مجله کشاورزی، جلد ۶، شماره ۲، سال ۱۳۸۳، صص ۲۶-۱۳

استفاده از روش خوشه‌بندی در تعیین همگنی هیدرولوژیک و ارزیابی آن توسط روشهای تحلیل ممیزی و منحنی‌های Andrew در حوضه آبریز کرخه

منیره بیابانکی * و سید سعید اسلامیان **

چکیده

در موارد زیادی داده‌های کافی از متغیرهای هیدرولوژیک برای تحلیل فراوانی وجود ندارد. این امر باعث می‌شود تا متخصصین هیدرولوژی از تحلیل منطقه‌ای استفاده کنند. از داده‌های نمونه برای برازش با توزیع‌های احتمال استفاده می‌شود و از تئوری‌های احتمالات برای جامعه استنتاج آماری می‌شود. نتایج حاصل از تحلیل منطقه‌ای درصورتی معتبر خواهد بود که منطقه مورد مطالعه از نظر عوامل اقلیمی، فیزیوگرافی، ویژگی‌های خاک و غیره همگن باشد. بنابراین قبل از استفاده روش تحلیل منطقه‌ای باید از همگنی منطقه اطمینان حاصل شود. در این تحقیق با استفاده از ۱۶ خصوصیت هیدرولوژیک از ۴۱ ایستگاه هیدرومتری در حوضه آبریز کرخه و استفاده از روش تحلیل عاملی و انتخاب عوامل فیزیوگرافی مستقل با روش خوشه‌بندی مناطق همگن انجام شد و سپس با استفاده از روش تحلیل ممیزی و منحنی‌های Andrew روش گروه‌بندی آزمون شد. نتایج نشان داد که با استفاده از روش تحلیل عاملی پنج عامل سطح حوضه، قطر دایره معادل، ارتفاع متوسط حوضه، شیب متوسط حوضه و نسبت انشعاب می‌توانند به عنوان عوامل فیزیوگرافی مستقل معرفی شوند. همچنین گروه‌بندی به روش الگوریتم خوشه‌ای با انتخاب فاصله اقلیدسی ۱۰ و روش گروه‌بندی Ward، شش زیرحوضه هولیان (سیمره)، تنگ‌سازبن، چم‌ژاب، جلوگیر (ماژین)، چم‌گز و پای‌پل از کل ۴۱ زیرحوضه مورد استفاده در تحقیق به عنوان یک گروه همگن جداگانه معرفی شد. منحنی‌های Andrew و تحلیل ممیزی نیز دلالت بر این گروه‌بندی دارد و درنتیجه منطقه به دو گروه همگن که شامل ۳۵ و شش زیرحوضه می‌باشند تقسیم‌بندی می‌شود.

واژه‌های کلیدی : الگوریتم خوشه‌ای؛ تحلیل عاملی؛ تحلیل ممیزی؛ حوضه آبریز کرخه؛ همگنی منطقه‌ای

---

* - کارشناس ارشد آبیاری، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه صنعتی اصفهان، اصفهان - ایران

** - دانشیار گروه مهندسی آب، دانشکده کشاورزی، دانشگاه صنعتی اصفهان، اصفهان - ایران

مقدمه

در تحلیل منطقه‌ای با استفاده از آمار ایستگاه‌های هیدرومتری موجود در منطقه موردنظر، روابط یا منحنی‌هایی برای کل منطقه حاصل می‌شود. با داشتن بعضی از ویژگی‌های حوضه در محل‌هایی‌که اندازه‌گیری جریان انجام نمی‌شود، مقدار جریان در محل موردنظر را می‌توان تعیین کرد. در واقع تحلیل منطقه‌ای شامل دو بخش است. در بخش نخست حوضه‌های همگن تعیین شده و در بخش دوم مدل‌های تحلیل منطقه‌ای برای مناطق همگن مشخص می‌شوند.

نخستین مرحله در تحلیل منطقه‌ای بررسی همگنی حوضه‌های آبریز می‌باشد. باتوجه به این‌که حوضه‌ها از لحاظ خصوصیات اکولوژیک دارای تفاوت‌هایی می‌باشند، باید در ابتدا وضعیت هیدرولوژیک آنها بررسی شود و حوضه‌های مورد مطالعه درقالب گروه‌های همگن از یکدیگر تفکیک شوند. روشهای متعددی برای بررسی همگنی حوضه‌های آبریز ارایه شده است که مبتنی بر یک یا تعدادی از خصوصیات حوضه‌های آبریز (نظیر خصوصیات زمین‌شناسی، اقلیمی، فیزیوگرافی و هیدرولوژیک) می‌باشد. قبل از تفکیک منطقه موردنظر به مناطق همگن می‌توان خصوصیات فیزیوگرافی مستقل حوضه‌های آبریز را با استفاده از روشهایی (نظیر تجزیه عاملی) تعیین نمود (۱).

به‌طورکلی در مجموع بررسی همگنی حوضه‌ها بر مبنای آزمون تحلیل خوشه‌ای [1] می‌باشند و حوضه‌های آبریز در فاصله اقلیدسی پارامترهای اکولوژیک از یکدیگر متمایز می‌شوند.

تفکیک منطقه مطالعاتی به مناطق همگن باعث دقت بیشتر و خطای کمتر در مدل‌های رگرسیون در مناطق فاقد آمار می‌شود. در یک تحقیق برای مطالعه ۲۲۱ ایستگاه آب‌سنجی در ایالت آریزونای آمریکا برای بررسی مناطق همگن از روش تحلیل خوشه‌ای استفاده شد. گروه‌بندی حوضه‌های مورد مطالعه بر اساس ویژگی‌های هیدرولوژیک منطقه و همچنین بر اساس ویژگی‌های فیزیوگرافی حوضه‌های مطالعاتی انجام شد (۱۰).

در یک تحقیق برای بررسی اهمیت تعیین مناطق بر اساس خصوصیات حوضه نسبت به تعیین مناطق همگن با روش تحلیل خوشه‌ای و بر اساس عواملی نظیر مساحت حوضه، تراکم آبراهه، شاخص بارندگی، طول آبراهه اصلی، شیب آبراهه اصلی و عامل کمبود رطوبت خاک در ۱۶۸ حوضه مطالعاتی اقدام شد. در این مطالعه از روش Ward (۵) و برای تعیین دسته‌های همگن از نسبت درست‌نمایی استفاده شد. نتایج نشان داد که در بین پنج گروه مورد مطالعه تعداد چهار گروه قابل قبول و تغییرات دبی را توجیه نمود (۵). از روش الگوریتم خوشه‌ای نیز برای شناخت

---

1 - Cluster Analysis

مناطق همگن استفاده شده است (۷).

پس از تعیین مناطق همگن می‌توان با استفاده از روشهایی مانند تحلیل ممیزی و منحنی‌های Andrew از صحت گروه‌بندی اطمینان حاصل نمود. روش منحنی‌های Andrew را می‌توان به عنوان یک روش جداگانه برای تفکیک منطقه مطالعاتی به مناطق همگن نیز مورد استفاده قرار داد. در این روش یک نقطه در فضای چندبعدی از پارامترها به وسیله یک منحنی دوبعدی نشان داده می‌شود. اختلاف بین دو منحنی با فاصله اقلیدسی بین پارامترها متناسب است (۶). از این روش برای ارزیابی شباهت بین ۱۲۰ حوضه واقع در شمال شرق ایالت ویکتوریای استرالیا استفاده شده است. برای هر حوضه نه پارامتر (سطح حوضه، ارتفاع متوسط حوضه، مساحت پوشیده از جنگل، طول آبراهه اصلی، شیب آبراهه اصلی، تراکم آبراهه، ضریب شکل حوضه، میانگین بارندگی سالانه و ضریب تغییرات بارندگی ماهانه) درنظر گرفته شده و با استفاده از رگرسیون چندگانه متغیرهای مهم مرتبط با دبی تعیین شد (۸).

در تحقیق حاضر با انتخاب ۱۶ پارامتر هیدرولوژیک و تعیین عوامل فیزیوگرافی مستقل با استفاده از روش تحلیل عاملی، از روش تحلیل خوشه‌ای برای تقسیم‌بندی ۴۱ زیرحوضه موجود در حوضه آبریز کرخه به مناطق همگن استفاده شد. سپس با روشهای تحلیل ممیزی و منحنی‌های Andrew صحت گروه‌بندی‌ها آزمون شد.

مواد و روش

خصوصیات حوضه آبریز

حوضه آبریز رودخانه کرخه در غرب کشور، در مناطق میانی و جنوبی کوه‌های زاگرس قرار دارد. این حوضه از نظر مشخصات جغرافیایی بین ′۵ °۴۶ الی ′۱۰ °۴۹ طول شرقی و ′۸ °۳۰ الی ′۵۶ °۳۴ عرض شمالی قرار گرفته است. مساحت این حوضه در داخل ایران حدود ۵۰۷۶۸ کیلومتر مربع می‌باشد که حدود ۳۳۶۷۴ کیلومتر مربع آن را کوهپایه و حدود ۱۷۰۹۴ کیلومتر مربع آن را دشت‌ها تشکیل داده است.

حوضه آبریز کرخه از نظر تقسیم‌بندی هیدرولوژیک ایران جزیی از حوضه آبریز خلیج فارس به شمار می‌رود و به دو قسمت کرخه علیا و کرخه سفلی تقسیم شده است. محدوده مورد تحقیق حوضه کرخه تا محل پای پل است که وسعت آن حدود ۴۲۹۶۴ کیلومتر مربع می‌باشد.

رودخانه کرخه مجموع رودخانه‌های اصلی گاماسیاب، قره‌سو، سیمره و کشکان به وجود می‌آید که هر یک از آنها دارای حوضه‌های آبریز هستند که زیرحوضه‌های اصلی کرخه را تشکیل می‌دهند.

به‌طورکلی مطالعات حوضه آبریز کرخه از دیرزمان به سبب وسعت حوضه، زیاد بودن دبی پایه و کنترل منابع آب سطحی با اهداف گوناگون مورد توجه بوده است. در این تحقیق از ۱۶ پارامتر فیزیوگرافی ۴۱ ایستگاه هیدرومتری در حوضه آبریز کرخه استفاده شده است که واحد و علایم آنها به شرح زیر است :

area : مساحت (کیلومتر مربع)، wsa : شیب متوسط حوضه (درصد)، rsa : شیب آبراهه اصلی (درصد)، dr : تراکم زهکشی، br : نسبت انشعاب، pay : بارندگی متوسط سالانه (میلی‌متر)، pyram : محیط حوضه (کیلومتر)، hm : ارتفاع متوسط حوضه (متر)، li : طول آبراهه اصلی (کیلومتر)، sumli : مجموع طول آبراهه‌ها، dm : قطر دایره معادل (کیلومتر)، rc : ضریب میلر، graw : ضریب گراولیوس، t_c : زمان تمرکز (ساعت)، lm : طول مستطیل معادل (کیلومتر) و bm : عرض مستطیل معادل (کیلومتر) (جدول ۱)

نام و مشخصات جغرافیایی ایستگاه‌های مورد استفاده در جدول (۱) پیوست ارایه شده است.

تحلیل عاملی

یک روش آماری است که بین چندین متغیر به ظاهر بی‌ارتباط، رابطه خاصی را تحت یک مدل فرضی برقرار می‌کند. یکی از اهداف اصلی این روش، کاهش ابعاد داده‌ها است. فرض اساسی در این روش، وجود یک الگوی زیربنایی یا مدل خاص برای تعیین مفاهیم پیچیده ارتباطی بین متغیرها می‌باشد. در این مدل فرضی این ارتباط درقالب یک عامل ارایه می‌شود (۱۱).

مدل تحلیل عاملی در حالت کلی برای P عامل و m متغیر ($P \leq m$) به صورت زیر است:

$$Y_i = \mu_i + \sum_{j=1}^{P} \lambda_{ij} f_j + e_i \quad i=1, 2, \ldots, m \quad (۱)$$

در این فرمول، μ_i میانگین متغیر i ام کلیه مشاهدات، λ_{ij} ضریب عامل j ام در ارتباط با متغیر i ام (یا درحقیقت کوواریانس بین متغیر i ام و عامل j ام)، f_i عوامل مؤثر بر متغیر وابسته و e_i اثر عوامل خطا می‌باشد که فرض می‌شود همبستگی بین آنها صفر و دارای توزیع نرمال می‌باشند. از ماتریس دورانی واریماکس که از ضرایب عامل متغیرهای مورد استفاده تشکیل می‌شود می‌توان بر اساس معیارهای کفایت اندازه نمونه MSA [1] و KMO [2] متغیرهای مستقل مناسب برای منظور شدن در مدل تجزیه عاملی را انتخاب نمود.

معیار کفایت اندازه نمونه (MSA) برای تعیین منظور شدن متغیر در مدل تجزیه عاملی محاسبه می‌شود. مقادیر بزرگ این معیار در منظور شدن متغیر در مدل مؤثر است. معیار MSA با رابطه زیر محاسبه می‌شود :

$$MSA_i = \frac{\sum_{j \neq i} r_{ij}^2}{\sum_{j \neq i} r_{ij}^2 + \sum_{j \neq i} a_{ij}^2} \quad i=1, \ldots, m \quad (۲)$$

در این فرمول، r_{ij} ضریب همبستگی ساده و a_{ij} ضریب همبستگی جزیی متغیرهای i و j به شرط ثابت بودن سایر متغیرها می‌باشد.

ضریب دیگر برای بررسی نتایج تجزیه عامل‌ها KMO می‌باشد که از رابطه زیر محاسبه می‌شود :

1 - Measure of Sampling Adequacy

2 - Kaiser-Meyer-Olkin

$$KMO = \frac{\sum_{i\neq j}\sum_{j\neq i} r_{ij}^2}{\sum_{i\neq j}\sum_{j\neq i} r_{ij}^2 + \sum_{i\neq j}\sum_{j\neq i} a_{ij}^2} \qquad (۳)$$

درصورتی‌که مقدار این ضریب حدود ۰/۹ باشد نتیجه تجزیه عاملی بسیار مناسب، مقدار ۰/۸ مناسب، ۰/۷ متعادل، ۰/۶ متوسط، ۰/۵ کم و کمتر از ۰/۵ نامناسب است (۱۱).

تحلیل خوشه‌ای

تحلیل خوشه‌ای جستجو و سازماندهی اطلاعات برای تعیین گروه‌هایی از موضوعات مختلف می‌باشد که افراد داخل یک گروه از جنبه‌هایی مشابه و با افراد گروه‌های دیگر نامتشابه می‌باشند. اگر حوضه‌ها دارای خواص اندازه‌گیری شده بسیار مشابه باشند در فضای n بعدی بسیار نزدیک به یکدیگر قرار می‌گیرند. میزان مشابهت این حوضه‌ها با اندازه‌گیری فاصله بین آنها بررسی و عدد حاصل به عنوان ضریب نزدیکی یا مشابهت عنوان می‌شود (۹). طبقه‌بندی گروه‌های (نمونه‌های) یک جامعه به معلوم یا مجهول بودن تعداد گروه‌ها، تعداد گروه‌های هدف و نوع متغیرهای مورد اندازه‌گیری بستگی دارد. اگر تعداد گروه‌ها از قبل معلوم باشند، از روش K میانگین و درصورت مجهول بودن گروه‌ها از روش سلسله مراتبی استفاده می‌شود.

اجرای روشها در چهار مرحله زیر انجام می‌شود :

۱ - انتخاب متغیرهای مستقل، ۲ - استاندارد کردن داده‌ها، ۳ - اندازه‌گیری فواصل بین متغیرها، ۴ - گروه‌بندی متغیرها

چون دامنه تغییرات متغیرها متفاوت است باید برای مقایسه آنها همگی به حالت استاندارد تبدیل شوند که بر اساس رابطه زیر بیان می‌شود :

$$Z_{if} = \frac{x_{if} - m_f}{S_f} \qquad (۴)$$

در این فرمول، x_{if} داده خام، Z_{if} داده استاندارد شده، m_f میانگین داده‌های خام و S_f انحراف معیار داده‌های خام می‌باشند.

برای مشخص کردن میزان ارتباط بین دو متغیر از فاصله اقلیدسی استفاده می‌شود که به صورت رابطه زیر تعریف می‌شود :

$$D = \sqrt{\sum_{l=1}^{K} (X_{il} - X_{jl})^2} \qquad (۵)$$

در این فرمول، X_{il} خصوصیت l مربوط به فرد i و X_{jl} خصوصیت l مربوط به فرد j می‌باشد.

برای گروه‌بندی متغیرها در این تحقیق از روش Ward استفاده شده است. در این روش میانگین هر متغیر در داخل هر خوشه محاسبه شده و برای هر مشاهده، مربع فاصله اقلیدسی میانگین خوشه‌ها محاسبه می‌شود. این فاصله برای کلیه مشاهدات جمع می‌شود. در هر مرحله دو خوشه‌ای ترکیب می‌شوند که کوچک‌ترین افزایش در مجموع مربعات فواصل داخل خوشه‌ای را داشته باشند.

در یک تحقیق نشان داده شد که تفکیک منطقه مطالعاتی به مناطق همگن باعث دقت بیشتر

و خطـای کمتـر در مـدل‌های رگرسیون در مناطق فاقـد آمـار مـی‌شـود (۱۰). در ایـن مطالعـه ۲۲۱ ایستگاه آب‌سنجی در ایالت آریزونای آمریکا برای بررسـی مـناطق همگـن از روش تحلیل خوشه‌ای اسـتفاده شـد. گروه‌بندی حوضه‌های مورد مطالعه بـر اسـاس ویژگـی‌هـای هـیدرولوژیک مـنطقه و همچنـین ویژگـی‌هـای فیزیوگرافـی حوضـه‌هـای مطالعاتی انجام شد (۱۰).

در مطالعـه دیگـری اهمـیت تعیـین مـناطق بـر اساس خصوصیات حوضه در تحلیل منطقه‌ای سـیلاب در اسـکاتلند بررسی شد (۵). برای تعیین مناطق همگن از روش تحلیل خوشه‌ای و براساس عـوامل مسـاحت حوضـه، تـراکم آبراهه، شاخص بارندگی، طول آبراهه اصلی، شیب آبراهه اصلی و عامـل کمـبود رطـوبت خـاک در ۱۶۸ حوضـه مطالعاتـی اسـتفاده شـد. در ایـن مطالعـه از روش Ward و بـرای تعیـین دسـته‌های همگن از نسبت درسـت‌نمایـی استفاده شد. از بین پنج گروه چهار گـروه قابـل قـبول و یـک گـروه قـادر به توضیح تغییرات دبی نبود (۵). همچنین از روش الگوریتم خوشـه‌ای برای شناخت مناطق همگن نیز استفاده شده است (۷).

منحنی‌های Andrew

یکی از روشهای تحلیل داده‌های یک جامعه آمـاری، نمـایش آنهـا بـر روی محـور دوبعـدی مـی‌باشـد (۶). در ایـن روش یک نقطه در فضای چـندبعدی از پارامتـرها از طـریق یـک منحنـی دوبعـدی بـا استفاده از معادله (۶) تعیین و نمایش داده می‌شود :

$$F(t) = \frac{X_1}{\sqrt{2}} + X_2 Sin(t) + X_3 Cos(t) + X_4 Sin(2t) + X_5 Cos(2t) + \ldots \quad (۶)$$

متغیـرهای $X_1, X_2, \ldots$ هریک از مشخصه‌های نقاط موردنظر و به ترتیب اهمیت را نشان می‌دهد و معادلـه فـوق در دامـنه $-\pi$ تـا $+\pi$ رسـم مـی‌شود. تفاوت بین دو منحنی با فاصله اقلیدسی بـین صـفات آنهـا متناسـب اسـت. بـه‌طـوری‌کـه متغیـرهای n بعدی که در فضای اقلیدسی نزدیک یکدیگـر باشـند، دارای منحنـی‌های مشابه هستند. بـه منظور رسم این منحنی‌ها خصوصیات فیزیکی و هـیدرولوژیک از حوضـه‌ها که مستقل باشند به کمـک روشـهایی (از جملـه روش تحلیل عاملی) انـتخاب و سـپس این متغیرهای مستقل استاندارد مـی‌شـوند. از ایـن منحنی‌ها می‌توان به عنوان یک روش مستقل برای گروه‌بندی حوضه مطالعاتی به مـناطق همگـن و بـرای اطمیـنان از صـحت گروه‌بندی به روش تحلیل خوشه‌ای استفاده نمود.

تحلیل ممیزی

بـرای گـروه‌بندی مشاهدات به یکی از چند گـروه معلـوم، براسـاس مشخصـه‌هـای مخـتلف استفاده می‌شود.

در تحلـیل ممیـزی نحوه تفکیک دو یا چند گـروه از افراد از لحاظ اندازه‌گیری‌های انجام شده روی چند متغیر مطرح می‌باشد. در این روش قبل از شـروع تحلـیل نـیازی به استاندارد کردن داده‌ها بـرای حصول میانگین صفر و واریانس یک وجود نـدارد. درواقـع نتیجه این عمل تحت تأثیر مقیاس متغیـرهای جداگانـه نیست. اغلب تعیین توابعی از

متغیرهای $X_1, X_2, ... X_P$ که بتوانند m گروه را تاحد ممکن به‌نحوی از یکدیگر متمایز کند مفید است.

از ساده‌ترین روشها در تعیین توابع ممیزی کانونی استفاده از ترکیب خطی است.

$$Z=a_1X_1+a_2X_2+...+a_PX_P \qquad (۷)$$

اگر تغییرات مقدار میانگین یک گروه به گروه دیگر زیاد باشد و مقادیر درون یک گروه نسبتاً ثابت باشند، گروه‌ها را با استفاده از Z به‌خوبی می‌توان متمایز کرد.

از ضریب همبستگی تابع ممیزی و کای اسکوئر و نیز حداقل شدن آماره‌هایی مانند آماره ویلکس لامبدا می‌توان به عنوان معیارهایی برای اطمینان از گروه‌بندی استفاده نمود.

نتایج

بررسی همگنی منطقه با روش آماری چندمتغیره

در این روش خصوصیات حوضه‌های آبریز با تحلیل عاملی و تحلیل خوشه‌ای مطالعه شد و تغییرپذیری فضایی امتیازات عاملی توصیف و تفسیر شد. سپس گروه‌بندی امتیازات عاملی با روش گروه‌بندی فاصله‌ای مناطق همگن مشخص شد. نتایج این روش در دو مورد ارایه می‌شود :

- نتایج روش تحلیل عاملی برای تعیین خصوصیات متمایزکننده مناطق همگن

این روش برای خلاصه نمودن اطلاعات زیاد بوده و نتایج حاصل از نظر مفهوم معنی‌دار می‌باشند. در این مطالعه از تحلیل عاملی برای متغیر اندازه‌گیری شده در حوضه‌های منتخب استفاده شد.

متغیرها شامل خصوصیات مختلف حوضه نظیر: مساحت، محیط، شیب متوسط حوضه، شیب آبراهه اصلی، ارتفاع متوسط حوضه، تراکم زهکشی، نسبت انشعاب، طول آبراهه اصلی، قطر دایره معادل، ضریب میلر، ضریب گراولیوس، طول مستطیل معادل، عرض مستطیل معادل، زمان تمرکز، مجموع طول آبراهه‌ها و بارندگی متوسط سالانه می‌باشند. باتوجه به این‌که واحدهای سنجش هرکدام از متغیرها متفاوت از دیگری بود، لذا برای مقایسه صحیح متغیرها با یکدیگر، کلیه آنها به نمرات Z استاندارد تبدیل و ماتریسی به ابعاد ۱۶×۴۱ تشکیل شد که طی آن ایستگاه در ردیف و متغیرها در ستون ماتریس قرار داده شد.

چون نتایج حاصل از تحلیل عاملی پیچیده بوده و راه حل بهینه‌ای حاصل نمی‌شود، لذا برای حداکثر نمودن واریانس هریک از عامل‌ها و تسهیل در تفسیر ساختار عاملی، محورهای عاملی با دوران واریماکس دوران داده شد تا به حالت مستقل درآیند، نام‌گذاری عامل‌ها نیز براساس عامل‌های دوران یافته انجام شد. سپس با استفاده از روش برآورد رگرسیون، ماتریس امتیازات عاملی ایستگاه‌ها استخراج شد. برای محدود نمودن تعداد عامل‌ها از آماره KMO که تعیین‌کننده میزان تناسب تعداد عامل‌های انتخابی است استفاده شد و KMO برابر با ۰/۷۴۶ حاصل شد.

جدول ۱ - ماتریس ضرایب خودهمبستگی در خروجی Anti-Image

پارامتر	area	wsa	rsa	sumli	dr	li	hm	pay	tc	pyram	graw	dm	rc	lm	bm	Br
area	۰/۷۴															
wsa	-۰/۰۳	۰/۶۴														
rsa	۰/۰۴	-۰/۸۷	۰/۶۰													
sumli	-۰/۹۲	-۰/۰۵	۰/۰۱	۰/۷۷												
dr	۰/۴۷	-۰/۲۲	۰/۰۶	-۰/۵۲	۰/۴۸											
li	-۰/۳۲	-۰/۱۹	-۰/۳۶	۰/۰۸	۰/۰۰	۰/۸۲										
hm	-۰/۲۴	۰/۴۳	-۰/۱۸	۰/۱۸	-۰/۳۲	-۰/۰۲	۰/۴۷									
pay	-۰/۲۱	-۰/۰۴	۰/۰۸	۰/۲۶	-۰/۳۳	-۰/۲۵	۰/۱۶	۰/۶۴								
tc	۰/۲۷	۰/۲۰	۰/۴۲	-۰/۰۶	-۰/۰۴	-۰/۹۶	۰/۰۷	۰/۲۳	۰/۸۲							
pyram	-۰/۶۳	۰/۰۴	۰/۰۸	۰/۵۵	-۰/۲۵	۰/۰۱	۰/۱۴	۰/۲۲	-۰/۰۰	۰/۸۱						
graw	۰/۳۳	-۰/۱۲	۰/۰۳	-۰/۲۴	۰/۱۵	-۰/۰۴	۰/۱۲	-۰/۲۰	۰/۰۷	-۰/۴۰	۰/۵۰					
dm	۰/۱۴	-۰/۰۳	۰/۱۷	-۰/۲۳	۰/۲۶	-۰/۰۶	۰/۰۸	۰/۰۴	۰/۰۴	-۰/۲۹	۰/۱۹	۰/۸۵				
rc	۰/۲۵	۰/۰۶	-۰/۰۵	-۰/۳۲	۰/۰۵	۰/۱۶	۰/۰۸	-۰/۰۳	-۰/۱۳	-۰/۳۱	۰/۰۲	۰/۲۹	۰/۸۳			
lm	۰/۵۱	۰/۱۰	-۰/۰۶	-۰/۳۹	۰/۰۵	-۰/۱۴	-۰/۲۰	۰/۰۲	۰/۰۷	-۰/۵۰	-۰/۰۱	-۰/۳۰	۰/۴۵	۰/۴۸		
bm	-۰/۲۴	۰/۲۳	۰/۰۷	۰/۲۱	-۰/۰۲	-۰/۲۸	۰/۰۴	۰/۱۶	۰/۲۱	۰/۱۷	-۰/۲۳	۰/۰۳	-۰/۷۴	-۰/۲۵	۰/۸۶	
br	-۰/۰۹	۰/۲۵	۰/۱۰	۰/۰۶	-۰/۱۸	-۰/۲۰	۰/۳۳	۰/۲۸	۰/۲۴	۰/۱۱	۰/۰۱	۰/۱۳	۰/۰۱	-۰/۰۹	۰/۱۴	۰/۵۳

در ماتریس فوق، مقادیر عناصر روی قطر پارامتر MSA را نشان می‌دهد.

جدول پیوست ۱ - مشخصات ایستگاه‌های آب‌سنجی مورد استفاده در تحقیق [۳]

ردیف	شماره ایستگاه	رودخانه	محل ایستگاه	زیرحوضه	طول جغرافیایی	عرض جغرافیایی
۱	۲۱-۱۰۳	آب ملایر	پل حاج علی مراد	۱	۴۸° ۱۴'	۳۴° ۱۸'
۲	۲۱-۱۰۵	سرآب گاماسیاب	سنگ سوراخ (غرب)	۱	۴۸° ۲۳'	۳۴° ۰۳'
۳	۲۱-۱۰۷	آب نهاوند	گوشه سعدوقاص	۱	۴۸° ۱۴'	۳۴° ۱۷'
۴	۲۱-۱۱۱	شهاب	آقاجان بلاغی	۱	۴۸° ۰۳'	۳۴° ۵۰'
۵	۲۱-۱۱۳	خرم‌رود	آران (غرب)	۱	۴۷° ۱۵'	۳۴° ۲۵'
۶	۲۱-۱۱۵	گاماسیاب	دوآب	۱	۴۷° ۵۴'	۳۴° ۲۲'
۷	۲۱-۱۱۷	آبشار	صحنه	۱	۴۷° ۴۲'	۳۴° ۲۹'
۸	۲۱-۱۲۱	جامیشان	میان‌راهان	۱	۴۷° ۲۶'	۳۴° ۳۵'
۹	۲۱-۱۲۳	حریم‌نگار	کله‌چوب	۱	۴۶° ۲۷'	۳۴° ۳۵'
۱۰	۲۱-۱۲۵	دین‌آور	بیستون	۱	۴۷° ۲۷'	۳۴° ۲۶'
۱۱	۲۱-۱۲۷	گاماسیاب	پل‌چهر	۱	۴۷° ۲۶'	۳۴° ۲۰'
۱۲	۲۱-۱۲۹	آب‌مرک	ماهی‌دشت	۲	۴۶° ۴۹'	۳۴° ۱۶'
۱۳	۲۱-۱۳۱	آب‌مرک	خرس‌آباد	۲	۴۶° ۴۴'	۳۴° ۳۱'
۱۴	۲۱-۱۳۳	قره‌سو	دوآب‌مرک	۲	۴۶° ۴۷'	۳۴° ۳۳'
۱۵	۲۱-۱۳۷	قره‌سو	بیارکامیاران	۲	۴۶° ۵۷'	۳۴° ۴۸'
۱۶	۲۱-۱۳۹	رازآور	حجت‌آباد	۲	۴۷° ۰۰'	۳۴° ۲۹'
۱۷	۲۱-۱۴۱	قره‌سو	پل‌کهنه	۲	۴۷° ۰۸'	۳۴° ۱۹'
۱۸	۲۱-۱۴۳	قره‌سو	قره باغستان	۲	۴۷° ۱۵'	۳۴° ۱۴'
۱۹	۲۱-۱۴۵	بادآور	نورآباد (غرب)	۳	۴۷° ۵۸'	۳۴° ۰۵'
۲۰	۲۱-۱۴۷	سیمره	هولیان	۳	۴۶° ۱۵'	۳۳° ۴۴'
۲۱	۲۱-۱۴۹	جزمان	هولیان	۳	۴۶° ۰۶'	۳۳° ۴۶'
۲۲	۲۱-۱۵۳	کرند	خسروآباد	۳	۴۶° ۲۳'	۳۴° ۱۰'
۲۳	۲۱-۱۵۵	چم‌راوند	کله‌چوب	۳	۴۶° ۳۸'	۳۴° ۰۲'
۲۴	۲۱-۱۵۷	آب چناره	دارتوت	۳	۴۶° ۴۰'	۳۳° ۴۵'
۲۵	۲۱-۱۵۹	سیمره	تنگ‌سازبن	۳	۴۶° ۵۱'	۳۳° ۳۴'
۲۶	۲۱-۱۶۱	آب سیروان	ورگچ	۳	۴۶° ۵۰'	۳۳° ۳۳'
۲۷	۲۱-۱۶۳	دره‌دزدان	تنگ‌سیاب	۳	۴۷° ۱۲'	۳۳° ۲۳'
۲۸	۲۱-۱۶۵	سیمره	چم‌زاب	۳	۴۷° ۲۸'	۳۳° ۱۰'
۲۹	۲۱-۱۶۷	هررود	دهنو	۴	۴۸° ۴۷'	۳۳° ۳۱'
۳۰	۲۱-۱۶۹	هررود	کاکارضا	۴	۴۸° ۱۶'	۳۳° ۴۳'
۳۱	۲۱-۱۷۱	دوآب‌اشتر	سراب‌سیدعلی	۴	۴۸° ۱۳'	۳۳° ۴۸'
۳۲	۲۱-۱۷۳	کشکان	پل‌کشکان	۴	۴۷° ۴۸'	۳۳° ۳۰'
۳۳	۲۱-۱۷۵	خرم‌آباد	چم‌انجیر	۴	۴۸° ۱۴'	۳۳° ۲۷'
۳۴	۲۱-۱۷۷	کشکان	آفرینه	۴	۴۷° ۵۴'	۳۳° ۲۰'

ادامه جدول پیوست ۱ - مشخصات ایستگاه‌های آب‌سنجی مورد استفاده در تحقیق [۳]

ردیف	شماره ایستگاه	رودخانه	محل ایستگاه	زیرحوضه	طول جغرافیایی	عرض جغرافیایی
۳۵	۲۱-۱۷۹	چلهول	آفرینه	۴	۴۷° ۵۳′	۳۳° ۱۸′
۳۶	۲۱-۱۸۱	مادیان‌رود	برآفتاب	۴	۴۷° ۴۹′	۳۳° ۱۹′
۳۷	۲۱-۱۸۳	کشکان	پل‌دختر	۴	۴۷° ۴۳′	۳۳° ۱۰′
۳۸	۲۱-۱۸۵	کرخه	جلوگیر (ماژین)	۳	۴۷° ۴۸′	۳۲° ۵۸′
۳۹	۲۱-۱۸۷	کرخه	چم‌گز	۳	۴۹° ۴۷′	۳۲° ۵۷′
۴۰	۲۱-۱۸۹	آب زال	پل زال	۳	۴۸° ۰۵′	۳۲° ۴۹′
۴۱	۲۱-۱۹۱	کرخه	پای‌پل	۳	۴۸° ۰۹′	۳۲° ۲۵′

برای خارج ساختن متغیرهای غیرضروری از ماتریس Anti-Image استفاده شد. معیار تشخیص این متغیرها، آماره MSA است که عناصر قطری ماتریس همبستگی می‌باشند. به این صورت که هر متغیر که دارای کمترین مقدار MSA است با درنظر گرفتن سطوح معنی‌دار ماتریس ضرایب همبستگی میان متغیرها، از مجموعه متغیرها حذف می‌شوند (جدول ۱).

باتوجه به جدول (۱) در مرحله اول باید متغیر hm (ارتفاع متوسط حوضه) را که دارای کمترین مقدار MSA است حذف نمود. با حذف این متغیر تعداد عامل‌ها از پنج به چهار کاهش می‌یابد. ولی میزان افزایش آماره KMO معنی‌دار نمی‌باشد (۰/۷۷۵). بنابراین چون ادامه روند تحلیل، با تعداد عامل‌های بیشتر، دارای دقت زیادتری می‌باشد، لزوم حذف متغیرها برطرف می‌شود. نتیجه نشان می‌دهد که درمجموع پنج عامل و ۱۶ متغیر ۷۷/۶ درصد از واریانس داده‌ها را توجیه می‌نمایند.

باتوجه به جداول مربوط به مقادیر واریانس توجیه شده و ماتریس دورانی واریماکس (جدول‌های ۲ و ۳) می‌توان دریافت که حدود ۲۲/۴ درصد از کل واریانس توجیه نشده است که می‌توانست با بررسی متغیرهای هیدرولوژیک اضافی افزایش یابد. همچنین سطح حوضه با بیشترین بار وزنی (۰/۹۸۹) روی عامل اول و توجیه بیشترین درصد از کل واریانس (۴۷/۴ درصد) به عنوان عامل اول انتخاب شد. همچنین عامل دوم باتوجه به بار وزنی ۰/۷۸۶ و ۱۱/۹ درصد از واریانس کل به قطر دایره معادل نسبت داده شد. به این ترتیب تعداد سه عامل دیگر یعنی ارتفاع متوسط حوضه، شیب متوسط حوضه و نسبت انشعاب به عنوان عامل‌های بعدی انتخاب شد.

- نتایج تحلیل خوشه‌ای برای شناسایی مناطق همگن

در این مرحله عامل‌های استخراج شده از طریق فرآیند عاملی، تحلیل خوشه‌ای و با استفاده از امتیازات عاملی ایستگاه‌های منتخب گروه‌بندی شدند. هدف از این کار، گروه‌بندی ایستگاه‌های انتخاب شده در داخل مناطق همگن بود. بر این اساس با استفاده از روشهای ریاضی، یک نمودار

n بعدی تنظیم شد که در آن هر حوضه به عنوان یک نقطه در فضای عاملی n بعدی درنظر گرفته شد. در این مدل ریاضی، نقاط یا حوضه‌های نزدیک‌تر در امتیازات عاملی مشابه هستند. معیارهای مورد استفاده برای گروه‌بندی شامل خصوصیات فیزیوگرافی حوضه (سطح حوضه، قطر دایره معادل، ارتفاع متوسط حوضه، شیب متوسط حوضه و نسبت انشعاب) می‌باشند. برای تعیین گروه‌ها از روش Ward استفاده شد. نتایج روش خوشه‌بندی سلسله مراتبی Ward با نمودار درختی در شکل (۱) نشان داده شده است.

معیار تعیین مناطق همگن، مقدار فاصله واقعی یا فاصله مقیاس‌بندی شده (در فاصله یک تا ۲۵) است. برای تعیین تعداد مناطق همگن فاصله ۱۰ انتخاب و با این فاصله منطقه مورد مطالعه به دو منطقه همگن تبدیل شد. با بررسی شکل (۱) حوضه‌هایی که در هر منطقه همگن قرار گرفتند به صورت زیر می‌باشد :

-گروه همگن یک : شامل ایستگاه‌های پل حاج‌علی‌مراد، سنگ‌سوراخ، گوشه سعدوقاص، آقاجان‌بلاغی، آران، دوآب، آبشار، میان‌راهان، کله‌چوب، بیستون، پل‌دختر، ماهی‌دشت، خرس آباد، دوآب‌مرک، بیارکامیاران، حجت‌آباد، پل‌کهنه، قره‌باغستان، نورآباد، هولیان (جزمان)، خسروآباد، کله‌چوب، دارتوت، ورگچ، تنگ‌سیاب، دهنو، کاکارضا، سراب‌سیدعلی، پل‌کشکان، چم انجیر، آفرینه (کشکان)، آفرینه (چلهول)، برآفتاب و پل‌زال می‌باشد.

-گروه همگن دو : شامل ایستگاه‌های هولیان (سیمره)، تنگ‌سازبن، چم‌ژاب، جلوگیر (ماژین)، چم‌گز و پای‌پل می‌باشد.

منحنی‌های Andrew

در این روش برای تعیین متغیرهای مستقل قابل استفاده در مدل Andrew (رابطه ۶) از نتایج حاصل از تحلیل عاملی استفاده و توابع مربوط به هر یک از ۴۱ ایستگاه ترسیم شد (شکل ۲). شکل نشان می‌دهد که ایستگاه‌های هولیان (سیمره)، تنگ‌سازبن، چم‌ژاب، جلوگیر (ماژین)، چم‌گز و پای‌پل به‌ترتیب با شماره‌های ۱۴۷، ۱۵۹، ۱۶۵، ۱۸۱، ۱۸۵، ۱۸۷ و ۱۹۱ متمایز بوده و مابقی ایستگاه‌ها به صورت کلاف متراکم بوده و می‌توانند در یک گروه قرار گیرند که با نتیجه حاصل از روش قبل متناسب است.

از این روش به عنوان یک روش تعیین همگنی و همچنین بررسی صحت نتیجه حاصل از روش تحلیل خوشه‌ای استفاده شد.

تحلیل ممیزی برای تعیین صحت گروه‌بندی

در این مرحله برای بررسی صحت تحلیل خوشه‌ای در تعیین منطقه همگن از تحلیل ممیزی استفاده شد. نتایج ضریب تابع متمایزکننده در حالت استاندارد نشان می‌دهد که در تابع ایجاد شده ضریب میلر (rc) که معادل ۰/۶۲۴ است مهمترین متغیر متمایزکننده گروه‌ها از یکدیگر می‌باشد. در این بخش حداکثر تعداد توابع ممیزی کانونی با پیش فرض g-1 (g تعداد گروه‌ها) حاصل شد. حداقل درصد تجمعی واریانس توابع ممیزی به صورت پیش فرض ۱۰۰ درصد است. حداکثر سطح معنی‌دار آماره ویلکس لامبدا نیز یک فرض شده است.

جدول ۲ - مقادیر واریانس توجیه شده

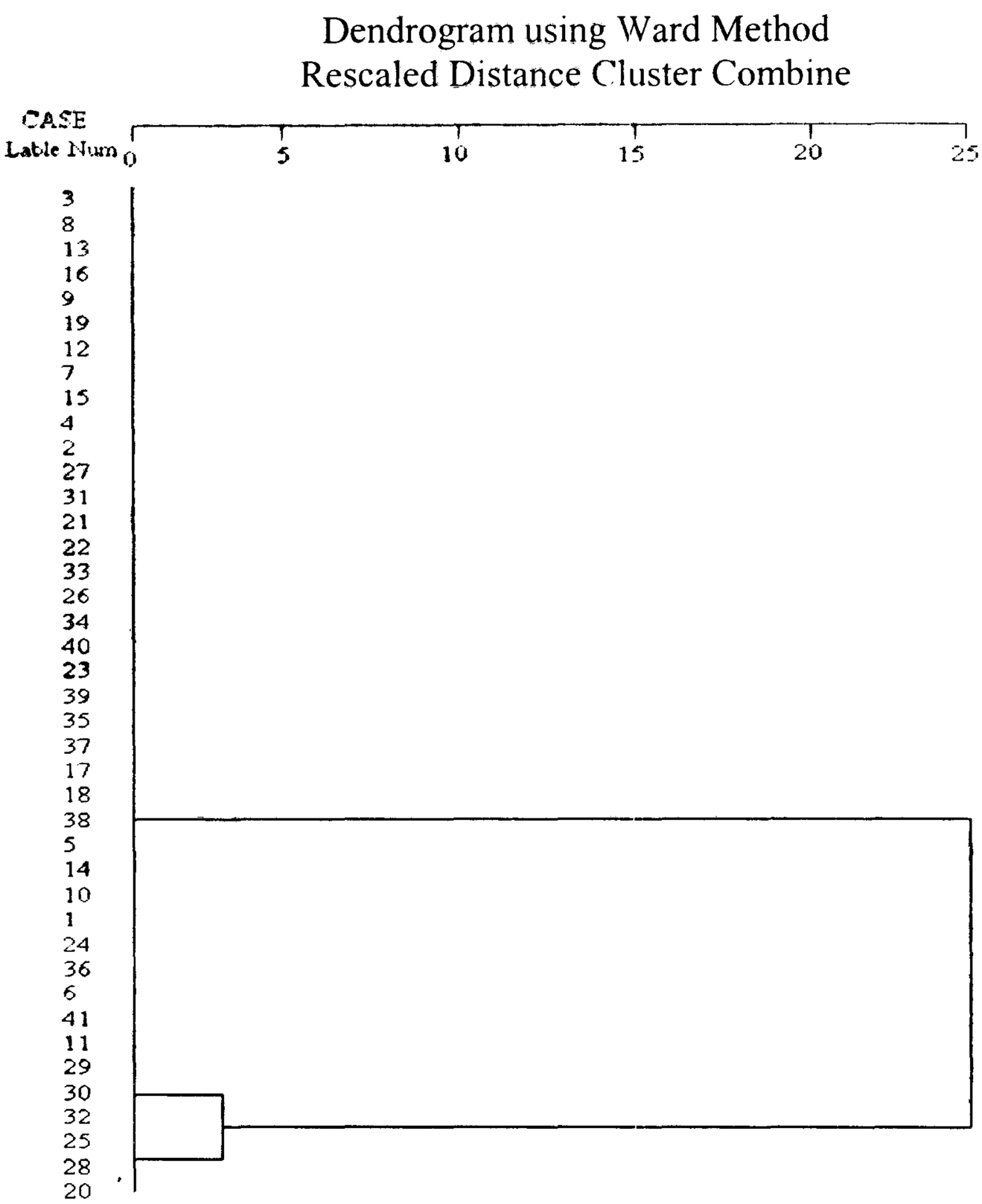
عامل	واریانس (درصد)	مقادیر تجمعی واریانس (درصد)
۱	۴۷/۴	۴۷/۴
۲	۴۷/۴	۵۹/۴
۳	۷/۰	۶۶/۴
۴	۶/۳	۷۲/۷
۵	۴/۹	۷۷/۶

جدول ۳ - ماتریس دورانی واریماکس

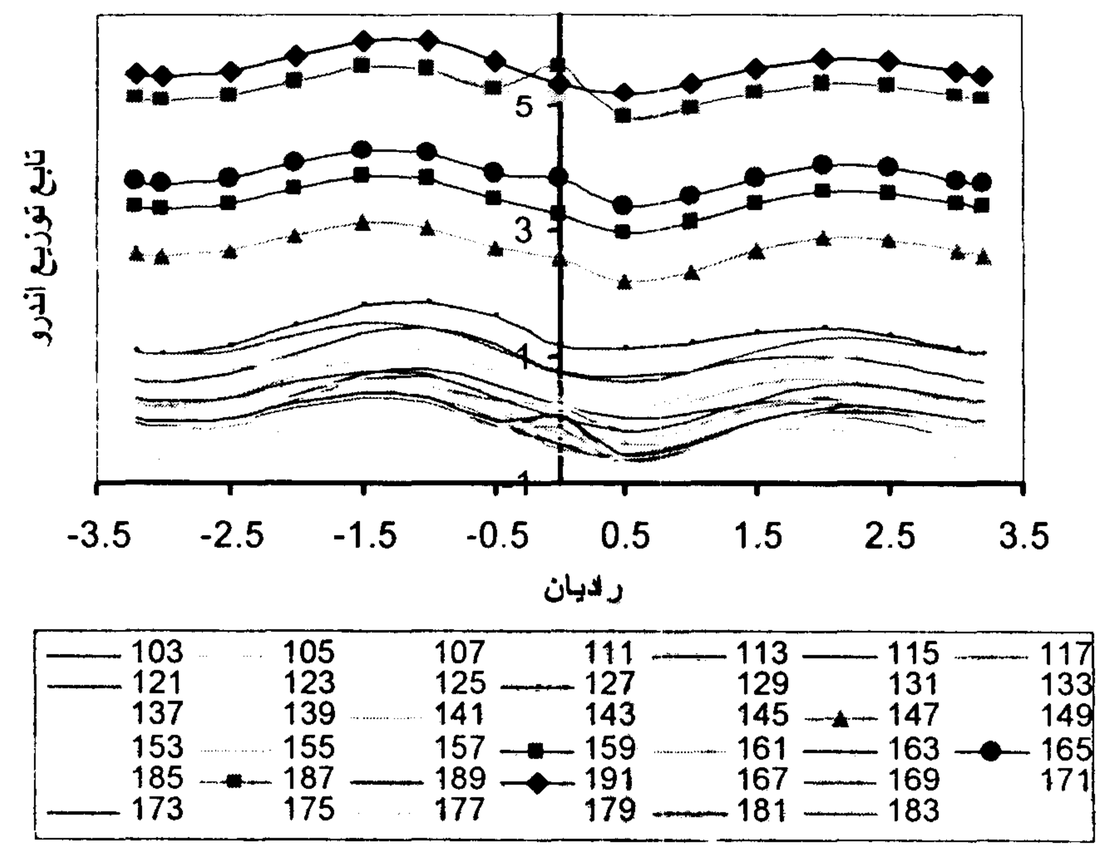
	عامل				
پارامتر استاندارد شده	۱	۲	۳	۴	۵
مساحت	۰/۹۸۹	-۵/۶۴۱ E -۰۲	۴/۶۰۳ E -۰۲	-۶/۱۳۷ E -۰۲	۹/۸۳۸ E -۰۲
طول آبراهه اصلی	۰/۹۸۸	۴/۸۹۵ E -۰۲	۱/۳۹۹ E -۰۲	-۱/۰۶۳ E -۰۲	۶/۱۳۸ E -۰۲
مجموع طول آبراهه‌ها	۰/۹۸۷	-۵/۱۷۳ E -۰۲	۴/۳۸۵ E -۰۲	-۴/۹۶۱ E -۰۲	۰/۱۱۵
محیط حوضه	۰/۹۷۸	۰/۱۰۷	-۱/۲۳۶ E -۰۲	۱/۰۰۵ E -۰۳	۱/۳۸۲ E -۰۲
عرض مستطیل معادل	۰/۹۷۷	-۰/۱۳۳	۲/۶۰۹ E -۰۲	-۰/۱۴۵	-۲/۲۶۸ E -۰۲
زمان تمرکز	۰/۹۷۵	۹/۸۳۲ E -۰۲	-۶/۰۴۱ E -۰۳	-۱/۵۴۰ E -۰۳	۲/۲۴۹ E -۰۲
ضریب میلر	۰/۹۳۴	-۰/۲۸۷	۳/۲۲۲ E -۰۲	-۰/۲۰۴	-۴/۷۸۲ E -۰۲
تراکم زهکشی	-۰/۳۲۱	-۰/۱۲۹	۰/۱۳۷	۷/۸۱۱ E -۰۲	۰/۲۸۴
طول مستطیل معادل	-۰/۱۸۳	۰/۷۰۷	-۹/۷۷۹ E -۰۲	۰/۲۶۶	۱/۷۶۰ E -۰۲
قطر دایره معادل	۰/۶۰۹	۰/۷۸۶	-۸/۷۷۶ E -۰۲	۰/۳۳۷	۹/۷۱۰ E -۰۲
بارندگی متوسط سالانه	-۰/۳۰۰	-۰/۴۰۴	۰/۱۶۱	۰/۲۳۳	۸/۲۹۲ E -۰۲
شیب آبراهه اصلی	-۰/۲۴۰	-۰/۲۸۹	۰/۲۵۸	-۴/۴۸۱ E -۰۳	۰/۲۲۴
ضریب گراولیوس	۰/۲۱۵	-۰/۲۵۱	-۰/۱۲۱	۴/۱۲۳ E -۰۲	-۰/۲۲۸
نسبت انشعاب	-۰/۲۷۹	۹/۱۹۸ E -۰۳	-۰/۷۰۰	-۰/۳۹۰	۰/۵۳۰
ارتفاع متوسط حوضه	-۰/۲۱۳	۰/۴۶۷	۰/۶۹۲	۰/۴۳۲	۰/۲۸۳
شیب متوسط حوضه	۰/۳۰۲	-۰/۳۵۴	۹/۱۶۷ E -۰۲	۰/۵۹۹	۰/۴۳۴

جدول ۴ - نتایج تحلیل ممیزی برای منطقه همگن

همبستگی کانونیک	آماره ویلکس لامبدا	سطح معنی‌دار	کای اسکوئر	درجه آزادی	تعداد توابع ممیزی مورد استفاده در تحلیل
۰/۹۹۶	۰/۰۰۷	۰/۰۰۰	۱۸۱/۳۹۱	۴	۱



شکل ۱ - نمودار درختی گروه‌بندی منطقه مطالعاتی



شکل ۲ - دسته منحنی‌های Andrew در منطقه مورد مطالعه

باتوجه به مقادیر همبستگی کانونیک، کای اسکوئر و آماره ویلکس لامبدا مشخص می‌شود که دو گروه همگن از یک دیگر متمایز می‌باشند.

بحث

در این تحقیق ۴۱ حوضه مورد استفاده از نظر خصوصیات هیدرولوژیک و استفاده از روشهای الگوریتم خوشه‌ای به دو گروه همگن تقسیم‌بندی شده‌اند. تحلیل ممیزی و منحنی‌های Andrew نیز مناسب بودن این تقسیم‌بندی را نشان می‌دهند. گروه همگن شماره دو شامل ایستگاه‌های موجود در آبراهه اصلی حوضه می‌باشد. برای بهبود گروه‌بندی منطقه از سایر خصوصیات هیدرولوژیک و فیزیوگرافی حوضه مانند درصد رسوبات درشت‌دانه و غیره نیز می‌توان استفاده نمود.

منابع مورد استفاده

۱ - ارقامی، ن. ر. و بزرگ‌نیا، ا. ۱۳۷۰. آمار چندمتغیره کاربردی. انتشارات آستان قدس رضوی (ترجمه)، مشهد.

۲ - اسلامیان، س.، زارعی، ع. و ابریشم‌چی، ا. ۱۳۸۳. برآورد منطقه‌ای جریان‌های کم رودخانه‌های حوضه آبریز مازندران. علوم و فنون کشاورزی و منابع طبیعی ۸(۱): ۲۷-۳۸.

۳ - جاماب (مهندسین مشاور). ۱۳۶۹. طرح جاماب حوضه آبریز کرخه. انتشارات وزارت نیرو، تهران.

۴ - سازمان تحقیقات منابع آب (تماب) (مهندسین مشاور). ۱۳۶۹. مجموعه گزارشات طرح جامع آب کشور. انتشارات وزارت نیرو.

5 . Acreman MC and Sinclair CD (1986) Classification of drainage basins according to their physical characteristics for flood frequency analysis in Scotland. Journal of Hydrology 84: 365-384.

6 . Andrew DF (1972) Plots of high-dimensional data. Biometrics 28: 125-36.

7 . Burn DH and Goel NK (2000) The formation of groups for regional floods frequency analysis. Hydrological Sciences Journal 45: 97-112.

8 . Nathan RJ and McMahon TA (1990) Identification of homogeneous regions for the purpose of regionalization. Journal of Hydrology 121: 217-238.

9 . Smakhtin VU (2001) Low flow hydrology a review. Journal of Hydrology 240: 147-186.

10 . Tasker GD (1982) Comparing methods of hydrology regionalization. Water Resources Bulletin 18(6): 965-970.

11 . Wiltshire SE (1985) Grouping basins for regional flood frequency analysis. Hydrological Sciences Journal 30: 151-159.