پژوهش در علوم ورزشی
سال سوم، شماره ششم، بهار ۱۳۸۴

پایایی و عینیت سنجی برای ابزار اندازه گیری در تربیت بدنی و ورزش

دکتر علی محمد امیرتاش
دانشیار دانشگاه تربیت معلم تهران

چکیده

خطای اندازه گیری به طور عام در همه رشته‌های علوم مشاهده می‌شود. روشهای ارزشیابی خطای اندازه‌گیری از سوی روانشناسان توسعه یافته است و کاربرد امروزه در بیشتر علوم به چشم می‌خورد. یکی از ویژگی‌های مهم ابزار اندازه‌گیری برای کاهش خطا در جمع آوری اطلاعات خام در تربیت بدنی و سایر علوم، پایایی آماری آنهاست. پایایی از نظر علمی، نسبت واریانس بدون خطا بر کل واریانس یک دسته نمرات است و معمولاً با ضریب همبستگی محاسبه می‌شود. در مقاله حاضر دو نوع پایایی مورد بررسی قرار گرفته است:

۱- پایایی زمانی آزمون در کاربردهای متعدد.

۲- ثبات درونی.

نوع اول از طریق محاسبه ضریب همبستگی بین تکرارهای آزمون با روشهای همبستگی و یا تحلیل واریانس در نمره‌های تکراری و نوع دوم توسط روشهای دو نیمه کردن کودر ریچاردسون و بالاخره ضریب آلفا ارزشیابی می‌شود.

عینیت نیز که پایایی بین آزمایشگران هم‌تراز است و رابطهٔ نزدیکی با پایایی دارد موضوع دیگر مقاله است. این مبحث نشان می‌دهد که برای محاسبه عینیت کم و بیش از همان روشهای محاسباتی پایایی استفاده می‌شود.

کلید واژه‌ها: پایایی، عینیت، روایی، ضریب آلفا، خطای اندازه‌گیری.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

مقدمه

نخستین گام در اندازه‌گیری شناخت نوع متغیرهای آن است. ابزار اندازه‌گیری این متغیرها یا از پیش ساخته شده و موجود است و یا پژوهشگر آن را اختراع می‌کند. انواع گوناگون متغیرهای تحقیق، برحسب تعریف محقق، با یکی از چهار مقیاس از دقیق به کم دقت تر صفر مأخذی[1]، فاصله‌ای[2]، رتبه‌ای[3] و یا اسمی[4] اندازه‌گیری می‌شوند. برای سهولت محاسبات آماری، مقیاس‌های دقیق‌تر را می‌توان به مقیاس‌های بعد از آن تبدیل کرد.

به عنوان نمونه، متغیری که دارای طبیعت صفر مأخذی است می‌تواند بر حسب نیاز به صورتی تعریف شود که مقیاس آن به شکل یکی از انواع سه گانه درآید (۵).

اهمیت تعریف دقیق متغیر در تعیین نوع روش آماری نهفته است که برای تجزیه و تحلیل اطلاعات خام حاصل از آن به کار می‌رود. به تبع نوع متغیر و تعریف عملیاتی[5]، که از آن شده، ابزارهای اندازه‌گیری نیز به صورتی طراحی و ساخته شده‌اند که اطلاعات خام حاصله از آنها یا کمّی یا رتبه‌ای و یا اسمی است تا با روشهای آماری که تنها با این سه نوع مقیاس سازگاری دارند قابل تجزیه و تحلیل و نتیجه گیری شوند[6].

پرسشنامه مصاحبه و مشاهده از انواع ابزار اندازه‌گیری هستند و بخش‌های اصلی که در این ابزار به کار گرفته می‌شود به ترتیب عبارت‌اند از: فکر و قلم، زبان و تکلم و بالاخره بینایی و دقت آن. هریک از این ابزار کاربرد، ویژگی و نقاط قوت و ضعف خود را دارد. مثلاً تنها راه کسب اطلاعات از افراد نابینا، مصاحبه و یا مشاهده رفتار آنهاست.

بخش مهمی از کیفیت و ارزش یافته‌های علمی در تربیت بدنی و ورزش در گرو دقت ابزار اندازه‌گیری و کم خطا بودن روشهای کاربرد در جمع‌آوری اطلاعات خام است. دقت ابزار اندازه‌گیری به درجه پایایی (ثبات اندازه‌گیری یا ثبات درونی)، عینیت و روایی

1. Ratio Scale
2. Interval Scale
3. Ordinal Scale
4. Nominal Scale
5. Operational Definoition

۶. برای کسب اطلاعات بیشتر به کتاب های آمار مقدماتی، اندازه گیری و ارزشیابی و یاجزوه های درسی نویسنده در این زمینه ها در دانشکده تربیت بدنی و علوم ورزشی دانشگاه تربیت معلم تهران مراجعه شود.

(اعتبار) آزمون‌ها از یک سو، و کم خطابودن روش اندازه‌گیری از سوی دیگر، و به یکسان و هماهنگی در جمع آوری اطلاعات بستگی دارد.

درجه پایایی ابزار اندازه‌گیری از جمله نکات مهمی است که در هر گزارش پژوهشی قید می‌شود. روشهای آماری متعددی برای تعیین میزان پایایی یک ابزار اندازه‌گیری، و یا درجهٔ توافق بین دو یا چند آزمایشگر (قاضی یا مصحح که عینیت نامیده می‌شود) در ارزش‌گذاری آزمودنی‌های معین روی یک متغیر وجود دارند که در تحقیقات کاربرد گسترده‌ای دارند. به طور مثال، با این روشها می‌توان درجه همخوانی و توافق بین دو یا چند آزمایشگر را در طبقه‌بندی یا امتیازدهی گروهی آزمودنی‌ها، میزان همسانی و ثبات بین سؤالات یک پرسشنامه (یا زیر مجموعه‌های یک آزمون مرکب مانند ایفرد یا آزمون کانادایی آمادگی‌های عمومی بدنی) یا درجه مشابهت بین پرسشنامه‌ها و یا ابزار اندازه‌گیری متفاوت را که برای اندازه‌گیری یک متغیر واحد ساخته شده‌اند، تعیین کرد. البته، انتخاب نوع روش آماری بستگی به نوع مقیاس اندازه گیری متغیر مورد نظر دارد، که می‌تواند از نوع کمّی و یا کیفی باشد.

به دلیل محدودیت در تعداد صفحات در مقاله حاضر، تنها مباحث پایایی و عینیت ابزار اندازه‌گیری به گونه عام مطرح می‌شود. در تعمیم مبانی نظری و راهکارهای عملی محاسباتی به ابزار اندازه‌گیری در تربیت بدنی و ورزش کم و بیش از همین قوانین صحبت می‌شود.

ضریب کاپا[1]، متداولترین شاخص آماری برای تعیین میزان توافق بین دو یا چند آزمایشگر، در خصوص متغیرهای کیفی است که از سوی کوهن[2] (۱۹۶۰) (۱۰) معرفی شده است. در سوی دیگر ضریب همبستگی پیرسون[3] شاخص آماری به کار رفته برای توافق بین دو آزمایشگر یا پایایی ابزار اندازه‌گیری، در دو مرتبه اندازه‌گیری، هنگام کار با متغیرهای کمّی است (۱۶). روش آماری متداول برای کسب توافق یا پایایی بین بیشتر از دو یا چند آزمایشگر و یا دسته نمره‌های هم بسته (اندازه‌گیری مکرر روی یک گروه)،

1. Kappa Coefficient
2. Cohen

برای اطلاعات کمّی، همبستگی درون گروهی نام دارد که توسط ابل[1] (۱۹۷۶) (۱۱) ارائه گردیده است.

کرونباخ[2] (۱۹۵۱) (۹) پژوهشگر دیگری است که روش آماری آلفا را برای محاسبه ثبات درونی سؤالات یک پرسشنامه، یا زیر مجموعه‌های یک آزمون مرکب، ساخته و گزارش داده است. علاوه بر محققان یاد شده، دانشمندان دیگری نیز درمورد پایایی و عینیت ابزار اندازه‌گیری کار کرده‌اند. که می‌توان از اسپیرمن[3] و براون[4] (۱۹، ۷) نام برد. از بین کارهای این دو دانشمند، دو نمونه در این نوشته آورده می‌شود که یکی از آنها مربوط به برآورد تعداد سؤالات، یا تعداد آزمایشگرها، برای دستیابی به پایایی مطلوب و دیگری، محاسبه ضریب لازم برای تعیین آن است.

در پایان مبحث فعلی روشی برای اصلاح میزان همبستگی بین دو متغیر که با استفاده از ابزارهایی با پایایی ناکافی اندازه‌گیری شده‌اند آورده می‌شود[5]. از این طریق می‌توان میزان مشابهت دو ابزار مختلف که یک متغیر مشترک را اندازه‌گیری می‌کنند، (محاسبه اعتبار معیاری ابزار اندازه‌گیری) نیز ارزشیابی کرد.

سرانجام باید از پژوهشگران آشنا در زمینه اعتباریابی ابزار اندازه‌گیری، کودر و ریچاردسون[6] نام برد که فرمول‌هایی را برای محاسبه پایایی ابزار اندازه‌گیری به نام خودشان تهیه و ارائه داده‌اند(۱۵). مباحث یادشده از نظر مفهوم، محاسبات و اندازه-گیری، به ترتیب ارائه و معرفی می‌شوند.

با توجه به اینکه کتاب‌های آماری روشهای مختلفی را برای محاسبات به کار می‌برند در مقاله حاضر در برخی موارد مباحث با ذکر مثال عددی ارائه شده‌اند.

---

۳. به ضریب همبستگی پیرسون به عنوان شاخص پایایی، انتقادهایی وارداست که بحث آن را می‌توان در کتاب دیگر (منبع۱۶) پیدا کرد.

1. Ebel
2. Cronbach
3. Spearman
4. Brown
5. Correction for-Attenuation
6. Kuder-Richardson

پایایی و عینیت

مبانی نظری

اگر هیچ گونه خطایی در اندازه‌گیری وجود نداشته باشد، نمره یا رکورد ورزشی واقعی فرد به دست می‌آید، اما در صورت اشکال در ابزار اندازه‌گیری، نمره‌ای که برای هر آزمودنی مشاهده می‌کنیم با توانایی ورزشی یا ویژگی واقعی اختلاف پیدا می‌کند. تفاوت بین نمره واقعی و نمره مشاهده شده نتیجه مقدار خطای اندازه‌گیری است. در این صورت، رکورد ورزشی یا نمره مشاهده شده (X) از دو بخش ترکیب شده است: بخش واقعی (T) و بخش خطای اندازه‌گیری(e): (X=T+e) به این ترتیب، خطای اندازه‌گیری عبارت است از انحراف نامنظم یا تصادفی نمره مشاهده شده آزمودنی از نمره مشاهده شده‌ای که به طور نظری انتظار آن می‌رود (e=x-T). خطاهای منظم در نظریه اندازه-گیری، خطای اندازه‌گیری محسوب نمی‌شود؛ برای مثال چنانچه استقامت عضلانی یک ورزشکار به طور ثابت در هر آزمون ۱۵ امتیاز کمتر از مقدار واقعی آن برآورد شود، ۱۵ امتیاز اختلاف یک خطای منظم است و در نظریه اندازه‌گیری به عنوان خطای اندازه-گیری بررسی نمی‌شود.

نظریه نمونه‌گیری‌های تصادفی هم تعداد (نظریه حدّ مرکزی) به ما می‌گوید که توزیع خطاهای تصادفی شکل یک ناقوس را دارد (توزیع طبیعی است) و در مرکز این توزیع رکورد واقعی قرار دارد. پراکندگی حول میانگین توزیع، مبین خطاهای نمونه-گیری است. کاربرد خط کش لاستیکی اندازه واقعی را نشان نمی‌دهد، اما با تکرار اندازه-گیری می‌توان با یافتن میانگین، نمره واقعی را برآورد کرد. البته هرچه دامنه‌های این توزیع طبیعی جمع‌تر باشد، خطای اندازه‌گیری کمتر خواهد بود.

براساس نظریه اندازه‌گیری، نمره واقعی فرد، به دلیل خطاهای تصادفی، با تکرار زیاد همان آزمون تغییر نمی‌کند(۱۰).

طبق این تئوری، از انحراف استاندارد خطای اندازه‌گیری می‌توان برای برآورد میانگین مانند هر انحراف استاندارد دیگر سود جست. نحوه محاسبه و استفاده از آن به شرح زیر است.[۱]

۱.روش‌های محاسباتی فقط به منظور درک و تفسیر داده‌های رایانه‌ای ارائه می‌شوند.

$$SE_m = S_x \sqrt{1 - r_{xx}}$$

در این فرمول SE_m یعنی انحراف استاندارد خطای اندازه‌گیری[۱]، S_x یعنی انحراف استاندارد اندازه‌های به دست آمده و r_{xx} مقدار پایایی آزمون با روش آلفای کرونباخ است.

حال، قابل پیش‌بینی است که با ۶۸ درصد اطمینان واقعی فرد در دامنه $X \pm 1SE_m$، با ۹۵ درصد اطمینان در دامنه $X \pm 2/5SE_m$ و با ۹۹ درصد اطمینان در دامنه $X \pm 3SE_m$ قرار می‌گیرد. مقدار پایایی یک ابزار اندازه‌گیری با افزایش میزان خطای اندازه گیری (e) کاهش می‌یابد . برای اثبات این نظریه می‌توان از این فرمول استفاده کرد:

$$\text{پایایی} = \frac{\text{واریانس نمره‌های بدون خطای اندازه‌گیری}}{\text{واریانس کل نمره‌ها}}$$

واریانس نمره‌های بدون خطا به شرح زیر قابل محاسبه است:

واریانس اشتباهات اندازه‌گیری – واریانس کل نمره‌ها= واریانس نمره‌های بدون خطا

یا به زبان آماری:

$$\delta^2_T = \delta^2_x - \delta^2_e$$

به این ترتیب، مشاهده می‌شود که هر چه مقدار واریانس خطا کوچک شود مقدار پایایی افزایش خواهد یافت[۲] (سپاسی، ۱۳۷۶) (۱).

با وجود اینکه پایایی و عینیت دو تعریف جداگانه و مفهوم‌های ویژه خود را دارند، به دلیل اینکه مقاله حاضر با دیدگاه راهنمای محاسباتی نگاشته شده است و محاسبه این دو مفهوم مشابه یکدیگر است، لذا تأکید مقاله بر تفکیک آنها از یکدیگر قرار ندارد.

---

۱. برای درک بهتر موضوع می توان به جزوه های درسی نویسنده در زمینه همبستگی و رگرسیون آماری در دانشکده تربیت بدنی و علوم ورزشی نیز مراجعه کرد.

منابع اشتباه در اندازه گیری

اشتباهات اندازه‌گیری از منابع متعددی سرچشمه می‌گیرند که تعدادی از آنها آشکار، ولی دیدن بعضی دیگر دشوار است. نویسندگان مختلف به صورت‌های متفاوتی آن‌ها را دسته بندی کرده‌اند (سپاسی ، ۱۳۷۶، ص ۱۵۸، کاپلان، ۱۹۸۷، ص۲۴۷) (۱۴،۱)، ولی جامع ترین فهرست را جانسون ونلسون (۱۹۷۹، صص۳۹-۴۸) (۱۳) از دیدگاه تربیت بدنی ارائه کرده‌اند. با توجه به اهمیتی که شناخت و کنترل این منابع خطا در افزایش پایایی آزمون دارد، آن‌ها به صورت کاملتر در زیر ارائه شده‌اند:

الف) متغیرهای وابسته به آزمودنی: از جمله نوسانات علاقه‌مندی یا همدلی با جریان سنجش، خستگی پذیری، از دست دادن کنجکاوی یا دقت و ....

ب) شرایط محیط و زمان سنجش: تغییر سالن، عوامل بروز حواس پرتی از قبیل سرو صدا، نوسانات نور، حرارت، شخصیت آزمونگران و .....

پ) ویژگی‌های ساختار آزمون: تعداد پرسش‌ها و طولانی بودن متن آن، جمله بندی آنها، ابهام یا دو پهلوبودن معانی آنها، سهل الوصول بودن آنها برای نگهداری در حافظه، غیر متجانس بودن آنها و ....

روشهای محاسبه پایایی

پایایی و ثبات در رسیدن به نتایج یکسان در شرایط مشابه (تعمیم آزمون)[1]، و انسجام و هماهنگی درونی (ثبات داخلی)[2] آن قابل بررسی است.

پایایی یا ثبات در این مقاله برگردان واژه Reliability است. در برخی از منابع فارسی این واژه را «اعتماد» ترجمه کرده‌اند (سرمد و همکاران، ۱۳۷۶،ص۱۶۶) (۲)؛ در منابع دیگر به آن پایایی گفته‌اند (نجفی زند، ۱۳۷۳، ص۱۷۸) (۴) که به نظر نویسنده هر دو واژه‌های مناسبی برای ترجمه واژه لاتین آن است؛ در مواردی نیز آن را اعتبار و روایی نوشته‌اند (شریفی، ۱۳۶۶، ص۲۰۰) (۳). دانشجویان تحصیلات تکمیلی تربیت‌بدنی هم بر حسب آنکه به کدام منبع مراجعه می‌کنند گاه آن را «پایایی» و «گاه اعتبار» می‌نامند.

1.R epeatability & Generalizability
2. Internal Consistency

به نسبت اینکه نوع پرسش‌ها چگونه بوده است و برای جمع آوری اطلاعات خام چند بار آزمایش به عمل آمده و از چه نوع مقیاس‌های اندازه‌گیری استفاده شده، روشهای محاسباتی متفاوتی برای برآورد پایایی به کار می‌رود.

پایایی آزمون در اندازه‌گیری‌های تکراری (تعمیم آزمون)

الف – داده‌های غیر کمّی در اندازه‌گیری‌های کیفی (ضریب کاپا):

در صورتی که اطلاعات خام به صورت دو ارزشی رد و قبول ورزشکار و غیر ورزشکار تمرین کرده و تمرین نکرده و .... از نوع غیر کمّی باشند، پایایی آزمون مربوط توسط محاسبه ضریب توافق کاپا (سافریت، ۱۹۸۶، کوهن، ۱۹۶۰، پیرسون، ۱۹۴۰) (۱۷، ۸، ۱۶) به دست می‌آید. به عبارت دیگر، چنانچه براساس نتایج یک ابزار اندازه‌گیری کیفی توسط دو آزمایشگر متفاوت و یا در دو روز و زمان مختلف، آزمودنی‌ها به یکسان طبقه-بندی شوند، آن ابزار دارای پایایی ملاکی یا معیاری خواهد بود.

میزان پایایی یا توافق بین دو آزمایشگر مختلف یا بیشتر (این مورد می‌تواند به صورت عینیت هم تفسیر شود) و یا نتایج حاصل در دو اندازه‌گیری متفاوت با ضریب استفاده از ضریب کاپا تعیین می‌شود. این ضریب می‌تواند به صورت زیر محاسبه شود:

مثال محاسبه : ضریب کاپا

$$K = \frac{P_a - P_c}{1 - P_c}$$

در این فرمول :

P_a= نسبت توافق مشاهده شده بین دو قاضی یا دو بار اندازه گیری است.

P_c= نسبت توافق مورد انتظار است.

این نسبت‌ها به صورت زیر محاسبه می‌شوند:

$$P_a = \frac{A + D}{A + B + C + D} \qquad P_c = \frac{(A+B)(A+C) + (C+D)(B+D)}{(A + B + C + D)^2}$$

فرض این است که اطلاعات در جدول ۱ تنظیم شده است:

جدول ۱: نتایج ردی و قبولی افراد دردو روز آزمایش

روز اول / روز دوم	رد	قبول	جمع
قبول	B=10	A=0	10
رد	D=30	C=60	90
جمع	40	60	100

ضریب کاپا چنانچه منفی یا کوچک به دست آید نشان می‌دهد که نمی‌توان آزمون یا ابزار اندازه‌گیری موردنظر را پایا دانست.

ضریب کاپا بین ۱و ۱- به دست می‌آید و ضریب صفر مبین آن است که توافق مشاهده شده برابر با توافق مورد انتظار تصادفی است (ارتباطی بین داوری‌ها وجود دارند) به همین ترتیب عدد ۱ بیانگر توافق کامل و عدد ۱- بیانگر عدم توافق کامل است. ضریب کاپا برابر با ۰/۷ و بیشتر را یک پایایی قابل قبول می‌دانند.

شایان ذکر است که ضریب کاپا یک ضریب توافق[1] است. در حالی که مجذور کای (برای طرح‌های ۲×۲ و بیشتر) و ضریب فای (برای طرح‌های ۲×۲) که آن نیز در حل این مسئله کارایی دارند هر دو شاخص‌های ارتباط و همبستگی هستند. اختلاف در این است که در شرایط عدم توافق کامل (سؤالات کیفی دو ارزشی) ضرایب مجذور کای و فای مثبت به دست خواهند آمد، در حالی که ضریب کاپا با نشان دادن علامت منفی این عدم توافق را نشان خواهد داد.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

ب) داده‌های کمّی در دو گروه نمره (ضریب همبستگی و آزمون F)

در صورتی‌که دو داور به نتایج یک آزمون به‌صورت کمّی نمره بدهند (در این مورد به‌صورت ضریب عینیت آزمون نیز قابل تفسیر است)، یا اینکه یک آزمون دوبار گرفته شود[2]، استفاده از ضریب همبستگی پیرسون برای محاسبات پایایی آزمون از متداولترین روشها است[3](۱۶).

1 . Index of Agreement

۲. آزمون- آزمون مجدد (در صورتی که متغیر مورد نظر با گذشت زمان خود به خود تغییر می کند)

۳. استفاده از ضریب همبستگی پیرسون به عنوان پایایی اشکالاتی دارد که بعداً در این مقاله به آن پرداخته خواهد شد.

راه دیگر برای محاسبه ضریب پایایی (همبستگی بین دو سلسله نمره کمّی) که مشکلات همبستگی پیرسون را ندارد، استفاده از روش تحلیل واریانس در نمره‌های هم بسته و تبدیل f حاصل به ضریب همبستگی است.

مثال عددی : فرض شود از ۱۰ نفر آزمودنی، با یک ابزار اندازه‌گیری معین از یک متغیر دوبار آزمایش به عمل آمده است. سؤال این است که پایایی آزمون با استفاده از روش تحلیل واریانس در نمره‌های هم بسته (اندازه‌گیری‌های تکراری) چقدر است.

حال اگر اطلاعات زیر در دست باشد[1]:

MS_B=۲۰ و ۲۰= SS_B= جمع مجذورات بین گروهی

MS_R=۷/۶۷ و ۶۹= SS_R= جمع مجذورات بین آزمودنی‌ها

MS_E=۱/۴۴ و ۱۳= SS_E= جمع مجذورات اشتباهات

برای تبدیل F در اندازه گیری‌های تکراری به ضریب پایایی (R) به شرح زیر عمل می‌شود:

$$R = \frac{MS_R - MS_E}{MS_R}$$

$$MS_E = \frac{SS_B + SS_E}{df_B + df_E} = \frac{۲۰ + ۱۳}{۱ + ۹} = ۳/۳ \qquad R^2 = (۰/۵۶۹۷)^۲ = ۰/۳۲$$

$$R = \frac{۷/۶۷ - ۳/۳}{۱ + ۹} = ۰/۵۶۹۷$$

$$R^2 = ۰/۳۲\ (۱۰۰) = ٪۳۲$$

شایان ذکر است که هرچه F کوچکتر باشد R عدد بزرگتری خواهد شد.

ج- داده های کمّی در بیشتر از دو گروه (آزمون F)

برای محاسبهٔ پایایی یک ابزار اندازه‌گیری که یک گروه آزمودنی را بیشتر از دوبار روی یک متغیر معین اندازه‌گیری کرده است، بهترین روش استفاده از تحلیل واریانس در نمره‌های تکراری روی یک گروه آزمودنی (نمره‌های هم بسته) است.

دلایل مختلفی در ارجحیت تحلیل واریانس به نسبت ضریب همبستگی ارائه شده است (سافریت، ۱۹۷۶. بوم کارتنر و جکسون، ۱۹۷۵. جانسون ونلسون) (۱۳،۶، ۱۷) که نمونه‌های آن عبارت‌اند از:

۱) پایایی و عینیت تنها با یک متغیر که چند بار اندازه‌گیری شده است سرو کار دارند . ولی ضریب همبستگی یک روش آماری دو متغیره است.

۲) ضریب همبستگی به دو دسته نمره محدود می‌شود، ولی برای بررسی پایایی و عینیت گاه لازم است بیش از دوبار اندازه‌گیری صورت گیرد. (به ویژه در متغیرهای دقت و مهارت).

۳) ضریب همبستگی اطلاعات محدودی درباره رابطه بین دو متغیر به‌دست می‌دهد، ولی روش تحلیل واریانس تغییرات از یک آزمایش به آزمایش دیگر، بین آزمودنی‌ها و تعامل بین آزمودنی‌ها و تکرار اندازه‌گیری‌ها را هم در نتیجه‌گیری نهایی دخالت می‌دهد.

ثبات درونی آزمون

در آزمون‌های مرکب که از چند مورد[1] (مانند پرسشنامه یا آزمون ایفرد در تربیت بدنی) تشکیل شده‌اند لازم است ثابت شود که این موارد در مجموع متغیر مورد نظر را اندازه‌گیری می‌کنند. روشهای آمار متفاوتی برای اثبات این موضوع وجود دارد که به نمونه‌هایی از آن اشاره می‌شود:

۱) روش دو نیمه کردن آزمون[2] : در این روش موارد آزمون یک در میان به دو بخش تقسیم می‌شوند و بین نمرات این دو بخش ضریب همبستگی محاسبه می‌شود[3]. البته چون یک آزمون را می‌توان به شکل‌های متعددی دو نیمه کرد و به صورت‌های مختلف

1. Item
2. Split-Halt

۳. ضرورت امتیاز گذاری به دو نیمه آزمون به صورت جداگانه و لذا احتمال متفاوت بودن واریانس نیمه ها

هم به آن نمره داد، لذا تفسیر و تعیین ثبات درونی آزمون از این طریق با مشکلاتی همراه است.

۲) روش کودر – ریچارسون: به منظور رفع مشکل تنوع در دو نیمه کردن یک آزمون، این دو پژوهشگر در سال ۱۹۳۷، یک فرمول محاسباتی ارائه دادند که قادر است همه انواع دو نیمه کردن یک آزمون را در یک مرتبه اجرا مورد توجه قرار دهد. این فرمول که KR-20 یا KR-21 نامیده می‌شود به جای تعیین همبستگی، ثبات درونی آزمون را از طریق ارتباط بین هریک از موارد آزمون با سایر موارد آن و همچنین کل آزمون به دست می‌دهد. این ثبات درونی در واقع میانگین ثبات درونی همه انواع دو نیم کردن‌های آزمون است. نقطه ضعف بزرگ این روش آن است که تنها برای سؤالات دو گزینه‌ای مناسب است(۱۵).

۳) روش فرم‌های موازی[1]: در صورتی که دو شکل کاملاً موازی از یک آزمون برای اندازه‌گیری یک متغیر موجود باشد، برای جلوگیری از یادگیری و حافظه آزمودنی‌ها در دو بار اندازه‌گیری می‌توان این دو آزمون را دو زمان نزدیک به یکدیگر در یک آزمودنی به اجرا در آورد و همبستگی بین آنها را به عنوان پایایی آزمون مورد نظر تفسیر کرد (گی[2]، ۱۹۸۱- ص ۱۱۸-۱۱۹) (۱۲) زیرا آزمون‌ها در همه ویژگی‌ها به جز نوع سؤالات که همگی یک متغیر را اندازه‌گیری می‌کردند (ثبات درونی) مشابه بوده‌اند (مشابه بودن دو آزمون قابل اثبات آماری است و بعداً به آن اشاره خواهد شد.)

۴) روش محاسبه آلفای کرونباخ[3]: اشکال فرمول‌های کودر- ریچاردسون برای سؤالات با گزینه‌های بیشتر از دو را محقق دیگری به نام کرونباخ در سال ۱۹۵۱ حل کرده است (۹). یکی از دو شرط اساسی برای استفاده از این روش، کافی بودن تعداد سؤالات (موارد) و دیگری مشابه و همگن بودن آن از نظر شکل و هدف است. در غیر این صورت، مقدار ثبات درونی مشاهده شده کمتر از میزان واقعی آن به دست خواهد آمد.

---

1. Equivalent Forms
2. Kuder - Richardson

۵ در بعضی منابع داخلی و خارجی فقط آلفا نامیده می شود.

برای محاسبه آلفا به روش کرونباخ راه‌های مختلفی وجود دارد که ساده‌ترین آن را می‌توان در مثال زیر زیر دید. این ضریب، ثبات درونی مجموعه‌ای از سؤالات (یا آزمایشگرهای مختلف برای عینیت) یک آزمون را تعیین می‌کند.

مثال عددی برای پرسشنامه‌های با سؤالات چند گزینه‌ای

پژوهشگری علاقمندی است ثبات درونی پرسشنامه را که ۴۰ سؤال دارد و به سبک لیکرت پنج گزینه‌ای ساخته است به دست آورد. این پرسشنامه را ۳۰ نفر از جامعه هدف کامل کرده‌اند و نتایج زیر به دست آمده است:

$\sum S_t^2$ = جمع واریانس‌های همه سؤالات (موارد) = ۱۸

S_x^2 = واریانس نمره‌های آزمودنی‌ها از آزمون = ۲۴

N= تعداد سؤالات (موارد) آزمون = ۴۰

برای محاسبه به شرح زیر عمل می‌شود:

$$r_{XX} = \text{آلفا} = \left[\frac{N}{N-1}\right]\left[1 - \frac{\sum S^2t}{S^2X}\right] = \frac{۴۰}{۴۰-۱}\left[۱ - \frac{۱۸}{۲۴}\right] = ۰/۲۶۵$$

راه دیگر برای محاسبه آلفا استفاده از روش تحلیل واریانس یک متغیره برای نمره‌های هم بسته است. در این حالت آلفا را می‌توان نسبتی از تفاوت آزمودنی‌ها روی متغیر مورد نظر که با استفاده از سؤالات (موارد) آزمون توصیف می‌شود دانست:

$$\text{آلفا} = ۱ - \frac{\text{واریانس خطا}}{\text{واریانس کل}}$$

تفاوت فرمول آلفا با فرمول همبستگی درون گروهی (که در مبحث ثبات آزمودن در اندازه‌گیری‌های مکرر توضیح داده شد)، در نحوه استخراج واریانس‌های خطاست.

واریانس خطا در همبستگی درون گروهی از تفریق واریانس‌های بین آزمودنی‌ها و بین آزمایشگران [درون گروهی (اشتباهات) و بین گروهی] از واریانس کل حاصل می‌شد، ولی در اینجا از تفریق جمع مجذورات درون سؤالی و بین سؤالی از واریانس کل به دست می‌آید.

مثال عددی برای سؤالات امتحانی یک جوابی

فرض کنیم پژوهشگری می‌خواهد ثبات درونی یک پرسشنامه با ۳ سؤال را که چهار آزمودنی به آن پاسخ داده‌اند به دست آورد. در جدول ۲ نمره‌های آزمودنی‌ها به سؤالات و اطلاعات آماری لازم ارائه شده است.

جدول ۲- خلاصه محاسبات نمره های چهار آزمودنی به ۳ سؤال

سؤالات / آزمودنی ها	سؤالات			جمع	میانگین
	اول	دوم	سوم		
۱	۱	۱	۱	۳	۱
۲	۲	۲	۲	۶	۲
۳	۱	۱	۱	۳	۱
۴	۲	۲	۱	۵	۱/۶۷
جمع	۶	۶	۵	۱۷	
میانگین	۱/۵	۱/۵	۱/۲۵		۱/۴۲

حال چنانچه در این نمره‌ها مجموع مجذورات کل برابر ۲/۹۱۷، مجموع مجذورات بین آزمودنی‌ها برابر ۲/۲۵، مجموع مجذورات بین سؤالات برابر ۰/۱۷، مجموع مجذورات اشتباهات (پسماند[1] یا خطا) برابر ۰/۵، میانگین مجذورات اشتباهات (واریانس خطا) برابر ۰/۰۸، میانگین مجذورات بین آزمودنی‌ها (واریانس کل) برابر ۰/۷۵ باشد، با استفاده از فرمول آلفا میزان ثبات درونی آزمون عبارت خواهد شد از[2]:

$$r_{xx} = ۱ - \frac{۰/۰۸}{۰/۷۵} \quad ۰/۸۹$$

1 . Residual

۲. چون همه این محاسبات توسط رایانه انجام می شود و جدول های خلاصه از آن طریق به دست می آید از حل کامل مسئله خودداری شده است.

فرمول پیشگویی اسپیرمن –براون برای افزایش درجه پایایی

یکی از شیوه‌های برآورد تعداد آزمایشگرها (به عنوان ابزار اندازه‌گیری) یا تعداد سؤال‌های مورد نیاز برای ساختن یک وسیله اندازه‌گیری با پایایی بیشتر استفاده از فرمول پیشگویی اسپیرمن – براون به این شرح است (۱۹):

$$\text{ضریب پایایی مورد انتظار} = \frac{(n) \times \text{پایایی معلوم}}{1 + (n - 1) \times (\text{پایایی معلوم})}$$

فرمول پیشگویی اسپیرمن- براون به این سؤال پاسخ می‌دهد که برای رسیدن به ضریب پایایی بیشتر، تعداد سؤالات یا آزمایشگرها چند برابر بیشتر شود. در این فرمول، n ضریبی است که نشان می‌دهد تعداد آزمایشگرها، یا سؤالات، چند برابر باید کم و یا افزوده شود. به عنوان مثال، از فرمول حاضر می توان برای دستیابی به پایایی شش سؤال به جای سه سؤال، که پایایی آنها ۰/۸۹ به دست آمده بود، برای اندازه گیری متغییر مورد نظر بهره گرفت. ضریب ۲ برای تعداد سؤالات، که مورد اشاره قرار گرفت، با استفاده از فرمول بالا باعث می‌شود تا ضریب پایایی به ۹۴٪ افزایش یابد.

مثال عددی :

$$\text{ضریب پایایی مورد انتظار} = \frac{۲ \times ۰/۸۹}{۱ + [(۲ - ۱) \times ۰/۸۹]} = \frac{۱/۷۸}{۱/۸۹} = ۰/۹۴$$

با این ترتیب، مشاهده می‌شود که با افزایش تعداد سؤالات به دو برابر، میزان پایایی درونی پرسشنامه از ۸۹٪ به ۹۴٪ افزایش یافته است.

برای اطمینان از نتیجه این فرمول می‌توان سه سؤال دیگر به مثال قبلی، مشابه با همان سطح ارزشیابی، اضافه کرد و با استفاده از جدول ۳ پایایی را برای آنها دوباره محاسبه کرد.

جدول ۳- نتیجه نمره دادن آزمودنی‌ها به ۶ سؤال

سؤالات آزمودنی‌ها	سؤالات ۱	۲	۳	۴	۵	۶	جمع	میانگین
۱	۱	۱	۱	۱	۱	۱	۶	۱
۲	۲	۲	۲	۲	۲	۲	۱۲	۶
۳	۱	۱	۱	۱	۱	۱	۶	۱
۴	۱	۲	۲	۱	۲	۲	۱۰	۱/۶۷
جمع	۵	۶	۶	۵	۶	۶	۳۴	
میانگین	۱/۲۵	۱/۵	۱/۵	۱/۲۵	۱/۵	۱/۵		۱/۴۲

با محاسبات رایانه‌ای، ضریب آلفا برای ۶ سؤال برابر ۰/۹۵۵۶ به دست آمده که از نتیجه حاصل توسط فرمول اسپیرمن – براون دور نیست. به طوری که مشاهده می‌شود یکی از اطلاعات اصلی برای استفاده از فرمول پیشگویی اسپیرمن – براون تعیین تعداد نفرات لازم (n) است. به منظور رسیدن به این تعداد نفر لازم فرمول زیر قابل استفاده است:

$$n = \frac{\text{(پایایی معلوم} - ۱) \text{ پایایی مورد انتظار}}{\text{(پایایی مورد انتظار} - ۱) \text{ پایایی معلوم}}$$

برای درک کاربرد فرمول بالا فرض کنید پرسشنامه‌ای با ۳ سؤال، آلفایی برابر ۰/۵۰ را به دست داده است (پایایی معلوم)، ولی کاربر این پرسشنامه می‌خواهد میزان این آلفا را به ۹۰٪ برساند (پایایی مورد انتظار). طبق فرمول داده شده، برای رسیدن به این منظور، تعداد سؤالات پرسشنامه باید ۹ برابر، یعنی ۲۷ سؤال شود:

$$n = \frac{۰/۹۰\ (۱ - ۰/۵)}{(۱ - ۰/۹۰)} = \frac{۰/۴۵}{۰/۰۵} = ۹ \times ۳ = ۲۷$$

اصلاح کوچک شدن همبستگی به علت پایایی ناکافی ابزار اندازه‌گیری

در تعداد زیادی از تحقیقات، متغیرها با ابزاری که پایایی آن‌ها کافی نیست، اندازه-گیری می‌شوند. نتیجه این است که همبستگی بین متغیرهایی که توسط ابزار با پایایی

ناکافی اندازه‌گیری شده‌اند همیشه ضعیف‌تر از مواردی است که با ابزاری که پایایی کافی دارند اندازه‌گیری شده‌اند. این کاهش همبستگی به نام کوچک شدن میزان همبستگی (Attenuation) مشهور است. در مواردی که همبستگی واقعی (بدون اشتباهات اندازه‌گیری) بین دو دسته نمره موردنظر باشد (اندازه‌گیری بدون اشتباه باشند) لازم است که برای جبران کمبود پایایی ابزار که جهت اندازه‌گیری به کار رفته از نوعی اصلاحیه استفاده شود. در فرمول زیر میزان همبستگی بین دو متغیر، با درنظرگرفتن پایایی ابزار اندازه‌گیری هریک ازآن محاسبه می‌شود:

$$\text{ضریب پایایی اصلاح شده} = \frac{\text{ضریب همبستگی بین متغیرهای ۱ و ۲}}{\text{(پایایی ابزار متغیر ۱) (پایایی ابزار متغیر ۲)}} = \text{همبستگی اصلاح شده} = rc$$

به عنوان مثال، چنانچه همبستگی بین ضریب هوشی و پرش طول جفتی برابر ۰/۳۰ و پایایی ابزارهای اندازه‌گیری مرتبط با ضریب هوشی و پرش طول جفتی به ترتیب ۰/۸۰ و ۰/۷۰ باشند، همبستگی اصلاح شده بین این دو متغیر عبارت خواهد بود از :

$$r = \frac{۰/۳۰}{۰/۸۰\ (۰/۷۰)} = \frac{۰/۳۰}{۰/۵۶} = ۰/۵۴$$

نکته قابل توجه این است که چنانچه همبستگی اصلاح شده بین دو ابزار اندازه-گیری نزدیک به ۱ باشد، می‌توان آن دو وسیله اندازه‌گیری را دو نوع متفاوت از یک ابزار اندازه‌گیری به حساب آورد (هردو یک متغیر را اندازه‌گیری می‌کنند)[۱].

نکته دیگر این است که از همبستگی اصلاح شده نمی‌توان برای مطالعات پیشگویی و رگرسیون استفاده کرد، زیرا در این محاسبات همبستگی واقعی، که به دست آمده، مورد نیاز است.

۱. کاربرد زیادی برای اثبات اعتبار معیاری و یا ساختن ابزار اندازه گیری مشابه دارد.

منابع

۱- سپاسی، حسن و پریوش نوربخش، (مترجمین)، (۱۳۷۶)، سنجش و اندازه‌گیری در تربیت بدنی (جلد اول)، انتشارات سمت، تهران، صص ۱۵۷-۱۶۳.

۲- سرمد، زهره و همکاران (۱۳۷۶)، روشهای تحقیق در علوم رفتاری، نشر آگه، تهران، صص ۱۳۹-۲۰۱.

۳- شریفی، پاشا و نرگس طالقانی (مترجمین)، (۱۳۶۶). روشهای تحقیق در علوم تربیتی، انتشارات رشد، تهران، صص ۱۹۹-۲۶۲.

۴- نجفی زند، جعفر(مترجم)، (۱۳۷۳)، مبانی پژوهش در علوم تربیتی، نشر قومس، تهران، صص ۱۷۵-۱۸۱.

۵- هومن، حیدر علی، (۱۳۶۶). اندازه‌گیری روانی و تربیتی و فن تهیه تست. نشر سلسله، تهران، فصل‌های دوم، سوم و چهارم.

6- Baumgartner T A, and Jakson As (1975). *Measurement for evaluation in physical Education*. Boston: Houghlon Mifflin Company.

7- Brown W (1910). *Some experimental results in the correlation in mental abilities*. British Journal of Psy, 3: 296-322.

8- Cohen J (1960). *A coefficient of agreement for nominal scales*. Educational and Psychological Measurement.

9- Cronbach L J(1951). *Coefficient alpha and the internal structure of tests*. Psychometrica, 16:297-334.

10- Duncan C(1994). *Introducing statistics for social research, step-by-step calculations and computer techniques using SPSS*. Rurledge, *New York, 296-280.*

11- Ebel RL (1976). *The Paradox of Educational Testing Measurement in Education.*

12-Gay LR (1981). *Educational Research: Competencies for Analysis and Application,* (2nd .ed) Charles E.Merrill Publishing Co. London.

13- Johnson BL and Nelson JK (1979*).practical Measurements for Evaluation in Physical Education (3rd.ed).* Burgess publishing Company, Minnesta, 39-48.

14- Kaplan RM (1981). *Basic Statistics for the Behavioral Sciences*. Alyn and Acon, Inc. Boston, pp.243-263.

15- Kuder GF and Richardson M W(1937). *The Theory and Estimation of Test Reliability.* Psychometrica, 2: 151-160.

16- Pearson K(1904). *On theTheory of Contingency and it's Relation to Association and Normal Correlation.* Drapers, company Res. Mem. Biometric Ser.I.

17- Safrit MJ (ed). (1976). *Reliability Theory,* Washington, D.C.: AAHPERD.

18- Safrit MJ (1986). *Introduction to Measurement in physical Education and Exercise Science.* St. Louis: Time Mirror/ Mosby.

19- Spearman C (1910).*Correlation Calculated from Faculty Data. British* Journal of Psychology, 3: 271-295.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی