

相關分析을 應用한 産業災害事例 要因의 考察

홍광수*, 정국삼**

한국산업안전공단*, 충북대학교 안전공학과**

1. 서론

본 연구에서는 산업재해 사례를 인구 대상으로 재해 발생의 여러 가지 요인들의 관련을 검토하고자 통계적 기법을 이용한 재해 요인별 상관분석, 또는 영향의 정도 파악, 재해 요인의 통세에 따른 기타 재해 요인에 대한 영향 분석을 시도하는 통계학적 분석 방법을 이용한 재해 발생의 중요 요인을 분석하고자 첫째, 산업재해 통계 자료의 내용을 분석하여 재해 관련 변수들을 파악하는데 불안정한 행동 및 불안전상태에 의한 재해 형태와 기타 변수들 간의 정성적 상관분석을 통한 상관계수를 고찰, 둘째, 명목척도인 범주형 변수 상호 간의 관련 여부를 파악하기 위해 카이제곱(chi-square)검정을 행하여 입원 일수를 종속 변수로 하는 기타 변수들의 독립성 여부와 변수 상호간 연관이 있다고 판단될 때 각 변수의 연관의 정도 비교, 셋째, 어떤 변수 상호간 일정한 관계를 가질 때 변수의 범주별로 반응변수(종속변수)에 미치는 영향을 회귀식 형태로 파악하고 비교하기 위하여 로짓(logit)모형을 적용하였다.

2. 이론적 배경

상관분석은 연구를 위한 변수 간에 존재하는 일정한 상관 관계의 파악을 위한 분석을 말한다. 다시 말해 상관분석은 그 연구 변수 간의 단순한 상관관계의 유무의 파악에서부터 그 관련의 정도 그리고 관계의 방향 등에 이르기까지 파악을 하려는 것이다. 변수간의 상호관계를 판단하는 주요 통계기법에는 Chi-square(χ^2)검정, Contingency coefficient(C), Kendall Coefficient of concordance(W), 분산 분석(Analysis of variance), Pearson's product moment correlation(γ) 등이 있으나, 본 연구에서는 명목 척도나 서열 척도를 갖는 데이터나 모집단의 분포에 대한 가정이 필요 없는 비모수검정(Non-parametric tests)^[7]에 쓰이는 상관도 분석, Chi-square(χ^2)검정, Logit 분석이다.

2.1. 상관도 분석

산업재해 사례에서 요인들 간의 상관도 계산 방법의 기본적인 개념은 다음과 같다^{[1][3]}. 인자 l과 인자 m이 동시에 일어나는 재해사례는 인자 l 또는 인자 m이 각각 일어난 사례에 대하여 어느 정도의 비율을 갖는가? 하는 문제로 인자 l이 일어난 재해사례 집합을 A, 인자 m이 발생한 재해 사례 집합을 B라 하고 각각의 요소 수를 n(.)로 표시하면 인자 l, 인자 m 사이의 상관도 ρ_{lm} 는

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

로 표시된다.

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

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

로 재해사례에 0 : 1로 대응하고 있다.

재해사례의 분류 인자를 순서대로 번호를 매기면, i 번째의 사례의 j 번째의 인자에 있어서의 Data는 다음과 같은 것이 된다.

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

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

여기에서

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

$$j = 1, 2, 3, \dots, K \text{ (인자의 총수)}$$

이다.

변수 X_{ij} 를 위와 같이 정의하기로 하면 인자 l 과 m 이 재해사례 총 건수 가운데 함께 일어나고 있는 경우를 A 로 하면

$$A = \sum_{i=1}^N X_{il}X_{im} \text{-----}(2.2)$$

그리고, 인자 l 이 일어나고, 인자 m 이 일어나지 않는 것의 수를 B , 인자 l 이 일어나지 않고, 인자 m 이 일어나는 것의 수를 C , 인자 l, m 이 모두 일어나지 않는 것의 수를 D 로 하면, 상관도 ρ_{lm} 은 다음 식으로 간단히 표시된다.

$$\rho_{l,m} = \frac{A}{A+B+C} \text{-----}(2.3)$$

여기에서 정의된 ρ_{lm} 은 인자 l 과 인자 m 이 어느 정도의 관련을 갖고 있는가를 나타내는 양이 된다.

2.2. 독립성 검정

빈도의 자료에 대한 검정 방법으로 쓰이는 χ^2 검정은 비모수 통계학의 영역이므로 모수 통계학에서 요구하는 까다로운 요건을 충족시킬 수 없는 상황에서도 적용할 수 있으며 그 활용 영역은 동질성의 검정, 독립성의 검정, 적합성의 검정이라는 세 가지로 볼 수 있는데 여기서는 독립성의 검정과 또한 독립성을 검정할 경우 두 변수의 관련성의 정도를 알 수 있는 척도(λ)를 이용한다. 어떤 변수가 범주(Category)를 값으로 갖는 경우 그 변수를 범주형 자료(Category data)라 하고 이러한 두 변수가 서로 상관이 있는지 독립인지 판단하는 피어슨의 카이제곱 통계량(Pearson's χ^2 -statistic)은

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \text{-----}(2.7)$$

여기서, O_i : i 번째 범주에 대한 관측빈도

E_i : i 번째 범주에 대한 기대빈도

k : 범주 수

으로 정의된다.

χ^2 -test는 관찰 도수가 기대 도수로부터 우연에 의한 빈도를 측정하는 통계적

수단으로 상관 계수를 계산하지 않고 상관의 유의성을 검정하는 것이며, χ^2 검증 결과 두 변수가 연관이 있을 때 두 변수의 연관성의 척도로 대표적인 람다(λ)는

$$\lambda_d = \frac{\sum f_i - F_d}{N - F_d} \text{-----}(2.8)$$

여기서, N ; 사례의 수
 F_d ; 종속 변수의 최빈도수(Modal frequency)
 f_i ; 독립변수의 각 카테고리내의 최빈도수

으로 정의된다.

λ 는 한 변수의 명목 카테고리의 도수가 다른 명목 변수의 도수에 대해 어느 정도의 예측 근거를 제공하는가를 측정하는 것^[6]으로 일반적으로 $0 < \lambda < 1$ 이며 이 값이 1에 가까울수록 두 변수의 결합도가 큰 것이며 이 값이 0에 가까울수록 두 변수의 결합도는 작은 것이다^[4].

2.3. 로짓(Logit;Log unit) 분석

교차 분할된 빈도 자료의 표집 과정에서 교차 분할표에서 반응 변수로 선정된 두 개의 변수가 범주로 이루어진 빈도 자료에 대한 일련의 가부 반응 분석(Quantal response analysis)기법^[6]의 하나인 로짓분석(Logit Analysis)에서 일반적으로 반응 변수(Response variable)라고 불리는 변수는 연속 변수에 대한 통계 검증에서 흔히 종속 변수라고 불리는 변수와 유사한 기능을 하는 변수이다.

이 분석 기법은 가능성 대비에 의한 독립 변수의 효과를 검증하는 것이다. 독립 변수의 각 범주에서 가능성을 검토하는 것이므로 수학적으로 “예측되는” 변수(Predicted variable)는 바로 이 가능성 대비이고, “예측하는” 변수(Predictor variable)는 독립 변수이다.

선형 로그분석에서는 효과 계수들에 의해 예측되는 것이 범주 빈도이나 여기서는 가능성 대비라는 것이 구체적인 의미이며 이 가능성 대비를 로그와 한 것을 로짓(logit)이라 부른다.

로짓 모형의 분석에 관한 통계적 이론은 일반적으로 공변량(설명변수)의 수가 P개인 경우를 생각하기로 한다. 즉 Y=1 혹은 2이고 공변량 x_1, \dots, x_p 가 연속형 변수라고 하면, 일반화 로짓모형(Generalized logit model)은

$$\log \frac{p(Y=1 | x_1, \dots, x_p)}{p(Y=2 | x_1, \dots, x_p)} = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \text{-----}(2.9)$$

로 정의된다. 이것은 다시

$$p(Y=1 | x_1, \dots, x_p) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \text{-----}(2.10)$$

로 쓸 수 있다^[7].

3. 분석 결과 및 고찰

3.1. 재해 요인변 상관도 분석

재해 사고의 과정상 중요 변수라 할 수 있는 불안정한 행동과 불안정한 상

태에 의해서 발생된 사고형 태와 기타 변수들 간의 상관 관계를 나타내 주는 정성적 상관도 분석을 의해서 우선 두 변수간 교차 빈도를 분석하고, 또 빈도분석 결과를 이용한 상관도 분석으로 변수간 높은 교차 빈도를 보일 경우 변수간 상관도도 높게 나타나는 경향을 보여 주었다. 그러나 빈도 수와 상관도가 직접적인 함수 관계를 나타내지는 않는데 그 이유는 전체 불안정한 행동에 의해 발생된 재해 형태 중 추락 재해의 비중이 어느 정도인가 하는 것도 고려되기 때문이라 판단 되었다. 산업 재해의 발생 과정상 중요 변수라 할 수 있는 불안정한 행동과 불안정한 상태에 의해서 발생된 재해 형태와 기타 변수들 간의 상관도 분석 결과, 각 변수간 상관의 정도를 판단할 수 있었으며 불안정한 행동에 의한 추락 재해와 기타 요인들의 상관도 분석 결과를 요약하면 재해시간은 14시~16시(0.101)에, 재해자 연령은 30~34세(0.099), 근무 작업 공정은 해체작업(0.349)이, 직·간접적인 영향은 작업자의 음주, 흡연, 피로 요인들(0.165)이 비교적 상관도가 높은 것으로 나타났으며 작업 내용은 작업 관련 이동 중(0.084), 작업자의 동작은 걷다(0.079) 요인이 상관도가 높게 나타났다.

재해 발생형태 중 중상(입원일수 200~250일)과 상관도가 높게 나타난 불안정한 행동에 의한 감전 재해와 기타 요인들간의 상관도가 높은 것으로 나타난 재해 시간은 22~24시(0.026), 공사 종류는 전기·통신공사(0.143), 재해자의 동업종 근무기간은 5~10년(0.312), 재해 당일 근무 시간 형태는 휴일근무(0.028)가 상관도가 높음을 알 수 있다. 또한, “불안정한 상태에 의한 추락”재해 발생과 상관도가 높은 인자를 살펴보면 재해 시간은 14~16시(0.089), 연령은 40~44세(0.099), 건설 단위 공정은 가설공사(0.177), 작업 내용은 조립, 설치 작업(0.125), 기인물은 가설 건축구조물(0.163), 기술직 방호설비는 없음(0.117)이 상관도가 가장 높은 것으로 나타나 건설 현장의 불안정한 행동에 의한 추락 재해는 사고 발생 과정의 건설 단위 공정과 작업 내용, 기술적 방호설비 및 안전장치 유·무가 직, 간접적인 영향이 있다는 것을 알 수 있었다.

3.2. 범주형 변수 간의 상관 및 독립성 검정

산업재해 발생시 재해정도(입원일수)에 영향을 주는 변인들이 어떤 것이며 가장 많은 영향을 미치는 변인이 무엇인가를 알아보기 위한 χ^2 -Test 결과 재해의 정도, 즉 입원 일수를 종속 변수로 하고 기타의 재해 관련 변수를 입원 일수에 영향을 주리라고 가정한 독립 변수로 할 때의 독립성 검정으로 입원 일수와 상관성이 있는 변수를 파악할 수 있었으며 또 입원 일수에 영향을 미치는 독립 변수들의 관련성의 정도를 판단하기 위하여 범주형 변수들 상호 관련성은 연관성의 척도 λ 값을 이용하여 파악하였는데 입원 일수를 종속 변수로 한 기타 독립 변수 상호간의 연관성을 비교한 결과 상관성이 있다고 평가된 변수들의 척도 λ 는 상해 종류, 재해자의 재해당시 작업내용, 재해 발생 형태가 연관성이 큰 것으로 나타났으며 연관성이 가장 큰 변수는 상해 종류인 것으로 판단되었다.

3.3 반응변수에 대한 범주형 변수의 효과

두 개의 값을 가진 반응변수에 대한 범주형 변수의 효과를 평가하는 모형을

관심을 가질 때, 특히 다른 변수들이 조정하기를 원하는 “실제” 변수 일 때 반응 변수에 대한 로짓 모형을 설정할 수 있었다.¹⁰⁾ 첫째, “건설:비건설”, “보호구 착용:미착용”, “추락, 붕괴, 도괴, 전도재해:기타 재해”로 구분된 범주형 변수와 “입원 일수(1개월 미만-경상, 1개월 이상-중상으로 구분한)”를 반응 변수로 한 기능상의 대비를 분석하기 위하여 설정된 모형에서 재해 형태와 건설 여부가 통제되었을 때 보호구 착용시 경상에 비해 중상이 일어날 가능성의 예측은 보호구 미착용시보다 약 1.2배 가량 높은 것으로 나타났다. 이는 우리가 예상할 수 있는 사실의 반대되는 결과라고 생각할 수 있으나 해석에 이용된 데이터가 재해가 발생한 사례만이었으며 장기간 입원을 요하는 상해 종류인 화상, 골절, 뇌진탕 등과 같은 재해를 일으키는 추락이나 감전 등의 재해 형태에서는 보호구 착용이 작업에 반드시 우선되었다고 볼 수 있기 때문이라고 판단된다.

또한, 보호구 착용, 건설 여부가 통제되었을 때 추락, 붕괴, 전도가 일어났을 때는 경상에 비해 중상이 일어날 가능성이 1.230배 높게 나타났다. 보호구 착용과 재해형태가 통제되었을 때 건설업종이 비건설 업종에 비해 경상에 비해 중상이 일어날 가능성이 1.161배 높게 나타났다.

또, 비건설업에서 재해자의 근속 기간이 입원 일수에 미치는 영향을 예측하기 위한 Logit 분석 결과 근속기간에 대한 재해발생 시 재해 정도에의 영향은 경상에 비해 중상이 일어날 가능성은 6개월~3년미만 근로자가 가장 작게 나타났으며 3년이상 근로자는 6개월~3년미만 근로자보다 경상에 비해 중상이 일어날 가능성이 2.99 배로 높게 나타났다. 또 6개월 미만 근로자는 6개월~3년미만 근로자보다 2.65 배로 나타났는데 이는 근속 기간이 짧은 근로자가 근속 기간이 긴 근로자 보다 재해 발생시 중상을 일으킬 수 있는 위험작업에 종사하기 때문이라고도 추정 할 수 있다.

4. 결 론

산업 재해 조사 항목별 내용을 산재발생에 영향을 미치는 변수(Variable)로 가정하고 각각의 변수 상호간 관련여부를 파악하기 위해 상관도 분석, 독립성 검정, 변수의 수준에 따른 효과 평가를 시도해 본 결과 재해 요인간 상관도 분석결과 재해 발생의 직접원인인 불안정한 상태와 불안정한 행동에 의한 여러 가지 재해 형태와 기타 요인들의 관련성을 상대적인 크기로 비교할 수 있는 상관도 계수를 산출할 수 있었다. 이러한 상관도는 산업 재해 원인에 관계되는 요인들의 연관성을 평가하여 사고를 미연에 방지하기 위한 재해 요인 관리 대상을 선정하는 방법으로 이용할 수 있다고 판단되었다. 이 결과를 이용하여 의해 재해 발생 형태별로 상관이 높은 불안전행동 및 불안전 상태 관련 요인을 집중 관리하되 유사 재해의 발생 가능성을 줄일 수 있을 것이다.

또, 재해 관련 요인들을 범주형 변수로 보고 재해정도, 즉 입원일수를 종속변수로 한 χ^2 -Test 결과, 재해자의 나이, 성별, 직종, 근속기간, 상해종류, 재해발생 형태 등이 입원 일수에 영향을 주는 것으로 나타났으며 그 중 가장 큰 영향을 주는 상관 요인은 상해 종류인 것으로 나타났다. 이 결과는 재해 요인들의 연관성 우선 순위를 측정하는 데 이용할 수 있다고 판단되었다.

한편 로짓분석을 통해 가능성 대비에 의한 독립 변인의 효과를 검증하였는데 건설업종이 비건설업종에 비하여 경상보다 중상의 발생 가능성이 큰 것으로 나타났으며, 재해형태나 건설업 여부에 관계없이 다른 재해형태에 비해서 추락, 붕괴, 전도가 발생하면 경상에 비해 중상이 일어날 가능성이 1.230배 높은 것으로 나타났다.

또한 비건설업의 경우 근무형태에 대한 재해원인이 작업자의 음주, 흡연, 피로 등의 개인적 내부요인인지 아니면 작업장 환경요인인지를 검토하였는데, 특히 시간외 근무 및 휴일근무는 정상근무시에 비해 1.384배만큼 작업장 환경요인이, 교대 근무는 정상근무에 비해 1.487배만큼 작업자 내부요인이 큰 것으로 나타났다.

參 考 文 獻

- [1] 鄭國三, 災害事故分析論, 忠北大學校 工科大學 安全工學科. (1993)
- [2] 安全工學協會, 化學工場における事故原因解析のためのエスパートシステムの構築, pp 175-181.(1990)
- [3] 忠北大學校 工科大學 産業科學技術研究所, 災害豫防을 위한 勤勞者 特性에 관한 研究, 勞動部, pp161-167, 465-478.(1993)
- [4] Marija J. Norusis/SPSS Inc, SPSS for windows, B1-B151.(1992)
- [5] 박광배, 頻度分析, 星苑社, pp32-42, pp99-118.(1992)
- [6] 허명희, SAS 범주형 데이터 分析, 高麗大學校 統計研究所 pp1-19, 23-39.(1992)
- [7] 李昶勳, 全永鎬, 工業統計學, 博英社, pp299, 371-379.(1992)
- [8] 李觀雨, 調査分析 方法論, 螢雪出版社, pp161-167, 465-478, 537-580 (1990)
- [9] 韓國産業安全公團, 社會統計로서의 産業災害統計學, pp1-20.(1994)
- [10] 남궁평, 홍중선, 범주형 資料의 統計分析, 자유아카데미, pp24-46. (1991)
- [11] 李昶勳, 全永鎬, 工業統計學, 博英社, pp299, 371-379.(1992)
- [12] '95産業災害分析, 勞動部 (1996)