

혼합모형을 이용한 특성화고 졸업생의 임금결정요인 분석[†]

류장수¹ · 조장식²

¹부경대학교 경제학부 · ²경성대학교 수학을융통계학부

접수 2016년 5월 17일, 수정 2016년 6월 26일, 게재확정 2016년 7월 7일

요 약

본 연구에서는 한국고용정보원에서 실시한 「2013 고졸자 취업진로조사」 자료를 활용하여 특성화고 졸업자의 임금결정요인을 분석하였다. 그런데 임금은 개인수준의 인적특성 (1-수준)과 취업지역 수준의 산업특성 (2-수준)에 의해 영향을 받는 다층구조를 가지게 된다. 이와 같이 다층구조 자료특성을 가지는 복수의 분석단위 구조가 되면, 전통적인 회귀분석과 같이 개인수준의 임금이 독립이라는 가정을 할 수 없게 된다. 따라서 개인수준의 임금에 영향을 미치는 다층구조의 특성을 가진 변수들의 영향력을 분석하기 위한 타당한 방법으로 위계적 선형모형을 이용하였다. 그리고 전통적인 회귀분석과 위계적 선형모형의 비교를 통하여 다음과 같은 주요 결과를 얻었다. 첫째, 다층구조를 갖는 위계적 선형모형이 전통적인 회귀모형보다 통계적으로 유의함을 알 수 있었다. 둘째, 2-수준의 총근로시간과 상용직의 평균임금이 개인수준의 임금에 통계적으로 유의하게 영향을 미치는 것으로 나타났다. 셋째, 마이스터고 졸업생이 그렇지 않은 졸업생에 비해서, 부모의 소득이 높을수록 통계적으로 유의하게 개인수준의 임금이 높아짐을 알 수 있다. 넷째, 부모의 교육수준이 높을수록, 학점이 높을수록, 학교 만족도가 높고 자격증 수가 많을수록 임금이 높아짐을 알 수 있다. 마지막으로 2-수준의 산업특성을 고려한 랜덤효과가 통계적으로 유의하게 나타났다.

주요용어: 고정효과, 다중대응분석, 랜덤효과, 위계적 선형모형, 특성화고.

1. 서론

최근 학력중심사회에서 능력중심사회로의 전환을 핵심적 정책과제로 선정하면서, 특성화고 학생들에 대한 지원방안이 크게 강화되고 있다. 특히 특성화고 졸업생의 취업과 이후 경력경로에서의 개선이 중요하다는 인식하에서 취업 관련 자료 구축과 연구가 예전에 비해 활발하게 진행되고 있는 실정이다. 특성화고 졸업생들에 대한 노동시장 성과 연구와 관련하여, Lee와 Jung (2010)은 특성화고등학교 학생들이 인식하는 취업장벽 수준을 측정하고 그와 관련된 변인과 그들의 설명력을 규명하는 연구를 했다. Jang과 Lee (2012)는 한국교육고용패널 자료를 활용하여 특성화고 학생의 취업선택을 결정하는 개인, 가정 및 학교 변인을 구명하고, 각 변인의 영향을 예측하였다. 그리고 Lee와 Joo (2013)는 한국교육고용패널 조사 자료를 활용하여 특성화고 졸업생 취업에 대한 학교 효과분석을 하였다. Cho와 Jung (2014)은 특성화고 졸업예정자의 취업결정요인분석을 하였다. 이외에도 다수의 특성화고 졸업생들의 노동시장성과에 관한 연구들이 있으나, 주된 연구대상은 취업실태 및 취업결정요인분석 중심이었으며

[†] 이 논문은 부경대학교 자율창의학술연구비 (2016)에 의하여 연구되었음.

¹ (48513) 부산광역시 남구 용소로 45, 부경대학교 경제학부, 교수.

² 교신저자: (48434) 부산광역시 남구 수영로 309번지, 경성대학교 수학을융통계학부, 교수.

E-mail: jscho@ks.ac.kr

임금결정요인에 관한 분석은 거의 없다고 해도 과언이 아니다. 물론 임금실태에 관한 기본통계분석 정도는 발견되고 있지만, 임금의 결정요인에 초점을 맞춘 통계적 분석은 부족한 것이 현실이다.

이러한 현실을 반영하여 본 연구에서는 특성화고 출신 취업자들의 임금결정요인을 분석하고자 한다. 그런데 임금은 개인의 인적특성 (1-수준)과 취업지역의 산업특성 (2-수준)에 의해 영향을 받는 다층구조의 복수 분석단위를 가지게 된다. 즉 특정 취업지역은 다수의 특성화고졸 취업자들을 포함하고, 다수의 취업자들은 동일한 취업 지역 안에 속하게 됨으로써 공유되는 특성이 존재하게 된다.

이와 같이 복수의 분석단위를 가지는 다층구조를 가지는 자료에 대해서 개인수준의 임금에 영향을 미치는 변수들의 영향력을 보다 타당하게 분석하기 위해서는 분석단위의 문제를 해결해야 한다는 지적이 많은 연구자들에 의해 제기되어 왔다 (Nasser와 Hagtvet, 2006). 그러나 회귀분석과 같은 전통적인 선형모형은 복수의 분석단위를 가지는 다층구조의 자료분석에서 분산을 하위 수준과 상위 수준으로 분해하지 못하며, 하위 수준의 개인특성 효과가 상위 수준의 집단에 따라 변하는 구조적 관계를 규명하지 못하는 등, 방법론적으로 한계를 갖게 된다 (Kang, 1998). 또한 다층구조의 속성을 지니는 위계적 자료를 전통적인 회귀모형으로 분석하는 경우, 집단수준의 자료를 인위적으로 개인수준에 분산시키게 되어 분산이 과대 추정됨으로써 가설검정의 오류가 커지게 된다. 또한 개인수준의 자료를 평균하여 집단수준의 자료로 활용하게 되면 집합화의 오류 (aggregation bias)를 범하게 된다. 따라서 위계적 구조를 갖는 자료를 전통적인 회귀분석 방법으로 분석한다면 집단의 분산이 주는 오차를 고려하지 않는 한계가 발생한다.

한편 위계선형모형 (hierarchical linear model; HLM)은 다층구조를 갖는 자료에 대해서 적절하게 분석할 수 있는 통계적 분석방법이다. 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. 그 외에도 Cho (2013), Im 등 (2014), 그리고 Park과 Cho (2015)는 위계적 선형모형을 이용한 분석을 하였다.

따라서 본 연구에서는 특성화고 졸업생들의 임금 결정요인을 분석하기 위해 개인수준의 인적특성과 취업지역 수준의 산업특성을 고려하는 다층구조 하에서 위계적 선형모형을 활용한다. 또한 위계적 선형모형을 이용하여 개인수준 및 지역수준의 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 변수들 간의 상호관련성을 분석한다. 제 4절에서는 다층구조를 갖는 복수의 분석단위 자료에서 전통적인 선형 회귀모형과 위계적 선형모형을 비교하고, 1-수준 특성변수와 2-수준 특성변수들이 개인별 임금에 미치는 영향력을 고정효과와 랜덤효과를 중심으로 분석한다. 그리고 마지막 제 5절에서는 결론을 제시한다.

2. 자료소개 및 기술통계

특성화고 졸업자의 임금결정요인을 분석하기 위해 본 연구에서는 한국고용정보원에서 2013년 5~7월 (조사기준일은 2013년 5월 1일)에 실시한 「2013 고졸자 취업진로조사」 (2013 High School Graduates Occupational Mobility Survey)의 원자료를 활용한다. 조사대상 모집단은 2013년 특성화고, 마이스터고, 종합고 (실업계열) 졸업자 중 비진학자로서 조사기준 시점 당시 이미 취업을 하였거나, 취업은 하지 않았으나 대학진학 (재수) 의사가 없는 졸업자 (군입대 제외)이다.

한편 개인수준의 임금은 개인수준의 인적 특성변수뿐만 아니라, 취업자들이 속해 있는 취업지역 수준

에도 영향을 받을 수 있다. 즉 데이터 구조는 동일한 지역 속에는 다수의 취업자들이 포함되어 있으므로 다층구조의 속성을 갖게 된다. 본 논문에서는 개인 수준의 인적 특성변수를 1-수준 (level-1), 취업지역 수준의 특성변수를 2-수준 (level-2)으로 설정하였다. 지역별 산업특성 데이터는 국가통계포털 웹사이트 (<http://kosis.kr/>)에서 구하였다.

분석에 사용된 1-수준 변수와 2-수준 변수는 Table 2.1과 같다.

Table 2.1 Variables explanation

level	variables	explanation
level-1	sex	whether or not male : 0='female', 1='male'
	maester	whether or not maester : 0='the others', 1='maester'
	engineer	whether or not engineer : 0='the others', 1='engineer'
	income	parent's income : 0='<100', 1='100~200', 2='200~300', 3='300~500', 4='>500'
	education	parent's education : 0='≤high school', 1='college', 2='≥university'
	score	graduation record : 0='middle', 1='very low', 2='low', 3='high', 4='very high'
	satisfy	satisfaction of high school
	license	number of licenses : 0='none', 1='1~2', 2='≥3'
level-2	region	high school region : 0='dongnam', 1='sudo', 2='chungchung', 3='junra', 4='daekyung', 5='kangwon', 6='jeju'
	EAR	Economic activity rates of a high school diploma
	time	total business hours
	wage	wage of regular job

0=reference category.

Table 2.2는 1-수준 변수별 임금에 대한 기술통계 분석결과이다.

Table 2.2 Descriptive statistic for level-1 variables

variables	n	mean	variables	n	mean
sex			education		
($p < 0.001$)			≤high	24,570	148.4
male	19,496	148.2	college	1,761	149.8
female	15,736	150.2	($p < 0.001$)		
maester			university	3,699	158
($p < 0.001$)			>university	287	157.3
etc	2,523	174.4			
engineer	32,709	147.1	none	6,005	141.5
($p < 0.001$)			1~2	18,390	146
engineer	19,247	148.8	($p < 0.001$)		
etc	15,985	149.4	3~5	9,527	159
<100	1,561	145.9	≥6	1,310	155.6
100~200	7,854	145.3			
200~300	9,783	147.4	sudo	16,683	143.3
300~500	12,224	151.5	chungchung	4,387	153.7
≥500	2,256	161.5	junra	3,658	149.8
very low	1,904	141.9	regions		
low	5,677	138.8	($p < 0.001$)		
middle	12,584	143.9	daekyung	4,450	164.1
high	9,560	154.2	dongnam	5,525	151.2
very high	5,507	165.2	kangwon	348	139.2
			jeju	181	137.4
			total	35,232	149.1

Table 2.2의 결과에 따르면, 평균임금이 남자에 비해서 여자가 다소 높게 나타났으며, 마이스터고 출신이 기타에 비해서 높게 나타났다. 기타계열이 공업계열에 비해서 약간 높게 나타났고, 부모의 소득이 높을수록, 졸업성적이 높을수록, 그리고 자격증 수가 많을수록 평균임금이 높아지는 경향을 보이고 있다. 지역적으로는 대경권이 가장 높은 임금을 보이고 있으며, 그 다음으로 충청권, 동남권, 전라권 등의 순으로 나타났다.

Table 2.3은 2-수준 변수에 대한 기술통계를 나타낸 것이다.

Table 2.3 Descriptive statistic for level-2 variables

regions	EAR	time	wage
BS (Busan)	59.6	187.0	140.6
CB (Chungbuk)	67.9	195.4	162.3
CN (Chungnam)	68.8	192.8	160.1
DG (Daegu)	60.2	195.7	143.6
DJ (Daejeon)	59.4	188.4	140.2
KW (Gangwon)	61.9	192.2	138.3
KJ (Gwangju)	59.0	184.6	145.4
KK (Gyeonggi)	63.1	188.0	145.3
KB (Gyeongbuk)	67.8	194.6	168.3
KN (Gyeongnam)	66.6	195.9	153.9
IC (Incheon)	67.1	193.1	141.4
JJ (Jeju)	74.8	187.4	125.4
JB (Jeonbuk)	63.3	191.9	138.5
JN (Jeonnam)	68.9	190.3	137.2
SU (Seoul)	60.2	179.5	141.1
US (Ulsan)	63.1	190.8	161.2

Table 2.3의 결과로부터 고졸출신의 경제활동비율은 제주도가 74.8로 가장 높게 나타났으며, 그 다음으로 전남이 68.9, 충남이 68.8, 충북이 67.9의 순으로 나타났다. 또한 총 근무시간은 경남이 195.9 시간으로 가장 높게 나타났으며, 그 다음으로 대구가 195.7 시간, 충북이 195.4 시간 순으로 나타났다. 상용직 평균임금은 경북이 168.3 만원으로 가장 높게 나타났으며, 그 다음으로 충북이 162.3 만원, 울산이 161.2 만원 순으로 나타났다.

3. 수준별 변수의 관련성 분석

이 절에서는 1-수준 변수 및 2-수준 변수들에 대한 임금과의 관련성을 각각 분석한다. 먼저 1-수준 변수들과 임금과의 상호관련성 분석을 위해, 다차원의 변수들의 복잡한 관계를 소수 몇 개의 축을 기준으로 시각화할 수 있는 방법인 다중대응분석 (multiple correspondence analysis)을 실시하였다. 아래 Figure 3.1은 1-수준 변수들에 대한 다중대응분석 결과를 나타낸다. 여기서 다중대응분석을 위해 종속 변수인 임금을 범주화를 해야 하는데, 여기서는 편의상 3등분하여 3개의 수준 (high, middle, low)으로 범주화를 했다.

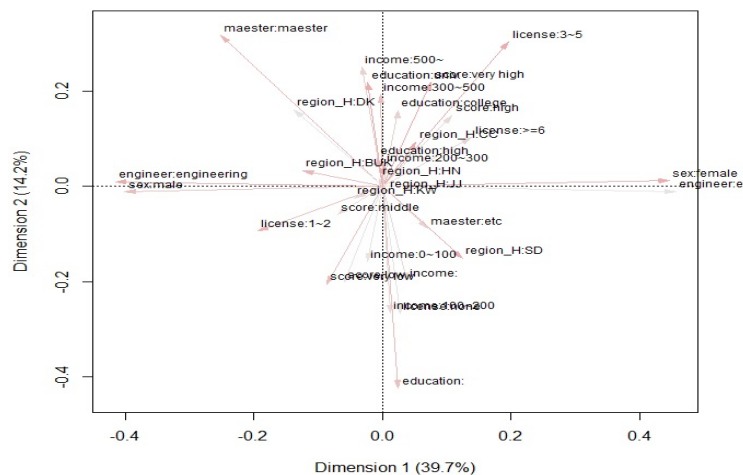
**Figure 3.1** The result of multiple correspondence analysis

Figure 3.1에 따르면, X축은 성별과 공업계열 더미 변수로 요약이 될 수 있고, Y축은 임금, 고교성적, 부모소득, 자격증 수 변수로 요약될 수 있다. 특히 Y축은 음의 방향에서 양의 방향으로 갈수록 임금과 고교성적은 낮아지고, 부모소득과 자격증 수는 감소하는 경향이 있다. 또한 고교지역이 수도권은 임금이 낮은 Y축의 양의 방향에 위치하고 있으며, 대경권과 충청권은 임금이 높은 음의 방향에 위치하고 있음을 알 수 있다.

따라서 성별과 공업계열 더미는 임금수준을 결정하는 Y축과 크게 관련이 없음을 알 수 있다. 또한 임금수준이 높은 경우는 고교성적이 최상위이면서, 자격증 수가 3개 이상이고, 부모소득이 높으며, 부모학력이 대졸이상과 상대적으로 가까운 거리에 위치하고 있어서 이들 범주들 간의 상호 관련성이 높음을 알 수 있다. 한편 임금수준이 낮은 경우는 자격증이 작고, 부모소득과 고교성적이 낮으며, 마이스터고가 아닌 경우와 고교지역이 수도권과 상대적으로 가까운 거리에 있어서 상호 관련성이 높은 것으로 나타났다.

Figure 3.2는 취업지역 수준의 2-수준 변수들에 대해서 산업특성 변수들과 지역과의 관련성을 알아보기 위해 버블차트 결과를 제시한 것이다. Figure 3.2의 결과로부터 총 근로시간이 많아질수록 평균임금과 고졸 경제활동비율이 증가하는 경향을 보임을 알 수 있다. 또한 총근로시간과 고졸 경제활동비율이 높은 지역은 충남, 충북, 경남, 경북으로 나타났으며, 광역시 (metropolitan city)에 비해서 도 (do)가 고졸 경제활동비율이 상대적으로 높고 총근로시간도 높음을 알 수 있다.

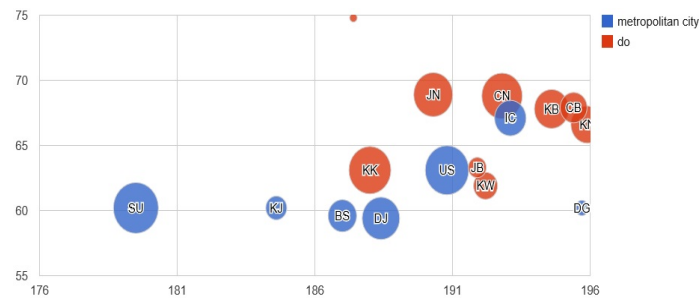


Figure 3.2 Bubble chart for level-2 variables (X-axis: time, Y-axis: EAR, size = wage)

4. 위계적 선형모형 분석결과

개인수준의 임금을 결정하는데 있어서 동일 취업지역에 속한 다수의 특성화고졸 취업자들이 동일한 지역 안에 속하게 됨으로써 공유되는 특성이 존재하게 된다. 즉 개인별 임금은 개인수준의 1-수준 변수와 취업지역 수준의 2-수준 변수에 의해 영향을 받는 다층구조로서, 복수의 분석단위를 가지는 통계모형으로 분석할 필요가 있다. 이를 위해 Table 2.1에 있는 1-수준 변수와 2-수준 변수들을 사용하여 아래의 두 가지의 모형을 고려한다.

첫째, 1-수준 변수만 고려한 전통적인 선형회귀모형 (Model I : OLS model)이다. 즉 Model I은 특성화고 졸업자들의 취업지역 특성은 모두 동일하다고 가정하며, 1-수준 변수들이 개인수준의 임금에 미치는 효과는 취업지역과 관계없이 고정되어 있는 상수로 설정하며 확률적 모형식은 다음과 같다.

$$y_{ij} = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_p x_{pj} + e_{ij}, \quad e_{ij} \sim N(0, \sigma^2), \quad i=1, \cdots, n_i, \quad j=1, \cdots, 16, \quad (4.1)$$

여기서 y_{ij} 는 j 번째 취업지역에 속해있는 i 번째 개인의 임금을 의미하고, x_{1j}, \cdots, x_{pj} 는 1-수준 변수이고 회귀계수 β_k , $k=1, 2, \cdots, p$ 는 고정된 상수이다.

둘째, 1-수준과 2-수준 변수들을 모두 포함한 위계적 선형회귀모형 (Model II : HLM model)이다. 위계적 선형회귀모형은 1-수준 변수들이 개인수준의 임금에 미치는 영향력은 취업지역에 따라 확률적으로 변할 수 있다고 가정하는 모형이다. 따라서 Model II에서 추정된 회귀계수는 고정된 상수가 아니라 확률변수로 가정하며, 확률적 모형식은 다음과 같다.

$$\begin{aligned} \text{1-수준 : } y_{ij} &= \beta_{0j} + \beta_{1j}x_{1j} + \beta_{2j}x_{2j} + \cdots + \beta_{pj}x_{pj} + e_{ij} \\ \text{2-수준 : } \beta_{0j} &= \gamma_{00} + \gamma_{01}z_1 + \gamma_{02}z_2 + \gamma_{03}z_3 + u_{0j}, \\ \beta_{kj} &= \gamma_{k0} + u_{kj}, \quad u_{kj} \sim N(0, \tau_{kk}^2), \quad k = 1, \dots, p, \quad j = 1, \dots, 16. \end{aligned} \quad (4.2)$$

여기서 z_1, z_2, z_3 는 2-수준 변수를 의미하고, 회귀계수 β_{kj} 는 랜덤효과를 나타내는 확률변수를 의미한다. 즉 u_{kj} 는 2-수준 j 에 따라서만 값이 달라지며, j 그룹 내에서는 모든 관측치 i 가 동일한 값을 갖는다. 오차항에 대한 분포는 $u_j|x_{ij}, z_j \sim N(0, \sigma_{u_j}^2)$ 이고 $e_{ij}|x_{ij}, z_j, u_j \sim N(0, \sigma_e^2)$ 으로 가정한다. 본 연구에서는 특성화고 졸업생들의 여러 특성 중에서 성별, 마이스터고 출신여부 및 공업계열 여부가 취업지역별 임금의 변동효과를 알아보고자 한다. 따라서 위계적 선형모형의 절편, 성별, 마이스터고 더미 및 공업계열 더미의 회귀계수에 대해서만 랜덤효과를 고려한다. 그 이유는 성별은 인적특성을 나타내는 대표적인 변수이며, 마이스터고 더미 및 공업계열 더미는 특성화고의 특성을 나타내는 대표적인 변수이기 때문이다.

Table 4.1~Table 4.3은 위계적 선형모형에서 랜덤효과를 나타내는 성별, 마이스터고 더미 및 공업계열 더미에 따른 개인수준의 임금을 취업지역별로 비교한 결과를 제시한 것이다.

Table 4.1~Table 4.3의 결과에 따르면 성별, 마이스터고 더미 및 공업계열 더미에 따른 개인수준의 임금분포가 취업지역별로 많은 차이가 있음을 알 수 있다. 즉 성별, 마이스터고 더미 및 공업계열 더미가 개인수준의 임금에 미치는 효과가 2-수준인 취업지역에 따라 다르다는 것을 의미하며, 이들 변수에 대한 랜덤효과를 분석할 필요성이 있음을 시사하는 것이다.

Table 4.1 Comparisons of wage for sex and regions

regions	male		female		Boxplot for male and regions
	mean	sd	mean	sd	
BS	143.7	1,257	140.8	972	
CB	149.4	637	170.4	988	
CN	162.5	965	158.7	901	
DG	137.3	971	161.9	666	
DJ	140.2	653	148.7	347	
IC	145.6	1,385	139.3	1,282	
JB	144.1	428	141.5	379	
JJ	141.9	75	115.2	75	
JN	138.8	563	144.0	478	
KB	162.9	1,606	182.0	961	
KJ	143.4	499	151.6	454	
KK	146.7	4,672	147.4	3,408	
KN	164.2	1,605	129.3	510	
KW	143.1	263	179.4	112	
SU	137.1	3,217	145.5	4,030	
US	167.6	774	154.0	243	
total	148.2	19,570	150.2	15,806	

Table 4.2 Comparisons of wage for maester and regions

regions	maester		etc		Boxplot for maester and regions
	mean	n	mean	n	
BS	163.4	241	139.9	1,988	
CB	179.3	48	161.7	1,577	
CN	199.8	101	158.4	1,765	
DG	186.5	84	145.2	1,553	
DJ	164.8	41	142.2	959	
IC	150.7	60	142.4	2,607	
JB	151.7	226	139.5	581	
JJ	-	0	128.6	150	
JN	220.0	10	140.4	1,031	
KB	189.4	472	165.7	2,095	
KJ	-	0	147.3	953	
KK	162.3	573	145.8	7,507	
KN	194.1	257	150.6	1,858	
KW	170.0	5	153.7	370	
SU	154.5	120	141.5	7,127	
US	183.4	292	156.9	725	
total	174.4	2,530	147.1	32,846	

Table 4.3 Comparisons of wage for engineering and regions

regions	engineering		etc		Boxplot for engineering and regions
	mean	n	mean	n	
BS	147.5	1,166	136.9	1,063	
CB	156.1	638	166.2	987	
CN	162.2	960	159.0	906	
DG	136.3	987	164.0	650	
DJ	140.9	656	147.5	344	
IC	139.8	1,606	146.7	1,061	
JB	146.1	444	138.9	363	
JJ	132.8	63	125.5	87	
JN	132.1	646	156.0	395	
KB	166.3	1,713	177.5	854	
KJ	148.4	530	146.0	423	
KK	150.8	4,423	142.5	3,657	
KN	163.4	1,539	135.7	576	
KW	143.8	281	184.4	94	
SU	134.2	2,990	147.1	4,257	
US	169.6	666	154.5	351	
total	148.8	19,308	149.4	16,068	

Table 4.4는 Model I의 선형회귀 모형식 (4.1)과 Model II의 위계적 선형회귀 모형식 (4.2)의 분석결과를 제시한 것이다.

Table 4.4의 결과로부터 다음과 같은 사실을 얻을 수 있다.

첫째, 두 모형을 비교한 결과 Model I과 Model II 모두 통계적으로 유의하지만 ($F = 134.840^{**}$, $\chi^2 = 2,116.810^{**}$), Model I에 비해서 Model II가 통계적으로 유의하게 적합도가 높음을 알 수 있다 ($\chi^2(10)=1,462.38^{**}$).

Table 4.4 Comparisons of OLS and HLM

variables		Model I (OLS)		Model II (HLM)	
		B	s.e.	B	s.e.
fixed-effects					
constant		3.894**	.227	1.618	1.000
sex (male dummy)		-.017**	.005	.010	.043
maester		.132**	.007	.151**	.028
engineer		-.003	.005	-.294	.040
income	100~200	.011	.009	.013	.009
	200~300	.033**	.009	.039**	.009
	300~500	.055	.009	.063**	.008
	>500	.093	.011	.089**	.011
education level	college	-.023	.008	-.018*	.008
	≥university	.037	.006	.045**	.006
GPA	very low	-.022	.008	-.028**	.008
	low	-.022	.005	-.017**	.005
	high	.056	.005	.059**	.005
	very high	.115	.006	.120**	.006
satisfaction		.030	.002	.029**	.002
no. of license	1~2	.015	.005	.012**	.005
	≥3	.053	.006	.047*	.006
region	sudo	-.064	.006	-.086**	.015
	chungchung	-.030	.007	-.113***	.016
	junra	-.029	.008	.043*	.017
	daekyung	.065	.007	.065**	.013
	kangwon	-.048	.018	-.156**	.030
	jeju	-.181	.032	.035	.056
EAR		.010**	.001	.003	.004
labor time		.000	.001	.012*	.005
regular wage		.001**	.000	.002**	.001
random-effects					
sd (sex)				.166**	.035
sd (maester)				.101**	.025
sd (engineer)				.152**	.032
sd (constant)				.138**	.028
sd (residual)				.317**	.001
F / χ^2		$F = 134.840^{**}$		$\chi^2 = 2,116.810^{**}$	
n		33,678		33,678	
H_0 : OLS v.s. HLM		$\chi^2(10)=1,462.38^{**}$			

* $p < 0.05$, ** $p < 0.01$.

둘째, 성별, 마이스터고 더미 및 공업계열 더미에 대한 랜덤효과의 결과를 살펴보면, 모두 통계적으로 유의함을 알 수 있다. 즉 취업지역별로 성별, 마이스터고 더미 및 공업계열 더미가 개인수준의 임금에 미치는 영향력은 동일하지 않음을 알 수 있다. Figure 4.1과 Figure 4.2는 각각 성별, 마이스터고 더미 및 공업계열 더미에 대한 취업지역별 랜덤효과와, 이들 변수들의 랜덤효과에 대한 95% 신뢰구간 (왼쪽) 및 표준정규분포의 분위수 (오른쪽)를 제시한 결과이다. 이 결과로부터 이들 변수에 대한 랜덤효과를 확인할 수 있다. 따라서 취업지역의 2-수준변수와 랜덤효과를 고려하는 위계적 선형모형이 타당함을 나타내는 결과라 할 수 있다.

셋째, 고정효과의 결과를 살펴보면, 먼저 2-수준 변수를 추가한 Model II에서 총근로시간과 상용직 평균임금은 개인수준의 임금에 통계적으로 유의하게 영향을 미침을 알 수 있다. 즉, 총근로시간과 상용직 평균임금이 높은 지역일수록 개인수준의 임금이 통계적으로 유의하게 증가한다는 것을 의미한다.

넷째, 성별에 대한 회귀계수를 살펴보면, 선형회귀분석에서는 남자가 여자에 비해서 통계적으로 유의하게 개인수준의 임금이 낮게 나타났지만, 위계적 선형회귀 모형결과에서는 통계적으로 유의하지는 않지만 남자가 여자에 비해서 개인수준의 임금이 높게 나타났다. 이러한 결과는 랜덤효과의 특성이 반영된 결과로 유추할 수 있다.

다섯째, 마이스터고 졸업생은 기타 졸업생에 비해서 개인수준의 임금이 더 높게 나타났으며, 부모소득이 높을수록 더 높게 나타났다. 그리고 부모의 학력은 대졸 이상인 경우가 대졸 미만인 경우보다 개인수준의 임금이 더 높게 나타났으며, 고교성적은 높을수록 임금이 더 높게 나타났다.

여섯째, 대학생활의 만족도가 높을수록, 취득한 자격증 수가 많을수록 개인수준의 임금은 더 높게 나타났다으며, 부산·울산·경남 지역에 비해서 수도권, 전라권, 강원 및 제주도는 개인수준의 임금이 더 낮게 나타났고, 대경권은 더 높게 나타났다.

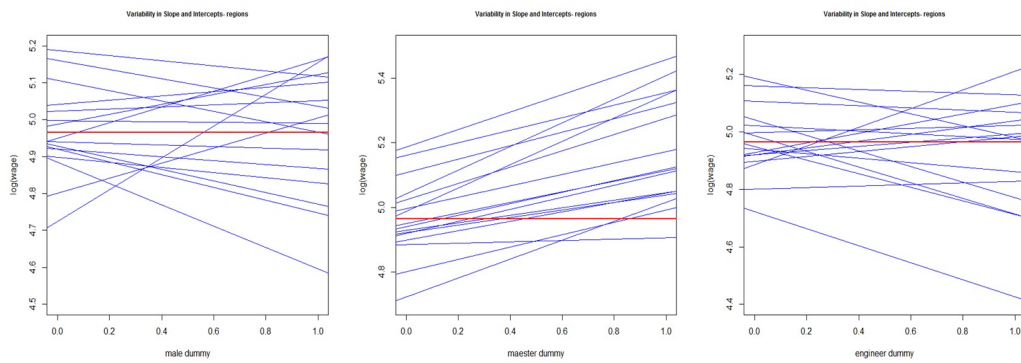


Figure 4.1 Plot of variability in slope and intercepts for regions

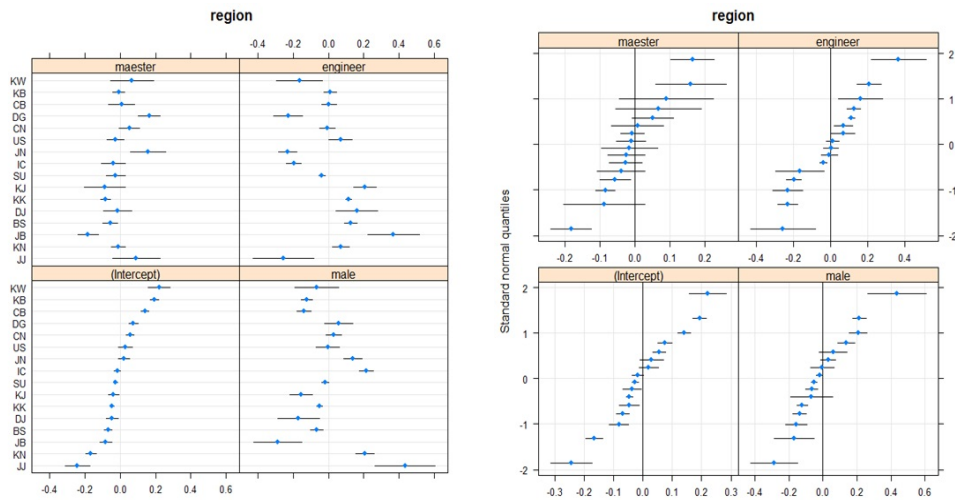


Figure 4.2 95% prediction intervals (left) and quantiles of the standard normal (right) on random effects

5. 결론

본 연구에서는 한국고용정보원에서 실시한 「2013 고졸자 취업진로조사」 자료를 활용하여 임금결정요인을 분석하였다. 특성화고 졸업자의 임금이 개인수준의 인적특성과 취업지역 수준의 산업특성에 의해 영향을 받는 다층구조를 가지게 된다는 점에 초점을 두고 분석하였다. 주요 분석결과를 요약하면 다음과 같다.

첫째, OLS 모형에 비해서 HLM 모형이 통계적으로 유의하게 적합도가 높음을 알 수 있었고, 성별, 마이스터고 더미 및 공업계열 더미에 대한 랜덤효과가 모두 유의함을 알 수 있었다.

둘째, 고정효과와 결과를 살펴보면, 총근로시간과 상용직 평균임금은 개인수준의 임금에 통계적으로 유의하게 영향을 미침을 알 수 있다.

셋째, 전통적인 선형회귀분석에서는 남자가 여자에 비해서 통계적으로 유의하게 개인수준의 임금이 낮게 나타났지만, 위계적 선형회귀 모형결과에서는 통계적으로 유의하지는 않지만 남자가 여자에 비해서 개인수준의 임금이 높게 나타났다.

넷째, 마이스터고 졸업생은 기타 졸업생에 비해서 개인수준의 임금이 더 높게 나타났으며, 부모소득이 높을수록 더 높게 나타났다. 그리고 부모의 학력은 대졸 이상인 경우가 대졸 미만인 경우보다 개인수준의 임금이 더 높게 나타났으며, 고교성적이 높을수록 개인수준의 임금이 더 높게 나타났다.

다섯째, 대학생활의 만족도가 높을수록, 취득한 자격증 수가 많을수록 개인수준의 임금은 더 높게 나타났으며, 부산·울산·경남 지역에 비해서 수도권, 전라권, 강원 및 제주도는 개인수준의 임금이 더 낮게 나타났다.

한편 본 연구에서는 질편, 성별, 마이스터고 여부 및 공업계열 여부에 대해서만 랜덤효과를 반영하여 위계적 선형모형 분석했다는 점과 2-수준 특성변수로 고졸 경제활동비율, 총근로시간 및 상용직 임금만 고려했다는 점에서 연구결과에 대한 지나친 일반화는 무리가 있음을 밝혀 둔다. 따라서 더 많은 개인수준의 1-수준 특성변수를 랜덤효과를 반영하는 부분과 더 많은 2-수준 특성변수를 포함하는 연구는 향후 과제로 남겨둔다.

References

- 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, K. H. and Jung, C. Y. (2014). Factors affecting the determinants of employment decision making of vocational high school senior students. *The Society of Korean Agricultural Education*, **46**, 73-102.
- Im, S. H. (2002). Multilevel models for cross-sectional approach to the developmental research. *Journal of Educational Evaluation*, **15**, 295-315.
- Im, S. H., Sung, J. H. and Won, S. H. (2014). Efficient strategy for the genetic analysis of related samples with a linear mixed model. *Journal of the Korean Data & Information Science Society*, **25**, 1025-1038.
- Jang, H. J. and Lee, J. H. (2012). Determinants of the specialized vocational high school student's career choice using hierarchical generalized linear model. *Journal of Vocational Education & Training*, **15**, 51-74.
- 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, J. H. and Jung, C. Y. (2010). Variables associated with employment barriers of specialized vocational high school students. *The Society of Korean Agricultural Education*, **42**, 25-47.
- Lee, S. C. and Joo, H. J. (2013). The effects of specialized vocational high school characteristics on student employment: Utilizing HGLM model. *Journal of Employment and Skills Development*, **16**, 59-82.
- Nasser, F. and Hagtvat, K. A. (2006). Multilevel analysis of the effects of student and instructor/course characteristics on student ratings. *Research in Higher Education*, **47**, 559-590.

- Park, S. I. and Cho, J. S. (2015). Determinants of employee's wage using hierarchical linear model. *Journal of the Korean Data & Information Science Society*, **26**, 65-75.
- Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical linear models : Applications and data analysis methods*, SAGE publication, San Francisco.

The wage determinants of the vocational high school graduates using mixed effects mode[†]

Jangsoo Ryu¹ · Jangsik Cho²

¹Division of Economics, Pukyong National University

²Division of Mathematics and Applied Statistics, Kyungsoo University

Received 17 May 2016, revised 26 June 2016, accepted 7 July 2016

Abstract

In this paper, we analyzed wage determinants of the vocational high school graduates utilizing both individual-level and work region-level variables. We formulate the models in the way wage determination has multi-level structure in the sense that individual wage is influenced by individual-level variables (level-1) and work region-level (level-2) variables. To incorporate dependency between individual wages into the model, we utilize hierarchical linear model (HLM). The major results are as follows. First, it is shown that the HLM model is better than the OLS regression models which do not take level-1 and level-2 variables simultaneously into account. Second, random effects on sex, master dummy and engineering dummy variables are statistically significant. Third, the fixed effects on business hours and mean wage of regular job for level-2 variables are statistically significant effect individual-level wages. Finally, parental education level, parental income, number of licenses and high school grade are statistically significant for higher individual-level wages.

Keywords: Fixed effects, hierarchical linear model, multiple correspondence analysis, random effects, vocational high school.

[†] This work was supported by a Research Grant of Pukyong National University (2016 year).

¹ Professor, Division of Economics, Pukyong National University, Busan 48513, Korea.

² Corresponding author: Professor, Division of Mathematics and Applied Statistics, Kyungsoo University, Busan 48434, Korea. E-mail: jscho@ks.ac.kr