1)(n-1)인 χ^2 분포를 따르게 된다.

결정원칙은 검정 통계량보다 더 큰 값이 나올 확률인 P 값 (P - value)을 구해서 유의수준 ($\alpha = 0.05$)보다 작으면 귀무가설을 기각하게 되어 두 변수는 서로 독립이 아니라는 결론을 얻게 된다.

2.2.2 관련성 척도 (λ)

χ^2 검정 결과는 두 변수간에 통계적으로 의미있는 관련성이 있는지는 알려주지만, 이를 근거로 두 변수간에 관련성이 어느 정도인지는 판단할 수 없다. χ^2 값이 크다고 반드시 관련성이 강하다고는 볼 수 없으며, 관련성은 약하더라도 표본이 크면 χ^2 값은 크게 나타날 수 있다. 그러므로, 관련성의 정도를 알기 위해서는 별도의 측정방법이 필요하다.

본 연구에서는 이 중에서 가장 무난하고 해석이 명확한 람다(λ) 를 채택하여 관련성을 측정하기로 한다.⁶⁾ 람다는 식 (2.8) 과 같이 정의되는데, 취할 수 있는 값의 범위는 $0 \leq \lambda \leq 1$ 으로, $\lambda = 0$ 일 때 두 변수는 완전히 독립적이고, $\lambda = 1$ 일 때 두 변수는 완전한 관계임을 의미한다.

$$\lambda = \frac{\sum (f_i - F_d)}{(N - F_d)} \quad \dots\dots\dots (2.6)$$

여기서, N 은 사례수이고, f_i 는 독립 변수의 각 카테고리 내의 최빈도수를 말하며, F_d 는 종속 변수의 최빈도수를 말한다.⁵⁾

2.3 Logistic 회귀분석

Logistic 회귀분석(Logistic Regression Analysis)은 자료의 분포형태가 비선형적이고, 종속변수가 이원화, 즉, 단지 두 개의 값만을 취할 때, 종속변수와 독립변수들간의 인과관계를 추정하는 기법이다.⁴⁾

Logistic 회귀분석으로 어떤 사건이 일어날 확률을 직접적으로 추정할 수 있다. 하나 이상의 독립변수에 대하여, 이 모델은 다음과 같이 나타낼 수 있다.³⁾

$$P(x) = \frac{e^z}{(1 + e^z)} \quad \dots\dots\dots (2-7)$$

여기서, z 는 선형결합을 나타내는데, 다음과 같이 나타낼 수 있다.

$$Z = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p \quad \dots\dots\dots (2-8)$$

3. 분석결과 및 고찰

분석을 위해 수집된 산업재해사례 data 들은 모두 25 개의 항목들로 구성되어 있는데, 신뢰성있는 분석을 위하여 10 개 항목으로 재분류하였고, 또 각각의 항목들에 대해서 통계분석기법을 처리하는데 있어서 용이하도록 code화 하였다.

또, 재해사례 조사 중 미기재 항목이나 기록이 불충분한 자료는 가용항목들만 code화 하고, 나머지 항목들은 분류불능으로 처리하였다.

3.1 상관도 분석

재해사고의 과정상 중요한 변수라고 할 수 있는 사고의 발생형태와 근로손실일수에 영향을 줄 수 있는 기타 변수들간의 상관관계를 상관도 분석을 통하여 그 결과를 도출해 내었다. 상관도 분석의 결과 예를 Table 1 에 나타내었다. Table에서 가장 높은 상관도를 보이는 것은 '61 일 ~ 90 일 - 6 개월 미만' 으로, 항목간 높은 교차빈도를 보일 경우 상관도도 높게 나타나는 경향을 보이고 있다.

3.2 독립성 검정

독립성 검정은 두 변수간의 관련성 여부와 만일 상호 관련성이 있다면 그 정도는 얼마나 되는지 알아보기 위한 것이다.

본 연구에서는 독립성 검정을 위해서 가장 잘 알려진 Pearson의 카이제곱 통계량을 이용하였고, 정량적인 관련성 척도로써 표본의 크기와 관계없이 해석이 명확한 λ 를 적용하였다. Table 2는 '근로손실일수'와 '발생월' 항목간의 χ^2 검정을 행한 결과이다. 양측검정의 접근 유의 수준이 설정 유의 수준($\alpha=0.05$)보다 작으므로 "두 변수가 상호 독립이다"라는 귀무가설을 기각하게 되어 대립가설, 즉 "두 변수가 상호 독립이 아니다."를 채택하게 된다. 즉, 근로 손실 일수는 재해가 발생하는 월에 따라 다르다는 결론을 얻게 된다. 또 우도비 검정(Likelihood Ratio Chi-Square) 결과도 같은 결과를 나타내고 있다. 연관성의 정량적 척도인 λ 값은 0.075 로써 두 변수간에는 다소 미약한 연관성이 있다.

3.3 Logistic 회귀분석

Logistic 회귀분석에서는 상관도 분석에서 나온 결과를 바탕으로 반응변수, 즉 '근로손실일수' 항목에 큰 영향을 미친다고 판단되는 독립변수 네 가지(재해형태, 기인물, 근속기간, 업종)를 채택하여 반응변수의 발생확률을 예측하였다. 이 분석기법에서는 반응변수가 2원화 되어야 하므로 반응변수, '근로손실일수' 항목을 경상(근로손실일수 56일 이하) 및 중상(근로손실일수 56일 초과)을 나누어 독립변수들에 따라 각각의 발생확률

Table 1. '근로손실일수' 항목과 '근속기간' 항목간의 상관도

손실일수 근속기간	1일 ~ 14일	15일 ~ 28일	29일 ~ 60일	61일 ~ 90일	91일 ~ 120일	121일 ~ 150일	151일 ~ 180일	181일 ~ 270일	271일 ~ 1년	1년 ~ 2년	사망
01	0.015	0.064	0.160	0.161	0.115	0.081	0.048	0.139	0.039	0.044	0.022
02	0.011	0.046	0.106	0.052	0.066	0.037	0.048	0.056	0.020	0.039	0.024
03	0.029	0.051	0.094	0.068	0.060	0.039	0.027	0.062	0.010	0.033	0.032
04	0.010	0.055	0.054	0.073	0.048	0.039	0.039	0.025	0.023	0.015	0.029
05	0.016	0.038	0.072	0.039	0.052	0.027	0.040	0.031	0.033	0.037	0.005
06	0.016	0.001	0.008	0.071	0.067	0.072	0.012	0.014	0.015	0.062	0.003
07	0.020	0.036	0.071	0.064	0.047	0.039	0.015	0.027	0.009	0.043	0.038
08	0.038	0.017	0.047	0.048	0.014	0.029	0.030	0.020	0.019	0.021	0.007
09	0.067	0.011	0.009	0.010	0.010		0.014	0.003	0.029	0.044	0.022

Table 2. Output χ^2 -test on Month by Lost working days

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	662.221 ^a	110	.000
Likelihood Ratio	659.289	110	.000
Linear-by-Linear Association	87.096	1	.000
N of Valid Cases	2133		

a 22 cells (16.7%) have expected count less than 5.
The minimum expected count is 1.39.

Directional Measures

			Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Nominal by Nominal	Lambda	Symmetric	.075	.010	7.761	.000
		발생월 Dependent	.103	.011	9.139	.000
		손실일수 Dependent	.043	.012	3.450	.001
	Goodman and Kruskal tau	발생월 Dependent	.030	.002		.000 ^c
		손실일수 Dependent	.033	.003		.000 ^c

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

c Based on chi-square approximation

을 계산하였다. Table 3은 로지스틱 회귀분석의 결과를 나타내는데, 유의 수준으로 볼 때 독립변수 중 반응변수에 영향을 주는 것은 '연령', '근속기간', '기인물' 항목이다.

Table 3. Output of Logistic Regression Analysis

· Number of selected cases : 2133 · Number rejected because of missing data : 0 · Number of cases included in the analysis : 2133 Dependent Variable.. 근로손실일수 -2 Log Likelihood 2635.8918 -2 Log Likelihood 2589.731 Goodness of Fit 2127.392 Cox & Snell - R ² .021 Nagelkerke - R ² .021							
		Chi-Square	df	Significance			
Model		46.161	4	.0000			
Block	46.161		4	.0000			
Step		46.161	4	.0000			
Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
형태	.0018	.0154	.0140	1	.9060	.0000	1.0018
기인물	-.0385	.0085	20.4146	1	.0000	-.0836	.9622
근속기간	-.0552	.0209	7.0163	1	.0081	-.0436	.9463
연령	.0745	.0202	13.5881	1	.0002	.0663	1.0774
Constant	.9717	.1979	24.0992	1	.0000		

4. 결론

각각의 분석에 대하여 다음과 같은 결론을 얻을 수 있었다.

1) χ^2 - Test 로 독립성 검정을 한 결과, 독립변수중 근속기간, 기인물, 재해발생월, 성별, 업종, 재해형태, 등이 종속변수인 근로손실일수와 상호관련이 있는 것으로 나타났고, λ 값을 적용시킨 결과 근로손실일수와 가장 관련성이 높은 요인은 재해발생월로 나타났다.

2) 근로손실일수를 경상 및 중상으로 나누어 반응변수와 독립변수의 인과관계를 추정해본 결과, 중상의 발생확률이 높게 나타날 경우, 작용하는 재해발생요인은 전도붕괴, 가설건축구조물, 6개월 미만 등으로 나타났다.

参 考 文 献

- 1) 鄭國三, 災害事故分析論, 忠北大學校 安全工學科, 1993
- 2) 洪光洙, 相關分析을 應用한 産業災害事例 要因의 考察, 충북대학교 석사학위논문, 1997.
- 3) 정충영, 최이규, SPSSWIN을 이용한 통계분석, 무역경영사, 1998.
- 4) 배일섭, 정영숙, SPSS 한글프로그램과 통계분석기법, 대구대학교출판부, 1998.
- 5) 홍중선의 2명, 조사방법과 통계자료분석, 전영사, 1996.
- 6) 김호정, 사회과학 통계분석, 삼영사, 1993.
- 7) 남궁 평, 홍중선, 범주형자료분석, 탐진, 1996.