magiran.com

نامهٔ آموزش عالی، دورهٔ جدید
سال اول، شمارهٔ دوم، تابستان ۱۳۸۷، ص ۷۷-۶۳

برآورد ویژگی‌های سؤال با استفاده از مدل همگنی مضاعف در رویکرد ناپارامتریک نظریهٔ سؤال پاسخ

علی مقدم‌زاده*
رضا پیروی **
سیده‌طیبه مطیعی لنگرودی***

چکیده

هدف از پژوهش حاضر، برآورد ویژگی‌های سؤال با استفاده از مدل همگنی مضاعف در رویکرد ناپارامتریک نظریهٔ سؤال پاسخ است. بدین منظور، پاسخنامه‌های داوطلبان رشتهٔ ریاضی فیزیک آزمون ورودی دانشگاه‌های کشور در سال ۱۳۸٤، در بخش آزمون اختصاصی درس ریاضی مورد استفاده قرار گرفت. از بین کلّیهٔ داوطلبان شرکت‌کننده در این رشته، با توجه به نرم‌افزار مورد استفاده در سازمان سنجش، به روش نمونه‌گیری سیستماتیک یک گروه نمونه ۳۰۰۰ نفری و سپس از میان آنها، دو نمونه ۸۰۰ و ۸۰ نفری به‌طور کاملاً تصادفی انتخاب شدند و دو سؤال این پژوهش، بر روی این دو گروه مورد بررسی قرار گرفت. به‌منظور مقایسه شاخص‌های برآوردشده سؤال‌ها در دو گروه، از آزمون t وابسته و ضریب همبستگی استفاده شد. نتایج تحقیق نشان داد که بین شاخص‌های محاسبه شده سؤال با استفاده از مدل همگنی مضاعف در رویکرد ناپارامتریک نظریه سؤال پاسخ، در دو نمونه تفاوتی وجود ندارد و این مدل در نمونه‌های کم نیز قابل استفاده است.

واژگان کلیدی: مدل همگنی مضاعف، رویکرد ناپارامتریک، نظریهٔ سؤال پاسخ، ویژگی‌های سؤال

---

*دانشجوی دکترای سنجش و اندازه‌گیری دانشگاه علامهٔ طباطبایی
**کارشناس‌ارشد دفتر آزمون‌سازی و روان‌سنجی سازمان سنجش آموزش کشور
***کارشناس‌ارشد دفتر آزمون‌سازی و روان‌سنجی سازمان سنجش آموزش کشور

magiran.com

مقدمه

تلاش‌های نخستین تکوین نظریهٔ کلاسیک اندازه‌گیری[1] در دههٔ ۱۸۹۰ آغاز شد. این نظریه یکی از روش‌های دیرینه در ساخت و توسعهٔ آزمون‌ها در حوزهٔ علوم انسانی است که از اوایل دههٔ ۱۹۰۰ برای ساخت و نمره‌گذاری ابزارهای اندازه‌گیری استفاده شده است. اوج تکامل این نظریه را بعد از اسپیرمن[2]، می‌توان در دو کتاب: (۱) «مبانی نظری آزمون‌های روانی[3]» گالیکسن[4] (۱۹۵۰) و «تئوری‌های آماری نمرات آزمون‌های روانی[5]» لرد[6] و ناویک[7] (۱۹٦۸) مشاهده کرد (همبلتون[8] و واندر لیندن[9]، ۱۹۸۲).

اگرچه نظریهٔ کلاسیک اندازه‌گیری مدت زمان طولانی به جامعهٔ روان‌سنجی خدمت کرده است، اما برخی مطالعات، نشان از وجود محدودیت‌هایی در این نظریه (از جمله گالیکسن، ۱۹۵۰؛ لرد و ناویک، ۱۹٦۸؛ همبلتون و سوامیناتان[10] و راجرز[11]، ۱۹۹۱) و آزمون‌های ساخته شده بر اساس آن دارند (به لرد، ۱۹۸۰؛ و همبلتون، ۱۹۸۹ مراجعه شود).

زمینهٔ ارائهٔ نظریه‌های جدید اندازه‌گیری از اوایل نیمه دوم قرن بیستم به وسیلهٔ افرادی چون لرد در فاصله (۱۹۵۳ تا ۱۹۵۲)، راش[12] (۱۹٦۸ تا ۱۹۵۸)، رایت[13] (۱۹۷۹)، همبلتون و جونز[14] (۱۹۸۳ و ۱۹۷۹) و غیره فراهم شد. نظریه‌های جدید اندازه‌گیری چه از لحاظ روش‌های آماری و به‌کارگیری توابع و مدل‌های ریاضی و چه از جهت مفروضه‌های نظری و نتایج کاربردی، تفاوت‌های چشمگیری با نظریهٔ کلاسیک اندازه‌گیری دارند. نظریه‌های جدید در دههٔ ۱۹۵۰ به‌عنوان یک جایگزین برای نظریهٔ کلاسیک اندازه‌گیری معرفی شدند.

نظریهٔ سؤال پاسخ[15]، یک نظریهٔ جامع آماری دربارهٔ عملکرد سؤال‌های آزمون، آزمودنی‌ها و چگونگی سنجش توانایی‌هایی است که به وسیلهٔ سؤال‌ها در آزمون

1. Classical Test Theory (CTT)
2. Spearman
3. Theory of Mental Tests (IRT)
4. Gulliksen
5. Statistical theories of mental test scores
6. Lord
7. Novick
8. Vander Linden
9. Hambleton
10. Swaminathan
11. Rojers
12. Rasch
13. Wright
14. Jones
15. Item response theory

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

magiran.com

اندازه‌گیری می‌شود. مقیاس اندازه‌گیری سؤال پاسخ‌ها در این نظریه می‌تواند گسسته یا پیوسته باشد و به صورت دو ارزشی و یا چند ارزشی نمره‌گذاری شود. طبقات نمرهٔ سؤال نیز می‌تواند منظم و یا نامنظم باشد؛ می‌تواند یک توانایی و یا چند توانایی در آزمون مستتر باشد؛ و همچنین در ارتباط بین سؤال پاسخ‌ها و توانایی یا توانایی‌هایی که اندازه‌گیری می‌شود، چند روش (یا مدل) وجود دارد (همبلتون و جونز، ۱۹۹۳). نظریهٔ سؤال پاسخ را می‌توان به دو رویکرد پارامتریک[1] و ناپارامتریک[2] تقسیم نمود. رویکرد ناپارامتریک سؤال پاسخ، این امکان را فراهم می‌سازد تا پارامترهای هر سؤال مثل ضریب دشواری و ضریب تمیز سؤال را با رتبه‌بندی پاسخ‌دهندگان بر اساس نمرهٔ آنان (تعداد پاسخ‌های درست به‌علاوه خطای تصادفی) که بر اساس توانایی (θ) مرتب شده‌اند، برآورد کنیم. مدل‌های رویکرد ناپارامتریک سؤال پاسخ عبارت‌اند از: (الف) مدل همگنی یکنواخت[3] و (ب) مدل همگنی مضاعف[4]. مدل همگنی مضاعف برمبنای مفروضه‌های تک‌بعدی بودن[5]، استقلال موضعی[6] و یکنواختی[7] سؤال‌ها قرار گرفته است (سیجت‌سما[8] و مولن‌آیر[9]، ۲۰۰۲/ به‌نقل از مقدم‌زاده، ۱۳۸۶):

الف) تک‌بعدی بودن

اولین مفروضه، تک‌بعدی بودن است؛ یعنی همهٔ سؤال‌های تست، یک صفت مکنون مشابه را اندازه می‌گیرند(سیجت‌سما و مولن‌آیر، ۲۰۰۲/ ترجمهٔ ذوالفقارنسب، ۱۳۸۵)؛ به تعبیر دیگر، تک‌بعدی بودن آزمون بدین معنی است که احتمال عملکرد موفقیت‌آمیز آزمودنی در مجموعه‌ای از سؤال‌ها می‌تواند به صورت یک مدل ریاضی که فقط یک پارامتر توانایی دارد، تعیین شود (درانز[10] و کینگستن[11]، ۱۹۵۰ به نقل از مطیعی‌لنگرودی، ۱۳۸۱). همبلتون (۱۹۸۹) می‌گوید که البته مفروضهٔ تک‌بعدی به صورت کامل صادق نخواهد بود؛ زیرا همیشه یک‌سری عوامل شناختی، شخصیتی، اجرایی و غیره وجود دارند که حداقل تا اندازه‌ای عملکرد آزمودنی در آزمون را تحت‌تأثیر قرار می‌دهند؛

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

1. Parametric approach
2. Nonparametric approach
3. Monotone homogeneity model
4. Double homogeneity model
5. Unidimensional
6. Local independence
7. Monotonic
8. Sijistma
9. Molenar
10. Dorans
11. Kingston

magiran.com

ب) استقلال موضعی

مفروضهٔ استقلال موضعی بدین معنی است که پاسخ فرد به سؤال i تحت‌تأثیر پاسخ‌های او به سؤال‌های دیگر تست نیست. این یک مفروضهٔ قوی است؛ برای مثال، استقلال موضعی ممکن است به وسیلهٔ یادگیری که بر اثر تمرین به وجود آمده نقض شود. این امر در طول اجرای آزمون هنگامی که نمره‌گذاری روی صفت مکنون صورت می‌گیرد، رخ می‌دهد (سیجت‌سما و مولن‌آیر، ۲۰۰۲/ ترجمهٔ ذوالفقارنسب، ۱۳۸۵). در صورتی‌که احتمال پاسخ آزمودنی A به سؤال معیّنی ۰/۲ باشد و احتمال پاسخگویی به همین سؤال برای آزمودنی B مساوی ۰/۹ باشد و چنانچه پاسخ‌های آزمودنی‌ها به صورت موضعی مستقل از یکدیگر باشند، در این صورت احتمال اینکه هر دوی آنها به آن سؤال پاسخ درست بدهند، مساوی است با ۰/۱۸= (۰/۹)(۰/۲) (آلن و ین، ۱۹۷۹/ ترجمهٔ دلاور، ۱۳۷٤)؛

ج) یکنواختی توابع سؤال پاسخ[1]

مفروضهٔ بعدی این‌است که احتمال شرطی $\mathbf{P_i(\theta)}$ به‌طور یکنواخت روی θ بدون کاهش است. این مفروضه در معادلهٔ(۱) نشان داده شده است. به‌روشنی می‌توان دید که احتمال $X_i = 0$ نیز به وسیلهٔ تابع سؤال پاسخ قابل توصیف است:

$$p(X_i = 0|\theta) = 1 - p(X_i = 1|\theta)$$ معادله(۱)

این مدل به صورت همگون ایجاد شده است و توابع سؤال پاسخ آنها به‌صورت یکنواخت با صفت مکنون ارتباط دارد. اهمیت کاربردی مدل همگنی یکنواخت این است که می‌توان پاسخ‌دهندگان را بر اساس نمرهٔ کل آنها در مقیاس θ رتبه‌بندی کرد؛ بنابراین، مدلی است که در آن افراد براساس یک مقیاس ترتیبی مورد اندازه‌گیری قرار می‌گیرند. اگر c را مقداری ثابت درنظر بگیریم و این مقدار ثابت را با s و t جمع کنیم، به شرط برقراری رابطه $o \le s \le t \le k$ خواهیم داشت:

$$\mathbf{P(\theta > c | X_+ = s) \le P(\theta > c | X_+ = t)}$$ معادله (۲)

مطابق این معادله، افراد را می‌توان به وسیلهٔ x_+ به‌صورت احتمالی روی θ رتبه‌بندی کرد. در سؤال‌های دو ارزشی که به‌صورت ۰ و ۱ نمره‌گذاری می‌شود، نمرهٔ شرطی مورد انتظار هر سؤال برابر است با مقادیر تابع سؤال پاسخ آن که می‌توان به‌صورت معادله زیر نوشت:

$$\mathbf{E(X_i|\theta) = 0 \times P(X_i = 0|\theta) + 1 \times P(X_i = 1|\theta) = P_i(\theta)}$$ معادله (۳)

1. Item response functions monotonic

magiran.com

به‌عبارت دیگر، تغییرناپذیری ترتیب قرار گرفتن سؤال‌های بر این دلالت دارد که احتمال پاسخگویی به سؤال i کوچک‌تر یا برابر احتمال پاسخگویی به سؤال j است ($P_i \leq P_j$) (سیجت‌سما و مولن‌آیر، ۲۰۰۲/ ترجَمهٔ ذوالفقارنسب، ۱۳۸۵).

برای تحلیل داده‌ها در مدل همگنی مضاعف ناپارامتریک سؤال پاسخ، دو بررسی صورت می‌گیرد:

الف) بررسی پذیرش یا محبوبیت سؤال[1] (P_i)

در نظریهٔ سؤال پاسخ، عموماً و در مدل ناپارامتریک آن خصوصاً، ویژگی‌های سؤال‌ها مستقل از ویژگی‌های آزمودنی‌های نمونه معیّن و در واقع، بر اساس جامعهٔ آزمودنی‌ها برآورد می‌شود (سیجت‌سما و مولن‌آیر، ۲۰۰۲/ به‌نقل از مقدم‌زاده، ۱۳۸۶). این پارامتر نشان می‌دهد که منحنی ویژگی سؤال در کجای مقیاس توانایی قرار دارد و معادل دشواری سؤال در نظریهٔ سؤال پاسخ در رابطه با سطح صفت زیربنایی θ است (مولن‌آیر، ۱۹۹۷)؛

ب) بررسی مقیاس‌پذیری[2] (H_i)

این پارامتر که با نماد H_i نشان داده می‌شود، قدرت سؤال را در تمایزگذاری یا تشخیص آزمودنی‌ها در سطوح مختلف توانایی نشان می‌دهد و بیان می‌کند که هر سؤال آزمون تا چه حد قادر به تفکیک و تمیز بین آزمودنی‌ها در سطوح مختلف توانایی است. در همهٔ مدل‌ها، ضریب H_i نقش اصلی را بازی می‌کند و برآوردی سریع و کاربردی در اندازه‌گیری‌های کیفی است. به‌نظر موکن[3] (۱۹۹۷) در این رابطه می‌توان اشاره نمود که معتقد است، مقیاسی مفید است که ضریب مقیاس‌پذیری آن برای همهٔ سؤال‌ها بزرگ‌تر از ۰/۳۰ باشد (H_i>۰/۳۰)؛ در واقع، این شاخص نشان‌دهندهٔ همسویی و هماهنگی سؤال با کل آزمون و یا بیانگر آن است که آیا سؤال همان ویژگی موردنظر آزمون را اندازه‌گیری می‌کند یا نه. این شاخص به نوعی با ضریب قدرت تشخیص سؤال در مدل پارامتریک نظریهٔ سؤال پاسخ همخوانی دارد. یک شیوه برای تعریف ضریب H_i، اصطلاح خطای احتمالاتی گاتمن[4] است. فرض می‌کنیم که احتمال خطای گاتمن برای جفت سؤال‌های (i , k) بدین صورت تعریف شده است:

$$e_{ik} = \pi_{ik}(1,0) \qquad e_{ik}^{(0)} = \pi_i(1-\pi_k) \; if \; i < k$$
$$e_{ik}^{(0)} = (1-\pi_i)\pi k_{if\,i} > k$$

1. Acceptance
2. Scalibity
3. Mokken
4. Guttman probability error

magiran.com

که احتمال‌هایی از نمرهٔ مشاهده شده وخطای مورد انتظار تحت استقلال کناری را نشان می‌دهد.

$$H_{ik} = ۱ - \frac{e_{ik}}{e_{ik}^{(\circ)}}$$

در مقیاس گاتمن، احتمال اینکه آزمودنی در یک سؤال بسیار آشنا (سؤال آسان)، نمره صفر و در یک سؤال ناآشنا (سؤال بسیار سخت) نمره ۱ بگیرد، خیلی کم است. برعکس‌شدن مطلب فوق را خطای گاتمن می‌نامند؛ برای مثال، اگر پاسخ‌دهنده‌ای در دو سؤال i و k، به سؤال i پاسخ صحیح و به سؤال k که آسان‌تر از سؤال i است، پاسخ غلط بدهد، مرتکب خطای گاتمن شده است. اگر $H_{ik} = ۱$ باشد، مقیاس‌پذیری کامل است و صفر زمانی به‌دست می‌آید که استقلال آماری وجود دارد و i و k کاملاً مستقل از یکدیگر هستند. از تقسیم دو خطا بر یکدیگر، مقدار مقیاس‌پذیری که به‌دست می‌آید همواره از یک کم‌تر خواهد بود. ارزش‌های H_i می‌تواند برای ارزیابی برازش سؤال i با توجه به سؤال‌های دیگر به‌صورت یک ماتریس همبستگی از مجموعه کل سؤال‌ها مورد استفاده قرارگیرد. در نهایت، یک ضریب می‌تواند از مجموعه کل سؤال‌ها به‌عنوان نسبت خطا و ترکیبی از H_i یا H_k تعریف شود. با فرض $e = \sum k < i e_{ki}$ و $e^{(\circ)} = \sum k < i e_{ik}^{(\circ)}$، ضریب برای مقیاس به‌صورت زیر بیان می‌شود:

$$H = ۱ - \frac{e}{e^{\circ}} = \frac{\sum_i^n < k \pi_{ik}^{(\circ)} H_i}{\sum_i^n < k \pi_{ik}^{(\circ)}} = \frac{\sum_i^n < k \pi_{ik}^{(\circ)} H_{ik}}{\sum_i^n < k \pi_{ik}^{(\circ)}}$$

بنابراین، برای تک‌تک سؤال‌ها، یک مقدار ثابت c فرض می‌شود؛

$$(H_i;\ i=1,\ ...,\ n) = c$$

پس $H \geq c, 0 \leq c \leq 1$ و ضریب مقیاس‌پذیری تک‌تک سؤال‌ها تا اندازه‌ای شباهت به ضریب مقیاس‌پذیری کل آزمون دارد. ضریب مقیاس‌پذیری می‌تواند به عنوان نسبت واریانس با توجه به نمره به حالت نرمال درآید (موکن، ۱۹۹۷). مقیاس همگنی یکنواخت مانند مدل همگنی مضاعف به‌صورت مجموعه سؤال‌هایی تعریف می‌شود که هر کدام همبستگی مثبت با ویژگی هر سؤال (صفر و یک) دارند که بزرگ‌تر یا برابر مقدار ثابت C ($c = ۰/۳$) است. ضریب مقیاس‌پذیری، مجموعه سؤال‌ها را نیز به‌صورت یک کل مورد آزمون قرار می‌دهد که باید بزرگ‌تر از C تک‌تک سؤال‌ها باشد. موکن در ادامه طبقه‌بندی زیر را برای پذیرش مقیاس‌ها پیشنهاد می‌کند:

$< H$ ۰/ ۵۰ مقیاس قوی

$۰/۵۰ < H < ۰/۴۰$ مقیاس متوسط

$۰/۴۰ < H < ۰/۳۰$ مقیاس ضعیف

magiran.com

روش پژوهش جامعهٔ آماری

به‌منظور پاسخ به پرسش‌های پژوهشی موردنظر، جامعه‌ای از کلیهٔ داوطلبان ورود به دانشگاه‌های کشور در رشتهٔ ریاضی ـ فیزیک در کنکور سراسری سال ۱۳۸٤ انتخاب شد. جدول (۱)، تعداد داوطلبان شرکت‌کننده در کنکور سراسری ۱۳۸٤ را در رشته ریاضی فیزیک نشان می‌دهد.

جدول(۱) تعداد داوطلبان شرکت‌کننده در کنکور سراسری ۱۳۸٤ در رشتهٔ ریاضی فیزیک

رشته	زن	درصد	مرد	درصد	کل
ریاضی فیزیک	۱۳۸۸٦۸	٤٥/۹۱	۱٦۳٦٤۳	٥٤/۰۹	۳۰۲٥۱۱

نمونه‌گیری بر اساس دسترسی به فهرست داوطلبان و با توجه به برنامهٔ کامپیوتری طراحی‌شده در سازمان سنجش آموزش کشور، با استفاده از روش نمونه‌گیری منظمانجام شد. از بین کلیهٔ داوطلبان شرکت‌کننده در گروه آزمایشی ریاضی ـ فیزیک، یک گروه ۳۰۰۰ نفری به‌طور تصادفی انتخاب شد و پس از حذف آزمودنی‌هایی که به‌درستی به سؤال‌ها پاسخ نگفته بودند (افرادی با نمرهٔ پایین‌تر از $\frac{۱}{۴}$ نمرهٔ کل یا حد شانس)، از بین افراد باقیمانده چند نمونه‌گیری صورت گرفت.

مطالعات اندکی وجود دارد که تأثیر حجم نمونه را بر برآوردهایی که برای پایایی پارامترها صورت می‌گیرد، به‌طور منظم بررسی کرده باشند. مطابق پژوهش‌های موجود در زمینهٔ آزمون‌سازی، کاربرد موفق مدل‌های پارامتریک سؤال پاسخ و مدل‌های ناپارامتریک سؤال پاسخ، مستلزم استفاده از سؤال‌ها و آزمودنی‌هایی با حجم بزرگ است تا بتوان به‌طور همزمان صفت مکنون و پارامترهای سؤال را برآورد کرد (لرد، ۱۹٦۸؛ همبلتون و سوآمیناتان، ۱۹۸٥)؛ بنابراین، بر اساس تحقیقات انجام شده در هر دو نظریه و با توجه به نرم‌افزار کامپیوتری به‌کار رفته، از بین داوطلبان گروه آزمایشی ریاضی فیزیک، در نهایت دو نمونهٔ بزرگ و کوچک (۸۰۰ نفری و ۸۰ نفری) انتخاب شد.

ابزار، روش اجراء و تحلیل داده‌ها

در آزمون ورودی دانشگاه‌ها در ایران از آزمون‌های پیشرفت تحصیلی استفاده می‌شود که در سازمان سنجش به‌وسیلهٔ استادان با تجربهٔ کشور طراحی و تهیه می‌شود. ابزار مورد استفاده برای جمع‌آوری داده‌های پژوهش، پاسخنامه‌های داوطلبان در آزمون اختصاصی درس ریاضی در رشتهٔ ریاضی فیزیک با ٥٥ سؤال بود. برای تجزیه و تحلیل داده‌ها از نرم‌افزارهای کامپیوتری SPSS و MSP استفاده شد. ابتدا تحلیل‌ها و مشخصه‌های کلاسیک سؤال‌ها و آزمون از طریق SPSS محاسبه و سپس داده‌ها با نرم‌افزار MSP تحلیل شد. به‌منظور تحلیل داده‌ها در مدل ناپارامتریک سؤال پاسخ، از روش‌های آماری مانند آزمون t وابسته، ضریب همبستگی پیرسون و آزمون‌های معنی‌داری آنها استفاده شد و سپس مقایسه‌های موردنظر انجام گرفت.

magiran.com

نتایج پژوهش

در پژوهش حاضر، سؤال اول تحقیق، یعنی تأثیر خصوصیات آزمودنی‌ها بر ویژگی‌های سؤال، بر اساس مدل ناپارامتریک سؤال پاسخ این سؤال‌ها تحلیل شد. به‌دنبال آن، آزمون برازندگی برای هر سؤال و کل آزمون با استفاده از ضریب مقیاس‌پذیری محاسبه شد. سؤال‌هایی که ضریب مقیاس‌پذیری کمتر از ۰/۳ یا منفی داشتند، حذف شدند.

در جدول (۱) (ضمیمهٔ یک)، نتایج تحلیل سؤال‌ها با نرم‌افزار ناپارامتریک سؤال پاسخ یعنی MSP، روی ۵۵ سؤال ریاضی آزمون سراسری سال ۱۳۸٤ نشان داده شده است. همان‌طور که در جدول(۱) ملاحظه می‌شود، تنها یک سؤال با توجه به ضریب منفی، کنار گذاشته شده است. لازم به ذکر است که در هر دو نمونه، همین ضرایب مقیاس‌پذیری و محبوبیت سؤال به‌دست آمده است.

جدول(۲) مقایسهٔ ضرایب همبستگی شاخص یا محبوبیت سؤال (دشواری سؤال) در دو گروه نمونه به روش همگنی مضاعف در مدل ناپارامتریک

روش مقایسه	تعداد سؤال‌ها	ضریب همبستگی r_{pbis}	سطح معنی‌داری
NIRT	۵۵	۰/۹۸٤	۰/۰۰۰۱

همچنین همان‌طور که جدول (۲) نشان می‌دهد، نتایج مقایسهٔ دو نمونهٔ ۸۰۰ نفری و ۸۰ نفری حاکی از آن است که در مدل ناپارامتریک سؤال پاسخ (NIRT)، همبستگی بسیار بالایی بین مقادیر برآورد شدهٔ این دو گروه نمونه وجود دارد. این نکته بیانگر آن است که برآورد این شاخص‌ها در مدل ناپارامتریک یک برآورد عالی است و مقدار همبستگی در حدود ۰/۹۹ می‌باشد که در سطح ۰/۰۰۰۱. هم معنی‌دار است. لازم به ذکر است که در نرم‌افزار MSP که برای تحلیل داده‌ها با مدل‌های ناپارامتریک به‌کار می‌رود، محبوبیت سؤال (P_i) معادل ضریب دشواری و مقیاس‌پذیری (H_i) معادل ضریب تشخیص است. از آزمون t وابسته نیز برای نشان دادن تفاوت دو نمونه در برآورد شاخص محبوبیت سؤال استفاده شد. جدول(۳)، نتایج مقایسه این دو نمونه را نشان می‌دهد:

جدول(۳) مقایسهٔ NIRT از نظر شاخص محبوبیت سؤال‌ها با استفاده از آزمون t وابسته در دو نمونهٔ ۸۰۰ و ۸۰ نفری

روش مورد مقایسه	تعداد سؤال‌ها	میانگین سطح دشواری (محبوبیت سؤال)		خطای استاندارد میانگین سطح دشواری		آزمون t	درجهٔ آزادی (d.f)	سطح معنی‌داری
		گروه A (۸۰۰نفر)	گروه B (۸۰نفر)	گروه A (۸۰۰نفر)	گروه B (۸۰نفر)			
NIRT	۵۵	۰/۷۲۳	۰/۷۲۳	۰/۲۵	۰/۲۵	۰/۷٤	۵۳	۰/٤۲۵

magiran.com

همان طور که ملاحظه می‌شود، میانگین سطح دشواری در دو نمونهٔ بزرگ و کوچک با یکدیگر متفاوت نبوده و تقریباً یکسان است. خطای استاندارد میانگین سطح دشواری در مدل ناپارامتریک سؤال پاسخ نیز یکسان است. آزمون t نشان می‌دهد که در مدل ناپارامتریک سؤال پاسخ، تفاوت بین دو گروه نمونه از نظر برآورد پارامتر دشواری معنی‌دار نیست و مقادیر برآورد شدهٔ پارامتر دشواری در دو گروه نمونهٔ متفاوت، یکسان یا تقریباً یکسان است. این مسئله بیانگر عدم تأثیرپذیری برآورد محبوبیت سؤال در مدل ناپارامتریک نظریهٔ سؤال پاسخ از ویژگی‌های آزمودنی‌ها می‌باشد.

جدول(٤) مقایسهٔ میانگین ضرایب همبستگی شاخص مقیاس‌پذیری (قدرت تشخیص) سؤال‌ها در دو گروه نمونه با روش همگنی مضاعف در مدل ناپارامتریک

روش مقایسه	تعداد سؤال‌ها	ضریب همبستگی (r_{pbis})	سطح معنی‌داری
NIRT	۵۵	۰/۹۹۵	۰/۰۰۰۱

در مدل ناپارامتریک سؤال پاسخ، مقیاس‌پذیری معادل شاخص یا پارامتر قدرت تشخیص یا شیب سؤال (a_i) است. جدول (٤)، ضریب همبستگی برآورد شاخص قدرت تشخیص سؤال‌های آزمون ریاضی را در دو گروه نمونه در مدل ناپارامتریک سؤال پاسخ نشان می‌دهد. همان‌طور که ملاحظه می‌شود، ضریب همبستگی برابر ۰/۹۹۵ شده است که در سطح ۰/۰۰۰۱ معنی‌دار می‌باشد که نشان‌دهندهٔ همبستگی بسیار بالا در مدل ناپارمتریک سؤال پاسخ در برآورد شاخص یا پارامتر قدرت تشخیص است.

جدول(۵) مقایسهٔ شاخص مقیاس‌پذیری (پارامتر قدرت تشخیص) سؤال در دو گروه نمونه با روش همگنی مضاعف در مدل ناپارامتریک با استفاده از آزمون t وابسته

روش مورد مقایسه	تعداد سؤال‌ها	میانگین مقیاس‌پذیری		خطای استاندارد میانگین سطح دشواری		آزمون t	درجهٔ آزادی (d.f)	سطح معناداری
		گروه A (۸۰۰ نفر)	گروه B (۸۰ نفر)	گروه A (۸۰۰ نفر)	گروه B (۸۰ نفر)			
NIRT	۵۵	۰/۸۳۲	۰/۸۳۱	۰/۰٤۱	۰/۰٤۲	-۰/۰۱	٥٤	۰/۸۹٦

همان‌طورکه در جدول (۵) ملاحظه می‌شود، در مدل همگنی مضاعف ناپارامتریک نظریهٔ سؤال پاسخ، میانگین قدرت تشخیص و انحراف استاندارد (SD) آنها در دو گروه آزمودنی بسیار نزدیک به یکدیگر و خطای استاندارد در مدل همگنی مضاعف ناپارامتریک تقریباً یکسان است. بر اساس نتایج آزمون t وابسته در مدل همگنی مضاعف ناپارامتریک سؤال پاسخ، تفاوت بین دو گروه در برآورد این شاخص معنی‌دار

magiran.com

نیست؛ به عبارت دیگر، در مدل ناپارامتریک نظریهٔ سؤال پاسخ، ضریب مقیاس‌پذیری (پارامتر قدرت تشخیص) تحت‌تأثیر ویژگی‌های گروه آزمودنی‌ها و حجم نمونه قرار نمی‌گیرد و از یک گروه به گروه دیگر، مقادیر آن مشابه است و در نتیجه مقادیر برآوردشده در دو گروه تفاوت معنی‌داری با یکدیگر ندارند.

نتیجه‌گیری

در بررسی سؤال اول پژوهش، تفاوت شاخص سطح دشواری یا محبوبیت سؤال در دو گروه نمونه ۸۰۰ و ۸۰ نفری در مدل ناپارامتریک سؤال پاسخ مورد مطالعه قرار گرفت و نتایج نشان داد که در مدل ناپارمتریک نظریهٔ سؤال پاسخ، ضریب مقیاس‌پذیری (H_i)، نامتغیر و تقریباً ثابت است. لازم به ذکر است که مزیت عمدهٔ این مدل نسبت به مدل‌های پارامتریک این است که در نمونه‌های خیلی کمتر نیز شاخص‌ها تحت‌تأثیر ویژگی‌های آزمودنی‌ها یا سؤال‌های دیگر تغییر نمی‌کند. در این زمینه، موکن (۱۹۹۷) به نتایج مشابه‌ای دست یافت که مدل‌های ناپارامتریک در آزمون‌های استخدامی و نمونه‌هایی با حجم کمتر به‌کار می‌رود، ولی در مدل ناپارمتریک نظریهٔ سؤال پاسخ، ضریب مقیاس‌پذیری بدون تأثیرپذیری از میزان توانایی افراد برآورد شده و ویژگی‌های سؤال‌های آزمون، تأثیری بر مقدار برآوردشده ندارد. رمسی[1] (۱۹۹۷)، تحلیل‌هایی را با استفاده از مدل‌های همگنی یکنواخت و همگنی مضاعف انجام داد و بررسی پژوهشی وی در این رابطه نیز، مطالب فوق را تأیید کرد.

در بررسی سؤال دوم تحقیق، یعنی تفاوت دو گروه از نظر شاخص مقیاس‌پذیری، نتایج زیر به‌دست آمد: شاخص مقیاس‌پذیری نیز مانند شاخص دشواری سؤال، از نمونه‌های مختلف آزمودنی تأثیر نمی‌پذیرد. در این مدل، ضریب مقیاس‌پذیری، شاخصی است که میزان آگاهی را که یک سؤال دربارهٔ سطح توانایی مورد سنجش ارائه می‌دهد، نشان می‌دهد. ضریب مقیاس‌پذیری در مدل همگنی مضاعف ناپارامتریک سؤال پاسخ نامتغیر بوده و به گروه‌های مختلف آزمودنی‌ها قابل تعمیم است. عدم تغییر، یکی از مهم‌ترین ویژگی‌های این مدل است (اونا[2]، ۲۰۰۴).

1. Ramsey
2. Onna

magiran.com

منابع

آلن، مری‌جی و ین، وندی‌ام (۱۹۷۹). *مقدمه‌ای بر نظریه‌های اندازه‌گیری (روان‌سنجی)*. ترجمهٔ علی دلاور(۱۳۷٤)، تهران: انتشارات سمت.

سیجت سیما، کلاس و مولن‌آیر، ایو دبلیو(۲۰۰۲). *مقدمه‌ای بر تئوری ناپارامتریک سؤال پاسخ*. ترجمهٔ سلیمان (سالار) ذالفقارنسب(۱۳۸۵). نشر رایانه.

مطیعی‌لنگرودی، سیده طیبه (۱۳۸۱). *بررسی مزایا و محدودیت‌های دو روش اندازه‌گیری: کلاسیک و سؤال پاسخ (IRT) از لحاظ نظری و عملی*. پایان‌نامهٔ کارشناسی‌ارشد، دانشگاه تهران: دانشکده روان‌شناسی و علوم‌تربیتی.

مقدم‌زاده، علی (۱۳۸٦). مقایسهٔ مدل‌های پارامتریک با مدل‌های غیرپارامتریک نظریهٔ سؤال پاسخ. پایان‌نامهٔ کارشناسی‌ارشد، دانشگاه علامهٔ طباطبایی: دانشکدهٔ روان‌شناسی وعلوم‌تربیتی.

Gulliksen, H. (1995). *Theory of mental tests*. Newyork: John Wiley & Sons.

Hambleton, R.K. (1989). *Principles and selected applications of Item Response Theory*. In R. Linn (Ed), *Educational Measurement (3rd end)*. Newyork: Macmillan. 147-200.

Hambleton, R.K. & Jones, R. W. (1993). Comparison of classical test theory and Item Response Theory and their applications to test development. *Educational Measurement: Issues and Practice*. 12(3),38-47.

Hambleton, R. K., Swaminathan, H. & Rojers, H. J. (1991). *Fundamentals of Item Response Theory*. Newburypark: Sage.

Hambleton, R. K. & Vander Linden, W. J. (1982). Advance in Item Response Theory and applications: An introduction. *Applied Psychplogical Measurement*. 6(4), 373-387.

Lord, F. M. (1980). *Applications of Item Response Theory to Practice Testing Problems*. Hillsdale, N.J: Lawrence Erlbaum.

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Mokken, R.J. (1997). Nonparametric models for dichotomous responses. In: Hambleton, R.K. and Van der Linden, W.J. (Ed's). *Handbook of Modern Item Response Theory*. New York-Berlin: Springer -Verlag, pp.351-367.

magiran.com

Molenaar, I. W. (1997). Nonparametric models for dichotomous responses. In: Hambleton, R.K. and Van der Linden, W.J. (Ed's). *Handbook of modern Item Response Theory*. New York-Berlin: Springer-Verlag, pp. 369-379.

Onna, M. J. H. (2004).Ordered latent class models in nonparametric Item Response Theory. Lawrence Erlbaum.

Ramsey, J. (1997). *Nonparametric models for dichotomous responses*. In: Hambleton, R.K. and Van der Linden, W.J. (Ed's). Handbook of Modern Item Response Theory. New York-Berlin: Springer -Verlag, pp. 369-379.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

magiran.com

ضمیمهٔ جدول (۱) نتایج تحلیل داده‌ها با نرم‌افزار ناپارامتریک (MSP) سؤال پاسخ

شمارهٔ سؤال	مقیاس‌پذیری سؤال (H_i) در مرحلهٔ انتخاب سؤال					
	ویژگی مقبولیت سؤال (p_i)	گام ۱	گام ۲	گام ۳	گام ٤	گام ٥
۱	۰/۹۷	۰/۷٦	۰/۸۹	۰/۸٥	۰/٦۱	۰/۹۱
۲	۰/۸٥	۰/۹۲	۰/۸۱	۰/۷۸	۰/٦۱	۰/۹۲
۳	۰/٤٥	۰/۸۲	۰/۸۷	۰/٥٤	۰/۳۹	۰/٦٦
٤	۰/۷٤	۰/۷۳	۰/۸٥	۰/٥٦	۰/٤٥	۰/۸۱
٥	۰/۸۹	۰/٥۲	۰/۸۸	۰/٥۲	۰/٤۷	۰/۸٤
٦	۰/۸۱	۰/۹۱	۰/۸۷	۰/۳٦	۰/٤۸	۰/۸۲
۷	۰/۸۷	۰/۹۲	۰/۷۸	۰/۳٥	۰/۸٥	۰/۷٤
۸	۰/۸٥	۰/۸۱	۰/۸٦	۰/۸۸	۰/۸۸	۰/۸۸
۹	۰/٦٥	۰/۸٦	۰/۸۷	۰/۸۷	۰/۸۷	۰/٦۹
۱۰	۰/٥٤		۰/۷٤	۰/۷۸	۰/۷۸	۰/٦۱
۱۱	۰/۹۸		۰/۸۹	۰/۸۲	۰/۸۱	۰/٦۱
۱۲	۰/۸۸		۰/۸۱	۰/٥۲	۰/۸۲	۰/۳۹
۱۳	۰/۸۷		۰/۸۷	۰/۸۱	۰/۸٤	۰/٤٥
۱٤	۰/۷۸		۰/۸٥	۰/۸۲	۰/۸٤	۰/٤۷
۱٥	۰/۸٦		۰/۷۲	۰/۸٤	۰/۸۲	۰/٤۸
۱٦	۰/۸۱		۰/۷٦	۰/۸٤	۰/٥۲	۰/٤۹
۱۷	۰/۸۲			۰/۸٦	۰/۹۱	۰/۹٥
۱۸	۰/۸٤			۰/۸۱	۰/۹۲	۰/۹۳
۱۹	۰/۸٤			۰/۸٤	۰/۸۲	۰/۸۷
۲۰	۰/۸۲			۰/۸۲	۰/٥۲	۰/۸٦
۲۱	۰/٥۲			۰/٥۸	۰/۸۱	۰/۹۱

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

magiran.com

ادامهٔ ضمیمهٔ جدول (۱)نتایج تحلیل داده‌ها با نرم‌افزار ناپارامتریک (MSP) سؤال پاسخ

شمارهٔ سؤال	مقیاس‌پذیری سؤال (H_i) در مرحلهٔ انتخاب سؤال					
	ویژگی مقبولیت سؤال (p_i)	گام ۱	گام ۲	گام ۳	گام ٤	گام ٥
۲۲	۰/۹۱			۰/۵۱	۰/۸۲	۰/۹۲
۲۳	۰/۹۲			۰/۸۷	۰/۸٤	۰/٦٦
۲٤	۰/٦٦			۰/۸۵	۰/۸٤	۰/۳۷
۲۵	۰/۷۵			۰/٦۵	۰/۸۲	۰/۵۸
۲٦	۰/۵۸			۰/۷٦	۰/۸۲	۰/۸۵
۲۷	۰/۸۵			۰/۷٦	۰/۷٤	۰/۷۸
۲۸	۰/۸۵			۰/۷۳	۰/۸۸	۰/۵٤
۲۹	۰/۵٤			۰/۸۷	۰/۸۷	۰/۵٦
۳۰	۰/٦۵				۰/۷۸	۰/۵۲
۳۱	۰/۸۱				۰/۸٦	۰/۳٦
۳۲	۰/۸٤				۰/۷۲	۰/۳۵
۳۳	۰/۸۲				۰/۸۱	۰/۳۷
۳٤	۰/۷٤				۰/۸۲	۰/٤۹
۳۵	۰/۷۳				۰/۸٤	۰/۵۸
۳٦	۰/۵۷				۰/۸٤	۰/۵۷
۳۷	۰/۳۵				۰/۸۲	۰/۵۹
۳۸	۰/٦۵				۰/۵۲	۰/۵۸
۳۹	۰/۵۲				۰/۹۱	۰/۵۹
٤۰	۰/۵٤				۰/۵۷	۰/۳۹
٤۱	۰/۷۵				۰/۳۵	۰/۵۹
٤۲	۰/۷٦					۰/٦۱

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

magiran.com

ادامهٔ ضمیمهٔ جدول (۱) نتایج تحلیل داده‌ها با نرم‌افزار ناپارامتریک (MSP) سؤال پاسخ

شمارهٔ سؤال	مقیاس‌پذیری سؤال (H_i) در مرحلهٔ انتخاب سؤال					
	ویژگی مقبولیت سؤال (p_i)	گام ۱	گام ۲	گام ۳	گام ٤	گام ٥
٤٣	۰/٦٣					۰/٥٧
٤٤	۰/٦١					۰/٦١
٤٥	۰/٧٢					۰/٦١
٤٦	۰/٥٣					۰/٦٤
٤٧	۰/٥٧					۰/٦١
٤٨	۰/٥٨					۰/٥٣
٤٩	۰/٥١					۰/٥٧
٥٠	۰/٦١					۰/٥٨
٥١	۰/٦٧					۰/٥١
٥٢	۰/٧٣					۰/٥٧
٥٣	۰/٦٩					۰/٣٥
٥٤	۰/٦٧					۰/٦٥
کل	۰/٧٢					۰/٥٦

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی