세미나자료집 11-S25

2011년 한국아동·청소년패널2010 데이터분석방법론세미나 자료집

▌ 일시 : 2011년 8월 30일(화) 13:00~18:00

▌ 장소 : 한국교원단체총연합회관 2층 다산홀

▌ 주최 : 한국청소년정책연구원

국무총리산하국책연구기관
한국청소년정책연구원

「2011년 한국아동 · 청소년패널2010 데이터분석방법론 세미나」

사회 : 이경상(한국청소년정책연구원 연구위원)

시간	내용	비고
12:30～13:00	등록	
13:00～13:10	개회 및 발표자 소개	이경상(사회자)
13:10～13:20	환영인사	이재연(한국청소년정책연구원장)
13:20～17:40	발표·강의	이기종 교수(발표자)
17:40～18:00	질의응답	참석자 전원
18:00	폐회	이경상(사회자)

잠재성장모형

이기종 교수
[국민대 교육학과]

잠재성장모형

이기종 교수(국민대 교육학과)

지금까지 연구자가 구조방정식모형을 사용해 분석한 자료는 어떤 특정한 하나의 시점에서 관찰된 것이었다. 주지하는 바와 같이, 어떤 특정한 시점에서 관찰된 자료는 그 특정한 시점에서 연구자가 관심을 갖는 현상에 관한 정보를 제공한다. 이렇게 현상의 어떤 특정 시점에서 일회적으로 관찰된 자료를 횡단(橫斷) 또는 횡단면(橫斷面) 자료라고 한다. 횡단면 자료는 한 시점에서 현상을 관찰한 것이기 때문에 그 특정 시점에서 현상을 나타내는 데 유용하다. 이렇듯 특정 시점에서 현상의 단면을 관찰하는 것에 연구자가 관심을 두는 경우도 많이 있다.

횡단면 자료는 그러나 특정 시점에서 일회적으로 현상을 관찰한 것이기 때문에 시간의 흐름에 따른 현상의 변화에 관해서는 아무 것도 알려 줄 수 없다. 연구자는 때로는 시간의 흐름에 따라 현상이 어떻게 변화하는지, 달리 말해, 변화 추세 또는 양상에 관심을 갖을 수도 있다. 횡단면 자료는 특정 시점에서 한 번 관찰된 것이므로 연구자의 이런 관심을 충족시켜 줄 수 없다는 데 취약점이 있다.

이렇게 시간의 흐름에 따른 변화의 모습을 관찰하려면 현상을 일회적이 아닌 여러 번 지속적으로 관찰해야 한다. 여러 번 관찰된 자료를 종단(縱斷) 자료라고 한다. 아주 간단한 형태이지만 앞의 7장에서 나온 예제가 종단자료의 한 예이다. 동일한 피험자가 5년 시간 간격을 두고 관찰되었기 때문이다. 일반적으로 종단자료를 수집하는 방법은 패널조사와 코호트조사이다. '종단'이라는 용어는 일반적으로

패널조사와 코호트조사를 포괄하는 의미로 사용되나, 엄밀하게는 패널조사에만 종단이라는 수식어를 사용해야 한다고 주장하기도 한다 (Glenn, 1977). 용어사용에서 이런 미묘한 입장 차이가 생기는 이유는 패널조사와 코호트조사는 여러 번 관찰된다는 점에서는 동일하나 연구대상을 표본하는 방법에서 차이가 있기 때문이다.

패널조사는 현상의 횡단면을 표본하고 그 횡단면을 일반적으로 동일한 간격을 두고 시간대를 달리 해 좇아가면서 관찰한다. 달리 표현하면, 패널조사는 횡단면에 포함된 동일한 피험자를 좇아가면서 주기적으로 여러 번 반복적으로 관찰한다. 코호트는 라틴어로 함께 뜰에 있었던 사람이란 뜻으로 원래 로마군의 보병대를 가리키는 용어이다. 이 용어에서 유래되어 코호트는 특정한 시간대에 어떤 의미 있는 사건을 경험한 집단을 의미한다. 예컨대, 비행기 사고에서 살아남은 집단, 담배를 피우는 30-40세 사이의 남자, 2008년 금융위기 경험집단 등 코호트는 다양하다. 코호트조사는 특정한 시간대에 어떤 중요한 사건을 경험한 코호트로부터 여러 번 서로 다른 시점에서 매번 피험자를 표본해 관찰한다. 따라서 코호트가 서로 다른 시간대에 여러 번 반복적으로 관찰되지만, 매번 관찰되는 피험자는 코호트로부터 표본되기 때문에 동일한 피험자가 뽑혀져 나올 개연성은 매우 낮다. 즉, 서로 다른 시점에서 관찰되는 피험자는 같은 코호트의 구성원이기는 하나, 표본되기 때문에 지금의 피험자가 다음 관찰에서도 표본으로 뽑혀 피험자가 될 개연성은 매우 희박하다.

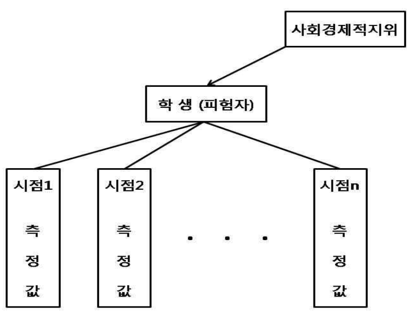
이런 이유로 엄밀하게는 동일한 피험자 집단이 여러 번 동일한 조건 또는 매번 다른 다양한 실험조건 하에서 관찰된 패널조사에만 국한해서 종단이라는 용어를 사용해야 한다는 것이다. 종단이라는 용어를 포괄적으로 사용하건 엄밀하게 사용하건 종단이라는 용어에서의 초점은, 관찰이 한 번이 아닌 여러 번이라는 데 있다. 종단자료는 피험자가 여러 번 반복적으로 관찰된다고 해서 반복측정이라고도 한다.

이런 유형의 연구설계를 종단연구 또는 반복측정 실험설계라고 한다. 여기서는 종단연구와 반복측정 실험설계를 호환적으로 사용한다.

1. 종단자료의 구성 및 제한점

반복측정이라는 또 다른 이름에서도 알 수 있듯이 종단자료는 동일한 피험자가 시간의 흐름에 따라 여러 번 관찰된 자료이다. 따라서 종단자료의 특징은 여러 시점에서 관찰된 측정값이 피험자에 소속되는 내재적 특성을 갖는다. 그러니까 종단자료는 피험자에게서 반복해서 관찰된 측정값이 피험자에 내재되는 2중구조를 갖게 된다.

이런 2중구조를 갖는 자료는 주변에서 많이 볼 수 있다. 예를 들면, 학교효과를 관찰하는 연구의 경우, 학생에게서 관찰된 측정값은 학생이 학교에 속한 관계로 학교에 내재된 2중구조이다. 비유를 하면, 종단자료에서의 각각 반복측정값은 피험자에 소속되어 피험자에 내재된 것이며, 학교효과에서의 각각 학생은 학교에 소속되어 학교에 내재된 것이다. 더 나아가 종단자료에서 피험자에 영향을 끼치는 사회경제적지위 같은 배경변수는 학교효과에서 학교에 영향을 미치는 학교소재지 같은 배경변수와 같은 것이다. 이를 그림으로 나타내면 아래 [그림 8-1]과 같다.



[그림 8-1] 2중구조 예

위 [그림8-1]에서 알 수 있듯이 반복측정값은 피험자에 내재된 2중 구조를 갖는다. 또한 피험자에 영향을 미치는 배경변수는 여럿 있을 수 있다. 2중구조보다 더 많은 중층구조를 갖는 경우도 가능하지만 해석이 쉽지 않아 여기서는 2중구조로 논의를 제한한다.

반복측정 실험설계의 주요 목적은 시간의 흐름에 따른 각 실험조건에서 종속변수의 변화 모습을 관찰하는 것이다. 그러나, 이런 목적이 충실하게 달성되기 위해서는 이월효과, 성숙, 역사 등의 내적 타당도를 저해하는 요인과 피험자 손실의 문제가 보완되어야 한다.

이월효과는 피험자 집단이 여러 번 관찰되는 데서 비롯되는 문제이다. 앞선 관찰에서의 모든 것이 그 다음 관찰로 이월되어, 그 다음 관찰에 영향을 미치는 것을 이월효과라고 한다. 동일한 피험자가 여러 번 실험조건에 노출되므로 연습이나 기억 또는 실험처치에 대한 순응이나 저항이 발생하게 된다. 예를 들면, 첫 번째 관찰에서 사용된 측정도구가 다음 관찰에서 사용되면, 첫 번째에서 문제를 풀었던 경험과 기억이 남아 있어 두 번째에서는 보다 빠른 속도로 문제를 풀 수 있다. 또한 앞서의 실험처치가 감각적으로 유쾌한 것이었다면 두 번째에서는 보다 적극적인 태도로 수용할 것이며, 이는 두 번째 관찰에서 보다 나은 결과를 빚게 된다. 반대로 불유쾌한 것이었다면, 두 번째에서는 첫 번째에서보다 더 저항을 불러일으키게 되고 이는 첫 번째 관찰에서보다 더 저조한 결과를 빚게 된다.

반복측정에서는 실험처치가 실제로는 효과 없었다고 하여도 시간이 흐름에 따라 앞에서의 관찰에 비해 그 다음 관찰에서 신체적 또는 심리적으로 더 나은 모습을 보이게 된다. 이렇게 피험자가 실험처치와 무관하게 앞서의 관찰에서 보다 후속관찰에서 더 나은 상태로 변화하는 것을 성숙(成熟)이라고 한다. 예컨대, 어린이에게 키를 크게 하는 운동을 시킨다고 해도 어린이는 성장기에 있기 때문에 따로 운동을 하지 않아도 키가 크게 된다. 따라서 어린이에게 키를 크

게 하는 운동을 시킨다 해도 어린이가 키가 크게 된 것이 실험처치인 운동효과에 의한 것인지의 여부를 판단하기 어렵다. 성숙은 독립변수의 종속변수에 대한 영향력을 희석시키는 것이어서 통제되어야 하나, 인간이 피험자인 경우 통제에 뒤따르는 윤리적 문제가 발생할 수 있다.

성숙이 피험자의 신체적·심리적 발달에 기인한 것이라면, 역사는 실험처치 중에 피험자가 경험하는 사건에서 비롯된다. 실험 도중에 피험자가 우연치 않게 경험하게 되는 일회적 또는 일련의 사건이 종속변수에 영향을 미치는 것을 역사라고 한다. 예를 들어, 재정적 문제로 야기된 우울증을 치료받는 피험자가 예기치 않게 거액의 복권에 당첨되었다면, 치료프로그램의 효과와 무관하게 우울증은 현저히 완화되게 된다. 이처럼 역사는 종속변수에서의 변화가 독립변수에 의한 것인지 아니면 예기치 않은 사건에 의한 것인지를 판단하기 어렵게 해 종속변수에 대한 독립변수의 영향력을 희석시킨다.

반복측정 실험설계는 피험자 집단 하나가 여러 번 관찰되기 때문에 독립표본의 실험설계에 비해 피험자 수가 상대적으로 적다. 예를 들어, 성별에 따른 학업성취도를 비교하는 경우 남자와 여자 두 집단이 필요하다. 그러나 반복측정의 경우 특별한 목적이 없는 한 남자집단 또는 여자집단 중 하나의 집단만 필요하다. 이렇게 상대적으로 적은 수의 피험자로도 연구가 가능하기 때문에 반복측정 실험설계를 선택하기도 하는 데, 여러 번 관찰되는 도중에 피험자가 손실된다면 적정크기 이하로 표본이 작아지기 때문에 타당한 결과를 얻기 어렵게 된다. 표본손실이 일어나는 경우는 피험자 학생이 전학을 간다거나 아니면 학교를 자퇴한다거나 하는 등 다양하다.

2. 구형성 가정

종단자료를 분석하는 하나의 방법은 각각의 실험조건이 독립변수가 되고 그 때마다 관찰된 종속변수의 실험조건별 평균차이를 비교하는 것이다. 마치 독립표본의 분산분석처럼 자료를 분석하는 것이다. 그러나 이렇게 자료를 분석하려면 구형성 가정을 충족시켜야 한다.

실험처치효과가 없다는 영가설 하에서 실험처치평균제곱에 대한 평균제곱오차의 비율로 정의되는 F비율, 다시 말해 $F = \frac{MSA}{MSE}$ 이 분자의 자유도 $\nu_1 = p-1$, 분모의 자유도 $\nu_2 = (n-1)(p-1)$ 인 F분포가 되기 위한 필요충분조건이 구형성 가정으로 다음과 같이 정의된다(Huynh & Feldt, 1970). 여기서 ν는 자유도를 나타내는 기호, p는 실험조건의 숫자, n은 실험조건의 피험자 숫자이다.

$$C' \Sigma C = \lambda I$$

위에서 C는 차수가 $(p-1) \times p$인 영가설을 나타내는 직교정규계수행렬, Σ는 차수가 $p \times p$인 모집단공분산행렬, λ는 영보다 큰 스칼라, I는 차수가 $(p-1) \times (p-1)$ 인 단위행렬이다.

직교행렬은 전위와 역행렬이 서로 같은 사각행렬--즉 어떤 행렬 A가 $A' = A^{-1}$인 행렬로 따라서 $A'A = AA' = I$가 되는--을 말한다. 이는 각 행 또는 열이 서로 90° 관계에 있으며 독립적이며 중복되지 않음을 뜻한다. 직교정규계수행렬은 직교관계를 나타내는 원소가 1인 행렬이다. 실험설계에서 서로 독립적이고 중복되지 않는 대비를 직교대비라고 부르는 것에서도 직교정규계수행렬의 의미를 유추해 볼 수 있다.

C가 직교정규계수행렬이 되어야 하는 것은 독립표본에서 분산분석의 영가설을 반영하기 때문이다. 독립적 집단의 집단평균이 서로 같다는 분산분석에서의 영가설--즉, $H_0 : \mu_1 = \mu_2 = \mu_3 \cdots = \mu_p$ --은, p개의 실험집단은 독립적이기 때문에 서로 상관이 없는 직교관계에 있으며, 집단평균이 서로 같기 때문에 표준화된 1로 표시되어도 무방함을 나타낸다. 따라서 C가 직교정규계수행렬이 된 것이다.

위의 정의에서 등호 오른쪽을 살펴보자. 단위행렬 I는 대각선 원소의 값이 1이고 그 외의 원소는 모두 영인 행렬이며, λ는 영보다 큰 어떤 값이기 때문에 단위행렬 I에 영인 아닌 어떤 값 λ을 곱해서 나온 λI는, 차수가 $(p-1) \times (p-1)$이며 대각선 원소의 값은 λ이며 그 외의 원소는 영인 행렬이다. 그러므로 구형성 가정은 독립변수인 실험처치의 하위수준인 각각의 실험조건에서 종속변수의 분산은 λ이며, 공분산은 영이 되어야 한다는 것이다. 즉, 구형성 가정은 반복측정에서 피험자가 서로 다른 실험조건 또는 동일한 실험조건에서 여러 번 관찰될 때, 각각의 조건에서 관찰된 종속변수의 분산이 같아야 한다는 것이다. 한 마디로, 독립변수인 각각의 실험조건에서 관찰된 종속변수의 분산이 같아야 한다는 것이 구형성 가정이다.

그러나 현실에서 구형성 가정은 지켜지기 힘들다. 그 이유는 동일한 피험자 집단이 여러 번 반복해서 관찰될 때마다 모든 조건이 동일하지 않기 때문에 측정오차의 문제가 발생하게 된다. 측정오차를 해결하기도 쉽지 않지만 아주 정교한 절차를 통해 측정오차 문제가 완화된다고 해도 이월효과가 발생하게 된다. 반복측정 때마다 달리 나타나게 되는 측정오차나 이월효과로 각각의 관찰에서 종속변수의 분산이 같아지기는 매우 어렵다.

3. 종단자료 분석틀

일반적 실험설계에서는 독립변수를 제외한 제3의 변수가 종속변수에 영향을 미치지 않도록 통제한 상태에서 종속변수를 관찰한다. 이렇듯이 횡단자료는 종속변수에 영향을 미칠 수 있는 제3의 변수를 통제하는 방식으로 독립변수의 종속변수에 대한 효과를 탐색하고 있다. 그러나 이렇게 하면 종속변수에 대한 독립변수의 효과가 일관되지 않거나 모호하게 해석될 여지를 많이 가지고 있다. 종속변수에 영향을 미치기는 하지만 측정되지 않은 이질적 변수가 매번 다르게 영향을 미치기 때문이다. 따라서 이러한 이질적 변수가 충분히 통제되지 않는다면 실재와는 다른 결과를 얻게 된다.

또한 이질적 변수가 통제된다 해도 종속변수에서의 변화를 어떻게 파악할 것인가의 문제가 대두된다. 횡단자료처럼 특정시점에서 현상을 관찰한 단면은 본질적으로 靜的이어서 그 시점에서 유용한 정보를 제공해준다. 그러나 반복측정된 종단자료처럼 서로 다른 여러 번의 시점에서 나타나는 動的인 움직임을 파악하는데, 靜的인 방법으로 자료를 분석하는 것은 분명 한계가 있다. 시간의 흐름에 따른 현상의 변화를 확인할 수 없기 때문에 그 결과 또한 횡단면에 대한 단편적인 적용에 그칠 수밖에 없다.

시간의 흐름에 따른 변화를 추적하기 위해서 과거에는 실험처치를 기준으로 前과 後의 변화를 추정하였다. 실험처치를 전후해서 종속변수에 어떤 차이가 있었는지를 관찰해 변화를 추정한 것이다. 그러나 이런 방식으로 변화를 추정하는 방식은 자료가 전과 후의 두 개밖에 없어서 언제나 직선으로 나타나게 된다. 예를 들어, 전의 평균이 후의 평균보다 낮으면 증가하는 직선이 될 것이며, 반대의 경우는 감소하는 직선으로 나타나게 된다. 그러나 만약 변화가 직선이 아닌 곡선궤적이라면 전후의 변화만을 포착하는 이런 방식은 최소한

의 정보만 제공하게 된다.

따라서 시간흐름에 따라 여러 번 관찰되는 반복측정에 기초해 역동적으로 움직이는 변화추이를 파악할 수 있어야 한다. 서로 다른 시점에서 여러 번 자료가 관찰된다면 변화양태를 파악할 수 있다. 게다가 더 나아가 변화양태가 正的인지 負的인지 확인할 수 있으며, 변화의 궤적이 선형인지 비선형인지를 확인할 수 있다.

누누이 언급했듯이, 독립표본처럼 종단자료를 분석하는 것은 종단자료의 본성에 어긋나는 일이다. 종단자료는 본질적으로 피험자로부터 여러 번 반복적으로 관찰된 것이어서, 측정값이 서로 독립적이지 않아 서로 상관이 있게 마련인데다 동시에 피험자에 내재된 2중구조를 갖고 있다. 이 같은 측정값의 비독립적 성질과 2중구조가 자료분석에 반영되어야 한다. 이런 비독립성이 무시되고 마치 독립적인 것처럼 자료를 분석하거나, 종단자료의 2중구조가 자료분석에 적절히 반영되지 않는다면 실재와는 다른 엉뚱한 결론에 도착하게 된다.

또한, 종단자료가 시간의 흐름에 따른 변화를 잘 반영하기 위해서는 매번 관찰되는 측정값이 공통척도와 단일차원성을 갖아야 한다. 공통척도는 여러 시점에서 관찰되는 측정값이 하나의 척도 상에서 해석 가능해야 한다는 것이다. 예컨대, 교육현장에서 피험자인 학생의 성장을 측정하는 데 주로 사용되는 것은 학업성취도이다. 그러나 매번 관찰되는 학업성취도는 서로 다른 내용을 담고 있는 때가 많다. 따라서 반복적으로 관찰되는 측정값이 하나의 척도에 위치되지 않는 한, 분석결과를 의미 있게 해석하기는 쉽지 않다. 심리측정에서 검사동등화의 여러 기법이 공통척도 문제의 해결에 일조할 수 있다. 단일차원성은 여러 시점의 측정값이 하나의 심리적 구인을 측정해야 한다는 것이다.

마지막으로 시간의 흐름에 따른 ‘변화’ 또는 ‘성장’을 어떻게 모형에 담아낼 것인가가 명료해질 필요가 있다. 변화 또는 성장이라고

할 때, 변화하고 성장하는 주체는 각각의 피험자 개인이다. 따라서 먼저 이 피험자 개개인의 개인내 변화를 설명할 수 있는 모형이 설정되어야 한다. 피험자 개개인이 출발할 때 어떤 상태에 놓여 있었는지의 초기상태와 시간의 흐름에 따라 초기상태에서 어떻게 변화하였는지의 차이에 대한 설명이 있어야 한다. 다음으로 피험자 개개인의 개인간 차이와 변화에 영향을 미치는 배경변수를 설명하는 모형이 설정되어야 한다. 피험자 개개인은 서로 동일하지 않다. 따라서 배경변수가 피험자 개개인에게 미치는 영향력도 다를뿐더러 변화의 모습도 서로 다르다. 이런 점을 설명하는 모형이 뒤따라야 한다. 한마디로, 변화 또는 성장의 개인내 그리고 개인간 차이를 설명하는 모형이 필요한 것이다.

변화 또는 성장이라는 용어는 혼용되어 사용되지만 어의에서 약간의 차이가 있다. 변화는 긍정적인 것과 부정적인 것을 모두 내포하나, 성장은 주로 긍정적인 것만을 지칭한다. 예컨대, 학생 성적이 올라가는 것도 떨어지는 것도 모두 변화이지만, 성장은 일반적으로 성적이 올라가는 것을 지칭한다. 변화라는 용어가 포괄적이어서 더 정확한 표현이나 여기서는 문맥에 따라 호환적으로 사용한다.

위와 같이 성장에 대한 궁금증을 타당하게 설명해주는 것이 잠재성장모형(Latent Growth Model, LGM)이다. 잠재성장모형에서 '잠재'라는 표현은 피험자 개개인의 초기상태와 변화율을 잠재변수로 만든다는 데서 비롯된 것이며, '성장'이란 표현은 시간의 흐름에 따른 피험자 개개인의 긍정적 변화를 나타내는 용어이며, '모형'이란 표현은 물론 종속변수에 대한 독립변수의 관계를 가리키는 용어이다. 잠재성장모형은 다층모형(multilevel model), 무선계수모형(random coefficient model), 위계선형모형(hierarchical linear model), 잠재성장곡선분석(latent growth curve analysis)이라고도 불리운다.

4. 성장을 나타내는 틀

앞서도 언급되었듯이 성장의 주체는 각각의 피험자 개인이며 이런 개개인의 성장이 서로 다른 여러 시점에서 주기적으로 반복해서 측정된다. 그 결과, 피험자 각각에서 반복되어 측정된 결과는 피험자 개개인에게 내재된다. 따라서 피험자 개개인의 성장은 2중구조를 갖는 위계모형으로 표현될 수 있다(Bryk & Raudenbush, 1992).

1) 위계모형 관점

위계모형에서 첫 번째 수준--흔히, 1수준모형이라고 하며 Level1, 또는 L1으로 표시되는--은 개인내 성장에 관한 것이다. 시간의 흐름에 따라 초기상태에서 시작해 여러 번 반복적으로 측정될 때마다, 각 개인은 변화하게 되는 데 이런 각 개인의 성장을 나타내는 것이 1수준모형이다. 달리 진술하면, 1수준모형은 피험자 개개인의 변화를 시간과 각 개인마다 고유한 성장의 함수로 표현한 것이다.

만약 개인내 성장을 나타내는 1수준모형이 직선궤적이라면, 1수준모형이 성장을 설명하기 위해서는 두 개의 정보만 있으면 된다. 하나는 피험자 개개인의 초기상태에 관한 것이고, 또 하나는 시간의 흐름에 따른 변화율에 관한 정보이다. 궤적이 직선이기 때문에 초기상태를 나타내는 절편과 변화율을 나타내는 기울기가 성장을 설명해주는 데 필요한 정보이다. 달리 말해, 절편과 기울기 두 개의 모수만 있으면 성장이 설명되는 것이다. 절편과 기울기를 더해 흔히 성장모수라고 한다. 여기서 모수란 용어는 정상분포에서 평균과 분산으로 정상분포가 설명되듯이, 절편과 기울기로 개인내 성장이 설명된다는 의미의 모수이다.

여기서 변화율이란 용어에 비율을 뜻하는 율이라는 표현이 들어

있어 변화한 비율이라는 느낌을 줄 수 있다. 그러나 변화율이 지칭하는 것은 비율이 아닌 변화궤적이다. 정확하게는 변화궤적이라는 용어를 사용하거나 또는 변화궤적의 결과물을 뜻하는 변화값, 아니면 이와 유사한 용어를 사용해야 하나, 변화율이라는 용어로 사용되는 경우가 많아 여기서도 이 용어를 그대로 사용한다.

만약 변화궤적이 직선이 아니라면 그에 대응되는 정보가 필요하게 된다. 예를 들어, 변화궤적이 곡선이면 2차항에 관한 모수가 1수준모형에 포함되어야 한다. 또한 곡선이 아니라면 3차항 또는 그 이상의 고차항이 필요하고 이에 상응하는 모수가 1수준모형에 포함되어야 한다.

위계모형에서 두 번째 수준--흔히, 2수준모형이라고 하며 Level2, 또는 L2로 표시되는--은 개인간 차이에 관한 것이다. 피험자 개개인은 시간이 흘러감에 따라 변화하지만, 그 변화의 내용은 서로 다르다. 동일한 초기상태에서 출발해도 어떤 피험자는 다른 피험자에 비해 더 빨리 또는 늦게 변화한다. 이런 이유로 인해 피험자 개인간에 차이가 생기게 된다. 달리 말해, 모든 피험자가 머리끝에서 발끝까지 서로 같지 않기 때문에 개인간에 변산이 발생하게 된다. 이 개인간 변산의 함수로 1수준모형에서 도출된 두 개의 모수를 설명한 것이 2수준모형이다.

부연하면, 피험자 개개인의 개인내 성장을 피험자마다 달라서 발생하는 개인간 차이의 함수로 표현한 것이 2수준모형이다. 개개인의 성장을 설명하는 1수준모형에서의 절편과 기울기가 각각 종속변수가 되고, 이 종속변수 각각을 개인간 차이에서 발생하는 변산이 독립변수가 되어 설명한 것이 2수준모형이다. 개인내 변화를 나타내는 절편이나 기울기에서 또는 절편과 기울기 모두에서의 차이로 인해 개인간에 변산이 발생하고, 이를 반영하는 것이 2수준모형인 것이다. 예를 들어, 같은 초기상태에서 출발했다고 하더라도 앞으로 나아가

는 진전의 과정은 피험자 개인마다 다르게 나타나 개인간 변산이 생기게 된다.

덧붙여, 어떤 영역, 예컨대 국어교과에서의 변화가 다른 영역, 예를 들면 수학교과에서의 변화에 연계되어 있다면, 개인내 성장모수는 국어와 수학교과와 같이 변하게 된다. 이 때 국어 및 수학교과와 개인내 성장모수 간의 이런 가설적 연계가 개인간 차이모형인 2수준모형에 표현된다.

지금까지의 논의를 요약하면 다음과 같다. 개인내 성장모형인 1수준모형은 반복해서 관찰된 측정값이 개인내 변산으로 회귀되는 회귀방정식과 같다. 만약 변화궤적이 직선이라면, 이 때 개인내 변산을 반영하는 절편과 기울기 두 개 모수가 발생하게 된다. 그리고 개인내 성장모수 두 개는 다시 종속변수가 되어 개인간 변산으로 회귀된다. 이렇게 하는 이유는 피험자 누구나 성장하지만 개인간 차이가 발생하기 때문이다. 이것이 개인간 차이모형이며 2수준모형인 것이다. 한 마디로, 개인내 성장에 관해 한 번의 회귀분석이 이루어지고, 이 결과가 종속변수가 되어 개인간 차이로 다시 한 번의 회귀분석이 이루어지는 것이 위계모형이다. 물론 변화궤적이 직선이 아니라면 이 부분을 반영하는 모수가 발생하게 된다.

2) 잠재성장모형 관점

시간의 흐름에 따른 피험자 개개인의 성장을 잠재변수로 만들고, 이 잠재변수 간의 관계를 밝히는 것이 잠재성장모형이다. 피험자 개개인으로부터 여러 번 반복해서 관찰된 측정값은 초기상태 및 변화율 잠재변수로 잠재화되어 측정모형을 통해 측정되고, 이 두 잠재변수 간의 관계는 이런 관계를 반영하는 잠재변수모형을 통해 밝혀지는 것이 잠재성장모형이다. 개인내 성장이 초기상태와 변화율 두 잠

재변수로 측정되고, 개인간 차이가 이 두 잠재변수 간의 관계로 나타나는 것이 잠재성장모형이다. 그러니까 한 마디로 잠재성장모형은 구조방정식모형의 틀로 성장궤적을 탐구하는 것이다. Byrk & Raudenbush(1992)의 표현처럼 구조방정식모형으로 종단자료를 분석하는 것이 잠재성장모형이다.

이 때, 초기상태 및 변화율 두 잠재변수는 피험자 개개인 간에 서로 다른 랜덤변수라고 가정된다. 잠재성장모형에서는 구조방정식모형에서 사용되는 공분산행렬 외에 평균을 분석에 사용한다. 이를 평균구조를 갖는 구조방정식모형이라고 부른다(Meredith & Tisak, 1990; McArdle, 1986; Muthen, 1991). 또한 잠재성장모형에서는 피험자 개개인의 반복측정값이 피험자에 내재된 2중구조라는 점과 그 결과 서로 독립적이지 않으며 자기상관이 있음이 반영된다. 그리고 위계모형에서처럼 개인의 성장모수가 개인간 차이를 일으키는 배경변수로 설명되는 2수준의 모형이 구사된다.

잠재성장모형은 시간에 따른 변화특성을 개인내 및 개인간 성장궤적을 추적하여 알아낸다. 또한 초기상태와 변화율에 다른 어떤 배경변수가 어떻게 영향을 미치는지도 확인할 수 있다. 예컨대, 학업성취 초기상태를 나타내는 절편과 학업성취 변화율을 나타내는 기울기에 공부시간이 어떻게 영향을 미치는지를 검증할 수 있는 것이다. 물론 공부시간 외의 여러 개의 배경변수가 동시적으로 영향을 미치는 것도 검증할 수 있다. 잠재성장모형은 이와 같이 개인내 및 개인간 성장과정을 밝혀낼 수 있어, 개인내 및 개인간 차이만을 확인할 수 있는 전통적 통계방법보다 유용하다.

시간의 흐름에 따른 여러 영역에서의 개인내 변화, 개인간 변화, 그리고 그런 변화 간의 관계(상관)--흔히, 1수준모형과 2수준모형으로 표현되는 위계모형--가 구조방정식모형으로 표현될 수 있는 것은 동일한 자료구조를 나타내는 각각의 수리적 표현이 서로 동등하기

때문이다. 개인의 성장에 대한 위계모형과 구조방정식모형은 서로 표현되는 방법은 달라도 모두 동일한 자료구조를 나타낸다. 피험자 개개인의 성장에 대한 위계모형은 여러 영역에서 여러 번 관찰된 자료의 모집단공분산행렬에 어떤 특정한 구조를 가정하는 것과 동등하다. 예를 들면, Θ_ϵ의 분산이 서로 같다던가 B가 영행렬이라던가 하는 식으로 모집단공분산행렬 Σ를 구성하는 모수행렬에 어떤 제약을 가한 것이, 여러 영역에서의 개인의 성장을 나타내는 위계모형의 수리적 표현과 동등하기 때문이다. 이렇기 때문에 평균구조를 갖는 구조방정식모형으로 성장궤적을 밝혀낼 수 있고. 이를 잠재성장모형이라고 부르는 것이다.

지금까지의 논의에서 유추할 수 있듯이, 잠재성장모형은 다음과 같은 상황에 적합하다. 첫째, 시간의 흐름에 따른 어떤 변수의 초기 상태와 변화율에 관심이 있을 때. 둘째, 초기상태 및 변화율이 배경 변수에 연계되어 배경변수의 영향력을 파악하고 싶을 때.

잠재성장모형을 사용하기 위해 필요한 요건은 다음과 같다. 첫째, 종속변수가 연속적이어야 한다. 기본적으로 자료분석에 공분산행렬이 사용되기 때문에 연속변수이어야 한다. 그러나 성공/실패 같은 비연속적 자료도 가중최소제곱법 같은 모수추정 방법을 통해 활용 가능하다(Kaplan, 2000). 둘째, 측정단위가 같아야 한다. 이는 공통척도 요건을 충족시켜 결과를 의미 있게 해석하기 위한 것이다. 셋째, 적어도 세 차례--흔히, wave라고 하는--의 측정이 있어야 한다. 성장궤적이 어떤 모습인지를 알 수 있는 최저 자료개수가 3개이므로, 적어도 세 차례 또는 그 이상의 측정이 있어야 한다. 비선형인 경우 적어도 4개 이상의 측정이 필요하다. 물론 측정시점에서 원칙적으로 모든 피험자가 동시에 측정되어야 한다. 피험자 개인 별로 서로 다른 시점에서 측정되면 보다 일반적인 틀 속에서 해결 가능하나 동시 측정이 바람직하다. 마지막으로, 시간구조가 원칙적으로 동일간격이

어야 한다. t_1과 t_2의 간격이 1년이라면 그 외가 이것보다 짧거나 길어서는 안 된다. 그러나 시간구조가 반드시 동일하지 않아도 장애가 되지 않는다.

4. 성장의 통계모형

그러면 실제 자료를 통해 성장을 수리적 모형으로 나타내보자. 예제 자료는 한국교육개발원이 수집한 한국교육종단연구-2005 중학생 패널 1~3차년도 자료이다(한국교육개발원, 2010). 한국교육종단연구 패널자료는 2005년에 시작되었으며 150개 중학교가 참여하고 있으며 표본크기는 출발 당시 6.908명이다. 1년 주기로 피험자 개개인에게서 반복적으로 측정된 값은 성별, 부모교육수준, 가계소득, 자아개념, 국어, 영어, 수학 등이다. 국어 등 학업성취도 측정값은 검사동등화가 이루어진 상태에서 관찰된 것이다.

<표 8-1> 여기로

(1) 위계모형 관점

1) 위계모형의 1수준모형

위 <표8-1>에 나타난 어떤 특정한 피험자 개인 p의 국어와 수학성적을 위계모형의 1수준모형으로 나타내보자. 개인은 i로 표시하고 사람은 p로 표기하는 것이 관행이나 개인과 사람이라는 용어가 동일한 것이기 때문에 p로 표시하고, 대신 시간을 t로 시간의 변화를 i로 표시한다. 그러면 사람(p)의 시간변화(i=1, 2, 3)에 따른 국어성적(y^r)과 수학성적(y^m)은 각각 다음과 같다. 측정값은 y로 표시되고, 국어와 수학 과목표시는 각각 위첨자 r과 m로 표시된다.

$$y_{ip}^{r} = \pi_{0p}^{r} + \pi_{1p}^{r} t_i + \epsilon_{ip}^{r} \qquad (1)$$

$$y_{ip}^{m} = \pi_{0p}^{m} + \pi_{1p}^{m} t_i + \epsilon_{ip}^{m} \qquad (2)$$

위에서 π_{0p}는 시간 t_i=0에서의 초기상태를 나타내며 수식의 절편이다. π_{1p}는 성장궤적을 나타내며 수식의 기울기이다. t_i는 모든 사람에게 동일한 시간차원으로, 예컨대 연령, 학년 등이 여기에 해당된다. ϵ_{ip}는 측정오차이다.

부연하면, 절편 π_{0p}는 시간 t_i=0에서의 사람(p)의 진점수로 출발점에서의 평균값이다. 기울기 π_{1p}는 t가 한 단위 변할 때 사람(p)의 진점수에서의 변화를 나타내며 시간의 흐름에 따른 변화율이다. 앞서도 언급했듯이, 절편과 기울기로 개인내 성장을 설명할 수 있어 이 둘이 모수가 된다.

위 식(1)과 (2)는 성장궤적이 직선인 경우를 나타내며, 만약 성장궤적이 곡선이라면 이 부분이 수리적 모형에 반영되어야 한다.

$$y_{ip} = \pi_{0p} + \pi_{1p} t_i + \pi_{2p} t_i^2 + \epsilon_{ip} \qquad (3)$$

위 식(3)에서 π_{2p}는 곡선성장궤적을 나타내는 항이다. 만약 성장궤적이 곡선이 아닌 다른 모습이라면 그에 상응되는 고차항이 모형에 반영되어야 한다. 예를 들면, 3차항이나 그 이상의 고차항이 모형에 포함될 수 있다.

지금까지의 1수준모형은 개인내 성장에 관한 것으로, 개인에게서 반복측정된 성적이 개인에 내재되어 있음이 반영되고 있다. 2수준모형에서는 식(1) 또는 (2), 그리고 (3)이 초기상태 및 성장궤적(변화율) 모수에서 개인간 차이를 예측해 줄 수 있는 배경변수를 취급할 수

있는 형태로 확장된다. 이렇게 함으로서 종단자료의 2중구조를 반영한다.

2) 위계모형의 2수준모형

위계모형의 2수준모형에서는 두 개의 모형이 설정된다. 하나는 초기상태 모수(π_{0p})에 관한 것이고, 또 하나는 성장궤적 모수(π_{1p})에 관한 것이다. 달리 말해, 만약 직선궤적이라면, 하나는 절편에 관한 것이고 또 하나는 기울기에 관한 것이다. 사람 p의 초기상태와 성장에 영향을 미치는 성별이나 사회경제적지위 같은 배경변수를 x_p라고 하면 2수준모형은 다음과 같게 된다.

$$\pi_{0p} = \mu_{\pi 0} + \gamma_{\pi 0} x_p + \zeta_{0p} \qquad (4)$$

및

$$\pi_{1p} = \mu_{\pi 1} + \gamma_{\pi 1} x_p + \zeta_{1p} \qquad (5)$$

위에서 $\mu_{\pi 0}$과 $\mu_{\pi 1}$는 $x_p = 0$일 때, 모집단에서의 상태와 성장을 나타내는 절편이며, $\gamma_{\pi 0}$와 $\gamma_{\pi 1}$는 x_p를 초기상태와 성장에 연결시키는 기울기이다.

덧붙여, 위에 설정된 모형에서 성별 같은 배경변수는 시간의 흐름과 무관한 배경변수이나, 공부에 대한 부모의 압력처럼 시간에 따라 변하는 배경변수도 가능하다. 성별은 시간이 흘러도 변하지 않지만 부모압력은 시간에 따라 변할 수 있다. 배경변수는 시간과 무관한 변수도, 시간에 따라 변할 수 있는 변수도 모두 가능하다. 따라서 학교현장을 예로 들면, 시간흐름에 따라 변하는 또는 변하지 않는 학

교 및 학생수준 특성이 초기상태와 성장에 미치는 영향이 파악될 수 있다.

지금까지의 논의를 위 <표 8-1>의 자료를 사용해 행렬로 표현해보자. 국어와 수학 두 과목에서 3번의 측정이 이루어졌음으로 각 사람은 각 영역에서 y_{1p}, y_{2p}, y_{3p} 3개의 측정자료를 갖는다. 따라서 개인 내 성장을 나타내는 1수준모형에서 식(1)과 (2)는 다음 행렬처럼 된다.

먼저 국어영역에서 각 사람 p에 대해 다음과 같은 3개의 측정값이 있다.

$$\begin{pmatrix} y^r_{1p} \\ y^r_{2p} \\ y^r_{3p} \end{pmatrix} = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \end{pmatrix} \begin{pmatrix} \pi^r_{0p} \\ \pi^r_{1p} \end{pmatrix} + \begin{pmatrix} \epsilon^r_{1p} \\ \epsilon^r_{2p} \\ \epsilon^r_{3p} \end{pmatrix}$$

다음으로 수학영역에서 각 사람 p에 대해 다음과 같은 3개의 측정값이 있다.

$$\begin{pmatrix} y^m_{1p} \\ y^m_{2p} \\ y^m_{3p} \end{pmatrix} = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \end{pmatrix} \begin{pmatrix} \pi^m_{0p} \\ \pi^m_{1p} \end{pmatrix} + \begin{pmatrix} \epsilon^m_{1p} \\ \epsilon^m_{2p} \\ \epsilon^m_{3p} \end{pmatrix}$$

여기서 국어와 수학영역을 따로 따로 하면 단일변수인 경우가 되며, 함께 붙이면 다중변수의 경우가 된다. 만약 수직연결을 통해 두 영역을 함께 붙여 하나로 만들면 다음처럼 된다. 수직연결이란 이미 있는 두 행렬을 위아래로 그대로 갖다 붙여 하나로 만든 것이다.

$$
\begin{bmatrix} y^r_{1p} \\ y^r_{2p} \\ y^r_{3p} \\ y^m_{1p} \\ y^m_{2p} \\ y^m_{3p} \end{bmatrix}
=
\begin{bmatrix} 1 & t_1 & 0 & 0 \\ 1 & t_2 & 0 & 0 \\ 1 & t_3 & 0 & 0 \\ 0 & 0 & 1 & t_1 \\ 0 & 0 & 1 & t_2 \\ 0 & 0 & 1 & t_3 \end{bmatrix}
\begin{bmatrix} \pi^r_{0p} \\ \pi^r_{1p} \\ \pi^m_{0p} \\ \pi^m_{1p} \end{bmatrix}
+
\begin{bmatrix} \epsilon^r_{1p} \\ \epsilon^r_{2p} \\ \epsilon^r_{3p} \\ \epsilon^m_{1p} \\ \epsilon^m_{2p} \\ \epsilon^m_{3p} \end{bmatrix}
$$

사람 p의 국어와 수학영역에서 반복적으로 관찰된 측정값이 한 데 합쳐진 위 행렬은 3개의 서로 구분되는 요소로 되어있다. 상수 1과 모든 사람에게 동일한 측정회수를 나타내는 부분인 t가, 사람마다 다른 미지의 개인내 성장에 관한 잠재모수벡터에 의해 뒤에서 곱해지고, 여기에 사람마다 다른 미지의 측정오차벡터가 더해진 것이다.

고전검사이론의 신뢰도에서 거론되는 신호와 잡음을 통해 위 행렬을 들여다보자. 주지하는 바와 같이, 송수신자가 서로 주고받은 정보의 총량 가운데 신호가 차지하는 부분이 커질수록 송수신이 정확하게 이루어지는 것이며, 반대로 잡음 부분이 커질수록 송수신은 불명료하게 된다. 위 행렬에서 잠재성장벡터는 시간의 흐름에 따른 진정한 변화를 나타내는 개인내 신호와 같은 것이며, 오차벡터는 변화의 측정을 방해하는 개인내 잡음과 같은 것이다.

식(1)과 (2)에서 방해오차 ϵ_{ip}를 설정했으나 ϵ_{ip}의 분포에 관해서는 논의되지 않았다. 어쩌면 ϵ_{ip}는 동분산이며 독립적일 수도 있으며, 아니면 이분산이거나 자기상관을 갖을 수도 있다. 오차 ϵ_{ip}에 관한 일반적 가정은 ϵ_{ip}가 동분산이며 서로 독립적이라는 것이다. 이 때 오차벡터는 다음과 같다.

$$\left|\begin{matrix} \epsilon_{1p}^{r} \\ \epsilon_{2p}^{r} \\ \epsilon_{3p}^{r} \\ \epsilon_{1p}^{m} \\ \epsilon_{2p}^{m} \\ \epsilon_{3p}^{m} \end{matrix}\right| \approx \mathrm{N}\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_{\epsilon^r}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{\epsilon^r}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{\epsilon^r}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{\epsilon^m}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{\epsilon^m}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{\epsilon^m}^2 \end{bmatrix}$$

오차벡터는 정상분포를 이루며, 이 때 오차벡터의 평균벡터는 영벡터($\mu = 0$), 공분산행렬은 대각선행렬($\Sigma = D$)이다. 즉, 오차의 평균은 0이고 공분산은 없고 분산만 있는 행렬이다. 원래는 $\mathrm{N}(\mu, \Sigma)$로 표기되어야 하나, 괄호를 만들 수 없어 위와 같이 표기되었다.

위 행렬에서 ϵ_{ip}는 한 영역에서는 동분산이나 다른 영역에서까지 동분산임을 가정할 수는 없다. 즉, 국어영역에서는 동분산이지만 수학영역에서도 동분산이라고 가정할 수는 없는 것이다. 그 이유는 오차분산 σ_{ϵ}^2이 두 영역에서 동일하다는 어떤 선험적 증거가 없기 때문이다. 만약 이런 것이 가정되면 국어와 수학영역에서의 오차분산 σ_{ϵ}^2을 같게 하고 이를 검증하면 된다.

모든 모집단구성원이 한 영역에서 변화에 관해 동일한 함수형태를 갖는다고 해도, 성장모수에서 개인간에 변산이 존재하기 때문에 성장궤적은 사람에 따라 여전히 다를 수 있다. 더욱이 한 영역과 또 다른 영역에서 개인내 성장모수간에 공변산이 존재하기 때문에, 개인내 변화는 모든 영역을 가로질러 일어나게 된다.

따라서 영역을 가로지르는 경우, 개인내 성장모수벡터의 개인간 분포에 관한 설명이 필요하게 된다. 개인내 성장모수벡터의 개인간 분포는 다음과 같다. 아래에서도 원래는 $\mathrm{N}(\mu, \Sigma)$로 표기되어야 하나 바깥괄호를 만들 수 없어 아래처럼 표기한다.

$$\begin{bmatrix} \pi_{0p}^{r} \\ \pi_{1p}^{r} \\ \pi_{0p}^{m} \\ \pi_{1p}^{m} \end{bmatrix} \approx \mathrm{N}\begin{bmatrix} \mu_{\pi 0}^{r} \\ \mu_{\pi 1}^{r} \\ \mu_{\pi 0}^{m} \\ \mu_{\pi 1}^{m} \end{bmatrix}, \begin{bmatrix} \sigma_{\pi_0^r}^2 & \sigma_{\pi_0^r \pi_1^r} & \sigma_{\pi_0^r \pi_0^m} & \sigma_{\pi_0^r \pi_1^m} \\ \sigma_{\pi_1^r \pi_0^r} & \sigma_{\pi_1^r}^2 & \sigma_{\pi_1^r \pi_0^m} & \sigma_{\pi_1^r \pi_1^m} \\ \sigma_{\pi_0^m \pi_0^r} & \sigma_{\pi_0^m \pi_1^r} & \sigma_{\pi_0^m}^2 & \sigma_{\pi_0^m \pi_1^m} \\ \sigma_{\pi_1^m \pi_0^r} & \sigma_{\pi_1^m \pi_1^r} & \sigma_{\pi_1^m \pi_0^m} & \sigma_{\pi_1^m}^2 \end{bmatrix}$$

위에서 보는 것처럼, 개인내 성장모수벡터는 평균벡터 μ와 공분산행렬 Σ를 갖는 정상분포이다. 평균벡터 μ는 국어와 수학영역에서 절편과 기울기 평균을 원소로 갖는 벡터이며, 공분산행렬 Σ는 대각선을 중심으로 절편과 기울기 간의 공분산을 원소로 하는 대칭행렬이다.

위에서 가정된 분포가 변화에서 개인간 차이를 나타내는 2수준모형이다. 위 모형은 모두 14개의 개인간 모수로 되어 있다. 구체적으로, $\mu_{\pi 0}^{r}$부터 $\mu_{\pi 1}^{m}$까지의 4개 모집단평균, $\sigma_{\pi_0^r}^2$에서부터 $\sigma_{\pi_1^m}^2$까지 잠재성장벡터의 4개 분산, 그리고 $\sigma_{\pi_0^r \pi_1^r}^2$부터 $\sigma_{\pi_0^m \pi_1^m}^2$까지 잠재성장벡터의 6개 공분산이다. 이것은 각각 각 영역에서 변화의 평균궤적, 각 영역에서 절편 및 기울기의 분산과 공분산, 그리고 절편 및 기울기의 영역간의 공분산이다. 이 14개 2수준모형 모수로 있을 수 있는 모든 문제에 답할 수 있다.

(2) 구조방정식모형 관점

1) 개인내 성장모형

구조방정식모형에서 개인내 성장모형은 측정모형으로 나타내어진다. 한 개인에서의 성장을 나타내는 초기상태와 변화율이 잠재변수가 되고, 주기적으로 관찰된 반복측정값이 이 두 잠재변수의 측정변수가 되는 측정모형이 구조방정식모형에서의 개인내 성장모형이다.

위 표<8−1>의 국어와 수학 두 영역에서 사람 p의 반복측정값은 다음처럼 쓸 수 있다.

$$\begin{vmatrix} y^r_{1p} \\ y^r_{2p} \\ y^r_{3p} \\ y^m_{1p} \\ y^m_{2p} \\ y^m_{3p} \end{vmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & t_1 & 0 & 0 \\ 1 & t_2 & 0 & 0 \\ 1 & t_3 & 0 & 0 \\ 0 & 0 & 1 & t_1 \\ 0 & 0 & 1 & t_2 \\ 0 & 0 & 1 & t_3 \end{bmatrix} \begin{vmatrix} \pi^r_{0p} \\ \pi^r_{1p} \\ \pi^m_{0p} \\ \pi^m_{1p} \end{vmatrix} + \begin{vmatrix} \epsilon^r_{1p} \\ \epsilon^r_{2p} \\ \epsilon^r_{3p} \\ \epsilon^m_{1p} \\ \epsilon^m_{2p} \\ \epsilon^m_{3p} \end{vmatrix}$$

위 행렬은 어딘가 낯이 익은 느낌이 들 것이다. 그 이유는, 앞의 위계모형에서 제시된 사람 p의 관찰자료 앞에 영벡터가 더해진 것이기 때문이다. 이 행렬은 구조방정식모형의 측정모형과 동일하다. 5장의 측정모형의 앞에 영벡터가 더해진 이 행렬이 측정모형과 같은 것은 더해진 것이 영벡터이기 때문이다. 모든 원소가 영으로 고정된 영벡터가 더해져도 영을 더한 것이기 때문에 아무 변동이 없는 것과 같은 이치이다. 위 행렬을 일반적 형태로 표시하면 아래와 같다.

$$y = \tau_y + \lambda_y \eta + \epsilon \qquad (6)$$

여기서 y는 사람 p의 측정자료를 담고 있는 벡터, τ_y는 원소가 0으로 고정된 영벡터, λ_y는 1로 된 행과 시간을 나타내는 상수로 된 행을 갖는 고정행렬, η는 초기상태(π_{0p})와 변화율(π_{1p}) 모수를 담고 있는 잠재변수벡터, ϵ는 측정오차를 담고 있는 벡터이다.

위 식(6)은 행렬로 표시되었지만 회귀방정식과 같은 꼴로 되어 있는 것을 알 수 있다. 따라서 τ_y와 λ_y는 각각 회귀방정식의 절편과 기울기에 해당된다. 즉, τ_y는 절편벡터이며 λ_y는 기울기행렬이다. 예를

들어, 첫 번째 측정시점이 중1이면 λ_y의 시간을 나타내는 상수는 0, 1, 2가 된다.

여기서 τ_y와 λ_y의 원소가 일반적인 경우에서의 구조방정식모형과는 달리, τ_y는 알고 있는 값인 0으로 λ_y는 상수 1과 시간을 나타내는 상수 t로 구성되도록 제약을 부과한 것에 주목할 필요가 있다. 이렇게 제약을 부과함으로서 위계모형과 구조방정식모형은 서로 표현되는 형태는 달라도 동일한 자료구조를 나타내는 동등한 모형이 된다.

또한 눈여겨 볼 것은, 초기상태(π_{0p})와 변화율(π_{1p}) 모수가 잠재변수벡터 η에 흡수되어 있다는 점이다. 그렇기 때문에 이 모형을 잠재(변수)성장모형이라고 부른다. 때로는 η는 잠재성장벡터라고도 불리운다. 여기서는 두 용어를 호환적으로 사용한다.

덧붙여, 측정오차 간의 공분산 $cov(\epsilon,\epsilon)$은 대각선행렬임을 가정한다. $cov(\epsilon,\epsilon)$가 대각선행렬이면 아래와 같다.

$$\Theta_\epsilon = cov(\epsilon,\epsilon) = \begin{bmatrix} \sigma^2_{\epsilon^r} & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{\epsilon^r} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{\epsilon^r} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{\epsilon^m} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{\epsilon^m} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{\epsilon^m} \end{bmatrix}$$

측정오차 공분산행렬 Θ_ϵ에 여러 형태의 제약을 부과할 수 있다. 위 가정은 자기상관이 없다는 제약을 부과한 것과 같다. 대각선 이외의 원소를 0으로 고정하지 않고 자유모수로 설정하면 자기상관을 허용하는 것이며 이를 검증하면 된다.

또한 $\sigma^2_{\epsilon^r}$에서부터 $\sigma^2_{\epsilon^m}$까지 오차분산이 동분산이라는 제약을 부과할

수 있다. 즉, 대각선 원소가 같다는 등가제약을 부과하는 것이다. 만약 대각선원소가 같다는 등가제약을 부과하지 않으면 이분산을 허용하는 것이 된다. 즉 측정오차가 동분산이 아닌 이분산이라고 판단되면 대각선에 등가제약을 안 주고 이를 검증하면 된다.

2) 개인간 차이모형

앞서의 개인내 성장이 측정모형으로 표현된 것이라면, 개인간 차이모형은 잠재변수모형으로 표현된다. 앞서 측정모형에서 계수행렬 λ_y가 상수 1과 시간을 나타내는 상수 t로 된 고정행렬로 설정된 바 있다. 계수행렬 λ_y가 고정행렬로 된 것은 1수준모형의 개인성장모수 $(\pi_{0p}^r \ \ \pi_{1p}^r \ \ \pi_{0p}^m \ \ \pi_{1p}^m)$를 구조방정식모형의 내생 잠재변수벡터 η로 옮기기 위한 것이다. 그렇게 설정함으로서 η가 관심대상인 개인간 수준의 모수를 포함하게 만들고, 이런 2수준모형 분석이 구조방정식모형의 잠재변수모형에서 이루어지게 하기 위한 것이다.

이렇게 하기 위해서 필요한 것은 단지 잠재성장벡터 η를 아래처럼 다시 쓰는 것이다.

$$\begin{bmatrix} \pi_{0p}^r \\ \pi_{1p}^r \\ \pi_{0p}^m \\ \pi_{1p}^m \end{bmatrix} = \begin{bmatrix} \mu_{\pi_0}^r \\ \mu_{\pi_1}^r \\ \mu_{\pi_0}^m \\ \mu_{\pi_1}^m \end{bmatrix} + \begin{bmatrix} 0\,0\,0\,0 \\ 0\,0\,0\,0 \\ 0\,0\,0\,0 \\ 0\,0\,0\,0 \end{bmatrix} \begin{bmatrix} \pi_{0p}^r \\ \pi_{1p}^r \\ \pi_{0p}^m \\ \pi_{1p}^m \end{bmatrix} + \begin{bmatrix} \pi_{0p}^r - \mu_{\pi_0}^r \\ \pi_{1p}^r - \mu_{\pi_1}^r \\ \pi_{0p}^m - \mu_{\pi_0}^m \\ \pi_{1p}^m - \mu_{\pi_1}^m \end{bmatrix}$$

위 행렬을 일반적 형태로 다시 쓰면 아래와 같다.

$$\eta = \alpha + B\eta + \zeta \qquad (7)$$

여기서 α는 초기상태$(\mu_{\pi 0})$ 및 변화율$(\mu_{\pi 1})$ 모수를 담고 있는 벡터,

B는 영행렬, ζ는 피험자 개인의 성장모수에서 각각의 모집단평균을 뺀 편차벡터이다. 즉, ζ는 η에서 α를 뺀 방해오차벡터로, $\zeta = (\eta - \alpha)$이다.

위 식(7)은 구조방정식모형에서 등호 오른쪽의 $B\eta$항을 왼쪽으로 옮기기 전의 잠재변수모형과 같은 꼴이다. 즉, 등호 왼쪽이 $\eta - B\eta$가 되어 $(I-B)\eta$로 환원되기 전의 형태와 같은 것이다. 위 식(7)의 특징은 α로 절편과 기울기를 옮긴 것이다. 이렇게 함으로써 α의 원소인 절편과 기울기가 추정될 수 있도록 한 것이다.

$(\eta - \alpha)$인 방해오차벡터 ζ는 영평균벡터 0, 공분산행렬 Ψ를 갖는다. Ψ는 초기상태와 변화율의 분산과 공분산을 담고 있다.

$$\Psi = cov(\zeta,\zeta) = \begin{bmatrix} \sigma^2_{\pi^r_0} & \sigma_{\pi^r_0\pi^r_1} & \sigma_{\pi^r_0\pi^m_0} & \sigma_{\pi^r_0\pi^m_1} \\ \sigma_{\pi^r_1\pi^r_0} & \sigma^2_{\pi^r_1} & \sigma_{\pi^r_1\pi^m_0} & \sigma_{\pi^r_1\pi^m_1} \\ \sigma_{\pi^m_0\pi^r_0} & \sigma_{\pi^m_0\pi^r_1} & \sigma^2_{\pi^m_0} & \sigma_{\pi^m_0\pi^m_1} \\ \sigma_{\pi^m_1\pi^r_0} & \sigma_{\pi^m_1\pi^r_1} & \sigma_{\pi^m_1\pi^m_0} & \sigma^2_{\pi^m_1} \end{bmatrix}$$

이 방해오차 공분산행렬 Ψ가 성장에서 개인간 차이를 설명해 줄 수 있는 많은 정보를 갖고 있다. 예를 들어, 국어영역에서 초기상태와 변화율 간에 상관이 있는가에 관한 정보는 $\sigma_{\pi^r_0\pi^r_1}$이 갖고 있으며, 국어와 수학영역의 초기상태는 서로 상관이 있는지에 관한 답변은 $\sigma_{\pi^r_0\pi^m_0}$이 해 줄 수 있다. 모든 경우를 예로 들지 않았지만, 한 영역에서 또는 영역을 가로지르는 여러 가지 상관조합에 관한 정보가 바로 방해오차 간의 공분산행렬 Ψ에 담겨 있다. 어쩌면 이미 알고 있겠지만, 이 공분산행렬 Ψ의 기능은 위계모형의 개인내 성장모수벡터의 개인간 분포에서의 공분산행렬과 같은 것이다. 따라서 앞의 위계모형에서 언급했듯이 종단자료에서 있을 수 있는 모든 궁금증에 대한

답이 바로 Ψ에 있다.

주목해야 할 것은 위에서 식(7)은 성장·변화에 영향을 미치는 배경변수를 포함하고 있지 않다. 그러나 피험자 개인에 영향을 미치는 배경변수는 많이 있다. 예를 들면, 성별, 학교문화나 부모의 학업에 대한 압력 등이 그것이다. 배경변수에는 성별처럼 시간이 지나도 변하지 않은 것도 있고, 부모의 학업에 대한 압력처럼 시간에 따라 변하는 것도 있다.

시간에 따라 변하는 것이건 또는 아니건 간에 배경변수는 피험자 개인에게 영향을 미친다. 따라서 피험자 개인의 성장·변화에서 배경변수가 고려되어야 한다. 필요하다면 피험자 개인의 초기상태와 변화율에 영향을 미치는 배경변수가 잠재성장모형에 포함되어야 한다. 이런 경우 측정모형은 아래와 같다.

$$x = \tau_x + \lambda_x \xi + \delta \quad (8)$$

위에서 x는 측정된 배경변수를 담고 있는 벡터, τ_x는 평균벡터, λ_x는 단위행렬, ξ는 편차점수로 된 배경변수벡터, δ는 영벡터이다. 여기서 δ가 영벡터라는 사실은 배경변수 x가 측정오차 없이 관찰된다는 것이고, 이점에서 x는 회귀분석에서의 독립변수와 같다. 그리고 λ_x가 단위행렬이라는 것은 x가 ξ에 1:1 대응으로 관찰된다는 뜻이다.

덧붙여 ξ가 편차점수로 되어 있다는 것은 배경변수를 중심화(centering)하는 효과가 있다. 중심화란 한 마디로 수집한 자료에서 어떤 값을 빼는 것을 말한다. 그러니까 원점수를 편차점수로 변환하는 것이 중심화이다. 원점수를 편차점수로 변환할 때, 빼는 값으로 평균이 사용될 수도 있고 그 외의 다른 값이 사용될 수도 있다. 주로 독립변수에 적용되는 중심화는 그렇게 해서 그 변수의 척도를 옮기고자 하는 것이다. 평균이 아닌 값을 사용한 중심화는 편차점수를

구한다는 점에서는 같으나 빼는 값이 달라 편차점수의 성질이 다르다.

위의 식(1)--$y_{ip}^{r} = \pi_{0p}^{r} + \pi_{1p}^{r} t_i + \epsilon_{ip}^{r}$--이나 식(2)에서, 절편의 의미 있는 해석을 위해 시간변수를 편차점수로 변환하는 경우를 생각해 보자. 위 식에서 절편 π_{0p}는 시간 t_i=0에서의 초기상태를 나타내는데, 이 절편이 진정한 의미의 초기상태를 나타내게 하기 위한 것이 중심화이다.

예를 들어, 피험자 개인에게서 반복적으로 측정된 값이 5년 주기로 5년, 10년, 15년, 20년에 있었다고 하자. 이 측정값이 편차점수화되지 않고 수식에 사용되면, π_{0p}는 초기상태인 5년에서의 상태가 아니라 t_i가 영이 되는 0년에서의 상태를 나타낸다. 그러나 0년에서의 상태는 현실적으로 측정할 수 없는 의미 없는 값이다. 따라서 잠재성장모형에서는 처음 측정시점으로 모든 시간변수를 중심화해서 π_{0p}가 진정한 초기상태를 나타내게 한다. 즉, 시간변수 5년, 10년, 15년, 20년에서 처음 측정시점 5년을 빼서 시간변수를 0년, 5년, 10년, 15년으로 만드는 것이 중심화이다. 이렇게 함으로서 π_{0p}가 진정한 초기상태를 나타내게 하고, π_{1p}가 5년에서부터 20년까지의 변화율을 나타내게 하는 것이다.

또한, 연구자가 출발시점에서 관심을 두지 않고 다른 시점에 관심을 갖는다면 그 시점을 기준으로 중심화할 수도 있다. 예컨대, 10년에 관심을 갖는다면 −5년, 0년, 5년, 10년으로 중심화된다.

구조방정식모형으로 성장을 접근할 때의 장점은, 오차에 관한 유연성이 허락된다는 것과 오차가 독립적이고 동분산성이 있어야 한다는 일반적 가정이 완화될 수 있다는 것이다. 또한 일반적 반복측정 설계에서 지켜져야 하는 구형성 가정도 잠재성장모형에서는 걸림돌이 되지 않는다(Hamagami & McArdle, 2001; Heck, 2001; Hess, 2000; Kaplan, 2000; Willett & Sayer, 1996; 곽수란·이기종, 2009 재인용).

<3개년 국어 수학 공분산행렬 여기로>

<이하에 예제자료 3과 4, 그 다음 5와 6을 붙임>

2. 모형공분산행렬

잠재성장모형에서의 모수 확인

3. 모수추정

4. 모수평가

예제

직선이 경우
곡선인 경우
자유추정인 경우