

## 위계적 선형모형을 이용한 대졸 신규취업자 임금 결정요인 분석<sup>†</sup>

박성익<sup>1</sup> · 조장식<sup>2</sup>

<sup>1</sup>경성대학교 국제무역통상학과 · <sup>2</sup>경성대학교 정보통계학과

접수 2014년 11월 25일, 수정 2014년 12월 8일, 게재확정 2014년 12월 17일

### 요약

본 논문에서는 전문대 및 4년제 대졸 신규취업자의 임금을 결정하는 요인을 분석하기 위해 개인 수준의 인적특성 변수들과 업종수준의 특성변수들을 사용하였다. 본 논문은 개인수준의 임금이 개인 수준의 인적특성 (1-수준)과 업종 수준의 산업특성 (2-수준)에 의해 영향을 받는 다층구조 (multi-level)를 가지게 된다는 점을 주목하였다. 이와 같이 위계적 자료 특성을 가지는 복수의 분석단위 구조가 되면, 전통적인 회귀분석에서와 같이 개인수준의 임금이 독립이라는 가정을 할 수 없게 된다. 따라서 본 논문에서는 개인수준의 임금이 영향을 미치는 다층구조의 특성을 가진 변수들의 영향력을 보다 타당하게 분석하기 위한 방법으로 위계선형모형 (HLM; hierarchical linear model)을 이용하였다. 주요 결과는 다음과 같다. 첫째, 개인수준과 업종수준 변수들 모두를 포함한 다중대응분석의 결과에 의하면, 개인수준의 임금이 서로 다른 그룹에 대응되는 개인 특성 변수값과 업종 특성 변수값이 그룹별로 서로 상이하여 개인특성 변수만이 아니라 업종특성 변수도 개인수준의 임금에 영향을 미치고 있다는 점이 발견되었다. 둘째, 개인수준과 업종수준 변수들 모두를 포함한 의사결정나무분석의 결과에 의하면, 개인수준의 임금에 가장 많은 영향을 미치는 변수가 업종별 임금이고, 그 다음으로는 업종별 근로시간, 연령, 성별 등의 순으로 나타났다. 이와 같이 개인수준의 임금을 결정하는 데 있어서 업종의 특성이 매우 중요한 것으로 나타났다는 점은 위계적 선형모형의 활용이 타당하다는 것을 시사하는 것이다. 셋째, 개인수준의 인적특성과 업종 수준의 산업특성 변수들을 모두 포함한 모형이 다른 모형들에 비해서 모형 적합도가 가장 개선되어 위계적 선형모형이 적합한 것으로 나타났다.

주요용어: 고정효과, 다층구조, 랜덤효과, 위계선형모형, 집단내 상관.

### 1. 서론

대졸 청년층의 노동시장 성과에 영향을 미치는 요인에 대하여 그동안 많은 연구가 수행되어 왔다. 이러한 연구들은 성, 연령, 학력, 학점평균, 대학수준 등과 같은 특성 외에도 부모 소득이나 부모 학력과 같은 귀속적 특성도 임금에 큰 영향을 미친다는 연구결과를 보고하고 있다 (Chai, 2007; Hwang과 Baek, 2008; Lee와 Kim, 2003). 그런데 이러한 연구들은 업종의 특성이 개인의 임금을 결정하는 경로를 간과하고 있다는 문제점이 존재한다. 특정 업종은 다수의 취업자들을 포함하므로, 다수의 취업자들은 동일한 업종 안에 속하게 됨으로써 공유되는 특성이 존재하게 된다. 이러한 특성이 존재할 경우에, 전통적인 회귀분석에서와 같이 개별 취업자들이 독립이라는 가정을 할 수 없게 된다. 즉 개인별 평균

<sup>†</sup> 이 논문은 2014학년도 경성대학교 학술연구비지원에 의해 연구되었음.

<sup>1</sup> (608-736) 부산광역시 남구 수영로 309 번지, 경성대학교 국제무역통상학과, 교수.

<sup>2</sup> 교신저자: (608-736) 부산광역시 남구 수영로 309 번지, 경성대학교 정보통계학과, 교수.

E-mail: jscho@ks.ac.kr

급여는 개인수준의 인적특성 (1-수준)과 산업 수준의 업종특성 (2-수준)에 의해 영향을 받는 다층구조 (multi-level)로서, 위계적 자료 특성을 가지는 복수의 분석단위의 구조가 된다.

따라서 개인의 임금에 영향을 미치는 변수들의 영향력을 보다 타당하게 분석하기 위한 방법으로 분석단위의 문제를 해결해야 한다는 지적이 제기되어 왔다 (Ethington, 1997; Nasser와 Hagtvet, 2006). 그러나 회귀분석과 같은 전통적인 선형모형들은 복수의 분석단위를 가지는 위계적 자료의 분석에서 분산을 하위 수준과 상위 수준의 분산으로 분해하지 못하며, 하위 수준의 개체특성 효과가 상위 수준의 집단에 따라 변하는 구조적 관계를 규명하지 못하는 등의 방법론적으로 한계를 갖게 된다 (Kang, 1998). 다층구조의 속성을 지니는 위계적 자료를 전통적인 회귀모형으로 분석하는 경우, 집단수준의 자료를 인위적으로 개인수준에 분산시키게 되면 분산이 과대 추정됨으로써 가설검정의 오류가 커지게 된다. 또한 개인수준의 자료를 평균하여 집단수준의 자료로 활용하게 되면 집합화의 오류 (aggregation bias)를 범하게 된다. 따라서 다층구조의 위계적 구조를 갖는 자료를 전통적인 회귀분석 방법으로 분석한다면 집단의 분산이 주는 오차를 고려하지 않는 한계가 있다.

한편 위계선형모형 (HLM; hierarchical linear model)은 여러 학문분야에서 널리 활용되는 통계분석 기법으로서 자료의 위계적 특성을 적절하게 분석할 수 있는 기법이다. Raudenbush와 Bryk (2002)은 위계적 선형모형이 횡단적 다층 자료구조의 통합모형, 다변량 모형, 잠재변수 모형, 베이지안 추론모형 등에 폭넓게 적용될 수 있음을 제시하고 있다. Jeon과 Kang (2005)은 다층자료의 구조적 특성에 따른 위계적 선형모형의 모수추정을 비교한 바 있으며, Im (2002)은 위계적 선형모형을 이용한 발달연구에 대한 횡단적 접근법에 대한 모형을 설정하고 분석하는 방안을 제시하였다. Cho (2013)은 위계적 선형모형을 이용한 강의평가 점수에 대한 결정요인 분석을 하였다. 그 외에도 Baek과 Shin (2007)은 강의 평가에 대한 연구를 위계적 선형모형을 이용하여 분석한 바 있다.

따라서 본 연구에서는 대졸 신규취업자의 임금을 결정하는 요인을 분석하기 위해 1-수준의 인적특성과 2-수준의 업종특성을 모두 고려하여 위계적 선형모형으로 분석하고자 한다. 이를 위해 다중대응분석 (multiple corresponding analysis)을 이용하여 개인별 임금과 1-수준 특성변수 및 2-수준 특성변수들 간의 상호 관련성을 분석한다. 또한 의사결정나무분석 (decision tree analysis)을 이용하여 개인별 임금에 영향을 미치는 1-수준 특성변수와 2-수준 특성변수들의 상호작용 효과를 분석하고자 한다. 또한 위계적 선형모형을 이용하여 4가지의 분석모형을 설정하여 1-수준 및 2-수준의 특성변수들이 개인별 임금에 미치는 영향을 분석하기 위해서 고정효과 (fixed effect)와 랜덤효과 (random effect)를 분석한다.

본 논문의 구성은 다음과 같다. 제 2절에서는 자료소개와 기술통계를 하고, 제 3절에서는 다중대응분석을 통해서 개인별 임금과 1-수준 및 2-수준 특성변수들 간의 상호관련성을 각각 분석한다. 그리고 의사결정나무분석을 이용하여 개인별 임금에 영향을 미치는 1-수준 및 2-수준 특성변수들의 상호작용효과를 각각 분석한다. 제 4절에서는 위계적 선형모형을 활용하여 4가지 분석모형을 설정하고 1-수준 특성변수와 2-수준 특성변수들이 개인별 임금에 미치는 영향력을 분석한다. 그리고 마지막 제 5절에서는 결론을 제시한다.

## 2. 자료 및 변수설명

본 연구에서 사용한 1-수준 데이터는 한국고용정보원 (<http://survey.keis.or.kr/>)에서 대졸자들의 직업이동경로 조사 (GOMS; graduates occupational mobility survey)를 위해 2011년 2월 졸업자들을 대상으로 2년간 추적 조사한 데이터이다. 전체 조사대상자 중 월평균급여가 50만 원 이상인 취업자 369,864명만을 분석대상으로 하였다.

한편 개인수준의 임금은 1-수준 (개인수준) 특성변수뿐만 아니라, 취업자들이 속해 있는 2-수준 (업종수준)에도 영향을 받을 수 있다. 즉 데이터 구조는 동일한 업종 속에는 다수의 취업자들이 포함되어 있으므로 다층구조 (multi-level)의 속성을 지니게 되는 위계적 자료구조를 갖게 된다. 본 논문에서는 개

인 수준의 특성변수를 1-수준 (level-1), 업종 수준의 특성변수를 2-수준 (level-2)으로 설정하였다. 2-수준의 업종별 데이터는 국가통계포털 웹사이트 (<http://kosis.kr/>)에서 구하였다.

분석에 사용된 독립변수는 1-수준 변수로는 성별 (*sex*), 연령 (*age*), 본인학력 (*edu*), 학점 (*gpa*), 취업목표 (*goal*), 부모소득 (*pincome*), 부모학력 (*pedu*) 등이며, 2-수준 변수로는 업종별 평균임금 (*ewage*), 업종별 근로시간 (*etime*), 업종별 근로자수 (*esize*) 등이며, 변수에 대한 설명은 아래 Table 2.1과 같다.

**Table 2.1** Variables explanation

level	variables	explanation	scale	role
level-1	<i>wage</i>	employee's wage	continuous	dependent
	<i>sex</i>	sex	nominal	independent
	<i>age</i>	age	continuous	independent
	<i>edu</i>	education level	ordinal	independent
	<i>gpa</i>	grade point average	continuous	independent
	<i>goal</i>	employment goal	nominal	independent
	<i>pincome</i>	parent's income	continuous	independent
level-2	<i>pedu</i>	parent's education level	ordinal	independent
	<i>ewage</i>	average salary in the industry	continuous	independent
	<i>etime</i>	average labor time in the industry	continuous	independent
	<i>esize</i>	average workers in the industry	continuous	independent

한편 1-수준 변수와 2-수준 변수별 기술통계와 분산분석 결과는 아래 Table 2.2에 제시하였다. 단 분석의 편의를 위해서 연령, 학점, 부모소득, 업종별 평균임금, 업종별 평균 근로시간, 업종별 평균 근로자수는 각각 세 개의 범주로 나누어 분석하였다.

**Table 2.2** Descriptive statistics

variables	categories	N	mean	s.d.	p-value	
<i>sex</i>	female*	184,486	175.7	74.8	<0.001	
	male	185,378	231.4	113.1		
<i>age</i>	≤25*	112,338	164.2	59.8	<0.001	
	26~29	135,968	219.0	93.3		
	>30	50,991	246.6	154.3		
<i>edu</i>	college*	145,739	185.7	98.9	<0.001	
	university	224,126	215.2	98.8		
level-1	<i>gpa</i>	155	207.8	100.4	<0.001	
	75~90	219,628	205.3	99.6		
	>90	148,608	196.7	100.3		
<i>goal</i>	no*	172.1	204.5	109.8	<0.001	
	yes	197.8	202.8	90.4		
<i>pincome</i>	<300*	79,098	202.6	112.0	<0.001	
	300~500	129,646	197.2	86.4		
	>500	82,968	224.1	115.7		
<i>pedu</i>	high*	255,005	200.0	101.1	<0.001	
	college	15,899	198.9	80.9		
	university	98,960	213.7	98.9		
level-2	<i>ewage</i>	low	124,806	183.06	96.161	<0.001
	middle	111,850	178.50	82.195		
	high	120,034	248.05	104.362		
<i>etime</i>	low	124,116	185.00	95.146	<0.001	
	middle	131,468	206.26	101.793		
	high	97,598	223.27	100.517		
<i>esize</i>	low	110,351	208.76	105.740	<0.001	
	middle	126,998	221.28	100.123		
	high	119,341	179.85	89.682		

\* : reference category in hierarchical linear model

위의 Table 2.2의 결과에서 모든 변수들은 범주별로 통계적으로 유의한 차이를 보이고 있다. 구체적으로 1-수준 변수에 대한 개인수준 임금을 비교해 보면 여자에 비해서 남자의 임금이 더 높게 나타났다. 그리고 연령이 증가할수록, 전문대에 비해서 4년제 대졸자가, 그리고 정규직일수록 개인수준의 임금이

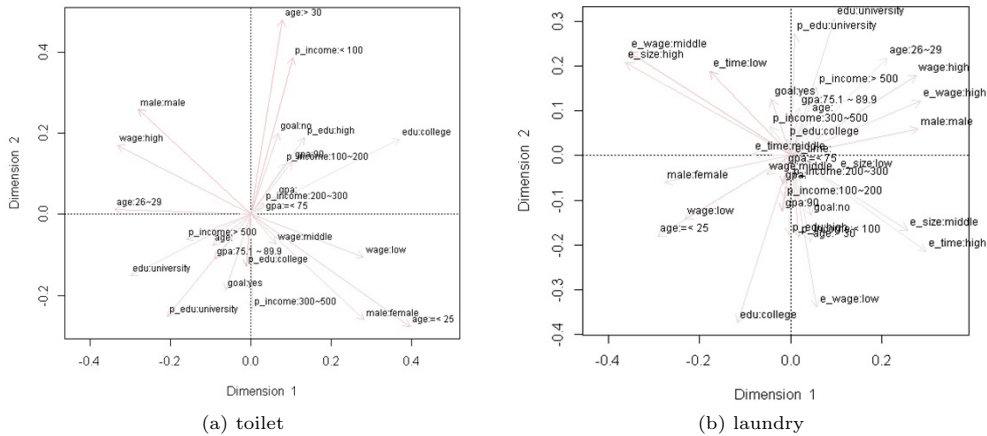
더 높게 나타났으며, 부모소득과 부모학력이 증가할수록 개인수준의 임금이 더 높게 나타났다.

2-수준 변수에 대한 월평균 임금을 비교해 보면, 업종별 임금과 업종별 근로시간이 증가할수록 개인수준의 임금이 증가함을 알 수 있다.

### 3. 수준별 특성변수들의 관련성 분석

이 절에서는 1-수준 (개인)과 2-수준 (업종)의 특성변수들 각각이 개인수준 임금과의 상호 관련성을 알아보고, 또한 개인수준의 임금이 미치는 각 수준별 특성변수들의 상호작용효과를 알아보고자 한다.

먼저 수준별 특성변수들이 개인수준 임금과의 상호관련성을 알아보기 위해 다중대응분석 (multiple correspondence analysis)을 실시하였다. 다중대응분석은 개체 (케이스)와 범주에 계량적 수치를 부여함으로써 범주형 데이터를 수량화하는 분석기법으로서, 내적 일관성의 원리로부터 범주의 수량화를 실시하는 분석기법이다. 여기서 편의상 개인수준의 임금을 상, 중, 하 등 3개의 그룹으로 범주화 하였다. 아래 Figure 3.1의 왼쪽 그림은 1-수준 특성변수만을, Figure 3.1의 오른쪽 그림은 1-수준 특성변수에 2-수준 특성변수를 추가해서 각각 다중대응분석을 실시한 결과이다.



**Figure 3.1** Multiple corresponding analysis for level-1 and level-2 variables (left : only level-1 variables, right : level-1 and level-2 variables)

위의 Figure 3.1의 결과로부터 다음과 같은 결과를 알 수 있다.

먼저 1-수준 변수만을 포함시킨 왼쪽 그림을 살펴보면, 높은 임금을 받는 그룹은 X축의 음의 방향에 위치하고 있으며, 상대적으로 가까운 거리에 있는 신규취업자의 특성은 남자이면서 연령은 26-29세이며, 부모의 소득이 500만 이상이고 본인과 부모의 학력이 4년제 이상이며, 취업목표가 있는 것으로 나타났다. 한편 낮은 임금을 받는 신규취업자의 특성은 X축의 양의 방향에 위치해 있으며, 여자이면서 연령이 25세 이하이고 학점이 낮으며 부모의 학력과 소득이 보통이하인 것으로 나타났다.

다음은 1-수준 변수와 2-수준 변수들을 모두 포함시킨 Figure 3.1의 오른쪽 그림을 살펴보면, 높은 임금을 받고 있는 그룹의 특성은 X축의 양의 방향에 위치하고 있으며, 남자, 평균임금이 높은 업종, 연령 26-29세, 본인학력이 4년제 이상이며, 부모소득과 부모학력이 높으며, 기업규모가 중규모인 경우와 상대적으로 가까운 거리에 있어서 이들 간의 상호 관련성이 높은 것으로 나타났다. 한편, 낮은 임금을 받고 있는 그룹의 특성은 X축의 음의 방향에 위치하고 있으며, 여자, 연령이 25세 이하, 평균임금이 낮은 업종, 본인학력이 전문대이고 평균 근로시간이 낮은 업종과 상대적으로 가까운 거리에 위치하고 있어서 이들 간의 상호관련성이 높은 것으로 나타났다.

다음으로는 1-수준 특성변수와 2-수준 특성변수들이 개인수준의 임금이 미치는 효과를 각각 분석하

기 위해 비모수적인 방법인 의사결정나무 분석을 이용하고자 한다. 의사결정나무분석을 수행하기 위해 지니지수 (Gini index)를 분리기준으로 사용하였으며, 이지분리를 수행하는 CART (classification and regression trees; Breiman 등, 1984) 알고리즘을 사용하였다. 관련된 선행연구로는 Cho 등 (2009)과 Cho (2010) 등이 있다.

본 논문에서는 변수들에 대한 정지규칙으로 최대나무깊이 (maximum tree depth)는 3으로 설정하였으며, 최소 케이스 수 (minimum number of cases)에서 부모마디 (parent node)는 20,000, 자식마디 (child node)는 5,000으로 설정하였으며, 가지치기 (pruning)를 병행하였다. 아래 Figure 3.2는 1-수준 (개인특성) 변수만을 모형에 투입했으며, Figure 3.3은 1-수준(개인특성) 변수와 2-수준 (업종특성) 변수를 모두 투입했을 때의 결과를 각각 제시한 것이다. 여기서 개인수준의 임금, 업종별 임금 및 업종별 근로자 수는 자연로그를 취한 값이다.

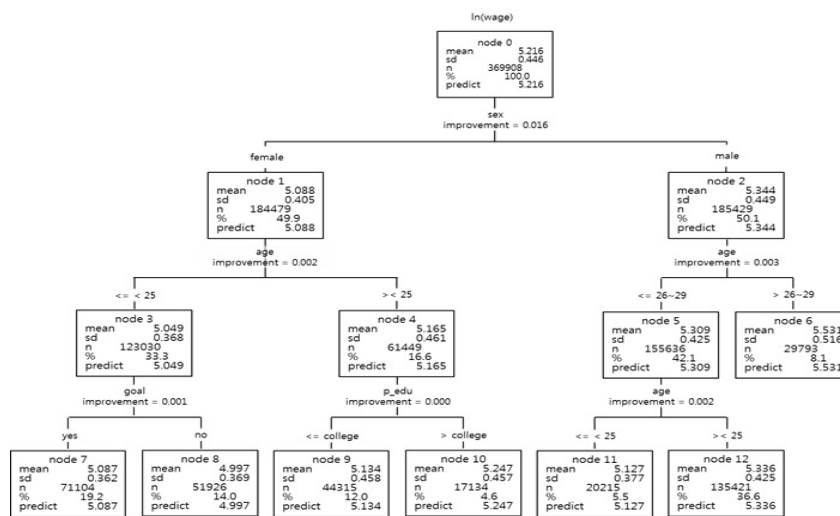


Figure 3.2 Decision tree analysis for only level-1 variables

먼저, Figure 3.2의 결과에서, 개인수준의 임금에 가장 큰 영향을 미치는 것은 성별로 나타났으며, 그 다음으로 연령, 취업목표, 부모학력의 순으로 나타났다. 먼저 남자 (5.344)인 경우는 여자 (5.088)에 비해서 임금이 더 높게 나타났으며, 남자인 경우는 연령이 29세 이하 (5.309)에 비해서 30세 이상 (5.531)에서 개인수준의 임금이 더 높게 나타났다. 또한 연령이 29세 이하인 경우는 연령이 많을수록 개인수준의 임금이 높게 나타났다.

한편 여자인 경우는 연령이 25세 미만 (5.049)보다 25세 초과 (5.165)인 집단에서 개인수준의 임금이 더 높게 나타났으며, 25세 초과인 경우는 부모학력이 높을수록 임금이 높게 나타났다. 또한 연령이 25세 이하인 집단은 취업목표가 있는 집단 (5.087)이 취업목표가 없는 집단 (4.997)보다 임금이 더 높게 나타났다.

다음으로 Figure 3.3의 결과에서는 개인수준의 임금에 가장 많은 영향을 미치는 변수로는 업종별 임금이고, 그 다음으로는 업종별 근로시간, 연령, 성별 등의 순으로 나타났다. 먼저 업종별 임금이 6,910을 초과하는 집단에서는 연령이 25세 미만 (5.182)인 집단에 비해서 25세 초과 (5.493)인 집단의 임금이 높게 나타났으며, 이들 중 25세 초과인 집단의 경우는 여자 (5.369)에 비해서 남자 (5.538)인 경우가 더 높게 나타났다. 또한 연령이 25세 미만인 집단에서는 업종별 근로시간이 189시간 이하 (5.125)에 비해서 190시간 이상 (5.310)인 경우가 개인수준의 임금이 더 높게 나타났다. 한편 업종별로

개인수준의 임금이 6,909이하인 집단에서는 업종별 근로시간이 173시간 이하 (4.985)에 비해서 173시간 초과 (5.185)인 경우가 더 높게 나타났다. 또한 업종별 근로시간이 173시간 초과인 집단에 대해서는 연령이 많을수록 임금이 더 높았으며, 업종별 근로시간이 173시간 이하인 경우는 업종별 근로시간이 150시간 초과 (4.937)에 비해서 150시간 이하 (5.171)가 더 높게 나타났다.

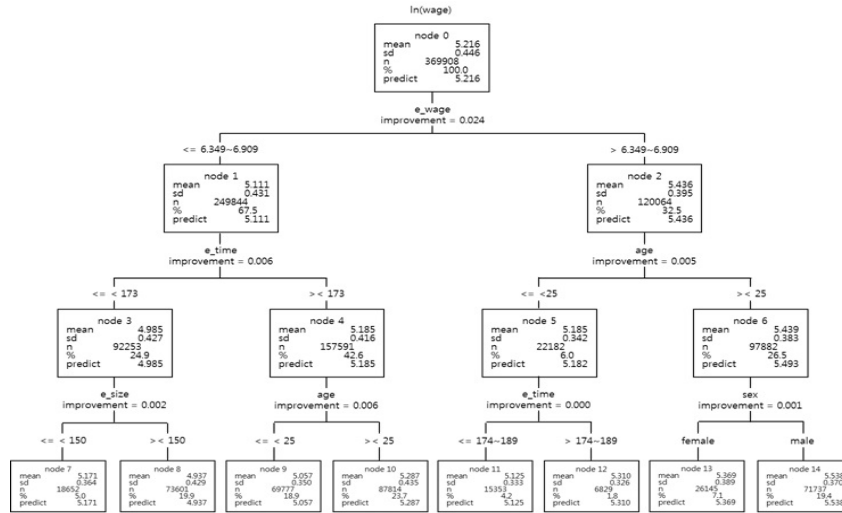


Figure 3.3 Decision tree analysis for level-1 and level-2 variables

## 4. 위계적 선형모형의 분석결과

### 4.1. 연구모형

개인수준의 임금에 대한 1-수준의 특성 변수들과 2-수준의 특성 변수들을 구분해서 각 수준에서 유의하게 영향을 미치는 변수들을 파악하고자 한다. 이를 위해서 Table 1에 있는 1-수준의 특성 변수와 2-수준의 특성 변수들을 사용하였다.

본 논문에서 고려한 모형으로는 (1) 기본모형 (Model I; null model), (2) 1-수준 변수만 포함한 모형 (Model II; random coefficient model), (3) 2-수준 변수만 포함한 모형 (Model III; mean as outcomes model), (4) 1-수준과 2-수준 변수들을 모두 포함한 모형 (Model IV; intercepts and slopes as outcomes model)을 설정하고 분석하였다. 각 분석 모형에 대한 수리적 표현은 다음과 같다.

먼저 기본모형인 Model I (null model)은 개인수준의 임금에 미치는 영향을 분석함에 있어서 1-수준 및 2-수준 특성변수들을 모두 포함하지 않은 모형이다. 즉 1-수준과 2-수준에 대한 기본모형은 다음과 같다.

$$- 1\text{-수준} : y_{ij} = \beta_{0j} + e_{ij}, e_{ij} \sim N(0, \sigma^2).$$

$$- 2\text{-수준} : \beta_{0j} = \gamma_{00} + u_{0j}, u_{0j} \sim N(0, \tau^2_{00}), i = 1, \dots, n_i, j = 1, \dots, m.$$

여기서  $y_{ij}$ 는  $j$ 번째 업종에 대한  $i$ 번째 개인의 임금을 나타내며,  $\beta_{0j}$ 는  $j$ 번째 업종의 평균임금을 나타낸다. 그리고  $e_{ij}$ 는  $i$ 번째 개인의 임금이  $j$ 번째 업종의 평균급여로부터 편차를 나타내는 것으로, 이 값의 분산인  $\sigma^2$ 이 개인 수준의 변량이 된다. 그리고  $\gamma_{00}$ 는 표본 전체 평균을 나타내며,  $u_{0j}$ 는  $j$ 번째 업종의 효과, 즉  $j$ 번째 업종 간의 차이를 나타낸다. 또한  $\tau^2_{00}$ 는 개인 수준에서의 변량을 의미한다.

두 번째 Model II (random coefficient model)는 기본모형에 1-수준 특성변수 만을 추가한 것으로서 분석모형은 다음과 같다.

- 1-수준 :  $y_{ij} = \beta_{0j} + \beta_{1j} \times sex + \beta_{2j} \times edu + \beta_{3j} \times gpa + \beta_{4j} \times age + \beta_{5j} \times goal + \beta_{6j} \times p_{edu} + \beta_{7j} \times p_{income} + e_{ij}, e_{ij} \sim N(0, \sigma^2).$
- 2-수준 :  $\beta_{0j} = \gamma_{00} + u_{0j}, u_{0j} \sim N(0, \tau_{00}^2).$   
 $\beta_{ij} = \gamma_{i0} + u_{ij}, u_{ij} \sim N(0, \tau_{ii}^2), i = 1, 2, \dots, n_i, j = 1, \dots, m.$

세 번째 Model III (mean as outcomes model)은 기본모형에 level-2 (업종 수준)의 특성변수들만을 추가한 것으로서 분석모형은 다음과 같다.

- 1-수준 :  $y_{ij} = \beta_{0j} + e_{ij}, e_{ij} \sim N(0, \sigma^2).$
- 2-수준 :  $\beta_{0j} = \gamma_{00} + \gamma_{01} \times e_{wage} + \gamma_{02} \times e_{time} + \gamma_{03} \times e_{size} + u_{0j}, u_{0j} \sim N(0, \tau_{00}^2), u_{0j} \sim N(0, \tau^2_{00}),$   
 $i = 1, \dots, n_i, j = 1, \dots, m.$

마지막으로 Model IV (intercepts and slopes as outcomes model)는 기본모형에 1-수준 변수와 2-수준 변수 모두 추가한 모형으로서 분석모형은 다음과 같다.

- 1-수준 :  $y_{ij} = \beta_{0j} + \beta_{1j} \times sex + \beta_{2j} \times edu + \beta_{3j} \times gpa + \beta_{4j} \times age + \beta_{5j} \times goal + \beta_{6j} \times p_{edu} + \beta_{7j} \times p_{income} + e_{ij}, e_{ij} \sim N(0, \sigma^2).$
- 2-수준 :  $\beta_{0j} = \gamma_{00} + \gamma_{01} \times e_{wage} + \gamma_{02} \times e_{time} + \gamma_{03} \times e_{size} + u_{0j}, u_{0j} \sim N(0, \tau_{00}^2).$   
 $\beta_{ij} = \gamma_{i0} + u_{ij}, u_{ij} \sim N(0, \tau_{ii}^2), i = 1, \dots, n_i, j = 1, \dots, m.$

#### 4.2. 분석결과 및 해석

본 논문의 관심은 1-수준의 특성 변수와 2-수준의 특성 변수들이 개인수준의 임금에 미치는 영향력을 알아보는 것이다. 4.1절에서 제안한 4가지의 분석모형에 대한 위계적 선형모형의 결과는 Table 4.1에 제시하였다.

**Table 4.1** The results of hierarchical linear model

effects	level	variables	Model I		Model II		Model III		Model IV	
			B	s.e. <sup>0</sup>	B	s.e.	B	s.e.	B	s.e.
		intercept	5.30***	0.03	5.04***	0.02	5.31***	0.02	5.05***	0.02
		<i>sex<sub>male</sub></i>			0.19***	0.02			0.18***	0.02
		<i>age</i>			0.02***	0.00			0.02***	0.00
		<i>edu.u<sup>2</sup></i>			0.13***	0.02			0.12***	0.02
		<i>gpa</i>			0.00***	0.00			0.00***	0.00
	level -1	<i>goal</i>			0.00	0.02			0.01	0.02
	fixed	<i>c<sup>1</sup></i>			0.00	0.03			0.00	0.03
		<i>P<sub>edu</sub></i>			0.07***	0.02			0.08***	0.02
		<300			0.02	0.01			0.01	0.01
		<i>P<sub>income</sub></i>			0.04**	0.02			0.04**	0.02
		>500			0.09***	0.02			0.09***	0.02
	level -2	<i>e<sub>wage</sub></i>					0.28**	0.04	0.17***	0.04
		<i>e<sub>size</sub></i>					-0.04**	0.02	0.00	0.02
		<i>e<sub>time</sub></i>					0.01***	0.00	0.00***	0.00
		<i>sex<sub>male</sub></i>			0.03***	0.01			0.03***	0.01
		<i>age</i>			0.00***	0.00			0.00***	0.00
		<i>edu.u<sup>2</sup></i>			0.01***	0.00			0.01***	0.00
		<i>gpa</i>			0.00***	0.00			0.00***	0.00
	random	<i>goal</i>			0.02***	0.00			0.02***	0.00
		<i>c<sup>1</sup></i>			0.04***	0.01			0.04***	0.01
		<i>P<sub>edu</sub></i>			0.04***	0.01			0.04***	0.01
		<300			0.01***	0.00			0.01***	0.00
		<i>P<sub>income</sub></i>			0.02***	0.00			0.02***	0.00
		>500			0.02***	0.00			0.02***	0.00
		$\sigma^2$	0.16***	0.00	0.12***	0.00	0.16***	0.00	0.12***	0.00
		$\tau^2_{00}$	0.06***	0.01	0.04***	0.01	0.02***	0.00	0.03***	0.01
		total	0.2135		0.1578		0.1804		0.1482	
		<i>ICC<sup>3</sup></i>	0.2663		0.2399		0.1350		0.1891	
		-2 <i>LL<sup>4</sup></i>	364,574.25		258,983.11		350,061.10		250,292.92	
		<i>AIC<sup>5</sup></i>	364,578.25		259,007.11		350,065.10		250,316.92	
		<i>BIC<sup>6</sup></i>	364,599.89		259,136.52		350,086.67		250,445.89	

\*:p<0.10, \*\*:p<0.05, \*\*\*:p<0.01. s.e.<sup>0</sup>:standard error. c<sup>1</sup>:college, u<sup>2</sup>:university. *ICC<sup>3</sup>*:intra class correlation. -2*LL<sup>4</sup>*:-2 log-likelihood. *AIC<sup>5</sup>*:Akaike information criterion. *BIC<sup>6</sup>*:Bayesian information criterion.

Table 4.1의 결과로부터 다음과 같은 사실을 얻을 수 있다.

먼저 기본모형인 Model I (null model)의 결과에서 개인수준의 임금은 5.30이며 개인수준에서 오차는 0.16이고 업종수준에서의 오차는 0.06의 추정치를 얻었다. 이때 개인수준의 오차 0.16은 집단(업종) 내 분산을, 업종수준의 오차 0.06은 집단(업종)간 분산을 나타낸다. 따라서 집단 내 상관(ICC; intra class correlation)이 약 26.6%로 나타났다. 이는 업종수준의 특성에 따른 차이가 설명할 수 있는 개인수준 임금의 최대 분산이 26.6%이며, 나머지 약 73.4% 정도는 개인수준의 특성들에 의한 개인수준의 차이라는 것을 의미한다.

다음으로 기본모형에 1-수준(개인) 변수들만 추가한 Model II (random coefficient model)의 분석 결과는 다음과 같다. 먼저 기본모형에 비해서 추가한 독립변수들은 취업목표, 부모학력 전문대 더미 및 부모소득 300만원미만 더미를 제외한 모든 변수들이 통계적으로 유의하게 나타났다. 또한 모형 I에 비해서 2-수준의 오차분산이 0.06에서 0.01로 감소했는데, 이것은 업종수준 임금의 73.4% 정도를 1-수준 독립변수들이 설명하고 있음을 의미한다. 그리고 모형 적합도를 나타내는  $-2LL$  ( $-2 \log$ -likelihood), AIC (Akaike information criterion), BIC (Bayesian information criterion) 역시 그 값이 작아졌는데 이는 모형 적합도가 전체적으로 개선되었음을 의미한다.

다음으로 기본모형에 2-수준(업종)의 특성 변수들만 투입한 Model III (mean as outcomes model)의 결과에서는 기본모형에 비해서 2-수준의 오차분산이 0.06에서 0.02로 감소했으며, 2-수준의 특성변수들이 개인수준의 임금에 통계적으로 유의한 영향을 미치는 것으로 나타났다. 또한 기본모형에 비해서 모형 적합도를 나타내는  $-2LL$ , AIC, BIC 등의 값들이 모두 감소한 것으로 나타났는데 이는 모형 적합도가 전체적으로 개선되었음을 알 수 있다.

마지막으로 기본모형에 1-수준과 2-수준의 특성 변수들을 추가한 Model IV (intercepts and slopes as outcomes model)의 결과에서는 1-수준의 취업목표, 부모학력의 전문대 더미, 부모소득의 300만원 이하 더미와 2-수준의 업종별 근로자수를 제외한 모든 변수들의 통계적으로 유의하게 나타났다. 기본모형에 비해서 1-수준의 오차분산(0.12)과 2-수준의 오차분산(0.03)으로 많이 감소했으며, 전체적인 모형 적합도를 나타내는  $-2LL$ , AIC 및 BIC의 값도 Model I, Model II 및 Model III에 비해서 많이 감소한 것으로 나타났다. 즉 Model IV가 다른 모형들에 비해서 모형 적합도가 가장 개선된 것으로 나타났다.

Model IV의 결과를 바탕으로 다음과 같은 결과를 알 수 있다. 먼저 1-수준 특성변수들에 대한 결과를 살펴보면, 여자에 비해서 남자가 임금이 높은 것으로 나타났으며, 연령이 많을수록 더 많은 임금을 받는 것으로 나타났다. 또한 전문대 졸업자에 비해서 4년제 졸업자의 임금이 더 많으며, 학점이 높을수록 임금이 더 높아지는 것을 알 수 있다. 그리고 부모의 학력이 높을수록, 부모의 소득이 높을수록 임금이 더 높게 나타났다. 2-수준 특성변수들에 대한 결과를 살펴보면 업종별 임금이 높을수록, 업종별 근로시간이 많을수록 임금이 더 높게 나타남을 알 수 있다.

한편, 랜덤효과(random effect)의 결과를 보면, 모든 변수들에 대해서 랜덤효과가 통계적으로 유의함을 알 수 있다. 즉 이들 변수들이 임금에 미치는 효과는 업종에 따라 통계적으로 유의한 차이가 있음을 의미한다.

이상의 위계적 선형모형의 결과는 3절의 다중대응분석과 의사결정나무분석의 결과들과 대체로 유사함을 알 수 있다.

## 5. 결론

본 논문은 위계선형모형을 이용하여 개인수준의 특성변수와 업종수준의 특성변수들이 개인별 임금에 미치는 영향력을 분석하였다. 주요 결과는 다음과 같다. 첫째, 개인수준과 업종수준 변수들 모두를 포함한 다중대응분석의 결과에 의하면, 개인수준의 임금이 서로 다른 그룹에 대응되는 개인 특성 변수값과



업종 특성 변수값이 그룹별로 서로 상이하여 개인특성 변수만이 아니라 업종특성 변수도 개인수준의 임금에 영향을 미치고 있다는 점이 발견되었다. 예를 들어, 임금이 높은 (300만원 이상) 그룹은 업종별 근로자수도 많고, 남자이면서 연령이 26~29세로 부모소득이 높고, 학력은 4년제 대졸자와 상대적으로 가까운 거리에 위치하고 있어서, 이들 간의 상호 관련성이 높은 것으로 나타났다. 한편, 개인수준의 임금이 낮은 (200만원 이하) 그룹은 여자이면서 연령이 25세 미만인 집단과 상대적으로 가까운 거리에 위치하고 있어서, 이들 간의 상호 관련성이 높은 것으로 나타났다.

둘째, 개인수준과 업종수준 변수들 모두를 포함한 의사결정나무분석의 결과에 의하면, 개인수준의 임금에 가장 많은 영향을 미치는 변수가 업종별 임금이고, 그 다음으로는 업종별 근로시간, 연령, 성별 등의 순으로 나타났다. 이와 같이 개인수준의 임금을 결정하는 데 있어서 업종의 특성이 매우 중요한 것으로 나타났다는 점은 위계적 선형모형의 활용이 타당하다는 것을 시사하는 것이다.

셋째, 위계적 선형모형에서는 개인수준 및 업종수준 특성변수들의 포함여부에 따라 4개의 모형을 설정하고 그 모형 간의 적합도를 비교하였다. 분석결과에 의하면, 개인수준과 업종수준 특성변수들을 모두 포함한 모형이 다른 모형들에 비해서 모형 적합도가 가장 개선되어 위계적 선형모형이 적합한 것을 알 수 있다. 먼저 개인수준 특성변수들에 대한 결과를 살펴보면, 여자에 비해서 남자가 임금이 높은 것으로 나타났으며, 연령이 많을수록 더 많은 임금을 받는 것으로 나타났다. 또한 전문대 졸업자에 비해서 4년제 졸업자의 임금이 더 많으며, 학점이 높을수록 임금이 더 높아지는 것을 알 수 있다. 그리고 부모의 학력이 높을수록, 부모의 소득이 높을수록 임금이 더 높게 나타났다. 업종수준 특성변수들에 대한 결과를 살펴보면 업종별 임금이 높을수록, 업종별 근로시간이 많을수록 임금이 더 높게 나타났다.

본 논문은 임금이 결정되는 데 있어서 개인 수준의 개체특성 효과가 상위 수준의 집단, 즉 업종에 따라 변하는 구조적 관계가 존재한다는 것을 입증하였다는 점에서 학문적 의의가 있다. 이러한 위계적 결정메커니즘은 많은 경제변수의 관계에서도 존재하는 것으로 추정된다. 임금 결정의 경우만 예를 들더라도, 개인 수준의 요인 외에도 지역 수준의 사회경제적 변수가 중요한 영향을 미치는 것으로 판단된다. 이러한 관점에서 보면, 위계적 선형모형을 활용하여 다양한 경제 변수의 결정 요인을 연구할 필요가 존재한다고 할 수 있다.

## References

- Baek, S. G. and Shin, H. J. (2008). Multilevel analysis of the effects of student and course characteristics on student course evaluation - Focused on the undergraduate liberal education program. *Journal of Educational Evaluation*, **21**, 1-24.
- Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984). *Classification and regression trees*, Chapman and Hall Press, Washington, D.C.
- Chai, G. M. (2007). An analysis of the determinants of employment and wage of new college graduates. *Korean Journal of Social Welfare*, **59**, 35-61.
- Cho, J. S. (2010). A study on equating method based on regression analysis. *Journal of the Korean Data Information & Science Society*, **21**, 513-521.
- Cho, J. S. (2013). Determinants of student course evaluation using hierarchical linear model. *Journal of the Korean Data Information & Science Society*, **24**, 1-12.
- Cho, J. S., Kang, C. W. and Choi, S. B. (2009). Comparison on equating methods for course evaluation. *Journal of the Korean Data Information & Science Society*, **20**, 65-75.
- Ethington, C. A. (1997). A hierarchical linear modeling approach to studying college effects. In *Higher Education: Handbook of Theory and Research* **12**, edited by C. A. Ethington, Agathon Press, New York, 165-194.
- Hwang, Y. J. and Baek, B. B. (2008). Determinants of employment status of university graduates youth. *Journal of Vocational Education and Training*, **11**, 1-23.
- Im, S. H. (2002). Multilevel models for cross-sectional approach to the developmental research. *Journal of Educational Evaluation*, **15**, 295-315.

- Jeon, M. J. and Kang, S. J. (2005). A comparison of multilevel models in their parameter estimation - The comparison of 2-level HLM, 3-level HLM, and CMM. *Journal of Education Evaluation*, **18**, 123-147.
- Kang, S. J. (1998). Analytical comparisons between classical linear models and multilevel models as educational and social research methods. *Journal of Educational Evaluation*, **11**, 207-258.
- Lee, K. Y. and Kim, Y. H. (2003). The determinants of youth labor market performance. *Quarterly Journal of Labor Policy*, **3**, 69-93.
- Nasser, F. and Hagtvet, K. A. (2006). Multilevel analysis of the effects of student and instructor/course characteristics on student ratings. *Research in Higher Education*, **47**, 559-590.
- Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical linear models : Applications and data analysis methods*, SAGE publication, San Francisco.

## Determinants of employee's wage using hierarchical linear model<sup>†</sup>

Sungik Park<sup>1</sup> · Jangsik Cho<sup>2</sup>

<sup>1</sup>Department of International Trade and Commerce, Kyungsoong University

<sup>2</sup>Department of Informational Statistics, Kyungsoong University

Received 25 November 2014, revised 8 December 2014, accepted 17 December 2014

### Abstract

This paper analyzes the determinants of wage for the college and university graduates utilizing both individual-level and industry-level variables. We note that wage determination has multi-level structure in the sense that individual wage is influenced by individual-level variables (level-1) and industry-level (level-2) variables. Then, the assumption that individual wage is independent in the classical regression is violated. Therefore, this paper utilizes the hierarchical linear model (HLM). The major results are the followings. First, the multiple correspondence analysis including level-1 and 2 variables reveals that both level 1 and level 2 variables affects individual wages judging from the fact that the values of level 1 and level 2 variables differ across the different level of individual wage groups. Second, the decision tree analysis including level-1 and 2 variables shows that the most influential variable in wage determination is industry-level wage and the next is industry-level working hour, ages and sex in the declining order in. This suggests that the utilization of the HLM is appropriate since the characteristics of industry is important in determining the individual wage. Third, it is shown that the HLM model is the best compared to the other models which do not take level-1 and level-2 variables simultaneously into account.

*Keywords:* Fixed effect, hierarchical linear model, intra class correlation, multi-level, random effect.

---

<sup>†</sup> This research was supported by Kyungsoong University Research Grants in 2014.

<sup>1</sup> Professor, Department of International Trade and Commerce, Kyungsoong University, Busan 608-736, Korea.

<sup>2</sup> Corresponding author: Professor, Department of Informational Statistics, Kyungsoong University, Busan 608-736, Korea. E-mail: jscho@ks.ac.kr