

کارگاه مجازی

تحلیل سؤالات آزمونهای چند گزینه ای (تحلیل بعد کمی)

هدف از تحلیل سؤال های آزمون واریسی تک تک سؤالات و تعیین میزان دقت و نارسایی آنها است. در تحلیل سؤال های آزمون نقاط قوت و ضعف یک آزمون و کیفیت همه سؤال های آن تعیین می شود. بنابراین لازم است که پس از اجرای آزمون سؤال های آن مورد تجزیه و تحلیل قرار گیرد و با استفاده از نتایج حاصله به تجدید نظر در آزمون و بهبود کیفیت سؤالات برای استفاده های بعدی اقدام شود. در تجزیه و تحلیل سؤالات آزمونها دو جنبه کیفی و کمی آن مورد بررسی قرار می گیرد، از نظر کیفی (شکل، روش تهیه سؤال، دقت در متن سؤال، سطح سؤال در حیطه شناختی و گزینه درست و گزینه های نادرست سؤال) مورد بررسی و ارزیابی قرار می گیرد و به گونه کلی شواهدی که در تعیین روایی محتوا و روایی سازه تست مؤثر است، مورد بررسی قرار می گیرد. از نظر کمی مواردی مثل (ضریب دشواری، ضریب تمیز، تحلیل گزینه های انحرافی و پایایی آزمون) مورد توجه و بررسی قرار می گیرند.

اطلاعات مورد نیاز برای تحلیل سؤال های یک آزمون، پاسخهایی است که آزمون شوندگان به هر سؤال داده اند. یعنی باید تعیین کنید چند نفر گزینه درست را انتخاب کرده اند و هر یک از گزینه های انحرافی چند نفر را به به خود جلب کرده است و چند نفر به آن پاسخ نداده اند. برای این منظور هر سؤال را در جدولی مانند جدول زیر وارد می کنیم:

سوال	گروهها	الف	ب	ج	د	بدون پاسخ	جمع
	گروه بالا						
	گروه پایین						
	ضریب دشواری: ...	ضریب تمیز: ...					گزینه های انحرافی: ...

محاسبه ضریب دشواری (difficulty index):

درصد کل آزمون شوندگانی که به یک سؤال پاسخ صحیح می دهند، ضریب دشواری آن سؤال است که با حرف P نشان داده می شود. اگر در تحلیل یک سؤال کلیه افراد یا همه برگه های امتحانی دخالت داشته باشند، برای محاسبه ضریب دشواری سؤال کافی است تعداد کل افرادی را که به آن سؤال پاسخ صحیح داده اند بر تعداد کل آزمون شوندگان تقسیم کنیم و نتیجه را در عدد ۱۰۰ ضرب نماییم. رقم حاصل ضریب دشواری سؤال است.

$$P = 100 \frac{(R)}{T}$$

P- ضریب دشواری سؤال:

R- تعداد کل افرادی که به سؤال پاسخ صحیح داده اند.

T- تعداد کل آزمون شوندگان.

در مواردی که تعداد آزمون شوندگان (تعداد برگه ها) زیاد است و اطلاعات ما به نحوه پاسخ دهی افراد گروه بالا و پایین محدود می شود از فرمول زیر استفاده می شود:

$$(P) = \frac{\text{انتخاب های درست گروه بالا} + \text{انتخاب های درست گروه پایین}}{\text{تعداد افراد گروه بالا} + \text{تعداد افراد گروه پایین}} * 100$$

ضریب دشواری (P) =

برای انتخاب گروه بالا و پایین به شکل زیر عمل می کنیم:

اگر تعداد آزمون شوندگان کمتر از ۲۰ نفر باشند، تمام برگه ها را بعد از تصحیح به ترتیب نمره کسب شده از بالا به پایین مرتب می کنیم و آنها را به دو گروه قوی و ضعیف تقسیم می کنیم، که گروه بالا مربوط به نمرات بالا و گروه پایین مربوط به نمرات پایین است.

اگر تعداد آزمون شوندگان بین ۲۰ تا ۴۰ نفر باشند ۱۰ نفر از گروه قوی و ۱۰ نفر از گروه ضعیف به ترتیب انتخاب می نماییم و افرادی که نمرات آنها در وسط قرار می گیرد از محاسبه خارج می کنیم.

اگر تعداد افراد از ۴۰ نفر بیشتر باشد، برای گروه بالا ۲۷ درصد افراد و برای گروه پایین ۲۷ درصد افراد را انتخاب می نماییم.

تفسیر ضریب دشواری:

هر چه ضریب دشواری سوالی بزرگتر (به صد نزدیکتر) باشد آن سوال، سوال آسانتری است و هر چه این ضریب کوچکتر (به صفر نزدیکتر) باشد سوال، سوال دشوارتری است. ضریب دشواری بین ۰/۳ تا ۰/۷ حداکثر اطلاعات را درباره تفاوت آزمودنی ها به دست می دهد. بهترین ضریب دشواری برای یک سوال ۵۰ درصد است.

(۱) سؤال آسان $\longrightarrow$ (۰) سؤال دشوار

محاسبه ضریب تمیز: (discrimination index)

ضریب تمیز، قدرت سوال را در تمایز گذاری یا تشخیص بین گروه قوی و گروه ضعیف آزمون شوندگان مشخص می کند.

ضریب تمیز با استفاده از این فرمول محاسبه می شود:

$$D = \frac{\text{انتخاب های درست گروه پایین} - \text{انتخاب های درست گروه بالا}}{\text{تعداد افراد یک گروه (بالا یا پایین)}} * 100$$

تفسیر ضریب تمیز:

هر قدر ضریب تمیز سوال بزرگتر (نزدیک تر به یک) باشد قوه تمیز آن بیشتر و هر قدر این ضریب کوچکتر (به صفر نزدیک تر) باشد قوه تمیز آن کمتر است. مثلاً اگر ضریب تمیز سؤالی ۰/۹۰ باشد، سؤال آزمون شوندگان قوی و ضعیف را به خوبی از هم متمایز خواهد کرد. اما اگر ضریب تمیز سؤالی ۰/۱۰ باشد، سؤال آزمون شوندگان را از هم متمایز نخواهد کرد.

ضریب تمیز صفر حاکی از آن است که سوال بین گروه قوی و ضعیف تمایز ایجاد ننموده است و اگر در محاسبه با ضریب تمیز منفی روبرو شویم به معنای آن است که در این سوال گروه قوی، ضعیف عمل کرده است و گروه پایین کلاس آن را بهتر پاسخ داده اند. بدیهی است این سوال اشکال اساسی دارد و باید به کلی کنار گذاشته شود و یا در آن تجدید نظر اساسی به عمل آید.

مقادیر مطلوب ضریب تمیز:

سؤالات بسیار خوب	۴/ و بالاتر
سؤالات تقریباً خوب	۳۹/ - ۳۰/
سؤالاتی که احتیاج به تجدید نظر دارند	۲۹/ - ۲۰/
سؤالات بسیار ضعیف، که نیاز به تجدید نظر اساسی دارند و یا باید کنار گذاشته شوند.	۱۹/ به پایین

تحلیل گزینه های انحرافی (Distractor Analysis)

تعریف: گزینه های انحرافی، گزینه های اشتباهی هستند که همراه با گزینه پاسخ یا صحیح آورده می شوند تا توجه آزمودنی های ناآگاه و کم اطلاع را از گزینه پاسخ یا صحیح منحرف کنند.

تفسیر گزینه های انحرافی:

هر گزینه ی انحرافی باید حداقل یک نفر از افراد گروه ضعیف را به جلب کند و اگر هردو گروه قوی و ضعیف را به خود جلب کند، تعداد افراد گروه ضعیف باید بیشتر از گروه قوی باشد. بنابراین در صورتی یک سوال به خوبی عمل می کند که افراد گروه ضعیف بیشتر از افراد گروه قوی گزینه های انحرافی آن سوال را انتخاب نمایند. چنانچه گزینه ی انحرافی خاصی وجود داشته باشد که هیچ یک از افراد گروه قوی وضعیف آن را انتخاب نکرده باشند، گزینه ی انحرافی مناسبی نیست و باید اصلاح شود. در سؤالات استاندارد افراد با دانش کافی گزینه صحیح را انتخاب می نمایند و افراد فاقد اطلاعات مناسب پاسخ های نادرست (گزینه های انحرافی) را به طور مساوی انتخاب می نمایند.

گزینه انحرافی قوی، گزینه ای است که از گروه قوی کسی را به خود جلب نکند ولی از گروه ضعیف چند نفر را به خود جلب نماید. اگر گزینه های انحرافی درصد بیشتری به خود اختصاص دادند. اولاً : این انتخاب به معلومات ناقص مربوط میشود. ثانیاً ممکن است ساختار سؤال ضعیف یا فریب دهنده باشد . گزینه های انحرافی فوق العاده مقبول، سبب کاهش روائی و اعتبار آزمون می شود.

ضریب اعتماد یا پایایی آزمون (Reliability):

قابلیت اعتماد که واژه هایی مانند پایایی، ثبات و اعتبار برای آن به کار برده می شود یکی از ویژگی های ابزار اندازه گیری است. مفهوم یاد شده با این امر سروکار دارد که ابزار اندازه گیری در شرایط یکسان تا چه اندازه نتایج یکسانی به دست می دهد. اگر آزمون در هر بار اجرا نتایج متفاوتی بدهد آن آزمون پایا نخواهد بود و در واقع چیزی را به درستی اندازه نخواهد گرفت. دامنه ضریب قابلیت اعتماد از صفر (عدم ارتباط) تا ۱+ (ارتباط کامل) است. قابلیت اعتماد در یک آزمون می تواند از موقعیتی به موقعیت دیگر و از گروهی به گروه دیگر متفاوت باشد. برای محاسبه ضریب قابلیت اعتماد ابزار اندازه گیری شیوه های مختلفی به کار برده می شود. از آن جمله می توان به الف) اجرای دوباره (روش بازآزمایی)، ب) روش موازی (همتا)، ج) روش تصنیف (دو نیمه کردن)، د) روش کودر- ریچاردسون و ضریب آلفای کرونباخ اشاره کرد.

الف) اجرای دوباره آزمون یا روش بازآزمایی

این روش عبارت است از ارائه یک آزمون بیش از یک بار در یک گروه آزمودنی تحت شرایط یکسان. برای محاسبه ضریب قابلیت اعتماد با این روش، ابتدا ابزار اندازه گیری بر روی یک گروه آزمودنی اجرا شده و سپس در فاصله زمانی کوتاهی دوباره در همان شرایط، آزمون بر روی همان گروه اجرا می شود. نمرات بدست آمده از دو آزمون مورد نظر قرار گرفته و ضریب همبستگی آنها محاسبه می شود. این ضریب نمایانگر قابلیت اعتماد (پایایی) ابزار است. روش بازآزمایی برای ارزیابی ثبات مولفه های یک ابزار اندازه گیری به کار می رود ولی با این اشکال نیز روبه رو است که نتایج حاصله از آزمون مجدد می تواند تحت تاثیر تمرین (تجربه) و حافظه آزمودنی قرار گرفته و بنابراین منجر به تغییر در قابلیت اعتماد ابزار اندازه گیری گردد.

ب) روش موازی یا استفاده از آزمون های همتا

یکی از روشهای متداول ارزیابی قابلیت اعتماد استفاده از آزمون های همتا است که شباهت زیادی با یکدیگر دارند ولی کاملاً یکسان نیستند. ضریب قابلیت اعتماد بر اساس همبستگی نمرات دو آزمون همتا به دست می آید. خطاهای اندازه گیری و کمبود یا عدم وجود همسانی میان دو فرم همتای آزمون (تفاوت در سؤالات یا محتوای آنها) قابلیت اعتماد را کاهش می دهد. ممکن است فرم های همتا به طور همزمان ارائه نگردد. در این صورت قابلیت اعتماد هم شامل ثبات و همسانی می شود. اگر فاصله اجرای فرم های همتا زیاد باشد و تغییراتی در ضریب قابلیت اعتماد مشاهده گردد، ممکن است که این تغییر مربوط به عوامل فردی (آزمودنی ها)، همتا نبودن آزمون ها و یا ترکیبی از هر دو باشد.

(ج) روش تنصیف (دو نیمه کردن)

این روش که به روش فرد- زوج نیز معروف است، برای تعیین هماهنگی درونی یک آزمون به کار می‌رود. بدین ترتیب که آزمون را در مورد تعداد زیادی آزمودنی بکار می‌برند. پس از اجرای آزمون، آن را به دو نیمه تقسیم می‌کنند. بدین ترتیب که سؤالهای فرد آن را به عنوان یک آزمون و سؤالهای زوج آن را به عنوان یک آزمون دیگر در نظر می‌گیرند. یعنی برای هر آزمودنی دو نمره به دست می‌آورند. یک نمره برای سؤالهای فرد و یک نمره برای سؤالهای زوج. بین دو سری نمره، ضریب همبستگی بدست می‌آورند. ضریب همبستگی بدست آمده اعتبار نصف آزمون به حساب می‌آید. برای پیدا کردن اعتبار کل آزمون از فرمول زیر، که به فرمول اسپیرمن براون معروف است، استفاده می‌کنند:

$$r_{11} = \frac{2r \frac{1}{2} \frac{1}{2}}{1+r \frac{1}{2} \frac{1}{2}}$$

در این فرمول:

r_{11} : ضریب اعتبار کل آزمون

r : ضریب اعتبار دو نیمه آزمون $\frac{1}{2} \frac{1}{2}$

مثال: اگر ضریب اعتماد دو نیمه آزمون ۰/۶۰ باشد، اعتبار کل آزمون برابر خواهد بود با:

$$r_{11} = \frac{2r \frac{1}{2} \frac{1}{2}}{1+r \frac{1}{2} \frac{1}{2}} = \frac{2 \times 0/60}{1+0/60} = 0/75$$

(د) روش کودر - ریچاردسون

کودر و ریچاردسون دو فرمول را برای محاسبه هماهنگی درونی آزمون ها ارائه نموده اند. فرمول اول برآوردی از میانگین ضرایب قابلیت اعتماد برای تمام طرق ممکن تنصیف (دو نیمه کردن) را به دست می دهد. این ضریب برابر است با:

$$r_1 = \frac{n}{n-1} \left(1 - \frac{\sum pq}{S^2} \right)$$

که در آن :

n تعداد سوال ها - p نسبت پاسخ های صحیح - q نسبت پاسخ های غلط - S^2 واریانس نمرات کل
فرمول دوم: برای استفاده از فرمول دوم باید سطح دشواری سوال های آزمون یکسان باشد:

$$r_2 = \frac{n}{n-1} \left(1 - \frac{\bar{x}(n-\bar{x})}{nS^2} \right)$$

که در آن :

n تعداد سوال ها - S^2 واریانس نمره های کل - $\bar{x}$ میانگین نمرات

برای محاسبه ضریب قابلیت اعتماد با این روش، نیازی به داشتن نسبت پاسخ های صحیح و غلط نیست، بلکه در دست داشتن واریانس نمره های کل و میانگین نمرات کفایت می کند.

برای مثال، آزمونی دارای ۴۰ سوال است. این آزمون به گروهی از آزمودنیها داده شده است. میانگین نمره های آزمون ۲۵ و واریانس کل ۲۶ محاسبه شده است. سطح دشواری سوال ها نیز یکسان است. ضریب قابلیت اعتماد این آزمون را محاسبه کنید:

$$r_2 = \frac{40}{39} \left(1 - \frac{25(40-25)}{40(26)} \right) = 1.036(1 - 0.36) = 0.66$$

ه) روش آلفای کرونباخ

یکی دیگر از روشهای محاسبه قابلیت اعتماد استفاده از فرمول کرونباخ است. این روش برای محاسبه هماهنگی درونی ابزار اندازه گیری از جمله پرسشنامه ها یا آزمونهایی که خصیصه های مختلف را اندازه گیری می کند بکار می رود. معمولاً ضریب اعتماد پرسشنامه هایی را که بصورت طیف لیکرت تنظیم شده اند، با این روش بدست می آورند. در این گونه ابزارها، پاسخ هر سوال می تواند مقادیر عددی مختلف را اختیار کند.

برای محاسبه ضریب آلفای کرونباخ ابتدا باید واریانس نمره های هر زیر مجموعه سوال های پرسشنامه (یا زیر آزمون) و واریانس کل را محاسبه کرد. سپس با استفاده از فرمول زیر مقدار ضریب آلفا را محاسبه کرد.

$$r_a = \frac{j}{j-1} \left(1 - \frac{\sum S_j^2}{S^2} \right)$$

که در آن:

j = تعداد زیر مجموعه سوال های پرسشنامه یا آزمون.

S_j^2 = واریانس زیر آزمون j ام.

S^2 = واریانس کل آزمون.

برای مثال اگر پرسشنامه دارای سه زیر مجموعه سوال باشد، به طوری که واریانس هر زیر مجموعه به ترتیب ۶، ۴ و ۷ بوده و واریانس کل برابر با ۳۲ باشد، مقدار ضریب آلفای کرونباخ با استفاده از فرمول یاد شده به صورت زیر بدست می آید:

$$r_a = \frac{3}{2} \left(1 - \frac{17}{32} \right) = 0.8$$

مقدار صفر این ضریب نشان دهنده عدم قابلیت اعتماد و ۱+ نشان دهنده قابلیت اعتماد کامل است.

عوامل موثر بر ضریب قابلیت اعتماد:

ضریب قابلیت اعتماد یک آزمون تحت تاثیر عوامل مختلفی قرار دارد که عبارتست از:

۱ - طول آزمون: علی الاصول آزمون های طولانی تر از قابلیت اعتماد بالاتری در مقایسه با آزمون های کوتاه برخوردارند. برای مثال ضریب قابلیت اعتماد یک آزمون ۶۰ سوالی می تواند بیشتر از یک آزمون ۴۰ سوالی باشد.

۲ - سوال های مشابه از نظر محتوا و با سطح دشواری متوسط، قابلیت اعتماد آزمون را افزایش می دهد.

۳ - ماهیت متغیر اندازه گیری شده نیز ضریب قابلیت اعتماد را تحت تاثیر قرار می دهد.

تجدید نظر، اصلاح و بازنویسی آزمون:

زمانی که ضریب دشواری و ضریب تمیز سوال های یک آزمون تعیین شد، باید به اصلاح سوالاتی که نیاز به تغییر دارند اقدام نمود. سوال هایی که ضریب دشواری پایین یا بالا دارند و یا ضریب تمیز آنها خیلی کم است باید مورد تجدید نظر قرار گیرند. این موارد شامل گزینه های انحرافی معیوب نیز می شود که باید اصلاح یا عوض شوند.

منابع:

- سیف، علی اکبر. اندازه گیری سنجش و ارزشیابی آموزشی. تهران. نشر دوران
- گنجی، حمزه ۱۳۲۱. روانسنجی. تهران. انتشارات دانشگاه پیام نور