

사학연금 퇴직률 산출 개선방안 연구

백혜연*

〈 초 록 〉

공적연금제도는 장기적 유지 및 운영을 위해 기금의 재정건전성 및 지속가능성 진단을 목적으로 재정계산제도를 운영하고 있다. 정확한 재정계산은 매우 중요하며 이를 위한 선행작업으로 재정계산에 요구되는 기본 가정들을 보다 합리적으로 추정해야 할 필요가 있다. 본 연구는 로지스틱 회귀분석(logistic regression)을 이용하여 사학연금의 재정계산에 적용되는 다양한 기초율들 중 퇴직률을 산출하는 것에 그 목적이 있다. 사학연금은 현재 퇴직률을 교원 및 직원에 대하여 각 성별로 총 4개 집단을 구분하여 각 집단별 가입연령과 재직기간에 따라 산출하고 있다. 그러나 본 연구에서는 학교급 등 퇴직률 산출에 있어 보다 유의한 집단 구분이 있는지를 확인하고 보정의 어려움을 피할 수 있는 하나의 대안으로서 로지스틱 회귀분석을 이용하여 퇴직률을 산출해 보았다. 또한 우수한 모형을 판별하기 위해 통계적으로 우수한 모형보다는 실무적으로 사학연금 재정추계에 적합한 모형을 찾는 것을 목표로 하여 퇴직률을 추정된 값을 제시하였다.

핵심용어 : 공적연금, 로지스틱 회귀분석, 재정계산, 퇴직률, 학교급

* 한국보건사회연구원 인구정책연구실 부연구위원

제1장 서론

1. 연구의 배경 및 목적

스튜어드십 코드 도입과 사회적 책임투자 강화 요구 등 최근 우리나라 공적연금제도에 대한 논의는 그 어느 때보다 다양하게 이루어지고 있다. 그 중 가장 큰 관심의 대상은 역시 공적연금제도의 재정건전성 및 지속가능성이라 할 수 있을 것이다. 사실 공적연금제도의 안정적 유지에 대한 우려와 의구심은 비단 우리나라만의 문제는 아니며 공적연금제도를 운영하고 있는 모든 국가가 공통적으로 경험하는 사항이다. 이러한 근심에 대해 우리나라를 비롯한 해외 많은 국가들은 공적연금의 재정상태 진단을 목적으로 재정계산제도를 운영하고 있다. 재정계산제도는 “연금재정에 영향을 미치는 다양한 요인들에 대한 가정을 설정하고 이를 기반으로 장래의 수입과 지출, 재정수지, 기금규모 등을 추정하는 일련의 과정”(김병률·김재경, 2016, p. 190)으로 이에 재정상태에 대한 점검 이외에도 장기적 재정균형을 유지하기 위한 운영계획 등이 포함되기도 한다(김순옥·권문일, 1998; 윤석명·강성호, 2001). 이러한 측면에서 정확한 재정계산은 공적연금제도의 장기적 유지 및 운영을 위해 매우 중요하며 이를 위한 선행작업으로 재정계산에 요구되는 기본 가정들을 보다 합리적으로 추정해야 할 필요가 있다.

본 연구는 로지스틱 회귀분석(logistic regression)을 이용하여 사학연금의 재정계산에 적용되는 다양한 가정 중 퇴직률을 산출하는 것에 그 목적이 있다. 퇴직률은 보수상승률 또는 운용수익률과 같은 거시경제변수처럼 사학연금기금에 직접적인 영향을 미치는 요소는 아니지만 퇴직률이 높아지면 가입기간의 단축으로 인해 퇴직급여액이 하락하여 재정지출이 감소하는 한편, 연금수급요건을 충족하는 퇴직자 수가 상대적으로 감소하여 연금 대비 일시금 지출액이 증가하게 된다. 이렇듯 퇴직률은 주요 변수 중의 하나로 전반적인 재정계산결과에 영향을 미치는 주요

1. 제도변수란 가입자 집단의 특성에 따라 우리나라 4대 공적연금별로 각각 개별적으로 산출·적용하는 기본 가정을 일컫는다. 퇴직률, 각종 연금선택률 등이 이에 해당한다. 반면 장래 인구전망, 그리고 물가상승률 등을 비롯한 거시경제변수는 공적연금 간 재정계산결과와의 정합성 향상을 위해 정부가 제공하는 수치를 동일하게 적용한다.

제도변수 중 하나로 산출방안에 있어 보다 높은 수준의 합리성과 정교함이 요구되기 때문이다.¹ 현행 사학연금 재정계산 목적의 퇴직률은 성별과 교원 및 사무직원의 두 직종별로 남자 교원과 여자 교원, 그리고 남자 사무직원과 여자 사무직원과 같이 4개 집단으로 구분되어 각 집단별로 가입연령과 재직기간에 따라 산출된다. 이 경우 초·중·고 교원과 대학 교원의 일반적인 정년연령이 다르다는 것을 반영하기 어렵다. 또한 퇴직자 수를 재직자 수로 나눈 것으로 정의되는 기초퇴직률에 기반하여 퇴직률이 산출되는 바, 4개 집단별로 가입연령 및 재직기간에 상응하는 재직자와 퇴직자가 존재하지 않거나 그 수가 적을 경우에는 기초퇴직률이 극단적으로 나타나 보정(graduation)에도 어려움이 많다.² 이에 본 연구에서는 학교급 등 퇴직률 산출에 있어 보다 유의한 집단 구분이 있는지를 확인하고 보정의 어려움을 피할 수 있는 하나의 대안으로서 로지스틱 회귀분석을 이용하여 퇴직률을 산출해 보고자 한다.

2. 공적연금의 재정계산을 위한 기초율

공적연금의 재정계산을 위해 사용되는 기초율들에 대해서 간단히 살펴보면 다음과 같다. 국민연금의 경우 재정계산을 위해 주요 가정으로 사용되는 변수들은 크게 3가지로 나눌 수 있다. 주요 가정을 위한 변수들은 인구변수, 경제변수, 제도변수로 구분된다. 인구변수는 출산, 사망, 국제이동을 포함한 인구변동요인에 관한 가정을 위해 사용되며, 경제변수는 장래인구추계를 기반으로 추정된 전망치들로 임금상승률, 기금투자수익률, 물가상승률 등 주요 거시경제 전망을 위한 변수들을 포함하고 있다. 마지막으로 제도변수는 국민연금 가입률, 지역가입자 비율, 납부예외자 비율, 소득수준 가정, 공적연금 연계신청률 가정 등이 있다(국민연금재정추계위원회, 2013). 사학연금에서 퇴직률이 가입기간, 퇴직급여 수준, 퇴직급여 종류들을 결정하는 기초율의 역할을 하는 것과 유사하게 국민연금에서는 가입종별 간 이동률이나 사망률이 유사한 값들을 결정하기 위해 사용되고 있다. 이와 같이 퇴직률의 개념은 국민연금보다는 오히려 퇴직급여 개념이 급여에 포함되어 있는 특수직역연금이나 사보험의 형태인 퇴직급여제도에서 중요하게 사용되는 개념이라고 볼 수 있다.

공무원연금에서는 사학연금과 동일하게 퇴직률을 재정계산 시 기초율로 사용하고 있다. 한국개발연구원(2016)에 따르면 “공무원연금 장기재정추계는 고령화가 예상보다 빠르게 진전되는

2. 보다 구체적인 퇴직률 산출방법은 사학연금 홈페이지(www.tp.or.kr)에 게시되어 있는 사학연금재정계산 보고서를 참조하기 바란다.

시점에서 합리적인 전망의 중요성이 입증되고” 있으며, “장기재정추계를 구성하는 중요한 한 축인 추계 가정(기초율)은 연금제도 변화와 재직자 및 수급자의 인구구조 노령화를 반영하여 편의(bias)없는 전제를 사용해야”한다고 강조하고 있다. 기존 공무원연금은 사학연금과 유사하게 성별·직종별로 퇴직률을 추정하였다. 기존 모형에서는 최근 3년 통계치를 기초로 하였고, Jenkins 5차 접촉보간법, Greville 3차 11항식을 활용한 보정 후 최근 퇴직률 변동추세를 감안하여 전체 총 퇴직률이 추계직전연도 실적수준이 되도록 조정하였었다. 그러나 공무원연금 관련 연구인 한국개발연구원(2016), 한국보건사회연구원(2018)의 기초율 개선 연구과제에서는 기존 모형과 같이 직종(비교육직, 교육직), 성(남성, 여성)으로 구분하여 퇴직률을 추정하되, 연령과 재직기간의 함수로 모형화 하였다. 또한 이때 추정 방법론으로는 비모수적 커널 추정법을 사용하여 각 연령과 재직기간 조합별 퇴직률을 추정한 결과를 제시하였다. 공무원연금의 경우 전체 공무원 가입자 중 교육 공무원 수가 차지하는 비중이 약 32%(2017년 기준) 수준이기 때문에 교육직 내에서 성별 이외에 다른 기준으로 분리하여 별도의 퇴직률을 산출하기에는 무리일 수 있다. 그러나 한국보건사회연구원(2017)에 따르면 사학연금의 “전체 재직 교직원 중 재직 교원이 차지하는 비중은 2000년대 중반까지는 증가하다가 후반 들어 감소하는 양상을 보였고, 2000년 재직 교원 비율은 62.6%에서 2004년 64.0%에 도달하였다”고 한다. 그리고 “이 비율이 2006년까지 유지되다가 2007년부터 감소하기 시작하여, 2016년 55.7%에 이르렀다”고 한다. 이와 같이 공무원연금과는 반대로 사학연금은 교원의 비중이 직원의 비중보다 많은 편이기 때문에 교원의 특성을 반영한 세분화된 퇴직률 산출 방법에 대해 고려해 볼만하다고 볼 수 있다. 따라서 본 연구에서는 기존의 퇴직률 추정 방법을 세분화하여 추정하는 방법으로 개선하는 것에 대해 논의해 보고자 한다.

본 연구의 구성은 다음과 같다. 공적연금제도의 재정진단이라는 특수한 목적에 의해 시행되는 재정계산의 특성상 이와 관련한, 특히 퇴직률 산출과 관련한 연구를 찾아보기는 매우 어려운 실정이다. 따라서 제II장에서는 먼저 본 연구에서 사용하는 데이터에 대한 설명과 더불어 퇴직자 분포도에 기반한 기초적인 분석을 제공한다. 이를 통해 성별 및 직종에 따른 4개 집단 구분 이외 다른 효과적 구분의 가능성에 대한 단초를 얻을 수 있다. 제III장에서는 로지스틱 회귀분석에 대한 간략한 소개 후 본 연구에서의 로짓 모형 및 분석결과를 제시한다. 제IV장은 현행 재정계산에서의 퇴직률과 제III장의 분석결과로부터 산출되는 퇴직률을 적용하여 각각 퇴직자 수를 추정한 후 실제 퇴직자 수와 비교한 결과를 주요 내용으로 한다.

이를 통해 로짓 모형에 기반한 퇴직률 산출방법의 장단점을 살펴본다. 마지막으로 결론에서는 본 연구의 내용을 요약함과 동시에 한계점과 향후의 과제에 대해 간략히 논하도록 한다.

제2장 사학연금 퇴직자에 대한 기본 분석

1. 이용 자료

본 연구에서는 사학연금 가입자의 퇴직률 산출을 위해 2010년부터 2017년까지 8개년 동안의 재직자 2,289,971명과 퇴직자 204,099명의 자료를 사용하였다.³ 해당 자료에는 성별, 교원과 사무직원의 직종 구분, 7개로 구분된 학교급, 출생일과 더불어 연령, 사학연금 가입기간 등이 포함되어 있다.⁴ 과거 8년 동안의 자료를 합한 것은 통계적 분석 및 퇴직률 모형설정에 충분한 크기의 퇴직자 표본을 얻기 위함이다. 그리고 퇴직자를 성별과 직종, 학교급에 따라 구분할 경우 총 28개 집단으로 나누어 볼 수 있으나 여기에 더해 연령과 가입기간을 추가하면 각 셀(cell)에 해당하는 퇴직자 수를 확보하기 어려울 수 있기 때문에 충분한 표본을 수집하고자 하였다. 2~3개년 자료를 이용할 경우 퇴직이라는 사건에서 나타나는 사학연금 가입자의 최근 특징을 반영하기는 용이할 것이나 이러한 추세가 향후에도 장기간에 걸쳐 지속되지 않는다면 장기 재정추계에서의 퇴직 관련 추계는 과소 또는 과대 추정될 가능성이 높다.⁵ 따라서 8개년 퇴직자 자료를 합침으로써 60년 이상의 기간을 대상으로 하는 장기 재정추계에 있어 보다 안정적인 퇴직률을 산출하고자 하였다.

3. 재직자는 매 연도의 6월 말 현재 사학연금에 가입 중인 재직 교직원을 의미한다. 참고로 월별 가입자 수는 2월에 큰 폭으로 감소한 후 3월에 급격한 반등을 나타내며 이후 7월에 최고점에 이르기까지 증가하다 8월에 약간의 감소를 나타낸 후 연말까지 다소간의 증가 또는 유사한 수준을 유지하는 추세를 나타낸다.

4. 7개 학교급은 유치원, 초등학교, 중학교, 고등학교, 전문대학, 대학, 특수학교 그리고 학교법인을 포함한 기타 사학연금법 적용대상기관을 말한다. 사학연금 가입기간은 사학연금법을 적용받는 재직기간으로서 사학연금법 적용대상기관에서의 재직기간뿐만 아니라 그 이전 공무원연금법 또는 군인연금법 적용기간의 합산, 소급통산 및 군 복무기간 산입 등을 포함한 기간을 말한다. 사학연금 최대 가입기간은 기존 33년에서 2016년 법 개정에 따라 36년으로 단계적으로 연장되었다.

5. 이러한 점에서 2010년 이전의 퇴직자 자료 역시 포함하는 것이 보다 합리적일것으나 몇 차례의 사학연금법 개정 중 2010년의 개정이 사학연금제대상 가장 중요한 개혁이었다는 점에서 2010년부터의 퇴직자 자료를 이용하였다. 2010년 시행된 사학연금법의 주요 개정 내용은 다음과 같다. 부담금 및 급여산정의 기준이 보수월액에서 기준소득월액으로 변경되었고, 동시에 퇴직연금액의 산정기준이 퇴직 전 보수월액 3년 평균에서 전 재직기간의 평균기준소득월액으로 대체되었다. 또한 1996년 도입된 연금수급개시연령이 60세에서 65세로 상향 조정되었으며, 유족연금의 지급률이 퇴직연금의 70%에서 60%로 낮아졌다.

또한 재정추계 목적에 따라 ‘퇴직자’를 사학연금의 퇴직급여를 신청한 퇴직 교직원 또는 재직 중 사망에 따라 유족급여가 신청된 사망 교직원으로 정의하였다. ‘재정추계에서의 퇴직’이란 급여지출을 발생시키는 사건으로서, 계약의 만료 후 다시 재임용되거나 퇴직 후 다른 사학연금 적용대상기관으로 재취업하여 퇴직급여의 수령 없는 ‘신분변동상의 퇴직’과는 구분되어야 한다.

2. 성별 · 직종 · 학교급 구분에 따른 퇴직자 특성

본 절에서는 성별, 직종 및 학교급 등으로 퇴직자를 구분한 후 각 집단별로 가입기간 및 퇴직연령에 따른 기초 통계량 및 분포도를 제시한다. 이를 통해 퇴직이라는 사건(event)에 있어 집단 간 차이를 쉽고 빠르게 살펴볼 수 있으며, 로지스틱 회귀분석을 이용한 퇴직률 산출모형에 어떠한 범주형(categorical) 설명변수를 고려해야 할지에 관한 직관적 정보를 얻을 수 있다.

〈표 1〉은 기존 퇴직률 산출방법에서처럼 성별과 직종으로 구분한 4개 집단의 가입기간 및 퇴직 시 연령의 기초 통계량을 정리한 것이다. 〈표 1〉에 따르면 여자 사무직원의 평균 가입기간(4.65년)은 남자 사무직원의 평균 가입기간(10.04년)의 46.3%이고, 특히 교원일 경우 남자 퇴직자의 평균 가입기간(18.99년) 대비 여자 퇴직자의 평균 가입기간(5.03년)은 26.5%에 불과하다. 퇴직연령 역시 남자 대비 여자 교원과 사무직원의 퇴직연령은 각각 19.53세와 8.41세가 차이나 직종을 통제(즉, 직종을 고정)하였을 때 성별에 따라 평균 가입기간과 퇴직연령에 상당한 차이를 나타내고 있음을 추측해 볼 수 있다. 반면 성별을 통제하였을 때에는 평균 가입기간에서 나타나는 직종의 영향은 상대적으로 더 작아 보인다. 남자의 경우 사무직원의 평균 가입기간(10.04년)은 교원(18.99년)의 52.9%, 여자의 경우 92.5%로 나타나 직종을 통제하였을 때 성별에 의한 차이보다 작게 나타나는 것이다.⁶

6. 참고로 〈표 1〉에서처럼 2개 범주형 변수(성별과 직종)에 따른 교차 구분 대신 성별로만 구분하면 남자의 평균 가입기간은 14.89년, 여자는 4.87년으로 남자 대비 여자 평균 가입기간의 비율은 32.7%이다. 직종으로만 구분할 경우에는 교원의 평균 가입기간은 9.86년, 직원은 6.69년 그리고 교원 대비 직원의 평균 가입기간 비율은 67.8%이다.

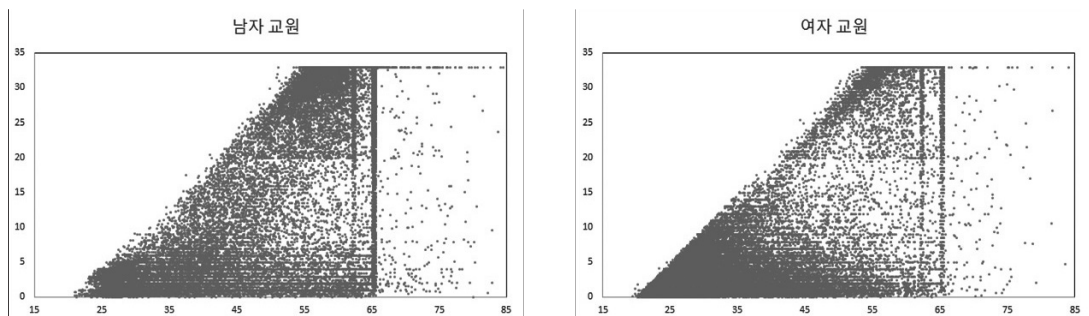
〈표 1〉 성별 및 직종 구분에 따른 가입기간과 퇴직연령의 기초 통계량 비교

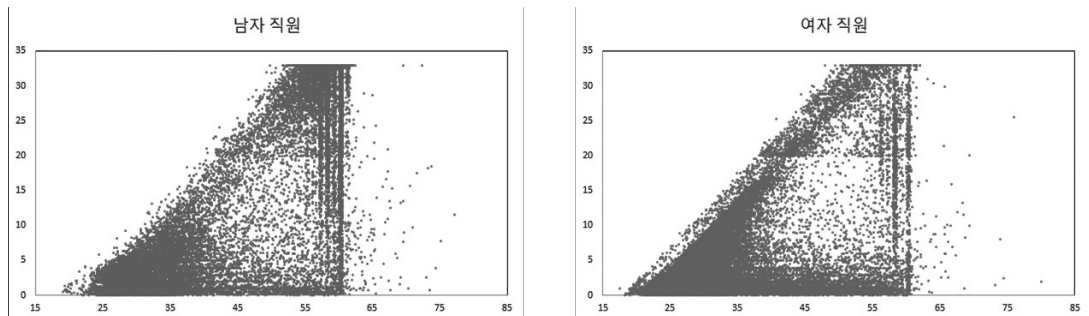
(단위: 명, 년, 세)

구 분		남자 교원	여자 교원	남자 직원	여자 직원
표 본 수		39,448	74,618	33,979	56,054
가입기간	평 균	18.99	5.03	10.04	4.65
	표준편차	13.67	7.77	11.13	6.55
퇴직연령	평 균	51.93	32.40	40.01	31.60
	표준편차	13.31	10.72	12.24	9.17

〈그림 1〉은 〈표 1〉의 4개 집단별로 퇴직연령(횡축)과 사학연금 가입기간(종축)에 따라 퇴직자 분포를 나타낸 것으로 앞서 살펴본 성별과 직종 구분에 따른 평균 가입기간에서의 차이를 시각적으로 보다 편리하게 살펴볼 수 있다. 낮은 연령대의 퇴직자는 긴 가입기간을 지닐 수 없으므로 좌상방에 해당하는 퇴직자는 존재할 수 없다. 그리고 해당 퇴직자에게 적용되는 사학연금법상 가입 가능한 최대 기간은 33년이므로 그 이상의 가입기간을 지닌 퇴직자 역시 발생하지 않으므로 사다리꼴 모양의 분포가 나타나게 된다. 이러한 공통점과 더불어 앞서 언급한 바와 같이 교원의 경우 성별에 따른 퇴직자 분포가 확연하게 구분된다. 여자 교원의 경우 퇴직자의 대다수가 좌하방에 모여 있어 가입기간과 퇴직연령이 낮음을 알 수 있다. 반면 남자 교원은 여자 교원에 비해 상대적으로 좌하방과 우상방에 고루 분포되어 있어 가입기간과 퇴직연령의 평균과 표준편차가 여자 교원에 비해 높을 것임을 쉽게 유추해 볼 수 있다. 그러나 성별을 통제할 경우 남자 교원과 남자 직원 간, 그리고 여자 교원과 여자 직원 간에는 분포도에 있어 이러한 정도의 명확한 차이를 찾아보기 어렵다.

〈그림 1〉 퇴직연령 및 가입기간에 따른 퇴직자 분포 : 성별 및 직종별 구분





그러나 <그림 1>에서는 <표 1>에서 확인하기 어려웠던 교원과 사무직원 간 한 가지 큰 특징적 차이가 발견된다. 퇴직에 있어 정년연령이 바로 그것이다. 이는 <그림 1>에서 사다리꼴의 우변처럼 보이는 수직선으로 확인된다. 사학연금 적용대상기관의 경우 공무원과 달리 정년이 모든 대상기관의 교직원에게 일률적으로 적용되는 것은 아니다. 그러나 유치원을 제외한 많은 사학연금 가입기관에서는 「국가공무원법」상 교육공무원과 사무직원의 정년을 그대로 적용하고 있는 것으로 알려져 있다. 「국가공무원법」에 따르면 사무직원의 정년은 60세이며, 교육공무원의 정년은 다시 학교급에 따라 대학 교원에게는 65세 그리고 초·중·고 교원에게는 62세의 정년이 적용된다. 이에 따라 <그림 1>의 남자 교원에서는 퇴직연령 65에서 모든 가입기간에 걸쳐 수직선이 뚜렷하게 형성되어 있으며, 바로 왼쪽 62세에서는 가입기간이 긴 위쪽 부분에 동일한 수직선이 형성되어 있다. 62세의 위 부분에 수직선이 뚜렷하다는 것은 퇴직자 중 장기 재직자가 많다는 것을 의미하는 반면 65세의 경우에는 장기 재직자뿐만 아니라 가입기간이 짧지만 정년연령에 의해 퇴직하게 된 단기 재직자 역시 많다는 것을 의미한다. 여자 교원의 경우 남자 교원에 비해 흐릿하기는 하지만 마찬가지로 퇴직연령 65세와 62세에 이것이 발견된다. 또한 사무직원 퇴직자 분포를 보여주는 <그림 1> 하단 패널에서는 60세 혹은 60세 근처의 퇴직연령에서 이러한 수직선을 볼 수 있다.

성별을 통제할 경우 직종 간 가입기간 및 퇴직연령에서의 차이가 뚜렷하지 않아도 직종은 정년연령의 차이로 인해 퇴직률 산출에 있어 의미있는 중요한 범주형 변수일 수 있다. 또한 정년연령은 직종 이외에도 학교급에 따라 다르게 적용되는 바, 현행 퇴직률의 범주형 변수인 성별과 직종에 더해 학교급을 고려해 볼 필요가 있을 것이다. 이에 <표 2>와 <그림 2>에서 학교급 구분에 따른 가입기간과 퇴직연령의 기초 통계량과 더불어 퇴직자 분포를 살펴보았다. <표 2>에 따르면 긴 가입기간을 나타내는 중학교와 고등학교 퇴직 교직원의 평균 가입기간은 각각 27.78년과 26.36년으로 대학 퇴직 교직원에 비해 3.82배와 3.63배, 유치원에 비해서는 무려 9.05배와 8.59배가 더 긴 것으로 나타나 학교급 차이에 따라 큰 차이가 나타남을 알 수 있다. 초등학교와 특수학교의

경우 그 중간으로 두 학교급 간에는 평균 가입기간에 큰 차이를 나타내지 않으며 사무직원으로만 구성된 학교법인 위주의 기타 적용대상기관이 그 뒤를 잇고 있다.

〈표 2〉 학교급 구분에 따른 가입기간과 퇴직연령의 기초 통계량 비교

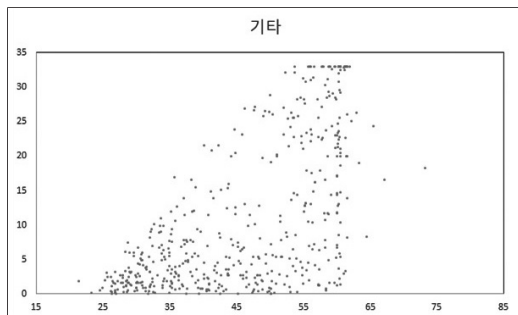
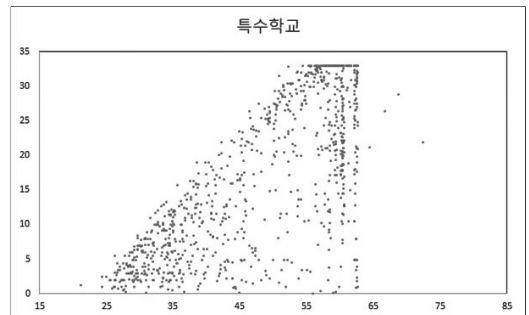
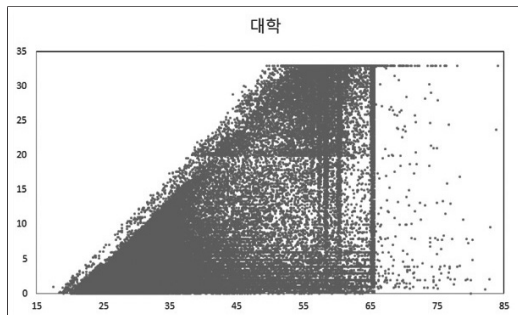
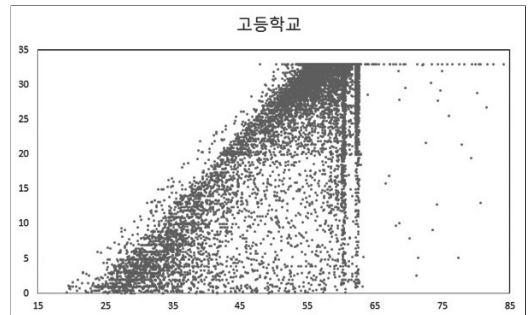
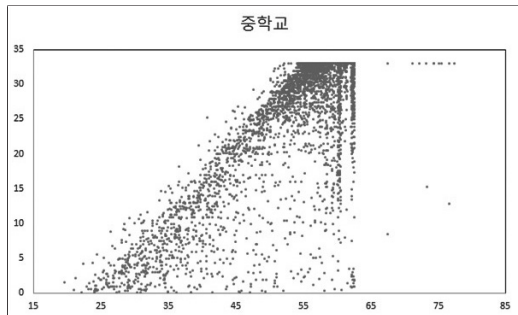
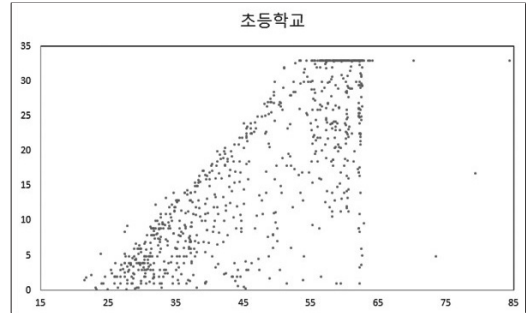
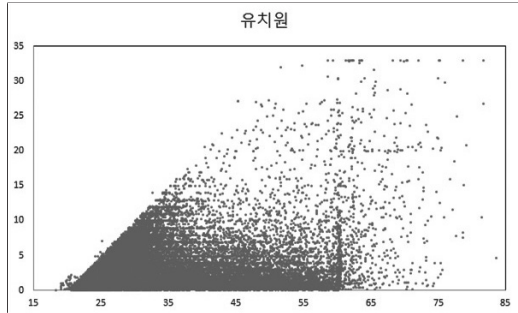
(단위 : 명, 년, 세)

구 분		유치원	초등	중	고등	대학*	특수	기타
표 본 수		54,242	880	6,561	16,492	124,455	964	505
가입기간	평 균	3.07	18.19	27.78	26.36	7.27	17.38	10.89
	표준편차	2.96	11.53	8.84	10.10	9.83	11.26	10.73
퇴직연령	평 균	31.32	48.10	55.79	54.96	36.27	48.27	45.76
	표준편차	9.01	12.53	8.51	9.55	13.06	12.03	11.65

* 원시 자료(raw data)의 전문대와 일반대학을 합한 것임

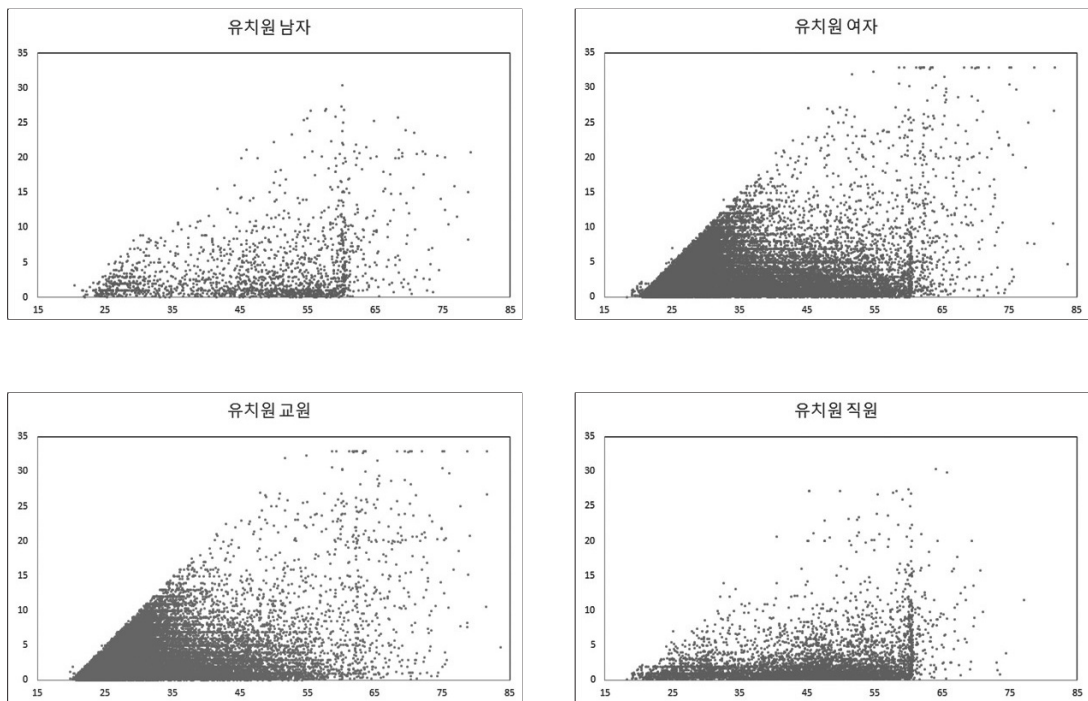
퇴직자의 가입기간 및 퇴직연령에 있어 학교급 구분에 따른 차이는 〈그림 2〉에서 보다 명확히 확인할 수 있다. 유치원 퇴직자는 장기 가입자가 드물다는 것이 특징이 있다. 퇴직자의 대다수가 하단에 몰려있을 뿐만 아니라 퇴직연령 60세에 수직선이 일부 형성되어 있으나 종축의 하단 부근에 집중되어 있는 것이다. 퇴직연령에서 가입기간을 차감한 수치를 가입 시 연령으로 본다면 유치원 교직원의 가입연령은 상당히 넓은 연령대 구간에 퍼져 있음을 추론해 볼 수 있다. 이에 비해 중학교와 고등학교는 장기 재직자가 매우 많은 것으로 파악된다. 특히 고등학교의 경우 좌하방에서 우상방에 걸쳐 퇴직자로 이루어진 하나의 직선이 형성된다. 이로부터 미루어 짐작하건대 이들의 가입연령은 20대 후반에서 30대 초반에 집중되어 있는 것으로 보인다. 또한 퇴직연령 60세와 62세 부근에서 수직선이 형성되어 있는 것을 알 수 있는데 이는 앞서 설명한 바와 같이 해당 학교급 교육공무원과 사무직 공무원의 정년에 따른 것으로 추측된다. 이에 비해 대학 퇴직자는 마치 우상방과 좌하방 퇴직자 집단으로 구분되는 것으로 나타나는 특징을 보이고 있다. 또한 퇴직연령 65세에 뚜렷한 수직선이 있어 〈그림 1〉의 남자 교원과 여자 교원에서 나타났던 동일한 수직선이 대학 교원 정년에 의한 것임을 알 수 있다. 초등학교와 특수학교 그리고 기타 기관은 해당 퇴직자 수가 크지 않아 다른 학교급 기관에서 나타난 것과 같은 명확한 특징을 찾아보기 쉽지 않다. 이는, 만약 학교급을 범주형 변수로서 분석모형에 감안하고자 한다면 충분한 표본수 확보를 위해 일부 학교급 퇴직자 자료를 합치는 것이 필요함을 시사한다.

〈그림 2〉 퇴직연령 및 가입기간에 따른 퇴직자 분포 : 학교급별 구분



〈그림 1〉과 〈그림 2〉는 성별, 직종 그리고 학교급에 따라 퇴직자의 가입기간 및 퇴직연령에 있어 차이가 있을 수 있음을 쉽고 직관적으로 알려 준다. 그러나 이러한 시각적 비교를 통해서 어떤 범주형 변수가 더 중요한 것인지, 그리고 어떠한 변수들을 동시에 감안해야 할지에 대한 판단기준을 제공해 주지는 못한다. 예를 들어 〈그림 2〉의 유치원 퇴직자 분포도를 성별에 따라서, 그리고 직종에 따라서 세분(〈그림 3〉 참조)해 보면 유치원 여자 퇴직자와 유치원 교원 퇴직자 분포도는 매우 유사하며 또한 〈그림 2〉의 유치원 전체 퇴직자 분포도와도 비슷함을 알 수 있다. 이때 퇴직률과 관련하여 유치원이라는 학교급에 대해 성별을 같이 고려해야 하는지 또는 직종을 감안해야 하는지, 아니면 성별과 직종 모두 중요한 변수인지 둘 다 중요하지 않은 변수인지를 결정하기란 매우 어려운 것이다. 이에 본 연구에서는 로짓 모형을 사용함으로써 퇴직률에 영향을 미치는 주요 변수들을 통계적으로 구분함과 동시에 해당 변수들을 이용하여 퇴직률을 산출해 보고자 하는 것이다.

〈그림 3〉 유치원 퇴직자의 성별 구분과 직종별 구분



제3장 로지스틱 회귀분석과 퇴직률 산출

1. 로지스틱 회귀분석⁷

로지스틱 회귀분석은 반응변수가 0과 1 둘 중 하나의 값을 갖는 이분법적(binary) 자료일 때 사용되는 일반화 선형모형의 한 종류이다. 일반화 선형모형(Generalized Linear Model, GLM)은 1972년 John Nelder와 Robert Wedderburn에 의해 소개되었다(Nelder and Wedderburn, 1972). 일반화 선형모형은 전통적인 선형모형(linear model)을 확장한 모형으로 보험 및 리스크 관리 분야에서 많이 활용되고 있다. 회귀분석이나 분산분석은 반응변수가 정규분포를 따르는 연속형 변수일 때 사용되지만 대부분의 경우 반응변수가 정규분포를 따른다는 가정을 사용할 수도 없고 범주형 변수가 반응변수일 경우도 있기 때문에 이런 경우에는 일반화 선형모형을 사용한다.

일반화 선형모형에서 반응변수 Y 와 설명변수 $X_1 = x_1, \dots, X_p = x_p$ 간의 연관성은 다음과 같이 정의된다.

$$f_Y(y) = c(y, \phi) \exp\left\{\frac{y\theta - \alpha(\theta)}{\phi}\right\}.$$

$$E[Y|x_1, \dots, x_p] = \mu, \text{Var}[Y|x_1, \dots, x_p] = \phi \cdot V(\mu).$$

즉, x_1, \dots, x_p 가 주어졌을 때 Y 는 평균이 μ 인 지수족 분포(exponential family)를 따른다고 가정한다. 지수족 확률분포의 확률밀도함수 $f_Y(y)$ 는 위 식과 같고, 정준모수(canonical parameter) θ , 산포모수(dispersion parameter) ϕ 상수들과 $\alpha(\theta)$, $c(y, \phi)$ 함수들에 의해 최종적인 형태가 결정된다. 그리고 연결함수(link function) $g(\mu)$ 는 미분가능 단조함수라 가정하고, 설명변수와

7. 본 절은 Agresti(2002)와「SAS/STAT 9.2 사용자 가이드 (SAS/STAT 9.2 User's Guide)」의 로지스틱 프로시저(logistic procedure) 부분을 참고하여 작성하였다. 로지스틱 회귀분석을 포함한 일반화 선형모형 전반에 대해서는 Agresti(2002)를 참조하기 바란다.

반응변수 Y의 기댓값인 μ 에 대한 함수값($g(\mu)$)이 선형관계를 갖는다고 가정한 것이 일반화 선형모형이다.

$$g(\mu) = x'_i \beta.$$

다음의 표는 지수족 분포와 그에 대한 연결함수(link function)들의 예시를 나타낸 것이다.

〈표 3〉 지수족 분포에 대한 연결함수

분포	link 이름	link 함수	평균
Normal	Identity	μ	$\mu = X\beta$
Exponential	Inverse	μ^{-1}	$\mu = (X\beta)^{-1}$
Gamma			
Inverse Gaussian	Inverse squared	μ^{-2}	$\mu = (X\beta)^{-1/2}$
Poisson	Log	$\ln(\mu)$	$\mu = \exp(X\beta)$
Bernoulli	Logit	$\ln\left(\frac{\mu}{1-\mu}\right)$	$\mu = \frac{\exp(X\beta)}{1 + \exp(X\beta)} = \frac{1}{1 + \exp(-X\beta)}$
Binomial			
Categorical			
Multinomial			

출처 : McCullagh and Nelder (1989).

일반화 선형모형에서 로짓 연결함수를 사용한 것이 로지스틱 회귀분석 모형이 된다. 로지스틱 회귀분석은 독립변수의 선형결합을 이용하여 사건의 발생가능성(발생확률)을 예측하는데 사용되는 통계기법이다. 독립변수의 선형결합으로 종속변수를 설명한다는 관점에서는 회귀분석, 판별분석과 유사하다. 로지스틱 회귀분석의 종속변수는 명목척도로서 binary data이어야 하며, p개의 독립변수로 사건 E가 발생할 확률을 예측하기 위한 모형을 의미한다.

$p(E)$: 사건 E가 발생할 확률

$$p(E) = \frac{e^Z}{1 + e^Z}, p(E) = \frac{1}{1 + e^{-Z}}, e^Z = \frac{p}{1-p}, Z = \ln\left(\frac{p}{1-p}\right)$$

$Z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$: 독립변수의 선형결합.

〈표 4〉 로지스틱 회귀분석 시 변수 선택 방법

	변수 선택 방법
전진 선택법 (forward selection)	<ul style="list-style-type: none"> • 입력변수를 각 변수의 기여도에 따라 하나씩 추가하면서 선택 • 한 번 선택된 변수는 다른 변수에 의해 중요성이 상실되더라도 회귀모형에서 제거되지 않음.
후진제거법 (backward elimination)	<ul style="list-style-type: none"> • 모든 변수를 포함하는 완전모형으로부터 불필요한 변수를 하나씩 제거하는 방법 • 중요한 변수가 모형에서 제외될 가능성이 적음. • 한 번 제외된 변수는 다시 선택되지 않음.
단계별 선택법 (stepwise method)	<ul style="list-style-type: none"> • 각 단계마다 변수의 선택과 제거를 반복하면서 중요한 변수를 찾아내는 방법 • 중요한 변수를 하나씩 선택하면서 이미 선택된 변수들이 제거될 수 있는 지 각 단계마다 검토 • 모든 가능한 회귀를 통해서 얻은 모형보다 적절치 못한 경우가 있음.

오즈비(Odds ratio)는 사건이 발생할 확률과 발생하지 않을 확률의 비율 또는 사건이 성공할 확률과 실패할 확률의 비율을 의미하고,

$$\text{Odds Ratio} = \frac{\exp(\alpha + \beta_1 x_1 + \dots + \beta_i (x_i + 1) + \dots + \beta_p x_p)}{\exp(\alpha + \beta_1 x_1 + \dots + \beta_i (x_i) + \dots + \beta_p x_p)} = \exp(\beta_i).$$

오즈비가 1보다 작다(계수가 음의 값을 갖는다)는 것은 입력변수 x 가 감소방향으로 영향을 미침을 의미하고, 반대로 오즈비가 1보다 크다(계수가 양의 값을 갖는다)는 것은 증가방향으로 영향을 미침을 의미한다. 〈표 4〉는 로지스틱 회귀분석 시 변수를 선택하는 방법에 대한 설명이며, 본 분석에서는 SAS를 이용하여 로지스틱 회귀분석을 진행하였다. 자세한 일반화 선형모형과 로지스틱 모형에 대한 설명은 Goldburd, Khare, Tevet (2016), Nelder, Wedderburn (1972), 이항석, 이가은, 이경아(2017), 이항석, 이민하, 백혜연(2018)을 참고하면 된다.

2. 모형 설정 및 분석결과

본 연구에서는 반응변수(response variable)인 퇴직 사건의 설명변수(explanatory variables)로 직종, 학교급별 가입기간과 퇴직연령, 그리고 가입기간과 연령(현재 연령)의 교호작용(interaction term)을 고려한다. 현행 퇴직률 산출방법상의 두 변수인 성별과 직종에 학교급을 추가한 이유는 명백하다. 〈표 2〉와 〈그림 2〉에서 확인하였듯 학교급에 따라 퇴직자의 가입기간 및 연령이 다른

유형을 나타내는 것으로 판단되며, 특히 유치원과 대학 그리고 중·고등학교 간에는 이러한 차이가 뚜렷하기 때문이다. 또한 사학연금 직종과 학교급에 따라 교육공무원과 일반 사무직 공무원의 정년이 대다수 사학연금 퇴직자에게도 적용되는 것으로 여겨지기 때문이다. 다만 3개의 범주형 변수와 2개의 연속형 변수를 모두 고려하기에는 유치원과 대학을 제외한 나머지 사학연금 적용대상기관(초·중·고등학교, 특수학교, 기타 기관)의 퇴직자 수가 충분하지 않을 수 있다. 이에 학교급을 유치원과 대학 그리고 나머지 다른 기관의 합으로 크게 3가지 수준(level)으로 분류하였다.⁸ <그림 2>에서 살펴볼 수 있듯 해당 기관, 특히 중학교와 고등학교의 퇴직자들은 가입기간과 퇴직연령에 따른 분포에서 어느 정도의 유사성이 있다고 판단하였다.

또한 <그림 2>의 중학교와 고등학교에서는 상당한 정도로 가입기간과 퇴직연령 간 양의 상관관계가 확인되는 반면 유치원의 경우 이러한 일정한 관계를 찾기 어려운 부분이 있다. 이에 로지스틱 회귀분석 모형에 가입기간과 연령에 더해 이 두 변수 간의 교호작용 역시 고려해 보았다. 마지막으로 로지스틱 회귀분석을 위해 입력하는 기초 자료에서 교육공무원과 일반 사무직 공무원에 적용되는 정년연령 이상의 가입자와 퇴직자는 제외하였다. 이는 현행 사학연금 재정계산에서의 퇴직률 산출방법과 마찬가지로 정년연령을 반영하기 위함이다. 다시 말해 본 연구에서 로지스틱 회귀분석을 이용하여 추정하고자 하는 퇴직률은 정년 이하 연령으로 한정하는 것이다. 이에 따라 정년연령 이상의 가입자와 퇴직자 자료를 제외함으로써 해당 자료가 추정결과에 미칠 수 있는 영향을 통제하고자 하였다. 이하에서는 다양한 모형에 대한 로지스틱 회귀분석 결과를 제시하도록 한다.

먼저 현행 사학연금 퇴직률 산출방법에서와 마찬가지로 성별과 직종을 범주형 변수로 설정하여 분석해 보았다. <표 5>는 그 결과로 모든 조건이 동일할 경우 교원이 직원에 비해 퇴직률이 낮으며 연령과 가입기간이 증가할수록 마찬가지로 퇴직률이 감소함을 알 수 있다. 다만 카이제곱(χ^2) 검정에 따르면 교원의 유의확률이 통계적으로 유의한 수준이 아님을 알 수 있다. 즉, 현행 방법에서와 같이 직종과 성별만을 고려할 때 그것이 의미 있는 구분이 아닐 수 있다는 것이다. 이에 앞서 언급한 바와 같은 이유로 학교급과 교호작용을 추가로 감안하였으며 각 모형의 분석결과를 <표 6>~<표 9>에 제시하였다.

8. 실제 학교급을 3개의 범주로 가공하지 않고 원래의 학교급을 이용하여 로지 분석을 해 본 결과 통계적으로 유의한 결과의 모형을 발견할 수 없었다.

〈표 5〉 로지스틱 회귀분석 결과 : 성별 · 직종 · 연령 · 가입기간

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	1.4650	0.010300	20403.130	<.0001
직 종	교 원	1	0.0084	0.004870	2.982	0.0842
성 별	남 자	1	0.2630	0.005440	2337.538	<.0001
연 령		1	0.0152	0.000347	1920.774	<.0001
가입기간		1	0.0262	0.000392	4451.460	<.0001

〈표 6〉은 〈표 5〉의 모형에서 성별을 학교급으로 대체한 분석결과이다. 카이제곱 검정 결과는 모두 변수가 통계적으로 유의미함을 알려준다. 이에 따르면 〈표 5〉에서와 마찬가지로 다른 모든 조건이 동일할 때 직원은 교원에 비해 퇴직률이 높으며 연령과 가입기간이 증가할수록 퇴직률은 감소한다. 또한 학교급의 경우 유치원의 퇴직률이 가장 높고 기타 기관이 가장 낮은 퇴직률을 나타낸다. 이러한 결과는 〈표 6〉의 모형에 성별 또는 교호작용을 추가(〈표 7〉과 〈표 8〉)하거나 성별과 교호작용 모두 추가(〈표 9〉)한 모형에서 동일하게 나타난다.

〈표 6〉 로지스틱 회귀분석 결과[모형 1] : 학교급 · 직종 · 연령 · 가입기간

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	1.6508	0.010700	23977.188	<.0001
직 종	교 원	1	0.1314	0.005940	489.618	<.0001
학 교 급	유치원	1	-0.5835	0.007140	6677.467	<.0001
	기 타	1	0.5232	0.007990	4286.120	<.0001
연 령		1	0.0152	0.000337	2032.555	<.0001
가입기간		1	0.0130	0.000412	995.652	<.0001

〈표 7〉 로지스틱 회귀분석 결과[모형 2] : 성별 · 학교급 · 직종 · 연령 · 가입기간

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	1.6674	0.010700	24427.911	<.0001
직 종	교 원	1	0.1210	0.005980	409.152	<.0001
성 별	남 자	1	0.1037	0.005800	319.081	<.0001
학 교 급	유치원	1	-0.5443	0.007450	5333.740	<.0001
	기 타	1	0.5193	0.008000	4208.384	<.0001
연 령		1	0.0135	0.000348	1501.848	<.0001
가입기간		1	0.0134	0.000411	1057.635	<.0001

〈표 8〉 로지스틱 회귀분석 결과[모형 3] : 학교급 · 직종 · 연령 · 가입기간 · 교호작용

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	0.1633	0.011800	192.996	<.0001
직 종	교 원	1	0.4007	0.006220	4153.162	<.0001
학 교 급	유치원	1	-0.5536	0.007340	5685.157	<.0001
	기 타	1	0.2588	0.008150	1008.762	<.0001
연 령		1	0.0451	0.000361	15584.155	<.0001
가입기간		1	0.4119	0.001630	63641.309	<.0001
연령×가입기간		1	-0.0076	0.000029	70732.939	<.0001

〈표 9〉 로지스틱 회귀분석 결과[모형 4] : 성별 · 학교급 · 직종 · 연령 · 가입기간 · 교호작용

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	0.1836	0.011700	244.401	<.0001
직 종	교 원	1	0.3914	0.006250	3922.795	<.0001
성 별	남 자	1	0.1445	0.006130	556.802	<.0001
학 교 급	유치원	1	-0.5058	0.007600	4430.902	<.0001
	기 타	1	0.2525	0.008160	956.267	<.0001
연 령		1	0.0428	0.000371	13326.847	<.0001
가입기간		1	0.4121	0.001630	64039.147	<.0001
연령×가입기간		1	-0.0076	0.000029	71021.838	<.0001

4개의 모든 모형은 통계적으로 유의한 카이제곱 검정결과를 보인다. 해당 모형은 동류 모형(nested model)으로 이 경우 아카이케 정보 기준(Akaike Information Criterion, AIC)에 의존, AIC 수치가 더 작은 것을 선호하게 된다. 그러나 본 연구에서는 AIC를 이용, 그 값이 제일 작은 [모형 4]를 선택하는 것을 지양한다. 왜냐하면 연구의 목적이 '통계적으로' 우수한 모형을 판별하는 것이 아니라 '실무적으로' 사학연금 재정추계에 적합한 모형을 찾는 것이기 때문이다. 물론 다양한 모형들 중 우수한 모형을 판별한 결과에 대해서 연구자들마다 이견이 존재할 수 있기 때문에 이러한 점이 본 연구의 한계라 볼 수 있다. 본 연구에서는 실무적으로 좀 더 적합한 모형을 판별하기 위해 〈표 6〉부터 〈표 9〉까지의 분석결과에 기반한 퇴직률을 사학연금 가입자에 적용하여 추정된 퇴직자와 실제 퇴직자를 비교해 봄으로써 재정추계 목적에 보다 적합한 모형을 찾기 위한 단초를 얻고자 한다. 이 과정에서 상기의 4가지 모형 이외 각 개별집단을 대상으로 가입기간과 연령을 설명변수로 하는 로지스틱 회귀분석을 각각 추가적으로 실시하고 그 결과를 이용하여 퇴직자를 추정해 본다. 즉, 학교급, 성별 및 직종에 따라 데이터를 분류하고 각 집단의 데이터만을 사용하여 로지스틱

회귀분석을 시행하는 것이다.⁹ 이는 <표 6>부터 <표 9>의 결과를 통해 학교급이 설명변수에 포함된 로짓 모형은 모두 통계적으로 유의한 결과를 제공한다는 사실에 기반한 것으로, 실제 퇴직자 자료와 보다 유사한 결과를 제공하는 퇴직률 모형을 찾기 위한 하나의 시도이다.

<표 10> 집단별 로지스틱 회귀분석 결과[모형 5]

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
유치원 교원					
Intercept	1	2.6801	0.026700	10078.760	<.0001
연 령	1	-0.0300	0.000887	1146.202	<.0001
가입기간	1	-0.4696	0.008080	3382.545	<.0001
연령×가입기간	1	0.0135	0.000238	3227.327	<.0001
유치원 직원					
Intercept	1	0.7639	0.080500	90.117	<.0001
연 령	1	0.0099	0.001810	30.482	<.0001
가입기간	1	0.2568	0.032600	62.162	<.0001
연령×가입기간	1	-0.0032	0.000656	23.330	<.0001
기타 교원					
Intercept	1	2.6718	0.095500	782.860	<.0001
연 령	1	0.0127	0.002570	24.296	<.0001
가입기간	1	0.6459	0.005370	14449.260	<.0001
연령×가입기간	1	-0.0119	0.000099	14450.090	<.0001
기타 직원					
Intercept	1	1.1216	0.103900	116.531	<.0001
연 령	1	0.0239	0.002510	90.646	<.0001
가입기간	1	0.6392	0.012300	2687.547	<.0001
연령×가입기간	1	-0.0116	0.000221	2768.774	<.0001
대학 교원					
Intercept	1	-1.8958	0.025500	5520.766	<.0001
연 령	1	0.0906	0.000721	15803.900	<.0001
가입기간	1	0.7924	0.004920	25988.560	<.0001
연령×가입기간	1	-0.0137	0.000078	30763.550	<.0001
대학 직원					
Intercept	1	1.3029	0.019600	4432.624	<.0001
연 령	1	0.0198	0.000683	841.984	<.0001
가입기간	1	0.0436	0.000773	3187.830	<.0001

9. 이는 앞선 4개 모형과는 완전히 다른 접근이다. 상기의 4개 모형이 '사학연금 가입자'의 퇴직률 추정모형이라면 이 방법은 모수추정에 사용된 집단별 데이터의 종류에 따라 예를 들어 사학연금 '유치원 교원' 가입자의 퇴직률 추정모형, 사학연금 '대학 교원 가입자'의 퇴직률 모형 등이 된다.

〈표 10〉은 학교급과 직종에 따라 각 집단별로 얻어낸 로지스틱 회귀분석 결과이다. 학교급과 직종에 더해 성별을 추가하였을 때에는 유치원 남자 교원의 경우 유의미한 분석결과를 얻을 수 없었다. 이에 따라 학교급과 직종에 따른 개별 집단에 대해서만 가입기간과 연령을 설명변수로 하는 로지스틱 분석을 시행하였으며 가입기간과 연령의 교호작용이 있는 모형과 제거한 모형의 비교를 통해 보다 유의한 결과를 선택해 놓은 것이다. 대부분의 집단에서는 교호작용이 포함된 모형이 보다 유의하였던 반면 대학 직원의 경우는 교호작용이 제거된 모형이 선택되었다는 특징을 보인다.

〈표 11〉 모형 간 퇴직률 비교 : 가입기간 10년, 40세

학교급	직종	성별	모형 1	모형 2	모형 3	모형 4	모형 5
유치원	교원	남자	0.12601	0.11693	0.05248	0.04712	0.10119
		여자	0.12601	0.12807	0.05248	0.05405	0.10119
	직원	남자	0.14121	0.13002	0.07637	0.06816	0.07846
		여자	0.14121	0.14220	0.07637	0.07793	0.07846
기 타	교원	남자	0.04550	0.04371	0.02399	0.02264	0.00755
		여자	0.04550	0.04826	0.02399	0.02607	0.00755
	직원	남자	0.05156	0.04906	0.03540	0.03313	0.02126
		여자	0.05156	0.05413	0.03540	0.03809	0.02126
대 학	교원	남자	0.07445	0.07135	0.03086	0.02896	0.01518
		여자	0.07445	0.07854	0.03086	0.03331	0.01518
	직원	남자	0.08403	0.07980	0.04538	0.04225	0.07372
		여자	0.08403	0.08775	0.04538	0.04849	0.07372

〈표 11〉은 〈표 6〉부터 〈표 10〉까지의 분석결과에 근거하여 하나의 예시로서 가입기간 10년인 40세 재직자의 퇴직률을 계산하여 비교해 놓은 것이다. 〈표 6〉부터 〈표 10〉까지의 모형을 편의상 모형 1, 모형 2, 모형 3 그리고 모형 4와 모형 5라 하였다. 모형 1부터 모형 4까지 학교급은 공통으로 포함되어 있으며, 모형 1은 직종이 추가로, 모형 2는 다시 성별이 추가된 것이다. 모형 3과 모형 4는 모형 1과 모형 2에 가입기간과 연령 간의 교호작용이 포함된 것이다. 모형 5의 경우 다른 4개 모형과 달리 개별 집단에 따라 분석한 〈표 10〉에 근거하여 계산한 확률이다. 모형 1, 3, 5는 성별이 제거되어 있으므로 남자와 여자의 퇴직률이 동일하다. 성별이 감안된 경우(모형 2, 모형 4) 〈표 7〉과 〈표 9〉에서 언급하였듯 남자의 퇴직률이 낮음을 알 수 있다. 또한 〈표 6〉부터 〈표 9〉의 분석결과에서처럼 모형 1부터 모형 4까지 유치원, 대학, 기타 기관의 순으로 퇴직률이 높다는 것이 확인된다. 특히 개별 집단을 대상으로 분석한 모형 5에서도 이러한 학교급의 특성은 그대로 발견된다. 다만 모형 1부터 모형 4까지 (즉, 〈표 6〉부터 〈표 9〉) 교원의 퇴직률이 직원보다 낮은

것으로 나타나나 모형 5에서는 유치원의 경우 다른 모형과 달리 교원의 퇴직률이 직원보다 높다는 차이를 보인다. 또한 <표 11>로부터 확인할 수 있는 한 가지 특이한 점은 연령과 가입기간 간 교호작용의 역할이다. 모형 3은 모형 1에 교호작용을 추가한 것으로 모형 3에서의 추정 퇴직률이 모형 1에 비해 매우 낮음을 알 수 있다. 모형 2에 교호작용을 추가한 모형 4에서도 이를 확인할 수 있다. 반면 개별 집단 기반의 모형 5에서는 반드시 그러한 것은 아니다. 교호작용이 제외된 대학 직원의 경우 교호작용이 포함된 모형 1과 모형 2와 유사하게 정도의 퇴직률을 보이나 교호작용이 포함된 유치원의 경우 그 반대로 교호작용이 포함된 모형 3 및 모형 4보다 교호작용이 제외된 모형 1과 모형 2에 더 근접한 퇴직률이 산출되는 것이다. 결국 이렇듯 모형에 따라 퇴직률의 산출결과가 상당한 차이를 보이고 있음을 감안할 때 실제 사학연금 퇴직자 자료와의 비교를 통해 목적에 보다 부합하는 모형을 찾는 것이 바람직한 방안일 수 있다.

제4장 퇴직자 추계결과의 비교

본 장에서는 앞서 살펴본 총 5개의 로짓 모형의 분석결과를 적용하여 퇴직자를 추계하고 그 결과를 비교한다. 이러한 비교방법을 사용하는 이유는 5개 로짓 모형 모두 카이제곱 검정결과 통계적으로 유의한 모형으로 판단되어 그 중에서 우월한 모형을 선택하기 어려울 뿐만 아니라 통계적으로 유의한 모형일지라도 이에 따른 예측결과는 실제와 큰 차이를 나타낼 수 있기 때문이다. 이에 5개 모형과 현행 사학연금 재정추계(이하 ‘현행’)에서의 퇴직률을 이용하여 각각 퇴직자를 추계하고 이를 실측치(이하 ‘실적’)와 비교함으로써 로지스틱 회귀분석을 이용한 모형의 특징을 살펴보고자 한다.

〈표 12〉 퇴직자 수 추계 결과의 비교 : 가입기간

가입기간	모형 1	모형 2	모형 3	모형 4	모형 5	현행	실적
n ≤ 5년	13,669	13,858	16,617	17,421	16,089	19,766	14,504
5년 < n ≤ 10년	3,068	3,346	1,948	2,158	1,108	2,018	2,503
10년 < n ≤ 15년	2,179	2,585	1,289	1,456	811	769	681
15년 < n ≤ 20년	1,909	2,307	1,341	1,498	909	696	463
20년 < n ≤ 25년	1,870	2,254	1,821	2,013	1,383	1,013	691
25년 < n ≤ 30년	1,683	2,352	2,196	2,659	2,068	1,421	1,036
n > 30년	3,291	3,725	4,441	5,035	5,019	3,766	3,446
합계	27,669	30,427	29,652	32,240	27,388	29,450	23,324

〈표 13〉 퇴직자 수 추계 결과의 비교 : 퇴직연령

퇴직연령	모형 1	모형 2	모형 3	모형 4	모형 5	현행	실적
$r < 22$ 세	67	69	140	144	23	15	94
$22 \leq r < 27$ 세	4,086	3,915	5,787	5,880	4,751	5,120	5,488
$27 \leq r < 32$ 세	3,781	3,796	4,345	4,573	3,732	6,509	5,047
$32 \leq r < 37$ 세	2,896	3,298	2,992	3,313	2,807	4,598	3,167
$37 \leq r < 42$ 세	2,640	3,151	2,068	2,322	2,090	2,226	1,502
$42 \leq r < 47$ 세	2,644	3,101	1,732	1,942	1,522	1,302	994
$47 \leq r < 52$ 세	2,319	2,798	1,552	1,804	1,259	1,104	881
$52 \leq r < 57$ 세	2,051	2,732	2,025	2,570	1,471	1,336	1,240
$57 \leq r < 62$ 세	2,853	3,243	4,030	4,680	4,024	3,344	2,300
$r \geq 62$ 세	4,333	4,326	4,982	5,013	5,709	3,896	2,611
합계	27,670	30,427	29,652	32,240	27,388	29,450	23,324

퇴직자 추계에는 2016년도 말 현재 가입자 313,156명과 2017년의 신규 가입자 30,623명 등 총 343,779명의 가입자가 사용되었으며, 이들을 대상으로 2017년 퇴직자를 추계하였다. 이에 대한 실적 자료로는 2017년에 퇴직하였으며 급여를 신청한, 사학연금 가입자였던 자 23,324명이다. 〈표 12〉와 〈표 13〉은 각 비교대상과 실제 퇴직자를 가입기간과 퇴직연령에 따라 비교해 놓은 것이다.¹⁰ 먼저 퇴직자 수에 있어 현행 퇴직률과 로짓 모형에 따른 퇴직률 모두 실측치보다 높은 추정치를 나타낸다는 특징이 있다. 실측치에 있어 한 가지 유의할 사항은 사학연금법상 퇴직급여의 시효는 퇴직 후 5년인 바, 2017년 퇴직자의 경우 2021년까지 급여를 신청할 수 있다는 것이다. 비교대상인 23,324명은 2017년에 퇴직하고 동일 연도에 급여를 신청한 자들이므로 향후 급여를 신청할 모든 퇴직자를 포함하는 것은 아니다. 2017년에 퇴직하여 급여를 신청하는 퇴직자의 정확한 수치는 향후 미래에 확인가능하며, 그 규모는 23,324명보다 클 것으로 예상된다.¹¹ 한 예로 본 연구에 사용된 퇴직자 자료 중 퇴직급여의 신청 시효가 만료된 2011년과 2012년 퇴직자 자료에 따르면 2011년 퇴직하고 급여를 신청한 총 28,489명 중 퇴직연도인 2011년에 87.3%가 신청하였으며 2012년의 경우 이 수치는 90.1%로 확인된다. 이 비율을 2017년 실측치에 적용하면 퇴직 직후 당해에 바로 급여를

10. 현 사학연금공단의 재정추계시스템은 퇴직자 수 추계결과를 가입기간과 퇴직연령으로만 분류할 수 있어 성별, 학교급, 직종에 따른 퇴직자 수 비교는 제시가 불가하였다.

11. 퇴직급여의 시효를 고려할 때 2010년 또는 2011년 가입자 등 현재보다 5년 이상 이전의 자료를 사용하여 퇴직자를 추계해 보고자 하였으나 2015년 사학연금법 개정 이후 기간에 대해서만 추계가 가능한 사학연금공단 사정에 따라 2017년 퇴직자를 추계하였다.

신청한 퇴직자가 각각 26,717명과 25,887명으로 예상된다. 그렇다 할지라도 모든 모형은 이러한 예상치보다 높은 추정치를 나타낸다. 다만 직종과 학교급을 범주형 변수로 삼는 모형 1(27,669명)과 직종 및 학교급 구분에 따른 개별 집단별로 로짓 모형을 구축하여 추계한 모형 5(27,388명)가 그 중 가장 유사한 수의 예측치를 나타낸다.

실측 퇴직자(실적)를 5년 간격의 가입기간과 5세 간격의 퇴직연령으로 분류할 때 나타나는 한 가지 특징은 5년 이하의 가입기간을 지닌 퇴직자가, 그리고 22세 이상 27세 미만의 젊은 퇴직자가 큰 비중을 차지하고 있다는 점이다. 사학연금 가입 시의 연령이 전 연령구간에 있어 고르게 퍼져있지 않는 한 가입기간이 짧으면 젊은 퇴직자가 많을 것임을 쉽게 유추해 볼 수 있다. 실제 퇴직자들의 가입기간과 퇴직연령에서 나타나는 이러한 특징이 각 모형에 얼마나 잘 나타나 있는지를 비교하기 위해 각 모형별로 전체 퇴직자 수 대비 가입기간 또는 퇴직연령 구간별 비중을 나타내 보았다. <표 14>와 <표 15>가 바로 그것이다. 이에 따르면 현행 퇴직률 모형과 개별 집단에 로짓 모형을 적용한 모형 5가 실측치와 가장 유사한 패턴을 나타내는 것으로 판단된다. 그러나 모형 5의 경우 가입기간에 있어 5년 이하 퇴직자가 실측치에 비해 상대적으로 적은 비율이 추계되고 그 결과 나머지 가입기간 구간에서 실측치에 비해 높은 비율을 나타내고 있다. 이러한 점은 현행 퇴직률 모형(현행 재정추계)도 유사하게 나타나는 바, 다만 현행 모형은 가입기간 5년 이하 퇴직자를 매우 크게 추계하고 있다는 차이가 있다. 또한 퇴직연령에서의 구간별 비율을 보여주고 있는 <표 15>에서도 이와 매우 유사한 패턴이 발견된다.

<표 14> 구간별 비율 비교 : 가입기간

가입기간(n)	모형 1	모형 2	모형 3	모형 4	모형 5	현행	실적
n ≤ 5년	49.40%	45.55%	56.04%	54.04%	58.75%	67.12%	62.18%
5년 < n ≤ 10년	11.09%	11.00%	6.57%	6.69%	4.05%	6.85%	10.73%
10년 < n ≤ 15년	7.88%	8.50%	4.35%	4.52%	2.96%	2.61%	2.92%
15년 < n ≤ 20년	6.90%	7.58%	4.52%	4.65%	3.32%	2.36%	1.99%
20년 < n ≤ 25년	6.76%	7.41%	6.14%	6.24%	5.05%	3.44%	2.96%
25년 < n ≤ 30년	6.08%	7.73%	7.41%	8.25%	7.55%	4.83%	4.44%
n > 30년	11.89%	12.24%	14.98%	15.62%	18.33%	12.79%	14.77%
합 계	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

제5장 결론

본 연구는 사학연금 재정추계 목적의 퇴직률 산출을 위한 하나의 대안으로 일반화 선형모형의 로지스틱 회귀분석을 적용해 보았다. 이는 현재 사학연금 재정추계에서 사용하고 있는 퇴직률 산출과정에서의 어려움을 해결하기 위함이다. 현행 방법에 따르면 직종 및 성별에 따라 가입자와 퇴직자를 4개 집단으로 구분하고 각 집단별로 가입기간 및 가입연령별 가입자 수 대비 퇴직자 수로 정의되는 퇴직률을 산출, 보정한 후 연내 퇴직자가 균등하게 분포되어 있다는 가정 하에 이를 퇴직률로 전환하고 있다. 이 과정에서 퇴직자 수가 존재하지 않거나 존재하더라도 그 수가 많지 않은 가입기간 및 가입연령 구간이 상당 부분 존재하여 퇴직률 산출이 어렵거나 보정을 거치더라도 그 결과의 신뢰성을 확보하기 어려운 단점이 존재한다. 또한 산출방법의 적용에 앞서 직종 및 성별에 따른 집단 구분의 근거가 미약하다는 문제가 있다. 이에 본 연구에서는 로지스틱 회귀분석을 통해 이러한 현행 산출방법의 약점을 극복해 보고자 하였다.

로지스틱 회귀분석 결과에 따르면 성별과 직종에 따른 집단 구분은 통계적으로 유의미하지 않은 것으로 나타난다. 또한 성별과 직종으로 집단을 구분할 경우 실제 사학연금 퇴직자가 나타내는 직종 및 학교급에 있어 정년연령을 반영하기 어렵다. 이에 본 연구에서는 학교급을 범주형 설명변수의 하나로 추가하되 유치원과 대학 그리고 기타 나머지 기관의 3개 범주로 나누었다. 그 결과 모든 로짓 모형에서 통계적으로 유의미한 분석결과를 얻었으며 학교급을 설명변수로 할 경우 성별이 추가되어도 모형의 유의성에는 변화가 없음을 확인하였다. 또한 본 연구결과로부터 현행과 같은 퇴직률 산출방법을 유지할 경우 집단 구분에 학교급을 고려하는 방법을 시도해 볼 필요가 있음을 확인할 수 있었다. 또는 보다 복잡한 과정을 거쳐야 하겠지만 가입기간 및 퇴직연령 구간에 따라 보다 적합한 모형을 취사선택하는 혼합(mixture) 모형도 하나의 대안이 될 수 있을 것으로 여겨진다. 그러나 실측치와의 비교 시 현행 방법 및 검토 대상의 모든 로지스틱 회귀분석 결과에 의한 퇴직자 추계는 상당 부분 과대 추계의 문제가 발생하는 것으로 여겨진다. 특히 실측치에 따르면 가입기간이 짧고 퇴직연령이 낮은 구간에서의 퇴직자 비중이 매우 크고 상대적으로 중간 구간의 퇴직자 비중이 작았던 바, 로지스틱 회귀모형에 의한 퇴직자 추계는 현행 퇴직률보다 이러한 특징을 보다 잘 반영하지 못하는 것으로 파악된다. 다만 실제와의 비교에 있어 현실적 제약에 따라 2017년 퇴직자

실측치만을 비교 대상으로 삼았던 바, 퇴직자 수 추계의 과대 추계 등 상기한 바와 같은 적합도의 문제는 추후 보다 큰 규모의 실제 퇴직자 자료와의 비교 검토가 반드시 필요할 것이다. 사학연금 재정추계는 70년 가량의 초장기 기간을 추계대상으로 삼고 있으므로 한 해의 실측치와의 비교만을 통해 모형을 선택할 경우 전반적이며 평균적인 퇴직 사건의 추세를 왜곡할 수 있다. 따라서 장기 재정추계에 적용하기 위해서는 합리적인 퇴직률 산출 모형 구축에 좀 더 고민할 필요가 있다. 비록 본 연구가 사학연금 재정추계를 위한 완전한 퇴직률 산출방법을 제공하지는 못하였으나 본 연구의 결과가 향후 보다 나은 대안 마련의 발판이 되길 바란다.

참고문헌

- 국민연금재정추계위원회 (2013), 「국민연금 장기재정추계」, 「국민연금 제도 및 기금운용 개선방안」.
- 김병률 · 김재경 (2016), 「사학연금 재정계산제도 개선방향」, 「사학연금연구」, 제1호, 189-241.
- 김순옥 · 권문일 (1998), 「각국의 재정계산제도 비교연구」, 국민연금연구센터.
- 사학연금재정재계산위원회 (2016), 「사학연금 장기재정추계」, 사학연금재정계산 보고서, 사립학교교직원연금공단.
- 윤석명 · 강성호 (2001), 「국민연금 재정계산제도 실행방안」, 국민연금연구센터.
- 이항석 · 이가은 · 이경아 (2017), 「실손의료보험에서 요율산정과 일반화선형모형(GLM)의 활용」, 리스크관리연구, 28(4): 37-71.
- 이항석 · 이민하 · 백혜연 (2018), 「실손의료보험 할인할증제도의 실증분석」, 「보험학회지」, 116 : 33-66
- 한국개발연구원 (2016), 「공무원연금 재정추계 기초율 개선 연구」.
- 한국보건사회연구원 (2017), 「사학연금 사망률 전망 및 가입자 추계」.
- 한국보건사회연구원 (2018), 「공무원 충원 계획에 따른 공무원연금 가입자 수 전망모형」.
- Agresti, Alan. (2002), Categorical Data Analysis (2nd ed.), Hoboken, NJ: John Wiley & Sons.
- Goldburd, M., Khare, A. and Tevet, D. (2016), Generalized Linear Models for Insurance Rating, Casualty Actuarial Society.
- McCullagh, P. and Nelder, J. A. (1989), Generalized Linear Models (2nd ed.), London: Chapman & Hall.
- Nelder, J. A. and Wedderburn, R. W. M. (1972), Generalized Linear Models, Journal of the Royal Statistical Society Series A(General), 135(3): 370-384.
- SAS Institute Inc. (2008), SAS/STAT 9.2 User's Guide, Cary, NC.