

대기 오염이 암에 의한 사망률에 미치는 영향에 관한 연구

韓 之 衣 * 우 치 수**

요 약

단계별 회귀 분석법을 사용하여 환경 및 대기 오염이 암에 의한 사망률에 미치는 영향을 연구하였다. 또한 SO₂와 흡연도 암 사망률에 대한 중요한 요소라는 것을 보였다.

1. 서 론

환경 오염이 암의 발생 및 사망률에 미치는 영향은 매우 크다. 또한 이러한 암은 자연 환경, 인위적인 환경 그리고 사회 환경의 종합적인 반영이라고 할 수 있다.

길림시 환경 보호국에서는 1980년부터 1985년까지 6년간, 암과 관련된 수치자료에 대한 통계를 내어 왔는데 <표1>과 <표2>에 그 자료를 제시해 두었다. 우리가 이들을 이용해 연구하려고 하는 내용은 다음과 같다.

첫째, <표1>에서 각 요소가 암 사망률에 미치는 영향의 크기

둘째, <표2>에서 어떤 종류의 오염 지수가 암에 의한 사망률에 많은 영향을 미치는가?

셋째, 이들을 이용하여 최적의 회귀 방정식을 도출하여 암 사망률을 예측하고 평가한다.

본 논문에서는 단계적 회귀 분석법을 사용하여 위에서 언급한 문제들에 대한 해석을 하고자 한다.

2. 수리 모형과 계산 과정

여기에서 사용하는 통계적 기법은 주로 참고 문헌[1-3]에서 따온 방법이다.

$$X = (X_{ij}) \quad i=1, 2, 3, \dots, n \\ j=1, 2, 3, \dots, p+1 \text{이라고 하자.}$$

* 中國 吉林化工學院

** 서울대학교 계산통계학과

단계 1. 단계별 회귀 분석법으로 표준 회귀 계수와 회귀 방정식을 구한다.

단계 1. 중상관계수를 계산한다.

$$R = \frac{\sigma_{\hat{y}y}}{\sigma_{\hat{y}}\sigma_y} \quad \dots\dots(1)$$

여기에서

$$\bar{\hat{y}} = \frac{1}{n} \sum_{i=1}^n \hat{y}_i$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$\sigma_{\hat{y}}^2 = \frac{1}{n} \sum_{i=1}^n \hat{y}_i^2 - (\bar{\hat{y}})^2$$

$$\sigma_y^2 = \frac{1}{n} \sum_{i=1}^n y_i^2 - (\bar{y})^2$$

$$\sigma_{\hat{y}y} = \frac{1}{n} \sum_{i=1}^n \hat{y}_i y_i - \bar{\hat{y}} \bar{y} \quad \text{로 정의한다.}$$

단계 3. 상관 행렬을 계산한다.

$$S_{uv} = \sum_{i=1}^n (X_{iu} - \bar{X}_u)(X_{iv} - \bar{X}_v)$$

$$u, v = 1, 2, 3, \dots, p+1$$

$$\bar{x}_u = \frac{1}{n} \sum_{i=1}^n X_{iu} \quad \text{에서}$$

$$r_{uv} = \frac{S_{uv}}{\sqrt{S_{uu}} \sqrt{S_{vv}}} \quad \dots\dots(2)$$

$$u, v = 1, 2, 3, \dots, p+1$$

을 계산하고 $p+1$ 개 변수간의 상관 행렬은

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p+1} \\ r_{21} & r_{22} & \cdots & r_{2p+1} \\ - & - & \cdots & - \\ - & - & \cdots & - \\ - & - & \cdots & - \\ r_{p+11} & r_{p+12} & \cdots & r_{p+1p+1} \end{bmatrix} \quad \dots\dots(3)$$

이 되고, r^{uv} 를 아래와 같이 정의를 하자.

$$R^{-1} = (r^{uv}) \quad u, v = 1, 2, 3, \dots, p+1 \quad \dots\dots(4)$$

그러면 r^{uv} 은 R^{-1} 행렬의 uv 번째 원소가 된다.

단계 4. 부분 상관계수를 아래와 같이 계산한다.

$$\rho_{y, u, 1, 2, \dots, u-1, u+1, u+2, \dots, p} = \frac{-r^{u, p+1}}{\sqrt{r^{uu}} \sqrt{r^{p+1, p+1}}} \dots\dots(5)$$

$$u = 1, 2, 3, \dots, p+1$$

<표1> 암과 그에 영향을 주는 요소의 값

요소 년도	평균음식 섭취량	대기오염 종합지수	담배판매 량(만갑)	평균공업 투자(원)	공업청의 오염지수	환경소음 오염지수	암중 사망률
1980	2352.2	3.01	15.377	491	0.77	1.18	4.90
1981	2341.7	4.12	20.377	515	0.98	1.20	5.85
1982	2283.0	4.79	20.028	572	0.85	1.20	5.13
1983	2349.0	2.82	22.824	643	0.95	1.21	5.12
1984	2408.0	2.02	27.190	814	0.98	1.27	5.04
1985	2525.0	2.32	34.891	887	1.06	1.23	5.01

<표2> 암과 그에 영향을 주는 오염지수의 값

요소 년도	SO ₂ 오염지수	NoX 오염지수	황진 오염지수	담배판매 량(만갑)	COD 오염지수	TSP 오염지수	암중 사망률
1980	0.49	0.62	5.35	15.377	2.31	5.57	4.90
1981	0.66	0.62	5.94	20.377	1.90	9.27	5.85
1982	0.82	0.62	3.30	20.028	2.83	6.40	5.13
1983	0.27	0.67	5.86	22.824	2.04	4.47	5.12
1984	0.28	0.55	3.80	27.190	2.25	3.43	5.04
1985	0.19	0.64	4.23	34.891	2.02	4.20	5.01

3. 대기 정합 요소가 암사망률에 미치는 영향

<표1>의 수치에 대해서, 사망률을 $y_i, i=1, 2, \dots, 6$ 로 나타내고 종속변수로 삼아보자. 또한, 한사람의 평균 음식 섭취량, 대기오염 종합 지수, 담배 판매량(만갑), 연평균 공업 투자액, 공업청의 오염지수, 그리고 환경 소음 오염 지수 등을 독립 변수 $X_i, i=1, 2, \dots, 6$ 로 두자.

각 단계의 회귀 방법을 사용하여 회귀 방정식을 구하면

$$\hat{y} = -596 + 0.62X_2 + 7.72X_6 \text{ 와 같다.} \dots\dots(6)$$

또한 증상관계수 $R=0.98$, 표준편차의 추정값 $\hat{\sigma} = 0.09$ 이며

유의 수준 $\alpha=0.05$ 에서 F 검정을 하면 유의차가 있다. 또한, 표준 회귀 계수는 $\beta_2^*=1.308$, $\beta_6^*=0.712$ 와 같다.

이 수치에 의하면 변수 X_2 (대기오염 지수)와 X_6 (환경 소음 오염지수)가 암사망률에 가장 큰 영향을 미치는 것을 알 수 있고, 이것은 길림시의 실제 현상과 유사하다.[2]

4. 대기 환경오염이 암사망률에 미치는 영향

대기 환경 오염이 암사망률에 미치는 영향을 연구하기 위해 <표2>에서 사망률을 종속 변수로 두고, SO_2 오염 지수, TSP오염 지수, 담배 판매량, COD오염 지수를 독립 변수로 두어 단계별 회귀 분석법을 사용하였고,

$$\hat{y} = 13.05 + 7.97X_1 - 0.59X_3 + 0.03X_5 - 3.97X_5 \quad \dots(7)$$

과 같은 회귀 방정식을 구하였다. 여기에서 중상관계수는 $R=0.99$ 이고, $\hat{\sigma}=0.056$ 이다. 그리고, 유의 수준 $\alpha=0.05$ 에서 F 검정을 하면 유의 차가 있다.

각 변수 X 들이 암사망률에 미치는 영향의 크기를 명확히 구하기 위해 부분 상관 계수를 구하면 다음의 표와 같다.

<표3> 암사망률에 대한 요소별 부분상관계수와 표준회귀계수

오염요인	부분 상관 계수	표준 회귀 계수
SO_2	0.987	$\beta_1^*= 5.83$
TSP	-0.979	$\beta_2^*= -3.63$
담배 함소량	0.985	$\beta_3^*= 0.538$
COD	-0.988	$\beta_4^*= -3.8$

이로부터 COD와 SO_2 가 사망률에 미치는 영향이 크며, TSP와 담배 판매량도 사망률에 상당한 영향을 미치는 것을 알 수 있다.

5. 결 론

첫째, 암 사망률에 가장 큰 영향을 미치는 요인은 대기 오염이고, 그 다음은 소음 오염이다. 대기 오염 문제는 이미 각계의 주시를 받고 있으나 소음오염 문제는 아직도 전 사회의 주시를 받고 있지는 못하다.

둘째, 길림시의 환경 오염원은 주로 석탄에 의해 발생하고, 길림시의 석탄 함소량은 비교적 낮다. 그러므로 대기중의 SO_2 의 함량은 높지 않다. 그러나, 연구 결과는 SO_2 가 암 사망률에 가장 크게 영향을 미친다는 것을 보이고 있다.

셋째, 그러므로 대기 중의 SO_2 함량을 줄이도록 노력해야 할 것이며, 또한 담배 판매량도 암 사망률에 많은 영향을 미치고 있으므로 흡연을 제어함으로써 환경 오염을 다스리고 국민의 건강을 유지하도록 노력해야 하겠다.

〈참 고 문 헌〉

- (1) 포박송 편저: 회귀 분석 및 시험 설계 <1983>
- (2) 장문광 편: 길림시 정중 생태 상황적 연구 <1986>
- (3) 동문천, 주광아: 수량화 이론 및 응용 <1979>

**Evaluation of the Major Sources of
Atmospheric Pollution in Jilin City by
Regression Diagnostics**

Han Zhi Nong, Wu Chi Su

Abstract

We study by stepwise regression method, the influence which atmospheric pollution in Jilin City has on the rate of death according to cancer. The extremes are discovered and we know that SO₂ and smoking are important factors too.