

روی آوردهای نوین در روان‌سنجی

قسمت دوم : مبانی و مفاهیم نظریه سؤال پاسخ

New Approaches to Psychometrics

Part Two : The Basics and Concepts of Item Response Theory (IRT)

Ali Asgari
PhD Candidate
Tehran University

علی عسگری
دانشجوی دکتری
دانشگاه تهران

نظریه سؤال پاسخ^۱ یکی از پیشرفته‌ترین نظریه‌های اندازه‌گیری و نقطه مقابل نظریه کلاسیک است. زیرا برخلاف این نظریه بر نمره سؤالها و نه نمره کل تست بنا شده است (بیگر، ۲۰۰۲/۱۳۸۱). در IRT برای تحلیل داده‌های به دست آمده از مقیاسها و پرسشنامه‌ها از مدل‌های ریاضی استفاده می‌شود. به مدل‌های IRT اغلب مدل‌های خصیصه مکنون^۲ نیز گفته می‌شود. اصطلاح مکنون برای تأکید بر پاسخهای پیوسته سؤالی به کار می‌رود که در واقع نشانگر مشاهده‌پذیر خصائص یا رگه‌هایی است که فرض می‌شود وجود دارد و باید از پاسخها استخراج شود.

مدل‌های IRT توابع ریاضی هستند که احتمال یک برونشد^۳ پیوسته، مانند پاسخ درست به یک سؤال را بر اساس پارامترهای آزمودنی (توانایی، θ) و پارامترهای سؤال (دشواری، قدرت تشخیص و حدس) مشخص می‌کند. به بیان دیگر، مدل‌های IRT مبنایی برای برآورد آماری پارامترهایی است که بیانگر جایگاه آزمودنیها بر روی پیوستار مکنون، یا به گونه دقیقتر، اندازه خصیصه مکنونی است که به آزمودنیها و سؤالها نسبت داده می‌شود. وقتی برآورد پارامترهای مربوط به دست آمد، برای تعیین اندازه و قدرت پیش‌بینی پاسخهای سؤال، بر اساس یک مدل معین، از آزمونهای آماری استفاده می‌شود. آزمونهای آماری برای تعیین این مطلب به کار می‌روند که مدل مورد نظر برای برآورد پارامترها تا چه حد می‌تواند ساختار الگوهای آماری موجود در داده‌ها را، یا به گونه کلی، و یا با در نظر گرفتن زیر مجموعه خاصی از داده‌ها مانند بردار پاسخ مربوط به سؤالها و آزمودنیهای منفرد اندازه‌گیری کند (امبرستون و رایس، ۲۰۰۰).

مفروضه‌های اساسی IRT

۱) نخستین مفروضه IRT اشاره به این مطلب دارد که یک عامل مشترک^۴، همه کوواریانسهای سؤال را توجیه می‌کند. این عامل مشترک در واقع خصیصه مکنون مورد نظر است که در ادبیات روان‌سنجی معمولاً به دو گونه بیان می‌شود: الف) تک‌بعدی بودن^۵؛ یعنی یک و فقط یک عامل مشترک یا خصیصه مکنون وجود دارد که همبستگیهای درونی سؤال را توجیه می‌کند، ب) استقلال مکانی^۶؛ یعنی اگر عامل مشترک تا اندازه‌ای از دو سؤال خارج شود، کوواریانس پسماند^۷ آنها برابر با صفر خواهد بود. به بیان دیگر، پس از کنترل یک یا چند عامل مشترک، پاسخهای سؤال، ناهمبسته یا مستقل از یکدیگر هستند (امبرستون، ۱۹۹۶؛ هامبلتون، سوامینتان و راجرز، ۱۹۹۱).

۲) بر پایه مفروضه دوم، مدل‌های پارامتریک IRT توابعی هستند که به گونه تکنوا^۸ افزایش می‌یابند (یعنی احتمال پاسخ درست به سؤال با افزایش سطح خصیصه مکنون بیشتر می‌شود). این مفروضه را می‌توان از طریق نمودار نمره‌های میانگین سؤال بر روی نمره‌های باقی‌مانده^۹ بررسی کرد (امبرستون و رایس، ۲۰۰۰؛ مولنار و سیجتسما، ۲۰۰۰).

1. Item Response Theory (IRT)
2. latent trait
3. output

4. common factor
5. unidimensionality
6. local independence

7. residual
8. monotonicity
9. rest scores

۳) رابطه بین خصیصه مکنون و پاسخ مشاهده شده شکل معینی دارد (لرد، ۱۹۸۰؛ امبرستون، ۱۹۹۶). خطی که خصیصه مکنون و احتمال پاسخ را به یکدیگر مربوط می‌سازد خم ویژه سؤال^۱ نام دارد. مدل استاندارد ریاضی که برای خم ویژه سؤال به کار می‌رود شکل تراکمی تابع منطقی^۲ است. این تابع، خانواده‌ای از خمها را تعریف می‌کند که به سبب سادگی کار با آن بر سایر مدلها ترجیح داده شد (بیکر، ۲۰۰۳/۱۳۸۱).

مفاهیم نظریه سؤال پاسخ

الف) خم ویژه سؤال: بر پایه نظریه سؤال پاسخ برای هر سطح از توانایی، احتمال معینی وجود دارد که به دست آوردن پاسخ درست به سؤال توسط امتحان‌شونده را در آن سطح مشخص می‌سازد. این احتمال که با نماد $p_{(0)}$ نشان داده می‌شود. چنانچه مقادیر $p_{(0)}$ به عنوان تابع توانایی رسم شود، نتیجه یک خم هموار خواهد بود (نمودارهای شکل ۱). احتمال پاسخ درست در پایین‌ترین سطوح توانایی نزدیک به صفر است، و بعد از آن افزایش می‌یابد تا وقتی که سطح توانایی به بالاترین حد خود می‌رسد. در این هنگام احتمال پاسخ درست نزدیک به ۱/۰ خواهد بود. هر یک از سؤالهای منفرد سازنده تست دارای یک خم ویژه خاص آن سؤال است.

ب) پارامترهای سؤال: خم ویژه سؤال دارای سه پارامتر است که برای توصیف آن به کار می‌رود. ۱) پارامتر دشواری سؤال (b) جایگاه سؤال را در مقیاس توانایی توصیف می‌کند و متناظر با نقطه‌ای است که برای آن $p_{(0)} = 0.5$ باشد. وقتی سؤال آسان باشد این مقدار در سطح پایین توانایی، و زمانی که سؤال دشوار باشد این مقدار در سطح بالای توانایی قرار دارد. مقادیر درجه دشواری سه سؤال مختلف در نمودار الف شکل ۱ نشان داده شده است، ۲) قدرت تشخیص سؤال (a) نشان می‌دهد سؤال تا چه اندازه می‌تواند بین امتحان‌شوندگانی که توانایی آنها پایین‌تر از جایگاه سؤال است با امتحان‌شوندگانی که توانایی آنها بالاتر از این جایگاه قرار دارد، تمایز ایجاد کند. این ویژگی اساساً منعکس‌کننده شیب خم ویژه سؤال در بخش میانی آن است. هر چه شیب خم زیادتر باشد، قدرت تشخیص سؤال بیشتر است و مقادیر قدرت تشخیص سه سؤال مختلف در نمودار ب شکل ۱ نشان داده شده است و ۳) پارامتر حدس (c) بیانگر به دست آوردن پاسخ درست سؤال از طریق حدس محض و در واقع مجانب پایین خم یا نقطه پایینی آن وقتی است که به سمت اعداد منفی در محور توانایی حرکت می‌کند. مقدار (c) به عنوان تابعی از سطح توانایی تغییر نمی‌کند. از این رو، احتمال به دست آوردن پاسخ درست از طریق حدس برای آزمودنیهایی که دارای بالاترین و پایین‌ترین سطح توانایی باشند، یکسان است. مقادیر حدس سه سؤال مختلف در نمودار پ شکل ۱ نشان داده شده است.

پ) پارامتر آزمودنی: پارامتر آزمودنی بیانگر مقدار خصیصه مکنون در فرد یا جایگاه وی در مقیاس توانایی است. برآورد پارامتر آزمودنی براساس نمره کل فرد در اندازه‌گیری به دست می‌آید. این نمره در واقع نمره وزن دار فرد بر پایه مدلی است که شامل پارامترهای تشخیص باشد (هامبلتون و دیگران، ۱۹۹۱). معمولاً در دو مورد نمی‌توان توانایی آزمودنی را برآورد کرد. نخست زمانی که آزمودنی به هیچ یک از سؤالها پاسخ درست ندهد که برآورد توانایی متناظر با آن بی‌نهایت منفی است. دوم زمانی که آزمودنی به همه سؤالها پاسخ درست دهد که برآورد توانایی متناظر با آن بی‌نهایت مثبت است (بیکر، ۲۰۰۲/۱۳۸۱).

ت) نامتغیر بودن^۳ پارامترها: بنا بر نظریه سؤال پاسخ، پارامترهای سؤال در بین گروههای مختلف نامتغیر است. به بیان دیگر، اگر این پارامترها در گروههای نمونه مختلف برآورد شود، باید مقادیر یکسانی داشته باشند. زیرا بنا بر تعریف، چنانچه قسمتی از خم در اختیار باشد، می‌توان بقیه آن (یعنی مقادیر مورد انتظار) را به دست آورد (هالین و دیگران، ۱۹۸۳). بدین ترتیب پارامترهای سؤال وابسته به سطح توانایی امتحان‌شوندگان نیست. از این رو، رها از گروه نمونه^۴ و در واقع ویژگی خاص خود سؤال و نه ویژگی گروهی است که به آن پاسخ داده‌اند (بیکر، ۲۰۰۲/۱۳۸۱). چنانکه لرد (۱۹۸۰) نیز بیان می‌کند ثبات و عدم تغییر پارامترهای سؤال در بین گروههای مختلف یکی از مهمترین ویژگیهای نظریه سؤال

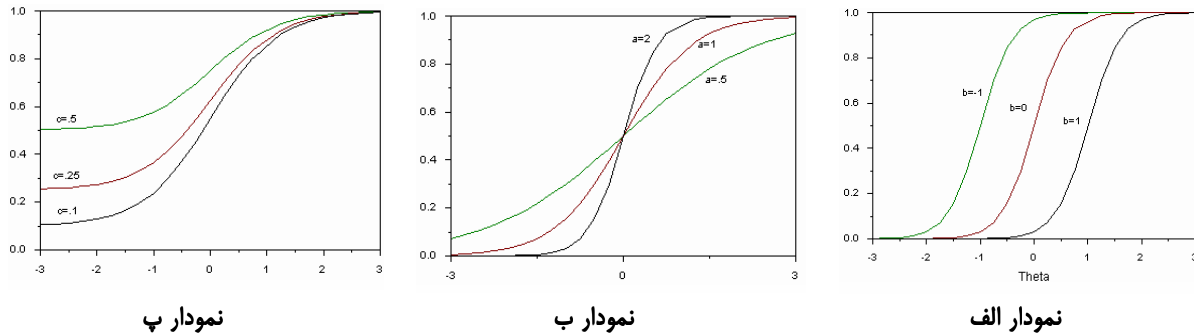
1. Item Characteristic Curve (ICC)

2. logic function

3. invariance

4. sample free

پاسخ است. زیرا این انتظار وجود دارد که پارامترها یا مشخصه‌های سؤال توصیف‌کننده سؤال و تست باشد، و گروه مورد سنجش از طریق پارامترهای توانائی توصیف و تبیین شوند. این اصل امکان بررسی سهم هر سؤال را به گونه انفرادی و نیز تصمیم‌گیری در بار حذف یا اضافه نمودن آن را در تست فراهم می‌آورد.



شکل ۱: نمودارهای خم ویژه سه سؤال بر پایه پارامترهای درجه دشواری، قدرت تشخیص و حدس مختلف

افزون بر این، توانایی آزمودنی نیز نسبت به سؤالهایی که برای برآورد آن به کار می‌رود، نامتغیر است. در اینجا برای روشن شدن مطلب بهتر است به مطالعه لرد (۱۹۸۰) اشاره شود. لرد دو آزمون سنجش خزانه واژگان را روی ۱۸۳۰ نفر اجرا و مشاهده کرد که رابطه بین نمره‌های توانائیهایی حاصل از این دو تقریباً یک خط راست است. اگر دو تست برای هر فرد توانائی یکسانی برآورد کند، قاعدتاً نتیجه یک خط راست با رابطه کامل را به نمایش می‌گذارند. یعنی توانائی آزمودنیها که براساس نظریه سؤال پاسخ به دست آمده است، تأثیر چندانی از ویژگیهای سؤالهای دو تست نمی‌پذیرد.

ث) آگاهی: یکی از ویژگیهای برجسته IRT مفهوم اعتبار است. به گونه سنتی، اعتبار به دقت اندازه‌گیری (میزانی از اندازه‌گیری که رها از خطاست) اشاره دارد. اما بر پایه IRT، دقت در بین دامنه نمره‌های تست یکسان نیست. برای نمونه، نمره‌ها در دامنه‌های کناری تست نسبت به نمره‌هایی که به دامنه میانی نزدیکتر است، به گونه کلی بیشتری دارد. آگاهی از لحاظ آماری به معنای مفهوم مقابل میزان دقت در برآورد یک پارامتر است. از این رو، IRT، مفهوم آگاهی (به معنای کاهش عدم حتمیت^۲) سؤال و تست را توسعه داده و جایگزین اعتبار کرده است (امبرستون و رایس، ۲۰۰۰؛ هامبلتون و دیگران، ۱۹۹۱؛ رایت، ۱۹۹۲). چون در این نظریه هر یک از سؤالهای تست، صفت مکتون زیربنایی را اندازه می‌گیرد، در نتیجه مقدار آگاهی مبتنی بر یک سؤال منفرد است و می‌توان آن را برای هر یک از سطوح توانایی محاسبه کرد. چون هر تست به منظور برآورد توانایی آزمودنی به کار می‌رود، مقدار آگاهی را که آن تست در هر یک از سطوح توانایی به دست می‌دهد نیز می‌توان محاسبه کرد. افزون بر این، چون هر تست شامل یک مجموعه سؤال است، بنابراین آگاهی تست در هر سطح معینی از توانایی برابر با جمع ساده آگاهیهای سؤال در آن سطح خواهد بود.

مزایا و کاربرد IRT

۱) IRT خطای استاندارد اندازه‌گیری را بر مبنای تابع آگاهی تست، و نه متوسط خطای استاندارد برای همه سطوح توانایی در نظریه کلاسیک اندازه‌گیری، برآورد می‌کند. این مطلب به متخصصان اجازه می‌دهد تا سؤالهایی را برگزینند که در دامنه خاصی از توانایی یا خصیصه را با بیشترین دقت اندازه‌گیری کنند.

۲) بر پایه IRT بین امکان وجود دارد که آزمونهای دقیقی برای اندازه‌گیری هم‌ارزی^۳ گروههای مختلف تجربی به عمل آید. این امکان، به ویژه در پژوهشهای بین فرهنگی که انتظار می‌رود خصیصه مورد اندازه‌گیری در گروههای مختلف، میانگین متفاوتی داشته باشند از اهمیت زیادی برخوردار است.

۳) روشهای IRT می‌تواند تورش^۱ سؤال را از تفاوت‌های حقیقی در خصیصه اندازه‌گیری شده متمایز سازد. اگر خم ویژه سؤال برای دو جامعه یکسان باشد، سؤال تورش‌دار نخواهد بود. اما اگر خم ویژه سؤال یکسان نباشد، به معنای آن است که رفتار و کارکرد سؤال در گروه‌های مختلف متفاوت است (کیم، کوهن و پارک، ۱۹۹۵).

۴) IRT امکان سنجش برازش یافته^۲ را فراهم می‌آورد که در آن برآورد نمره حقیقی بر مبنای تعداد پاسخهای درست قرار ندارد. این مطلب به پژوهشگران اجازه می‌دهد تا به افراد مختلف، سؤالهای متفاوتی بدهند اما جایگاه آنان کماکان در یک مقیاس قرار داشته باشد.

۵) IRT سنجش انطباقی^۳ رایانه‌ای را تسهیل می‌سازد. بدین ترتیب، سؤالهایی انتخاب می‌شوند که برای هر آزمودنی بیشینه آگاهی را به دست می‌دهد. این امکان می‌تواند به گونه چشمگیری هزینه و زمان اجرای تست را کاهش دهد (هالین و دیگران، ۱۹۸۳).

۶) نمره‌های حقیقی در نظریه کلاسیک اندازه‌گیری دامنه محدودی دارد و محصور به تعداد سؤالهای تست است. در حالی که در IRT، خصیصه مکنون (θ) دامنه‌ای از بی‌نهایت منفی تا بی‌نهایت مثبت را می‌پذیرد.

۷) در IRT امکان وزن‌دهی دقیق و اصولی به سؤالها وجود دارد. در نتیجه به همه سؤالها نمره یکسانی داده نمی‌شود.

۸) پارامترهای سؤال و آزمودنی در یک مقیاس مشترک و واحد بیان می‌شود، و در نتیجه تطبیق دشواری و تناسب هر سؤال با توانایی افراد آسان‌تر خواهد بود.

۹) در نظریه کلاسیک مقیاس اندازه‌گیری حداکثر فاصله‌ای و حتی احتمالاً در مواردی رتبه‌ای است. در حالی که روشهای IRT ممکن است حتی مقیاس نسبتی ایجاد کنند.

منابع

- بیکر، اف. بی. (۱۳۸۱). پایه‌های اساسی تئوری سؤال پاسخ (نظریه‌های جدید روان‌سنجی). ترجمه: حیدرعلی هومن و علی عسگری. تهران: نشر پارسا (تاریخ انتشار اثر اصلی، ۲۰۰۲).
- Emberston, S. E. (1996). Item response theory models and spurious interaction effects in factorial ANOVA designs. *Applied Psychological Measurement, 20*, 201-212.
- Emberston, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Laurence Erlbaum Associates, INC.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Press.
- Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). *Item response theory: Applications to psychological measurement*. Homewood IL: Dow Jones-Irwin.
- Kim, S.H., Cohen, A.S., & Park, T.H. (1993). *Comparison of item parameters in multiple groups*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Molenaar, I. W., & Sijtsma, K. (2000). MIP5 for windows. A program for Mokken scale analysis for polytomous item[computer software]. Groningen The Netherlands: iec ProGamma.
- Wright, B. D. (1992). IRT in the 1990s: Which models work best? *Rasch measurement transactions, 6* (1), 196-200.