فصلنامه اندازه‌گیری تربیتی
دانشگاه علامه طباطبائی
سال هشتم، شماره ۳۱، بهار ۹۷، ص ۹۹ تا ۱۱۲

Quarterly of Educational Measurement
Allameh Tabataba i University
Vol. 8, No. 31, Spring 2018

برآورد مدل‌های معادلات ساختاری در حضور داده‌های گمشده با استفاده از روش بیشینه درستنمایی با اطلاعات کامل: یک مطالعه‌ی شبیه‌سازی

شیدا مرادی[۱]، شبنم فانی*[۲]، فاطمه پورخاقان[۳]

تاریخ دریافت: ۹۶/۰۹/۰۵
تاریخ پذیرش: ۹۷/۰۲/۱۵

چکیده

مدل‌یابی معادلات ساختاری، یک رویکرد آماری چندمتغیری نیرومند جهت ارزیابی روابط پیچیده‌ی بین متغیرهای مکنون در بسیاری از حوزه‌های علوم انسانی و رفتاری است. یکی از چالش‌های رایج در برآورد مدل‌های معادلات ساختاری که بر مبنای آزمون فرضیه‌ها انجام می‌شود، وجود داده‌های گمشده است. شیوه‌ی معمول، حذف آزمودنی‌هایی با پاسخ‌های گمشده روی هر کدام از سوالات است که با افزایش درصد مقادیر گمشده در مجموعه داده‌ها، منجر به از دست رفتن حجم زیادی از اطلاعات نمونه و اریبی برآوردگرهای حاصل می‌شود. در تحلیل مدل‌های معادلات ساختاری با حضور مقادیر گمشده می‌توان از روش بیشینه درستنمایی با اطلاعات کامل استفاده کرد که در آن از تمام داده‌های موجود در نمونه به طور حداکثری استفاده می‌شود. در این مقاله، عملکرد روش بیشینه درستنمایی با اطلاعات کامل تحت هر سه مکانیسم داده‌های گمشده شامل گمشدن کاملاً تصادفی، تصادفی و غیرتصادفی در یک مطالعه‌ی شبیه‌سازی مورد بررسی قرار می‌گیرد. بدین منظور، دو مدل تحلیل عاملی تأییدی را در نظر گرفته، داده‌ها تحت هر سه مکانیسم مذکور تولید و تأثیر دو شاخص حجم نمونه (۱۰۰ و ۵۰۰) و درصد مقادیر گمشده (۲٪، ۵٪، ۱۰٪، ۱۵٪، ۲۰٪، ۲۵٪، ۳۰٪، ۳۵٪ و ۴۰٪) بر شاخص ریشه‌ی میانگین مجذورات خطاهای برآورد مورد بررسی قرار گرفته است. نتایج نشان می‌دهند که با افزایش درصد مقادیر گمشده، عملکرد مدل‌های معادلات ساختاری

. دانشجوی دکتری سنجش و اندازه گیری، دانشگاه علامه طباطبائی تهران ، sheidamoradi2016@gmail.com
.* استادیار گروه آمار، دانشگاه علامه طباطبائی تهران. s.fani@atu.ac.ir
. دانشجوی دکتری سنجش و اندازه گیری، دانشگاه علامه طباطبائی تهران، fatemepourkhaghan@yahoo.com

با استفاده از روش بیشینه درستنمایی با اطلاعات کامل در حضور مکانیسم‌های گمشدن کاملاً تصادفی، تصادفی و غیرتصادفی به طور کلی بهتر از عملکرد آن بدون استفاده از این روش بر حسب برخی شاخص‌های نیکویی برازش است.

واژگان کلیدی: مدل‌یابی معادلات ساختاری (SEM)، روش بیشینه درستنمایی با اطلاعات کامل (FIML)، مکانیسم گمشدن کاملاً تصادفی (MCAR)، مکانیسم گمشدن تصادفی (MAR)، مکانیسم گمشدن غیرتصادفی (MNAR)، تحلیل عاملی تاییدی (CFA)

مقدمه

در سال‌های اخیر، علاقه‌ی روبه‌رشدی برای استفاده از مدل‌یابی معادلات ساختاری[1] (SEM) در حوزه‌های مختلف علوم انسانی و رفتاری ایجاد شده است. عمومیت SEM از آنجا ناشی می‌شود که پژوهشگران با ارزیابی روابط پیچیده بین متغیرهای مشاهده‌پذیر و مکنون مواجه هستند که توسط برنامه‌های کامپیوتری مرتبط با SEM همچون لیزرل برای پژوهشگران فراهم می‌شود. ترکیبی از مزیت‌های روش‌شناختی SEM، قابلیت کاربرد آن برای گستره‌ی وسیعی از مسائل مورد بررسی در علوم انسانی و نرم‌افزارهای نیرومند منجر به استفاده وسیع از SEM شده است (هویل[2] ۲۰۱۲ ، کین[3] ۲۰۱۳).

داده‌های گمشده[4] یکی از چالش‌های رایج در اکثر روش‌های تحلیل آماری است و مدل‌یابی معادلات ساختاری از این قاعده مستثنا نیست. استفاده از روش‌های نامناسب در بررسی داده‌های گمشده منجر به سوگیری در برآورد پارامترها، خطاهای استاندارد و آماره‌های آزمون می‌شود (مازن و همکاران[5] ۱۹۸۷،آلیسون[6] ۲۰۰۳). دلایل متعددی برای ایجاد داده‌های گمشده وجود دارد. در هنگام جمع‌آوری داده‌ها به وسیله‌ی پرسشنامه ممکن است افراد به پاسخ دادن سوال‌هایی که آنها را شرمنده می‌سازند تمایل نداشته باشند. بعضی

1. Structural Equation Modeling (SEM)
2. Hoyle
3. Khine
4. Missing Values
y. Muthen et al
. . Allison

از سوالات ممکن است از دیدگاه بعضی‌ها خوب بیان نشده باشند و پاسخگو نداند که چطور پاسخ دهد. در صورتی که پرسشنامه معتبر بوده و مشکلات فوق را نداشته باشد، ممکن است برخی از سوالات در مورد برخی افراد مصداق نداشته باشند و باز با مقادیر گمشده مواجه شویم (لیتل و روبین[1]، ۲۰۰۲).

شیوه‌های رایج در برخورد با مقادیر گمشده، حذف و جانهی[2] داده‌های گمشده می‌باشد. از جمله‌ی این روش‌ها، روش حذف listwise یا تحلیل داده‌های کامل[3] است که در آن تمام آزمودنی‌هایی که دارای پاسخ گمشده روی هر کدام از سوالات هستند از مجموعه داده‌ها حذف می‌شوند و تحلیل فقط بر اساس آزمودنی‌هایی با پاسخ کامل انجام می‌گیرد. این روش زمانی به‌کار می‌رود که تعداد داده‌های گمشده کم باشد. حذف مقادیر گمشده سبب از دست رفتن بخشی از اطلاعات می‌شود که نتیجه آن کاهش دقت و ایجاد سوگیری است. یک روش رایج دیگر، روش حذف دوبه‌دو[4] یا تحلیل موارد موجود[5] است که در این روش تحلیل بر مبنای تمام داده‌های موجود انجام می‌گیرد و صرفا مقادیر گمشده از تحلیل حذف می‌گردند. معمولاً، استفاده از هر دو روش فوق منجر به برآوردگرهایی دارای سوگیری برای پارامترها می‌گردد. روش دیگر، که در کاربردها به‌طور گسترده مورد استفاده قرار می‌گیرد، روش جانهی است. در این روش، مقادیر جانهی جایگزین مقادیر ثبت نشده می‌شوند و سپس داده‌های کامل به وسیله روش‌های استاندارد آماری تحلیل می‌شوند. مثلاً در جانهی میانگین[6]، میانگین مقادیر مشاهده شده جانشین مقادیر گمشده می‌شود. در روش‌های مبتنی بر جانهی، به‌دست آوردن استنباط‌های معتبر نیازمند اصلاحاتی در روش‌های تحلیلی استاندارد است تا جایگاه متفاوت مقادیر واقعی و مقادیر جانهی‌شده مشخص گردد (لیتل و روبین، ۲۰۰۲).

1. Little and Rubin
2. Imputation
3. Complete case analysis
4. Pairwise deletion
5. Available case analysis
6. Mean imputation

یکی از مسائل مهمی که در بررسی داده‌های دارای مقادیر گمشده باید در نظر گرفت، مکانیسم داده‌های گمشده می‌باشد. مکانیسم داده‌های گمشده، ارتباط بین گمشدگی و مقادیر پاسخ‌ها را در ماتریس داده‌ها بیان می‌کند. در واقع، مکانیسم داده‌های گمشده به دنبال پاسخ این پرسش است که آیا گمشدن پاسخ‌ها به مقادیر پاسخ‌های موجود در مجموعه داده‌ها وابسته است یا خیر. روبین[1] (۱۹۷۶) و لیتل و روبین (۲۰۰۲) کلاسه‌بندی زیر را برای مکانیسم داده‌های گمشده ارائه داده‌اند:

۱. گمشدن کاملا تصادفی[2] (MCAR). مکانیسم MCAR بدین معنی است که مکانیسم مقادیر گمشده با مقادیر هیچ کدام از متغیرها، چه آنهایی که مشاهده شده‌اند و چه آنهایی که گمشده‌اند، رابطه‌ای ندارد. در مکانیسم MCAR ، داده‌های مشاهده شده یک نمونه‌ی تصادفی از کل داده‌ها هستند.

۲. گمشدن تصادفی[3] (MAR). مکانیسم MAR مستلزم این است که علت گمشدن داده‌ها با مقادیر گمشده رابطه‌ای نداشته باشد اما ممکن است به مقادیر مشاهده‌شده‌ی سایر متغیرها مرتبط باشد. در واقع، در حالت MAR ، احتمال گمشدن یک مشاهده به شرط پاسخ‌های مشاهده‌شده، مستقل از پاسخ‌های مشاهده نشده است.

۳. گمشدن غیر تصادفی[4] (MNAR). در این حالت، مکانیسم داده‌های گمشده علاوه بر مقادیر پاسخ‌های مشاهده‌شده به مقادیر پاسخ‌های مشاهده‌نشده نیز وابسته است.

معمولاً MCAR و MAR را گمشدن قابل چشم‌پوشی[5] و MNAR را گمشدن غیرقابل چشم‌پوشی[6] می‌نامند. در حالت قابل چشم‌پوشی، استنباط آماری را می‌توان فقط با تحلیل داده‌های مشاهده‌شده و بدون اشاره‌ی صریح به یک فرم پارامتری از مکانیسم داده‌های گمشده انجام داد. در حالت غیرقابل چشم‌پوشی، روش‌های آماری که مکانیسم داده‌های گمشده را در مدل اعمال نمی‌کنند منجر به سوگیری می‌شوند.

1. Rubin
2. Missing Completely At Random (MCAR)
3. Missing At Random (MAR)
4. Missing Not At Random (MNAR)
5. Ignorable
6. Nonignorable

در مدل‌یابی معادلات ساختاری جهت بررسی داده‌های گمشده معمولاً از روش‌های حذف listwise و حذف دوبه‌دو استفاده می‌شود که غالباً منجر به سوگیری می‌شوند. روش دیگر، روش بیشینه درستنمایی با اطلاعات کامل[1] (FIML) است که نسبت به دو روش ذکر شده دارای عملکرد بهتری است (اندرس و باندالوس[2] ۲۰۰۱ ، الینسکی[3] و همکاران ۲۰۰۳).

روش FIML، که نخستین بار توسط فینکباینر[4] (۱۹۷۹) مطرح شده است، روشی جهت استفاده‌ی حداکثری از تمام داده‌های موجود می‌باشد. در این روش، برآورد و تحلیل مدل‌های معادلات ساختاری، حتی زمانی که داده‌ها ناقص هستند، امکان‌پذیر است و امکان دخالت دادن تمامی داده‌های موجود برای برآورد پارامترها میسر می‌شود. در روش FIML داده های گمشده جایگذاری یا جانهی نمی‌شوند و برآورد پارامترها و خطاهای استاندارد به‌طور مستقیم با استفاده از تمام داده‌های مشاهده‌شده صورت می‌گیرد. در واقع، روش FIML با دخالت دادن موارد ناقص در روند برآورد، الگوریتم را به سمت یک مجموعه دقیق‌تر از برآورد پارامترها هدایت می‌کند(اندرس[5]، ۲۰۱۰). در نرم‌افزار لیزرل، روش FIML از طریق الگوریتم EM اجرا می‌شود. الگوریتم EM توسط دمپستر و همکاران[6] (۱۹۷۷) معرفی شده است. الگوریتم EM یک الگوریتم تکرارشونده‌ی[7] کلی جهت محاسبه‌ی برآوردهای بیشینه درستنمایی پارامترها در حالتی است که در مجموعه‌ی داده‌ها مقادیر گمشده وجود دارد. هر تکرار از الگوریتم EM شامل دو گام است: گام امید ریاضی گرفتن[8] (گام E) و گام بیشینه‌سازی[9] (گام M). از این رو، الگوریتم EM نامیده شده است. الگوریتم EM با یک مجموعه مقادیر آغازین[10] برای پارامترها شروع می‌شود. در گام E، امید ریاضی لگاریتم تابع درستنمایی داده‌های مشاهده‌شده به شرط داده‌های ناقص با استفاده

1. Full Information Maximum Likelihood (FIML)
2. Enders and Bandalos
3. Olinsky
4. Finkbeiner
5. Enders
6. Dempster et al
7. Iterative
8. Expectation step (E step)
9. Maximization step (M step)
00. Initial values

از مقادیر کنونی پارامترها برآورد می‌شود. در گام M، برآورد پارامترها به‌گونه‌ای محاسبه می‌شود که امید ریاضی لگاریتم تابع درستنماییِ مورد بررسی در گام E بیشینه شود. الگوریتم EM با اجرای متوالی گام‌های E و M تا رسیدن به همگرایی ادامه پیدا می‌کند (اندرس، ۲۰۰۱).

اندرس و باندالوس (۲۰۰۱) نشان داده‌اند که FIML در حالتی که مکانیسم داده های گمشده MCAR و MAR باشد دارای عملکرد بهتری نسبت به سایر روش‌های رایج برخورد با مقادیر گمشده می‌باشد. در واقع، بنیاد روش FIML برای مکانیسم MAR است (هان و گئو،[1] ۲۰۱۴). اکثر تحقیقات انجام شده عملکرد روش FIML را تحت مکانیسم‌های MCAR و MAR بررسی کرده‌اند. هان و گئو (۲۰۱۴) به بررسی حالتی در سنجش انطباقی چندبعدی پرداخته‌اند که در آن فرضیه MAR نقض شده است. آنها تاثیرات عدم برقراری MAR را بر عملکرد روش FIML مورد بررسی قرار داده‌اند.

در این مقاله، قصد داریم عملکرد روش FIML برای معادلات ساختاری را هم برای مکانیسم‌های گمشدن قابل چشم‌پوشی MCAR و MAR و هم گمشدن غیر قابل چشم‌پوشی MNAR مورد بررسی قرار دهیم. ارزیابی عملکرد FIML تحت سه مکانیسم داده‌های گمشده بر مبنای یک مطالعه‌ی شبیه‌سازی انجام می‌گیرد. هدف مقاله‌ی حاضر، بررسی عملکرد روش FIML تحت مکانیسم‌های مختلف داده‌های گمشده است. در این راستا، مدل‌سازی معادلات ساختاری در نرم‌افزار لیزرل یک بار بدون در نظر گرفتن روش FIML و یک بار با اعمال روش FIML انجام می‌گیرد و نتایج مورد مقایسه قرار می‌گیرند.

روش

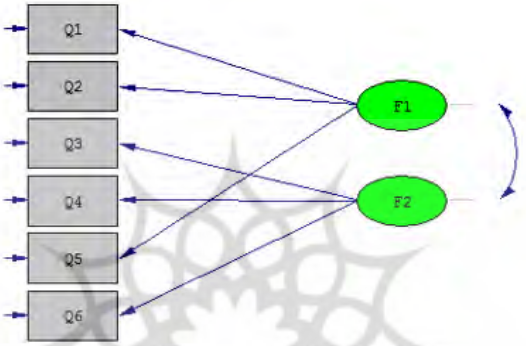
در مطالعه‌ی شبیه‌سازی از دو مدل تحلیل عاملی تاییدی، یک مدل دو عاملی با ۶ متغیر مشاهده‌شده و یک مدل سه عاملی با ۹ متغیر مشاهده شده، استفاده می‌شود (شکل ۱ و ۲). در این شکل‌ها، مدل‌های تحلیل عاملی تاییدی با سه نشانگر برای هر عامل مشخص شده‌اند؛

1. Han and Guo

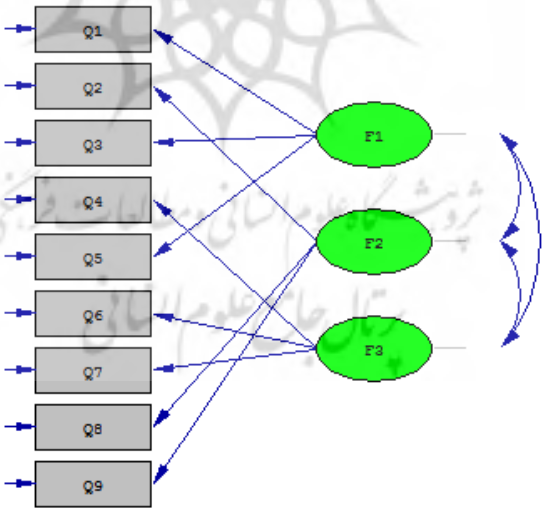
به‌طوری که هر نشانگر تنها روی یک عامل بارگذاری شده است. در مطالعه‌ی شبیه‌سازی برای مدل‌های تحلیل عاملی تاییدی دو شاخص زیر در نظر گرفته شده‌اند:

- اندازه‌ی نمونه با ۲ سطح (۱۰۰ و ۵۰۰)
- درصد مقادیر گمشده با ۹ سطح (۲٪، ۵٪، ۱۰٪، ۱۵٪، ۲۰٪، ۲۵٪، ۳۰٪، ۳۵٪ و ۴۰٪)

در هر کدام از مکانیسم‌های داده‌های گمشده تأثیر اندازه‌ی نمونه و درصد مقادیر گمشده بر شاخص ریشه میانگین مجذورات خطای برآورد[1] (RMSEA) مورد بررسی قرار می‌گیرد.



شکل ۱. مدل تحلیل عاملی تاییدی دو عاملی



شکل ۲. مدل تحلیل عاملی تاییدی سه عاملی

1. Root Mean Square Error of Approximation (RMSEA)

بر مبنای هر کدام از دو مدل تحلیل عاملی تاییدی ذکر شده یک ماتریس همبستگی متناظر با مدل‌های تحلیل عاملی تاییدی دو عاملی و سه عاملی برای متغیرهای مشاهده شده در نظر گرفته شده است. جهت تولید داده‌ها با ساختار همبستگی مناسب از تبدیل چولسکی[1] استفاده شده است. تولید و شبیه‌سازی داده‌ها با استفاده از نرم افزار آماری R صورت گرفته است. بعد از تولید ماتریس داده‌ها، مقادیر گمشده تحت هر کدام از سه مکانیسم MCAR، MAR و MNAR تولید شده‌اند. در حالت MCAR و در هر کدام از دو مدل، مقادیر گمشده با حذف تصادفی یک درصد معین از مشاهدات مربوط به هر سطر ماتریس داده‌ها شبیه‌سازی شده‌اند. در این راستا، مشاهدات ماتریس داده‌ها با یک ماتریس اعداد تصادفی یکنواخت متناظر شده‌اند و این اعداد تصادفی یکنواخت جهت انتخاب مشاهداتی که باید مطابق با نه درصد معین مقادیر گمشده حذف شوند مورد استفاده قرار گرفته‌اند. در حالت MAR، متغیرهای Q_4، Q_5 و Q_6 در مدل دو عاملی و متغیرهای Q_4 تا Q_9 در مدل سه عاملی دارای مقادیر گمشده هستند و در هر کدام از دو مدل متغیرهای Q_1، Q_2 و Q_3 مشاهده شده هستند. برای هر کدام از دو مدل مقادیر گمشده تحت مکانیسم MAR به متغیرهای مشاهده شده وابسته هستند اما مستقل از مقادیر متغیرهای گمشده می باشند. با اتخاذ روشی مشابه با روش اندرس و باندالوس (۲۰۰۱) مقادیر گمشده در هر کدام از دو مدل تحت مکانیسم MAR و برای نه درصد معین گمشدگی شبیه‌سازی شده‌اند. در حالت MNAR، گمشدن برای همه متغیرها و مطابق با نه درصد معین گمشدن رخ می‌دهد. تحت MNAR، مکانیسم گمشدن به مقادیر متغیرها، مشاهده شده و مشاهده‌نشده، وابسته است.

بعد از تولید مقادیر گمشده تحت سه مکانیسم، مجموعه داده‌های شبیه‌سازی شده در نرم‌افزار لیزرل جهت اجرای مدل‌های تحلیل عاملی تاییدی مورد استفاده قرار گرفته‌اند. مجموعه داده‌های هر مدل برای دو حجم نمونه (۱۰۰ و ۵۰۰) و نه درصد مقادیر گمشده شبیه‌سازی شده‌اند. در حالت کلی، مدل‌های تحلیل عاملی تاییدی برای هر کدام از سه مکانیسم در ۱۸ حالت برای مقادیر متفاوت حجم نمونه و درصدهای مقادیر گمشده اجرا

1. Cholesky transformation

شده اند. هر کدام از مدل‌ها در دو حالت، یکبار با در نظر گرفتن روش بیشینه درستنمایی با اطلاعات کامل (FIML) و یکبار بدون اعمال این روش در لیزرل اجرا شده‌اند. جهت بررسی عملکرد FIML، نتایج حالت‌های مختلف برای دو مدل تحلیل عاملی تاییدی با هم مورد مقایسه قرار می‌گیرند.

یافته‌ها

در راستای بررسی عملکرد FIML، تمرکز بر روی مقایسه‌ی نتایج مربوط به دو حالت اعمال روش FIML با استفاده از الگوریتم EM و عدم اعمال روش FIML (حالت بدون استفاده از الگوریتم EM) برای هر کدام از سه مکانیسم در دو حجم نمونه (۱۰۰ و ۵۰۰) و نه درصد متفاوت مقادیر گمشده (۲٪، ۵٪، ۱۰٪، ۱۵٪، ۲۰٪، ۲۵٪، ۳۰٪، ۳۵٪ و ۴۰٪) می‌باشد. شاخص ریشه میانگین مجذورات خطای برآورد (RMSEA) به عنوان ملاک اصلی برای مقایسه‌ی مدل‌ها مورد استفاده قرار می‌گیرد. شاخص RMSEA زمانی برازش بهتری به‌دست می‌دهد که مقدار کمتری داشته باشد. برای مدل‌هایی با برازش خوب مقدار RMSEA کمتر از ۰/۰۵ است و مدل‌هایی که شاخص RMSEA برای آنها بیشتر از ۰/۱ است برازش ضعیفی دارند.

جداول ۱ و ۲ نتایج مربوط به مدل‌های دو عاملی و سه عاملی را نشان می‌دهند. در جدول یک، برای مکانیسم MCAR در دو حجم نمونه ۱۰۰ و ۵۰۰ ، نتایج مقایسه شاخص RMSEA در حالت اعمال روش FIML نشان می‌دهد که با افزایش درصد مقادیر گمشده یک کاهش چشمگیر در شاخص RMSEA (از ۰/۵۵۶ به ۰/۱۸۳ در حجم نمونه ۱۰۰ و از ۰/۵۷۶ به ۰/۱۹۸ در حجم نمونه ۵۰۰) رخ می‌دهد، در صورتی که برای حالت عدم اعمال FIML کاهشِ ناچیز و همراه با نوسانات در شاخص RMSEA رخ می‌دهد. برای مکانیسم MAR در دو حجم نمونه ۱۰۰ و ۵۰۰ ، نتایج مقایسه شاخص‌ها مشابه با نتایج مکانیسم MCAR است؛ با این تفاوت که میزان کاهشی که در شاخص RMSEA رخ داده است کمتر از میزان کاهش متناظر آن در مکانیسم MCAR است. نتایج مکانیسم MNAR در دو حجم نمونه ۱۰۰ و ۵۰۰ ، همانند نتایج مکانیسم‌های MCAR و MAR ، کاهش شاخص‌ها

را در حالت اعمال FIML نشان می‌دهد. میزان کاهش شاخص RMSEA در مکانیسم MNAR نسبت به مکانیسم MAR بیشتر است. علاوه بر این در مکانیسم MNAR نسبت به مکانیسم‌های MCAR و MAR میزان کاهش شاخص‌ها در حجم نمونه ۵۰۰ نسبت به حجم نمونه ۱۰۰ بیشتر است و این مطلب بیانگر تاثیر افزایش حجم نمونه بر روی نتایج مکانیسم MNAR می‌باشد. همچنین بررسی نتایج حالت عدمِ اعمال روش FIML در هر سه مکانیسم به طور واضح نشان‌دهنده‌ی عدم ثبات -گاه افزایش و گاه کاهش- شاخص RMSEA می‌باشد که این مسئله در مقایسه با حالت‌هایی که FIML اعمال شده است، برتری FIML به‌ویژه برای مکانیسم MNAR را نشان می‌دهد.

یک نکته که باید درباره مدل دو عاملی در نظر گرفته شود مقدار شاخص RMSEA برای مدل‌ها است که حول ۰/۵ مقادیری را می‌گیرد؛ چنین مقادیری بیانگر برازش مناسب نیستند، با این حال نکته مورد توجه مقاله حاضر کاهش شاخص RMSEA در حضور FIML می‌باشد که مطابق با نتایج جداول یک تا سه این کاهش رخ داده است. این مسئله در مدل تحلیل عاملی تاییدی سه عاملی برطرف شده است.

در جدول دو نتایج مدل سه عاملی نمایش داده شده است. نتایج مکانیسم MCAR در حالت اعمال FIML نشان می‌دهد که با افزایش درصد مقادیر گمشده، شاخص RMSEA کاهش می‌یابد و مقادیر شاخص‌ها بیانگر برازش مناسب است. در حجم نمونه ۵۰۰ نسبت به حجم نمونه ۱۰۰ کاهش بیشتری در شاخص RMSEA رخ داده است. در حالت عدم اعمال FIML، با افزایش درصد مقادیر گمشده در حجم نمونه ۱۰۰، شاخص RMSEA افزایش می‌یابد و برای مقادیر گمشده بالاتر از ۲۰ درصد به دلیل تعداد اندک داده‌های موجود مدل اجرا نمی‌شود. در حجم نمونه ۵۰۰، عدم ثبات – گاه افزایش و گاه کاهش –در شاخص‌های RMSEA رخ داده است. برای حالت ۴۰ درصد به دلیل اینکه مقدار آماره خی دو کمتر از درجه آزادی بدست آمده است مقدار RMSEA برابر با صفر گزارش شده است. مقایسه دو حالت اعمال FIML و عدم اعمال آن در مکانیسم MCAR بیانگر عملکرد مطلوب FIML برای مدل سه عاملی می‌باشد. همچنین در حالت اعمال FIML ، در حجم نمونه ۵۰۰ نسبت به حجم نمونه ۱۰۰ میزان کاهش بیشتری در شاخص RMSEA رخ داده است.

نتایج مدل سه عاملی تحت مکانیسم MAR مشابه با نتایج مکانیسم MCAR است و بیانگر کاهش شاخص‌ها در حالت اعمال FIML و عدم ثبات شاخص‌ها در حالت عدم اعمال FIML می‌باشد. برای مکانیسم MNAR نتایجی مشابه با نتایج مکانیسم‌های MCAR و MAR بدست آمده است. میزان کاهش شاخص RMSEA در حجم نمونه ۵۰۰ نسبت به حجم نمونه ۱۰۰ و نیز نسبت به دو مکانیسم دیگر بیشتر است. این مطلب دلیل دیگری بر کارایی روش FIML برای مکانیسم MNAR می‌باشد. نتایج حالت عدم اعمال FIML همانند نتایج متناظر آن در دو مکانیسم دیگر بیانگر عدم ثبات شاخص RMSEA است.

جدول ۱. نتایج مدل دو عاملی برای مکانیسم‌های MCAR ، MAR و MNAR به‌ازای حجم نمونه‌های ۱۰۰ و ۵۰۰.

	درصد مقادیر گمشده	مکانیسم MCAR		مکانیسم MAR		مکانیسم MNAR	
		حجم نمونه ۱۰۰	حجم نمونه ۵۰۰	حجم نمونه ۱۰۰	حجم نمونه ۵۰۰	حجم نمونه ۱۰۰	حجم نمونه ۵۰۰
		RMSEA	RMSEA	RMSEA	RMSEA	RMSEA	RMSEA
با استفاده از الگوریتم EM (FIML)	٪۲	۰/۵۵۶	۰/۵۷۶	۰/۵۵۱	۰/۵۶۲	۰/۵۵۹	۰/۵۶۹
	٪۵	۰/۵۲۳	۰/۵۳۲	۰/۵۴۹	۰/۵۵۲	۰/۵۳۸	۰/۵۴۴
	٪۱۰	۰/۴۷۵	۰/۴۸۵	۰/۵۴۱	۰/۵۴۱	۰/۵۲۸	۰/۵۰۳
	٪۱۵	۰/۴۴۶	۰/۴۳۳	۰/۵۳۷	۰/۵۳۸	۰/۳۸۳	۰/۴۷۶
	٪۲۰	۰/۴۱۹	۰/۳۹۴	۰/۵۴۲	۰/۵۲۸	۰/۴۷۵	۰/۴۴۷
	٪۲۵	۰/۳۸۱	۰/۳۵۲	۰/۵۳۴	۰/۵۱۶	۰/۴۵۵	۰/۴۲۶
	٪۳۰	۰/۳۱۸	۰/۲۸۹	۰/۵۰۵	۰/۴۹۷	۰/۴۴۰	۰/۳۹۲
	٪۳۵	۰/۲۹۹	۰/۲۳۲	۰/۴۸۸	۰/۴۷۵	۰/۴۰۸	۰/۳۷۱
	٪۴۰	۰/۱۸۳	۰/۱۹۸	۰/۴۶۸	۰/۴۶۴	۰/۳۸۵	۰/۳۳۳
بدون استفاده از الگوریتم EM	٪۲	۰/۵۷۷	۰/۵۹۶	۰/۵۴۴	۰/۵۶۸	۰/۵۷	۰/۵۸۷
	٪۵	۰/۵۶۲	۰/۵۹۱	۰/۵۴۳	۰/۵۶۶	۰/۵۷۳	۰/۵۸
	٪۱۰	۰/۵۷۳	۰/۵۸۲	۰/۵۴۲	۰/۵۶	۰/۵۷۱	۰/۵۶۸
	٪۱۵	۰/۵۶۹	۰/۵۷۹	۰/۵۴۵	۰/۵۶۴	۰/۴۴۴	۰/۵۷۳
	٪۲۰	۰/۵۹۱	۰/۵۹۵	۰/۵۶۳	۰/۵۶۵	۰/۵۸۸	۰/۵۶۷
	٪۲۵	۰/۵۸۳	۰/۵۶۳	۰/۵۶	۰/۵۶۵	۰/۵۴۶	۰/۵۶۱
	٪۳۰	۰/۵۳۵	۰/۵۱۶	۰/۵۴۸	۰/۵۶۶	۰/۵۶۹	۰/۵۶
	٪۳۵	۰/۷۶۷	۰/۳۸۱	۰/۵۷	۰/۵۷۸	۰/۵۹۱	۰/۵۶۲
	٪۴۰	*	۰/۳۶۸	۰/۵۶۲	۰/۵۸۱	۰/۶۲۱	۰/۵۶۷

*Problem: Too few cases selected; program not run.

جدول ۲. نتایج مدل سه عاملی برای مکانیسم‌های MCAR ، MAR و MNAR به‌ازای حجم نمونه‌های ۱۰۰ و ۵۰۰.

	درصد مقادیر گمشده	مکانیسم MCAR		مکانیسم MAR		مکانیسم MNAR	
		حجم نمونه ۱۰۰	حجم نمونه ۵۰۰	حجم نمونه ۱۰۰	حجم نمونه ۵۰۰	حجم نمونه ۱۰۰	حجم نمونه ۵۰۰
		RMSEA	RMSEA	RMSEA	RMSEA	RMSEA	RMSEA
با استفاده از الگوریتم EM (FIML)	٪۲	۰/۱۰۳	۰/۱۰۷	۰/۱۰۱	۰/۱۰۵	۰/۱۷۸	۰/۱۰۵
	٪۵	۰/۱۰۳	۰/۱۰۱	۰/۰۸۸	۰/۰۹۳	۰/۰۹۱	۰/۰۹۶
	٪۱۰	۰/۰۹۶	۰/۰۸۹	۰/۰۸۹	۰/۰۹	۰/۰۱۰۶	۰/۰۸۱
	٪۱۵	۰/۱۰۵	۰/۰۸۷	۰/۰۸۵	۰/۰۸۶	۰/۱۱۴	۰/۰۷۶
	٪۲۰	۰/۰۷۹	۰/۰۷۴	۰/۰۸۷	۰/۰۸۶	۰/۱۰۹	۰/۰۶۴
	٪۲۵	۰/۰۶۹	۰/۰۶۸	۰/۰۸۳	۰/۰۸۱	۰/۱۰۲	۰/۰۶۴
	٪۳۰	۰/۰۶۹	۰/۰۶۶	۰/۰۶۷	۰/۰۷۶	۰/۰۸۹	۰/۰۵۸
	٪۳۵	۰/۰۶۹	۰/۰۵۵	۰/۰۶۶	۰/۰۷	۰/۰۹۸	۰/۰۵۴
	٪۴۰	۰/۰۹	۰/۰۵۱	۰/۰۶۴	۰/۰۶۸	۰/۰۸۹	۰/۰۴۸
بدون استفاده از الگوریتم EM	٪۲	۰/۰۹۶	۰/۱۱۳	۰/۱۱۱	۰/۱	۰/۰۹۷	۰/۱۱۰
	٪۵	۰/۱۱۰	۰/۰۹۷	۰/۰۹۸	۰/۰۹۶	۰/۱۲۳	۰/۱۰۹
	٪۱۰	۰/۱۰۳	۰/۰۸۷	۰/۰۹۹	۰/۰۹۷	۰/۱۳۲	۰/۱۲۱
	٪۱۵	۰/۱۷۲	۰/۱۲۰	۰/۱۲	۰/۰۹۹	۰/۱۱۹	۰/۱
	٪۲۰	۰/۳۵۹	۰/۱۴۶	۰/۱۹۲	۰/۱۰۳	۰/۱۵۱	۰/۰۹۶
	٪۲۵	*	۰/۱۳۷	۰/۱۴۱	۰/۱۱	۰/۱۳۵	۰/۱۲۵
	٪۳۰	*	۰/۱۷۲	۰/۱۵۷	۰/۱۳۲	۰/۱۶۱	۰/۱۲۷
	٪۳۵	*	۰/۳۶۸	۰/۱۲۶	۰/۱۲۱	۰/۱۷۲	۰/۱۳۴
	٪۴۰	*	۰	۰/۱۴	۰/۱۵	۰	۰/۱۳۹

*Problem: Too few cases selected; program not run.

بحث و نتیجه‌گیری

مدل‌یابی معادلات ساختاری (SEM) یکی از روش‌های آماری پرکاربرد در علوم انسانی و رفتاری می‌باشد که با استفاده از آن می‌توان به ارزیابی روابط پیچیده بین متغیرهای مشاهده شده و مکنون پرداخت. یکی از چالش‌های رایج در روش‌های تحلیل داده‌ها، از جمله SEM ، مسئله داده‌های گمشده است. روش‌های معمول در برخورد با مقادیر گمشده، در صورتی

که درصد مقادیر گمشده بالا باشد، می‌تواند منجر به سوگیری در نتایج شود. در این مقاله، FIML به عنوان روشی جهت بررسی داده‌های گمشده در تحلیل معادلات ساختاری از طریق نرم‌افزار لیزرل مورد بررسی قرار گرفت. با توجه به پیشینه پژوهش، تاکنون هیچ مطالعه جامعی با در نظر گرفتن هر سه مکانیسم مقادیر گمشده با استفاده از روش FIML انجام نشده بود.

در این مقاله، مقایسه‌ی نتایج در سه مکانیسم مقادیر گمشده و برای حجم نمونه و درصدهای متفاوت مقادیر گمشده نشان می‌دهد در حالت‌های اعمال کردن FIML نسبت به حالت‌هایی که FIML اعمال نشده است، شاخص RMSEA با افزایش درصد مقادیر گمشده کاهش می‌یابد. در مدل دو عاملی و برای حالت اعمال FIML، در مکانیسم MCAR نسبت به MAR و MNAR، شاخص RMSEA کاهش چشمگیرتری دارد. همچنین، در MNAR نسبت به MAR کاهش بیشتری در RMSEA رخ داده است. در مدل سه عاملی میزان کاهش RMSEA در مکانیسم MNAR در حجم نمونه ۵۰۰ نسبت به سایر حالتها بیشتر است. همچنین تاثیر افزایش حجم نمونه بر نتایج حاصل در مدل سه عاملی نسبت به مدل دو عاملی بیشتر است و نوساناتی که در حجم نمونه ۱۰۰ برای حالت اعمال FIML وجود دارد در حجم نمونه ۵۰۰ برطرف شده است. این مطالب بیانگر آن است که روش FIML می‌تواند برای مقادیر گمشده غیرقابل چشم‌پوشی (MNAR) که عدم لحاظ کردن آنها در تحلیل نسبت به مقادیر گمشده قابل چشم‌پوشی (MCAR و MAR) منجر به سوگیری‌های شدیدتری می‌شود به مثابه‌ی یک روش مطلوب مورد استفاده قرار گیرد. در هر سه مکانیسم، در حالت‌هایی که FIML اعمال نشده است، شاخص RMSEA با عدم ثبات، کاهش ناچیز و یا افزایش همراه است، که این عدم ثبات در مکانیسم MNAR بیش‌تر رخ می‌دهد. بنابراین مطابق با این نتایج می‌توان استدلال کرد در صورت وجود مقادیر گمشده و عدم لحاظ کردن آنها در تحلیل، نتایج ناپایا و دارای سوگیری خواهند بود؛ این سوگیری را می‌توان با استفاده از روش FIML برطرف کرد. در راستای بررسی‌های بیشتر می‌توان استفاده از FIML و بررسی عملکرد آن را در یک مدل معادلات ساختاری پیشنهاد

کرد. همچنین می‌توان عملکرد روش FIML را در مقایسه با روش‌های جانهی مورد بررسی قرار داد.

منابع

Allison, P. D. (2003). "Missing Data Techniques for Structural Equation Modeling". *Journal of Abnormal Psychology, 112(4), 545–557.*

Dempster, A. P., Laird, N. M., Rubin, D. B. (1977). "Maximum likelihood from incomplete data via the EM algorithm". *Journal of the Royal Statistical Society. Series B (Methodological), 39(1), 1-38.*

Enders, C. K. (2001). "A primer on maximum likelihood algorithms available for use with missing data". *Structrual Equation Modeling, 8(1), 128-141*.

Enders, C. K. (2010). *Applied missing data analysis.* The Guilford Press. New York, London.

Enders, C. K., Bandalos, D. L. (2001). "The relative performance of full information maximum likelihood estimation for missing data in structural equation models". *Structrual Equation Modeling, 8(3), 430–457*.

Finkbeiner, C. (1979). "Estimation for the multiple factor model when data are missing". *Psychometrika, 44(4), 409–420*.

Han, K. T., Guo, F. (2014). "Impact of violation of the missing-at-random assumption on full-information maximum likelihood method in multidimensional adaptive testing". *Practical Assessment, Research and Evaulation,* 19(2).

Hoyle, R. H. (2012). *Handbook of structural equation modeling*. The Guilford Press. New York, London.

Little, R. J. A., Rubin, D. B. (2002). *Statistical analysis with missing data,* 2nd Edition. New York: John Wiley.

Khine, M. S. (2013). *Application of structural equation modeling in educational research and practice*. Sense Publishers.

Muthén, B., Kaplan, D., Hollis, M. (1987). "On structural equation modeling with data that are not missing completely at random". *Psychometrika, 52(3), 431–462*.

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, 53–79.*

Rubin, D. B. (1976). "Inference and missing data". *Biometrika, 63, 581–59.*