

Differential Item Functioning (DIF) Detection Rate Using Rasch Trees Model: A Simulated and Real Data Study of the NAJA High Stakes Tests.

Hamid Reza Karami

Masoud Gramipour*

Asghar Minaei

PhD student of Measurement, Islamic Azad University, Saveh, Iran
Associate Professor of Curriculum Planning, Kharazmi University, Tehran, Iran.
(Corresponding Author)
Associate Professor of Measurement, Allameh Tabatabai University, Tehran, Iran

Abstract

The purpose of this study was to use the rasch tree model and to investigate the intervening factors in revealing the differential action of the test questions. To answer the research questions, the method of simulation studies and data of the specific test of Amin University of Law Enforcement Sciences were used. The present research is applied in terms of purpose and descriptive in terms of method called psychometric research. The statistical population of the special exam of Amin University of Law Enforcement Sciences in 1398 with the number of 2414 people has been analyzed in full. The DIFtree package in R software was used to determine the detection rate of the differentiation Item Functioning using the rasch tree model. The results of the simulation study showed that the rasch tree model detects the differentiation Item Functioning in samples with a volume of more than 1000 subjects as 100%. The results also showed that in the specific exam of the University of Law Enforcement Sciences, 9 questions have differentiation Item Functioning, in the most important of which, 7 questions in the mathematics group with 18 years of age (second category) and 6 questions in the mathematics group with 19 years of age (category First) has a bias towards the experimental sciences group (third category) and the orientation of the bias has been in favor of the first and second category and to the detriment of the third category.

Keywords: Beech tree model, Differential action of the question, Bias, simulation.



اندازه گیری تربیتی


دوره ۱۱، شماره ۴۴، تابستان، ۱۴۰۰، ۳۰-۱

atu.ac.ir


DOI:

نرخ آشکارسازی کنش افتراقی سؤال با استفاده از مدل درخت‌های راش یک مطالعه شبیه‌سازی شده و داده‌های واقعی آزمون‌های سرنوشت‌ساز ناجا^۱


دانشجوی دکتری سنجش و اندازه‌گیری، دانشگاه آزاد اسلامی ساوه، ساوه، ایران

حمیدرضا کرمی 

دانشیار برنامه‌ریزی درسی، دانشگاه خوارزمی، تهران، ایران. (نویسنده مسئول)

مسعود گرامی‌پور* 

دانشیار سنجش و اندازه‌گیری تربیتی دانشگاه علامه طباطبائی، تهران، ایران

اصغر مینایی 

چکیده

هدف از این تحقیق استفاده از مدل درخت‌های راش و بررسی عوامل مداخله‌گر در آشکارسازی کنش افتراقی سؤال‌های آزمون بود. برای پاسخگویی به سؤال‌های تحقیق از روش مطالعات شبیه‌سازی و داده‌های آزمون اختصاصی دانشگاه علوم انتظامی امین استفاده شد. پژوهش حاضر از نظر هدف کاربردی و از نظر روش، توصیفی موسوم به تحقیقات روان‌سنجی است. جامعه آماری آزمون اختصاصی سال ۱۳۹۸ دانشگاه علوم انتظامی امین 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 مطالعه شبیه‌سازی نشان داد که مدل درخت‌های راش، کنش افتراقی سؤال را در نمونه‌های با حجم بالای ۱۰۰۰ آزمودنی به صورت صد درصد شناسایی می‌کند. همچنین نتایج نشان داد که در آزمون اختصاصی دانشگاه علوم انتظامی 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 آن به نفع رسته اول و دوم و به ضرر رسته سوم بوده است.

کلیدواژه‌ها: مدل درخت‌های راش، کنش افتراقی سؤال، سوگیری، شبیه‌سازی.

۱. مقاله حاضر برگرفته از رساله دکتری رشته سنجش و اندازه‌گیری دانشگاه آزاد اسلامی واحد ساوه است.

* نویسنده مسئول: mgramipour@khu.ac.ir

مقدمه

در جمهوری اسلامی ایران، جوانانی که علاقه‌مند به تحصیل و راه‌یابی به دانشگاه علوم انتظامی هستند از طریق شرکت در آزمون سراسری با انتخاب دانشگاه و رشته موردنظر و کسب نمره رتبه قبولی موردنیاز پذیرفته‌شده و مشغول تحصیل می‌شوند. آزمون‌های بزرگ‌مقیاسی مانند آزمون‌های اختصاصی نیروی انتظامی جمهوری اسلامی ایران که دانشگاه علوم انتظامی امین 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 به‌صورت سالیانه توسط دانشگاه علوم انتظامی امین به‌صورت سراسری و متمرکز 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، سوامیناتان و راجرز، ۱۹۹۱). آزمون‌ها و حرفه آزمونگری موردتوجه 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 جهت تصمیم‌گیری برای آینده آزمودنی‌ها به کار می‌روند از اهمیت بیشتری برخوردار است (پاپهام، ۲۰۰۵).

یکی از مهم‌ترین تهدیدها برای اعتبار آزمون، سوگیری سؤال است که از آن تحت عنوان کنش افتراقی سؤال 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

۱ Hambleton, Swaminathan and Rogers

۲ Popham

۳ Differential item functioning (DIF)

آزمون‌های روان‌سنجی معمولاً فرض می‌شود که ویژگی‌های اندازه‌گیری برای تمامی افراد، پایدار است. این ویژگی تحت عنوان تغییرناپذیری اندازه‌گیری^۱ شناخته شده است (میلسپ، ۲۰۱۲). زمبو^۲ (۲۰۰۷) استدلال کرد که تجزیه و تحلیل کنش افتراقی ممکن است لزوماً نشان‌دهنده 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، به طور 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 آزمون‌هایشان قابل اتکا و اعتماد بوده و فاقد سوگیری باشد (کامیلی^۳، 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 به مرور 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 DIF را می‌توان با استفاده از روش‌های مبتنی بر نظریه کلاسیک^۴ (CTT) و روش‌های مبتنی بر نظریه پاسخ سؤال^۵ (IRT) طبقه‌بندی نمود. ایس و راجو^۶ (۲۰۰۳) اظهار داشتند که روش‌های منتل-هنسل^۷ (MH)، منتل هنزل تعمیم‌یافته^۸ (GMH)؛ رگرسیون لجستیک^۹ (LR) و نمودار دلتا^{۱۰} (SIBTEST) تحت روش‌های CTT قرار دارند، در حالی که روش‌های مجذور کای-دو لرد^{۱۱} راجو و

۱ measurement invariance

۲ Zumbo

۳ Camilli

۴ classical test theory

۵ item response theory

۶ Ellis and Raju

۷ Mantel-Haenszel

۸ generalized mantel-haenszel

۹ Logistic Regression

۱۰ delta plot

۱۱ Lord's chi-square

روش‌های آزمون نسبت درستنمایی^۱ تحت مدل‌های IRT انجام شده است. به‌طور کلی روش‌های موجود از چهار رویکرد، نمره مورد انتظار سؤال، روش‌های IRT، روش‌های ناپارامتریک و مدل خطی تعمیم‌یافته گسترش یافته‌اند. باین وجود هر یک از رویکردها مفروضات و خلاهایی دارند. به لحاظ تاریخی از اولین و مهم‌ترین روش‌های نمره مورد انتظار سؤال که در طول تاریخ مطالعه کنش افتراقی سؤال استفاده شده‌اند، روش‌های مبتنی بر تحلیل واریانس است. روش تحلیل واریانس هرچند روشی ساده در تحلیل کنش افتراقی سؤال است، اما به دلیل اینکه تنها با استفاده از میانگین نمرات، نسبت به تشخیص کنش افتراقی اقدام می‌شد به سرعت منسوخ و روش‌های جامع‌تر و بهتری جایگزین آن شد. روش نمودار دلتا توسط آنگوف معرفی شد. در این روش شاخص‌های دشواری سؤال به انحراف نرمال تبدیل می‌شود تا بتوان به‌وسیله آن‌ها به رسم نمودارهای دلتا پرداخت. روش دلتا هرچند ساده است اما به دلیل ناپایایی، محدودیت‌ها و عدم اطمینان نتایج دچار خدشه بود و کمتر مورد استفاده قرار گرفت (ماپرانگا، دورانز و میدلتون،^۲ ۲۰۰۸:۱۷). علی‌رغم سهولت استفاده، این روش‌ها به علت وجود مشکلات اشاره شده مورد تردید است و برای رفع این مشکلات روان‌سنج‌ها به رویکردهای IRT روی آوردند. استفاده از رویکردهای مبتنی بر IRT در تشخیص کنش افتراقی سؤال از اوایل دهه ۱۹۶۰ با مدل راش آغاز شد. اما مدل راش تنها متکی بر پارامتر دشواری بوده و در تشخیص کنش افتراقی سؤال کارایی لازم را ندارد، لذا منجر به قضاوت‌های نادرست می‌شود. سایر مدل‌های IRT هم خالی از اشکال نیست. برآش مناسب مدل IRT نیاز به حجم نمونه بالا دارد که هزینه‌بر و مقرون‌به‌صرفه نیست. ضرورت وجود مدل‌های تک‌بعدی و یکسانی منحنی سؤال از دیگر محدودیت‌های این نظریه است. در این زمینه مطالعاتی که توسط کورویل^۳ (۲۰۰۴) انجام شده است نشان می‌دهد کاربرد IRT در زمینه کنش افتراقی سؤال مزیت‌چندانی نسبت به نظریه کلاسیک اندازه‌گیری ندارد.

همه روش‌های تشخیص کنش افتراقی سؤال که مرور شد، می‌توانند تنها با یک متغیر گروهی دسته‌ای 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

^۱ likelihood ratio test

^۲ Mapuranga, Dorans, & Middleton

^۳ Courville

متغیرهای پی در پی این اجازه داده می شود که به عنوان یک منبع کنش افتراقی سؤال مورد تحلیل قرار بگیرند. البته درجایی که هیچ زیرگروه از پیش تعیین شده ای نیاز نباشد. اینها همان روش های درخت های سؤال متمرکز^۱ (IFT) هستند که ممکن است در تحلیل مدل راش یا رگرسیون لجستیک به کار برده شوند (تاتز و برگر، ۲۰۱۶: ۷۳۰)؛ بنابراین با توجه به مطالب فوق بیان گردید که روش های مختلفی جدیدی برای تشخیص DIF پدید آمده است. از جمله می توان به روش رگرسیون لجستیک تربیتی هیبریدی^۲ و درخت های راش اشاره نمود. در حالی که روش رگرسیون لجستیک تربیتی هیبریدی (IHOLR) بر اساس خواص IRT عمل می کند. لیکن از روش درخت های راش (RT) که در این تحقیق از آن استفاده می شود، به دلیل اینکه چندین متغیر با هم در نظر گرفته می شوند برای تعیین DIF استفاده می شود. مزیت دیگر RT این است که در مقایسه با اکثر روش های تعیین DIF نیازی به مشخص کردن گروه های تمرکز و مرجع به عنوان پیش نیاز نیست. لیکن از مدل درخت های راش^۳ که در این تحقیق از آن استفاده می شود، به دلیل اینکه چندین متغیر با هم در نظر گرفته می شوند برای تعیین کنش افتراقی استفاده می شود. مزیت دیگر درخت های راش این است که در مقایسه با اکثر روش های تعیین کنش افتراقی نیازی به مشخص کردن گروه های کانونی و مرجع به عنوان پیش نیاز نیست. در روش درخت های راش هنگام تعیین گروه ها، پارامترهای سؤال را در کلیه متغیرهای کمکی سؤال مورد بررسی قرار می دهند. گروه ها طبق متغیرهای کمکی که دارای قوی ترین بی ثباتی (عدم تناسب پارامترهای سؤال در گروه ها) باشند؛ مشخص می شوند. به عنوان مثال، اگر بخواهیم مشخص کنیم که آیا کنش افتراقی از نظر جنسیت و انگیزه ذاتی وجود دارد؟ می توان آن را با توجه به جنسیت تفکیک و سپس بر اساس نمره انگیزه ذاتی تمایز داد. در حالی که متغیر کمکی که قوی ترین بی ثباتی را به خود اختصاص می دهد جنسیت است و نمره انگیزه ذاتی دومین متغیر است.

علاوه بر این؛ روش درخت های راش نسبت به سایر روش ها در تعیین امتیاز (نمره) برش به عنوان یک ویژگی برتر، برخوردار است. در روش هایی که از گروه های از پیش تعریف شده با متغیرهای پیوسته استفاده می کنند، میانگین حسابی یا میان به عنوان نقاط برش

^۱Item-focused trees

^۲Hybrid Ordinal Logistic Regression

^۳Rasch trees (RT)

مورد استفاده قرار می‌گیرد. در روش درخت‌های راش درحالی‌که 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 میانه، میانگین نمرات انگیزه ذاتی افراد که اختلاف پارامتر 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، کوپف و زالایس، ۲۰۱۵).

در این تحقیق برای تشخیص کنش افتراقی سؤال از مدل درخت‌های راش با استفاده از یک مطالعه شبیه‌سازی شده و داده‌های واقعی اقدام خواهد شد. از این رو، در پژوهش حاضر سعی بر آن است که به سؤالات زیر پاسخ داده شود؟

- نحوه مداخله عواملی چون حجم نمونه؛ شدت کنش افتراقی و درصد سؤال‌های دارای کنش افتراقی در آشکارسازی آن با استفاده از مطالعه شبیه‌سازی چگونه است؟
- کاربرد مدل درخت‌های راش در آشکارسازی کنش افتراقی سؤال آزمون‌های سرنوشت‌ساز ناجا چیست؟

پیشینه پژوهش

گرامی پور^۲ (۲۰۲۰) در مقاله‌ای با عنوان «رویکرد درخت‌های سؤال 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 به‌صورت 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 کنش افتراقی یکنواخت که در آن تعداد ۲ تقسیم (۲ متغیر پیش‌بین مشترک)

^۱ Strobl, Kopf and Zeileis

^۲ gramipour

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. آریا دوست^۱ (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 برگشت‌پذیر 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^۲ و همکاران (۲۰۱۹) در مقاله‌ای با عنوان «مقایسه روش‌های مختلف تشخیص کنش افتراقی سؤال در مدل‌های تشخیصی شناختی» با استفاده از یک مطالعه مونت کارلو به بررسی عملکرد روش‌های مثل - هنسزل، رگرسون لجستیک و آزمون‌های والد مبتنی بر اطلاعات سؤال، اطلاعات مشاهده‌شده و ماتریس کوواریانس در خصوص تشخیص کنش افتراقی سؤال پرداختند. یوکسل^۳ و همکاران (۲۰۱۹) در مقاله‌ای با عنوان «کنش افتراقی سؤال شاخص‌های بهداشتی ناتینگهام 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 مدت‌زمان درد قابل توجه نبودند. در سایر سؤالات نیز به لحاظ جنسیت، سن و مدت درد کنش افتراقی وجود نداشت.

مرور تحقیقات گذشته نشان می‌دهد که هنوز در نیروی انتظامی در رابطه با کنش افتراقی آزمون‌های استاندارد، پژوهشی انجام نشده است. همچنین در تحقیقات گذشته داخلی، کنش افتراقی سؤال‌های آزمون تنها برحسب یک ویژگی و آن‌هم بیشتر برحسب جنسیت مورد مطالعه قرار گرفته است. اما در پژوهش حاضر اطلاعات در دسترس داوطلبان

^۱Aryadoust

^۲Liu

^۳Yüksel

آزمون‌های استخدامی نیروی انتظامی برحسب برخی از ویژگی‌های جمعیت شناختی (سن داوطلبان، گروه‌های آزمایشی) مورد استفاده قرار گرفته است. لذا جای خالی پژوهشی در رابطه با ویژگی‌های چندگانه مؤثر بر کنش افتراقی در آزمون‌های استخدامی ضروری به نظر می‌رسد. گفتنی است محقق درخواست سایر ویژگی‌هایی مانند استان متقاضی، معدل و غیره را ارائه که به دلیل رعایت برخی ملاحظات از دادن آمار فوق توسط مرکز گزینش و استخدام ناجا خودداری گردید.

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 فرایندها زیر برای رسیدن به یک ساختار درخت راش به‌صورت 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 برآزش بهبود دهید؛

۴) مراحل ۱ تا ۳ را به‌صورت 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 می شود. برای برآورد پارامترهای سؤال مدل درخت های راش، می توان از روش ماکزیمم درستنمایی شرطی به دلیل اینکه با روش ماکزیمم درستنمایی سازگاری دارد، استفاده نمود. فرض کنید $n = 1, \dots, n$ پارامترهای افراد و $m = 1, \dots, m$ پارامترهای مربوط به سؤال و u_{ij} مربوط به پاسخ موضوع i به سؤال j 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 می شود. پارامترهای سؤال نیز را می توان 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 γ_{r_i} تابع متوازن برای هر ترتیبی از r_i (فیشر و مولانرا، ۱۹۹۵) است. برای مشخص کردن مقیاس مشخص باید محدودیت هایی را اعمال کنیم. به عنوان مثال برای پارامترهای سؤال، مقدار صفر و برای سایر جاها، مقدار $m - 1$ لحاظ می شود. به منظور بررسی اینکه آیا پارامترهای سؤال بین گروه های افراد توسط متغیرهای کمکی تبیین می شوند؟ ما از روش آزمون تغییرات ساختاری بکار رفته در اقتصادسنجی، استفاده می کنیم. این آزمون ها معمولاً برای شناسایی، به عنوان مثال، افت بازده سهام با گذشت زمان، به کار می رود در حالی که در اینجا ما از همین روش استفاده می کنیم تا تغییرات پارامتر را نسبت به متغیر کمکی افراد را بررسی کنیم.

عملکرد نمره فردی $i = 1, \dots, n$ را $\psi(u_i, \hat{\beta})$ برای n مشاهده را در نظر بگیرید. مشتق این عملکرد فردی با لگاریتم درستنمایی $\Psi(u_i, \hat{\beta})$ است که با توجه به بردار پارامتر، یک اندازه گیری کلی از انحراف برای مدل های درستنمایی- پایه است. برای مدل راش این سهم مشارکت فردی را می توان به راحتی از طریق درستنمایی شرطی برآورد نمود. فرایند جمع تجمعی به صورت زیر تعریف می شود:

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

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

به همین ترتیب برای جمع تجمعی بیش از ℓ امین متغیر کمکی نیز داریم: با $n = 1, \dots, n$ انحراف فردی با توجه به ترتیب متغیر کمکی در هر مرحله تا عنصر t جمع می‌شوند. هنگامی که $W_{\ell}(t)$ 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 $W_{\ell}(t)$ برای ارزیابی متغیرهای کمکی نیاز به تعدیل دارد. برای گرفتن انحراف هماهنگ $W_{\ell}(\cdot)$ آماره آزمون مختلف، بسته به اینکه آیا ℓ امین متغیر، متغیر عددی به طبقه‌ای است، می‌تواند مورد استفاده قرار گیرد. اگر عددی باشد، گفته می‌شود که آمار آزمون زیر نرمال است:

$$S_{\ell} = \max_{i=i_1, \dots, i_{\ell}} \left(\frac{i}{n} \cdot \frac{n-i}{n} \right)^{-1} \left\| w_{\ell} \left(\frac{i}{n} \right) \right\|_{\Psi}^2 \quad (3)$$

این عبارت را می‌توان ماکزیمم لاگرانژ به‌عنوان یک گزینه جایگزین تکی برای همه نقاط برش $[\underline{i}, \bar{i}]$ تعبیر کرد. توزیع محدودکننده برتری یک فرایند گره پایین باسل^۱ است که از آن می‌توان مقادیر p را محاسبه کرد (زایلیس و همکاران، ۲۰۰۸ و مرکل^۲ و

^۱Lagrange

^۲Bessel

^۳Merkle

زالایس، ۲۰۱۳). از طرف دیگر، اگر ℓ امین متغیر کمکی رسته‌ای باشد (با مقدار $x_{i\ell}$ از $q = 1, \dots, Q$ متغیر رسته‌ای)؛ استفاده از آماره آزمون زیر نرمال خواهد بود:

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

که در آن Δ_q ارتقای به q امین متغیر رسته‌ای است. این آماره آزمون، برای مرتب‌سازی مجدد در مقوله‌های (رسته‌های) Q و افراد در هر گروه (رسته) تغییرناپذیر است. این آماره آزمون عدم ثبات را برای هر Q محاسبه می‌کند. این تابع توزیع دارای توزیع کای دو (χ^2) با $(m-1)(Q-1)$ درجه آزادی برای هر p 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 $L_C(\beta|r_i)$ ساخت

$$\Psi(u_i, \beta) = - \sum_{j=1}^m u_{ij} \cdot \beta_j - \log(\gamma_{r_i}(\beta)) \quad (5)$$

برای محاسبه آزمون‌های تغییرات ساختاری، سهم‌های فردی در عملکرد نمره از معادله شماره ۵ حاصل می‌شود. سهم ℓ امین موضوع برای ℓ امین پارامتر سؤال عبارت است از:

$$\Psi(u_i, \beta)_j = \frac{\delta \Psi(u_i, \beta)}{\delta \beta_j} = -u_{ij} - \frac{1}{\gamma_{r_i}(\beta)} \cdot \frac{\delta \gamma_{r_i}(\beta)}{\delta \beta_j} \quad (6)$$

گفتنی است مشتقات دوباره عبارت $\gamma_{r_i}(\beta)$ 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 تعیین‌شده توسط مدل و مشخص نمودن سؤالات سودار از روش بنجامین-هوچبرگ^۲ (۱۹۹۵) استفاده شده است. این روش بر اساس آزمون نرخ کشف

^۱Liou

^۲Benjamini and Hochberg

کاذب^۱ 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 معنی‌دار است، فاصله اطمینان برای P -مقدار با سطح معنی‌داری 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 $1 - (1 - \alpha)^k$ 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 H_1, H_2, \dots, H_m بر پایه P -مقدارهای P_1, P_2, \dots, P_m و q^* $= 0,05$ باشند. همچنین فرض کنید $P_{(1)} \leq P_{(2)} \leq \dots \leq P_{(m)}$ مقادیر مرتب شده P -مقدار و $H_{(i)}$ فرض صفر بر اساس مقدار $P_{(i)}$ باشد. آزمون فرض چندگانه به صورت زیر تعریف می‌شود:

$$P_{(i)} \leq \frac{i}{m} q^*$$

به عبارتی آزمون نرخ کشف کاذب، امید ریاضی نسبت رد به اشتباه فرض H را کنترل می‌کند. در این روش حساسیت به رد کردن فرض H کمتر از سایر روش‌ها مانند بونفرونی^۲ است و با مقایسه p -مقدارها؛ زمانی فرض H رد می‌شود که حداقل یک جفت رسته پیدا شود که تفاوت معنی‌داری باهم داشته باشند. (بنجامین و هوچبرگ، ۱۹۹۵).

روش

پژوهش حاضر از نظر هدف کاربردی و از نظر روش، توصیفی موسوم به تحقیقات روان‌سنجی است. به منظور تعیین عملکرد مدل درخت‌های راش و برای برآورد پارامترهای مدل با رهیافت بسامدی در مطالعه‌ای شبیه‌سازی مورد ارزیابی قرار می‌گیرد. چنین روشی به محقق اجازه می‌دهد که پارامترهای برآورد شده آزمودنی و سؤال را با مقادیر واقعی آن که در داده‌های واقعی غیرقابل مشاهده هستند، مقایسه کند. همچنین مطالعات شبیه‌سازی شده به

^۱ False discovery rate (FDR)

^۲ Bonferroni

محقق اجازه می‌دهد که نتایج نظری را در عمل تأیید کند. سرانجام این روش سریع‌تر، ارزان‌تر و آسان‌تر از جمع‌آوری اطلاعات از آزمودنی‌های واقعی است و محقق می‌تواند مدل‌ها و روش‌های جدید روان‌سنجی را به‌سرعت و با هزینه بسیار کمی مورد بررسی قرار دهند (هارول^۱ و همکاران، ۱۹۹۶؛ اسپنس^۲، ۱۹۹۳). در تحقیق حاضر، علاوه بر شبیه‌سازی؛ داده‌های واقعی (آزمون‌های استخدامی ناجا) نیز با استفاده از مدل درخت‌های راش مورد تجزیه و تحلیل قرار گرفته است. در شبیه‌سازی داده‌های مدل پرسش- پاسخ سه مرحله اساسی طی شد:

۱- تعیین شکل مدل پاسخ سؤال- پاسخ‌های سؤال غالباً به‌صورت تک‌بعدی، به لحاظ شرطی مستقل و به‌صورت مدل‌های درخت‌های راش تولید شد. در تحقیق حاضر، مدل درخت‌های راش انتخاب شد.

۲- تعیین پارامترهای مدل پاسخ سؤال- مدل‌های پاسخ سؤال با وجود ویژگی‌های عملیاتی متغیر برای تعیین مدل سؤال‌های آزمون بسیار انعطاف‌پذیر هستند؛ بنابراین ویژگی‌های اندازه‌گیری آزمونی شبیه‌سازی شده تا قبل از اینکه پارامترهای سؤال مشخص نشده است، ثابت نیست. در تحقیق حاضر پارامترها از تحقیقات قبلی که توسط پارشال و میلر^۳ (۱۹۹۵) روی آزمودنی‌های واقعی انجام شده است، انتخاب شدند. آن‌ها ابتدا پارامترهای سؤال را بر اساس مطالعات قبلی اجرای آزمون بر آزمودنی‌های واقعی برآورد کردند، سپس از آن‌ها برای شبیه‌سازی داده‌ها استفاده کردند.

به‌منظور آشکارسازی کنش افتراقی سؤال آزمون ۱۰۰ سؤالی با مدل درخت‌های راش در مقیاس بزرگ یعنی با تعداد زیادی از آزمودنی‌ها شبیه‌سازی شد. معیار شناسایی کنش افتراقی سؤال بر اساس سطح معنی‌داری در رد فرض صفر (با نرخ خطای نوع اول ۰,۰۵) بود. سایر عوامل مداخله‌گر که تأثیر آن بر عملکرد مدل درخت‌های راش در تشخیص کنش افتراقی سؤال مورد بررسی قرار گرفت عبارت‌اند از:

الف) حجم نمونه: چهار نمونه با حجم‌های ۱۵۰، ۵۰۰، ۱۰۰۰ و ۲۴۱۴ (برابر حجم نمونه داده‌های واقعی) آزمودنی که مختص آزمون‌های سرنوشت‌ساز است، شبیه‌سازی شد.

^۱Harwell

^۲Spence

^۳Parshall & Miller

ب) کنش افتراقی سؤال هماهنگ: بسته به این که گروه مرجع نسبت به گروه کانونی در کدام یک از پارامترهای دشواری و تمیز سؤال با یکدیگر متفاوت باشند، روش‌های مختلف تشخیص کنش افتراقی سؤال ممکن است در آشکارسازی آن متفاوت عمل کنند. برای بررسی تأثیر این عوامل به صورت شبیه‌سازی شده مقادیر بسیار سخت (۰/۵-); سخت (۰/۲۵-); متوسط (۰); آسان (۰/۲۵) و بسیار آسان (۰/۵) به‌عنوان شدت کنش افتراقی سؤال به پارامترهای دشواری در گروه مرجع اضافه شد. اگر تفاوت گروه مرجع با گروه کانونی در مقدار پارامتر دشواری سؤال باشد، سؤال دارای کنش افتراقی هماهنگ است.

ج) درصد سؤال‌های دارای کنش افتراقی سؤال: همان‌طور به بیان گردید برای هر کدام از موارد سختی در جدول فوق، به‌طور فرضی برای ۱۵۰، ۵۰۰، ۱۰۰۰ و ۲۴۱۴ نفر، پاسخ‌نامه مربوط به ۱۰۰ سؤال شبیه‌سازی گردید که در هر مورد از بین این ۱۰۰ سؤال، تعداد د ۱۰، ۲۰ و ۳۰ سؤال را دارای کنش افتراقی اختیار نمودیم. در مجموع هر حالت شبیه‌سازی، ۱۰۰ مرتبه تکرار شده است.

به‌منظور شبیه‌سازی پارامترهای دشواری برای یک آزمون ۱۰۰ سؤالی با مدل درخت‌های راش از نرم‌افزار R استفاده شد. شیوه تولید داده‌ها در نرم‌افزار R به این صورت بود که برای یک تکرار ابتدا تعداد آزمودنی‌ها مشخص شد. با توجه به فرضیه‌های تحقیق حاضر، تعداد نمونه در یکی از حجم‌های ۱۵۰، ۵۰۰، ۱۰۰۰ یا ۲۴۱۴ آزمودنی می‌توانست تعیین شود.

داده‌ها به دو بخش: افرادی که به‌درستی به یک سؤال پاسخ داده‌اند و افرادی که به سوال پاسخ درست نداده یا اصلاً پاسخ نداده‌اند؛ تقسیم شده‌اند.

پس از انجام محاسبات لازم نمره کل آزمون برای تمامی افراد برابر با ۰/۲۷۷۳۶۹۵ به دست آمد؛ بنابراین

$$\ln\left(\frac{۲۷۷۳۶۹۵}{۱ - ۰.۲۷۷۳۶۹۵}\right) = -۰.۹۵۷۵۴۷۴$$

محاسبه گردید.

پاسخ‌نامه مربوط به سؤالات بدون سوگیری را از توزیع برنولی با احتمال موفقیت (پاسخ درست) p که از معادله

$$\text{لو جیت} \left(p \left(\text{پاسخ درست} \right) \right) = -0.9575474 + 0.179 (\text{دشواری سؤال})$$

به دست می آید، استخراج نمودیم. در این مدل «دشواری سؤال» بیانگر سختی (شدت) کل سؤالات است. برای شبیه سازی پاسخ نامه سؤالات دارای سوگیری بر مبنای مشاهدات به عمل آمده، برای بر روی عامل سن متمرکز شدیم. برای این منظور تعداد ۵۳۸ نفر در گروه سنی ۰ (۱۸ سال) و تعداد ۱۸۷۶ نفر در گروه سنی ۱ (۱۹ سال) قرار گرفتند. بر این اساس اقدام به گروه بندی نمودیم که در نتیجه مدل لو جیت زیر استخراج گردید:

$$\begin{aligned} \text{لو جیت} \left(p \left(\text{پاسخ درست} \right) \right) \\ = -0.9575474 + 0.179 (\text{دشواری سؤال}) + 0.45 (\text{گروه}) \end{aligned}$$

در اینجا «گروه» برای گروه های سنی به کار رفته است. برای متغیر سختی، مقادیر (۰/۵، ۰/۲۵، ۰، -۰/۲۵، -۰/۵) را در نظر می گیریم. با جایگذاری هر کدام از این مقادیر احتمال پاسخ صحیح به سؤالات مشخص می شود.

برای تعیین کنش افتراقی سؤال مدل لو جیت، با توجه به عوامل مداخله گر مختلف، چهار حجم متفاوت، یک نوع کنش افتراقی سؤال هماهنگ، پنج مقدار با شدت متفاوت کنش افتراقی و سه سطح درصد سؤالات دارای کنش افتراقی، ۶۰ شرایط مختلف آزمایشی با ۱۰۰ تکرار شبیه سازی شد؛ بنابراین در تحقیق حاضر ۶۰۰۰ مجموعه داده پاسخ سؤال شبیه سازی شده است.

در قسمت تحلیل داده های واقعی؛ آزمون های اختصاصی دانشگاه علوم انتظامی امین را مورد تحلیل به منظور تشخیص کنش افتراقی سؤال قرار داده ایم. جامعه آماری آزمون ورودی سال ۱۳۹۸ دانشگاه علوم انتظامی امین به تعداد ۲۴۱۴ نفر و به صورت تمام شمار مورد تحلیل قرار گرفته است. داده های مربوط به آزمون از معاونت تربیت و آموزش دانشگاه علوم انتظامی تهیه گردیده است. برای رسیدن به اهداف این مطالعه، شرکت کنندگان به ترتیب بر اساس سن به دو گروه (گروه اول افراد دارای سن ۱۹ سال و گروه افراد دارای سن ۱۸ سال در نظر گرفته شد. گروه اول به تعداد ۱۸۷۶ نفر (۸۰٪) و گروه دوم به تعداد ۵۳۸ نفر (۲۰٪) و به سه گروه آزمایشی (ریاضی، علوم تجربی و علوم انسانی) تقسیم شدند. شرکت کنندگان آزمون ورودی، کسانی بودند که در سراسر کشور

مایل به ورود به دانشگاه علوم انتظامی بوده و در آزمون اختصاصی عمومی شامل ۱۰۰ سؤال از دروس زبان و ادبیات فارسی (۲۱ سؤال)؛ زبان عربی (۱۹ سؤال)؛ فرهنگ و معارف اسلامی (۱۷ سؤال)؛ آزمون هوش (۱۳ سؤال)؛ اطلاعات عمومی (۱۵ سؤال) و زبان انگلیسی (۱۵ سؤال) خود را مورد آزمون قرار داده‌اند. همچنین سه گروه آزمایشی ریاضی و فیزیک به تعداد ۹۵۱ نفر (۳۹٪) و تعداد ۱۶۰ نفر (۷٪) گروه علوم تجربی و تعداد ۱۳۰۳ نفر (۵۴٪) گروه علوم انسانی در تحلیل‌ها لحاظ گردیدند. جهت تشخیص نرخ آشکارسازی کنش افتراقی سؤال با استفاده از مدل درخت‌های راش از بسته DIFtree (موریتز، ۲۰۲۰) در نرم‌افزار R استفاده شد.

یافته‌ها

الف) مطالعه شبیه‌سازی: در این بخش با توجه به موارد گفته‌شده و پاسخ به سؤال تحقیق و بررسی عوامل مداخله‌گر چون حجم نمونه؛ شدت کنش افتراقی و درصد سؤال‌های دارای کنش افتراقی در آشکارسازی آن با استفاده از مطالعه شبیه‌سازی و مدل درخت‌های راش پرداخته می‌شود.

تعداد مواردی (از ۱۰۰ مورد) که توسط درخت راش به‌درستی مشخص شده‌اند در جدول شماره (۱) مشخص شده است.

جدول ۱. درصد آشکارسازی کنش افتراقی سؤال (از هر صد تکرار) در شرایط مختلف شبیه‌سازی شده

	تعداد سؤالات دارای سوگیری	سختی سؤالات		
		حجم نمونه	۱۰ سؤال	۲۰ سؤال
۱۵۰	بسیار سخت	۰	۰	۰
	سخت	۰	۰	۰
	متوسط	۰	۰	۰
	آسان	۰	۰	۰
	بسیار آسان	۰	۰	۰
۵۰۰	بسیار سخت	۵	۳۲	۴۷
	سخت	۱۱	۴۰	۶۸
	متوسط	۱۷	۵۵	۸۶

	آسان	۲۳	۷۰	۹۰
	بسیار آسان	۱۷	۶۱	۸۷
	بسیار سخت	۶۲	۹۳	۱۰۰
	سخت	۷۴	۹۹	۱۰۰
۱۰۰۰	متوسط	۹۱	۱۰۰	۱۰۰
	آسان	۹۲	۱۰۰	۱۰۰
	بسیار آسان	۹۱	۱۰۰	۱۰۰
	بسیار سخت	۱۰۰	۱۰۰	۱۰۰
	سخت	۱۰۰	۱۰۰	۱۰۰
۲۴۱۴	متوسط	۱۰۰	۱۰۰	۱۰۰
	آسان	۱۰۰	۱۰۰	۱۰۰
	بسیار آسان	۱۰۰	۱۰۰	۱۰۰

همان‌طور که در جدول شماره (۱) ملاحظه می‌شود، در همه شرایط آزمایشی، حجم نمونه به صورت نمونه به صورت خطی باعث افزایش نرخ آشکارسازی کنش افتراقی سؤال می‌شود. در حجم‌های نمونه مختلف نتایج به شرح زیر است:

الف) حجم نمونه دو هزار و چهارصد و چهارده آزمودنی: با این حجم نمونه، مدل درخت‌های راش در مقادیر DIF با تمامی سختی سؤالات (بسیار سخت؛ سخت؛ متوسط؛ آسان و بسیار آسان) می‌تواند صد درصد کنش افتراقی هماهنگ را مشخص کند.

ب) حجم نمونه هزار آزمودنی: با این حجم نمونه، هراندازه که تعداد سؤالات دارای DIF در مجموع ۱۰۰ سؤال بیشتر باشد، شانس تشخیص درست سؤالات دارای DIF نیز افزایش می‌یابد. همچنین درصد سؤالات دارای سوگیری تأثیر قابل ملاحظه‌ای در تشخیص کنش افتراقی دارد به‌عنوان مثال در درصد‌های ۲۰ و ۳۰ سؤالات دارای سوگیری و در سختی سؤالات متوسط و آسان و بسیار آسان؛ درخت‌های راش می‌تواند ۱۰۰ درصد کنش افتراقی را تشخیص دهد اما در سؤالات سخت و بسیار سخت این میزان کاهش یافته است. همچنین در ۱۰ درصد سؤالات دارای سوگیری و در تمامی حالات سختی سؤالات، میزان تشخیص کنش افتراقی به‌طور محسوسی کاهش یافته است.

ج) حجم نمونه ۵۰۰ آزمودنی: در این حجم نمونه، درخت‌های راش در تمامی درصد سؤالات دارای سوگیری و در تمامی حالات سختی سؤالات به‌خوبی عمل‌نکرده و این

روند در سؤالات بسیار سخت و در ۱۰ درصد سؤالات بسیار ضعیف در تشخیص کنش افتراقی عمل نموده است.

د) حجم نمونه ۱۵۰ آزمودنی: در این حجم نمونه، درخت‌های راش در تمامی حالات درصد سؤالات دارای سوگیری و سختی سؤالات نمی‌تواند کنش افتراقی سؤال را تشخیص دهد؛ بنابراین با توجه به نتایج می‌توان گفت که با کنترل عوامل مداخله‌گر به‌غیراز درصد سؤالات کنش افتراقی، درخت‌های راش در حجم نمونه‌های بالای ۱۰۰۰ آزمودنی بسیار خوب عمل کرده به‌نحوی که در داده‌های واقعی (تعداد ۲۴۱۴ آزمودنی) به‌صورت صد درصد در تشخیص کنش افتراقی عمل نموده است.

ب) تحلیل کنش افتراقی سؤال (DIF) با استفاده از مدل درخت‌های راش: در این بخش به‌منظور پاسخ به سؤال دوم تحقیق و بررسی کاربرد مدل درخت‌های راش در آشکارسازی کنش افتراقی سؤال آزمون‌های اختصاصی دانشگاه علوم انتظامی امین که در سال ۱۳۹۸ انجام‌شده است را موردبررسی و تحلیل قرار می‌دهیم. اطلاعاتی که از پاسخنامه‌ی تعداد ۲۴۱۴ نفر از آزمون دهندگان آزمون ورودی دانشگاه علوم انتظامی امین به دست آمد، به‌صورت زیر است:

۱- نمره میانگین این آزمون ۱۰۰ سؤالی، $27/73$ بود و انحراف معیار آن $8/66$ است.
 ۲- 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 گرفته‌شده بود.

۳- انحراف معیار محاسبه‌شده برای آزمون دهندگان گروه‌های سنی ۱۸ و ۱۹ سال به ترتیب $8/49$ و $8/70$ است که تفاوت اندک میانگین $(8/70 - 8/49 = 0/21)$ نشان‌دهنده تغییرپذیری پایین است.

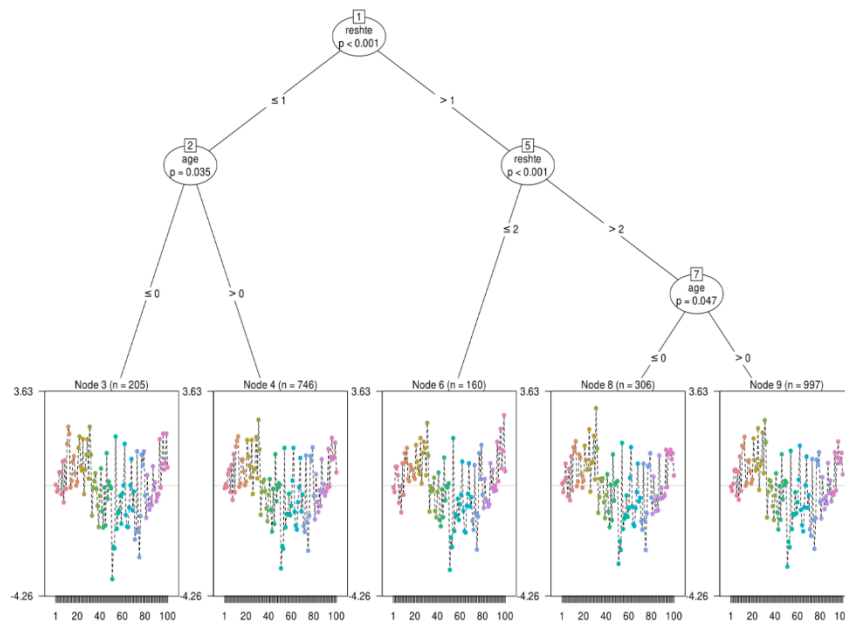
۴- میانگین نمره‌ی آزمون دهندگان گروه ریاضی، $27/97$ و گروه تجربی $25/41$ و گروه علوم انسانی $27/85$ بودند.

۵- نتایج نشان داد که تفاوت قابل‌توجهی میان نمره‌ی میانگین آزمون‌دهندگان گروه آزمایشی نیست. بااینکه 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 قابل‌مقایسه

هستند. گفتنی است که میانگین گروه علوم تجربی از گروه علوم انسانی بالاتر بوده و این تفاوت معنی دار بوده است.

۶- همان طور که بیان شد پاسخ های تعداد ۱۰۰ سؤال برای تعداد ۲۴۱۴ آزمودنی مورد بررسی واقع شد. در این مطالعه دو متغیر سن و گروه های آزمایشی (ریاضی و فیزیک، علوم تجربی و علوم انسانی) لحاظ گردید. داده های موصوف برای دو گروه مردان ۱۸ و ۱۹ سال بررسی گردید. ابتدا با استفاده از مدل درخت راش، رسته ها را به نحوی تعیین می کنیم که سؤالات دارای DIF را آشکار کند. برای این منظور نمودار درختی ۱ استخراج گردید.

نمودار ۱. درخت های راش برای سن و گروه های آزمایشی آزمودنی ها دانشگاه علوم انتظامی امین (منبع: نگارنده)



همان طور که در نمودار ۱ ملاحظه می شود در گروه های پایانه، برآوردهای سؤالات دشوار برای هر یک از ۱۰۰ سؤال نمایش داده شده است. در درخت های مشاهده می شود که پارامترهای سؤال برای آزمودنی های ۱۸ و ۱۹ سال و گروه های آزمایشی موجب ایجاد درخت های بالا و پایین گردیده است. همان طور که در نمودار ۱ مشاهده می شود تعداد ۹

گروه 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 (گره ۸) و افراد دارای ۱۹ سال (گره ۹) تقسیم شده‌اند. سن به‌تنهایی تعداد ۱۳۰۳ نفر (۵۴٪) آزمون دهندگان (تعداد ۳۰۶ نفر در گره ۸ و تعداد ۹۹۷ نفر در گره ۹) را تحت تأثیر قرار داده است که سطح معنی‌داری قابل توجهی دارد. به‌طور کلی؛ الگوی تقسیم‌بندی مشاهده شده در نمودار ۱ در دشواری سؤال نشان می‌دهد که عدم تغییرپذیری اندازه‌گیری در داده‌ها وجود ندارد و زیرگروه‌های داوطلبان (گروه‌های آزمایشی) شناسایی شده تحت تأثیر DIF متفاوت هستند. همچنین بر اساس نمودار ۱ رشته‌های تعیین شده به صورت جدول شماره (۲) تشکیل شده است.

جدول ۲. رشته‌بندی حاصل از درخت رانش

شماره رشته	رشته
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 $\binom{5}{2} = 10$ 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 DIF هستند؛ بنابراین از بین ۱۰۰ سؤال آزمون اختصاصی دانشگاه علوم انتظامی 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 شماره (۳) نشان داده شده است:

جدول ۳. 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	۷۱-۷۴-۸۱	۸۹-۹۹

در ادامه بر اساس رهیافت BH، در سطح $\alpha = 0.05$ ، به قضاوت در خصوص فرضیات زیر می‌پردازیم

$$\begin{cases} H_0: & \text{هیچ دو رسته‌ای در پاسخ‌گویی به سؤال ۳ متفاوت نیستند} \\ H_1: & \text{حداقل دو رسته در پاسخ‌گویی به سؤال ۳ متفاوت هستند} \end{cases}$$

در این رهیافت برای حصول اطمینان از آن‌که احتمال خطای نوع اول از ۰/۰۵ تجاوز نکند، سطح آزمون را برای هر زوج مقایسه برابر با $\frac{0.05}{10} = 0.005$ در نظر گرفته‌ایم. برای بررسی فرضیه بالا تفاضل رسته‌ها به ترتیب P-مقدار محاسبه و خروجی نتایج در جدول شماره (۴) نشان داده شده است.

جدول ۴. 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 شماره (۴)؛ تعداد سؤالات دارای سوگیری به نفع رسته نشان داده شده بالا (ردیف افقی) نسبت به رسته سمت چپ (ردیف عمودی) است. به عنوان 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 (۴) به خوبی می‌تواند نمودار درخت راش (نمودار ۱) را تبیین نماید.

بحث و نتیجه‌گیری

هدف از پژوهش حاضر، بررسی کنشی افتراقی سؤال با یک مطالعه شبیه‌سازی و داده‌های آزمون اختصاصی دانشگاه علوم انتظامی امین با استفاده از درخت‌های راش بود. یافته‌ها نشان می‌دهد که تحلیل درخت‌های راش می‌تواند ۱۰۰٪ کنش افتراقی سؤال شبیه‌سازی شده در نمونه‌های بزرگ (۲۴۱۴ آزمودنی)؛ ۹۳/۴٪ در نمونه‌های ۱۰۰۰ آزمودنی را در تمامی حالات سؤالات بسیار سخت تا بسیار آسان آشکار کند این میزان در نمونه ۵۰۰ آزمودنی به ۴۷٪ تقلیل پیدا می‌کند و در نمونه ۱۵۰ تایی مدل درخت‌های راش برای آشکارسازی کنش افتراقی به هیچ‌وجه مناسب نیست. برای مقایسه مطالعه شبیه‌سازی این تحقیق با تحقیقات گذشته می‌توان گفت که در مطالعه گرامی پور (۱۳۹۳) با استفاده از روش رگرسیون لجستیک به‌طور متوسط ۹۰/۱۸ درصد در نمونه‌های ۱۰۰۰ آزمودنی و در ۳۰ درصد سؤالات موفق به آشکارسازی کنش افتراقی سؤال در مطالعه شبیه‌سازی شده

است این مقدار در تحقیق هیدلاگو و لوپز-پینا^۱ (۲۰۰۴) که با استفاده از رگرسیون لجستیک صورت گرفته بود و نشان داده بودند که این روش نسبت به سایر روش‌های داده-های طبقه‌ای برتری دارد صرفاً ۷۸/۴۲ درصد بوده است. همچنین نتایج کاربرد مدل درخت‌های راش در تشخیص کنش افتراقی سؤال در مطالعه شبیه‌سازی نشان داد که حجم نمونه، شدت کنش افتراقی و درصد سؤالات بر نرخ آشکارسازی کنش افتراقی مؤثر است؛ بنابراین همان‌طور که گفته شد تحقیق حاضر نشان داد که مطالعه هم‌زمان متغیرهای تحقیق و کنترل آن‌ها می‌تواند نتایج قابل قبولی داشته باشد لذا از این حیث، تحقیق حاضر دارای کارکرد بسیار مطلوب در آشکارسازی کنش افتراقی سؤال است.

در بخش تحلیل داده‌های آزمون اختصاصی دانشگاه علوم انتظامی و کاربرد مدل درخت‌های راش در آشکارسازی کنش افتراقی سؤال و با در نظر گرفتن متغیرهای سن و گروه آزمایشی؛ مشخص شد سن و گروه‌های آزمایشی باعث ایجاد اختلاف در بین داوطلبان آزمون دانشگاه علوم انتظامی 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 اختصاصی دانشگاه علوم انتظامی 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 نتیجه‌گیری 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

^۱Hidalgo & López-Pina

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 انجام‌شده 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 به‌عمل‌آمده؛ برخلاف مطالعه آریا دوست (۲۰۱۸) که نشان داد 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 قابل‌نمایش 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 به‌خوبی 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.

مطالعه حاضر به بررسی سوگیری سؤالات آزمون اختصاصی دانشگاه علوم انتظامی امین با استفاده از مدل درخت‌های راش و یک مطالعه شبیه‌سازی شده پرداخت و نشان داده شد که مدل درخت‌های راش به‌خوبی می‌تواند سوگیری سؤالات را در مطالعه شبیه‌سازی همچنین سؤالات آزمون اختصاصی دانشگاه علوم انتظامی را مشخص نماید. لذا برای پژوهش‌های آینده پیشنهاد می‌شود که از مدل درخت‌های راش برای تشخیص کنش افتراقی هماهنگ در نمونه‌های بالای ۱۰۰۰ آزمودنی استفاده شود. البته قضاوت در مورد سوگیری سؤال در این نقطه به پایان نمی‌رسد بلکه مطالعات کیفی و اکتشافی بعد از تحلیل آماری کنش افتراقی شروع می‌شود. الدر^۱ و همکاران (۲۰۰۳) در مطالعه‌ای بعد از حذف سؤال‌های مظنون به سوگیری دریافتند که پاک‌سازی آزمون از این‌گونه سؤال‌ها به سود آزمون است، اما این نکته نیز قابل توجه است که حذف سؤال‌های دارای سوگیری بر پایایی

^۱Elder

آزمون تأثیر نمی گذارد ولی بر اعتبار محتوایی و ساختاری آزمون تأثیر می گذارد. به ویژه اگر سؤال‌های زیادی از آزمون حذف شوند باعث کم‌رنج شدن سازه مورد بررسی و ویران شدن جدول مشخصات آزمون می‌شود. همچنین 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 روتلج^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 هر زمینه مطالعاتی اصولاً نمی‌توان نتیجه‌ای کاملاً منطقی به دست آورد. چراکه توان جمع‌آوری اطلاعات و نیز ابزار جمع‌آوری اطلاعات محدود می‌باشد. در این پژوهش داده‌ها به دلیل محدودیت‌های سازمانی (ناجا) به صورت کامل که بتوان سایر متغیرها (جنسیت، استان، قومیت و ...) را در اختیار داشت در دسترس محقق قرار داده نشد و این مهم بر بررسی مدل درخت‌های راش با متغیرهای گفته شده میسر نگردید. در پایان پیشنهاد می‌شود آزمون اختصاصی دانشگاه علوم انتظامی 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.

تعارض منافع




تعارض منافع ندارم.

سپاسگزاری

پژوهشگران تحقیق حاضر بر خود لازم می‌دانند از معاونت تربیت و آموزش دانشگاه علوم انتظامی امین که در جمع‌آوری داده‌های مورد نیاز تحقیق، نویسندگان را حمایت کردند، کمال تشکر و قدرانی را دارد.

^۱Engelhard, Hansche & Rutledge

ORCID

-  <http://orcid.org/00000-0003-2490-9326>
-  <http://orcid.org/0000-0003-0056-4191>
-  <http://orcid.org/0000-0002-1197-9598>

منابع

گرامی پور، مسعود (۱۳۹۳). ارزیابی توان آماری تحلیل رگرسیون لجستیک در آشکارسازی کنش افتراقی سؤال‌های آزمون. فصلنامه مطالعات اندازه‌گیری و ارزشیابی آموزشی، سال چهارم، شماره ۸، ۱۸۷-۲۱۱. بازیابی از:

<http://jresearch.sanjesh.org/>.

همبلتون، رونالد ک؛ سوامیناتان، اچ؛ راجرز، اچ. جین (۱۹۹۱). مبانی نظریه پرسش - پاسخ. ترجمه محمدرضا فلسفی نژاد (۱۳۸۹). تهران: دانشگاه علامه طباطبایی.

References

- Aryadoust, Vahid. (2018). Using recursive partitioning Rasch trees to investigate differential item functioning in second language reading tests. *Studies in Educational Evaluation*, 56(2018). 197–204.
- 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)*, 57(1), 289-300.
- Camilli, G. (2006). *Test fairness*. In R. Brennan (Ed.), *Educational measurement* (pp. 221-256). New York: American Council on Education & Praeger series on higher education.
- Elder, C.; Mc Namara, T. & Congdon, P. (2003). Rasch techniques for detecting bias in performance tests: An example comparing the performance of native and non-native speakers on a test of academic English. *Journal of Applied Measurement*, 4(2): 181–197.
- Ellis, B. B. & Raju, N. S. (2003). *Test and item bias: What they are, what they aren't, and how to detect them*. Educational Resources information center (ERIC).
- Engelhard, G. Hansche, L. & Rutledge, K. (1990). Accuracy of bias review judges in identifying differential item functioning on teacher certification tests. *Applied Measurement in Education*, 3(4), 347–360
- Fischer G, Molenaar I (eds.) (1995). *Rasch Models: Foundations, Recent Developments and Applications*. Springer-Verlag, New York.
- Geramipour, Masoud. (2020). Item-Focused Trees Approach in Differential Item Functioning (DIF) Analysis: A Case Study of an EFL Reading Comprehension Test, *Journal of Research in English Language Studies*, Vol. 7, No. 2, 2020,123-147.
- Hambleton, R, K. Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications.

- Hidalgo, M. D. & López-Pina, J. P. (2004). Differential item functioning detection and effect size: A comparison between logistic regression and Mantel Haenszel procedures. *Educational and Psychological Measurement*, 7(11): 878–887.
- 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.
- Liu, Yanlou, Yin, Hao, Xin, Tao, Shao, Laicheng, Yuan, Lu (2019). A Comparison of Differential Item Functioning Detection Methods in Cognitive Diagnostic Models, doi: 10.3389/fpsyg.2019.01137.
- Merkle EC, Zeileis A (2013). Tests of Measurement Invariance without Subgroups: A Generalization of Classical Methods. *Psychometrika*, 78(1), 59–82.
- Millsap, R. E. (2012). *Statistical approaches to measurement invariance*. New York, NY: Routledge
- Moritz, B. (2020). *Item Focussed recursive Partitioning for Simultaneous Selection of items and variables that induce Differential Item Functioning (DIF) in dichotomous or Polytomous items*. Package DIFtree, Version 3.1.6.
- Popham, W. J. (2005). High-Stakes Tests: Harmful, Permanent, Fixable. *American Educational Research Journal*.6(4), p85.
- Strobl, Carolin, Kopf, Julia, Zeileis, Achim (2015). *Rasch Trees: A New Method for Detecting Differential Item Functioning in the Rasch Model*, Working Papers in Economics and Statistics.
- Tutz, G., & Berger, M. (2016). Item-focussed trees for the identification of items in differential item functioning. *Psychometrika*, 81(3), 727-750.
- Yüksel, Selcen., Halil Elhan, Atilla., Gökmen, Derya., Küçükdeveci, Ayşe Adile., & Kutlay, Şehim. (2019). Analyzing differential item functioning of the Nottingham Health Profile by Mixed Rasch Model. *Turk J Phys Med Rehab* 2018;64(4):300-307.
- 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.
- Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, 4(2), 223-233.

In Persian

- Gramipour, Massoud (1393). Evaluation of statistical power of logistic regression analysis in revealing the differential action of test questions. *Journal of Educational Measurement and Evaluation Studies*, Fourth Year, No. 8, 187-211. Retrieved from: <http://jresearch.sanjesh.org/>.

Hambleton, Ronald K.; Saminathan, H.; Rogers, H.. Jane (1991).
Fundamentals of Question-Answer Theory. Translated by Mohammad
Reza Filsafinejad (1389). Tehran: Allameh Tabatabai University.