Journal of the Korean Data & Information Science Society
2011, **22**(1), 1–8

한국데이터정보과학회지

한국노동패널 탈락 분석

이상협[1] · 박찬용[2] · 정성석[3] · 최혜미[4]

[1234]전북대학교 통계학과

접수 2010년 10월 23일, 수정 2010년 12월 19일, 게재확정 2010년 12월 24일

요 약

동일한 표본에 대하여 여러 시점에 걸쳐 반복적으로 조사되는 패널자료의 경우 표본의 탈락은 조사 결과의 신뢰성과 표본의 대표성 등에 문제를 야기할 수 있어 표본 탈락의 경향이나 요인에 대한 분석을 통해 표본의 탈락률을 낮추는 방안이 모색되어져야 할 것이다. 본 연구에서는 11년간 (1998∼2008) 한국노동패널 자료의 표본 탈락 시간에 대한 정보를 사용하여 생명표방법과 Cox 비례위험모형으로 표본 탈락에 영향을 주는 요인을 분석하였다.

주요용어: 생존분석, 패널탈락, 한국노동패널.

1. 서론

1960년대부터 시작한 미국의 Personal Survey of Income Dynamics (PSID)나 National Longitudinal Survey (NLS) 등에 비하면 우리나라는 패널조사의 역사가 그리 길지 않다. 최근 패널조사에 관한 수요가 급격히 증가하고 있어 한국노동연구원에서는 1998년부터 한국노동패널조사 (KLIPS)를 실시하고 있고, 고용정보원에서는 2001년부터 산업·직업별 고용구조조사 (OES)와 청년패널 (Youth Panel)을 실시하고 있으며 청소년정책연구원, 직업능력개발원, 보건사회연구원 등에서 여러 형태의 패널조사들이 이루어지고 있다. 패널조사는 특정정책이나 프로그램이 목표 집단에 미치는 인과적인 효과를 직접적으로 평가하거나 예측해 볼 수 있는 개인차원의 정보를 얻을 수 있다는 장점이 있지만 , 패널을 구축하기 위해 막대한 예산과 시간, 인원이 투입되며 선정된 표본에 대한 조사가 1회에 그치지 않고 이후에도 계속되기 때문에 패널탈락으로 인해 패널의 대표성 확보가 어렵다는 단점이 있다. 응답거절, 거주지 이전, 사망 등의 이유로 응답대상의 소실, 즉 표본탈락은 표본의 대표성과 표본유지율에 영향을 주고, 표본탈락이 무작위로 발생하는 것이 아니라 특정 집단에서 주로 발생하게 된다면 무응답으로 인한 편의가 발생하게 된다. 따라서 표본유지율 제고를 위해 탈락 요인을 파악하는 것은 중요한 문제라 할 수 있다.

패널자료에서 표본탈락 요인에 대표적인 선행 연구를 미국의 PSID자료 중심으로 살펴보면 Becketti 등 (1988), Fitzgerald 등 (1998), Lillard와 Panis (1998), Zabel (1998) 등이 있다. 이들 연구 대부분은 패널탈락 요인을 성별, 혼인상태, 학력 등과 같은 응답자의 인구통계학적 특성 측면에서 분석한 반면 Zabel (1998)은 인구통계학적 변수들의 영향보다는 면접원과 면접과정이 패널탈락에 많은 영향을 미친다고 주장하였다. 국내에서도 패널탈락요인에 대한 분석이 점점 활발해지고 있는데, 한국노동패널조

[1] (561-756) 전북 전주시 덕진구 덕진동 1가 664-14, 전북대학교 통계학과, 석사과정.
[2] (561-756) 전북 전주시 덕진구 덕진동 1가 664-14, 전북대학교 통계학과, 석사과정.
[3] (561-756) 전북 전주시 덕진구 덕진동 1가 664-14, 전북대학교 통계학과 (응용통계연구소), 교수.
[4] 교신저자: (561-756) 전북 전주시 덕진구 덕진동 1가 664-14, 전북대학교 통계학과 (응용통계연구소), 조교수. E-mail: hchoi@jbnu.ac.kr

사 (KLIPS)의 패널탈락에 관한 연구로는 이상호 (2005)가 프로빗 모형을 이용하여 가구주가 연령이 낮을수록, 미혼인 경우 탈락할 위험이 높게 나타났고, 성별에 대한 차이는 유의미하지 않으며, 고소득층과 실업자의 경우 탈락위험이 높게 나타나는 양극화 현상이 있는 것을 발견한 연구와 신선옥 (2008)의 면접원의 교육수준이 높고 경력이 많을수록 조사의 협조도에 긍정적인 영향을 끼치며, 숙련도가 높으면 조사 성공부수는 늘어나지만 다음해의 성공률에 큰 영향을 미치지는 않는다는 연구 결과가 있다. 이외에도 다수의 연구결과가 있으나, 주로 로짓이나 프로빗 모형 등 이산확률 모형을 이용하여 분석하였다. 본 논문에서는 KLIPS 11년간 (1998∼2008)의 자료를 이용하여 패널탈락요인을 살펴보았는데, 표본탈락이 관측될 때까지의 시간에 비례위험모형과 생명표방법 (life-table method)과 같은 생존분석 기법 (이재원 등, 2007; Cho 등, 2008)을 적용하여 분석하였다.

본 논문은 다음과 같이 구성되어있다. 2절에서는 KLIPS 자료에 대해 간단히 설명하고, 3절에서는 생존분석 기법을 이용한 분석결과를 정리하고 있다. 마지막으로 4절에서는 본 논문의 결과를 선행논문과 비교하며 간략하게 요약하였다.

2. 분석 자료

한국노동패널조사 (KLIPS)는 1998년에 1차 조사가 시작되어 1년 1회 실시되고 있으며, 경제활동 및 노동시장이동, 소득활동 및 소비, 교육 및 직업훈련, 사회생활 등에 관한 패널조사이다. 1995년 인구주택총조사의 10% 표본조사구 중 제주도를 제외한 전국의 도시지역만을 대상으로 층화한 후 5,000가구를 추출하여 추적조사하고 있다. KLIPS의 원표본 가구 유지율은 외국의 대표적인 패널조사들에 비해 상대적으로 높은 유지율을 보이고 있다. 예로 11차년도 KLIPS 원표본 유지율이 74.2%로 미국의 PSID (67%)나 영국의 BHPS (68.4%)보다 높다. 패널의 11년간 성별, 학력, 직업유무의 표본비율의 변화를 살펴보기 위해 표 2.1과 같이 1차년도와 11차년도의 표본비율을 정리하였다.

표 2.1 1차 조사와 11차 조사의 가구원 표본비율

특성	1차 조사		11차 조사	
남자	6,471	(49.0%)	5,671	(48.3%)
여자	6,850	(51.0%)	6,063	(51.7%)
무학	856	(6.0%)	544	(4.6%)
초등학교	1,741	(13.0%)	1,494	(12.7%)
중학교	1,765	(13.0%)	1,420	(12.1%)
고등학교	5,473	(41.0%)	4,026	(34.3%)
2년제 대학, 전문대학	1,015	(8.0%)	1,449	(12.3%)
4년제 대학 이상	2,460	(19.0%)	2,799	(23.9%)
기타	1	(0.0%)	2	(0.0%)
취업자	6,425	(48.0%)	6,374	(54.3%)
미취업자	6,896	(52.0%)	5,360	(45.7%)
전체	13,321	(100.0%)	11,734	(100.0%)

학력에 대하여 기타 범주는 제외하고 검정을 실시한 결과 유의확률이 0.001이하로 학력의 분포가 차이가 남을 확인할 수 있고, 직업유무도 마찬가지 결과를 보였다. 즉, 조사가 여러 해에 걸쳐 진행됨에 따라 특성별 분포에 변화가 나타났음을 알 수 있다. 일반적으로 패널조사의 경우에는 모집단의 변동과 패널의 무응답 때문에 원표본이 가지고 있던 대표성 문제가 점차 악화될 소지가 있는데, 표본의 대표성을 보완해 주기 위해 김영원 등 (2005), 김규성 등 (2005)가 KLIPS의 가중치에 대한 보정을 제안하고 있다. 그러나 본 논문에서는 KLIPS의 표본의 대표성 문제는 고려하지 않고, 인구통계학적 특성 (성별, 연령, 지역, 결혼상태, 학력), 직업유무, 거주형태와 조사방법 (표 2.2)이 패널탈락에 미치는 영향에 대

하여 KLIPS 11년간 (1998∼2008)의 가구주에 관한 자료만을 이용하여 살펴보았다. 패널탈락을 각 년차에서 응답하지 않은 경우로 정의하고, 탈락 후 재진입한 경우는 새로운 표본으로 간주하였다. 예를 들어 응답자가 1년차에서 3년차까지 응답하고 4년차에 응답하지 않으면 탈락이고, 5년차에 재진입하면 이 응답자는 새로운 응답자로 간주하였다. 또한, 시간 변수 값은 2008년에 탈락하지 않은 경우를 중도절단으로 정의하여, 1차에서 11차까지 모두 응답한 가구주 경우는 중도 절단된 11이고, 4차에 새로 진입하여 11차까지 모두 응답한 경우 한 가구주 경우는 중도 절단된 8이다.

표 2.2 사용된 변수 설명

변 수	변수의 값 설명
성 별	남성=0, 여성=1
연 령	30세 미만, 30세 이상∼60세 미만, 60세 이상
지 역	서울, 광역시, 경기도, 강원·충청도, 전라·제주도, 경상도
직업유무	취업=0, 미취업=1
결혼상태	미혼, 기혼, 이혼 및 별거
학 력	초등학교 이하, 중·고등학교, 전문대 이상
조사방법	면접, 유치, 전화·혼합
입주형태	자가, 전세, 월세, 기타
주거형태	단독주택, 아파트, 연립주택, 다세대주택, 상가주택·기타
탈락여부	응답=0, 탈락=1
시간	표본유지시간

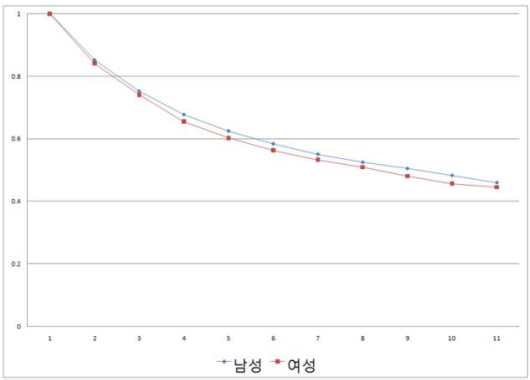
3. 패널탈락 분석

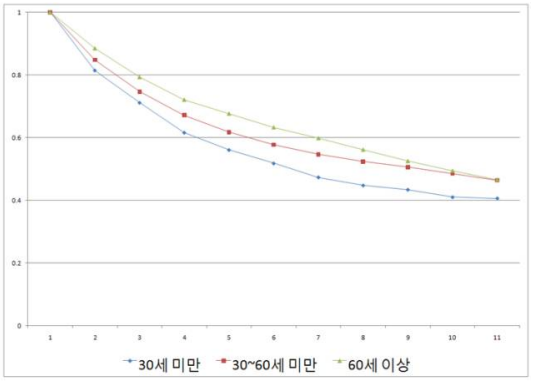
먼저 생명표방법을 이용하여 각 변수의 범주별로 패널탈락이 관측될 때까지의 시간 분포를 비교하고, Cox의 비례위험모형을 이용하여 탈락에 영향을 주는 변수를 찾아서 탈락 위험함수에 미치는 영향을 살펴보았다.

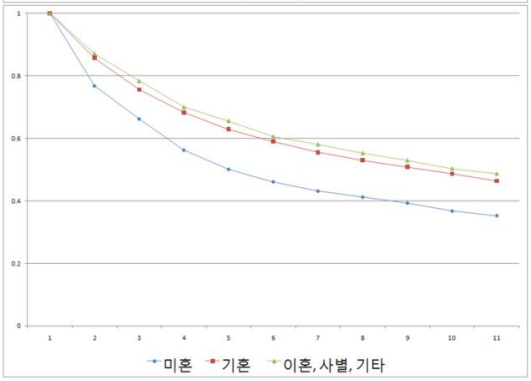
3.1. 생명표방법 분석 결과

생명표방법은 본 연구의 자료와 같이 표본의 크기가 큰 경우 생존곡선 추정에 자주 이용되는 방법으로 변수의 각 범주별 (1차조사 기준) 패널 생존곡선을 이 방법으로 추정하였고, 로그-순위 검정 (log-rank test)과 윌콕슨 (Wilcoxon) 검정을 이용하여 추정된 생존곡선의 동일성을 검정하였다. 김민수 등 (2010)은 기업의 부도율 비교를 위하여, 박희창 (2009)는 인터넷 고객 수명 분석을 위하여 생명표분석을 사용한바있다. 추정된 생존곡선의 교차여부를 참조하여 생존곡선이 평행한 경우 로그-순위 검정을 사용하고, 교차한 경우는 윌콕슨검정통계량을 사용하였다. 지역·조사방법·입주형태·주거형태의 생존곡선이 그룹별로 교차되어 윌콕슨검정통계량을 나머지 변수들은 로그순위검정통계량을 사용하였다. 그림 3.1은 여러변수 중 성별·연령·결혼상태의 범주별 추정된 생존곡선을 보여주고 있다. 그림 3.1의 표의 각 변수별 유의확률을 참조하여 생존곡선을 비교한 결과 성별을 제외한 모든 변수가 그룹별로 차이가 있음을 알 수 있다. 추정된 곡선들을 살펴보면 대체로 탈락위험률이 초기에 높았으나 점점 낮아짐을 알 수 있고, 즉 표본 패널이 안정화되는 경향을 보임을 알 수 있다. 그룹별 유의한 차이를 보인 연령·학력·지역·직업유무·결혼상태·조사방법·입주형태·주거형태를 자세히 살펴보면 다음과 같다.

연령별로는 60세 이상의 탈락률이 가장 낮게 나타났고 30세 미만의 탈락률이 가장 높게 나타났으나, 7차 이후부터 60세 이상의 탈락률과 30세∼60세미만의 탈락률 차이가 점차 줄어들어 11차에는 거의 같아지는 모습을 보이고 있다. 3차까지는 학력이 높을수록 탈락률이 높게 나타나 학력에 따른 탈락률의 차이가 커지는 것으로 보이지만, 이후로는 생존곡선이 거의 평행하게 나타났다. 지역별로 살펴보면 서







변수	log-rank 통계량(유의확률)		Wilcoxon 통계량(유의확률)	
성별	0.7941	0.3729	1.0741	0.3000
연령	21.5211	<.0001	31.4643	<.0001
학력	28.5242	<.0001	37.3384	<.0001
지역	62.3586	<.0001	60.6081	<.0001
직업유무	19.8528	<.0001	13.7077	0.0002
결혼상태	60.2432	<.0001	69.2852	<.0001
조사방법	14.8451	0.0006	15.0805	0.0005
입주형태	35.7829	<.0001	35.6321	<.0001
주거형태	18.0929	0.0012	17.6698	0.0014

그림 3.1 생존곡선과 검정결과

울의 탈락률이 가장 높고 경상도가 탈락률이 가장 낮은 것으로 나타났다. 서울·경기와 광역시·강원·충청의 경우 초기에는 탈락률이 유사하지만 5차년도 이후부터는 탈락률에 차이가 발생하고 경상·전라·제주는 탈락률이 9차년도 까지는 서로 교차하며 큰 차이를 보이지는 않지만 그 이후부터는 차이가 커지는 모습을 보이고 있다. 미취업자의 탈락률이 전체 조사기간에 걸쳐 높아 시간이 지날수록 취업자와 미취업자의 생존곡선이 더 큰 차이를 보이고 있다. 결혼경험이 있는 그룹 즉 기혼 그룹과 이혼 및 사별 그룹은 큰 차이를 보이고 있지 않으나, 미혼과 결혼경험 그룹은 큰 차이를 보임을 알 수 있다. 미혼 그룹은 그룹 특성상 이동성이 높아서 이러한 경향을 보인다고 짐작할 수 있다. 유치 조사의 경우 조사 초기에 탈락률이 가장 낮은 것으로 나타났으나 5차년도 이후부터 탈락률이 가장 높은 것으로 나타났고, 면접의 경우 3차년도 이후부터 가장 낮은 탈락률을 보이고 있다. 자가의 경우 모든 구간에서 가장 낮은 탈락률을 보이고 있고, 전세·월세의 경우는 탈락률의 차이가 커지다가 7차년도 이후부터 차이가 작아지는 모습을 보이고 있다. 단독주택·상가·기타와 아파트·연립주택·다세대주택의 초기 탈락률은 유사하지만 7차년도 이후부터 상가·기타의 탈락률이 완만해져 9차년도 이후에는 단독주택보다 탈락률이 낮아지고, 4차년도 이후부터 아파트·연립주택의 탈락률은 완만해져 다세대주택과 차이가 커지다가 7차년도 이후부터 다시 차이가 작아지는 모습을 보이고 있다.

3.2. 비례위험모형 분석 결과

패널생존시간을 T라고 하고 이에 영향을 주는 k개의 공변량 $x_1, x_2, \cdots, x_k$가 있을 때 Cox의 비례위험모형은 다음과 같이 나타낼 수 있다.

$$h(t) = h_0(t)exp(\beta_1 x_1 + \cdots + \beta_k x_k)$$

여기서 $h(t)$는 공변량들을 고려했을 때의 위험함수 (hazard function)이고 $h_0(t)$는 기저위험함수 (baseline hazard function)이다. 사용된 변수 중 성별을 제외한 나머지 변수들은 시간에 따라 그 값이 변하는 공변량으로 위 비례위험모형은 이러한 경우에도 분석할 수 있는 장점이 있다. Cox의 비례위험모형은 생존분포에 대한 가정이 없기 때문에 가능도함수를 이용하여 계수를 추정할 수가 없어서 부분우도함수를 이용하여 계수를 추정하는데 이 경우는 관측된 생존시간 사이에 동일한 값이 존재하지 않는 경우를 가정하고 있다. KLIPS와 같은 패널조사의 경우에는 동일한 생존시간을 갖는 관측대상이 많이 존재하여 적합하지 않아서 생존시간의 중복이 있는 경우를 가정하여 부분우도함수를 계산하는 근사적인 방법을 사용하는 것이 타당하다. 대표적인 근사적인 방법 중 Efron방법의 추정치가 Breslow방법보다 실제 계수에 더 근사하게 나타난다고 알려져 있어 (DeLong 등, 1994; Hertz-Picciotto 등, 1997) 본 논문에서는 통계패키지인 SAS의 PROC PHREG에서 Efron방법을 이용하여 분석하였다.

표 3.1 비례위험모형 회귀계수 추정치

변수 (기준범주)		β추정치	표준오차	e^{β}추정치
취업여부 (취업)	미취업	0.2471	0.0378	1.28**
연령 (30세미만)	30~60세	-0.0073	0.0528	0.993
	60세 이상	-0.3328	0.0909	0.717**
	60세 이상*log(time)	0.2722	0.0559	1.313**
지역 (서울)	광역시	-0.1760	0.0419	0.839**
	경기도	-0.0635	0.0448	0.938
	강원도, 충청도	-0.1419	0.0649	0.868*
	전라도, 제주도	-0.2949	0.0723	0.745**
	경상도	-0.2697	0.0586	0.764**
학력 (초등학교이하)	중·고등학교	0.0582	0.0493	1.060
	전문대 이상	0.1248	0.0560	1.133*
결혼상태 (미혼)	기혼	-0.2757	0.0553	0.759**
	이혼, 별거	-0.3503	0.0696	0.705**
조사방법 (면접)	유치	0.1209	0.0455	1.129**
	전화, 혼합	0.1060	0.0427	1.112*
입주형태 (자가)	전세	0.0951	0.0370	1.1*
	월세	0.2124	0.0533	1.237**
	기타	0.1320	0.0691	1.141
주거형태 (단독주택)	아파트	0.0701	0.0393	1.073
	연립주택	0.0277	0.0599	1.028
	다세대주택	-0.0129	0.0495	0.987
	상가주택,기타	-0.1635	0.0750	0.849*

*p<0.05, **p<0.01

비례위험모형을 적용하기 위해서는 먼저 포함될 각 공변량에 대해 비례성 가정이 (Lee 등, 2003) 성립해야 한다. 이를 위해 앞 절에서 구한 범주별 생존곡선이 교차되지 않은 변수, 성·학력·취업·결혼상태의 로그순위 검정 유의확률이 0.25이하인 학력·취업·결혼상태를 공변량으로 먼저 포함시켰다. 생존곡선이 교차하는 형태인 연령·지역·조사방법·입주형태·주거형태의 경우는 비례성 가정에 위배될 수 있기 때문에 모두 모형에 포함시키고, 응답기간과의 교호작용항을 포함시켜 비례성 가정을 검정한 결과, 연령

만이 교호작용이 유의하여 모형에 포함시켰다. 여기에서 응답기간은 자연로그를 취한 값을 사용하였고, 비례위험모형을 이용하여 분석한 결과가 표 3.3과 같다. 결과를 자세히 살펴보면 취업여부별로는 미취업자의 탈락위험은 취업자에 비해 1.28배 (e^{β}추정치 열 참조) 높은 것으로 나타났다. 연령별로는 30세 미만과 30~60세인 경우의 탈락위험은 차이가 없지만, 60세 이상인 경우는 0.717배로 고연령인 경우 탈락위험이 낮아지는 것으로 나타났으나, 응답기간이 길어질 경우 60세 이상의 탈락 위험은 증가하는 것을 관찰하였다. 지역별로는 서울에 비해 경기도 제외한 다른 지역에서 탈락 위험이 유의한 차이를 보이며 감소하는 것으로 나타났고, 남부지방보다 중부지방의 탈락 위험이 더 높은 모습을 보이고 있다. 학력별로는 초등학교 이하에 비해 중·고등학교의 경우 탈락 위험이 유의한 차이가 없으며, 전문대 이상인 경우 탈락할 위험이 1.133배 높은 것으로 나타나 학력이 높을수록 탈락할 위험이 높은 것으로 나타났다. 결혼상태별로는 미혼에 비해 기혼인 경우 탈락할 위험이 0.759배, 이혼·별거인 경우 0.705배로, 미혼인 경우에 탈락할 위험이 가장 높은 것으로 나타났다. 조사방법별로는 면접에 비해 유치의 경우 탈락할 위험이 1.129배 높아지는 것으로 나타났고, 전화·혼합의 경우 1.112배로 면접에 비해 탈락할 위험이 높은 것으로 나타났다. 입주형태별로는 자가에 비해 전세인 경우 탈락할 위험이 1.1배, 월세인 경우 1.237배로 나타났고, 기타는 차이가 없는 것으로 나타나 자가인 경우에 탈락할 위험이 가장 낮은 것으로 나타났다. 주거형태별로는 단독주택에 비해 아파트·연립주택·다세대주택인 경우 탈락할 위험의 차이가 없는 것으로 나타났고, 상가주택·기타인 경우 0.849배로 탈락할 위험이 가장 낮은 것으로 나타났다.

4. 결론

한국노동패널 가구주는 고학력일수록, 미혼의 경우 그리고 미취업자가 취업자에 비해 패널탈락위험이 높은 것으로 나타나 경제활동인구조사와 도시가계조사 자료를 이용한 이지연 등 (2006)의 연구결과와 유사하게 나타났다. 고연령인 경우의 탈락위험이 낮은 면은 이전의 선행연구들 (김대일 등, 2000; 이상호, 2005; 이지연 등, 2006)과 유사하나, 조사기간이 길어질수록 탈락위험이 높아지는 것을 볼 수 있는데 사망이나 건강상의 이유로 인한 탈락 때문이라고 짐작된다. 결혼상태 면에서는 미혼의 탈락위험이 높게 나타났는데 이것은 개인 상태의 변동 (이동) 가능성이 높은 응답자의 탈락위험이 높다는 천영민 등 (2009)의 연구결과와 일치한다. 지역적인 차이를 보면 대체적으로 서울·경기, 중부지방, 남부지방 순으로 탈락위험이 높게 나타났는데 이는 지역적인 특성에 따른 차이라고 생각된다. 입주형태의 경우 자가인 경우 낮은 탈락위험은 이지연 등 (2006)의 연구결과와 유사하게 나타났다. 선행연구들과 비교해보면 전체적인 결과는 유사하지만 부분적인 차이가 존재하는데 이는 분석대상의 차이로 인한 결과라고 생각된다. 이는 선행연구들에서도 지적한 내용으로 각각의 패널자료에 대한 개별적인 분석이 필요함을 알 수 있다.

비록 11회 조사된 패널자료를 사용하였지만, 생명표방법을 통한 각 변수의 범주별 패널탈락시간분포를 추정하여 시간에 따른 탈락률 변화를 시각화와 비례위험모형을 이용하여 생명표방법에서 확인한 결과들의 변수의 범주별 위험률 차이를 계산하고, 시간에 따른 탈락률변화의 유의성 검정 또한 실시하여 생명표방법에 의한 결과를 좀 더 구체화할 수 있었다. 패널조사기간이 길어질수록 생존분석기법을 적용한 분석이 더욱 다양하고 효율적으로 적용되리라 본다.

참고문헌

김규성, 황영은, 박진우 (2005). 패널조사에서 가중치 부여 방법 및 효과에 관한 연구, <제6회 한국노동패널발표논문집>, (한국노동연구원).

김대일, 남재량, 류근관 (2000). 한국노동패널 표본의 대표성과 패널조사 표본 이탈자의 특성 연구, <노동경제론집>, **23**, 1-33.

김영원, 김재광, 이기재, 조유미 (2005). 한국노동패널 표본의 대표성과 가중치 보정 방법, <제6회 한국노동패널발표논문집>, (한국노동연구원).

박진경, 오광호, 김민수 (2010). 생존분석기법을 이용한 건설업과 타 업종간의 부도율 비교 분석, <한국데이터정보과학회지>, **21**, 747-756.

박희창 (2009). 생명표를 이용한 회원유형별 인터넷 고객 수명 분석, <한국데이터정보과학회지>, **20** ,109-115.

신선옥 (2008). 한국노동패널조사의 응답자 태도에 면접원이 미치는 효과, <노동리뷰>, **1**, 74-82.

이상호 (2005). 한국노동패널의 표본이탈 분석 -가구소득을 중심으로-, <노동리뷰>, **11**, 66-80.

이재원, 박미라, 유한나 (2007). <생명과학연구를 위한 통계적 방법>, 자유 아카데미, 서울.

이지연, 김진 (2006). 횡단조사자료 종단화의 가치와 한계: 경제활동인구조사와 도시가계조사, <한국인구학>, **29**, 159-188.

천영민, 윤정혜, 오민홍 (2009). 대졸자 직업이동 경로조사에서 패널탈락분석, <응용통계연구>, **22**, 981-993.

Becketti, S., William, G., Lee, L. and Finis, W. (1988). The panel study of income dynamics after fourteen years : An evaluation. *Journal of Labor Economics*, **6**, 472-492.

Cho, Mi-Soon and Kim, Soon-kwi (2008). Comparative study on statistical packages analyzing survival model - SAS, SPSS, STATA-, *Journal of the Korean Data & Information Science Society*, **19**, 487-496.

DeLong, D. M., Guirguis G. H. and So, Y. C. (1994). Efficient computation of suset selection probabilities with application to Cox regression, *Biometrika*, **81**, 607-611.

Fitzgerald, J., Gottschalk, P. and Moffit, R. (1998). An analysis of sample attrition in panel data, *The Journal of Human Resources*, **33**, 251-299.

Hertz-Picciotto, I. and Rockhill, B. (1997). Validity and efficiency of approximation methods for tied survival times in Cox regression, *Biometrics*, **53**, 1151-1156.

Lee, SeongWon and Kim, JuSeong (2003). A test procedure for checking the proportionality between hazard functions, *Journal of the Korean Data & Information Science Society*, **14**, 561-570.

Lillard, L. A. and Panis, C. W. A. (1998). Panel attrition from the PSID, *The Journal of Human Resources*, **33**, 437-457.

Zabel, J. E. (1998). An analysis of attrition in the panel study of income dynamics and the survey of income and program participation with and application to a model of labor, *The Journal of Human Resources*, **33**, 479-506.

Journal of the Korean Data & Information Science Society
2011, **22**(1), 1–8

한국데이터정보과학회지

Panel attrition factors in Korean Labor and Income Panel Study

Sang-hyeop Lee[1] · Chan-yong Park[2] · Sung-suk Chung[3] · Hyemi Choi[4]

[1234]Department of Statistics, Chonbuk National University

Received 23 October 2010, revised 19 December 2010, accepted 24 December 2010

Abstract

In panel studies in which the same respondents are interviewed repeatedly over the long term, panel attrition may cause the problems in the reliability of the result and the representativeness of the sample in panel study. In this article, we explore the risk factors of sample attrition in the first 11 waves of the Korean Labor and Income Panel Study (KLIPS) data covering the years 1998-2008, for which the survival analysis techniques such as life-table method and Cox proportional hazard model based on the time to the attrition of each respondent as the survival time of the respondent are applied.

Keywords: Korean Labor and Income Panel Study, panel attrition, survival analysis.

[1] Graduate Student, Dept. of Statistics, Chonbuk National University, Jeonbuk 561-756, Korea.
[2] Graduate Student, Dept. of Statistics, Chonbuk National University, Jeonbuk 561-756, Korea.
[3] Professor, Dept. of Statistics (Institute of Applied Statistics), Chonbuk National University, Jeonbuk 561-756, Korea.
[4] Corresponding author: Assistant Professor, Dept. of Statistics (Institute of Applied Statistics), Chonbuk National University, Jeonbuk 561-756, Korea. E-mail: hchoi@jbnu.ac.kr