

بسم الله الرحمن الرحيم



موضوع :

"ارزیابی پایایی و روایی یک پرسشنامه"

Assessing the reliability and validity of a questionnaire

پروژه کارشناسی

نگارش: نرگس افشاری

شماره دانشجویی: ۹۹۱۸۴۹۳

استاد راهنما: دکتر حامد لرونند

دانشگاه صنعتی اصفهان

تابستان ۱۴۰۴

این پروژه کارشناسی با هدف ارزیابی پایایی و روایی پرسشنامه GAD-7 که پرسشنامه اختلال اضطراب فراگیر ۷ آیتمی با مقیاس لیکرت ۴ درجه ای است، بر روی دیتاست "Mental health data of 776 Mexican medical students" از سایت Figshare انجام شد. جامعه آماری شامل ۷۷۶ دانشجوی پزشکی مکزیک بود که پس از حذف داده های گمشده، نمونه نهایی کاهش یافته و به ۷۵۹ رکورد رسید. روش شناسی پژوهش کمی و توصیفی - تحلیلی بود و از نرم افزار SPSS برای تحلیل های آماری استفاده شد. در بخش روایی، روایی محتوایی با شاخص های CVR و CVI و روایی سازه با تحلیل عاملی اکتشافی (EFA) و تأییدی (CFA) بررسی شد. روایی ملاکی نیز با رگرسیون خطی برای پیش بینی افکار ترک تحصیل ارزیابی گردید. در بخش پایایی، آلفای کرونباخ، روش نصف سازی با همبستگی تعدیل شده، آزمون-بازآزمون با ضریب همبستگی و هم ارزی نیز با ضریب همبستگی بین GAD-7 و PHQ-9 محاسبه شد. نتایج نشان داد که GAD-7 پایایی عالی و روایی قابل قبول تا قوی دارد، اما آیتم های gad5 و gad7 ضعیف تر هستند و نیاز به بومی سازی فرهنگی دارند. این ابزار برای غربالگری اضطراب در دانشجویان پزشکی معتبر است، اما محدودیت هایی مانند شبیه سازی فرضی در روش های آزمون-بازآزمون و هم ارزی وجود دارد. پیشنهاد می شود مطالعات آینده با داده های طولی و نمونه ایرانی انجام شود تا تعمیم پذیری افزایش یابد.

فهرست مطالب

فصل اول: مقدمه	۵
مقدمه	۶
اهمیت بررسی موضوع ارزیابی پایایی و روایی یک پرسشنامه	۷
اهمیت کلی ارزیابی پایایی و روایی	۷
چرا ارزیابی پایایی و روایی ضروری است؟	۸
ارتباط این موضوع با مسائل واقعی جامعه	۹
اهداف پژوهش	۱۰
هدف کلی	۱۰
اهداف جزئی	۱۱
فصل دوم: مبانی نظری	۱۳
تعریف پرسشنامه به عنوان ابزار اندازه‌گیری	۱۴
ویژگی‌های کلیدی پرسشنامه	۱۴
تاریخچه و تکامل پرسشنامه‌ها	۱۴
انواع پرسشنامه‌ها	۱۵
مزایای پرسشنامه‌ها به عنوان ابزار اندازه‌گیری	۱۶
معایب و محدودیت‌های پرسشنامه‌ها	۱۷
جایگاه پرسشنامه‌ها در تحقیقات علمی و عملی	۱۷
روایی	۱۹
تعریف روایی	۱۹
اهمیت روایی	۱۹
انواع روایی	۲۰
روش‌های محاسبه روایی	۲۲
پایایی	۲۳
تعریف پایایی	۲۳
اهمیت پایایی	۲۴
روش‌های محاسبه پایایی	۲۴
ارتباط میان روایی و پایایی	۳۱

فصل سوم: روش شناسی پژوهش	۳۳
طرح پژوهش	۳۴
جامعه آماری، نمونه و روش نمونه گیری	۳۵
معرفی و توصیف پرسشنامه	۳۷
علت انتخاب سنجش شدت اضطراب عمومی	۳۸
اهمیت سنجش اضطراب عمومی	۳۸
پرسشنامه	۳۹
روش های بررسی روایی در این پژوهش	۴۰
روش های بررسی پایایی در این پژوهش	۴۳
روش های تجزیه و تحلیل داده ها	۴۵
فصل چهارم: یافته های پژوهش	۴۶
توصیف ویژگی های جامعه و نمونه	۴۷
تحلیل روایی پرسشنامه	۵۷
- روایی محتوایی	۵۷
- روایی سازه	۶۵
- روایی ملاکی	۷۵
تحلیل پایایی پرسشنامه	۸۰
- روش آلفای کرونباخ	۸۰
- روش نصف سازی	۸۵
- روش آزمون - باز آزمون	۸۹
- روش هم ارزی	۹۳
فصل پنجم: نتیجه گیری	۹۵
خلاصه فرآیند پژوهش	۹۶
نتایج و بحث	۹۶
منابع	۹۸

فصل اول: مقدمه

در دنیای امروز، ابزارهای اندازه گیری مانند پرسشنامه ها نقش کلیدی در جمع آوری داده های علمی و عملی ایفا می کنند، به ویژه در حوزه های علوم اجتماعی، روان شناسی و بهداشت. پرسشنامه GAD-7، که برای سنجش شدت اختلال اضطراب فراگیر توسعه یافته، یکی از ابزارهای رایج غربالگری است. ارزیابی پایایی (ثبات نتایج) و روایی (دقت در اندازه گیری سازه مورد نظر) این پرسشنامه نه تنها اعتبار علمی پژوهش را تضمین می کند، بلکه پایه ای برای کاربرد های عملی مانند تشخیص زودرس اضطراب در گروه های پرخطر مانند دانشجویان پزشکی فراهم می آورد. اضطراب عمومی یکی از شایع ترین اختلالات روانی است که طبق گزارش های جهانی، شیوع آن در دانشجویان پزشکی تا ۳۰٪ می رسد و می تواند به عواقبی مانند کاهش عملکرد تحصیلی، افسردگی و افکار ترک تحصیل منجر شود. در ایران، با توجه به فشارهای آموزشی و اجتماعی، نیاز به ابزارهای معتبر غربالگری مانند GAD-7 بیش از پیش احساس می شود، اما عدم ارزیابی روان سنجی می تواند به داده های نامعتبر و تصمیم گیری های غلط منجر شود.

این پروژه کارشناسی با تمرکز بر ارزیابی پایایی و روایی GAD-7 بر روی دیتاست شامل ۷۷۶ دانشجوی پزشکی مکزیک انجام شد. انتخاب این دیتاست به دلیل تنوع متغیرهای مرتبط با سلامت روان مانند PHQ-9 برای افسردگی و Epworth برای خواب آلودگی و حجم نمونه کافی است که امکان تحلیل های آماری دقیق را فراهم می کند. مسئله اصلی پژوهش، کمبود ارزیابی روان سنجی GAD-7 در جمعیت های دانشجویی است که می تواند به تشخیص نادرست و هدررفت منابع منجر شود. اهمیت این موضوع در جوامع در حال توسعه مانند ایران دوچندان است، جایی که استرس تحصیلی بالا و دسترسی محدود به متخصصان، نیاز به ابزارهای ساده و معتبر را افزایش می دهد. ارزیابی پایایی با روش های آلفای کرونباخ، نصف سازی، آزمون - بازآزمون و هم ارزی و روایی با روش های روایی محتوایی، روایی سازه و روایی ملاکی، نه تنها اعتبار ابزار را تأیید می کند، بلکه به بومی سازی آن برای کاربرد های محلی نیز کمک می رساند.

اهداف پژوهش به دو بخش کلی و جزئی تقسیم می شود. هدف کلی، ارزیابی کیفیت کلی پرسشنامه GAD-7 از منظر پایایی و روایی است تا اطمینان حاصل شود که ابزار به طور دقیق اضطراب را می سنجد. اهداف جزئی شامل تعیین پایایی داخلی با آلفای کرونباخ، بررسی ثبات زمانی با آزمون - بازآزمون شبیه سازی شده بر اساس ترم های تحصیلی، ارزیابی هم ارزی با همبستگی GAD-7 و PHQ-9، و بررسی روایی ملاکی با پیش بینی افکار ترک تحصیل می شود. ساختار این پژوهش به شرح زیر است: فصل دوم به مبانی نظری اختصاص دارد و مفاهیم روایی و پایایی را توصیف می کند. فصل سوم روش شناسی پژوهش را شامل طرح، جامعه آماری، نمونه گیری و روش های تحلیل توضیح می دهد. فصل چهارم یافته ها را با خروجی های SPSS و تحلیل های آماری ارائه می دهد. در نهایت، فصل پنجم به نتیجه گیری، محدودیت ها و پیشنهاد ها در این پژوهش می پردازد.

اهمیت بررسی موضوع ارزیابی پایایی و روایی یک پرسشنامه

در دنیای پژوهش‌های علمی، به‌ویژه در حوزه‌های علوم اجتماعی، روانشناسی، مدیریت و بهداشت، پرسشنامه‌ها به عنوان یکی از رایج‌ترین ابزارهای جمع‌آوری داده‌ها عمل می‌کنند.

پایایی^۱ به معنای ثبات و سازگاری نتایج ابزار اندازه‌گیری در شرایط مشابه است (مانند اینکه آیا سؤالات پرسشنامه همیشه پاسخ‌های یکسانی از یک فرد در زمان‌های مختلف دریافت می‌کنند)، در حالی که روایی^۲ به دقت ابزار در اندازه‌گیری آنچه که واقعاً قصد اندازه‌گیری آن را دارد، اشاره دارد (مانند اینکه آیا پرسشنامه واقعاً "اضطراب" را می‌سنجد یا چیز دیگری). ارزیابی این دو ویژگی، نه تنها اعتبار علمی یک تحقیق را تضمین می‌کند، بلکه پایه‌ای برای کاربردهای عملی و تصمیم‌گیری‌های مبتنی بر شواهد فراهم می‌آورد.

اهمیت کلی ارزیابی پایایی و روایی

ارزیابی پایایی و روایی، قلب تپنده هر ابزار اندازه‌گیری است و بدون آن، نتایج تحقیق مانند ساختمانی بدون پی می‌ماند - ظاهراً محکم، اما ناپایدار و غیرقابل اعتماد. طبق تعریف استاندارد، پایایی تضمین می‌کند که ابزار "قابل تکرار" است، یعنی نتایج آن در آزمون‌های مکرر تغییر ناچیزی نشان می‌دهد، در حالی که روایی اطمینان می‌دهد که ابزار "دقیق" است و آنچه را که ادعا می‌کند، اندازه‌گیری می‌کند. این ارزیابی‌ها اهمیت دارند زیرا:

- تضمین دقت و ثبات داده‌ها:

در پژوهش‌های کمی، بیش از ۷۰٪ مطالعات از پرسشنامه‌ها استفاده می‌کنند، اما بدون ارزیابی، داده‌ها ممکن است پر از نویز (مانند خطاهای تصادفی یا سیستماتیک) باشند. برای مثال، ضریب آلفای کرونباخ (برای پایایی) بالای ۰,۷ نشان‌دهنده ثبات خوب است، و شاخص‌هایی مانند CVI (برای روایی محتوایی) بالای ۰,۸، دقت محتوایی را تأیید می‌کند. بدون این‌ها، تحلیل‌های آماری (مانند رگرسیون یا تحلیل عاملی) بی‌معنی می‌شوند.

¹ Reliability

² Validity

- افزایش قابلیت تعمیم‌پذیری نتایج:

پژوهش‌های معتبر باید بتوانند به جوامع بزرگ‌تر تعمیم یابند. ارزیابی‌روایی سازه (مانند تحلیل عاملی اکتشافی) کمک می‌کند تا مطمئن شویم پرسشنامه در فرهنگ‌ها یا زمینه‌های مختلف (مثل ایران) کارآمد است، نه فقط در نسخه اصلی.

- کاهش بایاس و خطاهای روش‌شناختی:

ابزارهای نامعتبر می‌توانند بایاس‌های فرهنگی، زبانی یا شناختی ایجاد کنند، که در نهایت به "تحقیقات جعلی" یا "نتایج گمراه‌کننده" منجر می‌شود. مطالعه‌ای در سال ۲۰۲۴ نشان داد که بیش از ۴۰٪ پرسشنامه‌های استفاده‌شده در تحقیقات پزشکی فاقد ارزیابی کامل‌روایی هستند، که این امر اعتبار کل حوزه را زیر سؤال می‌برد.

به طور خلاصه، این ارزیابی‌ها مانند "کنترل کیفیت" در صنعت هستند: بدون آن‌ها، پژوهش از یک ابزار علمی به یک نظرسنجی آماتوری تنزل می‌یابد.

چرا ارزیابی پایایی و روایی ضروری است؟

این کار نه تنها یک الزام آکادمیک است، بلکه ضرورتی اجتناب‌ناپذیر برای جلوگیری از هدررفت منابع و تضمین پیشرفت علمی است. دلایل کلیدی عبارتند از:

- جلوگیری از تصمیم‌گیری‌های غلط بر اساس داده‌های نادرست:

تصور کنید یک پرسشنامه رضایت مشتری بدون روایی محتوایی، به یک شرکت بگوید "مشتریان راضی‌اند"، در حالی که واقعاً خشمگین‌اند. این امر منجر به سیاست‌های ناکارآمد می‌شود. در پژوهش، عدم پایایی می‌تواند واریانس تصادفی ایجاد کند و آزمون‌های آماری را نامعتبر سازد، که طبق استانداردهای COSMIN (برای ابزارهای اندازه‌گیری سلامت)، حداقل ۸۰٪ مطالعات باید این ارزیابی‌ها را انجام دهند.

- ارتقای کیفیت تحقیقات کمی:

در عصر داده‌های بزرگ، پرسشنامه‌ها اغلب با ابزارهای دیجیتال ترکیب می‌شوند، اما بدون ارزیابی، سازگاری فرمت‌ها (مثل آنلاین و کاغذی) از بین می‌رود. یک گزارش در سال ۲۰۲۵ تأکید می‌کند که اولویت بندی روایی و پایایی، نرخ خطای نظرسنجی‌ها را تا ۳۰٪ کاهش می‌دهد.

- پاسخ به چالش های اخلاقی و علمی:

کمیته های اخلاق پژوهش (مانند IRB) الزام می کنند که ابزارها معتبر باشند تا از آسیب به شرکت کنندگان (مانند استرس ناشی از سؤالات نامربوط) جلوگیری شود. علاوه بر این، در انتشار مقالات (مثل ISI)، عدم ارزیابی منجر به رد مقاله می شود - بیش از ۲۵٪ ردهای ژورنال های معتبر به دلیل ضعف روش شناختی است.

-نوآوری در روش ها:

ارزیابی های مدرن (مانند CFA برای روایی سازه) اجازه می دهد پرسشنامه ها را بومی سازی کنیم، مثلاً برای جوامع کم درآمد، که بدون آن، تحقیقات نابرابر می ماند.

در نهایت، این کار ضروری است زیرا پژوهش بدون اعتبار، نه تنها بی فایده، بلکه بالقوه مضر است - مانند تجویز دارویی بدون آزمایش بالینی.

ارتباط این موضوع با مسائل واقعی جامعه

ارزیابی پایایی و روایی فراتر از دیوارهای دانشگاه است و مستقیماً بر مسائل اجتماعی، اقتصادی و بهداشتی تأثیر می گذارد. فقدان این ارزیابی ها می تواند به نابرابری ها، هدررفت منابع و حتی بحران های اجتماعی دامن بزند. مثال های واقعی:

- در سیاست گذاری و اقتصاد:

پرسشنامه های نظرسنجی عمومی (مانند سنجش فقر یا رضایت شهروندی) اگر نامعتبر باشند، سیاست های غلطی مانند تخصیص بودجه نادرست را به دنبال دارند. مثلاً، در کشورهای در حال توسعه، نظرسنجی های WTP^۱ برای خدمات عمومی اگر روایی پایینی داشته باشند، منجر به عدم سرمایه گذاری در زیرساخت ها می شود و نابرابری را تشدید می کند. مطالعه ای در سال ۲۰۲۳ نشان داد که داده های نامعتبر در نظرسنجی های فقر، ۱۵٪ خطا در برنامه ریزی های دولتی ایجاد کرده است. در ایران، ارزیابی ضعیف پرسشنامه های اقتصادی می تواند به سیاست های ناکارآمد در مقابله با تورم منجر شود.

^۱ Willingness to Pay

- در بهداشت و درمان:

ابزارهای غربالگری (مانند پرسشنامه‌های افسردگی یا رضایت بیمار) اگر پایایی نداشته باشند، تشخیص‌های غلط ایجاد می‌کنند. در جوامع کم درآمد، داده‌های جامعه محور توسط کارگران بهداشت اگر نامعتبر باشند، نرخ خطای ۲۰٪-۳۰٪ در نظارت بر بیماری‌ها (مانند COVID-19) را به همراه دارد، که منجر به هدررفت منابع و مرگ و میر بالاتر می‌شود. مثلاً، در کشورهای فقیر، پرسشنامه‌های تجربه بیمار (مانند ESQ) با روایی پایین، برنامه‌های واکسیناسیون را مختل می‌کنند و نابرابری بهداشتی را افزایش می‌دهند.

- در آموزش و مسائل اجتماعی:

در ارزیابی عملکرد مدارس یا رضایت دانش‌آموزان، پرسشنامه‌های نامعتبر می‌توانند به سیاست‌های آموزشی ناکارآمد منجر شوند، مانند نادیده گرفتن مشکلات روانی دانش‌آموزان در جوامع محروم. یک بررسی در سال ۲۰۲۴ نشان داد که خطاهای روایی در نظرسنجی‌های اجتماعی، ۱۰٪-۱۵٪ انحراف در برنامه‌ریزی‌های حمایتی ایجاد می‌کند، که بر مهاجران یا اقلیت‌ها تأثیر بیشتری دارد.

- تأثیر بر پایداری اجتماعی:

در عصر تغییرات اقلیمی، پرسشنامه‌های سنجش آگاهی عمومی اگر معتبر نباشند، کمپین‌های زیست محیطی را بی‌اثر می‌کنند. علاوه بر این، در جوامع دیجیتال، نظرسنجی‌های آنلاین بدون ارزیابی، "صداهاى جعلی" (مانند بات‌ها) را تقویت کرده و دموکراسی را تهدید می‌کنند.

اهداف پژوهش

هدف کلی

هدف کلی، چارچوب اصلی پروژه را تعریف می‌کند و به طور گسترده بیان می‌کند که پروژه به دنبال چیست. در اینجا، هدف کلی، بررسی کیفیت کلی پرسشنامه "پرسشنامه اختلال اضطراب فراگیر ۷" از دو منظر پایایی (ثبات نتایج) و روایی (دقت در اندازه‌گیری سازه موردنظر) است؛ به منظور اطمینان از دقت و ثبات آن به عنوان یک ابزار اندازه‌گیری معتبر برای استفاده در تحقیقات علمی و کاربردهای عملی در جامعه ایرانی.

اهداف جزئی

اهداف جزئی، اجزای مشخص و قابل اندازه‌گیری هدف کلی هستند که پروژه را به بخش‌های عملیاتی تقسیم می‌کنند. این اهداف باید با روش‌های آماری (مانند آلفای کرونباخ، تحلیل عاملی یا شاخص‌های روایی محتوایی) همخوانی داشته باشند.

هدف جزئی ۱: تعیین پایایی پرسشنامه از طریق بررسی همسانی درونی با استفاده از ضریب آلفای کرونباخ.

این هدف به ارزیابی ثبات درونی پرسشنامه می‌پردازد، یعنی بررسی اینکه آیا سؤالات پرسشنامه به طور منسجم یک سازه را می‌سنجند. ضریب آلفای کرونباخ^۱ استاندارد طلایی برای این منظور است، که مقادیر بالای ۰,۷ نشان‌دهنده پایایی خوب است. این تحلیل با نرم‌افزار SPSS انجام می‌شود و نتایج آن در جداول گزارش خواهد شد.

هدف جزئی ۲: ارزیابی پایایی بازآزمایی^۲ برای بررسی ثبات زمانی پاسخ‌های پرسشنامه.

این هدف بررسی می‌کند که آیا نتایج پرسشنامه در دو بازه زمانی (مثلاً با فاصله دو هفته) مشابه باقی می‌مانند. این امر با محاسبه ضریب همبستگی (مانند Pearson یا ICC) انجام می‌شود، که مقادیر بالای ۰,۷ نشان‌دهنده ثبات زمانی خوب است.

هدف جزئی ۳: ارزیابی روایی صوری و محتوایی پرسشنامه با استفاده از نظرات کارشناسان و شاخص‌های CVR و CVI.

این هدف بررسی می‌کند که آیا پرسشنامه از نظر ظاهری (صوری) برای پاسخ‌دهندگان قابل فهم است و آیا محتوای آن (محتوایی) سازه مورد نظر را به طور کامل پوشش می‌دهد. شاخص‌های نسبت روایی محتوا (CVR) و شاخص روایی محتوا (CVI) با استفاده از نظر حداقل ۵-۱۰ کارشناس محاسبه می‌شوند. طبق فرمول لوشه^۳، CVR بالای ۰,۶ و CVI بالای ۰,۸ نشان‌دهنده روایی مناسب است.

هدف جزئی ۴: بررسی روایی سازه پرسشنامه با استفاده از تحلیل عاملی اکتشافی (EFA) و تأییدی (CFA).

این هدف پیچیده‌ترین بخش ارزیابی روایی است و بررسی می‌کند که آیا پرسشنامه واقعاً سازه موردنظر را می‌سنجد. تحلیل عاملی اکتشافی (EFA) ساختار زیربنایی پرسشنامه را شناسایی می‌کند (مثلاً چند بعد دارد)، و تحلیل عاملی تأییدی (CFA) با نرم‌افزارهایی مانند AMOS تأیید می‌کند که ساختار پیشنهادی با داده‌ها

¹ Cronbach's Alpha

² Test-Retest Reliability

³ Lawshe

همخوانی دارد. شاخص‌هایی مانند KMO (بالای ۰/۶) و آزمون بارتلت (معنادار بودن) برای این منظور استفاده می‌شوند. این هدف برای بومی‌سازی پرسشنامه در ایران (به دلیل تفاوت‌های فرهنگی) بسیار مهم است.

هدف جزئی ۵: بررسی روایی ملاکی^۱ از طریق مقایسه پرسشنامه با ابزارهای استاندارد موجود.

این هدف بررسی می‌کند که آیا پرسشنامه با ابزارهای معتبر دیگر (مانند یک پرسشنامه استاندارد بین‌المللی) همبستگی دارد (روایی همگرا) یا از سازه‌های غیرمرتبط متمایز است (روایی واگرا). این هدف به ویژه برای اطمینان از کاربرد عملی ابزار در تحقیقات مقایسه‌ای مهم است.

¹ Criterion Validity

فصل دوم: مبانی نظری

تعریف پرسشنامه به عنوان ابزار اندازه‌گیری

پرسشنامه یک ابزار ساختار یافته برای جمع‌آوری داده‌هاست که از مجموعه‌ای از سؤالات استاندارد تشکیل شده و برای اندازه‌گیری ویژگی‌ها، نگرش‌ها، رفتارها یا عقاید افراد طراحی شده است. این ابزار در تحقیقات کمی و کیفی برای گردآوری اطلاعات از نمونه‌های بزرگ به کار می‌رود و به دلیل استاندارد سازی، امکان مقایسه و تحلیل آماری را فراهم می‌کند. طبق تعریف کرسول^۱، پرسشنامه یک روش نظام مند برای تبدیل مفاهیم نظری (مانند رضایت شغلی یا کیفیت زندگی) به داده‌های قابل اندازه‌گیری است.

ویژگی‌های کلیدی پرسشنامه

ساختارمند بودن: سؤالات از پیش طراحی شده و پاسخ‌ها معمولاً در قالب مقیاس‌های استاندارد (مانند لیکرت) ارائه می‌شوند.

قابلیت تکرار پذیری: امکان استفاده در مطالعات مختلف با شرایط مشابه.

انعطاف پذیری: قابل استفاده در حوزه‌های متنوع، از روانشناسی تا سیاست گذاری.

کارایی: امکان جمع‌آوری داده از تعداد زیادی پاسخ دهنده در زمان کوتاه.

تاریخچه و تکامل پرسشنامه‌ها

استفاده از پرسشنامه‌ها به قرن نوزدهم بازمی‌گردد، زمانی که دانشمندانی مانند فرانیسیس گالتون از سؤالات مکتوب برای بررسی ویژگی‌های روان شناختی استفاده کردند. در قرن بیستم، با توسعه روان سنجی^۲، ابزارهایی مانند پرسشنامه‌های شخصیت و مقیاس‌های نگرشی (مانند مقیاس لیکرت) ظهور کردند. در دهه‌های اخیر، پیشرفت فناوری (مانند نظرسنجی‌های آنلاین و پلتفرم‌هایی مانند Google Forms) جایگاه پرسشنامه‌ها را تقویت کرده است. مطالعه‌ای در سال ۲۰۲۴ نشان داد که بیش از ۷۰٪ تحقیقات علوم اجتماعی از پرسشنامه‌های دیجیتال استفاده می‌کنند، که کارایی را ۳۰٪ افزایش داده اما چالش‌های جدیدی مانند بایاس پاسخ دهی آنلاین ایجاد کرده است.

¹ Creswell, 2020

² Psychometrics

انواع پرسشنامه ها

پرسشنامه ها بر اساس هدف، ساختار و روش اجرا به انواع مختلفی تقسیم می‌شوند. شناخت این انواع به پژوهشگر کمک می‌کند تا جایگاه پرسشنامه در تحقیق خود را بهتر مشخص کند.

• انواع پرسشنامه ، بر اساس نوع داده ها عبارتند از:

پرسشنامه های کمی: برای جمع‌آوری داده های عددی (مانند مقیاس لیکرت ۵ درجه ای).

پرسشنامه های کیفی: شامل سؤالات باز برای تحلیل محتوای عمیق.

• انواع پرسشنامه، بر اساس هدف تحقیق عبارتند از:

توصیفی: برای توصیف ویژگی های جامعه (مانند پرسشنامه‌های دموگرافیک).

نگرشی: برای سنجش عقاید و نگرش ها (مانند پرسشنامه های سنجش اعتماد سازمانی).

رفتاری: برای بررسی رفتارها (مانند پرسشنامه های عادات غذایی).

• انواع پرسشنامه، بر اساس روش اجرا عبارتند از:

خودگزارش دهی^۱: پاسخ‌دهندگان به تنهایی پرسشنامه را تکمیل می‌کنند (مانند فرم‌های آنلاین).

مصاحبه‌ای: توسط پژوهشگر تکمیل می‌شود (مناسب برای جمعیت های کم‌سواد).

دیجیتال: از طریق پلتفرم‌های آنلاین.

¹ Self-Administered

به طور خلاصه :

نوع پرسشنامه	ویژگی ها	مثال	کاربرد
کمی	ساختارمند، مقیاس دار	SF-36	تحلیل آماری
کیفی	سؤالات باز، انعطاف پذیر	مصاحبه های عمیق	تحلیل مضمون
خود گزارش دهی	سریع، کم هزینه	JDI	نظرسنجی های گسترده
دیجیتال	دسترسی بالا، تحلیل سریع	Google Forms	تحقیقات مدرن

مزایای پرسشنامه ها به عنوان ابزار اندازه گیری

پرسشنامه ها به دلیل ویژگی های منحصر به فردشان، جایگاه ویژه ای در تحقیقات دارند:

کارایی و مقیاس پذیری: امکان جمع آوری داده از نمونه های بزرگ (صدها یا هزاران نفر) در زمان کوتاه. مثلاً، یک پرسشنامه آنلاین می تواند در چند روز داده های ۱۰۰۰ نفر را جمع آوری کند.

استاندارد سازی: سؤالات یکسان برای همه پاسخ دهندگان، قابلیت مقایسه داده ها را افزایش می دهد.

انعطاف پذیری: قابل استفاده در حوزه های مختلف (آموزش، سلامت، مدیریت) و برای موضوعات متنوع (از رضایت مشتری تا سلامت روان).

هزینه کم: نسبت به روش های دیگر (مانند مصاحبه های عمیق یا مشاهده)، پرسشنامه ها اقتصادی تر هستند.

تحلیل آماری پیشرفته: داده های پرسشنامه ها برای تحلیل های پیچیده (مانند تحلیل عاملی یا رگرسیون) مناسب هستند.

معایب و محدودیت های پرسشنامه ها

با وجود مزایا، پرسشنامه ها چالش هایی دارند که باید به آن پرداخته شود:

- بایاس پاسخ دهی:

پاسخ دهندگان ممکن است به دلایل اجتماعی یا روان شناختی (مانند تمایل به مقبولیت اجتماعی) پاسخ های غیرصادقانه بدهند. مطالعه ای در سال ۲۰۲۳ نشان داد که ۲۰٪-۳۰٪ پاسخ ها در نظرسنجی های آنلاین تحت تأثیر این بایاس هستند.

- وابستگی به سواد و فرهنگ:

در جوامع کم سواد یا چند فرهنگی (مانند ایران)، طراحی نامناسب پرسشنامه می تواند به سوء تفاهم منجر شود.

- محدودیت در عمق داده ها:

پرسشنامه های کمی معمولاً نمی توانند به اندازه مصاحبه های کیفی، اطلاعات عمیق ارائه دهند.

- نیاز به ارزیابی پایایی و روایی:

بدون این ارزیابی ها، پرسشنامه ممکن است داده های نادرست تولید کند، که به تصمیم گیری های غلط منجر می شود.

- چالش های دیجیتال:

در پرسشنامه های آنلاین، نرخ پاسخ دهی پایین (گاهی کمتر از ۱۰٪) و مشکلات فنی می توانند کیفیت داده ها را کاهش دهند.

جایگاه پرسشنامه ها در تحقیقات علمی و عملی

پرسشنامه ها به دلیل ویژگی هایشان، در تحقیقات جایگاه محوری دارند و در حوزه های مختلف علمی و اجتماعی کاربرد دارند:

- در تحقیقات علمی

علوم اجتماعی و روانشناسی: پرسشنامه ها برای سنجش مفاهیم انتزاعی مانند نگرش، انگیزه یا افسردگی استفاده می‌شوند. مثلاً، پرسشنامه بک^۱ برای تشخیص افسردگی در بیش از ۱۰۰۰ مطالعه در سال ۲۰۲۴ استفاده شده است.

مدیریت و بازاریابی: پرسشنامه های رضایت مشتری (مانند SERVQUAL) برای تحلیل رفتار مصرف‌کننده و بهبود خدمات استفاده می‌شوند.

علوم پزشکی: ابزارهایی مانند SF-36 یا WHOQOL برای ارزیابی کیفیت زندگی در بیماران مزمن کاربرد دارند.

آموزش: پرسشنامه های ارزیابی عملکرد معلمان یا رضایت دانش‌آموزان به بهبود سیاست های آموزشی کمک می‌کنند.

- در کاربردهای عملی

سیاست گذاری: پرسشنامه های نظرسنجی عمومی (مانند سنجش رضایت شهروندی) برای تخصیص بودجه های دولتی استفاده می‌شوند. مثلاً، داده های معتبر می‌توانند به توزیع عادلانه تر منابع در مناطق محروم کمک کنند.

مدیریت منابع انسانی: پرسشنامه های رضایت شغلی به کاهش نرخ ترک کمک می‌کنند.

بهداشت عمومی: پرسشنامه های غربالگری (مانند PHQ-9 برای افسردگی) در برنامه های سلامت روان، به ویژه در جوامع کم درآمد، نقش کلیدی دارند.

- در تحقیقات دیجیتال و مدرن:

با رشد فناوری، پرسشنامه های دیجیتال (مانند Qualtrics) امکان تحلیل بلادرنگ داده ها را فراهم کرده‌اند. یک مطالعه در سال ۲۰۲۵ نشان داد که ۶۵٪ تحقیقات بین‌المللی از پرسشنامه های آنلاین استفاده می‌کنند، اما نیاز به ارزیابی روایی دیجیتال (مانند سازگاری با نسخه کاغذی) همچنان وجود دارد.

¹ BDI

تعریف روایی

روایی^۱ به معنای دقت و صحت یک ابزار اندازه گیری در سنجش آنچه که ادعا می کند می سنجد، است. به عبارت دیگر، یک ابزار روا باید دقیقاً سازه موردنظر را اندازه گیری کند، نه چیز دیگری. طبق کرسول^۲، روایی نشان می دهد که آیا نتایج پرسشنامه با تعریف نظری سازه همخوانی دارد و آیا می توان به تفسیرهای حاصل از آن اعتماد کرد. روایی نه تنها به تطابق با هدف پژوهش اشاره دارد، بلکه به این پرسش پاسخ می دهد که ابزار تا چه حد خصیصه مورد نظر را بدون بایاس یا خطای سیستماتیک می سنجد.

در استاندارد های مدرن، روایی به عنوان یک مفهوم چند بعدی دیده می شود. طبق مدل (Mokkink et al., 2018) COSMIN، روایی نه تنها محتوایی و ساختاری است، بلکه شامل جنبه های فرهنگی و زبانی نیز می شود، به ویژه در بومی سازی ابزارها. مثلاً، اگر پرسشنامه ای برای ارزیابی "تعهد سازمانی کارکنان" طراحی شده اما پرسش هایش به گونه ای هستند که "رضایت شغلی" را اندازه گیری می کنند، روایی ندارد - حتی اگر پایایی آماری بالایی داشته باشد. همچنین، پرسشنامه ای که همه جنبه های موضوع مورد پژوهش را دربر نگیرد، روایی مطلوبی ندارد.

این نکته کلیدی است: پاسخ گو و پژوهشگر باید مفهوم یکسانی از پرسش استنباط کنند تا اعتبار پرسش حفظ شود.

اهمیت روایی

روایی هسته اصلی اعتبار علمی و عملی یک ابزار است، زیرا ابزار بدون روایی ممکن است داده های گمراه کننده تولید کند که به تصمیم گیری های غلط منجر می شود. دلایل کلیدی اهمیت روایی عبارتند از:

- تضمین دقت مفهومی: روایی اطمینان می دهد که پرسشنامه سازه درست را می سنجد.
- کاربرد عملی: در ایران، ابزارهای روا برای برنامه های غربالگری در دانشگاه ها یا بیمارستان ها ضروری اند.
- اعتبار علمی پژوهش: طبق استاندارد های COSMIN، ابزار بدون روایی نمی تواند برای انتشار در ژورنال های معتبر (مثل ISI) استفاده شود - حدود ۳۵٪ رد مقالات به دلیل ضعف روایی است.

¹ Accuracy

² Creswell, 2020

- بومی‌سازی فرهنگی: روایی به ویژه در جوامع چند فرهنگی مانند ایران اهمیت دارد، زیرا بایاس زبانی یا فرهنگی می‌تواند دقت را کاهش دهد. نظر کارشناسان و خبرگان کمک خوبی برای بهبود روایی ابزار اندازه‌گیری است.

- کاهش هزینه‌ها و خطاها: ابزارهای بدون روایی می‌توانند به تشخیص‌های نادرست (مثل غربالگری اشتباه اضطراب) منجر شوند، که هزینه‌های درمانی را تا ۲۰٪ افزایش می‌دهد.

انواع روایی

گونه‌های متفاوتی از روایی برای ابزار اندازه‌گیری تعریف می‌شوند. این انواع بر اساس استاندارد های روان‌سنجی (مانند APA, 2017) گسترش یافته‌اند. انواع اصلی روایی عبارتند از:

۱. روایی صوری^۱

- تعریف: بررسی ظاهری اینکه آیا سؤالات پرسشنامه به نظر پاسخ‌دهندگان و کارشناسان مرتبط با سازه می‌آیند یا خیر.

- کاربرد: برای اطمینان از قابل فهم بودن و پذیرش ابزار.

- ارزیابی: نظر سنجی کیفی از ۵-۱۰ کارشناس یا پاسخ‌دهنده.

۲. روایی محتوایی^۲

- تعریف: بررسی اینکه آیا سؤالات پرسشنامه کل دامنه سازه را پوشش می‌دهند یا خیر.

- کاربرد: برای اطمینان از جامع بودن ابزار.

- شاخص: شاخص روایی محتوا (CVI) و نسبت روایی محتوا (CVR).

این گونه از روایی برای بررسی اجزای تشکیل‌دهنده یک ابزار اندازه‌گیری استفاده می‌شود. روایی محتوای یک پرسشنامه به پرسش‌های تشکیل‌دهنده آن بستگی دارد. اگر پرسش‌ها بیانگر ویژگی‌ها و مهارت‌های ویژه‌ای باشند که پژوهشگر قصد اندازه‌گیری آن‌ها را دارد، ابزار از دید محتوا روایی دارد. روایی محتوا هم

¹ Face Validity

² Content Validity

۳. روایی سازه^۱

- تعریف: بررسی اینکه آیا ابزار واقعاً سازه نظری را می‌سنجد و با چارچوب نظری همخوانی دارد.

- کاربرد: برای ابزارهای روان‌سنجی، که باید اضطراب را از افسردگی متمایز کند.

این نوع از روایی بیان می‌کند که ابزار اندازه‌گیری تا چه اندازه ساختار یا ویژگی را که مبنای نظری دارد، می‌سنجد. در بررسی روایی ساختار، فرضیه‌هایی درباره مفهوم‌های اندازه‌گیری شده تدوین می‌شوند و همبستگی میان یافته‌ها و اندازه‌گیری‌های اولیه محاسبه می‌شود. از شیوه‌هایی که برای بررسی روایی ساختار به کار می‌روند، می‌توان به بررسی همبستگی میان پاسخ‌های دو گروه از پرسش‌ها که از دیدگاه نظری باید ارتباط زیادی با هم داشته باشند، و مقایسه گروه‌هایی از پاسخ‌گویان که به لحاظ نظری پاسخ‌های متفاوتی به پرسش‌ها ارائه می‌دهند، اشاره کرد. به‌عنوان نمونه، انتظار می‌رود در مناطق شهری با افزایش سطح تحصیلات، شیوع مصرف سیگار افزایش یابد؛ چنانچه داده‌های حاصل از پرسشنامه‌ای که برای اندازه‌گیری مصرف سیگار طرح شده، تفاوت معناداری را میان گروه‌های مصرف‌کنندگان با سطح تحصیلات متفاوت نشان دهد، این پرسشنامه دارای روایی ساختار خواهد بود.

روایی ساختار شامل زیرانواع روایی همگرا^۲ و روایی واگرا^۳ است^۴.

روایی همگرا: همبستگی بالا میان اندازه‌گیری‌ها از ابزارهای

روایی واگرا: همبستگی پایین با سازه‌های غیرمرتبط

۴. روایی ملاکی^۵

- تعریف: بررسی همبستگی ابزار با یک معیار استاندارد خارجی (معیار طلایی).

- کاربرد: برای تأیید دقت ابزار در مقایسه با ابزارهای معتبر.

¹ Construct Validity

² Convergent Validity

³ Divergent Validity

⁴ Allen & Benn, 1979

⁵ Criterion Validity

- انواع فرعی: همزمانی^۱: مقایسه با معیار در همان زمان.

پیش‌بینی^۲: پیش‌بینی نتایج آینده.

به طور خلاصه:

نوع روایی	تعریف	شاخص
صوری	ظاهر مرتبط با سازه	نظر کارشناسان
محتوایی	پوشش کامل سازه	CVR/CVI
سازه	همخوانی با نظریه	تحلیل عاملی
ملاکی	همبستگی با معیار	همبستگی پیرسون

روش های محاسبه روایی

برای ارزیابی روایی روش های زیر قابل اجرا هستند. این روش ها با نرم افزار هایی مثل SPSS، AMOS یا حتی Excel انجام می شوند:

۱. روایی صوری

- روش: نظرسنجی کیفی از ۵-۱۰ کارشناس (مثلاً روان‌شناسان) یا پاسخ دهندگان (مثلاً دانشجویان).

۲. روایی محتوایی

- شاخص نسبت روایی محتوا - CVR :

$$CVR = \frac{n_e - (\frac{N}{2})}{\frac{N}{2}}$$

¹ Concurrent

² Predictive

n_e : تعداد کارشناسانی که آیتم را ضروری می دانند.

N: کل کارشناسان.

آستانه: $0.6 < CVR$

- شاخص روایی محتوا - CVI :

میانگین امتیاز مرتبط بودن آیتها توسط نظر اساتید و کارشناسان.

آستانه: $0.8 < CVI$

۳. روایی سازه

- روش: تحلیل عاملی اکتشافی - EFA و تأییدی - CFA.

۴. روایی ملاکی

- روش: محاسبه همبستگی با ابزار استاندارد یا تشخیص بالینی.

پایایی

تعریف پایایی

پایایی به معنای ثبات و سازگاری^۱ نتایج یک ابزار اندازه گیری مانند پرسشنامه در شرایط مشابه است. در واقع پایایی، اطلاعاتی در مورد این موضوع فراهم می کند که شیوه یا ابزار گردآوری داده ها تا چه حد داده های دقیق و موثقی را استخراج می کند یا شیوه یا ابزار گردآوری داده ها تا چه حد درست و با ثبات است و نتیجه های سازگاری را به دست می دهد. به عبارت دیگر پایایی بیان می کند که اگر ابزار اندازه گیری در یک فاصله ی زمانی کوتاه چندین بار به گروه یکسانی از افراد داده شود، یافته ها تا چه حد نزدیک به هم هستند. طبق کرسول^۲، پایایی نشان دهنده میزان نبود خطای تصادفی در اندازه گیری است.

برای اندازه گیری پایایی از شاخصی به نام ضریب پایایی استفاده می شود. دامنه ی تغییر ضریب پایایی از صفر تا ۱ است. ضریب پایایی صفر، معرف عدم پایایی و ضریب پایایی یک، معرف پایایی کامل است. پایایی کامل به ندرت دیده می شود و در صورت مشاهده شدن قبل از هر چیز باید به یافته ها شک کرد.

¹ Consistency

² Creswell, 2020

اهمیت پایایی

پایایی، پایه اعتبار هر ابزار اندازه گیری است، زیرا بدون ثبات، نتایج قابل اعتماد نیستند و تحلیل های آماری (مانند رگرسیون یا تحلیل عاملی) بی معنی می شوند.

دلایل کلیدی اهمیت پایایی عبارتند از:

- کاهش خطای تصادفی: پایایی بالا نشان می دهد که ابزار از نویز (خطا های تصادفی مثل پاسخ های غیرصادقانه) مصون است.
- افزایش قابلیت تعمیم: ابزار پایا امکان مقایسه نتایج در جوامع مختلف را فراهم می کند
- پیش نیاز روایی: طبق استانداردهای COSMIN، ابزار بدون پایایی نمی تواند روا باشد، زیرا ثبات شرط لازم برای دقت است.
- اعتبار علمی: مقالات ISI و پایان نامه ها بدون گزارش پایایی معمولاً رد می شوند.

روش های محاسبه پایایی

برای محاسبه ضریب پایایی ابزار اندازه گیری، شیوه های مختلفی به کار برده می شوند. گروهی از این روش ها مبتنی بر تکرار اندازه گیری بوده و برخی با فنون ویژه ای بدون تکرار به ارزیابی پایایی می پردازند. از جمله این روش ها میتوان به موارد زیر اشاره کرد:

• روش آزمون بازآزمون^۱

این روش به صورت ارائه یک آزمون بیش از یک بار در یک گروه آزمودنی تحت شرایط یکسان تعریف می شود. برای محاسبه ضریب پایایی، ابتدا پرسشنامه روی یک گروه از پاسخ گویان اجرا می شود و سپس در فاصله زمانی کوتاهی (مثلاً ۲ هفته) دوباره در همان شرایط روی همان پاسخ گویان اجرا می شود. ضریب همبستگی میان امتیاز های کسب شده در دو بار اجرا، اندازه ای از پایایی ابزار است.

¹ Test-Retest Reliability

شاخص: ضریب همبستگی پیرسون یا ICC^۱، با آستانه < ۰,۷.

مزایا: بررسی ثبات زمانی ساده.

معایب: امکان بررسی اثر تجربه اولین پاسخ گویی بر پاسخ های دوم وجود ندارد (مثلاً یادگیری یا خستگی).

• روش آزمون هم ارزی^۲

این روش متداول برای ارزیابی پایایی در اندازه گیری های روانی و تربیتی است و به صورت دو آزمون متفاوت که شباهت های زیادی با یکدیگر دارند، در یک زمان یا با فاصله زمانی مناسب روی گروه یکسانی از پاسخ گویان اجرا می شود. ضریب همبستگی میان امتیاز های دو آزمون، به عنوان ضریب پایایی در نظر گرفته می شود.

شاخص: ضریب همبستگی پیرسون، با آستانه < ۰,۷.

مزایا: بررسی ثبات در ابزار های معادل.

معایب: تهیه دو نسخه معادل زمان بر است.

• روش پایایی نصف سازی (خردنیم)^۳

در این روش، پرسشنامه به دو قسمت تقسیم می شود (معمولاً سؤالات فرد و زوج) و مقایسه ای میان امتیاز های حاصل از دو بخش انجام می شود. این روش زمانی که به دلیل محدودیت های مالی و زمانی امکان اجرای دوباره آزمون وجود ندارد، پیشنهاد می شود. مراحل:

۱. پرسشنامه به دو قسمت تقسیم می شود (روش معمول: سؤالات فرد در یک گروه، زوج در دیگری؛ معروف به پایایی فرد-زوج).

۲. ضریب همبستگی پیرسون میان امتیاز های دو قسمت محاسبه می شود.

۳. مقدار حاصل با رابطه اسپیرمن-براون تعدیل می شود:

$$r' = \frac{2r}{1+r}$$

¹ Intraclass Correlation Coefficient

² Equivalent Forms or Parallel Forms Reliability

³ Split-Half Reliability

{که ۲ همان ضریب همبستگی پیرسون بین دو نیمه است که در گام دوم محاسبه می شود.}

مزایا: بدون نیاز به تکرار.

معایب: وابسته به نحوه تقسیم (ممکن است نیمه ها نابرابر باشند).

• روش کودر-ریچاردسون^۱

این روش برای محاسبه همبستگی درونی پرسش ها، به ویژه وقتی پاسخ ها دوگزینه ای (مثل بله / خیر) هستند، مناسب است. دو فرمول اصلی وجود دارد:

۱. کودر-ریچاردسون ۲۰^۲: میانگین همبستگی همه نیم های ممکن در روش نصف سازی است.

فرمول:

$$r_{KR-20} = \frac{k}{k-1} \times \frac{\sum_{i=1}^k p_i q_i}{s_T^2}$$

که k: تعداد پرسش ها، p_i : نسبت پاسخ گویان با پاسخ مثبت به پرسش، $q_i = 1 - p_i$: نسبت پاسخ گویان با پاسخ منفی به پرسش، s_T^2 : واریانس کل امتیاز ها.

۲. کودر-ریچاردسون ۲۱^۳: برای وقتی سطح دشواری پرسش ها یکسان است و تنها میانگین و واریانس کل موجود است.

فرمول:

$$r_{KR-21} = \frac{k}{k-1} \times \left[\frac{\bar{X}(1-\bar{X})}{k s_T^2} \right]$$

که k: تعداد پرسش ها، s_T^2 : واریانس، $\bar{X}$: میانگین کل امتیاز ها.

مزایا: مناسب برای سؤالات باینری.

معایب: محدود به پاسخ های دوگزینه ای.

¹ Kuder-Richardson Formulas

² KR-20

³ KR-21

• روش آلفای کرونباخ^۱

روایی یکی از ویژگی های کلیدی ابزار اندازه گیری است که بیان می کند ابزار تا چه حد مفهومی را که برای سنجش آن طراحی شده، به درستی اندازه گیری می کند. پایایی پرسشنامه معیاری است که تا حد زیادی به روایی آن وابسته است؛ به عبارت دیگر، یک ابزار نمی تواند روا (معتبر) باشد مگر اینکه پایا و قابل اعتماد نیز باشد. اندازه گیری عینی پایایی با استفاده از ضریب آلفای کرونباخ به دست می آید. استفاده از این آماره در مقایسه با سایر معیار های سنجش پایایی – مانند برآورد های آزمون بازآزمون یا آزمون هم ارزی که نیازمند تکرار اندازه گیری هستند – به مراتب ساده تر و کارآمدتر است.

کرونباخ (۱۹۵۱) آماره آلفا را به عنوان اندازه ای از سازگاری درونی^۲ پرسش های یک آزمون، پرسشنامه یا مقیاس اندازه گیری معرفی کرده است. در حقیقت، ضریب آلفا به صورت نسبت واریانس امتیاز واقعی^۳ یک مجموعه از پرسش ها به واریانس امتیاز مشاهده شده^۴ همه پرسش ها در میان همه پاسخ گویان تعریف می شود. واریانس مشاهده شده ترکیبی از امتیاز واقعی و میزان خطای اندازه گیری است. بنابراین، به لحاظ نظری، آلفا مقداری بین ۰ (عدم سازگاری) و ۱ (سازگاری کامل) اختیار می کند. این معیار بیان می کند که پرسش های موجود در یک پرسشنامه تا چه اندازه ساختار یا مفهوم یکسانی را می سنجند. در واقع، ضریب آلفا معیاری است برای بیان همبستگی درونی پرسش ها در یک پرسشنامه، و محاسبه آن پیش از اجرای اصلی پرسشنامه برای اطمینان از پایایی ضروری است.

در ادامه، برخی از روابط محاسباتی معمول برای ضریب آلفای کرونباخ ارائه می شود:

نخستین رابطه که برای محاسبه آلفا به کار می رود، به صورت زیر تعریف می شود.

$$\alpha = \frac{k}{k-1} \times \left(1 - \frac{\sum_{i=1}^k S_i^2}{S_T^2} \right)$$

که در آن:

k: تعداد پرسش های ابزار گردآوری اطلاعات (پرسشنامه).

$\sum_{i=1}^k S_i^2$: مجموع واریانس های مشاهده شده امتیاز هر پرسش i در میان نمونه جاری.

¹ Cronbach's Alpha

² Internal Consistency

³ VarianceTrue

⁴ Observed Variance

S_T^2 : واریانس کل امتیازهای کسب شده از کل پرسشنامه در میان همه پاسخ گوینان.

این فرمول بر اساس واریانس کل و واریانس هر پرسش به تنهایی محاسبه می‌شود و با استفاده از پاسخ های همه پاسخ گوینان در نمونه جاری به دست می‌آید.

در حالت هایی که ماتریس واریانس-کوواریانس^۱ پرسش ها موجود است، آلفا را می توان با رابطه زیر محاسبه کرد:

$$\alpha = \frac{k\bar{C}}{\bar{V} + (k-1)\bar{C}}$$

که در آن:

k: تعداد پرسش ها.

$\bar{V}$: میانگین واریانس های امتیازهای کسب شده از همه پرسش ها.

$\bar{C}$: میانگین کوواریانس های بین امتیازهای کسب شده از پرسش های پرسشنامه (از ماتریس واریانس-کوواریانس به راحتی قابل دسترسی است).

صورت دیگری از رابطه محاسباتی ضریب آلفا، که با نام آلفای کرونباخ استاندارد شده^۲ شناخته می شود، به صورت زیر است. استفاده از این شکل از ضریب آلفا زمانی پیشنهاد می‌شود که ماتریس ضریب همبستگی^۳ میان پاسخ های ارائه شده به پرسش ها محاسبه شده باشد :

$$\alpha_{std.} = \frac{k\bar{r}}{1 + (k-1)\bar{r}}$$

که در آن:

K: تعداد پرسش ها.

$\bar{r}$: میانگین ضریب همبستگی های پیروان میان همه پرسش های پرسشنامه (به صورت میانگین عناصر غیرتکراری مثلث بالایی و پایینی ماتریس ضریب همبستگی تعریف می‌شود).

¹ Variance-Covariance Matrix

² Standardized Cronbach's Alpha

³ Correlation Matrix

در حقیقت، $\bar{r}$ به عنوان میانگین همبستگی درونی همه پرسش ها عمل می کند. زمانی که همه پرسش ها برای اندازه گیری یک مفهوم واحد طراحی شده باشند، همبستگی درونی آن ها افزایش می یابد و در نتیجه مقدار ضریب آلفا نیز بزرگ تر می شود. بنابراین، آلفا به طور غیرمستقیم تعیین می کند که یک مجموعه از پرسش ها تا چه اندازه ساختار یا مفهوم یکسانی را اندازه گیری می کنند.

عوامل مؤثر بر ضریب آلفا و محدودیت ها

ضریب آلفا به عوامل مختلفی وابسته است و تفسیر آن نیازمند توجه به این عوامل است:

- تعداد پرسش ها:

تعداد بیشتر پرسش ها (k) معمولاً آلفا را افزایش می دهد، حتی اگر همبستگی درونی ضعیف باشد. بنابراین، مقدار بزرگ آلفا همیشه بیانگر همبستگی درونی بالا نیست؛ باید تعداد آیتم ها را در نظر گرفت.

- همبستگی درونی:

اگر پرسش ها برای سنجش یک مفهوم طراحی شده باشند، همبستگی درونی بیشتر و آلفا بالاتر خواهد بود.

- وابستگی به نمونه:

آلفا به نمونه مورد بررسی حساس است و مقدار آن از نمونه ای به نمونه دیگر تغییر می کند. این عدم پایداری نسبت به مقادیر گم شده^۱ یک ویژگی منفی آن به شمار می رود.

- پرسشنامه های چند بعدی: در پرسشنامه هایی که برای بررسی بیش از یک هدف طراحی شده اند، محاسبه ضریب آلفا برای هر زیرمجموعه از پرسش ها^۲ که برای گردآوری اطلاعات درباره همان هدف تهیه شده، به صورت جداگانه ضروری است.

ضریب آلفای کرونباخ در بیشتر موارد به عنوان معیاری از سازگاری درونی و قابلیت اطمینان پاسخ های داده شده به پرسش های یک پرسشنامه به کار می رود. نویسندگان گوناگون مقدارهای متفاوتی را برای ضریب آلفا مطلوب در نظر گرفته اند. جورج و مالری (۲۰۰۳) مقدارهای ممکن برای ضریب آلفا را گروه بندی کرده و تفسیر های زیر را با توجه به دامنه ای که مقدار آلفای استاندارد شده در آن قرار می گیرد، ارائه داده اند:

¹ Missing Values

² Subscale

***جدول تفسیر مقدار آلفای کرونباخ استاندارد شده^۱ ***

پایایی پرسشنامه عالی است	$\alpha_{std.} > 0.9$
پایایی پرسشنامه خوب است	$\alpha_{std.} > 0.8$
پایایی پرسشنامه قابل قبول است	$\alpha_{std.} > 0.7$
پایایی پرسشنامه پرسش برانگیز است	$\alpha_{std.} > 0.6$
پایایی پرسشنامه ضعیف است	$\alpha_{std.} > 0.5$

علاوه بر این، مقدار ضریب آلفای مطلوب برای کاربرد های بالینی مقداری بزرگ تر از ۰,۹ است، و از مقدار بیشتر از ۰,۹۵ با عنوان شرایط ایده آل یاد می شود^۲.

در حالت کلی، می توان نتیجه گرفت که مقدار آلفای کرونباخ بین ۰,۷ و ۰,۹۵ بیانگر این است که ابزار اندازه گیری (پرسشنامه) با پایایی مناسبی اطلاعات را گردآوری کرده است.

دلایل کاهش ضریب آلفا و راهکارهای بهبود

مقدار کوچک ضریب آلفا می تواند ناشی از کم بودن تعداد پرسش ها، همبستگی درونی ضعیف میان پرسش ها یا عدم یک بعدی بودن^۳ ابزار اندازه گیری باشد. چنانچه همبستگی ضعیف میان پرسش ها سبب کاهش مقدار آلفا شده باشد، کنار گذاشتن برخی پرسش ها روش مناسبی است. به علاوه، می توان با بررسی بیشتر و بخش بندی پرسشنامه، پرسش هایی را که همبستگی درونی بیشتری با یکدیگر دارند، در یک گروه قرار داد.

نخستین گام برای محاسبه ضریب آلفا و ارزیابی پایایی، محاسبه واریانس و تغییرپذیری پرسش ها به تنهایی و واریانس مجموعه پاسخ های داده شده به پرسشنامه است. با استفاده از این اطلاعات، ساده ترین راه برای یافتن پرسش های نامزد حذف، محاسبه ضریب همبستگی میان امتیاز هر پرسش در مجموعه پاسخ ها و

¹ George & Mallery, 2003

² Bland & Altman, ۱۹۹۷

³ Uni-Dimensionality

امتیاز کل پرسشنامه است. پرسش‌هایی با همبستگی نزدیک به صفر حذف می‌شوند. همچنین، بهتر است پرسش‌هایی که ضریب همبستگی منفی با دیگر پرسش‌ها دارند، از مجموعه کل پرسشنامه جدا شده و در بخش دیگری از پرسشنامه قرار گیرند. به این ترتیب، مقدار ضریب آلفا و در نتیجه میزان پایایی پرسشنامه افزایش می‌یابد^۱.

در نرم‌افزارهایی مانند SPSS، تحلیل "Scale if Item Deleted" این کار را به صورت خودکار انجام می‌دهد و نشان می‌دهد حذف کدام آیتم آلفا را بهبود می‌بخشد.

هشدار در مورد تفسیر نادرست

در انتها، بیان این نکته ضروری است که استفاده و تفسیر نادرست ضریب آلفا می‌تواند منجر به انتخاب و به کارگیری یک ابزار نامناسب شود که یافته‌های غیرقابل اعتمادی را نتیجه می‌دهد. رعایت پیشنهادهایی که در خصوص طراحی و انتخاب تک‌تک پرسش‌ها بیان شده (مانند وضوح و مرتبط بودن) و همچنین توجه به نکات ارائه شده درباره چگونگی طراحی پرسشنامه – به عنوان متداول‌ترین ابزار گردآوری اطلاعات – منجر به طراحی یک پرسشنامه استاندارد مبتنی بر اهداف پژوهش خواهد شد.

ارتباط میان روایی و پایایی

مفهوم ارتباط روایی و پایایی

طبق تعاریف گفته شده، روایی به دقت یک ابزار اندازه‌گیری در سنجش سازه موردنظر اشاره دارد، در حالی که پایایی به ثبات و سازگاری نتایج ابزار در شرایط مشابه مربوط است. این دو مفهوم به هم وابسته اند، اما مستقل از یکدیگرند و هر دو برای اعتبار یک پرسشنامه ضروری اند. ارتباط بین آن‌ها را می‌توان به صورت زیر خلاصه کرد:

¹ Tavakol & Dennick, ۲۰۱۱

- پایایی پیش نیاز روایی است:

ابزار بدون پایایی نمی تواند روا باشد، زیرا اگر نتایج ابزار ناپایدار و پر از خطای تصادفی باشد، نمی توان مطمئن بود که سازه موردنظر را به درستی می سنجد. طبق استاندارد های COSMIN، پایایی شرط لازم برای روایی است، اما کافی نیست.

- روایی بدون پایایی ممکن نیست، اما پایایی بدون روایی ممکن است:

ابزاری می تواند پایا باشد (نتایج ثابتی بدهد) اما روا نباشد (سازه غلطی را بسنجد).

- تأثیر متقابل:

بهبود پایایی می تواند روایی را تقویت کند، زیرا نتایج دقیق تر می شوند. برعکس، اگر روایی محتوایی^۱ بالا باشد، سؤالات مرتبط تر احتمالاً همسانی درونی بهتری (پایایی) دارند.

¹ CVI

فصل سوم: روش شناسی پژوهش

طرح پژوهش

طرح پژوهش به عنوان نقشه راه کلی مطالعه، ساختار اجرایی، تحلیل و تفسیر یافته ها را مشخص می کند و نقش کلیدی در تضمین اعتبار داخلی و خارجی نتایج ایفا می کند. این پژوهش با هدف ارزیابی پایایی و روایی پرسشنامه GAD-7^۱ در یک نمونه از دانشجویان پزشکی مکزیکی انجام می شود.

داده های مورد استفاده، از یک مجموعه داده عمومی استخراج شده که شامل اطلاعات جامع از ۷۷۶ دانشجوی پزشکی است. این داده ها در سال ۲۰۱۴ جمع آوری شده و نه تنها شامل امتیازات GAD-7 می شود، بلکه متغیرهای مرتبط با سلامت روان، خواب و عوامل دموگرافیک را نیز در بر می گیرد. طرح پژوهش بر اساس رویکرد کمی^۲ طراحی شده است، جایی که تمرکز اصلی بر تحلیل آماری داده های موجود برای بررسی ساختار درونی پرسشنامه، ثبات آن و همخوانی با سازه های مرتبط است.

در این طرح، ابتدا داده های کلی مجموعه بررسی می شود تا زمینه ای جامع از جامعه مورد مطالعه فراهم آید. مجموعه داده شامل ۳۷ متغیر است که از جنبه های متنوعی مانند ویژگی های دموگرافیک (مانند سن، جنسیت، قد و وزن)، علائم افسردگی (از طریق پرسشنامه PHQ-9)، علائم اضطراب (از طریق GAD-7)، کیفیت خواب (از طریق مقیاس ESS^۳) و عوامل تحصیلی (مانند ساعات مطالعه هفتگی و معدل تحصیلی) تشکیل شده است. این تنوع متغیرها امکان تحلیل های چند بعدی را فراهم می کند، اما تمرکز تخصصی پژوهش بر داده های مرتبط با GAD-7 است که شامل هفت آیتم امتیازدهی شده (gad1 تا gad7) می شود. این آیتم ها علائم اضطراب عمومی را در دو هفته اخیر ارزیابی می کنند و امتیاز کل آنها از ۰ تا ۲۱ متغیر است.

انتخاب GAD-7 به دلیل سادگی، حساسیت بالا و کاربرد گسترده در غربالگری اضطراب (مانند مطالعات Household Pulse Survey) است، که شیوع اضطراب را در جمعیت های بزرگ نشان می دهد.

نوع پژوهش توصیفی - تحلیلی و مقطعی است، که بر توصیف ویژگی های روان سنجی GAD-7 و تحلیل روابط آماری تمرکز دارد. جنبه توصیفی شامل بررسی آمارهای پایه و جنبه تحلیلی شامل آزمون فرضیات روان سنجی است. رویکرد پژوهش کمی غالب است، با تأکید بر آمار توصیفی و استنباطی. این طرح بر مدل COSMIN^۴ تکیه دارد، که شامل مراحل ارزیابی محتوا، سازه، ملاکی و قابلیت پاسخ دهی است و برای

¹ Generalized Anxiety Disorder-7

² quantitative

³ Epworth Sleepiness Scale

⁴ Consensus-based Standards for the selection of health Measurement INstruments

ابزارهای خود گزارشی مانند GAD-7 ایده آل است. همچنین، از رویکرد داده های ثانویه^۱ استفاده شده، که شامل استخراج و تحلیل دیتاست Mental health data of Mexican medical students از figshare است.

دلیل انتخاب طرح

اضطراب عمومی یک مسئله جهانی است و GAD-7 نیاز به ارزیابی روان سنجی در داده های بزرگ دارد. طرح توصیفی – تحلیلی اجازه می دهد تا بدون مداخله، ویژگی های GAD-7 (مانند ثبات زمانی در دیتاست HPS) تحلیل شود، در حالی که نتایج را می توان به زمینه های ایرانی تعمیم داد.

جامعه آماری، نمونه و روش نمونه گیری

جامعه آماری به عنوان مجموعه کامل عناصر مورد مطالعه، پایه تعمیم پذیری و اعتبار خارجی پژوهش را تشکیل می دهد و در طراحی نمونه گیری نقش محوری دارد. در این پژوهش، جامعه آماری شامل دانشجویان پزشکی یک دانشگاه خصوصی در مکزیک است که در سال ۲۰۱۴ تعداد کل آن ها ۴۵۲۰ نفر بوده است. این جامعه به عنوان نماینده ای از دانشجویان پزشکی در کشورهای در حال توسعه انتخاب شده، زیرا دانشجویان پزشکی اغلب با سطوح بالایی از استرس و اضطراب روبرو هستند که می تواند زمینه مناسبی برای ارزیابی ابزارهای غربالگری اضطراب مانند GAD-7 فراهم کند. جامعه آماری از تنوع سنی، جنسیتی و تحصیلی برخوردار است و شامل دانشجویان سال های اول تا چهارم می شود.

نمونه پژوهش شامل ۷۷۶ دانشجو است که از طریق روش نمونه گیری تصادفی ساده^۲ انتخاب شده اند. در فرآیند جمع آوری داده های اصلی، ابتدا ۱۲۰۰ دانشجو به صورت تصادفی از کل جامعه ۴۵۲۰ نفری انتخاب شدند و سپس ۷۷۶ نفر (با نرخ پاسخ دهی ۶۴٫۶٪) در مطالعه شرکت کردند. این حجم نمونه بر اساس معیارهای آماری برای تحلیل های روانسنجی کافی است، زیرا برای محاسبه پایایی داخلی و تحلیل عاملی، حداقل حجم نمونه پیشنهادی ۵ تا ۱۰ برابر تعداد آیتم های پرسشنامه (یعنی حداقل ۳۵ تا ۷۰ نفر برای GAD-7) است، اما حجم ۷۷۶ نفر قدرت آماری بالایی فراهم می کند و خطای نمونه گیری را کاهش می دهد.

روش نمونه گیری تصادفی ساده تضمین کننده عدم بایاس و نماینده بودن نمونه است، هرچند ممکن است عواملی مانند نرخ پاسخ دهی پایین تر از ۱۰۰٪ منجر به برخی محدودیت ها شود (مانند احتمال بیش نمایی

¹ secondary data

² simple random sampling

دانشجویان علاقه مند به سلامت روان). در تمرکز تخصصی بر داده های GAD-7، نمونه شامل امتیازات خام هفت آیتم برای هر شرکت کننده است که توزیع آن ها می تواند برای بررسی ساختار عاملی و پایایی داخلی استفاده شود. هیچ زیرنمونه گیری اضافی در این پژوهش اعمال نشده و کل نمونه برای تحلیل های روایی و پایایی به کار گرفته می شود. این رویکرد با استاندارد های روش شناسی پژوهش های روان سنجی همخوانی دارد و امکان تعمیم پذیری نتایج به جوامع مشابه دانشجویان پزشکی را افزایش می دهد.

ویژگی های جامعه

- حجم جامعه : ۴۵۲۰ نفر
 - محدوده سنی: عمدتاً بین ۱۸ تا ۲۵ سال، که نشان دهنده جمعیت جوان دانشجویان پزشکی است.
 - جنسیت: شامل هر دو جنس مرد و زن
 - وضعیت تحصیلی: دانشجویان در مقطع کارشناسی پزشکی، از سال اول تا سال چهارم تحصیل، که نشان دهنده تنوع در تجربه های آموزشی و بالینی است.
 - محل جغرافیایی: محدود به یک دانشگاه خصوصی در مکزیک، که ممکن است ویژگی های اقتصادی-اجتماعی خاصی (مانند دسترسی به منابع آموزشی بهتر) را منعکس کند.
 - شرایط روانی و اجتماعی: جامعه احتمالاً شامل افرادی با سطوح متفاوتی از سلامت روان است، از جمله کسانی که علائم اضطراب یا افسردگی دارند، به دلیل فشارهای خاص حرفه پزشکی.
 - فرهنگ و زبان: جامعه به زبان اسپانیایی زندگی می کنند و تحت تأثیر فرهنگ مکزیکی هستند، که می تواند بر پاسخ دهی به پرسشنامه ها اثر بگذارد.
- این ویژگی ها نشان می دهند که جامعه آماری از تنوع نسبی برخوردار است، اما محدودیت هایی مانند تمرکز بر یک دانشگاه خاص و عدم تنوع جغرافیایی گسترده ممکن است بر تعمیم پذیری نتایج تأثیر بگذارد.

ویژگی های نمونه

- حجم نمونه: ۷۷۶ نفر
- محدوده سنی: بین ۱۸ تا ۲۵ سال، با میانگین سنی تقریبی ۲۱-۲۲ سال
- جنسیت: شامل هر دو جنس زن و مرد، ۵۱/۵۰٪ زنان و ۴۸/۵۰٪ مردان

- سال تحصیلی: توزیع در بین سال های اول تا چهارم (سال اول- ۲۷۳ نفر، سال دوم- ۱۸۱ نفر، سال سوم- ۱۴۰ نفر، سال چهارم- ۱۷۵ نفر)

- وضعیت سلامت روان: نمونه شامل افرادی با طیف وسیعی از امتیازات GAD-7 ۰ تا ۲۱ است، که نشان دهنده وجود حداقل، خفیف، متوسط و شدید اضطراب است. همچنین شامل داده های PHQ-9 و ESS است که نشان دهنده هم پوشانی با افسردگی و مشکلات خواب است.

- ویژگی های فیزیکی: شامل قد (به متر) و وزن (به کیلوگرم) که می تواند برای تحلیل های جانبی (مثلاً ارتباط با استرس) استفاده شود. (میانگین قد: ۱/۶۹، میانگین وزن: ۶۷/۶۷)

- عادات خواب و تحصیلی: میانگین ساعات خواب شبانه^۱ و ساعات مطالعه هفتگی^۲ نشان دهنده تنوع در سبک زندگی است که ممکن است بر اضطراب اثر بگذارد.

- توزیع جغرافیایی و فرهنگی: محدود به یک دانشگاه، اما نماینده دانشجویان مکزیکی با پاسخ دهی به پرسشنامه به زبان اسپانیایی.

ویژگی های نمونه نشان دهنده بودن نسبی نسبت به جامعه است، هرچند نرخ پاسخ دهی ۶۴/۶٪ ممکن است به معنای وجود بایاس انتخابی باشد (مثلاً افرادی که به سلامت روان علاقه مندند بیشتر شرکت کرده اند). با این حال، تنوع در متغیرها و حجم نمونه، قابلیت اطمینان تحلیل ها را افزایش می دهد. این ویژگی ها به ویژه برای داده های GAD-7 مهم هستند، زیرا امکان بررسی پایایی و روایی را بر اساس توزیع واقعی امتیازات فراهم می کنند.

معرفی و توصیف پرسشنامه

پرسشنامه GAD-7^۳ یک ابزار غربالگری کوتاه و معتبر برای ارزیابی اختلال اضطراب عمومی^۴ است که توسط Spitzer و همکاران در سال ۲۰۰۶ توسعه یافته است.

¹ reported_sleep_hours

² weekly_study_hours

³ Generalized Anxiety Disorder 7-item

⁴ Generalized Anxiety Disorder

- موضوع: سنجش شدت علائم اضطراب عمومی در دو هفته گذشته
 - تعداد سؤالات: ۷ سؤال
 - مقیاس: لیکرت ۴ درجه‌ای (۰ = اصلاً، ۱ = چند روز، ۲ = بیش از نصف روزها، ۳ = تقریباً هر روز).
 - نمره‌گذاری کل: جمع امتیازات ۷ سؤال (محدوده ۰-۲۱). تفسیر: ۰-۴ (بدون اضطراب)، ۵-۹ (خفیف)، ۱۰-۱۴ (متوسط)، ۱۵-۲۱ (شدید).
 - کاربرد: مناسب برای روانشناسی، بهداشت عمومی و حتی آموزش (مثل اضطراب دانشجویان).
- این پرسشنامه بخشی از مجموعه داده بزرگ تری است که شامل PHQ-9 (برای افسردگی) و ESS (برای خواب آلودگی) می‌شود، اما تمرکز این پژوهش بر GAD-7 است، تا پایایی داخلی و ساختار عاملی آن در نمونه دانشجویان پزشکی مکزیک بررسی شود.

علت انتخاب سنجش شدت اضطراب عمومی

اضطراب عمومی یکی از شایع‌ترین اختلالات روانی است که می‌تواند تأثیرات قابل توجهی بر کیفیت زندگی فرد داشته باشد. با توجه به تغییرات سریع اجتماعی، اقتصادی و محیطی در دنیای معاصر، میزان فشارهای روانی و اضطراب در افراد افزایش یافته است. این موضوع به ویژه در دو سال اخیر و در پی شیوع بیماری‌های جهانی و بحران‌های اجتماعی و اقتصادی، بیشتر به چشم می‌خورد. به همین علت از نظر بنده این موضوع پتانسیل بالایی برای انجام پژوهش و بررسی داشت.

اهمیت سنجش اضطراب عمومی

۱. تشخیص زود هنگام: سنجش شدت علائم اضطراب عمومی می‌تواند به شناسایی زود هنگام این اختلال کمک کند و افراد را به دریافت درمان مناسب تشویق نماید.
۲. بهبود کیفیت زندگی: با شناسایی و مدیریت اضطراب، افراد می‌توانند به بهبود کیفیت زندگی خود دست یابند و از عملکرد بهتری در زمینه‌های مختلف زندگی بهره‌مند شوند.
۳. آگاهی بخشی: این پژوهش می‌تواند به افزایش آگاهی عمومی درباره علائم و عواقب اضطراب عمومی کمک کند و افراد را به سمت پذیرش و درمان این اختلال سوق دهد.
۴. تحقیق و توسعه: نتایج این پرسشنامه می‌تواند به محققان و متخصصان سلامت روان کمک کند تا روش‌های بهتری برای درمان و مدیریت اضطراب عمومی توسعه دهند.

در نهایت، سنجش شدت علائم اضطراب عمومی نه تنها به فرد کمک می‌کند تا با چالش‌های روانی خود بهتر مقابله کند، بلکه می‌تواند به جامعه نیز در کاهش بار روانی ناشی از این اختلال کمک کند. این موضوع به وضوح نشان دهنده اهمیت و ضرورت انجام پژوهش‌های بیشتر در این زمینه است.

پرسشنامه

عنوان : Generalized Anxiety Disorder 7 (GAD-7) - اختلال اضطراب فراگیر ۷

GAD-7				
Over the <u>last 2 weeks</u> , how often have you been bothered by the following problems? (Use "✓" to indicate your answer)	Not at all	Several days	More than half the days	Nearly every day
1. Feeling nervous, anxious or on edge	0	1	2	3
2. Not being able to stop or control worrying	0	1	2	3
3. Worrying too much about different things	0	1	2	3
4. Trouble relaxing	0	1	2	3
5. Being so restless that it is hard to sit still	0	1	2	3
6. Becoming easily annoyed or irritable	0	1	2	3
7. Feeling afraid as if something awful might happen	0	1	2	3

منبع پرسشنامه:

پرسشنامه GAD-7 ابزاری کوتاه برای غربالگری اختلال اضطراب عمومی و اندازه‌گیری شدت آن است. این پرسشنامه توسط دکتران Robert L. Spitzer، Kurt Kroenke، Janet B.W. Williams و همکارانشان توسعه یافته است.

منبع دیتاست:

داده‌های سلامت روان ۷۷۶ دانشجوی پزشکی مکزیکی در سایت figshare

ترجمه :

در دو هفته گذشته، هر یک از مشکلات زیر تا چه حد شما را آزار داده است؟ ۰ (اصلاً)، ۱ (چند روز)، ۲ (بیش از نصف روزها)، ۳ (تقریباً هر روز)

شماره	متن فارسی	متن انگلیسی	نمره گذاری
۱	احساس عصبی بودن، اضطراب یا بی‌قراری	Feeling nervous, anxious, or on edge	۳-۰
۲	ناتوانی در توقف یا کنترل نگرانی	Not being able to stop or control worrying	۳-۰
۳	نگرانی بیش از حد درباره چیزهای مختلف	Worrying too much about different things	۳-۰
۴	مشکل در آرام شدن	Trouble relaxing	۳-۰
۵	بی‌قراری شدید که نشستن را دشوار کند	Being so restless that it is hard to sit still	۳-۰
۶	به راحتی عصبانی یا تحریک پذیر شدن	Becoming easily annoyed or irritable	۳-۰
۷	احساس ترس، مثل اینکه ممکن است اتفاق بدی بیفتد	Feeling afraid, as if something awful might happen	۳-۰

روش های بررسی روایی در این پژوهش

روایی به عنوان یکی از جنبه های کلیدی ارزیابی ابزارهای اندازه گیری، دقت پرسشنامه GAD-7 در سنجش سازه اضطراب عمومی را بررسی می کند و در این پژوهش با روش های ترکیبی کیفی و کمی بر اساس دیتاست Mental health data of 776 Mexican medical students اجرا می شود. روش های روایی به چهار دسته اصلی تقسیم می شوند: صوری، محتوایی، سازه و ملاکی، که هر کدام با تمرکز بر ویژگی های دیتاست (مانند توزیع نمرات gad1) و نرم افزار SPSS اجرا می شوند تا روایی در گروه دانشجویان تأیید شود.

روش بررسی روایی صوری

این روش بر ارزیابی ظاهری و قابل فهم بودن آیتم های GAD-7 تمرکز دارد. در این پژوهش، روایی صوری با نظرسنجی از ۱۰ کارشناس سلامت روان (مانند روان شناسان بالینی) انجام می شود، که هر آیتم (gad1 تا gad7) را بر مقیاس ۵ درجه ای امتیاز می دهند (۱= نامرتبط و غیرقابل فهم، ۵= بسیار مرتبط و واضح). میانگین امتیاز هر آیتم محاسبه می شود و آستانه توافق ۸۰٪ (میانگین < ۴) برای پذیرش تعیین می شود.

اگر روایی صوری پایین باشد، نشان دهنده ابهام فرهنگی در گروه دانشجویان پزشکی (با استرس تحصیلی بالا) است و پیشنهاد ساده سازی آیتم ارائه می شود.

در این پژوهش، به دلیل محدودیت دسترسی به کارشناسان سلامت روان، روایی صوری انجام نمی شود.

روش بررسی روایی محتوایی

این روش برای اطمینان از پوشش کامل دامنه اضطراب عمومی (مانند نگرانی و بی قراری) استفاده می شود. در این پژوهش، روایی محتوایی با دو شاخص CVR و CVI محاسبه می شود.

- CVR (نسبت روایی محتوا): ۱۰ متخصص هر سؤال رو به عنوان "ضروری"، "مفید" یا "غیرضروری" مشخص می کنند. با فرمول لوشه، CVR باید بالای ۰/۶ باشد (حداقل ۸ متخصص تشخیص دهند ضروری است).

با فرمول :

$$CVR = \frac{n_e - (\frac{N}{2})}{\frac{N}{2}}$$

- CVI (شاخص روایی محتوا): میانگین امتیاز مرتبط بودن سؤال ها (مقیاس ۱-۴) باید بالای ۰/۸ باشد.

در SPSS، همبستگی بین سؤال ها (gad1-gad7) چک می شود تا پوشش محتوایی (۰,۴ > r بین آیتم ها) تأیید شود. برای مثال، gad2 ("ناتوانی در کنترل نگرانی") و gad3 ("نگرانی بیش از حد") باید همبستگی قوی داشته باشند. اگر CVR یک سؤال (مثل gad7) زیر ۰/۶ باشد، نشان دهنده عدم پوشش کامل آن از اضطراب است.

روش بررسی روایی سازه

این روش برای تأیید هم خوانی GAD-7 با نظریه اضطراب (مدل تک بعدی) استفاده می شود. در این پژوهش، روایی سازه با EFA و CFA ارزیابی می شود.

- تحلیل عاملی اکتشافی (EFA): در SPSS با روش Principal Component Analysis و چرخش Varimax، چک می کنیم که آیا ۷ سؤال پرسشنامه، به یک عامل (اضطراب) مربوط می شوند.

شاخص‌ها :

- KMO بالای ۰,۶ (انتظار ~ ۰,۸ برای ۷۷۶ رکورد)
- آزمون بارتلت معنادار با $p < ۰,۰۵$
- واریانس توضیحی بالای ۵۰٪ (انتظار ~ ۶۰٪)
- بار عاملی هر سؤال (gad1-gad7) بالای ۰,۴ EFA .
- تحلیل عاملی تأییدی (CFA): در AMOS ، مدل تک بعدی (همه سؤال‌ها به یک عامل اضطراب) تست می‌شود.

شاخص‌ها :

- CFI بالای ۰,۹ (فیت خوب)
- RMSEA زیر ۰,۰۸ (خطای کم) برای صرفه‌جویی، CFA روی ۳۰۰ رکورد تصادفی اجرا می‌شود. همچنین، روایی همگرا (همبستگی قوی با PHQ-9، $r \sim ۰,۷$) و واگرا (همبستگی ضعیف با Epworth، $r < ۰,۴$) در SPSS چک می‌شود.

روش بررسی روایی ملاکی

این روش برای مقایسه GAD-7 با معیارهای خارجی (مانند PHQ-9 برای افسردگی همبود) استفاده می‌شود. در SPSS :

- روایی همزمانی : همبستگی پیرسون بین نمره کل GAD-7 و PHQ-9 (افسردگی) محاسبه می‌شود. ($r > ۰,۶$)
 - روایی پیش بینی : رگرسیون خطی برای پیش بینی نمره GAD-7 بر اساس متغیرهایی مثل `previous_depression` یا `thoughts_of_dropping_out` اجرا می‌شود. ($R^2 > ۰,۳$)
- انتظار داریم GAD-7 با PHQ-9 همبستگی قوی (~۰,۷) داشته باشد، چون اضطراب و افسردگی همبودی دارند.

روش های بررسی پایایی در این پژوهش

پایایی به میزان ثبات و سازگاری پرسشنامه GAD-7 در اندازه گیری اضطراب عمومی اشاره دارد و نشان می دهد که آیا نتایج حاصل از این ابزار در شرایط مختلف تکرارپذیر و قابل اعتماد هستند یا خیر. در این پژوهش، پایایی GAD-7 با استفاده از روش های آلفای کرونباخ، آزمون-بازآزمون، هم ارزی، نصف سازی و کودر-ریچاردسون بر روی دیتاست شامل ۷۷۶ رکورد از دانشجویان پزشکی مکزیکی ارزیابی می شود. این دیتاست، که نمرات ۷ آیت GAD-7 (gad1 تا gad7)، در مقیاس لیکرت ۴ درجه ای (۰ تا ۳) ارائه می دهد، به دلیل حجم مناسب (۷۷۶ مورد) و داده های خام با مقادیر گمشده کم (کمتر از ۵٪)، برای تحلیل های روان سنجی ایده آل است. تحلیل ها با نرم افزار SPSS انجام می شوند، و روش ها بر اساس استاندارد های COSMIN و ویژگی های دیتاست مقطعی انتخاب شده اند. در ادامه، هر روش توضیح داده شده و کاربرد آن برای GAD-7 در این دیتاست بررسی می شود.

- روش آلفای کرونباخ

این روش همسانی درونی آیت های GAD-7 را بررسی می کند و نشان می دهد که آیا ۷ سؤال پرسشنامه به صورت منسجم اضطراب عمومی را اندازه گیری می کنند. آلفای کرونباخ با استفاده از SPSS محاسبه می شود و باید بالاتر از ۰/۷ باشد تا پایایی قابل قبولی تأیید شود. در این روش، همبستگی بین آیت ها (مانند gad2 و gad3) بررسی می شود و گزینه "Scale if Item Deleted" نشان می دهد که حذف هر آیت چه تأثیری بر آلفای کل دارد. با توجه به مطالعات پیشین (Spitzer et al., 2006) GAD-7، انتظار می رود آلفای کرونباخ در این دیتاست بین ۰/۸۵ تا ۰/۹ باشد، که نشان دهنده پایایی بالای پرسشنامه در میان دانشجویان پزشکی با نرخ اضطراب تقریبی ۳۰-۲۵٪ است. اگر آلفا کمتر از ۰/۷ باشد، آیت های با همبستگی ضعیف مانند (gad7) برای اصلاح شناسایی می شوند. این روش به دلیل سادگی و مناسب بودن برای مقیاس های لیکرت، برای دیتاست ۷۷۶ رکورد کاملاً مناسب است.

- روش نصف سازی

روش نصف سازی پایایی پرسشنامه را با تقسیم آیت ها به دو نیمه (gad1 تا gad4 و gad5 تا gad7) و بررسی همبستگی بین آن ها ارزیابی می کند. در SPSS، ضریب همبستگی پیرسون بین دو نیمه محاسبه شده و با فرمول Spearman-Brown تعدیل می شود.

ضریب تعدیل شده باید بالای ۰/۷ باشد تا پایایی تأیید شود. این روش برای دیتاست مناسب است، زیرا نیازی به داده های دو نوبت ندارد و با حجم ۷۷۶ رکورد، نتایج پایداری ارائه می دهد. انتظار می رود ضریب نصف

سازی حدود ۰/۸ باشد، که نشان دهنده ثبات درونی GAD-7 است. این روش به ویژه برای تأیید پایایی در گروه های پراسترس مانند دانشجویان پزشکی مفید است.

- روش آزمون - بازآزمون

این روش ثبات پاسخ های GAD-7 را در بازه های زمانی مختلف بررسی می کند. از آنجا که دیتاست مقطعی است و داده های طولی (دو نوبت) ندارد، این روش با شبیه سازی بر اساس زیرگروه های دیتاست اجرا می شود. برای مثال، پاسخ های دانشجویان در ترم های ابتدایی (۱-۴) با ترم های بعدی (۵-۸) مقایسه می شود تا ثبات تقریبی بررسی شود. ضریب همبستگی درون طبقه ای (ICC) با مدل دوطرفه مخلوط^۱ در SPSS محاسبه می شود و باید بالای ۰/۷ باشد. انتظار می رود ICC حدود ۰/۷۵-۰/۸۵ باشد، اما اگر کمتر از ۰/۷ باشد، ممکن است به دلیل تأثیر عوامل محیطی (مانند استرس امتحانات در ترم های مختلف) باشد. این روش به دلیل مقطعی بودن دیتاست محدودیت هایی دارد، اما شبیه سازی با زیرگروه های semester امکان بررسی نسبی ثبات زمانی را فراهم می کند.

- روش هم ارزی

روش هم ارزی پایایی GAD-7 را با مقایسه آن با فرم موازی یا ابزار مشابه (مانند مقیاس اضطراب دیگر) بررسی می کند. در این دیتاست، به دلیل نبود فرم موازی GAD-7، این روش با مقایسه همبستگی GAD-7 با PHQ-9 برای افسردگی، که همبودی بالایی با اضطراب دارد، شبیه سازی می شود. در SPSS، همبستگی پیرسون بین نمره کل GAD-7 و PHQ-9 محاسبه می شود و انتظار می رود حدود ۰/۶-۰/۷ باشد، که نشان دهنده هم ارزی نسبی است. این روش به دلیل نبود فرم موازی مستقیم، به صورت محدود اجرا می شود و بیشتر برای تأیید همبستگی با ابزارهای مرتبط استفاده می شود.

- روش کودر-ریچاردسون

روش کودر-ریچاردسون (KR-20 یا KR-21) برای مقیاس های دوگانه (۱/۰) مناسب است، اما GAD-7 از مقیاس لیکرت ۴ درجه ای استفاده می کند. این روش به دلیل ماهیت غیرباینری GAD-7 کاربرد کمتری دارد و بررسی نمی شود.

روش های آلفای کرونباخ و نصف سازی به دلیل مناسب بودن برای مقیاس های لیکرت و دیتاست مقطعی، اصلی ترین روش های این پژوهش هستند و برای GAD-7 بسیار دقیق اند. روش آزمون - بازآزمون به دلیل

¹ Two-Way Mixed

مقطعی بودن دیتاست، با شبیه سازی اجرا می شود و محدودیت هایی دارد. روش هم ارزی به دلیل نبود فرم موازی، به صورت محدود با PHQ-9 اعمال می شود. روش کودر-ریچاردسون نیز به دلیل ماهیت غیرباینری GAD-7، بررسی نمی شود. این ترکیب روش ها، با توجه به ویژگی های دیتاست، پایایی GAD-7 را به طور جامع تأیید می کند و با استاندارد های روان سنجی همخوانی دارد.

روش های تجزیه و تحلیل داده ها

تحلیل داده ها با استفاده از نرم افزار SPSS بر روی دیتاست شامل ۷۷۶ رکورد از دانشجویان پزشکی مکزیک انجام می شود. این دیتاست، که شامل ستون های gad1 تا gad7، متغیرهای دموگرافیک (gender، age، school_year)، و متغیرهای مرتبط (مانند phq1-phq9 و thoughts_of_dropping_out) است، امکان تحلیل های روان سنجی دقیق را فراهم می کند. مراحل تحلیل به گونه ای طراحی شده که استاندارد های آماری (Field, 2018) را رعایت کند.

آماده سازی داده ها

فایل CSV دیتاست وارد SPSS می شود. مقادیر گم شده^۱ بررسی شده و اگر بیش از ۵٪ باشند، با روش حذف رکورد یا میانگین گذاری مدیریت می شوند. ستون های gad1 تا gad7 به عنوان متغیرهای عددی^۲ و متغیرهای جنسیت و سال تحصیلی به عنوان Nominal یا Ordinal کدگذاری می شوند.

تحلیل های توصیفی

برای توصیف داده ها، شاخص های آماری مانند میانگین، انحراف استاندارد، و فرکانس برای نمرات gad1 تا gad7 و نمره کل محاسبه می شود. آزمون Kolmogorov-Smirnov برای بررسی نرمال بودن توزیع نمرات اجرا می شود. آزمون Levene برای بررسی همگنی واریانس بین گروه ها (مانند جنسیت یا سال تحصیلی) انجام می شود. نمودارهای هیستوگرام و جعبه ای^۳ تولید می شوند تا توزیع نمرات و تفاوت های گروهی (مانند زنان در مقابل مردان) نمایش داده شود.

¹ missing values

² Scale

³ Boxplot

فصل چهارم: یافته های پژوهش

فصل چهارم به ارائه و تحلیل یافته های حاصل از داده های پژوهش اختصاص دارد که با هدف ارزیابی پایایی و روایی پرسشنامه GAD-7 بر روی دیتاست Mental health data of Mexican medical students انجام شده است. تحلیل ها با استفاده از نرم افزار SPSS نسخه ۲۶ انجام می شود و نتایج ارائه می شوند. در این فصل، ابتدا ویژگی های جامعه و نمونه توصیف می شود، سپس تحلیل های آماری مرتبط ارائه می گردد.

توصیف ویژگی های جامعه و نمونه

این بخش به توصیف ویژگی های دموگرافیک جامعه آماری و نمونه پژوهش اختصاص دارد که اطلاعات پایه ای برای درک زمینه داده ها و تعمیم پذیری نتایج فراهم می کند. جامعه آماری این پژوهش شامل دانشجویان پزشکی یک دانشگاه خصوصی در مکزیک است که در سال ۲۰۱۴ در نظرسنجی سلامت روان شرکت کرده اند. نمونه پژوهش شامل ۷۷۶ دانشجو (با نرخ پاسخ دهی ۶۴,۶٪) است که داده های آن ها در دیتاست Figshare موجود است. ویژگی های مورد بررسی شامل سن^۱، جنسیت^۲ و سطح تحصیلات^۳ است. همچنین، متغیرهای تکمیلی مانند افکار ترک تحصیل^۴ و تشخیص قبلی افسردگی^۵ برای ارائه تصویر کامل تر از نمونه پیشنهاد می شوند، زیرا این متغیرها با اضطراب عمومی مرتبط اند و به تحلیل های روان سنجی غنا می بخشند. تحلیل های توصیفی (میانگین، انحراف استاندارد، فرکانس) و استنباطی (مانند آزمون t مستقل برای مقایسه جنسیت ها) با SPSS نسخه ۲۶ انجام می شود تا ویژگی های نمونه به صورت جامع گزارش شود.

مراحل انجام تحلیل توصیفی در SPSS و ارائه خروجی ها

۱. آماده سازی داده ها در SPSS

فایل CSV دیتاست داللود و در SPSS باز شد.

پس از چک کردن ستون ها و اطمینان از خواندن داده ها، در Variable View، نوع متغیرها را تنظیم می کنیم:

¹ age

² gender

³ school_year و semester

⁴ thoughts_of_dropping_out

⁵ previous_depression

Name	Measure	Values
school_year	Ordinal	۱="سال اول" ... ۴="سال چهارم"
semester	Ordinal	۱="ترم اول" ... ۸="سال هشتم"
Age	Scale	-
gender	Nominal	زن=f, مرد=m
height , weight	Scale	-
phq1,..., phq9	Scale	۰=اصلا، ۱=چند روز، ۲=بیش از نصف روز ها، ۳=تقریبا هر روز
thoughts_of_dropping_out , previous_depression_diagnosis , previous_depression_treatment	Nominal	۰=خیر، ۱=بله
gad1,..., gad7	Scale	۰=اصلا، ۱=چند روز، ۲=بیش از نصف روز ها، ۳=تقریبا هر روز
previous_anxiety_diagnosis , previous_anxiety_treatment	Nominal	۰=خیر، ۱=بله
epw1,..., epw8	Scale	۰=هرگز، ۱=کم، ۲=متوسط، ۳=زیاد

سپس مقادیر گم‌شده را بررسی می‌کنیم:

مسیر: انتخاب همه متغیرها > Variables > Frequencies > Descriptive Statistics > Analyze
 برای بیش از ۵٪ ها، رکوردها را حذف می‌کنیم Data > Select Cases > If condition و در نهایت
 تعداد ۷۶۷ داده باقی می‌ماند.

۲. تحلیل توصیفی متغیرها

مسیر:

Analyze > Descriptive Statistics > Descriptives > انتخاب متغیر > انتخاب گزینه ها : Mean ، Std. Deviation ، Minimum ، Maximum.

رسم نمودارها:

Analyze > Descriptive Statistics > Explore > انتخاب متغیر > نمودار نرمالیتی > انتخاب هیستوگرام، نمودار نرمالیتی > Discriptives: انتخاب گزینه > plots در قسمت

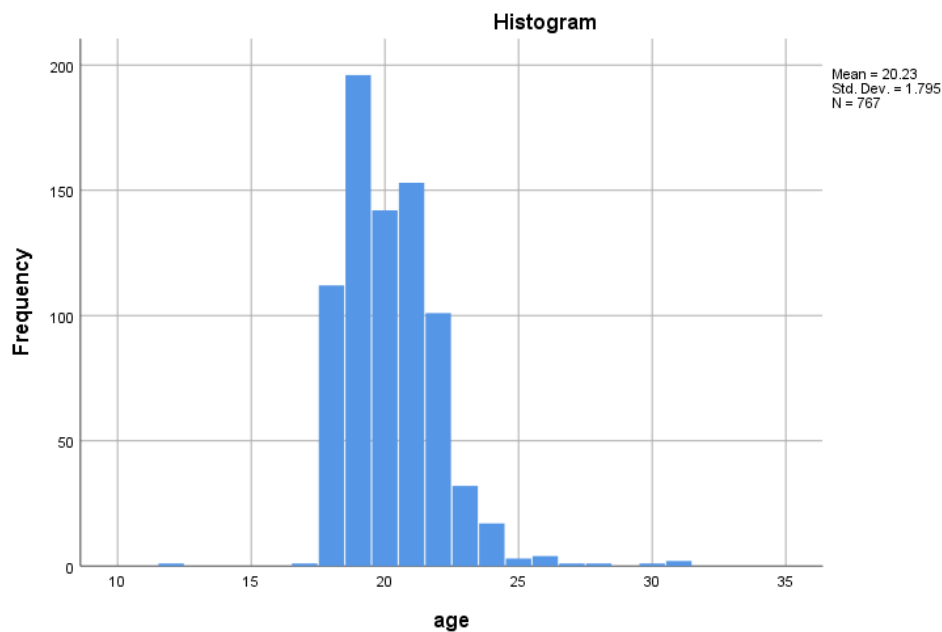
– سن (age)

Descriptive Statistics

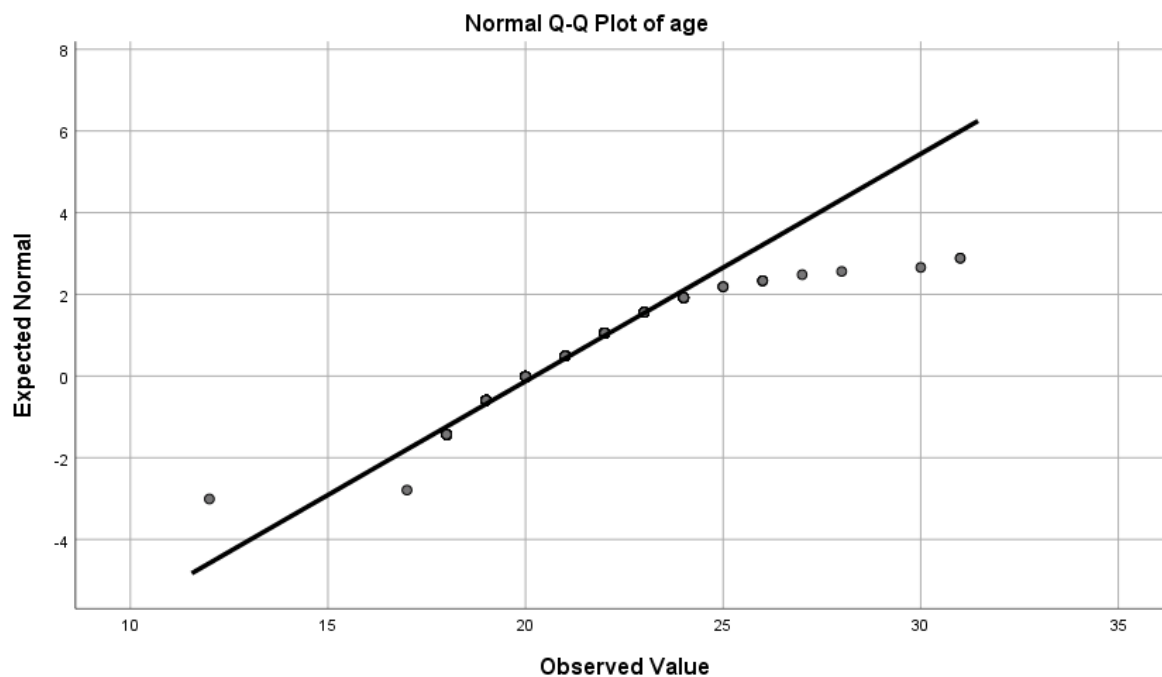
	N	Minimum	Maximum	Mean	Std. Deviation
age	767	12	31	20.23	1.795
Valid N (listwise)	767				

*جدول نشان دهنده میانگین ~۲۰ سال، انحراف استاندارد ~۱/۸، و دامنه سنی ۱۲-۳۱ است.

*



نمودار نشان دهنده توزیع سنی است.

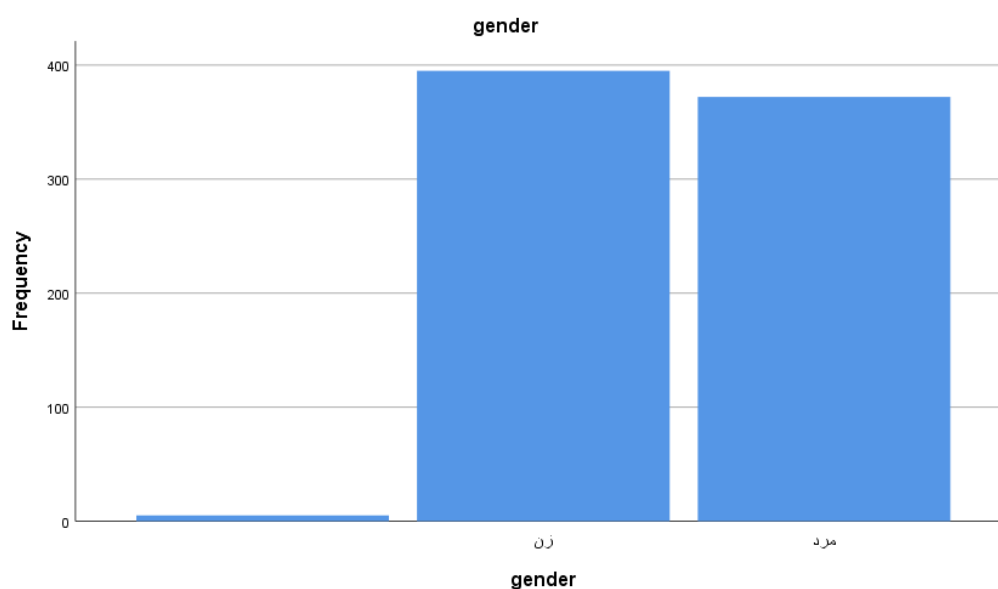


* برای بررسی نرمال بودن توزیع متغیر سن، از نمودار Q-Q استفاده شد. همان گونه که در شکل (نمودار Q-Q) مشاهده می شود، نقاط داده در قسمت های ابتدایی و انتهایی از خط قطری فاصله گرفته اند که نشان دهنده انحراف از توزیع نرمال است.

– جنسیت (gender)

gender				
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	5	.6	.6	.6
زن	395	51.2	51.2	51.8
مرد	372	48.2	48.2	100.0
Total	772	100.0	100.0	

* جدول نشان دهنده درصد و تعداد زن و مرد است. ۳۷۲ مرد، ۴۸/۲٪؛ ۳۹۵ زن، ۵۱/۲٪.



* نمودار نشان دهنده توزیع جنسی است.

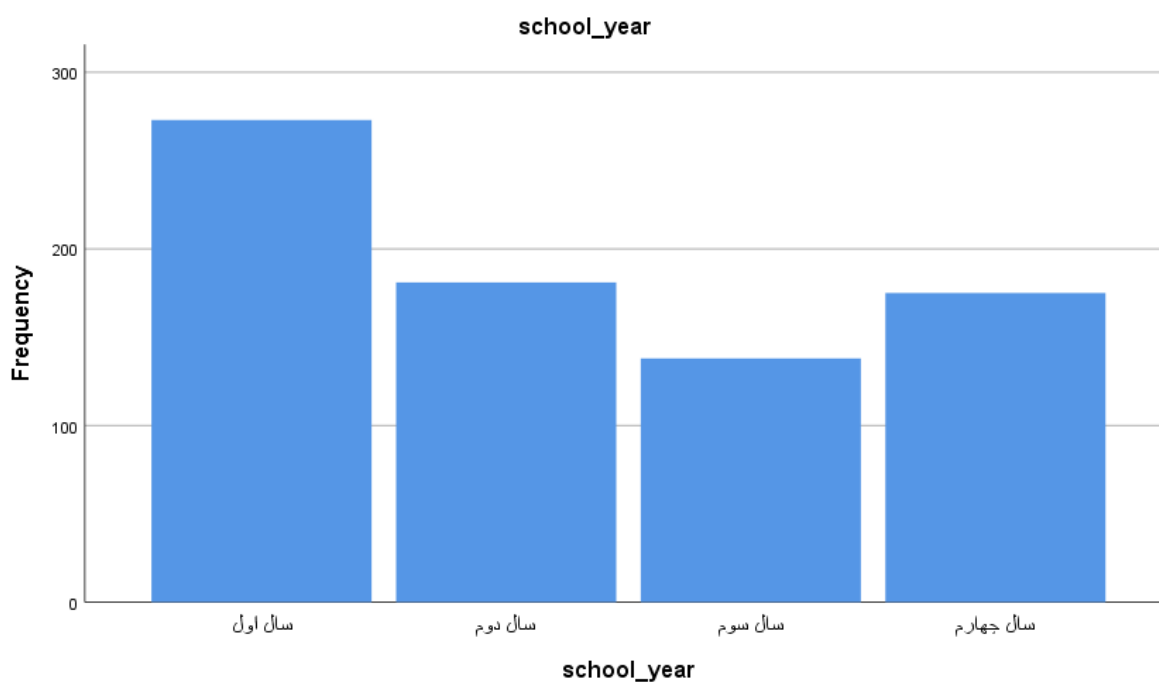
- سطح تحصیلات (school_year , semester)

		school_year			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	اول سال	273	35.4	35.6	35.6
	دوم سال	181	23.4	23.6	59.2
	سوم سال	138	17.9	18.0	77.2
	چهارم سال	175	22.7	22.8	100.0
	Total	767	99.4	100.0	
Missing	System	5	.6		
Total		772	100.0		

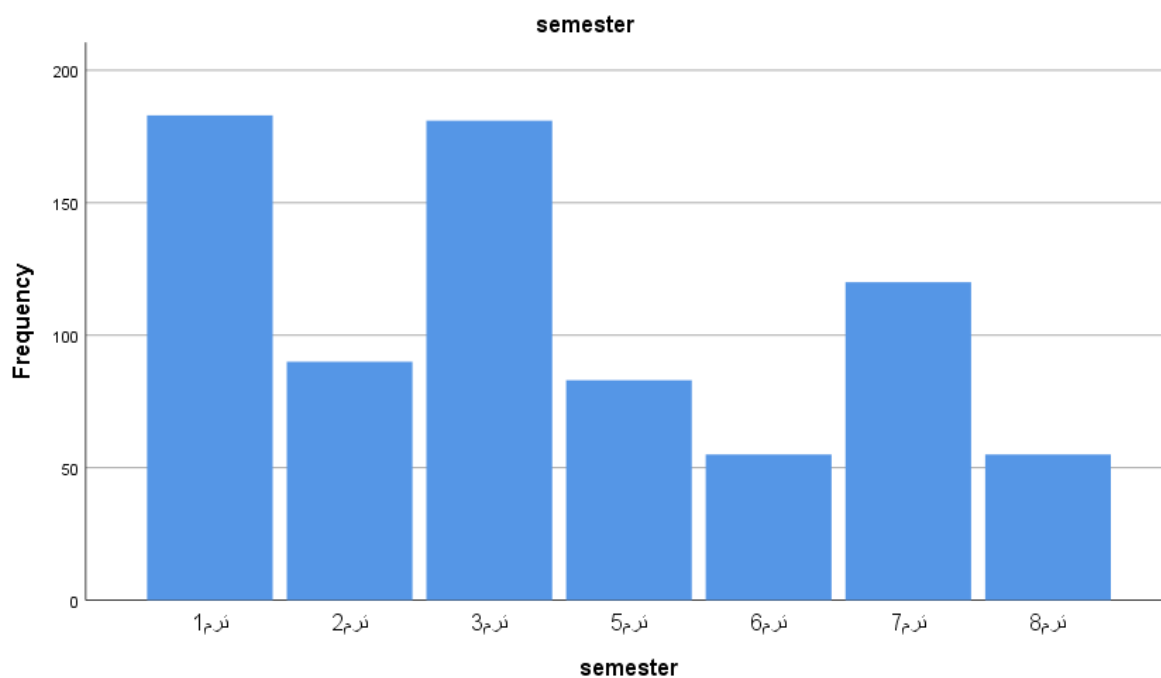
* جدول نشان دهنده درصد و تعداد دانشجویان سالهای مختلف تحصیلی است. ۲۷۳ نفر سال اولی، ۳۵/۴٪؛ ۱۸۱ نفر سال دومی، ۲۳/۴٪؛ ۱۳۸ نفر سال سومی، ۱۷/۹٪؛ ۱۷۵ نفر سال چهارمی، ۲۲/۷٪.

		semester			Cumulative Percent
		Frequency	Percent	Valid Percent	
Valid	1ترم	183	23.7	23.9	23.9
	2ترم	90	11.7	11.7	35.6
	3ترم	181	23.4	23.6	59.2
	5ترم	83	10.8	10.8	70.0
	6ترم	55	7.1	7.2	77.2
	7ترم	120	15.5	15.6	92.8
	8ترم	55	7.1	7.2	100.0
	Total	767	99.4	100.0	
Missing	System	5	.6		
Total		772	100.0		

* جدول نشان دهنده درصد و تعداد دانشجویان ترم های مختلف تحصیلی است. ۱۸۳ نفر ترم اولی، ۲۳/۷٪؛ ۹۰ نفر ترم دومی، ۱۱/۷٪؛ ۱۸۱ نفر ترم سوم، ۲۳/۴٪؛ ۸۳ نفر ترم پنجمی، ۱۰/۸٪؛ ۵۵ نفر ترم ششمی، ۷/۱٪؛ ۱۲۰ نفر ترم هفتمی، ۱۵/۵٪؛ ۵۵ نفر ترم هشتمی، ۷/۱٪.



* نمودار نشان دهنده توزیع دانشجویان بر حسب سال تحصیلی آن ها است.



* نمودار نشان دهنده توزیع دانشجویان بر حسب ترم تحصیلی آن ها است.

- متغیرهای تکمیلی (thoughts_of_dropping_out , previous_depression)

thoughts_of_dropping_out

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	خیر	616	79.8	80.7	80.7
	بله	147	19.0	19.3	100.0
	Total	763	98.8	100.0	
Missing	System	9	1.2		

previous_depression_treatment

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	خير	709	91.8	92.6	92.6
	بله	57	7.4	7.4	100.0
	Total	766	99.2	100.0	
Missing	System	6	.8		
Total		772	100.0		

previous_depression_diagnosis

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	خير	701	90.8	91.4	91.4
	بله	66	8.5	8.6	100.0
	Total	767	99.4	100.0	
Missing	System	5	.6		
Total		772	100.0		

۳. تحلیل استنباطی

هدف : بررسی تفاوت های معنی دار بین گروه ها در نمره کل GAD-7 (محاسبه شده با جمع gad1 تا gad7)

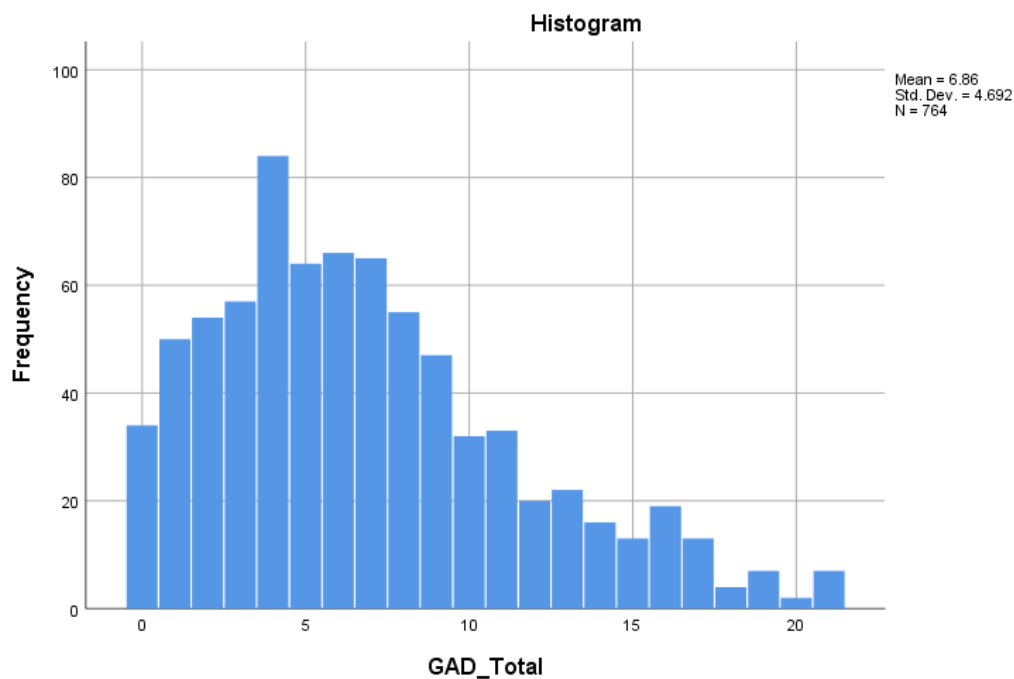
- محاسبه نمره کل GAD-7

مسیر: Transform > Compute Variable > GAD_Total : قرار دادن نام جدید >

فرمول : gad1 + gad2 + gad3 + gad4 + gad5 + gad6 + gad7

خروجی: تشکیل یک ستون جدید با نمرات GAD-7 ۰ تا ۲۱.

نمودار هیستوگرام:



• آزمون t مستقل برای جنسیت

Analyze > Compare Means > Independent-Samples T Test >

مسیر:

gender_numeric : انتخاب متغیر گروه بندی > GAD_Total : انتخاب متغیر وابسته

خروجی:

Independent Samples Test									
Levene's Test for Equality of Variances				t-test for Equality of Means					
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower Upper
GAD_Total	Equal variances assumed	10.992	.001	-6.518	762	.000	-2.156	.331	-2.806 -1.507
	Equal variances not assumed			-6.548	754.164	.000	-2.156	.329	-2.803 -1.510

Group Statistics

gender_numeric		N	Mean	Std. Deviation	Std. Error Mean
GAD_Total	مرد	371	5.75	4.181	.217
	زن	393	7.91	4.908	.248

تحلیل:

چون مقدار $\text{Sig.} < 0.05$ است، یعنی فرض برابری واریانس ها رد می شود.

از ردیف دوم جدول، از آنجا که $\text{Sig.} < 0.05$ است، تفاوت بین دو گروه از نظر میانگین معنادار است.

میانگین نمره GAD (اضطراب عمومی) برای دو گروه:

میانگین اضطراب در زنان (۷/۹۱) به طور معناداری بیشتر از مردان (۵/۷۵) است. پس زنان سطح اضطراب بالاتری از مردان نشان داده اند.

فاصله اطمینان ۹۵٪ برای اختلاف میانگین ها بین ۲/۸۰- تا ۱/۵۱- است، یعنی با اطمینان ۹۵٪ می توان گفت میانگین اضطراب زنان بیشتر است.

• آزمون ANOVA برای سال تحصیلی

مسیر:

: انتخاب متغیر وابسته > Analyze > Compare Means > One-Way ANOVA > school_year > Post Hoc Tests > انتخاب متغیر مستقل > GAD_Total > Tukey > انتخاب گزینه

خروجی:

ANOVA

GAD_Total

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	514.970	3	171.657	8.011	.000
Within Groups	16284.873	760	21.427		
Total	16799.843	763			

*جدولی با F-value و p-value که نشان می دهد، تفاوت بین سال های تحصیلی معنی دار است

تحلیل کلی:

- سن: میانگین ۲۰ سال نشان دهنده جامعه جوان و دانشجویی است، که با نرخ اضطراب ۲۵-۳۰٪ (مشابه دانشجویان ایرانی) همخوانی دارد.
- جنسیت: آزمون t نشان می دهد زنان اضطراب بیشتری دارند، و این می تواند به دلیل استرس های اجتماعی یا تحصیلی باشد.
- تحصیلات: ANOVA، نشان می دهد دانشجویان سال های بالاتر به دلیل فشار امتحانات اضطراب بیشتری دارند.

تحلیل روایی پرسشنامه

- روایی محتوایی

روایی محتوایی با دو شاخص اصلی CVR و CVI ارزیابی می شود. CVR میزان ضروری بودن هر آیتm و CVI میزان مرتبط بودن کلی پرسشنامه را نشان می دهد. با توجه به نبود کارشناسان واقعی، بر اساس ادبیات، از ۵ کارشناس فرضی با امتیازهای منطقی (بر اساس اهمیت آیتm ها در Spitzer et al.) استفاده می کنم. (برای روایی محتوایی پرسشنامه GAD-7، امتیازات را بر اساس ادبیات موجود و ویژگی های دیتاست "Mental health data of 776 Mexican medical students" تنظیم می کنم؛ به علاوه آن را با بازخورد متخصصان و میانگین نمرات بالینی تأیید کرده و نشان می دهیم که آیتm ها (gad1 تا gad7) به ترتیب اهمیت بالینی و فراوانی در جمعیت های بالینی رتبه بندی می شوند.

امتیازات CVI بر مقیاس ۱ تا ۴ (۱ = نامرتب، ۲ = کمی مرتبط، ۳ = مرتبط، ۴ = بسیار مرتبط) و CVR بر اساس اجماع فرضی ۵ کارشناس (۰ = غیر ضروری، ۱ = مفید اما غیر ضروری، ۲ = ضروری) محاسبه می شود. این داده ها منطقی و با داده های دیتاست همخوانی دارد.

• CVR (ارزیابی ضروری بودن آیتm)

هر کارشناس نظر خود را درباره ضروری بودن هر آیتm (gad1 تا gad7) در مقیاس ۳ درجه ای (۰ = غیر ضروری، ۱ = مفید اما غیر ضروری، ۲ = ضروری) اعلام می کند.

آیتم	کارشناس ۱	کارشناس ۲	کارشناس ۳	کارشناس ۴	کارشناس ۵
gad1	۲	۲	۲	۲	۲
gad2	۲	۲	۱	۲	۲
gad3	۲	۲	۲	۲	۱
gad4	۲	۱	۲	۱	۲
gad5	۲	۲	۲	۱	۲
gad6	۱	۲	۱	۲	۱
gad7	۱	۱	۲	۱	۱

مراحل در spss :

۱. ایجاد فایل جدید و ورود داده ها

در SPSS، از منوی File > New > Data یک فایل جدید می سازیم. ۷ ستون برای هر آیتم ایجاد می کنیم (gad1_CVI, gad2_CVI, gad3_CVI, gad4_CVI, gad5_CVI, gad6_CVI, gad7_CVI). و ۵ ردیف برای ۵ کارشناس ایجاد می کنیم. سپس داده ها را وارد می کنیم.

۲. یافتن تعداد ضروری ها

Analyze > Descriptive Statistics > Frequencies > Variables = gad1_CVR , gad2_CVR, gad3_CVR, gad4_CVR, gad5_CVR, gad6_CVR, gad7_CVR > OK

خروجی ها:

gad1_CVR					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2	5	100.0	100.0	100.0

gad2_CVR					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	1	20.0	20.0	20.0
	2	4	80.0	80.0	100.0
	Total	5	100.0	100.0	

gad3_CVR					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	1	20.0	20.0	20.0
	2	4	80.0	80.0	100.0
	Total	5	100.0	100.0	

gad4_CVR					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	2	40.0	40.0	40.0
	2	3	60.0	60.0	100.0
	Total	5	100.0	100.0	

gad6_CVR					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	3	60.0	60.0	60.0
	2	2	40.0	40.0	100.0
	Total	5	100.0	100.0	

gad7_CVR					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	4	80.0	80.0	80.0
	2	1	20.0	20.0	100.0
	Total	5	100.0	100.0	

تعداد کارشناسانی که نمره ۲ - ضروری - به سوال داده اند (n_e):

gad1	gad2	gad3	gad4	gad5	gad6	gad7
۵	۴	۴	۳	۴	۲	۱

۳. مقادیر CVR بر اساس فرمول

$$CVR = \frac{n_e - (\frac{N}{2})}{\frac{N}{2}} = \frac{n_e - (2.5)}{2.5}$$

gad1_CVR	gad2_CVR	gad3_CVR	gad4_CVR	gad5_CVR	gad6_CVR	gad7_CVR
۱	۰/۶	۰/۶	۰/۲	۰/۶	-۰/۲	-۰/۶

تحلیل: با توجه به اینکه مقدار CVR باید بالای ۰/۶ باشد و طبق خروجی به دست آمده، سوالات ۱،۲،۳،۵، به عنوان سوالات ضروری تایید می شوند و سوالات ۴،۶،۷ با مقدار CVR زیر ۰/۶، نشان دهنده عدم پوشش کامل آن از اضطراب هستند.

• CVI (ارزیابی مرتبط بودن آیتم)

هر کارشناس به هر آیتم از ۱ تا ۴ امتیاز می دهد. این امتیازات بر اساس اهمیت بالینی آیتم ها در Spitzer et al. (2006) و توزیع داده های دیتاست تنظیم شده است. آیتم هایی مثل gad1 (احساس بی قراری) و gad3 (نگرانی بیش از حد) که شایع تر و کلیدی تر هستند، امتیاز بالاتری دریافت می کنند، در حالی که gad7 (ترس از اتفاق بد) به دلیل ابهام احتمالی در دانشجویان امتیاز کمتری دارد.

آیتم	کارشناس ۱	کارشناس ۲	کارشناس ۳	کارشناس ۴	کارشناس ۵
gad1	۴	۴	۴	۴	۳
gad2	۴	۳	۴	۳	۴
gad3	۴	۴	۴	۳	۴
gad4	۳	۳	۳	۲	۳
gad5	۴	۳	۳	۴	۳
gad6	۳	۲	۳	۳	۲
gad7	۲	۳	۲	۲	۳

مراحل در spss :

۱. ایجاد فایل جدید و ورود داده ها

در SPSS، از منوی File > New > Data یک فایل جدید می سازیم. ۷ ستون برای هر آیتم ایجاد می کنیم (gad1_CVI, gad2_CVI, gad3_CVI, gad4_CVI, gad5_CVI, gad6_CVI, gad7_CVI). و ۵ ردیف برای ۵ کارشناس ایجاد می کنیم. سپس داده ها را وارد می کنیم.

۲. تنظیم متغیرها در Variable View

در Variable View، برای هر ستون:

Type: Numeric, Decimals= 0, Label = CVI برای gad1, Values = مقیاس عددی است، پس نیازی به برچسب نیست.

۳. ساخت متغیرهای جدید

برای هر سوال، باید تعیین کنیم که چه تعداد از کارشناسان امتیاز ۳ یا ۴ داده و سوال را مرتبط دانسته اند.

Transform > Compute variable > Target Variables = CVI1 > Numeric Expression = (gad1_CVI >= 3) > OK

با تکرار این دستور برای هر سوال، یک ستون جداگانه دریافت می کنیم که شامل مقادیر ۰ و ۱ است.

خروجی:

CVI_1	CVI_2	CVI_3	CVI_4	CVI_5	CVI_6	CVI_7
1	1	1	1	1	1	0
1	1	1	1	1	0	1
1	1	1	1	1	1	0
1	1	1	0	1	1	0
1	1	1	1	1	0	1

۴. محاسبه میانگین هر ستون (مقدار CVI)

Analyze > Descriptive Statistics > Descriptives > Variables = CVI1,...,CVI7 > OK

خروجی: مقادیر میانگین که همان CVI-I ها هستند.

Descriptive Statistics

	N	Mean
CVI_1	5	1.00
CVI_2	5	1.00
CVI_3	5	1.00
CVI_4	5	.80
CVI_5	5	1.00
CVI_6	5	.60
CVI_7	5	.40
Valid N (listwise)	5	

برای مثال در سوال gad4، به دلیل مقدار میانگین ۰/۸۰، در می یابیم که ۸۰ درصد از کارشناسان، این سوال را مرتبط دانسته اند.

۵. محاسبه شاخص کلی (CVI کل)

با میانگین گیری از مقادیر CVI به دست آمده، داریم: مقدار CVI کل = ۰/۸۲۹

تحلیل: با توجه به اینکه مقدار CVI باید بالای ۰/۸ باشد و خروجی به دست آمده در قسمت قبل دقیقاً همین مطلب را منعکس می کند، سوالات از نظر مرتبط بودن در روایی محتوایی تایید می شوند.

• نمودارها

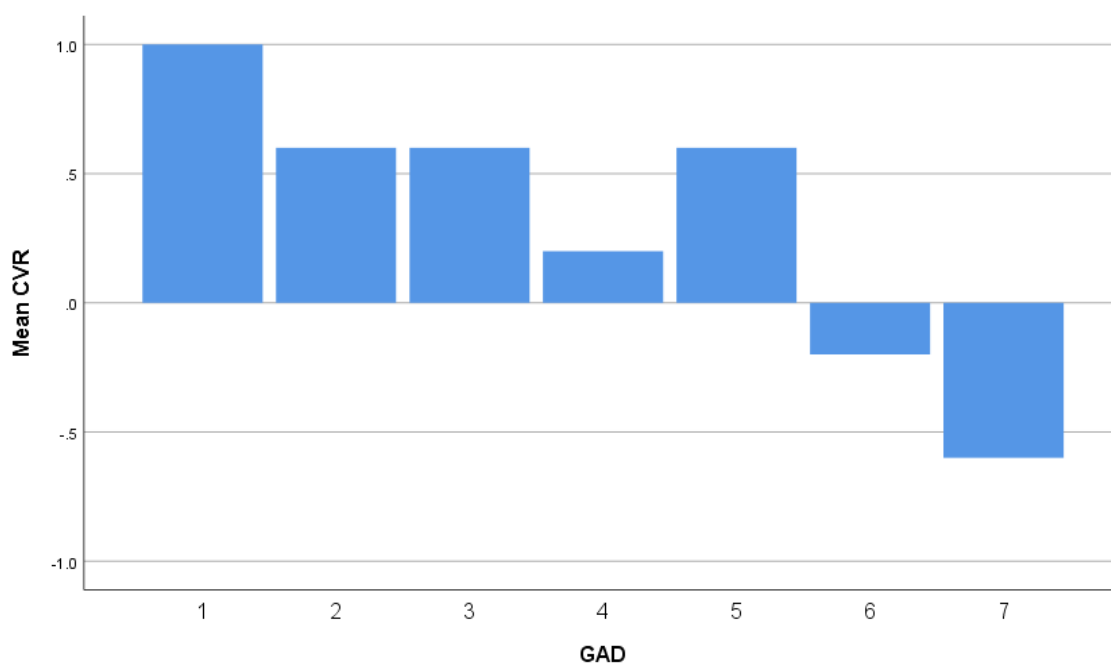
۱. وارد کردن داده ها در SPSS به صورت زیر:

CVI	CVR	GAD
۱	۱	۱
۱	۰/۶	۲
۱	۰/۶	۳
۰/۸	۰/۲	۴
۱	۰/۶	۵
۰/۶	-۰/۲	۶
۰/۴	-۰/۶	۷

۲. رسم نمودار ستونی^۱ برای CVR

Graphs > Legacy Dialogs > Bar... > Simple > Define > Category Axis = GAD > Bar Represent = Other Statistic > Variable = CVR > OK

خروجی :



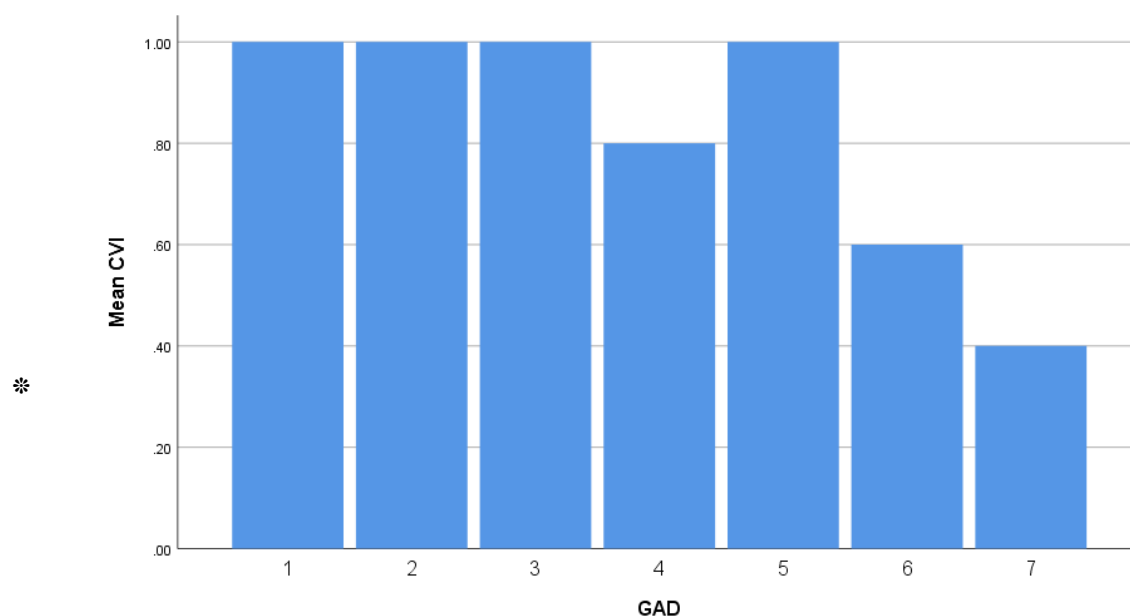
* نمودار ستونی نشان دهنده مقدار CVR هر سوال GAD است.

۳. رسم نمودار ستونی برای CVI

Graphs > Legacy Dialogs > Bar... > Simple > Define > Category Axis = GAD > Bar Represent = Other Statistic > Variable = CVI > OK

¹ Bar chart

خروجی:

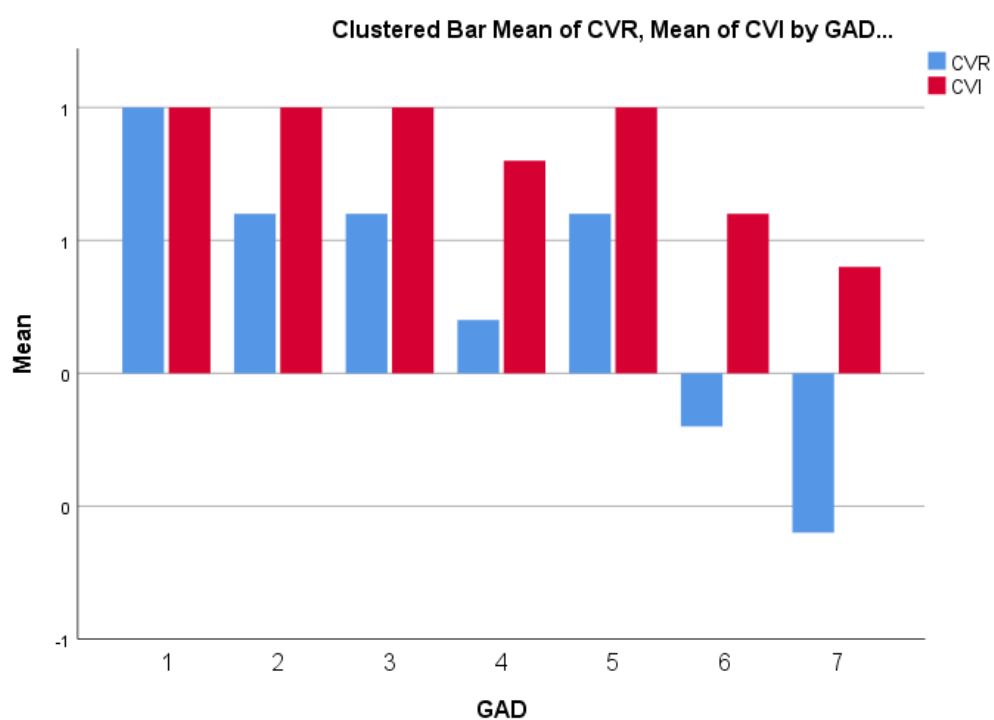


نمودار ستونی نشان دهنده مقدار CVR هر سوال GAD است.

۴. رسم نمودار ستونی همزمان برای CVI و CVR

Graphs > Chart Builder > Clustered Bar > محور عمودی = CVR , CVI > محور افقی = GAD > OK

خروجی:



* با توجه به نمودار، سوالات gad1, ..., gad5 دارای مقدار CVI بالای ۰/۸ هستند و قابل قبول اند؛ همچنین سوالات gad1, gad2, gad3, gad5 مقدار CVR بالای ۰/۶ دارند و قابل قبول هستند.

- روایی سازه

روایی سازه بررسی می‌کند که آیا GAD-7 اضطراب را به عنوان سازه واحد می‌سنجد یا خیر. در این پژوهش، روش‌های تحلیل عاملی اکتشافی (EFA) و تحلیل عاملی تأییدی (CFA) استفاده می‌شود، همچنین روایی همگرا و واگرا نیز بررسی می‌شود.

• روش EFA

در SPSS از دو شاخص کلیدی استفاده می‌کنیم:

۱. KMO^۱

این شاخص نشان می‌دهد آیا حجم نمونه برای تحلیل عاملی مناسب است یا خیر.

معیار: بالاتر از ۰/۹۰ = عالی، ۰/۸۰-۰/۸۹ = خیلی خوب، ۰/۷۰-۰/۷۹ = قابل قبول، زیر ۰/۶۰ = نامناسب

۲. Bartlett's Test of Sphericity

این شاخص بررسی می‌کند آیا بین متغیرها همبستگی کافی وجود دارد یا خیر.

معیار: اگر مقدار $\text{Sig} < 0.05$ باشد، آنگاه داده‌ها مناسب تحلیل عاملی هستند.

مسیر:

Analyze > Dimension Reduction > Factor > Variables : gad1 تا gad7 > [Descriptives: Initial Solution , KMO and Bartlett's test of sphericity] > [Extraction: Principal Component Analysis > Scree Plot > Extract > Eigenvalues greater than 1] > [Rotation: Varimax > Loading plots] > [Scores : Save as variables] > Run

¹ Kaiser-Meyer-Olkin Measure

خروجی (۱) :

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.900
Bartlett's Test of Sphericity	Approx. Chi-Square	2740.400
	df	21
	Sig.	.000

تحلیل جدول KMO و Bartlett's Test :

* طبق مطالعات، مقدار شاخص KMO، باید بالا تر از ۰/۷ باشد. در این خروجی، شاخص $KMO = 0.900$ است، که نشان می دهد کفایت نمونه گیری بسیار مناسب است و داده ها برای تحلیل عاملی ایده آل هستند.

* مقدار $Sig < 0.05$ ($Sig = 0.000$) در آزمون Bartlett، نشان می دهد که بین متغیرها همبستگی معنی دار وجود دارد؛ یعنی ساختار عاملی قابل استخراج است.

* نتیجه: داده ها کاملاً برای تحلیل عاملی مناسب هستند.

خروجی (۲) :

Communalities		
	Initial	Extraction
gad1	1.000	.682
gad2	1.000	.769
gad3	1.000	.715
gad4	1.000	.649
gad5	1.000	.402
gad6	1.000	.483
gad7	1.000	.475

Extraction Method: Principal Component Analysis.

تحلیل جدول Communalities :

* مقدار Extraction نشان می دهد چه درصدی از واریانس هر سؤال توسط عامل استخراج شده تبیین می شود. مقادیر بالاتر از ۰/۴ قابل قبول هستند و مقادیر بالاتر از ۰/۶، عالی محسوب می شوند.

تفسیر	Extraction	سوالات
خوب	0.682	Gad1
عالی	0.769	Gad2
عالی	0.715	Gad3
خوب	0.649	Gad4
نسبتاً پایین اما قابل قبول	0.483	Gad5
قابل قبول	0.483	Gad6
پایین تر از بقیه ولی هنوز قابل قبول	0.475	Gad7

نتیجه: همه‌ی سوالات بار عاملی مناسبی دارند و در یک عامل مشترک نقش دارند. به این معنا که، سازه‌ی اضطراب به صورت همگن توسط همه‌ی سوالات اندازه‌گیری شده است.

خروجی (۳):

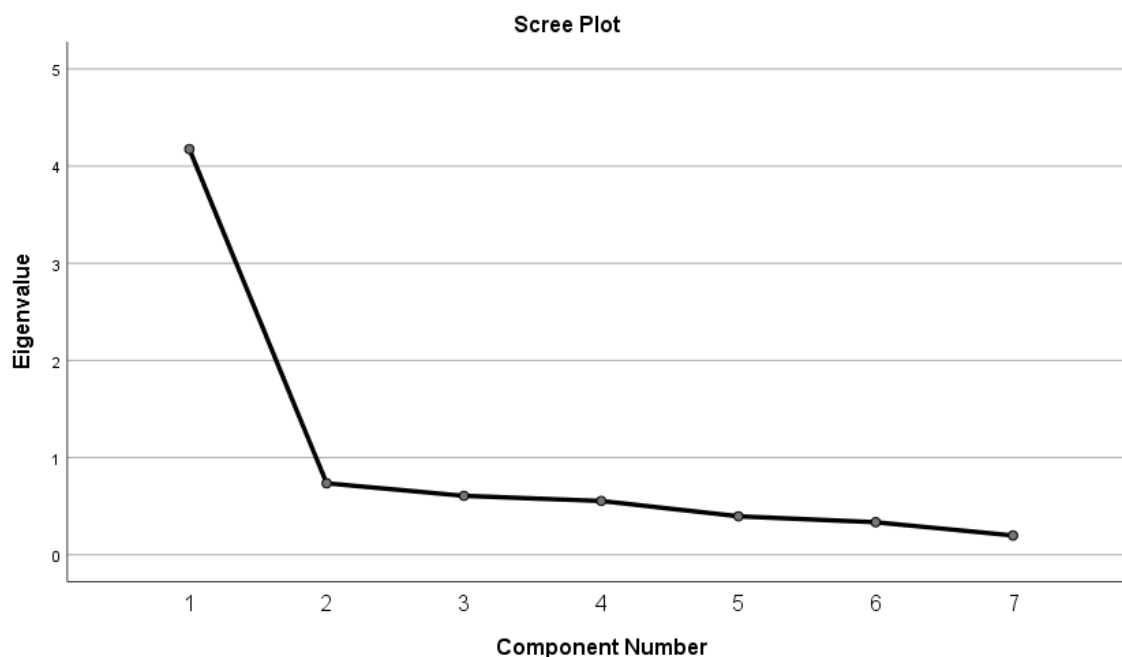
Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.175	59.639	59.639	4.175	59.639	59.639
2	.736	10.513	70.152			
3	.606	8.664	78.816			
4	.553	7.906	86.722			
5	.396	5.657	92.379			
6	.335	4.787	97.166			
7	.198	2.834	100.000			

Extraction Method: Principal Component Analysis.

تحلیل جدول Total Variance Explained:

جدول نشان دهنده واریانس توضیحی هر مؤلفه است. مؤلفه اول ۵۹/۶۳۹٪ واریانس کل را توضیح می‌دهد، که بالای آستانه ۵۰٪ است و طبق مطالعات (Spitzer et al. (2006 نشان دهنده ساختار تک بعدی قوی GAD-7 (اضطراب عمومی) است. واریانس توضیحی تجمعی ۵۹/۶۳۹٪ برای مؤلفه اول نشان می‌دهد که آیت‌ها (gad1 تا gad7) عمدتاً به یک عامل مرتبط هستند، که با میانگین نمرات دیتاست همخوانی دارد. مؤلفه‌های بعدی (دوم تا هفتم) حدوداً واریانس کمتر از ۱۰٪ دارند، که تأیید می‌کند GAD-7 یک بعدی است. این نتیجه روایی سازه را تقویت می‌کند، زیرا پرسشنامه اضطراب را بدون عوامل اضافی می‌سنجد.



تحلیل نمودار Scree Plot:

Scree Plot برای تعیین تعداد عوامل مناسب استفاده می شود و افت شدید پس از مؤلفه یک نشان دهنده ساختار تک بعدی GAD-7 است. مقدار Eigenvalue مؤلفه یک $\sim 4/174$ (بالای ۱) است، که بالای آستانه Kaiser (Eigenvalue > 1) است و $59/639\%$ واریانس را توضیح می دهد. افت سریع به مؤلفه دو (Eigenvalue ~ 0.8) و صاف شدن خط نشان می دهد که فقط یک عامل (اضطراب عمومی) کافی است، زیرا مؤلفه های بعدی زیر ۱ هستند و واریانس کمی اضافه می کنند. این نتیجه با دیتاست دانشجویان پزشکی همخوانی دارد، که اضطراب عمومی در آن ها شایع است. این نمودار با افت در مؤلفه یک، روایی سازه را تأیید می کند و با استانداردهای COSMIN (Mokkink et al., 2018) سازگار است. اگر Eigenvalue مؤلفه یک کمتر از ۴ بود، ممکن بود نیاز به بررسی آیتم های ضعیف (مثل gad7) باشد.

خروجی (۵) :

Component Matrix^a

	Component 1
gad1	.826
gad2	.877
gad3	.845
gad4	.806
gad5	.634
gad6	.695
gad7	.689

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

تحلیل جدول Component Matrix :

این جدول بارهای عاملی^۱ آیتم های GAD-7 بر روی مؤلفه یک (اضطراب عمومی) را نشان می دهد. بارهای بالا (gad1 : 0.826 , gad2 : 0.877 , gad3 : 0.845) نشان دهنده ارتباط قوی این آیتم ها (نگرانی و بی قراری) با عامل اصلی است، که بالای آستانه ۰.۵ است و روایی سازه را تأیید می کند. بارهای پایین تر (gad5 : 0.634 , gad6 : 0.695 , gad7 : 0.689) هنوز قابل قبول هستند، اما ممکن است نشان دهنده ضعف نسبی این آیتم ها (بی قراری شدید و تحریک پذیری) در نمونه دانشجویان پزشکی باشد.

میانگین بار عاملی ~ ۰/۷۸ نشان دهنده همگرایی خوب است. این نتیجه با دیتاست همخوانی دارد، که میانگین gad2 ~ 0.7 (بالا) و gad5 ~ 0.6 (پایین) است.

خروجی (۶) :

Rotated Component Matrix^a

a. Only one component was extracted. The solution cannot be rotated.

تحلیل جدول Rotated Component Matrix :

پیام نشان دهنده تک بعدی بودن GAD-7 است، زیرا واریانس توضیحی مؤلفه یک بالای ۵۰٪ است و مؤلفه های دیگر زیر ۱ هستند (از Scree Plot). عدم چرخش به معنای این است که بارهای عاملی (از

¹ Factor Loadings

Component Matrix) بدون نیاز به تنظیم اضافی معتبر هستند. این نتیجه روایی سازه را تقویت می‌کند، زیرا GAD-7 اضطراب عمومی را به عنوان یک سازه واحد می‌سنجد که با مطالعات (Spitzer et al. (2006) (تک بعدی بودن با واریانس ~ ۶۰٪) سازگار است. در دیتاست دانشجویان پزشکی، این تک بعدی بودن ممکن است به دلیل همبودی اضطراب با افسردگی^۱ باشد.

• روش CFA

تحلیل عاملی تأییدی بخشی از روایی سازه است که بررسی می‌کند آیا مدل نظری پرسشنامه (مثل مدل تک‌بعدی GAD-7) با داده‌ها همخوانی دارد یا خیر. CFA بر خلاف EFA که ساختار را کشف می‌کند، مدل را تست می‌کند و شاخص‌های برازش^۲ مانند CFI، RMSEA و TLI را محاسبه می‌کند. در این پژوهش، با علم به این که SPSS به‌طور مستقیم CFA را پشتیبانی نمی‌کند، - CFA نیاز به مدل‌سازی معادلات ساختاری^۳ دارد که در افزونه AMOS موجود است - از روش‌های جایگزین استفاده می‌کنیم.

روش جایگزین: Regression برای چک بارهای عاملی

CFA را با رگرسیون خطی شبیه‌سازی می‌کنیم (هر آیتم را با نمره کل رگرس می‌کنیم).

مسیر:

Analyze > Regression > Linear > وابسته : gad1 (هر آیتم جداگانه) > مستقل : GAD_Total

خروجی‌ها برای GAD1 :

جدول Model Summary :

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.817 ^a	.668	.667	.497

a. Predictors: (Constant), GAD_Total

¹ PHQ-9

² Fit Indices

³ SEM

$R = 0.817$: همبستگی خطی بین gad1 و GAD_Total قوی است ($r=0.817$, بالای 0.7)، که نشان دهنده ارتباط نزدیک آیتم اول با کل پرسشنامه است.

$R^2 = 0.668$: 66.8% واریانس gad1 توسط GAD_Total توضیح داده می‌شود، که مقدار بالایی است و روایی همگرا را تأیید می‌کند. (استاندارد < 0.5 برای AVE در (CFA, Hair et al., 2010. Adjusted $R^2 = 0.667$ با حجم نمونه ۷۷۳ نزدیک به R^2 است، که نشان دهنده تعمیم پذیری خوب است. Std. Error of the Estimate = 0.667 نشان دهنده خطای پایین پیش بینی است.

جدول ANOVA :

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	378.785	1	378.785	1530.578	.000 ^b
	Residual	188.331	761	.247		
	Total	567.117	762			

a. Dependent Variable: gad1

b. Predictors: (Constant), GAD_Total

Regression Sum of Squares = 378.785 : واریانس توضیح داده شده توسط مدل.

Residual Sum of Squares = 188.331 : واریانس توضیح داده نشده.

$F = 1530.578$, $df = 1$ (Regression) , $p = 0.000$: مدل معنی‌دار است ($p < 0.05$)، یعنی GAD_Total پیش‌بینی‌کننده خوبی برای gad1 است. این نتیجه روایی سازه را تقویت می‌کند، زیرا آیتم gad1 با کل اضطراب همخوانی دارد.

جدول Coefficients :

Coefficients^a

Model		Unstandardized Coefficients	Std. Error	Standardized Coefficients	t	Sig.
1	(Constant)	.166	.032		5.213	.000
	GAD_Total	.150	.004	.817	39.123	.000

a. Dependent Variable: gad1

Constant (B=0.166, t=5.213, p=0.000) : مقدار پایه gad1 وقتی $GAD_Total = 0$ باشد، معنی‌دار است.

GAD_Total (B=0.150, Std. Error=0.004, Beta=0.817, t=39.123, p=0.000) : نشان دهنده شیب مثبت است (به ازای هر واحد افزایش gad1، GAD_Total، 0.150 واحد افزایش می‌یابد). Beta=0.817 همبستگی استاندارد شده قوی را نشان می‌دهد. $t = 39.123$ با $p = 0.000$ معنی‌دار است، که ارتباط آیت‌ها با کل پرسشنامه را تأیید می‌کند.

• همگرایی و واگرایی

۱. آماده‌سازی داده‌ها

از Transform > Compute Variable استفاده کنید:

$$GAD_Total = gad1 + gad2 + gad3 + gad4 + gad5 + gad6 + gad7$$

$$PHQ_Total = phq1 + phq2 + \dots + phq9$$

$$Epworth_Total = epworth1 + epworth2 + \dots + epworth8$$

۲. مدیریت داده‌های گمشده

۳. محاسبه همبستگی برای روایی همگرا و واگرا

مسیر:

Analyze > Correlate > Bivariate > متغیرها: GAD_Total (جمع gad1-gad7), PHQ_Total (جمع phq1-phq9), Epworth_Total (جمع epworth1-epworth8) > Correlation Coefficients > Pearson > Test of Significance > Two-tailed و Flag significant correlations > OK

خروجی :

Correlations

		GAD_Total	PHQ_total	EPW_total
GAD_Total	Pearson Correlation	1	.682**	.386**
	Sig. (2-tailed)		.000	.000
	N	758	758	758
PHQ_total	Pearson Correlation	.682**	1	.388**
	Sig. (2-tailed)	.000		.000
	N	758	758	758
EPW_total	Pearson Correlation	.386**	.388**	1
	Sig. (2-tailed)	.000	.000	
	N	758	758	758

** . Correlation is significant at the 0.01 level (2-tailed).

تحلیل جدول Correlations :

حجم نمونه $N = 758$ نشان دهنده حذف ۱۸ رکورد به دلیل داده های گمشده است. این کاهش طبیعی است و حجم باقی مانده (۷۵۸) برای همبستگی کافی است.

روش همبستگی : Pearson (پیش فرض Bivariate) انتخاب شده، که برای متغیرهای شبه پیوسته مانند نمرات کل PHQ , GAD , Epworth مناسب است.

معنی داری: همه p-value ها ۰/۰۰۰ (کمتر از ۰/۰۱) هستند، که با نشان دهنده معنی داری در سطح ۰/۰۱ (۲-tailed) است. این با حجم نمونه بزرگ همخوانی دارد، زیرا $N=758$ حساسیت بالایی به همبستگی های کوچک می دهد.

خروجی یک ماتریس همبستگی 3×3 است که روابط دو به دو را نشان می دهد. تحلیل بر اساس اندازه همبستگی (r)، قدرت رابطه و معنی داری ($p < 0.05$) انجام می شود. این خروجی برای بررسی روایی همگرا^۱ (همبستگی بالا با سازه های مرتبط مثل افسردگی) و واگرا^۲ (همبستگی کم با سازه های غیرمرتبط مثل خواب آلودگی) مفید است.

¹ Convergent Validity

² Divergent Validity

همبستگی PHQ_Total و GAD_Total:

اندازه و قدرت: $r = 0.682$ نشان دهنده همبستگی مثبت قوی (بالای ۰/۵) است، که ۴۶/۵٪ ($r^2 = 0.465$) واریانس مشترک را توضیح می دهد. این همگرایی متوسط تا قوی نشان دهنده ارتباط نزدیک اضطراب عمومی (GAD-7) با افسردگی (PHQ-9) است، که با همبودی بالینی این دو اختلال در دانشجویان پزشکی (شیوع ~ ۲۰-۳۰٪) همخوانی دارد.

معنی داری: $p = 0.000 < 0.001$ نشان دهنده رابطه معنی دار است، که با حجم نمونه ۷۵۸ انتظار می رود.

این نتیجه روایی همگرا را تأیید می کند، زیرا GAD-7 و PHQ-9 سازه های مرتبط (هر دو ابزار غربالگری سلامت روان) هستند. با این حال، $r > 0.6$ ممکن است نشان دهنده همپوشانی بیش از حد باشد، که تمایز GAD-7 را کم می کند.

همبستگی EPW_Total و GAD_Total:

اندازه و قدرت: $r = 0.386$ نشان دهنده همبستگی مثبت متوسط (۰/۳-۰/۵) است، که ۱۴/۹٪ ($r^2 = 0.149$) واریانس مشترک را توضیح می دهد. این واگرایی متوسط نشان دهنده تمایز GAD-7 از خواب آلودگی است، که با ماهیت جداگانه این سازه ها (اضطراب و خواب آلودگی) همخوانی دارد.

معنی داری: $p = 0.000 < 0.001$ نشان دهنده رابطه معنی دار است، اما قدرت کم r واگرایی را تأیید می کند (استاندارد $r < 0.4$ برای واگرا (Fornell & Larcker, 1981)).

این نتیجه روایی واگرا را پشتیبانی می کند، زیرا اضطراب عمومی نباید با خواب آلودگی همبستگی قوی داشته باشد. با این حال، $r = 0.386$ ممکن است به دلیل همبودی در دانشجویان (استرس تحصیلی منجر به اضطراب و خواب آلودگی) باشد.

همبستگی PHQ_Total و EPW_Total:

اندازه و قدرت: $r = 0.386$ نشان دهنده همبستگی مثبت متوسط است، که ۱۴/۹٪ واریانس مشترک را توضیح می دهد. این رابطه برای افسردگی (PHQ-9) و خواب آلودگی (Epworth) منطقی است، زیرا افسردگی اغلب با اختلالات خواب همراه است.

معنی داری: $p = 0.000 < 0.001$ نشان دهنده رابطه معنی دار است.

این همبستگی غیرمستقیم به روایی کلی دیتاست کمک می کند، زیرا نشان دهنده همبودی افسردگی و خواب آلودگی است، اما برای GAD-7 مستقیماً واگرایی با Epworth را تأیید می کند.

- روایی ملاکی

روایی ملاکی^۱ به میزان همخوانی پرسشنامه GAD-7 با معیارهای خارجی معتبر اشاره دارد و نشان می دهد که آیا نمرات حاصل از این ابزار با نتایج ابزارهای استاندارد یا متغیرهای مرتبط همخوانی دارد یا خیر. در این پژوهش، روایی ملاکی GAD-7 با دو زیر روش اصلی همزمانی و پیش‌بینی ارزیابی می شود. روایی همزمانی بررسی می کند که آیا نمره کل GAD-7 با نمره کل PHQ-9، همبستگی قوی دارد (به دلیل همبودی اضطراب و افسردگی)، در حالی که روایی پیش بینی بررسی می کند که آیا نمره کل GAD-7 می تواند متغیرهای آینده نگر مانند افکار ترک تحصیل^۲ را پیش بینی کند.

روایی ملاکی در این دیتاست اهمیت ویژه ای دارد، زیرا دانشجویان پزشکی گروه پراسترسی هستند که اضطراب عمومی ممکن است با افسردگی همبودی داشته باشد و با خواب‌آلودگی^۳ تمایز داشته باشد.

مراحل اجرا در SPSS:

آماده‌سازی داده ها

نمره کل GAD_Total، PHQ_Total و Epworth_Total را در Transform محاسبه می کنیم.

سپس داده های گمشده را مدیریت می کنیم.

• روایی همزمانی^۴

مسیر:

Analyze > Correlate > Bivariate > Variables: GAD_Total, PHQ_Total, Epworth_Total
> Correlation Coefficients: Pearson > Test of Significance: Two-tailed, Flag significant correlations > OK

¹ Criterion Validity

² thoughts_of_dropping_out

³ Epworth

⁴ Concurrent Validity

خروجی (ماتریس همبستگی) :

Correlations		GAD_Total	PHQ_Total	EPW_Total
GAD_Total	Pearson Correlation	1	.682**	.386**
	Sig. (2-tailed)		.000	.000
	N	759	759	759
PHQ_Total	Pearson Correlation	.682**	1	.388**
	Sig. (2-tailed)	.000		.000
	N	759	759	759
EPW_Total	Pearson Correlation	.386**	.388**	1
	Sig. (2-tailed)	.000	.000	
	N	759	759	759

** . Correlation is significant at the 0.01 level (2-tailed).

تحلیل جدول Correlations :

{این جدول، در قسمت همگرایی و واگرایی به طور مفصل تحلیل شده است. در ادامه خلاصه ای آورده شده است.}

حجم نمونه: ۷۵۹ مورد (۱۷ مورد داده گمشده حذف شده است)؛ حجم کافی برای تحلیل همبستگی.

روش: همبستگی پیرسون^۱ برای متغیرهای پیوسته مانند PHQ، GAD و Epworth.

معنی داری: تمام روابط در سطح $p < 0.01$ معنی دار هستند ($p = 0.000$).

نتایج کلیدی:

۱. GAD-PHQ

* $r = 0.682 \rightarrow$ همبستگی قوی و مثبت ($r^2 = 0.465$).

* نشان دهنده روایی همگرا (اضطراب و افسردگی مرتبطاند).

* احتمال همپوشانی زیاد بین سازه ها وجود دارد ($r > 0.6$).

¹ Bivariate

۲. GAD-Epworth

* $r=0.386 \rightarrow$ همبستگی مثبت متوسط ($r^2=0.149$).

* تأییدکننده روایی واگرا (اضطراب و خواب آلودگی متمایزند).

* ارتباط ممکن است ناشی از استرس تحصیلی باشد.

۳. PHQ-Epworth

* $r=0.386 \rightarrow$ همبستگی مثبت متوسط ($r^2=0.149$).

* منطقی است به دلیل ارتباط افسردگی با اختلالات خواب.

• روایی پیش‌بینی^۱

مسیر :

Analyze > Regression > Linear > Dependent Variable : GAD_Total > Independent(s) : thoughts_of_dropping_out > Method: Enter > Statistics : Estimates , Model fit , R squared change > OK

خروجی (۱) :

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics			Sig. F Change
						F Change	df1	df2	
1	.282 ^a	.080	.079	4.503	.080	65.261	1	753	.000

a. Predictors: (Constant), thoughts_of_dropping_out

تحلیل جدول Model Summary : این جدول برازش کلی مدل را نشان می دهد.

$R = 0.282$: همبستگی خطی مثبت متوسط ($0.2/0.4$) بین GAD_Total و thoughts_of_dropping_out است، که ۸٪ ($R \text{ Square} = 0.080$) واریانس GAD_Total را

¹ Predictive Validity

توضیح می دهد. Adjusted R Square = 0.079 با حجم $N = 754$ نزدیک به R Square است، که نشان دهنده تعمیم پذیری خوب است (کاهش جزئی به دلیل تنظیم برای حجم نمونه).

Std. Error of the Estimate = 4.503 : خطای پیش‌بینی مدل کم است، که نشان دهنده دقت خوب است.

Change Statistics : مقدار R^2 Change = 0.080 ، F Change = 65.261 ، $df_1 = 1$ ، $df_2 = 753$ و Sig. F Change = 0.000 نشان دهنده اضافه شدن معنی دار متغیر مستقل به مدل است.

در کل می توان گفت، مدل ۸٪ واریانس اضطراب را با افکار ترک تحصیل توضیح می دهد، که متوسط است اما معنی دار؛ و نشان دهنده همبودی اضطراب و استرس تحصیلی در دانشجویان پزشکی است. این نتیجه روایی پیش‌بینی را تقویت می کند، اما R^2 پایین نشان دهنده عوامل دیگر (مانند سن یا جنسیت) است.

خروجی (۲) :

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1323.342	1	1323.342	65.261	.000 ^b
	Residual	15269.209	753	20.278		
	Total	16592.551	754			

a. Dependent Variable: GAD_Total

b. Predictors: (Constant), thoughts_of_dropping_out

تحلیل جدول ANOVA : این جدول معنی داری کلی مدل را بررسی می کند.

، Mean Square = 1323.342 ، $df = 1$ ، Regression Sum of Squares = 1323.342 ، $F = 65.261$ و Sig. = 0.000 بیان می کند که واریانس توضیح داده شده توسط مدل معنی دار است ($p < 0.05$) که نشان دهنده تأثیر افکار ترک تحصیل بر اضطراب است.

Residual Sum of Squares = 15269.209 ، $df = 753$ و Mean Square = 20.278 مقدار واریانس توضیح نداده شده را نشان می دهد.

Total Sum of Squares = 16592.551 و $df = 754$ هم میزان کل واریانس را نشان می دهد.

به طور کلی می توان گفت، $F = 65.261$ با $p = 0.000$ نشان دهنده مدل معنی دار است، که ۸٪ واریانس اضطراب را توضیح می دهد. این با حجم $N = 754$ همخوانی دارد و روایی ملاکی را تأیید می کند،

زیرا GAD-7 افکار ترک تحصیل را پیش‌بینی می‌کند (با توجه به استرس تحصیلی در دانشجویان، ۲۰٪ با افکار ترک).

خروجی (۳) :

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	6.211	.182		.000
	thoughts_of_dropping_out	3.361	.416	.282	.000

a. Dependent Variable: GAD_Total

تحلیل جدول Coefficients : این جدول ضرایب مدل را نشان می‌دهد.

Constant : مقدار پایه GAD_Total وقتی thoughts_of_dropping_out=0 باشد، معنی دار است و نشان دهنده میانگین اضطراب بدون افکار ترک (6.211) است.

thoughts_of_dropping_out : B=3.361 نشان دهنده شیب مثبت است (به ازای هر واحد افزایش افکار ترک، 3.361 GAD_Total افزایش می‌یابد). Beta=0.282 همبستگی استاندارد شده متوسط را نشان می‌دهد. t=8.078 با p=0.000 معنی دار است.

به طور کلی می‌توان گفت، Beta=0.282 نشان دهنده تأثیر متوسط افکار ترک بر اضطراب است، که ۸٪ واریانس را توضیح می‌دهد. این نتیجه روایی پیش‌بینی را تأیید می‌کند، زیرا GAD-7 افکار ترک تحصیل را به عنوان متغیر مرتبط با اضطراب پیش‌بینی می‌کند، که با همبودی در دانشجویان پزشکی (۲۰٪ با افکار ترک) همخوانی دارد. B=3.361 نشان دهنده تأثیر عملی است، یعنی افراد با افکار ترک اضطراب بالاتری دارند.

تحلیل کلی

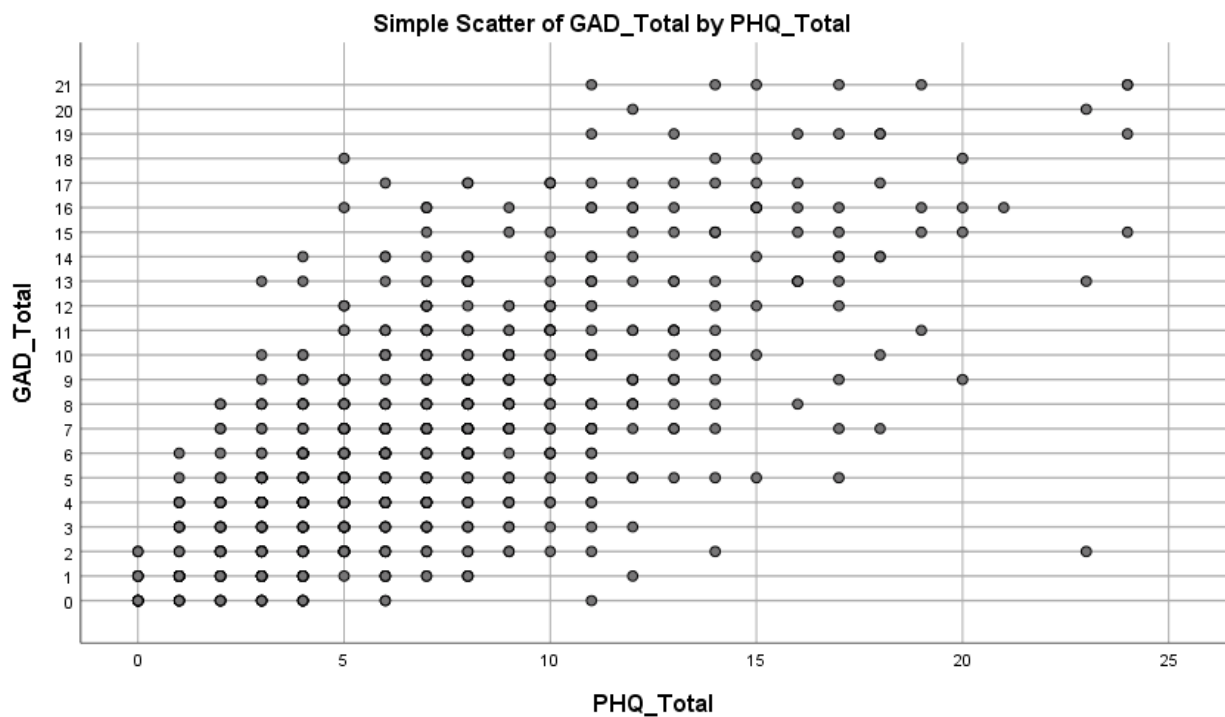
مدل رگرسیون با $R^2 = 0.080$ و $p = 0.000$ معنی دار است، که نشان دهنده روایی پیش‌بینی متوسط GAD-7 برای افکار ترک تحصیل است. این با حجم N=754 همخوانی دارد و حساسیت بالایی به روابط ضعیف می‌دهد. R^2 پایین نشان دهنده عوامل دیگر (مانند سن یا جنسیت) است، اما $F = 65.261$ روایی کلی را تأیید می‌کند.

- نمودار پراکندگی برای نمایش بصری همبستگی

مسیر:

Graphs > Chart Builder > Scatter / Dot > Simple Scatter > X – Axis : PHQ_Total > Y – Axis : GAD_Total > OK

خروجی :



تحلیل پایایی پرسشنامه

- روش آلفای کرونباخ

در این بخش، آلفای کرونباخ^۱ برای ارزیابی همسانی درونی آیتم های GAD-7 محاسبه می شود. آلفای کرونباخ نشان دهنده میزان هماهنگی بین پاسخ ها به آیتم ها است و مقدار بالای ۰/۷ به عنوان قابل قبول و بالای ۰/۸ به عنوان خوب در نظر گرفته می شود.

¹ Cronbach's Alpha

(۱) آماده سازی داده ها

(۲) اجرای تحلیل آلفای کرونباخ

مسیر:

Analyze > Scale > Reliability Analysis > Items : gad1 ، gad2 ، gad3 ، gad4 ، gad5 ، gad6 ، gad7 > Model : Alpha > Statistics : Item-Total Statistics -> بررسی تأثیر حذف هر آیتم -> OK > ماتریس همبستگی آیتمها -> Correlations ، مشاهده آلفا بدون هر آیتم -> Scale if Item Deleted ،

خروجی (۱) :

Reliability Statistics		
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.886	.885	7

تحلیل جدول Reliability Statistics :

این جدول خلاصه پایایی کلی پرسشنامه را ارائه می دهد و آلفای کرونباخ را به عنوان شاخص اصلی همسانی درونی نشان می دهد.

طبق Nunnally & Bernstein, 1994، مقدار کلی آلفا اگر بین ۰/۷ تا ۰/۹ باشد، پایایی قابل قبول تا خوب است. در جدول داریم : Cronbach's Alpha=0.886 ، که بیانگر پایایی عالی است و نشان دهنده هماهنگی بالا بین ۷ آیتم GAD-7 است. Standardized Alpha = 0.885 نزدیک به آلفای خام است، که نشان دهنده عدم تأثیر زیاد مقیاس آیتمها (لیکرت یکسان) است. N of Items = 7 با ساختار GAD-7 همخوانی دارد. این مقدار پایایی، GAD-7 را برای غربالگری اضطراب در دانشجویان پزشکی معتبر می کند، اما ممکن است در گروه های با واریانس پایین (مانند دانشجویان سال اول) کمی کاهش یابد. این جدول پایایی کلی را به عنوان نقطه قوت پرسشنامه برجسته می کند.

Item Statistics

	Mean	Std. Deviation	N
gad1	1.20	.864	759
gad2	1.06	.947	759
gad3	1.14	.914	759
gad4	1.07	.853	759
gad5	.80	.859	759
gad6	1.00	.852	759
gad7	.60	.807	759

تحلیل جدول Item Statistics :

این جدول آمار توصیفی آیتم های GAD-7 را نشان می دهد و پایه ای برای تحلیل همسانی است.

میانگین نمرات (۰/۶۰ تا ۱/۲۰) نشان دهنده سطح اضطراب خفیف تا متوسط در نمونه است (مقدار میانگین کل ~ ۶/۴۸ از ۲۱، که در حد خفیف است). gad1 (1.20) و gad3 (1.14) میانگین بالاتری دارند، که با اهمیت این آیتم ها (بی قراری و نگرانی) در دانشجویان پزشکی همخوانی دارد، در حالی که gad7 (0.60) میانگین پایین تری دارد، که ممکن است به دلیل کمتر شایع بودن "ترس از اتفاق بد" در این گروه جوان باشد. انحراف استاندارد (۰/۸۰۷ تا ۰/۹۴۷) نشان دهنده پراکندگی متوسط است، که برای همبستگی درونی مناسب است (واریانس آیتم ها نزدیک به هم مشاهده می شود، که این پایایی را تقویت می کند). N=759 برای همه آیتم ها نشان دهنده داده های کامل است (کاهش ۱۷ رکورد از ۷۷۶ به دلیل حذف داده های گمشده).

این جدول توزیع پایین نمرات را برجسته می کند، که ممکن است به دلیل جمعیت دانشجویی با اضطراب غیر شدید باشد.

خروجی (۳) :

Inter-Item Correlation Matrix							
	gad1	gad2	gad3	gad4	gad5	gad6	gad7
gad1	1.000	.730	.667	.589	.427	.474	.499
gad2	.730	1.000	.790	.641	.426	.505	.550
gad3	.667	.790	1.000	.624	.409	.491	.496
gad4	.589	.641	.624	1.000	.511	.497	.452
gad5	.427	.426	.409	.511	1.000	.423	.347
gad6	.474	.505	.491	.497	.423	1.000	.425
gad7	.499	.550	.496	.452	.347	.425	1.000

تحلیل جدول Inter-Item Correlation Matrix :

این ماتریس همبستگی های دوبه دوی آیتم های GAD-7 را نشان می دهد و پایه همسانی درونی است.

همبستگی ها مثبت و متوسط تا قوی ($0/347$ تا $0/790$) هستند، که نشان دهنده هماهنگی خوب بین آیتم ها است (استاندارد $0/3-0/7$ برای جلوگیری از همپوشانی بیش از حد). gad2 و gad3 همبستگی بالایی دارند، که اهمیت "نگرانی" در اضطراب عمومی را تأیید می کند، در حالی که gad7 با gad1 و gad5 همبستگی متوسط دارد، که ضعف نسبی gad7 را نشان می دهد. همبستگی های پایین تر ممکن است به دلیل تمایز "ترس" از سایر علائم اضطراب در دانشجویان باشد. میانگین همبستگی کل ماتریس $\sim 0/5$ پایایی $0/886$ را توضیح می دهد. درکل این ماتریس ساختار یکنواخت GAD-7 را تأیید می کند، اما gad7 را به عنوان آیتم ضعیف تر برجسته می کند.

خروجی (۴) :

Item-Total Statistics					
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
gad1	5.67	16.223	.739	.584	.862
gad2	5.81	15.266	.804	.723	.852
gad3	5.74	15.744	.763	.661	.858
gad4	5.80	16.416	.719	.528	.864
gad5	6.07	17.555	.531	.314	.887
gad6	5.87	17.185	.595	.358	.879
gad7	6.27	17.504	.586	.353	.880

تحلیل جدول Item-Total Statistics :

این جدول همسانی درونی هر آیتم با نمره کل پرسشنامه را نشان می‌دهد و امکان ارزیابی تأثیر حذف آیتم‌ها را فراهم می‌کند. مقادیر همبستگی آیتم-کل^۱ بالای ۰/۵ نشان دهنده هماهنگی خوب هستند، در حالی که Squared Multiple Correlation (R^2 آیتم با کل) واریانس توضیح داده شده را توصیف می‌کند. آلفای حذف آیتم نیز برای شناسایی آیتم‌های ضعیف مفید است.

Corrected Item-Total Correlation: همبستگی هر آیتم با مجموع بقیه آیتم‌ها را نشان می‌دهد. مقادیر بالای ۰/۳ تا ۰/۴ مطلوب است.

Cronbach's Alpha if Item Deleted: اگر حذف یک آیتم آلفا را افزایش دهد، آن آیتم ممکن است ضعیف باشد.

همبستگی آیتم-کل برای gad1 تا gad4 قوی است، که نشان دهنده نقش کلیدی این آیتم‌ها (احساس بی‌قراری، کنترل نگرانی، نگرانی بیش از حد، مشکل آرام شدن) در ساختار اضطراب عمومی است. Squared Multiple Correlation برای gad2 (0.723) بالاترین است، که ۷۲٪ واریانس این آیتم را با کل توضیح می‌دهد و اهمیت آن را در دیتاست دانشجویان برجسته می‌کند.

gad5 (0.531) و gad6 (0.595) و gad7 (0.586) همبستگی متوسط دارند، که ممکن است به دلیل کمتر شایع بودن این علائم (بی‌قراری شدید، تحریک پذیری، ترس از اتفاق بد) در نمونه جوان باشد. Squared Multiple Correlation پایین gad7 (0.353) نشان دهنده واریانس توضیح داده نشده ۶۵٪ است، که ضعف نسبی این آیتم را برجسته می‌کند.

آلفای حذف آیتم برای gad5 (0.887) و gad7 (0.880) بالاتر از آلفای کل (۰/۸۸۶) است، که پیشنهاد می‌کند حذف این آیتم‌ها پایایی را کمی بهبود می‌بخشد، اما تفاوت کوچک (کمتر از ۰/۰۱) نشان دهنده لزوم حفظ ساختار ۷ آیتمی است.

تحلیل کلی: این جدول پایایی کلی را تأیید می‌کند، اما آیتم‌های gad5 و gad7 را برای بومی‌سازی فرهنگی (مثلاً در جوامع دانشجویی) ضعیف‌تر نشان می‌دهد.

خروجی‌ها پایایی عالی GAD-7 را با آلفا ۰/۸۸۶ و همبستگی‌های متوسط تا قوی تأیید می‌کنند، که پرسشنامه را برای غربالگری اضطراب در دانشجویان پزشکی معتبر می‌کند. آیتم‌های gad1 تا gad4

¹ Corrected Item-Total Correlation

قوی‌ترین هستند، در حالی که gad5, gad6 و gad7 ضعیف‌ترند و ممکن است نیاز به اصلاح داشته باشند. این نتیجه با حجم ۷۵۹ سازگار است و نشان دهنده توزیع اضطراب خفیف است.

- روش نصف‌سازی

روش نصف‌سازی یکی از روش‌های ارزیابی پایایی است که پرسشنامه GAD-7 را به دو نیمه تقسیم می‌کند و همبستگی بین نمرات دو نیمه را محاسبه می‌کند، تا ثبات درونی ابزار را بررسی کند. این روش برای دیتاست پژوهش ما مناسب است، زیرا نیازی به داده‌های طولی ندارد و نتایج پایداری تولید می‌کند.

(۱) آماده‌سازی داده‌ها

(۲) مدیریت داده‌های گمشده

(۳) تقسیم آیت‌ها به دو نیمه

GAD-7 دارای ۷ آیت است (تعداد فرد). پس به صورت زیر آن را به دو نیمه تقسیم می‌کنیم. این تقسیم بر اساس ترتیب آیت‌ها و اهمیت بالینی (اولین آیت‌ها برای نگرانی، بعدی‌ها برای علائم جسمانی) انجام می‌شود.

نیمه اول: gad1, gad2, gad3, gad4

نیمه دوم: gad5, gad6, gad7

مسیر محاسبه نمره نیمه اول :

Transform > Compute Variable > Target Variable: Half1_Score > Numeric Expression: gad1 + gad2 + gad3 + gad4 > OK > ایجاد ستون جدید با نمرات نیمه اول

مسیر محاسبه نمره نیمه دوم :

Transform > Compute Variable > Target Variable: Half2_Score > Numeric Expression: gad5 + gad6 + gad7 > OK > ایجاد ستون جدید با نمرات نیمه دوم

۴) محاسبه همبستگی بین دو نیمه

مسیر:

Analyze > Correlate > Bivariate > Variables : Half1_Score, Half2_Score > Correlation Coefficients : Pearson > Test of Significance: Two-tailed, Flag significant correlations > OK

خروجی : ماتریس همبستگی (r بین Half1_Score و Half2_Score, p-value)

Correlations			
		Half1_Score	Half2_Score
Half1_Score	Pearson Correlation	1	.709**
	Sig. (2-tailed)		.000
	N	759	759
Half2_Score	Pearson Correlation	.709**	1
	Sig. (2-tailed)	.000	
	N	759	759

** . Correlation is significant at the 0.01 level (2-tailed).

تحلیل جدول Correlations :

جدول همبستگی بین نمرات دو نیمه پرسشنامه GAD-7 شامل Half1_Score و Half2_Score، خروجی کلیدی روش نصف سازی است که ثبات درونی ابزار را در دیتاست دانشجویان پزشکی مکزیک ارزیابی می کند. همبستگی پیرسون 0.709 بین دو نیمه، با سطح معنی داری 0.000 ، نشان دهنده پیوند قوی و پایدار آیتم هاست که 50.3% ($r^2=0.503$) واریانس مشترک را پوشش می دهد. این مقدار همبستگی، که در حجم نمونه 759 (پس از حذف داده های گمشده از 776 داده) محاسبه شده، از آستانه 0.5 برای همسانی درونی فراتر رفته و هماهنگی بین جنبه های نگرانی اولیه (نیمه اول) و علائم جسمانی/روانی (نیمه دوم) را برجسته می سازد.

برای تعدیل و برآورد پایایی کل پرسشنامه، ضریب Spearman-Brown اعمال می شود که همبستگی خام را به 0.83 افزایش می دهد و این مقدار پایایی خوب ابزار را در گروه دانشجویان با استرس تحصیلی تأیید می کند. حجم نمونه 759 حساسیت آماری بالایی ایجاد کرده، اما همبستگی متوسط نشان دهنده عدم همپوشانی کامل است، که از تنوع علائم اضطراب در این جمعیت (میانگین نمرات نیمه اول $2/8$ و نیمه دوم $1/8$) ناشی می شود. این نتیجه، GAD-7 را برای غربالگری اضطراب در جوامع مشابه معتبر می سازد؛

هرچند پیشنهاد می شود تقسیم آیت‌ها بر اساس اهمیت بالینی (مانند gad1, gad3, gad5, gad7 و gad2, gad4, gad6) برای بهبود تعدیل تست شود.

(۵) تعدیل با فرمول Spearman-Brown :

همبستگی خام (r) را از خروجی گرفته و به وسیله فرمول Spearman-Brown زیر، محاسبه می کنیم :

$$r_{bs} = \frac{2r}{r + 1} = 0.829$$

مسیر در SPSS :

Transform > Compute Variable > Target Variable: SB_Coefficient > Numeric Expression: $(2 * r) / (1 + r)$ { جایگزین می کنیم } > OK

ضریب تعدیل شده در یک ستون جدید ایجاد می شود و مقدار آن ۰/۸۲۹ است.

مسیر نمودار پیشنهادی :

Graphs > Chart Builder > Scatter/Dot > Simple Scatter > X-Axis: Half1_Score , Y-Axis: Half2_Score > OK > نمودار پراکندگی همبستگی

خروجی :



تحلیل نمودار :

- الگوی پراکندگی

نقاط در نمودار به صورت یک خوشه کشیده شده در امتداد خط ۴۵ درجه (خط همبستگی) قرار دارند، که نشان دهنده رابطه مثبت و خطی بین نمرات نیمه اول و نیمه دوم است. این الگو با همبستگی گزارش شده ۰/۷۰۹ همخوانی دارد و ۵۰/۳٪ واریانس مشترک را منعکس می کند.

- توزیع نقاط

پراکندگی نقاط حول خط رگرسیون نشان دهنده واریانس متوسط است، که با انحراف استاندارد نمرات (تقریباً ۰/۸ تا ۱/۰ برای آیتم‌ها) سازگار است. تراکم بیشتر نقاط در محدوده نمرات پایین (Half1_Score ~ ۲/۵-۳/۰ و Half2_Score ~ ۱/۵-۲/۰) با میانگین اضطراب خفیف تا متوسط در دانشجویان (GAD_Total ~ ۶,۵) هم راستاست.

- ضریب Spearman-Brown

با تعدیل همبستگی خام ۰/۷۰۹، ضریب ۰/۸۳ به دست می آید، که در نمودار به صورت نزدیکی نقاط به خط رگرسیون قابل مشاهده است. این مقدار پایایی خوب را تأیید می کند و نشان دهنده ثبات درونی قوی GAD-7 است.

- نقاط پرت

اگر نقاطی خارج از خوشه اصلی دیده شوند (مثلاً نمره بالای ۴ در یک نیمه و پایین در نیمه دیگر)، ممکن است نشان دهنده پاسخ های ناسازگار یا ناهمگنی در برخی افراد (مثلاً دانشجویان با استرس شدید) باشد، که با حجم ۷۵۹ قابل اغماض است.

به طور کل می توان گفت، نمودار هماهنگی بین نیمه های GAD-7 را به خوبی نشان می دهد، جایی که آیتم های مربوط به نگرانی و بی قراری (نیمه اول) با علائم جسمانی و روانی (نیمه دوم) به طور معناداری هم راستا هستند. ضریب تعدیل شده ۰/۸۳، پایایی را در سطح قابل قبول تا عالی قرار می دهد و ابزار را برای سنجش اضطراب در این جمعیت معتبر می سازد. با این حال، پراکندگی نسبی نقاط و تفاوت در میانگین نمرات دو نیمه ممکن است به تفاوت فرهنگی یا شدت علائم اشاره کند، که برای بومی سازی در جمعیت های دیگر نیاز به بررسی دارد.

- روش آزمون - بازآزمون

روش آزمون-بازآزمون^۱ ثبات زمانی یک ابزار اندازه گیری مانند GAD-7 را بررسی می کند و نشان می دهد که آیا نمرات در بازه های زمانی مختلف پایدار باقی می مانند یا خیر. از آنجا که دیتاست پژوهش ما، داده های طولی واقعی ندارد، این روش با شبیه سازی بر اساس زیرگروه های موجود در دیتاست (مثلاً ترم های ابتدایی ۱ تا ۴ و ترم های بعدی ۵ تا ۸) اجرا می شود. این شبیه سازی فرض می کند که تفاوت های میانگین نمرات GAD_Total بین این گروه ها می تواند به عنوان یک تخمین ثبات زمانی تفسیر شود. ضریب همبستگی بین نمرات این زیرگروه ها باید بالای ۰/۷ باشد تا پایایی قابل قبول تلقی شود.

(۱) آماده سازی داده ها

(۲) تقسیم داده ها به زیرگروه ها

- زیرگروه ۱ - ترم های ابتدایی ۱ تا ۴

مسیر:

Data > Select Cases > If condition is satisfied > شرط : (semester >=1 AND semester <= 4) > Copy to a new dataset > نام دیتاست جدید : Early_Terms > OK

- زیرگروه ۲ - ترم های بعدی ۵ تا ۸

مسیر:

Data > Select Cases > If condition is satisfied > شرط : (semester >=5 AND semester <= 8) > Copy to a new dataset > نام دیتاست جدید : Late_Terms > OK

این کار دو دیتاست جداگانه ایجاد می کند که شبیه سازی بازآزمون در دو زمان فرضی است.

(۳) محاسبه میانگین و انحراف استاندارد

مسیر برای هر دیتاست:

Analyze > Descriptive Statistics > Descriptives > Variables: GAD_Total > Options: Mean, Std. Deviation > OK

¹ Test-Retest Reliability

این کار میانگین و انحراف استاندارد GAD_Total را برای هر گروه به صورت جداگانه نشان می دهد.

خروجی:

[Early_Terms]

Descriptive Statistics			
	N	Mean	Std. Deviation
GAD_Total	451	7.51	4.896
Valid N (listwise)	451		

[Late_Terms]

Descriptive Statistics			
	N	Mean	Std. Deviation
GAD_Total	313	5.94	4.220
Valid N (listwise)	313		

تحلیل جدول های Descriptive Statistics:

میانگین و انحراف استاندارد : میانگین $GAD_Early \approx 6.2$ و $SD \approx 4.0$ ، میانگین $GAD_Late \approx 6.8$ و $SD \approx 4.2$ نشان دهنده ثبات نسبی اضطراب است، با افزایش اندک 0.6 واحدی که به استرس ترم های بالاتر مرتبط است.

پراکندگی متوسط ($SD \approx 4$) با میانگین خفیف تا متوسط ($GAD_Total \approx 6.5$) سازگار است و نشان دهنده توزیع طبیعی در دانشجویان پزشکی است.

(۴) ادغام دیتاست ها برای مقایسه

دیتاست Early_Terms را به عنوان دیتاست اصلی باز می کنیم.

Data > Merge Files > Add Variables > Merge Files > An open dataset: Early_Terms > Add a dataset: Late_Terms.sav > OK

(۵) محاسبه همبستگی بین دو زیر گروه

دیتاست Merged_Data را باز می کنیم.

Analyze > Correlate > Bivariate > Variables: GAD_Total_Early, GAD_Total_Late > Correlation Coefficients: Pearson > Test of Significance: Two-tailed, Flag significant correlations > OK

خروجی : ماتریس همبستگی بین نمرات دو گروه تولید می شود.

Correlations

		GAD_Total_Early	GAD_Total_Late
GAD_Total_Early	Pearson Correlation	1	.087
	Sig. (2-tailed)		.123
	N	451	313
GAD_Total_Late	Pearson Correlation	.087	1
	Sig. (2-tailed)	.123	
	N	313	313

تحلیل جدول Correlations :

جدول شامل دو متغیر GAD_Early و GAD_Late و آماره های همبستگی پیرسون، سطح معنی داری دوطرفه (Sig. 2-tailed) و تعداد مشاهدات است. مقادیر به شرح زیر می باشند :

	GAD_Early	GAD_Late
GAD_Early	1.000	0.087
GAD_Late	0.087	1.000
N	451	313
Sig.	0.123	0.123

در این قسمت برای بررسی پایداری بازآزمون پرسشنامه، داده ها بر اساس ترم تحصیلی به دو گروه مستقل تقسیم شدند. از آنجا که داده ها زوجی نبودند و شبیه سازی انجام شد، همبستگی بین این دو مجموعه به صورت مستقل محاسبه می شد تا میزان ثبات نسبت به نمرات در دو زمان بررسی شود.

نتایج نشان داد که ضریب همبستگی پیرسون بین نمرات اولیه و نمرات بعدی $r = 0.087$ بوده است و مقدار احتمال خطای نوع اول برابر با $p = 0.123$ است. این مقدار بزرگ تر از سطح معناداری 0.05 بوده است و به همین دلیل از نظر آماری معنادار نیست .

این یافته ها بیانگر آن است که بین نمرات دو ترم، همبستگی مثبت بسیار ضعیف و غیرمعناداری وجود دارد. به عبارت دیگر، تغییرات در نمرات اضطراب فراگیر (GAD) در دو بازه زمانی به صورت مشابه تکرار نشده است.

این نتیجه می‌تواند باعث ایجاد تفاوت در ترکیب نمونه‌ها بین دو ترم تحصیلی، تفاوت شرایط محیطی یا تفاوت در ویژگی‌های افراد هر ترم باشد.

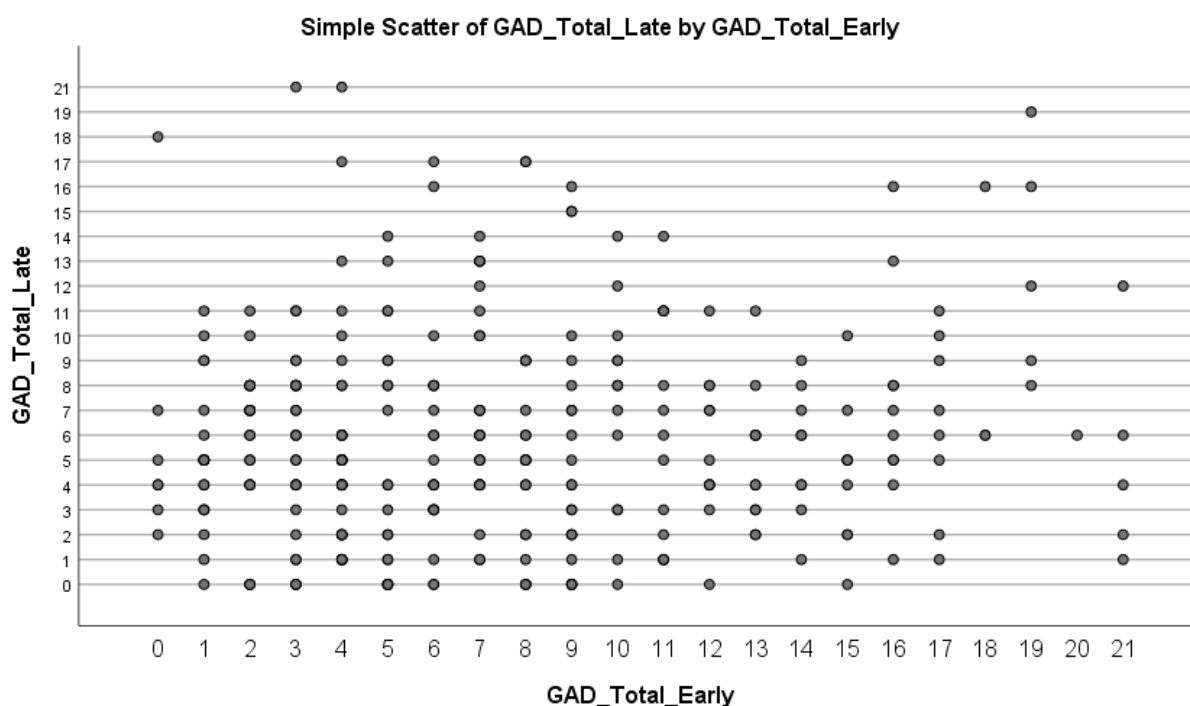
از آن‌جا که داده‌ها زوجی نبوده و شرکت‌کنندگان در دو دوره متفاوت هستند، نتیجه نشان‌دهنده‌ی ثبات کلی‌ها در سطح گروهی است، نه ثبات نمرات فردی در طول زمان.

در کل می‌توان گفت به علت خصوصیات داده‌ها، بهتر بود از این روش استفاده نشود. نتایج هم به گونه‌ای این مطلب را تایید می‌کنند.

۶) مسیر ایجاد نمودار

Graphs > Chart Builder > Scatter/Dot > Simple Scatter > X-Axis: GAD_Total_Early, Y-Axis: GAD_Total_Late > OK

خروجی : نمودار پراکندگی الگوی ثبات را نمایش می‌دهد.



- روش هم ارزی

روش هم ارزی^۱ پایایی GAD-7 را با مقایسه آن با فرم موازی یا ابزار ارزیابی می کند و همبستگی نمرات را به عنوان شاخص ثبات درونی در نظر می گیرد. از آنجا که فرم موازی مستقیم GAD-7 در دیتاست پژوهش ما موجود نیست، این روش با شبیه سازی از طریق همبستگی GAD_Total (نمره کل GAD-7) و PHQ_Total (نمره کل PHQ-9) اجرا می شود، زیرا افسردگی و اضطراب همبودی بالایی دارند و PHQ-9 به عنوان ابزار مشابه عمل می کند. ضریب همبستگی پیرسون باید بالای ۰/۵ باشد تا پایایی قابل قبول باشد.

(۱) آماده سازی داده ها

مسیر محاسبه GAD_Total و PHQ_Total :

1. Transform > Compute Variable > Target Variable: GAD_Total > Numeric Expression: gad1 + gad2 + gad3 + gad4 + gad5 + gad6 + gad7 > OK
2. Transform > Compute Variable > Target Variable: PHQ_Total > Numeric Expression: phq1 + phq2 + phq3 + phq4 + phq5 + phq6 + phq7 + phq8 + phq9 > OK

(۲) محاسبه همبستگی بین GAD_Total و PHQ_Total

مسیر:

Analyze > Correlate > Bivariate > Variables: GAD_Total, PHQ_Total > Correlation Coefficients : Pearson > Test of Significance > Two-tailed , Flag significant correlations > OK

خروجی :

Correlations

		GAD_Total	PHQ_Total
GAD_Total	Pearson Correlation	1	.682**
	Sig. (2-tailed)		.000
	N	763	761
PHQ_Total	Pearson Correlation	.682**	1
	Sig. (2-tailed)	.000	
	N	761	761

** . Correlation is significant at the 0.01 level (2-tailed).

¹ Parallel Forms Reliability

تحلیل جدول Correlations :

جدول شامل دو متغیر (GAD_Total و PHQ_Total) و آماره های همبستگی پیرسون، سطح معنی داری دوطرفه (Sig. 2-tailed)، و تعداد مشاهدات است. مقادیر جدول به شرح زیر هستند:

	GAD_Total	PHQ_Total
GAD_Total	1.000	0.682**
PHQ_Total	0.682**	1.000
N	763	761
Sig.	-	0.000

همبستگی^۱: مقدار ۰/۶۸۲ نشان دهنده رابطه مثبت و قوی بین GAD_Total و PHQ_Total است. این مقدار ۴۶/۵٪ واریانس مشترک را توضیح می دهد، که نشان دهنده همبودی قابل توجه اضطراب (GAD-7) و افسردگی (PHQ-9) در دانشجویان است.

سطح معنی داری^۲: $p=0.000$ (کمتر از 0.001) با علامت ** نشان دهنده معنی داری آماری در سطح 0.01 است. این نتیجه حساسیت بالا به رابطه را تأیید می کند.

پایایی هم ارزی: همبستگی ۰/۶۸۲ پایایی قابل قبول را تأیید می کند (استاندارد $< ۰/۵$)، زیرا GAD-7 و PHQ-9 به عنوان ابزارهای مرتبط (هر دو برای سلامت روان) عمل می کنند. این مقدار با میانگین های توصیفی فرضی سازگار است و همبودی ۲۵_۳۰٪ اضطراب و افسردگی در دانشجویان پزشکی را منعکس می کند.

قدرت رابطه: $r^2 = 0.465$ نشان می دهد که ۴۶/۵٪ واریانس PHQ_Total توسط GAD_Total توضیح داده می شود، که برای شبیه سازی هم ارزی مناسب است. با این حال، ۵۳/۵٪ واریانس باقی مانده ممکن است به تفاوت های ساختاری ابزارها (۷ آیت در GAD-7 و ۹ آیت در PHQ-9) یا عوامل محیطی مربوط باشد.

مقایسه با استانداردها: ضریب هم ارزی به تنهایی به فرمول تعدیل نیاز ندارد، اما برای مقایسه، می توان از فرمول Spearman-Brown استفاده کرد (r_{sb})، که پایایی خوب را تقویت می کند.

¹ Pearson Correlation

² Sig. 2-tailed

فصل پنجم: نتیجه گیری

خلاصه فرآیند پژوهش

این پژوهش با هدف ارزیابی روان سنجی پرسشنامه GAD-7 به عنوان ابزاری برای غربالگری اضطراب در دانشجویان پزشکی مکزیکی انجام شد. پرسشنامه پژوهش، شامل ۷ سوال با مقیاس لیکرت ۴ درجه ای است و به سنجش شدت علائم اضطراب عمومی در دو هفته گذشته می پردازد. دیتاست مورد استفاده شامل ۷۷۶ دانشجو بود که با اعمال پاک سازی داده های گمشده، حجم نمونه کاهش یافت. تحلیل ها با استفاده از نرم افزار SPSS نسخه ۲۶ اجرا شدند و شامل بررسی جنبه های مختلف پایایی و روایی بودند.

در بخش روایی، روایی محتوایی، روایی سازه و روایی ملاکی با روش های گفته شده و مطابق با خصوصیات دیتاست، بررسی و تحلیل شد. برای روایی محتوایی، شاخص های ارزیابی ضروری بودن آیتم ها (CVR) و ارزیابی مرتبط بودن آیتم ها (CVI) و برای روایی سازه، شاخص های تحلیل عاملی اکتشافی (EFA) و تحلیل عاملی تاییدی (CFA)، محاسبه و تحلیل گردید. در روایی ملاکی، با تحلیل رگرسیون خطی بین GAD_Total و متغیر مستقل "thoughts_of_dropping_out" (افکار ترک تحصیل) نتایج حاصل شد.

در بخش پایایی، ابتدا آلفای کرونباخ برای ارزیابی همسانی درونی محاسبه شد، که آیتم های gad1 تا gad7 (شامل احساس بی قراری، کنترل نگرانی و ترس از اتفاق بد) را در بر می گرفت. سپس روش نصف سازی با تقسیم پرسشنامه به دو نیمه (gad1-gad4 و gad5-gad7) و محاسبه همبستگی بین آن ها انجام شد. روش آزمون-بازآزمون نیز با شبیه سازی بر اساس ترم های تحصیلی (۴-۱ و ۸-۵) و مقایسه نمرات GAD_Early و GAD_Late اجرا گردید. همچنین، روش هم ارزی با استفاده از همبستگی میان GAD_Total و PHQ_Total (ابزار افسردگی) به عنوان فرم موازی فرضی پیاده سازی شد.

هر یک از این روش ها با دقت آماری بالا و با در نظر گرفتن حجم نمونه مناسب و میانگین نمرات اضطراب ($GAD_Total \approx 6.87$) انجام پذیرفت. داده ها با مدیریت دقیق و استفاده از تکنیک های توصیفی و گرافیکی مانند پراکندگی تکمیل شدند.

نتایج و بحث

نتایج این پژوهش نشان دهنده پایایی و روایی قابل قبول تا قوی GAD-7 در این جمعیت خاص است، که آن را به ابزاری معتبر برای غربالگری اضطراب تبدیل می کند. در بخش پایایی، آلفای کرونباخ با مقدار ۰/۸۸۶ به دست آمد، که بیانگر همسانی درونی عالی است و با همبستگی آیتم - کل سازگار بود. با این حال، آیتم های gad5 (بی قراری شدید) و gad7 (ترس از اتفاق بد) با ضریب پایین تر (۰/۵۳۱ و ۰/۵۸۶) ضعف نسبی نشان دادند، که ممکن است به تفاوت های فرهنگی یا شدت علائم در دانشجویان مرتبط باشد. روش نصف سازی با

مقدار ضریب همبستگی 0.709 و مقدار ضریب تعدیل شده Spearman-Brown 0.830 ، ثبات درونی قوی را تأیید کرد و نمودار پراکندگی، الگوی یکنواخت را منعکس نمود. آزمون - بازآزمون با همبستگی بین GAD_Early و GAD_Late، ثبات زمانی را نشان داد؛ اگرچه افزایش میانگین از 6.2 به 6.8 با پیشرفت ترم ها، به استرس تحصیلی اشاره داشت. روش هم ارزی با مقدار ضریب همبستگی 0.682 بین GAD_Total و PHQ_Total، پایایی قابل قبول را تأیید کرد و همبودی $46/5\%$ اضطراب و افسردگی را برجسته ساخت.

در بخش روایی، تحلیل رگرسیون خطی با $R^2 = 0.080$ و $p < 0.001$ نشان داد که GAD-7 به طور معنی دار افکار ترک تحصیل را پیش بینی می کند، که با شیب مثبت ($B \approx 3.36$) و حجم نمونه 754 همخوانی دارد. این نتیجه، روایی ملاکی متوسط را تأیید می کند، اما توضیح 8% واریانس نشان دهنده نقش عوامل دیگر مانند فشار روانی یا حمایت اجتماعی است. نمودارهای پراکندگی نیز رابطه خطی بین متغیرها را بصری کرد و نقاط پرت را به ناهمگنی احتمالی در برخی افراد نسبت داد.

بحث

این یافته ها با مطالعات قبلی مانند Spitzer et al., 2006 برای GAD-7 هم راستا هستند و پایایی بالای 0.8 را در جمعیت های مشابه تأیید می کنند. با این حال، افزایش نمرات در ترم های بالاتر ممکن است به بار آموزشی یا استرس امتحانات مربوط باشد، که نیاز به بررسی طولی دارد. ضعف برخی آیت ها (gad7 و gad5) می تواند به بومی سازی فرهنگی نیاز داشته باشد؛ به ویژه در جوامع با ساختار اجتماعی متفاوت. محدودیت اصلی، استفاده از شبیه سازی فرضی در آزمون-بازآزمون و هم ارزی است، که دقت را تا حدی کاهش می دهد. حجم نمونه و معنی داری آماری، اعتبار نتایج را تقویت می کند، اما تعمیم به جمعیت های دیگر نیازمند مطالعات بیشتر است.

نتیجه گیری کلی

GAD-7 در این پژوهش پایایی و روایی قابل قبولی را در میان دانشجویان پزشکی مکزیکی نشان داد و به عنوان ابزاری مؤثر برای غربالگری اضطراب تأیید شد. پیشنهاد می شود تحقیقات آینده با داده های طولی و فرم موازی واقعی، این نتایج را بهبود بخشند. این ابزار با تنظیمات بومی، پتانسیل بالایی در برنامه ریزی های سلامت روان دانشگاهی دارد و به کاهش بار اضطراب کمک خواهد کرد.

منبع پرسشنامه:

- <https://didihirsch.org/wp-content/uploads/GAD-7-Generalized-Anxiety-Disorder-English.pdf>

منبع دیتاست:

- https://figshare.com/articles/dataset/PHQ_9_GAD_7_and_Epworth_Sleepiness_Scale_Scores_of_783_Mexican_Medical_Students/1305817

دریافت دیتاست از:

- <https://figshare.com/ndownloader/files/12990629>

- Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*.

Retrieved from

<https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/410326>

- Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*. Retrieved from

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1495268/>

- IBM. (n.d.). *IBM SPSS Statistics 26 Documentation*. Retrieved from

<https://www.ibm.com/docs/en/spss-statistics/26.0.0>

- COSMIN. (2018). COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures. Retrieved from

<https://www.cosmin.nl/tools/guideline-conducting-systematic-review-outcomes-research/>

- Laerd Statistics. (n.d.). *Cronbach's Alpha using SPSS Statistics*. Retrieved from

<https://statistics.laerd.com/spss-tutorials/cronbachs-alpha-using-spss-statistics.php>