

보건분야 - 연구자료
연구원 2000-37-187
H-RD-I-2000-37-187

산업보건 데이터웨어하우스를 이용한 산업재해 추이 및 예측모델

1999



한국산업안전공단
산업안전보건연구원

제 출 문

한국산업안전공단 이사장 귀하

본 연구를 1999년도 산업안전보건연구원의 연구사업 중 “산업보건 데이터웨어하우스를 이용한 산업재해 추이예측 모델에 관한 연구”에 대한 최종 결과 보고서로 제출합니다.

1999년 12월 31일

제출자 : 산업안전보건연구원장 정 호 근

연구책임자 : 선임연구원 이 관 형

공동연구자 : 수석연구원 박 정 선

원장 정 호 근

요 약 문

1. 과 제 명 : 산업보건 데이터웨어하우스를 이용한 산업재해 추이 예측모델에 관한 연구
2. 연구기간 : 1999년 1월 - 12월
3. 연 구 자 : 산업보건위생연구실 선임연구원 이관형
4. 연구목적 : 산업재해 발생 추이 및 예측 모델 개발
5. 연구내용 : 과거에 발생한 산업재해 시계열 자료를 탐색과 분석을 통해 미래시점에 있어서 산업재해가 정량적으로 발생할 수 있는 규모를 예측하려 한다.
6. 활용계획 : 산업재해 예방 장.단기 계획 수립에 따른 전략적 대응방안을 마련하는데, 예측지표로 사용
7. 연구개요 : 본 연구는 재해자의 발생패턴과 향후 발생방향을 설명해준다. 연구의 기초자료는 산업재해가 발생한 시점을 기준으로 1986년 1월부터 1999년 7월까지의 월 산업재해 발생재해자수로 시계열 예측모형 선정 분석에서 지수평활모형에서 승법 로그 윈터스 모형이 가장 로버스트한 예측모형으로 선정되었다. 이 모형으로부터 향후 1년간 산업재해 발생이 정량적 규모로 어느 정도 재해가 발생될 것인지를 예측하였다.
8. 중 심 어 : 산업재해, 추이, 예측, 지표

차 례

I. 서론

1. 배경	1
2. 연구목적	3
3. 이론적 고찰	4

II. 연구설계 및 방법

1. 연구자료	8
2. 시계열 예측 목적과 방법 틀	9
3. 연구분석 결과	10

III. 결론

25

참고문헌

27

I. 서론

1. 배경

우리 나라의 산업재해 통계 집계 체계가 체계적으로 잡히기 시작한 것은 노동청 근로기준국 및 지방사무소에 산업안전담당관을 두면서 산업재해 업무가 제 제도에 들어선 시기로 1972년부터 1998년까지 27년간 발생한 총 재해자수는 2,942,478명이며, 이 가운데 산업재해 사망자수는 46,204명, 직업병자수는 35,311명으로 집계되고 있으며, 평균 1년에 10만 여명의 재해자와 1천 7백 명의 사망자가 산업현장에서 발생되고 있는 실정이다.

또한 산업재해로 인한 경제적 직접손실액(산재보상금 지급액)만도 10조 5천억원에 달하는 등 그 피해는 이루 말 할 수 없다.

산업화에 따라 꾸준히 증가되던 재해율은 1985년을 정점으로 감소되어 왔으며, 1995년에 산업재해가 기준 재해율 1%미만인 0.99%을 처음으로 달성하여, 1996년 0.88%로 떨어져 10년 전인 1986년(2.99%)의 3분의 1로 줄었다. 이런 감소 추세가 1997년 0.5%로 떨어졌으나, 1998년 산업재해 분석에 따르면 아직도 전국의 산업현장에서 하루 평균 141명(사망은 6명)의 재해자가 발생되고 있고, 이로 인한 경제적 직·간접적 손실 추정액은 7조 3천억원에 이르고, 노동 손실일수(41,511천일)도 노사분규에 의한 노동 손실일수(1,452천일)의 29여배이며, 노사분규로 인한 생산차질액 16,363억원보다 산업재해로 인한 총 경제적 손실 추정액의 4.4배에 이르는 등 산재예방 후진국을 면치 못하고 있는 실정이다.

또한 다른 유형의 재해로 경제적 손실액을 비교하면 교통사고 4,830억원(96년 기준), 화재사고 1,217억원(97년 기준), 풍수해 사고 4,830억원(96년 기준)의 경제적 손실액에 비해 16배~64배에 해당된다.

그러나 재해자 수는 6만6,770명으로 전년도 보다 4,778명이 감소했으며 이에 따라 재해율도 0.07% 포인트 낮은 0.88%였다. 재해율이 전반적으로 낮아지고 있

음에도 산재로 인한 사망자수는 1996년보다 72명이 증가한 2,742명이며, 10년 전인 1988년 사망자수 보다 981명 증가했으며, 매년 점진적으로 증가 추세에 있다.

이는 지난 30여년간 성장위주의 경제정책으로 새로운 생산구조와 제조공정이 도입되는 과정에서 경제적 효율성과 생산성만이 강조되었고, 생산과정에 참여하는 근로자의 안전과 보건은 상대적으로 경시된 탓이다. 경제성장의 가속화로 인하여 우리의 산업현장에서는 생산체계의 복잡 다변화, 신종기계·설비의 끊임없는 개발·도입·신규 화학물질과 에너지 사용의 증가, 기존 산업시설의 노후화 등이 급속히 진전되면서 재래형 반복 재해 뿐 만 아니라 신종재해와 직업병이 복합적이며, 지속적으로 발생되고 있어 많은 근로자의 생명과 건강이 크게 위협 받고 있다.

그럼에도 불구하고 수년 전에 발생한 “서해 페리호 침몰”, “대구 지하철 가스 폭발 사건”, “삼풍백화점 붕괴” 최근의 “화성 씨랜드 화재사건” 등에서 알 수 있듯이 우리 사회에 팽배해 있는 안전불감증과 인명경시 풍조는 개선되지 않고 있어 산업현장 어느 곳에서나 재해 위험이 잠재해 있다. 또한 이로 인한 인적·물적 손실의 피해로 국제사회에서는 “산재다발국”이라는 오명을 가질 수 밖에 없었다.

이에 1996년 발표된 산업선진화 3개년 계획은 국제 경제사회 구조개편에 따른 새로운 규범의 등장, 경제수준에 걸맞는 국제적 위상제고의 필요성이 강조되고, 앞으로 국제 환경의 변화는 국제적으로 향상된 지위와 경제 규모 등 국력에 상응하는 전략적 대응방안 차원에서의 안전·보건 기준을 국제적 시각에서 개선할 필요성이 시급하다고 할 수 있겠다.

또한 산업구조의 변화로 인한 산업재해의 다양화, 인력구성의 변화에 따른 새로운 예방 행정수요의 확대, 노동시장 수급상황의 변동에 따른 재해예방 수요 증가 등의 국내 환경의 변화는 정치적 민주화가 성숙되어감에 따라 산업민주주의도 폭과 깊이를 더해 가면서 산업현장에서의 인명존중 풍토가 확산되고 있고, 이에 소득수준 향상과 근로 조건의 개선으로 노동자의 생존권적 욕구가 생활권적 욕구로 발전하면서 재해예방에 대한 노동자 참여권 보장 등 안전·보건문제가 노사분규의 주요쟁점으로 부상할 것으로 예방된다. 또한 노동자 “삶은 질” 수준 향상으로

생활의 안전과 고학력화의 진전으로 작업환경개선과 생산성 향상을 동시에 추구하는 “노동의 인간화(Humanization of Work)”와 “노동 생활의 질(Quality of Working Life)”에 대한 관심이 고조될 것이다.

근본적으로 산업재해 예방하기 위한 전략적 장·단기적 대응방안이 없이는 노동자의 삶의 질 향상이나 21세기 기업의 경쟁력은 물론 산업안전선진국으로의 진입을 기대할 수 없다.

산업재해가 늘고 있는 것은 노동자의 부주의도 있지만 근본적으로 기업이나 정부의 산재예방에 대한 관심과 투자가 부족하기 때문이다. 특히 50인 미만의 사업장에서 전체 재해의 절반 이상 발생하고 입사 6개월 미만의 노동자가 40%나 되고 있다는 것으로 입증된다.(독일에서는 입사 1년 미만인 경우는 거의 산재발생이 없음)

중화학공업과 대형사업장의 증가에 따라 산재의 규모도 대형화되고 새로운 유형의 직업병도 발견되고 있으며, 또한 증가될 것으로 추정된다. 고령 및 여성노동자의 증가와 함께 최근 인력감축으로 노동자들의 노동강도가 높아지고 있는 점도 재해다발의 요인이란 점에서 관심을 높여야 할 때이다.

이와 함께 산재다발 사업장에 대해서는 사업주의 처벌 등 단호한 조치가 필요하다. 정부와 기업은 산재로 인한 인적, 물적 피해를 줄이는 일이 가시적인 인력양성이나 생산성 향상을 이루는 일 못지 않게 중요하다는 점을 명심해야 한다.

2. 연구목적

본 연구는 산업재해 예측은 복잡 다변하는 산업구조와 향후 미래에 발생될 재해자 예측을 통해 산업재해 예방 전략수립 사업계획 및 방향을 제시 자료로 제공이 가능하고, 예측 재해의 수치적 중요성 외에 중요한 산업재해 발생 요인의 직·간접적인 영향들을 체계적으로 분석 정리하여 산업안전보건 정책 결정권자에게 의사결정 방향제시 자료를 제공하기 위한 목적이 있다 하겠다.

3. 이론적 고찰

가. 의사결정 모형

(1) 계량 의사결정 모형

관심있는 변수와 관련변수에 대한 과거와 현재의 자료로부터 시간 의존적인(time-dependent) 또는 연관 변수간의 횡단면적인(cross-sectional) 상호관계에서 일정한 형태의 패턴을 찾아 수요예측모형을 구성하는 방법으로, 과거의 자료가 존재하지 않거나 부족한 경우에도 모형을 추정하는 데 어려움이 있다.

일반적인 시계열 모형은 자료의 형태, 분석의 용이성, 분석자료 해석의 이해 정도에 따라 회귀분석법, 박스-젠킨스(Box-Jenkins)방법, 지수 평활법(exponential smoothing), 시계열 분해방법 등 네 가지로 분류할 수 있다. 회귀분석 방법과 박스-젠킨스 방법은 수학적 이론을 바탕으로 한 체계적인 방법인 반면, 시계열 자료의 분해방법과 지수 평활법은 경험적이고 직관적인 방법이라 할 수 있다. 또한 다중 시계열 방법으로는 전이함수 모형, 개입모형, 상태공간 분석, 다변량 ARIMA 모형, 구조방정식 모형 등이 있는데, 본 연구에서 이용되는 모형은 분석하고자 하는 시계열 자료가 시간의 흐름에 따라 느리게 변동할 때 효과적인 예측방법으로 지수 평활법 모형으로 적용하였다.

(2) 비계량 의사결정 모형

미래에 전개될 기업환경변화, 기술변화 등을 예측하는데 유용한 모형이며 사용대상자나 전문가의 경험과 지식 그리고 직관력 등을 수렴하여 분석하는 기법으로 계량의사결정 모형에 비해 객관성이 떨어지고 비용이 더 소요되는 단점이 있는데, 널리 사용되는 비계량 의사결정 모형으로는 시장조사(market survey)법, 비교 유추법(유사서비스, 외국자료조사), Delphi법, 집단회의(Brain Storming)법, 의견 교환법, 행태학적(morphological) 분석, 관련지(relevance tree)법, 시나리오법, 선호서열 분석법 등이 있다.

나. 시계열분석에 대한 이론적 고찰

(1) 기본적 시계열 모형(Box-Jenkins model)

시계열 자료를 분석하는 주목적은 과거의 값들로부터 미래의 값을 예측하는 데 있다. 이러한 예측기법에는 자료의 종류, 예측치의 정확성에 대한 요구 정도, 기법 사용의 난이도 등에 따라 다양한 방법들이 있다. 이들 중 본 연구에서 사용될 지수 평활법(exponential smoothing) 중 윈터스 모형에 대해 살펴보면 다음과 같다.

먼저 시계열 모형구축 절차는 3단계로 구성된 반복절차로, 첫째, 시계열 자료로부터 얻은 여러 가지 모형 식별 통계량을 이용하여 잠정적인 모형을 선택하고, 선택된 시계열 모형의 모수를 추정한 다음, 선택된 모형의 적합성을 진단하여 부적합한 경우 첫 번째 단계로 돌아가 모형을 수정하고, 다시 추정, 진단과정을 반복하여 만족스러운 모형이 선택될 때까지 계속해 간다.

이들 단계를 자세히 살펴보면 다음과 같다.

(가) 모형식별

모형의 식별이란 시계열자료에 의해 그에 적합한 자기회귀이동평균모형을 찾기 위해 차수를 정하는 단계이다. 차수를 정하기 앞서 비정상 시계열 자료인 경우 분산성과 추세제거를 위해 변환과 적절한 차분을 하여 자료를 정상 시계열 자료로 변환시킨 후, 이 정상 시계열의 자기상관함수(ACF)와 부분자기상관함수(PACF)의 패턴을 검토하여 차수를 결정한다. 즉, 차수에 따라 ACF와 PACF가 일정한 패턴을 보이므로 이를 분석하여 차수를 정하게 된다.

(나) 모수추정

모형의 추정을 위해 사용되는 방법은 크게 3가지 방법이 있다. 첫째, 조건부 최소제곱(Conditional Least Squares)추정법은 모수 추정을 간편하게 하기 위해 처음 몇 개의 관측값을 조건화하여 잔차 제곱합을 최소화하는 방법이다. 둘째, 비조건부 최소제곱 (Least Squares)은 앞과 같은 조건 없이 잔차 제곱합을 최소화하는 방법이다.

며, 마지막으로 최우추정법(Maximum Likelihood)은 표본계열에 대한 우도를 수치적으로 최대화하는 방법으로 이 방법의 경우 많은 계산을 필요로 한다.

모수 추정에 있어 통상적으로 AIC(Akaike information criterion)와 SBC(Schwarz bayesian criterion)통계량으로 판단하여 결정하는데 이 수치가 적을수록 선정된 모형이 더 좋다고 판정한다.

(다) 모형진단

추정된 모형의 적합성을 검토하는 단계로 윈터스 지수 평활법 분석에서는 적합한 잔차의 임의성(random)여부를 확인하는 것이 진단의 기본이다. 잔차의 random성 여부를 검토하는 방법에는 잔차분석법과 포트만토 검정법(Portmanteau test)이 있다. 모형의 적합성 여부를 판단한 결과 모형이 적합하지 않을 경우, 모형 식별 단계로부터 다시 시작하여 최적의 모형을 찾을 때까지 반복적인 작업을 계속 수행하는 것이다.

(라) 예측

앞의 절차에 따라 시계열에 적합한 모형이 설정되면 설정된 모형을 이용하여 미래 시점의 시계열 값을 예측하게 된다. 이때, 시계열 예측은 시계열의 모형을 정확히 알고 있고 이 모형에 포함된 모수들의 추정오차는 시계열 예측에 크게 영향을 미치지 않는다는 가정 하에 이루어진다. 이를 그림으로 도식화 하면 다음과 같다.

(2) 개입모형(Intervention Model)

산업재해 시계열 예측시 산업구조, 정책 및 제도, IMF와 같은 등 여러 가지 경제정책 등과 같은 외적인 요인에 의해서 종종 영향을 받는다. 이와 같은 외적인 요인들을 개입이라 하는데 시계열 모형의 분석에 있어서 이러한 개입이 존재할 경우 이들의 효과를 평가하여 보다 정확한 시계열 모형을 설정하고 예측의 정도를 높이는 데 기여할 수 있다.

개입변수에는 두 가지 유형이 있는데 하나는 어떤 시점 T에서 발생한 개입이 그 시점 이후에도 지속적으로 효과가 유지되는 경우로서 변수는

$$S_t = \begin{cases} 0, t < T \\ 1, t \geq T \end{cases}$$

로 나타내며, 다른 하나는 천재지변처럼 시점 T에서만 충격을 나타내는 변수로

$$P_t = \begin{cases} 1, t = T \\ 0, t \neq T \end{cases}$$

로 나타낸다. 이들의 특징을 알아보기 위해 다음과 같은 개입모형을 가정해보자.

$$Z_t = w_0 + w_1 P_t + (\text{Noise})_t$$

여기서 $t < T$ 이면 $Z_{\text{sub } t} = w_{\text{sub } 0} + (\text{Noise})_{\text{sub } t}$ 이고, $t \geq T$ 이면 $Z_{\text{sub } t} = w_{\text{sub } 0} + w_{\text{sub } 1} + (\text{Noise})_{\text{sub } t}$ 이므로 시점 T에서 시계열의 수준이 w_0 에서 $(w_0 + w_1)$ 로 이동되었음을 알 수 있다.

II. 연구설계 및 방법

본 산업재해 예측 모형 연구에는 지수 평활법 중 윈터스 모형기법을 이용하여 산업재해 발생자수 예측을 위한 최적의 모형을 도출하고 이를 이용해 재해자수를 예측하려 한다.

1. 연구자료

본 연구에 사용된 자료는 1986년 1월부터 1999년 9월까지 산업현장에서 재해가 발생한 시점을 기준으로 월별로 누적된 산업재해자 수로 163개월 분(13년 7개월) 자료이다. 발생시점에 대한 월별 누적 산업재해자수에 대한 도표는 표 1과 같다.

표 1. 우리나라의 월별 산업재해 발생현황 자료(1986년 - 1999년)

시간	발생수	시간	발생수	시간	발생수	시간	발생수	시간	발생수	시간	발생수	시간	발생수
1	8768	25	10604	49	9147	73	9292	97	6736	121	6259	145	3393
2	7399	26	8872	50	9703	74	7569	98	5252	122	5447	146	3840
3	10319	27	11982	51	11725	75	9556	99	7654	123	7349	147	4144
4	10855	28	11520	52	11451	76	9884	100	7648	124	7336	148	3986
5	11383	29	12213	53	12084	77	10057	101	7663	125	7936	149	4049
6	11668	30	13068	54	11702	78	10090	102	7966	126	7504	150	3921
7	12448	31	13368	55	12197	79	9713	103	7933	127	7792	151	4088
8	12361	32	14036	56	11987	80	8603	104	7745	128	7476	152	4005
9	10372	33	11090	57	11387	81	7484	105	6517	129	6898	153	4492
10	11412	34	12536	58	9973	82	8717	106	7544	130	8390	154	4158
11	10807	35	12210	59	11255	83	8129	107	7486	131	7507	155	4369
12	10366	36	11739	60	11048	84	7495	108	7518	132	5592	156	4100
13	8957	37	9993	61	10288	85	5692	109	6270	133	5447	157	3317
14	9777	38	8249	62	7842	86	6717	110	6112	134	4423	158	2712
15	11820	39	11461	63	10862	87	7814	111	7758	135	6179	159	4109
16	12145	40	11407	64	11333	88	7841	112	7833	136	6053	160	4273
17	13071	41	11836	65	11329	89	7890	113	8285	137	6192	161	4360
18	13433	42	12068	66	11100	90	8243	114	8639	138	6531	162	4574
19	13616	43	12310	67	11403	91	8037	115	8185	139	6221	163	4258
20	12687	44	12527	68	11479	92	7518	116	8684	140	5964		
21	13398	45	10512	69	9580	93	7523	117	7957	141	5290		
22	10516	46	12125	70	11741	94	7482	118	9020	142	6324		
23	12242	47	11355	71	10830	95	8086	119	8036	143	5180		
24	11801	48	11505	72	9652	96	7603	120	6484	144	4867		

* 본 자료는 노동부에서 발표된 공식통계가 아니고, 발생시점이기 때문에 시점에 따라 자료의 변동이 있음.

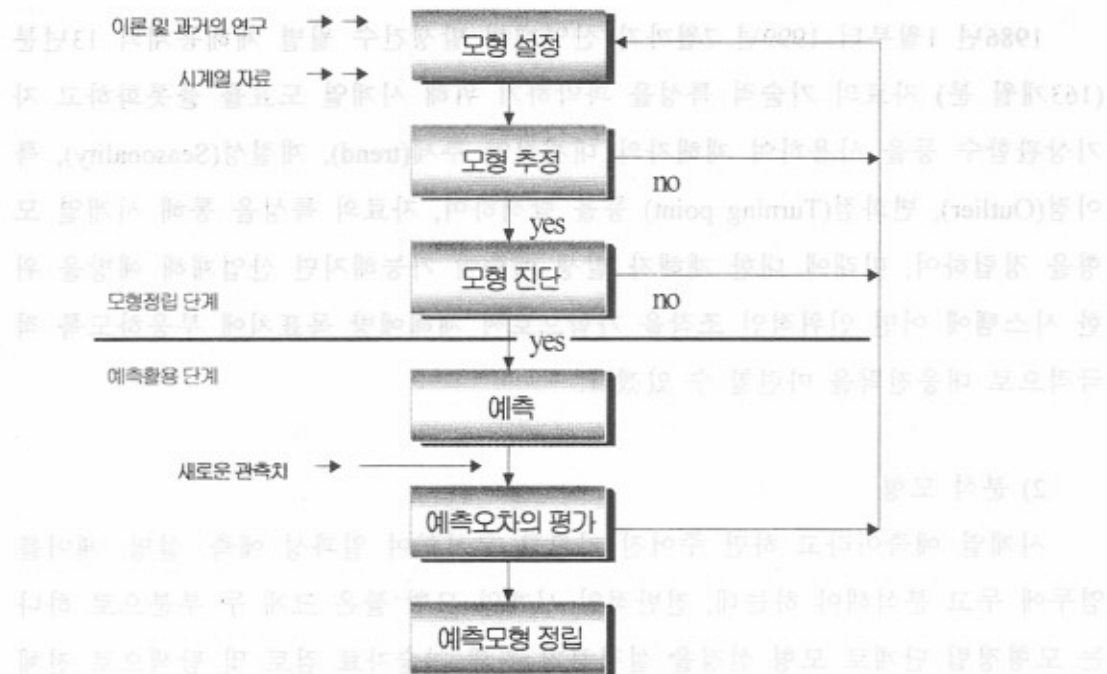
2. 시계열 예측 목적과 방법 틀

1) 분석 목적

1986년 1월부터 1999년 7월까지 산업재해 발생건수 월별 재해통계의 13년분 (163개월 분) 자료의 기술적 특성을 파악하기 위해 시계열 도표를 플롯화하고 자기상관함수 등을 사용하여 재해자의 대체적인 추세(trend), 계절성(Seasonality), 특이점(Outlier), 변화점(Turning point) 등을 탐색하여, 자료의 특성을 통해 시계열 모형을 정립하여, 미래에 대한 재해자 발생 예측이 가능해지면 산업재해 예방을 위한 시스템에 어떤 인위적인 조작을 가함으로써 재해예방 목표치에 부응하도록 적극적으로 대응전략을 마련할 수 있겠다.

2) 분석 모형

시계열 예측이라고 하면 주어진 자료를 분석하여 일과성 예측, 설명, 제어를 염두에 두고 분석해야 하는데, 전반적인 시계열 모형 틀은 크게 두 부분으로 하나는 모형정립 단계로 모형 설정을 설정하기 위해 기술자료 검토 및 탐색으로 전체 분석의 80% 정도 차지하며, 모형 설정 다음으로 모형추정과 모형진단을 통하여 두 번째 단계인 예측활용 단계로 모형 정립으로부터 예측하고 예측오차의 평가를 통해 모형을 최신화 시켰다. 이를 도식화하면 그림 1과 같다.



[그림 1] 산업재해 예측 모델 개념도

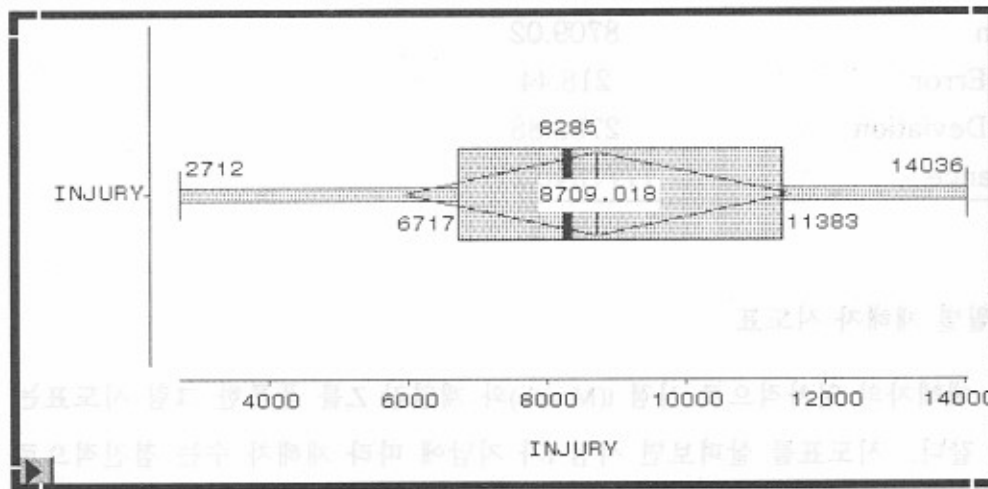
3. 연구분석 결과

1) 자료의 탐색

가) 월별 재해자수의 기술통계

본 연구의 산업재해 예측에 사용된 자료는 1986년 1월부터 1999년 7월까지 산업현장에서 재해가 발생된 월별 재해자수에 대한 기술통계량과 박스그래프는 다음과 같다(표 2;그림 2).

산업재해 발생 기준 시점인 1999년 7월까지의 월 평균 산업재해 발생은 8,709명이 발생되고 있으며(95% 신뢰구간: 8,277(하한값)명, 9,140(상한값)명), 사분위 통계량에서 살펴보면 중앙값(Q1: 50%)은 8,285명으로 분석되었다(Q3(75%): 11,383명; Q1(25%): 6,717명).



[그림 2] 월별 재해자 발생현황 박스그래프

<표 1> 산업재해 월별 재해자 4분위수 통계량

특성	통계량
100% Max	14036
75% Q3	11383
50% Med	8285
25% Q1	6717
0% Min	2712
Range	4666
Q3-Q1	5447

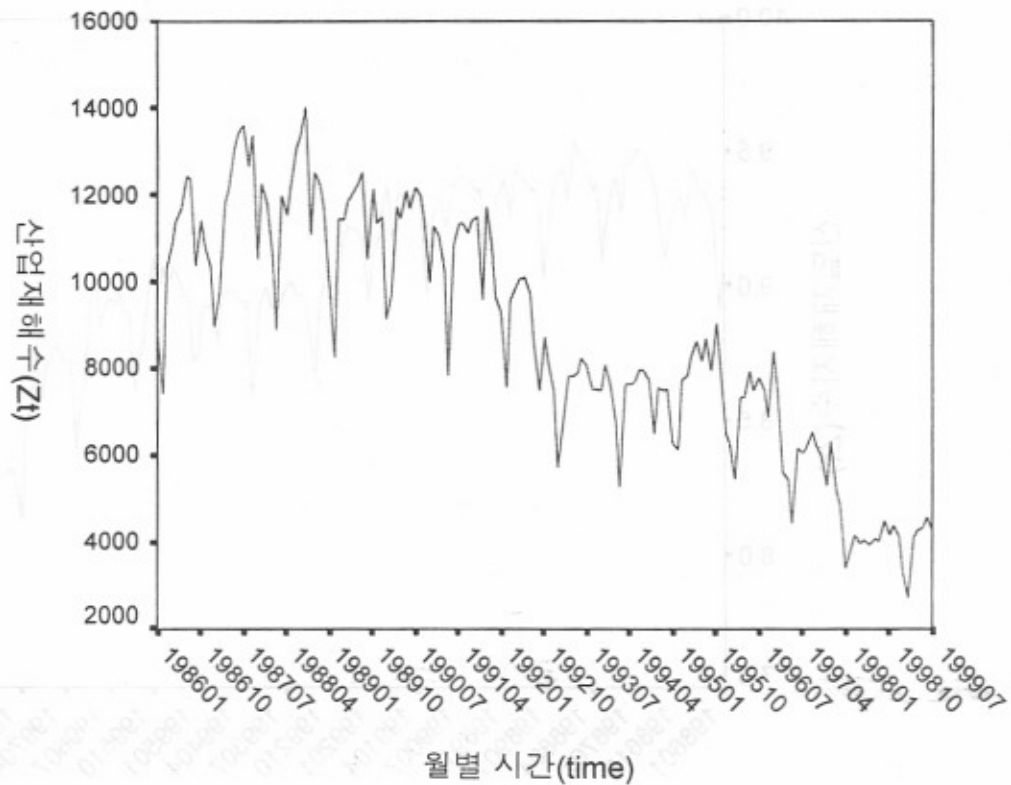
<표 2> 월별 산업재해 발생 재해자의 기술통계

특성	통계량
Range	11324
Minimum	2712
Maximum	14036
Mean	8709.02
Std. Error	218.44
Std. Deviation	2788.88
Variance	7777840.277

나) 월별 재해자 시도표

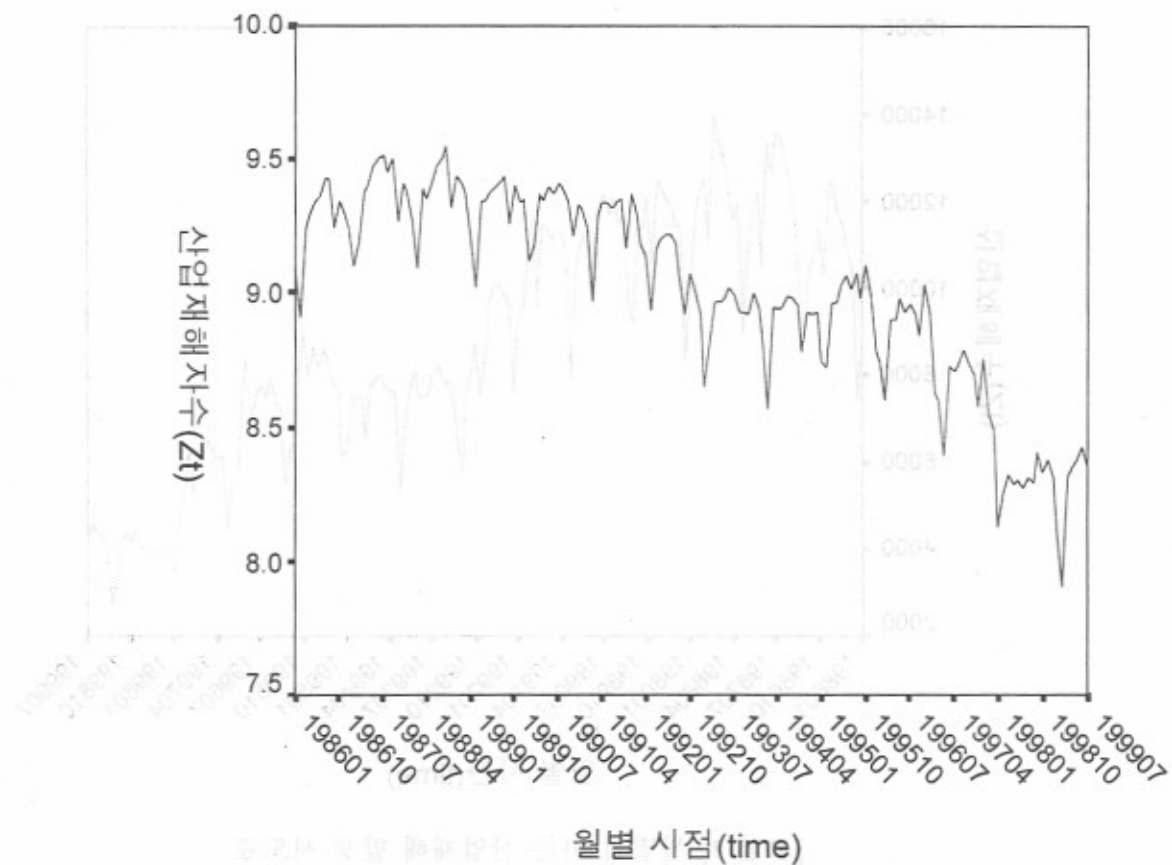
월별 재해자의 일차적으로 시점 t (Month)와 계열값 Z_t 를 플롯한 그림 시도표는 그림 2와 같다. 시도표를 살펴보면 시점 t 가 지남에 따라 재해자 수는 점진적으로 감소하는 추세를 보이고 있으며, 또 다른 특성을 보면 월별 계절 변동이 보이며, 순환 주기는 1년이며, 월별 재해자 시도표에서 변화점은 1992년 5월과 1997년 10월을 기점으로 월별 재해자 감소되는 부분에 있어, 시계열 패턴과 더불어 갑작스러운 변화가 있었다(그림 3).

통계량	통계량
14036	100% Max
11324	75% Q3
8709	50% Med
6217	25% Q1
2712	0% Min
11324	Range
7778	Q3-Q1



[그림 3] 월별에 따른 산업재해 발생 시도표

우선 월별 산업재해 발생재해자의 원자료의 정상성(stationarity) 여부를 파악하기 위해 시도표를 보면, 시간이 지남에 따라 분산의 변동이 심하고, 추세성이 있는 비정상 시계열임을 알 수 있기 때문에 시계열 자료의 정상성을 위해 변수변환과 차분을 시도하였다.

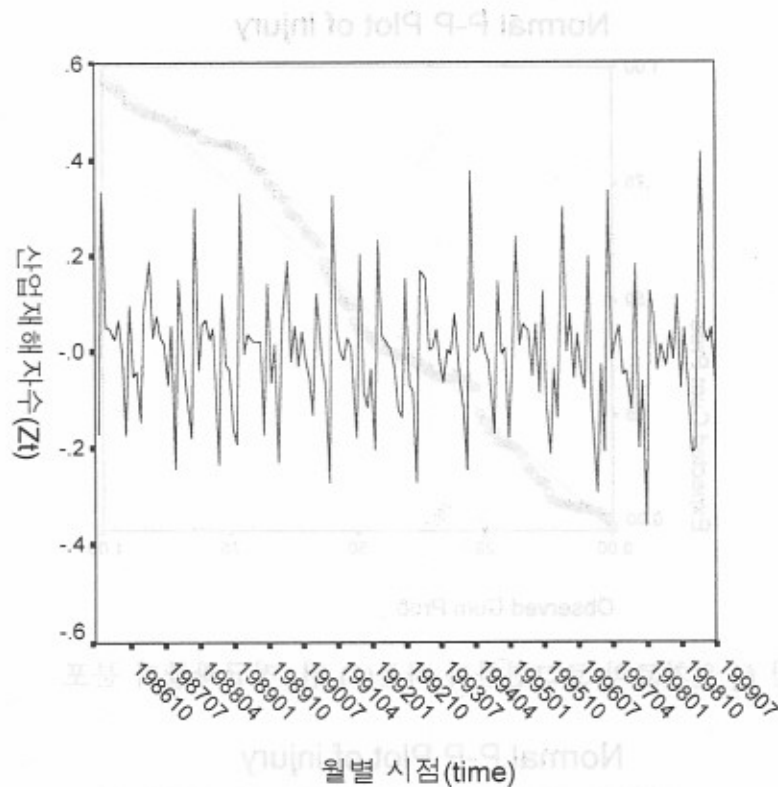


[그림 4] 원자료의 변수변환 후(natural log), 시도표

다) 자료의 변환과 차분

[그림 3]와 같이 시간이 경과함에 따라 분산이 일정하지 않고 감소하는 경향이 뚜렷하기 때문에 우선 분산을 안정화 시키기 위해 로그변환을 하였으나, 아직도 로그변환 한 경우에도 추세는 제거되지 않아(그림 4), 분산과 추세를 제거하기 위해 다시 1차 차분하였다.

[그림 5]에서 보면 1차 차분된 시계열자료의 시도표에서는 추세성분이 제거되어 있음을 알 수 있다. 따라서 1차 차분된 시계열자료는 정상성을 만족하고 있다고 할 수 있다.

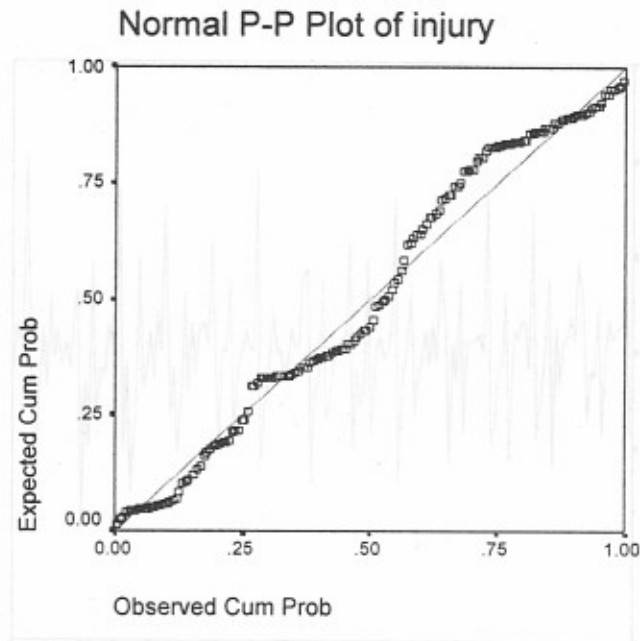


[그림 5] 로그변환과 1차 차분 후, 정상적인 시계열 시도표

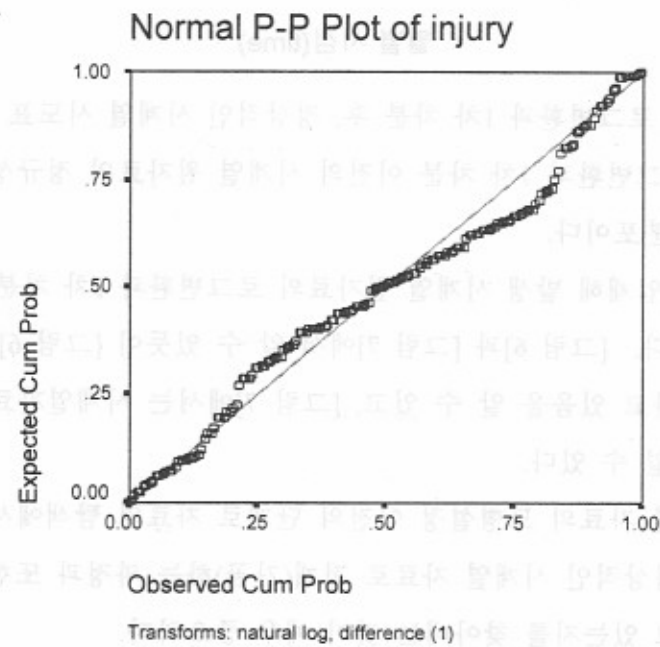
[그림 6]은 로그변환과 1차 차분 이전의 시계열 원자료의 정규성을 검증하기 위한 정규확률지 분포이다.

[그림 7]은 산업재해 발생 시계열 원자료의 로그변환과 1차 차분된 후, 정규분포를 보여주고 있다. [그림 6]과 [그림 7]에서 알 수 있듯이 [그림 6]은 P-P 플롯에서 정규성에 벗어나고 있음을 알 수 있고, [그림 7]에서는 시계열자료가 정규성을 만족하고 있음을 알 수 있다.

따라서 시계열 자료의 모형설정 이전의 단계로 자료의 탐색에서 비정상적인 시계열자료에서 정상적인 시계열 자료로 정제(가공)하는 과정과 또한 정제된 자료가 정규분포를 하고 있는지를 찾아내는 것이 매우 중요하다.



[그림 6] 원자료의 로그변환과 차분(1) 전, 정규확률지 분포



[그림 7] 원자료의 로그변환과 차분(1) 전, 정규확률지 분포

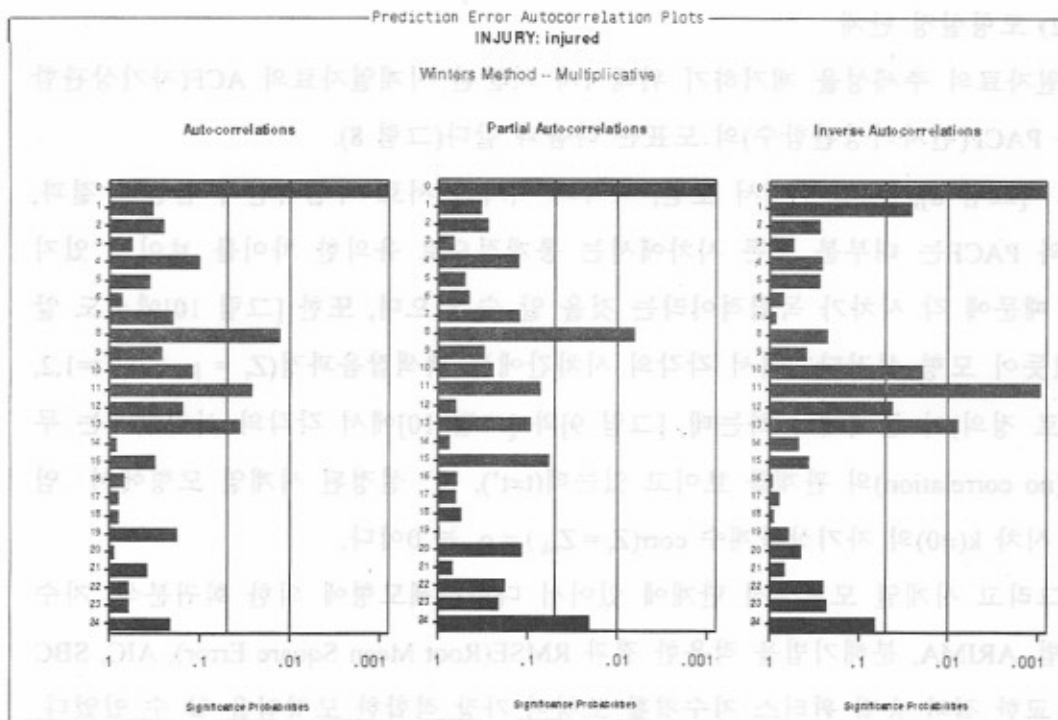
2) 모형설정 단계

원자료의 추세성을 제거하기 위해 1차 차분된 시계열자료의 ACF(자기상관함수)와 PACF(편자기상관함수)의 도표는 다음과 같다(그림 8).

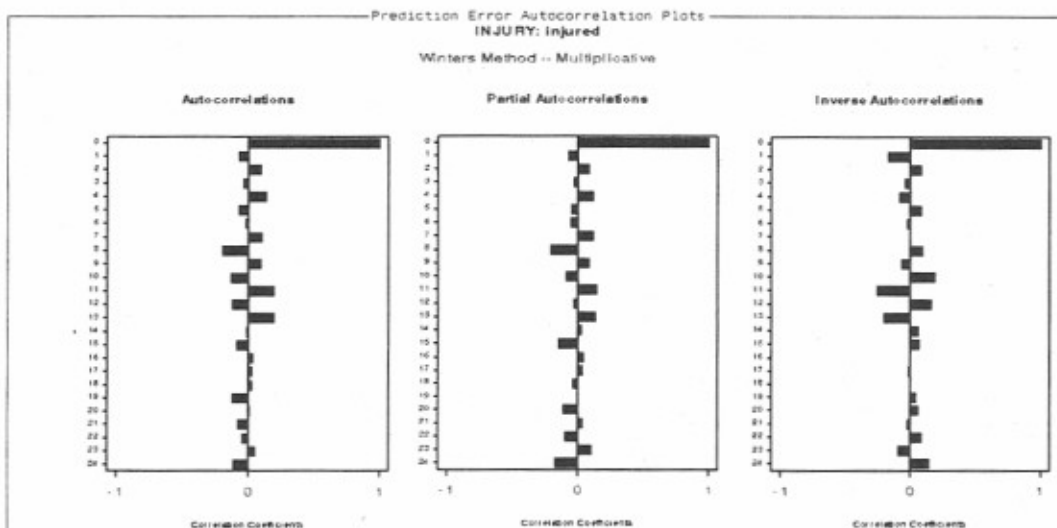
[그림 8], [그림 9]에서 보면, 각각의 시차가 서로 독립적인지 검증한 결과, ACF와 PACF는 대부분 모든 시차에서는 통계적으로 유의한 차이를 보이고 있지 않기 때문에 각 시차가 독립적이라는 것을 알 수 있으며, 또한 [그림 10]에서도 알 수 있듯이 모형 설정단계에서 각각의 시차간에는 백색잡음과정($Z_t = \mu + a_t$, $t=1,2, \dots, T$ 로 정의)이 존재해야 하는데, [그림 9]와 [그림 10]에서 각각의 시차간에는 무상관(no correlation)의 관계를 보이고 있는데($t \neq t'$), 즉 설정된 시계열 모형에서 임의의 시차 $k(k \neq 0)$ 의 자기상관계수 $\text{corr}(Z_t = Z_{t+k}) = \rho_t$ 는 0이다.

그리고 시계열 모형선정 단계에 있어서 다항추세모형에 의한 회귀분석, 지수평활법, ARIMA, 분해기법을 적용한 결과 RMSE(Root Mean Square Error), AIC, SBC를 비교한 결과 승법 윈터스 지수평활 모형이 가장 적합한 모형임을 알 수 있었다.

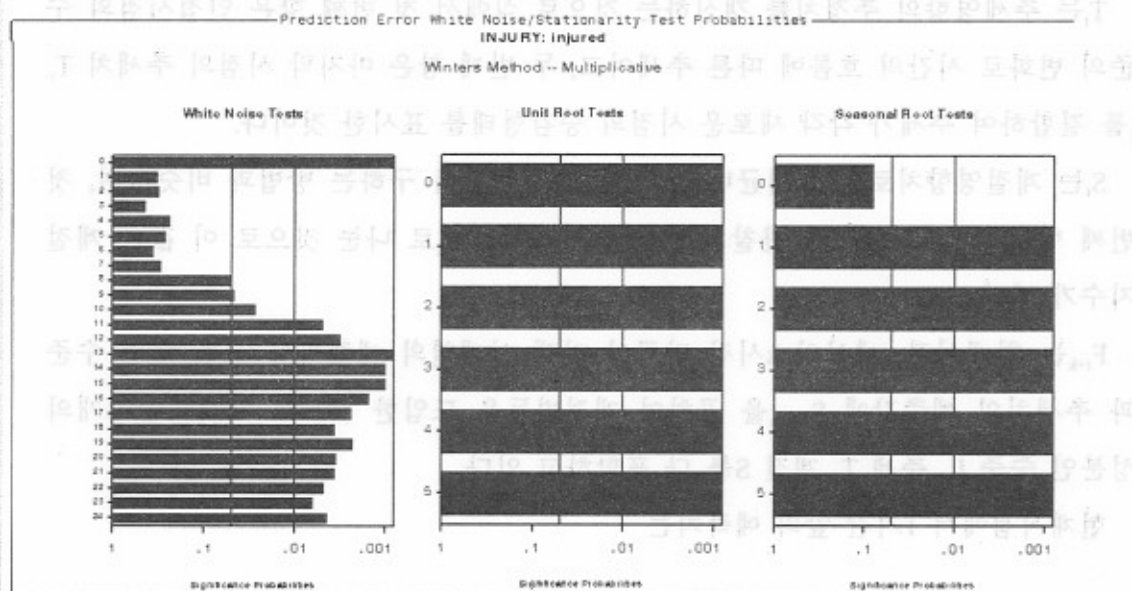




[그림 8] 산업재해 예측 모형설정을 위한 ACF와 PACF 통계적 유의확률 분포



[그림 9] 산업재해 예측 모형설정을 위한 ACF와 PACF의 자기상관계수



[그림 10] 산업재해 예측 모형 설정을 위한 백색잡음과정

3) 설정된 모형의 모수 추정결과

모형 탐색과 반복적인 모형설정으로 선정된 승법 윈터스 모형의 모수 추정에 대한 기본모형식은 다음과 같다.

$$\text{수준모형(Level)} : L_t = \alpha(Y_t / (S_{t-m})) + (1-\alpha)(L_{t-1} + T_{t-1})$$

$$\text{추세영향치(Trend)} : T_t = \beta(L_t + L_{t-1}) + (1-\beta)L_{t-1}$$

$$\text{계절영향치(Seasonal)} : S_t = \gamma(Y_t / L_t) + (1-\gamma)(S_{t-m})$$

$$\text{예측치} : F_{t+k} = (L_t + kT_t)S_{t+k-m}$$

여기서 Y_t 는 t 시점의 관찰치이고, L_t 는 t 시점의 계절수준평활치로 식에서 $Y_t / (S_{t-m})$ 은 현 시점의 자료를 m 시점 이전의 계절지수로 나눈 것으로 계절영향을 제거한 것이고, $(L_{t-1} + T_{t-1})$ 은 추세요인에 대한 수준을 수정한 것으로 추세의 변동영향을 포함시켜 수준을 갱신한다. 따라서 수준모형은 현재시점의 계절영향이 제거된 자료에 추세영향을 추가한 평활치라고 할 수 있다.

T_t 는 추세영향의 추정치를 갱신하는 것으로 식에서 첫 번째 항은 인접시점의 수준의 변화로 시간의 흐름에 따른 추세이고, 두 번째 항은 마지막 시점의 추세치 T_{t-1} 를 결합하여 추세가 각각 새로운 시점의 증감형태를 표시한 것이다.

S_t 는 계절영향치로 이동평균비율법에서 계절지수를 구하는 방법과 비슷하며, 첫 번째 항인 $\gamma(Y_t/L_t)$ 는 현재 관찰치 Y_t 를 평활화계열 L_t 로 나눈 것으로 이 값은 계절지수가 된다.

F_{t+k} 는 현재시점 t 에서의 k 시점 만큼의 미래 시계열의 예측치로 식을 보면 수준과 추세치의 예측값에 S_{t+k-m} 을 곱하여 계절변동을 도입한 것으로 예측치는 3개의 성분인 수준 L , 추세 T , 계절 S 를 다 포함하고 있다.

현재시점에서 1기간 앞의 예측치는

$$F_{t+1} = \{L_t + (1)T_t\}S_{t+k-m}$$

으로 현재시점의 수준추정치에 현재시점의 추세 추정치를 더한 값(현재시점의 수준의 증감을 나타냄)에 마지막 년도의 계절지수를 곱하므로 예측치는 계절변동에 따라 계절지수의 기준보다 높아지거나 낮아진다. 이러한 방법으로 현재시점의 k 기간 예측치는

$$F_{t+k} = \{L_t + (k)T_t\}S_{t+k-m}$$

으로 예측치는 k 시점 앞이므로 추세는 k 개월 동안 계속되었으므로 현재 추세요인 T_t 에 k 를 곱하여 현재 수준치에 더하고, 계절지수를 곱하여 구했다.

이상과 같이 승법 윈터스 모형의 모수추정치에 대한 검정통계량과 검증결과를 보면 다음과 같다(표 3). 수준(L), 추세(T), 계절성(S)의 모수추정치에 대한 검정통계량에 통계적으로 매우 유의한 결과를 보여줌을 알 수 있다.

따라서 본 모형에서 설정된 윈터스 모형에 대한 추정은 매우 적합하다고 말할 수 있겠다.

<표 3> 산업재해 발생 예측 모형의 모수 추정결과

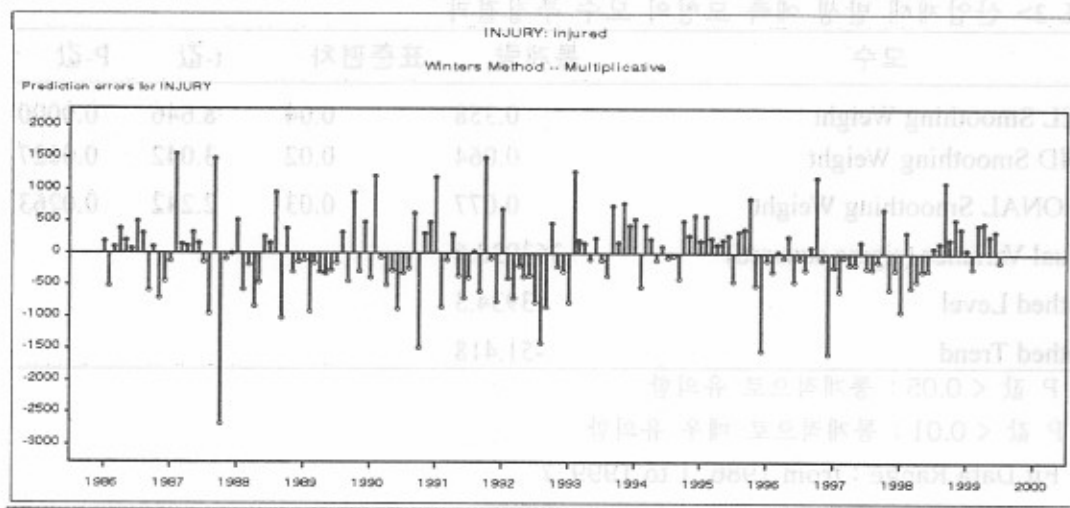
모수	통계량	표준편차	t-값	P-값
LEVEL Smoothing Weight	0.358	0.04	8.646	0.0000
TREND Smoothing Weight	0.064	0.02	3.042	0.0027
SEASONAL Smoothing Weight	0.077	0.03	2.242	0.0263
Residual Variance (sigma squared)	363938.6			
Smoothed Level	3934.3			
Smoothed Trend	-51.418			
+ P 값 < 0.05 : 통계적으로 유의함				
++ P 값 < 0.01 : 통계적으로 매우 유의함				
+++ Fit Data Range : from 1986. 1 to 1999. 7				

4) 설정된 모형의 적합성 검증결과

시계열 예측모형에서 설정된 승법 윈터스 모형의 적합성 검정을 위해 예측된 잔차를 분석한 결과, 추정된 잔차의 ACF와 PACF가 모든 시차에 대하여 통계적으로 유의한 값을 보이지 않았다. 즉, 자기상관계수와 편자기상관계수가 각각의 시차에 대하여 독립적인 상관계수를 갖고 있음을 알 수 있다[그림 11].

따라서 결론적으로 추정된 잔차는 백색잡음(white noise)의 성질을 잘 만족하였기에 설정된 모형이 잘 적합되었다고 말 할 수 있겠다[그림 10].

<표 3>에서 보면 설정된 윈터스 예측모형의 적합성 검증통계량에서 예측력이 95.4%(R-square)를 보이고 있다.



[그림 11] 설정된 승법 윈터스 모형의 예측 잔차 분포

<표 3> 설정된 윈터스 예측모형의 적합성 검증통계량

통계량 종류	통계량
Mean Square Error	357240.4
Root Mean Square Error	597.69
R-Square	0.954
Adjust R-Square	0.953
Akaike Information Criterion	2090.1
Schwarz Bayesian Information Criterion	2099.4

5) 설정된 윈터스 모형을 이용한 미래시점의 산업재해 월별 예측치

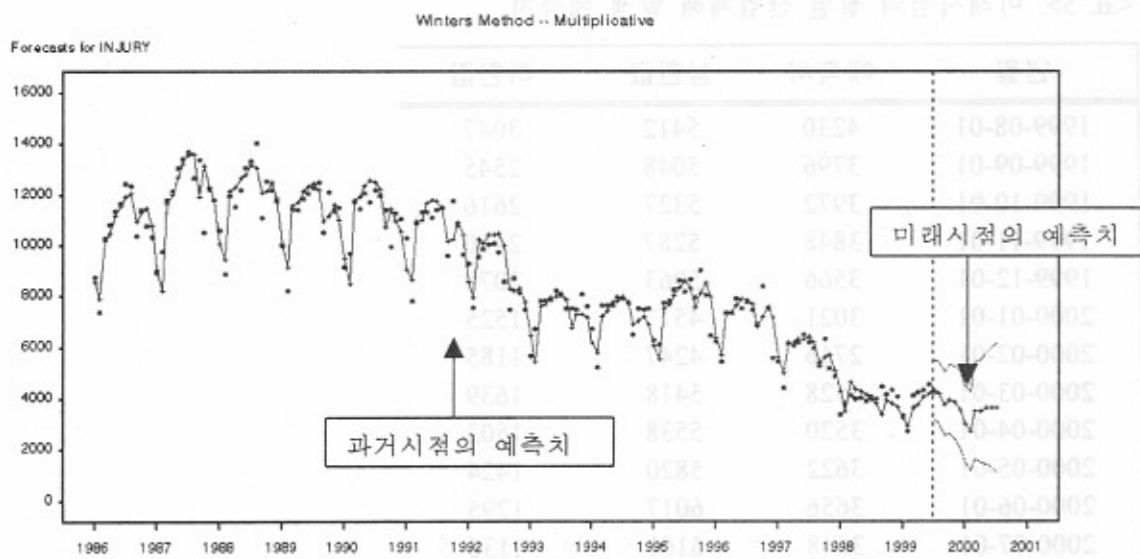
미래시점의 월별 산업재해 발생 예측치는 <표 5>와 같다. 예측치 보면 99년 12월에서 2000년 1월, 2월에 급격히 감소추세에서 2000년도 2/4분기에 다시 서서히 증가하는 추세를 보이고 있다.

<표 5>. 미래시점의 월별 산업재해 발생 예측치

년월	예측치	상한값	하한값
1999-08-01	4230	5412	3047
1999-09-01	3796	5048	2545
1999-10-01	3972	5327	2616
1999-11-01	3848	5287	2410
1999-12-01	3566	5063	2070
2000-01-01	3021	4517	1525
2000-02-01	2716	4247	1185
2000-03-01	3528	5418	1639
2000-04-01	3520	5538	1503
2000-05-01	3622	5820	1424
2000-06-01	3656	6017	1295
2000-07-01	3648	6161	1136

그리고 과거시점과 미래시점의 월별 산업재해 발생 실측치와 예측치 시도표는 [그림 12]와 같다. 전반적으로 보면 점진적으로 산업재해자수는 감소추세를 보이고 있으며, 미래시점에서 예측된 재해자수를 보면 2000년 1/4분에 급격히 감소하다가 다시 2/4분 시점을 정점으로 다시 재해자수가 증가추세를 보이고 있다. 또한 1998년에는 전반적인 발생추이 경향이 무너졌는데, 이는 한국 경제의 크나 큰 사건인 1997년 10월에 발표한 IMF에 의한 산업전반의 침체가 개입된 것으로 판단되며 1999년에는 경기침체에서 벗어나 경제가 활성화 국면이 된다면 10월, 11월에는 이전 보다 다소 재해자가 증가하리라 예상된다. 따라서 시간이 지남에 따라 새로운 재해 발생자수를 추가로 입력하여 계속적으로 예측모형을 향상시킬 필요가 있다.

또한 2000년에는 전반적으로 경제상황이 호전되면서 생산가동률과 노동력 증가로 예측된 재해자수보다 더 증가될 것으로 사료된다.



[그림 12] 월별 산업재해 실측치(점선)와 미래시점의 예측치(실선) 시도표

III. 결 론

본 연구에서는 월별 산업재해 발생자를 이용하여 미래시점의 시계열 예측을 실시한 결과 산업재해 발생자가 시간이 흐름에 따라 전체적으로 감소추세를 보이고 있음을 알 수 있다.

전체 산업재해 발생추이 패턴을 보면 순환주기를 1년 단위로 끊어 봤을 때 2월, 9월이 가장 낮게 발생하고 6,7월 10,11월이 높게 발생하고 있음을 알 수 있다. 그리고 산업재해 발생추이 패턴이 유사한 형태로 띄고 있는데, 특히 1992년, 1997년에 산업재해 발생추이의 유사한 경향이 무너졌는데, 이는 두 번의 큰 사건(event)이 있어 왔는데 대표적인 것으로 1992년 5월에 세계 경제 대공황으로 1997년 10월에 발표된 IMF로 경제 침체와 경기불황에 의한 산업전반에 영향을 미친 것으로 판단된다.

또한 전반적인 발생추이 경향이 무너졌는데, 이는 한국 경제의 크나 큰 사건인 1997년 10월에 발표한 IMF에 의한 산업전반의 침체에 의한 것으로 판단된다. 1999년에는 IMF와 같은 극단적인 영향권에서 벗어나면, 생산가동률과 노동력 증가가 예상되면서 1998년 산업재해 발생보다는 재해자가 증가하리라 예상된다.

그러나 본 연구에서 설정된 승법 윈터스 예측모형에서 갖는 몇 가지 한계점을 갖고 있는데, 더욱 의미있는 결과를 얻기 위해서는 추가적으로 보완되어야 할 것이다.

첫째, 윈터스 모형의 기본 알고리즘은 과거보다 현재 시점에 가중치를 더 부여하여 미래시점을 예측하는 것이며, 또한 수학적 이론 방법을 바탕으로 체계적인 방법이 아닌 경험적이고 직관적 방법으로, 시간의 흐름에 따라 느리게 변동할 때 효과적인 예측방법이다.

따라서 본 연구에서 모형설정과정에서부터 예측모형 정립단계까지 갖는 한계점으로는 시간의 흐름에 따른 산업재해수로 예측했기 때문에, 재해 발생에 직접적

인 요인이 반영되지 않았으며, 또한 외적으로 영향을 미치는 요인으로 경제적 상황 즉, 노동생산성 지수(부가, 상용직, 생산직), 월 평균 근로일수, 실업률, 이직률, 산업생산지수 등과 같은 노동경제 지표 용인을 간접요인 변수로 반영되지 않은 상태에서 예측했기 때문에, 다변량 시계열 예측방법으로 원인과 결과의 인과관계를 설명할 수 있는 직·간접 변수를 투입할 수 있는 전이함수 모형이나, 상태공간 분석 모형, 다변량 ARIMA 등과 같은 모형으로 예측하여 비교해야 할 필요가 있다.

그리고 산업재해 발생에 영향을 미치는 직접원인과 관리적 요인변수로 구분할 수 있는데, 먼저 직접원인 변수로 사업장내 작업환경, 안전 방호장치, 보호구의 결함, 경계표시 및 설비 결함 등과 같은 불안정한 상태, 또한 작업자의 불안정한 행동으로 안전장치 제거, 복장 및 보호구의 잘못 사용, 불안정한 작업자세, 기계·기구의 잘못 사용, 운전중인 기계손질, 위험장소 접근, 불안정한 속도조절, 불안정한 상태방치 등이며, 두 번째 관리적 원인변수로 1) 기술적 원인으로 건물 기계장치 부적합, 구조 재료의 부적합, 생산방법의 부적당, 점검·정비 보존 불량, 2) 교육적 원인으로 안전보건 지식의 부족, 안전보건 수칙의 오해, 경험훈련의 미숙, 작업방법의 교육 불충분, 유해작업 교육 불충분, 3) 작업관리상 원인으로 안전보건 관리조직 결함, 작업수칙 미 제정, 작업준비 불충분, 인원배치 부적당, 작업지시 부적당 등의 산업재해 원인변수를 코드화하는 작업이 필수적이면서 데이터베이스화 하여, 향후 후속으로 산업재해 예측모형에 원인변수로 사용되어 진행해야 할 것이다.

참고 문헌

- 고려대학교, 시계열 사례분석집1, 1994.
- 고려대학교, 시계열 사례분석집2, 1996.
- 김선경, 신규통신서비스 수요예측방법에 관한 연구, KAIST, 석사학위논문, 1995
- 김철홍, 통계적 방법과 인공신경망을 이용한 통신수요의 예측, KAIST, 1997
- 박광태, 김경희, 이동전화 서비스의 신규 가입자 수 예측을 위한 시계열모형, 경영
논총, 1997
- 염용섭, 정보통신사업 서비스 유형별 증장기 수요예측, 통신개발연구원, 1996
- 이덕기, 예측방법의 이해, 초판, 1999
- 이명호, 전기통신수요분석을 위한 계량모형구축, 통신개발연구원, 1993
- 최기현, 이종협, 시계열분석과 그 응용, 자유아카데미, 1994
- 허명희, 박유성, 시계열 자료분석, 초판, 자유아카데미, 1994
- Bowerman, B.L. and O'Connell, R.T. Forecasting and Time Series : An Applied Approach,
3rd ed, Duxbury Press, California. 1993
- Box, G.E. P and G.M Jenkins, Time Series Analysis : Forecasting and Control, 2nd ed,
Holden-Day, San Francisco. 1976
- Chatfield, C. The Analysis of Time Series : An Introduction, 2nd ed. Chapman and Hall,
London. 1980
- Cryer, J.D. Time Series Analysis. Duxbury Press, Boston. 1986.
- Hamilton, J.D, Time Series Analysis, Princeton University Press, 1994

**산업보건 데이터웨어하우스를 이용한
산업재해 추이 및 예측모델에 관한 연구**

연구자료(연구원 2000-37-187)

발 행 일 : 2000. 2.

발 행 인 : 산업안전보건연구원장 정호근

연구책임자 : 선임연구원 이관형

발 행 처 : 한국산업안전공단
산업안전보건연구원

주 소 : 인천광역시 부평구 구산동 34-6

전 화 : (032)5100-911

F A X : (032)518-0864

비매품