조사연구(2016. 2) 17권 1호: 31-53

특집논문

구조방정식모형의 통계적 쟁점

Statistical Issues in Structural Equation Models

김규성[a)]
Kyu-Seong Kim

구조방정식모형은 계량 데이터를 다루는 사회과학 분야에서 널리 쓰이고 있다. 이 모형은 잠재변수를 포함하고 변수들 간의 관계를 명시적인 인과관계로 설명할 수 있어 적용될 수 있는 범위가 매우 넓다. 또한 연구자가 공분산 데이터 행렬만으로 손쉽게 분석할 수 있기 때문에 구조방정식모형의 활용은 점점 늘어나는 추세이다. 그러나 적은 자유도를 갖는 데이터로 넓은 적용범위에서 많은 결과물을 산출하기 때문에, 그 이면에는 많은 모형 가정과 확률적 의사결정 과정, 그리고 결과의 해석에 대한 난점들이 숨어 있다.

본 논문에서는 구조방정식모형의 통계적 쟁점을 데이터 생성, 변수의 척도, 인과관계, 잠재변수 식별, 모형 식별 및 모수 추정, 모형 적합 및 검정의 6가지로 나누어 논의하였다. 주로 통계적인 관점에서 구조방정식모형의 구조 및 기본 개념에 초점을 맞춘 쟁점들이다.

주제어: 모형적합, 인과관계, 잠재변수

Structural equation models have been used widespreadly in the field of quantitative social science. These models include latent variables and can explain the causality among variables explicitly,

a) 서울시립대학교 통계학과 교수 김규성.
E-mail: kskim@uos.ac.kr

so could be applied to many fields widely. Also researchers are easily able to analyze the models on the basis of covariance data only. These lead structural equation models to rapid increase of usage. However, there are some difficulties such as lots of model assumptions, stochastic decision process and hard interpretation of the results due to lack of degree of freedom for covariance data in order to explain a somewhat complicated model.

In the paper, I discussed six statistical issues of structural equation models such as data generation, scale of variables, causality, identification of latent variables, model identification and model parameter estimation, and model fit and test. These are some issues mainly focused on structure and basic concepts of structural equation models in the statistical viewpoint.

Key words: casuality, latent variables, model fit

Ⅰ. 서론

구조방정식모형(structural equation model)은 계량 사회과학 분야에서 사용되는 가장 인기 있는 모형 중의 하나이다. 구조방정식모형이 사회과학 분야의 많은 연구자들에게 인기 있는 모형으로 자리잡은 배경에는 이 모형이 발전되어 온 이력과도 연관이 있다. 구조방정식모형은 단일한 모형이 아니라 여러 형태의 모형을 포함하는 모형의 범주(class)로 정의될 수 있는데, 이 범주에는 요인분석(factor analysis) 모형, 경로분석(path analysis) 모형, 연립방정식 모형(simultaneous equation modeling) 등이 포함된다. 경로분석은 20세기 초에 유전학 분야에서 개발되어(Wright 1918, 1934 등) 사회과학 분야로 발전해 갔고, 요인분석은 심리학 혹은 심리측정학 (psychometrics) 분야에서 지능의 구

조를 탐구하는 방법론으로 개발된(Spearman 1904) 후 다른 분야로 발전해 나갔다. 또한 연립방정식 모형론은 주로 계량경제학 분야에서 수요, 공급의 문제를 다루다가 개발되고 발전된 이론이다. 각각의 분야에서 필요에 의하여 개발되고 발전해 오던 개별 이론들이 1970년대에 들어 요즘 유행하는 말로 학제간 접근법(multi-disciplinary approach)에 의하여 구조방정식모형 방법론의 틀 안에서 하나의 융합방법론으로 자리를 잡게 된다(Jöreskog 1973 등). 각 분야에서 개발되던 모형을 구조(structure) 부분과 관측(measurement) 부분으로 구분한 후 두 부분이 통합된 일반이론으로 발전된 것이다. 1990년대를 지나면서 구조방정식모형을 활용하는 연구 분야는 점차 확대되어 종래의 유전학, 심리학, 경제학의 범위를 넘어 행정학, 경영학, 교육학, 사회학 등에서도 구조방정식모형 관련 연구결과가 발표되고 있다(예를 들어 국내 연구결과로는, 심준섭 2015; 정효채 외 2013; 오현복 외 2015; 김민석·민진 2014 등).

구조방정식모형이 널리 사용하게 된 현실적인 이유 중의 하나는 구동 가능한 소프트웨어가 개발된 것이다. 즉 구조방정식모형을 사용한 데이터 분석을 가능하게 하는 컴퓨터 프로그램이 속속 개발되어 연구자들이 손쉽게 데이터를 분석할 수 있게 된 것이다. LISREL(LInear Structural RELations), AMOS (Analysis of MOment Structure), EQS(EQustionS) 등은 구조방정식모형 분석을 위한 전용 소프트웨어이고, 범용 통계소프트웨어인 SAS, SPSS 등에는 구조방정식모형 분석을 위한 모듈이 내장되어 있다. 최근에는 공개 프로그램인 R에도 구조방정식모형 관련 모듈이 개발되어 사용자들에게 제공되고 있는 중이다.

구조방정식모형의 활용 범위가 넓어지게 된 데는 이유가 있다. 연구자가 탐구하고 싶어하는 내용에 대해 확률 데이터에 기초한 통상적인 통계모형은 그 내용을 충분히 제공해 주지 않는다. 그러나 구조방정식모형은 그러한 연구자들이 궁금해 하는 내용을 명시적으로 명료하게 제공해 주기 때문이다. 예를 들어 회귀모형, 분산분석모형, 패널모형 등을 다루는 통상적인 통계분석에서는 잠재변수, 상관성이 있는 오차항 및 인과관계(causality) 등을 확률 데이터에 근거하여 명시적인 수치로 설명하는 것은 매우 어렵다. 따라서 잠재변수보다

는 관측변수, 상관성 있는 오차항보다는 서로 독립적인 오차항, 인과관계보다는 연관성(association)을 다루는 것이 보통이다. 반면 구조방정식모형은 연구자가 알고자 하는 잠재변수, 인과관계, 상관성 있는 오차항 등을 명시적으로 다루기 때문에 사회과학 분야의 연구자들에게 인기가 있는 것이다.

구조방정식모형 관련 연구에서 나타나는 특이한 현상은 구조방정식모형에 대한 적절성, 한계, 적용방법, 메타분석, 장단점 등을 설명하는 연구결과가 많고, 또한 지속적으로 계속 발표되고 있다는 점이다(예를 들면, Hermida 2015; Iacobucci 2009; Bentler 2010; Fabrigar et al. 2010; Nachtigall et al. 2003; Bollen & Pearl 2013; Freedman 1987; Kline 2011; Chin 1998 등). 그 이유는 모형의 적용성, 모형에 부여되는 강한 가정, 그리고 분석결과의 해석 등과 연관이 있다. 구조방정식모형의 적용 범위가 넓기 때문에 개별 연구 분야에 맞는 적절한 모형을 선택해야 하고, 모형에 부여된 강한 가정들이 개별 연구자의 상황에 부합하는지를 검토해야 하며, 통계 소프트웨어가 제공하는 수많은 출력물들이 무엇을 의미하는지를 연구자가 올바로 이해해야 할 필요가 있기 때문에 이와 관련된 연구결과가 추가적으로 발표되는 것이다.

구조방정식모형과 관련한 연구결과 및 연구결과에 대한 문헌 연구, 그리고 구조방정식모형을 활용한 사례연구 결과는 무수히 많다. 또한 구조방정식모형 활용에 관한 지침서(guideline)나 소프트웨어 매뉴얼도 이미 많이 공개되어 있다. 본 본문에서 다루려고 하는 것은 구조방정식모형에 관한 새로운 이론이나 활용사례, 혹은 활용지침 등이 아니다. 대신 구조방정식모형에 본질적으로 내재하고 있는 통계적 쟁점들에 대하여 논의하고자 한다. 구조방정식모형을 활용하여 연구를 수행하는 연구자들이 구조방정식모형의 데이터 생성, 모형 식별, 추정, 검정 등에서의 통계적 쟁점들을 분명하게 인식하고 대처하지 않으면, 이는 고스란히 구조방정식모형 분석을 하는 연구자에게 불확실성으로 남고 더 나아가 구조방정식모형 분석의 오·남용을 초래하기 때문이다. 본 논문에서는 비록 부분적이긴 하지만 이러한 통계적 쟁점들을 데이터 생성, 변수의 척도, 인과관계, 잠재변수 식별, 모형 식별 및 모수 추정 그리고 모형 적합 및 검정의 6가지 소주제로 구분하여 쟁점들을 논의한다.

제2절에서는 구조방정식모형 이론에 대한 개괄적인 검토를 하고, 제3절에서는 앞 절에서 검토한 내용을 바탕으로 6가지 통계적 쟁점 사항을 논의한다. 마지막으로 제4절에서는 간단한 요약과 함께 향후 연구방향을 언급한다.

Ⅱ. 구조방정식모형 이론

1. 모형 정의

일반적으로 구조방정식모형은 구조(structure) 부분과 측정(measurement) 부분으로 구성된다. 구조 부분에서는 잠재변수(latent variable)들의 관계가 연립방정식을 통하여 정의되고, 측정 부분에서는 관측변수(observable variable)와 잠재변수 간의 관계가 확인적 요인모형(confirmatory factor model)을 통하여 정의된다.

구조 부분에서 $(m\times1)$차원의 잠재변수 벡터 η와 $(n\times1)$차원의 잠재변수 벡터 ξ는 다음의 관계가 있는 것으로 가정한다.

$$\eta = B\eta + \Gamma\xi + \zeta \tag{1}$$

여기에서 벡터 η는 내생적(endogenous) 혹은 종속적 잠재변수 벡터이고 벡터 ξ는 외생적(exogenous) 혹은 독립적 잠재변수 벡터이며, 행렬 $B_{m\times m}$과 $\Gamma_{m\times n}$은 회귀계수 행렬, ζ는 $(m\times1)$차원의 오차항(error term) 벡터이다. 행렬 B의 원소는 벡터 η의 각 성분들 간의 관계를 나타내고, 행렬 Γ의 원소는 잠재변수 벡터 ξ가 잠재변수 벡터 η에 대한 영향을 나타낸다. 여기에서 오차항 벡터 ζ는 잠재변수 벡터 ξ와 상관관계가 없는 것으로 가정하고, 행렬 $(I-B)$은 정칙(non-singular)인 것으로 가정한다. 벡터 ξ와 ζ는 모두 확률벡터로서 평균이 0이고 공분산행렬이 각각 $Cov(\xi)=\Phi(n\times n)$, $Cov(\zeta)=\Psi(m\times m)$인 것으로 가정한다. 이러한 가정으로부터 내생 잠재변수 η의 공분산행렬을 계산하면 다음과 같다.

$$Cov(\eta) = (I - B)^{-1}[\Gamma\Phi\Gamma' + \Psi][(I - B)^{-1}]'$$

여기서 정의한 잠재변수 η와 ξ는 값이 관측되는 변수가 아니다. 대신 관측변수 y와 x를 도입하고 이를 잠재변수 η와 ξ에 관련짓는다. 측정 부분에서 y를 $(p \times 1)$차원의 관측변수 벡터, x를 $(q \times 1)$차원의 관측변수 벡터라고 하고 다음을 가정한다.

$$y = \Lambda_y \eta + \epsilon, \ x = \Lambda_x \xi + \delta \tag{2}$$

여기에서 Λ_y와 Λ_x는 요인적재 행렬(factor loading matrix)이고 ϵ과 δ는 관측변수 벡터 y와 x에 대한 관측오차 벡터이다. 관측오차 벡터 ϵ와 δ는 확률벡터로 간주하여 평균은 0이고, 공분산행렬은 각각 $Cov(\epsilon) = \Theta_\epsilon$, $Cov(\delta) = \Theta_\delta$ 이라고 하자. 또한 벡터 η, ξ, ϵ과 δ는 서로 상관관계가 없는(uncorrelated) 것으로 간주한다.

위의 가정으로부터 관측변수 벡터 (y,x)의 공분산행렬 Σ을 구하면 다음과 같다.

$$\Sigma_{(p+q)\times(p+q)} = Cov(y,x) = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix} \tag{3}$$

여기에서

$$\Sigma_{yy} = \Lambda_y (I - B)^{-1}[\Gamma\Phi\Gamma' + \Psi][(I - B)^{-1}]'\Lambda_y{}' + \Theta_\epsilon$$
$$\Sigma_{xx} = \Lambda_x \Phi \Lambda_x{}' + \Theta_\delta$$
$$\Sigma_{yx} = \Lambda_y (I - B)^{-1}\Gamma\Phi\Lambda_x{}' = \Sigma_{xy}{}'$$

이다(Lee 2007; Jöreskog 1994 등). 공분산행렬 Σ의 원소는 모수행렬 Λ_y, Λ_x, B, Γ, Φ, Θ_δ, Θ_ϵ의 원소로 구성됨을 알 수 있다.

2. 모형 가정

구조방정식모형 방법론은 식(1)과 (2)에 제시된 일반적인 구조방정식모형 중에서 다루고자 하는 분야의 이론에 적합한 모형을 설계하는 방법론이다. 따라서 특정 분야의 이론에 따라 어떤 잠재변수는 다른 잠재변수에 영향이 없을 수도 있고, 어떤 관측변수는 다른 잠재변수의 영향이 없거나 그 반대일 수도 있으며, 관측오차 중에는 서로 상관이 없는 경우도 있을 수 있다. 따라서 특정 이론을 가정하면 모수행렬 Λ_y, Λ_x, B, Γ, Φ, Θ_δ, Θ_ϵ의 원소 중에 일부가 0으로 고정되는 것으로 나타난다. 만일 모수행렬 Λ_y, Λ_x, B, Γ, Φ, Θ_δ, Θ_ϵ의 원소 중에 0이 아닌 원소들로 벡터 θ를 구성하면, 특정 이론이 가정하는 구조방정식모형에서 관측변수의 공분산행렬은

$$\Sigma = \Sigma(\theta)$$

가 된다. 따라서 벡터 θ는 공분산행렬 $\Sigma(\theta)$의 특성을 결정하는 모수(parameter) 벡터가 된다.

3. 모형 식별

구조방정식모형의 공분산행렬 $\Sigma(\theta)$을 고려했을 때, 공분산행렬 $\Sigma(\theta)$로부터 모수벡터가 유일하게 표현될 수 있으면 모수벡터 θ는 식별이 가능하다고 한다. 즉, 만일 모수벡터 θ_1과 θ_2에 대하여 공분산행렬이 동일하면, $\Sigma(\theta_1) = = \Sigma(\theta_2)$, 두 모수벡터가 동일해질 때, $\theta_1 = \theta_2$, 두 모수벡터 θ_1, θ_2는 식별가능하다고 한다. 모수벡터 θ가 식별가능하면 공분산행렬 $\Sigma(\theta)$을 갖는 구조방정식모형도 식별가능하다고 한다. 만일 모수벡터 θ가 식별가능하지 않으면 공분산행렬 $\Sigma(\theta)$도 식별가능하지 않고, 따라서 공분산행렬 $\Sigma(\theta)$을 갖는 모형도 식별가능하지 않게 된다.

구조방정식모형의 모수식별 문제는 공분산행렬을 통한 모수식별 문제이다. 모집단 공분산행렬이 공분산 원소 σ_{ij}로 구성되고, $\Sigma = \{\sigma_{ij}\}$, 공분산 원소 σ_{ij}는 모수벡터 θ에 의해 규정되는 구조방정식모형을 따른다고 가정해 보자. 따라서 모수벡터 θ는 공분산행렬 Σ를 $\Sigma(\theta)$로 규정한다. 반대로 공분산행렬 $\Sigma = \{\sigma_{ij}\}$로부터 모수벡터 θ의 원소들을 원소$\{\sigma_{ij}\}$에 의하여 유일하게 표현할 수 있으면 모수벡터 θ의 원소는 식별이 가능한 것이 된다. 또한 모수벡터 θ의 모든 원소가 식별가능하면 구조방정식모형은 식별가능하게 된다.

4. 모수 추정

구조방정식모형에서 모수 추정은 모수벡터 θ를 추정하는 것을 의미한다. 만일 관측변수의 확률벡터 표본 $(y_i', x_i')'$, $i = 1, \ldots, n$,이 공분산행렬 $\Sigma(\theta)$을 갖는 구조방정식모형에서 얻어졌다고 하자. 그리고 이를 이용하여 표본 공분산행렬 S_n를 만들자:

$$S_n = \frac{1}{n-1} \sum_{i=1}^{n} \begin{pmatrix} y_i - \overline{y}_n \\ x_i - \overline{x}_n \end{pmatrix} \begin{pmatrix} y_i - \overline{y}_n \\ x_i - \overline{x}_n \end{pmatrix}', \tag{4}$$

여기에서 $\overline{y}_n = \sum_{i=1}^{n} y_i / n$, $\overline{x}_n = \sum_{i=1}^{n} x_i / n$은 각각 확률벡터 y와 x의 표본평균 벡터이다. 확률벡터 표본이 행렬 $\Sigma(\theta)$을 갖는 구조방정식모형에서 생성되었으므로 S_n은 공분산행렬 $\Sigma(\theta)$를 비편향추정하는 성질이 있다.

만일 모수벡터 θ가 정확하게 식별(just-identified)된다면 즉, θ에 포함된 원소의 수와 공분산행렬 S_n의 식별가능한 원소의 수가 동일하다면 방정식 $S_n = \Sigma(\theta)$을 풀어 유일 해 $\hat{\theta}$를 구할 수 있다. 그리고 그 해 $\hat{\theta}$는 모수벡터 θ의 추정량이 된다. 그리고 만일 모수벡터 θ가 과대식별(over-identified) 된다면, 즉, S_n의 식별가능한 공분산 원소의 수가 모수벡터 θ의 원소의 수보다 많다면 추정된 공분산행렬 $\hat{\Sigma} = \Sigma(\hat{\theta})$과 표본 공분산행렬 S_n의 차이를 최소화하는 값을 추정량으로 하게 된다. 행렬에서는 근접성의 측도가 유일한 것이 아니므로 근접

성을 나타내는 근접성 측도 $F(S_n, \Sigma(\theta))$를 정의한 후 근접성 측도 F를 최소화하는 추정량 $\hat{\theta}$를 구하는 것이 일반적이다. 근접성 측도 F의 형태는 모수추정 방법에 따라 달라진다. 만일 모수벡터 θ를 정확 식별이나 과대 식별하지 못하고 과소식별(under-identified)하게 된다면, 즉 S_n의 식별가능한 공분산 원소의 수가 모수벡터 θ의 원소의 수보다 적다면 모수벡터 θ는 유일하게 추정되지 않는다.

모수벡터를 과대 식별하는 경우에 널리 사용되는 추정방법으로 관측변수 벡터가 다변량 정규분포를 따른다고 가정한 후 우도함수를 최대로 하는 추정량을 구하는 최대우도추정법을 들 수 있다. 이 방법에 의하면 근접성 측도는 다음과 같은 형태가 된다.

$$F(S_n, \Sigma(\theta)) = \log|\Sigma(\theta)| + trace\,(S_n^{-1}\Sigma(\theta)) - \log|S_n| - (p+q) \tag{5}$$

다변량 정규분포 가정이 성립한다면 최대우도 추정량 $\hat{\theta}$는 점근적으로 최소분산을 갖는 효율적인 추정량임이 알려져 있다(Jöreskog 1967 등).

다른 방법 중 널리 쓰이는 모수 추정방법으로는 최소제곱추정법과 일반화 최소제곱추정법이 있다. 일반화 최소제곱추정법에서 사용되는 근접성 측도는 아래와 같다.

$$F(S_n, \Sigma(\theta)) = \frac{1}{2} trace\,[(S_n - \Sigma(\theta))\,W_n^{-1})^2] \tag{6}$$

여기에서 W_n행렬은 가중치행렬로서 양정치(positive definite) 행렬이고, $trace$는 행렬의 트레이스를 나타낸다. 만일 가중치행렬로서 단위행렬을 사용하면 최소제곱추정법이 된다. 이 방법에서 사용되는 근접성 측도는 다음과 같다.

$$F(S_n, \Sigma(\theta)) = \frac{1}{2} trace\,[(S_n - \Sigma(\theta))^2] \tag{7}$$

특정 구조방정식모형이 가정되었을 때 확률벡터 표본의 크기가 증가하면 일반화 최소제곱법으로 구한 모수 추정값은 모수에 점차 가까워진다(Lee 2007, 3장 등).

관측변수에 다변량 정규분포 가정이 적절하지 않는 경우에 사용할 수 있는 방법으로 점근분포무관(asymptotically distribution-free, Browne 1984) 방법 등이 있다.

앞의 식(5)~(7)에서 주어진 근접성 측도는 모수벡터 θ의 원소들의 선형 함수로 쉽게 표현되지 않기 때문에, 이러한 근접성 측도를 최소화하는 해를 구하는 것은 수리적으로 쉽지 않다. 이러한 경우 반복법(iteration method)을 사용하여 해를 구할 수 있다(Jöreskog & Sörbom 1996; Bentler & Wu 2002; Jennrich & Robonson 1969 등). 반복법에는 뉴톤-랩슨 방법, 점수화(scoring) 방법, 가우스-뉴턴 방법 등이 있다. 뉴톤-랩슨 방법은 일반적인 목적함수를 최소화하는 데 사용될 수 있고, 점수화 방법은 우도함수가 사용된 목적함수 최소화에 사용하기 적합하며, 가우스-뉴톤 방법은 일반화 최소제곱법에 사용하기 적합하다.

5. 모형 및 개별 모수 검정

1) 모형 검정

수집한 확률벡터 데이터 $(y_i', x_i')'$, $i = 1, \ldots, n,$가 연구대상으로 설정한 구조방정식모형에 적합한지 여부는 통계적 가설검정을 통하여 판정할 수 있다. 영가설은 '특정 구조방정식모형이 모집단에 부합한다'이다:

$$H_0 : \Sigma = \Sigma(\theta)$$

이에 대하여 대립가설은 '부합하지 않는다'이다: $H_1 : \Sigma \neq \Sigma(\theta)$. 다시 말하면, 영가설은 표본 공분산행렬 S_n이 공분산행렬 $\Sigma(\theta)$을 갖는 구조방정식모형에서 생성되었다는 것이고, 대립가설은 표본 공분산행렬 S_n이 임의의 공분산행렬 Σ을 갖는 구조방정식모형에서 생성되었다는 것이다. 만일 표본 공분산행렬

S_n에 대한 확률분포가 가정되면 우도비 검정(likelihood ratio test) 방법을 통하여 통계적 가설검정을 실시할 수 있다. 만일 표본 공분산행렬의 분포함수를 f라고 한다면, 위 가설에 대한 우도비 함수는 다음과 같이 주어진다.

$$LR(\hat{\theta}_n) = \frac{f(S_n|\Sigma(\hat{\theta}_n))}{f(S_n|\hat{\Sigma})}$$

그리고 만일 관측변수 벡터 (y,x)가 공분산행렬 $\Sigma(\theta)$를 갖는 다변량 정규분포에서 생성되었다고 가정하면 $-2\log(LR(\hat{\theta}_n))$은 점근적으로 카이제곱분포에 수렴함이 알려져 있다(Johnson & Wichern 2007 등). 즉, 표본수 n이 증가함에 따라 다음이 성립한다.

$$-2\log(LR(\hat{\theta}_n)) \rightarrow \chi^2_{df} \tag{8}$$

이때 자유도 df는 모집단 공분산행렬의 식별가능한 공분산 원소의 수에서 모수벡터 θ의 원소의 수를 뺀 값이다.

$$df = \frac{1}{2}(p+q)(p+q+1) - (\theta\text{의 원소의 수})$$

주어진 확률벡터 $(y_i{}', x_i{}')' i = 1, \ldots, n,$ 값을 이용해 검정통계량 $-2\log(LR(\hat{\theta}_n))$의 값을 계산한 후 카이제곱분포의 임계값과 비교하여 임계값보다 크면 대립가설을 받아들이고, 반대로 작으면 영가설을 받아들인다. 영가설을 받아들이게 되면 연구자가 설정한 모형, 즉, 모수벡터 θ를 갖는 구조방정식모형을 채택한다는 뜻이다.

2) 개별 모수 검정

모수벡터 θ에 속하는 특정 모수의 유의성을 검정하는 문제를 고려해 보자. 모수벡터 θ에서 i번째 위치의 원소를 뺀 모수벡터를 $\theta^{(i)}$라 하고 $-2\log(LR(\hat{\theta}_n^{(i)}))$를 모수벡터 $\theta^{(i)}$를 검정하는 검정통계량 값이라고 하자. 그러면 식(8)을 이용

하면 표본 수 n이 증가하면 점근적으로 다음이 성립한다.

$$-2\log(LR(\hat{\theta}_n^{(i)})) \rightarrow \chi^2_{df+1}$$

여기에서 자유도가 $(df+1)$로서 모수벡터 θ인 경우의 자유도 df보다 1이 더 큰 이유는 모수벡터 $\theta^{(i)}$의 원소의 수가 모수벡터 θ이 원소의 수보다 1이 작기 때문이다. 따라서 검정통계량 $-2\log(LR(\hat{\theta}_n^{(i)}))$의 값과 $-2\log(LR(\hat{\theta}_n))$의 값의 차이는 순수하게 특정 모수 때문에 생긴다. 결국 두 값의 차이가 크다는 것의 의미는 특정 모수의 모형에 대한 기여도가 크다는 것을 의미하게 된다. 또한 카이제곱 분포의 이론에 의하면 두 검정통계량의 차이는 자유도가 1인 카이제곱 분포를 따르므로,

$$-2\log(LR(\hat{\theta}_n^{(i)})) + 2\log(LR(\hat{\theta}_n)) = 2\log\left(\frac{LR(\hat{\theta}_n)}{LR(\hat{\theta}_n^{(i)})}\right) \rightarrow \chi^2_1 \tag{9}$$

가 된다. 두 값의 차이를 자유도 1인 카이제곱 분포의 임계값과 비교하여 특정 모수의 유의성을 검정할 수 있다. 즉, 식(9)의 통계량 값이 임계값보다 크면 검정의 대상이 되는 특정 모수는 모형에 대한 기여도가 큰 것이므로 모형에 포함시켜야 한다.

앞에서 설명한 우도비 함수의 차이를 이용한 개별 모수 검정법 이외에 라그랑쥐 승수(Laglange multiplier)를 이용한 검정법, 왈드 검정법(Wald 1943) 등이 있다.

Ⅲ. 통계적 쟁점

본 절에서는 앞 절에서 고찰한 구조방정식모형에 대한 고찰을 바탕으로 통계적 관점에서 쟁점들을 토의하고자 한다. 데이터 생성, 변수의 측도, 인과관계, 잠재변수 식별, 모형 식별 및 모수 추정, 모형 적합 및 검정 등에 관한 내용

을 토의한다.

1. 데이터 생성

구조방정식모형에서는 분석데이터를 다변량 확률데이터로 가정한다. 즉, 측정 부분에서 측정변수 벡터 $(y_i', x_i')'$, $i = 1, \ldots, n$,는 서로 독립이고 동일한 분포를 갖는 확률벡터로 가정되며 확률벡터의 공분산행렬은 식(3)에 주어진 바와 같이 가정된다. 확률 데이터가 가정되어야 이를 근거로 통계분포가 유도되며, 이 통계분포로부터 각종 확률이 계산된다. 데이터 생성 매카니즘에 대한 가정이 이러함에도 불구하고 많은 연구자들은 이에 대한 관심을 크게 갖지 않는 것이 현실이다(Freedman 1987 등). 연구자가 데이터 수집이나 생성에 관여하지 않고 다만 주어진 데이터를 분석하는 경우에는 데이터 생성 매카니즘에 대한 이해가 충분하지 않을 수도 있다. 그렇더라도 데이터를 LISREL 등 통계 소프트웨어에 입력하고 수행하면 결과가 출력되므로 구조방정식모형 분석결과를 얻는 데는 큰 불편함이 없다고 느낄 수 있다. 다만 문제는, 만일 분석 데이터가 구조방정식모형에서 가정하는 확률데이터가 아니라면 출력결과 수치들은 단지 계산결과에 불과한 수치라는 점이다. 이러한 수치들을 근거로 구조방정식모형 이론을 적용하여 어떤 현상을 설명하는 것은 적절하지 않다. 그 이유는 다음과 같다.

구조방정식모형 분석은 식(4)에 제시된 표본 공분산행렬 S_n을 기초로 한다. 그리고 구조방정식모형 분석 결과 수치들은 대부분 표본 공분산행렬 S_n에서 유도된 것이다. 분석대상인 식(3)의 모집단 공분산행렬 Σ는 미지의 행렬이므로 표본행렬 S_n을 대신 사용하는 것이고, 그러한 기저에는 확률표본 행렬 S_n은 모집단 공분산행렬의 비편향추정량이고 일치추정량이라는 이론적 사실이 있다. 따라서 만일 분석하고자 하는 데이터가 확률표본이 아니라면 이러한 사실은 성립하지 않으므로 표본행렬 S_n에 기초한 구조방정식모형 분석은 이론적 근거가 없는 분석이 되고 만다.

확률표본은 확률 매카니즘에 의한 데이터 수집이나 획득을 전제로 한다. 그러나 실제 많은 사회과학 데이터들이 모두 확률 매카니즘에 의해서 수집되었거나 획득되었다고 보기는 어렵다. 조사설계나 실험설계에 의한 조사나 실험에서 얻어진 데이터는 확률데이터로 간주할 수 있다. 그러나 설계 없이 비확률적으로 얻어진 데이터, 예를 들면 인터넷에서 얻어진 데이터, 행정데이터 등은 확률데이터로 보기 어렵기 때문에 구조방정식모형에 적합하기 위해서는 면밀한 검토를 거쳐야 한다.

확률데이터라 하더라도 횡단면 조사나 패널조사 등에서 얻어진 가중치가 부여된 데이터는 구조방정식모형의 가정에 곧바로 일치하지는 않는다. 포함확률이 다르기 때문에 데이터가 모두 동일한 확률로 획득됐다고 간주하기 어렵기 때문이다. 이와 같은 경우 포함확률의 역수로 가중치를 만들고, 가중치가 부여된 데이터로 표본 공분산행렬을 추정한 후, 분석을 진행하는 방법이 알려져 있다(예를 들면 Muthen 1995; Asparouhov 2005; Vermunt & Magidson 2007 등).

2. 변수의 척도

구조방정식모형 분석은 공분산 구조를 분석하는 방법론이다. 통상적으로 언급되는 공분산은 식(4)에 제시된 피어슨 공분산으로서 연속형 변수에 적합한 공분산이다. 그런데 심리학, 사회학, 교육학 등에서 사용되는 데이터에는 5점 척도, 7점 척도 데이터와 같은 순서형 데이터나 이항 데이터와 같은 명목형 데이터가 흔하다. 순서형 데이터는 척도 값은 있으나 척도 값 간의 거리가 정의되지 않기 때문에, 그리고 명목형 데이터는 척도 값 자체가 의미가 없기 때문에 순서형 데이터나 명목형 데이터에 대하여 식(4)의 피어슨 표본 공분산행렬을 구한 후 분석을 진행하는 것은 올바른 분석 방법이 아니다. 명목형 변수에 대해서는 이계열상관계수(二系列相關係數, biserial correlation coefficient)를 구하고, 순서형 변수에 대해서는 순서형 변수에 대응하는 연속형 잠재변수를 고려한 후 연속형 잠재변수로부터 공분산행렬을 유도하여 대신 사용하는 방법이 알려져 있다 (Jöreskog 1994; Xie 1989; Muthen 1983 등).

3. 인과관계

구조방정식모형에서는 변수들 간의 관계를 인과관계로 설명하는 것이 가능하다. 상관계수나 회귀모형의 회귀계수는 변수들 간의 연관성 측도이지 인과관계 측도는 아니므로 변수 간의 관계를 연관관계로만 설명하는 것이 타당하다. 단지 연구자가 필요에 따라 연관관계를 인과관계로 해석하는 것뿐이다. 이에 비하면 구조방정식모형은 인과관계를 자체적으로 포함하고 있으므로 모형에 근거한 인과관계 설명은 타당한 설명이고, 이 점은 구조방정식모형의 큰 특징이 된다.

예를 들어 x를 설명변수로 하고 y를 반응변수로 하는 회귀모형에서 회귀계수는 $\beta_{y|x}=\rho_{xy}\sigma_y/\sigma_x$ 이다. 만일 x와 y가 표준화된 변수라고 가정하면 회귀계수 $\beta_{y|x}$는 상관계수 ρ_{xy}가 된다. 회귀계수를 연관성 측도로 해석하면 두 변수의 연관 정도는 상관계수 ρ_{xy}로 설명된다. 반면 인과관계로 해석하면 $\beta_{y|x}$는 설명변수 x가 반응변수 y에 미치는 영향력으로 해석되고, 영향력의 정도는 ρ_{xy}가 된다. 이제 반대로 y를 설명변수, x를 반응변수로 하는 회귀모형을 고려하면 회귀계수는 $\beta_{x|y}=\rho_{xy}\sigma_x/\sigma_y$ 이고, 표준화된 변수에서 회귀계수는 $\beta_{y|x}=\rho_{xy}$ 가 된다. 즉, 표준화된 변수에서 상관계수 ρ_{xy}는 x를 원인, y를 결과로 하는 인과관계 수치가 되기도 하고, 혹은 그 반대가 되기도 한다. 바꿔 말하면 통계데이터는 연관성과 인과관계를 구분하여 설명하지 않는다. 다만 연구자가 모형을 근거로 동일한 수치를 보고 연관성 측도나 인과관계 측도로 활용하는 것이다. 회귀모형을 설정하면 연관성 측도로 활용하고, 경로분석모형을 설정하면 인과관계 측도로 사용하는 것뿐이다. 그러나 일반적으로 상관계수와 같은 연관성 측도는 인과관계를 함의하지 않는다(Guttman 1977 등).

통계모형에 근거한 분석에서 연관성과 인과관계가 뚜렷하게 구분되지 않는 이유는 통계모형이 콜모고로프의 확률공리 위에서 구축되었기 때문이다 (Suppes 1982). 콜모고로프의 확률공리에서 인과관계는 조건부 확률 $p(A|B)$로 설명될 수 있는데, 이때 사건 A와 사건 B의 의미를 살펴보아야 하다. 만일 사건이 시간적으로 A,B 순으로 발생했다면 베이즈 정리를 이용하여 $p(A|B)$

를 계산할 수는 있지만, 나중에 일어난 사건 B가 먼저 일어난 사건 A의 원인으로 해석될 수는 없기 때문이다.

4. 잠재변수 식별

교육학의 지능 연구나 사회학의 사회계층 연구 등에서와 같이 잠재변수를 고려하고, 고려된 잠재변수를 포함하는 모형을 구축하는 것은 매우 의미 있는 연구방법이다. 문제는 잠재변수를 실측한 데이터가 없기 때문에 잠재변수의 존재를 확인하고 이를 수치로 나타내기가 매우 어렵다는 데 있다. 식(2)의 요인모형을 고려하자: $y = \Lambda\eta + \epsilon$. 여기에서 y는 관측 가능한 변수이기 때문에 데이터가 있고, η는 잠재변수이기 때문에 관련 데이터가 없다. 위 요인모형에서 잠재변수 η와 요인적재행렬 Λ가 구분되어 식별되지 않으면 잠재변수 η와 요인적재행렬 Λ는 유일하게 추정되지 않는다. 만일 직교행렬 B를 사용하면 $\Lambda\eta = (\Lambda B)(B'\eta)$이 언제나 성립하기 때문에 요인적재행렬 Λ와 잠재변수 η가 유일하게 식별되지 않는 것이다. 이를 방지하기 위하여 잠재변수가 다변량 정규분포를 따른다는 강한 가정과 유일성 조건(Johnson & Wichern 2007: 496)을 부여하여 잠재변수를 식별하고 요인적재행렬을 유일하게 추정한다. 데이터가 생성되지 않는 잠재변수를 모형에 포함시키기 위해서는 데이터 대신 강한 가정을 부여해야 하는 상황이 결과적으로 초래된 것이다.

통계모형에 부여된 가정은 모형 적합 후 데이터를 통하여 가정 부여의 타당성을 점검해야 한다. 비록 정규분포 가정을 통하여 구조방정식모형에 잠재변수를 가정했지만 모형 적합 후에 부여된 가정의 적절성은 데이터를 통하여 사후적으로 점검해야 하는 것이다. 통상적으로 회귀모형에서는 잔차 검토를 통하여 정규분포 가정 검토를 수행하고, 요인분석에서는 잠재변수를 추정한 후에 그림을 그려 정규분포 가정의 적절성을 검토한다. 반면 구조방정식모형에서는 잠재변수의 정규분포 가정을 점검하는 것이 쉽지 않다. 통상적으로 잔차는 관측값에서 적합값을 뺀 것인데 잠재인자는 관측값이 없으니 잔차도 정의될 수 없다. 만일 잠재변수의 잔차를 정의할 수 있으면 다변량 정규분포 가정 검토를 실시할 수 있을 것이다(Johnson & Wichern 2007: 177).

5. 모형 식별 및 모수 추정

일반적으로 어떤 확률모형에 대해 이에 대응하는 공분산행렬은 유일하게 정의되지만, 반대로 그 역은 성립하지 않는다. 예를 들어 어떤 데이터의 분산이 3이라고 했을 때 이 데이터가 분산이 3인 정규분포에서 생성되었을 것이라고 주장하는 것은 무리이다. 분산이 3인 감마분포나 다른 분포에서도 생성될 수 있기 때문이다. 구조방정식모형 방법론에서도 이와 유사한 일이 발생한다. 구조방정식모형 방법론은 공분산행렬을 통해 이뤄진다. 따라서 구조방정식모형에서 유도된 공분산행렬로부터 결과를 얻은 후, 그 결과를 이용하여 원래의 구조방정식모형을 설명하는 것은 논리적으로 과장의 문제가 있다. 하나의 공분산행렬에 대응하는 구조방정식모형은 무수히 많기 때문이다. 따라서 이러한 문제를 방지하려면 구조방정식모형에 정규분포 가정과 같은 강한 가정을 계속 부여해야 한다. 정규분포에서는 평균벡터와 공분산행렬이 동일하면 같은 분포라는 사실이 성립하기 때문이다. 앞 소절에서도 언급한 바와 같이 모형에 부여된 가정은 사후적으로 점검되어야 한다. 불행하게도 아직까지는 다변량 정규분포 이외에는 다변량 변수 벡터에 활용가능한 분포가 마땅하지 않다.

구조방정식모형에서 모수는 구조방정식에서 직접 추정되는 것이 아니고 공분산행렬에서 간접적으로 추정된다. 그리고 식(5) 등에서와 같이 근접성 측도를 최소화하는 추정량을 찾게 된다. 그리고 근접성 측도가 선형이 아니므로 반복법을 통하여 근을 추정하게 된다. 반복법을 사용하면 구조방정식모형에서도 초기치 지정 문제, 추정치가 범위를 벗어나는 문제(Heywood 문제), 반복이 수렴을 안 하는 문제 등이 발생할 수 있다.

6. 모형 적합 및 검정

통계모형 적합에서는 적합도(goodness-of-fit)를 높이려는 적합도 원리와 불필요한 내용을 제거하고 모형을 간단하게 하려는 절약(parsimony)의 원리를 동시에 고려해야 한다. 모형이 정교해지면 적합도의 원리는 충족하지만 절

약의 원리는 위배하는 반면, 모형을 간단하게 하면 절약의 원리는 충족되지만 적합도는 떨어지게 된다. 구조방정식모형에서 모수의 수는 공분산행렬, $\Sigma(\theta)$, 에 포함된 모수벡터 θ의 원소의 수이다. 만일 θ의 원소의 수가 증가하면 적합도는 높아지게 된다. 모형 적합도를 나타내는 지표는 모형 카이제곱 값, 제곱근 근사 평균제곱오차, 적합도 통계값(goodness-of-fit statistics, GFI), 제곱근 평균제곱잔차(root mean square residual, RMR) 등이 있다. 이러한 지표들은 모두 모형 적합도를 표현하는 지표들이기 때문에 이러한 지표를 기초로 모형을 선택하다 보면 주어진 데이터에만 지나치게 잘 적합되는 모형을 선택하기 쉽다. 이른바 주어진 데이터에 집착하는 과대적합(over-fitting) 현상이 발생하는 것이다. 구조방정식모형에서 적합성 지표를 사용했을 때 발생하는 모형 과대적합과 관련된 구체적인 논의는 이기종(2016)을 참고할 수 있다.

추정에서와 마찬가지로 구조방정식모형에서 모형 검정은 공분산행렬을 통하여 이루어진다. 이때 귀무가설 혹은 영가설 H_0과 대립가설 H_1은 다음과 같다.

$$H_0 \ : \Sigma = \Sigma(\theta) \quad \text{v.s} \ H_1 : \Sigma \neq \Sigma(\theta)$$

구조방정식모형에 대한 가설검정이 통상적인 가설검정과 다른 점은 통상적인 가설검정에서는 영가설을 기각시키는 것이 목적인 반면, 구조방정식모형에서는 영가설을 채택하려는 것이 목적이라는 점이다. 영가설을 채택하기 위해서는 검정(power)값을 계산한 후 결정을 해야 하지만, 구조방정식모형에서 검정값을 계산하는 것은 매우 힘든 일이므로 보통 유의확률을 보고 양자택일을 한다. 유의확률을 보고 영가설을 채택하는 것은 주어진 모형이 옳다는 뜻이 아니고, 옳지 않다는 확신이 없어 채택한다는 의미이므로 주의를 해야 한다.

Ⅳ. 토의

1900년대 초에 심리학, 교육학, 경제학 등에서 개별이론으로 개발되어 발전되어 오다가 1970년대에 하나의 융합방법론으로 발전한 구조방정식모형 방법론

은 모형의 확장성, 유연성 그리고 소프트웨어의 개발에 힘입어 사용 편리성까지 획득하면서 이제는 계량 사회과학 분야에서는 가장 인기있고 가장 많이 활용되는 방법론이 되었다. 구조방정식모형은 잠재변수를 명시적으로 포함하고 인과관계 설명이 가능하기 때문에, 현상적인 데이터는 없지만 현상을 움직이는 동력을 찾아 원인과 결과를 규명하고자 하는 연구자의 욕구를 충족시켜주는 면이 있다. 또한 구조방정식모형 분석에서는 단지 공분산행렬 데이터만 필요하므로 소프트웨어를 활용한 분석 수행이 매우 수월하여 연구자가 결과를 도출하기가 매우 손쉬운 면도 있다. 이러한 면들이 복합적으로 작용하여 구조방정식모형의 활용 범위는 점점 넓어지고 있다.

이러한 밝은 면 뒤에는 정교하고 세밀하게 점검해야 하는 면이 동시에 존재한다. 구조방정식모형 분석에서는 공분산행렬 데이터만으로 많은 결과를 도출하는데, 공분산행렬의 자유도는 대체로 크지 않으므로 이를 보완하기 위하여 다변량 정규분포 가정이나 유일성 조건 가정 등 강한 가정을 구조방정식모형에 부여하게 된다. 따라서 이러한 가정이 성립하는 조건에서 구조방정식모형 분석은 타당성을 갖게 된다. 그러나 이러한 강한 가정을 데이터로 충분히 확인하기 어렵다는 점이 구조방정식모형 방법론이 갖는 단점이자 한계이다. 구조방정식모형의 단점과 한계를 명확하게 인식할 때 모형 활용을 좀 더 잘 하고, 오류의 길로 가지 않는 방어력을 가질 수 있다. 이러한 관점에서 본 연구에서는 구조방정식모형에 대한 통계적 쟁점을 데이터 생성, 변수의 척도, 인과관계, 잠재변수 식별, 모형 식별 및 모수 추정, 모형 적합 및 검정의 6가지로 나누어 논의하였다. 주로 통계적인 관점에서 구조방정식모형의 구조 및 기본 개념에 초점을 맞춘 쟁점들이다. 논의되지 않는 중요한 통계적 쟁점으로는 결측치 문제, 정규분포를 따르지 않는 데이터 문제, 데이터 변환 문제, 비선형 구조 문제, 비확률 데이터 문제, 종단면 데이터 문제, 조사데이터 가중치 문제 등이 남아 있다. 이러한 문제는 향후 연구로 남긴다.

연구자에 따라서는 통계적 쟁점 말고도 다른 쟁점들을 논의할 수 있을 것이다. 구조방정식모형 분석결과 보고도 그러한 점 중의 하나이다. 구조방정식 최종모형의 유일성과 잠재변수 식별성의 수치적 근거가 희박하면 모형에 부여된

가정으로 결론을 유도해야 하기 때문에 결과 분석 및 보고에 연구자들은 많은 신경을 쓰지 않을 수 없게 된다. 또한 데이터 획득 과정이 명료하게 설명되지 않으면 비록 구조방정식모형 분석으로 어떤 결과를 얻었다고 하더라도 이를 일반화하여 사회 현상을 설명하는 도구로 사용하는 데에는 무리가 따를 수 있다. 적은 자유도를 갖는 공분산행렬 데이터로 과도하게 모형을 적합하여 과대적합현상이 나타나면 이러한 불일치는 더 심해질 수 있다.

참고문헌

김민석·민진. 2014. “인사제도가 조직 효과성에 미치는 영향에 관한 연구: 제도내재화의 매개효과를 중심으로.” ≪한국인사행정학회보≫ 13: 89-118.

김종기·전진환. 2009. “국내 MIS 연구에서 구조방정식모형 활용에 관한 메타분석.” *Asia Pacific Journal of Information System* 19: 47-75.

심준섭. 2015. “행정학 연구에서 구조방정식모형 활용: 문제점 검토와 제언.” ≪한국행정학보≫ 49: 453-485.

오현복·남미향·이경숙·박인우. 2015. “교육서비스품질, 고객만족, 기관성과, 사회성과의 구조적 관계분석: A 공공기관을 중심으로.” *The Korean Journal of Human Resource Develpoment Quarterly* 17: 1-29.

이기종. 2016. “구조방정식모형의 모형평가 오·남용 문제와 교정.” ≪조사연구≫ 17: 71-83.

정효채·석진홍·박우성. 2013. “연령이 직무만족에 미치는 영향: 잠재성장모형을 이용한 종단연구.” ≪노동정책연구≫ 2: 67-93.

Asparouhov, T. 2005. “Sampling Weights in Latent Variable Modeling.” *Structural Equation Modeling* 12: 411-434.

Bentler, P.M. “SEM with Simplicity and Accuracy.” *Journal of Consumer Psychology* 20: 215-220.

Bentler, P.M. and Wu, E.J.C. 2002. *EQS 6 for Windows User's Guide*. Encino, CA: Multivariate Software.

Bollen, K.A. and Pearl, J. 2013. "Eight Myths about Causality and Structural Equation Models." In S.L. Morgan(ed.), *Handbook of Causal Analysis for Social Research*: 301-328.

Browne, M.W. 1984. "Asymptotically Distribution Free Methods in Analysis of Covariance Structure." *British Journal of Mathematical and Statistical Psychology* 37: 62-83.

Chin, W.W. 1998. "Commentary: Issues and Opinion on Structural Equation Modeling." *MIS Quaeterly* 22: vii-xvi.

Fabrigar, L.R., R.D. Porter, and M.E. Norris. 2010. "Some Things You should Know about Structural Equation Modeling but Never Thought to Ask." *Journal of Consumer Psychology* 20: 221-225.

Freefman, D.A. 1987. "As Others See Us: A Case Study in Path Analysis." *Journal of Eduational Statistics* 12: 101-128.

Guttman, L. 1977 "What Is not What in Statistics." *Journal of the Royal Statistical Society* 26: 81-107.

Hermida, R. 2015. "The Problem of Allowing Correlated Errors in Structural Equation Modeling: Concerns and Considerations." *Computational Metods in Social Sciences* 3: 1-13. (http://cmss.univnt.ro)

Iacobucci, D. 2009. "Everything You always Wanted to Know about SEM but Were Afraid to Ask." *Journal of Consumer Psychology* 19: 673-80.

Jennich, R. and S.M. Robinson. 1969. "A Newton-Raphson Algorithm for Maximum Likelihood Factor Analysis." *Psychometrika* 34: 111-123.

Johnson, R.A. and D.W. Wichern. 2007. *Applied Multivariate Statistical Analysis*(6th ed.). Pearson Education.

Jöreskog, K.G. 1967. "Some Contributions to Maximum Likelihood Factor Analysis." *Psychometrika* 32: 443-482.

Jöreskog, K.G. 1972. "Factor Analysis by Generalized Least Squares." *Psychometrika* 37: 243-260.

Jöreskog, K.G. 1973. "A General Method for Estimating a Linear Structural Equation System." In Goldberg A.S. and O.D. Duncan(eds.) *Structural Equation Models in the Social Sciences*. Academic Press, 85-112.

Jöreskog, K.G. 1994. “Structural Equation Modeling with Ordinal Variables.” *Multivariate Analysis and its Applications*: 297-310.

Jöreskog, K.G. and D. Sörbom. 1996. *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Hove and London: Scientific Software International.

Kline, R.B. 2011. *Principles and Practice of Structural Equation Modeling* (3rd ed.). The Guilford Press.

Lee, Sik-Yum 2007. *Structural Equation Modeling: A Bayesian Approach*. Wiley.

Muthen, B.O. 1983. “Latent Variable Structural Equation Modeling with Categorical Data.” *Journal of Econometrics* 22: 43-65.

Muthen, B.O. 1995. “Complex Sample Data in Structural Equation Modeling.” *Sociological Methodology* 25: 267-316.

Nachtigall, C., U. Froehne, F. Funke., and R. Steyer. 2003. “(Why) Should We Use SEM? Pros and Cons of Structural Equation Modeling.” *Methods of Psychological Research* 8: 1-22.

Spearman, C. 1904. “General Intelligence: Objectively Determined and Measured.” *The American Journal of Psychology* 15: 201-292.

Suppes, P. 1982. “Problems of Causal Analysis in the Social Science.” *Epistemologia, Special Issue* 239-250.

Vermunt, J.K. and Magidson, J. 2007. “Latent Class Analysis with Sampling Weights: A Maximum-likelihood Approach.” *Sociological Methods & Research* 36: 87-111.

Wald, A. 1943. “Tests of Statistical Hypothesis Concerning Several Parameters When the Number of Observations is Large.” *Transactions of the American Mathematical Society* 54: 426-482.

Wright, S. 1918. “On the Nature of Size Factors.” *Genetics* 3: 367-374.

Wright, S. 1934. “The Method of Path Coefficient.” *The Annals of Mathematical Statistics* 5: 161-215.

Xie, Y. 1989. “Structural Estimation Models for Ordinal Variables: An Analysis of Occupational Destination." *Sociological Methods & Research* 17: 325-251.

<접수 2016/1/31, 수정 2016/2/17, 게재확정 2016/2/22>