دانشکده منابع طبیعی

روش‌های تحلیل چندمتغیره در نرم‌افزار SPSS

تهیه و تنظیم

محمد علی زارع چاهوکی

عضو هیات علمی دانشکده منابع طبیعی

پاییز ۱۳۸۹

۱- تحلیل عاملی

۱-۱- مقدمه

یکی از روش‌های آماری برای تجزیهٔ اطلاعات موجود در مجموعهٔ داده‌ها روش تجزیه عامل‌ها یا تحلیل عاملی[1] است. این روش توسط کارل پیرسون[2] (۱۹۰۱) و چارلز اسپیرمن[3] (۱۹۰۴) برای اولین بار هنگام اندازه‌گیری هوش مطرح شد و برای تعیین تأثیرگذارترین متغیرها در زمانیکه تعداد متغیرهای مورد بررسی زیاد و روابط بین آنها ناشناخته باشد، استفاده می‌شود. در این روش متغیرها در عامل‌هایی قرار می‌گیرند، به‌طوریکه از عامل اول به عامل‌های بعدی درصد واریانس کاهش می‌یابد، از این‌رو متغیرهایی که در عامل‌های اولی قرار می‌گیرند، تأثیرگذارترین هستند.

تجزیهٔ عاملی در واقع گسترش تجزیهٔ مؤلفه‌های اصلی است. در هر دو روش تلاش بر آن است که ماتریس کوواریانس تقریب زده شود، اما این تقریب در مدل تحلیل عاملی از دقت و ظرافت بیشتری برخوردار است. به‌طور کلی هدف از تجزیهٔ عامل‌ها به شرح زیر خلاصه می‌شود:

الف) تفسیر وجود همبستگی درونی بین تعدادی صفت قابل مشاهده از طریق عواملی که قابل مشاهده نیستند و آنها را عامل گویند. در واقع این عوامل غیرقابل مشاهده دلیل مشترک همبستگی بین متغیرهای اصلی هستند؛

ب) ارائه روش ترکیب و خلاصه کردن تعداد زیادی از متغیرها در تعدادی گروه متمایز؛

ج) از بین متغیرهای مختلف تأثیرگذارترین آنها تعیین شده و در پژوهش‌های بعدی به‌طور جزیی‌تر متغیرهای تأثیرگذار را با تکرار بیشتری بررسی می کنند.

با توجه به موارد بالا، عمده‌ترین هدف استفاده از تحلیل عاملی، کاهش حجم داده‌ها و تعیین مهمترین متغیرهای مؤثر در شکل‌گیری پدیده‌هاست. از آنجا که پژوهش‌های منابع طبیعی اغلب در عرصه مراتع و جنگل‌ها انجام می‌شود و شرایط محیط تحت کنترل پژوهشگر نیست، از این‌رو اغلب با تعداد زیادی از متغیرها روبرو هستیم. در نتیجه برای کاهش حجم متغیرها می‌توان از تحلیل عاملی به‌عنوان یک روش مناسب استفاده کرد. این روش در دهه‌های اخیر به‌ویژه با پیشرفت استفاده از برنامه‌های آماری در رایانه در سطح وسیع مورد استفاده پژوهشگران قرار گرفته است.

تحلیل عاملی بر دو نوع تحلیل عاملی اکتشافی[4] و تحلیل عاملی تأییدی[5] است. در تحلیل عاملی اکتشافی، پژوهشگر در صدد کشف ساختار زیرنبایی مجموعهٔ نسبتاً بزرگی از متغیرهاست و پیش‌فرض اولیه آن است که هر متغیری ممکن است با هر عاملی ارتباط داشته باشد. به‌عبارت دیگر پژوهشگر در این روش هیچ نظریهٔ اولیه‌ای ندارد.

در تحلیل عاملی تأییدی پیش‌فرض اساسی آن است که هر عاملی با زیرمجموعهٔ خاصی از متغیرها ارتباط دارد. حداقل شرط لازم برای تحلیل عاملی تأییدی این است که پژوهشگر در مورد تعداد عامل‌های مدل، قبل از انجام تحلیل، پیش‌فرض معینی داشته باشد، ولی در عین حال پژوهشگر می‌تواند انتظارات خود مبنی بر روابط بین متغیرها و عامل‌ها را نیز در تحلیل وارد کند. کاربردهای دیگر تحلیل عاملی تأییدی عبارتند از:

- تعین اعتبار یک مدل عاملی؛
- مقایسهٔ توان دو مدل متفاوت که از داده‌ها مشابه ساخته شده‌اند؛
- آزمون معنی‌داری یک بار عاملی ویژه؛
- آزمون اینکه آیا مجموعهٔ عامل‌ها با یکدیگر همبستگی دارند یا خیر؟
- آزمون رابطهٔ بین دو یا چند بار عاملی.

دستور تحلیل عاملی تأییدی برخلاف تحلیل عاملی اکتشافی در نرم‌افزار SPSS وجود ندارد. این روش در نرم‌افزار لیزرل[6] قابل انجام است.

1- Factor analysis
2- Karl Pearson
3- Charles Spearman
1- Exploratory factor analysis
2- Confirmatory factor analysis
1- LISREL=Linear structural relationships

قبل از پرداختن به این روش آماری در نرم‌افزار SPSS، لازم است برخی از مفاهیم کلیدی این روش معرفی شوند:

- اشتراک[۷]: اشتراک عبارت از میزان واریانس مشترک بین یک متغیر با دیگر متغیرهای به‌کار گرفته شده در تحلیل است.
- مقدار ویژه[۸]: مقدار ویژه میزان واریانس تبیین‌شده به‌وسیلهٔ هر عامل را بیان می‌کند.
- عامل[۹]: عبارت است از ترکیب خطی متغیرهای اصلی که نشان‌دهندهٔ جنبه‌های خلاصه شده‌ای از متغیرهای مشاهده شده است. به عامل متغیر پنهان[۱۰] نیز گفته می‌شود.
- عامل مشترک[۱۱]: عاملی که دو یا چند متغیر بر روی آن بار می‌شوند. عامل مشترک عاملی است که حداقل بین دو متغیر مشاهده‌شده مشترک است، بنابراین، عامل مشترک در تعیین دو یا چند متغیر دخالت مستقیم دارد. به فرآیند تعیین عامل مشترک و تفسیر آن، تحلیل عاملی مشترک[۱۲] می‌گویند که نوعی روش آماری است که از همبستگی‌های بین متغیرهای مشاهده‌شده برای برآورد عامل‌های مشترک و روابط ساختاری استفاده می‌کند.
- بار عاملی[۱۳]: عبارت است از همبستگی بین متغیرهای اصلی و عوامل. اگر مقادیر بار عاملی مجذور شوند، نشان می‌دهند که چند درصد از واریانس در یک متغیر توسط آن عامل تبیین می‌شود.
- ماتریس عاملی[۱۴]: جدولی است که بارهای عاملی کلیهٔ متغیرها را در هر عامل نشان می‌دهد.
- چرخش عاملی[۱۵]: فرایندی برای تعدیل محور عاملی به‌منظور دستیابی به عامل‌های معنی‌دار و ساده است.
- نمرهٔ عاملی[۱۶]: یک مقدار ویژه برای یک عامل است که برای یک واحد نمونه‌گیری خاص محاسبه می‌شود. نمرهٔ عامل‌ها از حاصل جمع وزنی مقدار متغیرها برای آن واحد نمونه‌گیری بخصوص به‌دست می‌آید.

معادله‌های پایه

در تحلیل عاملی نمرهٔ فرد i در متغیر j را می‌توان به‌عنوان مجموع ضرایب نمره‌ها در تعداد کمتری از متغیرهای حاصل که عوامل نامیده می‌شوند، تعریف کرد. هر عامل، ترکیب خطی متغیرهاست و بر پایهٔ رابطهٔ زیر برآورد می‌شود:

$$Z_{j_i} = a_{j1}F_{1i} + \ldots + a_{jm}F_{mi} + d_j U_{ji}$$ معادلهٔ (۱)

که در آن Z_{ji}: نمرهٔ معیار فرد iام در متغیر jام است. F_{1i}: نمرهٔ معیار فرد i در اولین عامل مشترک و F_{mi}: نمرهٔ معیار وی در mامین عامل مشترک است. عبارت Uji: نمرهٔ معیار فرد i در چیزی است که عامل اختصاصی نامیده می‌شود؛ یعنی عاملی که تنها در یک متغیر واحد موجود است که در این مورد متغیر j است. ضرایب a_{jm}: بارهای عاملی هستند. اینها ضرایبی هستند که به نمره‌های عامل مشترک نسبت داده می‌شوند. ضریب d_j: وزنی است که به نمره‌های عامل اختصاصی اختصاص می‌یابد.

معادلهٔ (۱) به شکل نمرهٔ معیار است، بنابراین نمره‌های Z_{ji} و نمره‌های عاملی F_i میانگین صفر و واریانس واحد دارند. با مجذور کردن هر دو طرف معادلهٔ (۱) و سپس جمع آنها برای N مورد و تقسیم آن بر N و این فرض که نمره‌های عاملی ناهمبسته هستند، می‌توان نوشت:

$$S_j^2 = 1 = a_{j1}^2 + a_{j2}^2 + \ldots + a_{jm}^2 + d_j^2$$

این معادله نشان می‌دهد که واریانس کل را می‌توان به دو بخش جمع‌پذیر واریانس مشترک[۱۷] و واریانس منفرد[۱۸] تقسیم کرد.

1- Communality
2- Eigenvalue
3- Factor
4- Latent variable
6- Common factor
7- Common factor analysis
4- Factor loading
5- Factor matrix
1- Factor rotation
2- Factor score
3- Common variance
4- Uniqueness variance

میزان اشتراک یک متغیر که اغلب با علامت h_j^2 نشان داده می‌شود، برابر است با مجموع مجذورات بارهای عاملی مشترک.

$$h_j^2 = 1 = a_{j1}^2 + a_{j2}^2 + ... + a_{jm}^2$$

میزان اشتراک بخشی از واریانس است که می‌توان آن را به عوامل مشترک نسبت داد. بخشی از واریانس که باقی می‌ماند و نمی‌توان آن را به عوامل مشترک نسبت داد، واریانس منفرد نامیده می‌شود که با علامت d_j^2 نشان داده می‌شود. واریانس منفرد را گاهی به دو مؤلفه اختصاصی[۱۹] (b_j^2) و واریانس خطا (e_j^2) تقسیم می‌کنند. واریانس اختصاصی بخشی از واریانس کل است که به عواملی که به متغیر معینی اختصاص دارد و به خطای اندازه‌گیری ربطی ندارد، مربوط می‌شود. از آنجا که تمام اندازه‌گیری‌ها تا اندازه‌ای شامل خطا هستند، در تحلیل عاملی، بخشی از واریانس منفرد ناشی از خطای اندازه‌گیری خواهد بود.

۱-۲- مراحل اجرای تحلیل عاملی

۱-۲-۱- انتخاب متغیرهای مناسب

متغیرهایی برای تحلیل عاملی مناسب‌ترند که در سطح سنجش فاصله‌ای باشند، لکن در برخی موارد از متغیرهای رتبه‌ای و اسمی نیز استفاده می‌شود. لازم به ذکر است که پژوهشگر می‌تواند هر تعداد متغیر مرتبط با مسألهٔ تحقیق را در تحلیل وارد کند. مشروط بر آنکه متغیرها با روش درستی سنجیده شده باشند و ضریب اعتبار سنجش متغیرها در حد قابل قبولی باشد.

در مورد اندازهٔ حجم نمونه نیز به‌طورکلی در تحلیل عاملی انبوهی از داده‌ها به‌کار برده می‌شود. حداقل حجم نمونه نباید کمتر از ۵۰ باشد. هرچه حجم اندازهٔ نمونه زیادتر شود، صحت و دقت تحلیل عاملی بیشتر است. به‌عنوان یک قاعدهٔ کلی تعداد نمونه باید در حدود ۴ یا ۵ برابر تعداد متغیرهای مورد استفاده باشد. این نسبت تا حدودی محافظه‌کارانه است. در بسیاری از موارد پژوهشگر مجبور است تا با نسبت ۲ به یک نیز به تحلیل عاملی بپردازد. اما زمانی‌که این نسبت پایین و حجم نمونه نیز کم باشد، تفسیر نتایج باید با احتیاط بیشتری انجام شود.

یکی از روش‌های انتخاب متغیرهای مناسب برای تحلیل عاملی استفاده از ماتریس همبستگی است. از آنجا که روش تحلیل عاملی بر همبستگی بین متغیرها اما از نوع غیرعلّی استوار است، بنابراین در استفاده از این روش باید ماتریس همبستگی بین متغیرها نیز محاسبه شود. به‌طور معمول این گونه ماتریس‌های همبستگی وجود رابطهٔ بین برخی متغیرها و عدم ارتباط آن با برخی دیگر را نشان می‌دهند. این الگو در تحلیل عاملی موجب شکل‌گیری خوشه‌هایی می‌شود که متغیرهای درون خوشه با یکدیگر همبستگی و با متغیرهای خوشه‌های دیگر همبستگی نداشته باشند. توصیه می‌شود متغیرهایی که با هیچ متغیری همبستگی معنی‌دار نداشته باشند، از تحلیل حذف شوند.

آماره‌های دیگری نیز وجود دارند که پژوهشگر از طریق آنها نیز قادر به تعیین و تشخیص مناسب‌بودن داده‌ها برای تحلیل عاملی است. از جمله این روش‌ها استفاده از ضریب KMO[۲۰] است که مقدار آن همواره بین صفر و یک در نوسان است و از رابطهٔ زیر به‌دست می‌آید:

$$KMO = \frac{\sum\sum r_{ij}^2}{\sum\sum r_{ij}^2 + \sum\sum a_{ij}^2}$$

که در آن r_{ij}: ضریب همبستگی سادهٔ بین متغیرهای i و j و a_{ij}: ضریب همبستگی جزیی بین آنهاست. اگر مجموع ضرایب همبستگی جزیی بین همه زوج متغیرها در مقایسه با مجموع مجذورات ضرایب همبستگی کوچک باشد، اندازهٔ

1- Specificity
1- Kaiser Meyer Olkin

KMO نزدیک به یک خواهد بود. مقادیر کوچک KMO بیانگر آن است که همبستگی بین زوج متغیرها نمی‌تواند توسط متغیرهای دیگر تبیین شود، بنابراین کاربرد تحلیل عاملی متغیرها ممکن است قابل توجیه نباشد.

در صورتیکه مقدار KMO کمتر از ۰/۵ باشد، داده‌ها برای تحلیل عاملی مناسب نخواهند بود و اگر مقدار آن بین ۰/۵ تا ۰/۶۹ باشد می‌توان با احتیاط بیشتر به تحلیل عاملی پرداخت. اما در صورتیکه مقدار آن بزرگتر از ۰/۷ باشد، همبستگی‌های موجود در بین داده‌ها برای تحلیل عاملی مناسب خواهد بود (جدول ۱).

جدول ۱- قضاوت در مورد ضریب KMO

مقدار KMO	تناسب داده‌ها برای تحلیل عاملی
بزرگتر یا مساوی ۰/۹۰	عالی
۰/۸۰-۰/۸۹	خیلی خوب
۰/۷۰-۰/۷۹	خوب
۰/۶۰-۰/۶۹	متوسط
۰/۵۹-۰/۵	ضعیف
کمتر از ۰/۵۰	غیرقابل پذیرش

برای اطمینان از مناسب بودن داده‌ها برای تحلیل عاملی افزون بر اینکه ماتریس همبستگی‌هایی که پایه تحلیل قرار می‌گیرند در جامعه برابر صفر نیست، باید از آزمون کرویت بارتلت[۲۱] بر اساس فرمول زیر استفاده کرد:

$$\chi^2 = -(n-1-\frac{2p+5}{6})\ln|R|$$

که در آن n: معرف تعداد آزمودنی‌ها، p: تعداد متغیرها، $|R|$: قدر مطلق دترمینان ماتریس همبستگی است. این آماره که دارای توزیع مربع‌کای با $0.5p(p-1)$ درجهٔ آزادی است. مقدار اطلاعات موجود در $|R|$ را با بررسی رابطهٔ بین تعداد مشاهده‌ها و تعداد متغیرها ارزشیابی می‌کند و احتمال خطا را برای رد کردن فرضیهٔ صفر عدم وجود تفاوت از ماتریس همانی[۲۲] می‌آزماید. ماتریس همانی ماتریسی است که همه عناصر قطری آن یک و همه عناصر غیرقطری آن صفر باشد.

آزمون بارتلت این فرضیه را که ماتریس همبستگی‌های مشاهده شده متعلق به جامعه‌ای با متغیرهای ناهمبسته است می‌آزماید. برای آنکه یک مدل عاملی مفید و دارای معنا باشد، لازم است متغیرها همبسته باشند، در غیر این صورت دلیلی برای تبیین مدل عاملی وجود ندارد. اگر فرضیهٔ «متغیرها با هم رابطه ندارند» رد نشود، کاربرد تحلیل عاملی زیر سئوال خواهد رفت، بنابراین باید در آن تجدید نظر کرد. مربع‌کای معنی‌دار بیانگر حداقل شرایط لازم برای اجرای تحلیل عاملی است.

۱-۲-۲- استخراج عامل‌ها

همانطور که در قبل نیز گفته شد هدف تحلیل عاملی خلاصه کردن متغیرها در تعدادی عامل است. پس برای انجام تحلیل عاملی باید روش استخراج عامل‌ها و معیار تعیین آنها مشخص شود.

الف) روش استخراج عامل‌ها: برای استخراج عامل‌ها روش‌های مختلفی وجود دارد که برحسب مقدار و نوع واریانسی که توسط متغیرهای هر عامل در مدل توجیه می‌شود، متفاوتند. اساسی‌ترین این روش‌ها تجزیهٔ مؤلفه‌های اصلی است.

ذکر این نکته ضروری است که در تحلیل عاملی سه واریانس وجود دارد؛ واریانس مشترک که به نسبتی از واریانس گفته می‌شود که به‌وسیله عامل‌های مشترک تبیین می‌شود. واریانس خاص که به یک متغیر خاص مربوط می‌شود و واریانس خطا که ناشی از بی‌اعتباری و ناپایایی داده‌های جمع‌آوری شده است.

1- Bartlett's test of sphericity
2- Identify matrix

در روش تجزیهٔ مؤلفه‌های اصلی، عامل‌ها همهٔ واریانس هر متغیر از جمله واریانس مشترک با سایر متغیرهای مجموعه و نیز واریانس خاص متغیر را توجیه می‌کنند. پس تعداد عامل‌ها در این روش از نظر تئوری باید با تعداد متغیرها برابر باشد، زیرا همهٔ واریانس هر متغیر باید توسط عامل‌ها تبیین شود. به‌عبارت دیگر در تجزیهٔ مؤلفه‌های اصلی به تعداد متغیرها، مؤلفه وجود دارد، ولی عامل‌هایی استخراج می‌شوند که بیشترین مقدار واریانس را تبیین کنند.

ب) معیار تعیین عامل‌ها: استخراج عامل‌ها با توجه به معیارهای زیر انجام می‌شود:

۱) معیار مقدار ویژه[۲۳]: هر عامل شامل یک یا چند متغیر است. مجذورات بارهای یک عامل نشان‌دهندهٔ درصدی از واریانس ماتریس همبستگی است که به‌وسیلهٔ آن عامل تبیین می‌شود، این مقدار را مقدار ویژه نامند. برای محاسبهٔ آن کافی است ضریب همبستگی متغیرها را با یک عامل به توان برسانیم و با هم جمع کنیم تا مقدار ویژه آن عامل به‌دست آید. هر چه مقدار ویژهٔ یک عامل بیشتر باشد، آن عامل واریانس بیشتری را تبیین می‌کند.

بر این اساس تعداد عامل‌ها با توجه به مقدار ویژهٔ هر عامل مشخص می‌شود و عامل‌هایی که مقدار ویژه آنها بیشتر از یک باشد، به‌عنوان عامل‌های معنی‌دار در نظر گرفته می‌شود.

استفاده از این معیار زمانی که تعداد متغیرها بین ۲۰ تا ۵۰ باشد، قابل اعتماد به نظر می‌رسد، اما اگر تعداد متغیرها کمتر از ۲۰ باشد، استفاده از این معیار باید با محافظه‌کاری انجام شود. همچنین اگر تعداد متغیرها بیش از ۵۰ باشد، استفاده از این معیار موجب استخراج تعداد زیادی عامل می‌شود (Hair، ۱۹۹۰).

۲) معیار پیشین[۲۴]: این روش زمانی مورد استفاده قرار می‌گیرد که تعداد عامل‌ها را پژوهشگر مشخص می‌کند.

۳) معیار تست بریدگی: این معیار عامل‌ها را بر مبنایی تعیین می‌کند که هنوز میزان واریانس خاص بر واریانس مشترک غلبه نکرده باشد، بنابراین تا زمانی که مقدار واریانس مشترک بیشتر از مقدار واریانس خاص باشد، عامل‌های معنی‌دار استخراج می‌شود. برای تعیین تعداد عامل‌ها بر اساس این معیار، نمودار مقدار ویژه در برابر تعداد عامل‌ها رسم می‌شود.

۴) معیار درصد واریانس تجمعی: در این حالت درصد واریانس تبیین‌شده مبنای تصمیم‌گیری قرار می‌گیرد و عامل‌هایی استخراج می شوند که درصد واریانس بالایی را در بر داشته باشند. چنانچه مقدار واریانس کمتر از ۵۰ درصد باشد، باید متغیرهایی را که میزان اشتراک آنها کم است، حذف کرد.

ذکر این نکته لازم است که برای انتخاب تعداد عامل‌های مناسب علاوه بر معیارهای گفته‌شده، دو معیار حداقل میانگین همبستگی‌های جزئی ولیستر[۲۵] و تحلیل موازی[۲۶] نیز وجود دارند که گرچه نتایج آنها دقیق و استفاده از آنها نیز آسان است، اما در نرم‌افزار SPSS وجود ندارد، از این‌رو در این کتاب به آنها پرداخته نشده است.

۱-۲-۳- تعیین متغیرهای هر عامل (تفسیر ماتریس عاملی)

در ماتریس عاملی هر ستون معرف یک عامل است. مقادیر هر ستون نشان‌دهندهٔ بارهای عاملی هر متغیر با یک عامل هستند. در خروجی نرم‌افزار عامل‌ها به‌ترتیب از چپ به راست با شماره‌های ۱، ۲، ۳ و الی آخر قرار می‌گیرند. متغیرها نیز در ستون اول از بالا به پایین فهرست می‌شوند. برای شروع تفسیر، پژوهشگر باید از اولین متغیر شروع کند و مقادیر مربوط به آن را در عامل‌های مختلف بررسی کند. هر جا که بیشترین مقدار مطلق بار عاملی وجود داشته باشد و از نظر آماری نیز معنی‌دار باشد، زیر آن خط بکشد. به همین ترتیب مراحل باید برای متغیرهای دیگر نیز انجام شود. در برخی مواقع ممکن است یک متغیر بر بیش از یک عامل بار شده باشد که این از موارد پیچیده و بغرنج در تحلیل عاملی است. اگر چه در

1- Eigenvalue Criterion
1- A Prior Criterion
2- Velicer' minimum average partial correlations (MAP)
3- Parallel analysis

بسیاری از موارد چرخش عامل‌ها بخشی از این گونه مشکلات را مرتفع می‌کند، اما در برخی مواقع این گونه مشکلات هنوز بدون راه حل باقیمانده است.

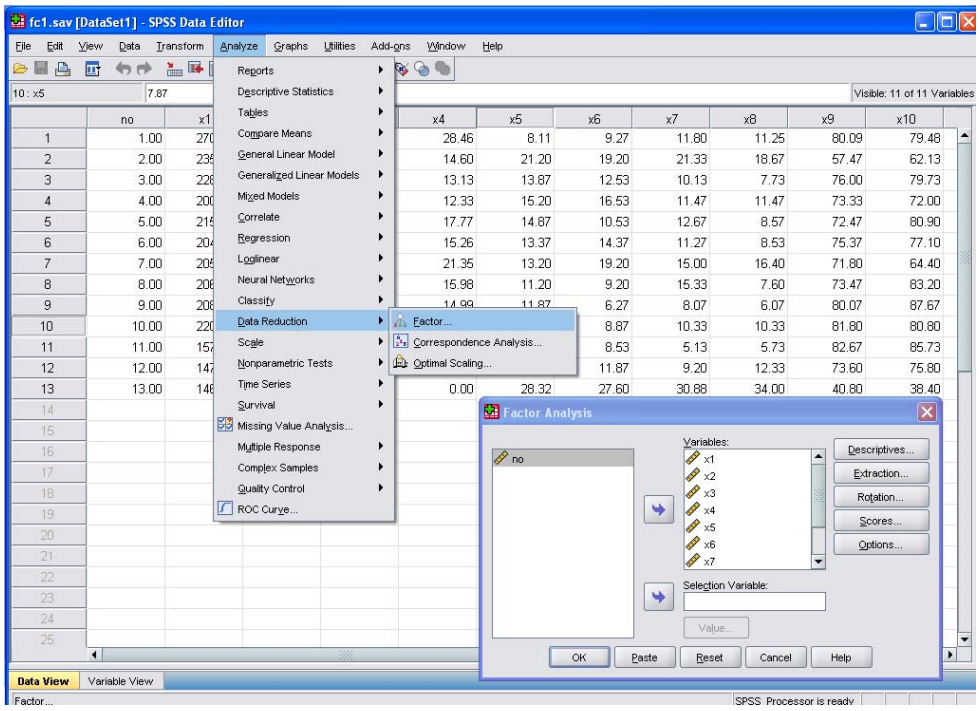
هنگامی که با بررسی ماتریس عاملی، بارهای عاملی معنی‌دار مشخص شدند، باید متغیرهایی که بر روی هیچ یک از عامل‌ها بار عاملی معنی‌دار ندارند نیز مشخص شوند. پژوهشگر می‌تواند به دو شیوه با متغیرهایی که با هیچ کدام از عامل‌ها همبستگی معنی‌دار ندارند، برخورد کند. شیوهٔ اول آن است که این متغیرها را به فراموشی سپرده و تنها متغیرهای معنی‌دار را تفسیر کند. شیوهٔ دوم آنکه پژوهشگر با این استدلال که همه متغیرها سهمی حتی کوچک در نتایج داشته‌اند، بنابراین برای رفع اثرات متغیرهایی که بار عاملی معنی‌دار نداشته‌اند، آنها را از تحلیل حذف و سپس تحلیل عاملی را بر اساس متغیرهای معنی‌دار تکرار کرده و نتایج را تفسیر کند. با تشخیص متغیرهای معنی‌دار هر عامل می‌توان نام مناسبی با توجه به نوع متغیرهای هر عامل و ضرایب آنها برای عامل‌ها تعیین کرد.

۱-۳- اجرای تحلیل عاملی در نرم‌افزار SPSS

برای انجام تحلیل عاملی در نرم‌افزار SPSS از روند زیر استفاده می‌شود:

Analyze>Data reduction>Factor analysis

بعد از انجام فرمان بالا، پنجرهٔ Factor Analysis ظاهر می‌شود (شکل ۱). متغیرها را در جعبهٔ Variables وارد می‌کنند.



شکل ۱- روند اجرای فرمان Factor Analysis

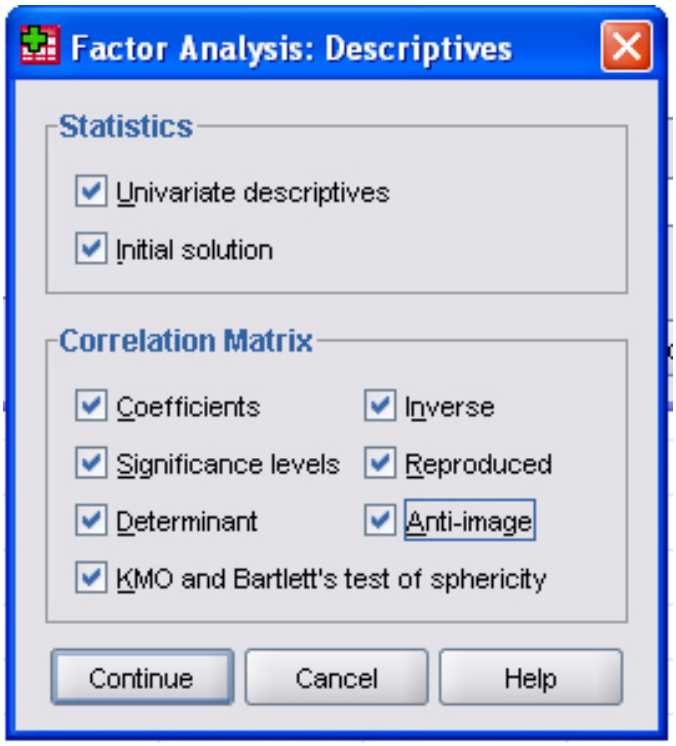
۱-۳-۱- محاسبهٔ آماره‌های توصیفی

با فعال کردن کلید Descriptive، امکان محاسبهٔ آماره‌های زیر فراهم می‌شود (شکل ۲):

- در قسمت بالای این کادر، در بخش Statistics با انتخاب گزینهٔ Univariate Descriptive آماره‌های تک متغیره از قبیل میانگین، انحراف معیار و تعداد مشاهده‌های مورد استفاده محاسبه می‌شود. با انتخاب گزینهٔ Initial Solution برآوردهای اولیه‌ای از عامل‌ها محاسبه می‌شود. خروجی این فرمان برآورد اولیه‌ای از میزان اشتراک‌ها، مقادیر ویژهٔ ماتریس همبستگی متغیرها، درصد کل واریانس توضیح داده شده به‌وسیلهٔ عامل‌های مشترک و نیز درصد تجمعی واریانس عامل‌ها را ارائه می‌دهد.

در قسمت پایین کادر، مستطیل دیگری تحت عنوان Correlation Matrix وجود دارد که گزینه‌های زیر در آن قرار دارند:

- با انتخاب گزینهٔ Coefficient ماتریس ضرایب همبستگی متغیرهای انتخابی به‌صورت یک ماتریس مثلثی شکل محاسبه می‌شود که عناصر قطر اصلی آن یک است. هر چه عناصر خارج از قطر اصلی این ماتریس به یک نزدیکتر باشد، مطلوبیت روش تحلیل عاملی بیشتر خواهد بود.
- با انتخاب گزینهٔ Significance Levels سطح معنی‌داری متناظر با ماتریس همبستگی به‌دست آمده محاسبه می‌شود.
- با انتخاب گزینهٔ Determinant دترمینان ماتریس ضرایب همبستگی محاسبه می‌شود. هر قدر مقدار به‌دست آمده کمتر باشد، انجام تحلیل عاملی معتبرتر خواهد بود.
- با انتخاب گزینهٔ KMO and Bartlett's test of sphericity کفایت اندازهٔ نمونه با استفاده از آزمون‌های KMO و بارتلت تعیین می‌شود. به‌طور کلی، این گزینه شاخصی برای مقایسهٔ مقادیر ضرایب همبستگی ساده و جزیی بر روی همه متغیرهاست. مقادیر بزرگ KMO بر رضایت‌بخش بودن تحلیل عاملی دلالت می‌کند و آزمون بارتلت نیز فرض یکّه‌بودن ماتریس ضرایب همبستگی را آزمون می‌کند، به‌طوریکه اگر آزمون بارتلت معنادار نباشد (احتمال مربوطه بزرگتر از ۰/۵ باشد)، این امکان برای ماتریس همبستگی وجود دارد که یک ماتریس یکّه باشد.این امر به معنای آن است که ماتریس مذکور برای تحلیل‌های بعدی مناسب نیست.
- با انتخاب گزینهٔ Inverse معکوس ماتریس ضرایب همبستگی متغیرها محاسبه می‌شود.
- با انتخاب گزینهٔ Reproduced ماتریسی محاسبه می‌شود که عناصر پایین قطر اصلی آن ضرایب همبستگی تبدیل‌یافتهٔ بین متغیرها، عناصر قطر اصلی میزان اشتراکات عامل‌ها و عناصر بالای قطر اصلی باقیمانده‌های روش تحلیل عاملی هستند. در مجموع روش تحلیل عاملی زمانی مفید است که مانده‌ها در این ماتریس کوچک و نزدیک به صفر باشند.
- با انتخاب گزینهٔ Anti-image ماتریسی محاسبه می‌شود که عناصر آن ضرایب همبستگی جزیی با علامت مخالف و عناصر قطر اصلی این ماتریس بیانگر دقت نمونه‌گیری هستند. این مقادیر ورود متغیرها را به مدل تأیید می‌کنند.



شکل ۲- کادر Descriptives فرمان Factor analysis

۱-۳-۲- استخراج عوامل

برای تعیین روش استخراج عوامل روی کلید Extraction در کادر اصلی کلیک کرده تا کادری به‌نام Factor Analysis: Extraction ظاهر شود (شکل ۳). در سمت چپ این کادر که با عنوان Method مشخص شده است، روش‌های مختلف استخراج عامل‌ها عبارتند از:

- مؤلفه‌های اصلی (PC)[۲۷]
- حداکثر درستنمایی (ML)[۲۸]
- عامل‌یابی محور اصلی (PAF)[۲۹]
- حداقل مربعات غیروزنی (ULS)[۳۰]
- عامل‌یابی آلفا (AF)[۳۱]
- عامل‌یابی تصویری (IF)[۳۲]

برای انتخاب هر کدام از روش‌های بالا باید بر روی علامت فلش که در سمت راست جعبه قرار دارد، کلیک کرد تا فهرست آنها نمایان شود. سپس بر روی روش مورد نظر کلیک تا به هنگام اجرای روش تحلیل عاملی از آن روش بهره گرفته شود. در این مثال روش مؤلفه‌های اصلی انتخاب می‌شود.

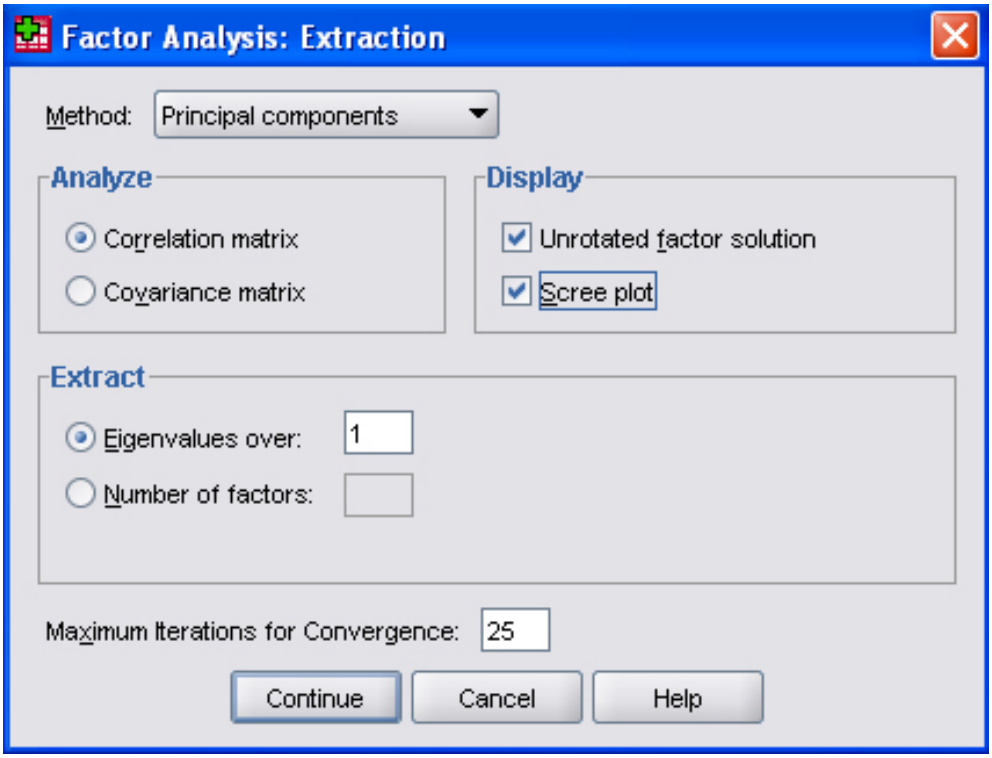
نکته: در صورتیکه هدف پژوهشگر خلاصه‌کردن متغیرها و دستیابی به تعداد محدودی عامل باشد، از روش مؤلفه‌های اصلی استفاده می‌شود.

در بخش Extract این کادر گزینه‌هایی وجود دارد که در آنها چگونگی انتخاب تعداد عامل‌ها مشخص می‌شود. با انتخاب گزینهٔ Eigenvalue over تعداد عامل‌هایی که مقادیر ویژهٔ متناظر آنها از عدد تعیین شده در جعبهٔ مقابل آن بیشتر است، استخراج می‌شوند. همچنین اگر گزینهٔNumber of Factors انتخاب شود، تعداد عامل‌ها را می‌توان به دلخواه در کادر مقابل این گزینه تعیین کرد.

در مستطیل سمت راست این کادر که با نام Display مشخص شده است، می‌توان ماتریس ضرایب عامل‌های غیردورانی را محاسبه و نموداری از مقادیر ویژه را برای عامل‌های مختلف رسم کرد. اگر گزینهٔ Unrotated factor solution انتخاب شود، ضرایب عامل‌ها قبل از دوران محاسبه می‌شود. با انتخاب گزینهٔ Scree plot نموداری رسم می‌شود که مقادیر ویژهٔ عوامل انتخابی از بزرگترین تا کوچکترین مقدار را شامل می‌شود.

آخرین گزینهٔ این کادر Maximum Iterations for Convergence نام دارد که با استفاده از آن می‌توان حداکثر تعداد دفعات تکرار برای همگرایی مدل را تعیین کرد.

1- Principle Components
2- Maximum Likelihood
1- Principle-axis Factoring
1- Unweighted Least Squares
3- Alpha Factoring
4- Image Factoring



شکل ۳- کادر Extraction فرمان Factor analysis

۱-۳-۳- دوران عامل‌ها

در این مرحله به‌منظور بهبود روابط بین متغیرها و عامل‌های اولیه و اعمال تبدیلات خاص بر روی عامل‌ها، عمل دوران انجام می‌شود. با فعال کردن کلید Rotation در کادر اصلی، کادرFactor Analysis: Rotation ظاهر می‌شود که شامل بخش‌های زیر است (شکل ۴):

بخش Method روش‌های مختلف دوران را شامل می‌شود. به‌طور کلی روش‌های دوران به دو دسته متعامد و غیرمتعامد (همبسته) تقسیم می‌شوند. اگر پژوهشگر بخواهد تعداد زیادی متغیر مورد بررسی را به یک مجموعه کوچکتر از متغیرهای غیرمرتبط با هم تقلیل دهد، روش متعامد مناسب‌تر است. اما اگر هدف اصلی تحلیل عاملی به‌دست آوردن چند عامل باشد که از نظر تئوریکی معنی‌دار باشد، روش غیرمتعامد مناسب‌تر است.

با انتخاب فرمان None هیچ نوع دورانی بر روی ماتریس ضرایب اعمال نمی‌شود.

فرمان Varimax از جمله متداول‌ترین روش‌های دوران متعامد است که استقلال میان عامل‌های استخراجی را حفظ می‌کند. این روش متغیرهای دارای بار عاملی بزرگتر را به کمترین تعداد تقلیل می‌دهد. این روش، جمع واریانس بارها در ماتریس عاملی را بیشترین مقدار می‌کند، به همین دلیل آن را واریماکس گویند. هنگامی از این روش استفاده می‌شود که هدف به‌دست آوردن عامل‌هایی است که دارای بار زیادی بر روی برخی از متغیرها و بار کم بر روی متغیرهای دیگر باشد.

در این روش تأکید بر ساده‌کردن ستون‌های ماتریس عاملی است، یعنی حداکثر امکان ساده کردن تا آنجایی حاصل می‌شود که بر روی یک ستون خاص ماتریس، فقط مقادیر (بارهای عاملی) صفر و یک قرار بگیرد. از این‌رو، مجموع تغییرات ایجاد شده در بارهای عاملی به حداکثر می‌رسد. در این حالت تفسیر عامل‌ها ساده می‌شود.

فرمان Quartimax هنگامی مورد استفاده قرار می‌گیرد که مدل تحلیل عاملی به تعداد عوامل کمتری وابسته باشد. این روش نیز از جمله روش‌های دوران متعامد است. هدف اصلی روش Quartimax ساده کردن سطرهای ماتریس عاملی است. یعنی اگر بار عاملی متغیر در یک عامل زیاد باشد، تا آنجاییکه ممکن است از تعلق متغیر به عامل‌های دیگر کاسته می‌شود. به‌عبارت دیگر بار عاملی‌اش در دیگر عامل‌ها پایین می‌آید. مشکل عمدهٔ این روش این است که در چرخش عامل‌ها تمایل به ایجاد یک عامل کلی و بزرگ وجود دارد.

فرمان Equamax نیز یکی از روش‌های متعامد است که در برگیرندهٔ اهداف هر دو روش فوق است. بدین معنی که هم سطرها و هم ستون‌ها ساده می‌شود.

در مجموع، روش‌های متعامد فوق مقادیر نسبتاً بزرگ (از نظر قدر مطلق) یا صفر به ستون‌های ماتریس ضرایب عامل‌ها اختصاص می‌دهند. در چنین شرایطی عوامل به‌دست آمده یا با متغیرهای انتخابی وابستگی زیادی داشته یا از آنها کاملاً مستقل خواهند بود.

فرمان Direct oblimin از جمله روش‌های دوران غیرمتعامد یا مورب است. ویژگی این روش ساده‌تر بودن تفسیر عامل‌هاست، در حالیکه عامل‌های به‌دست آمده از آن در این حالت مستقل نخواهند بود. با انتخاب این روش، مقدار دلتا در جعبه Delta مشخص می‌شود، به‌طوریکه در حالت دلتا برابر با صفر وابستگی بین عامل‌های تبدیل یافته به حداکثر مقدار خود خواهد رسید، در حالیکه به ازای مقادیر منفی با قدر مطلق بزرگ از وابستگی عامل‌ها کاسته شده و نتایج به‌دست آمده به نتایج روش‌های متعامد نزدیک می‌شود.

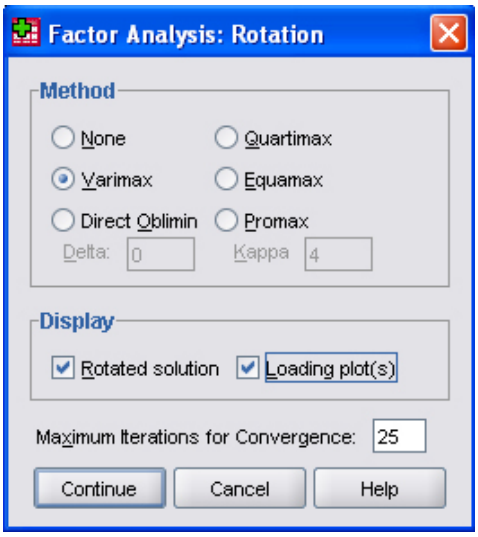
برای دستیابی به یک ساختار ساده، اگر راه حل متعامد انتخاب شود، روش Varimax بهترین روش است. اما در صورت انتخاب راه حل متمایل، چرخش Direct oblimin بهترین روش است. در این مثال روش Varimax انتخاب شد.

نکته: فیلد (۲۰۰۰) معتقد است که انتخاب نوع چرخش عاملی به این بستگی دارد که آیا دلیل نظری از یک طرف بر فرض همبستگی یا استقلال عاملها از همدیگر و از طرف دیگر برای چگونگی خوشه‌بندی متغیرها بر روی عاملها قبل از چرخش وجود دارد یا خیر؟ یک روش قابل قبول آن است که ابتدا هر دو روش اجرا شود. سپس نوع مناسب آن انتخاب شود.

در بخش Display گزینه‌های زیر وجود دارد:

- در صورت انتخاب یکی از روش‌های دوران، گزینهٔ Rotated Solution قابل انتخاب بوده و با انتخاب آن ماتریس ضرایب عامل‌های دوران یافته، ماتریس تبدیل عوامل، ماتریس الگو و ماتریس همبستگی بین برآوردهای دوران‌یافتهٔ عامل‌های مشترک محاسبه خواهد شد.
- با انتخاب گزینهٔ Loading plot(s) نمودار سه بعدی از ضرایب متغیرهای موجود در ماتریس الگو رسم خواهد شد.

در پایین این کادر گزینهٔ Maximum Iterations for Convergence وجود دارد که حداکثر تعداد دفعات تکرار فرآیند را برای همگرایی روش تحلیل عاملی مشخص می‌کند.



شکل ۴- کادر Rotation فرمان Factor analysis

۱-۳-۴- محاسبهٔ نمره‌های عاملی

برای محاسبهٔ نمره‌های عاملی بر روی کلید Scores در کادر اصلی کلیک تا کادری به‌نام Factor Analysis: Factor Scores ظاهر شود (شکل ۵). اولین گزینهٔ این کادر Save as Variables است که با انتخاب آن می‌توان نمره‌های عاملی به‌دست آمده برای مشاهده‌ها را به‌صورت متغیرهای جدید در کنار داده‌های اولیه ذخیره کرد.

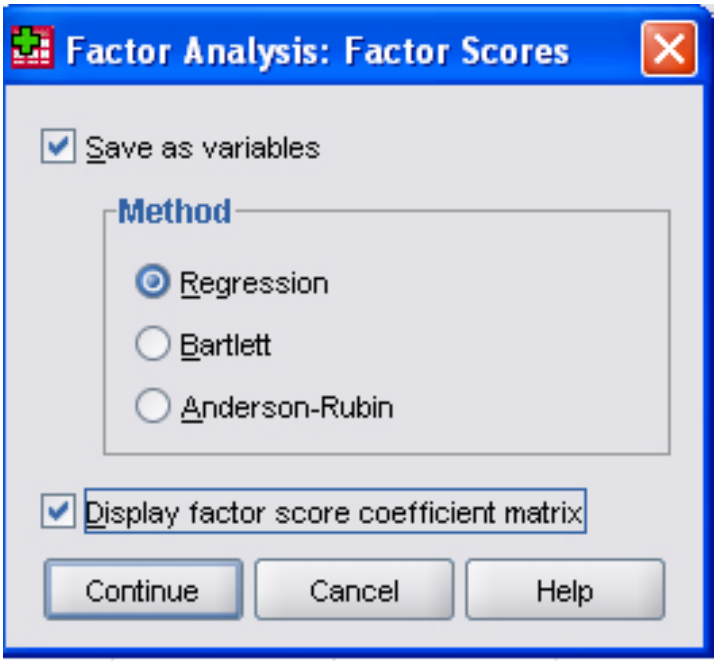
روش برآورد نمره‌های عاملی در بخش Method تعیین می‌شود. در برنامهٔ SPSS سه روش منظور شده است. در روش رگرسیون بدون توجه به نوع دوران انتخابی، تعداد عامل‌های وابسته برآورد می‌شوند. در این روش، حتی زمانی که عامل‌ها مستقل فرض می‌شوند، می‌توانند با همدیگر همبستگی داشته باشند. شایان ذکر است که اگر روش خاصی برای برآورد انتخاب نشود، روش رگرسیونی به‌صورت خودکار اعمال می‌شود.

در روش Bartlett برای برآورد نمره‌های عاملی از روش حداقل مربعات وزنی استفاده می‌شود. روش Anderson-Rubin برای هر عامل نمره‌هایی با انحراف معیار یک محاسبه می‌کند و عامل‌ها مستقل از یکدیگرند. در این روش، به منظور برآورد ضرایب، روش حداقل مربعات معمولی به‌کار گرفته می‌شود.

لازم به ذکر است که در صورت اعمال روش مؤلفه‌های اصلی، برای استخراج عوامل، نمره‌های عاملی حاصل از انتخاب هر سه روش یکسان خواهند بود.

آخرین گزینهٔ موجود در این کادر Display Factor Score Coefficient Matrix است که انتخاب آن برآورد ماتریس ضرایب عامل‌های به‌دست آمده را نشان می‌دهد.

در ادامه پس از انتخاب گزینه‌های مورد نیاز بر روی کلید Continue کلیک تا بار دیگر کادر اصلی ظاهر شود.



شکل ۵- کادر Factor Scores فرمان Factor analysis

با کلیک کردن بر روی گزینهٔ Options کادر اصلی شکل (۵-۶) ظاهر می‌شود که در بخشMissing Values آن چگونگی برخورد با داده‌های گمشده مشخص می‌شود. در این بخش سه گزینه زیر وجود دارد:

- گزینهٔ Exclude cases listwise برای حذف مشاهده‌هایی به‌کار می‌رود که در یکی از متغیرهای خود دارای دادهٔ گمشده هستند.

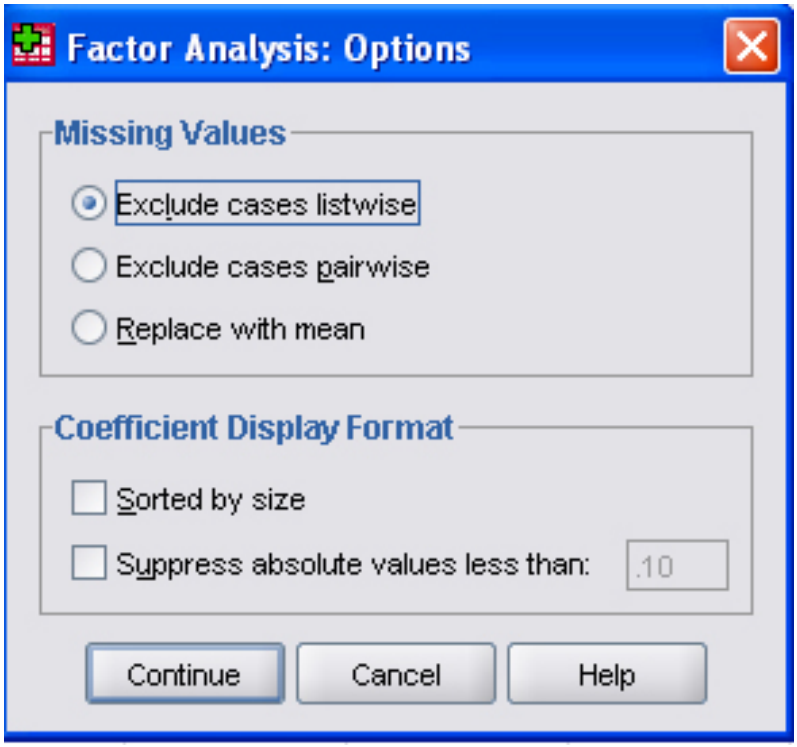
- گزینهٔ Exclude cases pairwise برای حذف مشاهده‌هایی به‌کار می‌رود که یک یا هر دو آنها دارای داده گمشده‌اند. این فرمان بر روی عملیاتی قابل اجراست که از دو متغیر به‌طور همزمان استفاده می‌شود.

- گزینهٔ Replace with Mean نه تنها مشاهده‌های دارای دادهٔ گمشده را حذف نمی‌کند، بلکه داده‌های گمشده را با میانگین دیگر مشاهده‌ها جایگزین می‌کند.

در بخش Coefficient Display Format دو گزینهٔ زیر وجود دارد:

- گزینهٔ Sorted by size متغیرهایی که ضرایب عامل یا همبستگی بالایی دارند، در یک گروه جای می‌دهد و آنها را به‌ترتیب از بزرگ به کوچک مرتب می‌کند.

- گزینهٔ Suppress absolute values less than ضرایبی را که قدر مطلق آنها از مقدار تعیین شده در جعبهٔ مقابل آن کوچکتر است، از ماتریس ضرایب حذف می‌کند. با توجه به آنکه ضرایب عاملی بین صفر و یک هستند، عدد مندرج در جعبهٔ مذکور باید در همین دامنه قرار داشته باشد.



شکل ۶- کادر Options فرمان Factor analysis

مثال ۱) در صورتیکه تحلیل عاملی بر روی ماتریسی شامل ۱۳ متغیر در ۱۵۷ واحد نمونه انجام شود و هدف تشخیص مهمترین متغیرها باشد، تفسیر نتایج به‌صورت زیر است:

با توجه به جدول (۲) چون مقدار آمارهٔ KMO برابر ۰/۸۰۸است پس داده‌ها برای انجام تحلیل عاملی مناسب‌اند. همچنین نتایج آزمون کرویت بارتلت نیز معنی‌دار است، به این مفهوم که فرض مخالف تأیید می‌شود، یعنی بین متغیرها همبستگی معنی‌دار وجود دارد.

جدول ۲- آمارهٔ KMO و نتایج آزمون کرویت بارتلت

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.808
Bartlett's Test of Sphericity	Approx. Chi-Square df Sig.	1759.665 78 .000

جدول (۳) میزان اشتراک متغیرها یا واریانس کل با میزان اشتراک عاملی متغیرها را نشان می‌دهد. برای مثال ملاحظه می‌شود که ۹۲/۶ درصد واریانس امتیازات متغیر x3، واریانس عامل مشترک است. Initial گویای تمامی اشتراک‌های قبل از

استخراج است، بنابراین تمامی آنها برابر با یک هستند. همان گونه که در جدول زیر مشاهده می‌شود بیشتر میزان اشتراک‌ها بالاتر از ۵۰ درصد است و بیانگر توانایی عامل‌های تعیین شده در تبیین واریانس متغیرهای مورد مطالعه است. با وجود این در بین مقادیر اشتراک، تفاوت‌هایی نیز مشاهده می‌شود. برای مثال مقدار اشتراک مربوط به متغیر x_1، ۰/۵۴ و برای متغیر x_3، ۰/۹۲۶ است.

جدول ۳- میزان اشتراک اولیه و بعد از استخراج عامل‌ها برای متغیرهای وارد شده در تحلیل عاملی

Communalities

	Initial	Extraction
x1	1.000	.540
x2	1.000	.808
x3	1.000	.926
x4	1.000	.870
x5	1.000	.801
x6	1.000	.885
x7	1.000	.825
x8	1.000	.758
x9	1.000	.865
x10	1.000	.888
x11	1.000	.867
x12	1.000	.853
x13	1.000	.698

Extraction Method: Principal Component Analysis.

جدول (۴) مقدار ویژه و واریانس متناظر با عامل‌ها را نشان می‌دهد. در ستونInitial Eigenvalues مقادیر ویژهٔ اولیه برای هر یک از عامل‌ها در قالب مجموع واریانس تبیین‌شده برآورد می‌شود. واریانس تبیین‌شده برحسب درصدی از کل واریانس و درصد تجمعی است.

مقدار ویژهٔ هر عامل، نسبتی از واریانس کل متغیرهاست که توسط آن عامل تبیین می‌شود. مقدار ویژه از طریق مجموع مجذورات بارهای عاملی مربوط به تمام متغیرها در آن عامل قابل محاسبه است، از این‌رو مقادیر ویژه، اهمیت اکتشافی عامل‌ها را در ارتباط با متغیرها نشان می‌دهد. پایین‌بودن این مقدار برای یک عامل به این معنی است که آن عامل نقش اندکی در تبیین واریانس متغیرها داشته است. در ستون Extraction Sums of Squared Loadings واریانس تبیین‌شدهٔ عامل‌هایی ارائه شده است که مقادیر ویژهٔ آنها بزرگتر از عدد یک باشد. ستون Rotation Sums of Squared Loadings مجموعهٔ مقادیر عامل‌های استخراج‌شده بعد از چرخش را نشان می‌دهد. همچنانکه مشاهده می‌شود سه عامل قابلیت تبیین واریانس‌ها را دارند. اگر عامل‌های به‌دست آمده را با روش Varimax چرخش دهیم، عامل‌های اول، دوم و سوم به‌ترتیب ۳۲/۴، ۲۹/۵ و ۱۹/۵ و در مجموع ۸۱/۴ درصد از واریانس را در بردارند.

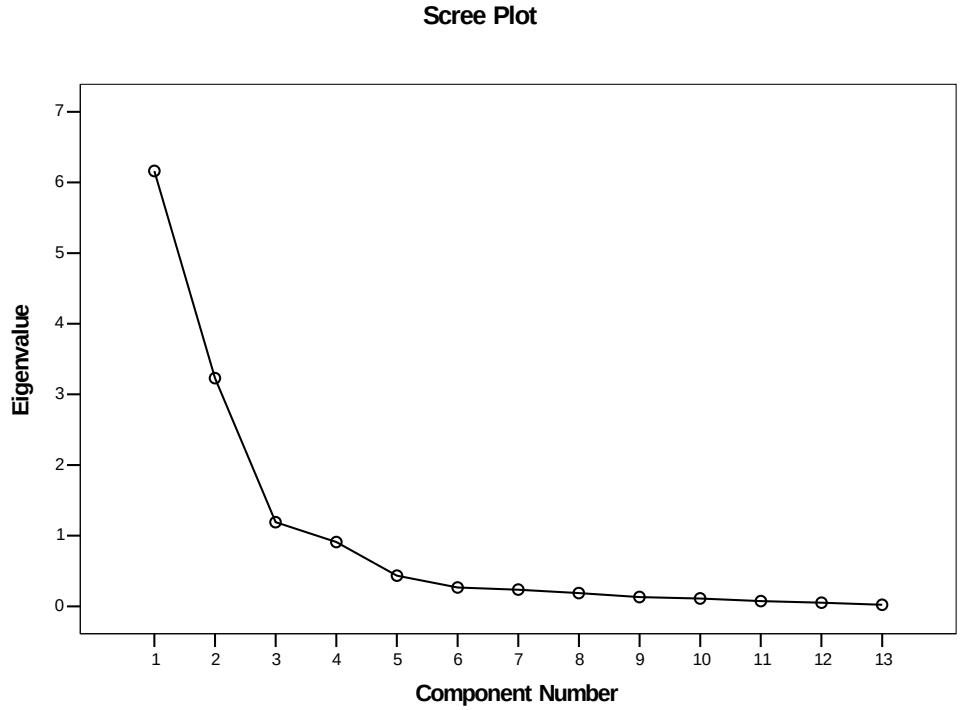
جدول ۴- درصد واریانس و مقادیر ویژه عامل‌های مختلف

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.162	47.403	47.403	6.162	47.403	47.403	4.213	32.411	32.411
2	3.230	24.843	72.247	3.230	24.843	72.247	3.839	29.534	61.944
3	1.190	9.156	81.403	1.190	9.156	81.403	2.530	19.459	81.403
4	.909	6.993	88.396						
5	.432	3.326	91.722						
6	.266	2.050	93.772						
7	.236	1.813	95.585						
8	.187	1.439	97.024						
9	.130	1.003	98.027						
10	.111	.850	98.877						
11	.074	.567	99.445						
12	.051	.389	99.833						
13	.022	.167	100.000						

Extraction Method: Principal Component Analysis.

شکل (۷) تغییرات مقادیر ویژه را در ارتباط با عامل‌ها نشان می‌دهد. این نمودار برای تعیین تعداد بهینه مؤلفه‌ها به‌کار می‌رود. با توجه به این نمودار مشاهده می‌شود که از عامل سوم به بعد تغییرات مقدار ویژه کم می‌شود، پس می‌توان سه عامل را به‌عنوان عوامل مهم که بیشترین نقش را در تبیین واریانس داده‌ها دارند، استخراج کرد.



شکل ۷- نمودار اسکری‌گراف برای تعیین تعداد عامل‌ها

جدول (۵) سهم متغیرها را در عامل‌ها قبل از چرخش نشان می‌دهد. اگر بارهای عاملی جلوی هر متغیر را به توان دو رسانده و با هم جمع کنیم، ارقام جدول (۳) ستون Extraction به‌دست می‌آید. این ضرایب از یک سو نشان‌دهندهٔ توانایی عامل‌های تعیین‌شده در تبیین واریانس متغیرهای مورد مطالعه و از سویی می‌تواند برای بررسی تناسب متغیرها برای تحلیل عاملی استفاده شود.

جدول ۵- ماتریس عاملی دوران نیافته

Component Matrix [a]

	Component		
	1	**2**	**3**
x1	.090	.713	.152
x2	.538	-.720	-.020
x3	.374	.469	-.752
x4	.659	-.658	.053
x5	.872	-.165	.117
x6	.785	-.480	.196
x7	.664	.587	.200
x8	.797	.246	.249
x9	.704	.399	.458
x10	.926	.132	-.112
x11	.860	.200	-.297
x12	-.857	-.037	.343
x13	-.110	.818	.127

Extraction Method: Principal Component Analysis.
a 3 components extracted.

جدول (۶) سهم متغیرها را در عامل‌ها بعد از چرخش نشان می‌دهد. هر متغیر در عاملی قرار می‌گیرد که با آن عامل همبستگی بالای معنی‌داری داشته باشد.

جدول ۶- ماتریس عاملی دوران یافته

Rotated Component Matrix [a]

	Component		
	1	**2**	**3**
x1	.419	**-.595**	.102
x2	.117	**.886**	.095
x3	.036	-.242	**.930**
x4	.271	**.886**	.109
x5	**.650**	.544	.289
x6	.511	**.784**	.099
x7	**.824**	-.221	.312
x8	**.822**	.143	.249
x9	**.927**	-.037	.069
x10	**.674**	.306	.583
x11	.548	.216	**.721**
x12	-.459	-.360	**-.716**
x13	.296	**-.779**	.051

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a Rotation converged in 4 iterations.

جدول (۷) ضرایب همبستگی بین عوامل را قبل و بعد از چرخش نشان می‌دهد.

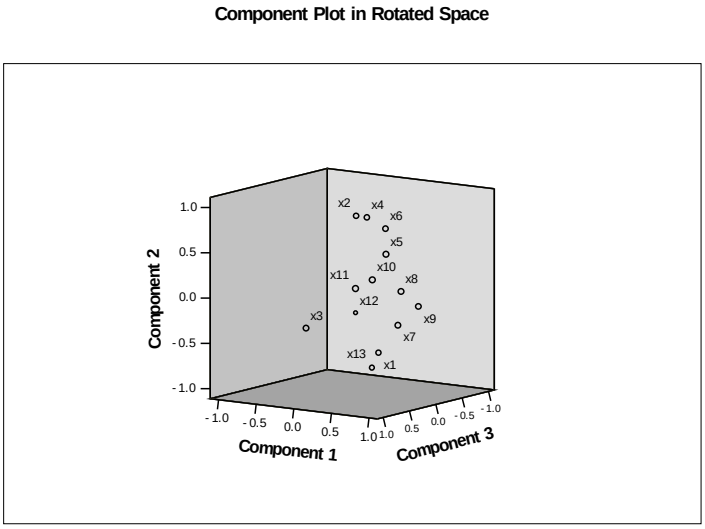
جدول ۷- ضریب همبستگی بین عوامل قبل و بعد از چرخش

Component Transformation Matrix

Component	**1**	**2**	**3**
1	.742	.456	.492
2	.376	-.890	.259
3	.556	-.007	-.831

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

شکل (۸) نمودار سه‌بعدی دوران‌یافته را نشان می‌دهد. در این نمودار پراکنش متغیرهای مورد بررسی نسبت به عامل‌های اول، دوم و سوم مشاهده می‌شود.



شکل ۸- نمودار سه بعدی پراکنش متغیرها نسبت به عامل‌های استخراج شده

۲- تحلیل خوشه‌ای

اصطلاح تحلیل خوشه‌ای[۳۳] اولین با توسط Tryon در سال ۱۹۳۹ برای روش‌های گروه‌بندی اشیائی که شبیه بودند مورد استفاده قرار گرفت. تجزیه خوشه‌ای ابزار میانبر تحلیل داده‌ها است که هدف آن نظم دادن به اشیاء مختلف به گروه‌هایی که درجه ارتباط بین دو شیء اگر آنها به یک گروه تعلق داشته باشند حداکثر و در غیر این صورت حداقل است. به‌عبارت دیگر تحلیل خوشه‌ای ساختار داده‌ها را بدون توضیح اینکه چه وجود دارد را نشان می دهد.

هدف از خوشه‌بندی داده‌ها آن است که مشاهدات را به گروه‌های متجانس تقسیم کنیم، به‌طوری که مشاهدات هر گروه بیشترین شباهت و مشاهدات گروه‌های مختلف کمترین شباهت را با هم داشته باشند.

تحلیل خوشه‌ای یک ابزار اکتشاف است و نتایج آن ممکن است ۱) در تعریف یک طرح طبقه‌بندی مانند رده‌بندی حیوانات، حشرات یا گیاهان مفید باشد. ۲) قواعدی برای اختصاص موارد جدید به طبقه‌ها به منظور شناسایی و تشخیص به‌دست دهد. ۳) حدود تعریف، اندازه و تنوع و تعریف برای آنچه قبلاً به شکل مفاهیم وسیعی بوده است، فراهم آورد. ۴) نمونه‌هایی برای معرفی طبقه‌ها بیاید. ۵) مدل‌های آماری برای توصیف جامعه‌ها ارائه دهد.

مفاهیم فاصله[۳۴] و تشابه[۳۵] از مفاهیم اساسی تحلیل خوشه‌ای است. فاصله اندازه‌ای است که نشان می‌دهد دو مشاهده تا چه حد جدا از یکدیگرند. در حالی که تشابه شاخص نزدیکی آنها با یکدیگر است. پژوهشگر قبل از تحلیل، نخست باید یک مقیاس کمی را که بر پایه همخوانی (تشابه) بین مشاهده‌ها اندازه گرفته می‌شود را انتخاب کند. این شاخص‌ها با توجه به الگوریتم تشکیل خوشه‌ها، ماهیت متغیرها (پیوسته، گسسته یا دو ارزشی) و مقیاس اندازه‌گیری انتخاب می‌شوند.

برای انجام این تجزیه خوشه‌ای در نرم‌افزار SPSS از روند زیر استفاده می‌شود:

Analyze>Classify>Cluster analysis

در محیط نرم‌افزار SPSS برای طبقه‌بندی روش‌های زیر وجود دارد:

۱- Two Step Cluster… (خوشه‌ای دو مرحله‌ای)

۲- K-means Cluster (خوشه‌ای k میانگین)

۳- Hierarchical Cluster … (خوشه‌ای سلسله مراتبی)

۴- Discriminate … (ممیزی)

۵- Tree … (ساختار درختی)

فرمان K-Means Cluster

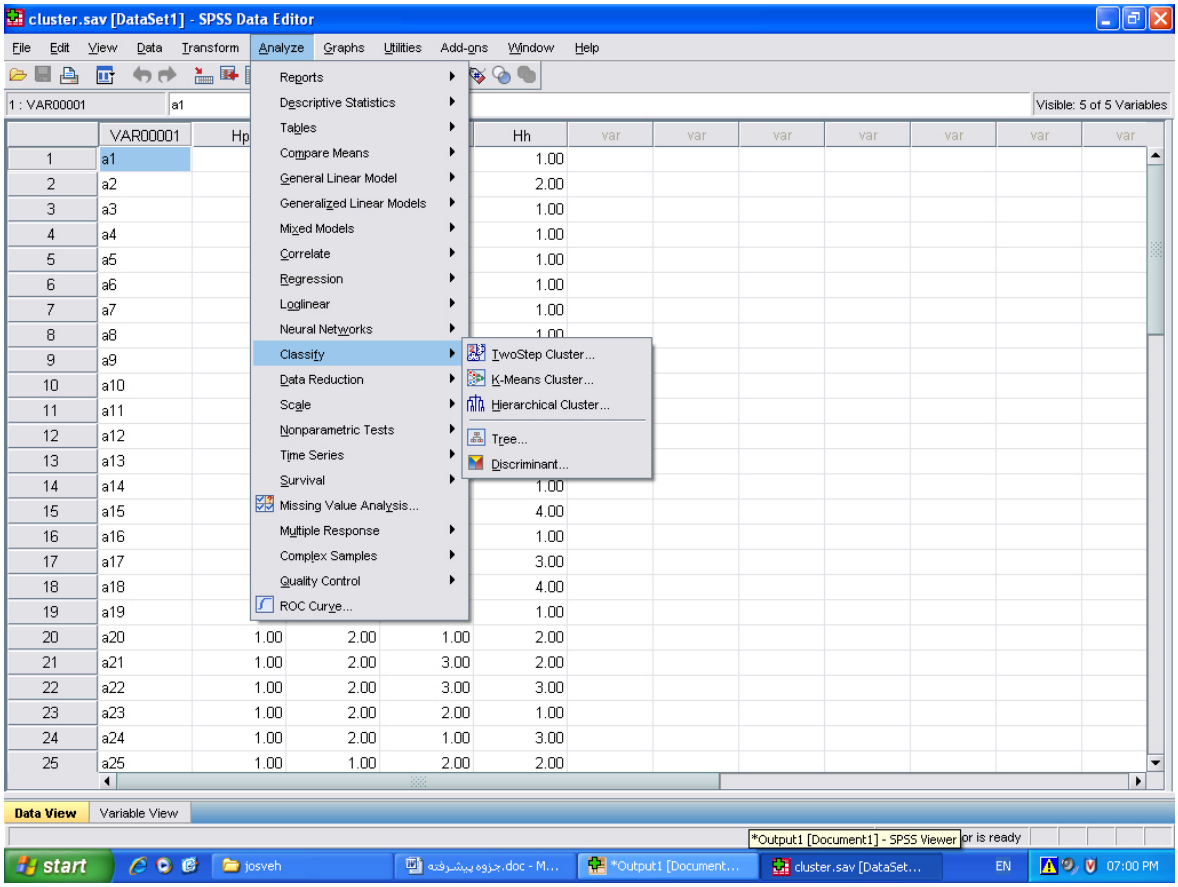
این فرمان برای گروه‌بندی مشاهدات هنگامی که تعداد گروه‌ها از قبل معین است، به کار می‌رود.

نام متغیرهایی که قصد گروه‌بندی آنها را داریم به جعبه Variables منتقل می‌کنیم. برچسب متغیرها در گروه‌بندی بر حسب مقادیر متغیری که در جعبه Label Cases by قرار می‌گیرد، تعیین می‌شود. تعداد گروه‌ها در جعبه Number of Clusters تعیین می‌شود. در بخش Method اگر گزینه Iterate and classify انتخاب شود، در هر بار تکرار مراکز خوشه‌ها تغییر می‌کند. اگر گزینه Classify only انتخاب شود، نسبت دادن مشاهدات بر اساس گروه‌بندی اولیه انجام می‌شود.

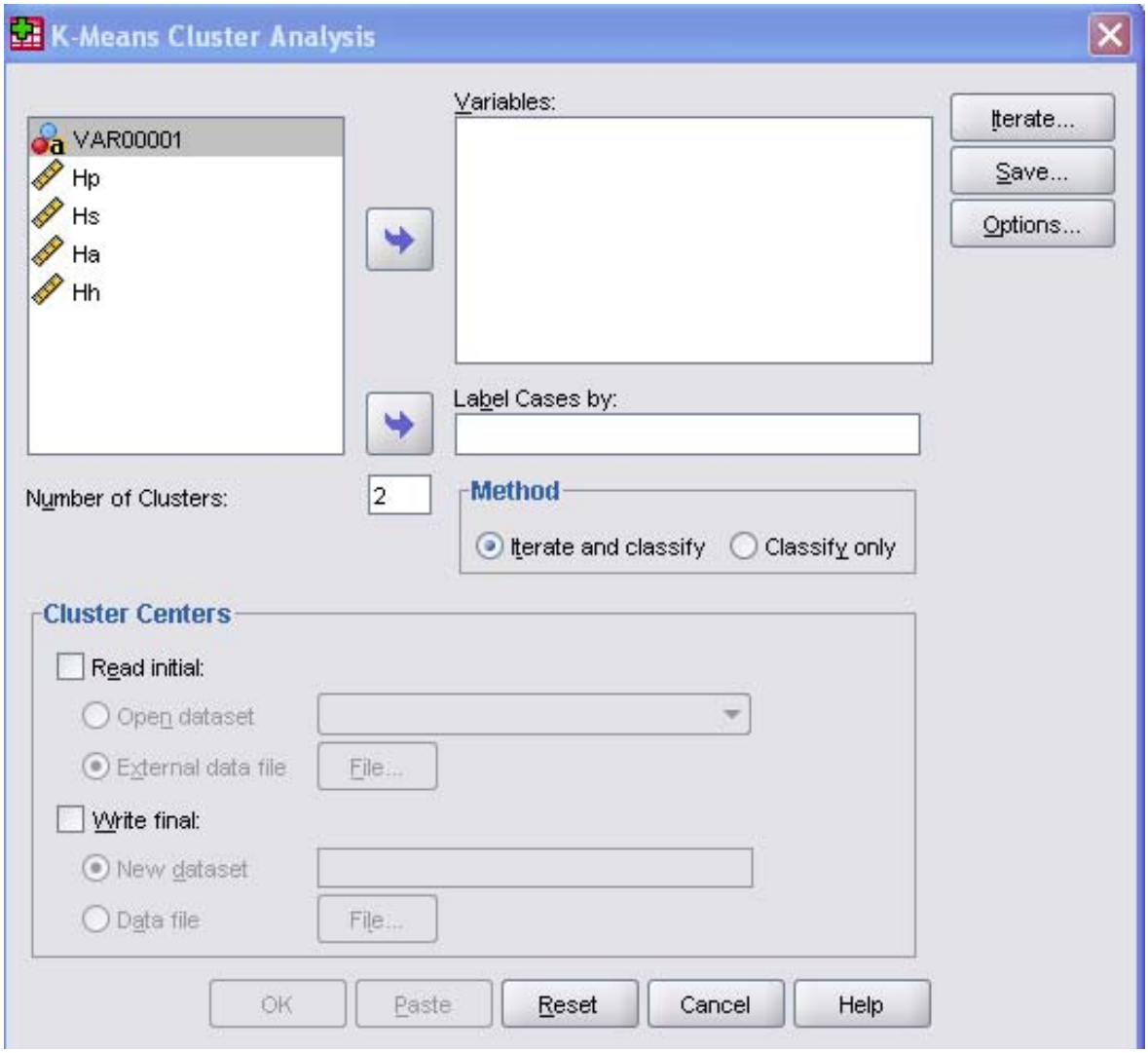
کلید Centers چگونگی تعیین تمرکز گروه‌ها را نشان می‌دهد. اگر تمرکز اولیه گروه‌ها از قبل در فایل داده خاصی قرار گرفته است، گزینه Read initial from فعال شود، تا کلید File برجسته شود. با فعال کردن کلید File پنجره گفتگوی خواندن فایل های داده باز می‌شود.

1- Cluster analysis
2- Distance
3- Similarity

اگر گزینه Write final as فعال شود، میانگین نهایی گروه‌ها را می‌توان در فایل خاصی ذخیره کرد. برای ذخیره فایل‌ها، پنجره گفتگوی ذخیره‌سازی فایل‌ها فعال می‌شود.



شکل ۹- روند انجام تجزیه خوشه‌ای در نرم‌افزار SPSS

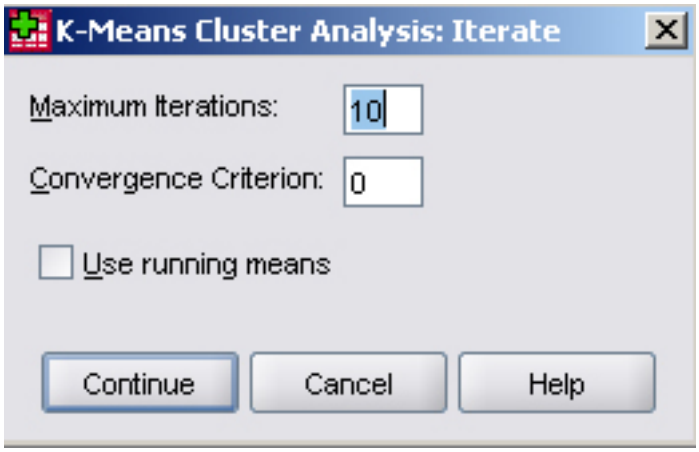


شکل ۱۰- کادر K-means cluster از منوی Classify

کلید Iterate به شرایط اجرای الگوریتم اشاره می‌کند و با فعال کردن آن جعبه گفتگوی زیر ظاهر می‌شود. حداکثر تعداد تکرار الگوریتم K-Means در جعبه Maximum Iterations تعیین می‌شود (پیش فرض آن ۱۰ تکرار است).

مقدار وارده شده در جعبه Convergence Criterion تعیین‌کننده معیار دیگر همگرایی الگوریتم است. این عدد نسبتی از حداقل فاصله بین مراکز اولیه گروه‌ها را نشان می‌دهد، بنابراین عددی بین صفر و یک است. برای مثال برای عدد ۰/۰۲ زمانی که فاصله بین مراکز خوشه‌ها از ۲ درصد کمترین اختلاف بین هر مرکز اولیه‌ای بیشتر نباشد، الگوریتم متوقف می‌شود.

گزینه Use running means سبب به هنگام شدن مراکز خوشه‌ها پس از اضافه شدن عضو جدید است. اگر این گزینه را انتخاب نکنید، مراکز جدید خوشه‌ها پس از آن که تمامی مشاهدات تخصیص داده شد، محاسبه می‌شوند.



شکل ۹- کادر Iterate از بخش K-means Cluster

کلید Save به تعیین نوع اطلاعات ذخیره شده بستگی دارد. گزینه Cluster membership وضعیت عضویت مشاهدات در گروه‌ها را ذخیره می‌کند. گزینه Distance from cluster center فاصله هر مشاهده را از مرکز گروه نشان داده و در یک فایل ذخیره می کند.

با فعال کردن کلید Options جعبه گفتگوی زیر ظاهر می شود.

در بخش Statistics آماره‌ها و خروجی‌های اختیاری تعیین می شود. پیش فرض این بخش انتخاب گزینه Initial cluster centers است که سبب نمایش مراکز خوشه‌های اولیه در گروه‌بندی می‌شود. این مراکز اولیه می‌توانند از یک فایل نیز خوانده شوند. گزینه دوم ANOVA table است که آزمون‌های F یک متغیره برای گروه‌بندی متغیرها را اجرا می‌کند.

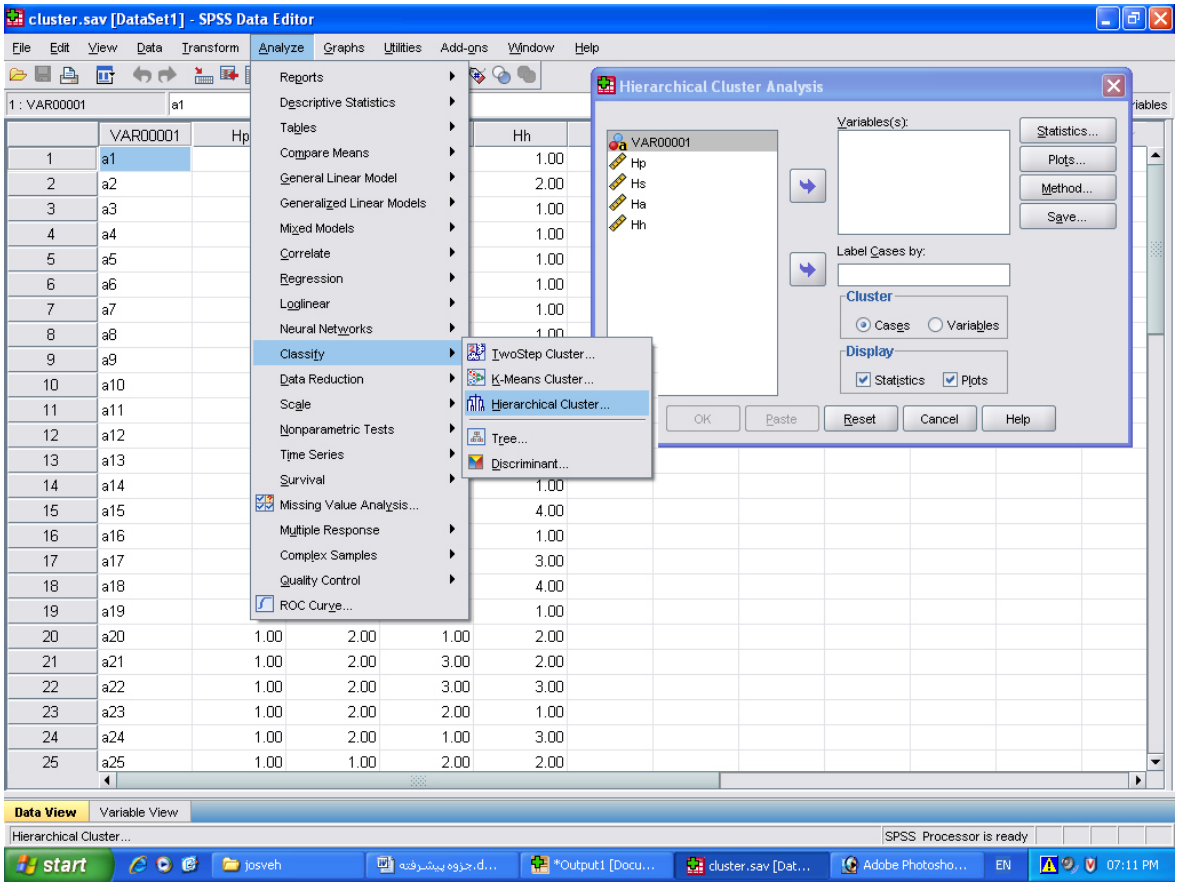
بخش Missing values دارای گزینه‌های زیر است:

اگر گزینه Exclude cases listwise انتخاب شود، مشاهده‌ای را که در یک متغیر خوشه‌بندی مقدار گمشده دارد، کامل حذف می‌شود.

اگر گزینه Exclude cases pairwise انتخاب شود، فقط متغیرهایی را که در همه متغیرهای خوشه‌بندی مقدار گمشده دارند، حذف می‌شوند. بنابراین مشاهده‌ها را بر اساس مقادیر متغیرهایی که مقدار گمشده ندارند، به نزدیکترین خوشه نسبت می‌دهد.

خوشه‌بندی سلسله مراتبی

اگر تعداد خوشه‌ها قبل از گروه‌بندی مشخص نباشد، از فرمان Hierarchical Cluster Analysis استفاده می‌شود.



شکل ۱۱- اجرای فرمان Hierarchical Cluster Analysis

بعد از ظاهر شدن جعبه اجرای فرمان، متغیرهایی که هدف گروه‌بندی مشاهدات بر اساس آنهاست، در بخش Variables قرار می دهند. اگر متغیرها برچسب خاصی دارند، نام متغیر دارای برچسب‌ها را به جعبه Label Cases منتقل می‌کنند. این متغیر اغلب کاراکتری است.

بخش Cluster نوع داده‌های مورد گروه‌بندی را مشخص می‌کند. گزینه Cases برای گروه‌بندی مشاهدات و گزینه Variables برای گروه‌بندی متغیرها انتخاب می‌شود.

تعیین چگونگی نمایش خروجی‌ها در بخش Display صورت می‌گیرد. گزینه Statistics نمایش خروجی‌های کمی آماری و گزینه Plots نمایش نمودارها را فراهم می‌کند.

روش نسبت دادن مشاهدات به خوشه‌های مختلف در جعبه گفتگوی کلید Method تعیین می‌شود.

در جعبه Cluster method می‌توان روش‌های مختلف خوشه‌بندی را انتخاب کرد.

- روش Between-groups-linkage، ترکیب خوشه‌ها از کمینه کردن متوسط فاصله بین تمام زوج مشاهداتی که در خوشه‌های مختلف قرار دارند، ایجاد می‌شود. در این روش از کلیه فواصل موجود بین نقاط خوشه‌ها استفاده می شود نه فقط نزدیکترین یا دورترین فاصله‌ها.

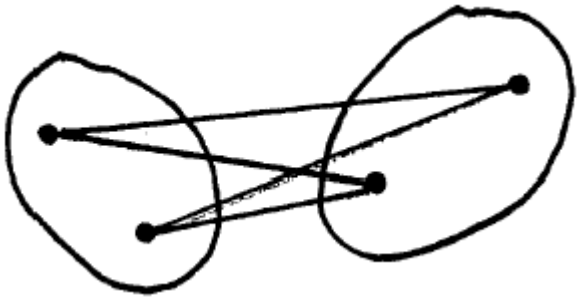
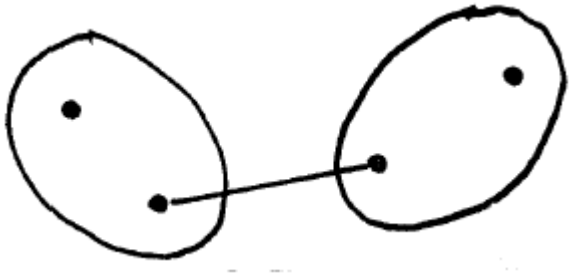


Figure 15.12
Cluster distance, average linkage method

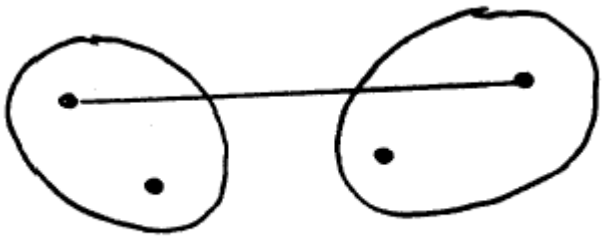
- روش Within-groups linkage: در این روش سعی می‌شود مشاهدات طوری در خوشه‌ها قرار گیرند که معدل فاصله نقاط داخل خوشه‌ها از یکدیگر به کمترین مقدار برسد.

- روش Nearest neighbor (Single linkage):



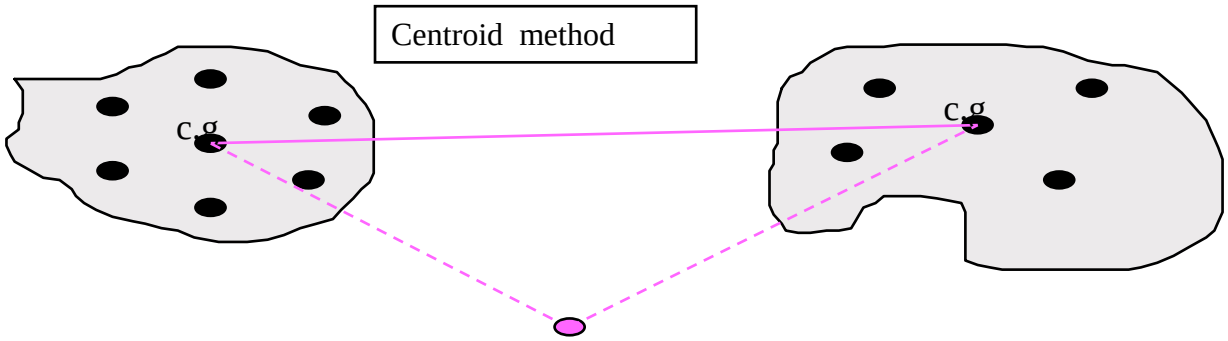
شکل – نزدیکترین همسایه

- روش Furthest neighbor (دورترین همسایه): در این روش فاصله بین دو خوشه را بر حسب فاصله بین دورترین نقاط آن محاسبه می‌شود.



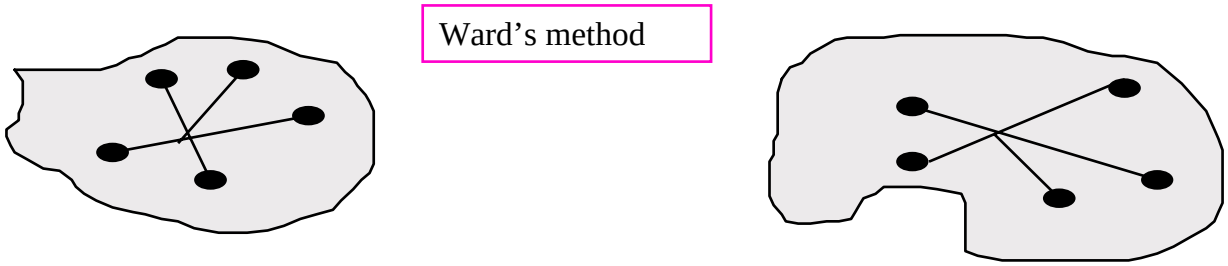
شکل – دورترین همسایه

- روش Centroid clustering (خوشه‌بندی متمرکز): فاصله بین دو خوشه، فاصله بین میانگین‌های آنهاست. فاصله‌ای که خوشه‌ها با یکدیگر ترکیب می‌شوند از مرحله‌ای به مرحله دیگر کاهش‌می‌یابد.



- روش Median clustering (خوشه‌بندی میانه): دراین روش به دو خوشه‌ای که ترکیب می‌شوند، وزن‌های یکسانی صرف نظر از تعداد نقاط آنها داده می‌شود. این عمل سبب می‌شود گروه‌های کوچک نسبت به دیگر گروه‌ها اثر مشابهی در ساختن خوشه‌های بزرگتر داشته باشند.

- روش Ward's: در این روش ابتدا میانگین‌های متغیرها در داخل هر خوشه محاسبه می‌شود. سپس برای هر مشاهده، مربع فاصله اقلیدسی میانگین‌های خوشه‌ها محاسبه می‌شود. این فاصله برای تمامی مشاهدات جمع می‌شود. در هر مرحله دو خوشه‌ای ترکیب می‌شوند که کوچکترین افزایش در مجموع مربعات فواصل داخل خوشه‌ای را داشته باشند.



بخش Measure روش‌های اندازه‌گیری فاصله دو نقطه را نشان می‌دهد. این روش‌ها عبارتند از:

الف) روش‌های فاصله‌ای (گزینه Interval): در این روش‌ها معیار تشابه یا عدم تشابه بین مشاهدات مختلف بر حسب میزان فاصله بین دو نقطه اندازه‌گیری می‌شود. مهمترین این معیارها عبارتند از:

- معیار فاصله اقلیدسی: اگر گزینه Euclidean distance انتخاب شود، مربع فاصله اقلیدسی برای دو بردار X_i و X_j با k متغیر از رابطه زیر تعیین می‌شود:

$$D = \sqrt{\sum_{i=1}^{k}(X_{il} - X_{jl})^2}$$

- معیار مربع فاصله اقلیدسی: اگر گزینه Squared Euclidean distance با نماد D2 که D همان فاصله اقلیدسی معمولی است.

$$D^2 = \sum_{i=1}^{k}(X_{il} - X_{jl})^2$$

- معیار Cosine: کسینوس زاویه بین دو بردار X_i و X_j است. مقدار این معیار بین ۱+ تا ۱- تغییر می‌کند.

- معیار Pearson: ضریب همبستگی خطی بین مقادیر دو بردار را اندازه می‌گیرد و بین ۱+ تا ۱- تغییر می‌کند.

- معیار Chebychev: فاصله بین دو مشاهده برابر بیشترین قدر مطلق اختلاف مقادیردر هر متغیر است.

$$Ch_{ij} = \underset{l}{Max}\left|X_{il} - X_{jl}\right|$$

- معیار Block: معیار فاصله، مجموع قدر مطلق انحراف مقادیر برای هر متغیر است:

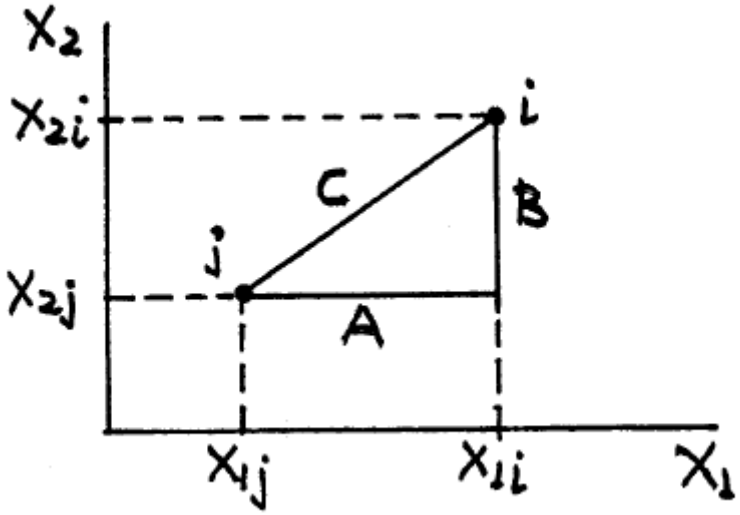
$$Bl_{ij} = \sum_{l} \left| X_{il} - X_{jl} \right|$$

- معیار Minkowski: معیار فاصله ریشه pام مجموع قدر مطلق اختلاف است که به توان p رسیده باشد. عدد p در جعبه Power تعیین می‌شود.

$$M_{ij} = \sqrt[p]{\sum_{l} \left| X_{il} - X_{jl} \right|^{p}}$$

- معیار Customized: معیار فاصله دو مشاهده، ریشه rام مجموع توان pام قدر مطلق اختلاف مقادیر متغیر در دو مشاهده است، بنابراین باید توان p و ریشه r را به عنوان ورودی در جعبه‌های Power و Root وارد کرد.

$$M_{ij} = \sqrt[r]{\sum_{l} \left| X_{il} - X_{jl} \right|^{p}}$$



ب) روش‌های شمارشی (گزینه Counts)

- معیار مربع کای (χ^2): با انتخاب گزینه Chi-square ... معیار عدم تشابه بر اساس آماره χ^2 محاسبه می‌شود.

- معیار Φ^2: با انتخاب گزینه Phi-square سعی می‌شود اندازه نمونه را برای کاهش اثر مقادیر فراوانی‌های مشاهده شده واقعی مقادیر محاسبه کند. Φ^2 مقدار χ^2 را با کل فراوانی نرمالیزه می‌کند ($\Phi^2 = \frac{\chi^2}{N}$).

ج- روش‌های Binary

در صورتی که داده‌ها به‌صورت Binary باشند، از معیارهای زیر استفاده می‌شود:

- فاصله اقلیدسی باینری مانند فاصله اقلیدسی معمولی است. حداقل آن صفر و حداکثر مقدار ندارد.
- مربع فاصله اقلیدسی باینری
- اندازه تفاوت حجم، معیار تشابهی با حداقل صفر و بدون حداکثر
- اندازه تفاوت الگوها
- اندازه‌های عدم تشابه و شکل

- Lance-Williams (یا ضریب غیر متریک Bray-Curtis) که مقداری بین صفر و یک دارد.

بخش Transform values وظیفه تبدیل مقادیر اولیه را بر عهده دارد. گزینه‌های موجود با فرض نماد Xi برای اندازه‌ها عبارتند از:

- None: تبدیلی صورت نمی‌گیرد.
- Z Scores: تبدیل استاندارد بر روی داده‌ها انجام می‌شود:

$$Z_i = \frac{X_i \overline{X}}{S_{\overline{X}}}$$

- Range -1 to +1: مقادیر خام به مقادیر با دامنه ۱- تا ۱+ تبدیل می‌شوند:

$$y_i = \frac{Z_i}{\text{Range}}$$

- Range 0 to +1: مقادیر خام به مقادیری با دامنه ۰ تا ۱+ تبدیل می‌شوند:

$$y_i = \frac{X_i - \text{Min}(X_i)}{\text{Range}}$$

- Max Mag. of 1: حداکثر مقدار یک خواهد بود.

$$y_i = \frac{X_i}{\text{Max}(X_i)}$$

- Mean of 1: میانگین مقادیر یک خواهد بود.

$$Y_i = X_i - \overline{X} + 1 \Rightarrow \overline{Y} = \frac{\sum Y_i}{N} = \frac{\sum (X_i - \overline{X}) + \sum 1}{N} = \frac{N}{N} = 1$$

- Standard dev of 1: انحراف معیار یک خواهد بود.

بخش Transform measures تبدیلات موجود بر روی اندازه نهایی (فاصله‌ها) را معرفی می‌کند. گزینه‌های این بخش عبارتند از:

- Absolute values: قدر مطلق مقادیر فاصله را نشان می‌دهد.
- Change sign: علامت فاصله را عوض می‌کند.
- Rescale to 0-1 Range: برد فاصله را به صفر تا یک تبدیل می‌کند.

کلید Statistics به تعیین چگونگی نمایش خروجی‌های آماری می‌پردازد. گزینهAgglomeration schedule، مشاهدات یا خوشه‌هایی را که در هر مرحله ترکیب می‌شوند و همچنین فواصل بین مشاهدات یا خوشه‌های ترکیبی و آخرین خوشه‌هایی را که مشاهده‌ای به آنها اضافه شده است، نشان می‌دهد.

گزینه Distance matrix: ماتریس فواصل یا شباهت بین مشاهدات (یا متغیرها) را نمایش می‌دهد. این ماتریس پایین مثلثی است و روی قطر آن صفر است.

بخش Cluster Membership وضعیت نهایی قرار گرفتن مشاهدات را در خوشه‌های مختلف نشان می‌دهد.

گزینه None از فهرست بندی شماره عضویت مشاهدات در خوشه‌ها جلوگیری می کند.

گزینه Single solution عضویت هر مشاهده را در مرحله‌ای تکی درخواست می‌کند. تعداد خوشه‌هایی را که می‌خواهید در جعبه Cluster وارد کنید. این عدد باید عددی بزرگتر از یک باشد.

گزینه Range of Solutions در هر مرحله، عضویت هر مشاهده را در دامنه تعداد خوشه‌های تعیین شده درخواست می‌کند. حداقل و حداکثر تعداد خوشه‌های مورد نظر را به‌ترتیب در جعبه‌های From و Through وارد کنید.

برای ترسیم نمودار خوشه‌ها، کلید Plots را فعال کنید.

گزینه Dendrogram نموداری از مشاهدات ترکیبی و مقادیر ضرایب خوشه‌بندی در هر خوشه را نشان می‌دهد. نحوه تنظیم نمودار دقیقاً با مراحل جدول Agglomeration Schedule یکسان است.

گزینه‌های بخش Icicle عبارتند از:

All clusters: سبب نمایش نمودار برای تمامی خوشه های ممکن می‌شود.

… Specified: نمودار را تنها برای دامنه معینی از خوشه‌ها ترسیم می‌کند. نقاط این دامنه و فاصله آنها با یکدیگر در جعبه‌های زیر این گزینه تعیین می‌شوند. جعبه Start نقطه شروع تعداد خوشه‌ها، جعبه Stop نقطه پایان و جعبه By فاصله بین شماره خوشه‌ها را نشان می‌دهد.

None: با انتخاب این گزینه، نموداری از نوع Icicle رسم می‌شود.

در بخش Orientation وضعیت قرار گرفتن محورهای نمودار تعیین می‌شود. گزینه Vertical شماره خوشه‌ها را در محور عمودی و گزینه Horizental شماره خوشه‌ها را در محور افقی نمایش می‌دهد.

مثال ۲- جدول زیر طبقات پوشش-فراوانی بر اساس مقیاس وان-در-مارل را برای ۱۱ گونه گیاهی در ۴۰ پلات نشان می‌دهد. گونه‌ها و پلات‌ها را طبقه‌بندی کنید.

	Ar.au	As.al	St.ba	Sc.or	Ar.si	Sa.sp	Ep.st	Zy.eu	Co.mo	St.pl	Se.ro
Plot1	3.75	0.1	0.5	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Plot2	30	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Plot3	10	6.25	3.75	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Plot4	15	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Plot5	0.1	3.75	0.1	10	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Plot6	0.1	6.25	0.1	6.25	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Plot7	0.1	1.75	0.1	10	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Plot8	0.1	10	0.1	25	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Plot9	0.1	6.25	0.1	3.75	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Plot10	0.1	0.1	3.75	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Plot11	0.1	0.5	1.75	3.75	3.75	0.1	0.1	0.1	0.1	0.1	0.1
Plot12	0.1	1.75	0.1	0.1	1.75	0.1	0.1	0.1	0.1	0.1	0.1
Plot13	0.1	0.1	0.5	0.1	6.25	0.1	0.1	0.1	0.1	0.1	0.1
Plot14	0.1	0.1	1.75	0.1	6.25	0.1	0.1	0.1	0.1	0.1	0.1
Plot15	0.1	0.1	1.75	3.75	3.75	0.1	0.1	0.1	0.1	0.1	0.1
Plot16	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Plot17	0.1	0.1	0.1	0.5	3.75	0.1	0.1	0.1	0.1	0.1	0.1
Plot18	0.1	1.75	0.1	0.1	10	0.1	0.1	0.1	0.1	0.1	0.1
Plot19	0.1	3.75	0.1	1.75	6.25	0.1	0.1	0.1	0.1	0.1	0.1
Plot20	0.1	1.75	0.1	0.1	6.25	0.1	0.1	0.1	0.1	0.1	0.1
Plot21	0.1	1.75	0.1	1.75	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Plot22	0.1	0.1	1.75	0.1	30	0.1	0.1	0.1	0.1	0.1	0.1
Plot23	0.1	6.25	0.1	0.1	10	0.1	0.1	0.1	0.1	0.1	0.1
Plot24	0.1	0.1	0.5	0.1	15	0.1	0.1	0.1	0.1	0.1	0.1
Plot25	0.1	0.1	0.1	0.1	0.1	1.75	15	0.1	0.1	0.1	0.1
Plot26	0.1	0.1	0.1	0.1	0.1	0.1	3.75	10	0.1	0.1	0.1
Plot27	0.1	0.1	0.1	0.1	0.1	1.75	0.1	0.1	0.1	0.1	0.1
Plot28	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Plot29	0.1	0.1	0.1	0.1	0.1	0.1	0.5	0.1	0.1	0.1	0.1
Plot30	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	6.25	1.75	0.1
Plot31	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	15	0.5	0.1
Plot32	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	10	0.1
Plot33	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	6.25	3.75	0.1
Plot34	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.75	0.1
Plot35	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.75
Plot36	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Plot37	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	6.25
Plot38	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Plot39	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.75
Plot40	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	6.25

برای طبقه‌بندی گونه‌های گیاهی یا پلات‌های نمونه‌برداری در نرم‌افزار SPSS فرمان زیر را انجام دهید:

Analysis>Classify>Hierarchical cluster

- بعد از باز شدن کادر اجرای فرمان، متغیرهای مورد نظر را از جعبه سمت چپ انتخاب کرده و به جعبه Variables منتقل کنید.
- در بخش Cluster گزینه Cases را برای گروهبندی پلات‌ها و گزینه Variables را برای گروهبندی گونه‌های گیاهی انتخاب کنید.

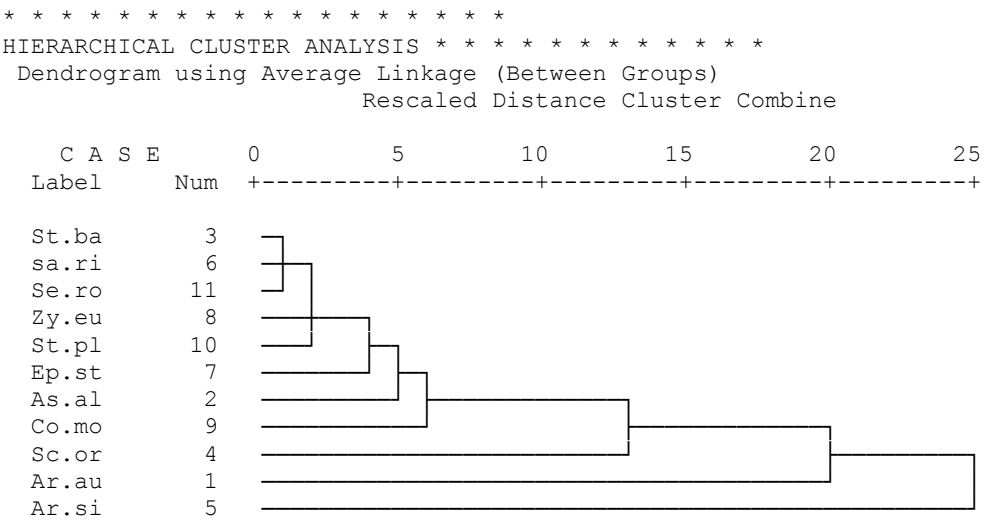
- گزینه Plots از بخش Display را انتخاب کنید. در صورت تمایل به نمایش خروجی‌های کمی آماری می‌توان گزینه Statistics را نیز فعال کرد.
- به منظور ترسیم دندروگرام، کلید Plots را فعال کنید. در جعبه ظاهر شده گزینه Dendrogram را انتخاب کنید.

بعد از تأیید موارد بالا، نتایج خروجی ظاهر خواهد شد که بخش مهم آن دندروگرام است. برای تعیین تعداد مطلوب خوشه‌ها (g) روش‌های زیر شده است:

الف) تعداد مطلوب خوشه‌های در روی نمودار دندروگرام جایی است که فاصله زیاد بین ادغام دو خوشه مشاهده می‌شود.

ب) اگر d_1، d_2، ...، d_{n-1} مقادیری باشند که بر اساس آنها ادغام خوشه‌ها صورت می‌گیرد (از مرحله دوم به بعد). اگر $\bar{d} = \frac{\sum d_i}{n-1}$ و $S_d^2 = \frac{(d_i - \bar{d})^2}{n-2}$ باشد، از d_1 شروع کرده و در اولین جایی که $d \rangle \bar{d} + \mu . S_d$ تعداد مطلوب خوشه‌ها به‌دست می‌آید. μ یک عدد اختیاری است که می‌تواند ۱/۹۶ باشد.

ج) از فرمول $\sqrt{\frac{N}{2}}$ برای تعیین تعداد خوشه‌ها استفاده می‌شود که در آن N تعداد کل افراد است.



۳- تحلیل ممیزی

تحلیل ممیزی یا آنالیز تشخیص[۳۶] توسط فیشر در سال ۱۹۳۶ ابداع شد و بر پایه روش‌شناسی مورد استفاده در رگرسیون خطی چند متغیره توسعه یافت. تحلیل ممیزی مشابه رگرسیون خطی چندگانه است با این تفاوت که متغیر وابسته نه تنها توزیع نرمال ندارد، بلکه یک متغیر کیفی با تعداد مقادیر اندک است. این روش زمانی مفید است که یک متغیر گروه‌بندی (کیفی) و چندین متغیر مستقل کمی وجود داشته باشد و هدف پژوهشگر به‌دست آوردن رابطه‌ای است تا بتواند با توجه به متغیرهای مستقل عضویت را در متغیر گروه‌بندی مشخص کند. تابع تشخیص معادله‌ای است که با داشتن مشخصات هر فرد جامعه می‌توان با قرار دادن این مشخصات در آن معادله پیش‌بینی کرد که فرد جامعه مورد نظر به کدام گروه تعلق دارد. این روش در مواقعی استفاده می‌شود که بخواهیم بر اساس صفات یا متغیرهای مشاهده شده مدلی برای پیش‌بینی عضویت گروهی بسازیم.

1- Discriminate analysis

در صورتی که متغیرهای X_1، X_2، ... و X_k در گروه‌های مختلف اندازه‌گیری شده باشند، شکل کلی تابع تشخیص XD به‌صورت زیر است:

$$X_D = b_1x_1 + b_2x_2 + ... + b_kx_k + b_0$$

این روش مانند رگرسیون چندمتغیره یک مدل خطی به دست می‌دهد که در آن متغیرهای تعیین کننده و مهم وارد مدل شده و متغیرهای نامناسب از آن خارج شده‌اند. از فرض‌های مهمی که باید وجود داشته باشد تا بتوان از این روش استفاده کرد، موارد زیر است:

- رابطه بین متغیرها باید خطی باشد؛
- متغیر وابسته باید به‌صورت یک متغیر دو یا چند مقوله‌ای باشد؛
- متغیرهای مستقل مورد استفاده در تحلیل ممیزی باید مقیاس سنجش فاصله‌ای یا نسبتی و توزیع نرمال داشته باشند؛
- بین متغیرهای مستقل نباید هم‌خطی وجود داشته باشد؛
- حجم متغیرها در طبقات خیلی با هم اختلاف نداشته باشد و اگر مساوی باشد، بهتر است؛
- حجم نمونه مورد مطالعه کمتر از ۳۰ نباشد.

مهمترین کاربردهای تحلیل ممیزی در موارد زیر خلاصه می‌شود:

- بررسی تفاوت های بین گروهی؛
- تعیین مناسب‌ترین روش برای تفاوت گذاری بین گروه‌ها؛
- تشخیص و حذف متغیرهایی که در ایجاد تمایز بین گروه‌ها نقشی ندارند؛
- طبقه‌بندی افراد مورد مطالعه در گروه‌های تعیین شده؛
- آزمودن میزان درستی و صحت طبقه‌بندی مشاهده شده با طبقه‌بندی پیش‌بینی شده.

با انجام تحلیل ممیزی یک تابع یا مجموعه‌ای از توابع ساخته می‌شود. برای k گروه k-1 تابع تشخیص ساخته می‌شود. اولین تابع بهترین ترکیب خطی برای پیش‌بینی عضویت در گروه‌ها به دست می‌دهد. برای تعیین بهترین تابع از شاخص لامبدای ویلکس استفاده می‌شود. مقدار این شاخص بین صفر و یک متغیر است. هر چه مقدار برای یک تابع کوچکتر باشد، آن تابع تفکیک‌کنندهٔ خوبی است. از آنجا که توزیع این شاخص به کای‌اسکور شبیه است، از این‌رو از طریق این آماره تعبیر می‌شود.

تحلیل ممیزی در نرم‌افزار SPSS از فرمان زیر انجام می‌شود:

Analyze>Cluster>Discriminate Analysis

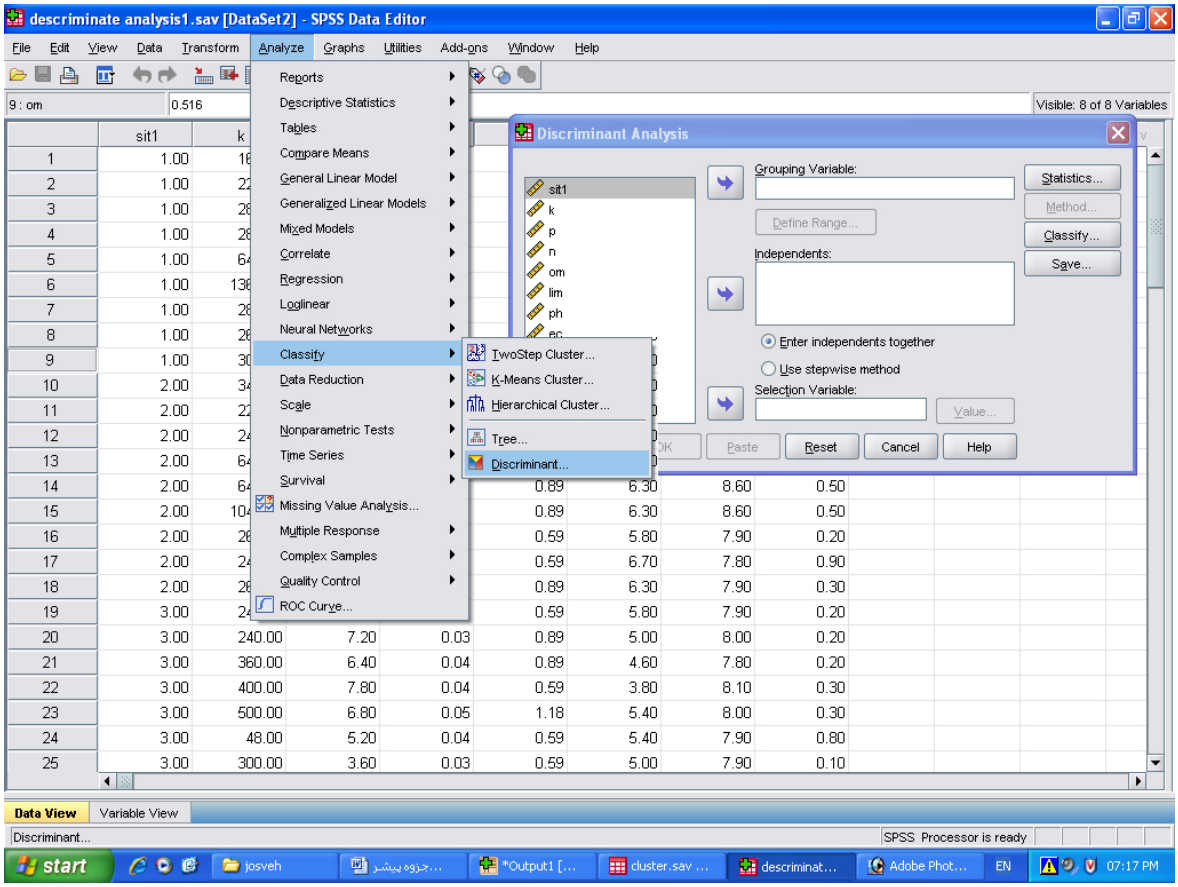
ابتدا متغیر گروه‌بندی، یعنی متغیری که وضعیت مشاهده‌ها را در گروه‌های از پیش تعیین شده‌ای نشان می‌دهد، به جعبه Grouping variable منتقل می‌کنیم. دامنه تغییرات شماره گروه‌ها را با فعال کردن کلید Define Range تعیین می‌کنیم.

در جعبه Independents متغیرهای مستقلی را که تغییرات آنها اساس گروه‌بندی مشاهدات است، وارد می‌کنیم. نحوه ورود این متغیرها در تحلیل ممیزی به دو صورت زیر است:

- اگر گزینه Enter Independents together انتخاب شود، همه متغیرها صرف نظر از مؤثر بودن یا نبودن در تمایز بین مشاهدات همزمان در تابع تحلیل ممیزی وارد می‌شوند.
- اگر گزینه Use stepwise method انتخاب شود، متغیرها در صورت مؤثر بودن به‌ترتیب اهمیت مرحله به مرحله وارد تحلیل ممیزی می‌شوند.

در جعبه Select variable امکان اجرای تحلیل ممیزی بر روی مشاهده‌هایی که مقدار خاصی از یک متغیر را اختیار کرده‌اند، فراهم می‌شود. مقدار خاص را با فشار دادن کلید Value که در انتهای این جعبه قرار دارد، وارد می‌کنند.

مشاهداتی که مقدار متغیر انتخابی آنها با این مقدار برابر باشد، در تحلیل ممیزی وارد می‌شوند. بدیهی است که متغیر انتخابی دیگر نمی‌تواند به‌عنوان متغیری مستقل در مجموعه متغیرهای تحلیل ممیزی وارد شود.



شکل ۱۲- اجرای فرمان Discriminate Analysis

با فشار دادن کلید Statistics نوع خروجی‌های آماره‌ای که پس از اجرای فرمان در صفحه خروجی ظاهر می‌شود، انتخاب می‌گردد.

در بخش Descriptives آماره‌های مختلف بر مبنای اطلاعات موجود مشاهدات انتخاب می‌شوند.

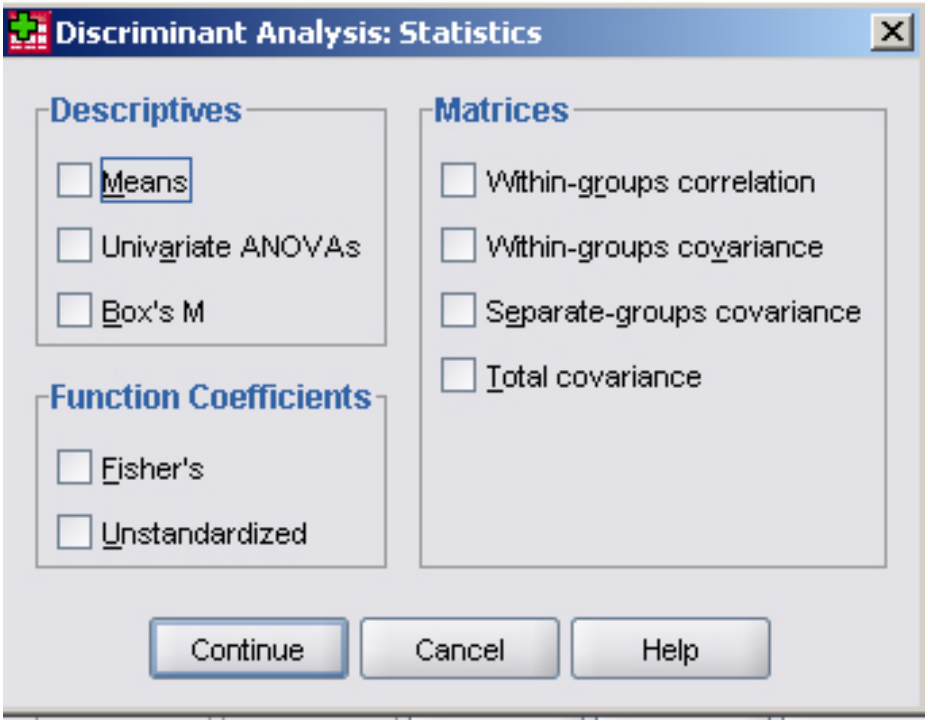
- گزینه Means سبب نمایش میانگین‌ها و انحراف معیارهای گروه‌ها به ازای تمام متغیرهای مستقل و همچنین کل مشاهدات می‌شود.
- گزینه Univariate ANOVAs فرض یکسانی میانگین گروه‌های مختلف را برای تمامی متغیرهای مستقل جداگانه آزمون می‌کند.
- گزینه Box'M فرض یکسانی ماتریس کواریانس‌های گروه‌ها را آزمون می‌کند.

بخش Function Coefficients ضرایب مختلف تابع تحلیل ممیزی را نشان می‌دهد. اگر گزینه Fisher's انتخاب شود، ضرایب تابع ممیزی فیشر برای گروه‌های موجود محاسبه می‌شود. اگر گزینه Unstandardized انتخاب شود، ضرایب استاندارد نشده تابع ممیزی برای محاسبه امتیازهای تابع ممیزی استفاده می شود.

بخش Matrices انواع ماتریس‌هایی که در نتایج امکان نمایش آنها وجود دارد، را نشان می‌دهد:

- گزینه Within-groups correlation ماتریس ضرایب همبستگی داخل گروه‌ها را نمایش می‌دهد.
- گزینه Within-groups covariance ماتریس واریانس-کوواریانس را نمایش می‌دهد.

- گزینه Separate-groups covariance ماتریس کوواریانس‌ها را به تفکیک گروه‌های مختلف و گزینه Total Cov. ماتریس واریانس-کوواریانس کل مشاهده‌ها را نمایش می‌دهد.



شکل ۱۳– جعبه گفتگوی کلید Statistics

کلید Method زمانی فعال خواهد شد که چگونگی ورود متغیرهای مستقل به تحلیل ممیزی گزینه Use Stepwise انتخاب شود. با فعال کردن این کلید جعبه گفتگوی شکل ۱۳ ظاهر می‌شود.

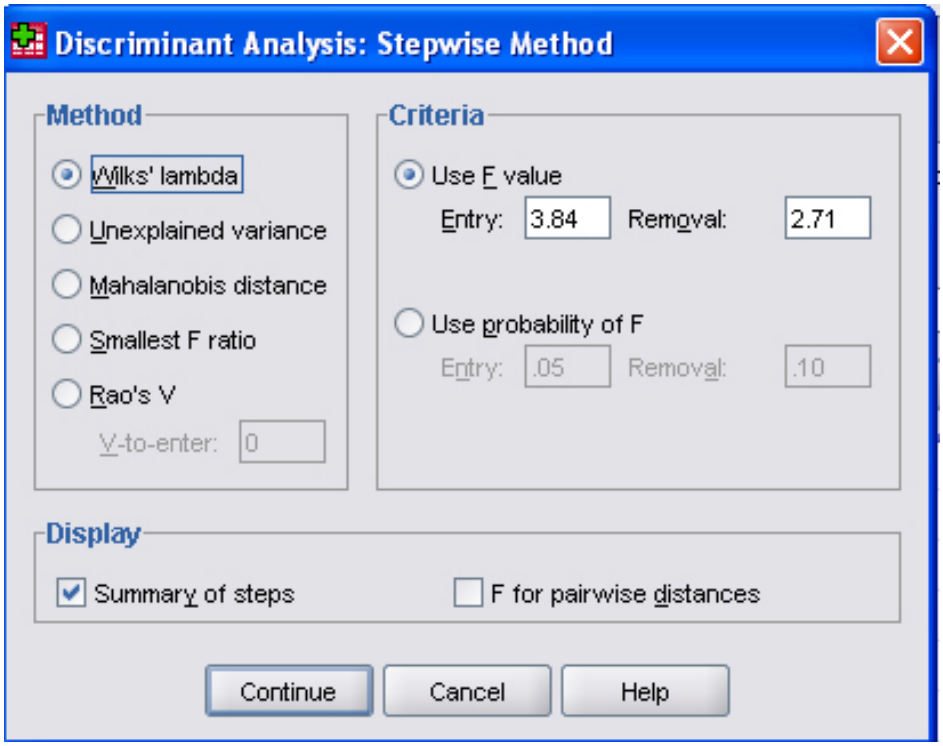
در بخش Method نوع آماره‌هایی که مقدار آنها مبنای ورود یا حذف متغیرها در تحلیل خوشه‌ای است، تعیین می‌شود. گزینه Wilk's Lambda بر اساس حداقل کردن آماره لامبدای ویلکس عمل می‌کند. گزینه Unexplained variance بر مبانی حداقل کردن مجموع واریانس تبیین نشده توسط متغیرها عمل می‌کند. گزینه Mahalanobis distance بر اساس بیشینه‌کردن فاصله ماهالانوبیس بین نزدیکترین گروه‌ها عمل می‌کند. گزینه Smallest F ratio کمترین نسبت F بین هر زوج گروهی را بیشترین می‌کند. گزینه Rao's V آماره V رائو را بیشترین می‌کند. مقدار V پس از فعال کردن این گزینه باید در جعبه V to enter قرار گیرد.

در بخش Criteria معیارهای عددی ورود و حذف متغیرها تعیین می‌شود. گزینه Use F value معیار ورود یا حذف متغیرها را مقدار آماره F فرض می‌کند. متغیرهایی به تحلیل ممیزی وارد می‌شوند که مقدار آماره F آنها از مقدار تعیین شده در جعبه Entry بیشتر باشد و آنهایی از تحلیل ممیزی حذف می‌شوند که مقدار آماره F آنها از مقدار تعیین شده در جعبه Removal کمتر باشد.

گزینه Use probability معیار ورود و خروج متغیرها را بر اساس احتمال متناظر با آماره F انجام می‌دهد. متغیرهایی به تحلیل ممیزی وارد می‌شوند که مقدار احتمال مؤثر نبودن آنها از مقدار تعیین شده در جعبه Entry بیشتر باشد و آنهایی از تحلیل ممیزی حذف می‌شوند که احتمال موثر نبودن آنها از مقدار تعیین شده در جعبه Removal بیشتر باشد. مقادیری که به هنگام ظهور جعبه گفتگو مشاهده می‌شوند، مقادیر پیش فرض این بخش هستند.

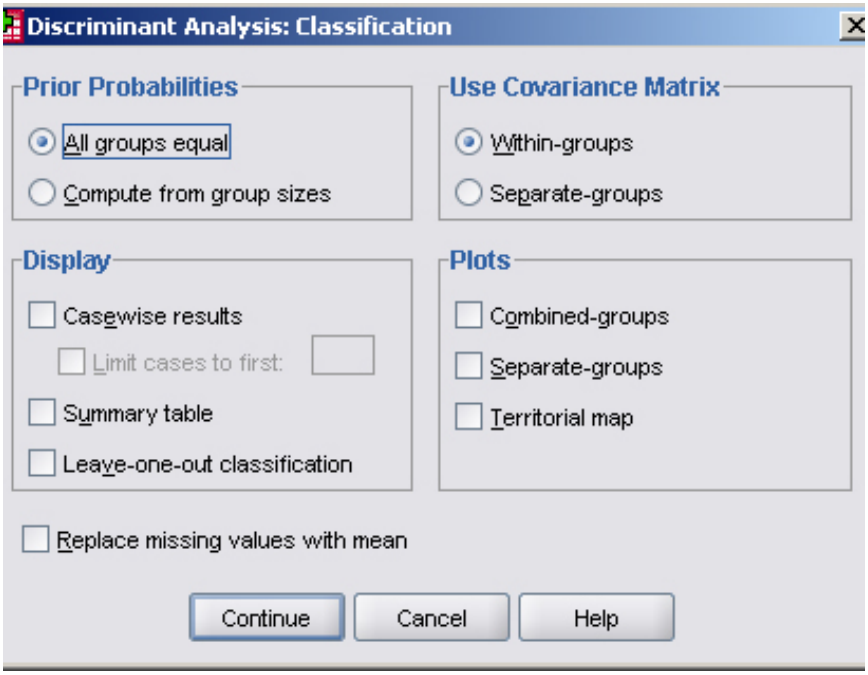
در بخش Display خروجی‌های پس از اجرای تحلیل ممیزی مرحله‌ای تعیین می‌شوند. در این بخش اگر گزینه Results at each step انتخاب شود، پس از اجرای هر مرحله کلیه نتایج عملیات در خروجی به نمایش در می‌آید. اگر گزینه Summary انتخاب شود، تنها عملیاتی که منجر به حذف یا ورود متغیرها در تحلیل ممیزی شده به نمایش در

می‌آید. همچنین اگر گزینه F for pairwise distances انتخاب شود، برای هر زوج فاصله میان گروه‌های موجود، آماره F متناظر به شکل یک ماتریس به نمایش در می‌آید.



شکل ۱۴– جعبه گفتگوی کلید Stepwise method

کلید Classify اطلاعات اولیه مورد نیاز برای گروه‌بندی و برخی خروجی‌های نهایی را تعیین می‌کند. با فعال کردن این کلید جعبه زیر ظاهر می‌شود.



شکل ۱۵– جعبه گفتگوی کلید Classification

در بخش Prior probabilities نحوه تخصیص احتمال‌های پیشین به گروه‌ها تعیین می‌شود. گزینه All groups equal به تمامی گروه‌های موجود وزن مساوی نسبت می‌دهد، اما گزینهCompute from group sizes به گروه‌ها وزنی متناسب با تعداد مشاهدات آن نسبت می‌دهد.

بخش Use Covariance Matrix نوع ماتریس-کوواریانسی که در تحلیل استفاده می‌شود، را نشان می‌دهد. گزینه Within-groups ماتریس مشترکی برای گروه‌ها فرض می‌کند، اما گزینه Separate-groups برای هر گروه، ماتریس واریانس جداگانه‌ای را بکار می‌برد.

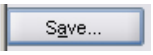
در بخش Plots نوع نمودارهای خروجی تحلیل ممیزی مشخص می‌شود. گزینه Combined-groups خروجی نهایی تحلیل را بر اساس نتیجه گروه‌بندی مشاهده‌ها به کلیه گروه‌های همگن به‌صورت یک بافت‌نگار نمایش می‌دهد. اگر تعداد گروه‌ها بیش از ۳ باشد، نمودار پراکنش مشاهده‌ها نسبت به دو تابع ممیزی اول ترسیم می‌شود.

گزینه Separate-groups برای هر گروه، نموداری از نقاطی که به آن منتسب شده‌اند را ترسیم می‌کند. با انتخاب گزینه Territorial map، اگر دو تابع ممیزی موجود باشد، امتیازات ممیزی هر مشاهده را ترسیم کرده و سعی می‌کند نمودار را به دو بخش مجزا بر اساس گروه‌های موجود تقسیم کند.

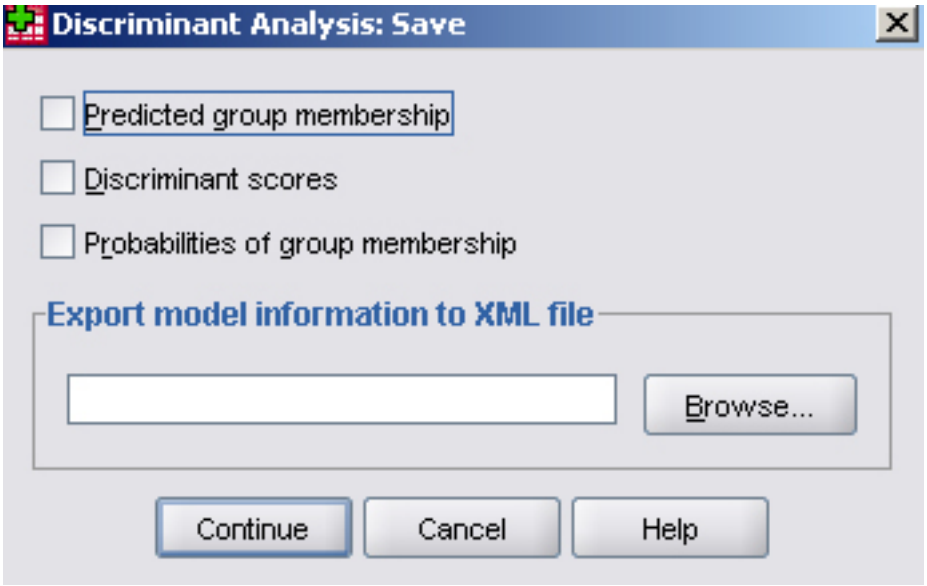
در بخش Display چگونگی نمایش اطلاعات در خروجی مشخص می‌شود. گزینه Results for each case نتیجه گروه‌بندی نهایی هر مشاهده، گروهی با بیشترین احتمال عضویت و گروهی پس از آن با بیشترین احتمال عضویت و امتیازات ممیزی هر مشاهده را در خروجی نمایش می‌دهد.

گزینه Summary table جدول نتایج گروه‌بندی را نمایش می‌دهد. این جدول شامل گروه‌بندی واقعی نمونه‌های انتخابی و گروه‌بندی پیشنهادی آنها بر اساس تحلیل ممیزی انجام شده در مقابل یکدیگر است.

با انتخاب گزینه Replace missing values with mean، میانگین متغیرهای مستقل جایگزین مشاهدهای گمشده خواهد شد. البته این جایگزینی در فایل داده جاری ذخیره نخواهد شد و فقط در طول تحلیل ممیزی معتبر است.

کلید 

برای نمایش اطلاعاتی که باید در فایل داده جاری پس از اجرای تحلیل ممیزی ذخیره شوند، در نظر گرفته شده است. پس از فعال کردن این کلید جعبه گفتگوی شکل ۱۵ ظاهر می‌شود.



شکل ۱۶- جعبه گفتگوی کلید Save

با انتخاب گزینه Predicted group membership شماره گروهی که تحلیل ممیزی برای هر مشاهده پیشنهاد می‌کند، در متغیر دیگری ذخیره می‌شود.

با انتخاب گزینه Discriminate scores، امتیازات ممیزی هر مشاهده در متغیر جدید ذخیره می شود.
گزینه Probabilities of group membership احتمال قرار گرفتن هر مشاهده در گروه‌های مختلف معرفی شده در تحلیل ممیزی را در متغیرهای جداگانه‌ای (به ازای هر گروه یک متغیر) را ذخیره می‌کند.

مثال ۳– در پژوهشی خصوصیات خاک در ۸ سایت مختلف اندازه‌گیری شده است. هدف این است که تابعی تفکیک کننده بر اساس خصوصیات خاک برای سایت های مختلف ارائه شود. بدین منظور در نرم‌افزار SPSS از روند زیر استفاده می‌شود:
Analyze>Classify>Discriminate
بعد از اجرای فرمان فوق، پنجره‌ای ظاهر می‌شود که متغیر گروه‌بندی را به قسمتGrouping variable وارد کرده و کدهای مربوط به آن را نیز تعریف می‌کنند. سپس متغیرهای مستقل را در بخش Independents قرار می‌دهند.
جدول زیر مقادیر ویژه و درصد واریانس را برای توابع مختلف نشان می‌دهد. مشاهده می‌شود که تابع اول ۶۲ درصد واریانس را در بر می‌گیرد.

Summary of Canonical Discriminant Functions

Eigenvalues				
Function	**Eigenvalue**	**% of Variance**	**Cumulative %**	**Canonical Correlation**
1	2.483(a)	62.0	62.0	.844
2	.890(a)	22.2	84.3	.686
3	.383(a)	9.6	93.8	.526
4	.110(a)	2.7	96.6	.315
5	.093(a)	2.3	98.9	.292
6	.038(a)	.9	99.9	.191
7	.005(a)	.1	100.0	.070
a First 7 canonical discriminant functions were used in the analysis.				

جدول زیر مقدار لامبدای ویلکس را توابع مختلف نشان می‌دهد. مشاهده می‌شود که مقدار این شاخص ار تابع اول به طرف تابع هفتم افزایش می‌یابد. گفته شد که هرچه این شاخص به صفر نزدیک‌تر باشد، بیانگر مناسب‌تر تابع برآوردی در تفکیک گروه‌هاست.

Wilks' Lambda				
Test of Function(s)	**Wilks' Lambda**	**Chi-square**	**df**	**Sig.**
1 through 7	.087	177.218	49	.000
2 through 7	.302	86.752	36	.000
3 through 7	.571	40.606	25	.025
4 through 7	.790	17.092	16	.380
5 through 7	.877	9.531	9	.390
6 through 7	.959	3.058	4	.548
7	.995	.357	1	.550

جدول زیر ضرایب کانونی استاندارد شده تابع‌ها را نشان می‌دهد. این ضرایب بیانگر اهمیت نسبی هر یک از متغیرها در تمایز بین گروه‌های مورد نظر در متغیر گروهبندی است.

Standardized Canonical Discriminant Function Coefficients							
	Function						
	1	2	3	4	5	6	7
k	-.788	.889	-1.270	-.732	-.409	-1.337	-.828
p	.017	.356	.711	.749	-.438	.127	-.390
n	2.524	.141	-.427	.713	-.297	-.017	.269
om	-2.190	-.170	-.482	.169	.959	.067	.174
lim	.162	.910	.142	-.186	.203	.296	.276
ph	.408	-.471	1.649	-.274	.448	.179	.877
ec	.203	-.551	.764	-.221	.827	.737	-.366

در جدول زیر ضرایب استاندارد نشده تابع‌ها ارائه شده است. با مقادیر ضرایب این جدول تابع تشخیص نوشته می‌شود.

Canonical Discriminant Function Coefficients							
	Function						
	1	2	3	4	5	6	7
k	-.002	.002	-.003	-.002	-.001	-.003	-.002
p	.004	.088	.176	.186	-.109	.031	-.097
n	178.367	9.945	-30.179	50.406	-20.985	-1.225	18.999
om	-5.148	-.399	-1.132	.397	2.255	.157	.410
lim	.067	.374	.058	-.076	.083	.122	.114
ph	1.495	-1.724	6.036	-1.002	1.638	.653	3.210
ec	.143	-.388	.538	-.156	.582	.519	-.258
(Constant)	-15.396	10.012	-47.477	5.111	-13.628	-5.105	-25.756
Unstandardized coefficients							

F1=-15.396-0.002k+0.004p+178.367n-5.148om+0.067lim+1.495ph+0.143ec

Structure Matrix							
	Function						
	1	2	3	4	5	6	7
lim	.017	.791(*)	-.025	-.244	.134	.498	.219
p	-.077	.290	.444	.662(*)	.078	-.314	-.412
om	-.104	.086	-.061	.555	.701(*)	-.409	.108
n	.311	.082	-.086	.486	.632(*)	-.502	-.008
ph	.072	.014	.445	.056	.387	-.762(*)	.253
k	.108	.197	.160	-.016	.479	-.760(*)	-.341
ec	.202	-.021	.120	-.097	.660	.042	-.706(*)
Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions Variables ordered by absolute size of correlation within function.							
* Largest absolute correlation between each variable and any discriminant function							

Functions at Group Centroids							
sit1	Function						
	1	2	3	4	5	6	7
1.00	1.607	-.219	-.726	.118	-.441	-.070	.109
2.00	-1.147	-.065	.721	-.623	-.051	.118	.083
3.00	-.874	-.701	-.644	-.206	-.317	.178	-.121
4.00	.173	.002	1.201	.460	-.337	-.008	-.038
5.00	-1.663	-.498	-.288	.465	.374	.240	.055
6.00	-.360	1.534	-.215	-.014	.087	-.082	-.019
7.00	-.807	-1.272	-.003	-.057	.221	-.414	-.016
8.00	3.431	-.316	.170	-.130	.377	.120	-.033
Unstandardized canonical discriminant functions evaluated at group means							

جدول زیر با انتخاب گزینه Summary table در بخش Classify ایجاد می‌شود. ستون Original به گروه‌های اولیه موجود در مسئله اشاره می‌کند. عبارت Predicted Group Membership به گروه‌های نسبت داده شده پس از تحلیل ممیزی اختصاص دارد. درصدهای فراوانی ارائه شده در جدول میزان تطبیق موارد مشاهده شده و برآوردی را نشان می‌دهد. برای مثال اگر فردی از گروه یک انتخاب شده و اطلاعات این فرد در تابع تشخیص قرار داده شود، در ۸۸/۹ درصد موارد تابع به درستی عضویت فرد را به گروه یک تشخیص می‌دهد.

Classification Results(a)

		sit1	Predicted Group Membership								Total
			1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	
Original	Count	1.00	8	0	0	1	0	0	0	0	9
		2.00	1	5	1	1	1	0	0	0	9
		3.00	2	0	5	1	1	0	0	0	9
		4.00	0	1	0	7	0	0	0	1	9
		5.00	1	4	0	0	4	0	0	0	9
		6.00	0	6	0	2	0	10	0	0	18
		7.00	0	0	4	1	0	0	4	0	9
		8.00	2	0	0	0	0	0	0	7	9
	%	1.00	88.9	.0	.0	11.1	.0	.0	.0	.0	100.0
		2.00	11.1	55.6	11.1	11.1	11.1	.0	.0	.0	100.0
		3.00	22.2	.0	55.6	11.1	11.1	.0	.0	.0	100.0
		4.00	.0	11.1	.0	77.8	.0	.0	.0	11.1	100.0
		5.00	11.1	44.4	.0	.0	44.4	.0	.0	.0	100.0
		6.00	.0	33.3	.0	11.1	.0	55.6	.0	.0	100.0
		7.00	.0	.0	44.4	11.1	.0	.0	44.4	.0	100.0
		8.00	22.2	.0	.0	.0	.0	.0	.0	77.8	100.0

a 61.7% of original grouped cases correctly classified.