

두 시점 유사종단자료를 이용한 학교수행 성장모형

최길찬(崔吉燦)

논문 요약

이 글은 같은 학년을 종단적으로 측정하는 유사종단설계에서 활용될 수 있는 학교수행 성장모형을 제시한다. 특히 국가수준 학업성취도 평가가 2년째 수행된 상황에서 두 시점 유사종단자료를 분석하는 모형의 제안을 이 글의 주목적으로 한다. 통계적으로 이 모형은 두 시점간의 성장률을 추정할 때, 초기상태를 통계적으로 통제하는 잠재변인 간의 회귀분석을 포함하는 3수준 잠재변인 위계모형이다. 실제 자료의 분석을 통하여, 이 모형이 어떻게 적용될 수 있는지를 보여주고, 정책적 활용에 대한 논의를 제시한다.

■ 주요어 : 유사종단자료, 국가수준 학업성취도 평가, 성장모형, 잠재변인 회귀분석 3수준 위계모형, 학교수행 평가

I. 서론

이 글은 국가수준 학업성취도 평가가 2년째 수행된 상황에서 각 학교의 성취수준의 향상도를 측정하는 통계적 모형을 제시하고, 그 예를 제시하는 것을 목적으로 한다. 이 모형은 다음의 몇 가지 큰 특징을 가지고 있다.

첫째, 국가수준 학업성취도 평가는 같은 학생을 종단적으로 측정하는 것이 아니라 같은 학년의 학생을 종단적으로 측정하는 유사종단 설계(quasi-longitudinal design) 혹은 코호트 설계(cohort design)에 기초한다(Willms & Raudenbush, 1989; 최길찬, 2005). 따라서 분석을 통해 얻어진 결과는 학생 개인의 성장을 통계적으로 추론하는 것이 아니라 학교의 평균적인 수행의 성장을 추론하는 것이다.

둘째, 이 모형은 측정시점이 두 개인 상황을 가정한다. 즉, 이 글에서 제안하는 통계적 모형은 국가수준 학업성취도 평가를 통해 수집된 2년간의 자료를 분석하는 데 적합한 통계적 모형이다.

측정시점이 들인 경우는 일반적으로 통계모형의 자유도(degree of freedom)가 부족한 문제가 발생한다. 즉, 자료를 통해 얻을 수 있는 정보의 수보다 추정해야 할 모수치의 수가 많은 문제가 발생하는데, 그 해결책은 추정하는 모수치의 수를 줄여서 자유도가 0 이상이 되도록 하는 것이다.

셋째, 이 모형은 각 학교의 첫 번째 연도(즉, 초기상태)의 평균, 첫 번째 연도와 두 번째 연도간의 평균의 차이(즉, 성장률)를 추정하는 것을 주 목적으로 한다. 이는 기본적으로 1989년에 Willms와 Raudenbush 가 제안한 모형과 기본적으로 동일하다. 다만, 그들이 제안한 모형에서는 각 학교의 초기상태 대신에 두 시점의 평균값을 추정하고, 두 시점간의 차이를 추정하였다.

넷째, 이 모형은 두 시점간의 성장률을 추정할 때, 초기상태를 통계적으로 통제한다. 다시 말해서, 학교의 첫 번째 연도의 상태가 성장률의 독립변인으로 모형에 투입된다. 이는 통계적으로는 잠재변인 간의 회귀분석(latent variable regression; Choi & Seltzer, 2010; Choi, Seltzer, Herman, & Yamashiro, 2007; Seltzer, Choi, & Thum, 2003; Choi, 2001)을 의미하며, 실제에 있어서 비슷한 초기상태를 가진 학교들간에 2년간의 학교 평균 성장률을 비교하는 것이 된다.

요약하자면, 이 글에 제시된 통계적 모형은 두 시점 유사종단자료 분석에 적합한 모형으로서, 초기상태와 성장률 간의 관계를 회귀분석 모형으로 모형 내에 포함한 3수준 잠재변인 위계모형이다. 이 모형을 이용하면 2년간의 학교의 수행수준을 평가할 때 초기상태를 통계적으로 통제한 후에 성장률을 비교할 수 있다. 다음 절에 두 시점 코호트 자료 성장모형을 제시하고, 이 모형을 실제 자료에 적용하여 분석한 결과를 제시한다. 그리고 이 모형의 적용 및 정책적 활용에 대한 논의를 마지막 절에 제시한다.

II. 두 시점 코호트 자료 성장모형

분석에 이용되는 자료는 3수준 위계구조를 가지고 있다. 즉, 학생이 연도(혹은 코호트)에 소속되어 있고, 종단적 코호트가 학교 내에 속해있는 구조이다. 따라서 1수준(학생간; 코호트내) 모형은 연도별 학교 평균을 규정하는 모형이다.

$$Y_{ijk} = \beta_{0jk} + r_{ijk} \quad r_{ijk} \sim N(0, \sigma^2) \quad (1)$$

종속변인 Y_{ijk} 는 학교 k ($k = 1, 2, \dots, K$), 연도 혹은 코호트 j ($j = 1, 2$)에 속한 학생 i 의 학업성취도 점수이다. K 는 전체 학교수이고, 연도 j 는 2년 자료를 가정하고 있기 때문에, 1 혹은 2의 값을 갖고, 수식 (1)에서 β_{0jk} 는 학교 k , 연도 j 에 속한 학생들의 평균이다. 1수준 모형은 각 학교의 해당 연도 j 의 기대값(평균값)을 추정한다. 한 가지 언급할 것은 각 시점(코호트)에 속한 학생은 서로 다르기 때문에 각 시점간 학생 배경변인의 차이를 통제할 수 있다는 점이다.

$$Y_{ijk} = \beta_{0jk} + (X_{ijk} - \bar{X}_k) + r_{ijk} \quad r_{ijk} \sim N(0, \sigma^2) \quad (1-1)$$

위의 수식 (1-1)은 하나의 학생배경변인, X_{ijk} , 을 포함한 모형이다. 만일 이 변인을 전체 평균에 의해 중심화(centering) 하면 β_{0jk} 은 학생배경변인의 영향력이 통계적으로 조정된 년도 j , 학교 k 의 평균값이다.

$$\beta_{0jk} = \theta_{0k} + \theta_{1k}Year_{jk} \quad (2)$$

$$\beta_{0jk} = \theta_{0k} + \theta_{1k}Year_{jk} + u_{0jk} \quad u_{0jk} \sim N(0, \tau_{\beta 0}) \quad (2-1)$$

위의 2수준(코호트 간; 학교 내) 모형에서 3가지 중요한 점이 있다. 첫째, 각 학교 k 는 오로지 2시점의 학교평균(β_{0jk})만이 있다. 즉, 첫 번째 연도 평균값(β_{01k})과 두 번째 연도 평균값(β_{02k})이다. 둘째, 이로 인해 모형의 무선효과(random effect)를 추정할 수 없다. 이는 각 학교별로 추정해야 할 모수는 3개 (θ_{0k} , θ_{1k} , u_{0jk}) 인데, 학교별로 우리가 가진 정보는 둘(1차년도 평균 과 2차년도 평균) 이기 때문에 각 학교별로 3개 모수치 중 적어도 하나는 고정시켜서 모형의 자유도가 최소한 0이 되도록 해야 한다. 따라서 무선효과가 제외된 수식 (2)의 모형만을 추정할 수 있다. 참고로 만일 3시점 이상의 자료가 있는 일반적인 경우 수식 (2-1)의 모형을 추정할 수 있다. 셋째, 어떤 연구문제를 추구하는냐에 따라 시간변인($Year_{jk}$)의 코딩값이 달라진다. 몇 가지 경우를 살펴보자.

만일 θ_{0k} 가 학교 k 의 첫 번째 시점에서의 상태를 나타내고, θ_{1k} 가 첫 번째 시점과 두 번째 시점간의 차이가 되도록 할 경우, 시간변인은 0 과 1로 코딩한다. 반면에 θ_{0k} 가 학교 k 의 2년간의 학교수행 수준의 평균을 나타내고, θ_{1k} 은 여전히 첫 번째 시점과 두 번째 시점간의 차이가 되도록 할 경우 시간변인은 1/2 과 1/2로 코딩한다. 마지막으로 θ_{0k} 가 학교 k 의 두번째 시점에서의 상태를 나타내고, θ_{1k} 는 첫번째 시점과 두번째 시점간의 차이가 되도록 할 경우, 시간변인은 1과 0으로 코딩한다.

$$\theta_{0k} = \Phi_{00} + V_{0k} \quad V_{0k} \sim N(0, \tau_{\Theta 0}) \quad (3a)$$

$$\theta_{1k} = \Phi_{10} + V_{1k} \quad V_{1k} \sim N(0, \tau_{\Theta 1}) \quad Cov(V_{0k}, V_{1k}) = \tau_{\Theta 0 \Theta 1} \quad (3b)$$

공식 (3a) 와 (3b) 는 3수준(학교 간) 모형을 나타낸다. 먼저, 앞 단락에서 제시한 시간변인 코딩방법 중 첫 번째를 택했다고 가정하면, Φ_{00} 는 첫 번째 년의 전체 학교의 평균이며, Φ_{10} 는 첫 번째 연도와 두 번째 연도 간의 전체 학교 평균 성장정도이다. 이 모수들은 고정효과(fixed effect) 계수라고 하며, 이에 반해 무선효과(random effect)인 V_{0k} 는 학교 k 의 첫 번째 연도 전체 평균에서의 편차이며, V_{1k} 는 학교 k 의 전체 학교 평균 성장으로 부터의 편차이다. 즉, 학교

k의 첫 번째 연도 평균과 성장 정도가 전체 평균으로 보다 얼마나 큰지 혹은 작은지를 나타내 준다. 이 무선효과들은 각각 평균이 0이고 변량이 τ_{e0} 와 τ_{e1} 인 정상분포를 따르며, 이 두 무선효과의 공변량은 τ_{e0e1} 이다.

학교의 수행정도를 평가할 때, 이 무선효과의 크기에 따라 순위를 정하는 방식이 있을 수 있는데, 이것이 최근 부가가치모형(Value Added Model: VAM; Sanders, Saxton, & Horn, 1997; Raudenbush & Bryk, 2002; McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004; Posnik & Bryk, 2005)이라 불리는 모형이 택하는 방법이다. 즉, 각 학교 k가 비교대상이 되는 가상적 근거(counterfactual)에 비해 얼마나 큰 정적 혹은 부적인 값을 갖느냐에 따라 그 학교의 효과를 판단하는데, 보통의 경우 전체 평균이 비교대상이 된다. 이것을 부가가치 추정치(Value-added estimate: VAE)라고 하며, 이 글에서 제시한 모형에서는 무선효과인 V_{0k} 와 V_{1k} 이다.

이 때, 아마도 세 가지 준거에서 학교수행 수준을 판단해 볼 수 있다. 첫째, 첫째 연도의 수행 수준에 따라 학교를 평가할 수 있다. 즉, 수식 (3a)에 있는 V_{0k} 의 크기에 따라 학교의 순위를 정할 수 있다. 둘째, 첫째 연도와 둘째 연도간의 성장 정도에 따라 학교를 평가할 수 있다. 즉, 수식 (3b)에 있는 V_{1k} 의 크기에 따라 학교의 순위를 정할 수 있다. 이 때 중요한 연구 문제 중 하나는 이들 두 평가 준거 간의 상관관계라고 할 수 있다. 즉, 첫 번째 연도에 잘 했던 학교가 더 많이 성장을 하는지 아니면, 반대로 첫 번째 연도에 잘하지 못했던 학교가 더 많이 성장을 하는지의 문제는 늘 흥미로운 연구문제라고 할 수 있다.

세 번째 평가준거는 학교의 1년 동안의 성장의 정도에 따라 평가하는데 있어, 첫 번째 연도의 수행수준이 비슷한 학교들끼리 비교하는 방식을 고려해 볼 수 있다. 예를 들어, 성장의 정도를 첫 번째 연도의 수행수준이 10 백분위에 있던 학교들끼리 비교하거나, 혹은 90 백분위에 있던 학교들끼리 비교할 수 있다. 이는 학교의 성장 기대 값이 표집학교 전체의 평균이 아니라 첫 번째 연도에 비슷한 수행수준을 가진 학교들의 평균 성장 정도가 된다. 통계방법론적 측면 볼 때 Wilms와 Raudenbush가 제안한 방법 이 점에서 차이가 난다. 통계적으로는 첫 번째 평가의 준거가 X축 (V_{0k})이고, 두 번째 평가 준거가 Y축 (V_{1k}) 이라고 할 때, 세 번째 평가 준거는 회귀선이 되며, 이는 회귀선에 의한 잔차(Residualized gain; $R(y|x)$) 이다. 세 번째 평가준거는 많은 경우 의미있는 준거가 될 수 있다. 왜냐하면, 학교의 수행정도를 비교할 때 초기 상태가 비슷한 학교들끼리 비교하는 것이며, 통계적으로 말하면 초기상태의 영향력을 통계적으로 통제한다는 의미이며, 이는 모든 학교가 초기상태에 관한한 동일한 조건하에서(leveling the playing field) 성장 정도를 비교하게 되기 때문이다.

세 번째 평가준거를 모형 내에서 구현하기 위해서는 초기상태를 수식 (3b)에 독립변인으로 포함해야 한다. 즉, 수식 (3a)와 (3b)는 수식 (3c)와 (3d)로 다시 쓰여져야 한다. 수식 (3d)에서 초기상태 (θ_{0k})는 전체평균에 의해 중심화되어 수식의 오른쪽에 포함되었다.

$$\theta_{0k} = \Phi_{00} + V_{0k} \quad V_{0k} \sim N(0, \tau_{\theta 0}) \quad (3c)$$

$$\theta_{1k} = \Phi_{10} + b \times (\theta_{0k} - \Phi_{00}) + V_{1k} \quad V_{1k} \sim N(0, \tau_{\theta 1}) \quad Cov(V_{0k}, V_{1k}) = 0 \quad (3d)$$

회귀 계수 b 는 초기상태와 성장 정도간의 관계를 나타내는 잠재변인 회귀계수(latent variable regression coefficient)이며, Φ_{10} 는 초기상태가 전체 초기상태 평균과 같은 학교의 성장 기대값이다. 무선효과 V_{1k} 는 초기상태를 통계적으로 통제한 후의 학교 k 의 전체 학교 평균 성장으로부터의 편차이다. 초기상태가 수식 (3d)의 독립변인으로 포함되었기 때문에 두 무선효과 간의 공변량은 0이다. 이 모형은 위계모형 내에 두 잠재변인들 간의 회귀모형을 통합한 모형이다. 한 가지 중요하게 언급할 것은 학교효과연구나 정책평가 연구에서 가장 중요하게 고려하는 학교수행수준이나 성장에 미치는 학교배경변인 혹은 정책관련 변인들이 위의 수식에 포함될 수 있다는 점이다. 이 경우 포함된 변인들의 효과는 학교의 초기상태를 통제하고 난 후의 각 변인들의 효과를 의미하게 된다.

III. 자료 분석 결과

1. 기술통계치

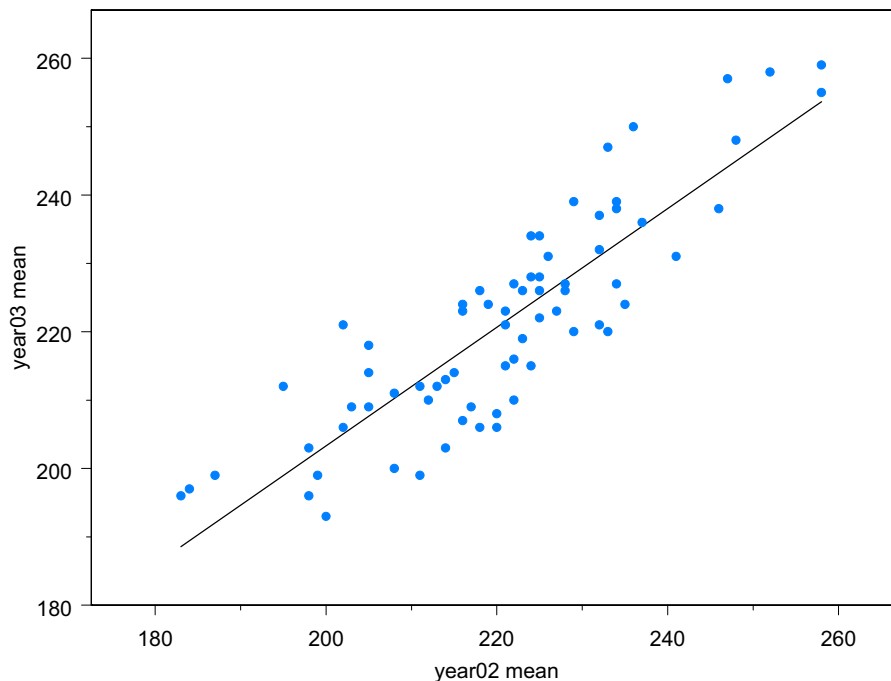
앞에 제시한 2시점 코호트 자료 성장모형을 이용하여 국가수준 학업성취도 평가에서 얻어진 자료와 같은 구조를 가진 자료를 분석하였다. 이 자료는 미국 북서부에 있는 한 교육구의 72개 초등학교 5학년의 2년간(2002년과 2003년도)의 기초학력검사 수학과목 성취도 자료이다. 참고로 이 교육구의 초등학교는 5학년이 마지막 학년이며, 이는 국가수준 학업성취도 평가가 초등학교 마지막 학년인 6학년을 대상으로 하는 것과 유사하다고 할 수 있다.

<표 1>에 보는 바와 같이 첫번째 연도의 교육구 전체 72개 학교의 평균은 약 220.75점이며, 두 번째 연도는 221.32로 거의 차이가 없었다. 표준편차도 2년간 약 25로 동일하였다. 2년간의 학교별 평균의 차이의 전체 평균은 0.56 정도 였으며, 가장 많이 성장한 학교는 전년도 대비 18.7 점 정도 성장하였고, 13.4점 정도 줄어든 학교가 부적으로 가장 많이 퇴보한 학교였다. 약 반 이상의 학교는 정적인 성장을 한 반면, 나머지 반 정도의 학교는 부적인 성장을 한 것으로 나타났다.

<표 1> 72개 학교의 수학성취도 점수의 2년간 기술통계치

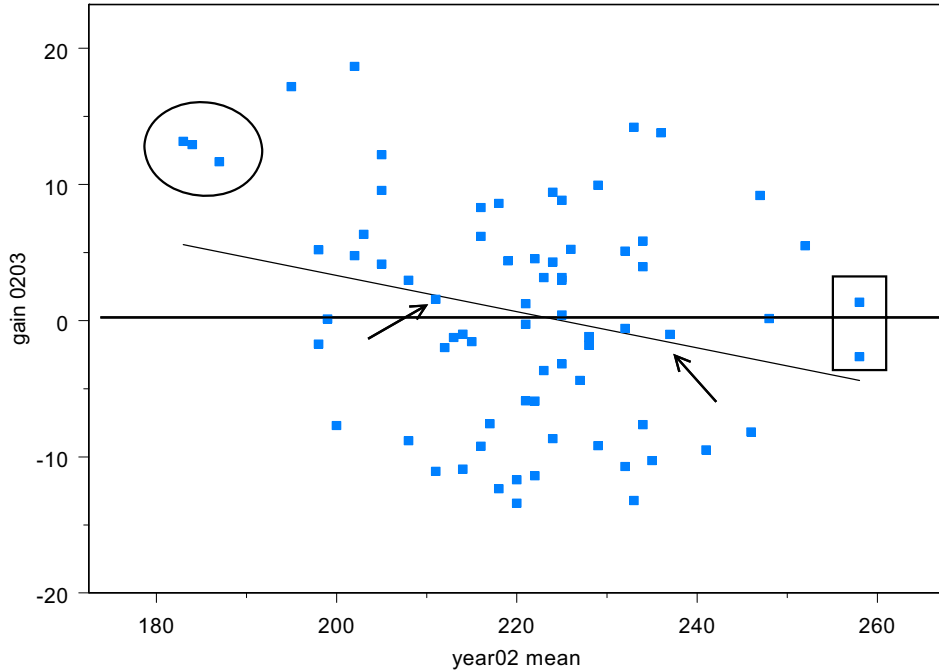
	사례수 (학생수)	평균	표준편차	최소값	최대값
2002년도	72 (2468)	220.75	25.68	139	286
2003년도	72 (2475)	221.32	25.27	152	281
2년간 성장	72	0.56	8.07	-13.4	18.66

[그림 1] 은 각 학교의 각 연도의 관찰점수 평균을 그림으로 나타낸 것이다. 여기서 보는 바와 같이 첫 번째 연도 학교 평균과 두 번째 연도 학교 평균 간에 높은 정적인 상관관계가 있다. 즉, 첫 번째 연도에서 높은 성취도를 가진 학교가 두 번째 연도에서도 높은 성취도를 보이는 것으로 나타났다. [그림 1]의 결과는 매년 학교의 수행수준을 횡단적인 관점에서 평가한다면, 학교의 수행수준의 순위가 2년간 크게 변하지 않을 것임을 함의한다.



[그림 1] 2002년과 2003 년 각 학교별 수학 평균 및 회귀선

이에 반해 [그림 2] 는 재미있는 두 가지 관점을 제시한다고 할 수 있다. 먼저, 앞서 학교를 평가하는 두 번째 준거로 제시한 학교의 1년간의 수행수준의 성장의 관점에서 그림을 살펴보자. 2년간의 학교 전체 성장 평균인 0.56이 평가준거가 될 수 있다([그림 2]의 수평선). 먼저 2002년에 평균이 거의 제일 낮았던 3개 학교([그림 2]의 좌측 상단의 표시된 학교들)의 학교 평균의 1년간의 변화가 10점 이상인 것이 두드러지게 나타난다. 반면에 2002년에 가장 평균이 높았던 두 학교([그림 2]의 우측 중간에 표시된 학교들) 중 하나는 교육구 전체 평균 성장 보다 아주 조금 더 높게 성장하였고, 다른 하나는 평균보다 저조한 성장을 나타냈다. 이처럼 1년간의 성장에서 볼 때 그림 위의 수평선 위에 위치한 학교들은 교육구 전체 평균 성장 보다 높게 성장한 학교들이고, 반대로 수평선 아래에 위치한 학교들은 교육구 전체 평균 성장 보다 낮게 성장한 학교들이다.



[그림 2] 2002년 수학 평균과 2002년 2003년간의 평균 성장 및 회귀선

마지막으로 [그림 2]에 나타난 회귀선을 기준으로 학교를 평가해 보자. 첫째, 회귀선이 Y축의 평균값을 나타내는 수평선과 그리 다르지 않은 것으로 판단하여 보면, 독립변인, 즉 첫 번째 연도의 관찰 평균과 종속변인, 즉 두시점 간 학교 평균의 차이는 관련성이 높지 않음을 알 수 있다. 회귀선을 기준으로 학교를 평가한다는 의미는 X축을 조건화 한 후에 Y값의 크기를 비교한다는 것이다. 그림에 나타난 학교들 중에서 전체 평균을 기준으로 보면 부적인 잔차를 가지지만, 회귀선을 기준으로 보면 정적인 잔차를 가지는 학교들이 있음을 볼 수 있다 (오른쪽 화살표 학교). 반대의 경우, 즉 전체 평균을 기준으로 보면 정적인 잔차를 가지지만, 회귀선을 기준으로 보면 부적인 잔차를 가지는 학교 (왼쪽 화살표 학교)도 있지만, 전체 평균을 나타내는 수평선과 회귀선이 크게 다르지 않기 때문에 이 두 다른 기준을 적용하더라도 학교의 평가는 크게 달라지지 않음을 알 수 있다.

2시점 코호트 자료 성장모형 적용 결과

<표 2>에 제시된 결과는 두시점 코호트 자료 성장 모형 (수식 (1), (2), (3c), (3d))을 베이지안 통계분석을 이용하여 분석한 결과이다. 이 분석은 WinBUGS1.4 (Spiegelhalter, Thomas, Best, & Lunn, 2003) 통계 프로그램을 이용하였다.

<표 2> 두시점 코호트 자료 성장 모형 분석 결과

모수	추정치	95% 신뢰구간
초기상태 효과(b)	-0.046	(-0.181, 0.089)
02년도 전체 평균(Φ_{00})	221.1	(217.4, 224.8)
1년간 성장 평균(Φ_{10})	0.059	(-1.797, 1.857)
1수준 모형 변량(σ^2)	452.2	(434.4, 470.7)
3수준 모형 변량 02년도 평균(τ_{00})	238.2	(164.8, 339.8)
1년간 성장 (τ_{01})	35.23	(17.3, 60.12)

표에서 제시된 고정효과 추정치는 각 모수의 사후분포(posterior distribution)의 평균값이며, 사후분포의 2.5% 와 97.5% 의 값(95% 신뢰구간)이다. 2002년도 표집학교 전체 평균 수학성적은 221.1 이었으며, 초기상태가 전체 평균과 같은 학교의 1년간 성장은 0.059로 거의 0에 가까웠다. 이 추정치의 신뢰구간 내에 0이 포함되어 있으므로, 통계적으로 0과 같다고 할 수 있다.

초기상태의 효과 또한 통계적으로 유의미 하지 않은 것으로 나타났다. 잠재변인 회귀계수 (b) 추정치는 0.046 이며, 사후분포의 95% 신뢰구간은 0.181과 0.089로 나타났다. 따라서 초기상태와 1년간의 성장은 부적인 관계를 보이지만, 이는 통계적으로 유의미하지 않았다. 이러한 결과는 앞서 제시한 기술 통계치와 기술통계치를 이용한 그림을 통해 간접적으로 살펴본 것과 일치하는 결과라고 할 수 있다.

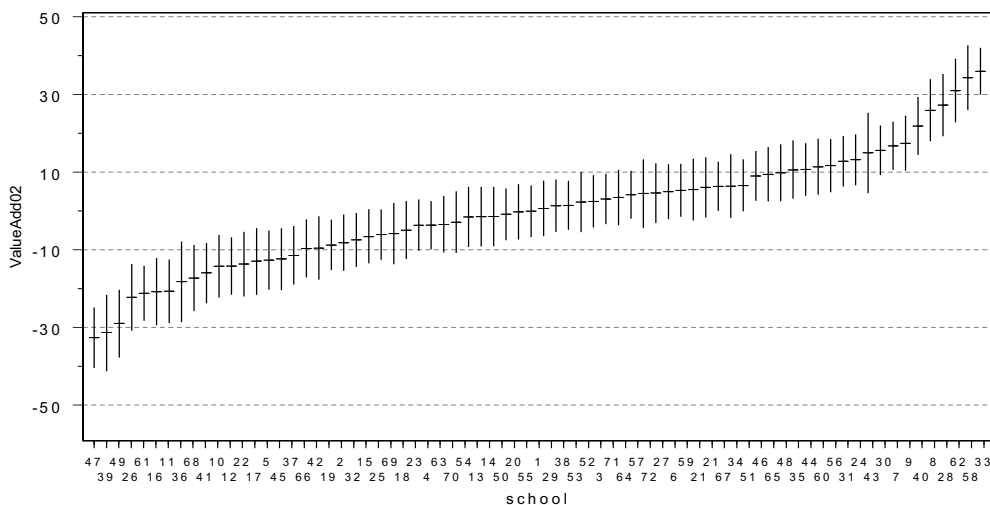
부가가치 추정치들에 관해 살펴보자. 앞서 제시한 통계모형의 모수 추정치들은 표집 전체 학교의 전체적인 현황을 요약해 준다고 할 수 있다. 즉, '전체 학교의 초기 상태의 평균은 얼마인가?', '초기 상태를 통제 한 후의 전체 학교 성장 평균은 얼마인가?', '초기상태와 성장간의 관계는 어떠한가?', '초기상태의 학교들 간 분포는 어떠한가?', '성장의 학교들 간 분포는 어떠한가?' 등이 주요 관심이다. 이에 반해, 부가가치 추정치는 개별 학교의 수행에 대한 정보와 이를 기초로 한 학교간의 비교에 주된 관심이 있다. 다시 말해, '학교 A는 학교 B와 초기상태에 있어 차이가 있는가?', '학교 C는 학교 D 보다 초기 상태를 통제 한 후 더 많이 성장하였는가?', 더 나아가 '학교의 순위는 어떠한가?' 등이 주된 질문이 된다.

부가가치 추정치를 이용한 학교수행 평가에서 [그림 3]은 수식 (3c)의 무선효과인 V_{0k} 의 사후 분포 평균과 95% 신뢰구간을 가장 작은 값을 가지는 학교부터 순서대로 나타낸 것이다. 앞서 언급한 것처럼 이 추정치의 전체 평균은 0 이다. 그림에서 보는 바와 같이 약 절반의 학교는 정적인 추정치를 갖고, 나머지 절반의 학교는 부적인 추정치를 갖는다. 특정 학교와 그 주변에 있는 학교들의 95% 신뢰구간은 서로 거의 많이 겹친다. 두 학교의 신뢰구간이 겹친다고 해서 반드시 그 두 학교의 부가가치 추정치가 통계적으로 다르지 않다고 결론내릴 순 없으나, 많은 학교들이

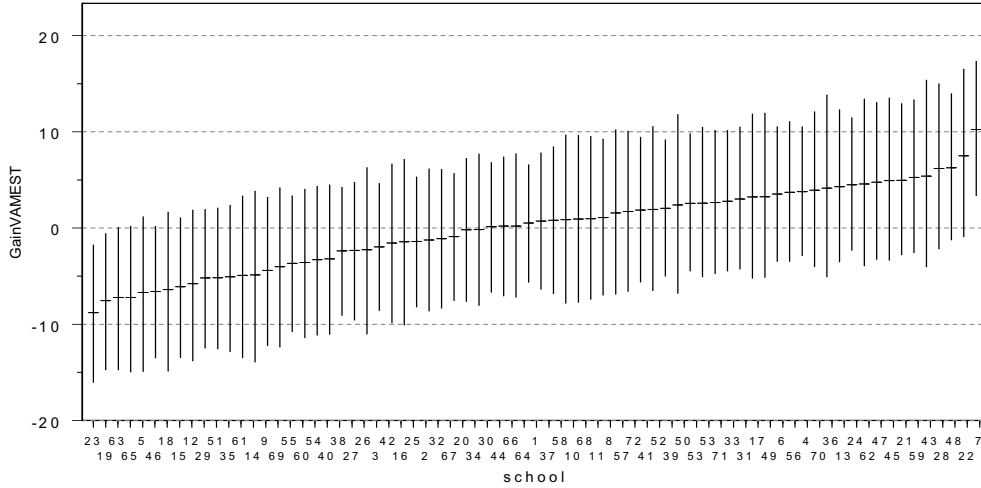
부가가치 추정치가 서로 다르지 않다는 것을 추론할 수 있다. 하지만, 부가가치 추정치가 가장 작은 몇 개 학교와 가장 큰 몇 개 학교는 통계적으로 2002년 학교 수행에 있어 차이가 있었다.

(초기 상태를 통제 한) 성장 부가가치 추정치 결과는 [그림 4]에 나타나 있다. 각 학교의 추정치는 수식 (3d)의 무선효과 V_{1k} 의 사후분포 평균과 95% 신뢰구간으로 표현되어 있고, 추정치의 전체 평균은 0이다. 또한 각 학교의 성장 부가가치 추정치는 초기상태를 통제 한 후의 값이다. 이 그림에서 두드러진 것은 거의 모든 학교 추정치의 95% 신뢰구간이 서로 겹친다는 것이다. 이는 거의 대부분의 학교는 1년간의 성장에 있어 통계적으로 차이가 없을 가능성이 많음을 나타낸다. 이같은 현상의 주된 이유는 추정오차(estimation error)가 크기 때문이며, 이는 무선효과의 정보 단위인 표집수, 즉 학교 내 학생수가 작기 때문이다. 하지만, 추정치가 가장 작은 학교 #23과 #19는 추정치가 가장 큰 학교 #22와 #7과 통계적으로 유의미하게 초기상태를 통제 한 후의 성장에 차이가 있었다.

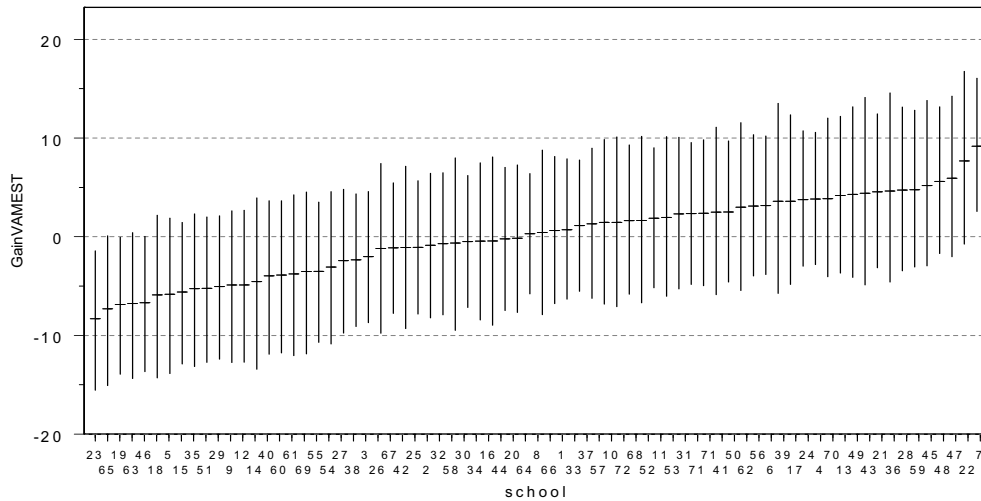
[그림 5]는 초기상태를 통제하지 않은 모형 (수식 (3a)와 (3b))의 결과를 제시한다. 초기상태와 성장간의 관계가 매우 약하고, 통계적으로 유의미하지 않은 결과를 통해 쉽게 유추해 볼 수 있는 것처럼 [그림 5]의 결과는 [그림 4]의 결과와 크게 다르지 않다. 즉, 초기상태의 통제여부와 무관하게, 학교의 성장 부가가치 추정치는 크게 변하지 않고 있다. 두 결과를 가장 쉽게 비교해 볼 수 있는 방법은 학교 순위를 비교해 보는 것이다. 초기상태를 통제 한 모형과 통제하지 않은 모형에 의해 정해진 학교 순위 간의 Spearman Brown 순위상관 계수 0.98 이었고, Kendal의 Tau 계수는 0.90 이었다. 만일 초기상태의 효과가 크고, 통계적으로 유의미하다면, 순위상관 계수는 훨씬 낮았을 것이다. 이론적으로 모형내에 통계적으로 유의미한 변인을 포함시키면 무선효과의 추정오차가 작아지는 효과가 있으나, 각 학교의 추정치는 전체 평균으로 축소되는(shrunk) 경향이 있다.



[그림3] 초기상태 (2002년) 학교 평균 부가가치 추정치



[그림 4] 초기상태를 통제한 후의 학교별 2002-2003년간의 성장 부가가치 추정치



[그림 5] 초기상태를 통제하지 않은 학교별 2002-2003년간의 성장 부가가치 추정치

IV. 결론 및 논의

이 글은 국가수준 학업성취도 평가를 통해 수집된 2년간의 성취도 자료를 분석하는 통계적 모형을 제시하고, 이 모형을 비슷한 자료 구조를 가진 미국의 한 교육구 기초학력 성취도 자료에 적용하고, 그 결과를 제시하였다. 가장 큰 자료상의 특징은 같은 학생을 반복적으로 측정하는 종단자료 구조가 아닌 같은 학년을 반복적으로 측정하는 유사종단 자료 구조이라는 것이며, 자료 시점에 있어서는 단지 2년간의 자료라는 점이다. 이러한 자료를 이용한 학교수행평가에 적합

한 통계모형으로 제시한 두시점 코호트 자료 성장모형은 통계적으로 잠재변인 간의 회귀분석을 포함한 모형이며, 이 연구에서는 학교의 초기상태를 통제한 후의 1년간의 성장을 추정하는 방법으로 제안되었다.

실제에 있어서 이 모형은 크게 2가지 측면에서 이용될 수 있다. 첫째로 모형 내의 고정효과 계수에 초점을 맞춰서 학교효과연구나 정책연구에서 학교수행이나 성장에 미치는 변인들의 영향력을 추정하는 모형으로 이용될 수 있다. 이 때 학교초기 상태를 독립변인으로 모형 내에 투입함으로써, 초기상태를 통제한 후의 학교변인이나 정책변인의 효과를 추정할 수 있다. 둘째, 무선효과에 초점을 맞춰서 최근에 학교나 교사효과를 추정하는 통계모형으로 많은 관심을 받고 있는 부가가치모형으로 이용될 수 있다. 이 글에서는 3개의 학교수행 평가 준거아래 부가가치 추정치의 크기에 따라 학교들을 비교하는 방법을 제시하였다.

마지막으로 이 연구에서 제시된 자료 분석 결과를 기초로 실제로 국가수준 학업성취도 평가의 분석에서 의미 있게 고려할만한 점을 제시한다. 첫째, 이 연구에서는 학교 초기 상태와 1년간의 학교수행수준의 성장과의 관계가 통계적으로 유의미하지 않은 것으로 나타났다. 이는 학생의 성취도를 종단적으로 수집하여 성장을 분석하는 패널자료 혹은 종단자료의 경우에 학생의 초기상태와 학생의 성장간의 높은 정적 상관관계가 있는(Choi & Seltzer, 2010) 것과는 불일치하는 결과이다. 이러한 결과는 이 연구에 국한되는 것인지 아니면 국가수준 학업성취도 평가 연구에서도 비슷한 결과가 나타나는지 여부는 흥미 있는 연구문제라고 할 수 있다.

둘째, 이 연구에 제시된 것처럼 일반적으로 무선효과에 기초한 부가가치 추정치의 추정의 정확성(precision)은 낮다. 부가가치 모형을 이용하여 학교나 교사의 효과를 추정하고 이를 근거로 학교 혹은 교사의 책무성을 평가하는 경우에 가장 중요한 2가지 통계적 요소는 추정의 정확성(accuracy)과 편향(bias)이다. 즉, 편향 정도를 줄이고, 정확성을 높이는 것이다. 추정의 정확성은 표집자료의 수에 관련이 있고, 추정의 편향은 무선택당 실험연구(cluster randomized experiment study)와 달리 학생이나 교사가 교실이나 학교에 무선택적으로 할당되지 않기 때문에 발생하는 것이다. 현재 미국에서는 비무선택당으로 인해 초래되는 추정의 오류를 극복하기 위해 경제통계학에서 많이 이용되는 고정효과모형(fixed effect model)에 대한 논의가 활발해 지고 있다. 하지만, 한국의 경우 교사의 순환근무제로 인해 이러한 무선택당의 문제가 어느 정도 해결된 아주 독특한 환경이기 때문에 추정의 오류의 문제에서 덜 심각한 위치에 있다고 할 수 있다. 따라서 국가수준 학업성취도 평가연구에서 수집된 자료를 이 모형에 적용하는 연구는 통계적 모형의 관점 뿐 아니라 실제 학교효과연구나 정책평가 연구의 관점에서 많은 함의점을 갖는다고 할 수 있다.

셋째, 국가수준 학업성취도 평가가 매년 실시됨에 따라 자료의 시점 수는 점점 늘어나며, 이에 따라 다양한 성장곡선을 모형에 적용할 수 있다. 예를 들어, 초등학교 6학년의 국어과목 성취도

의 연도별 변화추이가 어떤 모양을 가지는지를 추정할 수 있으며, 중요한 교육정책의 시행 혹은 학교배경변인의 변경에 의해 성장곡선이 어떻게 바뀌는지 여부도 탐구할 수 있다(예를 들어, interrupted time series model). 두 시점 코호트 자료 성장모형의 특징인 잠재변인 회귀분석의 아이디어를 이용하면, 다 시점에서 추정된 성장곡선들 간의 관계를 파악해 볼 수도 있을 것이다. 이와 같은 모형의 확장은 학교효과연구와 교육정책평가 연구에 지금보다 풍성한 시사점을 도출해 줄 수 있을 것으로 기대된다.

참고문헌

- 최길찬 (2005). 유사종단자료를 이용한 학교효과 추정 위계적 모형: 척도점수와 NCE 점수와의 비교. *아시아교육 연구*, 6(1), 59-81.
- Choi, K. (2001). Latent variable modeling in the hierarchical modeling framework in longitudinal studies: A fully Bayesian approach. *Asia Pacific Education Review*, 2(1), 44-55.
- Choi, K., & Seltzer, M. (2010). Modeling heterogeneity in relationships between initial status and rates of change: treating latent variable regression coefficients as random coefficients in a three level hierarchical model. *Journal of Educational and Behavioral Statistics*, 35(1), 54-91.
- Choi, K., Seltzer, M., Herman, J., & Yamashiro, K. (2007). Children Left Behind in AYP and Non AYP Schools: Using Student Progress and the Distribution of Student Gains to Validate AYP. *Educational Measurement: Issues and Practice*, 26(3), 21-32.
- McCaffrey, D., Lockwood, J. R., Koretz, D., Louis, T., & Hamilton, L. (2004). Models for value added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, 29 (1), 67-101.
- Ponisciak, S. M., & Bryk, A. (2005). Value added analysis of the Chicago public schools: An application of hierarchical models. In *Value added modeling: Issues with theory and applications*, ed. R. Lissitz, 40 - 81. Maple Grove, MN: JAM Press.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods, 2nd edition*. Newbury Park, CA: Sage Press.
- Sanders, W. L., Saxton, A. M., & Horn, S. P. (1997). The Tennessee value added assessment system: A quantitative outcomes based approach to educational assessment. *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* J. Millman. Thousand Oaks, CA, Corwin Press, Inc.: 137-162.
- Seltzer, M., Choi, K., & Thum, Y. M. (2003). Examining relationships between where students start and how rapidly they progress: Using new developments in growth modeling to gain insight into the distribution of achievement within schools. *Educational Evaluation and Policy Analysis*, 25(3), 263-286.

Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). *WinBUGS: windows version of Bayesian inference using Gibbs sampling, version 1.4, User Manual*. MRC Boistatistics Unit, Cambridge University.

Willms, D., & Raudenbush, S. (1989). A longitudinal hierarchical linear model for estimating school effects and their stability. *Journal of Educational Measurement*, 26(3), 209-232.

* 논문접수 2010년 2월 1일 / 1차 심사 2010년 2월 26일 / 게재승인 2010년 6월 11일

* 최길찬: 선임 연구원, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, 300 Charles E Young Drive, GSE&IS Mailbox 951522, Los Angeles, CA 90095; 주요 관심연구 분야는 잠재변인 회귀분석 위계모형, value-added models, 학교 및 교사 효과 연구, 베이지안 통계방법, 성장모형임.

* e-mail: kcchoi@ucla.edu.

Abstract

Latent Variable Regression 3-Level Hierarchical Model using 2 time-point cohort longitudinal data for monitoring school performance

Kilchan Choi

This paper attempts to propose a new statistical model that is readily applied to a setting where we wish to measure school performance between two points of time with a quasi longitudinal design. The Korean National level achievement test has been implemented for last two years using this design in which a particular grade of students is repeatedly measured over years. The proposed statistical model is called latent variable regression 3 level hierarchical model. It models school's performance growth between two years explicitly controlling for initial year's status. This approach involves latent variable regression at level 2 (between cohort; within school), i.e., regressing school's growth on initial status. Through analyses of a real dataset, it is illustrated its use and implications in monitoring school performance.

Key Words: quasi-longitudinal design, National-level achievement test, growth model, value-added estimate, latent variable regression hierarchical model, school performance evaluation