

한국인의 사망원인 질병에 대한 통계적 분석*

오만숙**·신우영***

요 약

통계청에서는 매년 전년도 사망한 모든 사망자들을 대상으로 사망원인이 되는 사고 및 질병 등의 요인, 사망자의 주소지, 연령, 교육수준 등의 자료를 조사하여 발표하고 있다. 이는 사망원인이 되는 요인에 대한 분석이 한국인의 무병장수와 정부적 보건의료정책의 수립에 도움이 될 것으로 판단하기 때문이다. 그러나 그동안 사망통계자료에 대한 다변적인 분석이 미흡하여 귀중한 자료를 적극적으로 활용하지 못하였다.

본 논문에서는 2008년 사망원인 통계자료를 대상으로, 한국인의 사망원인 중 1순위인 '신생물, 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정장애'와 앞으로 한국인에게 확산될 것으로 우려되는 '내분비, 영양 및 대사 질환'의 두 질병에 대하여 적합한 모형을 찾고, 유의한 영향을 미치는 설명변수들을 판별한 뒤 구체적인 영향의 분석을 수행한다.

분석결과 '신생물, 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정장애'는 대도시 보다는 지방주민에게, 1차 산업 종사자나 전문가 보다는 도시 노동자에게, 그리고 교육 수준이 낮은 사람에게 더 많이 사망원인이 됨을 보여 주었다. '내분비, 영양 및 대사 질환'의 경우는 여자 보다는 남자에게, 평균적으로 스트레스를 많이 직업군에서 더 많이 사망원인이 됨을 보여 주었다. 이 질병은 특히 직업의 영향을 매우 크게 받는 것으로 나타났는데, 기술공의 경우 학생·가사·무직에 비하여 오즈가 164% 증가함을 보여 주었다.

핵심주제어 : 사망통계, 로지스틱 회귀모형, 합성효과

I. 서 론

한국인의 수명이 과거에 비하여 현저히 길어지고 생활수준이 높아지면서 건강에 대한 관

* 논문접수일 2010년 11월 16일, 게재확정일 12월 21일

본 연구는 학술진흥재단과 한국산업경제저널에서 정한 연구윤리규정을 준수함

** 이화여자대학교 통계학과 교수, 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행한 기초연구사업임 (No. 2010-0010285). 교신저자.

*** 이화여자대학교 통계학과 석사과정.

심도 증가하고 있다. 건강하게 오래 사는 것에 대한 사람들의 바람과 관심으로 건강과 관련된 통계자료들이 많이 수집되고 있으며 이들 통계자료로부터 건강장수에 영향을 미치는 요인들을 알고자 하는 욕구가 증대되고 있다.

사망원인 통계자료는 매년 통계청에서 발표하는 자료이다. 사망원인 통계는 매년 9월에 작년 1월~12월까지 사망자의 사망원인 통계를 집계하여 홈페이지의 KOSIS(통계정보시스템)를 통해 국내외에 제공한다¹⁾. 사망원인으로 질병이 가장 많은데, 사망원인이 되는 질병을 보다 구체적으로 분류하기 위하여 한국표준질병사인분류(KCD, Korean Classification of Diseases)²⁾에 의거 분류 집계한다. 또한 사망자주소, 사망자생년월일, 사망년월일, 사망장소, 사망당시직업, 사망원인진단자, 혼인상태, 교육정도 등 인구통계적 변수들도 자료에 포함되어 있다. 사망원인 통계자료는 그 해 사망한 사망자들 모두를 대상으로 하고 있어 분실값이 거의 없는 자료로서 특정 사망원인에 영향을 미치는 요인들을 파악하기에 매우 중요한 자료이다. 한국인을 사망에 이르게 하는 주요 원인은 무엇이며 주요 사망원인에 영향을 미치는 요인들을 파악한다면, 우리가 바라는 수명연장과 건강한 삶에 도움이 될 수 있기 때문이다.

한국인의 사망원인에는 여러 가지가 있지만, 운수사고와 자살 등을 제외하고는 모두 질병과 관련된 것들로, 사망원인 1위는 암이고 그 뒤를 뇌혈관 질환과 심장 질환이 잇고 있다 <그림 1>³⁾. 이들 질병에는 사망자가 타고난 유전적 요인도 있겠지만 이외에도 사망자가 기본적으로 가지고 있던 환경적 요인도 상당한 영향을 미칠 것으로 추정된다. 왜냐하면, 사람들이 처해 있는 시대 및 사회 경제적 구조에 따라 사망의 양상이 달라질 수 있기 때문이다. 따라서 시대적, 사회적, 경제적 구조를 반영하는 환경적 요인들이 한국인의 사망에 어떤 영향을 미치는지 파악한다면, 건강장수를 위한 개인적 노력의 방향 설정에 도움이 될 것이며 또한 정부적 차원에서 국가복지 및 보건의료 정책수립을 위한 정보로 매우 유용하게 사용될 수 있을 것이다.

본 연구에서는 한국인의 사망원인 질병 중 큰 비율을 차지하며 또한 많은 국민들의 관심이 되는 두 가지 질병인 ‘신생물, 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정장애’와 ‘내분비, 영양 및 대사 질환’에 대하여 분석하고자 한다. 각 질병에 대하여 로지스틱 회귀 모형을 가정하고, 설명변수로는 사망원인 통계자료에서 제시하는 사망자의 주소지, 성

1) 사망원인통계연보(2008):전국편 = Annual report on the cause of death statistics, 통계청.

2) 한국질병분류사인코드(제 5차 개정,2005.5.10), www.kostat.go.kr, 통계청.

<http://kostat.go.kr/kssc/stclass/StClassAction.do?method=dis&classKind=5&kssc=popup>.

3) 사망원인통계(2008), 통계청.

http://blog.naver.com/hypomania9?Redirect=Log&logNo=20100064610&topReferer=http://cafeblog.search.naver.com&imgsrc=20100211_146/hypomania9_1265847287021qovn7_jpg/%C6%F7%B8%CB%BA%AF%C8%AF_s_hypomania9.jpg

별, 직업, 혼인상태, 교육정도, 사망시 연령을 선정하였다.

본 논문의 제 II장에서는 각 설명변수의 분포에 대하여 간단한 그림을 통한 기술적 추론을 실시하였다. 이 추론은 전체 사망자를 대상으로 한 것으로 사망자들의 각 변수별 분포를 이해하고자 하는 목적이다. 제 III장에서는 각 종속변수에 대하여 로지스틱 회귀분석을 수행하였다. 먼저 자료에 적합한 모형을 찾는 적합도 검정을 실시하고, 종속변수에 유의한 영향을 미치는 변수들을 찾아낸 다음, 최종 모형을 적합시킨 결과를 제시하였다. 설명변수들이 대부분 범주형 변수들이기 때문에 각 변수의 범주들 마다 기준 범주에 대한 상대적 크기로 계수가 주어지는데, 이 계수들의 유의성에 대하여 살펴보고, 비슷한 범주들끼리 묶은 다음, 범주들의 영향을 비교 해석하는 분석을 수행하였다. 마지막으로 제 IV장에서는 분석결과를 요약하여 결론을 제시하였다.



<그림 1> 주요 사망원인별 사망률 변화

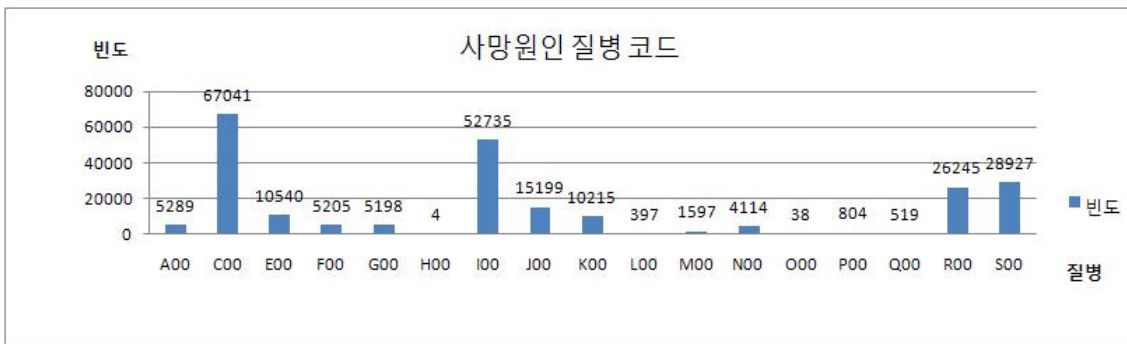
II. 자료의 구성과 기초분석

본 연구의 분석에 사용된 데이터는 통계청의 '2008년 사망원인 조사 자료'이다. 데이터의 구성은 변수가 총 10개이고 데이터의 개수는 총 24,775개(외국인 포함)이다. 변수들 중 관심 종속 변수인 사망원인 변수의 분류는 '한국표준질병사인분류(제5차 개정)'를 참고한 것이다. 사망한 년도의 자료이므로 모두 2008년에 사망한 경우의 데이터이지만 신고한 년도는 2008년과 2009년으로 나뉘어져 있다. 신고한 년도의 데이터를 모두 합치면 24,775개가 된다. <표 1>은 이 데이터에 포함된 10개 변수의 세부 범주별 빈도를 정리해 놓은 것이다.

<표 1> 2008 사망원인 조사 자료

변수	세부항목
신고년도	2008년, 2009년
주소(시도)	11-서울특별시, 21-부산광역시, 22-대구광역시, 23-인천광역시, 24-광주광역시, 25-대전광역시, 26-울산광역시, 31-경기도, 32-강원도, 33-충청북도, 34-충청남도, 35-전라북도, 36-전라남도, 37-경상북도, 38-경상남도, 39-제주도, 기타
성별	1-남자, 2-여자
사망자직업	1-관리자, 2-전문가 및 관련 종사자, 3-기술공 및 준 전문가, 5-서비스 종사자 및 판매 종사자, 6-농림어업 숙련 종사자, 7-기능원 및 관련 기능 종사자, 8-장치, 기계조작 및 조립 종사자, 9-단순노무 종사자, 13-학생, 가사, 무직, 99-미상, 군인 (사병제외)
혼인상태	1-미혼, 2-배우자 있음, 3-이혼, 4-사별, 9-미상
교육정도	1-무학, 2-초등학교, 3-중학교, 4-고등학교, 6-대학(교), 7-대학원 이상, 9-미상
사망원인 질병1	A00-B99 : 특정 감염성 및 기생충성 질환 C00-D48 : 신생물 D50-D89 : 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정장애 E00-E90 : 내분비, 영양 및 대사 질환 F00-F99 : 정신 및 행동 장애 G00-G99 : 신경계통의 질환 H00-H59 : 눈 및 눈 부속기의 질환 H60-H95 : 귀 및 꼭지돌기의 질환 I00-I99 : 순환기계통의 질환 J00-J99 : 호흡기계통의 질환 K00-K93 : 소화기계통의 질환 L00-L99 : 피부 및 피부 밑 조직의 질환 M00-M99 : 근육골격계통 및 결합조직의 질환 N00-N99 : 비뇨생식계통의 질환 O00-O99 : 임신, 출산 및 산후기 P00-P96 : 출생전후기에 기원한 특정 병태 Q00-Q99 : 선천기형, 변형 및 염색체 이상 R00-R99 : 달리 분류되지 않은 증상, 징후와 임상 및 검사의 이상 소견 S00-T98 : 손상, 중독 및 외인에 의한 특정 기타 결과 V01-Y98 : 질병이환 및 사망의 외인 Z00-Z99 : 건강상태 및 보건서비스 접촉에 영향을 주는 요인 U00-U99 : 특수목적 코드 U20-U99 : 한의분류
사망원인질병2	사망원인 1과 같음
사망시연령	0~111, 999(미상, 모름으로 응답)

<표 1>에서 사망원인질병은 '한국 표준 질병 사인 분류'를 참고로 분류한 것인데, 위의 분류 코드를 간편하게 앞의 세 자리 알파벳 기호와 숫자로 표기하기로 한다. 예를 들면, A00-B99 대신 A00를 사용하고, C00-D48 대신 C00를 사용하여 표기하는 것이다. 그리고 편의를 위해 공통항목을 나타내는 질병끼리 묶는데, 위의 질병들 중 '신생물 C00-D48' 과 '혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정장애 D50-D89', 이 두 질병을 하나로 묶어 '신생물 및 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정장애' 라 부르고 이를 C00로 표시한다. 또한 H00-H59, H60-H95를 하나로 묶어 '눈과 귀 및 부속기관들의 질환'으로 부르고 이를 H00로 표시한다. 이렇게 정리된 각 사망원인 질병의 빈도수를 살펴보면 <그림 2>와 같다.



<그림 2> 사망원인 질병의 dots 분포

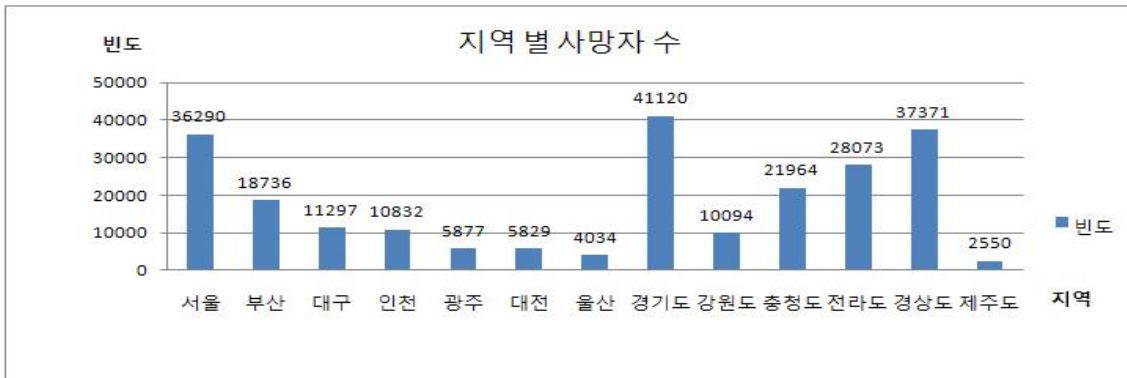
사망원인질병의 빈도수를 살펴보면 '신생물, 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정장애'를 나타내는 C00의 비율이 가장 높게 나타나고, '순환기계통의 질환'을 나타내는 I00의 비율도 비교적 높게 나타난다. 또한 '달리 분류되지 않은 증상, 징후와 임상 및 검사의 이상 소견 (R00)' 와 '손상, 중독 및 외인에 의한 특정 기타 결과(S00)' 의 빈도도 높다.

총 17개의 질병에 대해 모두 분석하는 것은 지면 관계상 불가능하고, 본 연구에서 우리는 가장 빈도가 높은 질병인 C00와 현재 빈도는 높지 않지만 미래 한국인에게 크게 증가할 것으로 예상되는 '내분비, 영양 및 대사 질환 (E00)'를 분석 대상으로 하겠다. 다른 질병들에 대해서도 같은 방법을 사용하면 되겠다.

위의 두 질병들 각각에 대하여 설명변수로 사용될 인구통계적 변수들인 주소지, 연령 등의 상관관계를 살펴본 결과 모든 변수에 대하여 유의한 상관관계를 보여주었다. 따라서 앞

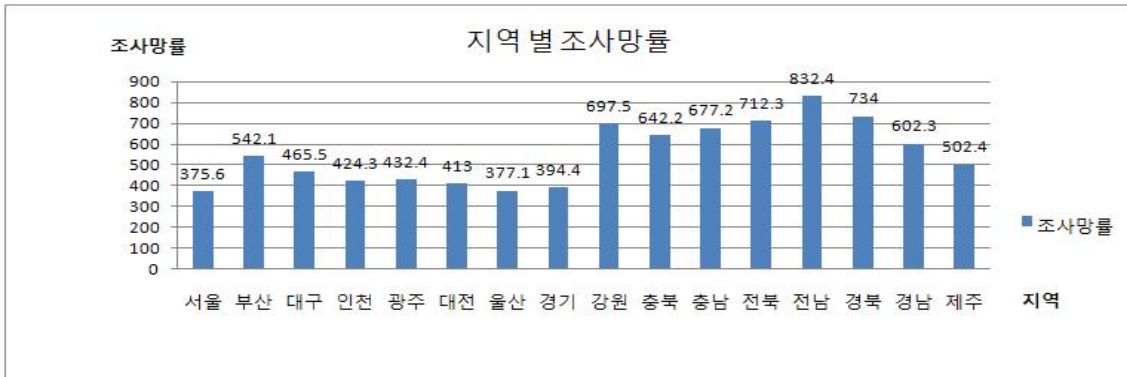
서 언급한 변수들을 회귀 모형에서 설명변수로 사용하는 것이 문제가 없다고 판단된다.

종속변수에 대한 본격적인 분석에 앞서, 자료에 포함된 설명변수들의 범주들에 대하여 사망자들이 어떻게 분포되어 있는지 알아보기로 한다. 여기에서는 설명변수와 사망과의 대략적인 관계를 알아보기 위한 탐색적 단계로서 위의 두 질병 뿐 아니라 다른 원인으로 사망한 사람들도 포함하였다.

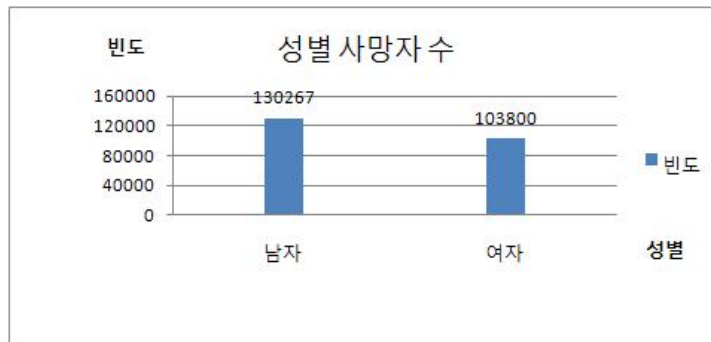


<그림 3> 사망자의 주소지별 dots 분포

먼저 <그림 3>에서 주소지별 사망자의 분포를 살펴보자. 예상대로 '서울', '경기도', '경상도'의 빈도수가 가장 높게 분포되어 있고 '제주도'의 빈도수가 가장 낮다. 참고로, '충청도', '전라도', '경상도'의 경우 남도와 북도를 합한 것이다. '서울', '경기도', '경상도'의 사망자 빈도수가 높게 나타나는 것은 이들 지역에 인구가 밀집되어 있기 때문일 것이므로, 절대 빈도수가 아닌 지역별 사망비율을 살펴볼 필요가 있다. <그림 4>은 지역별 조사망률의 분포를 나타낸 것이다. 조사망률이란 1년간의 사망자수를 그 해의 인구로 나눈 것이다. 그림을 보면 서울, 부산 등 대도시보다 강원, 충북 등 지방의 사망률이 높은 것을 알 수 있다. 대도시보다 지방의 사망률이 높은 것은 노년층이 주로 지방에 거주하기 때문이거나, 대도시 거주민들이 보다 더 건강한 생활을 하기 때문이거나, 혹은 이밖에 다른 원인들이 합성된 결과일 수도 있다. 따라서 다른 변수들을 무시하고 <그림 4>와 같은 지역과 사망률만의 주변적 관계만을 고려하는 것은 영향분석에 한계를 지닌다. 제 III장에서 우리는 다른 요인들을 제어한 상태에서 지역이 질병사망에 미치는 영향을 분석할 것이다.

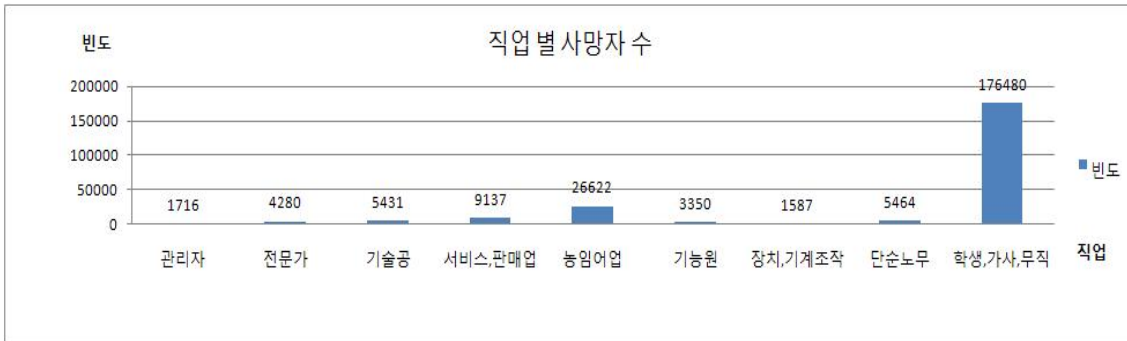


<그림 4> 사망자의 주소지별 조사망률 분포



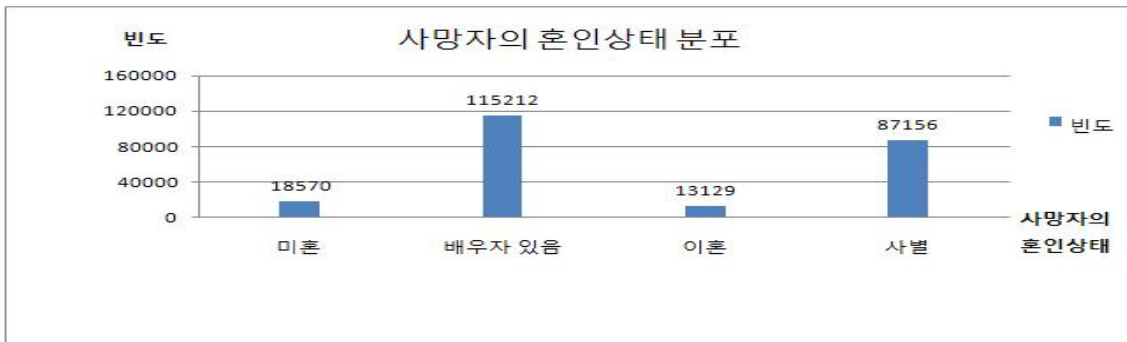
<그림 5> 사망자의 성별 분포

성별의 분포 <그림 5>를 살펴보면, '남자'의 수가 3만 명 정도 많아 2008년도 사망자 중 '남자'와 '여자'의 비율이 약 1.27:1 정도로 '남자'가 '여자'보다 약 27% 더 많음을 알 수 있다. 한국인의 남자와 여자의 성비 보다 사망자의 남자 비율이 높은 것은 남자의 수명이 짧다는 것을 의미한다.



<그림 6> 사망자의 직업 분포

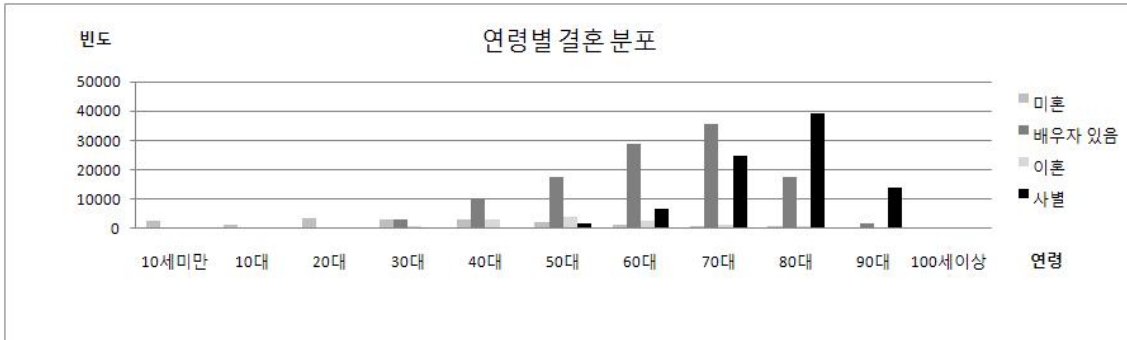
사망자의 직업 분포 <그림 6>를 살펴보면, ‘학생·가사·무직’의 빈도가 월등히 높게 나타난다. 사망 시 연령분포에서 고령자의 비율이 높은 점을 감안할 때, 사망 시 직업이 ‘무직’인 경우가 가장 높기 때문으로 해석된다. 직업이 사망원인 질병에 미치는 영향을 파악하기 위해서는 사망당시의 직업 보다는 생애 전체 중에서 가장 많은 기간 동안 가졌던 직업, 또는 성인이 된 이후 노년기를 맞이하기 이전까지의 주요 직업을 조사하는 것이 더 바람직할 것으로 여겨진다. 또한 농림어업 종사자가 많은데 이는 고령자들이 전문직보다는 농업, 어업 등 1차 산업 종사자가 많기 때문일 것으로 추측된다.



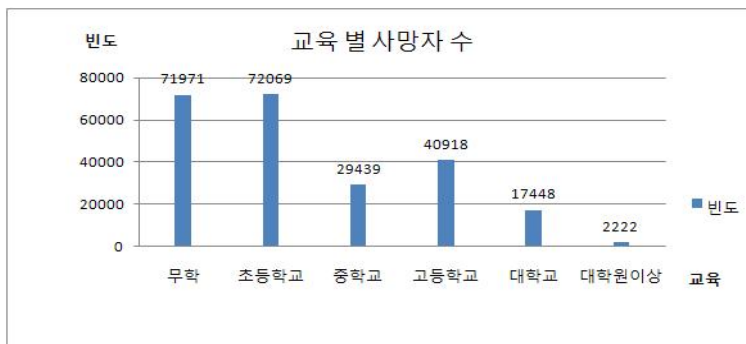
<그림 7> 사망자의 혼인상태 분포

사망자의 혼인상태 분포 <그림 7>을 살펴보면, 고령자의 사망이 많은데 사망 당시 혼인상태를 조사하였기 때문에 ‘배우자 있음’과 ‘사별’의 경우가 많은 빈도수를 차지하고 있다. 연령과 혼인상태는 매우 밀접한 관련을 가지고 있을 것으로 판단되기 때문에, 구체적으로 혼인상태와 연령의 관련성을 알아보기 위하여 연령별로 혼인상태의 세부항목을 조사하였다

<그림 8>. 이 분포를 살펴보면, 40대~70대까지는 '배우자 있음'의 분포가 가장 높고, 80세 이상은 '사별'의 분포가 가장 높다. 사망 시 연령을 살펴보면, 약 76%가 60세 이상으로 고령자의 비율이 높기 때문에 사망 시 결혼 여부에 '배우자 있음'과 '사별'의 빈도가 높게 나타난다.

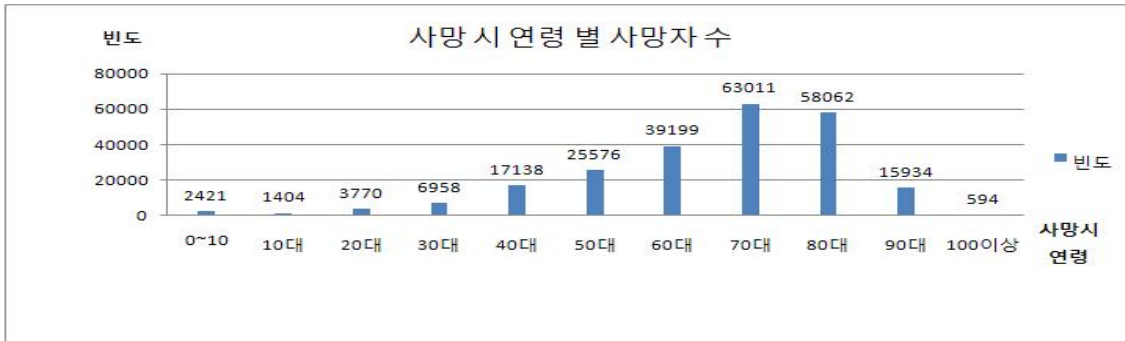


<그림 8> 연령별 혼인상태 분포



<그림 9> 사망자의 교육 분포

<그림 9>는 사망 당시 조사대상의 학력을 조사한 히스토그램이다. 사망자 중 고령자가 많기 때문에 저학력자 (무학, 초등학교 졸업)의 빈도수가 많고 특히 무학의 빈도수가 높게 나타났다 추측해 볼 수 있다.



<그림 10> 사망 시 연령 분포

<그림 10>은 사망 시 연령 분포를 살펴본 것이다. 70대와 80대에 사망한 연령이 예상대로 가장 많은 빈도수를 차지하였고, 100세 이상으로 사망한 경우도 거의 600명에 가깝게 나타난다. 10세 미만에 사망하는 빈도수도 약 1%로 10대 사망률의 2배정도 높은 수치를 나타내고 있는데 이는 영유아 사망이 많기 때문으로 추측된다.

이상의 결과를 보면, 사망자의 분포가 각 설명변수의 범주들에 대하여 균일하지 않은 양상을 보인다. 이는 각 설명변수의 범주와 사망간의 상관성을 보여주는 것이다. 그러나 앞서 언급한 바와 같이 위의 탐색적 추론은 설명변수와 사망과의 주변적 관계를 보는 것이기 때문에 다른 요인들과 합성될 가능성이 있어 최종적인 결론을 짓기에는 무리가 있다. 따라서 자료를 잘 설명하기에 충분한 설명변수들을 동시에 고려하는 다변량 분석이 요구된다. 다변량 분석으로부터 우리는 다른 변수들이 동일한 수준을 지닐 때 주어진 설명변수의 범주들이 어떠한 영향을 미치는지 파악할 수 있어 합성효과를 제어할 수 있다.

III. 로지스틱 회귀모형을 이용한 통계분석

조사된 17가지 사망원인 질병 중 우리는 두 개의 사망원인 질병들의 각각에 대하여 로지스틱 회귀모형을 적합해보고자 한다. 만약 다른 질병에 관심이 있다면 동일한 방법으로 분석이 가능할 것이다. 선택된 사망원인 질병들은 ‘신생물, 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정장애’, ‘내분비, 영양 및 대사 질환’이다. 각 질병으로 인한 사망유무를 나타내는 이항자료에 대하여 로지스틱 회귀모형을 적합시키고 적합도에 대한 검정을 수행한 후, 각 질병에 영향을 미치는 유의한 설명변수들에 대하여 자세한 분석을 수행하고자 한다.

3.1. 신생물, 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정 장애

신생물이란 세포조직이 자율적으로 비정상적이고도 지나치게 증식하는 것을 말하며, 신생물이 생체에 미치는 악영향의 정도에 따라 악성과 양성으로 구별한다. 상피세포로 이루어지는 악성신생물을 암종이라고 하며, 결합조직 및 혈액의 종양(백혈병, 림프종 등)을 육종이라고 분류하고 있다. 악성종양이라고도 한다. 또한 ‘혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정 장애’는 빈혈이나 응고결핍, 자색 반증 및 기타 출혈성 병태를 말한다.

C00(신생물, 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정 장애) 질병원인으로 사망하는 경우를 1, 그렇지 않으면 0으로 두는 이항변수를 종속변수로 하고 나머지 변수들, 주소지, 성별, 직업, 결혼여부, 교육정도, 나이를 설명변수로 사용하여 로지스틱 회귀모형을 적합해보았다. 참고로, 이 논문의 모든 분석은 SAS 9.2 (SAS, 2008) 를 사용하였다.

먼저 설명변수들의 교호작용 없는 주효과만을 가진 로지스틱 회귀모형을 적합시킨 결과 <표 2>에서와 같이 여러 적합 통계량에 대한 Value/Df값이 모두 작으므로 주효과만을 가진 로지스틱 회귀모형이 잘 적합됨을 알 수 있다. 다음으로, 각 설명변수에 대한 유의성 검정 결과인 <표 3> 를 살펴보면, 주소지, 성별, 직업, 결혼여부, 교육정도, 나이의 유의확률 ‘Pr > ChiSq’ 값이 모두 유의수준 0.05보다 작으므로 각 설명변수가 ‘C00(신생물, 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정 장애)’에 유의한 영향을 미치는 것을 알 수 있다.

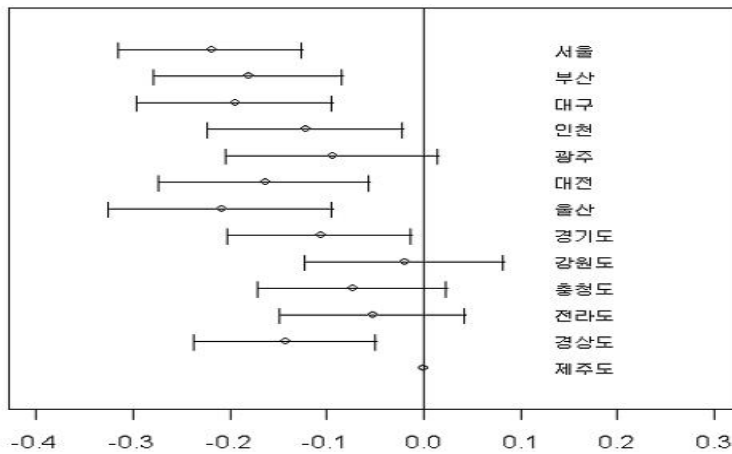
<표 2> C00 에 대한 로지스틱 회귀모형 적합도 검정

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	23E4	266323.9326	1.1380
Scaled Deviance	23E4	266323.9326	1.1380
Pearson Chi-Square	23E4	234275.7479	1.0010
Scaled Pearson X2	23E4	234275.7479	1.0010
Log Likelihood		-133161.9663	

<표 3> 각 독립변수에 대한 유의성 검정

LR Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
address	12	153.04	<.0001
sex	1	4.48	0.0343
job	8	522.95	<.0001
marry	3	6342.53	<.0001
edu	5	1586.98	<.0001
age	1	879.75	<.0001

고려한 6개의 설명변수가 모두 유의하므로, 6개의 설명변수의 주효과를 지닌 로지스틱 회귀모형을 적합시켜 보았다. 설명변수가 범주형이면 하나의 기준범주를 정하고 이 기준 범주에 대한 상대적 크기로 회귀계수가 주어진다. 변수 마다 범주의 수가 많아 회귀계수 추정치를 모두 나열하기에는 어려움이 있으므로 각 변수에 대하여 범주의 회귀계수의 추정치와 95% 신뢰구간을 그림으로 표시하기로 한다.



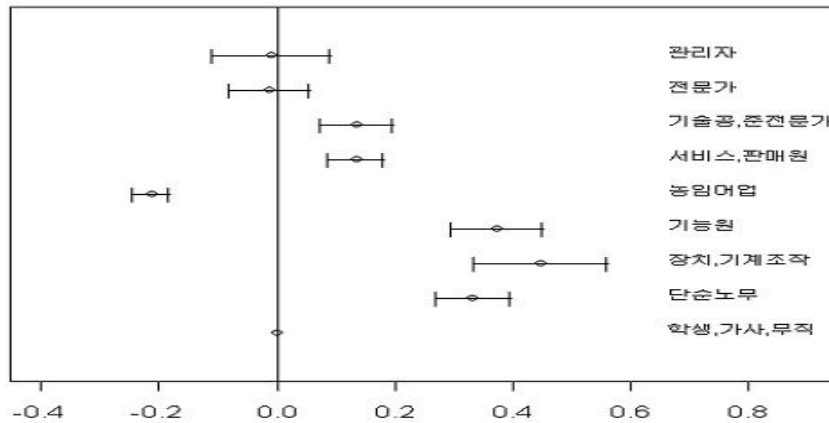
<그림 11> 주소지별 회귀계수추정치

<그림 11>은 로지스틱 회귀모형에서 조수지(지역)변수에 대해 제주도를 기준범주로 할 때 각 범주의 회귀계수의 95% 신뢰구간이다. 그림을 보면, 서울, 부산, 대구, 대전, 울산은 유의하게 제주도 보다 낮은 계수추정치 값을 갖고 나머지 지역은 대략 제주도와 비슷한 추정치를 갖는다고 할 수 있다. 서울, 부산, 대구, 대전, 울산의 계수추정치의 평균은 대략 -0.2 이다. 이를 해석하면, 다른 변수들을 일정한 수준으로 제어했을 때 서울, 부산, 대구, 대전, 울산 지역은 '신생물, 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정 장애'로 사망할 오즈가 제주도를 비롯한 다른 지역의 오즈의 82% 이다.

이상을 보면, 다른 변수들이 같을 때 대도시 거주가 '신생물, 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정 장애'로 사망할 가능성을 낮춘다고 할 수 있다. 대도시가 지닌 설명변수에 포함되지 않은 다른 요인들, 예를 들면 병원의 접근성, 생활수준, 건강검진 횟수 등이 영향을 미칠 수 있으리라 짐작된다.

성별에서 여자를 기준으로 할 때 남자의 회귀계수는 0.03 이다. 유의확률을 보면 0.034 로 0.05 보다 크기는 하나 그리 차이가 크지 않으므로, 다른 변수들이 동일할 때 남자와 여자의

‘신생물, 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정 장애’로 인한 사망 가능성은 거의 차이가 없다고 볼 수 있다. 이는 II장에서 주변적 분포를 보았을 때 사망자 중 남자가 여자보다 27% 더 많았음과 대조적인 결과이다. 즉, 전체적으로 남자의 사망비율이 높게 나타나지만 다른 변수들을 제어했을 때는 C00 로 인한 사망에서 남녀의 차이가 거의 없다는 결과를 보인다.

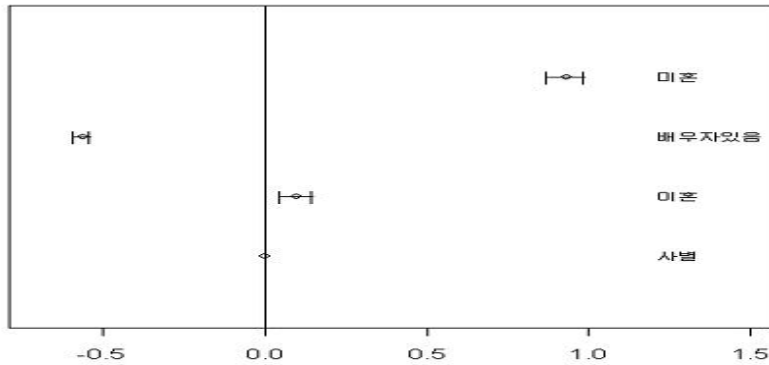


<그림 12> 직업별 회귀계수추정치

<그림 12>에서 직업을 보면, 학생·가사·무직을 기준으로 하였을 때 관리자와 전문가 범주는 유의하지 않다. 이를 제외한 직업 중 농림어업은 현저하게 작은 계수값을 가지고, 기술공·준전문가와 서비스·판매업은 비슷하게 양의 계수값을, 그리고 기능원, 단순노무, 장치기계조작은 매우 큰 양의 계수값을 가진다. 따라서 직업을 대략 네 그룹으로 나눌 수 있겠다. 농림어업 종사자는 ‘신생물, 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정 장애’로 인해 사망할 가능성이 다른 직업에 비해 낮다. 구체적으로, 다른 변수들이 동일할 때 농림어업 종사자는 ‘신생물, 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정 장애’로 인해 사망할 오즈가 학생·가사·무직, 관리자, 전문가 그룹의 오즈의 약 81%이다. 반면에 기술공·준전문가와 서비스·판매업의 오즈는 학생·가사·무직, 관리자, 전문가 그룹의 약 114.7%, 그리고 기능원, 단순노무, 장치 및 기계조작 종사자의 ‘신생물, 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정 장애’로 인한 사망 가능성은 다른 직업에 비해 매우 높다.

혼인상태 변수에 대한 <그림 13>을 보면 모든 범주가 유의하다. 다른 변수가 동일하게 제어되었을 때 미혼자가 ‘신생물, 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정 장애’

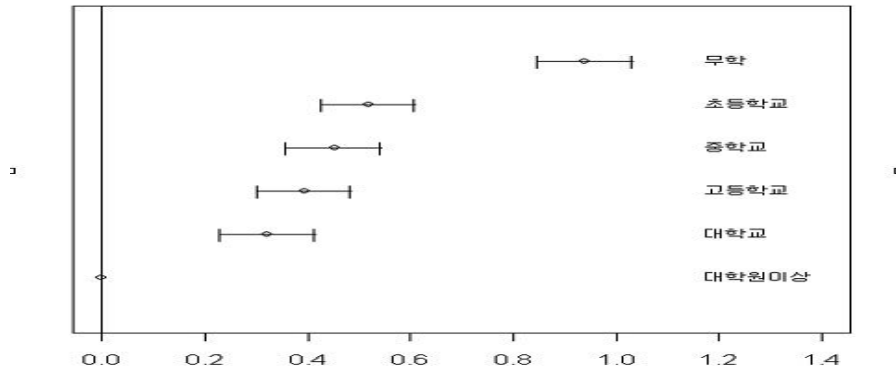
로 인해 사망할 오즈는 사별의 2.53배, 이혼의 2.29배, 배우자 있음의 4.48배 이다. 배우자 있음의 오즈는 '사별'의 오즈의 57%로 매우 낮다.



<그림 13> 혼인상태별 회귀계수추정치

혼인상태 변수에 대한 <그림 13>을 보면 모든 범주가 유의하다. 다른 변수가 동일하게 제어되었을 때 미혼자가 '신생물, 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정 장애'로 인해 사망할 오즈는 사별의 2.53배, 이혼의 2.29배, 배우자 있음의 4.48배 이다. 배우자 있음의 오즈는 '사별'의 오즈의 57%로 매우 낮다.

교육에 대한 <그림 14>를 살펴보면, 대학원을 기준으로 했을 때 초등학교, 중학교, 고등학교, 대학교가 비슷하고 무학은 현저하게 큰 양의 계수값을 갖는다. 다른 변수들이 제어되었을 때 초등학교, 중학교, 고등학교, 대학교 졸업자가 '신생물, 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정 장애'로 인해 사망할 오즈는 대학원 졸업자의 약 1.65배, 무학의 경우는 대학원 졸업자의 약 2.55배 이다. 학력이 낮을수록 위의 질병으로 인해 사망할 가능성이 높은 것으로 나타난다. 학력이 높을수록 건강에 대한 정보도 많고 또 수입이 많아 보다 더 건강관리에 신경쓰거나 치료에 적극적으로 임할 수 있기 때문으로 추측할 수 있다.



<그림 14> 교육별 회귀계수 추정치

마지막으로, 유일한 연속형 변수인 사망 당시 연령에 대한 계수 추정값은 0.01이고 표준오차는 0.0004 이다. 다른 변수들이 제어되었을 때 연령이 1년 증가하면 '신생물, 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정 장애'로 인해 사망할 오즈가 1% 증가하고 연령이 10년 증가하면 10% 증가한다. 연령에 대한 계수 추정값이 유의하긴 하지만 추정값이 작아 위 질병사망에 미치는 영향이 미미하다고 해석할 수 있다. 즉, 고려된 다른 설명변수들이 제어되었을 때 연령은 그다지 큰 영향을 미치지 않는다.

3.2. 내분비, 영양 및 대사 질환

내분비, 영양 및 대사 질환(E00)은 갑상샘, 당뇨병, 포도당 조절 장애, 영양실조, 비만 및 기타 과영양을 말하고 특히 대사 질환은 생체 내 물질대사 장애에 의해서 발생하는 질환의 총칭이며, 당뇨병은 당질대사의 이상이고, 통풍은 요산대사의 장애에 의해서 생기는 질병이다.

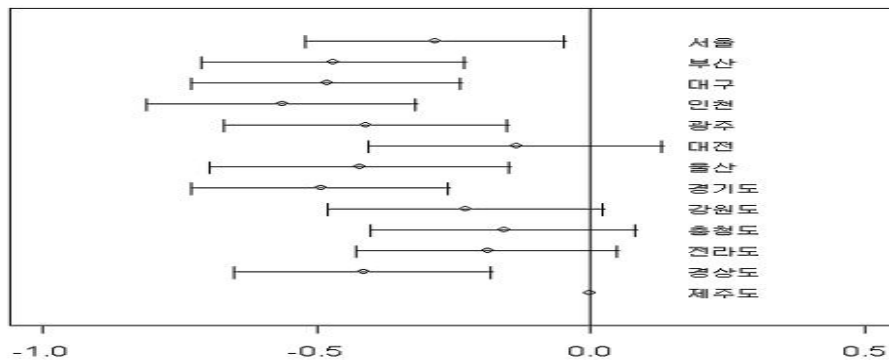
<표 4> E00 에 대한 로지스틱 회귀모형 적합도 검정

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	23E4	84855.2405	0.3626
Scaled Deviance	23E4	84855.2405	0.3626
Pearson Chi-Square	23E4	234324.3497	1.0012
Scaled Pearson X2	23E4	234324.3497	1.0012
Log Likelihood		-42427.6203	

<표 5> 각 독립변수에 대한 유의성 검정

LR Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
address	12	177.47	<.0001
sex	1	22.17	<.0001
job	8	494.89	<.0001
marry	3	85.07	<.0001
edu	5	97.35	<.0001

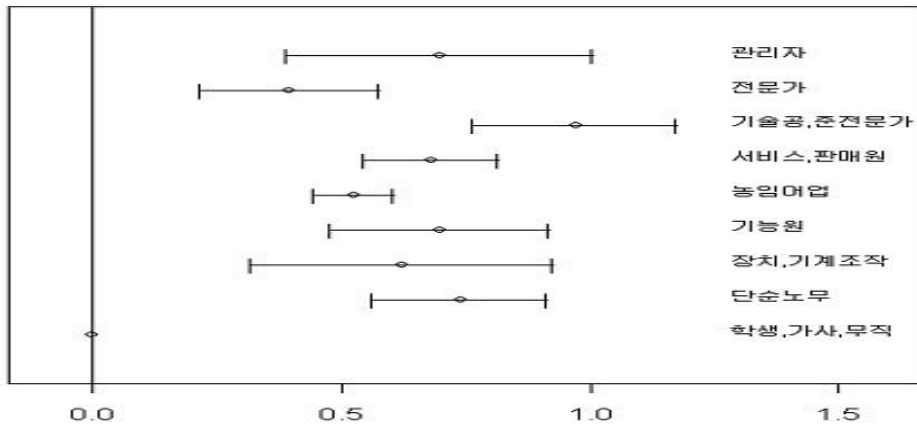
E00에 대하여 로지스틱 회귀모형을 적합해보았다. <표 4>를 살펴보면, Value/Df값이 모두 작으므로 적합도 검정을 만족한다. 또한 <표 5>에서와 같이 각 설명변수에 대한 유의성 검정 결과를 살펴보면, 주소지, 성별, 직업, 결혼여부, 교육정도의 유의확률 Pr값이 모두 유의수준 0.05보다 작으므로 각 계수를 회귀식에 적합시킬 수 있다. 그러나 연령은 유의하지 않아 모형의 설명변수에 포함시키지 않았다.



<그림 15> 주소지별 회귀계수추정치

<그림 15>에서 주소지 변수의 범주들에 대한 계수의 신뢰구간을 살펴보면, 강원도, 대전, 전라도, 충청도, 제주도와 나머지 지역으로 구분지을 수 있는데, 위의 지역과 비교하여 나머지 지역에서 E00 로 인한 사망 가능성이 낮음을 알 수 있다. 이 결과를 보면, ‘내분비, 영양 및 대사 질환’은 앞서 ‘신생물, 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정 장애 (C00)’ 만큼 뚜렷하게 대도시와 지방으로 구분되지는 않지만, 지방 중 경기도와 경상도를 제외한 나머지 지역들이 상대적으로 높은 사망 가능성을 보이는 것을 알 수 있다. 또한가지 특징은 서울의 경우는 신뢰구간의 상한이 0에 가까워 다른 대도시와 약간 다른 양상을 보이는 것을 알 수 있다.

성별 변수를 살펴보면, 남자의 계수는 0.11로, 다른 변수들이 동일한 조건으로 제어되었을 때 남자가 ‘내분비, 영양 및 대사 질환’으로 사망할 오즈는 여자의 오즈의 $e^{0.11}=1.12$ 배 높다. 앞서의 C00 질병에 비하여 E00 질병은 남녀의 차이가 뚜렷함을 알 수 있다.



<그림 16> 직업별 회귀계수추정치

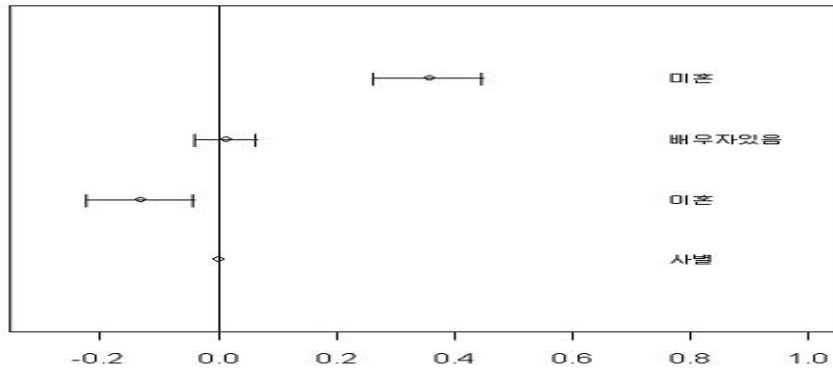
직업 변수 <그림 16>을 보면, 직업의 모든 범주가 ‘학생·가사·무직’ 보다 유의하게 큰 계수를 갖는다. 특히 기술공의 경우는 E00 로 사망할 오즈가 ‘학생·가사·무직’ 오즈의 무려 2.64배 높다. 즉, 164% 증가한다. 관리자, 기능원, 단순노무, 서비스, 장치 및 기계조작 이 다음 그룹으로 이들은 E00 로 사망할 오즈가 ‘학생·가사·무직’ 오즈의 약 2배이다. 농림어업과 전문가 직업이 그 뒤를 따르는데 이들의 오즈는 ‘학생·가사·무직’ 오즈의 약 1.6배

이다. 이상을 보면 ‘내분비, 영양 및 대사 질환’은 직업적 스트레스와 밀접한 관련이 있는 것으로 보인다. 기술공, 장치 및 기계조작 등 스트레스를 많이 받는 그룹이 ‘내분비, 영양 및 대사 질환’으로 사망할 가능성이 제일 높으며 그 뒤를 비교적 덜 스트레스를 받는 직업군이 라 할 수 있는 농림어업과 전문가가 따른다. ‘학생·가사·무직’은 가장 가능성이 낮은 그룹이다.

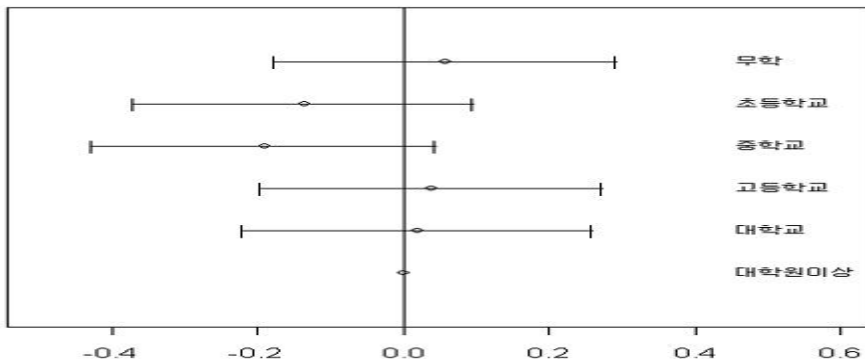
혼인상태 변수에 대한 <그림 17>을 보면 미혼이 높은 가능성을 보이는데, 미혼의 오즈는 사별의오즈의 1.43 배이다. 이혼의 경우 다소 낮기는 하지만 뚜렷이 낮다고 할 수는 없다.

교육 변수의 신뢰구간을 <그림 18>에서 살펴보면 유의한 범주가 없음을 알 수 있다. 이는 앞서 C00 질병과는 매우 다른 양상이다. C00 의 경우 교육수준이 높아지면 이 질병으로 인한 사망가능성이 현저히 낮아지는데 반해 E00 질병은 교육 수준과 무관하다. 일반적으로 교

육수준이 그 사람의 사회경제적 지위, 소득수준, 그리고 건강에 대한 지식과 관리를 상당히 반영하는 것을 고려할 때, '내분비, 영양 및 대사 질환'은 이들 요인들과는 무관한 질병으로 판단된다.



<그림 17> 혼인상태별 회귀계수추정치



<그림 18> 교육별 회귀계수추정치

IV. 결 론

사망원인 통계자료는 매년 통계청에서 전년도에 사망한 모든 사망자를 대상으로 수집하는 자료로 사망원인이 되는 요인과 함께 사망자의 주소지, 연령, 직업 등의 인구통계적 변수들

도 포함하고 있어 한국인의 건강관련 자료로 매우 중요한 의미를 지닌다. 그러나 그동안 이 자료에 대하여 단순한 기술적 통계만이 발표되어 자료의 활용이 제한적이었다. 본 논문에서는 자료에 포함된 모든 변수들을 고려한 다변량 회귀 분석을 실시하여, 다른 변수들이 제어되었을 때 각 변수가 미치는 영향에 대한 분석을 수행하였다.

관심있는 질병으로 가장 많은 경우의 사망원인이 되는 '신생물, 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정장애'와 앞으로 한국인에게 중요한 질병이 될 것으로 예상되는 '내분비, 영양 및 대사 질환'을 선택하여 이 두 질병 각각에 대하여 로지스틱 회귀분석을 실시하였다.

'신생물, 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정장애' 질병으로 인해 사망할 가능성은 지방 거주민이 대도시 주민 보다 더 높았다. 남녀의 차이는 거의 없었다. 직업별로는 (기능공, 단순노무, 장치 및 기계조작) > (기술공, 서비스업) > (관리자, 전문가, 학생·가사·무직) 순이었다. 혼인상태를 보면, 미혼>(사별, 이혼)> 배우자 있음 의 순으로 이 질병으로 인해 사망할 가능성이 높았다. 교육수준이 높을수록 이 질병으로 인해 사망할 가능성은 낮아짐을 볼 수 있었다. 끝으로 연령은 그다지 큰 영향을 미치지 못하는 것으로 나타났다. 이상의 결과는 모두 다른 변수들이 동일한 수준으로 제어되었을 때 주어진 변수의 영향으로 주어진 자료 내에서 합성효과를 최대한 제어한 결과이다.

'내분비, 영양 및 대사 질환'의 경우는 약간 다른 결과를 보였다. 첫째로, 지역적 구분이 '신생물, 혈액 및 조혈기관의 질환과 면역기전을 침범하는 특정장애' 만큼 뚜렷하게 나타나지 않았다. 특히 서울은 대도시와 지방의 중간적인 영향을 보였다. 성별의 영향이 뚜렷하게 나타나 남성이 여성보다 12% 더 증가된 오즈를 보였다. 직업의 영향을 보면, 직업적 스트레스의 영향이 매우 뚜렷하게 나타나 이 질병은 스트레스에 민감한 영향을 받는 질병으로 판단된다. 모든 직업이 학생·가사·무직에 비해 1.5배~2.6배 큰 오즈를 보였다. 반면에 한 사람의 사회경제적 지위, 소득수준 등을 반영한다고 볼 수 있는 교육수준은 이 질병에 유의한 영향을 미치지 못하였다. 혼인 상태는 미혼의 경우 이 질병으로 인한 사망가능성이 높고 나머지는 비슷하였다. 연령은 유의한 영향을 미치지 못하였다.

이상과 같이 사망원인 통계 자료 분석으로부터 사망에 이르게 하는 요인이 되는 질병에 영향을 미치는 여러 변수들을 살펴 보았다. 여러 변수들을 동시에 고려함으로써 단변량적 분석이 가지는 한계점을 극복하여 가능한 합성효과를 제어한 결과들을 볼 수 있었다.

사망원인 통계 자료에서 아쉬운 점은, 앞서 언급한 바와 같이 직업의 경우 생애 마지막 직업이 아니라 성인 이후에 가장 많은 시간을 보낸 직업을 조사하는 것이 더 바람직할 것이다. 또한 직업이나 교육 수준 외에 한 사람의 사회경제적 지위를 가장 잘 대변한다고 보는 수입에 대한 자료가 포함되어 있으면 더 좋은 분석 결과를 얻을 수 있을 것이다.

참 고 문 헌

한국질병분류사인코드(제 5차 개정,2005.5.10), www.kostat.go.kr, 통계청.

<http://kostat.go.kr/kssc/stclass/StClassAction.do?method=dis&classKind=5&kssc=popup>.

사망원인통계연보(2008):전국편 = Annual report on the cause of death statistics, 통계청.

SAS (2008), The SAS System for Windows V9.1.2, SAS Institute Inc.

사망원인통계 (2008), 통계청.

http://blog.naver.com/hypomania9?Redirect=Log&logNo=20100064610&topReferer=http://cafeblog.search.naver.com&imgsrc=20100211_146/hypomania9_1265847287021qovn7_jpg/%C6%F7%B8%CB%BA%AF%C8%AF_s_hypomania9.jpg

Statistical Analysis of the Cause of Death Statistics

Man-Suk OH* · Woo-Young SHIN**

Abstract

Bureau of Statistics, Korea, has been collecting data from deaths each year about the cause of death and other demographic variables. Statistical analysis of such data would give not only useful guidelines for individual health care but also establishing governmental policies for health care system. However, only univariate analyses have been done and there is a need for multivariate analysis of the data to control confounding factors.

In this paper, we analyze the Cause of Death Statistics, Korea, 2008. We especially focus on two diseases, 'Neoplasm, blood and blood diseases and certain disorders involving the immune mechanism' which is the most frequent cause of deaths and 'Endocrine, nutritional and metabolic diseases' which is rapidly increasing among Koreans. We use a logistic model for each of the two response variables. We first search for an appropriate model for the data, perform test of significance for each explanatory variable, and then interpret the effect of each explanatory variable to the response variables.

The analysis results show that for 'Neoplasm, blood and blood diseases and certain disorders involving the immune mechanism', there is a higher chance that this disease would be a cause of death for residents in large cities, for city laborers, and people with a low education level. For 'Endocrine, nutritional and metabolic diseases', there is a higher chance that this disease would be a cause of death for men, for people with more stressful jobs. Especially, the effect of job on this disease is very strong. For instance, the odds of technicians would die of 'Endocrine, nutritional and metabolic diseases' is 2.64 times larger than that of students·houseworkers·unemployed.

Keywords: Cause of Death Statistics, Logistic regression model, Confounding factor.

* Professor, Statistics Dept., Ewha Womans University, This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology (No. 2010-0010285). Corresponding author.

** Graduate student, Statistics Dept., Ewha Womans University.

