

مقاله پژوهشی

کاربرد مدل درخت‌های راش در آشکارسازی کنش افتراقی سؤال (مورد

مطالعه: آزمون‌های استخدامی نیروی انتظامی ج.ا. ایران)^۱

حمیدرضا کرمی^۲، مسعود گرامی پور^۳، اصغر مینایی^۴

تاریخ دریافت: ۱۴۰۰/۰۲/۱۵

تاریخ پذیرش: ۱۴۰۰/۰۴/۲۰

از صفحه ۳۱ تا ۵۴

چکیده

این مطالعه قصد دارد تا در آزمون بزرگی چون آزمون استخدامی نیروی انتظامی، سؤالاتی که شامل کنش افتراقی سنی و گروه آزمایشی باشند را با استفاده از مدل درخت‌های راش پیدا کند. روش پژوهش حاضر از نظر هدف کاربردی و از نظر روش، توصیفی موسوم به تحقیقات روان‌سنجی است. همچنین با توجه به اینکه این پژوهش آزمون‌های استخدامی را مورد بررسی قرار می‌دهد از حیث روش، از نوع ارزشیابی است. جامعه آماری آزمون استخدامی ورودی سال ۱۳۹۸ به تعداد ۲۴۱۴ نفر به صورت تمام شمار مورد تحلیل قرار گرفته است. جهت تشخیص نرخ آشکارسازی کنش افتراقی سؤال با استفاده از مدل درخت‌های راش از بسته DIFtree در نرم‌افزار R استفاده شد. نتایج حاصل از مدل درخت‌های راش نشان داد که تعداد ۹ سؤال دارای کنش افتراقی سؤال هستند که در مهم‌ترین آن، تعداد ۷ سؤال در گروه ریاضی با ۱۸ سال سن (رسته دوم) و ۶ سؤال در گروه ریاضی با ۱۹ سال سن (رسته اول) دارای سوگیری نسبت به گروه علوم تجربی (رسته سوم) بوده و جهت گیری سوگیری به نفع رسته اول و دوم و به ضرر رسته سوم بوده است. مدل درخت‌های راش به خوبی توانسته است سوگیری سؤالات آزمون استخدامی را مشخص نماید.

واژگان کلیدی: مدل درخت‌های راش، کنش افتراقی سؤال، سوگیری، نیروی انتظامی.

۱- این مقاله برگرفته از رساله دکتری با عنوان «نرخ آشکارسازی کنش افتراقی سؤال (DIF) با استفاده از مدل درخت‌های راش: یک مطالعه شبیه‌سازی شده و داده‌های واقعی آزمون‌های سرنوشت‌ساز ناجا» به راهنمایی دکتر مسعود گرامی پور می‌باشد.

۲- دانشجوی دکتری رشته سنجش و اندازه‌گیری دانشگاه آزاد اسلامی، ساوه، ایران. h.karami7662@gmail.com

۳- دانشیار گروه آموزشی برنامه‌ریزی درسی، دانشکده روان‌شناسی و علوم تربیتی دانشگاه خوارزمی، تهران، ایران، mgramipour@khu.ac.ir

۴- دانشیار گروه سنجش و اندازه‌گیری، دانشکده سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران، asghar.minaei@yahoo.com

مقدمه

در جمهوری اسلامی ایران، جوانانی که علاقه‌مند به استخدام در نیروی انتظامی هستند از طریق شرکت در آزمون ورودی و کسب نمره رتبه قبولی مورد نیاز پذیرفته شده و در مقاطع درجه‌داری یا افسری مشغول به آموزش می‌شوند که به این قبیل افراد، فراگیر اطلاق می‌شود؛ اما علاوه بر شیوه جذب دانشجویان از طریق یادشده، تعداد زیادی از افرادی که جذب نیروی انتظامی می‌شوند، شامل فارغ‌التحصیلان سایر دانشگاه‌های کشور و مراکز آموزش عالی کشور در مقاطع مختلف (حداقل فوق‌دیپلم) هستند که پس از طی مراحل جذب و گزینش، برای طی دوره آموزش‌های عمومی و تخصصی رشته شغلی خود پرداخته و به خدمت پلیس کشور درمی‌آیند و یا پس از اتمام آموزش‌های اولیه برای تکمیل آموزش‌های تخصصی به سایر دانشکده‌های دانشگاه علوم انتظامی امین ناجا معرفی شده و دوره تخصصی را طی می‌نمایند؛ این قبیل افراد را دانش‌آموخته (فراگیر) می‌نامند.

آزمون‌های بزرگ‌مقیاسی مانند آزمون‌های استخدامی نیروی انتظامی جمهوری اسلامی ایران که مرکز گزینش و استخدام ناجا برگزار می‌کند، نقش کلیدی در آینده جوانان دارند. داوطلبان بسیار زیادی بعد از گذراندن دوره متوسطه برای ورود به دانشگاه باید با یکدیگر در کنکور دیگری رقابت کنند. رقابت در چنین آزمونی مشکل بوده و شاید بتوان گفت مهم‌ترین آزمون پس از آزمون سراسری در ایران است که به صورت سالیانه به صورت سراسری برگزار می‌شود. داوطلبان و متقاضیان زیادی در سراسر کشور وجود دارد که علاقه‌مند به خدمت در نیروی انتظامی بوده و نگران پذیرفته نشدن در ناجا در همان مرحله اول (آزمون) می‌باشند.

موفقیت افراد در سؤالات آزمون می‌تواند تحت تأثیر عوامل ثانویه باشد. در دهه ۱۹۶۰، نه تنها مراکز آموزشی بلکه مردم عادی نیز به دنبال عادلانه بودن آزمون‌ها بودند (همبلتون^۱، سوامیناتان و راجرز^۲، ۱۹۹۱). هم‌اکنون آزمون‌ها و حرفه آزمون‌گری مورد توجه دقیق مردم قرار گرفته و اکنون منتشرکنندگان و مجریان آزمون‌ها باید نشان دهند که آزمون‌هایشان در برابر اقلیت‌ها، عاری از سوگیری هستند (همبلتون، سوامیناتان و راجرز، ۱۹۹۱؛ ترجمه فلسفی نژاد، ۱۳۸۹). از این رو عادلانه بودن برای آزمون‌های سرنوشت‌ساز به‌عنوان آزمون‌هایی که

1- Hambleton

2- Swaminathan and Rogers

جهت تصمیم‌گیری برای آینده آزمودنی‌ها به کار می‌روند از اهمیت بیشتری برخوردار است (پاپهام^۱، ۲۰۰۵).

یکی از مهم‌ترین تهدیدها برای اعتبار آزمون، سوگیری سؤال است که از آن تحت عنوان کنش افتراقی سؤال^۲ نیز یاد می‌شود. وجود کنش افتراقی گواهی علیه عادلانه بودن آزمون است، اگر پاسخ به سؤال‌های آزمون با کنترل توانایی آزمودنی‌ها تابعی از ویژگی‌های گروهی که آزمودنی‌ها به آن تعلق دارند، مانند (جنسیت، پایگاه اجتماعی و اقتصادی و زبان و ...) باشند، سؤال‌های آزمون نسبت به آن‌ها کنش افتراقی دارند. در آزمون‌های روانسنجی معمولاً فرض می‌شود که ویژگی‌های اندازه‌گیری برای تمامی افراد، پایدار است. این ویژگی تحت عنوان تغییرناپذیری اندازه‌گیری^۳ شناخته شده است (میلسپ، ۲۰۱۲). زمبو^۴ (۲۰۰۷) استدلال کرد که تجزیه و تحلیل کنش افتراقی ممکن است لزوماً نشان دهنده بی‌انصافی یا سوگیری در آزمون نباشد بلکه به دلیل اختلافات مشاهده شده در طبیعت جمعیت مورد مطالعه باشد. با این حال ممکن است که گروه‌های مختلف افراد به همان آزمایش، به‌طور متفاوتی واکنش نشان دهند که در این صورت اعتبار اندازه‌گیری مورد تهدید قرار می‌گیرد. همچنین اگر آزمون‌ها نتایج مختلفی برای گروه‌های متجانس (متمايز) از افراد ارائه دهد، عادلانه بودن آزمون نقض می‌شود.

سوگیری اندازه‌گیری در سالیان اخیر مورد توجه زیادی قرار گرفته است و روان‌سنج‌ها تحت فشار افکار عمومی و نظام‌های حقوقی مجبور به ارائه شواهدی هستند که نشان دهند آزمون‌های شان قابل اتکا و اعتماد بوده و فاقد سوگیری باشد (کامیلی^۵، ۲۰۰۶). تلاش‌ها در جهت مقابله با این چالش منجر به پیدایش مدل‌ها و روش‌های مختلفی شده است. مدل‌های مختلف بر اساس نیاز آزمون‌گران و تأثیر مؤلفه‌های مختلف به‌مرور بهبود یافته‌اند اما با این وجود هنوز توافق قابل قبولی در بین روان‌سنج‌ها و آزمون‌سازان در استفاده از رویکرد مناسب در شرایط تغییر پذیر بافتی تحقیق وجود ندارد.

روش‌های مختلف برای تشخیص کنش افتراقی را می‌توان با استفاده از روش‌های

-
- 1- Popham
 - 2- Differential Item Functioning (DIF)
 - 3- measurement invariance
 - 4- Zumbo
 - 5- Camilli

مبتنی بر نظریه کلاسیک^۱ و روش‌های مبتنی بر نظریه پاسخ سؤال^۲ طبقه‌بندی نمود. ایس و راجو^۳ (۲۰۰۳) اظهار داشتند که روش‌های منتل-هنسل^۴، منتل هنزل تعمیم‌یافته^۵؛ رگرسیون لجستیک^۶ و نمودار دلنا^۷ تحت روش‌های نظریه کلاسیک قرار دارند، درحالی‌که روش‌های مجذور کای-دو لرد^۸، راجو و روش‌های آزمون نسبت درست‌نمایی^۹ تحت مدل‌های نظریه پاسخ سؤال انجام شده است.

به‌طور کلی روش‌های موجود از چهار رویکرد، نمره مورد انتظار سؤال، روش‌های نظریه پاسخ سؤال، روش‌های نا پارامتریک و مدل خطی تعمیم‌یافته گسترش یافته‌اند. با این وجود هر یک از رویکردها مفروضات و خلأهایی دارند. مطالعه ادبیات تحقیق نشان می‌دهد، در ۶۰ سال اخیر روش‌های متفاوتی معرفی شده است که این امر نشان‌دهنده اهمیت مسئله است. درباره هیچ‌یک از روش‌ها و رویکردها توافق و دانش کلی برقرار نیست و محقق باید دست به انتخاب بزند.

همه روش‌های تشخیص کنش افتراقی سؤال، می‌توانند تنها با یک متغیر گروهی رسته‌ای سروکار داشته باشند. اگرچه علوم جدید کنش افتراقی سؤال که پدیدار شده‌اند، هم‌زمان می‌توانند بر روی تأثیر بیش از یک متغیر گروهی تحقیق کنند. به‌علاوه به متغیرهای پی‌درپی این اجازه داده می‌شود که به‌عنوان یک منبع کنش افتراقی سؤال مورد تحلیل قرار بگیرند. البته درجایی که هیچ زیرگروه از پیش تعیین‌شده‌ای نیاز نباشد. این‌ها همان روش‌های درخت‌های سؤال متمرکز^{۱۰} هستند که ممکن است در تحلیل مدل راش یا رگرسیون لجستیک به کاربرده شوند (تاتز و برگر^{۱۱}، ۲۰۱۶).

بنابراین با توجه به مطالب فوق بیان گردید که روش‌های مختلفی جدیدی برای تشخیص کنش افتراقی پدید آمده است. از جمله می‌توان به روش رگرسیون لجستیک ترتیبی

-
- 1- Classical Test Theory (CTT)
 - 2- Item Response Theory (IRT)
 - 3- Ellis and Raju
 - 4- Mantel-Haenszel (MH)
 - 5- Generalized Mantel-Haenszel (GMH)
 - 6- Logistic Regression (LR)
 - 7- delta plot (SIBTEST)
 - 8- Lord's chi-square
 - 9- likelihood ratio test
 - 10- Item-Focused Trees (IFT)
 - 11- Tutz and Berger

هیبریدی^۱ و درخت‌های راش اشاره نمود. درحالی که روش رگرسیون لجستیک ترتیبی هیبریدی بر اساس خواص نظریه پاسخ سؤال عمل می‌کند. لیکن از مدل درخت‌های راش که در این تحقیق از آن استفاده می‌شود، به دلیل اینکه چندین متغیر باهم در نظر گرفته می‌شوند برای تعیین کنش افتراقی استفاده می‌شود. مزیت دیگر درخت‌های راش این است که در مقایسه با اکثر روش‌های تعیین کنش افتراقی نیازی به مشخص کردن گروه‌های کانونی و مرجع به‌عنوان پیش‌نیاز نیست. در روش درخت‌های راش هنگام تعیین گروه‌ها، پارامترهای سؤال را در کلیه متغیرهای کمکی سؤال مورد بررسی قرار می‌دهند. گروه‌ها طبق متغیرهای کمکی که دارای قویترین بی‌ثباتی (عدم تناسب پارامترهای سؤال در گروه‌ها) باشند؛ مشخص می‌شوند. به‌عنوان مثال، اگر بخواهیم مشخص کنیم که آیا کنش افتراقی از نظر جنسیت و انگیزه ذاتی وجود دارد؟ می‌توان آن را با توجه به جنسیت تفکیک و سپس بر اساس نمره انگیزه ذاتی تمایز داد. درحالی که متغیر کمکی که قویترین بی‌ثباتی را به خود اختصاص می‌دهد جنسیت است و نمره انگیزه ذاتی دومین متغیر است.

علاوه بر این؛ روش درخت‌های راش نسبت به سایر روش‌ها در تعیین امتیاز (نمره) برش به‌عنوان یک ویژگی برتر، برخوردار است. در روش‌هایی که از گروه‌های از پیش تعریف شده با متغیرهای پیوسته استفاده می‌کنند، میانگین حسابی یا میان به‌عنوان نقاط برش مورد استفاده قرار می‌گیرد. در روش درخت‌های راش درحالی که متغیرهای پیوسته را گروه‌بندی می‌شوند، مقداری که بیشترین اختلاف پارامتر را دارد به‌عنوان نقطه برش در نظر گرفته می‌شود. به‌عنوان مثال برخی سؤالات کنش افتراقی را با نمرات انگیزه ذاتی دانشجویان نشان می‌دهند. هنگام تعیین گروه کانونی و مرجع به‌جای استفاده از میانگین حسابی یا میان، میانگین نمرات انگیزه ذاتی افراد که اختلاف پارامتر در روش درخت‌های راش بالاترین است به‌عنوان نقطه برش در نظر گرفته می‌شود. این انتخاب یک انتخاب دلخواه است و ممکن است با پارامتر واقعی متفاوت باشد که نشان‌دهنده قوی‌ترین تغییر پارامتر است. این عمل ممکن است باعث شود اختلاف پارامتر واقعی توسط یک نقطه برش دیگر پنهان شود (استرویل، کویف و زالایس، ۲۰۱۵).

1- Hybrid Ordinal Logistic Regression (HOLR)

2- Strobl, Kopf and Zeileis

گرامی پور^۱ (۲۰۲۰) در مقاله‌ای با عنوان «رویکرد درخت‌های سوال- متمرکز در تحلیل کنش افتراقی سؤال مورد مطالعه: آزمون درک مطلب EFL» آزمون درک مطلب بخش سؤالات عمومی انگلیسی شامل تعداد ۲۰ سؤال را مورد تحلیل قرار داد. وی متغیر کمکی و پیوسته را جنسیت و سوابق تحصیلی به صورت هم‌زمان برای تحلیل درخت‌های سؤال متمرکز در نظر گرفت. همچنین ایشان برای تشخیص کنش افتراقی سؤال از رگرسیون لجستیک از بسته DIFtree از نرم‌افزار R برای تشخیص کنش افتراقی یکنواخت استفاده نمود. نتایج پژوهش وی نشان داد که از تعداد ۲۰ سؤال بخش درک مطلب، تعداد ۱۰ سؤال دارای DIF یکنواخت که در آن تعداد ۲ تقسیم (۲ متغیر پیشین مشترک) و ۸ سؤال فقط یک تقسیم داشته‌اند. علاوه بر این ۶ تقسیم و ۵ سؤال از کنش افتراقی غیریکنواخت، در تحلیل کنش افتراقی سؤال غیریکنواخت یافت شده‌اند که تنها ۱ سؤال از آن‌ها دارای دو منبع متغیر دارای کنش افتراقی حاصل از شبیه‌سازی بوده است. گرامی پور؛ رضایی صدر؛ رضایی؛ نوروزی و مختاریان (۱۳۹۶) در مقاله‌ای با عنوان «کنش افتراقی سوال (DIF) و آزمون (DIF) مرتبط با جنسیت در آزمون‌های کنکور سراسری سازمان سنجش آموزش کشور» به بررسی کنش افتراقی سوال بر حسب جنسیت پرداختند در این تحقیق که با استفاده از رگرسیون لجستیک انجام شده است به این نتیجه رسیدند که حدود ۱۴٪ آزمون‌های کنکور سراسری از سال ۱۳۸۷ تا ۱۳۹۰ دارای کنش افتراقی جنسیتی هستند و سوالات این آزمون‌ها به جنس مؤنث سوگیری دارند. آریا دوست^۲ (۲۰۱۸) در مقاله‌ای با عنوان «استفاده از درخت‌های راش پارتیشن‌بندی بازگشتی برای بررسی کنش افتراقی سؤال در آزمون خواندن زبان دوم» یا استفاده از روش افراز بازگشتی درخت‌های راش برای شناسایی منابع کنش افتراقی درخت‌های راش نمونه را به پراکندگی غیرخطی برگشت پذیر و برآورد دشواری سؤال در هر پارتیشن تقسیم نمود. داده‌های او عبارت‌اند بود از واژگان و دانش گرامر و جنسیت تست شوندگان، وی نشان داد که آزمون دهندگان دارای نمرات گرامری بالا بوده و وجود کنش افتراقی ناشی از متغیرهای آشکار تنها بر زیرگروه‌های خاصی از آزمون گیرندگان با مشخصات خاص توانایی تأثیر می‌گذارد. لیو^۳ و همکاران (۲۰۱۹) در مقاله‌ای با عنوان «مقایسه روش‌های مختلف تشخیص کنش افتراقی سؤال در مدل‌های

1- Gramipour

2- Aryadoust

3- Liu

تشخیصی شناختی» با استفاده از یک مطالعه مونت کارلو به بررسی عملکرد روش‌های متل-هنسزل، رگرسیون لجستیک و آزمون‌های والد مبتنی بر اطلاعات سؤال، اطلاعات مشاهده‌شده و ماتریس کوواریانس در خصوص تشخیص کنش افتراقی سؤال پرداختند. یوکسل^۱ و همکاران (۲۰۱۹) در مقاله‌ای با عنوان «کنش افتراقی سؤال شاخص‌های بهداشتی نائینگهام با استفاده از مدل راش آمیخته» با استفاده یک متغیر پاسخ اسمی، بهترین مدل را برای متغیرهای تحرک بدنی، خواب و انزوای اجتماعی معرفی نمودند. نتایج آنان نشان داد که برای بعد انزوای اجتماعی در برخی سؤالات کنش افتراقی از لحاظ جنسیت وجود دارد. همچنین سؤالات از نظر خواب به لحاظ سن، جنسیت و مدت زمان درد قابل توجه نبودند. در سایر سؤالات نیز به لحاظ جنسیت، سن و مدت درد کنش افتراقی وجود نداشت. در این تحقیق برای تشخیص کنش افتراقی سؤال از مدل درخت‌های راش با استفاده از داده‌های واقعی استفاده می‌شود. از این رو، در پژوهش حاضر سعی بر آن است که اعتبار آزمون‌های استفاده‌ی از منظر کنش افتراقی سؤال‌های آزمون در درس مصوب آزمون، مورد بررسی و تحلیل قرار گیرد. لذا هدف اصلی پژوهش نرخ آشکارسازی کنش افتراقی سؤال با استفاده از مدل درخت‌های راش برای آزمون‌های سرنوشت‌ساز ناجا است.

مبانی نظری

ایده اصلی مفهوم درخت‌های راش، توسط استروبل و همکاران (۲۰۱۵) ارائه شد. تفسیر کنش افتراقی سؤال توسط مدل درخت‌های راش نسبت به سایر روش‌ها آسانتر است. این روش برای گروه‌هایی که از پیش مشخص نشده‌اند کاربرد دارد؛ بنابراین؛ می‌توان درک درستی از منابع روان‌شناختی کنش افتراقی سؤال به دست آورد. درخت‌های راش یک درخت واحد برای کل ترکیبات برای شناسایی سؤالات سودار ایجاد می‌کند. درخت‌های راش ابتدا پارامترهای سؤال، برآورد شده و به‌طور مشترک برای نمونه کل استفاده می‌شود. سپس پایداری پارامترهای سؤال با توجه به هر متغیر ارزیابی می‌شود. مدل توأم با محاسبه انحرافات از مدل مشترک برای هر یک از آزمودنی‌ها انجام می‌شود. انحرافات بر اساس متغیرهای کمکی مرتب می‌شود (به‌عنوان مثال؛ متغیر کمکی سن، از جوان‌ترین به پیرترین). اگر ترتیب متغیرهای کمکی به‌جای نوسان از یک تغییر منظم و سیستماتیک در دامنه مقادیر احتمال برخوردار باشد گفته می‌شود که کنش افتراقی رخ داده است.

به لحاظ فنی، مراحل و فرایند زیر برای رسیدن به یک ساختار از یک درخت راش به صورت زیر است:

۱) با شروع نمونه کامل، پارامترهای حاشیه‌ای سؤال برای همه افراد در نمونه را برآورد نمایید؛

۲) ثبات (پایداری) پارامترهای سؤال را با توجه به متغیرهای کمکی ارزیابی کنید؛

۳) اگر ناپایداری قابل توجهی وجود دارد، نمونه را در میان متغیرهای کمکی با بزرگ‌ترین بی‌ثباتی تقسیم کنید و مدل را در نقطه برش به بالاترین برازش بهبود دهید؛

۴) مراحل ۱ تا ۳ را به صورت بازگشتی در زیرمجموعه‌های حاصل تکرار کنید تا زمانی که هیچ ناپایداری قابل توجهی مشاهده نشود.

این ۴ مرحله با جزئیات بیشتری در زیر توضیح داده می‌شود.

برای برآورد پارامترهای سؤال مدل درخت‌های راش، می‌توان از روش ماکزیمم درست‌نمایی شرطی به دلیل اینکه با روش ماکزیمم درست‌نمایی سازگاری دارد، استفاده نمود. فرض کنید پارامترهای افراد $\theta_i, i = 1, \dots, n$ و $\beta_j, j = 1, \dots, m$ پارامترهای مربوط به سؤال و مربوط به پاسخ موضوع به سؤال باشد. تحت مدل راش داریم:

$$P(U_{ij} = u_{ij} | \theta_i, \beta_j) = \frac{e^{u_{ij}(\theta_i - \beta_j)}}{1 + e^{\theta_i - \beta_j}}$$

پارامترهای افراد به فرم آماره بسنده $r_i = \sum_{j=1}^m u_{ij}$ محاسبه می‌شود. پارامترهای سؤال نیز را می‌توان از روش تکرار درست‌نمایی شرطی به صورت زیر برآورد نمود:

$$L_C(\beta | r_1, \dots, r_m) = \prod_{i=1}^n L_C(\beta | r_i) = \prod_{i=1}^n \frac{e^{-\sum_{j=1}^m u_{ij} \cdot \beta_j}}{\gamma_{r_i}(\beta)} \quad (1)$$

که در آن تابع متوازن برای هر ترتیبی از (فیشر و مولانر^۱، ۱۹۹۵) است. برای مشخص کردن مقیاس مشخص باید محدودیت‌هایی را اعمال کنیم. به‌عنوان مثال برای پارامترهای سؤال، مقدار صفر و برای سایر جاها، مقدار لحاظ می‌شود.

به‌منظور بررسی اینکه آیا پارامترهای سؤال بین گروه‌های افراد توسط متغیرهای کمکی

تیین می‌شوند؟ ما از روش آزمون تغییرات ساختاری بکار رفته در اقتصادسنجی، استفاده می‌کنیم. این آزمون‌ها معمولاً برای شناسایی، به‌عنوان مثال، افت بازده سهام با گذشت زمان، به کار می‌رود در حالی که در اینجا ما از همین روش استفاده می‌کنیم تا تغییرات پارامتر را نسبت به متغیر کمکی افراد را بررسی کنیم.

دلایل منطقی استفاده از آزمون‌های تغییرات ساختاری به‌صورت زیر است:

- ۱) ابتدا برآورد حاشیه‌ای پارامترهای سؤال برای کل نمونه، به دست می‌آید؛
- ۲) انحرافات فردی از این مدل مشترک با توجه به متغیرهای کمکی مانند سن، محاسبه می‌شود؛
- ۳) اگر کنش افتراقی هماهنگ در رابطه با گروه‌هایی که توسط متغیرهای مختلف ایجاد شده‌اند، وجود داشته باشد، آنگاه تغییرات منظم در انحرافات فردی ارائه می‌شود؛
- ۴) اگر از طرف دیگر، هیچ کنش افتراقی وجود نداشته باشد، مقادیر صرفاً به‌طور تصادفی در حال نوسان است.

برای آزمون آماری تغییر ساختاری در پارامترهای مدل، از روش آزمون M بی‌ثباتی تعمیم‌یافته (زالایس و هورنیک، ۲۰۰۷) استفاده می‌شود.

عملکرد نمره فردی $i = 1, \dots, n$ را برای مشاهده را در نظر بگیرد. مشتق این عملکرد فردی با لگاریتم درست‌نمایی است $\Psi(u_i, \beta)$ که با توجه به بردار پارامتر، یک اندازه‌گیری کلی از انحراف برای مدل‌های درست‌نمایی-پایه است. برای مدل راش این سهم مشارکت فردی را می‌توان به راحتی از طریق درست‌نمایی شرطی برآورد نمود. فرایند جمع تجمعی به صورت زیر تعریف می‌شود:

$$W_\ell(t) = \hat{V}^{-1/2} n^{-1/2} \sum_{i=1}^{[n.t]} \psi(u_{i|\ell}, \hat{\beta}) \quad (0 \leq t \leq 1) \quad (2)$$

که در آن $(i | 1)$ ؛ i امین مشاهده ترتیبی با I امین متغیر کمکی و $[.]$ قسمت صحیح عبارت $(i) = \sum_{\ell=1}^L \psi(u_{i|\ell}, \hat{\beta})$ برآورد حاصل ضرب بیرونی شیب ماتریس کوواریانس است و t کسری از اندازه نمونه است. تحت فرض صفر (پایدار بودن پارامتر)، فرایند جمع تجمعی W_1 (.) به یک پل $1-m$ بعدی ابزاری بروانی که می‌تواند به‌عنوان پایه‌ای برای استنباط آماری استفاده شود، همگرا گردد (زالایس و هورنیک، ۲۰۰۷).

به همین ترتیب برای جمع تجمعی بیش از امین متغیر کمکی نیز داریم:
 با $i=1, \dots, n$ انحراف فردی با توجه به ترتیب متغیر کمکی در هر مرحله تا عنصر t جمع می‌شوند. هنگامی که $W I(t)$ به‌عنوان تابعی از اندازه نمونه در نظر گرفته می‌شود، تحت فرض صفر (پایدار بودن پارامتر)، روند جمع تجمعی مسیر یک فرایند تصادفی با میانگین صفر ثابت را دنبال می‌کند (درحالی که تحت فرض مقابل، عدم ثبات پارامتر، یک مسیر نوسان کاملاً تصادفی منحرف می‌شود. مزیت این رویکرد این است که مدل مجبور نیست انشعابات همه متغیرها را مجدداً مورد ارزیابی قرار دهد، زیرا انحرافات فردی یکسان باقی می‌مانند و فقط مسیر مربوط به $W_{-}(I)(t)$ برای ارزیابی متغیرهای کمکی نیاز به تعدیل دارد. برای گرفتن انحراف هماهنگ $W_{-} I$ (.) آماره آزمون مختلف، بسته به اینکه آیا امین متغیر، متغیر عددی به طبقه‌ای است، می‌تواند مورد استفاده قرار گیرد. اگر عددی باشد، گفته می‌شود که آمار آزمون زیر نرمال است:

$$S_{\ell} = \max_{i=\bar{i}, \dots, \bar{i}} \left(\frac{i}{n} \cdot \frac{n-i}{n} \right)^{-1} \left\| w_{\ell} \left(\frac{i}{n} \right) \right\|_2^2 \quad (3)$$

این عبارت را می‌توان ماکزیمم لاگرانژ^۱ به‌عنوان یک گزینه جایگزین تکی برای همه نقاط برش تعبیر کرد. توزیع محدود کننده برتری یک فرایند گره پایین باسل^۲ است که از آن می‌توان مقادیر را محاسبه کرد (زایلیس و همکاران، ۲۰۰۸ و مرکل^۳ و زالیس، ۲۰۱۳). از طرف دیگر، اگر امین متغیر کمکی رسته‌های باشد (با مقدار از متغیر رسته‌های)؛ استفاده از آماره آزمون زیر نرمال خواهد بود:

$$S_{\ell} = \sum_{q=1}^Q n \left(\sum_{i=1}^n I(x_{i\ell} = q) \right)^{-1} \left\| \Delta_q w_{\ell} \left(\frac{i}{n} \right) \right\|_2^2 \quad (4)$$

که در آن ارتقای به امین متغیر رسته‌های است. این آماره آزمون، برای مرتب‌سازی مجدد در مقوله‌های (رسته‌های) و افراد در هر گروه (رسته) تغییرناپذیر است. این آماره آزمون عدم ثبات را برای هر محاسبه می‌کند. این تابع توزیع دارای توزیع کای دو () با درجه آزادی

1- Lagrange
 2- Bessel
 3- Merkle

برای هر است. این آزمون برای هر دو متغیر طبقه‌ای اسمی و ترتیبی استفاده می‌شود. برای مدل راش، تابع هدف مورد استفاده برای برآورد پارامتر، لگاریتم درست‌نمایی شرطی است. کمکی‌های فردی لگاریتم درست‌نمایی شرطی را می‌توان به آسانی از لگاریتم ساخت برای محاسبه آزمون‌های تغییرات ساختاری، سهم‌های فردی در عملکرد نمره از معادله شماره ۵ حاصل می‌شود. سهم امین موضوع برای امین پارامتر سؤال عبارت است از: گفتنی است مشتقات دوباره عبارت با حذف گردیده است (فیشر و مولنار، ۱۹۹۵). در اجرای درخت راش، از الگوریتم لیو^۱ (۱۹۹۴) برای محاسبه مشتقات استفاده می‌شود. هنگامی که سهم‌های فردی به عملکرد نمره مدل راش از معادله شماره ۶ محاسبه شد و در معادله شماره ۲ درج می‌شود.

در این پژوهش برای آشکارسازی کنش افتراقی سؤال از مدل درخت‌های راش و برای مقایسه رسته‌های تعیین شده توسط مدل و مشخص نمودن سؤالات سودار از روش بنجامین-هوچبرگ^۲ (۱۹۹۵) استفاده شده است. این روش بر اساس آزمون نرخ کشف کاذب^۳ عمل مینماید. آزمون نرخ کشف کاذب؛ یک رویکرد آماری است که در آزمون چندین فرضیه برای تصحیح مقایسه‌های متعدد مورد استفاده قرار می‌گیرد. در این آزمون؛ هنگام آزمایش فرضیه صفر برای تعیین اینکه آیا نمره مشاهده شده از نظر آماری معنادار است، فاصله اطمینان برای P -مقدار با سطح معناداری α محاسبه می‌شود. هنگامی که فرضیه‌های K به‌طور هم‌زمان با سطح اطمینان α مورد آزمایش قرار می‌گیرند، احتمال وقوع نادرست مثبت (رد فرضیه صفر در صورت واقعی بودن) برابر است با که می‌تواند منجر به میزان خطای زیاد در آزمایش شود؛ برای تعدیل اقدامات اطمینان آماری بر اساس تعداد آزمایش‌های انجام شده، به تصحیح آزمایش چندگانه مانند نرخ کشف کاذب نیاز است. اساس روش بنجامین-هوچبرگ به صورت ریاضی زیر است:

فرض کنید آزمون بر پایه P -مقدارهای $\alpha = 0.05$ باشند. همچنین فرض کنید مقادیر مرتب شده P مقدار و فرض صفر بر اساس مقدار باشد. آزمون فرض چندگانه به صورت زیر تعریف می‌شود:

به عبارتی آزمون نرخ کشف کاذب، امید ریاضی نسبت رد به اشتباه فرض را کمتر

- 1- Liou
- 2- Benjamini and Hochberg (BH)
- 3- False Discovery Rate (FDR)

می‌کند. در این روش حساسیت به رد کردن فرض کمتر از سایر روش‌ها مانند بونفرونی^۱ است و با مقایسه p -مقادیرها؛ زمانی فرض رد می‌شود که حداقل یک جفت رسته پیدا شود که تفاوت معناداری باهم داشته باشند. (بنجامین و هوچبرگ؛ ۱۹۹۵: ۲۹۳).

یافته‌های تحقیق

اطلاعاتی که از پاسخنامه‌ی تعداد ۲۴۱۴ نفر از آزمون دهندگان آزمون ورودی به دست آمد، نشان داد که نمره میانگین این آزمون ۱۰۰ سؤالی، ۲۷/۷۳ بود که انحراف معیار آن ۸/۶۶ به دست آمد. جدول ۱ برخی از آمارهای توصیفی این آزمون را بدون در نظر گرفتن سن و سایر مشخصات ملاحظه می‌کند.

جدول ۱: آمارهای توصیفی

| انحراف معیار | خطای استاندارد میانگین‌ها | میانگین | حداکثر | حداقل | تعداد کل |
|--------------|---------------------------|---------|--------|-------|----------|
| ۸/۶۶ | ۰/۱۷۶ | ۷۲/۷۳ | ۵۸ | ۱ | ۲۴۱۴ |
| | | | | | نمره |

همان‌طور که در جدول ۱ دیده می‌شود، بیشترین نمره کسب‌شده توسط آزمون دهندگان ۵۸ از ۱۰۰ پاسخ درست بوده و کمترین نمره ۱ بوده است، زیرا برای جواب‌های نادرست نمره‌ی منفی در نظر گرفته شده بود.

متغیرهایی که در مطالعه‌ی سؤال‌های جداشده‌ی این تحقیق مورد بررسی قرار گرفتند، سن و گروه آزمایشی (ریاضی و فیزیک، علوم انسانی و علوم تجربی) بود. از میان ۲۴۱۴ آزمون دهنده، تعداد ۵۳۸ نفر (۲۲٪) دارای سن ۱۸ سال و تعداد ۱۸۷۶ نفر (۷۸٪) دارای سن ۱۹ سال بوده‌اند. میانگین نمره‌ی آزمون دهندگان دارای ۱۸ سال، ۲۷/۵۸ و میانگین نمره‌ی آزمون دهندگان دارای ۱۹ سال، ۲۷/۷۰ بود. جدول ۲ بخشی از آمارهای توصیفی گروه سنی را نشان می‌دهد.

جدول ۲: آمارهای گروه سنی

| سن | تعداد | نمره میانگین | انحراف معیار | خطای استاندارد میانگینها |
|--------|-------|--------------|--------------|--------------------------|
| ۱۸ سال | ۵۳۸ | ۲۷/۸۵ | ۸/۴۹ | ۰/۳۶۵ |
| ۱۹ سال | ۱۸۷۶ | ۷۰/۲۷ | ۸/۷۰ | ۰/۲۰۱ |

جدول ۲ به وضوح نشان می‌دهد که انحراف معیار محاسبه شده برای آزمون دهندگان گروه‌های سنی ۱۸ و ۱۹ سال به ترتیب ۸/۴۹ و ۸/۷۰ است که تفاوت اندک میانگین (۰/۲۱ = ۸/۴۹ - ۸/۷۰) نشان‌دهنده تغییرپذیری پایین است. برای مشخص کردن این که گروه سنی دارای ۱۸ سال و ۱۹ سال در توانایی پنهان مورد آزمون به شکل معناداری تفاوت دارند یا خیر، آزمون تی نمونه‌های مستقل اجرا شد. جدول ۳ نتایج به دست آمده از آزمون تی را نشان می‌دهد.

جدول ۳: آزمون تی نمونه‌های مستقل برای گروه سنی

| | | آزمون لون برای برابری واریانس‌ها | | آزمون تی برای برابری میانگین‌ها | | |
|--------|-----------------------|----------------------------------|--------------|---------------------------------|------------|--------------|
| | | F | سطح معناداری | t | درجه آزادی | سطح معناداری |
| امتیاز | فرض برابری واریانس‌ها | ۱/۳۵ | ۰/۲۴۴ | ۰/۳۷۰ | ۲۴۱۲ | ۰/۷۱۱ |

نتایج آزمون تی نمونه‌های مستقل که در جدول ۳ نشان داده شده است، نشان دهنده عدم تفاوت قابل توجهی میان نمره‌ی میانگین آزمون دهندگان گروه سنی است. با اینکه به نظر می‌آید گروه سنی ۱۸ سال نمره‌ی میانگین بالاتری نسبت به گروه سنی ۱۹ سال دارند، اندازه اثر چنین تفاوت اندکی بین میانگین‌ها (۰/۰۱) است که تأثیر اندازهای کمی محسوب می‌شود.

با توجه اندازه اثر بسیار پایین می‌توان از تفاوت نمره‌ی میانگین بین آزمون دهندگان گروه سنی ۱۸ و ۱۹ سال چشم‌پوشی کرد؛ بنابراین این دو گروه قابل مقایسه هستند. همچنین از میان ۲۴۱۴ آزمون دهنده، تعداد ۹۵۱ نفر (۳۹٪) گروه ریاضی و فیزیک و تعداد ۱۶۰ نفر (۷٪) گروه علوم تجربی و تعداد ۱۳۰۳ نفر (۵۴٪) گروه علوم انسانی بوده‌اند. میانگین نمره‌ی آزمون دهندگان گروه ریاضی، ۲۷/۹۷ و گروه تجربی ۲۵/۴۱ گروه علوم انسانی ۲۷/۸۵ بودند. جدول ۴ بخشی از آمارهای توصیفی گروه آزمایشی را نشان می‌دهد.

جدول ۴: آمارهای گروه سنی

| گروه | تعداد | نمره میانگین | انحراف معیار | خطای استاندارد میانگینها |
|--------|-------|--------------|--------------|--------------------------|
| ریاضی | ۹۵۱ | ۲۷/۹۷ | ۸/۷۴ | ۰/۲۸ |
| تجربی | ۱۶۰ | ۲۵/۴۱ | ۸/۸۶ | ۰/۷۰ |
| انسانی | ۱۳۰۳ | ۲۷/۸۵ | ۸/۵۳ | ۰/۲۴ |

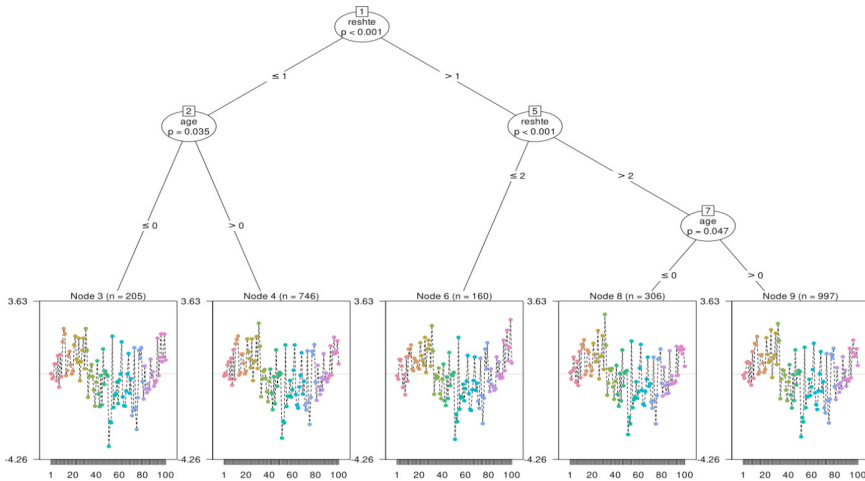
جدول ۴ به وضوح نشان می‌دهد که انحراف معیار محاسبه شده برای آزمون دهندگان گروه‌های آزمایشی به ترتیب ۸/۷۴، ۸/۸۶ و ۸/۵۳ است که تفاوت اندک میانگین باهم دارند و نشان‌دهنده تغییرپذیری پایین آنها است. برای مشخص کردن این که گروه آزمایشی در توانایی پنهان مورد آزمون به شکل معناداری تفاوت دارند یا خیر، آزمون تی نمونه‌های مستقل اجرا شد. جدول ۵ نتایج به دست آمده از آزمون تی را نشان می‌دهد.

جدول ۵: آزمون تی نمونه‌های مستقل برای گروه آزمایشی

| | | آزمون لون برای برابری واریانس‌ها | | آزمون تی برای برابری میانگین‌ها | | |
|--------|---------------|----------------------------------|--------------|---------------------------------|------------|--------------|
| | | F | سطح معناداری | t | درجه آزادی | سطح معناداری |
| امتیاز | ریاضی، انسانی | ۰/۳۹۳ | ۰/۵۳۱ | ۰/۲۸۳ | ۲۲۵۲ | ۰/۷۷۷ |
| | ریاضی، تجربی | ۰/۱۳۷ | ۰/۷۱۱ | ۳/۴۰ | ۱۱۰۹ | ۰/۰۰۱ |
| | تجربی، انسانی | ۰/۵۰۶ | ۰/۴۷۷ | -۳/۴۰ | ۱۴۶۱ | ۰/۰۰۱ |

نتایج آزمون تی نمونه‌های مستقل که در جدول ۵ نشان داده شده است، نشان‌دهنده عدم تفاوت قابل توجهی میان نمره‌ی میانگین آزمون دهندگان گروه آزمایشی است. با اینکه به نظر می‌آید گروه ریاضی نمره‌ی میانگین بالاتری نسبت به گروه تجربی و انسانی دارند. با توجه اندازه اثر بسیار پایین می‌توان از تفاوت نمره‌ی میانگین بین آزمون دهندگان گروه‌های آزمایشی چشم‌پوشی کرد؛ بنابراین این سه گروه قابل مقایسه هستند. گفتنی است که میانگین گروه علوم تجربی از گروه علوم انسانی بالاتر بوده و این تفاوت معنادار است. همان‌طور که بیان شد پاسخ‌های تعداد ۱۰۰ سؤال برای تعداد ۲۴۱۴ آزمودنی مورد بررسی واقع شد. در این مطالعه دو متغیر سن و گروه‌های آزمایشی (ریاضی و فیزیک، علوم تجربی و علوم انسانی) لحاظ گردید. داده‌های موصوف برای دو گروه مردان ۱۸ و ۱۹ سال بررسی گردید.

ابتدا با استفاده از مدل درخت راش، رسته‌ها را به نحوی تعیین می‌کنیم که سؤالات دارای کنش افتراقی را آشکار کند. برای این منظور نمودار درختی استخراج گردید.



نمودار ۱: درخت‌های راش برای سن و گروه‌های آزمایشی آزمودنی‌ها (منبع: نگارنده)

همان‌طور که در نمودار ۱ ملاحظه می‌شود در گروه‌های پایانه، برآوردهای سؤالات دشوار برای هر یک از ۱۰۰ سؤال نمایش داده شده است. در درخت‌های مشاهده می‌شود که پارامترهای سؤال برای آزمودنی‌های ۱۸ و ۱۹ سال و گروه‌های آزمایشی موجب ایجاد درخت‌های بالا و پایین گردیده است. همان‌طور که در نمودار ۱ مشاهده می‌شود تعداد ۹ گره وجود دارد که ۵ مورد از آنها گره‌های انتهایی هستند که توسط مستطیل‌های پایین درخت نشان داده شده است (گره‌های ۳، ۴، ۶، ۸ و ۹) که در آن پارامترهای دشواری سؤال برای ۱۰۰ سؤال آزمون به صورت گرافیکی آمده است. در گره بالا، تغییر در رشته (گروه آزمایشی) منجر به پارامترهای دشواری سؤال‌های مختلف برای گروه آزمایشی برای نمرات گروه ریاضی (≤ 1) و دارای علوم تجربی و علوم انسانی (> 1) است ($P < 0/001$). به عبارت دیگر، پارامترهای دشواری سؤال داوطلبان با نمرات رشته (گروه آزمایشی) با افراد گروه ریاضی با گروه‌های تجربی و انسانی متفاوت است. در مرحله بعد، داوطلبان با نمرات آزمون در گروه‌های آزمایشی دارای سن ۱۸ سال خود به تعداد ۲۰۵ نفر داوطلب در گره ۳ و تعداد ۷۴۶ نفر دارای ۱۹ سال سن در گره ۴ تقسیم می‌شوند ($P < 0/035$). به همین ترتیب در گره‌های ۵ و ۷ داوطلبان به گروه‌های آزمایشی علوم تجربی (گره ۶) و علوم

انسانی و سن افراد دارای ۱۸ سال (گروه ۸) و افراد دارای ۱۹ سال (گروه ۹) تقسیم شده‌اند. سن به تنهایی تعداد ۱۳۰۳ نفر (۵۴٪) آزمون دهندگان (تعداد ۳۰۶ نفر در گروه ۸ و تعداد ۹۹۷ نفر در گروه ۹) را تحت تأثیر قرار داده است که سطح معناداری قابل توجهی دارد. به طور کلی؛ الگوی تقسیم‌بندی مشاهده شده در نمودار ۱ در دشواری سؤال نشان می‌دهد که عدم تغییرپذیری اندازه گیری در داده‌ها وجود ندارد و زیرگروه‌های داوطلبان (گروه‌های آزمایشی) شناسایی شده تحت تأثیر DIF متفاوت هستند. همچنین بر اساس نمودار ۱ رسته‌های تعیین شده به صورت جدول ۶ تشکیل شده است.

جدول ۶: رسته‌بندی حاصل از درخت راش

| شماره رسته | رسته |
|------------|-------------------------------|
| ۱ | گروه ریاضی با ۱۸ سال سن |
| ۲ | گروه ریاضی با ۱۹ سال سن |
| ۳ | گروه علوم تجربی |
| ۴ | گروه علوم انسانی با ۱۸ سال سن |
| ۵ | گروه علوم انسانی با ۱۹ سال سن |

همان‌طور که بیان شد در این پژوهش برای پیدا کردن سؤالات دارای DIF بر اساس مدل درخت‌های راش از روش بنجامین-هوچبرگ (۱۹۹۵) استفاده شده است. این روش بر اساس آزمون FDR در بسته نرم‌افزار R انجام شد. از آنجاکه برای هر سؤال تعداد ۵ رسته موجود است، پس با زوج مقایسه برای هر سؤال روبرو هستیم. پس از محاسبه P -مقدارها برای هر ۱۰۰ سؤال دریافتیم که سؤالات ۳، ۷، ۲۸، ۳۷، ۷۱، ۷۴، ۸۱، ۸۹ و ۹۹ دارای هستند. بنابراین از بین ۱۰۰ سؤال آزمون استخدامی تعداد ۹ سؤال (۹٪ سؤالات) کنش افتراقی نسبت به آزمون دهندگان سنین ۱۸ و ۱۹ سؤال با گروه‌های آزمایشی ریاضی فیزیک؛ علوم تجربی و علوم انسانی داشتند؛ یعنی به‌عنوان سؤالاتی دارای DIF هماهنگ در درخت‌های راش شناخته شدند. در واقع در سؤالات تست هوش و فرهنگ و معارف اسلامی هیچ سؤال شامل DIF یافت نشده است. نتیجه در جدول ۶ نشان داده شده است:

جدول ۶: سؤالات دارای DIF هماهنگ

| نام درس | زبان و ادبیات فارسی | زبان عربی | اطلاعات عمومی | زبان انگلیسی |
|----------------------|---------------------|-----------|---------------|--------------|
| تعداد سؤال | ۲۱ | ۱۸ | ۱۴ | ۱۵ |
| شماره سؤال دارای DIF | ۷-۳ | ۳۷-۲۸ | ۸۱-۷۴-۷۱ | ۹۹-۸۹ |

در ادامه بر اساس رهیافت BH، در سطح، به قضاوت در خصوص فرضیات زیر می‌پردازیم. در این رهیافت برای حصول اطمینان از آن که احتمال خطای نوع اول از ۰/۰۵ تجاوز نکند، سطح آزمون را برای هر زوج مقایسه برابر با در نظر گرفته‌ایم. برای بررسی فرضیه بالا تفاضل رسته‌ها به ترتیب P- مقدار محاسبه که در نتایج در نمودار ۲ آمده است. در محور افقی نمودارهای شکل ۲، تفاضل رسته‌ها مشخص شده است. به‌عنوان مثال بیان‌گر تفاضل نتایج رسته سوم از رسته اول است (جهت‌گیری سؤال ۳ در مقایسه با سؤال ۱). تفاضل رسته‌ها نیز به ترتیب P- مقدارها از چپ به‌راست مرتب‌شده‌اند. همچنین میانگین نمونه‌ای تفاضل رسته‌ها نیز ارائه گردیده است.

نمودار ۲: نمودار مقایسه‌های زوجی به روش BH، بین رسته‌های مختلف

سؤال ۳

| | | | | | | | | | | | |
|--------|---------|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Mean | -0.154 | -0.128 | 0.111 | 0.09 | -0.064 | -0.043 | -0.037 | -0.027 | -0.017 | 0.021 | |
| Pvalue | 0.0047 | 0.0047 | 0.0152 | 0.1899 | 0.544 | 0.881 | 0.881 | 1 | 1 | 1 | |
| | C3 - C1 | | C3 - C2 | C5 - C3 | C4 - C3 | C4 - C1 | C5 - C1 | C4 - C2 | C2 - C1 | C5 - C2 | C5 - C4 |

سؤال ۷

| | | | | | | | | | | | |
|--------|---------|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Mean | 0.091 | 0.102 | -0.075 | -0.061 | 0.03 | 0.04 | 0.01 | 0.027 | 0.016 | 0.014 | |
| Pvalue | 0.0305 | 0.0583 | 0.2395 | 0.2943 | 0.4718 | 0.6826 | 1 | 1 | 1 | 1 | |
| | C3 - C2 | | C3 - C1 | C4 - C3 | C5 - C3 | C5 - C2 | C5 - C1 | C2 - C1 | C4 - C1 | C4 - C2 | C5 - C4 |

سؤال ۲۸

| | | | | | | | | | | | |
|--------|---------|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Mean | -0.093 | 0.061 | 0.051 | -0.032 | 0.064 | -0.043 | -0.029 | 0.022 | 0.019 | -0.003 | |
| Pvalue | 0.0081 | 0.2191 | 0.3052 | 0.3052 | 0.3052 | 0.9037 | 0.9037 | 1 | 1 | 1 | |
| | C3 - C2 | | C5 - C3 | C2 - C1 | C5 - C2 | C4 - C3 | C3 - C1 | C4 - C2 | C4 - C1 | C5 - C1 | C5 - C4 |

سوال ۳۷

| | | | | | | | | | | |
|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Mean | -0.109 | -0.085 | -0.1 | 0.081 | 0.057 | -0.075 | -0.028 | 0.009 | -0.019 | 0.024 |
| Pvalue | 0.0379 | 0.0379 | 0.2183 | 0.2183 | 0.2589 | 0.2612 | 0.7342 | 1 | 1 | 1 |
| | C3 - C2 | C4 - C2 | C3 - C1 | C5 - C3 | C5 - C4 | C4 - C1 | C5 - C2 | C2 - C1 | C5 - C1 | C4 - C3 |

سوال ۷۱

| | | | | | | | |
|--------|--------|--------|--------|--------|--------|--------|--------|
| Mean | -0.204 | -0.158 | -0.12 | 0.116 | -0.088 | 0.071 | -0.084 |
| Pvalue | 8e-04 | 0.0032 | 0.0156 | 0.0382 | 0.1179 | 0.1378 | 0.197 |

سوال ۷۴

| | | | | | | | | | | |
|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Mean | -0.19 | -0.119 | 0.105 | -0.085 | 0.104 | -0.085 | -0.07 | -0.015 | -0.015 | 0 |
| Pvalue | 0.0031 | 0.0543 | 0.1098 | 0.1822 | 0.1903 | 0.2902 | 0.2925 | 1 | 1 | 1 |
| | C3 - C1 | C3 - C2 | C5 - C3 | C5 - C1 | C4 - C3 | C4 - C1 | C2 - C1 | C4 - C2 | C5 - C2 | C5 - C4 |

سوال ۸۱

| | | | | | | | | | | |
|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Mean | -0.139 | 0.132 | -0.131 | 0.089 | -0.05 | 0.043 | 0.008 | -0.042 | 0.001 | -0.007 |
| Pvalue | 0.0143 | 0.017 | 0.1038 | 0.4673 | 0.8648 | 0.9408 | 1 | 1 | 1 | 1 |
| | C3 - C2 | C5 - C3 | C3 - C1 | C4 - C3 | C4 - C2 | C5 - C4 | C2 - C1 | C4 - C1 | C5 - C1 | C5 - C2 |

سوال ۸۹

| | | | | | | | | | | |
|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Mean | 0.124 | -0.098 | 0.077 | -0.093 | -0.046 | 0.005 | 0.031 | -0.016 | 0.026 | -0.021 |
| Pvalue | 0.0211 | 0.0572 | 0.2196 | 0.2248 | 0.5137 | 1 | 1 | 1 | 1 | 1 |
| | C4 - C3 | C3 - C2 | C5 - C3 | C3 - C1 | C5 - C4 | C2 - C1 | C4 - C1 | C5 - C1 | C4 - C2 | C5 - C2 |

سوال ۹۹

| | | | | | | | | | | |
|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Mean | 0.067 | 0.066 | 0.041 | 0.026 | 0.041 | -0.04 | 0.026 | 0.015 | -0.025 | 0 |
| Pvalue | 0.0207 | 0.0667 | 0.2576 | 0.2576 | 0.4102 | 0.4102 | 0.5231 | 1 | 1 | 1 |
| | C5 - C3 | C4 - C3 | C5 - C1 | C5 - C2 | C4 - C1 | C3 - C2 | C4 - C2 | C2 - C1 | C3 - C1 | C5 - C4 |

در نمودارهای ۲؛ اختلاف‌های معنادار پیرنگ شده‌اند. در سؤال شماره ۳ شاهد هستیم که رسته سوم (گروه علوم تجربی) به‌طور معناداری با دیگر رسته‌ها متفاوت است. به‌عنوان مثال؛ جهت‌گیری رسته سوم (گروه علوم تجربی) به نفع رسته اول (گروه ریاضی با ۱۸ سال سن) است چراکه اختلاف میانگین بین آن‌ها منفی شده است. به‌عنوان مثال دیگر در سؤال شماره ۹۹ داریم که رسته ۵ (گروه علوم انسانی با ۱۹ سال سن) با رسته سوم (گروه علوم تجربی) همچنین رسته ۴ (گروه علوم انسانی با ۱۸ سال سن) با رسته سوم (گروه علوم تجربی) به‌طور معناداری اختلاف دارند و با توجه به مقدار میانگین محاسبه‌شده؛ جهت‌گیری به ترتیب به سمت رسته ۵ (گروه علوم انسانی با ۱۹ سال سن) و رسته ۴ (گروه علوم انسانی با ۱۸ سال سن) بوده است. در سایر سؤالات نیز می‌توان به طریق مشابه، خروجی‌ها را تحلیل کرد. به‌منظور سادگی تحلیل و اینکه بتوانیم در یک نگاه کلی تمام جهت‌گیری را مشاهده کرد خلاصه نمودار ۲ در جدول ۷ نشان داده شده است.

جدول ۷: تعداد موارد DIF برای مقایسه دوبه‌دوی رسته‌ها به روش BH

| رسته‌ها | ۱ | ۲ | ۳ | ۴ | ۵ |
|---------|---|---|---|---|---|
| ۱ | | ۰ | ۰ | ۰ | ۰ |
| ۲ | ۲ | | ۱ | ۰ | ۰ |
| ۳ | ۶ | ۷ | | ۴ | ۵ |
| ۴ | ۲ | ۱ | ۰ | | ۱ |
| ۵ | ۲ | ۰ | ۰ | ۰ | |

اعداد نشان‌دهنده در جدول ۷؛ تعداد سؤالات دارای سوگیری به نفع رسته نشان داده شده بالا (ردیف افقی) نسبت به رسته سمت چپ (ردیف عمودی) است. به‌عنوان مثال؛ تعداد سؤالات دارای سوگیری رسته دوم نسبت به رسته اول؛ صفر است اما تعداد ۲ سؤال دارای سوگیری در رسته اول نسبت به رسته دوم به نفع رسته اول و به ضرر رسته دوم است. همچنین تعداد ۷ سؤال دارای سوگیری نسبت به رسته دوم نسبت به رسته سوم و به نفع رسته دوم و به ضرر رسته سوم است. همان‌طور که ملاحظه می‌شود جدول شماره ۷ به‌خوبی می‌تواند نمودار درخت راش (نمودار ۱) را تبیین نماید.

بحث و نتیجه گیری

هدف از پژوهش حاضر، بررسی کنشی افتراقی سؤال در آزمون استخدami ناجا با استفاده از درخت‌های راش بود. متغیرهای مورد استفاده برای بررسی DIF سن و گروه آزمایشی بودند. مشخص شد سن و گروه‌های آزمایشی باعث ایجاد اختلاف در بین داوطلبان آزمون ورودی به صورت غیرخطی و بازگشتی می‌شوند. آزمون دهندگان در ابتدا به گروه‌هایی با رشته (گروه آزمایشی) ریاضی و علوم انسانی، تجربی (گروه ۱) تقسیم شدند. بعد، داوطلبان گروه آزمایشی ریاضی به داوطلبان با گروه سنی ۱۸ و ۱۹ سال تقسیم شدند (گروه ۲). به همین ترتیب گروه‌های ۵ و ۷ به ترتیب برای گروه آزمایشی تجربی و انسانی (دارای ۱۸ و ۱۹ سال) ایجاد گردید. این بدین معنی است که پارامترهای دشواری سؤال برای داوطلبان آزمون ورودی در هر گروه بسته به متغیر متفاوت است. یافته‌های پژوهش نشان داد که سؤالات چه به لحاظ سن و چه به لحاظ گروه آزمایشی مورد مطالعه دارای DIF بوده که تقریباً ۵۴٪ از سؤالات تحت تأثیر سن افراد قرار گرفته است؛ بنابراین از یافته‌های به دست آمده در این مطالعه این نتیجه گیری حاصل می‌شود که متغیر سن و گروه آزمایشی نقش اساسی در پاسخ‌گویی افراد به سؤالات آزمون ایفا میکند و می‌تواند باعث تفاوت پاسخ‌گویی آزمون دهندگان آزمون‌های سرنوشت‌ساز و بزرگ شود. سوگیری سؤالات به سن و گروه آزمایشی جزو عوامل شناخته شده در این تحقیق به شمار می‌رود که باعث می‌شود آزمون فاقد عدالت باشد. چنین عواملی نامربوطی باید شناسایی و از سؤالات آزمون حذف شوند. با توجه به اینکه سؤالات دروس اطلاعات عمومی، زبان انگلیسی، زبان عربی، زبان و ادبیات فارسی در تحقیق به عنوان دروس دارای سؤالات سودار شناخته شدند لذا لازم است طراحان سؤال نسبت به طراحی سؤالاتی مجزا برای گروه‌های علوم انسانی و تجربی اقدام نمایند. مطالعه حاضر نقش متغیر آشکار سن را پیدا نمودن DIF نشان داد. در خصوص مقایسه نتایج حاصل از تحقیق حاضر با دیگر پژوهش‌های انجام شده نیز بایستی عنوان نمود که با توجه به نبود پژوهش‌های مشابه هم به لحاظ موضوعی و هم به لحاظ روش تحقیق، مقایسه با تحقیقات پیشین میسر نبوده لیکن با بررسی به عمل آمده؛ برخلاف مطالعه آریا دوست (۲۰۱۸) و گرامی پور؛ رضانی صدر؛ رضایی؛ نوروزی و مختاریان (۱۳۹۶) که نشان دادند که متغیر جنسیت بخش کوچکی از نمونه را تحت تأثیر قرار می‌دهد اما تحقیق حاضر

بررسی نمود که بخش قابل توجهی (۵۴٪) عملکرد کنش افتراقی سؤال از طریق درخت‌های راش قابل نمایش است. همچنین بوکسل و همکاران (۲۰۱۸) که توسط مدل راش آمیخته نشان دادند که برای بعد انزوای اجتماعی در برخی سؤالات DIF از لحاظ جنسیت وجود دارد و در سایر سؤالات نیز به لحاظ سن و مدت درد DIF وجود ندارد ولی تحقیق فعلی نقش متغیر سن را در تبیین سؤالات سودار توسط مدل درخت‌های راش به خوبی نشان داد و در پایان لازم به ذکر است که در تحقیق گرامی پور (۲۰۲۰) که با شبیه‌سازی نشان داد که متغیر سؤالات آزمون زبان انگلیسی (آزمون درک مطلب) دارای سوگیری به لحاظ جنسیت و سوابق تحصیلی هست که در تحقیق حاضر داده‌های واقعی آزمون مورد مطالعه قرار گرفت که علاوه بر شناسایی سؤالات دارای کنش افتراقی توانست نسبت به رسته‌بندی و ارتباط سؤالات نسبت به هم نیز بررسی و نتایج قابل استنادی را ارائه نماید.

پیشنادهای تحقیق

مطالعه حاضر به بررسی سوگیری سؤالات آزمون استخدامی نابجا با استفاده از مدل درخت‌های راش پرداخت و نشان داده شد که مدل درخت‌های راش به خوبی می‌تواند سوگیری سؤالات این آزمون‌ها را مشخص نماید. لذا برای پژوهش‌های آینده پیشنهاد می‌شود که بدون در نظر گرفتن اطلاعات آماری مربوط به کنش افتراقی سؤال مرتبط به سن و گروه آزمایشی، پرسش‌های آزمون صرفاً به لحاظ محتوایی و به صورت کیفی توسط گروه خبرگان بررسی و نتایج حاصل با یافته‌های این پژوهش، مقایسه شوند. انگل‌هارد، هانش و روتلیج^۱ (۱۹۹۰) چنین رویکردی را برای مطالعات کنش افتراقی سؤال پیشنهاد کرده است. همچنین یافته‌های پژوهش نشان داد که سن افراد دارای ۱۸ سال و رشته تحصیلی ریاضی فیزیک سهم قابل توجهی در قبولی آزمون دارد؛ بنابراین اختصاص دادن سهمیه بیشتری به این گروه از داوطلبان می‌تواند در به کارگیری دانشجویان با استعداد و باهوش بالاتر مفید باشد. در پایان پیشنهاد می‌شود آزمون‌های مرکز گزینش و استخدام نجا نیز توسط مدل راش لجستیک، مدل راش با معادلات ساختاری برای آشکار نمودن DIF و استفاده روش شبیه‌سازی داده‌ها با حجم نمونه‌های بیشتر، مورد استفاده قرار گیرند. همچنین از این مدل می‌توان برای ارزیابی آزمون‌های پایان‌ترم دانشجویان دانشگاه علوم انتظامی نیز استفاده نمود. لذا این مهم می‌تواند در تحقیقات آینده مورد استفاده قرار گیرد.

منابع

- گرامی پور، مسعود؛ رضایی صدر، اعظم؛ رضایی، احمد؛ نوروزی، لیلا و مختاریان، فرانک (۱۳۹۶). کنش افتراقی سؤال (DIF) و آزمون (DIF) مرتبط با جنسیت در آزمون‌های کنکور سراسری سازمان سنجش آموزش کشور، **فصلنامه مطالعات اندازه‌گیری و ارزشیابی آموزشی**، سال هفتم، شماره ۱۹، ۶۳-۳۵.

- Benjamini, Y. Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, **Journal of the Royal Statistical Society, Series B (Methodological)**, Vol.57, No. 1, 289-300.

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

- Liou M (1994). "More on the Computation of Higher-Order Derivatives on the Elementary Symmetric Functions in the Rasch Model." *Applied Psychological Measurement*, 18 (1), 53-62.

- Merkle EC, Zeileis A (2013). "Tests of Measurement Invariance without Subgroups: A Generalization of Classical Methods." *Psychometrika*, 78 (1), 59-82.

- Popham, W. J. (2005). High-Stakes Tests: Harmful, Permanent, Fixable, **American Educational Research Journal**. 6, P85.

- Zeileis A, Hornik K (2007). "Generalized M-Fluctuation Tests for Parameter Instability." *Statistica Neerlandica*, 61 (4), 488-508.

- Zeileis A, Hothorn T, Hornik K (2008). "Model-Based Recursive Partitioning." **Journal of Computational and Graphical Statistics**, 17 (2), 492-514.