

생존분석을 통한 구직기간 결정요인 연구

최지현*

2018년 기준 한 해 동안 구직활동에 참여한 사람은 전체 취업자의 약 1/3로 파악되었다. 구직 행위는 개인 차원에서 더 나은 자리로 나아가는 기회이자 도전이다. 그러나 사회 전반적으로 구직기간이 장기화될 경우 사회경제적 손실이 늘어나고, 다양한 사회 문제를 초래하게 된다. 본고는 다양한 개인적 특성이 구직기간의 차이를 결정한다고 보고 12년치(2008~2019) 한국노동패널조사(KLIPS)의 개인 단위 데이터를 바탕으로 생존분석을 수행하였다. 생존분석 기법으로 개별 범주의 생존함수를 비모수적으로 추정하는 카플란-마이어(K-M) 분석법과 함께 여러 요인들을 고려 가능한 Cox 비례위험 모형을 사용하되, 가속실패시간(AFT) 모형을 추가로 도입하여 비례위험(PH) 가정이 위배되는 경우를 극복하였다. 분석 결과, 로그-정규 분포를 가정한 AFT 모형의 적합도가 가장 높게 나타났다. 그리고 여성이 남성보다, 비수도권 거주자가 수도권 거주자보다, 사회보험급여 비수급자가 계속 및 당해 수급자에 비해, 그리고 고졸자가 대졸자가 비해 구직기간이 상대적으로 짧은 것으로 나타나 일부 결과는 통념과 달랐다. 하지만 생존분석을 통해 집단간 구직기간 차이를 결정짓는 구체적 배경 원인을 확인할 수 없으며, 취업의 질적 측면을 설명하지 못하기 때문에 해석에 있어 주의가 요구된다. 아울러 생존분석은 내생성 문제를 내포하고 있기 때문에 변수들간 인과관계보다는 상관관계로 이해하는 것이 보다 적절하다.

주요 용어 : 구직기간, 구직, 취업, 생존분석, AFT 모형

* 한국노동연구원 연구보조원. 경제학 석사.

1. 서론

구직활동은 단순히 실업 상태에서 이루어질뿐만 아니라 현업에 종사 중이지만 임금, 직급, 근로 환경 혹은 적성 등 여러가지 이유로 이직을 모색하는 활동까지 아우른다. 통계청이 발표한 일자리 이동통계(2018년 기준)에 따르면 한국의 등록취업자(임금 근로자 및 비임금 근로자) 인구는 약 2,384만 명으로 집계되었다(통계청, 2020). 그리고 이 중에서 전년(2017년)과 비교해 일자리를 새로 얻은 사람은 약 392만 명, 일자리를 옮긴 사람은 약 394만 명으로 파악되었다. 두 수치를 합치면 전체 취업자의 1/3 정도 되는 비중으로 결코 적지 않은 사람들이 한 해 동안 구직활동에 참여했음을 보여준다. 일반적으로 구직자들은 구직경쟁에서 유리한 위치를 점하기 위해 상위 교육기관 진학, 직업훈련, 자격증 취득 등의 노력을 보인다. 그렇지만 구직은 단순하게 설명되는 문제가 아니다. 거주지, 혼인 여부, 가족관계 등의 생활 환경상 요인, 성별 등의 태생적 요인, 나이와 신체 건강 같은 불가항력적 요인, 심리 상태, 그리고 가치관과 내적 동기 등이 복합적으로 작용하기 때문이다. 따라서 개인마다 새 일자리를 얻는 데까지 걸리는 기간에 있어 분명 차이를 보일 것이다.

이에 따라 본고는 구직기간 차이를 결정짓는 다양한 개인적 요인들을 구체적으로 검증하고자 한다. 이를 위해 한국노동패널조사(KLIPS)가 제공하는 다년도 개인 단위 데이터를 바탕으로 생존분석을 수행하였다. 본문은 다음과 같이 세 부분으로 구성된다. 제 2장은 국내외 선행연구를 통해 구직기간 혹은 구직 성공에 영향을 주는 요인의 유형들을 확인하고자 한다. 제 3장은 본격적인 분석에 앞서 생존분석 이론의 요점에 대해 대략적으로 개괄하고, KLIPS 데이터에서 무엇을 선별해 변수로 구성했는지, 그리고 실증 모형을 어떻게 설정하였는지 설명하고자 한다. 제 4장은 분석 결과를 정리한 부분으로, 설정한 모형을 바탕으로 생존분석을 수행한 후 개인 특성에 따라 구분된 구직자 그룹간 구직기간의 차이가 확인되었는지, 또 각 특성이 구직기간에 어느 정도로 유의한 효과를 나타냈는지 자세히 밝혔다.

II. 선행연구 검토

인구 집단의 구직활동은 국가 경제에 파급 효과를 미칠뿐만 아니라 직간접으로 다양한 사회문제를 유발하므로 여러 사회과학 분야에서 중요하게 취급되는 연구 주제다. 아울러 글로벌 금융위기 이후 잠재성장률 하락, 무역마찰 심화, 자국 우선주의 기류 확산 등에 따른 불확실성으로 인해 고용침체가 심해지는 상황에서 각국 정부에서 구직자를 대상으로 많은 행재정적 지원책을 시행하였다. 이후 그 효과성을 검증하고자 다양한 평가가 이루어졌다. 따라서 국내외적으로 관련 연구 결과가 많이 축적되었다.

여기서는 구직 혹은 구직기간에 영향을 미치는 요인을 주제로 다룬 연구들을 소개한다. 일반적으로 전체 구직자보다는 대졸자, 청년, 여성, 노인, 장애인 등 특정 인구집단을 표적으로 삼은 연구가 주를 이룬다. Miyamoto·Suphaphiphat(2020)는 25개 유럽 국가들을 대상으로 2000년부터 2018년까지의 장기실업자 데이터를 바탕으로 실업 유발 요인과 대응 정책에 대해 분석하였다. 그 결과, 능력상 미스매치, 노동시장의 미스매치, 매칭 과정에서의 효율성, 그리고 노동시장 정책이 장기실업을 초래하였음을 확인하였다. 또한 구직자에 대한 직업훈련 및 창업 인센티브등의 정책적 요인이 장기실업을 낮추는데 효과가 있음을 지적하였다. Berg·Lomwel·Ours(1998)는 프랑스의 연령별 실업자를 연구하였는데, 결과적으로 경기변동 요인이 중장년층보다 청년층에 더 큰 영향을 주었다고 시사했다. Caliendo·Tatsiramos·Uhlendorff(2009)는 비레위험 모형을 통해 독일의 실업보험(UI) 혜택 증가에 따른 영향을 분석하였으며, 오히려 실업보험이 구직자들의 실업 기간을 늘렸다고 지적하였다. 노르웨이의 Bratberg·Espen·Nilsen(1998)은 고학력일수록, 그리고 여성일수록 구직기간이 상대적으로 짧았고 재직기간이 길었음을 밝혔다. 다만 취업한 여성은 남성보다 임금이 낮은 것으로 나타났다. 안태현·고영근(2016)은 개인의 성격, 태도, 가치관같은 비인지적 요인이 구직에 영향을 준다고 가설을 세우고 한국노동패널의 부가조사 데이터를 분석하였다. 분석 결과에 따르면 big5 성격 요인 중 개방성, 성실성, 외향성, 신경증, 그리고 호혜성 지표가 취업 성공률과 밀접한 관련이 있다고 결론지었다. 유사한 연구를 진행한 임찬영(2018)은 big5 중 성실성 수준이 높을 경우 오히려 무기력과 심리적 위축, 불안감에서 벗어나기 어려워지므로 결단성 결여를 야기해 실업 탈출이 지연된다고 지적하였다. 아울러 이상록(2000)은 효과적인 구직방법과 구직 태도가 구직성공에 유의미한 영향을 준다고 보고 질적 요인의 중요성을 강조하였다.

한편 청년실업의 만성화가 주요 사회문제로 떠오르다 보니 신규 졸업자 내지는 청년층에 초점을 맞춘 연구가 상당 부분을 차지한다. Lynch(1985)는 영국의 청년을 대상으로 비레위험 모형을 통해 취업 결정요인을 분석해보았는데, 남성일수록, 백인종일수록, 고학력일수록 구직 성공확률이 높은 것으로 나타났다. 미국의 청년 코호트 자료(NLSY)를 분석한 Eckstein·Wolpin(1995)는 구직 보조금이 구직기간을 증가시키는 효과가 있음을 확인하였다. Berg·Ours(1999)는 비모수적 추정법을 통해 프랑스 청년 실업자 데이터를 분석하였는데 구직기간이 길어질수록 남성보다 여성이 실업 탈출확률이 낮아짐을 확인하였다. 국내에도 유사한 연구가 다수 진행되었다. 한국가구패널조사(KHPS) 자료를 분석한 조우현(1995)는 저소득가구일수록, 낮은 학력 수준일수록 청년들이 실업 탈출에 상대

적으로 어려움을 겪는다는 것을 입증하였다. 채창균·김태기(2009)는 청년들의 취업에 있어 부모의 사회경제적 지위와 가구소득이 유의미한 영향을 준다는 것을 확인하였다. 강주연·오유·김기승(2015)는 대졸자직업이동경로조사(GOMS) 자료를 가지고 카플란-마이어 및 Cox의 비례위험 모형을 통해 취업 소요기간을 분석하였는데, 구직기간이 길어질수록 취업률이 낮아지고, 자격증과 학점 같은 일부 인적자본 요소들이 구직기간을 단축시키는 효과가 있음을 지적하였다. 이규용·김용현(2003)은 대학생의 미취업 탈출확률에 가장 유의미한 영향을 주는 요소는 대학 소재지라고 밝혔다. 류장수(2005) 역시 지방대학이 수도권대학 졸업자보다 취업 소요기간이 짧고 정규직 취업률이 높았음을 확인하였다. 한국교육고용패널(KEEP)을 연구한 김민석·안은비·유경숙·정호숙(2018)은 졸업자의 전공별로 취업 소요기간에 차이가 있음을 증명하였다. 한편 한국노동패널(KLIPS)의 청년용 부가조사 자료를 분석한 이병희(2002)는 직업교육훈련을 받은 청년에게 구직기간이 짧아지는 효과가 있음을 보였다. 반면에 대졸자 직업이동 경로조사 자료를 분석한 성지미, 안주엽(2012)는 과도한 직업교육훈련이 오히려 취업에 부정적인 영향을 미쳤음을 시사하였다.

아울러 고령화 진전과 함께 사회적 약자에 대한 지원책이 늘어나면서 노인과 장애인을 대상으로 하는 연구도 늘어나는 추세다. 성지미·안주엽(2006)은 한국노동패널의 중고령자 부가조사 자료에 대한 분석을 통해 연령, 교육수준, 혼인상태, 건강, 기타 소득원, 자산 및 부채, 일자리 특성 등이 중고령자의 취업 결정에 큰 영향을 미친다는 것을 밝혔다. 독일, 미국, 스웨덴, 한국 등 4개국을 비교 연구한 장지연·신현구(2008)는 한국과 미국에서 소득 수준이 낮을수록 고령자의 경제활동참여율이 높다는 것을 지적하였다. 그러나 높은 학력 수준은 취업에 도움을 주지 못한 것으로 나타났음을 확인하였다. 국민노후보장패널(KRels)을 연구한 송일호·박명호(2012)는 중고령자가 가구주이고, 배우자가 없으며, 건강 수준이 나쁘면 취업확률이 낮아지며, 성별이 여성일 경우와 가구 소득과 지출이 많을수록 취업확률이 높아진다는 것을 실증하였다. 그러나 연금소득은 취업에 유의미한 영향을 주지 못한 것으로 나타났다. 한편 장애인의 취업률 자료를 분석한 전이상(2002)은 장애 요인과 장애 정도가 취업에 유의미한 영향을 미쳤음을 밝혔다. 나아가 여성보다 남성 장애인이 취업 가능성이 더 높았음을 확인하였다. 장애인고용패널(PSED) 자료 바탕으로 생존분석을 수행한 홍성표·정진철(2015)은 자격증을 취득한 장애인이 그렇지 않은 장애인보다 실업 상태에서 빨리 벗어날 수 있다고 결론내렸다. 동일한 자료를 분석한 강봉석·염동문(2015)는 경증 장애일수록, 근로외 소득이 낮을수록, 그리고 주관적 사회경제적 지위가 높을수록 취업이 빨랐다는 것을 증명하였다. 한편 여성 장애인의 경제활동실태조사를 분석한 양정빈(2015)은 가구주이며, 기초생활수급자가 아닌 장애인이 오히려 취업 가능성이 높다는 것을 확인하였다.

이상으로 구직자를 전체, 청년 및 신규 대졸자, 노인, 장애인 집단으로 구분하여 여러 선행연구 성과를 살펴보았다. 결과적으로 인구 사회학적 요인, 인적자본 요인, 신체적 요인, 심리적 요인, 제도적 요인 등이 복합적으로 구직활동에 영향을 미쳤다는 것을 확인할 수 있으며, 동일한 요인이지만 연구에 따라 그 영향의 방향성이 다르게 나타나는 경우도 있었다. 아울러 구직기간을 다룬 연구는 생존분석 기법, 특히 비례위험(PH) 모형을 사용해 분석한 경우가 많았다. 후술하겠지만 비례위험 모형이 높은 적합성을 갖기 위해서 시간에 따라 위험비가 일정하다는 비례위험 가정을 만족

해야 한다. 그러나 많은 연구에서 해당 가정을 충족하는지 검증하는 절차가 생략되는 경우가 많아 엄밀성에 대한 의구심이 든다. 본 연구 역시 생존분석 기법을 사용하였다. 다만 비례위험 가정에 대한 검정을 수행하였고, 만족하지 않은 것으로 나타나 대안적 분석법인 가속실패시간(AFT) 모형을 추가로 도입하여 분석 결과의 적합성을 높였다.

III. 분석 방법 및 분석 자료, 모형 설정

1. 분석 방법

가. 생존분석 개념

본고에서 다양한 개인적 특성이 구직기간에 미치는 영향을 검증하기 위해 생존분석(survival analysis) 기법을 활용하였다. 생존분석은 ‘관찰 대상’이 특정 ‘사건(event)’을 경험할 때까지 걸리는 시간(time-to-event), 즉 ‘생존시간(survival time)’을 모형화하여 시간 흐름에 따른 사건 발생확률(사망확률)을 추정하고, 사건 발생에 유의미한 영향을 주는 요인들을 밝히는 통계분석 기법이다. 따라서 본 연구의 관심 주제인 구직기간은 생존시간에 해당되며, 구직자는 관찰 대상, 취업 성공은 사건이다.

생존분석 데이터는 일반적인 통계 자료와 구별되는 특징들을 갖는다. 이상적인 연구 결과를 얻기 위해 모든 관찰 대상에게 사건이 발생할 때까지 관찰해야 한다. 그러나 보통 생존분석 데이터는 중도절단(censoring)이 빈번하게 발생한다. 이는 일종의 결측치로 탈락, 이사, 사망, 기타 다양한 사유에 따른 추적 실패로 인해 관찰기간 동안 관찰 대상에게서 사건이 발생이 확인되지 않은 것을 의미한다. 현실적으로 관찰자는 연구 대상을 일정 기간 이상 관찰할 수 없으므로 미리 연구 종료 시점을 설정한다. 연구 종료 시점 혹은 그 이전 어느 시점까지 관찰 대상자가 사건을 경험하지 않은 것을 알고 있으나, 이후로 사건 발생 여부를 알 수 없는 경우를 우측 중도절단(right-censoring)이라 일컫는다. 반면 좌측 중도절단은 관찰 시작 시점을 모르는 경우에 발생한다. 거의 모든 생존분석 데이터는 우측 중도절단된 형태에 해당된다.

생존분석 데이터의 또 다른 특징은 생존시간은 항상 양수 값을 가지며 정규분포를 따르지 않는다는 점이다. 사건이 발생할 때까지 관찰 대상간 시간 차이가 존재하고 연구 종료 시점까지 사건 발생을 확인할 수 없는 중도절단으로 인해 생존시간은 일반적으로 오른쪽으로 긴 꼬리를 가진 분포를 나타낸다. 따라서 정규분포를 전제로 하는 통계분석 기법(평균차이 검정, 분산분석, OLS 선형 회귀분석 등)은 사용하기 적절치 않다.

나. 생존함수 및 위험함수

상술했듯이 생존분석의 핵심은 관찰 대상의 생존시간을 모형화한 것이다. 분석의 결과로 생존시간의 함수인 ‘생존함수(survival function)’와 ‘위험함수(hazard function)’를 추정할 수 있다. 생존시간을 나타내는 확률변수를 $T(T>0)$, 사건이 발생하는 특정 시점을 $t(t \in T, t \geq 0)$ 라 한다면 생존함수 $S(t)$ 는 t 까지 생존확률(사건이 발생하지 않을 확률)을 의미한다. 특정 시점 t 에서 사건이 발생할 확률을 나타내는 확률밀도함수를 $f(t)$, t 까지 사건이 발생할 확률을 의미하는 누적확률분포함수를 $F(t)$ 라 한다면 생존함수 $S(t)$ 는 다음과 같이 정의된다.

$$F(t) = \Pr(T \leq t) = \int_0^t f(t)dt$$
$$S(t) = \Pr(T > t) = 1 - F(t)$$

한편 위험함수 $h(t)$ 는 시점 t 직후 사건이 발생할 확률(위험, hazard)을 의미한다. 따라서 t 까지 생존하였다는 조건 아래 시점 $t+\Delta t$ (Δt 는 매우 짧은 시간)에 사건을 경험하게 되는 조건부 확률로 표현할 수 있다.

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr[t + \Delta t > T > t \mid T > t]}{\Delta t}$$
$$= \frac{1}{S(t)} \cdot \frac{d}{dt} F(t) = \frac{f(t)}{S(t)}$$

그리고 시점 t 까지 누적 위험률을 의미하는 누적위험함수 $H(t)$ 및 생존함수 $S(t)$, 누적확률분포함수 $F(t)$, 확률밀도함수 $f(t)$ 와 다음과 같은 관계를 갖는다. 결론적으로 데이터를 통해 생존함수만 추정하면 관계식에 따라 다양한 함수들을 도출할 수 있다.

$$H(t) = \int_0^t h(t)dt = \int_0^t \frac{f(s)}{S(s)}ds = -\log S(t)$$
$$S(t) = \exp^{-H(t)}$$
$$F(t) = 1 - \exp^{-H(t)}$$
$$f(t) = h(t) \cdot \exp^{-H(t)}$$

다. 카플란-마이어 분석법(비모수적 방법)

카플란-마이어(Kaplan-Meier, 이하 K-M) 분석법은 생존함수가 어떤 이론적 분포를 띠지 않는다는 전제 하에서 특정 관찰 대상의 생존시간을 바탕으로 생존확률을 추정하는 비모수적(non-parametric) 접근법이다(Kaplan·Meier, 1958). 이 방법에 따르면 생존시간에 대한 모수적 가정 없이 각 시점에서 생존함수를 추정한다. 관찰 시작 시점부터 t번째 시점까지의 추정 생존함수 $\hat{S}(t)$ 는 각 시점의 생존비 p_i 를 곱하여 도출할 수 있다. 구간 생존비 p_i 는 시점 i에서 사건 비경험자 수(생존자 수) n_i 와 사건 경험자 수(사망자 수) d_i 의 차이를 n_i 으로 나눈 값과 같다. 특정 시점에서 중도절단된 관찰 대상자가 발생한 경우 생존자 수에 포함하지 않으며, 해당 시점의 p_i 은 계산하지 않는다. 다만 다음 p_i 를 계산할 때 중도절단된 수만큼 n_i 를 감소시키는 방식으로 중도절단 데이터가 생존확률에 미치는 영향을 반영할 수 있다. 이상 설명한 $\hat{S}(t)$ 도출 과정은 아래 식과 같이 정리할 수 있다.

$$p_i = 1 - \frac{d_i}{n_i} = \frac{n_i - d_i}{n_i}, i = 1, 2, \dots, t$$

$$\hat{S}(t) = \prod_{i \leq t} \left(1 - \frac{d_i}{n_i}\right) = p_1 \times \dots \times p_{t-1} \times p_t = S(t-1) \times p_t$$

도출된 $\hat{S}(t)$ 를 바탕으로 생존확률 곡선을 그릴 수 있는데 사건 경험자(사망자)가 발생할 때마다 구간 생존비 p_i 가 감소하므로 계단(step) 모양의 후하향 생존곡선을 얻을 수 있다.

아울러 K-M 분석법은 관찰 대상들을 독립된 집단(범주)로 구분하여 집단 간 생존함수의 차이를 파악할 수 있다. 그 차이가 통계적으로 유의미한지 확인하기 위해 주로 로그-순위 검정법(log-rank test)이 이용되는데, 이 방법은 각 집단의 관찰된 사건 경험자 수와 예측된 사건 경험자 수를 바탕으로 비교한다. 가령 분석 대상이 되는 집단 수가 2개이며, 각 생존함수를 $S_1(t)$, $S_2(t)$ 라 한다면 귀무가설은 ‘모든 시점에서 두 집단간 생존함수의 차이는 없다’($H_0: S_1(t) = S_2(t)$)이다. 두 집단에 대한 로그-순위 검정통계량(X^2_{LR})의 계산식은 다음과 같이 나타낼 수 있다. E_1 와 E_2 는 생존함수를 통해 추정된 각 집단의 기대 사망자 수이며, n_{1i} 와 n_{2i} 는 i 시점에서 각 집단의 생존자 수이다. O_1 와 O_2 는 관찰을 통해 확인된 각 집단의 실제 사망자 수다.

$$E_1 = \sum \frac{d_1 n_{1i}}{n_i}, E_2 = \sum \frac{d_2 n_{2i}}{n_i}$$

$$X^2_{LR} = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

라. Cox 비례위험 모형(준모수적 방법)

범주형 변수만 고려하여 집단간 생존함수 차이를 확인하는 K-M 분석법과 달리 Cox 비례위험(Cox proportional hazards, 이하 Cox PH) 모형은 사건 발생에 잠재적인 영향을 미칠 수 있는 여러 변수(연속형 및 범주형 변수)들을 포함 가능한 분석법이다. 그러므로 다양한 개인 특성들 중 구직기간에 유의미한 영향을 보이는 것이 무엇인지 확인하려는 본 연구의 목표에 부합한다. Cox PH 모형의 설명변수는 보통 공변량(covariate)이라 일컬어지며, 종속변수는 시간 흐름에 따른 위험률(사건 발생확률)이다. 일반적인 Cox PH 모형은 다음과 같이 정의된다(Cox, 1972).

$$\begin{aligned}h(t) &= h_0(t) \cdot e^{(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)} \\ &= h_0(t) \cdot \exp(X'B)\end{aligned}$$

위 식에서 X는 p개의 공변량이 포함된 열 벡터, B는 공변량과 선형결합하는 회귀계수의 열 벡터다. $h_0(t)$ 는 기저 위험함수(baseline hazard function)로 특정 시점 t에서 모든 공변량 값이 0일 경우 갖게 되는 기본적 위험인데, 시간이 지남에 따라 값이 변화하는 시간 의존적(time-dependent) 함수이며, 함수 분포에 관한 가정이 주어지지 않은 비모수적 부분이다. 위험함수 h(t) 역시 시간 의존적인 함수이므로 $h_0(t)$ 의 영향에 따라 비모수적이다. 반면에 $\exp(X'B)$ 는 시점 t 변화에 따라 값이 달라지지 않는 상수 부분으로, 공변량 X의 값이 주어질 때 기본적 위험에 미치는 추가적인 영향으로 이해할 수 있으며, 시점 t에서 기저위험 함수 $h_0(t)$ 의 높낮이를 결정한다. 그리고 회귀계수 벡터 B는 추정이 요구되는 모수다. 이에 따라 위험함수 분포에 대한 가정 없이 위험함수의 모수를 추정해야 하므로 Cox PH 모형을 준모수적(semi-parametric) 모형이라고도 일컬어진다.

한편 분석 대상이 되는 두 개체(집단) i, j가 존재한다고 가정해볼 때 개체(집단)간 위험비(hazard ratio, 이하 HR)는 다음과 같이 표현할 수 있다.

$$\begin{aligned}HR &= \frac{h_i(t)}{h_j(t)} = \frac{h_0(t) \exp(X_i' B)}{h_0(t) \exp(X_j' B)} \\ &= \frac{\exp(X_i' B)}{\exp(X_j' B)} = \exp[(X_i' - X_j') B] \\ &= \lambda (\text{constant})\end{aligned}$$

HR은 두 개체(범주)간 위험이 상대적으로 어느 정도 차이나는지를 나타내는 상대적 위험도(relative risk)로 해석할 수 있다. 또한 HR은 시간의 흐름과 무관하게(time-independent) 일정하며 두 개체의 공변량 차이에 비례하는 값(상수 λ)을 가진다. 이를 비례위험(proportional hazard, 이하 PH)이라고 하는데, Cox PH 모형이 높은 적합성을 갖기 위해 요구되는 조건이다. 따라서 모수 추

정 후 비례위험 가정을 위배하는지 검증 절차가 필요하다. 만일 PH 가정을 충족하지 않는 경우 과대추정으로 인해 정확한 추정량을 얻기 어렵다.

아울러 공변량 효과의 크기(탄력성)는 공변량의 계수가 지수이며, 밑이 자연상수 e인 거듭제곱 값인 $\exp(B)$ 다. 다만 공변량의 변수 유형에 따라 해석이 다소 다르다. 공변량이 연속형 변수일 경우 시점 t에서 공변량 값 1단위가 변화할 때 달라지는 위험비이며, 범주형 변수라면 기준 범주(reference category) 대비 관심 범주의 상대적 위험도¹⁾로 이해할 수 있다.

위 식의 양변에 자연로그를 취하면 회귀계수 추정에 필요한 선형식이 도출된다.

$$\ln(HR) = \ln\left[\frac{h_i(t)}{h_j(t)}\right] = (X_i' - X_j')B$$

그렇지만 두 개체 위험함수 비율로 정의된 $\ln(HR)$ 역시 비모수적이므로 최우추정법(MLE) 같은 모수 추정법은 사용할 수 없다. 따라서 Cox는 모수 추정을 위해 시점 t에서 사건 한 차례 발생했을 때 얼마나 많은 개체가 사건을 경험했는가는 조건부 확률을 바탕으로 부분우도함수(partial likelihood function)를 제시하였다. 다만 생존시간 표본 내에 동일한 생존시간을 가진 개체(tied value, 동점)가 없어야 한다는 다소 현실성이 떨어지는 가정이 있어 그대로 사용하기에 한계가 있다. 이에 따라 적확법(exact method), Breslow의 근사법(approximation method), 그리고 Efron의 근사법 등을 사용해 부분우도함수를 계산한 후 모수를 추정한다.

PH 성립 여부를 확인하는 방법으로는 LML(log-minus-log) 변환된 생존함수 그래프를 이용하는 방법²⁾, Schoenfeld 잔차를 이용한 적합도 검정(goodness of fitness, GOF test)³⁾ 등이 있다. PH 가정이 위배될 경우 시간 의존적(time-dependent) 공변량이 포함된 Cox PH 모형이나 층화된(stratified) Cox PH 모형처럼 기존의 모형을 변형한 방식으로 접근할 수 있으나 본고에서 다루지 않는다. 대신 PH 가정이 요구되지 않는 가속실패시간(AFT) 모형을 활용하는 방안을 택한다.

마. 가속실패시간 모형(모수적 방법)

위험함수를 직접 추정하는 Cox PH 모형과 달리 가속실패시간(accelerated failure time, 이하 AFT) 모형⁴⁾은 생존시간 T에 미치는 각 공변량의 영향을 바탕으로 생존함수 및 위험함수를 추정

- 1) 더미변수처럼 해당 범주에 속하면 1, 그렇지 않으면 0 값을 갖는 범주형 예측변수(공변량)가 1개 존재하한다고 가정해보자. 그러면 로그변환된 HR 모형은 $\ln(HR) = \beta_1 \cdot X_1$ 같이 나타낼 수 있다. 공변량의 계수를 추정하여 적합값을 구해보면 전자는 $\exp(\beta_1 \times 1) = \exp(\beta_1)$ 이며, 후자는 $\exp(\beta_1 \times 0) = 1$ 이다. 따라서 해당 범주에 속한 구성원은 그렇지 않은 구성원에 비해 사건 발생확률(위험률)이 $\exp(\beta_1)$ 배 높다고 판단할 수 있다.
- 2) 비례위험 가정 위배가 의심되는 범주형 공변량의 $\log[-\log(s(t))]$ 그래프를 그린 후 범주 간 그래프가 평행하다면(기울기가 같다면) 비례위험 가정이 성립함.
- 3) 비례위험 가정 위배가 의심되는 공변량 및 시간 t의 상호작용이 반영된 모형과 그렇지 않은 모형 간 우도비(likelihood ratio, LR)를 검정하는 방법으로 귀무가설은 공변량과 t간 상호작용이 없음(비례위험 가정이 성립함).

한다. 또한 위험함수의 분포를 특정하지 않는 Cox PH 모형과 다르게 분포를 가지는 확률변수로 취급하여 추정하는 모수적(parametric) 접근법이다. 그러므로 AFT 모형은 PH 가정에서 자유롭다는 이점이 있지만(Orbe-Ferreira-Núñez-Antón, 2002)⁵⁾, 선형적으로 특정 사건에 관한 생존함수의 분포를 알지 못하는 경우 정확한 추정이 어렵다는 한계가 있다(Klein-Goel, 2013). 일반적인 AFT 모형은 다음과 같이 정의할 수 있다.

$$\begin{aligned} T &= T_0 \cdot \exp(\alpha_1 x_1 + \dots + \alpha_q x_q) \cdot \epsilon \\ &= T_0 \cdot \exp(X'A) \cdot \epsilon \end{aligned}$$

AFT 모형은 모든 공변량 값이 0일 때 생존시간인 기저 생존시간(baseline survival time) T_0 , q 개 공변량이 포함된 열 벡터 X , 공변량과 선형결합하는 회귀계수의 열 벡터 A , 오차항 ϵ 을 포함한다. ϵ 는 확률변수로 다양한 분포⁶⁾를 가정할 수 있다. 이에 따라 확률변수인 T 의 분포도 결정된다. 한편 AFT모형의 양변에 자연로그를 취하면 다음과 같이 추정에 이용 가능한 선형식을 얻을 수 있다.

$$\begin{aligned} \log T &= \log T_0 + (X'A) + \log \epsilon \\ &= u + X'A + \sigma W \\ &= u + X'A + e \end{aligned}$$

모형의 종속변수는 로그변환된 생존시간 T 이며 오차항은 $e = \sigma W$ 이다. 오차항에 포함된 σ 는 양수 값($\sigma > 0$)을 가지는 척도 모수(scale parameter)로 생존함수 및 위험함수 분포의 높낮이를 결정한다. 공변량 효과의 크기(탄력성)는 $\exp(A)$ 와 같다. 그리고 $\exp(X'A)$ 는 공변량 X 의 값이 주어질 때 생존시간 T 에 미치는 추가적 영향을 의미하는 가속요인(acceleration factor)⁷⁾으로 시간 변화와 무관하게 일정한 값을 갖는다. 예컨대 분석 대상이 되는 두 개체(집단) i, j 가 존재하며, 개체 i 에 해당하는 경우 1 값을, 아닐 경우 0 값을 갖는 범주형 변수(공변량) 1개가 포함된 AFT 모형을 가정해 보자. 그리고 각각의 생존시간을 T_i, T_j 라 하고, 로그변환된 AFT 모형과 생존함수를 적절히 정리하면 가속요인을 통해 두 개체의 생존시간 차이를 설명할 수 있다.

-
- 4) 실패(failure)는 사건을 나타내며, 가속된(accelerated)이란 수식어는 실패(사건) 발생까지의 시간, 즉 생존 시간을 변화시키는 요인을 의미하는 가속요인(acceleration factor)에 대응되는 표현이다.
 - 5) 와이불(weibull) 및 지수(exponential) 분포를 가정해 위험함수를 추정하는 경우 분포의 성질에 의해 위험비(HR)는 시간 흐름과 무관하게 일정하다(Cox-Oakes, 1984). 따라서 두 분포를 가정한 AFT 모형은 모수적(parametric) PH 모형이라고도 불린다.
 - 6) 와이불, 지수 분포 외에도 로그-로지스틱(log-logistic), 로그-정규(log-normal), 감마(gamma), 고펜르츠(gompertz) 분포 등을 가정할 수 있다.
 - 7) Cox PH 모형의 상대적 위험도(relative risk)와 유사한 개념으로 공변량이 범주형 변수일 경우 기준 범주 대비 해당 범주의 생존시간 T 가 몇 배 더 긴가를 나타낸다.

$$T_i = \exp(u + \alpha \cdot 1 + e) = \exp(u + \alpha) \cdot \exp(e) = \gamma \cdot \exp(e)$$

$$T_j = \exp(u + \alpha \cdot 0 + e) = \exp(u) \cdot \exp(e) = \gamma_0 \cdot \exp(e)$$

$$S_i(t) = \Pr(T_i > t) = \Pr[\gamma \cdot \exp(e) > t] = \Pr[\exp(e) > \gamma^{-1}t], t \in T_i$$

$$S_j(t) = \Pr(T_j > t) = \Pr[\gamma_0 \cdot \exp(e) > t] = \Pr[\exp(e) > \gamma_0^{-1}t], t \in T_j$$

$$\text{Note that } \gamma = \gamma_0 \cdot \exp(\alpha),$$

$$S_i[\exp(\alpha) \cdot t] = \Pr[\exp(e) > \gamma^{-1} \cdot \exp(\alpha) \cdot t] = \Pr[\exp(e) > \gamma_0^{-1}t] = S_j(t)$$

$$S_j(t) = S_i(\phi t), \text{ where } \phi = \exp(\alpha)$$

In terms of random variables for survival time T,

$$T_i = \phi T_j$$

만일 가속요인(ϕ)이 $\phi > 1$ 이면 공변량 변화로 인해 생존시간 증가하고($T_i > T_j$), 반대로 $\phi < 1$ 이면 생존시간이 감소한다($T_i < T_j$). $\phi = 1$ 일 경우 공변량은 생존시간에 아무런 영향을 미치지 않는다($T_i = T_j$).

한편 생존함수와 위험함수는 AFT 모형 기반으로 추정할 수 있다. 생존시간 T가 와이불(weibull) 분포를 따른다고 가정할 때 특정 시점 $t(t \in T, t \geq 0)$ 에 대한 생존함수 S(t)와 위험함수 h(t), 그리고 확률밀도함수 f(t)는 아래 식처럼 같이 나타낼 수 있다(Lawless, 2011). 세 함수 모두 추정이 요구되는 공통의 모수 γ, p 를 가지고 있다. 전자는 함수 분포의 시작점을 결정하는 위치 모수(location parameter)인데, 아래 식에서 볼 수 있듯이 AFT 모형의 적합값과 동일한 형태다. 후자는 분포의 모양을 결정하는 형태 모수(shape parameter)다. 아울러 Cox PH 모형과 다르게 기저 위험함수 $h_0(t)$ 는 모수 γ_0, p 에 의해 영향을 받는 확률변수이므로 모수적 가정이 전제된다.

$$S(t) = \exp[-(\gamma t)^p] = \exp[-t^p \cdot (\exp(u + X'A))^p] = \exp[-t^p \cdot (\gamma_0 \phi)^p]$$

$$\text{, Where } \gamma = \exp(u + X'A), \gamma_0 = \exp(u), \phi = \exp(X'A)$$

$$h(t) = p \cdot \exp(u + X'A)^p \cdot t^{p-1} = p \cdot (\gamma_0 \phi)^p \cdot t^{p-1} = h_0(t) \cdot \phi^p$$

$$\text{, Where } h_0(t) = p(\gamma_0)^p \cdot t^{p-1}$$

$$f(t) = S(t) \cdot h(t) = \exp[-t^p (\gamma_0 \phi)^p] \cdot (\gamma_0 \phi p) \cdot t^{p-1}$$

일반적으로 최우추정법(MLE)을 이용해 확률밀도함수 f(t)를 추정한다. 그 결과로 회귀계수 추정량과 함께 형태 모수 p의 추정량을 얻게 된다(위치 모수 γ 를 적합하는 과정에서 척도 모수 σ 추정

값도 도출됨). 추정 후 회귀계수에 대한 우도비 검정(LR test, H_0 : 모형의 모든 회귀계수는 0 값을 갖는다)을 통해 모형의 적합성을 확인한다. 나아가 여러 분포를 가정하여 추정한 AFT 모형 중 가장 타당한 것을 선택하는데 AIC(Akaike's information criterion) 혹은 BIC(Bayesian information criterion) 척도를 활용할 수 있다. 두 척도 모두 로그화된 우도값(log-likelihood)을 이용하는데, 그 계산값이 작을수록 바람직한 모형이다.

바. 생존분석을 위한 R 패키지

본 연구를 위해 통계 및 빅데이터 분석에 있어 광범위하게 활용되는 오픈소스 소프트웨어인 R을 사용하였다. 생존분석을 위해 많은 패키지가 개발되었고, 이론의 발전 및 활용 분야의 확대에 발맞추어 명령어가 거듭 업데이트되었다. 관련 패키지들 중에서 가장 많이 활용되는 것은 survival 패키지(Therneau, 2020)다. 카플란-마이어 분석을 위한 survfit 함수, Cox PH 모형의 적합을 위한 coxph 함수, 그리고 AFT 모형을 추정하는 survreg 함수 등 거의 모든 생존분석 기법을 구현한 패키지다. 이 밖에도 flexsurv 패키지(Jackson, 2019), eha 패키지(Broström, 2015), rsm 패키지(Harrel, 2018) 등도 모수적 추정에 자주 이용된다. 분석 결과를 바탕으로 유려한 그래프를 생성하는 surminer(Kassambara, 2020) 패키지도 언급하지 않을 수 없다. 열거한 패키지들의 매뉴얼(pdf 파일)은 인터넷을 통해 손쉽게 구할 수 있다. 본고의 분석에서는 survival 패키지를 우선 활용하되, 결과를 교차 검증하거나 조금 더 자세한 정보를 얻기 위해 기타 패키지도 병행하여 사용하였다.

2. 분석 자료

본 연구의 구체적 목표는 임금 근로자의 구직기간에 미치는 개인적 요인들을 확인하는 것이며, 분석 방법은 앞서 설명한 생존분석 기법이다. 이를 위해 한국노동연구원(KLI)에서 생산하는 한국노동패널조사(KLIPS) 중 개인 단위 데이터를 활용하였다. 해당 데이터는 1999년부터 매년 공표되지만 분석에 필요한 데이터를 완전하게 확보할 수 있는 11~22차년도(2008~2019), 총 12년치 자료로 사용을 한정하였다.

분석 대상자는 각 조사에서 당해 연도에 구직활동이 있었던 사람이다. KLIPS 데이터에는 취업한 사람(관련 문항: p_2701)과 미취업 상태인 사람(p_2843)으로 구분해 구직활동 유무를 조사하였고, 구직활동이 있는 경우에 추가 문항을 통해 구직기간을 주 단위로 기입하였다(각 p_2706, p_2846). 본 연구는 취업자와 미취업자 구직기간을 모두 아울러다. 다만 취업자의 경우에는 각 조사 연도에 취업에서 성공한 임금 근로자로 한정하였다. 비임금 근로자(자영업자 및 무급 가족 종사자) 경우 개인의 특성이나 자질보다는 혈연·친분 관계에 따라 취업이 결정되는 경우가 많으며, 장기간 구직활동 끝에 창업을 선택하는 이들도 있어 일종의 편의(bias)로 작용할 소지가 높기 때문이다. 더군다나 KLIPS 데이터를 선별하는 과정에서 해당 요인들을 통제할 마땅한 방법이 없다. 아울러 미취업자의 구직기간은 조사 시점에서 구직활동이 계속 진행 중인 경우에 한하여 분석에 이용

하였다.⁸⁾ 미취업자 중 구직활동이 종료되었다고 응답한 이들은 취업에 성공에 한 것이 아니라 구직활동을 중도 포기한 비경제활동인구으로 간주되므로 분석에 불필요한 대상이다. 한편 다년도 데이터를 통합할 경우 전년도 미취업자들의 구직기간이 해를 넘겨 익년도 조사 때 다시 포착되는 중복 집계 우려가 있다. 그러나 통계를 내보면 1년(52주) 이상을 넘겨 구직활동을 하는 사람들은 극히 일부(구직기간 중간값은 4주, 하위 99.6%가 52주)이므로 분석에 있어 커다란 편의를 초래할 가능성은 아주 낮은 것으로 생각된다. 덧붙여 오직 실업 상태에서 진행되는 미취업자와 다르게 취업자의 구직활동은 재직 중 이직을 준비하였던 사례도 포함하기 때문에 해석에 유의할 필요가 있다. 원 데이터의 한계로 분석에 필요한 데이터를 정제하는 과정에서 이 차이를 통제할 수 없었다. 결론적으로 구직활동 기간은 생존기간, 취업은 사건에 해당된다. 그리고 취업자(임금 근로자)는 사건이 발생한 사람, 구직활동을 지속하는 미취업자는 그렇지 않은 사람이다.

목표변수인 구직기간에 잠재적으로 영향을 미칠 것으로 생각되는 다양한 개인의 특성을 통제하기 위해 선행연구와 KLIPS 설문지의 문항을 바탕으로 12가지 예측변수를 선정하였다(<표 1> 참조). 또한 분석 편의를 위해 일부 문항(학력, 거주지 등)의 응답 범주를 단순화하였으며, 원 데이터에서 연속형 변수인 연령은 범주형 변수인 연령대로 재구성하여서 분석에 이용하였다.

<표 1> 분석 변수

변수명	활용	유형	관련 노동패널조사 문항	비고
구직기간(주)	목표변수 (종속변수)	수치형 (연속형) 변수	p_2706, p_2846	원 자료의 당해 취업자 중 임금 근로자의 구직기간 및 당해 미취업자 중 지속적으로 구직활동하는 사람의 구직기간 통합
성별	예측변수 (공변량)	범주형 변수	p_0101	-
가구주			p_0102	범주 단순화
연령대 ⁹⁾			p_0107	범주형 변수로 변경
학력			p_0110	범주 단순화
거주지			p_0121	
사회보험급여 ¹⁰⁾			p_2142, p_2155, p_2165	
직업훈련 ¹¹⁾			p_4501	
자격증 ¹²⁾			p_4701	
싱글			p_5501	
건강상태			p_6103	
가족 수입 만족도			p_6501	
사회경제적 지위 향상 가능성			p_6601	

8) 조사 이후에 구직 성공 여부를 확인할 수 없으므로 우측 중도절단에 해당된다.

9) 만 나이 기준

10) 연금(국민연금, 사학연금, 공무원 연금, 군인 연금 등) 혹은 고용보험, 산재보험, 보훈연금 등 사회보험금

11) 직장연수, 학원수강, 교양강좌 및 통신강좌를 제외한 취·창업 또는 업무능력 향상을 목적으로 하는 교육이나 직업 훈련

3. 모형 설정

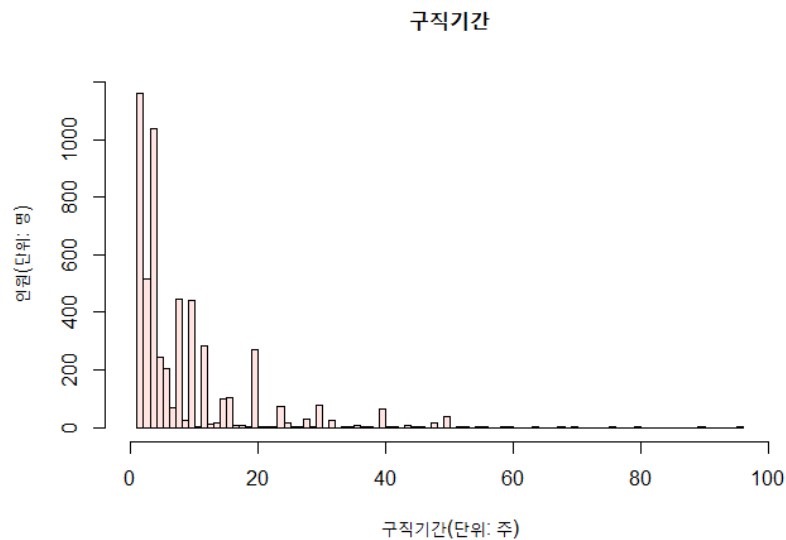
Cox PH 모형(준모수적 방법)와 AFT 모형(모수적 방법)은 K-M 분석법(비모수적 방법)과 달리 둘 이상의 예측변수를 포함한 분석이 가능하다. 앞서 설명한 바와 같이 두 모형 모두 종속변수(목표변수)와 공변량(예측변수)간에 함수 관계를 갖기 때문에 로그변환을 통해 얻은 선형식의 모수를 추정하면 각 공변량의 효과를 확인할 수 있다.

본고에서 분석에 사용할 3가지 유형의 분석 모형을 도입하였다: ①모든 연도의 데이터를 사용하며, 각 연도의 이질적인 특성을 통제하기 위한 연도 더미변수가 포함된 모형, ②각 연도의 데이터만 사용하는 모형, ③공변량의 상호작용(교호작용, interaction)을 반영하는 교차항이 포함된 모형. 교차항은 공변량(범주형 변수)의 곱으로 나타내며, 각 공변량 범주에 동시에 해당하는 경우 1값을 가지는 일종의 더미변수다. 교차항은 성별×연령대, 성별×학력, 성별×싱글, 싱글×연령대, 거주지×연령대, 가구주×연령대, 사회보장급여×연령대, 직업훈련×연령대 등 총 8가지를 설정하고 각각을 분석 모형 ①에 포함하여 추정하였다.

IV. 분석 결과

1. 기초 통계량

[그림 1] 구직기간 분포



12) 운전면허증을 제외한 국가자격증(기술사, 기능사, 전문직 등), 민간 자격증 또는 외국 자격증

<표 2> 구직기간 10분위수

구직기간(단위: 주) 10분위수									
1분위	2분위	3분위	4분위	5분위	6분위	7분위	8분위	9분위	10분위
2	2	3	4	4	7	10	12	20	96

본 연구 대상자(구직활동을 한 사람) 수는 총 5,355명으로, 특정 해에 집계된 수치가 아니라 12년 치(2009~2020) 총 인원이다. 구직기간의 분포를 히스토그램을 통해 나타내보면 좌측 구간에 관측치의 상당 부분이 몰려있는 형상(right-skewed)을 띠고 있다([그림 1], <표 2> 참고). 물론 분석에 앞서 극소수의 결측치와 이상치 제거 작업을 거쳤다. 각 연도에 사건(취업)이 발생한 인원과 중도절단 인원에 대한 수치는 <표 3>에 정리하였다. 기타 변수들의 통계량은 <표 4>에 나열하였다.

<표 3> 구직기간 중도단절

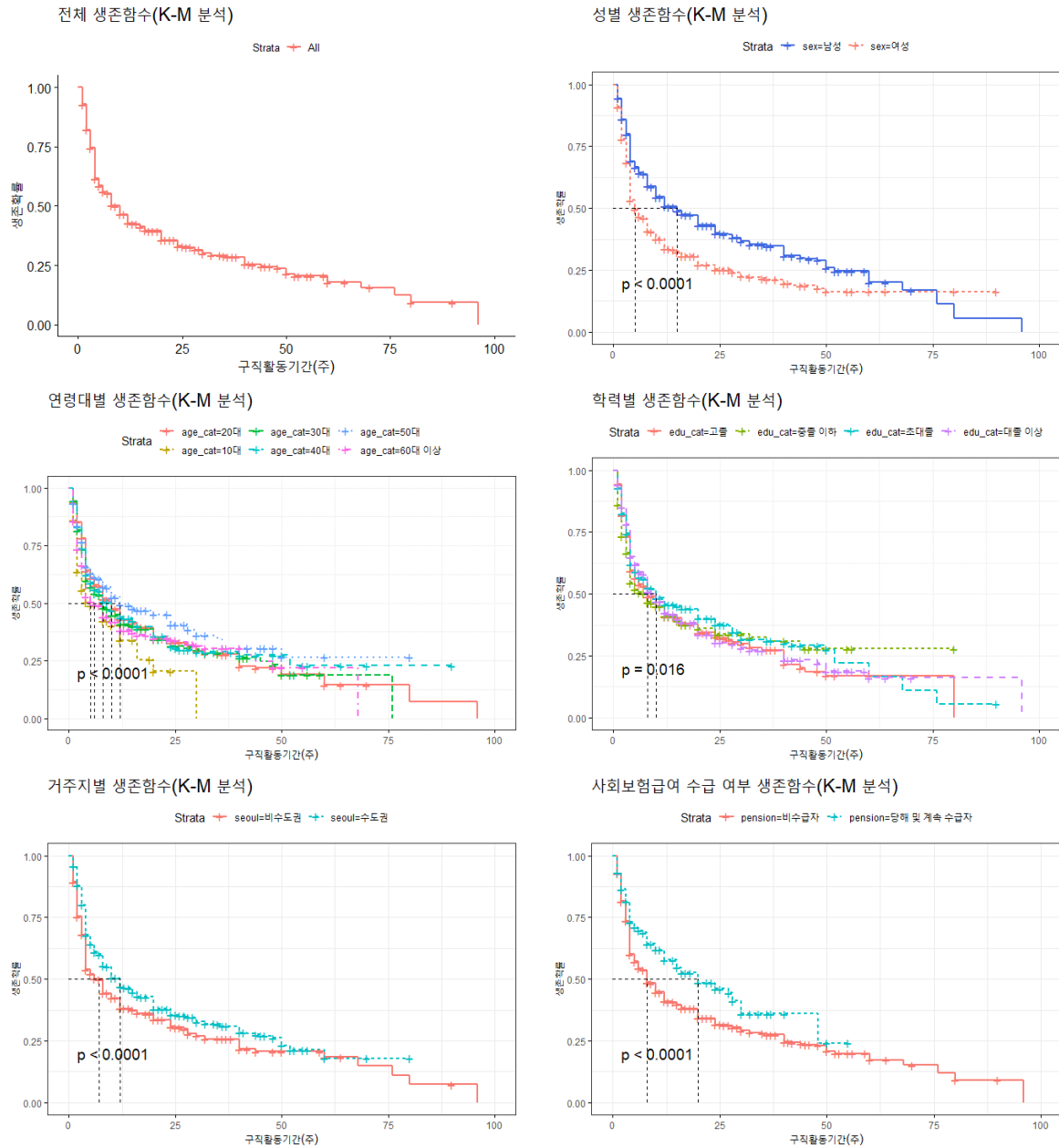
연도	총 인원(명)	사건(취업) 발생 인원(명)	중도단절 인원(명)	연도	총 인원(명)	사건(취업) 발생 인원(명)	중도단절 인원(명)
전체	5,355	2,890	2,465	2013	409	245	164
2019	650	344	306	2012	434	231	203
2018	722	474	248	2011	458	272	186
2017	347	191	156	2010	453	255	198
2016	362	197	165	2009	501	187	314
2015	353	180	173	2008	338	150	188
2014	328	164	164				

<표 4> 변수의 기초 통계

연속형 변수	관측 수(명)	평균	표준편차	중위수	최솟값	최댓값
구직기간(주)	5,355	8.55	9.64	4.00	1.00	96.00
범주형 변수	빈도 (비율)	범주 값	범주(측정 수준)			
성별	2,783 (51.97%)	0	남성			
	2,572 (48.03%)	1	여성			
가구주	3,232 (60.35%)	0	비가구주			
	2,123 (39.65%)	1	가구주			
연령대	107 (2.00%)	1	10대			
	1,837 (34.30%)	2	20대			
	1,235 (23.06%)	3	30대			
	924 (17.03%)	4	40대			
	653 (11.86%)	5	50대			
	617 (11.52%)	6	60대 이상			
학력	661 (12.34%)	1	중졸 이하(무학 포함)			
	1,220 (22.78%)	2	고졸			
	1,711 (31.95%)	3	초대졸(2년제 대학·전문대학)			
	1,763 (32.92%)	4	대졸 이상			
거주지	2,488 (46.46%)	0	비수도권			
	2,867 (53.54%)	1	수도권(서울·인천·경기)			
사회보험급여	4,777 (89.21%)	0	비수급자(비당해 수급자·수급 경험없음)			
	578 (10.79%)	1	당해 및 계속 수급자			
직업훈련	5,042 (94.16%)	0	없음			
	313 (5.85%)	1	있음(경험 있음 및 현재 받고 있음)			
자격증	5,078 (94.83%)	0	없음			
	277 (5.17%)	1	있음			
싱글	2,312 (43.17%)	0	기혼			
	3,043 (56.83%)	1	싱글(미혼·별거·이혼·사별)			
건강상태	312 (5.83%)	1	나쁨(건강하지 않는 편·아주 좋지 않음)			
	2,375 (44.35%)	2	보통			
	2,668 (49.82%)	3	양호(건강한 편·아주 좋음)			
가족 수입 만족도	1,786 (53.48%)	1	불만족(불만족·매우 불만족)			
	2,864 (33.35%)	2	보통			
	705 (13.17%)	3	만족(만족·매우 만족)			
사회경제적 지위 향상 가능성	2,875 (53.69%)	0	부정(별로 그렇지 않다·전혀 그렇지 않다) 및 모름			
	2,480 (46.31%)	1	긍정(대체로 그렇다 및 매우 그렇다)			

2. K-M 분석 결과

[그림 2] 전체 및 변수별 생존함수 그래프¹³⁾



우선 카플란-마이ер(K-M) 분석법을 통해 시간 흐름에 따른 생존확률(사건 미발생률)의 추이를 확인하였다. 해당 기법은 생존확률이 특정 분포를 따르지 않는다는 가정 하에서 생존함수를 추정하는 비모수적 방법이다. 이를 통해 전체 구직자들에 대한 구직 성공확률(사건 발생확률, 위험률)

13) 생존함수 그래프 상 + 표시는 해당 시점에서 중도단절이 발생했다는 의미다.

의 추이를 파악할 수 있을 뿐만 아니라 성별, 연령대, 거주지 등 각 예측변수에 속하는 범주간 구직 성공확률에 유의미한 차이가 있는지 확인할 수 있다. [그림 2]처럼 범주별 생존함수 곡선을 그려 그 차이를 밝히는 방법이 있으나, 로그-순위 검정(log-rank test, H_0 : 범주간 생존함수에 차이가 없음) 같은 통계적 방법을 통해 검증하는 것이 엄밀성 측면에서 바람직하다.

전체 구직자의 생존확률 곡선은 앞서 그려본 구직기간 히스토그램과 비슷한 양상을 보였다. 전체 구직 성공자의 70%가 나타난 10주 무렵까지 생존확률이 빠르게 감소(구직 성공률 증가)하였으며, 이후로는 감소 폭이 점차 완화되었다. 아울러 범주별 생존확률 곡선 사이에 상이함이 관찰되었다. <표 5>는 K-M 분석법에 따른 각 예측변수의 분석 결과다. 로그-순위 검정 결과, ‘가구주’와 ‘건강상태’를 제외한 나머지 변수의 경우에 범주간 생존함수의 유의미한 차이가 통계적으로 확인되었다($p < 0.1$). 그 차이를 중위 구직기간(해당 범주 구성원 절반이 구직에 성공하는데 소요되는 시간)을 통해 자세히 설명해보자면 이하와 같다. ‘성별’의 경우 여성이 남성보다, ‘학력’의 경우 고졸자가 대졸자와 비교해, ‘거주지’ 경우 비수도권이 수도권에 대비해 중위 구직기간이 상대적으로 짧은 것으로 나타나 통념과 다소 상반된다. 한편 ‘사회보험급여’의 경우 비수급자가 구직활동 당해 및 계속 수급자보다, ‘직업훈련’은 경험이 있는 구직자가 경험이 없는 구직자보다, ‘자격증’은 취득자가 미취득자와 비교해 중위 구직기간이 짧은 것으로 드러났다. 아울러 ‘싱글’이라 답한 사람은 기혼자보다 중위 구직기간이 비교적 길었으며, ‘가족 수입’에 더 만족할수록, 그리고 ‘사회경제적 지위 향상 가능성’을 긍정하는 경우 중위 구직기간이 상대적으로 짧아 주관적 요소와 가치관도 구직활동에 유의미한 영향을 미친 것으로 나타났다.

다만 범주간 생존기간의 차이를 초래하는 구체적 배경 원인은 알 수 없으므로 분석 결과 해석에 유의할 필요가 있다. 예를 들어 ‘학력’의 경우 초대졸자, 대졸자 이상이 고졸자보다 중위 구직기간이 상대적으로 긴 것으로 나타났는데, 이를 구직에 있어 학력 프리미엄이 작용하지 않는다고 단언할 수 없다. 고학력자일수록 기대치가 높기 때문에 적절한 일자리를 모색하고 준비하는데 많은 시간을 보내는 것일 수도 있다. 어쩌면 다른 원인이 존재할 수도 있다. 마찬가지로 ‘사회보험급여’의 당해 및 계속 수급자가 비수급자보다 중위 구직기간이 길었는데, 수급으로 인해 구직 의욕이 저하되었는지 아니면 사회보험 자체가 구직활동에 충분한 도움을 주지 못하였는지 분석 결과 자체만으로 판단할 수 없다.

<표 5> K-M 분석 결과 (유의수준: * p<0.1, ** p<0.05, *** p<0.01)

예측변수		인원(명)	실제 취업자 수(명)	기대 취업자 수(명)	추정 중위 구직기간(주)	로그순위 검정 (Chi-sq)
없음		5,355	2,890	-	9	-
성별	남성	2,783	1,330	1,649	15	162***
	여성	2,572	1,560	1,241	5	
가구주	비가구주	3,232	1,738	1,743	9	0
	가구주	2,123	1,152	1,147	10	
연령대	10대	107	63	43	5	31.8***
	20대	1,837	1,032	1,067	10	
	30대	1,235	673	650	8	
	40대	924	475	482	10	
	50대	635	293	346	12	
학력	60대 이상	617	354	302	6	10.3**
	중졸 이하	661	359	319	8	
	고졸	1,220	685	657	8	
	초대졸	1,711	839	886	10	
거주지	대졸 이상	1,763	1,007	1,028	10	70.5***
	비수도권	2,488	1,397	1,188	7	
사회보험급여	수도권	2,867	1,493	1,702	12	37.7***
	비수급자	4,777	2,671	2,573	8	
직업훈련	당해계속 수급자	578	219	317	20	27.3***
	없음	5,042	2,667	2,728	10	
자격증	있음	313	223	162	4	12.3***
	없음	5,078	2,715	2,753	10	
싱글	있음	277	175	137	6	31.1***
	기혼	2,312	1,305	1,166	8	
건강상태	싱글	3,043	1,585	1,724	10	2.8
	나쁨	312	140	159	12	
	보통	2,375	1,283	1,268	8	
가족 수입 만족도	양호	2,668	1,467	1,463	9	251***
	불만족	1,786	675	1,023	28	
	보통	2,864	1,697	1,520	8	
사회경제적 지위 향상 가능성	만족	705	518	346	4	62.9***
	부정 및 모름	2,875	1,410	1,610	12	
연도	긍정	2,480	1,480	1,280	8	124***
	2019	650	344	346	8	
	2018	722	474	354	5	
	2017	347	191	179	8	
	2016	362	197	177	8	
	2015	353	180	195	12	
	2014	328	164	180	12	
	2013	409	245	222	8	
	2012	434	231	241	10	
	2011	458	272	230	6	
	2010	453	255	274	10	
2009	501	187	303	28		
2008	338	150	188	15		

3. Cox PH 및 AFT 모형 분석 결과

구직기간에 유의미한 영향을 주는 여러 공변량을 파악하기 위해 도입한 세 가지 모형에 대해 Cox PH 모형(준모수적 방법) 및 AFT 모형(모수적 방법)을 전제로 분석하였다. 상술한 바와 같이 Cox PH 모형 추정 시 위험함수 분포에 대한 가정을 요구하지 않는다. 다만 위험비(HR)가 시간 흐름과 무관한 상수값을 갖는, 이른바 비례위험(PH)를 충족한다는 전제 하에서 공변량의 회귀계수를 추정한다. 하지만 AFT 모형 추정 시 생존시간이 특정 분포를 따른다고 가정하고, 시간의 함수로 표현되는 위험함수 역시 가정한 분포를 따르므로 PH 가정을 준수하지 않아도 된다.

<표 6>는 모든 공변량이 포함된 분석 모형 ①을 Cox PH 모형으로 간주하여 추정한 후 적합도 검정(GOF test) 실시한 결과를 정리한 것이다. 검정 결과, ‘사회보험급여’, ‘자격증’, ‘사회경제적 지위 향상 가능성’을 제외한 나머지 공변량의 검정 통계량은 통계적으로 유의한 것으로 나타나 ($p < 0.01$) 귀무가설(H_0 : HR는 시간 흐름과 관계없이 일정한 값을 갖는다)이 기각되어 PH 가정이 성립하지 않았다. 모형 전체에 대한 적합도 검정 결과 역시 PH 가정을 만족하지 않았다. 따라서 Cox PH 모형을 사용해 추정하면 잘못된 모수 추정량을 얻을 가능성이 매우 높다.

<표 6> 적합도 검정 (유의수준: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)

공변량	검정통계량(Chi-sq)
성별	3.681*
가구주	3.452*
연령대	51.676***
학력	57.282***
거주지	57.447***
사회보험급여	0.006
직업훈련	10.314***
자격증	0.643
싱글	8.378***
건강상태	15.212***
가족 수입 만족도	19.449***
사회경제적 지위 향상 가능성	0.731
모형 전체	203.971***

이에 따라 PH 가정이 불필요한 AFT 모형을 가정하여 추가 분석을 실시하였다. 생존시간 분포는 와이불(weibull), 지수(exponential), 로그-로직(log-logistic), 로그-정규(log-normal) 등 총 4가지를 가정하였다. <표 7>는 분석 모형 ①을 Cox 모형 및 4가지 AFT 모형으로 간주하여 추정한 각 공변량의 회귀계수를 정리한 것이다. 참고로 두 추정 모두 선형 회귀분석을 위해 로그변환한 모형

을 추정한 것으로 회귀계수 해석 시 유의해야 한다. 추정에 사용된 Cox 모형의 종속변수는 로그화된 HR이므로 기준 범주(reference category) 대비 해당 범주의 HR가 몇 배 더 높은지¹⁴⁾ 판단하기 위해서 회귀계수를 해야 한다. 마찬가지로 AFT 모형의 종속변수는 로그변환된 생존시간 T이며, 기준 범주 대비 해당 범주의 T가 몇 배나 긴지¹⁵⁾ 비교하기 위해 밀이 자연상수이며, 회귀계수가 지수를 갖는 거듭제곱 값으로 변환해야 한다.

아울러 AIC(Akaike's information criterion) 척도를 통해 여러 분포를 가정한 AFT 모형 중 가장 적합도가 높은 모형을 판별하였다. 그 결과, 분포들간 AIC 값에 큰 차이가 없었지만 로그-정규 분포를 가정할 때 가장 작은 것(AIC=20,149)으로 나타나 적합도가 제일 높았다.

로그-정규 분포를 가정한 AFT 모형의 분석 결과, '자격증'과 '건강상태' 제외한 나머지 공변량의 회귀계수가 통계적으로 유의미한 것으로 확인되었다($p < 0.1$). 회귀계수가 양수 값을 가지면 종속변수인 $\ln(T)$ 에 정(+)의 효과, 반대로 음수 값이면 부(-)의 효과를 미친다. 그리고 상술했듯이 거듭제곱 값으로 변환하면 범주간 비교가 가능하다. 이 경우 변환한 값이 1을 초과하면 해당 범주는 기준 범주에 비해 생존시간이 상대적으로 길며, 1 미만이면 기준 범주보다 생존시간이 상대적으로 짧다고 해석할 수 있다. [그림 3]은 유의미한 회귀계수들을 바탕으로 각 범주의 상대적인 구직기간(생존기간)을 보기 쉽게 정리한 것이다. 각 공변량 효과를 구체적으로 풀이해보면 이하와 같다. '성별'의 경우 여성이 남성보다 구직기간이 약 0.570배($\exp(-0.562)$) 길었다. 여성의 구직기간이 남성에 비해 절반 남짓 짧았다는 의미다. '가구주' 경우에 가구주에 해당하는 사람은 비가구주에 비해 구직기간이 약 0.746배($\exp(-0.294)$)에 불과하였다. '연령'의 경우 10대는 20대와 비교해 약 0.575배($\exp(-0.553)$), 50대는 약 1.366배($\exp(0.312)$), 60대 이상은 약 0.787배($\exp(-0.239)$)로 나타나 10대 및 60대 이상은 20대 대비 구직기간이 짧았다. '학력'의 경우 중졸자 이하가 고졸자에 비해 약 0.839배($\exp(-0.175)$), 대졸자 이상은 약 1.134배($\exp(0.126)$)의 구직기간을 보였다. '거주지'의 경우 수도권 거주자가 비수도권 거주자와 비교해 구직기간이 약 1.480배($\exp(0.392)$) 길었다. '사회보험급여'의 경우 구직활동 당해 및 계속 수급자가 비수급자보다 구직기간이 약 1.911배($\exp(0.647)$) 길었다. '직업훈련'의 경우 유경험자가 무경험자에 비해 구직기간이 약 0.788배($\exp(-0.238)$)에 불과하였다. '싱글'에 해당할 경우 기혼자보다 구직기간이 약 1.329배($\exp(0.285)$) 길었다. 또한 '가족 수입'에 만족하는 이들은 보통 수준의 만족에 비해 약 0.793배($\exp(-0.232)$), 불만족하다고 느낀 이들은 약 1.554배($\exp(0.441)$) 구직기간이 길었다. 마지막으로 '사회경제적 지위 향상 가능성'을 긍정하는 사람들은 그렇지 않거나 모른다고 답한 이들보다 약 0.871배($\exp(-0.138)$)의 구직기간을 보였다.

아울러 공변량 효과의 방향성(부호)이 해마다 달라지는지, 또 추세를 가지는지 확인하고자 개별 연도의 자료만 사용해(분석 모형 ②) 추가 분석을 수행하였다. 모형은 로그-정규 분포를 따르는 AFT 모형을 가정하였으며, 추정 결과는 <표 8>에 정리하였다. 분석 결과, 각 연도 추정에서 확인된 유의미한 회귀계수는 절대적 크기가 다를지언정 방향성(부호)은 통합 모형(분석 모형 ①)의 그것과 동일하게 도출되었다. 공변량 중 '성별', '가구주', '거주지', '사회보험급여', '싱글', '가족 수입

14) 상대적 위험도(relative risk)

15) 가속요인(acceleration factor)

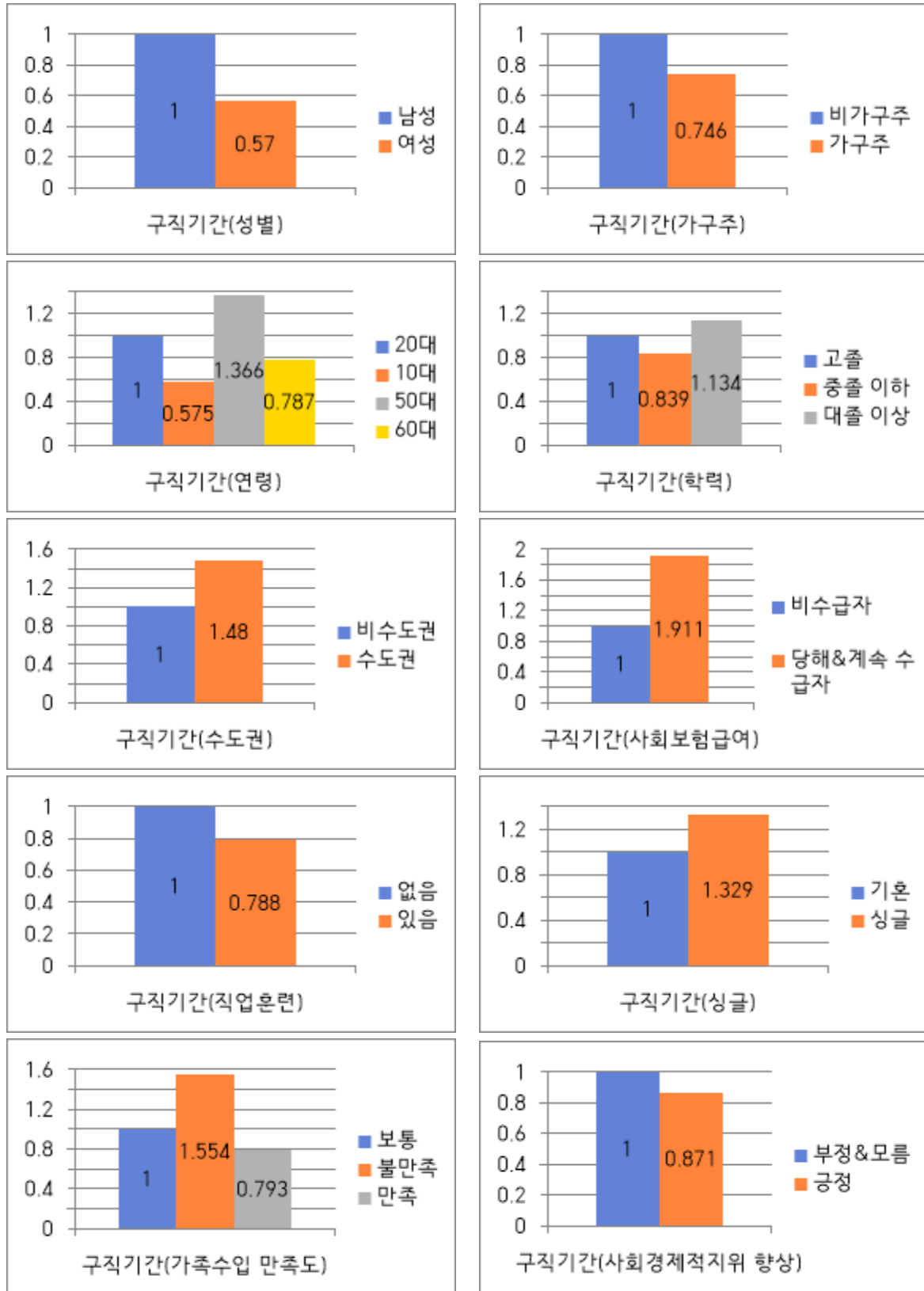
만족도'의 추정 회귀계수는 많은 연도에서 유의한 값이 도출되었으나, 시계열적으로 감소 혹은 증가 추세가 확인되지 않았다([그림 4] 참고). 그리고 '학력', '자격증', '건강상태'의 일부 범주에서 회귀계수의 유의미함이 추가로 확인되었으나, 1~2개 연도에서만 도출되었으므로 공변량의 효과를 일반화하기 어렵다.

마지막으로 공변량간 상호작용(교호작용)이 있을 것이라 가정하고 교차항이 포함된 모형을 분석하였다(분석 모형 ③). 교차항의 유형은 모두 8가지를 설정하였고, 로그-정규 분포를 따르는 AFT 모형을 가정하여 각각 추정하였다. <표 9>에 그 추정 결과를 정리하였다(지면 부족으로 관련된 공변량과 교차항의 회귀계수만 기입하였음). 분석을 통해 확인된 유의미한 회귀계수를 보면 방향성(부호) 측면에서 교차항이 포함되지 않는 모형(분석 모형①)과 동일하였다. 한편 '성별'와 '연령대'의 상호작용을 고려한 모형(③-1)에서 남성보다 상대적인 구직기간이 짧은 여성임에도 교차항을 보면 연령대별로 효과가 달리 나타났는데, 여성이면서 고연령대일수록 구직기간이 단축되는 효과가 나타났다. '성별'와 '싱글'의 상호작용을 반영한 모형(③-3)에서 여성이 남성보다 구직기간이 상대적으로 짧았지만 여성 및 싱글 범주에 동시 해당하는 경우 오히려 구직기간이 증가하는 효과가 나타났다. '성별'와 '연령대'의 상호작용을 고려한 모형(③-4)에서 40대 이상 한정으로 싱글에 해당하면 구직기간이 감소하는 효과를 보였다. '가구주'와 '연령대'의 상호작용을 고려한 모형(③-6)에서 40대 이상 연령대인 가구주는 구직기간이 늘어나는 효과가 확인되었다. '사회보장급여'와 '연령대'의 상호작용을 고려한 모형(③-7)에서 10대와 60대 이상 연령대이면서 당해 및 계속 수급자 경우 구직기간이 줄어드는 효과가 나타났다. 한편 '성별'와 '학력'의 상호작용을 고려한 모형(③-2), '거주지'와 '연령대'의 상호작용을 반영한 모형(③-5), 그리고 '직업훈련'와 '연령'의 교차항이 포함된 모형(③-8)은 공변량간 상호작용 효과가 극히 일부만 유의하게 나타났거나 아예 확인되지 않았다.

<표 7> 분석 모형 ① 추정 결과 (유의수준: * p<0.1, ** p<0.05, *** p<0.01, Ref.: 기준 범주)

	Cox PH 모형	가속실패시간(AFT) 모형				
		Weibull	Exponential	Log-logistic	Log-normal	
종속변수	ln(HR)	ln(T)				
상수항	-	2.630*** (0.102)	2.600*** (0.096)	2.110*** (0.099)	2.183*** (0.098)	
성별 (Ref.: 남성)	0.546*** (0.042)	-0.671*** (0.045)	-0.653*** (0.042)	-0.607*** (0.044)	-0.562*** (0.043)	
가구주 (Ref.: 비가구주)	0.340*** (0.046)	-0.389*** (0.049)	-0.372*** (0.046)	-0.317*** (0.048)	-0.294*** (0.047)	
연령대 (Ref.: 20대)	10대	0.437*** (0.135)	-0.505*** (0.144)	-0.497*** (0.135)	-0.541*** (0.143)	-0.553*** (0.139)
	30대	0.026 (0.058)	-0.040 (0.062)	-0.045 (0.058)	-0.030 (0.060)	-0.028 (0.059)
	40대	-0.098 (0.071)	0.116 (0.076)	0.111 (0.072)	0.060 (0.072)	0.057 (0.071)
	50대	-0.332*** (0.087)	0.388*** (0.094)	0.371*** (0.088)	0.338*** (0.088)	0.312*** (0.086)
	60대 이상	0.133 (0.096)	-0.051 (0.103)	-0.039 (0.096)	-0.252** (0.103)	-0.239** (0.101)
학력 (Ref.: 고졸)	중졸 이하	0.150* (0.086)	-0.178* (0.092)	-0.174** (0.086)	-0.195** (0.092)	-0.175* (0.090)
	초대졸	-0.055 (0.056)	0.044 (0.060)	0.040 (0.056)	0.025 (0.058)	0.021 (0.057)
	대졸 이상	-0.105** (0.050)	0.156*** (0.053)	0.158*** (0.050)	0.128** (0.052)	0.126** (0.052)
거주지 (Ref.: 비수도권)	-0.321*** (0.039)	0.409*** (0.041)	0.404*** (0.039)	0.404*** (0.040)	0.392*** (0.040)	
사회보험급여 (Ref.: 비수급자)	-0.648*** (0.075)	0.620*** (0.081)	0.579*** (0.075)	0.685*** (0.076)	0.647*** (0.075)	
직업훈련 (Ref.: 없음)	0.279*** (0.072)	-0.287*** (0.077)	-0.269*** (0.072)	-0.243*** (0.079)	-0.238*** (0.080)	
자격증 (Ref.: 없음)	0.058 (0.082)	-0.047 (0.087)	-0.045 (0.082)	-0.044 (0.087)	-0.047 (0.088)	
싱글 (Ref.: 기혼)	-0.304*** (0.050)	0.353*** (0.054)	0.341*** (0.051)	0.325*** (0.051)	0.285*** (0.050)	
건강상태 (Ref.: 보통)	나쁨	-0.097 (0.092)	0.120 (0.098)	0.116 (0.092)	0.065 (0.097)	0.024 (0.091)
	좋음	0.037 (0.040)	-0.042 (0.042)	-0.041 (0.040)	-0.038 (0.041)	-0.037 (0.041)
가족 수입 만족도 (Ref.: 보통)	불만족	-0.514*** (0.048)	0.558*** (0.051)	0.532*** (0.048)	0.482*** (0.048)	0.441*** (0.046)
	만족함	0.242*** (0.052)	-0.253*** (0.056)	-0.241*** (0.052)	-0.224*** (0.056)	-0.232*** (0.057)
사회경제적 지위 향상 가능성(Ref. 부정-모름)	0.121*** (0.039)	-0.129*** (0.042)	-0.123** (0.040)	-0.141** (0.041)	-0.138** (0.040)	
연도 더미	있음	있음	있음	있음	있음	
LR test	792.0***	911.3***	1,002.0***	866.0***	781.0***	
Log-likelihood	-22,604	-10,363	-10,374	-10,107	-10,041	
AIC	45,270	20,793	20,812	20,279	20,149	
관측치 수(명)	5,355	5,355	5,355	5,355	5,355	

[그림 3] AFT 모형(로그-정규 분포 가정) 추정에 따른 상대적 구직기간(기준 범주=1)



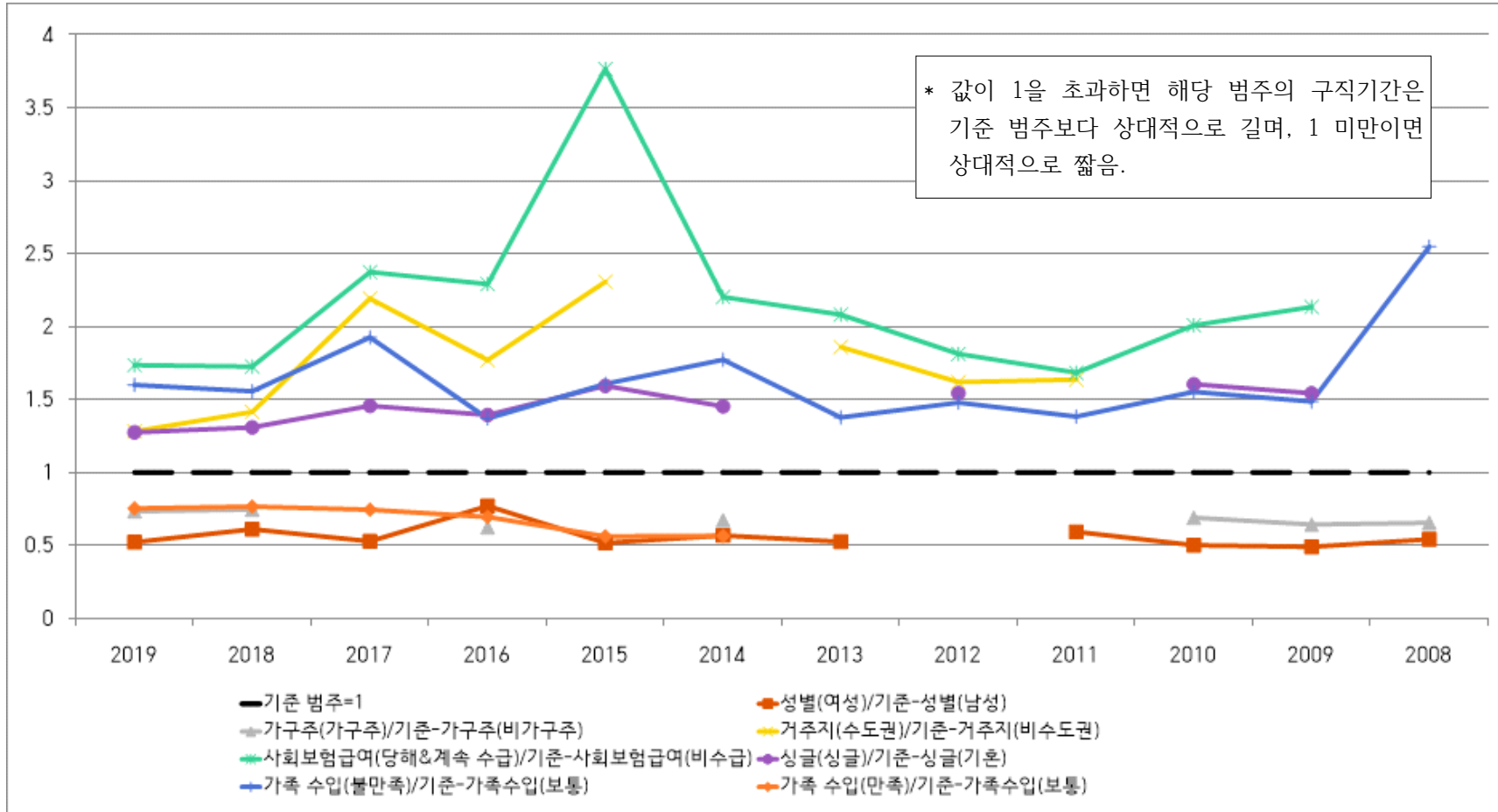
<표 8> 분석 모형 ② 추정 결과 (유의수준: * p<0.1, ** p<0.05, *** p<0.01, Ref.: 기준 범주)

		Log-normal 가속실패시간(AFT) 모형					
		2019	2018	2017	2016	2015	2014
종속변수		ln(T)					
상수항		2.620*** (0.223)	1.860*** (0.176)	1.750*** (0.261)	1.850*** (0.272)	2.080*** (0.358)	2.430*** (0.336)
성별 (Ref.: 남성)		-0.650*** (0.114)	-0.492*** (0.093)	-0.639*** (0.140)	-0.261* (0.141)	-0.659*** (0.180)	-0.566*** (0.180)
가구주 (Ref.: 비가구주)		-0.311** (0.131)	-0.293*** (0.103)	-0.255 (0.158)	-0.470*** (0.161)	-0.269 (0.197)	-0.397** (0.196)
연령대 (Ref.: 20대)	10대	-0.777** (0.333)	-0.279 (0.335)	-0.353 (0.472)	-0.255 (0.491)	-0.497 (0.451)	-2.220*** (0.650)
	30대	-0.123 (0.165)	0.167 (0.141)	-0.086 (0.201)	-0.147 (0.192)	-0.124 (0.243)	-0.155 (0.229)
	40대	0.118 (0.193)	0.040 (0.155)	0.023 (0.239)	0.111 (0.227)	0.036 (0.275)	0.146 (0.295)
	50대	0.267 (0.217)	0.265 (0.179)	0.102 (0.302)	0.027 (0.269)	0.309 (0.360)	-0.097 (0.370)
	60대 이상	-0.337 (0.259)	-0.066 (0.212)	-0.510 (0.362)	-0.362 (0.338)	-0.841* (0.474)	-0.652 (0.429)
학력 (Ref.: 고졸)	중졸 이하	-0.864*** (0.237)	-0.519*** (0.194)	-0.577 (0.368)	-0.182 (0.348)	-0.333 (0.404)	0.291 (0.390)
	초대졸	-0.215 (0.153)	-0.144 (0.129)	0.004 (0.194)	-0.263 (0.186)	-0.155 (0.226)	0.076 (0.224)
	대졸 이상	-0.169 (0.141)	-0.010 (0.114)	0.361** (0.164)	0.209 (0.166)	0.015 (0.207)	0.204 (0.212)
거주지 (Ref.: 비수도권)		0.249** (0.105)	0.347*** (0.088)	0.785*** (0.135)	0.571*** (0.137)	0.836*** (0.172)	0.134 (0.166)
사회보험급여 (Ref.: 비수급자)		0.551*** (0.174)	0.546*** (0.158)	0.864*** (0.291)	0.829*** (0.268)	1.330*** (0.341)	0.790** (0.315)
직업훈련 (Ref.: 없음)		-0.318 (0.205)	-0.204 (0.166)	-0.101 (0.274)	-0.152 (0.286)	-0.463 (0.318)	-0.125 (0.475)
자격증 (Ref.: 없음)		-0.222 (0.279)	-0.099 (0.124)	0.020 (0.275)	0.713* (0.414)	0.616 (0.526)	-0.139 (0.471)
싱글 (Ref.: 기혼)		0.243* (0.132)	0.269** (0.115)	0.376** (0.174)	0.332** (0.165)	0.466** (0.210)	0.373* (0.195)
건강상태 (Ref.: 보통)	나쁨	0.370 (0.241)	0.031 (0.190)	0.433 (0.364)	0.482 (0.335)	-0.435 (0.409)	0.782 (0.534)
	좋음	-0.125 (0.107)	0.047 (0.091)	0.014 (0.133)	-0.100 (0.137)	-0.126 (0.175)	0.067 (0.165)
가족수입 만족도 (Ref.: 보통)	불만족	0.470*** (0.127)	0.443*** (0.111)	0.656*** (0.179)	0.315* (0.164)	0.474** (0.197)	0.573*** (0.192)
	만족함	-0.283** (0.143)	-0.266** (0.112)	-0.294* (0.175)	-0.364* (0.190)	-0.576** (0.237)	-0.571** (0.251)
사회경제적 지위 향상 가능성 (Ref.: 부장모임)		-0.073 (0.110)	-0.021 (0.088)	-0.318** (0.137)	-0.155 (0.138)	-0.032 (0.169)	-0.322* (0.171)
LR test		140.0***	113.0***	130.0***	74.0***	83.8***	63.6***
Log-likelihood		-1,153	-1,469	-603	-630	-628	-574
AIC		2,349	2,981	1,251	1,304	1,300	1,193
관측치 수(명)		650	722	347	362	353	328

<표 8> 계속

		Log-normal 가속실패시간(AFT) 모형					
		2013	2012	2011	2010	2009	2008
종속변수		ln(T)					
상수항		2.280*** (0.308)	1.420*** (0.323)	1.780*** (0.263)	2.240*** (0.274)	2.670*** (0.368)	2.420*** (0.409)
성별 (Ref.: 남성)		-0.644*** (0.163)	-0.258 (0.166)	-0.523*** (0.138)	-0.691*** (0.148)	-0.714*** (0.189)	-0.613*** (0.202)
가구주 (Ref.: 비가구주)		-0.169 (0.173)	-0.179 (0.185)	-0.096 (0.149)	-0.372** (0.161)	-0.441** (0.218)	-0.422* (0.237)
연령대 (Ref.: 20대)	10대	0.285 (0.570)	-0.166 (0.572)	-0.901* (0.477)	0.182 (0.502)	-0.855 (0.585)	-1.360** (0.563)
	30대	-0.369* (0.200)	0.363 (0.223)	-0.175 (0.188)	0.108 (0.196)	-0.137 (0.259)	0.397 (0.279)
	40대	-0.200 (0.271)	0.388 (0.280)	0.081 (0.243)	-0.148 (0.250)	0.198 (0.316)	0.049 (0.347)
	50대	0.098 (0.314)	0.718** (0.354)	-0.123 (0.270)	0.737** (0.314)	0.794* (0.408)	0.853 (0.533)
	60대 이상	-0.615 (0.377)	0.647 (0.394)	0.218 (0.315)	0.123 (0.354)	-0.322 (0.471)	-0.054 (0.616)
학력 (Ref.: 고졸)	중졸 이하	0.270 (0.332)	-0.563* (0.339)	0.401 (0.291)	-0.053 (0.313)	0.403 (0.368)	-0.125 (0.452)
	초대졸	0.106 (0.207)	0.161 (0.223)	0.244 (0.181)	0.140 (0.199)	0.459* (0.254)	-0.027 (0.279)
	대졸 이상	0.142 (0.181)	0.141 (0.202)	0.410** (0.170)	0.158 (0.181)	0.146 (0.249)	0.247 (0.256)
거주지 (Ref.: 비수도권)		0.621*** (0.148)	0.482*** (0.149)	0.493*** (0.130)	0.089 (0.138)	0.139 (0.174)	0.073 (0.192)
사회보험급여 (Ref.: 비수급자)		0.733** (0.292)	0.595* (0.322)	0.521** (0.245)	0.698** (0.290)	0.758** (0.306)	0.343 (0.386)
직업훈련 (Ref.: 없음)		-0.419* (0.222)	0.026 (0.470)	-0.142 (0.307)	-0.380 (0.249)	-0.121 (0.398)	-0.107 (0.427)
자격증 (Ref.: 없음)		0.292 (0.390)	-0.626 (0.486)	0.077 (0.376)	-0.300 (0.309)	-0.151 (0.327)	0.085 (0.596)
싱글 (Ref.: 기혼)		-0.067 (0.181)	0.434** (0.193)	0.144 (0.159)	0.473*** (0.172)	0.434** (0.214)	-0.001 (0.249)
건강상태 (Ref.: 보통)	나쁨	-0.194 (0.330)	0.378 (0.366)	-0.138 (0.291)	-0.115 (0.323)	-0.686** (0.329)	-0.321 (0.414)
	좋음	-0.158 (0.150)	0.165 (0.157)	-0.371*** (0.137)	-0.129 (0.144)	-0.056 (0.186)	0.122 (0.208)
가족수입 만족도 (Ref.: 보통)	불만족	0.320* (0.167)	0.391** (0.165)	0.324** (0.145)	0.440*** (0.151)	0.396** (0.190)	0.935*** (0.218)
	만족함	-0.136 (0.203)	-0.022 (0.221)	-0.103 (0.196)	0.033 (0.215)	0.044 (0.301)	-0.326 (0.340)
사회경제적 지위 향상 가능성(Ref.: 부장 모됨)		-0.144 (0.149)	-0.106 (0.152)	-0.028 (0.131)	0.029 (0.142)	-0.144 (0.174)	-0.167 (0.204)
LR test		63.4***	51.7***	69.5***	68.7***	49.7***	48.6***
Log-likelihood		-843	-824	-892	-937	-757	-558
AIC		1,730	1,693	1,829	1,918	1,559	1,160
관측치 수(명)		409	434	458	453	501	338

[그림 4] 범주별 상대적 구직기간 추이(6개년 이상 유의미한 결과를 보인 범주만 표기)



<표 9> 분석 모형 ③ 추정 결과 (유의수준: * p<0.1, ** p<0.05, *** p<0.01, Ref.: 기준 범주)

Log-normal 가속실패시간(AFT) 모형, 종속변수: ln(T)						
③-1. 성별×연령대 교차항 포함 모형						
성별 (ref.: 남성)	-0.386*** (0.065)	연령대 (ref.: 20대)	10대	-0.965*** (0.214)	여성×10대	0.629** (0.276)
LR test: 817.5***			30대	0.113 (0.077)	여성×30대	-0.227** (0.105)
			40대	0.217** (0.097)	여성×40대	-0.244** (0.118)
			50대	0.679*** (0.114)	여성×50대	-0.659*** (0.137)
			60대 이상	-0.008 (0.120)	여성×60대 이상	-0.442*** (0.134)
③-2. 성별×학력 교차항 포함 모형						
성별 (ref.: 남성)	-0.537*** (0.081)	학력 (ref.: 고졸)	중졸	-0.005 (0.117)	여성×중졸	-0.305** (0.138)
LR test: 789.6***			초대졸	0.049 (0.079)	여성×초대졸	-0.055 (0.106)
			대졸 이상	0.095 (0.073)	여성×대졸 이상	0.072 (0.104)
③-3. 성별×싱글 교차항 포함 모형						
성별 (ref.: 남성)	-0.858*** (0.074)	싱글 (ref.: 기혼)	-0.027 (0.063)	여성×싱글	0.457*** (0.090)	
LR test: 746.2***						
③-4. 싱글×연령대 교차항 포함 모형						
싱글 (ref.: 기혼)	0.659*** (0.138)	연령대 (ref.: 20대)	10대	-0.555 (1.250)	10대×싱글	-0.016 (1.260)
LR test: 803.5***			30대	0.245* (0.143)	30대×싱글	-0.221 (0.159)
			40대	0.404*** (0.144)	40대×싱글	-0.440** (0.176)
			50대	0.686*** (0.151)	50대×싱글	-0.633*** (0.198)
			60대 이상	0.181 (0.161)	60대 이상×싱글	-0.740*** (0.184)
③-5. 거주지×연령대 교차항 포함 모형						
거주지 (ref.: 비수도권)	0.323*** (0.066)	연령대 (ref.: 20대)	10대	-0.640*** (0.180)	10대×수도권	0.185 (0.275)
LR test: 810.8***			30대	0.018 (0.080)	30대×수도권	-0.080 (0.103)
			40대	0.089 (0.095)	40대×수도권	-0.048 (0.114)
			50대	0.223** (0.110)	50대×수도권	0.160 (0.132)
			60대 이상	-0.569*** (0.123)	60대 이상×수도권	0.603*** (0.130)

<표 9> 계속

Log-normal 가속실패시간(AFT) 모형, 종속변수: ln(T)						
③-6. 가구주×연령대 교차항 포함 모형						
가구주 (ref.: 비가구주)	-0.477*** (0.083)	연령대 (ref.: 20대)	10대	-0.579*** (0.142)	가구주×10대	-0.116 (0.644)
LR test: 797.5***			30대	-0.048 (0.068)	가구주×30대	0.105 (0.117)
			40대	-0.079 (0.088)	가구주×40대	0.340*** (0.129)
			50대	0.048 (0.115)	가구주×50대	0.514*** (0.148)
			60대	-0.420*** (0.144)	가구주×	0.343**
			이상		60대 이상	(0.158)
③-7. 사회보장급여×연령대 교차항 포함 모형						
사회보장급여 (ref.: 비수급자)	0.998*** (0.277)	연령대 (ref.: 20대)	10대	-0.517*** (0.139)	10대× 당해계속 수급자	-2.290** (0.998)
LR test: 844.4***			30대	-0.061 (0.059)	30대× 당해계속 수급자	0.370 (0.333)
			40대	0.030 (0.072)	40대× 당해계속 수급자	0.155 (0.336)
			50대	0.289*** (0.087)	50대× 당해계속 수급자	0.063 (0.356)
			60대	0.038 (0.108)	60대 이상 당해계속 수급자	-0.972*** (0.298)
			이상			
③-8. 직업훈련×연령대 교차항 포함 모형						
직업훈련 (ref.: 없음)	-0.420*** (0.123)	연령대 (ref.: 20대)	10대	-0.581*** (0.143)	10대× 직업훈련 있음	0.380 (0.539)
LR test: 785.1***			30대	-0.050 (0.060)	30대× 직업훈련 있음	0.293 (0.194)
			40대	0.034 (0.073)	40대× 직업훈련 있음	0.392 (0.253)
			50대	0.293*** (0.088)	50대× 직업훈련 있음	0.252 (0.265)
			60대	-0.258** (0.102)	60대 이상× 직업훈련 있음	0.328 (0.354)
			이상			

V. 결론 및 시사점

본 연구는 구직자 개인 특성이 구직기간의 차이를 결정한다는 가설에서 출발해 한국노동패널조사(KLIPS)의 다년도 개인 수준 데이터를 바탕으로 잠재적 결정요인들을 검증하는 것을 목적을 가진다. 목표변수인 구직기간은 항상 양수 값을 갖고, 정규분포를 띠지 않으며, 중도절단이 빈번하게 발생하기 때문에, 이를 다루기 적합한 생존분석을 주요 분석 기법으로 활용하였다.

생존분석 수행 시 범주간 생존확률을 비모수적으로 추정해 그 차이를 확인하는 카플란-마이어(K-M) 분석법과 함께 여러 요인들을 함께 고려 가능한 Cox 비례위험(PH) 모형을 전제로 한 분석법을 이용하였다. 다만 Cox PH 모형을 이용하고자 할 때 시간의 흐름에 따라 위험비가 일정하다는 비례위험 가정을 만족해야 높은 적합성을 가진다. 하지만 다수의 선행연구에서 해당 가정에 대한 검정 과정이 생략되었다. 이에 따라 비례위험 가정이 불필요한 가속실패시간(AFT) 모형을 전제로 한 분석법을 추가로 도입해 기존 연구들과 차별화하여 분석 결과의 적합성을 높였다. 나아가 다년도 데이터를 통합한 모형, 각 년도 데이터만 이용한 모형, 교차항이 포함된 모형 등 3가지 모형을 도입하여 다각적인 분석을 시도하였다.

검정 결과에 따르면 비례위험 가정은 성립되지 않았으며, 로그-정규 분포를 가정한 AFT 모형의 적합도가 가장 높게 나타났다. 일부 공변량의 효과는 통념과 다소 다르게 나타났다. 가령 여성이 남성보다, 비수도권 거주자가 수도권 거주자에 비해, 고졸자가 대졸자보다, 그리고 사회보험급여 비수급자가 구직기간 당해 및 계속 수급한 사람과 비교해 구직기간이 상대적으로 짧아지는 효과를 보였다. 하지만 생존분석 결과는 구직기간이라는 양적 부문을 설명해줄 뿐 임금수준, 고용 안전성, 근로여건 등 취업의 질적 부문을 설명해주지 못한다. 다시 말해 구직기간이 짧다고 해서 해당 범주에 속한 구직자들이 반드시 더 나은 일자리를 갖는다는 의미가 아니다. 또한 구직자가 특정 요인의 범주에 해당하는지 여부에 따른 생존기간의 차이를 비교할 수 있어도 그 차이를 발생시키는 구체적 원인은 추측의 영역에 속한다. 예를 들어 대졸자가 고졸자보다 상대적으로 구직기간이 길게 나타났는데 학력 프리미엄이 작용하지 않았다고 단정지을 수 없다. 따라서 결과 해석 시 유의해야 할 필요가 있다.

아울러 생존분석의 또다른 한계점으로 내생성(endogeneity) 문제를 제기할 수 있다. 비록 대부분 구직기간이 몇 주에 불과해 구직기간 중 구직자 특성이 변화할 가능성이 낮을 것으로 생각되지만, 원인과 결과가 뒤바뀌는 이른바 역의 인과관계(reverse causality)에 따른 내생성이 고질적으로 잠재되어 있다. 예컨대 구직기간이 충분히 길어질수록 자격증 취득 혹은 직업훈련을 받게 될 가능성이 높아질 것이다. 중요한 요인인데 분석 모형에 반영하지 못한 이른바 누락변수(omitted variable) 역시 내생성을 유발시키는 또 다른 요소다. 유감스럽게도 생존분석에서 내생성 문제를 확실하게 해결할 방법은 아직까지 고안되지 않았다. 그러므로 엄밀히 보자면 생존분석 결과는 변수들간 인과관계가 아닌 상관관계(correlation relationship)를 나타낸다고 할 수 있다. 내생성 해결은 후속 연구의 중요한 과제라 판단된다.

참고문헌

- 강봉석, 염동문. (2015). 「장애인 실업자의 취업에 영향을 미치는 결정요인에 관한 연구」. 재활복지 공학회논문지, 9(2), 115-120.
- 강주연, 오유, 김기승. (2015). 「대졸자 특성과 취업 소요기간」. 산업관계연구, 25(2), 31-49.
- 고영근, 안태현. (2018). 경기변동과 고용 동학에 대한 분석: 집단 간 취업-미취업 이행확률 비교를 중심으로」. 노동경제논집, 41(2), 31-59.
- 김민석, 안은비, 유경숙, 정효숙. (2018). 「대학생 취업 소요기간에 영향을 미치는 요인 분석」. 한국콘텐츠학회논문지, 18(8), 27-35.
- 류장수. (2005). 「지방대학 졸업생의 노동시장 성과 분석: 수도권대학 졸업생과의 비교」. 노동경제논집, 28(2), 1-27.
- 송일호, 박명호. (2012). 「고령자와 준고령자의 취업결정요인 비교분석」. 사회과학연구, 19(2), 7-26.
- 성지미, 안주엽. (2006). 「중고령자 취업 결정요인」. 노동정책연구, 6(1), 39-74.
- _____. (2012). 「취업사교육과 첫 일자리」. 한국경제연구, 30(3), 5-46.
- 양정빈. (2015). 「여성장애인의 취업 및 임금수준 결정요인 분석」. 보건사회연구, 35(2), 226-253.
- 이규용, 김용현. (2003). 「대졸 청년층의 노동시장 성과 결정요인」. 노동정책연구, 3(2), 69-93.
- 이상록. (2000). 「실직기간 구직활동이 실직자의 재취업에 미치는 영향 분석」. 한국사회복지학, 43, 299-327.
- 이병희. (2002). 「노동시장 이행 초기 경험의 지속성에 관한 연구」. 노동정책연구, 2(1), 1-18.
- 임찬영. (2018). 「비인지적 요인이 취업에 미치는 영향: 구직기간과 근속기간 분석을 중심으로」. Journal of The Korean Data Analysis Society, 20(6), 3069-3085.
- 장지연, 신현구. (2008). 「중고령자 취업결정요인의 국가간 비교: 한국, 미국, 스웨덴, 독일. 중고령자 노동시장 국제비교」. 서울: 한국노동연구원, 43-84.
- 조우현. (1995). 「청년층 노동자의 고용문제와 실업확율의 결정요인 분석」. 노동경제논집, 18(1), 107-128.
- 전이상. (2002). 「장애인 취업의 결정요인에 관한 연구」. 박사학위논문, 전남대학교.
- 채창균, 김태기. (2009). 「대졸 청년층의 취업 성과 결정 요인 분석」. 직업교육연구, 28(2), 89-107.
- 통계청. (2020). 『2018년 일자리이동통계』.
- 홍성표, 정진철. (2015). 「자격취득이 장애인 재취업 및 구직기간에 미치는 영향」. 장애와 고용, 25(3), 169-196.
- Berg, G. J., Lomwel, A. G. C., & Ours, J. C. V. (1998). "Unemployment dynamics and age" (No. 0048)".
- Bratberg, E., & Nilsen, Ø. A. (1998). "Transition from school to work: Search time and job duration" (No. 27). *IZA Discussion Papers*.

- Caliendo, M., Tatsiramos, K., & Uhlenhorff, A. (2013). "Benefit duration, unemployment duration and job match quality: a regression discontinuity approach". *Journal of applied econometrics*, 28(4), 604-627.
- Cox, D. R. (1972). "Regression models and life tables. *Journal of the Royal Statistical Society: Series B*" (Methodological), 34(2), 187-202.
- Cox, D. R., & Oakes, D. (1984). *Analysis of survival data (Vol. 21)*. CRC Press.
- Eckstein, Z., & Wolpin, K. I. (1995). "Duration to first job and the return to schooling: Estimates from a search-matching model". *The Review of Economic Studies*, 62(2), 263-286.
- Kaplan, E. L., & Meier, P. (1958). "Nonparametric estimation from incomplete observations". *Journal of the American statistical association*, 53(282), 457-481.
- Klein, J. P., & Goel, P. K. (Eds.). (2013). *Survival analysis: state of the art (Vol. 211)*. Springer Science & Business Media.
- Lawless, J. F. (2011). *Statistical models and methods for lifetime data (Vol. 362)*. John Wiley & Sons.
- Lynch, L. M. (1985). "State dependency in youth unemployment: A lost generation?". *Journal of Econometrics*, 28(1), 71-84.
- Miyamoto, Suphaphiphat. (2020). "Mitigating Long-term Unemployment in Europe", *IMF Working Papare*.
- Orbe, J., Ferreira, E., & Núñez Antón, V. (2002). "Comparing proportional hazards and accelerated failure time models for survival analysis". *Statistics in medicine*, 21(22), 3493-3510.
- Therneau, T. M. (2020). "coxme: Mixed effects Cox models. R package version 2.2-5. 2015".
- Van Den Berg, G. J., & Van Ours, J. C. (1999). "Duration dependence and heterogeneity in French youth unemployment durations". *Journal of Population Economics*, 12(2), 273-285.