

با سلام و احترام،

فصل بیستم و پنجم: تحلیل عامل

عملاً در پژوهش‌ها با حجم زیادی از متغیرها مواجه هستیم. بنابراین یکی از اهداف پژوهشگر در جهت نیل به نتایج بهتر، کاهش حجم داده‌ها می‌باشد و در نقطه مقابل صورت‌بندی یک ساختار جدید، تا در چارچوب این ساختار جدید بتواند به تحلیل و تفسیر این داده‌ها بپردازد.

هدف اصلی روش‌شناسی تحلیل عاملی، مطالعه منظم و ساختار موجود در داده‌های چند متغیره است. یعنی تعیین تعداد و ماهیت عامل‌های مشترک و انگاره تأثیر آنها بر ویژگی‌های سطحی، تحلیل عاملی با آزمون انگاره همبستگی (یا کوواریانس) بین متغیرهای مشاهده شده اجرا می‌شود. در این تحلیل، متغیرهایی که همبستگی بالایی (چه مثبت و چه منفی) با هم دارند، احتمالاً تحت تأثیر عامل‌های یکسان هستند تمام متغیرهایی که نسبت به هم تقریباً همبستگی ندارند، از عامل‌های متفاوتی تأثیر می‌پذیرند. تحلیل عاملی تکنیکی آماری است که برای برآورد عامل‌ها یا متغیرهای پنهان از یک طرف و کاهش تعداد زیادی متغیر به تعداد کمتری عامل از طرف دیگر به کار می‌رود. بنابراین، روش تحلیل عاملی با این هدف به کار برده می‌شود که حتی المقدور از تعداد زیادی متغیر مشاهده شده، شمار معدودی عامل بیرون کشیده شود که هر یک از این عوامل از روی متغیرها و معنی آنها تفسیر می‌شوند. تحلیل عاملی شمار زیادی از متغیرها را به یک سری عامل‌های با حجم کمتر کاهش می‌دهد. به همین خاطر، روش تحلیل عاملی به عنوان یک روش مستقل و غیر وابسته قلمداد می‌شود. یعنی درصدد فرض شناسایی یک متغیر وابسته نیست.

روش‌های تحلیل عامل

طرز کار تحلیل عامل، چندین روش اضافی برای تحلیل سازه‌ها دارد.

تحلیل داده‌ها: روش مؤلفه‌های اصلی، به وسیله یافتن یک ترکیب خطی از متغیرها (یک مؤلفه) که برای چنین تغییری در متغیرهای اصلی محاسبه می‌شوند آغاز می‌گردد. سپس مؤلفه دیگری را پیدا می‌کند که برای بیشتر تغییرات باقی‌مانده ممکن محاسبه می‌شود و با مؤلفه قبلی، همبستگی ندارد و برای ادامه این روش هنوز مؤلفه‌های زیادی مثل متغیرهای اصلی وجود دارد. معمولاً برای بیشتر تغییرات، چند مؤلفه محاسبه خواهد شد و این مؤلفه‌ها

می‌توانند به عنوان جایگزین متغیرهای اصلی استفاده شوند. این روش‌ها، اکثر اوقات برای کاهش تعداد متغیرها در فایل داده‌ها استفاده می‌شوند.

تشخیص ساختار (ساختاریابی) (Structure detection): روش دیگر تحلیل عامل به مرحله بالاتر می‌رود با اضافه کردن این فرض که برخی تغییرات در داده‌ها نمی‌توانند به وسیله مؤلفه‌ها تبیین گردند. در نتیجه واریانس کل که تبیین می‌شود کوچکتر هست بنابراین جمع این ساختارها با مدل‌های عامل، روش‌های مطلوب برای بررسی همبستگی و رابطه بین متغیرها را می‌سازد. با هر متد اضافی باید سعی کنیم دو سؤال را با راه‌حل‌های خوبی پاسخ دهیم. اول اینکه چند مؤلفه (عامل) برای نمایش متغیرها نیاز است و دوم اینکه این مؤلفه‌ها چه کاری را انجام می‌دهند؟

استفاده از تحلیل عامل برای کاهش داده‌ها

یک تحلیل‌گر صنعتی می‌خواهد میزان فروش اتومبیل را با استفاده از یک مجموعه پیشگوها پیش‌بینی نماید. با این وجود، خیلی از پیشگوها همبسته هستند، و تحلیل‌گر می‌ترسد که این روش روی نتایجش تأثیر نامطلوب بگذارد. از **Factor Analysis** با استخراج مؤلفه‌های اصلی برای تمرکز تحلیل زیر مجموعه قابل کنترل پیشگوها استفاده نمایید.

شروع تحلیل

۱. برای اجرای تحلیل عامل مؤلفه‌های اصلی، مسیر زیر را از منوی اصلی برگزینید:

Analyze > Dimenion Reduction > Factor

۲. از **Vehicle Type** تا **Fuel Efficiency** را به عنوان متغیرهای تحلیل برگزینید.

۳. دکمه **Extraction** را کلیک کنید.

۴. گزینه **Screen Plot** را فعال نمایید.

۵. دکمه **Continue** را کلیک کنید.

۶. **Rotation** را در کادر محاوره **Factor Analysis** کلیک کنید.

۷. **Varimax** را در مجموعه **Method** انتخاب نمایید.

۸. دکمه **Continue** را کلیک کنید.

۹. **Scores** را در کادر محاوره **Factor Analysis** کلیک کنید.

۱۰. گزینه‌های **Save as Variables** و **Display Factor Score Coefficient matrix** را فعال نمایید.



۱۱. دکمه Continue را کلیک کنید.

۱۲. Ok را در کادر محاوره Factor Analysis کلیک کنید.

جداول و نمودارهای حاصله در خروجی ظاهر می شود. مؤلفه های با مقادیر ویژه بزرگ تر از 1 برای فایل کاری ذخیره شده اند.

ضرائب اشتراک

ضرائب اشتراک مقدار واریانس را در هر متغیری که در نظر گرفته شده است نشان می دهد. ضرائب اشتراک اولیه (Initial) برآوردهای واریانس در هر متغیر می باشند که با تمام مؤلفه ها یا عوامل در نظر گرفته شده اند. مقادیر این ستون، نشان دهنده مقدار کل واریانس هر متغیر می باشد که مجموعه عوامل می توانند (نه اینکه توانسته اند) آن را تبیین کنند. بنابراین از آنجا که عامل ها می توانند کلیه واریانس یک متغیر (۱۰۰ درصد) را تبیین کنند در جدول همانطور که دیده می شود مقدار این واریانس برای تمام گویه ها (متغیرها) برابر با عدد ۱ است. ضرائب اشتراک استخراج (Extraction) برآوردهای واریانس در هر متغیر می باشند که با مؤلفه ها در نظر گرفته شده است. مقادیر این ستون نشان دهنده مقداری از واریانس هر متغیر می باشد که مجموعه عوامل موردنظر توانسته اند (نه اینکه می توانند) آن را تبیین کنند. مقدار این واریانس بین صفر تا یک در نوسان است. هر چه مقادیر به یک نزدیک تر باشد، بهتر است و مقادیر کوچکتر هر متغیر نشان می دهد که متغیر موردنظر به اندازه کافی برای تحلیل عاملی مناسب نمی باشد و بنابراین باید از تحلیل خارج شود. به عنوان یک قاعده کلی، متغیرهایی را که عامل ها نتوانسته اند بالاتر از ۵۰ درصد از تغییرات آنها را تعیین کنند تعدیل و یا از مجموعه متغیرها حذف می کنیم تا بعداً در انتخاب و دسته بندی عامل ها مشکلی ایجاد نکنند.



	Initial	Extraction
Vehicle type	1.000	.930
Price in thousands	1.000	.876
Engine size	1.000	.843
Horsepower	1.000	.933
Wheelbase	1.000	.881
Width	1.000	.776
Length	1.000	.919
Curb weight	1.000	.891
Fuel capacity	1.000	.861
Fuel efficiency	1.000	.860

ضرائب در این جدول همه بزرگ هستند، که این نشان می دهد مؤلفه های استخراج شده متغیرها را به خوبی نمایش می دهند. اگر هر ضریب اشتراکی در استخراج مؤلفه های اصلی خیلی کوچک باشد، ممکن است لازم باشد که مؤلفه دیگر را استخراج نمایند.

واریانس بیان شده با حل اولیه، مؤلفه‌های استخراجی، و مؤلفه‌های چرخشی در جدول نمایش داده شده‌اند. قسمت اولیه جدول، مقادیر ویژه اولیه یا **Initial Eigenvalues** را نشان می‌دهند. مقدار ویژه، درصد از واریانس و درصد تجمعی از واریانس را برای هر کدام از عامل‌ها نشان می‌دهد.



Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	5.994	59.938	59.938
2	1.654	16.545	76.482
3	1.123	11.227	87.709
4	.339	3.369	91.098
5	.254	2.541	93.640
6	.199	1.994	95.633
7	.155	1.547	97.181
8	.130	1.299	98.480
9	9.055E-02	.905	99.385
10	6.149E-02	.615	100.000

داده‌های قسمت **Extraction sums of squared Loading**، توزیع واریانس را بعد از استخراج عامل‌ها نشان می‌دهد. تعداد ردیف‌ها در این قسمت با تعداد عامل‌های باقیمانده برابر است. در مثال ما، چون ۳ عامل باقی ماندند، بنابراین ۳ ردیف وجود دارد که هر ردیف برای یک عامل باقیمانده است. مقادیر این بخش نیز همانند بخش اول محاسبه می‌شوند و تنها تفاوت آنها در این است که در این بخش مقادیر بر حسب واریانس مشترک به دست می‌آیند و به همین خاطر مقادیر آن همیشه از مقادیر بخش اول (که از واریانس کل به دست می‌آیند) کوچکتر است. داده‌های قسمت **Rotation sums of squared Loading** توزیع واریانس را بعد از چرخش عامل‌ها نشان می‌دهد.

Component	Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	5.994	59.938	59.938
2	1.654	16.545	76.482
3	1.123	11.227	87.709

Component	Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	3.220	32.199	32.199
2	3.134	31.344	63.543
3	2.417	24.166	87.709

ستون **Total** حاوی مقادیر ویژه، یا مقدار واریانس در متغیرهای اصلی می‌باشد که با هر مؤلفه در نظر گرفته شده‌اند. به عبارت دیگر مقدار ویژه عبارت است از واریانس مجموعه متغیرهای مشاهده شده که توسط هر عامل تبیین شده است. اولین عامل همیشه بیشترین واریانس متغیرها را تبیین می‌کند و بنابراین بزرگترین مقدار ویژه را

دارد. عامل دوم، بیشترین واریانس بعدی را تبیین می‌کند و همین‌طور الی آخر. بر اساس ملاک کیزر، عامل‌هایی که مقدار ویژه آنها کمتر از یک است در انتخاب عامل‌ها به حساب نمی‌آیند. در مثال ما یعنی عوامل ۴ تا ۱۰.

Component	Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	5.994	59.938	59.938
2	1.654	16.545	76.482
3	1.123	11.227	87.709

ستون **% of Variance** نسبت واریانس در نظر گرفته شده با هر مؤلفه به جمع کل واریانس در تمام متغیرها را بر حسب درصد می‌دهد. یعنی مقدار واریانسی که هر عامل توانسته است نسبت به کل واریانس مجموعه متغیرها، آن را تبیین کند باید توجه داشته باشیم که یک تحلیل عاملی خوب همواره تحلیل عاملی است که در آن تعداد کمی عامل بتوانند مقدار زیادی از واریانس مجموعه متغیرها را تبیین کنند.

ستون **Cumulative %** درصد واریانس در نظر گرفته شده با **n** مؤلفه اول را می‌دهد. مثلاً، درصد تراکمی عبارتست از جمع درصد واریانس مؤلفه‌های اول و دوم.

بنابراین، مقادیر این ستون، به درصد تراکمی و تجمعی هر مقدار واریانسی اشاره دارد که هر عامل توانسته است آن را تبیین کند.

برای حل اولیه، تعدادی مؤلفه‌های متغیر وجود دارند، و در یک تحلیل همبستگی، جمع مقادیر ویژه برابر با تعداد مؤلفه‌ها است. شما می‌خواهید که مقادیر ویژه بزرگتر از ۱ استخراج شوند.

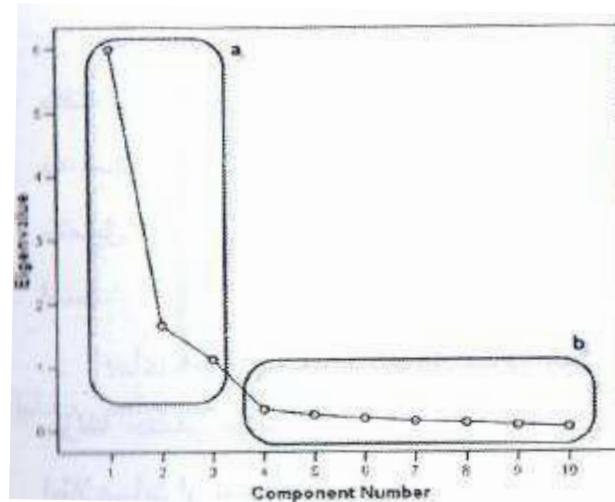
قسمت دوم جدول، مؤلفه‌های استخراجی را نشان می‌دهد. اینها تقریباً **88%** از تغییرپذیری در ده متغیر اصلی را توضیح می‌دهند، بنابراین به طور معنی‌داری می‌توانید پیچیدگی مجموعه داده‌ها را با استفاده از این مؤلفه‌ها، با از دست دادن فقط **12%** از اطلاعات کاهش دهید.

چرخش، درصد تراکمی تغییر توضیح داده شده با مؤلفه‌های استخراجی را حفظ می‌کند. اما آن تغییر اکنون با یکنواختی بیشتری در اطراف مؤلفه‌ها گسترده می‌شود. تغییرات بزرگ در جمع کل‌های خاص حکایت از آن دارد که ماتریس مؤلفه چرخشی آسان‌تر از ماتریس غیرچرخشی تفسیر خواهد شد.

نمودار Scree

نمودار **Scree** نتایج جدول قبلی مبنی بر تعداد مناسب عامل را به صورت تصویری نشان می‌دهد و برای تعیین تعداد بهینه مؤلفه‌ها به کار می‌رود. مقدار ویژه هر مؤلفه در حل اولیه به صورت نمودار درآمده است. در حالت

کلی، شما می‌خواهید مؤلفه‌ها را در شیب تند استخراج نمایید (a). مؤلفه‌های روی شیب کم کمتر در تحلیل کمک می‌کنند (b).



ماتریس مؤلفه چرخشی

با ماتریس مؤلفه چرخشی مشخص می‌کنید که چه مؤلفه‌ای نمایش داده شود. این جدول، ماتریس همبستگی بین گویه‌ها و عامل‌ها را بعد از چرخش نشان می‌دهد که در آن مقدار همبستگی بین -1 و $+1$ در نوسان است. بر اساس این جدول محقق بر اساس بزرگترین بار عاملی تک تک گویه‌ها، به دسته‌بندی آنها با توجه به میزان همبستگی با یکدیگر می‌پردازد.

اولین مؤلفه همبستگی بالایی با **Price in thousands** و **Horsepower** دارد. **Price in thousands** نمایشگر بهتری است، چون همبستگی کمتری با دو مؤلفه دیگر دارد (a). دومین مؤلفه همبستگی بالایی با **Length** دارد (b). سومین مؤلفه با **Vehicle Type** همبستگی زیادی دارد (c).

	Component		
	1	2	3
Vehicle type	101	9.478E-02	.954
Price in thousands	.935	-3.45E-03	4.136E-02
Engine size	.753	.436	.292
Horsepower	.933	.242	5.565E-02
Wheelbase	3.616E-02	.884	.314
Width	.384	.759	.231
Length	.155	.943	6.882E-02
Curb weight	.519	.533	.581
Fuel capacity	.398	.495	.676
Fuel efficiency	-.543	-.318	-.681

این به این معنی است که می‌توانید روی **Price in thousands**، **Length** و **Vehicle Type** در تحلیل چهارم تمرکز نمایید، اما حتی می‌توانید با ذخیره کردن رتبه‌های مؤلفه این کار را بهتر انجام دهید.

ماتریس ضرب رتبه مؤلفه

جدول روبرو، ماتریس تبدیل عامل نام دارد. این ماتریس برای محاسبه ماتریس عامل چرخش یافته از ماتریس عامل اصلی (چرخش نیافته) به کار می‌رود. در این جدول، اگر عناصر غیر متمایل به صفر نزدیک باشند نتیجه می‌گیریم که چرخش نسبتاً کوچک است. اما اگر مقدار این عناصر متمایل بزرگتر از **0.5** باشند در آن صورت نتیجه می‌گیریم که از چرخش بزرگتر استفاده شده است.

برای هر مورد و هر مؤلفه، رتبه مؤلفه با ضرب مقادیر متغیر اصلی مورد در ضرائب رتبه مؤلفه حساب می‌شود. نتیجه متغیرهای رتبه سه مؤلفه بیانگر ده متغیر اصلی هستند که فقط **12%** اطلاعات از دست رفته است. با استفاده از مؤلفه‌های ذخیره شده نیز ترجیح داده می‌شود که از **Price in thousands**، **Length** و **Vehicle Type** استفاده شود چون مؤلفه‌ها بیانگر تمام ده متغیر اصلی هستند، و مؤلفه‌ها رابطه‌شان با دیگران همبستگی خطی نیست. هر چند برآورد می‌شود که همبستگی خطی بین مؤلفه‌ها **0** باشد، باید به نمودارهای رتبه‌های مؤلفه نگاه کنید تا مقادیر پرت و رابطه‌های غیرخطی بین مؤلفه‌ها را بررسی نمایید.



	Component		
	1	2	3
Vehicle type	-173	-194	615
Price in thousands	414	-179	-081
Engine size	226	028	-016
Horsepower	368	-046	-139
Wheelbase	-177	397	-042
Width	011	289	-102
Length	-105	477	-234
Curb weight	070	043	175
Fuel capacity	012	017	262
Fuel efficiency	-107	108	-298

ماتریس نمودار پراکنش رتبه‌های مؤلفه

۱. مسیر **Scatter / Dot > Legacy Dialogs > Graphs** را برای ایجاد یک ماتریس نمودار پراکنش

از رتبه‌های مؤلفه انتخاب نمایید.

۲. گزینه **matrix Scatter** را انتخاب نمایید.

۳. دکمه **Define** را کلیک کنید.

۴. از **1 for analysis REGR Factor Score** تا **3 for analysis REGR Factor Score**

را به عنوان متغیرهای ماتریس برگزینید.

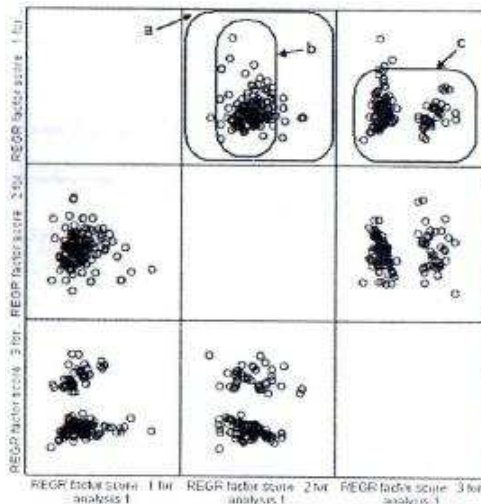
۵. Ok را کلیک کنید.

نمودار اول در ردیف اول، مؤلفه اول را در محور عمودی در مقایسه با مؤلفه دوم در محور افقی نشان می دهد، و ترتیب بقیه نمودارها به همین منوال است. (a)

ماتریس نمودار پراکنش نشان می دهد که مؤلفه اول دارای توزیع چوله است، و این به دلیل چوله بودن Price in thousands می باشد (b).

استخراج مؤلفه های اصلی با استفاده از لگاریتم تبدیل قیمت ممکن است نتایج بهتری در برداشته باشد. جدایی که در مؤلفه سوم دیده می شود با این واقعیت توجیه می شود که Vehicle Type یک متغیر دو تایی (binary) می باشد (c).

یک رابطه بین مؤلفه اول و سوم وجود دارد، که به دلیل این واقعیت است که چندین اتومبیل گران قیمت وجود دارد اما هیچ کامیون لوکسی وجود ندارد. این مسئله را با استفاده از لگاریتم تبدیل قیمت می توان تا اندازه ای حل کرد، اما اگر این طریق نیز مسئله را حل نکرد، ممکن است بخواهید که فایل را روی Vehicle Type بشکنید.



ادامه فصل ۲۵ را در مرجع کاربردی SPSS 20 (41) دنبال نمایید.



Telegram.me/iepnu
کانال تخصصی مهندسی صنایع دانشگاه پیام نور