Higher Education Letter

Institute for Research & Planning in Higher Education

National Organization for Educational Testing

Print ISSN: 2008-4617

Methods of Determining Factor Structure of Test Data Based on Traditional and New Measurement Approaches

Balal Izanloo[1]

1. Assistant Professor, Faculty of Psychology, Kharazmi University, Tehran, Iran; (Corresponding Author), Email: izan.b@khu.ac.ir

Article Info

Article Type: Research Article

Received: 2021/04/27

Revised: 2022/12/29

Accepted: 2023/01/04

Published online 2023/01/05

ABSTRACT

Objective: Usually, the dimensions or factors of a test are examined through the analysis of data obtained from its implementation by using the statistical methods of exploratory and confirmatory factor analysis. Over time, depending on different theoretical models, different methods have been presented to determine the number of dimensions or factors. The purpose of this article is to review the most wildly used major methods for this purpose and also examine their strengths and weaknesses.

Methods: To achieve this goal, the methods of specifying the number of data dimensions of tests, their description and performance in real data are discussed and finally, the conditions of using each of them are described.

Results: The application of different methods of determining dimensions or factors requires the researcher's insight and understanding of the basics and principles of these methods, the nature of the data and the conditions of them.

Conclusion: Different approaches to determine the dimensions provide reliable results only in conditions that are appropriate to their nature. Otherwise, the analysis performed with different approaches is not reliable.

Keywords: latent trait theory, item response theory, factor, dimension, construct, factor analysis.

Cite this article: : Izanloo, Balal (2022). Methods of Determining Factor Structure of Test Data Based on Traditional and new Measurement Approaches. *Higher Education Letter,* 15 (59): 59-90 Pages.

Publisher: Institute for Research & Planning in Higher Education & National Organization of Educational Testing

نامه آموزش عالی
شاپا چاپی: ۴۶۱۷-۲۰۰۸
سازمان سنجش آموزش کشور
مؤسسه پژوهش و برنامه‌ریزی آموزش عالی

روش‌های تعیین ساختار عاملی (ابعاد) داده‌های آزمون بر اساس رویکردهای اندازه‌گیری سنتی و جدید

بلال ایزانلو[۱]

۱. استادیار دانشکده روان‌شناسی، دانشگاه خوارزمی، تهران، ایران؛ (نویسنده مسئول)، پست الکترونیک: izan.b@khu.ac.ir

اطلاعات مقاله	چکیده

نوع مقاله:
مقاله پژوهشی

دریافت:
۱۴۰۰/۰۲/۰۷
اصلاح:
۱۴۰۱/۱۰/۰۸
پذیرش:
۱۴۰۱/۱۰/۱۴
انتشار :
۱۴۰۱/۱۰/۱۵

هدف: معمولاً ابعاد یا عامل‌های یک آزمون از طریق تحلیل داده‌های حاصل از اجرای آن به کمک روش‌های آماری تحلیل عاملی اکتشافی و تأییدی بررسی می‌شود. در طی زمان، بسته به مدل‌های نظری مختلف، روش‌های مختلفی برای تعیین تعداد ابعاد یا عامل‌ها ارائه شده است. هدف از نگارش مقاله حاضر، مرور منسجم روش‌های عمده پرکاربرد برای این منظور و بررسی نقاط قوت و ضعف آنها است.

روش پژوهش: برای رسیدن به این هدف، روش‌های مشخص کردن تعداد ابعاد داده‌های آزمون‌ها توصیف و عملکرد آنها در داده‌های واقعی مورد بحث قرار گرفته و درنهایت، شرایط استفاده از هر یک از آنها توصیف شده است.

یافته‌ها: کاربرد روش‌های مختلف تعیین ابعاد یا عامل‌ها نیازمند بینش و درک پژوهشگر از مبانی و اصول این روش‌ها، ماهیت داده‌ها و شرایط موجود در آنها است.

نتیجه‌گیری: رویکردهای مختلف تعیین ابعاد تنها در شرایط متناسب با ماهیت آنها نتایج قابل اطمینان ارائه می‌کنند در غیر این صورت، تحلیل انجام گرفته با رویکردهای مختلف قابل اطمینان نیست.

واژگان کلیدی: نظریه صفت پنهان، نظریه پاسخ سؤال، عامل، بعد، سازه، تحلیل عاملی.

اســـتناد: ایزانلو، بلال (۱۴۰۱). روش‌های تعیین ساختار عاملی (ابعاد) داده‌های آزمون بر اساس رویکردهای اندازه‌گیری سنتی و جدید. نامه آموزش عالی، ۱۵(۵۹)، صفحه ۵۹-۹۰

مقدمه

یکی از مهم‌ترین مسائل اندازه‌گیری مبتنی بر آزمون، تعیین تعداد عامل‌ها یا ابعاد موجود در داده‌ها است. دلیل تأکید بر تشخیص درست تعداد ابعاد یا عامل‌ها آن است که تفسیر داده‌ها به کمک عامل‌ها ساده‌تر می‌شود. معمولاً عامل‌ها همان سازه‌های پنهان هستند که بر اساس آنها می‌توان عملکرد افراد در آزمون‌ها را راحت‌تر خلاصه، تنظیم و توصیف کرد و در بسط و آزمون نظریه‌های جدید از آنها بهره گرفت. به‌علاوه، اهمیت ایجاد و توسعه ابزارهای اندازه‌گیری سازه‌ها در حوزه روان‌شناسی و علوم تربیتی نیز توجه به مبانی نظری و کاربردی این حوزه را بیش‌ازپیش روشن می‌سازد (ولیسر و همکاران[1]، ۲۰۰۰). موضوع تعیین تعداد عامل‌ها از سه جنبه آماری، روان‌سنجی و ریاضی قابل بررسی است. متأسفانه تمایز بین این سه رویکرد حتی در منابع اصلی تحلیل عاملی نیز چندان مورد توجه قرار نمی‌گیرد. تمایز بین این سه رویکرد در خصوص مسئله تعداد عامل‌ها را می‌توان به‌صورت زیر خلاصه کرد: الف) رویکردهای ریاضی بر ملاحظات وابستگی خطی بین بردارها استوارند و تعداد عناصر مستقلی که باید در داده‌ها وجود داشته باشند تا عامل و ضرایب کواریانس-عاملی برآورد شوند را بررسی می‌کنند. چنین ملاحظاتی در کارهای اولیه آلبرت و لیدرمن[2] و در برخوردهای جدید مباحث مربوط به شناسایی نشان داده شده‌اند؛ ب) رویکردهای روان‌سنجی روش‌هایی را مدنظر قرار می‌دهند که در آنها خطای اندازه‌گیری و نمونه‌گیری متغیرها تعداد عامل‌ها را تحت تأثیر قرار می‌دهد. چنین موضوعاتی در کارهای گاتمن و همکاران[3]، که آزمون معروف ریشه یک بر آنها استوار است، به‌خوبی نشان داده شده است؛ ج) آزمون‌های آماری به شیوه‌هایی تعلق دارند که در آنها نمونه‌گیری از آزمودنی‌ها بر تعداد عامل‌ها تأثیر دارد. شاید کارهای پیشگامانه لاولی و ماکس‌ول[4] نمونه‌های خوبی از این رویکرد باشند. امروزه این رویکرد عمدتاً با کارهای جورسکگ و همکاران وی[5] در ارتباط است. می‌توان روش‌های مختلف تعیین تعداد ابعاد را بر اساس این سه رویکرد دسته‌بندی کرد. مثلاً آزمون اسکری و آزمون کای دوی بارتلت[6] که برای تعیین تعداد عامل‌ها استفاده می‌شوند، گرچه منطق یکسانی دارند، ولی اولی منعکس‌کننده تغییرپذیری آماری (نمونه‌گیری آزمودنی) و دومی معمولاً دربرگیرنده تأثیرات روان‌سنجی (نمونه‌گیری متغیر) است. ازاین‌رو، لازم است هنگام استفاده از روش‌های مختلف برای تعیین تعداد عامل‌ها به نوع آزمون مورد نظر و اینکه بر منطق ریاضی، آماری یا روان‌سنجی برای تعیین تعداد عامل‌ها استوار است، توجه کرد (هورن و انگسترام[7]، ۱۹۷۹). در طول زمان، روش‌های زیادی برای مشخص کردن تعداد عامل‌ها ارائه شده است؛ به همین دلیل، در ادامه نخست روش‌های

1. Velicer et al
2. Albert & Lederman
3. Guttman et al
4. Lawley & Maxwell
5. Jöreskog et al
6. Bartlett›s Chi-Square Test
7. Horn & Engstrom

عمده‌ای که در عمل برای بررسی تعداد عامل‌ها یا ابعاد استفاده می‌شوند، توصیف و نکات عمده مربوط به هر یک از آنها ذکر شده و درنهایت طوری دسته‌بندی شده‌اند که فرد بتواند با توجه به ماهیت داده‌ها و روش مد نظر در خصوص انتخاب روش مناسب، تصمیم‌گیری کرده و دست به انتخاب بزند. در اینجا فقط بر روش‌های عمده پرکاربرد تأکید شده است. این روش‌ها عبارت‌اند از: ۱-مقادیر ویژه بزرگ‌تر از یک گاتمن-کایزر؛ ۲-نمودار اسکری کتل؛ ۳- روش مبتنی بر معنی‌داری ماتریس مانده‌ها و نیز مقایسه ساختارهای n عاملی با n+۱ عاملی؛ ۴-قاعده استخراج عامل‌ها تا جایی که تفسیرپذیر هستند و نیز قاعده مبتنی بر واریانس تبیین شده؛ ۵-ساختار بسیار ساده (VSS) ۶) کمینه متوسط همبستگی تفکیکی (MAP) ۷) تحلیل موازی؛ ۸- رویکرد مقایسه داده (CD)؛ ۹- مختصات بهینه و عامل شتاب؛ ۱۰- تحلیل خوشه؛ ۱۱- روش نموداری؛ ۱۲- روش کیفی؛ ۱۳- رویکرد برازش مدل؛ ۱۴- تحلیل عاملی غیرخطی؛ ۱۵- روش DIMTEST؛ ۱۶- روش‌های مبتنی بر استقلال موضعی. هر یک از این روش‌ها در ادامه به‌صورت خلاصه توصیف شده‌اند.

مبانی نظری و پیشینه پژوهش

۱) مقادیر ویژه بزرگ‌تر از یک گاتمن-کایزر[1]: گرچه به لحاظ ریاضی **مقادیر ویژه مثبت** یک ماتریس تعداد ابعاد لازم برای بازنمایی نمره‌های اصلی را بدون از دست دادن اطلاعات، تعیین می‌کنند و می‌توان مقادیر ویژه صفر را نادیده گرفت، ولی در تحلیل‌های مربوط به داده‌های حاصل از موقعیت‌های واقعی به‌ندرت مقادیر ویژه دقیقاً صفر می‌شوند و همین موضوع، مسئله تعیین تعداد عامل‌ها با این روش را دشوار می‌کند (هرچند که نادیده گرفتن مقادیر ویژه نزدیک صفر وسوسه‌انگیز است). روش مقادیر ویژه بزرگ‌تر از یک که از قدیم برای تعیین تعداد عامل‌ها استفاده می‌شود بر سه دلیل استوار است. اول آنکه طبق اثبات گاتمن، این روش در جامعه حد پایینی برای تعداد عامل‌های مشترک فراهم می‌کند؛ دومین دلیل این استدلال غیررسمی است: مؤلفه‌ای که میزان واریانس تبیین شده آن کوچک‌تر از واریانس یک متغیر (در شکل استاندارد) است، کمتر مورد علاقه است؛ سومین دلیل این جمله کایزر[2] است «اگر مقدار ویژه کمتر از یک باشد، پایایی نمره مؤلفه منفی می‌شود». شواهد حاکی از آن است که صرف نظر از دو مورد اول، مورد اخیر مبنای منطقی ندارد (کلیف[3]، ۱۹۸۸). درحالی‌که طبق ادعای گاتمن، قاعده مقادیر ویژه بزرگ‌تر از یک در جامعه حد پایین تعداد عامل‌های مشترک را فراهم می‌کند، مطالعات شبیه‌سازی‌شده، نشان‌دهنده آن است که استفاده از این قاعده تعداد عامل‌ها را بیش از آنچه هست برآورد می‌سازد (زویک و ولیسر[4]، ۱۹۸۲، ۱۹۸۶). دلیل این تناقض آن است که این روش، بسته به اینکه ماتریس حاصل مربوط به جامعه یا نمونه باشد نتایج آن به ترتیب به کم‌برآورد و بیش‌برآورد شدن عامل‌ها منجر می‌شود. ازاین‌رو، این روش به‌طور منظم دارای سوگیری است و بدون منطق است. چون خطای

1. Guttman-Kaiser criteri
2. kaiser
3. Cliff
4. Zwick & Velicer

نمونه‌گیری باعث می‌شود تعداد مقادیر ویژه بزرگ‌تر از یک بیشتر شوند، پس زمانی که در ماتریس‌های حاصل از نمونه از این روش برای تعیین تعداد عامل‌ها اســـتفاده می‌شـــود، تعداد عامل‌ها بیش از مقدار موجود برآورد خواهند شد. همچنین، منطق کایزر در خصوص ارتباط دادن پایایی مؤلفه‌ها به مقادیر ویژه بزرگ‌تر از یک بر اساس کاربرد نادرست یک فرمول متداول برای پایایی یک ترکیب است. این در حالی است که پایایی یک مؤلفه اصلی به پایایی نشانگرها یا اندازه‌های آن بستگی دارد نه مقادیر ویژه آن. به همین دلیل، منطق کایزر نمی‌تواند دارای مبنای منطقی باشد (برای بحث بیشتر به کلیف، ۱۹۸۸مراجعه کنید). به‌علاوه، وقتی حجم نمونه زیاد باشد استفاده از قاعده مقادیر ویژه تعداد عامل‌ها را بسیار کم برآورد می‌کند (هامفریز[1]، ۱۹۶۴). کایزر، کسی که ایده حفظ مؤلفه‌های دارای مقادیر ویژه بزرگ‌تر از یک را در یک تحلیل مؤلفه اصلی مطرح کرد، قطعاً هیچ‌وقت ادعا نکرده که این مؤلفه‌ها تنها مؤلفه‌هایی هستند که به لحاظ آماری اهمیت دارند. وی این اصل را تنها به‌عنوان یک قاعده سرانگشـــتی و مفید برای این منظور ارائه کرد، نه یک قاعده مصون از خطا. خلاصه‌ای از اظهار نظر وی به‌صورت زیر است:

«تصمیم گرفتم تنها مؤلفه‌های اصلی را به‌عنوان مؤلفه‌های مفید انتخاب کنم که نســـبت به یک آزمون سهم بیشتری از واریانس را تبیین می‌کنند... . در حال حاضر، بر اساس دیدگاه تحلیل عاملی، منظور من از تحلیل عاملی در اینجا مطالعه روابط متقابل ساختاری بین متغیرها است، چنین قاعده رفتاری مشکوک است. اما ... به نظر می‌رسد من در آنچه ظاهراً یک قاعده عالی برای تعیین تعداد عامل‌ها باشد، که می‌تواند به‌طور مناسب توسط مدل تحلیل عاملی دقیق تعیین شود، اشتباه کرده باشم ... من انکار نمی‌کنم که اطلاعات قبلی مربوط به یک مطالعه می‌تواند ... به یک پاسخ بهتر برای این مطالعه خاص منجر شود (کایزر، ۱۹۶۰ ص ۱؛ به نقل از شی‌کافت[2]، ۱۹۷۰)»

هیچ‌چیز در این عبارت بیان نمی‌کند که قاعده مقادیر ویژه بزرگ‌تر از یک، قانونی مقدس اســـت. با این حال، احتمالاً آنچه باعث گسترش این ایده شده که این اصل دارای مبنای ریاضی است، این بیان کایزر است «برای آنکه مؤلفه اصلی دارای پایایی کودر-ریچاردســـون مثبت باشد، لازم و کافی است که مقدار ویژه مربوط به آن بزرگ‌تر از یک باشـــد کایزر، ۱۹۶۰ ص ۲؛ همان منبع». این در حالی اســـت که روش کودر-ریچاردسون برای تعیین پایایی مؤلفه‌ها کاملاً نامناســـب است و پیش‌فرض‌های زیربنایی آن زیاد، محدود و تحقق آنها در مورد نمره‌های مؤلفه اصلی قطعاً حاصل نمی‌شـــود. شواهد تجربی، نشان‌دهنده آن است که مقادیر ویژه متغیرهای تصادفی که همراه تعدادی متغیر اندازه‌گیری شـــده تحلیل شده‌اند نیز بزرگ‌تر از یک می‌شود. پس این اصل نادرست است و نباید صرفاً بر اساس مقادیر ویژه بزرگ‌تر از یک تعداد عامل‌ها را مشخص کرد. بلکه عوامل دارای مقادیر ویژه کوچک‌تر از یک نیز می‌توانند دارای پایایی مناسب، مهم و تفسیرپذیر باشند. ازاین‌رو، به هیچ دلیل

1. humphreys
2. Shaycoft

نباید صرفاً بر اساس مقادیر ویژه در مورد اهمیت، معنی‌داری آماری و پایایی مؤلفه‌ها قضاوت کرد؛ چراکه وقتی متغیرهای تصادفی که فقط متشکل از واریانس خطا هستند و حتی واریانس اختصاصی نیز ندارند می‌توانند به مقادیر ویژه بزرگ‌تر از یک منجر شـــوند، استفاده از این قاعده به‌شدت نفی می‌شود. به‌علاوه، هیچ‌وقت قبل از بررسی ماتریس بارهای چرخش یافته (متعامد و مایل یا هر دو) در مورد تعداد عامل‌ها صرفاً بر اساس مقادیر ویژه بزرگ‌تر از یک، قبل از چرخش قضاوت نکنید، چراکه این نوع تصمیم‌گیری درنهایت به نتایج و به دنبال آن تفسیرهای نادرستی منجر می‌شود.

در تحلیل مؤلفه‌های اصلی به‌طور اجتناب‌ناپذیری واریانس یگانه[1] (که متشکل از واریانس اختصاصی + واریانس خطا است) با واریانس مشترک ترکیب می‌شود. یکی از راهکارها برای برطرف کردن این مشکل این است که در ماتریس همبستگی تحلیل شده به جای یک روی قطر اصلی، ضرایب پایایی متغیرها یا برآورد مناسبی از اشـــتراک آنها مثل SMC قرار گیرد. استفاده از این روش مشـــکل واریانس خطا را که در متغیرهای تصادفی وجود دارد حل می‌کند (رایت‌ولت و ون‌هوت[2]، ۱۹۹۳). در این شرایط اگر بارهای عاملی هر متغیری نزدیک به صفر باشد، پژوهشگر را قانع می‌کند که آن متغیر در اصل یا یک متغیر تصادفی است یا متغیری است که فقط حاوی واریانس یگانه است و باید از تحلیل حذف شود. استفاده از مقادیر اشتراک روی قطر اصلی به جای یک، که در تحلیل مؤلفه‌های اصلی استفاده می‌شود، در تحلیل محورهای اصلی، که به آن تحلیل عاملی مشترک نیز می‌گویند، کاربرد دارد. در این روش، پژوهشـــگر در مورد تعداد عامل‌ها تصمیم می‌گیرد نه مقادیر ویژه، و نظریه‌ای که مطالعه بر اساس آن طراحی و هدایت می‌شود نقش بارزی در این خصوص دارد. البته توجه کنیم که استفاده از اشتراک به جای یک روی قطر اصلی باز هم چیزی در مورد پایایی عامل‌ها به ما نمی‌گوید. اکنون این پرسش مطرح می‌شود که آیا این امکان وجود دارد که چون خطاهای اندازه‌گیری بر مقادیر ویژه تأثیر ندارند، ولی خطای نمونه‌گیری (منظور خطای مربوط به نمونه افراد است نه خطای مربوط به نمونه سؤال‌ها) روی آنها تأثیرگذار باشد؟ برخی از افراد تمایل دارند مقادیر ویژه نزدیک صفر را به خطای نمونه‌گیری نسبت دهند. اما واقعاً منظور از خطای نمونه‌گیری مقادیر ویژه چیست؟ در پاسخ باید گفت که کاهش بعد در صورتی رخ می‌دهد که مقدار ویژه دقیقاً صفر باشد. این حالت، زمانی رخ می‌دهد که یکی از متغیرها ترکیب خطی یک یا چند متغیر دیگر باشد. در این شرایط، ضریب همبستگی مرتبه صفر یا چندگانه در مجموعه متغیرهای موجود در تحلیل برابر یک خواهد بود. اگر انحراف یک مقدار ویژه از صفر به نمونه‌گیری قابل نسبت باشد، به این معنی است که مقدار آن در جامعه صفر اســـت، ولی نمونه‌ای که آماره‌ها بر اســـاس آن محاسبه شده‌اند طوری است که مقدار ویژه دقیقاً صفر نیست. به بیان دیگر، درحالی‌که همبستگی دو متغیر در جامعه یک است به‌طوری‌که هر فرد در جامعه دقیقاً روی خط رگرسیون قرار دارد، ولی به‌محض اینکه شما زیرمجموعه‌ای از این جامعه را انتخاب

1. Unique
2. Rietveld & Van Hout

کنید و به نمره‌های آنها در تمام متغیرها دست نزنید، چون در این حالت صحبت از خطای نمونه‌گیری است نه خطای اندازه‌گیری پس برخی از افراد در زیرمجموعه انتخاب‌شده دیگر روی خط رگرسیون قرار نمی‌گیرند. این در حالی است که برای هیچ نمونه‌ای، صرف نمونه‌گیری و خطای نمونه‌گیری همراه آن، همبستگی یک در جامعه کمتر از یک نمی‌شود. احتمالاً خطای نمونه‌گیری نمی‌تواند همبستگی کامل را کاهش داده و به مقادیر ویژه بالای صفر منجر شود. اصطلاح اهمیت یا معنی‌داری آماری یک مؤلفه به معنی آن است که مقدار ویژه آن به‌طور معنی‌داری از صفر بزرگ‌تر است. ولی اگر مؤلفه‌ای وجود داشته باشد پس مقدار ویژه آن نیز باید بزرگ‌تر از صفر باشد و زمانی که مقدار ویژه مربوط به یک مؤلفه در نمونه از صفر بزرگ‌تر است پس باید مقدار آن نیز در جامعه بزرگ‌تر باشد. به بیان دیگر، مؤلفه‌ای که در نمونه وجود دارد باید در جامعه نیز وجود داشته باشد. نه خطای نمونه‌گیری و نه خطای اندازه‌گیری نمی‌توانند بعد جدیدی ایجاد کنند که قبلاً در جامعه وجود نداشته باشد. با این حال، این دو نوع خطا بر روابط بین مؤلفه‌ها و متغیرهای اصلی که مؤلفه‌ها از آنها استخراج می‌شود تأثیـــر کمی و کیفی دارند. به‌علاوه، این احتمال وجود دارد که در اثر خطای نمونه‌گیری ابعادی که در جامعه حضور دارند در نمونه وجود نداشته باشند. توجه کنیم که مقادیر ویژه نشانه قطعی در مورد این انحراف نیز به ما ارائه نمی‌کنند. اگر مقادیر ویژه، چیز زیادی در مورد پایایی و معنی‌داری آماری مؤلفه یا عامل نمی‌دهند آیا دست‌کم چیزی در مورد اهمیت مؤلفه یا عامل ارائه می‌دهند؟ باز جواب خیر است؛ چراکه ممکن است مؤلفه یا عاملی با مقدار ویژه بسیار کوچک باز هم ارزشمند باشد ولو اینکه درصد بسیار کوچکی از واریانس را تبیین کند. همچنین، توجه شود که زیاد بودن متغیرها در تحلیل می‌تواند تعداد عامل‌ها را افزایش داده و درنتیجه هر عامل فقط بخش کوچکی از واریانس تمام متغیرها را تبیین کند. با توجه به این مطالب، تصمیم‌گیری در خصوص تعداد عامل‌ها باید بر اساس دانش فرد در مورد تعداد و ماهیت متغیرهای موجود در تحلیل، درک هدف مطالعه، بینش آماری و عقل ســـلیم باشد. به بیان دیگر این پژوهشگر است که باید مسئولیت تصمیم خود را بر عهده بگیرد تا اینکه اجازه دهد مقادیر ویژه این کار را برای وی انجام دهند (شی‌کافت، ۱۹۷۰).

با توجه به نقطه‌ضعف‌های روش مقادیر ویژه بزرگ‌تر از یک، یکی از روش‌هایی که امروزه با گسترش نرم‌افزارهای کامپیوتری در کنار سایر روش‌ها به‌طور مکرر توصیه شده، تحلیل موازی است (سیلورستاین[1]، ۱۹۸۷). در این روش، داده‌های تصادفی با همان تعداد متغیر و آزمودنی که در داده‌های واقعی وجود دارند، ایجاد شده و مقادیر ویژه داده‌های واقعی و شبیه‌سازی‌شـــده، مشـــخص می‌شـــود. آنگاه مقادیر ویژه داده‌های واقعی که مقدار آنها بزرگ‌تر از مقادیر ویژه داده‌های تصادفی است به‌عنوان تعداد عامل‌ها در نظر گرفته می‌شوند. درصورتی‌که برای تعیین تعداد عامل‌ها از تحلیل موازی مبتنی بر تحلیل مؤلفه‌های اصلی (که مد نظر هورن[2] (۱۹۶۵) بوده و در آن عدد یک روی قطر اصلی ماتریس همبستگی داده‌ها قرار می‌گیرد) به جای تحلیل موازی مبتنی بر تحلیل

1. Silverstein
2. Horn

عاملی مشترک (که در آن اشتراک متغیرها، که به روش‌های مختلف قابل برآورد است، روی قطر اصلی ماتریس همبستگی داده‌ها قرار می‌گیرد) استفاده شود، تعداد عامل‌ها کم‌برآورد[1] می‌شود. استفاده از مجذور همبستگی چندگانه به‌عنوان برآورد اشتراک، نتایجی به دست می‌دهد که بسیار نزدیک به نتایج حاصل از تحلیل بیشینه درســـت‌نمایی اســـت. درحالی‌که تحلیل مؤلفه‌های اصلی به دنبال تبیین کل واریانس توسط مؤلفه‌ها است و مؤلفه‌ها بین اجزای مختلف واریانس تمایز قائل نمی‌شـــوند، تحلیل عاملی مشترک به دنبال تفکیک واریانس مشترک از واریانس یگانه (واریانس اختصاصی+ واریانس خطا) است. بر اساس رویکرد اخیر، عامل‌ها باید واریانس مشترک را تبیین کنند. تحلیل عاملی مشترک بنا به دلایل زیر بر تحلیل مؤلفه‌های اصلی برتری دارد (رایت‌ولت و ون‌هوت، ۱۹۹۳): الف) تنها در مواردی خاص و نادر می‌توان فرض کرد که متغیرها از خطا آزادند. همچنین، انتظـــار نداریم که بتوان تمام متغیرهـــا را از روی عامل‌ها پیش‌بینی کرد. به بیان دیگر، متغیرها معمولاً دارای مقداری واریانس اختصاصی یا غیرمشـــترک نیز هســـتند؛ ب) بار متغیرها روی مؤلفه‌ها نسبت به بار آنها روی عامل‌ها بزرگ‌تر است. این اثر که تورم نامیده می‌شود، باعث شده افراد بارهای حاصل از تحلیل عاملی را ترجیح دهند. زیرا در مواردی یک ماتریس از همبستگی‌های غیر معنی‌دار درحالی‌که بارهای عاملی آن پایین یا غیر معنی‌دار است به بارهای مؤلفه‌ای معنی‌دار منجر شده است؛ ج) اگر اشتراک‌ها برابر یک باشند، که نقطه شروع برای تحلیل مؤلفه است، تحلیل عاملی به تحلیل مؤلفه منجر می‌شود.

۲) نمودار اسکری کتل: در این روش، مقادیر ویژه ماتریس همبستگی کاهش‌نیافته یا ماتریس کاهش‌یافته (که روی قطر اصلی آن به جای یک از اشـــتراک متغیرها اســـتفاده شده) به ترتیب نزولی روی یک نمودار ترسیم می‌شوند. ســـپس این نمودار برای مشخص کردن آخرین افت قابل توجه در مقادیر ویژه بررسی می‌شود. این بررسی به‌صورت چشمی بوده و جایی که منحنی دارای آرنج یا وضعیت لولامانند است به‌عنوان نقطه مورد نظر، در نظر گرفته می‌شود. درنهایت مقادیر ویژه، قبل از این نقطه به‌عنوان تعداد عامل‌ها در نظر گرفته می‌شود که بزرگ‌ترین مقادیر ویژه ماتریس مورد نظر هستند. منطق این آزمون، آن است که چندعاملی اصلی اکثر واریانس را تبیین می‌کند و واریانس خطا توسط عامل‌های جزئی توصیف می‌شود. گاهی عامل‌های اصلی را به یک صخره و عامل‌های جزئی را به ســـنگریزه‌های پایین صخره تشبیه می‌کنند (کورتنی و گوردون[2]، ۲۰۱۳). اگرچه این روش، بیشتر برای مقادیر ویژه ماتریس‌های همبستگی کاهش‌نیافته (که روی قطر اصلی آنها مقادیر یک قرار دارند) اســـتفاده می‌شود، ولی در ارتباط با تحلیل عاملی برای ماتریس‌های کاهش‌یافته نیز مورد استفاده قرار می‌گیرد. صرف نظر از نوع ماتریس، مشکل اصلی این روش، ذهنی بودن تفسیر آن است؛ به این معنی که تعریف مشخص و عینی در خصوص اینکه افت قابل توجه از نظر مقدار به چه معنی است، وجود ندارد، به‌خصوص زمانی که نمودار دارای نقطه زانو یا لولامانندی نباشد. به‌علاوه، گاهی ماتریس‌هایی وجود دارند که ترسیم مقادیر ویژه

1. underestimate
2. Courtney & Gordon

آنها افت قابل توجه مشخصی را نشان نمی‌دهد. عملکرد خوب این روش در شرایطی است که عامل‌های مشترک قوی در داده‌ها وجود داشـــته باشند (فابریگر و همکاران[1]، ۱۹۹۹). نتایج مطالعات شبیه‌سازی، نشان‌دهنده آن اســت که این روش نسبت به شاخص کایزر دقیق‌تر بوده و تغییرپذیری کمتری دارد. مقایسه نتایج این روش با میانگین دو داور آموزش‌دیده در داده‌های متشـــکل از ۴۸۰ مجموعه داده، نشـــان داد که این روش در ۴۱/۷ درصد از موارد تعداد عامل‌ها را به‌درستی مشخص می‌کند. درحالی‌که ملاک کایزر در هیچ‌یک از موارد تعداد عامل‌ها را به‌درستی نشان نداده و به بیش‌برآورد شدن تعداد عامل‌ها در نمونه منجر می‌شد (کورتنی و گوردون، ۲۰۱۳). برای یکسان شدن نتایج آزمون اسکری با آزمون بارتلت، باید سطح آلفای آزمون بارتلت را به جای سطح قراردادی فرضاً ۰/۰۱ به سطح پایین‌تر (مثلاً ۰/۰۰۳ یا پایین‌تر) کشاند. در نمونه‌ای از متغیرها که در یک طرح خاص استفاده شده‌اند، آزمون اسکری در نمونه‌های کوچک نسبت به نمونه‌های بزرگ معمولاً تعداد عامل‌ها را بیش از آنچه هست نشان می‌دهد. در مقابل آزمون بارتلت در نمونه‌های کوچک تعداد عامل‌ها را کمتر از آنچه در نمونه بزرگ هست برآورد می‌کند.

۳) روش مبتنی بر معنی‌داری ماتریس مانده‌ها و نیز مقایســـه ســـاختارهای n عاملی و n+۱ عاملی: در حالت اول، استخراج عامل‌ها تا جایی صورت می‌گیرد که کای دوی ماتریس مانده‌ها معنی‌دار نشود و در حالت دوم، استخراج عامل‌ها تا جایی انجام می‌شود که کای دوی مدل n عاملی نسبت به مدل n+۱ عاملی معنی‌دار نشود (رول[2]، ۲۰۱۸).

۴) قاعده اســـتخراج عامل‌ها تا جایی که تفسیرپذیر هستند و نیز قاعده مبتنی بر واریانس تبیین شده: قاعده اســـتخراج تعداد عامل‌ها تا جایی که تفسیرپذیر هســـتند به این معنی است که پژوهشگر تا زمانی که بتواند عامل‌های تفسیرپذیر پیدا کند به استخراج عامل‌ها ادامه می‌دهد. روشن است که سوگیری پژوهشگر در اینجا دخالت داشته و می‌تواند بر نتایج مؤثر باشد، به همین دلیل، بهتر است همراه با سایر روش‌ها مورد استفاده قرار گیرد. قاعده مبتنی بر واریانس تبیین شده عامل‌هایی که با هم در حدود ۷۰ تا ۸۰ درصد از واریانس را تبیین کنند به‌عنوان تعداد عامل‌ها در نظر می‌گیرد (همان منبع). با این حال پذیرش راه حل‌هایی که دست‌کم ۶۰ درصد (هیر و همکاران[3]، ۲۰۱۴) یا حتی ۵۰ درصد (استراینر[4]، ۱۹۹۴؛ تابکنیک و فیدل[5]، ۲۰۱۴) از واریانس را تبیین کنند نیز مورد تأکید قرار گرفته است. نتایج یک مطالعه فراتحلیل (پترسون[6]، ۲۰۰۰) روی ۸۰۳ مقاله در خصوص واریانس تبیین شده و بارهای عاملی در تحلیل عاملی نشان داد که متوسط واریانس تبیین شده ۵۶/۶٪ و متوسط ارزش مطلق بارهای عاملی ۰/۳۲ است. در مقابل تحلیل عاملی داده‌های مصنوعی (تصادفی) دارای

1. Fabrigar et al
2. Revelle
3. Hair et al
4. Streiner
5. Tabachnick & Fidell
6. Peterson

ساختار کلی مشابه داده‌های مشاهده‌شده در فراتحلیل ٪۵۰ از واریانس داده‌ها را تبیین کرد و به متوسط ارزش مطلق بار عاملی ۰/۲۱ (متوسط ارزش مطلق بار عاملی در تحلیل مؤلفه ۰/۲۲ و در تحلیل عاملی مشترک ۰/۱۸ بود) منجر شد. مقادیر اخیر نشان‌دهنده آن است که نتایج حاصل از بسیاری از تحلیل عاملی‌های انجام‌شده روی داده‌های واقعی مشکوک است و معنی‌دار نیست.

۵) ساختار بسیار ساده[1] (VSS): این روش که توسط رول و روکلین[2] (۱۹۷۹) پیشنهاد شده، راه‌حل‌های مختلف را به‌ازای سطوح مختلف پیچیدگی متغیر و تعداد عامل‌های متفاوت با استفاده از برازش تحلیل عاملی‌های مختلف که فقط از نظر تعداد عامل با هم فرق دارند بر اساس شاخص خوبی برازش ساختار بسیار ساده[3] مقایسه می‌کند. راه حل عاملی که به‌ازای آن شاخص ساختار بسیار ساده بیشینه می‌شود به‌عنوان تعداد عامل‌های بهینه برای استخراج در نظر گرفته می‌شوند. معمولاً اکثر پژوهشگران تحلیل عاملی اکتشافی تمایل دارند برون‌داد عاملی را با تمرکز بر بزرگ‌ترین بارها در ماتریس الگوی عاملی مربوط به متغیرها تفسیر کنند و بارهای کوچک را نادیده بگیرند. روش ساختار بسیار ساده این تمایل پژوهشی را عملیاتی کرده است. فرض کنید در ماتریس بارهای عاملی متغیرها، به‌جز بزرگ‌ترین بارهای عاملی هر متغیر، که می‌تواند بین یک تا تعداد عامل‌ها باشد، بقیه بارهای عاملی صفر شوند. سپس در این شرایط برازش عامل‌های مختلف که فقط از نظر تعداد عامل باهم فرق دارند را بر اساس شاخص خوبی برازش ساختار بسیار ساده مقایسه کنیم. راه حل عاملی که به‌ازای آن شاخص ساختار بسیار ساده بیشینه می‌شود، به‌عنوان تعداد عامل‌های بهینه برای استخراج در نظر گرفته می‌شوند. در شرایطی که هر سؤال فقط روی یک عامل دارای بار باشد ساختار از نوع ساده است، اگر تعداد بارهای عاملی بزرگ‌تر هر متغیر بیشتر از یک مورد در نظر گرفته شوند، ساختار به سمت پیچیده شدن پیش می‌رود، روشن است که در این شرایط هر متغیر با بیش از یک عامل ارتباط داشته و درنتیجه شاخص پیچیدگی نیز افزایش می‌یابد.

۶) کمینه متوسط همبستگی تفکیکی[4] (MAP): متوسط مجذور همبستگی‌های غیرقطری یک ماتریس همبستگی قبل از اینکه عاملی از آن استخراج شود را در نظر بگیرید. در گام اول، اولین مؤلفه یا عامل استخراج می‌شود و متوسط مجذور همبستگی‌های غیرقطری ماتریس (در بسته psych در R با MAPr2 نشان داده می‌شود) محاسبه می‌شود. در گام دوم با استخراج دومین مؤلفه یا عامل دوباره متوسط مجذور همبستگی‌های غیرقطری محاسبه می‌شود. این فرایند می‌تواند به ازای k-۱ مرحله تکرار شود (k برابر با تعداد متغیرهای موجود در ماتریس است). متوسط مجذور همبستگی‌های غیرقطری ماتریس در هر مرحله نوعی همبستگی تفکیکی است. با محاسبه تمام مقادیر متوسط مجذور همبستگی‌های غیرقطری در هر مرحله و مرتب کردن آنها

1. Very Simple Structure (VSS)
2. Revelle & Rocklin
3. Very Simple Structure index of goodness of fit
4. Velicer's Minimum Average Partial correlation

مرحله‌ای که متوسط مجذور همبستگی تفکیکی آن نسبت به بقیه کوچک‌تر باشد تعیین‌کننده تعداد عامل‌ها است. بر اساس این روش، تا زمانی که واریانس موجود در ماتریس همبستگی نشان‌دهنده واریانس منظم باشد تا خطا، مؤلفه استخراج می‌شود (کورتنی و گوردون، ۲۰۱۳). اگرچه این شیوه به لحاظ روش‌شناسی با تحلیل مؤلفه‌های اصلی در ارتباط است، ولی در چندین مطالعه شبیه‌سازی، نشان داده شده که روش MAP در تعیین تعداد عامل‌هایی (احتمالاً منظور عامل‌های مشترک است) که باید حفظ شوند نیز کاملاً خوب عمل می‌کند. به‌منظور بهبود دقت این روش در مطالعات شبیه‌سازی، اصلاحات گوناگونی روی شاخص‌های این روش صورت گرفته که یکی از این اصلاحات توان چهارم آن اســت. یعنی متوســط مجذور همبستگی غیرقطری به توان چهار ($MAPr^4$، که در بســته psych در R با MAPr4 نشان داده می‌شود). با وجود این تعدیل نتایج مطالعات نشـــان‌دهنده آن است که نسخه $MAPr^2$ آن نسبت به نسخه $MAPr^4$ آن در داده‌های پیوسته عملکرد بهتری دارد، در این حالت میزان دقت $MAPr^2$ در برابر $MAPr^4$ ۶۵٪ در برابر ۵۵٪ است (همان). کاربرد آزمون MAP در داده‌های رتبه‌ای نیز نشـــان می‌دهد که وقتی ماتریس استفاده شده برای این آزمون در داده‌های رتبه‌ای به جای همبستگی پیرسون، که رابطه بین متغیرهای طبقه‌ای را کمتر از آنچه هست برآورد می‌کند، همبستگی پلی کوریک باشـــد، دقت برآورد $MAPr^2$ از $MAPr^4$ بیشـــتر است. برآورد تعداد ابعاد بر اساس این روش در دو وضعیت دارای سوگیری خواهد بود؛ اول، زمانی که بارهای عاملی در حد پایین (۰/۴) و متوسط (۰/۵۵) باشند و دوم، زمانی که تعداد متغیرهای هر عامل کم (کوچک‌تر مســـاوی شـــش) باشد (گریدو و همکاران[1]، ۲۰۱۱). به‌علاوه، در ســـاختارهای عاملی مایل به متوسط و کم، توان و دقت MAP بهتر از روش تحلیل موازی (که در ادامه توصیف شده) است، ولی کارایی هر دو روش در تعیین تعداد عامل‌ها، در ساختارهای عاملی به‌شدت مایل به ضعیف است (کارُن[2]، ۲۰۱۹).

۷) تحلیل موازی[3]: یکی از روش‌های تعیین تعداد ابعاد که به‌شدت برای این منظور توصیه شده، تحلیل موازی هورن (۱۹۶۵) اســـت. در این روش، تعداد زیادی ماتریـــس داده تصادفی که هر یک به اندازه داده‌های واقعی آزمودنی و متغیر دارد ایجاد می‌شود. عامل‌های موجود در داده‌های واقعی تا جایی حفظ می‌شوند که مقادیر ویژه آنها از میانگین مقادیر ویژه حاصل از ماتریس‌های داده‌های تصادفی بزرگ‌تر باشند. بر اساس ایده هورن، برآورد کران پایین ریشه پنهان یک گاتمن برای تعیین رتبه یک ماتریس همبستگی، به‌عنوان کران بالای روان‌سنجی در تعیین تعداد عامل‌ها پذیرفته شده است. ازاین‌رو، برای تعیین رتبه یک ماتریس نمونه باید ریشه‌های پنهانی که به خطای نمونه‌گیری قابل اختصاص هستند را از تعداد مؤلفه‌ها کم کرد. هورن معتقد بود به دلیل خطای نمونه‌گیری در محاســـبه ریشـــه‌های پنهان، برخی مؤلفه‌های حاصل از متغیرهای ناهمبسته در جامعه واقعی

1. Garrido et al
2. Caron
3. Parallel analysis

می‌توانند مقدار ویژه بزرگ‌تر از یک داشته باشند. درنتیجه، وی روش تحلیل موازی را پیشنهاد کرد که واریانس ناشی از خطای نمونه‌گیری را لحاظ می‌کند. پس تحلیل موازی جایگزین نمونه‌ای قاعده مقادیر ویژه بزرگ‌تر از یک است. برای افزایش دقت این روش در مطالعات شبیه‌سازی اصلاحات مختلفی پیشنهاد شده است. مثلاً اجرای تحلیل موازی با استفاده از استخراج تحلیل مؤلفه، بر اساس همبستگی‌های گشتاوری پیرسون و معیار میانگین مقادیر ویژه، در دامنه وسیعی از داده‌های شرایط مختلف به‌خوبی عمل می‌کند، به‌طوری‌که میزان دقت ۷۶/۴۲٪ است (گریدو و همکاران، ۲۰۱۱). به‌علاوه می‌توان شیوه تولید متغیرهای ملاک تصادفی را با استفاده از جایگشت‌های تصادفی ستون‌های ماتریس داده‌های واقعی بهبود بخشید. این اصلاح مناسب‌تر به نظر می‌رسد؛ چراکه همان سطح کجی و تعداد مقوله که در داده‌های واقعی وجود دارند در داده‌های تصادفی نیز حفظ می‌شوند. این تعدیل در تحلیل موازی باعث شده که بتوان این روش را با استفاده از برآورد مؤلفه‌های اصلی، همبستگی‌های پُلی‌کُریک و ملاک مقدار ویژه متوسط برای داده‌های رتبه‌ای نیز اجرا کرد. توصیه در خصوص استفاده از همبستگی پلی کوریک به جای همبستگی پیرسون به این خاطر است که استفاده از همبستگی پیرسون برای داده‌های رتبه‌ای باعث حفظ عامل‌های حاصل از دشواری شده و در سطوح زیاد کجی نتیجه حاصل از آن دقیق نیست (همان). برخلاف ادعای پژوهشگران مختلف در خصوص حساسیت تحلیل موازی نسبت به پیش فرض‌های توزیعی، نتایج نشان‌دهنده عدم حساسیت این روش نسبت به این پیش‌فرض‌هاست (دینو[1]، ۲۰۰۹). به‌علاوه، در بین روش‌های مختلف تعیین تعداد عامل‌ها (شامل کمینه متوسط همبستگی تفکیکی، آزمون نسبت درست‌نمایی، ملاک آگاهی آکایک[2] (AIC)، ملاک آگاهی بیزین[3] (BIC) شوارتز[4]، تحلیل موازی مبتنی بر تحلیل مؤلفه، تحلیل موازی مبتنی بر عامل مشترک، آزمون خطای استاندارد اسکری مبتنی بر تحلیل مؤلفه و تحلیل عامل مشترک و مقادیر ویژه بزرگ‌تر از یک و مقادیر ویژه بزرگ‌تر از متوسط)، تحلیل موازی مبتنی بر عامل مشترک در شرایط مختلف (شامل تعداد عامل‌ها، نسبت متغیر به عامل، میزان اشتراک، حجم نمونه و همبستگی متفاوت بین عامل‌ها) نسبت به بقیه روش‌های ذکر شده، همواره دقیق‌ترین روش است و به‌عنوان اولین گزینه پیشنهاد می‌شود (پیرسون و همکاران[5]، ۲۰۱۳).

۸) رویکرد مقایسه داده[6] (CD): تحلیل موازی برای غلبه بر مشکل خطای نمونه‌گیری توسعه یافت. راشو و راش[7] (۲۰۱۲) برای بهبود روش تحلیل موازی رویکرد مقایسه داده‌ها را بسط دادند. منطق این روش، آن است که به جای ایجاد مجموعه داده‌های تصادفی که فقط خطای نمونه‌گیری را لحاظ می‌کنند، چندین مجموعه داده

1. Dinno
2. Akaike information criterion
3. Bayesian information criterion
4. Schwartz
5. Pearson et al
6. Comparison data
7. Ruscio & Roche

با ساختار عاملی معین تحلیل می‌شوند تا مشخص شود کدام‌یک از آنها نیمرخ مقادیر ویژه داده‌های واقعی را بازتولید می‌کند. این روش درواقع نسخه به‌روز شده تحلیل موازی است که تلاش می‌کند به جای ایجاد داده‌های تصادفی ماتریس همبســـتگی داده‌های مشاهده‌شـــده را بازتولید کند. ابتدا یک مجموعه داده با فرض ساختار یک‌عاملی ایجاد می‌شود. توجه کنیم که در این شرایط، مقادیر ویژه حاصل از این روش با مقادیر ویژه داده‌های واقعی قابل مقایســـه است، اینکه این ساختار تک‌بعدی تا چه حد مقادیر ویژه را بازتولید می‌کند با روش‌های آماری قابل بررسی است. به همین صورت می‌توان تعداد عامل‌ها را اضافه کرد تا زمانی که بازتولید مقادیر ویژه مشاهده‌شده به‌طور معنی‌داری بهبود پیدا نکند. بر این اساس، می‌توان تعداد عامل‌ها را مشخص کرد (راشو و راش، ۲۰۱۲). نتایج مطالعات شبیه‌سازی، نشان‌دهنده عملکرد خوب این روش نسبت به سایر روش‌ها، ازجمله تحلیل موازی معمولی است. برتری این روش نه‌تنها در لحاظ کردن خطای تصادفی است، بلکه ساختار عاملی و توزیع چندمتغیری متغیرها (سؤال‌ها) را نیز لحاظ می‌کند. بر اساس این روش می‌توان در ۸۷/۱۴٪ از موارد، تعداد عامل‌ها را به‌درستی پیش‌بینی کرد. البته باید خاطرنشان کرد که ضریب همبستگی استفاده شده در این مطالعه همبستگی پیرسون و ساختار عاملی داده‌های شبیه‌سازی‌شده حداکثر پنج‌عاملی بوده است. همچنین، راشو (به نقل از کورتنی و گوردون، ۲۰۱۳) معتقد بود به دلیل پیش‌فرض نرمال بودن در همبستگی پلی‌کوریک باید به جای این ضریب از همبســـتگی رتبه‌ای اسپیرمن برای این منظور اســـتفاده شود. از دیدگاه وی، برای اجرای روش مقایسه داده در شرایطی که متغیرها رتبه‌ای هستند باید از همبستگی رتبه‌ای اسپیرمن به جای همبستگی پلی‌کوریک برای این منظور استفاده شود، چون این همبستگی بر نرمال بودن توزیع متغیرها استوار نیست.

۹) مختصات بهینه و عامل شـــتاب[1]: تفســـیر آزمون اســـکری کتل به دلیل ماهیت نموداری آن ذهنی است. درنتیجه، نمی‌توان تعداد مؤلفه‌ها را با این روش به‌طور قطع تعیین کرد. برای کاهش این نقطه‌ضعف، دو روش غیرنموداری برای این آزمون ارائه شده است. الف) روش مختصات بهینه تلاش می‌کند i امین مقدار ویژه را (p، ... ،۲ ،۱=i) بر اساس رگرسیون خطی با استفاده از مختصات آخرین مقدار ویژه (p) و مختصات i+۱ امین مقدار ویژه پیش‌بینی کند. به‌این‌ترتیب p-۲ خط رگرسیون به دست می‌آیند (رایچی و همکاران[2]، ۲۰۱۳). به بیان دیگر در این روش تلاش می‌شود با استفاده از مختصات آخرین مقدار ویژه و i+۱ امین مقدار ویژه، مختصات i امین مقدار ویژه درون‌یابی شود (رایچی و مگیس[3]، ۲۰۱۰). سپس مقادیر ویژه پیش‌بینی شده و مشاهده‌شده مقایسه می‌شوند و آخرین مقدار ویژه مشاهده‌شده‌ای که بزرگ‌تر مساوی مقدار ویژه پیش‌بینی شده باشد، تعیین‌کننده تعداد مؤلفه‌ها خواهد بود. همچنین، ارزش عددی این مقدار ویژه باید یا بر اساس ملاک کایزر بزرگ‌تر مساوی

1. Optimal Coordinate and Acceleration Factor
2. Raîche et al
3. Raiche & Magis

یک باشد یا باید از ملاک آماره جایگاه[1] در تحلیل موازی که معمولاً مقدار ویژه متوسط یا یکی از صدک‌های ۵، ۵۰ یا ۹۵ مقادیر ویژه مربوط به مجموعه داده‌های تصادفی (مثلاً ۱۰۰ مجموعه داده تصادفی) است، بزرگ‌تر مساوی باشد؛ ب) عامل شتاب بر مختصات نقطه‌ای در نمودار اسکری تأکید دارد که در آن شیب منحنی به‌طور ناگهانی تغییر می‌کند. درواقع، هدف این روش پیدا کردن زانوی[2] منحنی اسکری است، زیرا بیشترین تغییر در شیب منحنی اسکری در نقطه زانو شکل آن رخ می‌دهد. جهت تقعر منحنی اسکری در این بخش از منحنی تغییر می‌کند. پیدا کردن این نقطه از طریق پیدا کردن مشــتق دوم یک معادله به دست می‌آید، که جزئیات آن در رایچی و همکاران (۲۰۱۳) ارائه شده است. قاعده تصمیم‌گیری در خصوص تعداد عامل‌ها بر اساس این روش به این صورت است که مقادیر ویژه مشاهده‌شده قبل از این نقطه به‌عنوان تعداد مؤلفه‌ها در نظر گرفته می‌شوند. افزون بر این، مقادیر ویژه قبل از این نقطه باید دارای یکی از این دو ملاک (یا هردو) نیز باشند. یعنی مقادیر ویژه قبل از این نقطه یا باید دارای مقدار ویژه بزرگ‌تر مساوی یک باشند یا از آماره‌های جایگاه در تحلیل موازی که معمولاً مقدار ویژه متوسط یا یکی از صدک‌های ۵، ۵۰ یا ۹۵ مقادیر ویژه مربوط به مجموعه داده‌های تصادفی (مثلاً ۱۰۰ مجموعه داده تصادفی) است، بزرگ‌تر مساوی باشند. نتایج این دو روش، در مقایسه با قاعده کایزر-گاتمن، همانند تحلیل موازی به راه حل‌های مبتنی بر امساک منجر می‌شود. هر دو روش در داده‌های مربوط به مقوله‌های پاسخ رتبه‌ای ۲ تا ۷ طبقه و نیز مقوله‌های پاسخ ۱۰ تا ۲۰ طبقه، که شبیه پیوسته محسوب می‌شوند، عملکرد خوبی دارند. در مطالعات شبیه‌سازی عملکرد هر دو روش بهتر از ملاک کایزر-گاتمن است. روش مختصات بهینه در ۷۴/۰۳ درصد از موارد در رقابت با تحلیل موازی، که در ۷۶/۴۲ درصد از موارد تعداد عامل‌ها را درست تشخیص می‌دهد، تعداد عامل‌ها را درست مشخص می‌کند. این میزان برای روش عامل شتاب ۴۵/۹۱ درصد است. تمایل این روش به کم‌برآورد کردن عامل‌ها است (کورتنی و گوردون، ۲۰۱۳).

۱۰) تحلیل خوشه: در این رویکرد بر اساس اصل خوشه‌بندی متغیرها، تعداد خوشه‌های سؤال که حد بالایی برای تعداد ابعاد است، تعیین می‌شود (رکاسی[3]، ۲۰۰۹). تحلیل خوشه به شیوه سلسله‌مراتبی خوشه‌هایی که نسبت به هم دارای ساختار آشیانه‌ای هستند را پیدا می‌کند. نتایج نشان‌دهنده آن است که در بین روش‌های تحلیل خوشـــه در حوزه مدل‌های چندبعدی روش خوشه‌بندی سلســـله‌مراتبی وارد[4] (۱۹۶۳) نسبت به سایر روش‌های رقیب، ساختار موجود در آزمون را بهتر بازنمایی می‌کند، به‌شرط آنکه درون‌داد تحلیل، زاویه بین بردار متغیرها باشد، نه ماتریس همبستگی یا ماتریس فاصله اقلیدسی. روش وارد در هر گام دوخوشه‌ای که مجموع مجذور درون‌خوشـــه‌ای برای خوشـــه حاصل از ترکیب آنها حداقل مقدار ممکن است را با هم ترکیب می‌کند. همبستگی بین هر زوج سؤال برابر با کسینوس زاویه بین سؤال‌ها در فضایی با ابعادی برابر با سؤال‌های آزمون

1. location statistics criteria
2. elbow
3. Reckase
4. Ward

است. پس در مقایسه با هر فضایی با ابعاد کمتر (مثلاً دو یا سه بعد) بهترین شرایط را برای بررسی زاویه بین بردار سؤال‌ها فراهم می‌کند. دومین روش تحلیل خوشه برای دسته‌بندی سؤال‌ها آن است که زاویه بین بردار سؤال‌ها در فضای دو و سه‌بعدی نیز برای بررسی تعداد ابعاد استفاده می‌شود. برای این کار می‌توان از بارهای حاصل از چرخش وریماکس سؤال‌ها استفاده کرد. ضرایب تشخیص حاصل از چرخش وریماکس تمایز بین خوشه‌های سؤال را به‌وضوح نشان می‌دهند؛ بنابراین، استفاده از آنها برای تحلیل خوشه بهتر از ضرایب تشخیص مایل یا چرخش نیافته است (کیم[1]، ۲۰۰۱). می‌توان ابتدا بارهای عاملی سؤال‌ها در فضای دو و سه‌بعدی را به شیب، تبدیل کرد و سپس تحلیل خوشه را بر اساس هر یک از آنها به‌صورت جداگانه انجام داد. سومین روش برای اجرای تحلیل خوشه، روش رول (۱۹۷۹) است. این روش که به‌اختصار iclust نام دارد، مناسب‌ترین روش برای اجرای تحلیل خوشه در روان‌سنجی و ساخت مقیاس است (رول، ۲۰۱۷). تابع مربوط به این روش به همین نام در بسته psych که جزو بسته‌های معروف R در حوزه روان‌سنجی است، وجود دارد و از این طریق قابل اجرا است. این روش با استفاده از الگوریتم خوشه‌بندی سلسله‌مراتبی تا جایی خوشه‌های سؤال را با هم ترکیب می‌کند که یکی از دو ملاک همسانی استفاده شده در این روش افزایش نیابد. نخستین ملاک، همان آلفای کرانباخ (α، متوسط پایایی به شیوه دو نیمه‌کردن) است و دومین ملاک، بدترین پایایی به شیوه دو نیمه کردن[2] (β) است. پایایی اخیر برآوردی از اشباع عامل کلی مقیاس حاصل است (رول، ۲۰۱۸).

۱۱) روش نموداری: در این روش، که پیش‌تر به‌عنوان یکی از روش‌های تحلیل خوشه توصیف شد، بردار سؤال‌ها در فضای دو و سه‌بعدی ترسیم و بررسی می‌شود تا بر اساس جهت بردار سؤال‌ها و پراکندگی آنها در مورد تعداد ابعاد لازم تصمیم‌گیری شود. استفاده از این رویکرد برای مشاهده عینی وضعیت موجود در داده‌ها است، به‌طوری‌که با این کار استنباط‌های مبتنی بر سایر روش‌ها به‌طور عینی بررسی و مشاهده می‌شوند (رکاسی، ۲۰۰۹).

۱۲) روش کیفی: در کنار روش‌های کمی مختلف نباید تحلیل محتوای کیفی آزمون توسط متخصصان محتوایی و طراحان را نادیده گرفت (چیو[3]، ۲۰۱۳). تحلیل کیفی محتوای آزمون توسط متخصصان آشنا به محتوای آن یا خود طراحان سؤال برای تعیین تعداد عامل‌ها مفید بوده و استفاده از آن مورد تأکید قرار گرفته است. می‌توان از این رویکرد در کنار سایر روش‌های کمی استفاده کرد. امروز در مدل‌های سنجش شناختی از این رویکرد به‌عنوان مرحله مقدماتی برای ساخت ماتریس کیو و اجرای تحلیل‌های کمی استفاده می‌شود. همچنین، می‌توان برازش مدل نظری حاصل از تحلیل محتوای کارشناسان را با تحلیل عاملی تأییدی بررسی کرد.

۱۳) رویکرد برازش مدل: این رویکرد از شاخص‌های مختلف برازش مدل به داده‌ها برای مقایسه مدل‌ها و انتخاب

1. Kim
2. the worst split half reliability
3. Chiu

یکی از آنها استفاده می‌کند. مسئله اصلی در تحلیل عاملی اکتشافی، تصمیم‌گیری در خصوص تعداد عامل‌هایی است که باید حفظ شوند. اگرچه این مسئله ذاتاً یک مسئله انتخاب مدل است، ولی رویکرد انتخاب مدل به‌ندرت برای این کار استفاده شده است. در انتخاب مدل بر اساس این رویکرد پژوهشگران باید ابتدا هدف تحلیل خود را شناسایی کنند: آیا هدف شناسایی تقریباً درست m عامل است یا شناسایی تکرارپذیرترین m عامل. در مرحله دوم پژوهشگران باید شاخص‌های برازش موافق با هدف خود را انتخاب کنند. نتایج مطالعات شبیه‌سازی در راستای مبانی نظری نشان می‌دهد که شاخص‌های برازش مختلف برای هدف‌های مختلف مناسب هستند. انتخاب مدل بر اساس یک هدف در ذهن (شناسایی تقریباً درست m عامل) لزوماً بر اساس هدف ذهنی دیگری (شناسایی تکرارپذیرترین m عامل) به همان تعداد عامل منجر نمی‌شود. بنابراین، بهتر است پژوهشگران قبل از انتخاب شاخص‌های برازش مناسب، منظور خود از «تعداد درست عامل‌ها» را دقیق مشخص کنند. این ایده رایج است که مدل دارای بهترین تعمیم‌پذیری به نمونه‌های آتی نزدیک‌ترین مدل به واقعیت نیز هست. شواهد زیادی حاکی از آن است که دست‌کم با توجه به اندازه نمونه‌هایی که احتمالاً در پژوهش‌های روان‌شناسی با آنها مواجه خواهیم بود، این ایده لزوماً درست نیست. برای بررسی این ادعا ابتدا لازم است چارچوب خاصی را مد نظر قرار دهیم. اختلاف نمونه[1] (SD) به تفاوت بین داده‌های مشاهده‌شده (S) و پیش‌بینی‌های مدل ($\hat{\Sigma}$) اشاره دارد. اختلاف کلی[2] (OD) به تفاوت بین ماتریس کواریانس جامعه (Σ_0) و پیش‌بینی‌های مدل ($\hat{\Sigma}$) اشاره دارد. اختلاف ناشی از تخمین[3] (DA)، یا خطای مدل، به اختلاف بین ماتریس کواریانس جامعه (Σ_0) و پیش‌بینی‌های مدل در جامعه ($\hat{\Sigma}_0$) اشاره دارد. در یک مدل مشخص DA یک کمیت ثابت ولی غیرقابل مشاهده است، زیرا پژوهشگران مدل‌ها را به جای جوامع به نمونه‌ها برازش می‌دهد. سرانجام، اختلاف ناشی از برآورد[4] (DE) نشان‌دهنده تغییرپذیری نمونه‌گیری است. بر این اساس در کل می‌توان تغییرپذیری کلی را به‌صورت $OD = DA + DE + o(N^{-1})$ تعریف کرد، که در آن با افزایش N، $o(N^{-1})$ کوچک و قابل اغماض خواهد بود و DE نیز کاهش می‌یابد. در حالت حدی $SD = \Sigma_0$ و $\hat{\Sigma} = \hat{\Sigma}_0$ و بنابراین $OD = DA$. ازآنجاکه نمونه‌ها همیشه محدود و مدل واقعی هرگز در خزانه مدل‌های رقیب تحت بررسی وجود ندارد، پس $OD > DA > 0$. در چارچوب این رویکرد ملاک‌های خاصی برای ارزیابی اختلاف ناشی از تخمین (DA) و اختلاف کلی (OD) ارائه شده است. شاخص RMSEA[5] برای ارزیابی اختلاف ناشی از تخمین (DA) و شاخص‌های AIC و BIC برای ارزیابی اختلاف کلی (OD) معرفی شده‌اند. زمانی که هدف بیشینه‌سازی راستی‌نمایی (واقعیت) است، ملاک RMSEA نسبت به سایر ملاک‌ها عملکرد بهتری دارد. در این حالت، مدلی با کوچک‌ترین تعداد عامل (m)

1. Sample Discrepancy
2. Overall Discrepancy
3. Discrepancy Due To Approximation
4. Discrepancy Due To Estimation
5. root mean square error of approximation

که RMSEA آن کمتر از ۰/۰۵ باشد، مناسب‌ترین تعداد عامل‌ها است. کرانه پایین فاصله اطمینان ۹۰ درصد اطراف RMSEA را نیز می‌توان برای انتخاب تعداد عامل‌هایی که باید حفظ شـــوند استفاده کرد. بر اساس این ملاک کوچک‌ترین تعداد عامل (m)، که برای آن کرانه پایین فاصله اطمینان ۹۰ درصد اطراف RMSEA زیر ۰/۰۵ باشد، را می‌توان برای حفظ تعداد عامل‌ها در نظر گرفت. بر اساس پیشینه این روش کوچک‌ترین تعداد عاملی که آزمون برازش نزدیک برای آن رد نشود را انتخاب می‌کنند. درصورتی‌که هدف پیدا کردن یک مدل از بین مدل‌های رقیب است، که فرایند واقعی تولید داده‌ها را به‌طور عینی به بهترین شکل تخمین بزند (یعنی مدل کمینه‌کننده DA)، حتی اگر آن مدل احتمالاً به نمونه‌های آتی قابل تعمیم نباشد، می‌توان از کرانه پایین فاصله اطمینان ۹۰ درصد اطراف RMSEA استفاده کرد، ولی اگر هدف تعمیم‌پذیری است برای انتخاب و حفظ تعداد عامل‌ها AIC توصیه می‌شود. شاخص AIC معمولاً در کنار BIC برای انتخاب مدل استفاده می‌شود. برای BIC، $n \geq 8$ نسبت به AIC جریمه بیشتری بر پیچیدگی مدل اعمال می‌کند و نسبت به AIC معمولاً مدل‌های دارای پارامترهای کمتر را ترجیح می‌دهد. شاخصAIC در نمونه‌های کوچک در انتخاب تعداد واقعی عامل‌ها، در صورت وجود، خوب عمل می‌کند. شـــاخص BIC نسبت به AIC در بازیابی m عامل واقعی عملکرد بهتری دارد. شاخص AIC همراه با افزایش نمونه مدل‌های پیچیده‌تر را انتخاب می‌کند، زیرا نرخ افزایش در عبارت بدی برازش همراه با n افزایش می‌یابد ولی عبارت جریمه به همان صورت باقی می‌ماند. این موضوع منشأ سؤال‌هایی در خصوص مناسب بودن AIC برای انتخاب مدل شده است. این گرایش AIC شاید به این خاطر مسئله‌ساز به نظر می‌رسد که AIC با هدف تعمیم‌پذیری (یعنی کمینه‌سازی اختلاف کلی) طراحی شده ولی پژوهشگران از آن برای پیگیری راستی‌نمایی (کمینه‌سازی اختلاف ناشی از برآورد) استفاده می‌کنند. چون $OD \approx DA + DE$ و DE تغییرپذیری نمونه‌گیری را منعکس می‌کند، پس تعجب‌آور نخواهد بود که N بر انتخاب مدل با استفاده از شاخص‌های AIC و BIC مؤثر باشد. در خصوص رتبه‌بندی مدل‌ها به‌منظور توانایی آنها از نظر وارسی روایی در داده‌های جدید با استفاده از AIC و BIC نیز می‌توان گفت AIC همانند سایر شاخص‌های وارسی روایی [مثل شاخص وارسی روایی مورد انتظار[1] (ECVI)]، طبق انتظار همراه با افزایش نمونه، مدل‌های پیچیده‌تر را انتخاب می‌کند (پریچر و همکاران[2]، ۲۰۱۳).

شاخص‌های AIC و BIC بسته به هدفشان معمولاً در پاسخ به سؤال «مدل خوب» برای داده‌ها پاسخ متفاوتی می‌دهند. با فرض اینکه یکی از مدل‌های تحت بررسی مدل واقعی است، هدف BIC شناسایی مدل‌هایی است که بیشـــترین احتمال مدل واقعی در داده‌ها را دارا هستند. در مقابل AIC وجود مدل واقعی قابل شناسایی را به‌طور روشن رد کرده و به جای آن پیش‌بینی مورد انتظار داده‌های آتی را به‌عنوان ملاک کلیدی کفایت مدل استفاده می‌کند. چون هدف BIC شناسایی مدل واقعی (در داده‌های فعلی) و هدف AIC پیش‌بینی داده‌های

1. Expected Cross-Validation Index
2. Preacher et al

آتی است، بی‌دلیل نیست که با توجه به تعداد نمونه به نتایج مختلفی منجر شوند. با توجه به اینکه تأکید BIC بر انتخاب مدل واقعی از بین مدل‌های موجود است، در شناسایی مدل واقعی، مفهوم کلیدی ثبات[1] است، که در اینجا به معنی آن است که احتمال انتخاب مدل واقعی از بین مدل‌های نامزد همراه به افزایش نمونه به یک میل کند، با فرض اینکه مدل واقعی یکی از مدل‌های تحت بررسی بوده و مقدار پارامتر نیز ثابت فرض شود. شاخص BIC معمولاً یک انتخاب‌کننده مدل دارای ثبات است ولی AIC یک انتخاب‌کننده مدل دارای ثبات نیست (کوها[2]، ۲۰۰۴). شاخص BIC در نمونه‌های کوچک خوب عمل می‌کند ولی در نمونه‌های بزرگ تعداد عامل‌ها را بیش‌برآورد می‌سازد (پریچر و همکاران، ۲۰۱۳). با این حال، دیزیک و همکاران[3] (۲۰۱۲) نشان داده‌اند که اگر حجم نمونه بیشتر از ۲۰۰۰ مورد باشد بیش‌برآورد BIC از تعداد عامل‌ها کاهش و برآورد درست تعداد عامل‌ها افزایش می‌یابد. بر اساس مطالعه پیرسون و همکاران[4] (۲۰۱۳) AIC تقریباً در تمام شرایط یا به نتایج درست یا به بیش‌برآورد یک عامل منجر می‌شود، بنابراین می‌تواند به‌عنوان یک جایگزین معقول در صورت دسترسی نداشتن به تحلیل موازی مبتنی بر تحلیل عاملی مشترک یا همراه با آن برای انتخاب تعداد عامل‌ها در نظر گرفته شود. استفاده از AIC برای تعیین بهترین مدل به این شرط پیشنهاد می‌شود که مدل حاوی یک عامل بیشتر نیز بررسی شود. معمولاً مدل‌های دارای مقدار AIC و BIC کوچک‌تر بهتر قلمداد می‌شوند. به همین منظور از تفاوت مقادیر این آماره‌ها در مدل‌های مختلف برای انتخاب مدل نیز استفاده می‌شود. برای مقایسه و انتخاب مدل با استفاده از AIC و BIC می‌توان از راهکار زیر استفاده کرد (فبوزی و همکاران[5]، ۲۰۱۴):

AIC: ابتدا برای m مدل (m=1,…,M)، مقدار آماره AIC محاسبه می‌شود. اگر برای هر یک از m مدل فرضی، آماره را با AIC_m و آماره مربوط به مدل دارای کوچک‌ترین مقدار را با AIC^* نشان دهیم و تفاوت بین هر یک از مقادیر AIC_m و AIC^* را به‌صورت $\Delta AIC_m = AIC_m - AIC^*$ محاسبه کنیم. می‌توان مقادیر $AIC\Delta$ را برای حمایت از هر مدل نامزد به‌صورت زیر تفسیر کرد:

- مقادیر کوچک‌تر از ۲ نشان می‌دهد که شواهد قابل توجهی در حمایت از مدل نامزد شده وجود دارد. یعنی این مدل، به‌خوبی بهترین مدل عمل می‌کند.
- مقادیر بین ۴ تا ۷ نشان می‌دهد که مدل نامزد شده به‌طور قابل توجهی از حمایت کمتری برخوردار است.
- مقادیر بزرگ‌تر از ۱۰ نشان می‌دهد که هیچ نوع حمایت قابل توجهی برای مدل وجود ندارد.

BIC: ابتدا برای m مدل، مقدار آماره BIC محاسبه می‌شود. اگر برای هر یک از m مدل فرضی، آماره را با BIC_m و آماره مربوط به مدل دارای کوچک‌ترین مقدار را با BIC^* نشان دهیم و تفاوت بین هر یک از مقادیر BIC_m و

1. Consistency
2. Kuha
3. Dziak et al
4. Pearson et al
5. Fabozzi et al

BIC^* را به‌صورت $\Delta BIC_m = BIC_m - BIC^*$ محاسبه کنیم. سپس می‌توان مقادیر $BIC\Delta$ را برای حمایت از هر مدل نامزد به‌صورت زیر تفسیر کرد:

- مقادیر کوچک‌تر از ۲ نشان می‌دهد که دو مدل عملکرد مشابهی دارند. به عبارتی این مدل به‌خوبی بهترین مدل عمل می‌کند.
- مقادیر بین ۲ تا ۶ نشان می‌دهد شواهد بر علیه مدل نامزد شده، مثبت است.
- مقادیر بین ۶ تا ۱۰ نشان می‌دهد شواهد بر علیه مدل نامزد شده، قوی است.
- مقادیر بزرگ‌تر از ۱۰ نشان می‌دهد که شواهد بر علیه مدل، بسیار قوی است.

تفسیرهای بالا، قواعد سرانگشتی محسوب شده و می‌توان آنها را برای آماره‌های اصلاح‌شده AIC و BIC نیز استفاده کرد. آماره اصلاح‌شده AIC که به AIC_c نشان داده می‌شود زمانی استفاده می‌شود که حجم نمونه کم (n) یا تعداد پارامترها (k) زیاد باشد. عبارت سوم در رابطه زیر برای خنثی کردن حجم نمونه و تعداد پارامترها به رابطه AIC اصلی اضافه شده پس با حذف آنها معادله AIC به دست می‌آید. حجم نمونه زمانی کوچک در نظر گرفته می‌شود که نسبت n/k کوچک‌تر از ۴۰ باشد. همراه با افزایش نمونه عبارت سوم به صفر نزدیک شده و مقدار AIC_c به AIC نزدیک می‌شود.

$$AIC_c = -2\log L(\hat{\theta}) + 2k + (2k+1) / (n-k-1)$$

تفاوت بین AIC و BIC در آن است که آماره اخیر جریمه بیشتری برای تعداد پارامترها نسبت به AIC در نظر می‌گیرد، که در رابطه آن در زیر مشهود است.

$$BIC = -2\log L(\hat{\theta}) + k\log n$$

در این عبارت‌ها n اندازه نمونه، k تعداد پارامترهای برآورد شده مدل داوطلب، θ بردار پارامترهای مدل و $L(\hat{\theta})$ درست‌نمایی مدل داوطلب به‌شرط داده‌ها در برآورد بیشینه درست‌نمایی θ ارزیابی شده است. در بافت نظریه سؤال‌پاسخ نیز روش‌هایی برای بررسی ابعاد شکل گرفته و گسترش یافته‌اند که در ادامه به مهم‌ترین آنها اشاره شده است.

۱۴) تحلیل عاملی غیرخطی: از ملاک‌های مختلف برازش مدل (مثل کای دو، تقسیم کای دو بر درجه آزادی و RMSE) برای تعیین تعداد عامل‌ها استفاده می‌کند. در تحلیل عاملی غیرخطی رابطه متغیرهای اندازه‌گیری شده و صفات پنهان یا عامل‌ها که توسط متغیرها اندازه‌گیری می‌شوند غیرخطی فرض می‌شود. استفاده از تحلیل عاملی غیرخطی برای بررسی بعدیت سؤال، به‌خصوص در بافت نظریه سؤال‌پاسخ جذاب است؛ چراکه در این نظریه فرض می‌شود فرم ریاضی منحنی مشخصه سؤال غیرخطی است. یعنی بین عملکرد افراد در سؤال و خصیصه مورد نظر رابطه غیرخطی وجود دارد (همبلتون و روینلی[1]، ۱۹۸۶). با توجه به دوارزشی بودن

1. Hambleton & Rovinelli

داده‌های حاصل از اجرای آزمون‌های چندگزینه‌ای و ایجاد عامل‌های ناشی از دشواری سؤال و تناسب تحلیل عاملی غیرخطی برای در نظر گرفتن این عامل، می‌توان از تحلیل عاملی غیرخطی هم در شکل اکتشافی و هم در شکل تأییدی برای این منظور استفاده کرد. این روش از طریق نرم‌افزار NOHARM قابل اجرا است (فراسر و مک‌دونالد[1]، ۲۰۱۲). البته امروزه این رویکرد به کمک بسته‌های مختلف R مثل mirt (چالمرز[2]، ۲۰۱۲) و sirt (روبیتز[3]، ۲۰۲۱) نیز قابل اجرا است.

۱۵) روش DIMTEST: جزو روش‌های ناپارامتریک اســـت که در آن بر اســـاس معنی‌دار شـــدن آماره زیر در خصوص بعدیت ضروری آزمون تصمیم‌گیری می‌شود.

$$T = \frac{T_L - \bar{T}_G}{\sqrt{1 + \frac{1}{N}}}$$

معنی‌دار شـــدن آماره به رد فرض صفر مبنی بر تک‌بعدی بودن ضروری آزمون منجر می‌شود (عناصر موجود در این رابطه در ادامه توصیف شـــده‌اند). اجرای این روش، نخســـتین گام در تعیین تعداد ابعاد بر اساس این روش است (رکاسی، ۲۰۰۹). بر اساس فرض استقلال موضعی ضروری[4] اگر سؤال‌های یک آزمون یک ویژگی یا صفت را اندازه‌گیری کنند آنگاه میانگین کواریانس شرطی بین زوج‌های سؤال به سمت صفر میل می‌کند. برای اجرای این روش، سؤال‌های آزمون باید به دو دسته تقسیم شوند یک دسته سؤال‌هایی که گمان می‌رود بهترین اندازه‌گیری آنها در جهت مشخص شده به‌وسیله نمره توانایی ترکیبی مرجع[5] است. نمره توانایی ترکیبی مرجع، ترکیب عناصر بردار توانایی است که بهترین جهت اندازه‌گیری آن در جهت بیشترین تشخیص است. این دسته از سؤال‌ها را آزمون افراز شده[6] (PT) گویند. سایر سؤال‌های آزمون که به نظر می‌رسد بهترین جهت اندازه‌گیری آنها با ســـؤال‌های آزمون افراز شـــده بیشـــترین تفاوت را دارد و به‌اصطلاح چیزی متفاوت از آنچه ســـؤال‌های افراز شده اندازه‌گیری می‌کنند را می‌سنجند به آزمون سنجش[7] (AT) موسوم‌اند. سؤال‌های AT از نظر جهت اندازه‌گیری بیشترین تفاوت را با سؤال‌های PT دارند. انتخاب آزمون AT به روش‌های مختلفی انجام می‌شـــود که عبارت‌اند از: استفاده از نظر متخصصان، استفاده از تحلیل عاملی و خوشه‌بندی سؤال‌ها بر اساس بهترین جهت اندازه‌گیری آنها در فضای توانایی‌ها. نتایج حاصل، بســـته به روش‌های مختلف، متفاوت خواهد بود. بهتر است اجازه دهیم خود این نرم‌افزار با استفاده از الگوریتم خوشه‌بندی سؤال‌های AT را پیدا کند. برای رســـیدن به یک نتیجه متعادل بهتر اســـت از یک نمونه برای پیدا کردن سؤال‌های AT و از نمونه دیگر برای

1. Fraser & McDonald
2. Chalmers
3. Robitzsch
4. Essential local independence
5. reference composite ability
6. Partitioned Test (PT)
7. Assessment Test (AT)

محاسبه آزمون آماری استفاده شود. در این شرایط، میانگین کواریانس زوجهای سؤال آزمون AT به‌شرط نمره افراد در آزمون PT محاسبه می‌شود. آماره حاصل که T_L نام دارد دارای سوگیری مثبت است (یعنی آزمون‌های چندبعدی را تک‌بعدی نشان می‌دهد)، بنابراین با استفاده از تکرار مجموعه داده‌های شبیه‌سازی‌شده‌ای که با داده‌های مشاهده‌شده همخوان هستند ولی با مدل تک‌بعدی ایجاد شده‌اند، اصلاح می‌شود. آماره‌ای که از این نوع داده‌های شبیه‌سازی به دست می‌آید T_G نام دارد. برای اصلاح سوگیری مثبت، T_G برای N مجموعه داده شبیه‌سازی‌شده، محاسبه و سپس میانگین آن $\bar{T}_G$ محاسبه می‌شود.

۱۶) روش DETECT: این روش نیز جزو روش‌های ناپارامتریک اســـت که در آن بر اســـاس نتایج سه شاخص که در ادامه توصیف شـــده‌اند در مورد تعداد ابعاد و میزان پیچیدگی تصمیم گرفته می‌شود (رکاسی، ۲۰۰۹). این روش مبتنی بر چند پیش‌فرض اســـت که به‌جز یک اســـتثنا همان پیش‌فرض‌هـــای مربوط به مدل‌های پاسخ‌سؤال چندبعدی هستند. این پیش‌فرض‌ها عبارت‌اند از: افزایش یکنواخت تابع پاسخ‌سؤال همراه با افزایش هر یک از توانایی‌ها. برقراری اســـتقلال موضعی به‌شرط توانایی، با این تفاوت که این استقلال از نوع استقلال موضعی زوجی[1] است. این همان فرض مستثنا شده است که قبلاً ذکر شد. استقلال شرطی زوجی به این معنی اســـت که کُواریانس بین هر زوج از ســـؤال‌های موجود در آزمون به‌شرط هر مقدار از توانایی برابر با صفر است (به‌عبارتی برای هر یک از گروهها یا افراد دارای توانایی یکســـان، کُواریانس بین هر زوج از سؤال‌ها برابر با صفر است (لازم به ذکر است که نرم‌افزار NOHARM نیز بر اساس پیش‌فرض استقلال موضعی ضعیف پایه‌گذاری شـــده ولی TESTFACT بر اساس پیش‌فرض اســـتقلال موضعی قوی[2] پایه‌گذاری شده است). تعریف بعد در روش DETECT عبارت اســـت از تعداد ابعاد لازم برای برقرار شدن استقلال موضعی زوجی با استفاده از تابع یکنواخت افزایشـــی مدل‌های پاسخ‌سؤال چندبعدی. تعداد ابعادی که به تحقق یافتن این ویژگی منجر شوند ابعاد لازم برای بازنمایی ماتریس نمره‌های خام حاصل از تعامل افراد با ســـؤال‌ها هستند. روش DETECT به دنبال تعیین خوشـــه‌های سؤال در فضای چندبعدی است. جهت اندازه‌گیری این خوشه‌های همگون با جهت اندازه‌گیری شـــده به وسیله کل آزمون متفاوت است. سؤال‌های موجود در یک آزمون به خوشه‌های مختلفی تقســـیم می‌شوند. تقسیم‌بندی که ارزش DETECT را بیشینه کند، ملاک دسته‌بندی سؤال‌ها به خوشه‌های مختلف خواهد بود. مقادیر نزدیک به صفر ارزش DETECT حاکی از تک‌بعدی بودن و مقادیر نزدیک یک یا بزرگ‌تر از آن به چندبعدی بودن اشاره دارند. دامنه مقادیر مشاهده‌شده این شاخص معمولاً بین صفر تا مقادیر مثبت کوچک‌تر از پنج اســـت. مقادیر یک یا بزرگ‌تر از آن به چندبعدی بودن زیاد اشاره دارند، مقادیر ۰/۴ تا یک نشان‌دهنده چندبعدی بودن متوسط تا زیاد هستند. مقادیر زیر ۰/۴ نشان‌دهنده چندبعدی بودن متوسط تا ضعیف و مقادیر زیر ۰/۲ نشان‌دهنده تک‌بعدی بودن هستند. همراه با این شاخص، DETECT دو شاخص دیگر

1. pairwise local independence
2. strong local independence

نیز برای تعیین تعداد ابعاد گزارش می‌کند که عبارت‌اند از: نسبت Dref به Dmax و شاخص IDN. نسبت Dref به Dmax: بر اســـاس وارسی روایی[1] شاخص DETEC محاسبه می‌شود. به این صورت که داده‌ها به دو نمونه تصادفی تقسیم می‌شوند. سپس تقسیم‌بندی که شاخص DETEC را برای نمونه اول بیشینه می‌کند مشخص می‌شـــود، شاخص DETEC برای این نمونه با Dmax مشخص می‌شود. تقسیم‌بندی که شاخص DETEC را برای نمونه دوم بیشینه می‌کند نیز مشخص می‌شود. سپس شاخص DETEC با استفاده از داده‌های نمونه اول بر اساس تقسیم‌بندی به دست آمده برای نمونه دوم محاسبه می‌شود که به آن Dref می‌گویند. نسبت Dref به Dmax شـــاخصی برای تعیین تک‌بعدی بودن اســـت. مقادیر یک برای این نسبت نشان‌دهنده مشابه بودن تقســـیم‌بندی‌های به دســـت آمده در هر دو نمونه و ثبات راه حل‌های به دست آمده است. مقادیر کوچک‌تر از یک نشان‌دهنده متفاوت بودن تقسیم‌بندی‌ها است. مقادیر بسیار کوچک این نسبت، نشان‌دهنده این است که تقسیم‌بندی‌ها ناشی از شانس است و احتمالاً به تک‌بعدی بودن ضروری اشاره دارد. شاخص IDN: این شاخص بر اســـاس تقسیم‌بندی سؤال‌های آزمون به دسته‌های مختلف، نسبتی از کواریانس برآورد شده سؤال‌ها که از خوشه‌بندی سؤال‌ها پیروی می‌کنند را مشـــخص می‌سازد. آنچه انتظار می‌رود این است که سؤال‌های داخل یک خوشه با یکدیگر دارای کُواریانس مثبت و سؤال‌های خوشه‌های مختلف با یکدیگر کُواریانس منفی داشته باشند. مقادیر یک برای این شاخص، نشان‌دهنده آن است که الگوی کُواریانس موجود بین خوشه‌های مختلف از نظر علامت با آنچه مورد انتظار است هم‌خوانی دارد، یعنی از خوشه‌بندی حاصل تبعیت می‌کند؛ به‌شرط آنکه خوشه‌بندی حاصل (چندبعدی بودن) از ساختار ساده یا تقریباً ساده پیروی کند.

۱۷) اســـتقلال موضعی: در نظریه صفت پنهان، بعدیت آزمون با اصل اســـتقلال موضعی (شرطی) گره خورده اســـت. این اصل در ســـه شـــکل قوی (لرد و نویک[2]، ۱۹۶۸)، ضعیف (مک‌دونالد، ۱۹۸۱) و ضروری (استوت[3]، ۱۹۸۷ و ۱۹۹۰) آن تعیین‌کننده تعاریف مختلف از بعدیت است. به همین دلیل در طی زمان، بسته به تعریف استقلال موضعی روش‌های زیادی برای بررسی تعداد ابعاد موجود در آزمون‌ها به وجود آمده (هتی[4]، ۱۹۸۵)، به‌تدریج به تعداد این روش‌ها اضافه شـــده (تات[5]، ۲۰۰۳) و روزبه‌روز نیز در حال افزایش هستند (استالفسن و هانیک[6]، ۲۰۰۸؛ لوی و اسویتنا[7]، ۲۰۱۱؛ اسویتنا و لوی[8]، ۲۰۱۲). خلاصه‌ای از روش‌های عمده تعیین بعدیت، شرایط استفاده، همراه با ویژگی‌های عمده و نرم‌افزارهای مربوطه را اسویتنا و لوی (۲۰۱۴) ارائه کرده‌اند. ارتباط پیش‌فرض استقلال موضعی با بحث بعدیت و نقش آن به هنگام استفاده از مدل‌های نظریه سؤال‌پاسخ بر تمام

1. cross-validation
2. Lord & Novick
3. Stout
4. Hattie
5. Tate
6. Stellefson & Hanik
7. Levy & Svetina
8. Svetina & Levy

برآوردهای حاصل از یک مدل خاص تأثیر منفی دارد. دو یا چند سؤال با وابستگی موضعی مثبت صفات خاصی را اندازه‌گیری می‌کنند که در ســـایر ســـؤال‌های آزمون وجود ندارد و دو مجموعه سؤال دارای وابستگی منفی صفات متفاوتی را اندازه‌گیری می‌کنند (ین[۱]، ۱۹۸۴). شاخص‌ها و آماره‌های مختلفی برای کمک به تشخیص وابســـتگی موضعی ارائه شده‌اند. ازاین‌بین، دو شاخص کای‌دو (χ^2) و جی‌دو (G^2) بیشتر مورد توجه قرار دارند. بررســـی عملکرد این دو شاخص در شـــرایط فرض صفر و تحت شرایطی که وابستگی موضعی وجود دارد، در داده‌های شبیه‌سازی‌شده نشان می‌دهد که تحت شرایط فرض صفر استقلال موضعی هم کای‌دو و هم جی‌دو دارای توزیعی مشابه با کای دو با یک درجه آزادی هستند. تحت شرایط وجود وابستگی موضعی ظاهراً هر دو شـــاخص به تشخیص وابستگی موضعی یا چندبعدی بودن در بین سؤال‌ها حساس هستند. در مقایسه با Q_3، که شاخص دیگری برای تشخیص وابستگی موضعی است، کای‌دو و جی‌دو در حالت وابستگی موضعی اساسی توان کمتری دارند، اگر وابستگی موضعی سطحی باشد توانی یکسان با Q_3 دارند و رفتارشان در شرایط فرض صفر بهتر از Q_3 است (چن و تیسن[۲]، ۱۹۹۷). افزون بر سه شاخص فوق، تبدیل r به z فیشر بر روی Q_3[۳]، آماره نسبت درست‌نمایی[۴] (G^2)، شاخص شیب جک‌نایف[۵] (JSI) و شاخص‌های مبتنی بر ساختار کواریانس (شامل $\chi^2_{G/D}$ مبتنی بر NOHARM و شاخص‌های اصلاح مبتنی بر تحلیل عاملی تأییدی[۶]) نیز برای بررسی استقلال موضعی به کار می‌روند. نتایج یک مطالعات شبیه‌ســـازی به شیوه مونت‌کارلو[۷] در شرایط مختلف نشان داد که شاخص شیب جک‌نایف و آماره نسبت درست‌نمایی G^2 در مدل دوپارامتری و مدل پاسخ مدرج سمجیما که این شاخص‌ها در آنها در دسترس هستند، نسبت به بقیه در اکثر شرایط عملکرد کافی تا خوبی دارند و استفاده از هر دوی آنها پیشنهاد شده است (هاتس و ادواردز[۸]، ۲۰۱۳).

بحث و نتیجه‌گیری

استخراج عامل‌های تفسیرپذیر، که مهم‌ترین هدف در بررسی تعداد عامل‌ها و ابعاد موجود در داده‌ها است، به این معنی اســـت که تعداد عامل‌ها بیشتر منعکس‌کننده خلاقیت پژوهشگر است تا داده‌ها. روش‌های مختلف استخراج عامل فقط جنبه سرنخ داشته و هیچ‌وقت نباید صرفاً به آنها تکیه کرد. وجود روش‌های متعدد برای بررســـی و مشخص کردن ابعاد داده‌ها نشان‌دهنده پیچیدگی این موضوع است. مشخص کردن ابعاد داده‌های حاصل از اجرای آزمون‌ها (به‌خصوص داده‌های دوارزشی) هیچ روش منحصربه‌فردی ندارد. همگام با گسترش فناوری و توسعه نرم‌افزارها روش‌های مختلفی برای این منظور پیشنهاد شده است (هتی، ۱۹۸۵؛ تات، ۲۰۰۳؛

1. Yen
2. Chen & Thissen
3. Fisher's r-to-z transformed Q3
4. Likelihood Ratio Statistic
5. Jackknife Slope Index
6. Confirmatory factor analysis (CFA) modification indices (MIs)
7. Monte Carlo
8. Houts & Edwards

استالفسن و هانوک، ۲۰۰۸؛ لوی و اسویتنا، ۲۰۱۱). روش‌های عمده‌ای که تا به امروز برای این منظور استفاده شده‌اند توسط پژوهشگران مختلف (ازجمله اسویتنا و لوی، ۲۰۱۴) خلاصه شده‌اند.

یکی از مشکلات عمده در تعیین تعداد عامل‌ها، چیزی است که به‌عنوان عامل در نظر گرفته می‌شود. بسیاری از روش‌های تحلیل عاملی فرض می‌کنند که ماتریس همبستگی مانده‌ها پس از استخراج عامل‌های مورد نظر متشـــکل از خطای تصادفی صرف است. در مقابل، در برخی از روش‌ها معتقدند گرچه ماتریس همبستگی از عامل‌های اصلی مورد نظر تشکیل شده ولی عامل‌های کوچک متعددی نیز وجود دارند که اساساً مورد توجه نیستند اما بخشی از کواریانس مشترک بین متغیرها را تبیین می‌کنند. حضور چنین عامل‌های کوچکی می‌تواند فرد را به اســـتخراج تعداد زیادی عامل سوق دهد و صرفاً بر اساس زمینه‌های آماری عدم برازش، راه حل‌هایی کـــه درواقع برازش خوبی با داده‌ها دارند را رد کنـــد (رول، ۲۰۱۷). با توجه به روش‌های مختلف تعیین ابعاد مشخص است که این مسئله هنوز بدون هیچ راه حل قطعی وجود دارد و هیچ‌یک از روش‌های فوق به‌طور قطع تعیین‌کننده درست تعداد عامل‌ها نیست. این گفته هنری کایزر[1] (به نقل از هورن و انگسترام، ۱۹۷۹) دلیلی بر این موضوع است. «راه حل مشکل تعداد عامل‌ها در تحلیل عاملی ساده است. به‌طوری‌که وی هر روز صبح پیش از صبحانه یک راه حل پیدا می‌کرد. اما مشکل پیدا کردن راه حلی است که دیگران، اگرچه آن را به‌عنوان بهترین راه حل نشمارند، ولی دست‌کم با آن موافق باشند».

هر یک از روش‌های ذکر شده تعیین بعد مزایا و معایب خاص خود را دارند (رایت‌ولت و ون‌هوت، ۱۹۹۳؛ رول، ۲۰۱۷). آزمون کای دو یا تغییر در آزمون کای دو به تعداد آزمودنی‌ها حســـاس اســـت و به شرایط غیرعملی منجر می‌شـــود؛ به این معنی که اگر فردی به دنبال عامل‌های بیشـــتری باشد می‌تواند به‌راحتی تحلیل را با آزمودنی‌های بیشتری انجام دهد. تحلیل موازی نیز تا حدودی به اندازه نمونه حساس است، از این نظر که در نمونه‌های بزرگ ($N>1000$) تمام مقادیر ویژه عامل‌های تصادفی بســـیار به یک نزدیک می‌شوند. طبق اظهار نظر پیرسون و همکاران (۲۰۱۳) برای بررسی تعداد عامل‌ها این روش باید اولین گام باشد. به هنگام استفاده از این روش اگر این شـــک وجود دارد که یک عامل با تعداد کمی متغیر بازنمایی شـــده (پنج مورد یا کمتر)، پژوهشگر باید مدل حاوی یک عامل بیشتر را نیز بررسی کند. به‌علاوه، اگر پژوهشگر به سایر روش‌های معتبر تعیین ابعاد مثل تحلیل موازی و AIC دسترسی ندارد، تعداد متغیرهای مشاهده‌شده زیاد نیست و اشتراک‌های برآورد شده همگی بزرگ هستند، پژوهشگر باید تعداد مقادیر ویژه ماتریس همبستگی کاهش‌یافته بزرگ‌تر از متوسط را استفاده کند. این ملاک تصمیم‌گیری در یک مطالعه از قاعده مقادیر ویژه بزرگ‌تر از یک دقیق‌تر بود. آزمون اســـکری کاملاً جذاب است ولی می‌تواند به اختلاف‌های تفسیری در خصوص موقعیت شکست نمودار اسکری منجر شود. استخراج عامل‌های تفسیرپذیر به این معنی است که تعداد عامل‌ها بیشتر منعکس‌کننده

1. Henry Kaiser

خلاقیت پژوهشگران است تا دادهها. با اینکه درک روش ساختار بسیار ساده (VSS) خیلی راحت است، ولی اگر دادهها به لحاظ عاملی خیلی پیچیده باشند، روش ساختار بسیار ساده خیلی خوب کار نمی‌کند (شبیه‌سازی‌ها نشـــان‌دهنده این است که این روش زمانی خوب کار می‌کند که پیچیدگی برخی سؤال‌ها بیشتر از ۲ نباشد). قاعده مقدار ویژه بیشتر از یک (ملاک گاتمن-کایزر)، اگرچه پیش‌فرض برنامه‌های زیادی است، ولی به نظر یک شیوه خشن از تقسیم تعداد متغیرها بر ۳ است و احتمالاً بدترین همه ملاک‌ها است. روش مقادیر ویژه بیشتر از یک، که از قدیمی‌ترین و پرکاربردترین روش‌های تعیین ابعاد است، همراه با نمودار اسکری[1] پیش‌فرض بسیاری از نرم‌افزارها اســـت (چکلوک و کُچاک[2]، ۲۰۱۶). در روش مقادیر ویژه بیشتر از یک، که بر جنبه ریاضی بحث تعداد ابعاد استوار است، مقادیر ویژه کمتر از یک ماتریس همبستگی به عوامل جزئی بی‌اهمیت و خطا نسبت داده می‌شود (فابریگر و همکاران، ۱۹۹۹) و در اصل برای تحلیل مؤلفه (یک ماتریس کاهش‌نیافته) ارائه شده نه تحلیل عاملی مشترک (یک ماتریس کاهش‌یافته). در تحلیل مؤلفه، مقادیر ویژه از ماتریسی استخراج می‌شوند که روی قطر اصلی آن مقادیر یک قرار گرفته‌اند. ازاین‌رو، روش مقادیر ویژه بیشـــتر از یک برای تحلیل عاملی مشترک که در آن از اشتراک متغیرها به جای یک روی قطر اصلی استفاده می‌شود، مناسب نیست. همانند هر روش مکانیکی، روش مقادیر ویژه بیشتر از یک نیز تا حد زیادی دلبخواهی است. مثلاً این ادعا که مقدار ویژه ۱/۱ را یک عامل مهم و مقدار ویژه ۰/۹۹ را به‌عنوان عامل بی‌اهمیت در نظر بگیریم واقعاً بی‌معنی است. در مطالعاتی که هم شامل تحلیل مؤلفه و هم تحلیل عاملی مشترک بوده، نتایج حاکی از آن است که روش مقادیر ویژه بیشتر از یک معمولاً به بیش‌برآورد شدن تعداد عامل‌ها منجر شده و به‌ندرت به کم‌برآورد شدن تعداد عامل‌ها منتهی می‌شود. تقریباً هیچ مطالعه‌ای نشان نداده که روش مقادیر ویژه بیشتر از یک، خوب عمل می‌کند. همچنین، نتایج مطالعات نشـــان می‌دهد که این روش تنها در ۸/۷۷ درصد از موارد تعداد عامل‌ها را به‌درستی مشخص می‌کند (کورتنی و گوردون، ۲۰۱۳). اســـتفاده از این ملاک در شرایطی است که تعداد متغیرهای اصلی از ۳۰ بیشتر نبوده و اشتراک متغیرها از ۰/۷ بیشتر باشد. با این حال، باید تأکید شود که مؤلفه‌ها یا عامل‌های مربوط به مقادیر ویژه کوچک‌تر از یک نیز ممکن است به‌شدت تفسیرپذیر بوده و نباید آنها را همیشه نادیده گرفت. قاعده حفظ عامل‌هایی که مقدار ویژه آنها از یک بیشتر است تنها در بافت تحلیل مؤلفه منطقی به نظر می‌رسد. همان‌طور که ذکر شد در این تحلیل، واریانس هر یک از متغیرها به اجزای مختلف تقسیم نشده و همه متغیرها در فرم استاندارد خود دارای اشتراک یک هستند. بر این اساس، می‌توان استدلال کرد که نگه‌داشتن مؤلفه‌هایی که کمتر از واریانس یک متغیر استاندارد شده (یعنی یک) را تبیین می‌کنند، ارزشمند نیست. استفاده از این قاعده در تحلیل عاملی منطقی به نظر نمی‌رسد؛ چراکه در این تحلیل از واریانس مشترک یک متغیر با سایر متغیرها به جای یک روی قطر اصلی ماتریس همبستگی استفاده می‌شود و بیشتر مواقع این اشتراک پایین‌تر

1. Scree test
2. Çokluk & Koçak

از یک است. به‌علاوه، باید توجه کرد که مقدار ارزش‌های ویژه مربوط به مؤلفه‌ها یا عامل‌ها تابع تعداد متغیرهای موجود در تحلیل است. بر این اساس، در مجموعه کوچکی از متغیرها ممکن است فرد عامل‌های بسیار کم و در مجموعه زیادی از متغیرها ممکن است عامل‌های بیشتری را استخراج کند. همچنین، با استفاده از این روش تعداد عامل‌های حفظ شده یک‌سوم تا یک‌پنجم تعداد متغیرها خواهند بود. بنابراین، گرچه این روش رایج است و پیش‌فرض بســـیاری از بسته‌های آماری است ولی ترجیحاً نباید در تحلیل عاملی مورد استفاده قرار گیرد یا به‌عنوان تنها روش برای تعیین تعداد عامل‌ها استفاده شود. طبق اظهار نظر پیرسون و همکاران (۲۰۱۳) قاعده مقدار ویژه به‌هیچ‌وجه خوب عمل نمی‌کند، ولی نسخه مبتنی بر تحلیل عامل مشترک آن بهتر از نسخه مبتنی بر تحلیل مؤلفه آن است.

توســـعه مدل‌های چندبعدی نیز اهمیت بررسی تعداد ابعاد را کاهش نداده و کاربرد این مدل‌ها نیز همیشه با مســـئله برازش و تحلیل ابعاد همراه اســـت. تعریف بعدیت داده‌ها بر اساس حوزه مدل‌سازی که فرد در آن کار می‌کند (مثلاً، تحلیل عاملی یا نظریه سؤال‌پاســـخ) متفاوت است. با این حال، به‌طور سنتی در تعریف بعدیت داده‌ها تلاش می‌کنیم به سؤال چه تعداد متغیر پنهان، تبیین‌کننده کواریانس بین سؤال‌های آزمون است پاسخ دهیم. علاقه ما درک و تبیین آماری واریانس و کواریانس بین سؤال‌های موجود در یک آزمون است. در چارچوب تحلیل عاملی، سؤال این است که چه تعداد عامل عملکرد افراد در یک آزمون خاص را تبیین می‌کند. درحالی‌که در چارچوب نظریه سؤال‌پاســـخ، سؤال این است چه تعداد متغیر پنهان برای دست‌یابی به استقلال موضعی و یکنواختی[1] لازم است.

1. monotonicity

جدول (۱) چارچوب تصمیم‌گیری تعداد ابعاد بر اساس ویژگی‌های داده‌ها، فرایند فرضی پاسخ‌سؤال و رویکردهای ارزیابی بعدیت (اقتباس از اسوتینا و لوی، ۲۰۱۴)

داده‌های دوارزشی		رویکرد اکتشافی		رویکرد تأییدی	
داده ازدست‌رفته	مجانب پایین	پارامتریک	ناپارامتریک	پارامتریک	ناپارامتریک
بله	بله	●تحلیل عاملی اکتشافی[a] ●تحلیل موازی ●آزمون تفاوت کای دو[a]		●GDDM[f] ●شاخص‌های وابستگی موضعی[b]	
	خیر	●تحلیل عاملی اکتشافی[a,b] ●تحلیل موازی ●آزمون تفاوت کای دو[a]		●کای دوی تعدیل‌شده[b] ●WRMR[b] ●GDDM[f] ●شاخص‌های وابستگی موضعی[f]	
خیر	بله	●تحلیل عاملی اکتشافی[a,c] ●تحلیل موازی ●آزمون تفاوت کای دو[a] ●تغییر در RMSR[c]		● تغییر در RMSR[c] ●$\chi^2_{G/D}$[c] ●ALR[c] ●GDDM[f] ●شاخص‌های وابستگی موضعی[f]	●آزمون تک‌بعدی بودن برای ساختار ساده[g]
	خیر	●تحلیل عاملی اکتشافی[a,b,c] ●تحلیل موازی ●آزمون تفاوت کای دو[a] ●تغییر در RMSR[c]	●برآورد ساختار ساده چندبعدی[d]	●کای دوی تعدیل‌شده[b] ●WRMR[b] ● تغییر در RMSR[c] ●$\chi^2_{G/D}$[c] ●ALR[c] ●GDDM[f] ●شاخص‌های وابستگی موضعی[f]	●آزمون تک‌بعدی بودن ضروری برای ساختار ساده[g] ●ارزیابی ساختار ساده چندبعدی[d]
داده‌های چندارزشی رتبه‌ای		رویکرد اکتشافی		رویکرد تأییدی	
داده ازدست‌رفته		پارامتریک	ناپارامتریک	پارامتریک	ناپارامتریک
بله		●تحلیل عاملی اکتشافی[b] ●تحلیل موازی		●کای دوی تعدیل‌شده[b] ●WRMR[f] ●GDDM[f] ●شاخص‌های وابستگی موضعی[f]	
خیر		●تحلیل عاملی اکتشافی ●تحلیل موازی[b]	●برآورد ساختار ساده چندبعدی[e]	●کای دوی اصلاح‌شده[b] ●WRMR[b] ●GDDM[f] ●شاخص‌های وابستگی موضعی[f]	●آزمون تک‌بعدی برای ساختار ساده[h] ●ارزیابی ساختار ساده چندبعدی[i]

Note. Entries indicate the examples of such approaches discussed. GDDM: generalized dimensionality discrepancy measure; WRMR: weighted root mean square residual; RMSR root mean square residual; ALR: approximate likelihood ratio.[a] TESTFACT, which supports inputting of lower asymptotes. [b]Mplus. [c]Based on NOHARM, which supports inputting of lower asymptotes. [d]DETECT. [e]Nonegligible PolyDETECT. [f]Via PPMC. [g]DIMTEST, which supports inputting of a single lower asymptote. [h]PolyDIMTEST. [i]PolyDETECT.

صرف نظر از رویکرد اتخاذشده، می‌توان بعدیت را به‌عنوان کمترین تعداد صفات پنهان (عامل‌ها) در نظر گرفت که به‌طور کافی عملکرد زیربنایی آزمون‌دهندگان را تبیین می‌کنند. اصطلاح بعدیت تابه‌حال، به‌طور نادقیقی اســتفاده شده و لازم است که تصریح شود. به لحاظ تخصصی اشـاره به بعدیت آزمون می‌تواند گمراه‌کننده باشد. یک مفهوم ظریف در این خصوص، تشخیص این مطلب است که پاسخ سؤال‌ها، حاصل تعامل پیچیده سؤال‌های آزمون و آزمون‌دهندگان است که در آن ویژگی‌های هر دو در تولید ماتریس داده‌های مشاهده‌شده، مؤثر است. گاهی منظور از بعدیت گروه‌بندی معنی‌دار سؤال‌ها بر اساس ابعاد کمتری است، به‌طوری‌که سؤال‌ها به تفاوت‌های موجود در این ابعاد حسـاس باشـند. این نوع نگرش به بعدیت، نقش آزمون‌دهندگان را نادیده می‌گیرد. با توجه به آنچه ذکر شد بعدیت داده‌های حاصل از اجرای یک آزمون روی گروهی از افراد (ماتریس داده‌ها) از تعداد ابعاد آزمون که سـؤال‌ها به آن حساس هستند و تعداد ابعادی که افراد در آنها متفاوت بوده و تغییر می‌کنند، متأثر است و بعدیت ماتریس داده‌ها، کمینه تعداد ابعادی است که سؤال‌ها به آنها حساس‌اند یا آزمون‌دهندگان در آنها متفاوت بوده و تغییر می‌کنند. با در نظر گرفتن این نکته، تفسیر درست بعدیت عبارت است از: بعدیت یک مجموعه سؤال به‌شرط یک جامعه خاص. به‌این‌ترتیب، تعامل پیچیده فرد و سؤال در تولید ماتریس داده‌های مشاهده‌شده، لحاظ شده و این نکته نیز در نظر گرفته می‌شود که بعدیت یک مجموعه سؤال از یک جامعه به جامعه دیگر متفاوت خواهد بود (اسوتینا و لوی، ۲۰۱۴). شیوه‌های مختلف ارزیابی بعدیت از نظر شیوه عملیاتی کردن بعدیت بسنده[1] با هم فرق دارند. روش‌های مختلف ارزیابی بعدیت، بسته به ارزیابی شکل‌های مختلف پیش‌فرض استقلال موضعی با هم متفاوت بوده و عملیاتی کردن بعدیت به تعاریف مختلف از استقلال موضعی وابسته است. استقلال موضعی در شکل سنتی آن به این معنی است که پس از مشخص کردن ابعاد پنهان لازم، الگوی وابسـتگی بین سؤال‌ها در ابعاد ناپدید خواهد شد. یعنی صرف مشخص کردن درست تعداد ابعاد پنهان کافی است. این ادعا لزوماً درست نیست؛ چراکه اگر الگوی وابستگی سؤال‌ها در ابعاد نادرست باشد (مثلاً وابستگی یک سؤال به یک بعد نادیده گرفته شود، یا پارامترهای سؤال‌ها به‌طور نادرستی به مساوی بودن محدود شده باشند) استقلال موضعی به دست نخواهد آمد، حتی اگر تعداد ابعاد پنهان به‌درستی مشخص شده باشد. پس باید هم تعداد ابعاد و هم ساختار فرضی در نظر گرفته شوند. شکل ضعیف‌تر استقلال موضعی نیازمند استقلال زوج‌های سؤال است. یعنی کواریانس هر زوج سؤال به‌شرط توزیع توانایی برابر صفر است. استقلال اساسی که از استقلال موضعی زوجی نیز ضعیف‌تر است در صورتی برقرار است که به‌ازای تمام مقادیر توانایی، قدر مطلق کواریانس شرطی زوج‌های سؤال با توجه به تعداد سؤال‌ها مجانباً به صفر میل کند. اگر استقلال ضروری[2] برقرار باشد، تعداد ابعاد پنهان بعدیت اساسی داده‌ها را نشان می‌دهند. در داده‌های دوارزشی،

1. adequate dimensionality
2. Essential independence

بسته به بود و نبود داده‌های ازدست‌رفته[1]، بود و نبود مجانب پایین و اکتشافی یا تأییدی بودن تحلیل، روش‌های مختلفی برای ارزیابی بعدیت پیشنهاد شده است (اسوتینا و لوی، ۲۰۱۲). جدول (۱) تصمیم‌گیری در خصوص روش برآورد تعداد ابعاد را در حالت کلی نشان می‌دهد. بر اساس این جدول، می‌توان روش‌های ارزیابی بعدیت را از چند جنبه مورد توجه قرار داد و تقسیم‌بندی کرد. **طیف اکتشافی در برابر تأییدی**: در حالت اکتشافی، نه تعداد عامل‌ها از قبل مشخص است و نه الگوی وابستگی سؤال‌ها، در مقابل در حالت تأییدی، تعداد عامل‌ها و الگوی روابط بین متغیرها از قبل مشخص است. تمایز بین تحلیل تأییدی و تحلیل اکتشافی دقیق نیست و تا حدی بین این روش‌ها همپوشی وجود دارد. **طیف پارامتریک در برابر ناپارامتریک**: در حالت پارامتریک، فرم تابع ارتباط‌دهنده سؤال‌ها به ابعاد مشخص است، در مقابل، در حالت ناپارامتری تابعی مشخص نمی‌شود. در این حالت نیز در برخی موارد عدم اطمینان وجود دارد. **طیف داده‌های پاسخ**: می‌تواند دوارزشی یا چندارزشی باشـــد. **پارامتر مجانب پایین**: در سؤال‌های پاسخ انتخابی که به‌صورت دوارزشی نمره‌گذاری می‌شوند، لحاظ می‌شود. روش‌های بررسی بعدیت، بسته به اینکه چنین پارامتری در ارزیابی بعدیت لحاظ شود یا لحاظ نشود یا آن را برآورد کنند، متفاوت هستند. **داده‌های ازدست‌رفته**: روش‌های بررسی بعدیت از نظر برخورد با این داده‌ها متفاوت هستند. روش‌هایی که برخورد با داده‌های ازدست‌رفته را لحاظ می‌کنند، فقط موقعیت‌هایی را لحاظ می‌کنند که داده‌های ازدست‌رفته تصادفی یا کاملاً تصادفی است. از دست رفتگی داده‌ها در این دو حالت قابل چشم‌پوشی است (اسوتینا و لوی، ۲۰۱۴).

References

Caron, P. O. (2019). Minimum average partial correlation and parallel analysis: The influence of oblique structures. *Communications in Statistics-Simulation & Computation*, *48*(7), 2110-2117.

Chen, W. H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational & Behavioral Statistics*, *22*(3), 265-289.

Chiu, C. Y. (2013). Statistical Refinement of the Q-matrix in Cognitive Diagnosis. *Applied Psychological Measurement*, *37*(8), 598-618.

Cliff, N. (1988). The eigenvalues-greater-than-one rule and the reliability of components. *Psychological Bulletin*, *103*(2), 276.

Çokluk, Ö., & Koçak, D. (2016). Using Horn's Parallel Analysis Method in Exploratory Factor Analysis for Determining the Number of Factors. *Educational Sciences: Theory & Practice*, *16*(2).

Courtney, M. G. R., & Gordon, M. (2013). Determining the number of factors to retain in EFA: Using the SPSS R-Menu v2. 0 to make more judicious estimations. *Practical Assessment, Research & Evaluation*, *18*(8), 1-14.

1. Missing Data

Dinno, A. (2009). Exploring the Sensitivity of Horn's Parallel Analysis to the Distributional Form of Random Data. *Multivariate Behavioral Research*, *44*(3), 362–388.

Dziak, J. J., Coffman, D. L., Lanza, S. T., & Li, R. (2012). Sensitivity and specificity of information criteria. The Methodology Center and Department of Statistics. *Penn State, the Pennsylvania State University*, *16*(30), 140.

Fabozzi, F. J., Focardi, S. M., Rachev, S. T., & Arshanapalli, B. G. (2014). *The Basics of Financial Econometrics: Tools, Concepts, and Asset Management Applications*. John Wiley & Sons.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272.

Fraser C., McDonald R. P. (2012). *NOHARM 4: A Windows program for fitting both unidimensional and multidimensional normal ogive models of latent trait theory*.

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2011). Performance of Velicer's minimum average partial factor retention method with categorical variables. *Educational & Psychological Measurement*, *71*(3), 551-570.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2014). *Pearson new international edition.* In *Multivariate data analysis*, Seventh Edition. Pearson Education Limited Harlow Essex.

Hattie, J. (1985). Methodology review: assessing unidimensionality of tests and ltenls. *Applied Psychological Measurement*, *9*(2), 139-164.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179-185.

Horn, J. L., & Engstrom, R. (1979). Cattell's scree test in relation to Bartlett's chi-square test and other observations on the number of factors problem. *Multivariate Behavioral Research*, *14*(3), 283-300.

Houts, C. R., & Edwards, M. C. (2013). The performance of local dependence measures with psychological data. *Applied Psychological Measurement, 37*(7), 541-562.

Humphreys, L. G. (1964). Number of cases and number of factors: An example where N is very large. *Educational & Psychological Measurement, 24*(3), 457-466.

Kim, J. P. (2001). *Proximity measures and cluster analyses in multidimensional item response theory*. Unpublished doctoral dissertation, Michigan State University, East Lansing, MI.

Kuha, J. (2004). AIC and BIC: Comparisons of Assumptions and Performance. *Sociological Methods & Research*, *33*(2), 188-229.

Levy, R., & Svetina, D. (2011). A generalized dimensionality discrepancy measure for dimensionality assessment in multidimensional item response theory. *British Journal of Mathematical*

& Statistical Psychology, *64*(2), 208-232.

Lord, F. M., & Novick, M. R. (1968). statistical theories of mental test scores: Addison Wesley.

McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical & Statistical Psychology*, *34*(1), 100-117.

Pearson, R., Mundfrom, D., & Piccone, A. (2013). A comparison of ten methods for determining the number of factors in exploratory factor analysis. *Multiple Linear Regression Viewpoints*, *39*(1), 1-15.

Peterson, R. A. (2000). A meta-analysis of variance accounted for and factor loadings in exploratory factor analysis. *Marketing Letters*, *11*(3), 261-275.

Philip Chalmers (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, *48*(6), 1-29. doi:10.18637/jss.v048.i06

Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research*, *48*(1), 28-56.

Raiche, G., & Magis, D. (2010). *nFactors: Parallel Analysis and Non-Graphical Solutions to the Cattell Scree Test*. R Package Version 2.3. 3.

Raîche, G., Walls, T. A., Magis, D., Riopel, M., & Blais, J. G. (2013). Non-graphical solutions for Cattell's scree test. *Methodology*: *European Journal of Research Methods for the Behavioral and Social Sciences*, *9*(1), 23.

Reckase, M. D. (2009). *Multidimensional item response theory models*. In *Multidimensional item response theory* (pp. 79-112). Springer, New York, NY.

Revelle, W. (1979). Hierarchical cluster analysis and the internal structure of tests. *Multivariate Behavioral Research*, *14*(1), 57-74.

Revelle, W. (2017). *How To Use the psych package for Factor Analysis and data reduction.*

Revelle, W. (2018) *psych: Procedures for Personality and Psychological Research,* Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version = 1.8.12.

Revelle, W., & Rocklin, T. (1979). Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research*, *14*(4), 403-414.

Rietveld, T., & Van Hout, R. (1993). *Statistical techniques for the study of language behaviour*. Berlin: Mouton de Gruyter.

Robitzsch, A. (2021). *sirt: Supplementary Item Response Theory Models*. R package version 3.10-118. https://CRAN.R-project.org/package=sirt

Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor

analysis using comparison data of known factorial structure. *Psychological Assessment*, *24*(2), 282.
Shaycoft, M. F. (1970, March). *The eigenvalue myth and the dimension-reduction fallacy*. In meeting of the American Educational Research Association, Minneapolis.
Silverstein, A. B. (1987). Note on the parallel analysis criterion for determining the number of common factor or principal components. *Psychological Reports,* 61, 351–354.
Stellefson, M., & Hanik, B. (2008). *Strategies for Determining the Number of Factors to Retain in Exploratory Factor Analysis*. Online Submission.
Stout, W. F. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, *52*(4), 589-617.
Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, *55*(2), 293-325.
Streiner, D. L. (1994). Figuring out factors: the use and misuse of factor analysis. *The Canadian Journal of Psychiatry*, *39*(3), 135-140.
Svetina, D., & Levy, R. (2012). An Overview of Software for Conducting Dimensionality Assessment in Multidimensional Models. *Applied Psychological Measurement*, *36*(8), 659-669. doi: 10.1177/0146621612454593
Svetina, D., & Levy, R. (2014). A framework for dimensionality assessment for multidimensional item response models. *Educational Assessment*, *19*(1), 35-57.
Tabachnick, B. G. and Fidell, L. S. (2014) *Using Multivariate Statistics*. 6th edn. Harlow: Pearson.
Tate, R. (2003). A comparison of selected empirical methods for assessing theStructure of responses to test items. *Applied Psychological Measurement*, 27, 159-203.
Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. *Problems & Solutions in Human Assessment*, 41-71.
Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*(301), 236-244.
Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, *8*(2), 125-145.
Zwick, W. R., & Velicer, W. F. (1982). Factors influencing four rules for determining the number of components to retain. *Multivariate Behavioral Research*, *17*(2), 253-269.
Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*(3), 432.