روی‌آوردهای نوین در روان‌سنجی

قسمت چهارم: مدلهای نظریه سؤال ـ پاسخ، مدلهای دو ارزشی

New Approaches to Psychometrics

Part Four: Models of Item Response Theory, Dichotomous Models

دکتر علی عسگری
دانشگاه علوم بهزیستی
و توانبخشی

Ali Asgari, PhD
University of Welfare and
Rehabilitation

برای بررسی داده‌های دو ارزشی مدلهای متعددی در *IRT* وجود دارد. اما انتخاب یک مدل باید بر پایه ملاحظات نظری و تجربی، مانند برازش داده ـ مدل[1] صورت گیرد. هر مدل با استفاده از ارزش عددی یک یا چند پارامتر، خم ویژه بخصوصی را معین می‌کند و از لحاظ انتقال اطلاعات درباره ویژگیهای فنی سؤالها ابزار مفیدی به‌شمار می‌آید (بیکر، ۲۰۰۲/۱۳۸۱). به منظور محاسبه احتمال پاسخ درست در سطوح مختلف توانایی، هر یک از مدلها از یک معادله ریاضی استفاده می‌کنند.

به گونه کلی، مدلهای *IRT* را می‌توان به دو خانواده مدلهای تک‌بعُدی[2] و چندبعُدی[3] تقسیم کرد (تیسن و اورلاندو، ۲۰۰۱، رایت، ۱۹۹۲). مدلهای تک‌بعُدی مستلزم تنها یک بعد واحد از خصیصه (توانایی) است. در حالی که، در مدلهای چند بعُدی فرض بر آن است که داده‌ها از خصایص چندوجهی یا چندگانه[4] به دست می‌آیند. با وجود این، به دلیل پیچیدگی فزاینده مدلهای چند بعُدی، در بیشتر پژوهشها و کاربردهای *IRT* از مدلهای تک‌بعُدی استفاده می‌شود.

افزون بر این، مدلهای *IRT* بر پایه تعداد پاسخهای نمره‌گذاری شده نیز طبقه‌بندی می‌شود. یک سؤال چند گزینه‌ای در واقع یک مدل دو ارزشی است حتی اگر دارای چهار یا پنج گزینه باشد. زیرا فقط به گونه درست / نادرست (صحیح / غلط) نمره‌گذاری می‌شود. طبقه دیگر مدلها، برای داده‌های چند ارزشی به کار می‌روند که در آن مقدار نمره هر پاسخ متفاوت است. برای نمونه، چنانکه در قسمت قبل بیان شد، مدل چند ارزشی راش در واقع تعمیم این مدل به داده‌های مربوط به دو یا چند طبقه مرتب شده است. مدلهای دو ارزشی *IRT* را می‌توان بر پایه تعداد پارامترهایی که در آنها به کار می‌رود نیز توصیف کرد. بدین ترتیب، در مدل سه پارامتری، هر سه پارامتر دشواری (*b*، جایگاه سؤال)، قدرت تشخیص (*a*، شیب سؤال) و حدس (*c*، مجانب[5] پایین خم ویژه) به‌کارمی‌روند. در حالی که در مدل دو پارامتری فرض بر این است که داده‌ها کمینه حدس را دارند اما بر اساس جایگاه و قدرت تشخیص سؤال متغیرند، و سرانجام در مدل تک پارامتری مقدار حدس و قدرت تشخیص سؤالها برابر فرض می‌شوند و تنها جایگاه سؤال برآورد می‌شود. افزون بر این، از لحاظ نظری یک مدل چهار پارامتری نیز وجود دارد که پارامتر چهارم آن مجانب بالایی خم ویژه است. اما این مدل به‌ندرت به‌کار برده می‌شود (هامبلتون و سوامیناتان،۱۹۹۱؛ تیسن و اورلاندو، ۲۰۰۱).

مدل تک‌پارامتری

مدل منطقی تک‌پارامتری ساده‌ترین مدل *IRT* است و همانگونه که از نام آن پیداست برای کاربرد نظریه سؤال ـ پاسخ تنها برآورد یک پارامتر، یعنی درجه دشواری (*b*)، برای آن لازم است. بر پایه این مدل، پارامتر قدرت تشخیص برای همه سؤالها ثابت و برابر با ۱/۰ قرار داده می‌شود؛ یعنی تنها پارامتر دشواری می‌تواند مقادیر مختلفی اختیار کند. معادله این مدل به قرار زیر است :

$$P_i(\theta)=\frac{1}{1+e^{-(\theta-b_i)}},$$

که در آن: $p(\theta)$ = احتمال پاسخ درست برای سطح معینی از توانایی، e برابر با ۲/۷۱۸، $(\theta- b)-$ = انحراف منطقی است.

1. model-data fit
2. unidimensional
3. multidimensional
4. multiple
5. asymptote

www.SID.ir

Archive of SID

برای نمونه، بر پایه فرمول بالا، احتمال پاسخ درست به سؤالی با درجه دشواری ۱/۰ برای فردی که در سطح توانایی ۳/- = *(θ)* قرار دارد، برابر با ۰/۰۲ خواهد بود. چون مقدار پارامتر تشخیص در معادله بالا که همواره برابر با ۰/۱ است، معمولاً در فرمول نشان داده نمی‌شود. بنا بر نظر هاروی و هامر (۱۹۹۹)، تک‌پارامتر، از لحاظ اجرایی، به نمره‌ای از توانایی اطلاق می‌شود که با احتمال ۵۰ درصدی پاسخ به یک سؤال انشایی سروکار دارد. چون مقدار پارامتر قدرت تشخیص در مدل راش برابر با مقدار ثابت ۱/۰ است، بیشینه مقدار تابع آگاهی حاصل از آن جایگاه محدودی دارد. به بیان دیگر، بیشینه مقدار تابع آگاهی سؤال برابر با ۰/۲۵ است. زیرا وقتی $p_{(\theta)}$ =۰/۵ باشد، $p_{(\theta)}\ q_{(\theta)}$ = ۰/۲۵ خواهد بود. از این رو، بیشینه مقدار نظری آگاهی برای یک تست در مدل تک‌پارامتری برابر با حاصلضرب تعداد سؤالها در عدد ۰/۲۵ خواهد بود.

مدل دو پارامتری

در این مدل، افزون بر جایگاه سؤال، قدرت تشخیص *(a)* نیز برآورد می‌شود و برای خم ویژه سؤال *(ICC)* این امکان فراهم می‌آید تا برای سؤالهای مختلف، شیبهای مختلفی را به نمایش بگذارد. پارامتر تشخیص این حقیقت را آشکار می‌کند که برخی سؤالها رابطه نیرومندتر (یا ضعیف[1]تری) با سازه بنیادی مورد اندازه‌گیری (توانایی) دارند. بنابراین، مقادیر بالاتر در این پارامتر، نشان‌دهنده رابطه محکم‌تر و سخت‌تر است (تورنتون، ۲۰۰۲). بنابر نظر هاروی و هامر (۱۹۹۹) قدرت تشخیص از اهمیت زیادی در *IRT* برخوردار است. زیرا مقدار آگاهی حاصل از یک سؤال را به گونه مستقیم تعیین می‌کند. بدین‌ترتیب، سؤالهایی که قدرت تشخیص بیشتری دارند، در صورت برابربودن سایر عوامل، آگاهی بیشتری در باره توانایی *(θ)* به دست می‌دهند. معادله تابع منطقی دو پارامتری به قرار زیر است :

$$P_i(\theta_j) = \frac{1}{1 + \exp[-Da_i(\theta_j - b_i)]},$$

برای نمونه، در ادامه مثال مربوط به مدل تک‌پارامتری، احتمال پاسخ درست برای سؤالی با همان درجه دشواری و قدرت تشخیص ۰/۵ برای فردی در سطح توانایی ۰/۳- = *(θ)*، برابر با ۰/۱۲ است. در اینجا ملاحظه می‌شود که احتمال پاسخ درست با احتساب قدرت تشخیص، افزایش یافته است. در مدل دو پارامتری، مانند مدل راش، دشواری سؤال نقطه‌ای در مقیاس توانایی است که احتمال پاسخ درست برای افرادی که در آن سطح از توانایی هستند برابر با ۰/۵ است. افزون بر این در این مدل (مانند مدل سه پارامتری) قبل از آنکه خم ویژه سؤال شیب زیادی پیدا کند، باید مقدار پارامتر تشخیص کاملاً بزرگ، دست کم (۱/۷) باشد (بیکر، ۲۰۰۲/۱۳۸۱). مقدار مثبت و بزرگ در مدل‌های راش و دو پارامتری موجب می‌شود که دنباله پایین خم به صفر نزدیک گردد.

مدل سه پارامتری

یکی از رایج‌ترین مدلهایی که روان‌شناسان از آن استفاده می‌کنند مدل منطقی سه‌پارامتری است. این مدل در ابتدا برای مدلسازی[2] داده‌های حاصل از توانایی شناختی به کار می‌رفت، اما به تازگی برای داده‌های شخصیتی نیز به‌کار می‌رود (امبرستون و رایس، ۲۰۰۰). مدل سه‌پارامتری کلی‌تر از مدلهای تک و دو پارامتری است. گرچه مدل دو پارامتری با یکی از حیاتی‌ترین ویژگیهای مدل راش، یعنی این اصل قطعی سروکار دارد که همه سؤالها از لحاظ قدرت تشخیص شبیه به یکدیگرند، اما واقعیت مهم دیگری را که ممکن است در بین سؤالهای مختلف وجود داشته باشد نادیده می‌گیرد. واقعیت مهمی که در اندازه‌گیریهای تستی وجود دارد این است که آزمودنیها می‌توانند پاسخ درست را از طریق حدس به دست آورند. بنابراین، احتمال پاسخ درست، مؤلفه کوچکی را که ناشی از حدس است نیز در بر می‌گیرد. هیچ‌یک از دو مدل خم ویژه سؤال که بحث آن گذشت پدیده حدس را به حساب نمی‌آورند.

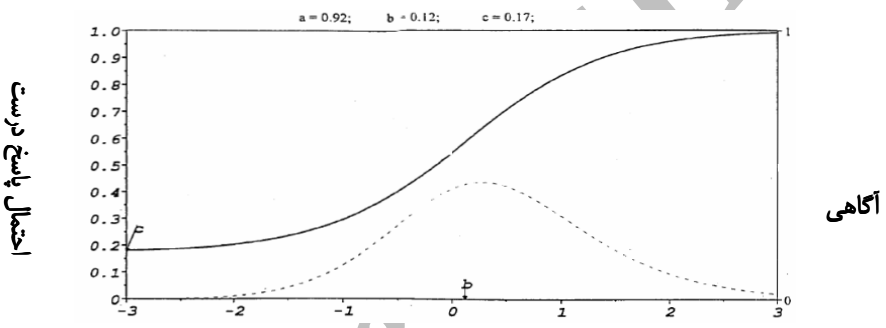
برن‌بام[3] (۱۹۶۸، نقل از بیکر، ۲۰۰۲/۱۳۸۱) در مدل منطقی دو پارامتری تغییری ایجاد کرده است که بر پایه آن

1. feeble 2. modeling 3. Birnbuam

Archive of SID
www.SID.ir

پارامتر دیگری را که بیانگر سهم حدس در احتمال پاسخ درست است نیز در بر می‌گیرد. بنابراین، مدل سه پارامتری این واقعیت را که مجانب پایین خم ویژه سؤال با احتساب حدس، ممکن است مستلزم پذیرش مقادیر غیرصفر[1] باشد، آشکار می‌کند (هاروی و هامر، ۱۹۹۹). یکی از اثرات جانبی کاربرد پارامتر حدس، تغییر تعریف پارامتر دشواری است. بر پایه دو مدل پیشین، پارامتر b نقطه‌ای در مقیاس توانایی است که احتمال به‌دست آوردن پاسخ درست در آن برابر با ۰/۵ است. اما اکنون حد پایین خم ویژه سؤال نه عدد صفر بلکه مقدار c است. از این رو، می‌توان پارامتر c را به عنوان پایین‌ترین سطح مقدار احتمال پاسخ درست تعریف‌کرد. بنابراین، پارامتر دشواری معرف نقطه‌ای در مقیاس توانایی است که احتمال پاسخ درست آن دقیقاً برابر با متوسط c و ۰/۱ است.

نمونه‌ای از خم ویژه و تابع آگاهی سؤال بر پایه مدل سه پارامتری در شکل ۱ نشان داده شده است. در این نمودار می‌توان به راحتی مشاهده کرد که چگونه پارامترهای اول و دوم در متأثر از پارامتر سوم (حدس) هستند. سطح دشواری یا بتا برابر با ۰/۱۲ است. در حالی که پارامتر حدس برابر با ۰/۱۷ و قدرت تشخیص نیز در سطح ۰/۹۲ قرار دارد. ویژگیهای این سؤال نشان‌می‌دهد که برای یک آزمون توانایی بسیار مناسب است. معادله تابع منطقی سه پارامتری به قرار زیر است:



شکل ۱ : نمودار خم ویژه و تابع آگاهی سؤال در مدل سه پارامتری

$$P_i(\theta_j) = c_i + \frac{1 - c_i}{1 + \exp[-Da_i(\theta_j - b_i)]},$$

در ادامه مثالهای بالا، احتمال پاسخ درست برای سؤالی با همان مقادیر دشواری و قدرت تشخیص، که احتمال حدس در آن برابر با ۰/۲ باشد برای فردی در سطح توانایی ۰/۳- = (θ)، برابر با ۰/۲۹۵ خواهد بود. در اینجا نیز مشاهده می‌شود که چگونه احتمال پاسخ درست با محاسبه پارامتر سوم (حدس) تغییر می‌کند.

در مورد کارکرد و مبانی مدل سه پارامتری بین صاحبنظران اختلاف‌نظر وجود دارد (موناهان، لی و بانکمن، ۲۰۰۷؛ بیکر، ۱۳۸۱/۲۰۰۲؛ تورنتون، ۲۰۰۰). به اعتقاد بیکر (۱۳۸۱/۲۰۰۲) تغییری که برن باوم در مدل دو پارامتری، به منظور گنجاندن عامل حدس ایجاد کرده، موجب شده است برخی از ویژگیهای خوب ریاضی تابع منطقی از دست برود و واقعیت این است که مدل سه پارامتری از نظر فنی یک مدل منطقی (لوجستیک) به حساب نمی‌آید. افزون بر این در مدل سه پارامتری به سبب وجود پارامتر حدس، خم ویژه تست در مقایسه با مدل دو پارامتری، با پارامترهای دشواری و قدرت تشخیص یکسان، خطی‌تر و سطح کلی تابع آگاهی آن پایین‌تر است. در شرایطی که مقادیر a و b در هر دو مدل یکسان باشد، تابع آگاهی مدل دو پارامتری برابر با حد بالایی تابع آگاهی در مدل سه پارامتری خواهد بود. اما نتایج پژوهش هاسکنز و دی بوئک (۲۰۰۱) کارکرد مدل سه پارامتری را به گونه دیگری نشان می‌دهد. به اعتقاد این پژوهشگران:

> بر پایه مدل‌سه‌پارامتری چهارچوبی برای مدلسازی داده‌های مؤلفه‌ای[2] ارائه می‌شود که از مدلهای نظریه سؤال ـ پاسخ برای سؤالهای چند ارزشی استفاده می‌کند. این چهارچوب دقتهای پاسخ[3] را نسبت به تکالیف پیچیده شناختی، که براساس عناصر اساسی‌تر مانند ساختارهای دانش، فرایندهای شناختی و راهبردها تجزیه می‌شود، مدلسازی می‌کند (ص ۱۹).

1. nonzero
2. componential data
3. response accuracies

Archive of SID
www.SID.ir

مدل چهار پارامتری

یکی دیگر از مدلهای کمتر شناخته‌شده *IRT*، مدل چهار پارامتری است. در این مدل افزون‌بر سه پارامتر *a*، *b* و *c*، پارامتر چهارم، یعنی مجانب بالای خم ویژه سؤال (ζ) نیز برآورد می‌شود. بر پایه این مدل، احتمال عدم‌موفقیت حتی برای تواناترین آزمودنیها در سؤالهای بسیار دشوار قابل ملاحظه و برآورد خواهد بود. با آنکه مدل چهار پارامتری از لحاظ مفهومی بسیار جالب است، از لحاظ عملی مطالعات کمی در مورد آن انجام شده است. زیرا تعداد پارامترهایی که باید برای آن برآورد شود بسیار زیاد است (سیجتسما و همکر، ۲۰۰۰). در *IRT* مدلهای دیگری مانند مدل منطقی یک ـ پارامتری[3] (OPLM)، بتای دو جمله‌ای مرکب چهار پارامتری[4] (4PBCB) توسعه یافته که شرح آنها در این مقاله کوتاه نمی‌گنجد. از خوانندگان علاقمند دعوت می‌شود تا برای مطالعه بیشتر به منابعی معتبر در این زمینه مانند موناهان، لی و بانکمن (۲۰۰۷)، سیجتسما و همکر (۲۰۰۰)، لرد (۱۹۸۰) مراجعه کنند.

با آنکه مدلهای نظریه سؤال ـ پاسخ امکانات گسترده‌ای در اختیار تست‌سازان و روان‌شناسان قرار داده است، به اعتقاد برخی صاحبنظران (هاروی و هامر، ۱۹۹۹؛ تورنتون، ۲۰۰۲) دانستن محدودیتهای آن از اهمیت زیادی برخوردار است. پژوهشگران باید همواره به یاد داشته باشند که هنگام اندازه‌گیری، در واقع در حال برازش یک مدل ریاضی با مفروضه‌ها و محدودیتهای معینی هستند و هیچ تضمینی وجود ندارد که مدلهایی که برای یک راهبرد معین در *IRT* به کار می‌رود برازش کافی با داده‌ها داشته باشد.

منابع

بیکر، اف. بی. (۱۳۸۱). *پایه‌های اساسی در تئوری سؤال پاسخ (نظریه‌های جدید روان‌سنجی)*. ترجمه حیدرعلی هومن و علی عسگری، تهران، نشر پارسا (تاریخ انتشار اثر اصلی، ۲۰۰۲).

Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Press.

Harvey, R., & Hammer, A. (1999). Item response theory. *Counseling Psychologist, 27*, 353-383.

Hoskens, M., & DeBoeck, P. (2001). Multidimensional componential item response theory models for polytomous items. *Applied Psychological Measurement, 25*, 19-37.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum

Monahan, P. O., Lee, W., & Ankenmann, R. D. (2007). Generating Dichotomous Item Scores with the Four-Parameter Beta Compound Binomial Model. *Journal of Educational Measurement, 44*, 211–225.

Sijtsma, K., & Hemker, B. T. (2000). A taxonomy of IRT models for ordering persons. *Journal of Educational and Behavioral Statistics, 25*, 391-415.

Thornton, A. (2002). *A primer on 2- and 3- parameter item response theory models*. Paper presented in Annual Meeting of the College of Education, University of North Texas (2nd, Denton, TX, February 1).

Thissen, D. & Orlando, M. (2001). Item response theory for items scored in two categories. In D. Thissen & Wainer, H. (Eds.), *Test scoring* (pp. 73-140). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Wright, B. D. (1992). IRT in the 1990s: Which models work best? *Rasch Measurement Transactions, 6 (1)*, 196-200.

1. One-Parameter Logistic Model
2. Four Parameter Beta Compound Binomial

www.SID.ir