

## 공분산 구조모형의 적합도 평가에 있어서 결측치 처리 방법 비교

### 완전정보최대우도, 다중대체, 베이지안 접근법을 중심으로

A Comparison of Full Information Maximum Likelihood, Multiple Imputation, and Bayesian Approach in Overall Goodness of Fit Assessment of Structural Equation Modeling with Missing Data

---

저자 (Authors)	이태헌 Tae Hun Lee
출처 (Source)	<a href="#">한국심리학회지: 일반 33(2)</a> , 2014.6, 507-536(30 pages) <a href="#">THE KOREAN JOURNAL OF PSYCHOLOGY : GENERAL 33(2)</a> , 2014.6, 507-536(30 pages)
발행처 (Publisher)	<a href="#">한국심리학회</a> The Korean Psychological Association
URL	<a href="http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE06368373">http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE06368373</a>
APA Style	이태헌 (2014). 공분산 구조모형의 적합도 평가에 있어서 결측치 처리 방법 비교. 한국심리학회지: 일반, 33(2), 507-536
이용정보 (Accessed)	중앙대학교 165.***.103.13 2019/06/24 16:13 (KST)

---

#### 저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

#### Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

## 공분산 구조모형의 적합도 평가에 있어서 결측치 처리 방법 비교: 완전정보최대우도, 다중대체, 베이지안 접근법을 중심으로

이 태 현<sup>†</sup>

University of Oklahoma

통계 모형을 이용한 데이터 분석에서 적절한 결측치 처리가 최종적인 통계적 결론 도출에 결정적인 영향을 미칠 수 있다는 것은 이미 잘 알려져 있다. 따라서 현재까지 다양한 결측치 처리 기법이 제안되어왔고, 그 중 최근 활발히 연구되고 있는 기법으로서 완전정보최대우도(full information maximum likelihood), 다중대체(multiple imputation), 그리고 베이지안(Bayesian) 접근법이 있다. 이 중 완전정보최대우도법은 사회과학 통계에서 가장 많이 사용되는 방법으로서 많은 심리학 연구자들에게도 이미 알려져 있는 방법이나, 다른 두 방법, 즉 다중대체 및 베이지안 접근법은 아직은 많은 심리학 연구자들에게 생소한 방법이다. 따라서 본 논문에서는 이 세 가지 방법을 공분산 구조모형의 맥락에서 소개하고 주요 특징을 비교함으로써 심리학 연구자들의 결측치 처리에 관한 이해를 돕고자 하였다. 공분산구조모형의 적용 과정은 다른 통계모형과 달리 자료와 모형간의 적합도 지수를 계산하고 이를 바탕으로 모형의 적절성을 판단하는 과정을 포함한다. 최대우도법에서는 카이제곱 통계량을 기준으로 다양한 적합도 지수가 파생되어 제안되었으며, 다중대체법은 D2와 D3 통계량이, 그리고 베이지안 접근법에서는 사후예측모형검증(posterior predictive model checking) 기법이 사용된다. 최대우도법의 카이제곱 통계량은 다양한 맥락에서 독립적으로 연구된 결과가 있으나, 다중대체법 및 베이지안 접근법에서 제안된 방법은 공분산 구조모형의 맥락에서 그 성능이 평가된 적이 없으며 서로 비교된 적도 없다. 따라서, 본 논문에서는 동일한 구조방정식모형의 자료-모형간 적합성 판단에 있어서 세 가지 다른 결측치 처리 방법이 어떤 영향을 미치는지 비교 분석하였다. 구체적으로, 본 논문에서는 모의실험(simulation) 기법을 사용하여, 모집단에서 데이터가 종단적 측정 불변성을 지지하지 않는 경우(longitudinal measurement non-invariance)를 가정하고, 옳은 모형을 사용한 경우의 제 1 종 오류율과 부분측정불변성(partial measurement invariance)을 가정한 모형을 적용한 경우의 검정력을 최대우도, 다중대체, 그리고 베이지안 접근법 별로 추정/비교하였다. 본 연구의 결과는 결측치가 존재하는 데이터를 공분산 구조모형을 이용하여 분석하고자 하는 연구자들에게 결측치 처리 기법의 특성을 이해하고 최종 모형의 적합성을 판단하는데 적절한 지침을 제공할 것으로 기대된다.

주요어 : 결측치, 공분산구조모형, 최대우도, 다중대체, 베이지안

<sup>†</sup> 교신저자: 이태현, University of Oklahoma, 455 West Lindsey Street, Norman, Oklahoma 73072  
Tel: +1-405-325-1604, E-mail: taehun@ou.edu

공분산 구조분석 모형을 이용하여 분석하고자 하는 데이터의 거의 대부분은 결측치(missing values)를 포함하고 있다고 해도 과언이 아니다. 통계 모형을 이용한 데이터 분석에서 적절한 결측치 처리가 최종적인 통계적 결론 도출에 결정적인 영향을 미칠 수 있다는 것은 이미 잘 알려져 있다(Schafer & Graham, 2002). 따라서 결측치를 올바른 방법으로 처리하는 것은 이후 연구자가 제안한 공분산구조분석 모형의 통계적 추론 결과를 신뢰할 수 있는 선행 조건이 된다. 지금까지 결측치 처리를 위한 다양한 방법이 제안되어 왔으나, 공분산구조모형의 맥락에서 현재 가장 널리 알려져 있고 다양한 통계소프트웨어에서 쉽게 이용할 수 있는 방법은 완전정보(full information) - 최대우도법이다(Finkelstein, 1979; Allison, 1987; Arbuckle, 1996; Wilks, 1932). 이에 더하여 최근에는 다중대체법(Rubin, 1976, 1977, 1987), 그리고 베이지안 접근법(Lee, 2007)이 많은 연구자들의 관심을 받고 있다.

이 세 가지 방법들은 모두 확률모형에 기반하여 결측치 및 데이터의 발생 과정을 설명하려는 통계적 기법들로서, 이론적 근거는 부족하지만 손쉽게 결측치를 처리할 수 있어 전통적으로 그리고 아직도 널리 사용되고 있는 줄지우기(listwise deletion) 혹은 짝지우기(pairwise deletion) 기법들에 비해 훨씬 나은 통계적 특성을 지니고 있음이 잘 알려져 있다(Little & Rubin, 2002). 즉, 확률모형에 기반하여 결측치를 처리한 뒤 연구자가 제안한 통계모형에 포함된 모수를 추정할 경우 그 추정치가 훨씬 덜 편향(bias)되고 훨씬 더 효율적임(efficiency)이 잘 알려져 있다. 회귀분석의 맥락에서는 Enders & Bandalos(2001) 그리고 중단연구의 맥락에서는 Newman(2003)이 완전정보-최대우도법과 다

중대체법이 전통적 방법에 비해 훨씬 덜 편향되고 더욱 효율적인 추정치를 제공한다는 것을 보여주었다.

Rubin(1976)이 최초로 체계화하고 현재 널리 받아들여지고 있는 결측치 분류체계에 따르면, 확률모형에 기반한 결측치 처리기법은 결측치가 완전무선적(missing completely at random, MCAR)으로 발생하거나 혹은 조건무선적(missing at random, MAR)으로 발생한 경우 모두 모수 추정치의 편향 및 효율성 면에서 더 우수한 것으로 알려져 있다. 간단히 설명하자면<sup>1)</sup> MCAR에서는 결측치를 포함한 데이터는 결측치가 발생하기 이전에 얻을 수 있음직한 가상적인 완전자료(complete data)로부터 무선 추출된 하나의 표본이라 간주한다. 따라서 MCAR의 가정 하에서는 결측치를 포함하고 있는 관찰된 데이터는 결측치에 대해 특별한 통계적 처리를 할 필요가 없는 무선 표본 중 하나일 따름이다.

반면에, MAR에서는 한 변인에 발생할 수 있는 결측치는 완전무선적으로 발생하는 것이 아니라, 다른 관찰된 변인의 값에 따라 그 확률이 정해진다고 가정한다. 따라서, 완전무선적 과정이 아니라 조건부 확률에 의해 결측치의 발생이 결정된다는 가정이다. 만약 결측치를 포함하는 데이터가 원자료의 완전무선 표본이 아닐 경우 MCAR의 가정은 위배되며 전통적인 결측치 처리 기법은 대표성이 결여된 부분적인 데이터만을 이용함으로써 편향되고 비효율적인 방식으로 모수를 추정할 수밖에 없게 된다. 그러나, 이 경우에도 확률모형에 기반한 기법들은 덜 편향되고 효율적인 모수

1) 결측치 발생과정에 대한 확률적 메커니즘에 대한 자세한 설명은 Little & Rubin(2002)를 참조하기 바란다.

추정치를 제공하는 것으로 널리 알려져 있다 (Little & Rubin, 2002).

이 중 완전정보최대우도법은 사회과학 통계에서 가장 많이 사용되는 방법으로서 많은 심리학 연구자들에게도 이미 알려져 있는 방법이나, 다른 두 방법, 즉 다중대체 및 베이지안 접근법은 아직은 많은 심리학 연구자들에게 생소한 방법이다. 따라서 본 논문에서는 이제 가지 방법을 공분산구조모형의 맥락에서 소개하고 주요 특징을 비교함으로써 심리학 연구자들의 결측치 처리에 관한 이해를 돕고자 하였다.

또한, 위에서도 언급하였듯이 현재까지 결측치 처리기법에 대한 연구는 전통적 처리기법에 비해 완전정보최대우도법이나 혹은 다중대체법 혹은 베이지안 접근법 등이 얼마나 통계적으로 얼마나 우수한 특성을 지닌 모수 추정치를 제공하느냐하는데 초점이 맞추어져왔다 (Yuan, Yang-Wallentin, & Bentler, 2012; Larsen, 2011; Olinsku, Chen, & Harlow, 2003; Gold & Bentler, 2000). 그러나, 선행 연구들도 세 가지 기법들 간 상대적 성능에 대한 비교는 수행하지 않았고, 대부분의 연구들이 회귀분석의 맥락에서 진행되었으며, 특히 데이터와 모형간 적합도를 판단하는 기준이 되는 통계량들에 대한 성능을 추정하고 비교한 연구는, 특히 공분산구조모형의 맥락에서는 존재하지 않는다.

공분산구조모형의 적용 과정은 다른 통계모형과 달리 자료와 모형간의 적합도 지수를 계산하고 이를 바탕으로 모형의 적절성을 판단하는 과정을 포함한다. 완전정보최대우도법에서는 카이제곱 통계량을 기준으로 다양한 적합도 지수가 파생되어 제안되었으며, 다중대체법은  $D_2$ 와  $D_3$  통계량이, 그리고 베이지

안 접근법에서는 사후예측모형검증 (posterior predictive model checking, PPMC) 기법이 사용된다.

최대우도법의 카이제곱 통계량은 다양한 맥락에서 독립적으로 연구된 결과가 있으나, 다중대체법 및 베이지안 접근법에서 제안된 데이터-모형 적합도 평가 방법은 공분산 구조모형의 맥락에서 그 성능이 평가된 적이 없으며 서로 비교된 적도 없다. 따라서, 본 논문에서는 구조방정식모형의 자료-모형간 적합성 판단에 있어서 세 가지 다른 결측치 처리 방법이 어떤 영향을 미치는지 비교 분석하였다.

구체적으로, 본 논문에서는 모의실험 (simulation) 기법을 사용하여, 모집단에서 데이터가 종단적 측정 불변성을 지지하지 않는 경우를 가정하고, 옳은 모형을 사용한 경우의 제 1 종 오류율과 부분측정불변성 (partial measurement invariance)을 가정한 모형을 적용한 경우의 검정력 (power)을 최대우도, 다중대체, 그리고 베이지안 접근법 별로 추정/비교하였다.

## 공분산구조모형에 대한 이론적 배경

### 공분산구조모형의 모수 추정과 모형 평가: 최대우도법

공분산구조모형의 적용에는 일반적으로 다음과 같은 가정을 전제로 한다. 우선,  $N$ 과  $p$ 를 각각 표본의 크기와 변인의 수라고 표기할 때, 크기가  $N \times p$ 인 자료행렬  $Y$ 는 평균벡터와 분산-공분산행렬이 각각  $\mu_0$ 와  $\Sigma_0$ 인 다변량정규분포로부터 추출되었다고 가정한다. 또한, 크기가  $q \times 1$ 인 모수벡터  $\vartheta_0$ 가 존재하고, 이 벡터와 연구자가 제안한 공분산구조모형을 이용하면, 모형-기반 평균벡터  $\mu(\vartheta_0)$  및 분산-공분산행렬  $\Sigma(\vartheta_0)$ 이  $\mu_0$  및  $\Sigma_0$ 와 각각 정확히 일

치한다고 가정한다. 공분산구조모형을 이용한 통계적 분석의 최종목적은 위 조건을 만족하는 모수 벡터  $\vartheta_0$ 를 추정하는 것이다. 이러한 모수는 자료행렬로부터 표본평균벡터  $\bar{y}$  및 표본분산공분산행렬  $S$ 를 구한다음,  $\mu(\vartheta_0)$ 와  $\Sigma(\vartheta_0)$ 이  $\bar{y}$ 와  $S$ 에 가능한 근접하도록 하는  $\vartheta_0$  값을 찾음으로써 추정한다.

다변량정규분포를 가정한 경우 아래 제시된 로그우도(log-likelihood)함수를 최대로 하는  $\vartheta_0$  값을 계산함으로써 최대우도추정치를 구할 수 있다.

$$\ell(\vartheta|Y) \propto -\frac{N}{2} \{ \ln|\Sigma(\vartheta)| + \text{tr}[S\Sigma^{-1}(\vartheta)] + [\bar{y} - \mu(\vartheta)]' \Sigma^{-1}(\vartheta) [\bar{y} - \mu(\vartheta)] \}$$

연구자가 제안한 공분산구조모형이 데이터 생성모형과 일치 할 경우 최대우도추정치는 표본의 크기가 커짐에 따라 점점 편향이 줄고, 더욱 일관적이고, 효율적으로 됨이 알려져 있다. 또한 최대우도추정치는 표본의 크기가 커짐에 따라 평균이  $\vartheta_0$ 이고 분산-공분산행렬이  $\Gamma^{-1}(\vartheta_0)$ 인 다변량정규분포에 접근함이 잘 알려져 있다(Van der Vaart, 2000).  $I(\vartheta_0)$  행렬은 관찰정보(observed information)행렬로 알려져 있으며 아래에 제시된 바와 같이 로그우도함수의 2차편미분을 통해 계산할 수 있다.

$$I(\vartheta) = -\frac{1}{N} \frac{\partial^2 \ell(\vartheta|\bar{y}, S)}{\partial \vartheta \partial \vartheta'}$$

공분산구조모형은 본질적으로, 관찰된 데이터를 생성한 확률모형에 대한 연구자의 가설을 반영하는 것이므로, 모수의 추정뿐만 아니

라, 데이터와 모형간 적합도를 계산하고 평가하는 과정을 포함한다. 아래 제시된 검정통계량은 우도비 검정에 이용되는 통계량으로 연구자가 제안한 구조방정식모형에 기반한 평균벡터와 분산-공분산행렬이 데이터에 기반한 평균벡터와 분산-공분산행렬과 얼마나 유사한지를 평가한다.

$$T_{ML} = -2[\ell(\hat{\vartheta}|Y) - \ell(\bar{y}, S|Y)]$$

다변량정규분포가정과 제안된 공분산구조모형이 생성모형과 일치할 경우, 표본의 크기가 커짐에 따라  $T_{ML}$ 은 자유도가  $d = p + p(p+1)/2 - q$ 인 카이제곱분포에 근접하는 것으로 알려져 있다. 다양한 모형평가지수들이 위의 카이제곱통계량을 바탕으로 제안되었으며(Browne & Cudeck, 1993; Bentler & Bonett, 1980; Tucker & Lewis, 1976) 그 성능 역시 다양한 맥락에서 평가되었다(Hu & Bentler, 1999).

만약 MCAR 혹은 MAR 혹은 어떤 다른 이유로 데이터에 결측치가 발생할 경우, 위에 제시된 로그우도함수는 계산이 불가능하며, 따라서 이 함수를 최대화하기 위해 개발된 기법들은 사용할 수가 없게 된다.

본 논문에서는 확률모형에 기반한 세 가지 결측치 처리 기법을 소개하고 공분산구조모형의 적용에 있어서 가장 중요한 과정 중 하나인 데이터-모형 적합도 평가에 있어서 이 세 가지 방법의 성능을 추정 및 비교하였다.

이후부터는 결측치와 관찰치를 구분하기 위해서 다음과 같은 표기를 사용하기로 한다.  $y_i'$ 를 크기가  $N \times p$ 인 자료행렬  $Y$ 의  $i$  번째

행이라고 할 때,  $y'_{i,obs}$  와  $y'_{i,mis}$  는 각각  $i$  번째 행의 관찰치와 결측치를 나타내기로 한다. 즉,  $y'_i = (y_{i,obs}, y_{i,mis})$  를 나타낸다. 그리고, 데이터 행렬  $Y$ 에서 관찰된 데이터만 모아 놓은 것을  $Y_{obs}$ 로 표기하기로 한다. 즉,  $Y_{obs} = y'_{i,obs} (i = 1, 2, \dots, N)$ 를 나타낸다.

### 완전정보최대우도법

앞에서도 언급했듯이, 완전정보최대우도법은 현재 공분산구조모형을 사용하는 연구자로서는 가장 사용하기 쉽고 널리 알려진 방법이다. 이 방법은 결측치를 포함하는 관찰된 데이터의 로그우도함수를 계산할 때, 데이터 행렬  $Y$ 의 개별 행마다 로그우도함수를 각각 구한 다음 이들의 합을 관찰된 데이터  $Y_{obs}$ 의 로그우도함수로 정의하고 이를 최대화하는 방식으로 모수 추정치를 구한다. 데이터의  $i$ 번째 행에 대한 로그우도함수는 다음과 같이 정의된다.

$$\log L(\vartheta | y_{i,obs}) \propto -\frac{1}{2} \{ \ln |\Sigma_{i,obs}(\vartheta)| + r'_{i,obs} \Sigma_{i,obs}^{-1}(\vartheta) r_{i,obs} \}$$

여기서  $r_{i,obs}(\vartheta) = y_{i,obs} - \mu_{i,obs}(\vartheta)$ 를 가리키고  $\mu_{i,obs}(\vartheta)$  및  $\Sigma_{i,obs}(\vartheta)$ 는 모형-기반 평균 벡터  $\mu(\vartheta)$  및 분산-공분산행렬  $\Sigma(\vartheta)$ 에서 관찰치  $y'_{i,obs}$ 에 해당하는 부분 행렬을 가리킨다<sup>2)</sup>. 관찰된 데이터 전체의 로그우도함수는 아래와 같이 계산한다.

$$\ell(\vartheta | Y_{obs}) \propto \sum_{i=1}^N \log L(\vartheta | y_{i,obs})$$

따라서 완전정보최대우도 추정치는 위 함수를 최대화하는 값으로 정의되며  $\check{\vartheta}$ 로 표기할 수 있다. 결측치가 없을 경우 위에서 계산된 결과는 최대우도법의 로그우도함수와 동일하며 따라서  $\check{\vartheta}$ 와  $\hat{\vartheta}$ 은 같아진다.

완전정보최대우도 추정치의 관찰정보행렬은 위 완전정보-로그우도함수의 2차편미분을 통해 계산할 수 있고 아래에서와 같이 주어진다.

$$\mathcal{J}(\vartheta) = -\frac{1}{N} \sum_{i=1}^N \frac{\partial^2 \log L(\vartheta | y_i)}{\partial \vartheta \partial \vartheta'}$$

위 행렬의 역행렬을 계산하면  $\check{\vartheta}$  분산-공분산행렬의 추정치를 얻을 수 있다.

앞에서도 언급했듯이, 데이터에 결측치가 존재할 경우 데이터 행렬  $Y$ 의 평균 벡터 및 분산-공분산 행렬은 구할 수 없고 따라서 모형의 적합도 평가를 위해 사용되는  $T_{ML}$  통계량 역시 계산할 수가 없다. 그러나, 데이터 행렬  $Y$ 의 평균벡터 및 분산-공분산행렬 자체를 포화 모형(saturated model)으로 간주하고 관찰된 데이터  $Y_{obs}$ 에 기초한 완전정보-로그우도함수를 최대화함으로써 이들을 추정할 수 있다. 이렇게 추정된 평균벡터 및 분산-공분산 행렬을 각각  $\check{\mu}$  그리고  $\check{\Sigma}$ 으로 표기하면 다음에 제시된 통계량은 제안된 공분산구조모형의 적합도를 평가할 수 있는 검정통계량으로 이용될 수 있다.

2) 더 자세한 설명과 예들은 Arbuckle(1996)에서 찾을 수 있다.

$$T_{FIML} = -2[\ell(\hat{\vartheta}|Y) - \ell(\tilde{\mu}, \tilde{\Sigma}|Y)]$$

다변량정규분포가정이 성립하고 제안된 공분산구조모형이 데이터의 생성모형과 일치할 경우, 표본의 크기가 커짐에 따라  $T_{FIML}$ 은 자유도가  $d$ 인 카이제곱분포를 따르는 것으로 알려져 있다.

### 베이지안 접근법

베이지안 접근법은 모든 통계적 추론을 사후분포(posterior distribution)에 바탕을 둔다. 사후분포는 모수의 선험적/사전 분포(prior distribution)와 데이터를 얻은 후 계산할 수 있는 우도함수를 베이지 정리를 이용하여 결합함으로써 얻을 수 있다.

$$p(\vartheta|Y) \propto p(\vartheta)L(\vartheta|Y)$$

모수에 대한 선험적 분포  $p(\vartheta)$ 는 연구자가 지닌 모수에 대한 사전 정보의 수준에 따라 정보적(informative) 사전 분포를 이용할 수도 있고, 비정보적(noninformative) 사전분포를 사용할 수도 있다. 우도함수  $L(\vartheta|Y)$ 는 최대우도법에서 사용되는 우도함수와 동일하다. 따라서, 비정보적 사전분포를 사용할 경우나, 표본의 크기가 매우 커서 우도함수가 사전분포에 비해 월등히 높은 정보를 제공할 경우 사후분포는 실질적으로 우도함수에 의해 결정된다고 볼 수 있다. 따라서 이러한 조건하에서는 최대우도함수법과 베이지안 접근법에서 구한 모수 추정치가 거의 동일한 특성을 가진다는 것이 잘 알려져 있다(Gelman, Carlin, Stern, Dunson, Vehtari, & Rubin, 2013).

그러나 베이지안 접근법의 한 가지 중요한 특징 중 하나는 모수에 대한 점추정치와 표준오차를 구하는 것 뿐 아니라, 데이터가 제공하는 정보를 바탕으로 모수가 취할 수 있는 값의 분포 전체를 확률적으로 계산하는데 있다. 대부분의 경우 사후분포는 분석적 방법으로 닫힌형 해(closed-form solution)를 구할 수가 없으며, 따라서 자료 확대(data augmentation, Tanner & Wong, 1987), 깃스 표집기(Gibb's sampler, Geman & Geman, 1984) 등의 마르코프사슬-몬테칼로(Markov Chain Monte Carlo, MCMC) 기법을 통한 모의시행(simulation)을 이용하여 사후분포를 추정한다. 이러한 다양한 기법들을 통하여 사후분포로부터 모수값들을 표집하고, 표집된 값들을 이용하여 모수 및 모형에 관한 통계적 분석을 진행하게 된다.

베이지안 접근법의 또다른 중요한 특징으로는 결측치를 처리하는 특별한 장치가 있는 것이 아니라, 결측치 역시 추정해야할 모수로 간주한다는 것이다. 따라서, 결측치가 존재할 경우에는 베이지안 접근법에서는 모수와 결측치의 결합사후확률분포를 계산함으로써 통계적 추론을 진행한다.

베이지안 접근법에서는 데이터와 모형간 적합도 판단 역시 일종의 사후확률분포를 이용한다. 이 때 사용되는 분포는 사후예측확률분포(posterior predictive distribution)로서 제안된 공분산구조모형과 실제 관찰된 데이터  $y_{obs}$ 를 통해 예측한 미래의 관측치  $y^{rep}$ 의 확률분포이며  $p(y^{rep}|y_{obs})$  표기할 수 있다. 만약  $y^{rep}$ 와  $y_{obs}$ 간의 차이가 크다면 제안된 공분산구조모형은 관찰된 데이터를 잘 설명하지 못한다는 증거가 되며 모형을 기각할 때의 근거가 된다.  $y^{rep}$ 와  $y_{obs}$ 간의 차이를 측정하기 위해서는 불일치측정치(discrepancy measure, Gelman, Meng,

& Stern, 1996)를 정의해야 하는데, 연구자가 데이터와 모형간의 적합도를 평가하는데 관심이 있다면 최대우도법에서 정의되었던 로그우도 함수를 그대로 이용하는 것도 한 가지 방법이 될 수 있다. 단, 불일치 측정치는 일반적인 통계량과는 달리 데이터  $Y$ 와 모수  $\vartheta$  두 가지 모두를 변수로 취하는 함수이다. 이를 분명히 나타내기 위하여 불일치 측정치는  $T(Y, \vartheta)$ 으로 표기된다. 불일치 측정치의 참조분포(reference distribution)는 모수와  $y^{rep}$ 의 결합사후확률분포에 의해 결정되며, 아래에 제시된 확률을 계산함으로써 베이지안 p-값을 구할 수 있다.

$$p_B = \Pr(T(y^{rep}, \vartheta) > T(y_{obs}, \vartheta) | y_{obs})$$

즉, 베이지안 p-값은 관찰된 데이터에 기반한 불일치 측정치  $T(y_{obs}, \vartheta)$ 가 제안된 모형에 예측하는 데이터에 기반한 불일치 측정치  $T(y^{rep}, \vartheta)$ 에 비해 얼마나 상대적으로 그 불일치의 정도가 더 큰지를 계산함으로써 얻을 수 있다. 즉, 모수와  $y^{rep}$ 의 결합사후확률분포하에서 계산된  $T(y_{obs}, \vartheta)$ 가  $T(y^{rep}, \vartheta)$ 에 비해 더 클수록 베이지안 p-값은 더 작은 값을 가진다.

최대우도법에서 사용되는 카이제곱통계량의 p-값을 계산하는 과정과의 가장 중요한 차이점은, 베이지안 p-값을 계산할 때에는 모수를 최대우도추정치 등의 특정한 값으로 대체하지 않고 사후분포전체를 이용한다는 것이다. 따라서, 베이지안 p-값은 모수를 추정하는 과정에서 발생할 수밖에 없는 불확실성을 사후분포를 이용하여 통합하는 과정을 포함하게 된다. 만약 모수를 최대우도추정치로 대체한 후 베이지안 p-값을 구한다면 이는 이론적으로 카이제곱의 p-값을 구하는 과정과 동일해진다.

즉, 카이제곱통계량의 p-값은 아래와 같이 표기할 수도 있다.

$$p_C = \Pr(T(y^{rep}, \check{\vartheta}) > T(y_{obs}, \check{\vartheta}))$$

이때,  $T(y^{rep}, \check{\vartheta})$ 는 카이제곱분포를 따르고  $T(y_{obs}, \check{\vartheta})$ 는  $T_{ML}$ 이 됨을 쉽게 보일 수 있다. 사후예측모형검증법의 큰 장점 중 하나는 카이제곱분포와 같은 참조분포를 이론적으로 도출할 필요가 없다는 것이다. 불일치 통계량의 참조분포는 모수와  $y^{rep}$ 의 결합사후확률분포에 의해 자동적으로 결정되기 때문이다(Gelman, Meng, & Stern, 1996). 또한 불일치추정치로서  $\ell(\vartheta | Y_{obs})$ 을 사용할 경우 모형의 적합도 평가를 위해 데이터에 기반한 평균벡터와 분산-공분산행렬을 계산 혹은 추정할 필요가 없다.

### 다중대체법

다중대체법의 기본적인 틀은 결측치를 두 가지 이상의 다른 값으로 대체하는 방식<sup>3)</sup>으로 결측치가 없는 것처럼 보이도록 만들어진 완성된(completed) 혹은 대체된(imputed) 자료들을 생성한 뒤 결측치가 존재하지 않을 때를 가정하고 개발된 기존의 통계적 기법(예: 최대우도법)을 그대로 적용하고, 그 결과들을 적절한 방식으로 재결합하는 데 있다.

공분산 구조 분석의 맥락에서 설명을 하면, 수 개 혹은 수십 개의 대체-자료들(imputed datasets)로부터 각각 평균벡터와 분산-공분산행렬을 구한 뒤, 로그우도함수를 최대화하는 과정을 통해 수 개 혹은 수십 개의 모수추정치

3) 다양한 방식의 대체법에 대한 보다 자세한 설명은 Van Buuren(2012)에서 찾을 수 있다.



를 구한다. 만약 10개의 대체-자료를 생성하였다라고 하면, 연구자가 제안한 공분산구조모형을 10개의 대체-자료에 각각 합치시키는 과정을 거쳐 10 세트의 모수 추정치와 10 세트의 모수 분산공분산 추정행렬, 그리고 10개의 카이제곱값을 얻을 수 있게 된다. 만약, 총  $M$  번의 대체를 실행하였고, 그 중  $m$  번째 대체-자료를  $Y_m$ , 그리고 이로부터 얻어진 모수추정치와 모수분산-공분산행렬을 각각  $\hat{\vartheta}_m$  그리고  $U_m$  이라고 표기한다면, 다중대체법에 기반한 모수 추정치  $\bar{\theta}$ 와 분산공분산행렬의 추정치  $var(\bar{\theta})$ 는 아래 공식을 이용하여 계산한다 (Schafer, 1997).

$$\begin{aligned} \bar{\theta} &= \frac{1}{M} \sum_{m=1}^M \hat{\vartheta}_m \\ var(\bar{\theta}) &= \bar{U} + \frac{M+1}{M} V \\ \bar{U} &= \frac{1}{M} \sum_{m=1}^M U_m \\ V &= \frac{1}{M-1} \sum_{m=1}^M (\hat{\vartheta}_m - \bar{\vartheta})(\hat{\vartheta}_m - \bar{\vartheta})' \end{aligned}$$

우도비검정을 위한 통계량을 통합하는 과정은 약간 더 복잡하며 현재까지 두 가지 방법이 제안되었다. 우선 Meng & Rubin(1992)이 제안한  $D_3$  통계량은  $M$ 개의 우도비통계량의 평균( $\bar{L}_1$ ) 뿐만 아니라, 다중대체법에 의해 추정된 모수추정치들을 대체된 데이터에서 각각 평가한 우도비들의 평균( $\bar{L}_2$ ) 역시 필요로 한다. 즉,  $D_3$ 는 다음과 같이 정의된다.

$$D_3 = \frac{\bar{L}_2}{k(1+r_3)}$$

$$\begin{aligned} r_3 &= \frac{M+1}{k(M-1)} (\bar{L}_1 - \bar{L}_2) \\ \bar{L}_2 &= \frac{1}{M} \sum_{m=1}^M 2\{\ell_1(\hat{\vartheta}_1|Y_m) - \ell_0(\hat{\vartheta}_0|Y_m)\} \\ \bar{L}_1 &= \frac{1}{M} \sum_{m=1}^M 2\{\ell_1(\hat{\vartheta}_1^{(m)}|Y_m) - \ell_0(\hat{\vartheta}_0^{(m)}|Y_m)\} \end{aligned}$$

여기서  $\vartheta_0$ 와  $\vartheta_1$ 의 아래첨자인 0과 1은 주어진 모수가 연구자가 제안한 공분산구조모형과 포화모형에 각각 관련된 것임 지칭한다.  $D_3$  통계량은  $F(d, \nu_3)$ -분포를 따르는 것으로 알려져 있으며 p-값은 아래와 같이 구할 수 있다.

$$\begin{aligned} p_{D_3} &= \Pr(F_{d, \nu_3} \geq D_3) \\ \nu_3 &= 4 + (t-4)[1 + (1-2t^{-1})r_3^{-1}]^2 \\ t &= d(M-1) \end{aligned}$$

$D_2$  통계량은 Li, Meng, Raghunathan, & Rubin (1991)에 의해 제안된 것으로  $D_3$ 에 비해 좀 더 간단히 구할 수 있으며, 공분산구조모형의 맥락에서는 유일하게  $D_2$ 와 그 p-값을 계산해주는 SAS Macro가 개발되어 있다(Allison, 2003).  $D_2$  통계량은 아래와 같이 구할 수 있다.

$$D_2 = \frac{\bar{L}_1/k - (1-M^{-1})\nu}{1 + (1+M^{-1})\nu}$$

여기서  $\nu$ 는  $\sqrt{\ell_1(\hat{\vartheta}_1^{(m)}|Y_m) - \ell_0(\hat{\vartheta}_0^{(m)}|Y_m)}$ ,  $m=1, 2, \dots, M$ 의 분산을 나타낸다.  $D_2$  통계량은  $F(d, b)$ -분포를 따르는 것으로 알려져 있으며 p-값은 아래와 같이 구할 수 있다.

$$\begin{aligned} p_{D_2} &= \Pr(F_{d, b} \geq D_2) \\ b &= d^{-1/M} (M-1) \{1 + [(1+M^{-1})\nu]^{-1}\}^2 \end{aligned}$$

$D_2$ 와  $D_3$  통계량은 공분산구조모형의 맥락에서는 자동적으로 제공되는 통계프로그램이 아직 없으며, 현재까지 그 성능이 연구된 적도, 그리고 다른 통계량과 상대적인 성능이 비교된 적도 없다.

## 방 법

### 결측치 생성을 위한 모형 및 연구 설계

#### 연구대상

표본 자료를 생성하기 위하여 하나의 구성개념(예: 우울증)이 종단적으로 측정되는 상황을 설정하였다. 자료생성모형으로서는 하나의 요인(latent factor)이 5개의 관찰 변인(indicator)에 의해 측정되고 같은 요인이 세 번에 걸쳐 측정되는 3요인 종단적 요인분석을 사용하였다. 5개의 관찰 변인의 요인 계수(factor loading)는 측정 시점마다 변하는 것으로 가정하였다. 즉, 측정 대상이 되는 구성개념의 종단적 측정불변성이 성립하지 않는 상황을 가정하고 이를 따르는 모형으로부터 자료를 생성하였다<sup>4)</sup>. 따라서 총 15개의 관찰 변인( $Y_1, Y_2, \dots, Y_{15}$ )을 가지는 자료가 생성되었다. 생성된 자료로부터 결측치를 생성하기 위해서는

4) 종단연구의 특성을 반영하여 오차항간의 상관도 작은 양의 값을 가지는 것으로 가정하였다. 본 논문의 주요 목적은 결측치 처리 방법에 따른 자료-모형간 적합도 판단을 위한 다양한 통계량의 성능을 추정하고 비교하는데 있으므로, 자료 생성을 위해 사용된 모형에서 지정된 모수값에 대한 자세한 기술은 본문에서 생략하고 부록에 추가하였다. 이에 대한 추가적 정보가 필요한 독자는 언제든지 저자에게 연락하여 구체적인 정보를 제공받을 수 있다.

다음과 같은 절차를 적용하였다. 자료행렬에서  $i$ 번째 개인의  $j$ 번째 변인 점수를  $Y_{ij}$ 라고 표기할 때,

1) 우선 첫 시점에 측정된 5개의 변인들의 개인별 평균 점수

$$\bar{Y}_i = \frac{1}{5}(Y_{i1} + Y_{i2} + Y_{i3} + Y_{i4} + Y_{i5})$$

를 구하고, 그 평균의 1사분위수, 2사분위수, 3사분위수를 구한다.

2) 위에서 구한 개인별 평균점수  $\bar{Y}_i$ 가 1사분위수보다 작은 개인들 중 임의적으로 40%를 추출한 뒤 여기에 해당하는  $Y_{i6}$ 의 값을 결측치로 변환한다. 그리고 동일한 과정을 나머지 9개의 변인  $Y_{i7}, \dots, Y_{i15}$ 에 대해 각각 적용하여 40%의 결측치를 각각의 변인으로부터 생성한다.

3) 다음으로, 개인별 평균점수  $\bar{Y}_i$ 가 1사분위수보다 크고 2사분위수보다 작은 값을 가지는 개인들 중 10%를 임의로 추출하여 해당  $Y_{i6}$ 의 관찰값을 결측치로 변환하고,  $Y_{i7}, \dots, Y_{i15}$  역시 동일한 과정을 개별적으로 거쳐 각각의 10%를 결측치로 변환하였다.

4) 개인별 평균 점수  $\bar{Y}_i$ 가 2사분위수 보다 크고 3사분위수 보다 작은 개인들은 3)과 동일한 과정을 거쳐 10%의 결측치를 생성하였고, 개인별 평균점수  $\bar{Y}_i$ 가 3사분위보다 큰 개인들은 2)와 동일한 과정을 거쳐 40%의 결측치를 생성하였다. 위에 기술된 방법에 따라 결측치를 생성할 경우 전체자료의 25%가 결측치로 변환됨을 알 수 있다<sup>5)</sup>.

5) 각 구간별 결측치 비율을 10%, 40%, 40%, 10% 혹은 10%, 20%, 30%, 40% 등으로 변화시킨 추가 연구에서도 본 논문에서 보고된 결론과 동일한 결론을 내릴 수 있었음을 미리 밝힌다.

각 구간별 결측치 비율을 60%, 20%, 20%, 60%로 변화시켜 전체자료의 40%가 결측치로 변환되게 한 조건, 그리고 각 구간별 결측치 비율을 15%, 5%, 5%, 15%로 변화시켜 전체자료의 10%가 결측치로 변환되게 한 조건 등을 추가한 연구에서도 본 논문에서 보고된 결론과 동일한 결론을 내릴 수 있었음을 미리 밝히며 그 결과는 부록의 표에 정리되어 있다.

위에서 기술한 결측치 생성방법은 조건무선적(MAR) 과정을 따른 것임을 쉽게 알 수 있다. 즉, 관찰값이 결측치로 변환되느냐 여부는 관찰된 데이터  $Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4}, Y_{i5}$ 의 값에 따라 결정되었으며, 주어진 조건 내에서 결측치의 발생은 완전무선적이다. 이러한 과정을 거쳐 생성된 데이터는 줄지우기 혹은 결레지우기 기법 등의 전통적 방법들을 이용하여 분석할 경우 편향된 모수 추정치를 얻을 수 있고, 자료와 모형간 적합도에 대한 판단 역시 잘못될 수 있다는 것이 잘 알려져 있다.

### 측정

본 논문의 주요한 목적 중 하나인 완전정보 최대우도법, 다중대체법, 그리고 베이지안 점근법에서 각각 제공되는 데이터-공분산구조모형 간 적합도 판단 통계량의 성능 추정 및 비교를 위해 다음과 같은 절차와 기준을 사용하였다.

연구자가 자료생성모형과 동일한 공분산구조모형을 제안하였을 경우에는 적합도 판단에 사용된 통계량은 연구자가 제안한 모형이 데이터를 잘 설명한다는 결론을 내리도록 하는 정보를 제공하여야 한다. 만약, 그렇지 않고, 모형에 문제가 있다는 결론을 내리도록 유도한다면, 연구자는 제안한 모형을 기각하게 됨으로써 제 1 종 오류를 범하게 된다. 따라서,

생성모형으로부터 추출된 가상 표본을 대상으로 생성모형과 동일한 모형을 적합시킴으로써, 세 가지 결측치 처리 방법이 제공하는 적합도 통계량, 즉 카이제곱통계량, 베이지안  $p$ -값,  $D_2$ ,  $D_3$  통계량의 제 1 종 오류율을 계산하고 비교함으로써 연구를 진행하였다.

베이지안  $p$ -값을 제외한 다른 통계량은 모두 점근적(asymptotic) 검정통계량으로서 그 성능이 표본의 크기( $N$ )와 밀접한 관련이 있다. 따라서, 표본의 크기를 점진적으로 증가시킴으로써( $N=100, 200, 350, 500, 1000$ ) 그 성능을 추정/비교하였다. 다중대체법은 대체를 몇 번 실행하느냐, 즉 다중대체를 통해 생성된 대체자료의 수( $M$ )가 최종적인 통계적 추론에 영향을 미치는 것으로 알려져 있다. 따라서 대체의 수 역시 점진적으로 증가시킴으로써( $M=5, 10, 20, 50, 100$ ) 성능을 추정하였다.

베이지안의 모형적합도 검정에 사용되는 베이지안  $p$ -값은 제안된 모형이 생성모형과 동일할 경우에도 균등(uniform) 분포를 따르지 않고 그 값들이 .5를 향하여 좀 더 밀집하여 있는 것으로 알려져 있다(Robins, van der Vaart, & Ventura, 2000). 따라서 일반적으로 사용되는 명목기각률을 사용하여 모형의 기각 여부를 결정할 경우 매우 낮은 기각률을 보이는 것으로 알려져 있다(Dahl, 2006). 그러나 공분산구조모형의 맥락에서는 베이지안  $p$ -값 분포의 특성이 연구된 적이 없다. 따라서, 본 연구에서는 베이지안  $p$ -값의 분포를 공분산구조모형의 맥락에서 탐색하는데 주안점을 두었고, 명목기각률을 사용하여 제 1 종 오류를 검증하는 것에는 큰 의미를 두지 않았다. 그보다는, 명목기각률에 근접하는 경험기각률을 보이는 베이지안  $p$ -값을 탐색하고 대강의 지침을 제공하는 것에 더 큰 중점을 두었다. 그리고 본

표 1. 결측치 처리방법에 따른 제 1 종 오류율

	표본 크기 = 100			표본 크기 = 200			표본 크기 = 350			표본 크기 = 500			표본 크기 = 1000		
	명목 기각률 ( $\alpha$ )														
	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
카이자승	0.09	0.26	0.39	0.03	0.13	0.23	0.02	0.09	0.16	0.02	0.08	0.13	0.02	0.06	0.12
불일치 측정치	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.00	0.01	0.00	0.00	0.01
대체수(M) = 5															
$D_3$	0.07	0.16	0.24	0.04	0.12	0.19	0.02	0.09	0.14	0.02	0.09	0.15	0.02	0.06	0.11
$D_2$	0.08	0.12	0.14	0.06	0.10	0.15	0.05	0.10	0.12	0.05	0.09	0.12	0.04	0.08	0.13
대체수(M) = 10															
$D_3$	0.06	0.15	0.26	0.04	0.13	0.21	0.01	0.08	0.15	0.02	0.09	0.15	0.02	0.07	0.12
$D_2$	0.07	0.14	0.18	0.08	0.17	0.25	0.07	0.15	0.21	0.06	0.12	0.20	0.07	0.15	0.22
대체수(M) = 20															
$D_3$	0.06	0.18	0.28	0.04	0.13	0.21	0.02	0.09	0.16	0.02	0.09	0.14	0.02	0.07	0.12
$D_2$	0.14	0.25	0.33	0.11	0.24	0.34	0.08	0.18	0.28	0.06	0.17	0.26	0.05	0.15	0.22
대체수(M) = 50															
$D_3$	0.06	0.17	0.28	0.04	0.13	0.22	0.02	0.09	0.16	0.02	0.09	0.15	0.02	0.07	0.12
$D_2$	0.21	0.38	0.48	0.10	0.24	0.34	0.06	0.19	0.28	0.05	0.15	0.25	0.04	0.12	0.18
대체수(M) = 100															
$D_3$	0.05	0.18	0.29	0.04	0.13	0.22	0.02	0.09	0.17	0.02	0.08	0.15	0.02	0.07	0.12
$D_2$	0.24	0.42	0.55	0.08	0.24	0.34	0.06	0.16	0.25	0.05	0.14	0.23	0.03	0.09	0.16

연구에서 찾아진 대강의 지침을 이용하여 검정력을 추정하고 비교하였다.

제 2 종 오류는 연구자가 제안한 모형이 자료생성모형과 다름에도 불구하고 제안된 통계 모형이 자료와 적합하다는 결론을 내릴 때 발생한다. 이 경우 적합도 판단을 위해 사용되는 검정통계량이 제안된 모형이 자료를 잘 설명하지 못한다는 결론을 내리도록하는 정보를 제공한다면 그 검정통계량은 검정력이 높은 통계량이라 불린다.

본 연구에서는 여러 통계량들의 검정력을

비교하기 위하여, 실제 자료생성모형은 종단적측정불변성을 따르지 않지만, 연구자가 부분측정불변성이 성립하는 모형을 제안한 상황을 설정하였다. 즉, 요인을 측정하는 5개의 측정 변인 중 3개는 종단적측정불변성이 성립한다는 가정을 하지만, 나머지 2 변인은 그렇지 못한 것으로 가정하는 모형을 제안하는 상황을 설정하고, 여러 검정통계량들의 검정력을 비교하였다<sup>6)</sup>.

6) 이 모형은 결과적으로 .04 정도의 RMSEA를 보이는데 이 값은 자료생성모형에서 오차항간의 상

검정통계량의 검정력은 표본의 크기와 밀접 관련이 있는 것으로 알려져 있으므로, 표본의 크기를 점진적으로 증가시키면서 검정력을 비교하였다. 다중대체법에서는 대체의 수를 증가시켜 검정력에 미치는 영향을 탐색하였다. 베이지안기법의 검정력은 제 1 종 오류의 비교 연구에서 추정된 베이지안 p-값을 사용하여 모형의 기각률을 계산함으로써 다른 기법들에서 사용되는 검정 통계량과의 검정력을 비교하였다.

## 결 과

### 완전정보최대우도법의 카이자승통계량

그림 1과 그림 2는 각각 카이자승통계량의 제 1 종 오류율과 검정력을 보여준다. 우선 카이자승통계량의 이론적 특성과 일치하게도 표본의 크기에 따라 경험기각률이 명목기각률  $\alpha$ 에 점점 근사함을 알 수 있다. 즉, 표본의 크기가 100에서부터 200, 350, 500, 1000으로 증가함에 따라, 경험 기각률이 .256에서 각

관의 정도 및 요인계수를 변화시킴으로써 얻어진 결과이다. 이 정도 수준의 적합도를 가지도록 생성모형의 모수 및 제한된 모형의 부분측정불변성의 정도를 결정한 이유는 RMSEA=.04 정도 수준의 적합도를 지닌 모형을 사용할 경우에 여러 가지 통계량의 검정력을 비교하기가 더 용이하기 때문이다, 만약 RMSEA가 .10 정도 되었다면, 대부분의 조건에서 모든 통계량의 검정력이 1.0에 가까워져 비교 자체가 어려워진다. 참고로 오차항간의 상관의 정도, 요인 계수 및 부분측정불변성의 정도를 변화시킴으로써 실행한 추가 연구에서 얻은 결론 역시 본 논문에서 보고된 결론과 동일하였음을 미리 밝힌다.

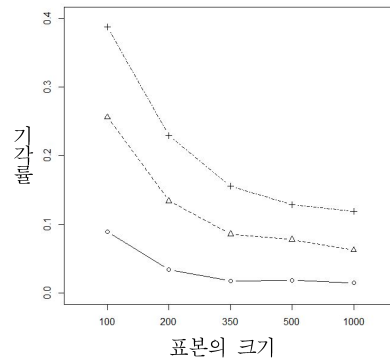


그림 1. 카이자승통계량의 제 1 종 오류, ○:  $\alpha = .01$ , △:  $\alpha = .05$ , +:  $\alpha = .10$ .

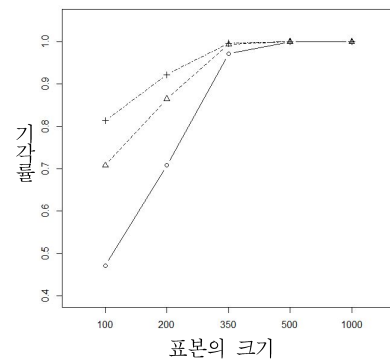


그림 2. 카이자승통계량의 검정력, ○:  $\alpha = .01$ , △:  $\alpha = .05$ , +:  $\alpha = .10$

각 .134, .086, .078, .063으로 명목기각률인  $\alpha = .05$ 에 점점 더 가까워짐을 알 수 있다. 명목기각률이 .01 그리고 .10일 때의 결과 역시 표본의 크기가 커짐에 따라 경험기각률이 명목기각률에 근사한 성능을 보임이 확인되었다. 자세한 수치는 표 1에서 확인할 수 있다.

그림 2에서 알 수 있듯이, 카이자승통계량의 검정력 역시 표본의 크기에 따라 증가함을 알 수 있다. 구체적으로는 명목기각률을 .05에 고정시켰을 때, 표본의 크기가 100에서 200, 350, 500, 1000 으로 증가함에 따라, 모형을 기

각할 확률이, .708에서 .865, .993, .988, 1.000으로 각각 증가하였다. 따라서, 카이제곱통계량은 표본의 크기가 작을 때에도 충분한 검정력을 갖춘 것으로 보인다. 그러나 이러한 결과는 조심스럽게 해석되어야 하는데 그 이유는, 표본의 크기가 작을 때는 제 1 종 오류가 명목기각률에 비해 상당히 높은 값을 보였기 때문이다. 즉, 표본의 크기가 작을 때에는 카이제곱통계량은 제안된 모형을 기각하는 비율을 증가시켜 검정력을 일정 수준으로 높이기도 하지만 제 1 종 오류율 역시 높아지는 경향을 보였다.

### 다중대체법의 $D_3$ 와 $D_2$ 통계량

그림 3과 4는 명목기각률이 .05일 때,  $D_3$ 와  $D_2$  통계량의 제 1 종 오류율을 보여준다. 그림에서 알 수 있듯이, 두 통계량 모두 표본의 크기가 커짐에 따라 경험기각률이 명목기각률 .05에 접근하는 것을 알 수 있다.

한 가지 주목할 만한 특성은  $D_3$  통계량은 다중대체의 수(M)와 크게 관계없이 표본의 크기에 따라 제 1 종 오류율이 변화하는데 반해,  $D_2$  통계량은 다중대체의 수가 제 1 종 오류율에 상당한 영향을 미치는 것으로 드러났다. 특히 표본의 수가 작을수록 다중대체의 수가 경험기각률과 상당한 관련이 있음이 드러났다. 한 가지 재미있는 결과는, 명목기각률이 .05로 고정되었을 경우, 표본의 수가 100일 때 M이 5에서 10, 20, 50, 100으로 증가함에 따라 제 1 종 오류율이 .119에서 .137, .247, .376, .421로 증가하는 양상을 보였다. 명목기각률이 .01 혹은 .10으로 고정되어 있을 경우에도  $D_2$ 의 제 1 종 오류율은 비슷한 양상을 보였다. 즉, 표

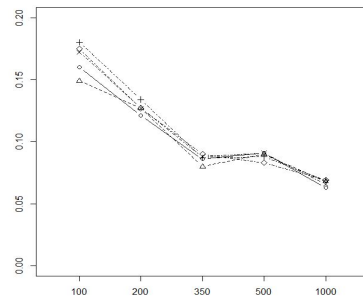


그림 3.  $D_3$  통계량의 제 1 종 오류,  $\alpha = .05$ ,  $\circ$ :  $M=5$ ,  $\triangle$ :  $M=10$ ,  $+$ :  $M=20$ ,  $\times$ :  $M=50$ ,  $\diamond$ :  $M=100$ .

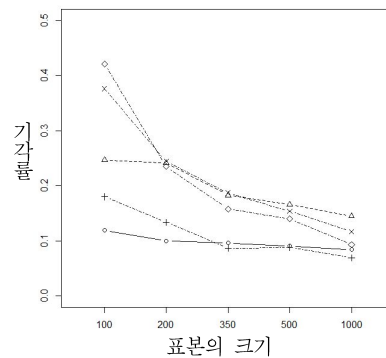


그림 4. 다중대체  $D_2$  통계량의 제 1 종 오류,  $\alpha = .05$ ,  $\circ$ :  $M=5$ ,  $\triangle$ :  $M=10$ ,  $+$ :  $M=20$ ,  $\times$ :  $M=50$ ,  $\diamond$ :  $M=100$ .

본의 크기가 커짐에 따라 전반적으로 경험기각률은 명목기각률에 접근하는 양상을 보이나, 주어진 표본의 크기가 고정되어 있을 경우 대체의 수가 증가함에 따라  $D_2$  통계량의 제 1 종 오류율이 증가하였다. 구체적인 기각률은 표 1에서 확인할 수 있다.

이러한 결과는 다중대체법을 사용하여 결측치를 처리한 뒤 제안된 공분산구조모형의 적합도를 판단하는 경우 현재 가장 쉽게 구할 수 있는  $D_2$  통계량에 의존한 판단은 제 1 종

오류율을 높일 가능성을 시사하며, 특히 대체의 수가 증가함에 따라 오류를 범할 확률이 더욱 커짐을 의미한다.

일반적으로 대체의 수가 증가함에 따라 다중대체법은 여러 가지 면에서 바람직한 통계적 결과를 산출하는 것으로 알려져 있는데 (Graham, Olchowki, & Gilreath, 2007; Bodner, 2008) 본 연구의 결과는 이러한 일반적 결과가  $D_2$  통계량에는 적용되지 않는 것으로 드러났다. 따라서 연구자들은 공분산구조모형 분야에서 현재 유일하게 이용 가능한  $D_2$  통계량을 이용하여 모형을 판단하는데 매우 조심해야 할 것이다.

그림 5와 그림 6은 명목기각률을 .05로 고정하였을 때  $D_3$ 와  $D_2$  통계량의 검정력을 보여준다.  $D_3$ 와  $D_2$  통계량 모두 표본의 크기가 증가함에 따라 검정력이 증가하는 양상을 보였으나, 표본의 크기가 고정되었을 때, 대체의 수에 따른 검정력에 관해서는 매우 다른 양상을 보였다. 즉,  $D_3$  통계량은 대체의 수가 증가함에 따라 검정력이 증가하는 패턴을 보이긴 하지만, 그 증가의 양이 크지 않고, 가장 큰 차이를 보이는 조건인  $N=100$ 의 경우에도  $M=5$ 와 100의 차이가 .10 미만이었다.

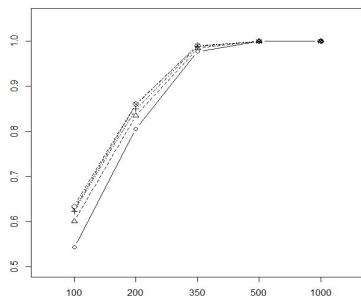


그림 5.  $D_3$  통계량의 검정력;  $\alpha = .05$ ,  
 $\circ$ :  $M=5$ ,  $\triangle$ :  $M=10$ ,  $+$ :  $M=20$ ,  
 $\times$ :  $M=50$ ,  $\diamond$ :  $M=100$ .

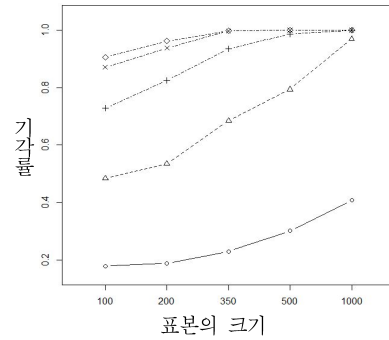


그림 6. 다중대체  $D_2$  통계량의 검정력;  
 $\alpha = .05$ ,  $\circ$ :  $M=5$ ,  $\triangle$ :  $M=10$ ,  
 $+$ :  $M=20$ ,  $\times$ :  $M=50$ ,  $\diamond$ :  $M=100$ .

반면  $D_2$  통계량은 대체의 수가 작을 경우 표본의 크기가 1000일 경우에도 충분한 검정력을 확보하지 못하는 것으로 드러났다. 예를 들면,  $M=5$ 이고  $N=1000$ 일 경우에도 검정력은 .408에 머물렀다. 대체의 수가 특히 5나 10일 경우에는 표본의 크기가 1000 정도로 상당히 크지 않을 경우에는 검정력이 매우 떨어지는 것으로 확인되었다.

표 2에서 알 수 있듯이  $D_2$ 을 이용할 때 충분한 검정력을 확보하기 위해서는 대체의 수를 최소한 20 이상으로 늘리는 것이 필요한 것으로 보인다. 그러나 이는  $D_2$  통계량이 보인 제 1종 오류의 양상을 고려한다면 조심스럽게 결정할 문제이다. 즉  $D_2$  통계량은 대체의 수가 증가함에 따라 제 1종 오류 역시 증가하는 패턴을 보였기 때문이다.

이러한 결과에 비추어 본다면, 대체의 수를 증가시키므로써 얻을 수 있는 장점들(예, 모수 추정치의 정확성 증가)을 유지하면서 모형판단의 정확성을 증가시키기 위해서는  $D_2$  보다는  $D_3$ 를 사용하는 것이 더욱 바람직해 보인다.

표 2. 결측치 처리방법에 따른 검정력

	표본 크기 = 100			표본 크기 = 200			표본 크기 = 350			표본 크기 = 500			표본 크기 = 1,000		
	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
명목 기각률 ( $\alpha$ )															
카이자승	0.48	0.71	0.81	0.71	0.87	0.92	0.97	0.99	1.00	0.92	0.99	1.00	1.00	1.00	1.00
불일치 추정치	0.19	0.44	0.55	0.49	0.69	0.77	0.96	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00
대체수(M) = 5															
$D_3$	0.30	0.54	0.66	0.58	0.81	0.88	0.92	0.98	0.99	0.99	1.00	1.00	1.00	1.00	1.00
$D_2$	0.11	0.18	0.24	0.11	0.19	0.26	0.13	0.23	0.31	0.19	0.30	0.39	0.28	0.41	0.52
대체수(M) = 10															
$D_3$	0.35	0.60	0.73	0.64	0.84	0.90	0.95	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00
$D_2$	0.28	0.48	0.61	0.34	0.54	0.66	0.47	0.69	0.79	0.59	0.79	0.89	0.86	0.97	0.99
대체수(M) = 20															
$D_3$	0.36	0.62	0.74	0.67	0.85	0.91	0.96	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$D_2$	0.55	0.73	0.82	0.65	0.83	0.89	0.82	0.94	0.97	0.95	0.99	1.00	1.00	1.00	1.00
대체수(M) = 50															
$D_3$	0.38	0.63	0.75	0.68	0.86	0.92	0.96	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$D_2$	0.72	0.87	0.92	0.83	0.94	0.97	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
대체수(M) = 100															
$D_3$	0.38	0.63	0.76	0.68	0.86	0.92	0.96	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$D_2$	0.79	0.91	0.95	0.88	0.96	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

베이지안 접근법: 베이지스 p-값

베이지안 접근법에서 사용되는 예측사후모형평가법에서 계산되는 베이지스 p-값은 빈도론자의 p-값과 달리 균일분포를 따르지 않고 많은 값들이 .5주변에 분포되어 있는 특성을 가지고 있다. 이러한 특성으로 베이지스 p-값을 이용하여 모형의 적합도를 판단하기 위해 명목기각률을 사용하는 것은 적절한 방법이 아니다. 본 연구에서 사용된 공분산구조모형의 적합도 판단에서도 표 1과 그림 8에서도 볼 수 있듯이, .01, .05, .10의 명목기각률을 사용하여 모

형의 제 1 종 오류율을 계산하면, 모든 조건에서 .01의 기각률을 넘지 않는 것을 알 수 있다. 이 결과만 놓고 본다면 베이지스 p-값을 이용한 모형적합도 판단은 제 1 종 오류율이 매우 낮은 것으로 판단할 수 있다. 그림 7은 제안된 공분산구조모형이 자료생성모형과 일치할 경우 베이지스 p-값의 분포를 보여준다. 그림에서 확인할 수 있듯이 .5 주변에 훨씬 많은 p-값들이 분포되어 있음을 알 수 있다.

사실 베이지안 통계의 관점에서는 베이지스 p-값을 빈도론자들이 사용하는 명목기각률을 기준으로 판단해야 할 이유는 없다. 왜냐하면,



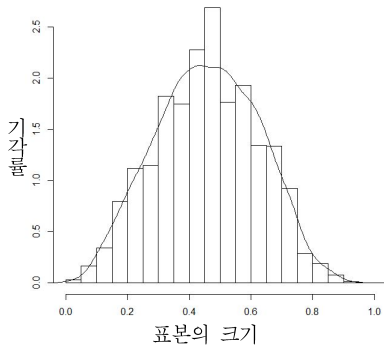


그림 7. 베이지스 p-값의 분포

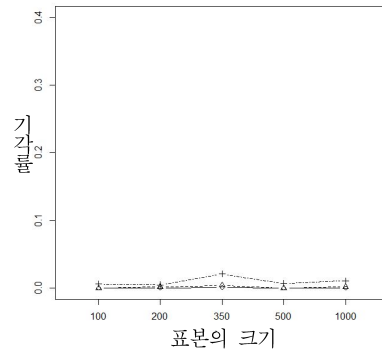


그림 8. 베이지스 접근법 제 1 종 오류,  
○:  $\alpha = .01$ ,  $\Delta$ :  $\alpha = .05$ , +:  $\alpha = .10$ .

베이지안의 입장에서는 어떤 사건의 확률을 반복 시행에서의 상대빈도로 해석할 필요가 없고, 확률값을 주관적으로 해석하는 것에 아무런 문제가 없기 때문이다. 그러나, 공분산구조모형을 실제 연구에 사용하고 빈도론자의 통계기법에 더욱 익숙하며 모형의 적합도를 판단해야하는 연구자의 입장에서는 베이지스 p-값이 제공되었을 때 이를 판단할 수 있도록 하는 대강의 지침이 있는 것이 실질적 도움이 되는 것도 사실이다. 따라서 본 논문에서는 베이지스 p-값이 어느 정도 작은 값을 가질 때, 일반적으로 사용되는 명목기각률의 기준을 충족하는 정도의 성능을 보이는지 탐색하였다. 그 결과, 베이지스 p-값이 .10, .20, 또는 .25의 보다 작은 값을 가질 때 모형을 기각한 결과, 모든 표본의 크기 조건에서 대체적으로 1%, 5%, 또는 10%의 경험기각률을 보이는 것으로 나타났다. 이러한 결과에 근거하여 본 논문에서는 임의적으로 .10, .20, 그리고 .25의 값을 기준으로 모형을 기각하기로 정하고 이후 검정력 계산에 적용하였다<sup>7)</sup>.

베이지스 p-값이 .10, .20, 그리고 .25 보다 작

7) 이 기준을 일반화하기 위해서는 더 많은 연구가 반드시 필요하다.

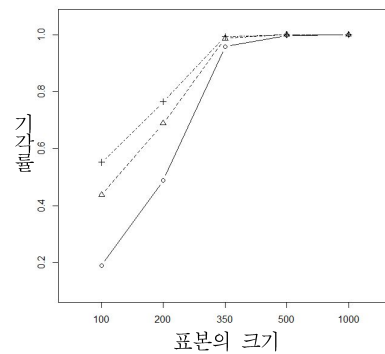


그림 9. 베이지스 접근법의 검정력,  
○:  $\alpha = .10$ ,  $\Delta$ :  $\alpha = .20$ , +:  $\alpha = .25$ .

을 경우 모형을 기각함으로써 얻은 사후예측 모형검증법의 검정력은 그림 9에서 볼 수 있다. 그림에서 알 수 있듯이 표본의 크기가 증가할수록 검정력 역시 증가함을 알 수 있고, 베이지스 p-값이 .25보다 작을 때 모형을 기각할 경우 표본의 크기가 상대적으로 작다 하더라도 합리적인 수준의 검정력을 확보하는 것으로 드러났다. 따라서, 물론 더 많은 연구가 필요하겠으나, 베이지스 p-값이 .25에서 .30보다 작은 값을 보일때는 모형을 기각하는 것이 옳은 판단이 될 것으로 보인다.

세 방법의 비교

마지막으로 완전정보최대우도법의 카이제곱 통계량, M=100일 경우 다중대체법의  $D_3$ 와  $D_2$  통계량, 베이지안 접근법의 사후예측모형 평가법의 불일치추정치를 제 1 종 오류와 검정력 측면에서 비교하였다.

그림 10은 명목기각률이 .05일 때 위 네 가지 접근법의 제 1 종 오류율을 보여준다. 모든 접근법에서 표본크기가 증가함에 따라 명목기각률에 접근하는 것을 알 수 있다. 한 가

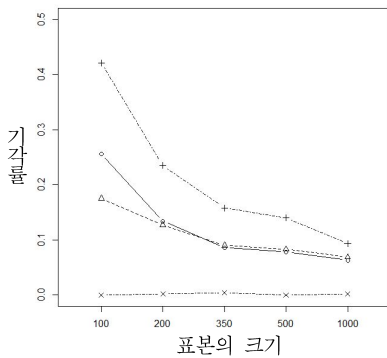


그림 10. 제 1 종 오류 비교,  $\alpha = .05$ , 카이제승 (O), 베이즈-불일치추정치(x),  $D_3$ ( $\Delta$ ),  $D_2$ (+).

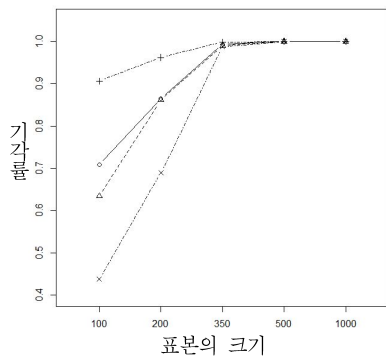


그림 11. 검정력 비교,  $\alpha = .05$ , 카이제승(O), 베이즈-불일치추정치(x),  $D_3$ ( $\Delta$ ),  $D_2$ (+).

지 재미있는 특징은 카이제곱 통계량과 M=100일 경우의  $D_3$ 통계량은 매우 비슷한 성능을 보였으나,  $D_2$  통계량은 같은 조건에서 일관적으로 매우 높은 제 1 종 오류율을 보였으며 표본의 크기가 작을 때에는 매우 심각한 수준의 오류율을 보이는 것으로 드러났다. 이러한 패턴은 명목기각률이 .01과 .10일때에도 마찬가지로 발생하였으며 자세한 값은 표 1에서 확인할 수 있다. 앞서서도 지적하였듯이 사후예측모형평가법은 제 1 종 오류율이 명목기각률에 비해 매우 낮았다.

그림 11은 검정력을 비교한 결과를 보여준다. 표본의 크기가 증가함에 따라 모든 방법의 검정력이 증가함을 알 수 있다.  $D_2$ 통계량은 표본의 크기가 작을 때에도 매우 높은 검정력을 보였다. 즉 명목기각률이 .05이고, M=100이고 N=100 일 경우 .906의 검정력을 보였는데 이는 다른 방법에 비해 월등히 높은 검정력이다. 그러나 이 결과는 대체의 수가 높을 경우  $D_2$ 의 제 1 종 오류율이 매우 높다는 것을 감안하면 매우 조심스럽게 해석해야 할 것이다. 즉,  $D_2$  통계량을 이용하여 모형이 기각되었을 경우, 이 결과는 모형이 자료를 설명하지 못하는 것인지, 아니면 제 1 종 오류인지 구분하기가 힘들다는 것을 의미하기 때문이다. 다른 방법들은 매우 합리적인 검정력 패턴을 보이고 있다. 베이지안 접근법은 앞서서 기술한 대로, 본 연구의 결과를 이용하여 .10, .20, .25의 값보다 작을 경우의 경험적 기각비율을 계산하여 검정력을 추정하였다. 표본의 크기가 작을 때 다른 방법들에 비해 검정력이 약간 떨어지는 것으로 보이나, 이는 본 연구에서 사용한 값들의 독특성에 기인한 것일 수 있다. 앞으로 더 많은 연구가 필요한 영역이다.

## 논 의

본 연구에서는 확률모형에 기반한 세 가지 결측치 처리 기법을 공분산구조모형의 맥락에서 소개하고 그 특징을 비교하였다. 공분산구조모형의 적용에서 가장 중요한 과정에 포함되는 데이터-모형간 적합도 평가에 관하여 세 가지 방법이 제공하는 통계량 및 불일치 측정치의 성능도 제 1 종 오류 및 검정력 측면에서 비교하였다. 본 논문의 목적 및 중요한 결과를 요약하면 다음과 같다.

첫째, 데이터에 결측치가 존재할 경우 현재 공분산구조모형을 이용하는 심리학 연구에서 가장 많이 사용되고 있는 방법은 완전정보최대우도방법이다. 그러나, 다중대체법 및 베이즈안 접근법이라는 이론적으로 매우 잘 확립된 기법 역시 이용가능하다.

둘째, 최대우도법의 카이제곱통계량은 여러 맥락에서 독립적으로 연구된 적이 있으나, 다른 기법에서 제공하는  $D_3$ ,  $D_2$  통계량, 그리고 베이즈접근법의 베이즈 p-값은 공분산구조분석의 맥락에서 연구된 바가 없었다. 본 논문에서는 그 성능을 비교한 결과, 현재 공분산구조모형의 사용자에게 가장 쉽게 이용가능한 통계량인  $D_2$ 는 우려할 만한 수준의 제 1 종 오류율을 보이는 것으로 드러났고, 이로 인해 모형 기각률이 다른 기법들에 비해 비정상적으로 높은 것으로 드러났다. 셋째, 본 연구에서는 베이즈 p-값을 해석하는 지침을 제공하기 위한 탐색적 연구를 통해, 연구자가 제안한 공분산구조모형이 자료생성모형과 일치할 경우 경험적기각률이 .01, .05, 그리고 .10에 근접할만한 수행을 보이도록 하는 값인 .10, .20, 그리고 .25등의 값을 제안하였다. 좀 더 나은 지침을 제공하는 값들을 찾기 위해서는 후속

연구가 더 필요한 것으로 보인다.

마지막으로, 본 논문은 세 가지 결측치 처리기법의 모형평가성능을 서로 비교하였다. 그 결과  $D_3$ 와 카이제곱통계량이 서로 비슷한 성능을 보이는 것으로 나타났고,  $D_2$ 의 사용은 추천되지 않는 것으로 보인다. 베이즈접근법은 다른 방법과 직접 비교를 위하여 로그우도함수를 불일치 측정치로 선택하고 그 특성을 탐색하였다.

본 논문은 최신의 결측치 처리기법을 심리학 연구자들에게 소개하고 공분산구조모형의 데이터-모형 적합도 판단의 맥락에서 서로 비교하였다는데 의의가 있을 것이다. 본 연구의 내용과 연구의 틀은 후속연구를 진행하는 바탕을 제공할 것으로 기대된다. 후속연구에서는 다양한 형태와 복잡성을 지닌 모형을 가지고, 실제 연구 상황과 실제 자료의 특성(예: 비정규성)을 보다 더 잘 반영한 상황을 설정하여 연구를 진행하는 것이 필요할 것으로 보인다. 축적된 연구 결과들은 결측치 처리에 대한 이해를 높이고, 올바른 기준을 가지로 모형을 판단하는 관행이 정착되는 데 기여할 것이다.

## 참고문헌

- Allison, P. D. (1987). Estimation of linear models with incomplete data. *Sociological methodology*, 17, 71-103.
- Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of abnormal psychology*, 112(4), 545-557.
- Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A.

- Marcoulides & R. E. Schumaker (Eds.), *Advanced structural equation modeling* (pp. 243-277). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, 88(3), 588-606.
- Bodner, T. E. (2008). What improves with increased missing data imputations?. *Structural Equation Modeling*, 15(4), 651-675.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sage Focus Editions*, 154, 136-136.
- Dahl, F. A. (2006). On the conservativeness of posterior predictive  $p$ -values. *Statistics & Probability Letters*, 76, 1170-1174.
- Enders, C. K. (2001). The performance of the full information maximum likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement*, 61(5), 713-740.
- Finkelstein, C. (1979). Estimation for the multiple factor model when data are missing. *Psychometrika*, 44(4), 409-420.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.
- Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6), 721-741.
- Gelman, A., Meng, X. L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica sinica*, 6(4), 733-760.
- Gold, M. S., & Bentler, P. M. (2000). Treatments of missing data: A Monte Carlo comparison of RBHDI, iterative stochastic regression imputation, and expectation-maximization. *Structural Equation Modeling*, 7(3), 319-355.
- Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, 8(3), 206-213.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.
- Larson, R. (2011). Missing data imputation versus full information maximum likelihood with second-level dependencies. *Structural Equation Modeling*, 18(4), 649-662.
- Lee, S. Y. (2007). *Structural equation modeling: A Bayesian approach* (Vol. 711). John Wiley & Sons.
- Li, K. H., Meng, X. L., Raghunathan, T. E., & Rubin, D. B. (1991). Significance levels from repeated  $p$ -values with multiply-imputed data. *Statistica Sinica*, 1(1), 65-92.
- Little, R. J. & Rubin, D. B. (2002). *Statistical analysis with missing data*. Wiley.
- Meng, X. L., & Rubin, D. B. (1992). Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika*, 79(1), 103-111.
- Olinsky, A. Chen, S. & Harlow, L. (2003). The comparative efficacy of imputation methods for

- missing data in structural equation modeling. *European Journal of Operational Research*, 151(1), 53-79.
- Robins, J. M., van der Vaart, A., & Ventura, V. (2000). Asymptotic distribution of  $p$  values in composite null models. *Journal of the American Statistical Association*, 95, 1143-1156.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581-592.
- Rubin, D. B. (1977). Formalizing subjective notions about the effect of nonrespondents in sample surveys. *Journal of the American Statistical Association*, 72(359), 538-543.
- Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. New York, New York: Wiley.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. CRC press.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological methods*, 7(2), 147.
- Sinharay, S., Stern, H. S., & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological methods*, 6(4), 317-329.
- Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, 82(398), 528-540.
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1), 1-10.
- Yuan, K-H., Yang-Wallentin, F. & Bentler, P. M. (2012). ML versus MI for missing data with violation of distribution conditions. *Sociological Methods & Research*, 41(4). 598-629.
- Van der Vaart, A. W. (2000). *Asymptotic statistics* (Vol. 3). Cambridge university press.
- Van Buuren, S. (2012). *Flexible imputation of missing data*. CRC press.
- Wilks, S. S. (1932). Moments and distributions of estimates of population parameters from fragmentary samples. *The annals of Mathematical Statistics*, 3(3), 163-195.
- 1차원고접수 : 2014. 03. 01.  
수정원고접수 : 2014. 05. 01.  
최종게재결정 : 2014. 05. 20.

## A Comparison of Full Information Maximum Likelihood, Multiple Imputation, and Bayesian Approach in Overall Goodness of Fit Assessment of Structural Equation Modeling with Missing Data

Lee, Taehun

University of Oklahoma

In practical applications of any statistical modeling, including structural equation modeling(SEM), virtually every data set contains missing values. It is a well known fact that improper handling of missing data can exert harmful impact on subsequent statistical inferences in a variety of ways to varying degrees. In the context of SEM, the full information maximum likelihood(FIML) has been arguably the most popular method for addressing missing data. Despite of being yet less widely known to majority of applied researchers as flexible alternatives to FIML, multiple imputation (MI) procedures and Bayesian approaches have recently begun to emerge as viable solutions among many applied researchers. An important objective of this article is to introduce these methods to applied researchers in an accessible manner using SEM as the context. Structural equation modeling actually involves the process of proposing, estimating, and evaluating the researcher's hypothesis that is believed to be underlying and purported in generating the observed data. Therefore, it is essential to evaluate the overall goodness-of-fit of the posited model in any given application. FIML, MI and Bayesian approaches, respectively, yield the chi-square,  $D_2$ ,  $D_3$ , and the posterior predictive modeling checking (PPMC) p-value as statistical tools for the assessment of data-model fit. Another important objective of this article is to study performance of these model evaluation tools in the context of SEM. Further, relative performance of these data-model fit assessment tools is to be evaluated with respect to their Type I error rates and power. The performance of these assessment tools, except the chi-square statistics, has never been evaluated nor been compared within the context of SEM. The initial results provided in the present article is believed to not only enhance the knowledge base regarding the characteristics of these assessment tools under missing data, but also provide an initial guideline for the proper use of these assessment tools in the real-world data analysis especially in the application of SEM with missing data.

*Key words* : missing data, structural equation modeling, full information maximum likelihood, multiple imputation, Bayesian

부 록

본 논문에서 제시된 결과를 얻기 위해 사용된 자료의 생성모형은 다음에 제시된 측정모형과 구조방정식모형에 기반 하였다. 즉, 관찰변인과 잠재변인을 각각  $Y_i$ 와  $\eta_i$ 라고 할 때, 측정모형과 구조방정식모형은 각각

$$Y_i = \tau + A\eta_i + \epsilon_i$$

$$\eta_i = \alpha + B\eta_i + \zeta_i$$

와 같이 주어진다. 이 모형으로부터 본문에 제시된 가상연구결과를 얻기 위해 사용된 벡터값은 아래와 같다.

$$\tau = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)'$$

$$A' = \begin{bmatrix} 1.8 & .8 & .8 & .8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.8 & .6 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.8 & .4 & .2 & .2 & 0 \end{bmatrix}$$

$$\alpha = (0.0, 1.0, 2.0)'$$

$$B = \begin{pmatrix} 1.0 & .0 & .0 \\ .0 & 1.0 & .0 \\ .0 & .0 & 1.0 \end{pmatrix}$$

$$\Phi = \begin{pmatrix} 1.0 & .6 & .3 \\ .6 & 1.0 & .6 \\ .3 & .6 & 1.0 \end{pmatrix}$$

$$\Psi = \text{diag}\{.3, \dots, .3\}.$$

여기서,  $\Phi$  및  $\Psi$  는 각각  $\epsilon_i$  및  $\zeta_i$ 의 분산-공분산 행렬을 나타내며  $\text{diag}.3, \dots, .3$ 는 크기가  $15 \times 15$ 이고 대각원소가 .3인 대각행렬을 의미한다. 종단연구의 특성을 반영하여 오차항간의 상관도 작은 양의 값을 가지는 것으로 지정하였다. 즉,  $\Psi$  행렬의 (1,6), (6,11), (1,11)의 원소값이 각각 .10, .10, .03의 값을 가지도록, 그리고 (6,1), (11,6), (11,1) 원소값이 각각 .10, .10, .03의 값을 가지도록 지정하였다. 이와 같이 지정된 모수벡터  $\vartheta_0$ 로부터 아래 제시된 모형-기반 평균벡터 및 분산-공분산행렬을 유도한 뒤,

$$\begin{aligned}\mu(\vartheta_0) &= \tau + \Lambda(I - B)^{-1}\alpha \\ \Sigma(\vartheta_0) &= \Lambda(I - B)^{-1}\Phi(I - B)^{-1}\Lambda'\end{aligned}$$

평균이  $\mu(\vartheta_0)$  이고 분산-공분산행렬이  $\Sigma(\vartheta_0)$ 인 다변량정규분포를 따르는 크기가  $15 \times 1$ 인 확률벡터로부터 크기가  $N$ 인 자료를 생성하였다.

자료 및 결측치의 생성, 모형의 적합, 적합도지수의 계산 등 모든 가상 연구 절차의 필요한 계산은 통계소프트웨어 R을 이용하였다.

아래 제시된 표는 각 구간별 결측치 비율을 15%, 5%, 5%, 15% 로 변화시켜 전체자료의 10%가 결측치로 변환되게 한 조건하에서의 제 1 종 오류 (표A1) 및 검정력 (표A2), 그리고 각 구간별 결측치 비율을 60%, 20%, 20%, 60% 로 변화시켜 전체자료의 40%가 결측치로 변환되게 한 조건하에서의 제 1 종 오류 (표A3) 및 검정력 (표A4) 등을 보여준다. 제시된 결과는 결측치 처리 방법의 적합도 판단에 대한 상대적 성능은 결측치의 수준 변화(10%, 25%, 40%)에 따라 큰 차이가 없음을 보여준다.



표 A1. 결측치 처리방법에 따른 제 1 종 오류율: 결측치 수준 10%

	표본 크기 = 100			표본 크기 = 200			표본 크기 = 350			표본 크기 = 500			표본 크기 = 1000		
	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
명목 기각률 ( $\alpha$ )															
카이자승	0.06	0.18	0.29	0.02	0.10	0.17	0.02	0.08	0.17	0.01	0.05	0.11	0.01	0.06	0.11
불일치 측정치	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.01
대체수(M) = 5															
$D_3$	0.05	0.16	0.27	0.02	0.11	0.18	0.02	0.09	0.16	0.01	0.05	0.11	0.02	0.06	0.11
$D_2$	0.06	0.13	0.21	0.03	0.10	0.16	0.03	0.09	0.14	0.02	0.08	0.13	0.02	0.07	0.12
대체수(M) =10															
$D_3$	0.05	0.17	0.26	0.02	0.10	0.17	0.02	0.08	0.16	0.01	0.05	0.12	0.02	0.06	0.11
$D_2$	0.08	0.21	0.32	0.05	0.15	0.24	0.04	0.13	0.21	0.03	0.10	0.18	0.02	0.09	0.15
대체수(M) =20															
$D_3$	0.05	0.17	0.27	0.02	0.10	0.17	0.02	0.09	0.17	0.01	0.06	0.13	0.01	0.06	0.11
$D_2$	0.10	0.24	0.35	0.05	0.15	0.24	0.04	0.12	0.21	0.02	0.10	0.18	0.02	0.08	0.16
대체수(M) =50															
$D_3$	0.05	0.18	0.26	0.02	0.10	0.17	0.02	0.08	0.17	0.01	0.06	0.12	0.01	0.06	0.11
$D_2$	0.10	0.24	0.34	0.05	0.14	0.22	0.03	0.11	0.19	0.02	0.08	0.16	0.02	0.07	0.14
대체수(M) =100															
$D_3$	0.05	0.17	0.26	0.02	0.10	0.18	0.02	0.08	0.17	0.01	0.06	0.12	0.01	0.06	0.11
$D_2$	0.10	0.25	0.35	0.05	0.13	0.22	0.03	0.11	0.20	0.01	0.08	0.15	0.02	0.07	0.13

표 A2. 결측치 처리방법에 따른 검정력: 결측치 수준 10%

	표본 크기 = 100			표본 크기 = 200			표본 크기 = 350			표본 크기 = 500			표본 크기 = 1,000		
	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
명목 기각률 ( $\alpha$ )															
카이자승	0.51	0.71	0.81	0.80	0.92	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
불일치 측정치	0.42	0.64	0.72	0.76	0.86	0.89	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
대체수(M) = 5															
$D_3$	0.53	0.76	0.85	0.84	0.94	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$D_2$	0.26	0.41	0.52	0.37	0.54	0.65	0.54	0.70	0.79	0.61	0.77	0.86			
대체수(M) = 10															
$D_3$	0.55	0.77	0.86	0.85	0.95	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$D_2$	0.53	0.75	0.85	0.75	0.89	0.94	0.93	0.98	1.00	0.98	1.00	1.00	1.00	1.00	1.00
대체수(M) = 20															
$D_3$	0.55	0.78	0.87	0.86	0.95	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$D_2$	0.67	0.85	0.91	0.88	0.95	0.97	0.99	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
대체수(M) = 50															
$D_3$	0.56	0.77	0.87	0.86	0.95	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$D_2$	0.70	0.88	0.93	0.90	0.97	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
대체수(M) = 100															
$D_3$	0.56	0.77	0.87	0.86	0.95	0.97	1.00	1.00	1.00	0.96	0.99	1.00	1.00	1.00	1.00
$D_2$	0.71	0.89	0.94	0.91	0.97	0.99	1.00	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00

표 A3. 결측치 처리방법에 따른 제 1 종 오류율: 결측치 수준 40%

	표본 크기 = 100			표본 크기 = 200			표본 크기 = 350			표본 크기 = 500			표본 크기 = 1000		
	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
	명목 기각률 ( $\alpha$ )														
카이제승	0.43	0.58	0.66	0.06	0.17	0.27	0.02	0.11	0.18	0.01	0.08	0.15	0.02	0.06	0.13
불일치 측정치	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.01
	대체수(M) =5														
$D_3$	0.07	0.21	0.27	0.04	0.12	0.20	0.03	0.10	0.16	0.03	0.08	0.14	0.03	0.08	0.14
$D_2$	0.08	0.10	0.14	0.04	0.06	0.09	0.03	0.05	0.08	0.03	0.04	0.08	0.02	0.05	0.07
	대체수(M) =10														
$D_3$	0.07	0.17	0.26	0.04	0.14	0.21	0.03	0.09	0.17	0.02	0.09	0.15	0.02	0.09	0.15
$D_2$	0.08	0.13	0.17	0.04	0.11	0.17	0.06	0.13	0.20	0.04	0.09	0.17	0.04	0.11	0.17
	대체수(M) =20														
$D_3$	0.07	0.17	0.24	0.04	0.13	0.21	0.03	0.11	0.19	0.02	0.10	0.16	0.02	0.07	0.14
$D_2$	0.13	0.20	0.24	0.10	0.22	0.33	0.10	0.22	0.31	0.07	0.19	0.27	0.08	0.16	0.23
	대체수(M) =50														
$D_3$	0.07	0.19	0.28	0.04	0.14	0.20	0.03	0.11	0.18	0.02	0.10	0.16	0.02	0.07	0.14
$D_2$	0.17	0.28	0.37	0.15	0.30	0.41	0.10	0.24	0.31	0.08	0.20	0.30	0.06	0.17	0.26
	대체수(M) =100														
$D_3$	0.07	0.18	0.27	0.04	0.15	0.23	0.03	0.11	0.19	0.02	0.10	0.16	0.02	0.07	0.14
$D_2$	0.35	0.56	0.68	0.18	0.33	0.44	0.10	0.22	0.30	0.06	0.19	0.29	0.06	0.15	0.24

표 A4. 결측치 처리방법에 따른 검정력: 결측치 수준 40%

	표본 크기 = 100			표본 크기 = 200			표본 크기 = 350			표본 크기 = 500			표본 크기 = 1,000		
	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
명목 기각률 ( $\alpha$ )															
카이자승	0.63	0.80	0.87	0.57	0.79	0.89	0.90	0.97	0.98	0.98	1.00	1.00	1.00	1.00	1.00
불일치 측정치	0.04	0.17	0.27	0.16	0.42	0.52	0.70	0.91	0.94	0.90	0.98	0.99	0.99	1.00	1.00
대체수(M) = 5															
$D_3$	0.06	0.19	0.29	0.27	0.52	0.65	0.63	0.84	0.91	0.86	0.96	0.98	1.00	1.00	1.00
$D_2$	0.05	0.08	0.09	0.06	0.09	0.13	0.08	0.12	0.15	0.06	0.11	0.16	0.10	0.18	0.24
대체수(M) = 20															
$D_3$	0.08	0.19	0.30	0.33	0.57	0.69	0.71	0.90	0.95	0.91	0.98	0.99	1.00	1.00	1.00
$D_2$	0.10	0.20	0.28	0.15	0.28	0.41	0.19	0.38	0.51	0.25	0.47	0.61	0.44	0.69	0.82
대체수(M) = 20															
$D_3$	0.08	0.22	0.35	0.34	0.59	0.72	0.75	0.92	0.95	0.94	0.99	1.00	1.00	1.00	1.00
$D_2$	0.40	0.65	0.75	0.39	0.61	0.72	0.53	0.76	0.86	0.67	0.86	0.92	0.94	0.99	1.00
대체수(M) = 50															
$D_3$	0.08	0.25	0.37	0.37	0.63	0.75	0.77	0.92	0.96	0.95	0.99	1.00	1.00	1.00	1.00
$D_2$	0.77	0.88	0.91	0.68	0.83	0.89	0.85	0.95	0.97	0.94	0.98	0.99	1.00	1.00	1.00
대체수(M) = 100															
$D_3$	0.08	0.26	0.37	0.38	0.63	0.76	0.77	0.92	0.96	0.96	0.99	1.00	1.00	1.00	1.00
$D_2$	0.85	0.92	0.95	0.78	0.90	0.94	0.94	0.98	0.99	0.98	1.00	1.00	1.00	1.00	1.00