روی‌آوردهای نوین در روان‌سنجی

قسمت سوم: مدلهای نظریه سؤال ـ پاسخ، مدلهای راش

New Approaches to Psychometrics
Part Three: Models of Item Response Theory, Rasch Models

علی عسگری
دانشجوی دکتری
دانشگاه تهران

Ali Asgari
PhD Candidate
Tehran University

در هفتاد سال گذشته نظریه‌پردازان متعددی تلاش کرده‌اند تا نشان دهند که چگونه می‌توان از اندازه‌ها و فراوانیهای عینی[1]، اندازه‌های انتزاعی[2] به دست آورد. یکی از عملی‌ترین و رایج‌ترین روی‌آوردهایی که برای این منظور به‌کار می‌رود، مدل راش[3] است. جورج راش، ریاضیدان دانمارکی، این روی‌آورد را در سال ۱۹۵۳ و به منظور تحلیل پاسخهای یک رشته از آزمونهای خواندن به وجود آورد. با آنکه وی را پدر تحلیل راش می‌دانند، اما بنجامین رایت[4] را باید قیم قانونی آن دانست. رایت و همکارانش در دانشگاه شیکاگو روشهای پیشرفته و ابزارهای تحلیل راش را توسعه، و کاربرد آن را در حوزه‌های مختلف علمی ارتقا بخشیدند (ماسوف و فیشر، ۲۰۰۲).

مدلهای راش در واقع روی‌آوردی ریاضی برای آزمون این فرضیه است که اندازه‌های مربوط به معنا[5] و واحد یک سازه را می‌توان از ابزاری که برای آن خصیصه تهیه شده است به دست آورد. وقتی داده‌ها با این مدلها برازش پیدا می‌کنند به معنای آن است که ابزار اندازه‌گیری و اندازه‌ها در یک واحد فاصله‌ای مشترک مقیاس‌بندی شده‌اند و می‌توانند در انواع یا شکلهای مختلف آن ابزار و نیز در بین نمونه‌های مختلف یک جامعه ثابت باقی بمانند (رایت و استون، ۱۹۷۹).

مدلهای راش، در واقع نوعی آزمون همسانی درونی[6] در نظریه سؤال ـ پاسخ‌اند که برای داده‌های دوارزشی و چند ارزشی به کار می‌روند. در این مدلها نیز مانند مقیاسهای گاتمن[7]، فرض براین است که همه سؤالها و مواد یک آزمون که یک سازه را اندازه‌گیری می‌کنند، یک نوع رابطه مرتب شده[8] را تشکیل می‌دهند. یک آزمون ممکن است دارای همسانی درونی مرتب‌شده‌ای باشد، حتی اگر مجموعه سؤالهای آن همبستگی بالایی با هم نداشته باشند (همسانی درونی جمع‌پذیر[9]، مانند آنچه از طریق آلفای کرونباخ[10] یا تحلیل عاملی[11] آزمون می‌شود). همسانی درونی مرتب شده بیانگر وجود عامد دشواری است. بدین ترتیب، یک سؤال دشوار می‌تواند پاسخ به سؤالهای با دشواری کمتر را پیش‌بینی کند اما عکس آن امکان‌پذیر نیست (رایت، ۱۹۹۶).

وقتی پژوهشگران برای رواسازی یک مجموعه از متغیرهای نشانگر در یک مقیاس از تحلیل عاملی استفاده می‌کنند، فرض را بر این قرار می‌دهند که با یک مدل خطی و جمع‌پذیر روبه‌رو هستند. خطی بودن بخشی از همبستگی و مبنایی برای خوشه‌بندی[12] متغیرهای نشانگر در یک عامل است. در جمع‌پذیری نیز فرض بر این است فقط زمانی معنای همه سؤالها دارای همسانی درونی است، که همبستگی بالایی با یکدیگر داشته باشند. با وجود این، ممکن است که سؤالها فاقد همبستگی درونی بالا، اما دارای رابطه مرتب‌شده نیرومندی باشند (رایت، ۱۹۸۵). به همین دلیل بسیاری از پژوهشگران ترجیح می‌دهند برای ساخت و توسعه مقیاسها به جای مدلهای جمع‌پذیر مانند آلفای کرونباخ و تحلیل عاملی، از مدلهای راش استفاده کنند. زیرا این مدلها نه‌تنها روابط جمع‌پذیر بین متغیرهای نشانگر، بلکه رابطه ترتیبی سؤالها (مانند ترتیب

1. concrete
2. abstract
3. Rasch
4. Wright, B. D.
5. meaning
6. internal consistency
7. Guttmann
8. ordered relationship
9. additively
10. cronbach
11. factor analysis
12. clustering

Archive of SID
www.SID.ir

دشواری) را نیز به حساب می‌آورند (تنورگرت، گیلپسی و کینگما، ۱۹۹۳). نظریه زیربنایی مدلهای راش در بسیاری جنبه‌ها شبیه به نظریه سؤال ـ پاسخ است. به بیان دیگر، مدل راش برای داده‌های دو ارزشی اغلب به عنوان مدل‌تک پارامتری نظریه سؤال ـ پاسخ در نظر گرفته می‌شود. اما هواداران این مدل، آن را دارای ویژگی خاصی می‌دانند که از مدلهای IRT متمایز است. به گونه اختصاصی، ویژگی معرف مدلهای راش صورتبندی انتزاعی[1] و ریاضی مقایسه نامتغیر است که می‌تواند برای اندازه‌گیری موفقیت‌آمیز سازه‌ها یک ملاک معتبر فراهم کند (سادوس، گارمندی، کیوز و الیوت، ۲۰۰۴). این ویژگی انتزاعی، مدلهای راش را از سایر مدلهایی که برای پاسخ به سؤالها یا مواد آزمون به‌کار می‌روند متمایز و آن را به عنوان مدلهای ایده‌آل یا استاندارد مطرح می‌سازد.

بنا بر نظر آندریش (۲۰۰۴) دیدگاه[2] یا پارادایم[3] مدلهای راش به گونه بارزی با سایر مدلهای اندازه‌گیری تفاوت دارد. در اغلب مدلها هدف اصلی توصیف مجموعه‌ای از داده‌هاست. به همین منظور پارامترها تعدیل می‌شوند و برپایه اینکه چگونه با داده‌ها برازش می‌یابند، رد یا پذیرفته می‌شوند. اما هدف از به‌کار بردن مدل راش به دست‌آوردن داده‌هایی است که با مدل برازش داشته باشد. منطق زیربنایی این دیدگاه آن است که مدلهای راش مستلزم شرایطی هستند که برای اندازه‌گیری باید برآورده شوند. درست همانگونه که عموماً در اندازه‌گیریهای علم فیزیک وجود دارد.

برای درک این منطق زیربنایی بیان مثالی در اندازه‌گیری وزن می‌تواند مفید باشد. فرض کنید وزن شئ A در یک موقعیت به گونه قابل ملاحظه‌ای بیشتر از وزن شئ B اندازه‌گیری شده است. سپس بلافاصله در یک موقعیت دیگر، این وزن شئ B است که بیشتر از وزن A به دست می‌آید. در اینجا شرط اساسی اندازه‌گیری، یعنی یکسان و نامتغیربودن نتایج حاصل از مقایسه دو اندازه‌گیری، صرف‌نظر از سایر عوامل، برآورده نشده است. این شرط اساسی در ساختار انتزاعی مدل راش است. بنابراین، مدلهای راش برای تناسب و برازش یافتن با داده‌ها، تغییر و تعدیل نمی‌شوند. بلکه روش اندازه‌گیری باید تغییر یابد تا این شرط را برآورده سازد. درست همانگونه که در مثال بالا مقیاس وزن باید تغییر کند. زیرا بین دو شئ در دو اندازه‌گیری جداگانه نتایج متفاوتی به دست داده است. علاوه بر این، در پارادایم مدلهای راش تأکید بر مطالعه و تعیین بی‌نظمی[4] در داده‌هاست که از طریق این مدل آشکار می‌شود (رایت، ۱۹۹۶).

خانواده مدلهای راش

لاینرس (۲۰۰۶) مدلهای راش را در دو طبقه کلی دو ارزشی و چند ارزشی به شرح زیر تقسیم‌بندی می‌کند:

مدل دو ارزشی: این مدل که در آن پاسخها به دو طبقه (بلی ـ خیر، درست ـ نادرست) تقسیم می‌شوند، شناخته‌شده‌ترین و رایج‌ترین مدل راش و دارای تابع ساده منطقی است. برای داده‌های دو ارزشی جایگاه یک سؤال در یک مقیاس، متناظر با جایگاه آزمودنی در نقطه‌ای است که احتمال موفقیت برابر با ۰/۵ است. به گونه کلی، احتمال پاسخ درست آزمودنی به یک سؤال با درجه دشواری کمتر از جایگاه آزمودنی، بیشتر از ۰/۵ و احتمال پاسخ درست آزمودنی به یک سؤال با درجه دشواری بالاتر از جایگاه آزمودنی، کمتر از ۰/۵است. وقتی پاسخ فرد بر پایه دشواری سؤال از کمترین تا بیشترین فهرست شود، بیشترین شباهت را به الگوی گاتمن دارد. با این فرمول: $\text{Loge}\ (P_{ni1} / P_{ni0}) = B_n - D_i$

که در آن:

P_{ni}= احتمال آنکه آزمودنی n که با سؤال i روبه‌رو می‌شود در طبقه j اندازه‌گیری شود.

B_n = توانایی فرد n

D_{ij} = دشواری سؤال i. نقطه‌ای که در آن بالاترین و پایین‌ترین طبقه‌های سؤال احتمال برابر دارند.

F_{ij} = اندازه‌مدرج‌کردن طبقه j-1 . نقطه‌ای که در آن طبقه‌های j-1 و j نسبت به اندازه سؤال احتمال برابر دارند.

مدلهای چندارزشی: مدلهای چندارزشی راش نخستین بار توسط آندریش (۱۹۷۸، ۲۰۰۴) و به منظور کاربرد مدل راش

1. formal 2. perspective 3. paradigm 4. anomalies

Archive of SID
www.SID.ir

برای داده‌های حاصل از مقیاس لیکرت ارائه شد. این مدلها در واقع تعمیم مدلهای دوارزشی و نوعی مدل اندازه‌گیری است که در زمینه‌هایی به کار می‌رود که هدف از آن اندازه‌گیری صفت یا توانایی از طریق فرایندی است که در آن پاسخ به سؤالها با اعداد صحیح متوالی نمره‌گذاری شود این مدل را می‌توان در مقیاسهای لیکرت، درجه‌بندی و نیز سؤالهای مربوط به اندازه‌گیریهای ترتیبی که در آنها نمره‌های متوالی بالاتر بیانگر سطح فزاینده پیشرفت و توانمندی است به کار برد.

از سوی دیگر، مدلهای چند ارزشی یک اندازه‌گیری احتمالی کلی و دارای این ویژگی متمایز است که برای کاربرد نمره‌های عددی متوالی یک بنیان نظری محکم فراهم آورده است. افزون براین ویژگی، مدلهای چند ارزشی امکان آزمون جدی این فرضیه را فراهم می‌آورد که طبقه‌های پاسخ، معرف سطح افزایشی یک خصیصه یا صفت مکنون است. از این رو داده‌ها، مرتب شده به حساب می‌آیند. در این مدل، نمره یک سؤال معین در واقع فراوانی تعداد جایگاه آستانه[1] در صفت مکنونی است که آزمودنی از آن بالاتر قرار دارد. جایگاه آستانه بر روی پیوستار مکنون معمولاً از ماتریس سؤال ـ پاسخ و از طریق فرایند برآورد بیشینه احتمال شرطی[2] استنباط می‌شود.

به گونه کلی، شاخص اصلی فرایند اندازه‌گیری در این مدل آن است که آزمودنیها در یک مجموعه طبقه‌های مرتب شده مجاور[3] گروه‌بندی شوند. شکل‌بندی پاسخهایی که در یک زمینه آزمایشی معین به‌کار می‌روند، می‌تواند از طریق روشهای مختلفی به این شاخص دست یابد. برای نمونه، ممکن است آزمودنی طبقه‌ای را انتخاب کند که به نظر وی به بهترین صورت سطح حمایت وی را از سؤال یا عبارت نشان می‌دهد. افزون بر این، امکان دارد داوران آزمودنیها را بر پایه ملاکهایی که به خوبی تعریف شده‌اند در طبقه‌های مختلف قرار دهند، و سرانجام ممکن است آزمودنی یک محرک فیزیکی را بر پایه شباهتی که به مجموعه محرکهای مرجع دارد، طبقه‌بندی کند. وقتی پاسخها فقط در دو طبقه قرار داشته باشند، مدل چند ارزشی راش به مدلی برای داده‌های دوارزشی تبدیل می‌شود. در این مدل خاص، دشواری سؤال و آستانه (منفرد) یکسان خواهد بود. انواع مدلهای چندارزشی به قرار زیرند:

۱) مدل مقیاس درجه‌بندی[4]: این مدل زمانی به کار می‌رود که تعداد آستانه سؤالها یکسان و تفاوت بین جایگاه هر آستانه معین با میانگین جایگاه آستانه‌ها برابر یا بین همه سؤالها یکسان باشد. فرمول این مدل به قرار زیر است:

$$\text{Log}\,(Pn_{ij} / Pn_{i\,(j-1)}) = B_n - D_i - F_j$$

۲) مدل امتیاز جزئی[5]: از این مدل اختصاصاً در زمینه‌های آموزشی و تربیتی استفاده می‌شود (مسترز، ۱۹۸۲). هر چند ساختار ریاضی این مدل با مدل مقیاس درجه‌بندی یکسان است، اما امکان محاسبه آستانه‌های مختلف را برای سؤالهای مختلف فراهم می‌آورد. فرمول این مدل عبارت است از:

$$\text{Log}\,(Pn_{ij}/ Pn_{i\,(j-1)}) = B_n - D_i - F_{ij} = B_n - D_{ij}$$

۳) مدل ساختار پاسخ گروه‌بندی شده[6]: این مدل با فرمول زیر وقتی به کار می‌رود که سؤالها بر اساس سهمی که در ساختار پاسخ دارند، یا به زیرمقیاسهای یک یا چند سؤال که در یک ساختار پاسخ سهیم هستند گروه‌بندی شوند.

$$\text{Log}\,(Pn_{ij}/ Pn_{i(j-1)}) = B_n - D_ig - Fg_j$$

به گونه کلی، مدلهای اندازه‌گیری راش به پژوهشگران امکان می‌دهد تا مشکلات زیربنایی اندازه‌گیریهای مدل کلاسیک و مقیاسهای خودسنجی، خودارزیابی و خود درجه‌بندی را حل کنند. این مدلها نمونه کاملی از اندازه‌گیری جمع‌پذیر زوجی است که دو شرط لازم برای تبدیل خصیصه به کمیت، یعنی جمع‌پذیر بودن و ترتیب را برآورده می‌سازد. مدل راش جمع‌پذیر است زیرا تفاوت بین سطح مشاهده شده و سطح مکنون، مستلزم اندازه‌گیری جمع‌پذیر دو متغیر مکنون متفاوت یعنی متغیرهای آزمودنی و سؤال است. افزون‌براین، مدل راش دارای ترتیب است زیرا بر پایه آن می‌توان متغیرهای آزمودنی و سؤال را در سطح مکنون و از طریق بالاتر یا پایین‌تر بودن نسبت به هم با یکدیگر مقایسه کرد (اکتون،

1. threshold location
2. conditional maximum likelihood
3. contiguous
4. rating scale
5. partial credit
6. grouped response-structure

Archive of SID
www.SID.ir

۲۰۰۳). برخی از مزایای کاربرد مدلهای اندازه‌گیری عبارتند:

۱) از پاسخهایی که در قالب مقیاس طبقه‌ای مرتب یا ترتیبی ارائه شوند، می‌توان یک اندازه فاصله‌ای حقیقی تولید کرد (رایت و لایرنس، ۱۹۸۹؛ مربیتز، موریس و گریپ، ۱۹۸۹).

۲) مشخص می‌شود هر سؤال تا چه حد می‌تواند سازه مورد نظر را اندازه‌گیری کند. به بیان‌دیگر، این مدل نشان می‌دهد که آیا سؤالهای مقیاس، یک سازه زیربنایی یا یک بعد واحد را تشکیل می‌دهند. این فرایند در واقع تک بعدی بودن مقیاس را آزمون می‌کند (رایت و استون، ۱۹۹۶).

۳) می‌توان نشان داد که هر سؤال چه جایگاهی در پیوستار اندازه‌گیری دارد. تعیین ترتیب سؤالها در پیوستار اندازه‌گیری از اهمیت زیادی در ارزیابی روایی مقیاس برخوردار است. زیرا توزیع سؤالها در طول پیوستار باید معنادار باشد تا نشان دهد سازه مورد نظر به خوبی اندازه‌گیری شده است. افزون بر این، شواهد مربوط به همسانی نسبی این توزیع در طول زمان یا در بین نمونه‌های مختلف، نشان می‌دهد که سازه مورد اندازه‌گیری پایایی دارد (اسمیت، ۲۰۰۱).

۴) می‌توان تعیین کرد که مقیاس تا چه اندازه توانسته است آزمودنیها را اندازه‌گیری کند. مدل راش افزون‌بر اینکه نشان می‌دهد آیا مقیاس برای اندازه‌گیری آزمودنیها به گونه مناسب تهیه شده، مشخص می‌کند که آیا هر آزمودنی نیز به گونه معتبری اندازه‌گیری شده است (آیا نمره افراد مطابق با الگوی مورد انتظار است). به بیان دیگر، روشهای راش نه‌تنها برای بررسی ویژگیهای آزمون مفیدند بلکه می‌توانند راهنمای مناسبی برای توسعه مقیاس نیز باشند.

منابع

Acton, G. S. (2003). What is good about Rasch measurement? *Rasch Measurement Transactions, 16*, 902-903.

Andrich, D. (2004). Controversy and the Rasch model: A characteristic of incompatible paradigms? *Medical Care, 42*, 1-16.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*, 357-74.

Linacre, J. M. (2006). W I N S T E P S: Rasch-model computer programs, Assessment System Cooperation.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

Massof, R., Fisher, W. P. (2002). The measurement of vision disability. *Rasch Measurement Transactions, 16*, 874-6.

Merbitz, C., Morris, J., & Grip, J. C. (1989). Ordinal scales and foundations of misinference. *Archives of Physical Medicine and Rehabilitation, 70,* 308-313.

Pesudovs, K., Garamendi, E., Keeves, J. P., & Elliott, D. B. (2003). The activities of daily vision scale for cataract surgery outcomes: Re-evaluating validity with Rasch analysis. *Investigative Ophthalmology and Visual Science, 44*, 2892-2899.

Smith, E. V., Jr. (2001). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. *Journal of Applied Measurement, 2 (3)*, 281-311.

Tenvergert, E., Gillespie, M., & Kingma, J. (1993). Testing the assumptions and interpreting the results of the Rasch model using log-linear procedures in SPSS. *Behavior Research Methods*, *Instruments & Computers,25 (3)*, 350-359.

Wright, B. D. (1996). Comparing Rasch measurement with factor analysis. *Structural Equation Modeling 3*, 3-24.

Wright, B. D. (1985). Additivity in psychological measurement. In E. Roskam (Ed.), *Measurement and personality assessment*. North Holland: Elsevier Science.

Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago, IL: MESA Press.

Archive of SID

www.SID.ir