مجله تخصصی اپیدمیولوژی ایران؛ ۱۳۹۵، دوره ۱۲، شماره ٤: صفحات ۷۱-٦٤.

مقاله مروری

سوء استفاده و سوء برداشت از ضریب آلفای کرونباخ به عنوان شاخصی برای ثبات درونی ابزار سنجش

رضا پاکزاد[۱]، فرشید علاءالدینی[۲]

[۱] دانشجوی دوره دکترای تخصصی اپیدمیولوژی، گروه اپیدمیولوژی و آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی تهران، تهران، ایران

[۲] مدیر کل پژوهش، توسعه و تجاری‌سازی، سازمان تدارکات پزشکی، جمعیت هلال احمر جمهوری اسلامی ایران، تهران، ایران

نویسنده رابط: فرشید علاء الدینی، نشانی: تهران، خیابان آیت‌الله طالقانی، تقاطع سپهبد قرنی، دفتر مرکزی سازمان تدارکات پزشکی هلال احمر، تلفن: ۹-۸۸۸۰۳۸۷۶

پست الکترونیک: falaedini@yahoo.com

تاریخ دریافت: ۹۴/۱۱/۲۱؛ پذیرش: ۹۵/۰۲/۰۴

مقدمه و اهداف: با وجود محاسبه راحت آلفای کرونباخ که قابلیت اطمینان درونیدرونی ابزار را می‌سنجد، استفاده نادرست و سوء برداشت از آن به صورت گسترده دیده شده است. عدم تفکیک تک‌بعدی و با ثبات بودن ابزار و نیز عدم توجه به تفاوت بین آلفای خام و آلفای استاندارد شده از موارد سوءبرداشت این ضریب است. هم‌چنین نقض پیش‌فرض‌های آلفا می‌تواند باعث افزایش یا کاهش آلفای محاسبه شده شود. در این مقاله نشان داده شده است که، یک ضریب آلفا بسیار بزرگ اما گمراه کننده می‌تواند هنگامی که فرضیه تک‌بعدی بودن نقض می‌شود، به‌دست آید. هم‌چنین استفاده از خطای استاندارد اندازه‌گیری به عنوان یک شاخص قابلیت اطمینان نشان داده شده است. در نهایت بحث تعمیم‌پذیری آلفا به طور مختصر بیان شده است.

واژگان کلیدی: قابلیت اطمینان، ثبات درونی درونی، ضریب آلفای کرونباخ، استفاده نادرست از آلفا، ضریب آلفای طبقه‌بندی شده

مقدمه

قابلیت اطمینان[۱] یا پایایی یک واژه کلی است که به هنگام تعیین اعتبار نتایج به دست آمده از ابزارها و اندازه‌گیری‌ها استفاده می‌شود و نشان دهنده ثبات در نتایج ابزار در هر بار اندازه‌گیری است (۲،۱). در این نوشته از واژه «ابزار» برای اشاره به هرگونه اندازه‌گیری کمی و چند آیتمی مثل آزمون‌ها و مقیاس‌ها استفاده شده است. هم‌چنین قابلیت اطمینان، پایایی و ثبات[۲] به عنوان مترداف هم به کار رفته شده است. تعاریف و انواع مختلفی از قابلیت اطمینان وجود دارد که از بین آن‌ها می‌توان به ثبات درونی[۳]، آزمون مجدد و هماهنگی را نام برد. هم‌چنین راه‌های متعددی برای به دست آوردن ضریب‌ها یا شاخص‌های قابلیت اطمینان وجود دارد، که از بین آن‌ها می‌توان به ماتریس‌های همبستگی یا کوواریانس و نیز آزمون باز آزمون و نیم‌تکه کردن نام برد. واژه پایایی را می‌توان در سه واژه ایستایی[۴]، هم‌ارزی[۵] و ثبات بیان کرد. با ثبات بودن یک ابزار را می‌توان با استفاده از شاخص معروف آلفای کرونباخ[۶] محاسبه کرد. استفاده از آلفای کرونباخ برای بیان ثبات یک ابزار بر خلاف روش آزمون- باز آزمون نیاز به یک بار اندازه‌گیری دارد. نتیجه این ویژگی اندازه‌گیری ثبات درونی توسط آلفا و استفاده راحت‌تر از آن نسبت به سایر شاخص‌ها است. با این وجود سوءبرداشت و سوءاستفاده از آلفای کرونباخ به فراوانی دیده شده است که از مهم‌ترین آن‌ها به موارد زیر می‌توان اشاره کرد:

۱- معمولاً پیش‌فرض‌های آلفای کرونباخ در مطالعه‌های مختلف رعایت نمی‌شود و این باعث کم‌برآوردی یا بیش‌برآوردی آلفا می‌شود؛

۲ - بسیاری از پژوهشگران آلفای استانداردشده را به آلفای خام ترجیح می‌دهند. توجیه آن‌ها این است که استفاده از شاخص استاندارد شده، وجود توزیع چوله را از بین می‌برند. در مورد آلفا این توجیه قابل اتکا نیست؛ به این دلیل که آلفا مبتنی بر ماتریکس کوواریانس است تا ماتریکس همبستگی؛

۳- در برخی شرایط پژوهشگران یک آلفا را برای کل ابزار

[۱] Reliability
[۲] Consistency
[۳] Internal consistency
[٤] Stability
[٥] Equivalence
[٦] Cronbach's alpha

[Downloaded from irje.tums.ac.ir on 2024-07-17]

استفاده می‌کنند و در صورت وجود ابعاد پنهان در ابزار آن را در نظر نمی‌گیرند؛

۴- وقتی یک پژوهشگر برای یک ابزار آلفا را گزارش می‌کنند در صورتی که مقدار عددی آن بالا باشد، به تک‌بعدی بودن ابزار اذعان می‌کند. دو مورد اخیر در ادامه توضیح داده شده است؛

۵- بسیاری از پژوهشگران دامنه‌ی عددی آلفای کرونباخ را ۰ تا ۱ می‌دانند، اما حقیقتاً وقتی کوواریانس آیتم‌ها خیلی کم باشد، آلفا می‌تواند کم‌تر از صفر یا منفی شود؛ و

۶- بسیاری از پژوهشگران در مطالعه خود از ابزارهای که توسط دیگران روا و پایا شده است؛ استفاده می‌کنند و به همین دلیل آلفای کرونباخ را برای ابزار بر اساس نمونه خود محاسبه نمی‌کنند. این سوءبرداشت از آن‌جا منشأ می‌شود، که آن‌ها نمی‌دانند که قابلیت اطمینان بر اساس نمره ابزار تعیین می‌شود و نه خود ابزار و در نتیجه مقدار آن از یک نمونه به نمونه دیگر متغیر است.

هر چند معیارهای متفاوتی برای تعیین حد قابل قبول آلفا وجود دارد، اما بیش‌تر پژوهشگران طبقه‌بندی که توسط Salvucci ارایه شده است را قبول دارند. بر اساس این طبقه‌بندی، اگر ضریب آلفا کوچک‌تر از ۰/۵۰، باشد؛ قابلیت اطمینان ابزار کم است و در صورتی که ضریب آلفا بین ۰/۸۰-۰/۵۰ باشد؛ قابلیت اطمینان متوسط و اگر بیش‌تر از ۰/۸۰ باشد؛ قابلیت اطمینان ابزار بالا است (۳). با این وجود راه‌های متفاوتی وجود دارد که بتوان یک مقدار آلفایی که برای یک ابزار محاسبه می‌شود؛ عددی بالا باشد. هنگامی که یک پژوهشگر ضریبی برای آلفا به‌دست می‌آورد که براساس معیارهای موجود به مقدار کافی قابل اطمینان و معتبر نباشد؛ سعی می‌کند که این مشکل را با حذف برخی از آیتم‌های ابزار اندازه‌گیری برطرف کند. به عبارت دیگر آیتم‌های که باعث کاهش قابلیت اطمینان می‌شود را حذف می‌کنند. برنامه‌های کامپیوتری پیشنهادهایی را برای حذف برخی از آیتم‌ها به منظور دست‌یابی به یک ضریب با قابلیت اطمینان درونی بالاتر ارایه می‌دهند. این روش برای افزایش برآوردهای قابلیت اطمینان امتیازهای به‌دست آمده نیز استفاده می‌شود. باید توجه داشت که این روش وقتی مناسب است که پژوهشگر قصد توسعه ابزار را ندارد و عدم توجه به این مسأله منجر به رخداد برخی از مشکلات جدی‌تر در مورد قابلیت اطمینان ابزار می‌شود. به همین دلیل نقدهایی بر آن نیز وجود دارد. برای مثال حذف آیتم‌های که کامپیوتر پیشنهاد به حذف آن‌ها می‌کند، می‌توانند منجر به کاهش روایی ابزار شود. Enders و همکاران میزان تأثیرپذیری ضریب آلفای کرونباخ از آیتم‌هایی که از لحاظ توزیع[1] دارای شکل‌های مختلف بودند را بررسی کردند. پژوهش‌های آن‌ها نشان داد که زمانی که میزان همبستگی بین آیتمی بالا و تعداد گروه‌ها زیاد باشد، قابلیت طمینان یک ابزار حداقل تأثیرپذیری را دارد و یادآور شدند که عمل حذف بخش‌هایی از یک ابزار بر اساس ناهنجاری‌های توزیعی، ممکن است تأثیر بدی بر میزان اعتبار ابزار بدون هیچ افزایش قابل ملاحظه‌ای در قابلیت اطمینان، داشته باشد (۴). Nunnally و همکاران نیز در مطالعه‌ی خود اظهار کردند که حذف برخی از آیتم‌هایی که همبستگی کلی کم یا انحراف معیار کوچک دارند، ممکن است میزان اعتبار ابزار را کاهش دهند. بنابراین، خارج کردن برخی از آیتم‌ها از ابزار به منظور بهبود ثبات درونی ممکن است به شدت منجر به کاهش اعتبار و محتوای سازه شود (۵).

پیش‌فرض آلفا

به طور معمول از آلفا به صورت حد پایینی محدوده قابلیت اطمینان یاد می‌شود، با این حال با دانستن فرضیه‌های دقیق و احتمال نقض آن‌ها و نیز بیان این‌که چقدر آسان آلفا می‌تواند متورم شود، می‌تواند اعتراض پژوهشگران را نسبت به این‌که چرا آلفا برای مدت زمان زیادی به عنوان یک استاندارد در علوم اجتماعی به کار می‌رفته است، برانگیزد (۶).

ضریب آلفا دارای سه پیش‌فرض اصلی است. پیش‌فرض نخست که به پیش‌فرض تئوری آزمون کلاسیک مشهور است؛ این است که نمره‌ی مشاهده شده برای هر آیتم، نتیجه جمع نمره واقعی و یک مقدار خطای آن در آیتم است. پیش‌فرض دوم این است که آلفا، معادل یا هم‌ارز تاو[2] می‌باشد که نشان می‌دهد تمامی آیتم‌ها بارگیری یکسانی دارند (یعنی نمره‌های واقعی مشابه به طور یکسان در نمره‌های تمامی آیتم‌ها مشارکت می‌کنند) و تمامی آیتم‌ها واریانس یکسانی دارند. پیش‌فرض سوم آلفا نیز این است که مقدارهای خطا، با هم همبستگی ندارند به عبارت دیگر مقدارهای خطاها غیر همبسته‌اند. در عمل ممکن است که هر سه این پیش‌فرض‌ها تا حدودی نقض شوند و بدین ترتیب صحت ضریب همبستگی آلفا به عنوان برآوردی از قابلیت اطمینان با مشکلاتی همراه شود (۷). بسیاری از ابزارهای پژوهشی که در علوم اجتماعی استفاده می‌شود؛ پیش‌فرض دوم را نقض می‌کنند و این یک مشکل اصلی برای آلفا است. پیش‌فرض دوم معمولاً به سه

[1] Distribution
[2] Tau-equivalent

[Downloaded from irje.tums.ac.ir on 2024-07-17]

دلیل نقص می‌شود: ۱- بیش‌تر از یک عامل پنهان در مقدار مشاهده شده یک آیتم (یا آیتم‌ها) مشارکت می‌کند؛ ۲- آیتم‌ها اغلب بارگیری‌های معادل با یک عامل پنهان ندارند؛ ۳- آیتم‌ها واریانس یکسانی ندارند. نقض هر کدام از این فرضیه‌ها منجر به برآورد تورشی از قابلیت اطمینان خواهد شد (۸). هم‌چنین مقدار عددی آلفا هنگامی که هر کدام از شرایط زیر وجود داشته باشد؛ بیش‌تر می‌شود: افزایش تعداد آیتم‌های یک ابزار، افزایش فاکتورهای پنهان مربوط به هر آیتم، افزایش مقادیر تکراری آیتم‌ها و افزایش آیتم‌های مشترک (۹،۱۰). این مسأله می‌تواند باعث سوءبرداشت در استفاده از آلفا شود؛ به‌ویژه هنگامی که ابعاد پنهان نیز در ابزار وجود دارد، اما مورد توجه قرار نگرفته است.

آلفای کرونباخ چه چیزی را می‌سنجد؟

شاخص آلفای کرونباخ مقدار واریانس مشترک بین آیتم‌های یک ابزار را می‌سنجند. مقدار عددی آلفای کرونباخ هنگامی عدد بالایی را نشان می‌دهد، که هر آیتم حداقل با یک آیتم دیگر در ابزار واریانس مشترک داشته باشد، این یعنی نیاز نیست با سایر آیتم‌های موجود در ابزار واریانس مشترک داشته باشد (۱۱). به عبارت دیگر آلفا وقتی بالا می‌شود که آیتم‌های یک ابزار، با هم همبستگی بالایی داشته باشند. از لحاظ ریاضی نیز قابل اثبات است که اگر ضریب همبستگی به توان دو رسانده شود، آلفا مقدار واریانس قابل پیشگویی (واریانس تبیین شده) نشان خواهد داد. اگر دو آیتم از لحاظ واریانس با هم هم‌پوشانی داشته باشند؛ به عبارت دیگر واریانس مشترک داشته باشند؛ تغییرات یکی را می‌توان بر اساس تغییرات دیگری تشخیص داد. این ویژگی واریانس مشترک یا واریانس تبیین شده است.

آلفای کرونباخ و فرضیه هم‌ارزی تاو

معمولاً آلفای کرونباخ شاخص مناسبی از قابلیت اطمینان درونی برای ابزارهایی است که هر آیتم آن بازه‌ای از پاسخ‌های ممکن را در نظر می‌گیرد (۱۲). در این مورد Feldt و همکاران بیان کردند «یک شرط لازم و کافی ضریب آلفا، برای این‌که برآوردی از قابلیت اطمینان یک آزمون باشد، آن است که امتیازها صرفاً معادل تاو باشند. این شرط دلالت بر آن دارد که آزمون از نظر تحلیلی تک‌بعدی باشد و تمامی بخش‌ها باید یک صفت واحد را اندازه‌گیری کنند» (۱۳).

در هنگام محاسبه قابلیت اطمینان، باید فرضیه تک‌بعدی بودن ابزار مورد نظر باشد. با این وجود، اگر فرضیه تک‌بعدی بودن نقض شود، دست‌یابی به قابلیت اطمینان بسیار بالا با استفاده از آلفا، رایج است. حتی زمانی که ابعاد مختلف را به طور کامل با یک ابزار اندازه‌گیری سنجیده شود، ممکن است باز هم ضریب قابلیت اطمینان آلفای کرونباخ عدد بسیار بالایی را نشان دهد. در این مورد، تنها گزارش ضریب آلفا به عنوان یک ضریب ثبات، شاخص ارزشمندی نیست و انجام این کار درست نیست. این وضع معمولاً در برخی از گزارش‌های پژوهش‌های مهم دیده می‌شود. Fedlt بیان کرده بود که هنگام استفاده از ابزارهای مختلف، فرضیه هم‌ارزی تاو یا تک‌بعدی[1] بودن ممکن است کاربرد ضریب قابلیت اطمینان را به شدت محدود کنند و از طرفی ممکن است فرضیه هم‌ارزی تاو یا تک‌بعدی بودن تا حدی در چنین آزمون‌هایی نقض شوند (۱۳).

استفاده‌های نادرست از ضریب آلفا

همان‌طور که در بالا اشاره شد، در برخی از گزارش‌های پژوهش‌ها از ضریب ثبات درونی استفاده‌های نادرستی می‌شود. برخی از مثال‌های مربوط به آن عبارت‌اند از: ۱- عدم گزارش قابلیت اطمینان امتیازهای جمع‌آوری شده داده‌های حقیقی یک مطالعه؛ ۲- در نظر نگرفتن ساختار عاملی داده‌ها؛ ۳- گزارش خطای استاندارد اندازه‌گیری به عنون قابلیت اطمینان در زمانی که فرضیه تک‌بعدی بودن نقض شده است؛ ۴- استفاده از ضریب آلفای کرونباخ به عنوان شاخص هم ارز بودن. هر کدام از این موارد به ریز در ادامه توضیح داده شده است.

استفاده نادرست ۱- عدم گزارش قابلیت اطمینان امتیازهای جمع شده در مطالعه حقیقی

زمانی که یک ابزار در مطالعه‌ای مورد استفاده قرار می‌گیرد، قابلیت اطمینان امتیازها باید گزارش شوند. قابلیت اطمینان امتیازها برای یک ابزار ممکن است مقدارهای مختلفی را در مطالعه‌های متفاوت داشته باشد. برای مثال یک ابزار می‌تواند در یک مطالعه قابل اطمینان باشد، اما در مطالعه‌ای دیگر، قابل اطمینان نباشد (۱۴). بنا به نظر Rudner قابلیت اطمینان هر ابزار از یک مطالعه به مطالعه‌ای دیگر و حتی از یک گروه به گروه دیگر متفاوت است (۱۵). از این‌رو، مطالعه‌ای مناسب است که قابلیت اطمینان امتیازهای جمع شده ابزار را گزارش کند.

هم‌چنین باید به این مورد توجه کرد که زمانی که پژوهشگران

[1] Uni-dimensional

[Downloaded from irje.tums.ac.ir on 2024-07-17]

از یک ابزار بسیار قابل اطمینان و ارزشمند استفاده می‌کنند، امتیازهای جمع‌آوری شده از این ابزار برای گروه مورد استفاده در مطالعه ممکن است به دلیل خطای ناشی از شرایط به کارگیری و امتیازدهی قابل اطمینان نباشد. به همین دلیل، هر زمانی که یک ابزار اندازه‌گیری در یک گروه مورد استفاده قرار می‌گیرد، قابلیت اطمینان امتیازهای این گروه نیز باید محاسبه و گزارش شود. بدون انجام این کار، تنها گزارش قابلیت اطمینان ابزار مناسب نیست و این مثالی از استفاده نادرست از ضریب قابلیت اطمینان و به ویژه ضریب آلفا می‌باشد.

استفاده نادرست ۲- در نظر نگرفتن ساختار عاملی داده‌ها

در صورتی که یک مقیاس دارای آیتم‌های کافی باشد (یعنی بیش از ۲۰)، می‌تواند آلفای بزرگ‌تر از ۰/۷۰ حتی زمانی که همبستگی میان آیتم‌ها بسیار کم است؛ داشته باشد. Cortina در مطالعه خود نشان داد که در صورتی که یک ابزار دارای بیش از ۱۴ آیتم بوده و همبستگی بین آیتم‌ها متوسط باشد؛ آلفای آن می‌تواند ۰/۷۰ یا بیش‌تر باشد. هم‌چنین نشان داد در صورتی که تعداد آیتم‌های یک ابزار از لحاظ تعداد کافی بوده و همبستگی بین آیتم‌ها بالا باشد، مقدار کلی آلفا ۰/۸۰ یا بیش از آن خواهد داشت؛ حتی اگر ابزار دارای چند بعد جداگانه باشد (۱۶). Tan این مسأله را نیز در مطالعه خود نشان داد و یادآور شد زمانی که ابزاری دارای فاکتورهای متعدد باشد، باز هم یک ضریب قابلیت اطمینان بزرگ می‌تواند با استفاده از یک ضریب آلفای کرونباخ به عنوان یک ضریب ثبات درونی به دست آورد (۲).

در این هنگام، آیتم‌ها می‌توانند به صورتی که از نظر درونی ثابت و همسان باشند؛ تفسیر شوند، اما این موضوع دقیقاً صحیح نمی‌باشد. تنها گزارش ضریب آلفای کرونباخ به صورتی که از نظر درونی ثابت و همسان باشد ممکن است پژوهشگران را گمراه کرده و این امر مثالی از استفاده نادرست از ضریب آلفای کرونباخ است. بنابرین، زمانی که لازم است یک ضریب ثبات درونی گزارش شود، ارایه تنها یک ضریب آلفای کرونباخ کافی نیست. Schmittt نیز بر اهمیت این موضوع تأکید کرد و پیشنهاد داد در صورتی که خواننده به طور کافی آگاه به یافته‌های به‌دست آمده باشد، هم همبستگی و هم آلفا باید گزارش شوند (۱۷).

زمانی که یک ضریب آلفای کرونباخ گزارش می‌شود، ساختار عاملی داده‌ها باید همراه با آن گزارش شوند. نه تنها باید میزان تناسب ضریب آلفای کرونباخ تشریح شود، بلکه تناسب ساختار عاملی داده‌ها نیز باید به بحث گذاشته شود. هم‌چنین زمانی که ابزار مورد استفاده دارای چند بعد[1] است، باید ضریب آلفا برای هر بعد آن جداگانه محاسبه شود (۱۸).

یک روش گزارش برآورد قابلیت اطمینان آلفا هنگامی که مقیاس دارای بعدهای متعدد است، استفاده از ضریب آلفای طبقه‌بندی شده[2] است (۱۳،۱۹). فرمول ضریب آلفای طبقه‌بندی شده به ترتیب زیر می‌باشد.

$$\text{Strat}\alpha\,\rho_{xx'} = 1 - \frac{\sum^{c}\sigma^2_{x_j}(1-\alpha\rho_{x_j x'_j})}{\sigma^2_{X_{tot}}}.$$

XJ نمره هر بعد، σ2Xj واریانس نمرات بعد j، αρxjxj´ ضریب آلفا برای بعد j، C تعداد بعدها و Xtot مجموع امتیازهای به‌دست آمده از مجموع X1+...+Xc است. هنگامی که ابزار دارای بعدهای متعدد است، استفاده از ضریب آلفای طبقه‌بندی شده بهترین روش برای محاسبه و گزارش قابلیت اطمینان ابزار می‌باشد.

هنگامی که فرضیه‌های ضریب آلفای کرونباخ نقض می‌شود، می‌توان از ضریب تتا[3] (θ) که یک شاخص دیگر برای محاسبه ثبات درونی است؛ استفاده کرد. این شاخص شامل یک آنالیز مؤلفه‌های اصلی آیتم‌های ابزار مورد استفاده است و آیتم‌هایی را که بیش‌تر با یکدیگر همبستگی دارند را برای به حداکثر رساندن آلفا، توزین می‌کند (۲۰). فرمول ضریب تتا براساس زیر است:

$$\text{Coefficient theta }(\theta) = \frac{N}{N-1}\left(1-\frac{1}{\lambda}\right)$$

که در آن N = تعداد آیتم‌های ابزار و λ = بیش‌ترین مقدار ویژه که از آنالیز مؤلفه‌های اصلی روی ماتریکس همبستگی آیتم‌های مورد بررسی به دست می‌آید (۲۱).

استفاده نادرست ۳- گزارش خطای استاندارد اندازه‌گیری به عنوان یک شاخص قابلیت اطمینان هنگامی که فرضیه تک‌بعدی بودن نقض می‌شود.

خطای استاندارد اندازه‌گیری ضریب قابلیت اطیمنان، روشی دیگر برای بیان قابلیت اطمینان ابزار هستند. خطای استاندارد اندازه‌گیری محدوده امتیاز احتمالی است که ممکن است در آن امتیاز واقعی یک فرد تغییر کند و این مقدار تابع قابلیت اطمینان ابزار است (۲۲). معمولاً وقتی آلفای کرونباخ گزارش می‌شود، همراه آن خطای استاندارد اندازه‌گیری نیز گزارش می‌شود. گزارش

[1] Dimension
[2] Stratified alpha coefficient
[3] Theta coefficient

[Downloaded from irje.tums.ac.ir on 2024-07-17]

خطای استاندارد اندازه‌گیری به همراه ضریب آلفای کرونباخ، فرصتی برای مشاهده خطای امتیازهای قابلیت اطمینان برای ابزار مورد استفاده است. در صورتی که فرضیه تک‌بعدی بودن در مجموعه داده‌های این مطالعه بررسی نشود، ممکن است استفاده از خطای استاندارد اندازه‌گیری، ژوهشگران را نیز گمراه کند.

زمانی که فرضیه تک‌بعدی بودن نقض می‌شود، استفاده از خطای استاندارد اندازه‌گیری برای دست‌یابی به نظراتی در مورد امتیاز خطا می‌تواند پژوهشگران را گمراه کند و این مثالی از استفاده نادرست از ضریب آلفای کرونباخ است. باید توجه داشت که مقدار خطای استاندارد اندازه‌گیری در درجه‌ی اول بستگی به مقدار برآورد قابلیت اطمینان ضریب آلفا دارد. در صورتی که آلفای کرونباخ سنجش مناسبی برای یک ابزار نباشند، خطای استاندارد اندازه‌گیری نیز سنجش مناسبی برای تفسیر قابلیت اطمینان یا خطاهای تصادفی امتیازات ابزار نیستند.

استفاده نادرست ۴- استفاده از ضریب آلفای کرونباخ به عنوان شاخص همسان بودن

یک رویکرد به مسأله همسانی آیتم، انجام یک آنالیز فاکتوریال برای یک ابزار است. در صورتی که ابزار تنها ۱ بعد داشته باشد؛ آیتم‌ها همسان‌اند. اگر یافته‌های آنالیز نشان دهد که ابزار دارای بیش از ۱ بعد است، ممکن است تقسیم ابزار به چند بخش که هر کدام یک بعد خاص را نشان می‌دهند؛ مطلوب باشد (۲۳). تلاش‌های زیادی برای تفسیر ضریب آلفا از نظر دیگر ویژگی‌های ابزار مانند همسان بودن یا تک‌بعدی بودن آن در طول سالیان متمادی، به عمل آمده است (۲۴). براساس نظر پژوهشگران مختلف، معادله‌ی ضریب قابلیت اطیمنان آلفا، تنها بر اساس این فرضیه است که میانگین واریانس مشترک میان آیتم‌های غیر موازی برابر با میانگین واریانس مشترک میان آیتم‌های موازی است (۲۳). این موضوع به این معنی است که آیتم‌های موجود در یک ابزار به طوری با هم همسان‌اند که هر آیتم، ابعاد کلی مشابهی را مانند دیگر آیتم‌ها، اندازه‌گیری می‌کند. با این وجود این پرسش پیش می‌آید که رابطه‌ی میان آلفا و همسان بودن چیست؟ باید توجه داشت که آلفا تابع میزان اشتراک بالا و تفاوت کم آیتم‌های موجود در یک ابزار است. هم‌چنین تابع ارتباط درونی نیز هست، اما باید توجه داشت که این امر دلالت بر تک‌بعدی یا همسان بودن نیست (۱۶).

در واقع مقدار ضریب آلفا تابع واریانس مشترک میان آیتم‌ها است. واریانس مشترک میان آیتم‌ها میانگین مقدار ضریب آلفا را تعیین می‌کند (۱۱). زمانی که ابعاد مختلفی در یک ابزار با هم ترکیب می‌شوند، ممکن است مقدار واریانس مشترک میان آیتم‌ها نیز افزایش یابد. در این مواقع، می‌توان برآورد بالاتری از قابلیت اطمینان در کل ابزار نسبت به سایر بعدها آن به‌دست آورد. به عبارتی دیگر وقتی که همبستگی ابعاد اندازه‌گیری نشوند، شرایط آزمون و عدم شفافیت آیتم‌ها، ممکن است فاکتور مشترکی در امتیازهای ابزار ایجاد کند. در این حالت واریانس مشترک میان آیتم‌ها ممکن است افزایش یابد و منجر به داشتن برآورد ثبات درونی بالاتر شود.

بنابراین، پرسش مهمی که در این شرایط ایجاد می‌شود؛ آن است که آیا ترکیب امتیازهای ابعاد یک ابزار از نظر تئوری درست است؟ در صورتی که پاسخ این پرسش «بله» باشد، نباید نگران تک‌بعدی یا همسان بودن ابزار بود و می‌توان از ضریب آلفا به عنوان برآورد قابلیت اطیمنان استفاده کرد (۱۳). با این وجود، اگر پاسخ «خیر» باشد، ممکن است استفاده از ضریب آلفا گمراه کننده باشد. چون ضریب آلفا معمولاً ضریب ثبات درونی نامیده می‌شوند و بسیاری از پژوهشگران آن را به عنوان مقدار همسانی آیتم‌ها یا تک‌بعدی بودن بررسی کرده‌اند و یافته‌ها نشان داده است که آلفا، شاخص خوبی برای این ویژگی ابزارنیست (۲۵،۹).

ثبات و ابعاد

چیزی که واضح است و بسیاری از پژوهشگران بر آن هم‌نظر هستند این است که آلفا ثبات یا قابلیت اطمینان یک ابزار را می‌سنجد. برای این‌که یک ابزار ثبات داشته باشد یک نقطه برش تعیین شده است؛ به طوری که بیش‌تر یک ابزار را هنگامی با ثبات می‌دانند که آلفای آن بیش‌تر از ۰/۷۰ باشد (۵). با این وجود در برخی شرایط آلفای کم‌تر از ۰/۷۰ را به عنوان بی‌ثباتی ابزار می‌دانند. متأسفانه این یک درک نادرست است؛ به این دلیل که ممکن است یک ابزار بیش از یک حیطه یا بعد را بسنجد و اگر این ابعاد متفاوت باشند و این مسأله در هنگام بررسی آلفا در نظر گرفته نشود؛ باعث کاهش آلفای محاسبه شده شود. برای مثال نمره‌های GRE-Verbal، GRE-Quantitative و GRE-Analytical با هم همبستگی بالایی ندارند؛ چون عموماً بعدهای مختلفی را می‌سنجند؛ در نتیجه اگر این مسأله را در نظر نگیرید و برای آزمون GRE یک آلفا گزارش شود، به طور قطعی مقدار آلفا کم خواهد بود، اما این آلفای کم به خاطر بی‌ثبات بودن آزمون GRE نیست، بلکه به خاطر ادغام بعدهایی است که اصولاً با هم متفاوت هستند.

[Downloaded from irje.tums.ac.ir on 2024-07-17]

اگر هنگام بررسی یک ابزار، ثبات درونی آن کم بود، باید وجود مشکل فوق را در نظر بگیرید. راه‌حل این مسأله آن است که از روش‌های آنالیز فاکتوریال یا تجزیه و تحلیل مؤلفه‌های اصلی استفاده کرد، تا وجود بعد در ابزار را تشخیص داد و یا اجزایی که شبیه هم هستند را به عنوان یک بعد در نظر بگیرید. در برخی مواقع نیز می‌توان آن آیتمی که تأثیر منفی بر شاخص ثبات درون (آلفا) دارد را حذف کرد. با این وجود اگر از وجود ابعاد مختلف در یک ابزار اطلاع داشته باشید باید این مسأله را در نظر گرفته و آلفا را برای هر بعد جداگانه سنجید (۱۱).

از سویی دیگر ممکن است هنگام سنجش ثبات درونی، مقدار آلفا بیش‌تر از ۰/۷۰ باشد. بیش‌تر بودن آلفا از ۰/۷۰ به‌طور حتمی نشان دهنده این نیست که ابزار دارای یک بعد است. Cortina در مقاله خود به این مسأله اشاره کرد و بیان کرده بود که بسیاری از پژوهشگران هنگامی از آلفا برای ثبات درونی ابزار خود استفاده می‌کنند، اگر مقدار آن بیش‌تر از ۰/۷۰ باشد، دیگر وجود بعدهای دیگر در ابزار را مورد بررسی قرار نمی‌دهند. او این مورد را مثال بارزی از سوءاستفاده از آزمون‌های آماری و استفاده نامناسب از آلفا دانسته بود (۱۶).

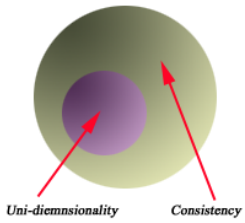
آن‌چه مسلم است این است که وجود آلفای پایین ممکن است نشان دهنده‌ی وجود ابعاد مختلف در ابزار مورد سنجش باشد، اما وجود آلفای بالا نشان دهنده عدم وجود ابعاد چندگانه در ابزار مورد سنجش نیست. با این وجود برخی از افراد این مسأله را قبول ندارند و می‌گویند که آلفای بالا نشان دهنده تک‌بعدی بودن ابزار است. در نتیجه نیاز با بررسی بیش‌تر برای وجود بعدهای دیگر در ابزار نیست. این یک سوء برداشت شایع است؛ چون عموماً ثبات و ابعاد دو جزء کاملاً مجزای یک ابزار هستند و باید به طور جداگانه بررسی شوند (۱۷).

آن‌چه که مسلم است این است که تک‌بعدی بودن زیر مجموعه با ثبات بودن است؛ اگر یک آزمون تک‌بعدی باشد آن با ثبات است. شکل شماره ۱ این موضوع را نشان می‌دهد، اما با ثبات بودن نشان دهنده‌ی تک‌بعدی بودن نیست (۱۱). Kelley این مسأله را با این مثال منطقی توضیح می‌دهد «اگر من یک مرد هستم، پس من انسان هستم، اما اگر انسان باشم ممکن است مرد نباشم؛ شاید زن باشم» (۲۶). متأسفانه این اشتباه و سوء برداشت اغلب در استفاده و تفسیر آلفا رخ می‌دهد.

همان‌طور که بیان شد، آلفای کرونباخ مقدار واریانس مشترک را می‌سنجد. هر وقت واریانس هر آیتم با حداقل یکی دیگر از سایر آیتم‌ها مشترک باشد، مقدار آلفای کرونباخ بیش‌تر است (۱۱).

لازم به ذکر است نیاز نیست هر آیتم با همه آیتم‌های دیگر واریانس مشترک داشته باشد. Gardner برای نشان دادن ابعاد و ثبات یک مقیاس ۹ آیتمی را طراحی کرد که در شکل‌های شماره ۴-۲ نشان داده شده است (۲۷). در شکل شماره ۲ هیچ‌کدام از آیتم‌ها با سایر آیتم‌ها واریانس مشترک ندارند. به عبارت دیگر، هیچ هم‌پوشانی بین آیتم‌ها وجود ندارد. در این حالت نه ثبات درونی و نه تک‌بعدی بودن مقیاس، وجود ندارد. نتیجه این شرایط یک آلفای کم و عدم وجود تک‌بعدی بودن مقیاس است.

در شکل شماره ۳، نشان داده شده است که همه‌ی آیتم‌ها حداقل با سایر آیتم‌ها واریانس مشترک دارند. این نشان دهنده‌ی ثبات درونی بالا و تک‌بعدی بودن مقیاس است. اما در شکل شماره ۴، هر آیتم فقط با برخی از آیتم‌ها واریانس مشترک دارند. این نشان دهنده‌ی ثبات درونی بالا است، اما مقیاس اصلاً تک‌بعدی نیست.



شکل ۱- رابطه تک بعدی و ثبات

شکل ۲- عدم وجود ثبات درونی و تک بعدی بودن مقیاس

شکل ۳- وجود ثبات درونی و تک بعدی بودن مقیاس

شکل ۴- وجود ثبات درونی و عدم تک بعدی بودن مقیاس

تعمیم‌پذیری آلفا

در برخی مواقع یک پژوهشگر، ابزاری را در مطالعه خود استفاده کرده است و روا و پایای آن را پس از بررسی قبول می‌کند. در این شرایط اگر کسی دیگر از این ابزار استفاده کند؛ از انجام روایی و پایایی ابزار روی داده‌های مطالعه خود صرف‌نظر می‌کند. منطق آن هم این است که چون ابزار در مطالعه‌ای دیگر روا و پایا شده است؛ دیگر نیاز به بررسی آن بر روی داده‌های مطالعه خود نیست. این

[Downloaded from irje.tums.ac.ir on 2024-07-17]

یک سوء برداشت از آلفا است (۲۸). Henson این مسأله را مورد نقد و بررسی قرار داده و بیان کرده است که اگر کسی از یک ابزاری استفاده می‌کند، صرف نظر از روا و پایا بودن ابزار، باید آن را در مطالعه خود و روی داده‌های خویش مورد بررسی قرار دهد چون ممکن است برخی آیتم‌های آن نیاز به تغییر داشته باشد. او این مسأله را مطرح می‌کند که چون ویژگی‌های نمونه‌های مختلف متفاوت است، تعمیم‌پذیری آلفا می‌تواند کار درستی نباشد (۲۹).

جمع‌بندی

ضریب آلفای کرونباخ یک شاخصی است که برای تعیین قابلیت اطمینان یا ثبات درونی ابزار استفاده می‌شود، اما معمولاً به خاطر چگونگی محاسبه‌ی آن و پیچیدگی‌های خاص خود ممکن است مورد سوء برداشت و استفاده‌ی نادرست قرار گیرد. باید هنگام محاسبه آلفای کرونباخ به پیش‌فرض‌های آن، وجود ابعاد پنهان و قابلیت تعمیم آن توجه کرد و در صورت نقص پیش‌فرض‌های آن می‌توان از شاخص های جایگزین شامل تتا استفاده کرد.

منابع

1. Starkweather J. Step out of the past: Stop using coefficient alpha; there are better ways to calculate reliability. University of North Texas Research and statistical support Retrieved September. 2012; 3.
2. Tan S. Misuses of KR-20 and Cronbach's alpha reliability coefficients. Education and Science. 2009; 34: 101-12.
3. Salvucci S, Walter E, Conley V, Fink S, Saba M. Measurement Error Studies at the National Center for Education Statistics. . ERIC Digest, Washington 1997. 115.
4. Enders CK, Bandalos DL. The effects of heterogeneous item distributions on reliability. Applied Measurement in Education. 1999; 12: 133-50.
5. Nunnally J. Psychometric Theory, 2nd edition (McGraw-Hill Psychology Ser). 1978; 115
6. Sijtsma K. On the use, the misuse, and the very limited usefulness of Cronbach's alpha. Psychometrika. 2009; 74: 107-20.
7. Yang Y, Green SB. Coefficient alpha: A reliability coefficient for the 21st century? Journal of Psychoeducational Assessment. 2011; 29: 4377-92.
8. Shevlin M, Miles J, Davies M, Walker S. Coefficient alpha: a useful indicator of reliability? Personality and Individual Differences. 2000; 28: 229-37.
9. Hattie J. An empirical study of various indices for determining unidimensionality. Multivariate Behavioral Research. 1984; 19: 49-78.
10. Raykov T. Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. Multivariate Behavioral Research. 1997; 32: 329-53.
11. Gardner PL. The dimensionality of attitude scales: a widely misunderstood idea. International Journal of Science Education. 1996; 18: 913-9.
12. McMillan JH. Educational research: Fundamentals for the consumer: HarperCollins College Publishers, New York, 1996; 363.
13. Feldt LS, Qualls AL. Bias in coefficient alpha arising from heterogeneity. Applied Measurement in Education. 1996; 9: 277-86.
14. Vockell E, Asher J. educational Research, 2nd Edition. Pearson publication, United State. 1995; 150.
15. Rudner LM, Schafer WD. Reliability. ERIC Development Team. ERIC Digest. 2001; 2-3.
16. Cortina JM. What is coefficient alpha? An examination of theory and applications. Journal of applied psychology. 1993; 78: 98-104.
17. Schmitt N. Uses and abuses of coefficient alpha. Psychological assessment. 1996; 8: 350-3.
18. Rae G. A note on using alpha and stratified alpha to estimate the reliability of a test composed of item parcels. British Journal of Mathematical and Statistical Psychology. 2008; 61: 515-25.
19. Cronbach LJ, Schönemann P, McKie D. Alpha coefficients for Stratified-Parallel Tests. Educational and Psychological Measurement. 196; 25: 291-312.
20. Carmines EG, Zeller RA. Reliability and validity assessment: Sage publications; Londan. 1979; 41
21. Zumbo BD, Gadermann AM, Zeisser C. Ordinal versions of coefficients alpha and theta for Likert rating scales. Journal of Modern Applied Statistical Methods. 2007; 6: 21-9.
22. Drummond RJ. Appraisal procedures for counselors and helping professionals: Prentice Hall; New York. 2004.
23. Gulliksen H. Theory of mental tests: Routledge, New York. 2013; 84
24. Zimmerman DW, Zumbo BD, Lalonde C. Coefficient alpha as an estimate of test reliability under violation of two assumptions. Educational and Psychological Measurement. 1993; 53: 33-49.
25. Leedy P, Ormrod JE. Practical research, Planning and design, 9th edition. Pearson Education, Inc. publishing as Merrill, New Jersey. 2005; 29
26. Kelley D. The Art of Reasoning, 3rd Edition: W. W. Norton & Company; New Jersey. 1998; 320
27. Gardner PL. Measuring attitudes to science: Unidimensionality and internal consistency revisited. Research in science education. 1995; 25: 283-9.
28. Yu CH.(editor). An introduction to computing and interpreting Cronbach Coefficient Alpha in SAS. Proceedings of 26th SAS User Group International Conference. p: 1-5. April 22-25, 2001 United states of America, Long Beach, California.
29. Henson RK, Thompson B. Characterizing Measurement Error in Test Scores across Studies: A Tutorial on Conducting" Reliability Generalization Analyses. Annual Meeting of the American Educational Research Association. Seattle, Washington, April 10-14, 2001.

[Downloaded from irje.tums.ac.ir on 2024-07-17]

Iranian Journal of Epidemiology 2017; 12(4): 64-71.
Review Article

Misuse and Misconception of Cronbach's Alpha Coefficient as an Index of Internal Consistency of Measuring Tools

Pakzad R[1], Alaeddini F[2]

1- PhD Student in Epidemiology, Department of Epidemiology, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran
2- General Manager of Research, Development and Commercialization, Medical Procurement Organization, Red Crescent Society of the Islamic Republic of Iran, Tehran, Iran
Corresponding author: Alaedini F, falaedini@yahoo.com

Calculation of Cronbach's Alpha which measures the internal reliability of a tool is easy, but misuse and misconception are widely observed. Common misunderstandings of the Cronbach's Alpha include inability to differentiate between being one-dimensional and consistency, and between raw and standardized alpha. Violations of the assumptions of alpha could also cause an increase or decrease in the calculated alpha. In this paper, we showed that a very large but confusing alpha coefficient could be obtained, particularly when the one-dimensional theory is violated. Also, the use of standard error measurement as an index of the reliability has been presented. Finally, the generalizability of alpha coefficient is also discussed.

Keywords: Reliability, Internal consistency, Cronbach's alpha coefficient, Misuses of cronbach's alpha, Stratified alpha coefficient

[Downloaded from irje.tums.ac.ir on 2024-07-17]