خوشه‌بندی: ابزاری برای آنالیز داده‌ها در مطالعات کمی و آمیخته

فرزاد رادمهر*[۱]، سیدحسن علم الهدائی[۲]

تاریخ دریافت: ۱۳۹۳/۰۲/۱۴ تاریخ پذیرش: ۱۳۹۳/۰۳/۲۱

چکیده

در این مطالعه خوشه بندی به‌عنوان یکی از روش‌های داده کاوی توصیفی جهت استفاده در مطالعات آموزشی و روانشناختی مطرح می‌گردد. گروه‌بندی داده‌های پژوهش در مطالعات قبلی در حوزه‌های آموزشی و روانشناختی توسط میانگین، میانه و برد صورت گرفته است که دارای نواقص و محدودیت‌هایی می‌باشد. لذا در این مطالعه خوشه‌بندی به‌عنوان ابزاری جایگزین با قابلیت‌های متعدد در مطالعات کمی و آمیخته معرفی می‌گردد. بدین منظور، ابتدا تعریف اجمالی از خوشه‌بندی ارائه گردیده و سپس مراحل ششگانه اجرای فرایند خوشه‌بندی(انتخاب متغیرهای خوشه‌بندی، انتخاب روش خوشه‌بندی، انتخاب معیار شباهت/تفاوت، انتخاب الگوریتم خوشه‌بندی، انتخاب تعداد خوشه و بررسی اعتبار و تفسیر نتایج خوشه‌بندی) توسط داده‌های کمی مرتبط با روانشناسی یادگیری ریاضیات بیان می‌گردد تا ضمن آشنایی بیشتر پژوهشگران حوزه‌ی مطالعات آموزشی و روانشناختی با این ابزار، نتایج جدیدی در زمینه‌ی روانشناسی یادگیری ریاضیات مطرح گردد. نتایج اعمال خوشه‌بندی بر روی داده‌ها نشان می‌دهد که از خوشه‌بندی می‌توان برای بررسی ارتباط متغیرهای مختلف بر متغیر وابسته در مطالعات کمی و شناسایی کاندیداهای شرکت در مصاحبه در مطالعات آمیخته استفاده کرد.

واژه‌های کلیدی: خوشه‌بندی، داده کاوی، روانشناسی یادگیری ریاضی.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

۱- دانشجوی دکتری آموزش ریاضی دانشکده علوم ریاضی دانشگاه فردوسی مشهد.
۲- استاد آموزش ریاضی دانشکده علوم ریاضی دانشگاه فردوسی مشهد.
* نویسنده‌ی مسئول مقاله، farzad.radmehr@vuw.ac.nz

مقدمه

داده کاوی، هنر و علم آنالیز هوشمندانه داده‌ها است و هدف آن پیدا کردن بینش و دانش نسبت به داده‌های پژوهش می‌باشد. این کنکاش گاهی به‌عنوان فرایند ساختن مدل نیز تعریف می‌شود و مدل ساخته شده برای فهم بیشتر دانش نهفته در داده‌ها و پیش‌بینی وقایع استفاده می‌گردد. با توجه به افزایش روزافزون داده‌ها، داده کاوی در اکثر حوزه‌های پژوهش از جمله مدیریت، اقتصاد، مهندسی، علوم پزشکی و زیست استفاده گردیده است(ویلیامز[1]،۲۰۱۱). در ایران نیز اخیرا این ابزار در علوم مهندسی استفاده شده است(به‌عنوان مثال کشاورزی، پیروی و رحمانی،۱۳۸۹). اما مطالعه‌ای در ایران که از این ابزار در پژوهش‌های آموزشی و روانشناختی استفاده کرده باشند، از سوی نویسندگان این مقاله یافت نگریده است. لذا در این پژوهش خوشه‌بندی به‌عنوان یکی از روش‌های داده کاوی برای پژوهشگران حوزه‌ی مطالعات آموزشی و روانشناختی معرفی می‌گردد تا بتوانند در مطالعات کمی[2] و آمیخته[3] خود برای کشف دانش نهفته در مجموعه داده‌ها استفاده نمایند.

خوشه‌بندی به‌عنوان یکی از روش‌های داده کاوی توصیفی، تکنیکی برای گروه‌بندی مشاهدات به ***k*** خوشه(گروه) مختلف می‌باشد؛ به‌طوری که مشاهداتی که در یک خوشه قرار می‌گیرند، بیشترین شباهت و مشاهدات خوشه‌های مختلف بیشترین تفاوت ممکن را با یکدیگر داشته باشند(کافمن و راسیو[4]،۲۰۰۵). برای انجام فرایند خوشه‌بندی بر روی مجموعه‌ای از داده‌ها، مراحل شش‌گانه‌ی موجود در شکل ۱ در نظر گرفته می‌شود(موئی و سارستد[5]،۲۰۱۱) که هر کدام در این مقاله به اختصار توضیح داده خواهند شد. همچنین برای فهم بیشتر و نحوه‌ی استفاده از آنان، فرایند خوشه‌بندی برای مجموعه‌ای از داده‌های کمی مرتبط با روانشناسی یادگیری ریاضی استفاده می‌گردد. بنابراین بر خلاف شیوه‌های معمول مقالات آموزشی و روانشناختی، در این مقاله از ابتدا برای فهم بیشتر این ابزار، نتایج برای داده‌های در نظر گرفته شده عنوان می‌شود تا خواننده بتواند هر کدام از مراحل خوشه‌بندی را بهتر درک نماید. داده‌های این پژوهش برگرفته از مطالعه‌ی حاجی‌بابا، رادمهر و علم‌الهدائی(۲۰۱۳) در رابطه‌ی بین متغیرهای اضطراب، نگرش، دقت ریاضی، ظرفیت حافظه فعال و سبک یادگیری فراگیران با عملکرد ریاضی آنان می‌باشد که به‌صورت کمی مورد مطالعه قرار گرفته است. نرم افزاری که برای انجام فرایند خوشه‌بندی در این مقاله استفاده

1 Williams
2 quantitative
3 Mixed methods
4 Kaufman & Rousseeuw
5 Mooi & Sarstedt

گردیده است، نرم افزار SPSS نسخه ۲۲ می‌باشد. شایان ذکر است که فرایند خوشه‌بندی توسط نرم‌افزارهای دیگر مانند R نیز قابل انجام است(براک،پیهور،داتتا،و داتتا[1]،۲۰۰۸).

لازم به ذکر است که گروه‌بندی یا خوشه‌بندی داده‌های پژوهش در مطالعات قبلی در حوزه‌های آموزشی و روانشناختی توسط میانگین، میانه و برد(به‌عنوان مثال علم‌الهدایی،۲۰۰۹) صورت می‌گرفته است که دارای نواقص و محدودیت‌هایی می‌باشد. برای نمونه می‌توان به محدود بودن آن به گروه‌بندی فقط بر اساس یک متغیر و نبود معیار مشخص در مورد تعداد گروهی که برای خوشه‌بندی داده‌ها باید در نظر گرفت، اشاره کرد. درحالی‌که در خوشه‌بندی می‌توان گروه‌بندی را بر اساس چندین متغیر صورت داد و بر اساس معیارهایی که در ادامه معرفی می‌شود، تعداد خوشه را نیز تعیین نمود.

روش

انتخاب متغیرها برای خوشه‌بندی

الگوریتم‌های خوشه‌بندی مانند بسیاری از آنالیزهای آماری برای هر نوع داده ورودی، نتایج را فارغ از اینکه داده‌ها پیش فرض‌های لازم برای خوشه‌بندی را دارا می‌باشند، در اختیار کاربر قرار می‌دهند(پاستور[2]،۲۰۱۰). لذا انتخاب متغیرهای صحیح مرتبط با دیسیپلین برای کسب نتایج معتبر حائز اهمیت می‌باشد. در بعضی از موارد، متغیرهایی که برای خوشه‌بندی باید انتخاب گردند از محتوای دیسیپلین قابل تشخیص است. به‌عنوان مثال، پژوهشگری که علاقه‌مند است تاثیر متغیرهای هیجانی و عاطفی را بر عملکرد ریاضی فراگیران بررسی کند، متغیرهایی نظیر اضطراب ریاضی، نگرش ریاضی، خودکارآمدی ریاضی را برای خوشه‌بندی مشاهدات خود به‌کار می‌برد؛ زیرا در ادبیات تحقیق این متغیرها به‌عنوان پیش‌بینی کننده‌ی عملکرد ریاضی در سطوح ریاضیات مدرسه‌ای مطرح شده‌اند(ساها[3]،۲۰۰۷؛ علم‌الهدائی،۲۰۰۹؛ موسوی، رادمهر و علم الهدایی،۲۰۱۲). در مورد داده‌های این پژوهش نیز در ادبیات تحقیق، به رابطه‌ی بین عملکرد ریاضی فراگیران و متغیرهای اضطراب ریاضی(علم‌الهدائی،۲۰۰۹)، نگرش ریاضی(ساها،۲۰۰۷)، دقت ریاضی(امانی، علم‌الهدائی و رادمهر،۲۰۱۱)، ظرفیت حافظه فعال(راقبار، بارنز و هچت[4]،۲۰۱۰) و سبک‌های یادگیری(علم‌الهدائی،۲۰۰۹؛ موسوی، رادمهر و علم الهدایی،۲۰۱۲) اشاره نموده است. در صورت نبود ادبیات تحقیق درمورد متغیرهای انتخابی جهت خوشه‌بندی، توصیه می‌شود که پژوهشگر علت

1 Brock, Pihur, Datta, and Datta
2 Pastor
3 Saha
4 Raghubar , Barnes & Hecht

انتخاب متغیرها را برای پژوهش خود تبیین نماید تا نتایج حاصل از خوشه‌بندی قابل اتکا گردد (پاستور، ۲۰۱۰).

از دیگر نکاتی که در این مرحله باید بدان توجه ویژه داشت، این است که نباید متغیرهای زیادی را برای خوشه‌بندی مشاهدات استفاده نمود؛ زیرا انتخاب تعداد زیادی متغیر جهت خوشه‌بندی، احتمال اینکه بعضی از متغیرها ویژگی‌های تقریبا یکسانی را اندازه‌گیری کنند، بالا می‌برد(موئی و سارستد،۲۰۱۱). از طرف دیگر متغیرهایی که همبستگی بالایی با یکدیگر دارند(همبستگی بالای ۰.۹۰) برای خوشه‌بندی همزمان نباید استفاده گردند؛ زیرا سبب می‌شوند که نتایج خوشه‌بندی بیشتر متاثر از این فاکتورها شود(پاستور، ۲۰۱۰؛ کتچن و شک[1]،۱۹۹۶؛ موئی و سارستد، ۲۰۱۱). به‌عنوان نمونه اگر متغیرهای باور ریاضی، نگرش ریاضی و اضطراب ریاضی برای خوشه‌بندی استفاده گردند، نتیجه‌ی خوشه‌بندی بیشتر متاثر از فاکتورهای باور و نگرش ریاضی خواهد بود. زیرا این دو فاکتور همبستگی بالایی با یکدیگر دارند. در ارتباط با پنج فاکتور در نظر گرفته شده برای خوشه‌بندی در مطالعه‌ی حاضر(متغیرهای اضطراب، نگرش، دقت ریاضی، ظرفیت حافظه فعال و نمره سبک یادگیری فراگیران) با توجه به کمی بودن فاکتورها ضریب همبستگی پیرسون محاسبه گردید. میزان این ضریب برای آنها بین منفی ۰.۳۱۱ و مثبت ۰.۳۴۶ متغیر بوده است که نشان می‌دهد همبستگی بالایی بین آنها وجود ندارد. در صورتی که ضریب همبستگی بالایی بین متغیرها وجود داشته باشد، می‌توان از تحلیل عاملی جهت کاهش متغیرها استفاده کرد(پاستور، ۲۰۱۰؛ موئی و سارستد، ۲۰۱۱).

نکته دیگری که محققان در انتخاب متغیرهای خوشه بندی باید رعایت نمایند، مرتبط با متغیرهایی است که شرکت‌کنندگان در پژوهش همگی به آن پاسخ یکسان داده‌اند. متغیرهایی از این قبیل برای خوشه‌بندی نباید استفاده گردند(بنت[2]،۱۹۷۵؛ پاستور،۲۰۱۰). زیرا این متغیرها شاخص‌های مناسبی برای تفکیک شرکت‌کنندگان در پژوهش نمی‌باشند. در این زمینه، داده‌های پژوهش حاضر در مورد پرسشنامه‌ی دقت ریاضی که شامل ۲۵ سوال با مقیاس لیکرت ۵ گزینه‌ای(خیلی کم، کم، متوسط، زیاد و خیلی زیاد) بوده است؛ شرکت‌کنندگان در پژوهش همه گزینه‌ها را در پاسخ‌های خود انتخاب نموده‌اند. به عبارت دیگر در جدول فراوانی که برای هر سوال رسم گردید، فراوانی صفر برای هیچکدام از گزینه‌های سوالات دیده نشد. در مورد متغیر اضطراب ریاضی نیز که از ۲۵ سوال با مقیاس لیکرت ۵ گزینه‌ای(خیلی کم، کم، متوسط، زیاد و خیلی زیاد) تشکیل یافته بود، شرایط مشابه بود. در مورد نگرش ریاضی که از ۴۷ سوال با مقیاس مشابه تشکیل شده بود، به جز سوال

1 Ketchen, & Shook
2 Bennett

اول که شرکت‌کنندگان به گزینه‌ی خیلی کم و کم پاسخ نداده بودند، در بقیه‌ی سوالات این پرسشنامه فراوانی صفر برای گزینه‌ها مشاهده نشد. در مورد ظرفیت حافظه فعال و نمره آزمون سبک یادیگری که توسط دو آزمون اندازه‌گیری گردیدند، نمرات فراگیران در این دو متغیر متفاوت بوده است. به عبارت دقیق‌تر، در مورد آزمون سبک‌های یادگیری که نمره‌ی آن بین ۰ تا ۲۰ متغیر می‌باشد، فراگیران نمره‌های متفاوتی از ۱ تا ۲۰ کسب کرده‌اند و در مورد آزمون ظرفیت حافظه‌ی فعال که نمره‌ی آن بین ۳ تا ۷ متغیر می‌باشد، فراگیران نمرات متفاوتی در این بازه کسب نموده‌اند. لذا تمامی این ۵ متغیر شرایط لازم برای انجام خوشه‌بندی را دارا می‌باشند[۱].

لازم به ذکر است که تعداد متغیرهایی که برای خوشه‌بندی در نظر گرفته می‌شود، بر روی حداقل نمونه‌ی انتخابی جهت خوشه‌بندی نیز تاثیر می‌گذارد. به عبارت دیگر، با افزایش تعداد متغیرها حجم نمونه‌ی انتخابی نیز باید افزایش یابد. فورمن[۲](۱۹۸۴)، تعداد 2^x مشاهده را به‌عنوان حداقل نمونه جهت خوشه‌بندی توصیه می‌کند که در آن x تعداد متغیر جهت خوشه‌بندی می‌باشد. به‌عنوان مثال در داده‌های پژوهش حاضر، ۵ متغیر جهت خوشه‌بندی انتخاب گردیده است. لذا حداقل نمونه جهت خوشه‌بندی در این مطالعه ۳۲ نفر می‌باشد. تعداد نمونه‌ی مطالعه‌ی حاجی بابا، رادمهر و علم‌الهدایی(۲۰۱۳)، ۱۶۹ دانش‌آموز سال سوم دبیرستان از شهرستان مشهد گزارش شده است که ۱۱۲ نفر به همه‌ی پرسشنامه‌ها پاسخ داده و در آزمون‌ها شرکت نموده‌اند. لذا این مجموعه داده، حداقل نمونه برای خوشه‌بندی را داراست.

انتخاب روش خوشه‌بندی

گام دوم در فرایند خوشه‌بندی داده‌ها، انتخاب روش خوشه‌بندی می‌باشد. با انتخاب روش خوشه‌بندی، پژوهشگران روشی که خوشه‌ها تشکیل می‌شوند را مشخص می‌نماید که معمولا شامل بهینه کردن مشخصه‌ای مانند حداقل کردن واریانس داخل خوشه‌ها یا حداکثر کردن فاصله‌ی بین خوشه‌ها می‌باشد(موئی و سارستد،۲۰۱۱). روش‌های خوشه‌بندی به دو دسته‌ی قطعی(هر مشاهده تنها در یک خوشه قرار می‌گیرد) و فازی(یک مشاهده می‌تواند به بیش از یک خوشه تعلق داشته باشد) تقسیم‌بندی می‌شوند(حسینی،۱۳۹۱). خوشه‌بندی قطعی که مورد نظر این مقاله می‌باشد، به سه دسته افرازی، سلسله مراتبی و مبتنی بر چگالی گروه‌بندی می‌شود(اوریت[۳] و همکاران، ۲۰۱۱). در بین این سه دسته، در مطالعه‌ی حاضر روش‌های افرازی و سلسله مراتبی مورد توجه

---

۱ برای کسب اطلاعات بیشتر در مورد پرسشنامه ها و آزمونهای استفاده شده به حاجی بابا، رادمهر و علم الهدائی (۲۰۱۳) مراجعه نمایید.

2 Formann

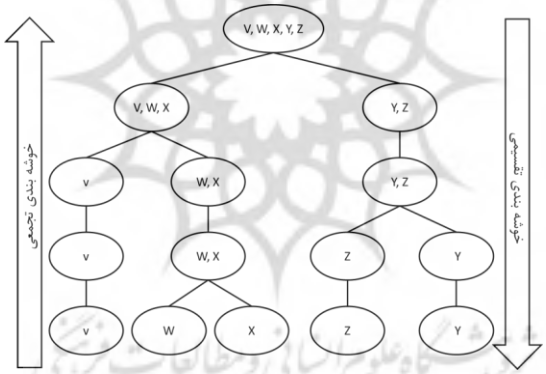
3 Everitt et al.

قرار می گیرند، زیرا این تکنیک‌ها برای خوشه‌بندی داده‌های پژوهش‌های آموزشی مورد استفاده قرار گرفته‌اند(به‌عنوان مثال: شاولسون[1]،۱۹۷۹؛ هاتی[2]، ۲۰۰۲؛ اویلید[3] و همکاران ، ۲۰۱۰).

خوشه‌بندی سلسله مراتبی

روش‌های خوشه‌بندی سلسله مراتبی به دو دسته‌ی کلی تجمیعی[4] و تقسیمی[5] گروه‌بندی می‌شوند که بر خلاف روش‌های افرازی نیازی به تعیین تعداد خوشه در آنها وجود ندارد. همان‌طور که در شکل ۲ نمایش داده شده است، در روش تجمیعی هر مشاهده در ابتدا در یک خوشه مجزا قرار می‌گیرد و سپس دو خوشه‌ای که بیشترین شباهت را با یکدیگر دارند، با هم ادغام می‌شوند. این فرایند مرتبا تکرار می‌شود تا زمانی که در نهایت تمامی مشاهدات در یک خوشه قرار بگیرند. در روش تقسیمی، فرایند برعکس می‌باشد و ابتدا تمامی مشاهدات در یک خوشه قرار می‌گیرند و فرایند خوشه‌بندی آغاز می‌گردد و تا زمانی که هر مشاهده در یک خوشه مجزا قرار نگیرد، ادامه پیدا می‌کند(اوریت و همکاران،۲۰۱۱؛ کافمن و راسیو،۲۰۰۵).

شکل ۱- مراحل فرایند خوشه‌بندی(اقتباس از موئی و سارستد،۲۰۱۱)



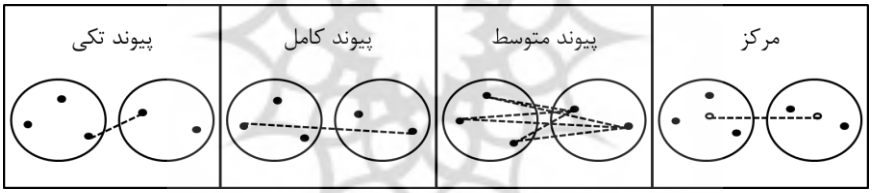
شکل ۲- خوشه‌بندی تجمیعی و تقسیمی

1 Shavelson
2 Hattie
3 Oyelade
4 Agglomerative
5 Divisible

همان‌طور که در شکل ۲ مشاهده می‌شود، در هر مرحله فقط دو خوشه با یکدیگر ادغام می‌شوند؛ لذا فرایند خوشه‌بندی توسط روش‌های سلسله مراتبی زمان بر می‌باشد. در نتیجه این روش برای مجموعه داده‌های بالای ۵۰۰ مشاهده توصیه نمی‌شود(موئی و سارستد،۲۰۱۱). با توجه به اینکه روش‌های خوشه‌بندی تقسیمی به‌علت پیچیدگی محاسباتی، کمتر برای برای خوشه‌بندی مشاهدات استفاده شده و بعضی از نرم‌افزارهای آماری این روش‌ها را پوشش نمی‌دهند(حسینی،۱۳۹۱)، در این پژوهش فقط الگوریتم‌های پرکاربرد خوشه‌بندی تجمیعی را به اجمال معرفی می‌نماییم.

انتخاب الگوریتم خوشه‌بندی

الگوریتم‌های متفاوتی برای انجام فرایند خوشه‌بندی سلسله مراتبی تجمیعی وجود دارد که وجه تمایز آنها بر اساس نحوه‌ی تعریف فاصله بین دو مشاهده و نحوه‌ی شکل‌گیری خوشه‌ها می‌باشد. پرکاربردترین الگوریتم‌های تجمعی عبارتند از الگوریتم پیوند تکی[1](فاصله‌ی بین دو خوشه توسط کمترین فاصله ممکن بین دو مشاهده خوشه تعیین می‌شود)، پیوند کامل[2](فاصله‌ی بین دو خوشه توسط بیشترین فاصله‌ی ممکن بین دو مشاهده خوشه تبیین می‌گردد)، پیوند متوسط[3](میانگین فاصله بین مشاهدات دو خوشه به‌عنوان فاصله‌ی دو خوشه تعریف می‌شود)، الگوریتم مرکز[4](در ابتدا مرکز هندسی هر خوشه مشخص شده و فاصله‌ی بین دو خوشه توسط فاصله‌ی بین دو مرکز خوشه تعریف می‌شود)(شکل ۳) و الگوریتم وارد[5] (مشاهداتی با یکدیگر ادغام می‌شوند که واریانس داخل گروهی حداقل شود)(موئی و سارستد،۲۰۱۱).



شکل۳- الگوریتم‌های پیوند تکی، کامل، متوسط و مرکز

هرکدام از پنج روش فوق کاربردهای متفاوت و نقاط ضعف و قوت خاص خود را دارا می‌باشند(برای آشنایی بیشتر به اوریت و همکاران،۲۰۱۱ مراجعه نمایید). به‌عنوان مثال الگوریتم وارد که در مطالعات آموزشی(به‌عنوان مثال: راک، برد و لین[6]،۱۹۷۲) مورد استفاده قرار گرفته است. درصورتی

1 Single linkage
2 Complete linkage
3 Average linkage
4 Centroid
5 Ward
6 Rock, Baird, and Linn

درصورتی قویا توصیه می‌گردد(اوریت و همکاران،۲۰۱۱؛ موئی و سارستد،۲۰۱۱) که داده‌های مطالعه دارای داده‌های پرت[1] نباشد و محقق پیش‌بینی کند که تعداد مشاهدات در خوشه‌ها یکسان می‌باشد. درغیر این صورت، روش مرکز می‌تواند جایگزین مناسبی باشد(اوریت و همکاران،۲۰۱۱). در صورت انجام خوشه‌بندی به‌کمک الگوریتم پیوند تکی، اکثریت داده‌ها در یک خوشه قرار گرفته و خوشه‌های دیگر دارای یک یا دو مشاهده می‌باشند. لذا این الگوریتم برای شناسایی داده‌های پرت بسیار مناسب است(موئی و سارستد،۲۰۱۱) زیرا داده‌های پرت در خوشه‌های مجزا قرار می-گیرند.

این خصوصیت پیوند تکی می‌تواند برای مطالعات آموزشی مفید باشد. در صورتی که محقق از روش کمی استفاده کند، می‌تواند ابتدا داده‌های پرت را به کمک این روش شناسایی کرده و آزمون‌های فرضیه‌ی خود را پس از خروج این داده‌ها جداگانه انجام دهد. زیرا داده‌های پرت در یک مطالعه‌ی کمی تاثیر به‌سزایی در رد یا عدم رد فرضیات تحقیق ایفا می‌نمایند. در روش‌های تلفیقی، شناسایی داده‌های پرت برای بعضی از محققان با اهمیت می‌باشد. به‌عنوان نمونه ممکن است محققی علاقه‌مند باشد تا مشکلات ریاضی فراگیران با اضطراب ریاضی بسیار بالا را شناسایی کند. محقق، این پژوهش را می‌تواند به‌صورت آمیخته متوالی[2] انجام دهد. بدین طریق که ابتدا فراگیران با اضطراب ریاضی بسیار بالا را در بین مجموعه داده‌ی هدف خود شناسایی کرده و سپس با انجام مصاحبه‌های نیمه ساختاری مشکلات ریاضی آنها را مورد موشکافی قرار دهد. در این صورت، توزیع پرسشنامه‌ی اضطراب ریاضی بین دانش‌آموزان چندین کلاس در مقطع مورد نظر و سپس انجام خوشه‌بندی تجمیعی به‌کمک پیوند تکی برای نتایج پرسشنامه، راهکار مناسبی برای شناسایی کاندیداهای مناسب برای مصاحبه به نظر می‌رسد.

خوشه‌بندی افرازی

در ادبیات تحقیق الگوریتم‌های مختلفی برای خوشه‌بندی افرازی وجود دارد. اما در پژوهش حاضر الگوریتم k-mean از خوشه‌بندی افرازی برای تحلیل داده‌های حوزه‌ی مطالعات آموزشی و روانشناختی انتخاب گردیده است. زیرا توسط نرم‌افزار SPSS قابل انجام است و به‌عنوان یکی از الگوریتم‌های ساده و سریع که مسائل معروف خوشه‌بندی را حل می‌کند، گزارش شده است (ولموروگان و سانتهانام[3]، ۲۰۱۰). در حوزه‌های مختلف نیز از جمله مطالعات علوم انسانی مورد استفاده قرار گرفته است(به‌عنوان نمونه کرنتز،کورن و منینگر[4]، ۲۰۰۹؛ تان، استینباخ و کومار[1]،

1 Outlier
2 Sequential mixed method
3 Velmurugan & Santhanam
4 Krantz, Korn, & Menninger

۲۰۰۶). در این روش تعداد خوشه‌ها(k) می‌تواند قبل از خوشه‌بندی مشخص شود و مجموعه داده‌ها در k خوشه‌ی اولیه قرار می‌گیرند. پس از آن مشاهدات از خوشه‌ای به خوشه‌ی دیگر منتقل می‌شوند. در صورتی که واریانس داخل گروهی خوشه‌ها(فاصله بین مشاهده و مرکز خوشه) کاهش یابد(حسینی،۱۳۹۱). فرایند خوشه‌بندی زمانی توقف می‌شود که تغییری در عضویت مشاهدات صورت نگیرد یا تعداد دفعات مشخصی فرایند خوشه‌بندی تکرار شده باشد(موئی و سارستد،۲۰۱۱).

خوشه‌بندی دو بخشی[۲]

در این روش خوشه‌بندی که توسط نرم‌افزار SPSS قابل انجام است، ابتدا الگوریتمی مشابه با الگوریتم k-mean بر روی داده‌ها انجام شده و سپس بر اساس نتایج به‌دست آمده، الگوریتمی تجمیعی بر روی داده‌ها اعمال می‌گردد. این روش می‌تواند همزمان داده‌های کمی(داده‌های برآمده از آزمون‌ها و پرسشنامه‌ها) و داده‌های اسمی(جنسیت) را خوشه‌بندی کند. همچنین به کاربر این امکان را می‌دهد که تعداد خوشه را مانند روش k-mean انتخاب نماید. از دیگر امکانات این روش این است که می‌تواند حداکثر تعداد خوشه را مشخص نماید و روش دوبخشی بر اساس استانداردهای آماری تعداد خوشه را برای کاربر انتخاب می‌نماید(موئی و سارستد،۲۰۱۱).

انتخاب معیار تشابه/ تفاوت

گام بعدی برای انجام فرایند خوشه‌بندی، انتخاب معیار تشابه/تفاوت است تا بتوان فاصله‌ی دو مشاهده را به‌صورت یک مقدار عددی مشخص نمود(کافمن و راسیو،۲۰۰۵). معیارهای متفاوتی برای خوشه‌بندی مجموعه داده‌ها وجود دارد که وابسته به ماهیت داده‌ها(مانند کمی، اسمی و...) می‌باشد(اوریت و همکاران،۲۰۱۱). داده‌هایی که معمولا در مطالعات کمی استفاده می‌شود، برگرفته از پرسشنامه‌ها و آزمون‌ها می‌باشد که به‌صورت داده‌های فاصله‌ای یا نسبی در نظر گرفته می‌شوند. درمیان معیارهای تشابه برای داده‌ی کمی(متر اقلیدسی[۳]، مینکوسکی[۴]، ضریب همبستگی پیرسون)، متر اقلیدسی توجه بیشتری را به خود جذب کرده است(ابونی و فیل[۵]،۲۰۰۷؛ اوریت و همکاران،۲۰۱۱؛ پاستور،۲۰۱۰؛ کافمن و راسیو،۲۰۰۵ ؛ موئی و سارستد،۲۰۱۱). در این مطالعه متر اقلیدسی برای الگوریتم پیوندتکی و توان دوم آن برای خوشه‌بندی به روش وارد اعمال می‌شود.

پرتال جامع علوم انسانی

1 Tan, Steinbach, & Kumar
2 Two step clustering
3 Euclidean
4 Minkowski
5 Abonyi & Feil

همچنین برای روش دوبخشی علاوه بر متر اقلیدسی، لگاریتم نسبت درست‌نمایی[1] که قابلیت محاسبه‌ی معیار تشابه/تفاوت را برای داده‌های کمی و گروه‌بندی را داراست، استفاده می‌شود.

تبدیل داده‌ها

تبدیل کردن داده‌ها هنگامی که متغیرهای پژوهش دارای مقیاس‌های متفاوتی هستند یا دارای برد متفاوت می‌باشند، صورت می‌پذیرد تا نتایج خوشه‌بندی متاثر از متغیرها با واریانس بالا نباشد (پاستور، ۲۰۱۰؛ جانسون و ویچرن[2]، ۲۰۰۷). روش‌های متفاوتی مانند z-score و استاندارد کردن توسط برد(تبدیل داده‌ها از مقیاس اصلی به مقیاس صفر تا یک یا از مقیاس اصلی به مقیاس ۱- تا ۱) وجود دارد(موئی و سارستد،۲۰۱۱) که در این پژوهش از روش z-score که در آن هر متغیر را به متغیری با میانگین صفر و انحراف از معیار یک تبدیل می‌کنند، استفاده می‌شود.

انتخاب تعداد خوشه

معیارهای اعتبار مختلفی برای ارزیابی خوشه‌بندی و انتخاب تعداد خوشه براساس محاسبات آماری وجود دارد(برای اطلاعات بیشتر رجوع کنید به اوریت و همکاران،۲۰۱۱؛ براک، پیهور، داتتا و داتتا، ۲۰۰۸) . از بین معیارهای عنوان شده در ادبیات تحقیق در این مطالعه معیارهای ضریب نیم رخ[3] (روسیوف[4]،۱۹۸۷)، استاندارد زانو[5] (تورنیدک[6]،۱۹۵۳)، استاندارد نسبت واریانس[7] (کالینسکی و هاراباس[8]،۱۹۷۴) و معیارهای آکائیک[9](آکائیک،۱۹۷۴) و بایس[10](شوارز[11]،۱۹۷۸) برای بررسی اعتبار خوشه‌بندی و انتخاب تعداد خوشه(به‌علت قابل محاسبه بودن توسط نرم‌افزار SPSS) انتخاب گردیده‌اند که ابتدا هر کدام از آنها توضیح داده می‌شوند و سپس برای مجموعه داده در نظر گرفته شده در بخش‌های بعدی به‌کار گرفته خواهند شد.

در روش‌های سلسله مراتبی از فاصله‌ی دو مشاهده‌ای که با یکدیگر ادغام و در یک خوشه قرار می‌گیرند، به‌عنوان معیاری برای انتخاب تعداد خوشه استفاده می‌نمایند. این روش که توسط تورنیدک(۱۹۵۳) معرفی گردید، به استاندارد زانو معروف می‌باشد. برای محاسبه‌ی آن باید جدول

---

1 Log-likelihood
2 Johnson & Wichern
3 Silhouette measure
4 Rousseeuw
5 Elbow criteria
6 Thorndike
7 Variance Ratio Criterion(VRC)
8 Calinski and Harabasz
9 Akaike
10 Bayes
11 Schwarz

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

تجمیعی[1] برای خروجی SPSS فراخوانی شود. دو ستون ضرایب[2] و مراحل[3] در یک فایل اکسل ذخیره شده و نموداری خطی برای این دو مقادیر به‌طوری رسم می‌شود که ستون مراحل، محور افقی و ستون ضرایب، محور عمودی را نشان دهند. در هر مکان که تفاوت قابل ملاحظه‌ای نسبت به مرحله‌ی قبل بین فاصله‌ی دو مشاهده‌ای که با هم ادغام می‌شوند، دیده شود؛ شماره‌ی مرحله‌ی در نظر گرفته شده و آن عدد از تعداد نمونه تفریق می‌گردد. عدد حاصل تعداد خوشه‌ی پیشنهادی توسط استاندارد زانو می‌باشد. مشکلی که برای این روش وجود دارد، این است که همیشه تفاوت قابل تشخیص نمی‌باشد و تفاوت قابل ملاحظه" یک صفت نسبی است و هر فرد برداشت خاص خود نسبت به این مفهوم را دارد(موئی و سارستد،۲۰۱۱).

معیار دیگری که در تحقیقات گذشته به آن اشاره شده و کارایی آن ثابت گشته است(میلیگان و کوپر[4]،۱۹۸۵) استاندارد نسبت واریانس(کالینسکی و هاراباس،۱۹۷۴) می‌باشد که توسط انجام خوشه‌بندی به روش k-mean قابل محاسبه است. برای محاسبه‌ی آن هنگام انجام فرایند خوشه‌بندی باید جدول ANOVA نیز فراخوانی شود و مقادیر F برای متغیرها با یکدیگر جمع گردند. جمع مقادیر F ، ضریب نسبت واریانس(VRC_k) را برای خوشه‌ی در نظر گرفته‌ی K نشان می‌دهد. پس از محاسبه‌ی VRC_k برای Kهای مختلف،K نظیر خوشه‌ای که VRC_k آن ماکزیمم باشد، به‌عنوان تعداد خوشه در نظر گرفته می‌شود.

استاندارد سوم، ضریب نیمرخ می‌باشد که معیاری برای نیکویی برازش[5] داده‌ها در خوشه‌ها است که که مقداری بین ۱ و ۱- انتخاب می‌نماید و هرچه میزان آن برای تعداد خوشه (***k***) بیشتر باشد، خوشه‌بندی از اعتبار بالاتری برخوردار است(براک،پیهور،داتتا و داتتا[6]، ۲۰۰۸). نتایج حاصل از این معیار را می‌توان در سه دسته ضعیف(کمتر از ۰.۲)، قابل قبول[7] (بین ۰.۲ تا ۰.۵) و خوب (بالای۰.۵) تقسیم‌بندی نمود(موئی و سارستد،۲۰۱۱). لازم به ذکر است که ضریب نیمرخ برای روش دو بخشی توسط نرم‌افزار SPSS محاسبه و در صفحه‌ی خروجی نمایان می‌گردد.

در نهایت در روش دو بخشی، در صورتی که تعداد خوشه‌ی مورد نظر برای محقق مشخص نباشد، پژوهشگر می‌تواند از دو معیار آکائیک(آکائیک،۱۹۷۴) و بایس(شوارز،۱۹۷۸) نیز برای انتخاب تعداد خوشه استفاده نماید. این دو معیار که به‌طور نسبی نیکویی برازش را آزمون می‌نمایند، بر اساس

1 Agglomeration schedule
2 Coefficients
3 Stage
4 Milligan and Cooper
5 Goodness of fit
6 Brock, Pihur, Datta, and Datta
7 fair

مقایسه‌ی نتایج حاصل از خوشه‌بندی با تعداد مختلف خوشه به‌دست می‌آیند و هرچه میزان نسبی این معیارها پایین‌تر باشد، نشان‌دهنده‌ی این است که تعداد خوشه‌ی در نظر گرفته شده برای داده‌های پژوهش مناسب‌تر است.

به نظر نویسندگان مقاله، توصیه می‌شود که پژوهشگران با احتیاط از شاخص‌های آماری ذکر شده استفاده نمایند، زیرا نرم افزار در صورت حتی اختلاف اندک عدد کمتر را انتخاب می نماید. بنابراین پیشنهاد می‌شود که در قسمت خروجی، جداول پایوت[1] فراخوانی شود تا مقدار معیار انتخابی (آکائیک یا بایس) برای تعداد خوشه‌های مختلف در خروجی قابل مشاهده باشد و کاربر برای مقادیر نزدیک به هم، ضریب نیمرخ را به همراه در نظر گرفتن ادبیات تحقیق و سوال پژوهش ملاک انتخاب تعداد خوشه‌ی نهایی قرار دهد. به عبارت دیگر، اگر ضریب نیم رخ برای دو و سه خوشه‌ی یکسان، معیار آکائیک برای $k=2$ کمی از $k=3$ کمتر و هدف محقق بررسی عملکرد ریاضی فراگیران بر اساس متغیرهای اضطراب و نگرش ریاضی باشد، انتخاب $k=3$ منطقی‌تر به نظر می‌رسد. زیرا پژوهشگر می‌تواند با نامگذاری صحیح خوشه‌ها بر اساس مشاهداتی که در آن قرار گرفته‌اند، فرضیات بیشتری را مورد بررسی قرار دهد. به عنوان مثال در صورتی که تعداد خوشه دو باشد، بررسی عملکرد ریاضی بر اساس اضطراب و نگرش بالا/پایین صورت می‌گیرد. درحالی‌که در صورتی‌که مشاهدات در سه خوشه قرار بگیرند، می‌توان عملکرد ریاضی را در سه گروه(به‌عنوان مثال کم، متوسط و بالا) بررسی نمود و اطلاعات بیشتری را در مورد رابطه‌ی بین متغیرهای در نظر گرفته شده برای خوشه‌بندی و متغیر هدف، کسب کرد.

بررسی اعتبار و تفسیر نتایج خوشه‌بندی

همانطور که بیان شد، الگوریتم‌های خوشه‌بندی مانند بسیاری از آنالیزهای آماری، برای هر نوع داده ورودی نتایج را فارغ از اینکه داده‌ها پیش‌فرض‌های لازم برای خوشه‌بندی را دارا می‌باشند، در اختیار کاربر قرار می‌دهند(پاستور، ۲۰۱۰). لذا بررسی اعتبار خوشه‌بندی برای جلوگیری از ایجاد نتایج نامعتبر و خوشه‌های غیر حقیقی لازم می‌باشد. در این بخش ملاحظاتی که برای استفاده از خوشه‌بندی در مطالعات آموزشی باید در نظر گرفت، به‌طور اجمال مرور می‌شود.

به پژوهشگری که علاقه‌مند است خوشه‌بندی را برای داده‌های پژوهش خود به‌کار برد، توصیه می‌شود که الگوریتم‌های مختلف خوشه‌بندی را برای داده‌های خود انجام دهد و نتایج حاصل را بر اساس معیارهای اعتبار مقایسه نماید تا بتواند با اطمینان بیشتری خوشه‌های داده‌های خود را انتخاب نماید. همچنین توصیه می‌شود که تعداد خوشه‌هایی که برای داده‌های خود در نظر می‌گیرد، تغییر دهد و نتایج حاصل را بر اساس معیارهای اعتبار مقایسه نماید(اوریت و همکاران،

1 Pivot tables

۲۰۱۱؛ کافمن و راسیو،۲۰۰۵). همچنین معیارهای اعتبار خوشه‌بندی را برای خوشه‌بندی با تعداد مختلف خوشه مثلا ۲، ۳ و ۴ خوشه بررسی نماید.

از سوی دیگر توصیه می‌شود که پس از انجام خوشه‌بندی بر روی مجموعه داده، نتایج به‌دست آمده با پژوهش‌های گذشته مقایسه شود که نتایج به‌دست آمده در راستای ادبیات تحقیق قرار گیرد (پاستور،۲۰۱۰). به‌عنوان نمونه اگر به کمک خوشه‌بندی دانش‌آموزان را در سه گروه بر اساس نمرات اضطراب و نگرش ریاضی آنها گروه‌بندی نماییم، آنالیزهای آماری نشان می‌دهد دانش‌آموزانی که در گروه اضطراب بالا و نگرش منفی قرار دارند، عملکرد ریاضی آنها در سطح دانش‌آموزان با اضطراب پایین و نگرش مثبت به ریاضی می‌باشد. نتایج حاصل از خوشه‌بندی با دقت بیشتری باید بررسی گردد، زیرا ادبیات تحقیق پیشنهاد می‌کند که نگرش ریاضی به‌طور مثبت و اضطراب به طوری منفی رابطه معناداری با عملکرد ریاضی دارا می‌باشند(ساها،۲۰۰۷؛ علم الهدائی،۲۰۰۹).

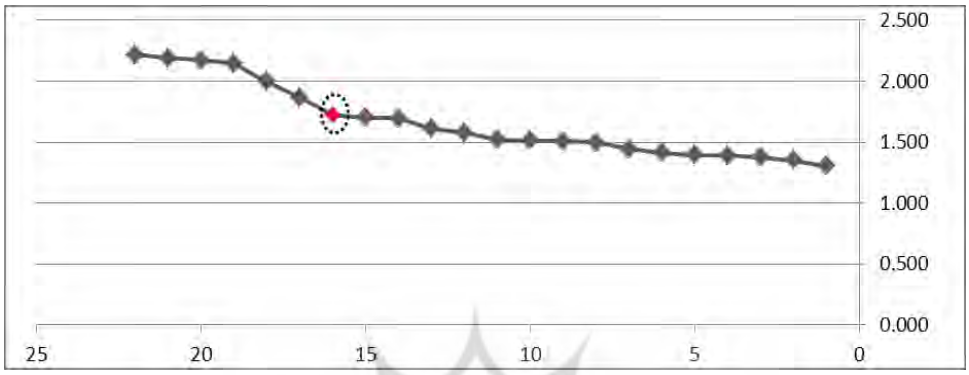
روش دیگر بررسی اعتبار خوشه بندی، تقسیم حجم نمونه به دو قسمت یا جمع‌آوری دو نمونه‌ی مستقل و انجام جداگانه‌ی خوشه‌بندی برای هر بخش و بررسی وجود تفاوت بین نتایج خوشه‌بندی دو قسمت می‌باشد(پاستور،۲۰۱۰) تا مشخص شود که نتایج خوشه‌بندی فارغ از نمونه‌های انتخابی می‌باشد. در مورد خوشه‌بندی سلسله مراتبی، توصیه می‌شود تا محقق بررسی نماید که نتایج خوشه‌بندی فارغ از ترتیب داده‌ها باشد. بدین منظور کاربر باید ترتیب داده‌ها را در مجموعه‌ی داده‌ها تغییر داده و خوشه‌بندی را مجدد انجام دهد و بررسی نماید که نتایج تفاوت معناداری با یکدیگر نداشته باشند. در صورتی که تفاوت معناداری مشاهده شود، توصیه می‌شود تا پژوهشگر اطمینان حاصل نماید که داده‌ی پرت در مجموعه داده‌ها وجود ندارد(موئی و سارستد،۲۰۱۱).

پس از بررسی اعتبار خوشه‌بندی توسط روش‌های فوق جهت تفسیر نتایج، محقق باید بررسی نماید که مرکز خوشه‌ها(میانگین تمام مشاهدات داخل هر خوشه) آیا تفاوت معناداری را بر اساس متغیرهایی که جهت خوشه‌بندی انتخاب گردیده‌اند، دارا می‌باشند. این مهم توسط انجام آزمون آنالیز واریانس یک طرفه صورت می‌پذیرد. بر اساس نتایج آن همچنین محقق می‌تواند خوشه‌های مختلف را نام گذاری نماید؛ به‌طوری که نام خوشه‌ی نمایش‌دهنده داده‌های قرار گرفته در خوشه باشد(موئی و سارستد،۲۰۱۱). به عنوان مثال اگر خوشه بندی توسط متغیر اضطراب ریاضی صورت گرفت و مراکز خوشه‌ها به‌ترتیب ۴۳ و ۵۵ و۶۷ گزارش شدند، می‌توان خوشه‌ها را به‌ترتیب به صورت اضطراب ریاضی کم، متوسط و زیاد نامگذری نمود.

نتایج تحقیق

خوشه‌بندی برای داده‌های پژوهش

خوشه‌بندی توسط پیوند تکی، وارد k-mean و دو بخشی برای داده‌های در نظر گرفته شده صورت می‌پذیرد تا ضمن کمک به فهم بیشتر گام‌های چهارم تا ششم فرایند خوشه‌بندی، یافته‌های جدیدی در زمینه‌ی روانشناسی یادگیری ریاضی برای متغیرهای در نظر گرفته شده به‌دست آید. استاندارد زانو برای الگوریتم پیوند تکی با در نظر گرفتن متر اقلیدسی تعداد k=7 را پیشنهاد می‌کند. شکل ۴ که نمایشگر ضرایب در ۲۲ مرحله پایانی است، زانو را در مرحله‌ی ۱۰۵ نشان می‌دهد. همانطور که بیان شد، برای محاسبه‌ی تعداد خوشه تعداد نمونه را از شماره‌ی مرحله



تفریق نماییم؛ لذا ۷=۱۰۵-۱۱۲. پس از ذخیره عضویت مشاهدات و تشکیل جدول فراوانی (همانطور که در ادبیات تحقیق بیان شده بود) اکثر مشاهدات در یک خوشه قرار گرفته(۱۰۶ مشاهده) و ۶ مشاهده‌ی دیگر هر کدام یک خوشه را تشکیل می‌دهند(جدول ۱).

شکل ۴- نمودار ضرایب خوشه‌بندی توسط الگوریتم پیوند تکی برای متغیرهای پژوهش

جدول ۱- میانگین متغیرهای پژوهش بر اساس خوشه‌های به‌دست آمده از الگوریتم پیوند تکی

	تعداد مشاهدات در خوشه	اضطراب ریاضی	دقت ریاضی	نگرش ریاضی	نمره آزمون سبک یادگیری	ظرفیت حافظه فعال	عملکرد ریاضی
خوشه ۱	۱۰۶	۷۷.۴۷	۸۳.۰۵	۱۷۱.۲۸	۸.۰۶	۴.۰۳	۴۳.۰۴
خوشه ۲	۱	۴۰	۶۱	۱۹۸	۱۴	۵	۳۷.۵۰
خوشه ۳	۱	۸۳	۷۲	۱۸۳	۲۰	۷	۴۰.۲۵
خوشه ۴	۱	۴۱	۹۹	۱۹۱	۱۸	۷	۵۴.۵۰
خوشه ۵	۱	۶۹	۱۰۴	۱۷۸	۱۰	۶	۶۹.۵۰
خوشه ۶	۱	۴۴	۱۰۸	۱۷۸	۱۰	۶	۷۰.۷۵
خوشه ۷	۱	۵۰	۱۱۴	۲۳۰	۱۲	۴	۷۳

۴۳.۹۱	۴.۱۰	۸.۴۳	۱۷۲.۵۱	۸۳.۵۸	۷۶.۲۴	۱۱۲	کل نمونه

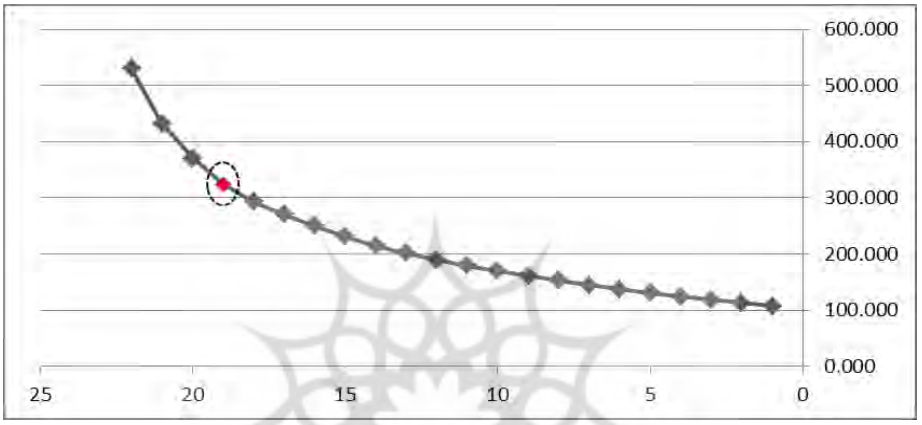
همانطور که بیان شد نتایج حاصل از خوشه‌بندی پیوند تکی برای انجام مطالعات آمیخته‌ی متوالی بسیار مناسب می‌باشد. با نگاهی به جدول ۱ می‌توان مشاهده نمود که فراگیری که در خوشه‌ی شماره‌ی ۲ قرار گرفته است، در حالی‌که نمرات آزمون ظرفیت حافظه‌ی فعال و سبک یادگیری آن بالاتر از میانگین می‌باشد و همچنین دارای اضطراب ریاضی پایین و نگرش مثبت به ریاضی می‌باشد و براساس این ۴ متغیر پیش‌بینی می‌شود که از عملکرد ریاضی مناسبی برخوردار باشد. اما هنگامی که نمره‌ی عملکرد ریاضی این دانش‌آموز را بررسی می‌نماییم، مشاهده می‌کنیم که نمره‌ی آن پایین‌تر از میانگین نمونه است و دلیل آن را می‌توان در نمره‌ی پایین دقت ریاضی او جستجو نمود. لذا با انجام مصاحبه با این دانش‌آموز می‌توان اطلاعات مفیدی در مورد تاثیر این ۵ فاکتور به‌خصوص دقت ریاضی بر عملکرد ریاضی به‌دست آورد. تحلیل‌های مشابهی را می‌توان برای هر یک از فراگیران قرار گرفته در خوشه‌های شماره ۳ تا ۷ بیان نمود و آنها را به‌خصوص در مورد متغیرهایی که سبب شده است خارج از خوشه اصلی قرار بگیرند، مورد مصاحبه قرار داد (متغیرهایی که برای فراگیر مورد نظر در جدول برجسته شده‌اند).

همانطور که قبلا بیان گردید، انتخاب محل زانو توسط کاربر صورت گرفته و ممکن است محقق دیگری برای نمودار فوق زانو را برای مرحله‌ی ۱۰۲ در نظر بگیرد و تعداد ۱۰ خوشه را انتخاب نماید و در آن صورت یک خوشه با ۱۰۲ مشاهده، یک خوشه با ۲ مشاهده و ۸ خوشه با یک مشاهده خواهیم داشت و می‌توان تفاسیر نظیر آنچه برای ۷ خوشه بیان شد، به‌دست آورد.

در مورد الگوریتم وارد نرم‌افزار SPSS توان دوم متر اقلیدسی را به‌عنوان معیار تشابه/تفاوت پیشنهاد می‌کند و لذا توان دوم آن برای خوشه‌بندی به روش وارد در نظر گرفته شد. استاندارد زانو k=4 را پیشنهاد می‌نماید(شکل ۵)[1] زیرا تفاوت قابل توجهی در مرحله‌ی ۱۰۸ دیده می‌شود(۴۵ واحد تفاوت درحالی‌که در مراحل قبلی میزان تفاوت تقریبا ۲۰ واحد بوده است) لذا ۴=۱۰۸-۱۱۲ . جدول ۲ خوشه‌های شکل گرفته به روش وارد را نشان می‌دهد. نتایج آزمون آنالیز واریانس یک طرفه برای شش متغیر جدول ۲ تفاوت معناداری را در بین خوشه‌های مختلف گزارش نمود(اضطراب، دقت، نگرش ریاضی، نمره آزمون سبک یادگیری و ظرفیت حافظه فعال:

۱-لازم به ذکر است علت رسم نمودار ضرایب برای ۲۲ مرحله نهایی به این دلیل است در نموداری که تمامی مراحل را نشان می‌دهد، نیاز به بزرگنمایی نمودار می‌باشد تا محل زانو قابل تشخیص باشد. لذا برای نمایش در مقاله معمولا مراحل پایانی فقط نمایش داده می‌شود. ولی محقق باید با دقت فراوان تمامی مراحل را بررسی نماید تا محل زانو را تشخیص دهد.

p-مقدار= کمتر از ۰.۰۰۱؛ عملکرد ریاضی:***p*** -مقدار=۰.۰۰۵) لذا خوشه‌های شکل گرفته توسط الگوریتم وارد به‌طور معناداری با یکدیگر تفاوت دارند. همچنین داده‌های پژوهش به‌طور تصادفی به دو دسته تقسیم گردید و نمودار ضرایب برای هر بخش از داده‌های تصادفی وجود تغییر در شیب نمودار و در نتیجه تشکیل زانو در k=4 را تایید کرد. لازم به ذکر است که کاربر باید دقت نماید تا در نگاه اولیه به نمودار ضرایب، غالبا تفاوت شیب در نمودار و در نتیجه زانو در مرحله‌ی نهایی (k=2) قابل تشخیص است. اما توصیه می‌شود که محقق با دقت بیشتری ضرایب قبل از k=2 را بررسی نماید تا در صورت افزایش شیب نمودار ضرایب، محل زانو در مراحل قبل از (k=2) تشخیص دهد.



شکل ۵- نمودار ضرایب خوشه‌بندی توسط الگوریتم وارد برای متغیرهای پژوهش

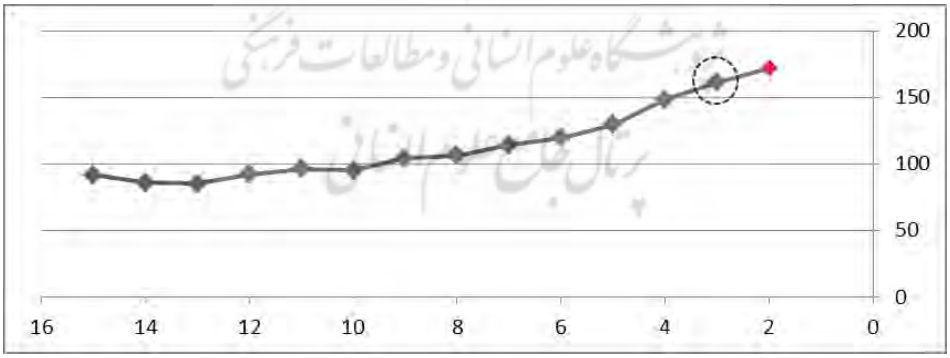
جدول ۲ اطلاعات مفیدی در مورد ارتباط ۵ متغیر در نظر گرفته شده و عملکرد ریاضی فراگیران بیان می‌نماید. به‌عنوان نمونه، فراگیران قرار گرفته در خوشه ۱، عملکرد ریاضی پایین‌تری از میانگین نمونه نشان داده‌اند که می‌توان آن را وابسته به اضطراب ریاضی بالاتر از میانگین، دقت ریاضی، نگرش ریاضی، نمره‌ی آزمون سبک یادگیری و ظرفیت حافظه فعال پایین‌تر از میانگین دانست. با انجام آزمون ***t*** برای مقایسه‌ی مقادیر متغیرهای فوق برای خوشه‌های ۱ و ۲ می‌توان مشاهده نمود که این دو خوشه تنها در دقت ریاضی و نمره آزمون سبک یادگیری تفاوت معناداری دارند(***p***-مقدار= کمتر از ۰.۰۰۱). اما این تفاوت معنادار برای فراگیران خوشه ۲ (دقت ریاضی و نمره آزمون سبک یادگیری بالاتر) سبب نشده است که به‌طور معناداری عملکرد ریاضی بالاتری نسبت به خوشه ۱ داشته باشند و داشتن اضطراب ریاضی بالا و نگرش ریاضی پایین‌تر از میانگین موجب شده که همچنان عملکرد ریاضی پایین‌تر از میانگین از خود بروز دهند. در مورد خوشه ۳ و ۴ که فراگیران این دو خوشه، عملکرد ریاضی بالاتر از میانگین و همچنین خوشه ۱ و ۲ از خود نشان داده‌اند؛ دارای اضطراب ریاضی کمتر، نگرش و دقت ریاضی بالاتر از میانگین و دو خوشه ۱ و

۲ بوده‌اند. هر چند که ظرفیت حافظه‌ی فعال و نمره‌ی آزمون سبک یادگیری برای خوشه‌ی ۴ پایین‌تر از میانگین و خوشه‌ی ۲ بوده است. لذا نتایج جدول ۲ نشان می‌دهد که اثرات منفی داشتن اضطراب ریاضی بالا، نگرش منفی به ریاضی و دقت ریاضی پایین بیشتر از داشتن ظرفیت حافظه‌ی فعال و نمره‌ی آزمون سبک یادگیری پایین است.

جدول ۲- میانگین متغیرهای پژوهش بر اساس خوشه‌های به‌دست آمده از الگوریتم وارد

	تعداد مشاهدات در خوشه	اضطراب ریاضی	دقت ریاضی	نگرش ریاضی	نمره آزمون سبک یادگیری	ظرفیت حافظه فعال	عملکرد ریاضی
خوشه ۱	۴۹	۷۹.۹۷	۷۶.۸۳	۱۶۶.۹۵	۵.۵۵	۳.۹۸	۴۱.۳۹
خوشه ۲	۴۵	۸۰.۱۷	۸۷.۸۸	۱۷۱.۳۱	۱۰.۴۲	۳.۹۸	۴۲.۷۰
خوشه ۳	۹	۶۰.۴۴	۸۲.۴۴	۱۸۶.۴۴	۱۴.۷۷	۵.۸۹	۴۹.۴۴
خوشه ۴	۹	۵۲	۱۰۰	۱۹۴	۷.۸۸	۳.۵۶	۵۵.۵۵
کل نمونه	۱۱۲	۷۶.۲۴	۸۳.۵۸	۱۷۲.۵۱	۸.۴۳	۴.۱۰	۴۳.۹۱

الگوریتم سومی که اعمال گردید، الگوریتم k-mean می‌باشد. این الگوریتم برای k=2 الی k=15 اجرا شده و با فراخوانی جدول ***ANOVA*** استاندارد نسبت واریانس (***VRC***) نیز محاسبه گردیده است(شکل ۶). جهت انتخاب تعداد خوشه، ***k*** نظیر ***VRC*** ماکزیمم را مد نظر قرار می‌دهیم که با توجه به شکل ۶، بالاترین مقادیر ***VRC*** به‌ترتیب متناظر با دو و سه خوشه می‌باشد. همچنین پس از تقسیم تصادفی داده‌ها به دو بخش، در یک بخش بالاترین ***VRC*** به‌ترتیب مربوط به دو و سه و در بخش دیگر به‌ترتیب متناظر به سه و دو خوشه بوده است. در نتیجه می‌توان نتیجه گرفت به روش k-mean تعداد خوشه‌ها برای مجموعه داده‌ی حاضر باید k=2 یا k=3 در نظر گرفته شود. جداول ۳ و ۴ میانگین متغیرهای پژوهش بر حسب خوشه‌های به‌دست آمده از الگوریتم k-mean برای k=2 و k=3 را نشان می‌دهد.



شکل ۶- نمودار مقادیر *VRC* برای k=2 الی15=k برای متغیرهای پژوهش

نتایج آزمون ***t*** تفاوت معنادار بین فراگیران قرار گرفته در دو خوشه‌ی جدول ۳ در ارتباط با متغیرهای پژوهش را نشان می‌دهد(***p***-مقدار:کمتر از ۰.۰۰۱). لذا خوشه‌های شکل گرفته توسط الگوریتم ***k-mean*** تفاوت معناداری با یکدیگر دارند. در مورد رابطه‌ی ۵ ، متغیر در نظر گرفته شده و عملکرد ریاضی همان‌طور که مشاهده می‌شود؛ فراگیران قرار گرفته در خوشه‌ی ۱ به‌طور معناداری دارای اضطراب ریاضی بالاتر، نگرش ریاضی منفی‌تر، دقت ریاضی، ظرفیت حافظه‌ی فعال و نمره‌ی آزمون سبک یادگیری پایین‌تری نسبت به فراگیران خوشه‌ی ۲ دارا می‌باشند. لذا فاکتورهای فوق سبب گردیده است که دانش‌آموزان قرار گرفته در خوشه‌ی ۲ به‌طور معناداری عملکرد ریاضی بالاتری نسبت به خوشه‌ی ۱ از خود بروز نمایند.

جدول ۳- میانگین متغیرهای پژوهش بر اساس خوشه‌های به‌دست آمده از الگوریتم *k-mean* برای *k=2*

	تعداد مشاهدات در خوشه	اضطراب ریاضی	دقت ریاضی	نگرش ریاضی	نمره آزمون سبک یادگیری	ظرفیت حافظه فعال	عملکرد ریاضی
خوشه ۱	۶۷	۸۳.۳۴	۷۹.۱۰	۱۶۶.۷۳	۶.۶۴	۳.۸۲	۴۰.۳۴
خوشه ۲	۴۵	۶۷.۱۵	۹۰.۲۶	۱۸۲.۶۲	۱۱.۱۱	۴.۵۱	۴۹.۰۵
کل نمونه	۱۱۲	۷۶.۲۴	۸۳.۵۸	۱۷۲.۵۱	۸.۴۳	۴.۱۰	۴۳.۹۱

آزمون آنالیز واریانس یک طرفه تفاوت معناداری را برای متغیرهای پژوهش برحسب سه خوشه‌ی به‌دست آمده از الگوریتم ***k-mean*** (جدول۴) گزارش می‌کند(اضطراب، دقت، نگرش ریاضی، نمره آزمون سبک یادگیری و ظرفیت حافظه‌ی فعال: ***p***-مقدار= کمتر از ۰.۰۰۱؛ عملکرد ریاضی: ***p***-مقدار=۰.۰۰۱) که مبین تفاوت معنادار این سه خوشه با یکدیگر می‌باشد. فراگیران خوشه‌ی ۱ که عملکرد ریاضی پایین‌تر از میانگین و دو خوشه‌ی دیگر دارند، دارای ظرفیت حافظه‌ی فعال، نمره‌ی آزمون سبک یادگیری، دقت و نگرش ریاضی پایین‌تر از میانگین می‌باشند و همچنین از اضطراب ریاضی بالاتر از میانگین نیز رنج می‌برند. لذا ۵ فاکتور انتخاب شده به خوبی علت پایین بودن عملکرد ریاضی آنها را توجیه می‌نماید. در مورد خوشه‌های ۲ و۳ ، آزمون ***t*** نشان داد که این دو گروه تفاوت معناداری با یکدیگر در ارتباط با متغیرهای دقت ریاضی، نگرش ریاضی، نمره‌ی آزمون سبک یادگیری و حافظه‌ی فعال نشان دادند. دقت ریاضی و نگرش ریاضی برای فراگیران خوشه‌ی ۳ بیشتر از فراگیران خوشه‌ی ۲ و ظرفیت حافظه‌ی فعال و نمره‌ی آزمون سبک یادگیری

برای فراگیران خوشه‌ی ۲ بالاتر از فراگیران خوشه‌ی ۳ بوده است. همان‌طور که نتایج خوشه‌بندی به روش وارد نشان داد، تاثیر دقت و نگرش ریاضی بر روی عملکرد ریاضی بیشتر از نمره‌ی آزمون سبک یادگیری و ظرفیت حافظه فعال بوده است و این امر در تفاوت ۴ واحدی در نمره‌ی عملکرد ریاضی این دو خوشه قابل مشاهده است(هر چند که این تفاوت معنادار نمی‌باشد).

جدول ۴- میانگین متغیرهای پژوهش بر اساس خوشه‌های به‌دست آمده از الگوریتم *k-mean* برای $k=3$

	تعداد مشاهدات در خوشه	اضطراب ریاضی	دقت ریاضی	نگرش ریاضی	نمره آزمون سبک یادگیری	ظرفیت حافظه فعال	عملکرد ریاضی
خوشه ۱	۵۴	۸۴.۱۱	۷۶.۵۵	۱۶۰.۸۳	۶.۷۵	۳.۸۹	۳۹.۶۳
خوشه ۲	۱۸	۶۷.۲۷	۸۲.۸۸	۱۷۷.۳۸۸	۱۳.۶۱	۵.۲۸	۴۵.۵۵
خوشه ۳	۴۰	۶۹.۶۵	۹۳.۴۰	۱۸۶.۱۰۰	۸.۳۷	۳.۸۵	۴۹.۲۵
کل نمونه	۱۱۲	۷۶.۲۴	۸۳.۵۸	۱۷۲.۵۱	۸.۴۳	۴.۱۰	۴۳.۹۱

آخرین روشی که بر روی داده‌های پژوهش حاضر اعمال شده‌ی الگوریتم دو بخشی با معیارهای متر اقلیدسی و لگاریتم نسبت درست‌نمایی می‌باشد، روش دو بخشی را برای k=2 الی k=15 بر روی داده‌های استاندارد شده‌ی پژوهش اجرا نمودیم و نرم‌افزار ***SPSS*** برای متر اقلیدسی بر اساس معیار اعتبار ***AIC*** داده‌ها را در دو خوشه قرار داده است و ضریب نیمرخ برابر ۰.۴ گزارش شده است که نشان می‌دهد داده‌ها به‌طور قابل در دو خوشه برازش شده است. میانگین متغیرهای پژوهش بر اساس خوشه‌بندی دو بخشی توسط متر اقلیدسی در جدول ۵ و نمودار مقادیر ***AIC*** برای k=2 الی k=15 برای متغیرهای پژوهش در شکل ۷ گزارش شده است. همانطور که در شکل نشان داده شده است، مقادیر ***AIC*** برای ۲ الی ۴ خوشه بسیار به یکدیگر نزدیک است. اما پس از محاسبه‌ی ضریب نیمرخ برای این سه خوشه، بالاترین ضریب نیمرخ برای دو خوشه به‌دست آمد و لذا نتایج برای دو خوشه گزارش شده است.

شکل ۷- نمودار مقادیر *AIC* برای k=2 الی k=15 توسط روش دوبخشی (توسط متر اقلیدسی)

جدول ۵- میانگین متغیرهای پژوهش بر اساس خوشه‌های به‌دست آمده از روش دو بخشی برای k=2

	تعداد مشاهدات در خوشه	اضطراب ریاضی	دقت ریاضی	نگرش ریاضی	نمره آزمون سبک یادگیری	ظرفیت حافظه فعال	عملکرد ریاضی
خوشه ۱	۱۳	۵۸.۸۴	۸۵.۹۲	۱۸۷۰۸۴	۱۳.۴۶	۵.۵۴	۵۳.۲۶
خوشه ۲	۹۹	۷۸.۵۲	۸۳.۲۸	۱۷۰.۵۰	۷.۷۷	۳.۹۱	۴۲.۵۱
کل نمونه	۱۱۲	۷۶.۲۴	۸۳.۵۸	۱۷۲.۵۱	۸.۴۳	۴.۱۰	۴۳.۹۱

نتایج آزمون ***t*** نشان داد که دو خوشه‌ی فوق در تمامی متغیرهای پژوهش به استثناء دقت ریاضی (***p***-مقدار=۰.۴۴۷) با یکدیگر تفاوت معنادار دارند(اضطراب ریاضی، نمره‌ی آزمون سبک یادگیری، ظرفیت حافظه فعال: ***p***-مقدار: کمتر از ۰.۰۰۱، نگرش ریاضی: ***p***-مقدار=۰.۰۰۱ و عملکرد ریاضی: ***p***-مقدار=۰.۰۰۳) که نشان‌دهنده‌ی این مطلب است که خوشه‌های شکل گرفته به‌طور معنادار با یکدیگر تفاوت دارند. بر اساس جدول فوق می‌توان مشاهده کرد که ۱۳ نفر قرار گرفته در خوشه‌ی ۱ به علت داشتن تفاوت معنادار با خوشه‌ی ۲ در نگرش ریاضی، ظرفیت حافظه‌ی فعال، نمره‌ی آزمون سبک یادگیری و اضطراب ریاضی، به‌طور معناداری عملکرد ریاضی بالاتری نسبت به خوشه‌ی ۲ از خود نشان داده‌اند. لازم به ذکر است که همچنین دقت ریاضی خوشه‌ی ۱ به‌طور غیر معناداری بالاتر از خوشه‌ی ۲ می‌باشد.

روش دو بخشی توسط معیار لگاریتم نسبت درست‌نمایی برای k=2 الی k=15 بر اساس معیار اعتبار ***AIC*** داده‌ها را در چهار خوشه قرار داده و ضریب نیمرخ برابر آن ۰.۳ گزارش شده است که نشان می‌دهد داده‌ها به‌طور قابل قبولی در ۴ خوشه برازش شده‌اند. میانگین متغیرهای پژوهش در جدول ۶ و نمودار مقادیر ***AIC*** برای k=2 الی k=15 در شکل ۸ گزارش شده است. آزمون آنالیز واریانس یک طرفه تفاوت معناداری را در ارتباط با میانگین متغیرهای پژوهش بر اساس خوشه‌های شکل گرفته گزارش نمود(***p***-مقدار: کمتر از ۰.۰۰۱ برای تمامی متغیرها) لذا خوشه‌های شکل گرفته توسط این الگوریتم نیز تفاوت معناداری با یکدیگر دارا می‌باشند.

شکل ۸- نمودار مقادیر *AIC* برای *k=2* الی *k=15* توسط روش دوبخشی (توسط معیار لگاریتم نسبت درست نمایی)

جدول ۶- میانگین متغیرهای پژوهش بر اساس خوشه‌های به‌دست آمده از روش دو بخشی برای k=4 الی k=15 (معیار لگاریتم نسبت درست نمایی)

	تعداد مشاهدات در خوشه	اضطراب ریاضی	دقت ریاضی	نگرش ریاضی	نمره آزمون سبک یادگیری	ظرفیت حافظه فعال	عملکرد ریاضی
خوشه ۱	۳۹	۸۳.۲۰	۷۳.۸۲	۱۶۰.۱۲	۵.۴۸	۳.۹۷	۳۹.۶۱
خوشه ۲	۳۳	۸۲.۵۴	۸۶.۷۲	۱۶۷.۰۹	۱۰.۶۰	۳.۹۴	۴۰.۲۵
خوشه ۳	۱۰	۶۰.۱۰	۸۲.۱۰	۱۸۴.۸۰	۱۴.۴۰	۵.۸۰	۴۹.۷۵
خوشه ۴	۳۰	۶۵.۶۳	۹۳.۳۳	۱۹۰.۵۰	۷.۹۰	۳.۸۷	۵۱.۵۳
کل نمونه	۱۱۲	۷۶.۲۴	۸۳.۵۸	۱۷۲.۵۱	۸.۴۳	۴.۱۰	۴۳.۹۱

با توجه به جدول ۶ ، دو خوشه‌ی ۱ و ۲ که دارای عملکرد ریاضی کمتر از میانگین می‌باشند، دارای ظرفیت حافظه فعال، نگرش ریاضی پایین‌تر از میانگین می‌باشند. همچنین دارای اضطراب ریاضی بالاتر از میانگین می‌باشند. در ارتباط با خوشه‌ی ۱، دقت ریاضی و نمره‌ی آزمون سبک یادگیری نیز از میانگین نمونه پایین‌تر است و سبب شده تا فراگیران در این گروه پایین‌ترین عملکرد ریاضی را از خود نشان دهند. در خوشه ۲ هر چند که دقت و نمره‌ی آزمون سبک یادگیری بالاتر از میانگین است، ولی همچنان به‌علت تاثیر ۳ متغیر دیگر همچنان عملکرد ریاضی پایین‌تر از میانگین می‌باشد. در مورد خوشه‌های ۳ و ۴ ، آزمون ***t*** نشان می‌دهد که این دو خوشه تفاوت معناداری در دقت ریاضی، نمره‌ی آزمون سبک یادگیری و ظرفیت حافظه‌ی فعال دارند. بدین صورت که ظرفیت حافظه‌ی فعال و نمره‌ی آزمون سبک یادگیری خوشه‌ی ۳ بالاتر از خوشه‌ی ۴ می‌باشد و از طرف دیگر دقت ریاضی خوشه‌ی ۴ بالاتر از خوشه‌ی ۳ می‌باشد. اما عملکرد ریاضی خوشه‌ی ۴ به‌طور غیر معناداری از خوشه‌ی ۳ بیشتر است. نکته‌ی جالب توجه این است که همانطور که از نتایج خوشه‌بندی به روش‌های دیگر و این روش استنباط می‌شود، تاثیر عوامل نگرش و دقت ریاضی بر عملکرد ریاضی بیشتر از نمره‌ی آزمون سبک یادگیری و ظرفیت حافظه‌ی فعال می‌باشد.

در مورد تقسیم نمونه به دو بخش و بررسی پایداری نتایج حاصل از خوشه‌بندی توسط اجرا مجدد خوشه‌بندی برای هر بخش، در مورد متر اقلیدسی در هر دو بخش نتایج همانند نتایج حاصل از خوشه‌بندی بر روی کل نمونه بوده است. در مورد لگاریتم نسبت درست‌نمایی در هر کدام از بخش‌ها با توجه به معیار ***AIC***، k=2 را الگوریتم دو بخشی پیشنهاد می‌کند؛ ولی اختلاف اندکی بین ***AIC*** k=2 و k=4 وجود دارد(تقریبا ۴ واحد) و ضریب نیمرخ برای k=4 نیز در محدوده خوشه‌بندی قابل قبول قرار گرفته است(۰.۴). لذا نتایج حاصل از خوشه‌بندی به روش لگاریتم

نسبت درست‌نمایی نیز تا حدودی پایدار می‌باشد و می‌توان k=4 را برای داده‌های این پژوهش انتخاب نمود.

بحث و نتیجه‌گیری

در این مطالعه خوشه‌بندی به‌عنوان ابزاری برای آنالیز داده‌ها در پژوهش‌های کمی و آمیخته‌ی حوزه‌ی مطالعات آموزشی و روانشناختی معرفی گردید. همانطور که بیان شد، گروه‌بندی داده‌های پژوهش در مطالعات قبلی در ایران در حوزه‌های آموزشی و روانشناختی توسط میانگین، میانه و برد صورت گرفته است که دارای نواقص و محدودیت‌هایی می‌باشد. لذا خوشه‌بندی به‌عنوان ابزاری جایگزین با قابلیت‌های متعدد در مطالعات کمی و آمیخته معرفی گشته و مراحل ششگانه‌ی اجرای فرایند خوشه‌بندی توسط مثال‌های گوناگون بیان گردید.

در این مطالعه دو کاربرد خوشه‌بندی برای مطالعات آموزشی و روانشناختی(شناسایی کاندیداهای مصاحبه در مطالعات آمیخته و بررسی رابطه بین متغیرهای مستقل و متغیر وابسته در مطالعات کمی) به نمایش گذاشته و بر روی داده‌هایی مرتبط با روانشناسی یادگیری ریاضی به‌کار گرفته شد. نتایج خوشه‌بندی در راستای تحقیقات گذشته در این حوزه بود که عملکرد ریاضی رابطه‌ی معنادار مثبتی با نمره‌ی آزمون سبک یادگیری، ظرفیت حافظه فعال، دقت ریاضی و نگرش ریاضی و همچنین رابطه‌ی منفی با اضطراب ریاضی دارد(امانی، علم الهدائی و رادمهر، ۲۰۱۱؛ ساها، ۲۰۰۷؛ علم الهدائی، ۲۰۰۹؛ موسوی، رادمهر، و علم الهدائی، ۲۰۱۲). علاوه بر تایید مطالب فوق خوشه‌بندی داده‌های پژوهش حاضر نتایج جدیدی را در این حوزه پیشنهاد نمود. به‌عنوان نمونه نشان داد که تاثیر منفی داشتن اضطراب ریاضی بالا و نگرش ریاضی منفی بیشتر از تاثیر منفی ظرفیت حافظه‌ی فعال پایین و نمره‌ی آزمون سبک‌های یادگیری پایین بر عملکرد ریاضی فراگیران می‌باشد.

لازم به ذکر است که کاربردهای خوشه‌بندی برای مطالعات آموزشی و روانشناختی به مواردی که در مقاله‌ی حاضر معرفی گردید، محدود نمی‌شود و کاربردهای دیگری همچون خوشه‌بندی مفاهیم(شاولسون، ۱۹۷۹) و سبک‌های آموزش(بنت، ۱۹۷۵) در ادبیات تحقیق اشاره شده است که توصیه می‌شود در مطالعات دیگر به آنها پرداخته شود.

تشکر و قدرانی

نویسندگان مقاله از خانم مریم حاجی بابا جهت در اختیار قرار دادن داده‌های پایان نامه‌ی ارشدشان کمال تشکر را دارند.

منابع

کشاورزی، امین، پیروی، فرزاد و رحمانی، امیرمسعود(۱۳۸۹) ، استفاده از تکنیک خوشه بندی در جداسازی رکوردهای داده صفحات وب نیمه ساختیافته، چهارمین کنفرانس داده کاوی، تهران.

حسینی، روح الله (۱۳۹۱)، مباحثی در داده کاوی، پایان نامه کارشناسی ارشد، دانشگاه فردوسی مشهد، مشهد، ایران.

Abonyi, J., & Feil, B. (2007). Cluster analysis for data mining and system identification. Springer.

Akaike, H. (1974). A new look at the statistical model identification. Automatic Control, IEEE Transactions on, 19(6), 716-723.

Alamolhodaei, H. (2009). A working memory model applied to mathematical word problem solving. Asia Pacific Education Review, 10(2), 183-192.

Amani, A., Alamolhodaei, H. & Radmehr, F. (2011). A gender study on predictive factors of mathematical performance of University students. Educational Research. 2(6), 1179˚ 1192.

Bennett, N. (1975). Cluster analysis in educational research: a non-statistical introduction. Research Intelligence, 1(1), 64-70.

Brock, G., Pihur, S., Datta, S., & Datta, S. (2008). ClValid: an R package for cluster validation. Journal of Statistical Software, 25(4), 1-22.

Cali ski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. Communications in Statistics-theory and Methods, 3(1), 1-27.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). Cluster Analysis: Wiley Series in Probability and Statistics (5th edition).Wiley.

Formann, A. K. (1984). Die Latent-Class-Analyse: Einführung in Theorie und Anwendung. Beltz.

Hajibaba, M., Radmehr, F., & Alamolhodaei, H. (2013). A psychological model for mathematical problem solving based on revised Bloom taxonomy for high school girl students. Journal of the Korean Society of Mathematical Education Series D: Research in Mathematical Education. Vol. 17, No. 3, 199˚ 220.

Hattie (2002). Schools Like Mine: Cluster Analysis of New Zealand Schools. Technical Report 14, Project asTTle. University of Auckland.

Johnson, R. A., & Wichern, D. W. (2007). Applied multivariate statistical analysis. Upper Saddle River, NJ: Prentice hall.

Kaufman, L., & Rousseeuw, P. J. (2005). Finding groups in data: an introduction to cluster analysis. Hoboken, NJ: Wiley.

Ketchen, D. J., & Shook, C. L. (1996). The application of cluster analysis in strategic management research: an analysis and critique. Strategic Management Journal, 17(6), 441-458.

Krantz, A., Korn, R., & Menninger, M. (2009). Rethinking museum visitors: Using K- means cluster analysis to explore a museum's audience. Curator: The Museum Journal, 52(4), 363-374.

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. Psychometrika, 50(2), 159-179.

Mooi, E., & Sarstedt, M. (2011). A concise guide to market research: The process, data, and methods using IBM SPSS statistics. Springer.

Mousavi, S., Radmehr, F., & Alamolhodaei, H. (2012). The role of mathematical homework and prior knowledge on the relationship between students mathematical performance, cognitive style and working memory capacity. Electronic Journal of Research in Educational Psychology, 10(3), 1223-48.

Oyelade, O. J., Oladipupo, O. O., & Obagbuwa, I. C. (2010). Application of k-means clustering algorithm for prediction of students academic performance. International Journal of Computer Science and Information Security, 7(1), 292-295.

Pastor, D. A. (2010). Cluster Analysis. In G. R. Hancock & R. O. Mueller (Eds.), The reviewer s guide to quantitative methods in the social sciences. (pp.41-54). Routledge.

Raghubar, K. P., Barnes, M. A., & Hecht, S. A.(2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. Learning and Individual Differences 20(2), 110˚122.

Rock, D. A., Baird, L. L., & Linn, R. L. (1972). Interaction between college effects and students' aptitudes. American Educational Research Journal, 149-161.

Rousseeuw, P. J. (1987). Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. Journal of Computational and Applied Mathematics, 20, 53-65.

Saha, S. (2007). A study of Gender Attitude to Mathematics, Cognitive style and Achievement in mathematics. Experiments in Education. 35, 61˚67.

Schwarz, G. (1978). Estimating the dimension of a model. The annals of statistics, 6(2), 461-464.

Shavelson, R. J. (1979). Applications of cluster analysis in educational research: looking for a needle in a haystack. British Educational Research Journal, 5(1), 45-53.

Tan, P., Steinbach, M., & Kumar, V. (2006). Introduction to Data Mining. Boston: Pearson Addison-Wesley.

Thorndike, R. L. (1953). Who belongs in the family?. Psychometrika, 18(4), 267-276.

Velmurugan, T., & Santhanam, T. (2010). Computational complexity between K-means and K-medoids clustering algorithms for normal and uniform distributions of data points. Journal of Computer Science, 6(3), 363-368.

Williams, G. (2011). Data mining with Rattle and R: the art of excavating data for knowledge discovery. Springer.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی